Molecular Mechanisms of Hedgehog Signal Transduction by the G-protein Coupled Receptor Smoothened

Eamon F. X. Byrne

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

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

Trinity 2017

"If our whole lives had not been spent in the midst of it, it would astound us."

— Loren Eiseley

"As the crystal grows the constituent atoms stack together in a lopsided way, and do not allow other stray atoms to intrude to cloud its mineral exactitude. Layer builds on layer to reveal the crystalline form, the macrocosm of the gem reflecting exactly the microcosm of atomic structure. As with the handiwork of a master mason, there is no mistake permissable in the atomic brickwork."

— Richard Fortey

Molecular Mechanisms of Hedgehog Signal Transduction by the G-protein Coupled Receptor Smoothened

Eamon F. X. Byrne

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

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

ABSTRACT

The Hedgehog signalling pathway is an essential developmental pathway present in all bilaterians that is involved in embryogenesis, body patterning and stem cell homeostasis. Dysregulation of the Hh pathway leads to various kinds of cancer, such as basal cell carcinoma and medulloblastoma. Smoothened (SMO), a Frizzled-type G-protein coupled receptor (GPCR), is the essential transmembrane signal transducer within the Hh pathway, conveying the signal from the upstream transmembrane protein, Patched1 (Ptc1), to the downstream intracellular proteins. The mechanisms by which SMO transmits the Hh signal from the extracellular environment, through the plasma membrane and to the intracellular proteins are not known. In this thesis, I present my work into the structural and functional characterisation of the extracellular and transmembrane domains (TMD) of human SMO in order to better understand the molecular mechanisms of its signal transduction. The extracellular region of SMO contains a highly conserved cysteine-rich domain (CRD) and a linker domain (LD). I present the first crystal structure of the CRD, LD and TMD of SMO, which is also the first crystal structure of a GPCR with a large functional extracellular domain. This structure revealed a domain architecture for SMO that enables regulation of its transmembrane domain by its extracellular domains. It also revealed a cholesterol molecule bound to the CRD, which we subsequently determined to be a new endogenous small-molecule agonist for SMO. I present five further structures of SMO bound to different small molecule agonists and antagonists. Together, these structures demonstrate that the position of the CRD relative to the TMD reflects the activation state of SMO. We also generated nanobodies against the extracellular region of SMO in order to stabilise its conformation. These studies not only improve our understanding of the workings of a key transmembrane protein within a fundamental signalling pathway but will also aid efforts to develop better therapeutics for an important cancer target.

Acknowledgements

First and foremost, a sincere thank you to my supervisor, Christian Siebold, for his buoyant enthusiasm, ever-present encouragement and guidance on my road to understanding crystallography and all things Hedgehog. It has been an exciting journey so far and I feel privileged to have been a part of the STRUBI community.

Thank you to my fellow DPhil students for all the good times, especially through Prosecco Fridays, Formal Halls, STRUBI Student Sessions and, of course, extended tea/coffee breaks: particularly Lea Sefer, Christiane Kowatsch, Eleanor Healey, Sam Griffiths, Joy Shu Fen Ho, Steinar Haldorsson, Nayab Malik, Keshalini Sabaratnam, Katharina Jungnickel, Rhys Pryce, Liz Allen, Alice Stelfox, Serban Ilca, Max Renner, James Hillier and Rachel Woolley. I feel incredibly fortunate to have worked in such a supportive environment.

Thank you to all of the postdocs who have been an enormous aid to me, either formally or informally, over the years: particularly Paul Miller, Ben Bishop, Ross Robinson, Amalie Rudolf, Rebekka Siebold-Schwab, Jonathan Elegheert and Tomas Malinauskas. Thank you to Margaret Jones, Weixian Lu and Oliver Hayward for their invaluable lab support. Thank you to Simon Newstead, E. Yvonne Jones, Radu Aricescu and Jon Grimes for many helpful discussions and feedback.

Thank you to the funders of my DPhil scholarship, without whom I would not have been able to come to Oxford: the Nuffield Department of Medicine, the Clarendon Fund and University College. Thank you also to Cancer Research UK, who funded many of the consumables that were essential to my research.

Thank you to my friends in Oxford: particularly Graeme, Mandy, Jeff, Elise, Cate and Samara. Finally, I would like to thank my partner Yien Li and my parents, Donna and Gerard, for their love and support over many years and great distances. It has made all the difference.

Technical Acknowledgements

I would like to acknowledge and sincerely thank all of those people whose specific technical expertise has been essential to me in the completion of my DPhil studies. Experimental results from several of them I have reproduced within this thesis in full or in part (more details are provided within the thesis where applicable).

Associate Professor Rajat Rohatgi, Dr Ria Sircar, Dr Giovanni Luchetti and Dr Sigrid Nachtergaele at the Stanford School of Medicine - for binding studies, functional and *in vivo* assays.

Professor Simon Newstead in the Biochemistry Department, University of Oxford - for fishing of LCP crystals.

Dr Robert Rambo and Dr Mark Tully at the B21 SAXS Beamline, Diamond Light Source - for the running of SAXS experiments and analysis of SAXS data.

Professor Mark Sansom and George Hedger in the Biochemistry Department, University of Oxford - for molecular dynamics simulations.

Professor Jan Steyaert, Dr Els Pardon and Eva Beke at Structural Biology Brussels, Vrije Universiteit Brussel - for llama inoculation and phage-display screening of nanobodies.

Dr Rebekka Siebold-Schwab at STRUBI, University of Oxford - for cell surface fluorescence microscopy using nanobodies.

Jesper Levring and Rachel Woolley at STRUBI, University of Oxford - for assistance with cloning and production of nanobodies.

Dr Robin Owen, Dr Danny Axford and Dr Darren Sherrell at MX I24 Microfocus Beamline, Diamond Light Source - for technical assistance with the beamline.

Abbreviations

1D4-beads	sepharose beads conjugated with anti-Rho1D4 epitope antibody
1D4-peptide	peptide containing the Rho1D4 epitope
β_2AR	β_2 adrenergic receptor
βOG	n-octyl-β-D-glucoside (detergent)
A280	absorbance at 280 nm
A_{2a}AR	A_{2a} adenosine receptor
BCC	basal cell carcinoma
BLI	bio-layer interferometry
BMP	bone morphogenetic protein
BRIL	thermostabilised apocytochrome b_{562}RIL
CDR	complementarity determining region (e.g. of nanobodies)
cAMP	cyclic AMP
cGMP	cyclic GMP
CLR	cholesterol
CHS	cholesteryl hemisuccinate
CMC	critical micelle concentration
CRD	cysteine-rich domain (e.g. of SMO)
D2R	dopamine D_2 receptor
DDM	n-dodecyl-β-D-maltopyranoside (detergent)
Dhh	Desert Hedgehog
DMNG	decyl maltose neopentyl glycol (detergent)
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribose nucleoside triphosphate
dSMO	Smoothened from *Drosophila melanogaster*
ECD	extracellular domains (e.g. of SMO)
ECL	extracellular loop
EndoF1	endoglycosidase F1
FCS	foetal calf serum

FSEC	fluorescence-detection size exclusion chromatography
Fzd	Frizzled
GlcNAc	N-acetyl-glucosamine (sugar)
Gli	glioma-associated oncogene family member
GPCR	G-protein coupled receptor
Hh	Hedgehog
hSMO	Smoothened from *Homo sapiens*
hpt	hours post-transfection
ICD	intracellular domain (e.g. of SMO)
ICL	intracellular loop
Ihh	Indian Hedgehog
IMAC	immobilised metal ion affinity chromatography
K_D	dissociation constant
kDa	kilodalton
LCP	lipidic cubic phase (crystallisation)
LD	linker domain (e.g. of SMO)
MβCD	methyl-β-cyclodextrin
MALS	multi-angle light scattering
MWCO	molecular weight cut-off
NAG	N-acetylglucosamine (sugar)
Nb	nanobody
Ni-NTA	nickel-nitrilotriacetic acid
NPC	Niemann-Pick type-C proteins
NMR	nuclear magnetic resonance
OGNG	octyl glucose neopentyl glycol (detergent)
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	poly-ethylene glycol
PFA	paraformaldehyde
Ptc1	Patched1
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RMSD	root mean squared deviation
RND	resistance-nodulation-division family proteins
SAG	Smoothened agonist
SANT	Smoothened antagonist

SAXS	small angle X-ray scattering
SDS-PAGE	sodium dodecyl sulfate poly-acrylamide gel electrophoresis
SEC	size exclusion chromatography
SFM	serum-free medium
Shh	Sonic Hedgehog
SLS	static light scattering
SMO	Smoothened (generally human Smoothened)
SPR	surface plasmon resonance
SSD	sterol-sensing domain (e.g. of Ptc1)
SuFu	Suppressor of Fused
SUMO	small ubiquitin-related modifier
TMD	transmembrane domain (e.g. of SMO)
VD	vapour diffusion (crystallisation)
vismo	vismodegib
VPA	valproic acid
WT	wild-type
XFEL	X-ray free electron laser
xSMO	Smoothened from *Xenopus laevis*

*Amino acid residues are generally referred to using the standard three-letter codes or, within a construct name, using the standard one-letter codes. Point mutations are indicated by the native residue code followed by the number position of that residue followed by the new residue code (e.g. 'V329F' or 'Val329Phe' refers to the residue valine at position 329 mutated to phenylalanine).

**N.B. For the sake of clarity, throughout this thesis the nomenclature used for *constructs* is distinct from the nomenclature used for *structures*. For instance, the crystallisation construct presented within is referred to as "SMOΔC-BRIL-V329F", in order to emphasise the additional features that were required for expression and stabilisation, whereas the crystal structures determined from this construct are referred to as "SMO CRD-LD-TMD", in order to highlight the difference between these structures and other SMO structures.

Contents

1 Introduction **1**

1.1 Signalling in embryonic development 2

1.2 The Hedgehog signalling pathway 4

1.2.1 Hh in development and disease 4

1.2.2 Hh pathway components 5

1.2.3 Hh and the primary cilium 8

1.3 G-protein coupled receptors (GPCRs) 11

1.3.1 Roles of GPCRs 11

1.3.2 Structure & function of GPCRs 11

1.3.3 Small molecules modulating GPCRs 12

1.4 Smoothened 15

1.4.1 Role of SMO 15

1.4.2 Interactions of SMO 15

1.4.3 SMO as GPCR 16

1.4.4 Structure of SMO 16

1.4.5 Regulation of SMO by Ptc1 18

1.4.6 Small molecules modulating SMO 20

1.5 Thesis outline 22

2 Materials & Methods **23**

2.1 Molecular Biology 24

2.1.1 cDNA and expression vectors 24

2.1.2 Construct design 25

2.1.3 PCR and site-directed mutagenesis 26

2.1.4 Restriction digests, ligations and transformations 28

2.1.5 SDS-PAGE and Western blots 29

2.2 Protein expression and purification 31

2.2.1 Small-scale mammalian expression trials 31

2.2.2 Large-scale mammalian expression and purification 32
2.2.3 Thermostability assay using analytical size exclusion chromatography (SEC) 34
2.2.4 Thermostability assay using intrinsic fluorescence 35
2.2.5 Design of membrane protein additive screen 36
2.3 Protein crystallisation and X-ray data collection 37
2.3.1 Basic principles of X-ray crystallography 37
2.3.2 Methods of crystallisation 38
2.3.3 Vapour diffusion trials 39
2.3.4 Addition of small molecules 39
2.3.5 Lipidic cubic phase (LCP) trials 40
2.3.6 Crystallisation conditions 40
2.3.7 Crystal fishing and data collection 42
2.4 Structure solution pipeline 45
2.4.1 Data processing 45
2.4.2 Molecular replacement 46
2.4.3 Model building and refinement 47
2.4.4 Assessing model quality 50
2.5 Small-angle X-ray scattering (SAXS) 53
2.5.1 Basic principles of SAXS 53
2.5.2 SAXS data collection and processing 54
2.6 Multi-angle light scattering (MALS) 55
2.7 Generation of nanobodies against Smoothened 57
2.7.1 Protein for llama immunisation and screening of nanobodies . 57
2.7.2 Periplasmic expression and purification of nanobodies in E. coli 58
2.8 Binding studies of nanobodies 59
2.8.1 Bio-layer interferometry (BLI) 59
2.8.2 Fluorescence microscopy 60

3 SMO is regulated by its extracellular domains 62
3.1 Abstract 63
3.2 Introduction 65
3.2.1 Structure of the Smoothened TMD 65
3.2.2 Structure of the Smoothened CRD 68
3.2.3 Cholesterol in GPCRs and Hh signalling 69
3.2.4 Chapter outline 72

3.3 Results . . . 75
3.3.1 BRIL fusion increases protein expression . . . 75
3.3.2 Detergent DDM thermostabilises SMO constructs . . . 77
3.3.3 Inactivating point mutation V329F boosts expression . . . 79
3.3.4 Optimisation of expression duration and additives . . . 83
3.3.5 Large scale purification of SMO for crysallisation . . . 84
3.3.6 SMO crystallises in LCP at high concentration . . . 88
3.3.7 Structure solution from multiple isomorphous crystals . . . 90
3.3.8 Crystal structure of human SMO CRD-LD-TMD . . . 93
3.3.9 Structure reveals cholesterol molecule bound to CRD . . . 94
3.3.10 SMO in solution binds cholesterol via CRD binding site . . . 99
3.3.11 Cholesterol is an activator of SMO . . . 104
3.3.12 Cholesterol stabilises CRD in MD simulations . . . 106
3.4 Discussion . . . 110
3.4.1 Domain arrangement suggests signal transduction mechanism between CRD and TMD . . . 110
3.4.2 Cholesterol: a new ligand for SMO . . . 112
3.4.3 Evolutionary divergence between vertebrate and *Drosophila* SMO 115
3.4.4 Identical TMD conformation in all SMO structures . . . 117
3.5 Summary . . . 118

4 Small molecule ligands can induce different conformations of SMO 119
4.1 Abstract . . . 120
4.2 Introduction . . . 122
4.2.1 SMO contains two distinct ligand binding sites . . . 122
4.2.2 Antagonist-bound SMO . . . 124
4.2.3 Agonist-bound SMO . . . 126
4.2.4 Chapter outline . . . 127
4.3 Results . . . 129
4.3.1 Cyclopamine thermostabilises SMO . . . 129
4.3.2 Crystallisation and structures of SMO-ligand complexes . . . 132
4.3.3 Vismodegib induces a conformational change . . . 133
4.3.4 Cyclopamine molecules occupy both binding sites . . . 140
4.3.5 Cholesterol remains bound in SAG structure . . . 144
4.3.6 Oxysterol structure unchanged from cholesterol structure . . . 145
4.3.7 TMD binding site contains an unidentified molecule . . . 151

4.4 Discussion . . . 154
4.4.1 A conserved binding mode for CRD ligands . . . 154
4.4.2 Simultaneous binding of ligands at two sites . . . 157
4.4.3 Cyclopamine and ciliary localisation . . . 160
4.4.4 Multiple CRD positions suggest a continuum of states . . . 161
4.5 Summary . . . 164

5 Nanobodies bind to the extracellular domains of SMO 165
5.1 Abstract . . . 166
5.2 Introduction . . . 167
5.2.1 Brief history of nanobodies . . . 167
5.2.2 Nanobodies as structural biology tools . . . 169
5.2.3 Rationale behind nanobodies for SMO . . . 170
5.2.4 Chapter outline . . . 172
5.3 Results . . . 173
5.3.1 SMO protein production for nanobody generation . . . 173
5.3.2 Nanobody production . . . 175
5.3.3 Nb4 binds SMO with nanomolar affinity . . . 177
5.3.4 Nb4 and Nb7 bind strongly to the cell surface . . . 180
5.3.5 Towards a SMO-nanobody crystal structure . . . 182
5.4 Discussion . . . 186
5.4.1 Nanobody characteristics . . . 186
5.4.2 Further steps towards a crystal structure . . . 187
5.5 Summary . . . 190

6 Discussion 191
6.1 Activation and inhibition of human SMO . . . 192
6.1.1 Cholesterol as the endogenous ligand for SMO . . . 192
6.1.2 Membrane composition as localisation indicator . . . 196
6.1.3 Synergy between two allosterically-linked binding sites . . . 197
6.1.4 Domain conformations and activation states . . . 199
6.1.5 Towards an active state structure of SMO . . . 204
6.1.6 Models of SMO activation and Ptc interaction . . . 207
6.2 Therapeutic implications of SMO structures . . . 210
6.3 SMO and the related Frizzled-type GPCRs . . . 211
6.4 Hh signalling downstream of SMO . . . 214
6.4.1 The intracellular domain of SMO . . . 214

6.4.2 SMO interaction partners 217
6.5 Conclusions 219

A Membrane Protein Additive Screen **220**

B Plasmids & Constructs **222**

C Details of Small Molecule Ligands **228**

D Details of Detergents **231**

Bibliography **233**

List of Figures

1.1 Morphogen gradients in development 3
1.2 Disease from dysregulation of Hh signalling 6
1.3 Schematic of Hh pathway components 8
1.4 Role and structure of the primary cilia 10
1.5 The GPCR superfamily and canonical GPCR function 13
1.6 Schematic of SMO function and architecture 17
1.7 FDA-approved anti-cancer drugs that target SMO 21

2.1 SMO construct design 26
2.2 Site-directed mutagenesis by overlapping PCR 28
2.3 Lipidic cubic phase crystallisation 41
2.4 Microfocus beamline I24 at the Diamond Light Source 44
2.5 Crystallographic Data Table 52

3.1 Structures of SMO LD-TMD with synthetic small molecule ligands . 67
3.2 Conformational changes of GPCRs associated with activation 68
3.3 Structures of CRDs from SMO and Fzd 70
3.4 Various roles for cholesterol 72
3.5 Model of SMO CRD-LD-TMD structure 74
3.6 Sequence alignment of SMO orthologues 76
3.7 Initial screening of three constructs 78
3.8 Thermostability screening of SMO constructs using FSEC 80
3.9 Expression screening of point mutations using FSEC 82
3.10 Characterisation of the SMO Val329Phe mutation 83
3.11 Optimisation of expression and stabilisation 85
3.12 Behaviour of SMOΔC-BRIL during and after purification procedure . 87
3.13 Large scale expression and purification of SMOΔ-BRIL-V329F 89
3.14 Crystallisation and structure solution 92

3.15 Structure of human SMO 95
3.16 Features of human SMO CRD-LD-TMD structure 96
3.17 Superposition of the SMO CRD-LD-TMD structure with previously determined LD-TMD structures 97
3.18 The cholesterol binding site in SMO 98
3.19 SMO binds to cholesterol-conjugated beads 100
3.20 Disruption of cholesterol interactions affects SMO stability 101
3.21 Point mutations in the cholesterol binding site disrupt bead-binding and signalling 103
3.22 Cholesterol activates Hh signalling via SMO CRD binding site 105
3.23 Molecular dynamics simulations of SMO in a lipid bilayer 107
3.24 Domain interface mutations disrupt signalling ability 109
3.25 Chemical structures of sterols relevant to SMO 111
3.26 Model of cholesterol in membrane with SMO 114
3.27 Sequence alignment of SMO orthologues with cholesterol binding site highlighted 116

4.1 SMO CRD-LD-TMD structure with two distinct binding sites 123
4.2 Binding modes of SMO ligands from LD-TMD structures 125
4.3 Thermostability of SMO with small molecule additives by SEC 130
4.4 Thermostability of SMO with small molecule additives by intrinsic fluorescence 131
4.5 Pearson correlation coefficient (CC) analysis of SMO-ligand structures 134
4.6 Crystal packing of SMO-ligand structures 135
4.7 Structure of human SMO bound to vismodegib 137
4.8 Structural analysis of mutations found in vismodegib-resistant tumours 138
4.9 Comparison of cholesterol- and vismogedib-bound SMO 139
4.10 Structure of human SMO bound to two cyclopamine molecules 142
4.11 Cyclopamine binding compared to other ligands 143
4.12 Structure of human SMO bound to SAG and cholesterol 146
4.13 SAG binding compared to other ligands 147
4.14 Structure of human SMO bound to 20(S)-OHC 149
4.15 Comparison between 20(S)-OHC and cholesterol structures 150
4.16 SAXS analysis of SMOΔC-BRIL-V329F 152
4.17 Unidentified density in TMD binding site 153
4.18 Comparison of ligand-bound CRD structures 155

4.19 Comparison of ligand-bound SMO CRD structures 156
4.20 Simultaneous binding of ligands at two sites 159
4.21 Ligand-bound structures reveal continuum of CRD states 162

5.1 Nanobodies are single-chain antigen-binding fragments 168
5.2 Nanobody-aided crystal structures of transmembrane proteins 171
5.3 Production of SMO for llama immunisation and nanobody screening . 174
5.4 Production of nanobodies raised against SMO 176
5.5 Nanobody binding studies using BLI 179
5.6 Fluorescence microscopy cellular assay of nanobodies binding to SMO 181
5.7 Initial mix tests of SMOΔC-BRIL-V329F with nanobodies 183
5.8 Large-scale purification of SMO with addition of Nb4 185

6.1 No covalent bond between Asp95 and cholesterol in crystal structure 194
6.2 Independently solved antagonist-bound structure of SMO closely resembles vismodegib structure 201
6.3 Distinct CRD position in antagonist-bound structures 203
6.4 Model for SMO activation mechanism 208
6.5 Models for Ptc1 function 209
6.6 Relative lengths of domains for Frizzled family receptors 212
6.7 Sequence alignment of Frizzled family receptors 213
6.8 SMO intracellular domain 216

C.1 Chemical structures of small molecules used in this thesis 229

D.1 Chemical structures of detergents used in this thesis 232

Chapter 1

Introduction

1.1 Signalling in embryonic development

The establishment of cell axis orientation and tissue patterning is essential in the early embryonic development of multicellular organisms. Key to this process are morphogens, biochemical substances whose local concentrations govern differential gene expression. Tissue polarity and patterning are organised along opposing concentration gradients of these morphogens, which emanate from different poles of the nascent embryo or tissue[30]. This system enables the conversion of continuous morphogen gradients into precise spatial patterns containing multiple discrete cell types.

Morphogen gradient formation is widespread in developmental contexts but two archetypal examples have dominated studies. The first is anterior-posterior patterning of the early *Drosophila* embryo (Fig. 1.1, panel a). This system involves the opposing gradients of two homeodomain proteins: Bicoid (Bcd) and Caudal (Cad). The second is ventral-dorsal patterning of the vertebrate neural tube. This system involves the opposing gradients of Sonic Hedgehog (Shh) emanating from the ventral pole (notochord) and bone morphogenetic protein (BMP) and Wnt from the dorsal boundary (Fig. 1.1, panel b). While Shh, BMP and Wnt are secreted messenger proteins that activate downstream signalling on receiving cells, Bcd and Cad are transcription factors that diffuse within the syncytial blastoderm prior to the establishment of cell membranes between its multiple nuclei. In both the former and latter cases, morphogens demonstrate two types of activity: repressing cell fates that generally occur at the opposite pole and activating genes for cell fates near to the origin of emanation. Relative concentrations of morphogens as well as the timing and duration of their presence all contribute to cell differentiation and positioning of tissue boundaries.

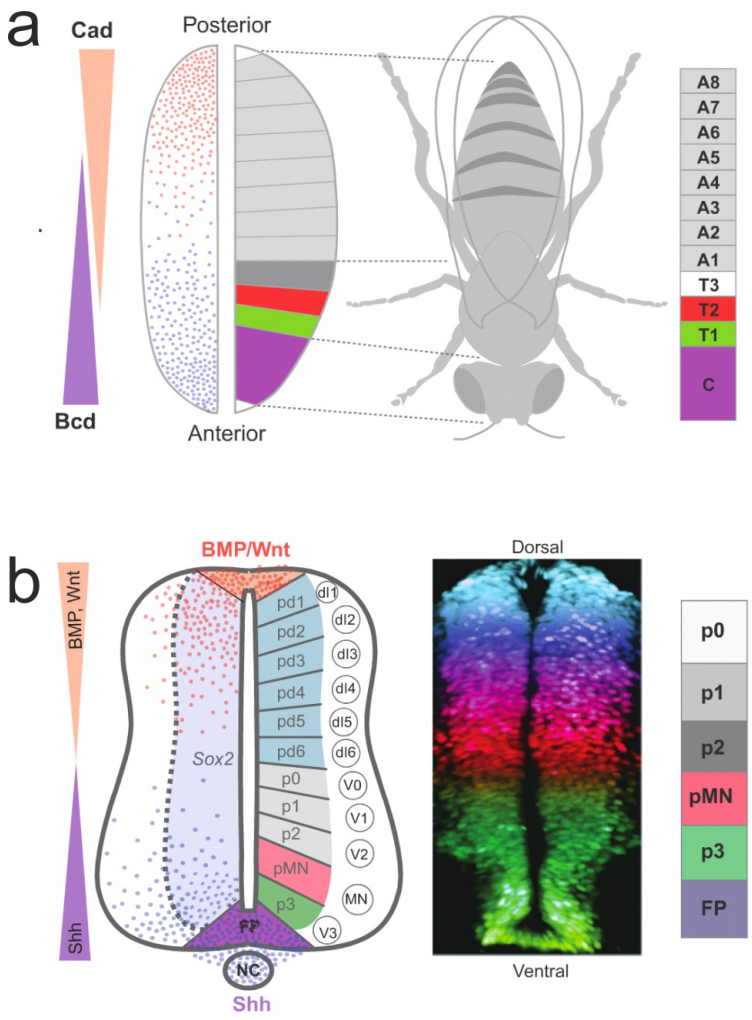


Figure 1.1: Morphogen gradients in development. a, Anterior-posterior patterning of the *Drosophila melanogaster* body plan with opposing gradients of Cad and Bcd. **b,** Dorsal-ventral patterning of the vertebrate neural tube with opposing gradients of BMP/Wnt and Shh. Figure adapted from [30], boxes 1 & 2.

1.2 The Hedgehog signalling pathway

1.2.1 Hh in development and disease

The Hedgehog (Hh) morphogen was first identified in *Drosophila melanogaster*[267, 118] as a segment polarity gene involved in positional signalling[212, 158, 145, 162, 356, 351, 247] and cell-to-cell communication[248, 354]. In vertebrate development, Hh signalling is most closely linked to early patterning of the central nervous system and limb formation[198, 115].

As a signalling pathway critical for embryogenesis, dysregulation of the Hh pathway is associated with body-patterning problems, often to do with ventral neural tube development, which in extreme cases can lead to holoprosencephaly or cyclopism (Fig. 1.2, panel a)[31, 312, 17, 64]. Issues with morphogen production, trafficking or response can all lead to a lack of required growth and differentiation in Hh-receiving cells, ultimately resulting in developmental abnormalities. Embryos with such issues are seldom brought to full term. When they are, the newborns often do not survive long due to breathing and swallowing difficulties associated with their deformations. Such developmental defects can also be caused by exogenous teratogenic compounds. For instance, cyclopamine, which is found in the corn lily plant, inhibits the Hh pathway[155, 77] and if ingested at a certain point during pregnancy can give rise to cyclopic animals[178, 177].

The Hh pathway has also been implicated in stem cell homeostasis, particularly in the skin[282]. In mature cells, aberrant signalling can either cause or contribute to various kinds of cancer. Basal cell carcinoma, the most common form of skin cancer (Fig. 1.2, panel b), and medulloblastoma, a devastating childhood brain cancer, have been most closely linked to specific mutations in the Hh pathway, with point

mutations in either of the two essential transmembrane proteins, Patched1 (Ptc1) and Smoothened (SMO), being the most common[375, 403, 301]. These oncogenic mutations have been shown to either deactivate Ptc1 or render SMO unresponsive to Ptc1 regulation, both cases leading to uncontrolled SMO activation and ultimately to unregulated cell proliferation (Fig. 1.2, panel c). Other types of cancer, including lung, liver, prostate, breast and pancreatic, have been shown in some cases to cause overproduction of Hh morphogens, hijacking the Hh signalling pathway in order to promote uncontrolled tumour growth (Fig. 1.2, panel c)[335, 333, 172, 201, 20, 357, 387, 73].

1.2.2 Hh pathway components

The Hh signalling pathway is conserved across all bilaterians, although there are important differences between vertebrate and invertebrate schemas[369, 159]. For the sake of brevity, details of the invertebrate pathway will not be included in this thesis except where the discussion directly requires it. In mammals, there are three Hh morphogen proteins: Sonic, Indian and Desert Hedgehog (Shh, Ihh, Dhh, respectively)[160, 304, 107, 47]. Shh supports pattern formation of the limbs and neural tube[74, 166], Ihh regulates cartilage and bone development, and Dhh is key for germ cell development in the testis and peripheral nerve sheath formation[16]. Shh has been more widely studied than either Ihh or Dhh.

Within the Hh-producing cell, full-length Shh is processed by autoproteolytic cleavage into two peptides (Fig. 1.3)[211]. The shorter and highly conserved N-terminal peptide (Shh-N) mediates all downstream signalling activities[116, 235] while the C-terminal region is essential for catalysing the autoproteolytic cleavage event[232]. Mature Shh-N is twice lipid-modified: covalently linked to a cholesterol molecule at its C-terminus[290] and palmitoylated on the first cysteine residue near

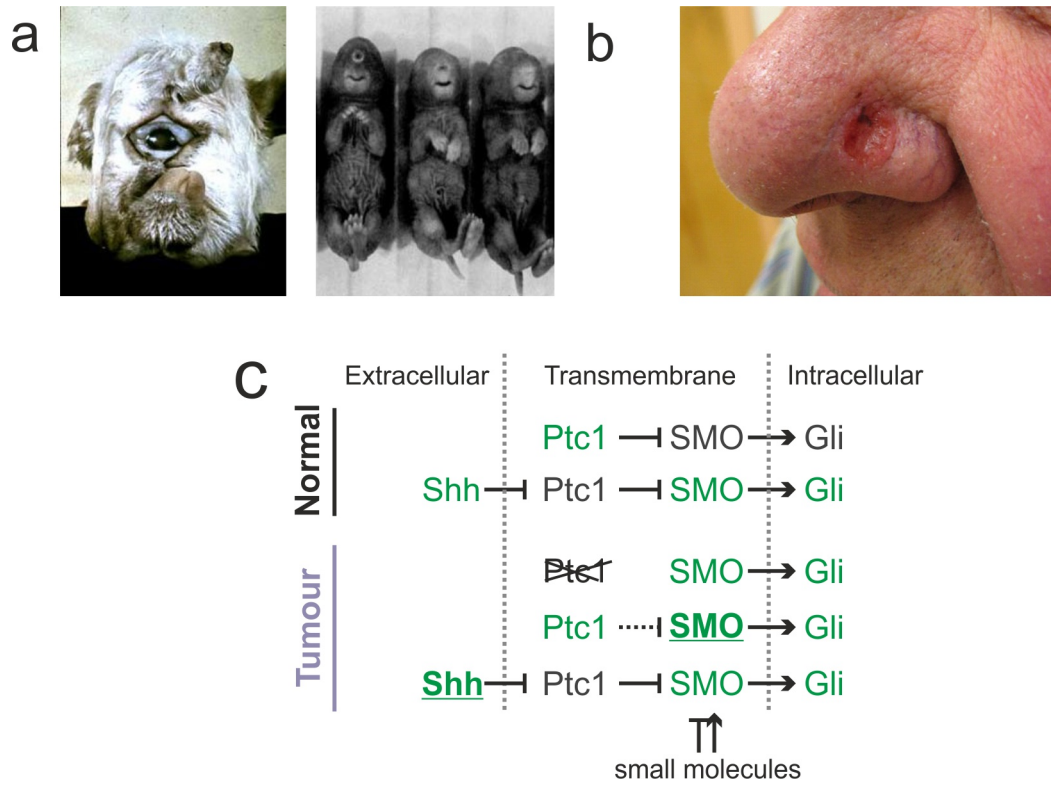


Figure 1.2: Disease from dysregulation of Hh signalling. **a,** Cyclopia faciocranial deformation in offspring induced by consumption of cyclopamine by the pregnant mother, shown in a lamb (left) and rabbits (right). Images from Wikipedia Commons and [177]. **b,** A mild case of basal cell carcinoma on a patient's nose. Image from Wikipedia Commons. **c,** Schematic of Hh pathway activation under normal circumstances (top) and dysregulation leading to tumours (bottom). Green colouring indicates an activated step in the pathway, bold and underlined indicates overactive step, black indicates an inactive step. Normal arrowheads indicate activating influence, flat arrowheads indicate inhibitory influence. In tumours, Ptc1 can be inactivated or SMO can be constitutively activated or Shh can be inappropriately applied. Most small molecules act at the level of SMO.

to its N-terminus[280]. The cholesterol adduct is not absolutely required for signal transduction but influences tissue distribution by acting as a lipid anchor to control the spatial mobility of Shh-N[232, 215, 125, 281, 124]. The palmitoyl adduct also has an important role in tissue distribution[54]. Dually lipidated Shh-N is secreted with the assistance of transmembrane protein Dispatched and secreted glycoprotein Scube2 (ref.[35, 362]). It is then transported or actively diffused to Hh-receiving cells, demonstrating both short- and long-range signalling effects[289, 358].

On Hh-responding cells, Shh-N binds directly to Ptc1, a twelve-pass α-helical transmembrane protein (Fig. 1.3)[148, 260, 349, 234, 122, 58, 162]. Upon Shh-N binding, Ptc1 releases its constitutive inhibition of SMO, a seven-pass α-helical transmembrane protein belonging to the G-protein coupled receptor (GPCR) superfamily[366, 4, 253, 59]. At this point Ptc1 exits the primary cilium[314], a singular antenna-like protrusion from the cell surface (see next section for more details), and SMO accumulates within it[79, 95], entering from the plasma membrane[241] and from intracellular vesicles[385]. However, SMO ciliary accumulation alone is insufficient to activate downstream signalling, implying that an additional transduction step within the primary cilium is required[313].

Intracellular Hh signalling culminates in the regulation of the Gli (glioma-associated oncogene family member) transcription factors after their dissociation from another intracellular protein called Suppressor of Fused (SuFu) (Fig. 1.3)[153, 363, 62, 266]. In the absence of signal, the Gli2 and Gli3 proteins are proteolytically processed in the cilia and then act as transcriptional repressors once they are transported to the nucleus. When the Hh pathway is activated, Gli2 and Gli3 retain their full-length, active form and act as transcriptional activators[31]. The overall effect of Hh signalling is a shift in the balance between the activator and repressor forms of the Gli proteins through regulation of their post-translational proteolytic processing[31]. An additional level of complexity is provided by arrays of interacting transcription factors and modular regulatory elements that govern gene transcription to generate coherent tissue patterning[30].

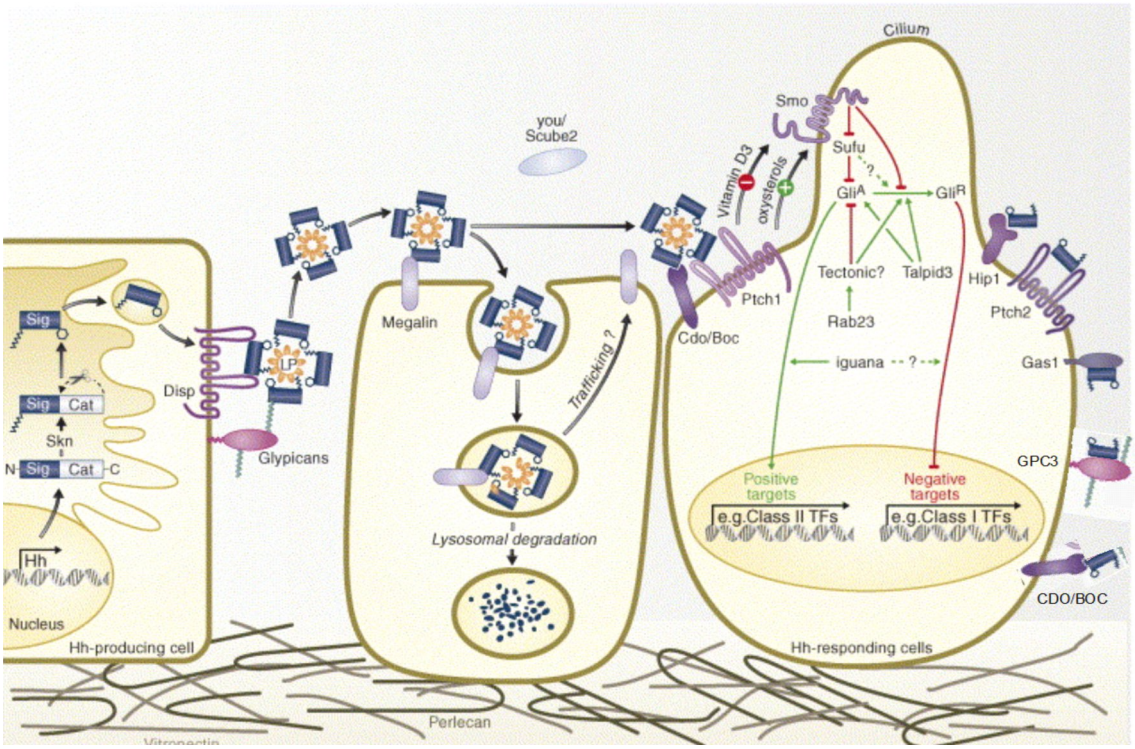


Figure 1.3: Schematic of Hh pathway components. Shh is translated, processed and exported from the Hh-producing cell (left). It is then transported or diffuses to the Hh-responding cell (right) where it binds to Ptc1 (here labelled 'Ptch1'). Being released from inhibition by Ptc1, SMO moves into the primary cilium to activate downstream signalling that results in altered gene transcription. Adapted from [384], figure 1.

1.2.3 Hh and the primary cilium

The primary cilium is a non-motile antenna-like protrusion from the cell surface of almost every vertebrate cell. These cilia are single transient organelles, appearing when the cell is in stage G0 and G1 of the cell cycle[285]. While their role is not yet fully defined, they are generally considered to be a signal processing centre and have been implicated in calcium-sensing[94] as well as in Hh, Wnt and platelet-derived growth factor signalling. Primary cilia are essential for Hh signalling in vertebrates[152, 314, 132, 176]. SMO localises to the primary cilium in response to Hh signal activation and this localisation is necessary for SMO function (Fig. 1.4, panel a)[79]. Further studies revealed that the cytoplasmic tail[188] and the extra-

cellular domains[1] of SMO regulate ciliary localisation. SMO can accumulate at the primary cilium without initiating downstream signalling in the presence of the antagonist cyclopamine, indicating that ciliary localisation of SMO is not sufficient for activation[313, 385]. This has led to the realisation that the activation of SMO must require distinct steps[313]. However, the mechanisms that control these steps remain unknown.

In unstimulated cells, Ptc1 is present in cilia and around the ciliary base (Fig. 1.4, panel b)[314]. Upon Shh stimulation, Ptc1 translocates out of the primary cilium. However, this translocation is not an essential step in releasing its inhibition of SMO as a Ptc1 mutant that could not move out of the primary cilium was otherwise able to operate normally[188]. In the absence of the Hh ligand, SMO can diffuse into the primary cilium but spends prolonged periods at various sites at the base of the cilium[242]. Upon activation, the affinity of SMO for these sites decreases substantially, allowing SMO to move into the cilium. Gli2 and Gli3, which govern most signalling, are not detectable in cilia in the absence of the Hh ligand but they seem to transit through it[389, 363, 189, 153, 31]. Suggestions have been made about the primary cilium containing a special lipid environment that may enable clustering leading to signalling or an abundance of a particular second messenger small molecule[110]. It is not yet clear what about the primary cilium makes it significant for Hh signalling[132, 315].

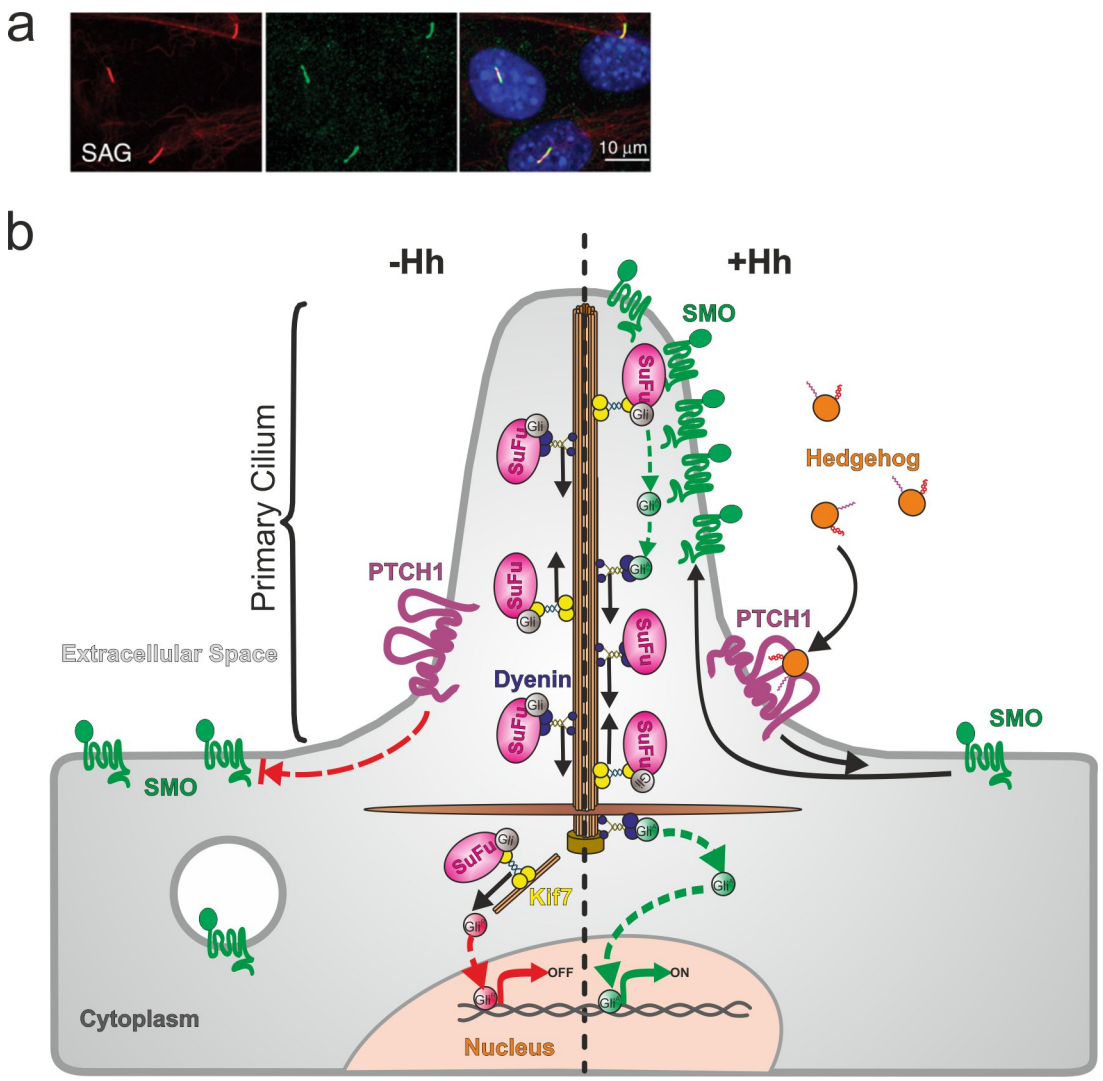


Figure 1.4: Role and structure of the primary cilia. a, Primary cilia in NIH-3T3 cells. Immunoblots using antibodies against tubulin (red), SMO (green) and DAPI (blue). SAG treatment induces ciliary localisation of SMO. Images reproduced from [314], figure 1. **b,** Schematic of Hh pathway components in and around the primary cilium.

1.3 G-protein coupled receptors (GPCRs)

1.3.1 Roles of GPCRs

GPCRs are a large family of transmembrane proteins that are identified by their heptahelical transmembrane bundle. There are over 800 different GPCRs in humans (Fig. 1.5, panel a) and they respond to a wide variety of extracellular signals ranging from photons to small molecules to proteins, allowing sensing of the external environment with key roles in vision, taste and smell, as well as mediating cell-to-cell communication in many contexts. GPCR signal transduction is essential for most physiological processes. GPCRs mediate fundamental signalling pathways within the body, including neurological, cardiovascular, endocrine, and reproductive functions.

1.3.2 Structure & function of GPCRs

For many years, GPCRs defied efforts to be expressed in a stable form and crystallised for structural studies. The invention and development of various techniques, such as stabilising point mutations, expression-boosting fusion proteins and lipid-based crystallisation[128], have led to the structural characterisation of several dozen distinct GPCRs, many of which have been solved in multiple conformations and in different complexes (for a full list see gpcrdb.org). Despite these advances, GPCRs remain challenging targets for structural biology.

Typically, GPCRs are activated by a small molecule ligand binding to the orthosteric pocket on the outer leaflet side of the protein, which induces a conformational change through several helices to the cytoplasmic side[283]. Canonical signalling entails the binding of a heterotrimeric G-protein complex to the intracellular

face of the GPCR (Fig. 1.5, panel b). Upon activation of the GPCR, the G-protein exchanges its bound GDP molecule for a GTP molecule, leading to release of the G-protein which splits into two subunits. The Gα subunit goes on to activate adenylate cyclase and induce the production of cyclic AMP (cAMP)[343]. The G$\beta\gamma$ dimer subunit activates various ion channels. The activated GPCR can bind to other GDP-bound G-proteins in order to maintain the signal. Activated GPCRs can also bind to β-arrestin which prevents binding to G-proteins. GPCR signalling is inactivated when its cytoplasmic tail is phosphorylated by GPCR-related kinases (GRK)[359].

The most widely studied GPCRs consist of a heptahelical transmembrane domain (TMD) with little extra bulk. On the extracellular or lumen side of the TMD bundle there is the so-called "orthosteric" ligand binding site for small molecules[298]. On the cytoplasmic face there is a binding site for downstream effector molecules, such as heterotrimeric G-proteins and β-arrestin[298]. One or more of the GPCR's extracellular loops may act as a lid on its orthosteric site[390, 10]. One or more "allosteric" sites on the outside of or within the GPCR's TM bundle may also modulate the signal of the orthosteric site[99, 200, 181, 180].

1.3.3 Small molecules modulating GPCRs

GPCRs are readily modulated by small molecules that bind to their relatively accessible orthosteric binding site, making them excellent drug targets. The drug does not need to pass through a membrane and the GPCR is already integral to a signalling function. The β_2 adrenergic receptor (β_2AR) is one GPCR that mediates a range of physiological responses, including smooth muscle relaxation. Its native small molecule ligand is the hormone and neurotransmitter, epinephrine (also known as adrenaline). A host of synthetic small molecule ligands for β_2AR have also been

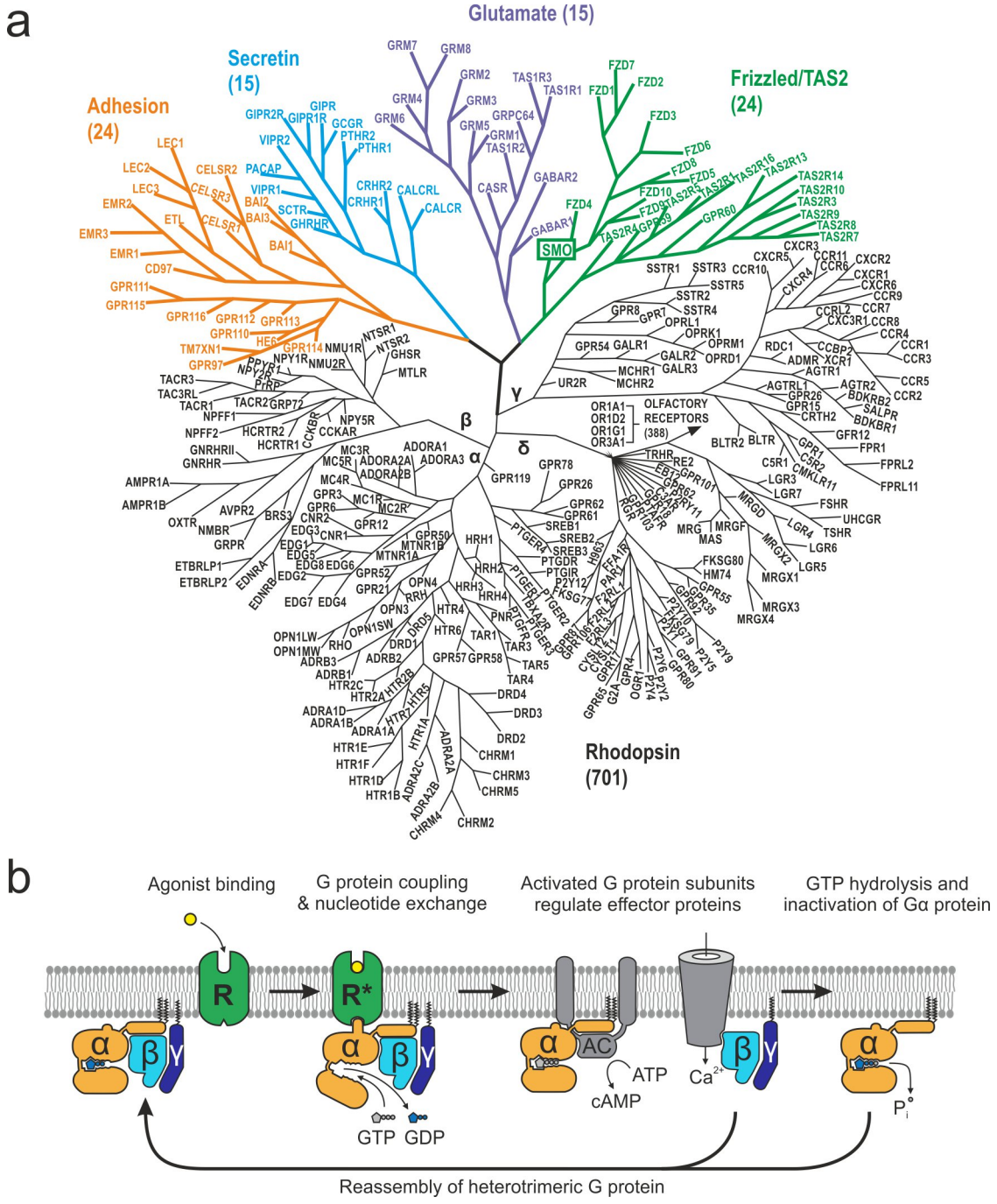


Figure 1.5: The GPCR superfamily and canonical function. a, Phylogenetic tree of GPCR superfamily. Different families are shown in different colours. SMO is highlighted with a box within the green section. Adapted from [346], figure 1. **b,** Canonical functioning of GPCRs. R, receptor; α, Gα subunit; β, Gβ subunit; γ, Gγ subunit; AC, adenylyl cyclase. Reproduced from [298], figure 1.

developed including the agonist salbutamol, marketed as Ventolin and used to treat asthma, and beta blockers, including propanolol, a class of antagonists used to treat cardiac arrhythmia and hypertension. Dopamine receptors are another type of GPCR that respond to the neurotransmitter dopamine and are prominent in the vertebrate central nervous system. The dopamine D_2 receptor (D2R) is the main target for all antipsychotic drugs. Recently, several small molecule ligands of GPCRs have been discovered that can preference one downstream signalling pathway over another, such as G protein-mediated signalling over β-arrestin mediated signalling[376, 48]. These small molecules are called "biased" ligands and offer hope for generating therapeutics with fewer side effects, such as pain reduction without substance addiction[231].

1.4 Smoothened

1.4.1 Role of SMO

Unlike canonical GPCR signalling, SMO does not act as the primary receptor for an extracellular ligand, with Ptc1 instead functioning as the receptor for all Hh morphogens[349, 234]. Shh-N binds to Ptc1 with high affinity in order to activate the pathway and does not interact at all with SMO (Fig. 1.6, panel a)[29, 349]. The primary role of SMO is transduction of the Hh signal across the plasma membrane[253], mediating the interaction between Ptc1 and the intracellular components of the pathway. Crucial details remain unknown: What is the nature of Ptc1's interaction with SMO? How does SMO transduce the signal across the plasma membrane? What is the nature of SMO's interaction with its downstream signalling partners?

1.4.2 Interactions of SMO

SMO has been reported to interact with a range of other proteins. SMO is constitutively inhibited by Ptc1 through an unknown catalytic mechanism (Fig. 1.6, panel a)[161, 353]. Inactive SMO appears to interact with something at the ciliary base, which restricts its movement into the primary cilium[242]. SMO activation involves phosphorylation at multiple sites on its intracellular C-terminal tail by the serine/threonine kinases Ck1α and Grk2(ref.[57]). β-arrestin and the anterograde motor protein Kif3A facilitate the ciliary accumulation of active, phosphorylated SMO[56, 197]. Phospholipase A2 enhances SMO ciliary translocalisation[6]. Integrin-linked kinase is permissive for ciliary accumulation of SMO and interacts with β-arrestin[13]. Pitchfork and Gprasp2 are essential for Hh-induced ciliary localisation of SMO. Gprasp2 may form a direct interaction with SMO and act as a bridge

between SMO and Pitchfork[169]. Activated SMO interacts with EVC and EVC2, two single-pass type I transmembrane proteins that are confined to the base of the primary cilium[103].

1.4.3 SMO as GPCR

SMO also interacts with the heterotrimeric G_i protein[307, 286, 373]. The designation of SMO as a "G-protein coupled" receptor is disputed, however, because several of its characteristics do not accord with that of a typical GPCR[11]. SMO contains seven transmembrane α-helices that are arranged in an architecture very similar to other GPCRs[379]. It also responds to extracellular small molecules and can interact with G proteins[307, 286, 373], both hallmarks of a GPCR. In canonical GPCR signalling, G-proteins act as the primary downstream effector molecules that facilitate changes in the concentration of second messengers such as Ca^{2+} and the primary cilium has been shown to be a specialised Ca^{2+} signalling organelle[94, 91], although a direct link between SMO and Ca^{2+} concentration has yet to be demonstrated. Links have also been made between G_i protein activation and downstream Gli signalling[410, 268]. However, it seems that Hh signalling can operate at least partially independently of G protein effectors[11, 307, 114]. This is based primarily on the observation that the pertussis toxin, which prevents G_i from coupling to GPCRs, does not impact Gli-dependent development in chick neural tube and zebrafish patterning[224, 140]. This is as yet an unresolved area of debate and will require further examination.

1.4.4 Structure of SMO

SMO is a Class F (Frizzled-type) GPCR (Fig. 1.5, panel a, green branch)[366]. Unlike Class A GPCRs, SMO contains not just the stereotypical heptahelical transmembrane

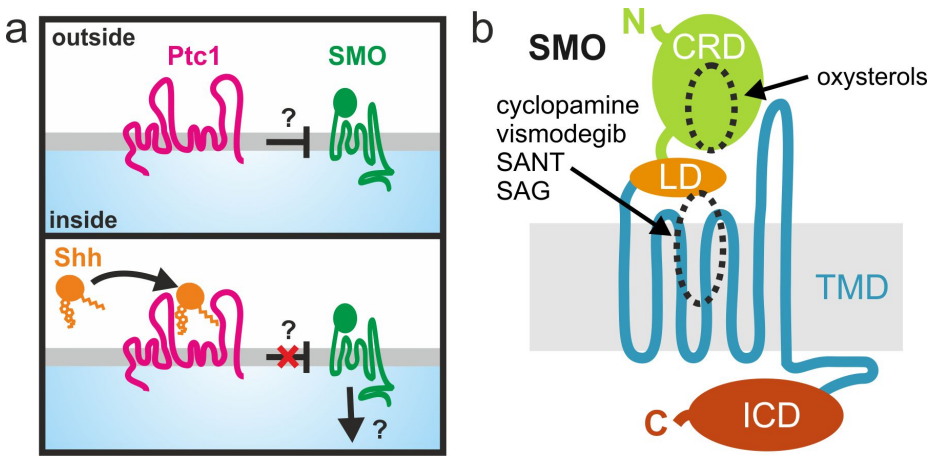


Figure 1.6: Schematic of SMO function and architecture. a, Unknown mechanism of interaction between the two essential transmembrane proteins, Ptc1 and SMO. Ptc1 releases its constitutive inhibition of SMO upon Shh binding, allowing SMO to activate downstream signalling. **b,** Domain architecture of SMO. Domains are individually coloured and labelled. Black dotted ovals indicate the two ligand binding sites with several small molecule ligands listed for each. N- and C-termini are labelled.

domain bundle but also sizeable extracellular and intracellular domains (Fig. 1.6, panel b). SMO contains at least two small-molecule binding sites: one in the TMD, which was classically defined as the cyclopamine binding site[52], and one in the extracellular cysteine-rich domain (CRD), which was recently determined to bind a different class of small molecules called oxysterols (Fig. 1.6, panel b)[263, 258].

Several crystal structures of the human SMO TMD were recently solved[379, 378, 388]. These were all determined in complex with synthetic small molecules, revealing a binding site very similar to the orthosteric binding sites of canonical GPCRs. Four of the structures were solved with antagonists and one was solved with an agonist. However, there were no major conformational differences between the agonist and antagonist structures, suggesting that only one signalling state of SMO had been captured regardless of the small molecule ligands used. This is discussed further in Chapter 3.2.1 (page 65).

The extracellular region of SMO consists primarily of the CRD, named for

its conserved disulfide bonding architecture, which has homologues in the Frizzled receptors (Fzd), Niemann-Pick type-C proteins 1 and 2 (NPC1 and 2), and riboflavin-binding protein[15]. A crystal structure of the zebrafish SMO CRD and a solution NMR structure of the *Drosophila* SMO CRD were recently determined[258, 296]. While neither structure contained a small molecule, in both instances a small molecule binding site was mapped to one particular location that corresponded well to the binding pocket of the CRD of Fzd in which the lipid moiety of Wnt binds[165]. This is discussed further in Chapter 3.2.2 (page 68).

Without any structural information, relatively little is known about the intracellular domain (ICD) of SMO. Various regions of the ICD are important for localisation to the primary cilium, although how they facilitate this is unknown. It has been proposed that signal activation is regulated by a conformational switch of SMO's ICD[413]. However, this study was conducted using *Drosophila* SMO, which operates at the cell surface without a primary cilium (*Drosophila* do not have primary cilia). It is not clear to what extent the *Drosophila* studies inform on mammalian SMO given the substantial divergence between invertebrate and vertebrate Hh signalling[369]. Several regions of the vertebrate SMO ICD can be phosphorylated and other regions have been shown to be essential for ciliary localisation[262].

1.4.5 Regulation of SMO by Ptc1

SMO is constitutively active in cells in which Ptc1 is missing or deactivated[253, 193]. In initial co-immunoprecipitation assays Ptc1 was demonstrated to form a complex with SMO[349, 253]. However, the discoveries that Hh causes Ptc1 to be removed from the cell surface and SMO to be stabilised at the cell surface[95], and that Ptc1 regulates SMO at sub-stoichiometric ratios[353, 161] led to the reasoning

that Ptc1 regulates SMO via an unknown catalytic mechanism. Given that a catalytic mechanism would also be transient, it is likely that the interaction initially observed between the two membrane proteins using co-immunoprecipitation assays was at least somewhat forced through overexpression of both Ptc1 and SMO[349, 253].

Ptc1 shows sequence similarity to NPC1 proteins, which are multi-pass transmembrane proteins with long extracellular loops and are involved in endosomal cholesterol transport[87, 205]. Both contain a conserved structural motif called the sterol-sensing domain (SSD)[204]. First discovered in HMG-CoA reductase[129], this five transmembrane helix motif has been identified in proteins with a range of functions and across phyla. It was initially hypothesised that Ptc1 directly interacted with the cholesterol adduct of Shh-N via its SSD[35]. However, Shh-N without the cholesterol modification retains its high affinity for Ptc1 *in vitro*[280] and mutations in the SSD of *Drosophila* Ptc do not disrupt Hh binding *in vivo*[236, 350], suggesting that the SSD of Ptc1 is dispensable for Shh-N binding.

Ptc1 and NPC1 show sequence similarity to the resistance-nodulation-division (RND) permease superfamily of bacterial efflux pumps[361], suggesting a potential role for Ptc1 in small molecule transport. This has led to speculation that Ptc1 may regulate SMO via a small molecule second messenger, possibly a sterol[353]. The recent discovery that SMO can be regulated by oxysterol molecules added weight to this hypothesis[257, 263]. If Ptc1 is a small molecule transporter, it is not clear whether it mediates influx or efflux, or whether the small molecule is an agonist or antagonist. Due to the elusiveness of this theorised second messenger compound, it has been dubbed "Molecule X" by some commentators.

1.4.6 Small molecules modulating SMO

The role of Hh signalling in tumorigenesis has prompted intense research into small molecules that target the Hh pathway and can be used as anti-cancer drugs. Efforts have focused primarily on SMO, because of its homology to other GPCR drug targets. The discovery of cyclopamine, a naturally occurring small molecule antagonist for SMO, provided a promising starting point for drug discovery efforts[52].

There are now a plethora of SMO-targeting small molecules available. Two small molecule antagonists have been approved by the US Food and Drug Administration (FDA) for treatment of advanced basal cell carcinoma: vismogedib (produced by Genentech, Roche, and marketed as Erivedge)[335] and sonidegib (produced by Novartis and marketed as Odomzo) (Fig. 1.7, panels a & b)[44]. These drugs have been investigated for the treatment of other types of cancer but various tumours have been seen to rapidly develop resistance to these drugs (Fig. 1.7, panel c)[320]. As a result, more effective therapies are being sought. A range of other SMO antagonists have been discovered or developed for functional and structural studies, including SANT1 (also, SANTs 2, 3 and 4)(ref.[53]), Anta XV (ref.[245]), LY2940680 (ref.[18]), MRT-92 (ref.[146]), itraconazole[190], budesonide[380, 383], and 22-NHC[263]. A range of agonists have also been discovered or developed, including SAG[53] (also, SAGs 1.1, 1.2, 1.3, 1.4 and 1.5 (ref.[119])), purmorphamine[336], 20(S)-OHC and 20(S)-yne[257]. Additional small molecule agonist and antagonists continue to be developed (e.g. [50, 206, 21, 225, 202, 5]). For more details of the small molecule ligands used in this thesis, please refer to Appendix C (page 228).

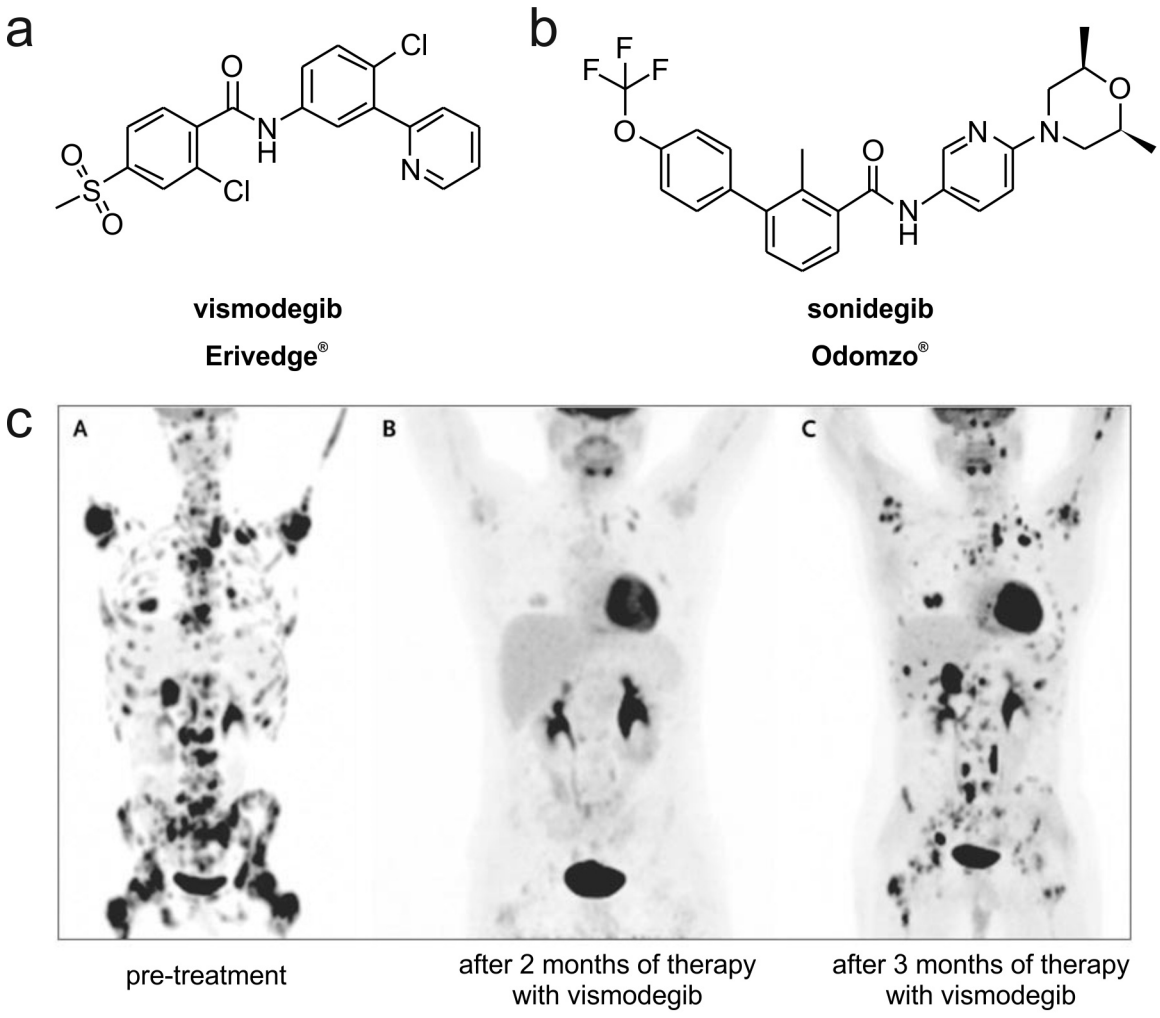


Figure 1.7: SMO-targeted drugs that are FDA-approved for the treatment of advanced basal cell carcinoma. **a,** Chemical structure of vismodegib from Genentech, Roche. **b,** Chemical structure of sonidegib from Novartis. **c,** Positron-emission tomographic (PET) scans of tumour response to vismodegib treatment in a patient with severe metastatic medulloblastoma. Whole-body projections from ^{18}F-fluorodeoxyglucose-PET scans at different time points. Dark patches indicate tissues with high metabolic activity, which includes tumours. Panel reproduced from [320], figure 1.

1.5 Thesis outline

SMO is an important, interesting and appropriate target for structural characterisation because it is a key component of an essential signalling pathway, its signal transduction mechanism is not known, it is part of a distinct class of GPCRs with an extracellular domain (no structures of which have yet been solved with both the transmembrane and extracellular regions), it is a clinically relevant drug target, and many small molecule ligands (both agonists and antagonists) for it are known.

Materials & Methods (Chapter 2, from page 23) are detailed in a single chapter immediately following the Introduction.

The results of this thesis are arranged in three chapters, each containing a focused introduction, reporting of experimental results and discussion of significant points. The first results chapter (Chapter 3, from page 62) details the efforts to express, stabilise and crystallise SMO, which resulted in the first structure of SMO with its full extracellular and transmembrane domains (CRD-LD-TMD) as well as the unexpected discovery of cholesterol as a new endogenous ligand for SMO. The second results chapter (Chapter 4, from page 119) extends the work by determining multiple additional structures of SMO CRD-LD-TMD bound to different small molecule ligands. The third results chapter (Chapter 5, from page 165) details efforts to generate nanobodies against the extracellular domains of SMO in order to stabilise different conformational states.

The final chapter (Chapter 6, from page 191) contains discussions on a range of areas related to the results described herein.

Chapter 2

Materials & Methods

2.1 Molecular Biology

2.1.1 cDNA and expression vectors

For expression in mammalian cells, constructs of human SMO (UniProt Id. Q99835) were cloned into the pHLSec vector[7] in frame to either a C-terminal Rho1D4 antibody epitope tag (1D4-tag)[249, 269] or a monoVenus tag[409, 259] followed by a Rho1D4 antibody epitope tag (for plasmid map, see Appendix B, page 222). Variants of this vector included an N-terminal HA tag (YPYDVPDYA) or a C-terminal Avi tag (GLNDIFEAQKIEWHE) for biotin ligase (BirA) recognition[113, 85]. The pHLSec vectors contain an optimised Kozak consensus sequence for initiation of translation, an optimised signal peptide derived from chicken receptor tyrosine phosphatase (GILPSPGMPALLSLVSLLSVLLMGCVA), which trafficks the translated protein to the cell membrane, and a suite of unique restriction sites (EcoRI, AgeI, KpnI, XhoI) for efficient modular editing of the vector (for plasmid map, see Appendix B, page 222)[7].

Nanobody constructs were cloned into the bacterial expression vector pMESy4 [274], in frame to an N-terminal PelB signal peptide (KYLLPTAAAGLLLLAAQPAMA) for periplasmic localisation and a C-terminal 6xHis tag (for plasmid map, see Appendix B, page 222). For fluorescence microscopy experiments, nanobody constructs were cloned into the bacterial expression vector pET22b (Novagen® #69744), in frame to an N-terminal PelB signal peptide followed by a FLAG tag (DYKDDDDK) with a C-terminal 6xHis tag (for plasmid map, see Appendix B, page 222).

2.1.2 Construct design

Three human SMO constructs were designed for initial expression screening: BRIL-ΔCRD-SMOΔC, SMOΔC and SMOΔC-BRIL (Fig. 2.1). The first, BRIL-ΔCRD-SMOΔC, was truncated at both N- and C- termini to remove the CRD and intracellular domains, leaving only the linker and transmembrane domains (UniProt Id. Q99835: residues Ser190-Gln555). A thermostabilised apocytochrome b_{562} four-helical bundle (BRIL; UniProt Id. P0ABE7 residues Ala23-Tyr128)[72, 71] was fused to the N-terminus of the truncated SMO. This construct was designed to exactly replicate the crystallisation construct used by Wang *et al.*[379]. The second construct, SMOΔC, was truncated at the C-terminus to remove the intracellular domain, leaving the extracellular and transmembrane domains (residues Ser32-Gln555). The N-terminus was also shortened slightly to remove the potentially flexible tail. Aside from the signal peptide and the 1D4-tag incorporated by the pHLSec vector, this construct had no additional modifications. The third construct, SMOΔC-BRIL, was made with the same truncation boundaries as SMOΔC but with intracellular loop 3 (ICL3), located between transmembrane helices V and VI (residues Ile429-Ala442), replaced with BRIL (residues Ala23-Tyr128). The insertion point of BRIL was determined by modelling the relative positions of the previously-solved structure of the SMO TMD (PDB 4JKV[379]) with the BRIL bundle (PDB 4EIY[222]) on the intracellular face of the TMD, taking into consideration the register of the connecting α-helices. The goal was to place the BRIL in such a way that helices V and VI of the SMO TMD extended directly into the connecting helices of BRIL with no break in the α-helical geometry, thereby reducing flexibility of the 7TM bundle and potentially increasing thermostability. Synthetic genes for each construct were obtained from GeneArt^{TM}(Thermo Fisher Scientific).

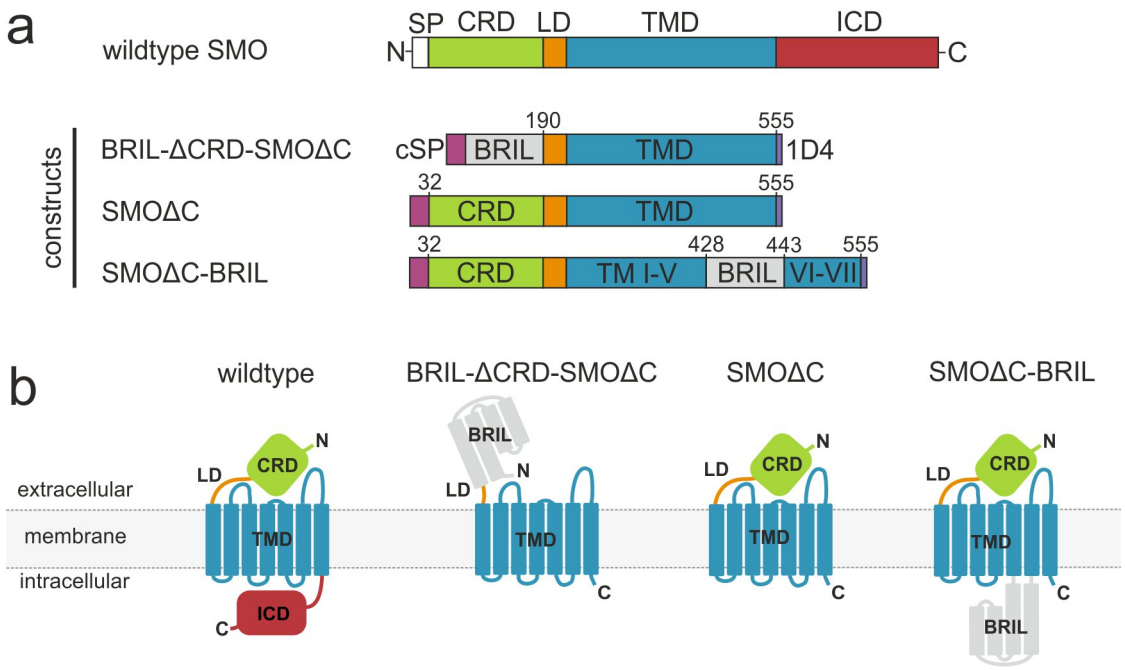


Figure 2.1: SMO construct design. a, Linear domain schematic of wildtype SMO and the primary three constructs. Numbers indicate domain boundaries of SMO. (SP, signal peptide; CRD, cysteine-rich domain; LD, linker domain; TMD, transmembrane domain; ICD, intracellular domain; BRIL, thermostabilised apocytochrome b_{562}RIL fusion; cSP, chicken signal peptide; 1D4, Rho1D4 epitope tag.) **b,** Cartoon schematic representation of wildtype SMO and the three primary constructs. Domains coloured as in panel a.

2.1.3 PCR and site-directed mutagenesis

PCR reactions were carried out in a 50 μL reaction volume containing: 5 μL 10X Pyrobest^{TM}Buffer II (Takara), 5 μL dNTP mixture (2.5 mM, Takara), 1 μL each of forward and reverse primers (10 μM), 1 μL template DNA (10 μg/mL) and 0.2 μL Pyrobest^{TM}DNA Polymerase (5 U/μL, Takara) made up to 50 μL with MilliQ^{TM}H_2O. PCR reactions were carried out in a T100 thermal cycler (BIORAD) as follows:

1. 95°C for 180 s (initial denaturation).
2. 95°C for 30 s (denaturation).
3. 55°C for 30 s (annealing).

4. 72°C for 60 s per kilobase of PCR product (elongation).

5. Repeat steps 2-4 34 times.

6. 72°C for 300 s (final elongation).

7. 4°C for ∞.

PCR products were purified by gel extraction using a 1-2% agarose gel in Tris-acetate-EDTA (TAE) buffer, excised from the gel and extracted using the QIAquick Gel Extraction Kit (Qiagen).

Overlapping PCR was used to introduce point mutations into the constructs. This method required four primers: one each to define the 5′ and 3′ ends of the construct, and a pair of complementary primers that spanned the target site ($\sim$21 nucleotides long) harbouring the required mutation at their centre (Fig. 2.2). In the first step, two PCRs were carried out to produce two DNA fragments comprising the sequence from the mutation site in the centre to either the 3′- or 5′-terminus. In the second PCR step, the two gel-purified overlapping PCR products from the first step were used as the template (1 μL each) and together were amplified using the 3′- and 5′-terminus primers. This extends the two fragments into a single fragment containing the entire length of the expression construct with the required mutation in the centre (Fig. 2.2). This DNA fragment was then gel-purified and cloned into the required vector. The same recipe and thermal cycle procedure was followed for overlapping PCR reactions as for normal PCR reactions except with 1 μL of each of the overlapping DNA templates (10-20 μg/mL) in place of the single DNA template. PCR primers were obtained from Eurofins (Eurofins Genomics©).

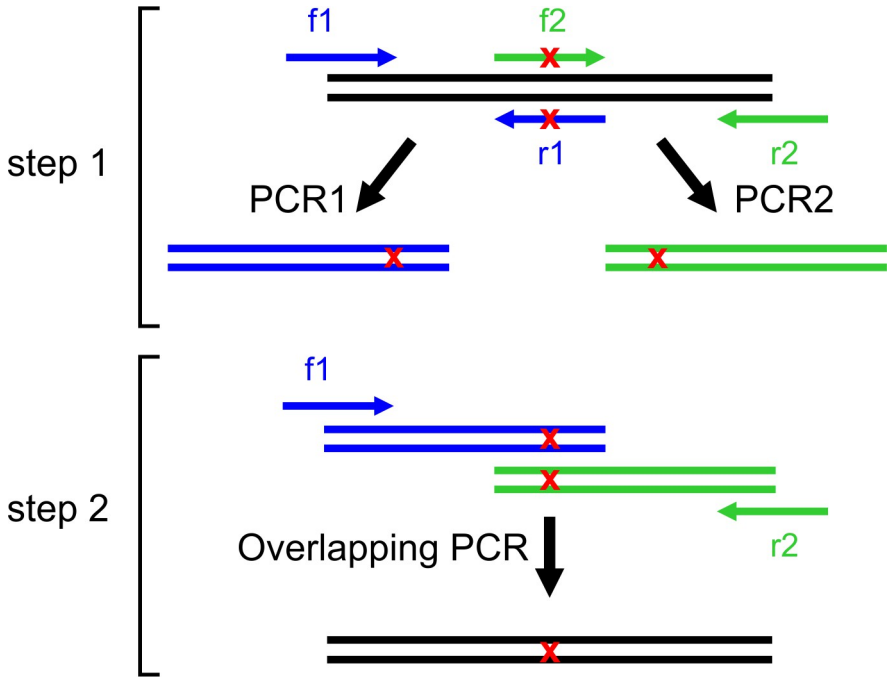


Figure 2.2: Site-directed mutagenesis by overlapping PCR. Step 1: the original double-stranded DNA template (black) is amplified in two fragments (blue and green) with the point mutation (red 'X') introduced into both by the central primers (f2 and r1). Step 2: the two fragments from step 1 along with the terminal primers (f1 and r2) are combined in one reaction to produce a full-length double-stranded product including the point mutation.

2.1.4 Restriction digests, ligations and transformations

Purified PCR products and 3 μg of the target vector were separately digested with 0.5-1.0 μL of the required restriction enzymes using the CutSmart® Buffer system for 1 h at 37°C (New England Biolabs). Digested products were then loaded onto a 1-2% agarose/TAE gel. Bands of the expected size were excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen).

Ligation reactions were carried out using the Quick Ligation™Kit (New England Biolabs) in a final volume of 20 μL according to the manufacturer's instructions, using a vector:insert molar ratio of approximately 1:4. Subsequently, 1 μL of this ligation mixture was used to transform 20 μL of Library Efficiency

DH5α^{TM}Competent Cells (Thermo Fisher Scientific). The DNA-cell mixture was incubated on ice for 30 min before being heat-shocked at 42°C for 45 s. After incubation on ice for 5 min, 200 μL Super Optimal Broth with Catabolite repression (SOC) medium was added and the mixture was incubated at 37°C for 30 min with 200 rpm shaking. Cells were then spread onto LB-Agar plates containing 100 μg/mL ampicillin and incubated at 37°C overnight. Clonal colonies were used to inoculate overnight cultures (5 mL LB containing 100 μg/mL ampicillin) for purification of the amplified plasmid DNA using the QIAprep Spin Miniprep Kit (Qiagen). Successful ligations of the cloned plasmids were checked by analytical restriction enzyme digestion and finally the clones were verified by Sanger sequencing (GATC Biotech). All DNA was stored at -20°C.

Larger quantities of DNA required for large scale transient transfection of HEK293 cells were obtained by re-transforming the cloned plasmid (transformation carried out as above), inoculating a starter culture (5 mL LB containing 100 μg/mL ampicillin) and expanding to 2 L before purifying the plasmid DNA using the HiSpeed Plasmid Giga EF Kit (Qiagen).

2.1.5 SDS-PAGE and Western blots

For analysis, proteins were separated according to their size by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples were mixed with SDS-loading dye (including β-mercaptoethanol for reducing conditions), loaded onto a 4-12% Bis-Tris NuPAGE® gel and run at 165 V for approximately 45 min with either MES or MOPS buffer, depending on the desired separation range (Thermo Fisher Scientific). Distinct from common SDS-PAGE procedures, samples of membrane proteins were not boiled prior to loading onto the gel as boiling tended to

increase aggregation. BenchMark^{TM}Protein Ladder (Thermo Fisher Scientific) was included for molecular weight estimation.

For Western blotting, proteins run on SDS-PAGE were transferred onto a nitrocellulose membrane (Hybond-ECL, GE Healthcare) using NuPAGE Transfer Buffer (Thermo Fisher Scientific) and run at 30 V for 90 min. The membrane was blocked with 5% skim-milk in Tris-buffered saline plus Tween-20 (TBST; 50 mM Tris pH7.6, 150 mM NaCl, 0.1% Tween-20) for 30 min and then probed with primary antibody diluted in 5% milk/TBST for 1 h at room temperature. Following a brief TBST wash step, the membrane was probed with secondary antibody conjugated to horseradish peroxidase (HRP) diluted in 5% milk/TBST for 1 h at room temperature. The membrane was then washed 5 times for 5 min each in TBST. Luminescent signal was detected using an enhanced chemiluminescence (ECL) kit (GE Healthcare) and imaged on a ChemiDoc XRS+ imager (Bio-Rad).

The primary antibody was either mouse anti-His6 (1:1000 dilution; Qiagen), mouse anti-1D4 (1:10,000 dilution; University of British Colombia) or mouse anti-HA (1:10,000 dilution; Sigma). The secondary antibody was HRP-conjugated goat anti-mouse IgG (1:2000 dilution; Sigma). BenchMark^{TM}Pre-stained Protein Ladder (Thermo Fisher Scientific) was used to monitor SDS-PAGE running and transfer efficiency and was supplemented with BenchMark^{TM}Protein Ladder for molecular weight estimation. BenchMark^{TM}Protein Ladder was visualised on Western blot by using an anti-HA antibody.

2.2 Protein expression and purification

2.2.1 Small-scale mammalian expression trials

Proteins were expressed in HEK-293S-GnTI$^-$ (ATCC® CRL-3022™) cells to reduce N-linked glycosylation heterogeneity[300, 8]. Screening constructs were tagged with a C-terminal monoVenus fluorescent tag[259] to enable efficient detection of sub-nanogram quantities of membrane protein.

For small-scale screening, cells in adherent format were grown to confluency and transiently transfected using Lipofectamine® 2000 (Thermo Fisher Scientific). After 48-72 h, cells were resuspended in 10 mM HEPES pH7.5, 300 mM NaCl, 1% (v:v) mammalian protease inhibitor cocktail (P8340, Sigma-Aldrich) buffer, then solubilised with 1.5% detergent and rotated for 1.5 h at 4°C. Insoluble material was removed by centrifugation (29,000 xg, 10 min, 4°C). Clarified supernatant was then incubated for 2 h at 4°C with purified Rho1D4 antibody (University of British Columbia) coupled to CNBr-activated sepharose beads (GE Healthcare). Protein-bound beads were then washed with 10 mM HEPES pH7.5, 300 mM NaCl buffer containing detergent at ∼3x critical micelle concentration (CMC) and eluted overnight in the same buffer supplemented with 500 μM TETSQVAPA peptide (Genscript).

For analysis, samples were loaded onto a Superose 6 3.2/300 column (GE Healthcare) equilibrated with 10 mM HEPES pH7.5, 300 mM NaCl, 0.007% (w/v) decyl maltose neopentyl glycol (DMNG, Anatrace), 0.0005% (w/v) cholesterol hemisuccinate (CHS, Anatrace) buffer, attached to a high-performance liquid chromatography system with automated micro-volume loader and in-line fluorescence detection (FSEC)[175]. Switching the column into different running buffers was very time-consuming due to the low maximum-allowable flow-rate. As a consequence, 0.007%

DMNG with 0.0005% CHS was used as the standard FSEC running buffer detergent supplement across all samples. DMNG has the lowest CMC of the detergents available (∼0.036 mM or 0.0034%), meaning a smaller quantity was required to maintain detergent micelles in solution[45]. Due to its uncharged nature, its large hydrophilic head group and its long hydrophobic tail, DMNG is also a very mild detergent, meaning it does not tend to denature proteins (for its chemical structure, see Appendix D, page 231)[291, 355, 221]. CHS had been previously observed to improve stability of detergent-solubilised membrane proteins in general[415, 25].

As an alternative procedure, the clarified supernatant from the detergent-solubilised cells may be loaded directly onto the FSEC without performing the intervening affinity purification step detailed above[175]. The resulting protein samples are therefore not pure but this does not matter because only the mVenus-tagged protein will be picked up by the fluorescence detector. This procedure requires fewer purification steps but has the potential to leave more precipitant on the column. Close monitoring of column health is advised.

2.2.2 Large-scale mammalian expression and purification

SMO constructs were expressed by transient transfection in HEK-293S-GnTI$^-$ (ATCC® CRL-3022™) cells in a typical batch volume of 9.6 L. Cells were grown in suspension at 37°C, 8.0% CO_2, 130 rpm to densities of 2-3 x 10^6 cells mL^{-1} in protein expression media (PEM, Invitrogen) supplemented with 2 mM L-glutamine (L-glut, Gibco), 0.1 mM non-essential amino-acids (NEAA, Gibco) and 1% foetal calf serum (FCS, Sigma-Aldrich). Cells from 0.8 L cultures were collected by centrifugation (1000 xg, 7 min) and resuspended in 120 mL Freestyle293 medium (Invitrogen) containing 1.2 mg PEI Max (Polysciences), 0.4 mg plasmid DNA and 5 mM valproic acid (VPA, Sigma-

Aldrich), followed by a 3-6 h incubation at 160 rpm. Culture media were subsequently topped up to 0.8 L with PEM and returned to 130 rpm. 48-72 h post-transfection cell pellets were harvested by centrifugation (1400 xg, 5 min), snap-frozen in liquid N_2 and stored at -80°C, resulting in a total of 120-180 g of cell mass per 9.6 L suspension medium, depending on the initial cell concentration[243].

Frozen cell pellets were thawed, resuspended in 10 mM HEPES pH7.5, 300 mM NaCl buffer supplemented with a 1:100 (v:v) dilution of mammalian protease inhibitor cocktail (P8340, Sigma-Aldrich) and solubilised with 1.0-1.5% (w/v) n-dodecyl-β-D-maltopyranoside (DDM, Anatrace) and 0.2-0.3% (w/v) CHS (Anatrace), then stirred vigorously for 1.5 h at 4°C. Insoluble material was removed by centrifugation (16,000 xg, 15 min, 4°C) and the supernatant incubated for 2 h at 4°C with purified Rho1D4 antibody (University of British Columbia) coupled to CNBr-activated sepharose beads (GE Healthcare). Protein-bound beads were washed extensively with 50 mM HEPES pH7.5, 300 mM NaCl, 10% (v/v) glycerol, 0.1% DDM, 0.02% CHS buffer and then with 50 mM HEPES pH7.5, 300 mM NaCl, 10% glycerol, 0.05% DDM, 0.01% CHS buffer and eluted overnight in 50 mM HEPES pH7.5, 300 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS, 500 μM TETSQVAPA peptide (Genscript). Eluate was concentrated to approximately 500 μL using a Vivaspin® Turbo 4 (or 15) PES 100 kDa MWCO concentrator (Sigma-Aldrich) and loaded onto a Superose 6 10/300 Increase column (GE Healthcare) equilibrated with 10 mM HEPES pH7.5, 150 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS. Peak fractions were pooled and concentrated using a Vivaspin® 500 PES 100 kDa MWCO concentrator (Sigma-Aldrich) to ∼3.0 mg/mL for vapour diffusion crystallisation or to ∼30 mg/mL for LCP crystallisation.

For crystallisation, samples were treated with Endoglycosidase F1 (EndoF1) for 1 h at room temperature to trim N-linked sugars down to a single N-acetyl-

glucosamine (GlcNAc) moiety bound to the corresponding asparagine side chain. This enzymatic cleavage was only possible because the protein was expressed in HEK-293S-GnTI$^-$ cells, which are deficient in N-acetylglucosaminyl transferase I activity and therefore lack complex N-glycans. The relatively simple and homogeneous N-linked glycans produced by these cells, such as $Man_5GlcNAc_2$ (Man, mannose), can be efficiently cleaved by EndoF1[300, 8, 88].

2.2.3 Thermostability assay using analytical size exclusion chromatography (SEC)

The thermostability of detergent-solubilised constructs was determined by heating them over a range of temperatures for equal time periods and then measuring the reduction in the peak intensity of the SEC profile for each sample[143]. Purified protein was distributed into 50 μL aliquots and each heated to a temperature between 30-80°C for 1 h before being run on SEC. An unheated control sample (4°C) was included for comparison. Maximum intensity of the monodisperse peak (peak height) for each sample was plotted against temperature to produce a 'melting curve' for the protein. Less reduction in peak intensity at increased temperatures indicates greater thermostability. For additive screens, purified protein was aliquoted out, supplemented with an additive and then heated for 1 h at 42°C. This is the temperature at which the peak height of SMO without any additives was reduced to approximately 70% of its unheated intensity. Samples were then run on SEC. Additives observed to increase the stability of the protein at 42°C relative to a no-additive control were tested against the full range of temperatures.

For cholesterol depletion assays, the pooled peak fractions from a SEC purification were re-applied to 1D4-antibody beads. Equal volumes of beads were treated

with different quantities of methyl-β-cyclodextrin (MBCD) and incubated with gentle rocking for 1 h at 15°C before extensive washing with 50 mM HEPES pH7.5, 300 mM NaCl, 0.03% DDM and elution with 50 mM HEPES pH7.5, 300 mM NaCl, 0.03% DDM, 500 μM TETSQVAPA peptide. Thermostability of the different samples was assessed by heating aliquots of their eluates to the indicated temperatures for 1 h before being loaded onto a Superose 6 3.2/300 column on a Shimadzu system and monitored by absorbance at 280 nm. Samples were kept at the baseline temperature of 20°C when not heated. The construct used in this assay was not fluorescently tagged in order to avoid the potentially confounding effects of the fluorescent tag on overall stability.

2.2.4 Thermostability assay using intrinsic fluorescence

The thermostability of detergent-solubilised constructs was also determined using a thermal ramp with continuous broad-spectrum monitoring on an UNCLE system (Unchained Labs). This system allows dye-free detection of intrinsic fluorescence and static light scattering (SLS) to continuously monitor protein unfolding and aggregation, respectively. Purified protein was separated into 10 μL aliquots and 0.5 μL of highly concentrated small-molecule/DMSO stocks were added to each. After thoroughly mixing, 9 μL of each was loaded into the sample cuvette and run on the UNCLE. A typical run included 16-48 samples, monitoring both intrinsic fluorescence (primarily tryptophan: ex266/em310-380) and SLS (266 nm) at 1°C intervals, ramping at 0.5-1.0°C.min^{-1} over a temperature range of 20-90°C.

2.2.5 Design of membrane protein additive screen

In order to screen a wide array of conditions for stabilisation of membrane proteins, an additive screen for high-throughput thermostability assessment was designed and produced. This 96-condition screen included a selection of:

- Detergents
- Salts
- Lipids
- Amino acids
- Sugars
- pH screen (with and without salt)
- Alcohols
- Non-detergent sulfobetaines (NDSBs)
- Denaturants and reducing agents

Screening a large number of conditions using SEC on a Shimadzu system, where samples are run sequentially, means that some samples are incubated for much longer prior to running than others. MilliQ^{TM}H_2O control conditions were included within the screen at regular intervals to serve as time controls for protein stability. For a full list of conditions, see Appendix A: Membrane Protein Additive Screen (page 220).

2.3 Protein crystallisation and X-ray data collection

2.3.1 Basic principles of X-ray crystallography

X-ray crystallography is based upon two fundamental ideas. Firstly, that light interacting with objects or apertures that are similar in size to the light's wavelength will bend or diffract around them. Diffraction around multiple objects causes the rays of light to interact constructively or destructively, generating an interference pattern. Secondly, that a crystal is the macroscopic manifestation of a microscopic lattice made up of an enormous number of identical units arranged in three dimensions. Combining these two ideas, when light of a similar wavelength to the distance between the lattice planes of a crystal interacts with that crystal it produces a diffraction pattern. The diffraction pattern contains information about the arrangement of atoms within the crystal's repeating unit (the unit cell) and is mathematically related to the three-dimensional position of its constituent atoms (by Fourier transformation). The diffraction pattern can therefore be used to calculate the precise structure of the individual molecules that comprise the unit cell.

Bragg's Law (see equation below) describes the condition of the scattering angle θ, the angle between the incident light and the lattice planes, at which constructive interference is at its maximum, producing intense spots, so-called Bragg peaks.

$$2d \sin \theta = n\lambda$$

where d is the interplanar distance, λ is the wavelength of incident light and n is a positive integer. X-rays are used for crystallographic data collection because their wavelength is small enough to probe inter-atomic distances ($\sim$1 Å). Unfortunately,

the short wavelength of x-rays means that they are high energy and cause radiation damage to the crystal, reducing the quality of its diffraction pattern over time. X-rays interact with the electrons within the crystal rather than with the atomic nuclei so the resulting geometric information about the molecule reflects the localisation of the electrons (the electron density). The known geometries of electrons within atoms, atoms within molecules and molecular components within macromolecular assemblies (such as amino acid residues within proteins) allows structural models of macromolecules to be determined (built) from electron density.

2.3.2 Methods of crystallisation

The first requirement in x-ray crystallography is to generate crystals of the target molecule. Crystallisation is a form of controlled precipitation of a solute molecule out of a solvent. Generally, to grow crystals the solute must be forced out of solution slowly enough to form the regular arrays that make up a crystal. The most common method for protein crystallisation is vapour diffusion, in which a solution of purified protein and precipitant (often a poly-ethylene glycol) is equilibrated against a larger reservoir solution containing a higher concentration of precipitant in an airtight container. The concentration of protein is gradually increased in the crystallisation drop as water vapour transfers away from it and into the reservoir. A range of salt, precipitant, pH, temperature and additive conditions are generally tested for their suitability in growing crystals of a particular protein.

An alternative crystallisation approach that has been instrumental in producing crystals of GPCRs and other membrane proteins is lipidic cubic phase (LCP). In this method, the detergent-solubilised protein at very high concentration (20-100 mg/mL) is reconstituted into a lipid mixture (often 90% monoolein, 10% cholesterol)

at a fixed ratio that produces a lipid-based superstructure called the cubic phase (Fig. 2.3, panel a). This lipidic cubic phase acts as a structured medium in which membrane protein crystals may form. Small globules (boluses) of the protein-lipid mixture are applied to a glass plate and covered with a precipitant solution before sealing the plate (Fig. 2.3, panels c-e). Here, the equilibration step proceeds as the precipitant solution soaks into the protein-lipid bolus, slowly increasing the concentration of precipitant and also altering the phase of the lipid.

2.3.3 Vapour diffusion trials

Crystallisation trials, using three protein:reservoir solution ratios of 1:1, 1:2 and 2:1 (200 nL total solution for each), in sitting drop vapour diffusion format were set up in 96 x 3 well crystallisation plates (Swissci, Hampton Research) using a Cartesian Technologies robot[377]. Crystallisation plates were maintained at 6.5°C in a TAP HomeBase storage vault and imaged via a Veeco visualisation system[238].

2.3.4 Addition of small molecules

Small molecules dissolved at high concentration in DMSO were added directly to the purified protein after deglycosylation. Depending on the solubility of the small molecule in DMSO and the desired final concentration, a dilution factor of 20 or 40 was used, resulting in final DMSO concentrations of 5% or 2.5%, respectively. Final concentrations of small molecules and their approximate molar excesses with respect to purified SMO protein at ∼30 mg/mL (∼465 μM) were as follows:

- 10 mM vismodegib, 25x molar excess (Selleck Chem)
- 2.0 mM cyclopamine, 5x molar excess (Sigma-Aldrich)

- 8.35 mM SAG, 18x molar excess (EMD Millipore)
- 6.85 mM 20(S)-OHC, 14x molar excess (Sigma-Aldrich)
- 6.75 mM 20(S)-yne, 14x molar excess (Cambridge Bioscience)

2.3.5 Lipidic cubic phase (LCP) trials

Protein samples were reconstituted into lipidic cubic phase (LCP) by mixing with molten lipid in a mechanical syringe mixer (Fig. 2.3, panels a & b)[37]. Lipid, consisting of 10% cholesterol (Sigma-Aldrich) and 90% 9.9 monoacylglycerol (monoolein; Sigma-Aldrich), was heated to 42°C before mixing with detergent-solubilised protein (either apo or with a small molecule in DMSO) at ∼30 mg/mL in a ratio of 3:2 (w/v). A Gryphon robot (Art Robbins Instruments) was used to dispense 50 nL boluses of protein-laden mesophase followed by 0.8 μL of precipitant solution onto each of 96 positions on a siliconised glass plate and then sealed with a glass coverslip in a "sandwich-plate" format (Fig. 2.3, panels c-e). Crystals were grown at 20°C and monitored by eye, using a microscope fitted with cross-polarisers, and subsequently imaged using a UV-imaging system (Rigaku Minstrel).

2.3.6 Crystallisation conditions

All crystals were produced from the same SMOΔC-BRIL-V329F construct, hereafter referred to as "SMO". Apo-SMO crystallised in 0.1 M MES pH6.0, 30% (v/v) PEG500 DME, 0.1 M sodium acetate, 0.5 mM zinc chloride, 0.1 M ammonium fluoride. Vismodegib-SMO crystallised in 0.09 M sodium acetate pH4.0, 0.09 M sodium malonate, 27% (v/v) PEG500 DME, 0.5 mM zinc chloride, 0.1 M ammonium fluoride. Cyclopamine-SMO crystallised in 0.1 M MES pH6.0, 0.100 M potassium formate or

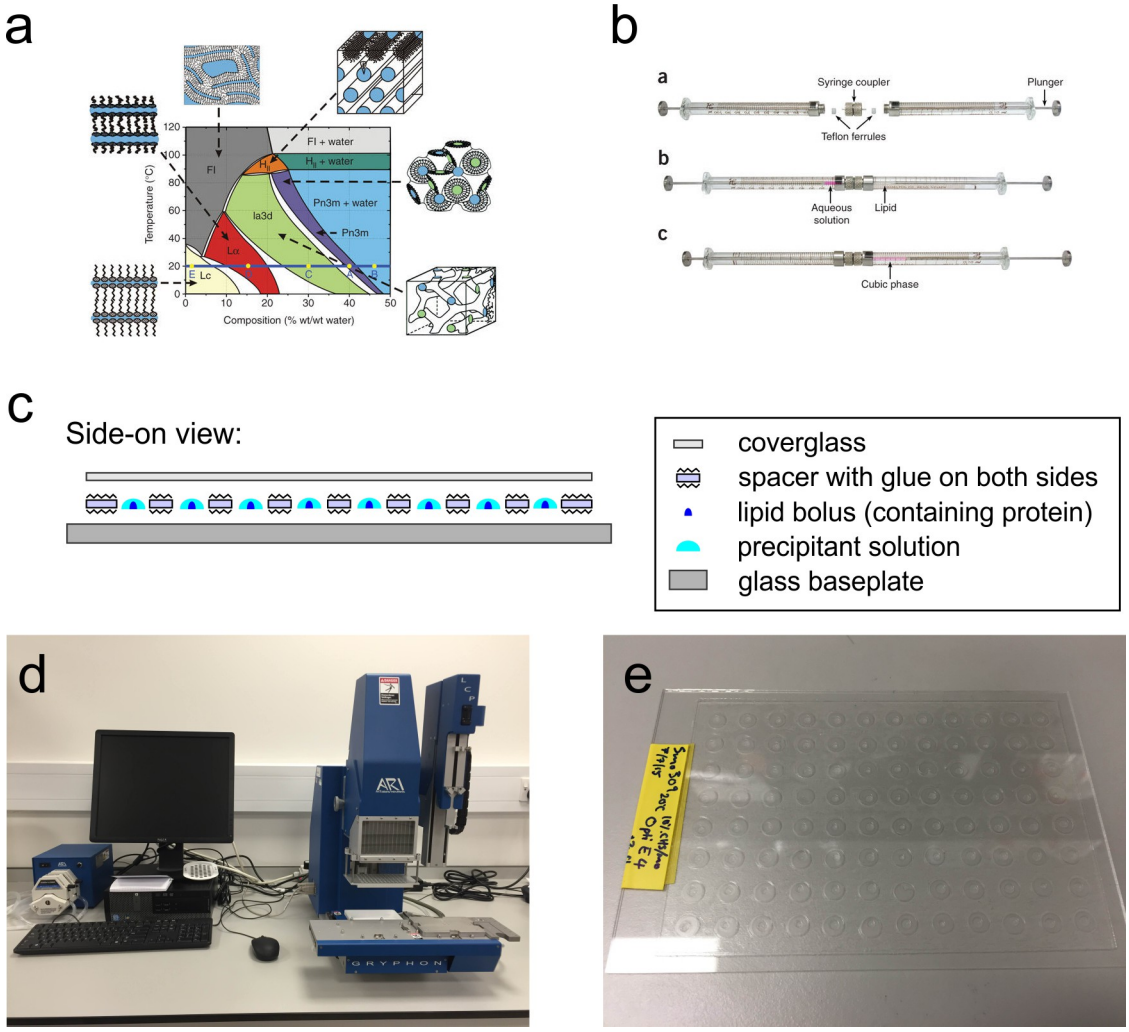


Figure 2.3: Lipidic cubic phase crystallisation. **a,** Temperature-composition phase diagram of the monoolein/water system. Cubic phase corresponds to the purple and blue areas (labelled 'Pn3m'). Changes in temperature or water/lipid ratio result in different phases. Reproduced from [37], figure 1. **b,** Double syringe lipid-mixing set-up. Reproduced from [37], figure 6. **c,** Side-on view schematic of plate set-up. **d,** Gryphon robot (Art Robbins Instruments) for 96-well LCP crystallisation plate set-ups. **e,** Completed LCP plate with 96 conditions.

0.100 M sodium acetate, 24-29% (v/v) PEG500 DME, 0.5 mM zinc chloride, 0.1 M ammonium fluoride. SAG-SMO crystallised in 0.1 M MES pH6.0, 0.09-0.12 M potassium formate, 24-27% (v/v) PEG500 DME, 0.5 mM zinc chloride, 0.1 M ammonium fluoride. 20(S)-OHC-SMO crystallised in 0.1 M MES pH6.0, 0.136 M potassium formate, 23.6% (v/v) PEG500 DME, 0.5 mM zinc chloride, 0.1 M ammonium fluoride. 20(S)-yne-SMO crystallised in 0.1 M MES pH6.0, 0.150 M sodium acetate, 26-29% (v/v) PEG500 DME, 0.5 mM zinc chloride, 0.1 M ammonium fluoride.

2.3.7 Crystal fishing and data collection

In order to fish crystals grown in glass sandwich-plates, wells were first opened by cutting through the coverslip in a circle around the well with a diamond-tipped glass-cutter and removing large pieces of glass using fine-tipped forceps, making sure to avoid disturbing the lipid bolus containing the crystals. Once the coverslip was removed, $\sim$1 μL of precipitant solution was pipetted onto the bolus to prevent it from immediately drying out. Crystals were then scooped out of the bolus using 20-100 μm Dual-Thickness MicroLoops^{TM}LD (MiTeGen) and flash-frozen at 105 K. No cryoprotectant was required as the lipid mixture itself acts as a cryoprotectant.

Fishing LCP crystals in this manner is extremely time-sensitive. Once the coverslip is removed, the well is exposed to air and the equilibrium between lipid and precipitant solution is disturbed. Even ensuring that the lipid bolus remains covered with precipitant solution, the lipid mixture will completely change phase from lipidic cubic to sponge within 2-3 min. The crystals that diffracted best were consistently those harvested first from newly opened wells, suggesting that the molecular ordering of the crystals is almost immediately disrupted by opening the well. As a result, the most promising crystals from each well were harvested as quickly as possible after

opening the well, ideally within 30 s. In addition, crystals that were left in the sealed sandwich-plates for too long a time (for instance, in order to grow more) were both more difficult to fish and more poorly diffracting because the lipid had transitioned to sponge-phase. This transition seemed to occur after 3-4 weeks at room temperature. As the LCP crystals often appeared after 2-3 days, this allowed a period of 2-3 weeks in which to fish them.

Data collection was conducted at the tunable microfocus beamline I24 at the Diamond Light Source (Harwell, UK)(Fig. 2.4). Upon cryo-cooling the LCP mixture becomes opaque, making it impossible to locate the crystals within the loops by eye. Grid-scan software was used to locate the crystals in three dimensions and position the beam accordingly[60]. A typical grid-scan run would have the following parameters: wavelength 0.9686 Å, exposure time 0.100 s, beamsize 9.0 x 6.0 μm, transmission 10-40%, detector resolution 3.5 Å, and take 20-500 images from a single orientation. Once a crystal had been located, a typical data collection run would have the following parameters: wavelength 0.9686 Å, exposure time 0.050 s, beamsize 9.0 x 6.0 μm, transmission 30-50%, detector resolution 3.0 Å, rotation per image 0.1°, number of images 900, to cover 90° total rotation.

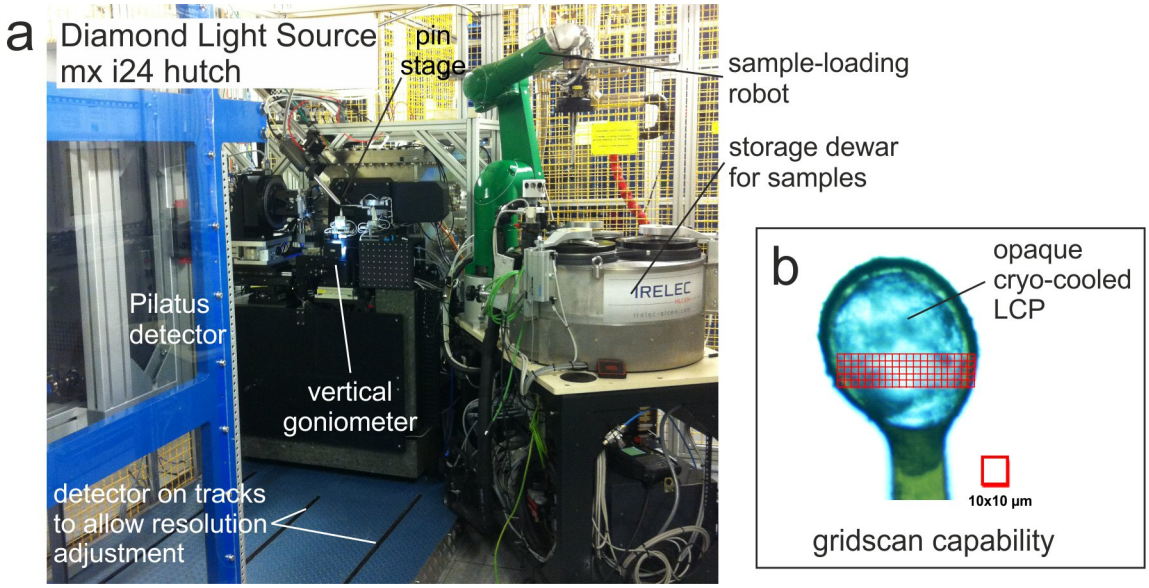


Figure 2.4: Microfocus beamline I24 at the Diamond Light Source. a, Photograph of I24 hutch with various equipment labelled (2015). **b,** Example of micro-loop with LCP mixture ready for grid-scanning (red grid). Photograph taken using the beamline's pin-stage camera.

2.4 Structure solution pipeline

2.4.1 Data processing

Data processing entails the interpretation and conversion of a set of two-dimensional diffraction patterns into a three-dimensional dataset with internally consistent unit cell dimensions and symmetry. The raw data are grey-scale images with many spots (reflections) arranged along straight or curved axes. The parameters that must be determined for each reflection are position and intensity. The positioning of the spots can be used to calculate the unit cell dimensions and the symmetry of the crystal (space group). Once the unit cell and space group of the crystal are known, this information can be used to look for reflections in particular positions (perhaps to push resolution despite low contrast) and also to infer reflections that are missing from the dataset. A full dataset contains all the unique reflections from the crystal, the total number of which will differ depending on the space group. Low symmetry space groups require more data to be collected (i.e. from a greater total angle) than high symmetry space groups do for a full dataset. The intensity of the reflections corresponds to information about the contents of the unit cell. These must be normalised (scaled) to account for radiation damage across the duration of the data collection and multiple observations of the same reflection must be combined and averaged (merged) to produce a final dataset.

X-ray data were generally processed using the automated processing software Xia2[170, 395]. Where data quality was not good enough for automatic processing, spotfinding, indexing and integration were performed using XDS[170], datasets were scaled using XSCALE[170] and then merged using Aimless[112, 394]. Typical individual SMO crystals generated only 20-45° wedges of usable data due to radiation damage. All diffracting crystals of SMO were in low symmetry spacegroups (P1 or

C2). These two points necessitated the merging of data from multiple crystals in order to generate complete datasets. The number of crystals required for final datasets for structure solution and refinement for each SMO structure were as follows: seven for apo, two for vismodegib, seven for cyclopamine, three for SAG, two for 20(S)-OHC, and three for 20(S)-yne. Data collection and processing statistics for all structures solved in this thesis are shown in the Crystallographic Data Table on page 52.

2.4.2 Molecular replacement

Crystallographic diffraction does not contain all of the information about the position of atoms within the unit cell because phase information is lost. This is called the "phase problem". Light waves can be described precisely using three components: wavelength, amplitude and phase. The wavelength describes the smallest distance between repeating features of the waveform (e.g. between subsequent peaks). The amplitude describes the magnitude of displacement of the wave from its direction of travel. The phase describes the translational offset of the waveform from a standard starting point. In x-ray crystallography, a known wavelength is used for data collection and amplitudes can be back-calculated from the intensities of the Bragg peaks. However, relative phases cannot be reproduced from a simple diffraction pattern. It is not possible to generate a model of the electron density without knowing the correct phases. This is particularly an issue for macromolecular crystallography where the number of atoms in the unit cell is very large (1,000s – 100,000s). Various strategies have been devised to get around this phase problem, which generally involve collecting additional datasets from very similar crystals containing a small number of extra atoms that disturb the diffraction pattern (but not the overall structure) to a known degree and then comparing these datasets to find a single solution for the phases. The most common method for determining phases, however, is to use the

previously-determined phases of a similar structure as a starting point. This is called molecular replacement. Once the approximate initial phases have been established, the remainder of the new structural model is then iteratively built and refined whilst progressively releasing the initial phase restraints in order to avoid excessive model bias.

The apo-SMO structure was solved by molecular replacement in PHASER [239] using the structure of human SMO TMD (PDB 4QIM[378]), zebrafish SMO CRD (PDB 4C79[258]) and BRIL (PDB 4EIY[222]) as search models. Two copies of the TMD were found initially within the asymmetric unit, followed by the CRD from one chain and the BRIL domain of the other. In order to unambiguously identify the missing CRD and BRIL domains, the identified TMD and CRD from one chain were together used as a search model for the other chain, for which the BRIL domain had already been identified. Once this full chain was identified, all three of these domains (CRD, TMD, BRIL) were together used as a search model for the first chain. Ultimately, the three domains of both chains within the asymmetric unit were identified with reliable statistics. The difficulty in locating the additional CRD and BRIL domains was reflective of a periodic disorder observed along one axis of the crystal (Fig. 3.14, page 92 and Fig. 4.6, page 135). Subsequent SMO-ligand complex structures were solved by molecular replacement using the apo-SMO structure.

2.4.3 Model building and refinement

Once initial phases have been estimated, they can be applied to the experimental diffraction data to generate an initial real-space electron density map. Two types of maps are generally used at this point: $2F_o - F_c$ and $F_o - F_c$, where F_o represents the observed structure factor amplitudes and F_c represents the calculated structure

factor amplitudes. The observed and calculated structure factor amplitudes are used in combination for the two maps in order to limit model bias. The first map, $2F_o - F_c$, displays a bias-dampened overall electron density whereas the second, $F_o - F_c$, displays positive and negative electron density corresponding to areas where atoms can be added or removed from the model, respectively. An initial map will likely contain only some of the electron density for the molecule in the asymmetric unit and may appear very messy. Density modification (also known as solvent flattening) can be performed in order to increase the contrast between low-signal and high-signal regions of the map. This may help to remove excessive noise from the initial map.

At this stage, more residues can be added to the model (built) either manually or using automated model-building software (such as BUCCANEER[82]). Prior knowledge of the primary amino acid sequence of the crystallised protein is essential for building into low resolution maps. As a greater proportion of the protein molecule is included in the model, the phases can be re-calculated from the model and the maps in turn can be re-calculated from the new phases. The reciprocal relationship between structure factors and real-space coordinates facilitates this iterative process, because every atom in the model affects every structure factor and vice versa. In this way, the model is gradually improved to the point that no more atoms can be added.

For the apo-SMO structure, extra electron density accounting for the regions between the CRD and LD, and between BRIL and the TMD were immediately discernible after density modification in PARROT (Fig. 3.14, page 92)[83]. Surprisingly, we also observed very strong extra density within the CRD ligand binding pocket (Fig. 3.18, page 98) and assigned this to cholesterol bound in a stereo-specific orientation based on shape, coordination and refinement statistics (cholesterol addition improved the R-factors by over 1%), which is also in agreement with the markedly lower B-factor of the refined cholesterol compared to the protein backbone. Cholesterol may be de-

rived from the cells (mM concentrations within the cell membrane) or from the LCP crystallisation mix (that contained 10% (w/v) cholesterol). The apo-SMO polypeptide chain was traced using iterative rounds of BUCCANEER[82], manual building in COOT[111] and refinement in autoBUSTER [28] and PHENIX[3]. This resulted in a well-defined model for the apo-SMO structure that included two molecules of SMO (residues Pro59-Trp549) with a BRIL protein segment inserted between SMO TMD helices V and VI, two N-linked glycans and a cholesterol molecule. Interestingly, we observed a systematic disorder along the c axis (Fig. 3.14, page 92 and Fig. 4.6, page 135), resulting in alternating ordered and less ordered hydrophilic layers within the LCP-grown crystals. This was not caused because of crystal non-isomorphy or pseudo-symmetry, because reducing the space group from C2 to P1 had no effect on the disorder.

For the vismodegib-SMO structure, extra electron density accounting for vismodegib and two well-ordered monoolein molecules was immediately apparent after molecular replacement. For the SAG-SMO structure, extra electron density accounting for SAG, cholesterol, a NAG moiety attached to Asn188 in the connector region and one well-ordered monoolein molecule was apparent after molecular replacement. SAG1.5, from the previous SMO LD-TMD structure (PDB: 4QIN)[378], was originally built into the TMD binding site and the superfluous density from the two extra fluorine atoms of SAG1.5 was immediately apparent in the $F_o - F_c$ map (Fig. 4.12, panel e, page 146), unambiguously confirming the identity of SAG. The structures were refined using autoBUSTER[28] and PHENIX[3] using non crystallographic and secondary structure restraints.

2.4.4 Assessing model quality

A statistical value called R_{work}, which is calculated by comparing the F_o and F_c for each reflection, is used to measure the agreement between the observed data and the calculated model. A lower R_{work} corresponds to closer agreement, with a value of 0.0 representing perfect agreement with the data and near to 0.6 representing a random model[392]. However, with sufficient parameters to overfit the data, it is possible for R_{work} to be arbitrarily low. In order to monitor this, a small set of randomly-selected reflections (the test set, usually $\sim$5%) are kept separate from the majority of the data and used to cross-validate the model. This value is called R_{free}[32]. An appropriate R_{free} value will usually be several percentage points higher than the R_{work} value. If this is not the case, then there is likely something amiss with the model or the data. Various other factors, primarily to do with the expected geometry of protein molecules, are also used to assess the quality of the model. Ramachandran statistics reflect the preferred dihedral angles of atoms across a peptide bond[294]. Based upon previously solved structures, each sidechain has certain preferred rotational orientations (rotamers) that can be optimised in the model. Individual bonds have expected lengths and preferred angles that can also be evaluated with respect to the model.

Pearson correlation coefficient (CC) analysis can also be used to guide the appropriate low resolution cutoff for the data ($CC_{1/2}$) and also to determine the quality of the model with respect to the experimental data (CC*, CC_{work}, CC_{free})[173]. For $CC_{1/2}$, the data are randomly divided into two equal sets and the average intensity of reflections is compared. High correlation ($CC_{1/2}$ near 1.0) occurs at low resolution and drops as resolution is increased, due to weaker signal at higher resolution. $CC_{1/2}$ underestimates the true signal (CC_{true}) in the data because the data are split in two at the beginning of the analysis. CC* is a value calculated from $CC_{1/2}$ in order

to estimate CC_{true}. CC_{work} and CC_{free} are analogous to R_{work} and R_{free}, described above, but measure correlations of the experimental intensities with the calculated intensities derived from the model. CC_{work} and CC_{free} can be directly compared to CC*. For a model that accounts for most of the true signal in the data, the CC_{work} and CC_{free} should be close to but not higher than CC*. A CC_{work} higher than CC* implies that the model is overfit and a CC_{free} lower than CC* implies that the model does not account for all of the signal.

Crystallographic and Ramachandran statistics are given in the Crystallographic Data Table on page 52. Graphs of CC analysis are given in Fig. 3.14 on page 92 and in Fig. 4.5 on page 134. Stereochemical properties were assessed by MOLPROBITY[55]. Superpositions were calculated using the program COOT[111] and electrostatic potentials were generated using APBS[12] and hydrophobicity was calculated according to the Eisenberg hydrophobicity scale[109], as implemented in PyMOL[325]. Buried surface areas of protein-protein interfaces were calculated using the PISA webserver[199] for a probe radius of 1.4 Å. Sequence alignment were performed using MULTALIN[81] and formatted with ESPRIPT[309]. Program Caver was used with default settings to visualize the SMO TMD ligand binding pocket[68].

SMOΔC-BRIL-V329F +	*cholesterol*	*vismodegib*	*cyclopamine*	*SAG*	*20(S)-OHC*	*20(S)-yne*
Data collection						
Beamline	DIAMOND-I24	DIAMOND-I24	DIAMOND-I24	DIAMOND-I24	DIAMOND-I24	DIAMOND-I24
Wavelength (Å)	0.9686	0.9686	0.9686	0.9686	0.9686	0.9686
No. crystals	7	2	7	3	2	3
Resolution (Å)	61.0-3.2 (3.41-3.20)	68.4-3.3 (3.39-3.30)	61.5-3.7 (3.80-3.70)	60.4-3.0 (3.08-3.00)	54.9-3.2 (3.28-3.20)	61.0-3.9 (4.00-3.90)
Space group	C2	P1	C2	C2	C2	C2
Cell dimensions						
a, b, c (Å)	122.9, 63.0, 208.6	52.8, 71.3, 107.0	123.7, 64.4, 208.7	122.9, 63.2, 208.1	122.8, 63.2, 208.5	123.0, 63.6, 207.4
α, β, γ (°)	90.0, 96.6, 90.0	91.5, 98.2, 105.9	90.0, 96.3, 90.0	90.0, 96.3, 90.0	90.0, 96.9, 90.0	90.0, 97.2, 90.0
Unique reflections	25214 (3815)	21097 (2380)	16991 (965)	31218 (2012)	26096 (1926)	14166 (675)
Completeness (%)	93.8 (78.8)	94.7 (85.8)	95.3 (74.4)	97.1 (87.5)	98.5 (97.8)	95.8 (63.2)
R_{merge} (%)	30.5 (>100)	15.2 (>100)	37.9 (>100)	9.7 (79.7)	13.7 (>100)	20.7 (>100)
R_{pim} (%)	13.8 (>100)	9.9 (>100)	28.1 (>100)	8.2 (79.6)	10.5 (>100)	18.2 (>100)
$CC_{1/2}$	0.980 (0.237)	0.953 (0.357)	0.955 (0.089)	0.992 (0.363)	0.992 (0.188)	0.966 (0.230)
CC*	0.995 (0.580)	0.988 (0.460)	0.988 (0.405)	0.998 (0.730)	0.998 (0.563)	0.991 (0.611)
CC_{work}	0.923 (0.354)	0.703 (0.340)	0.903 (0.386)	0.925 (0.511)	0.893 (0.407)	0.883 (0.383)
CC_{free}	0.941 (0.333)	0.941 (0.315)	0.918 (0.182)	0.758 (0.612)	0.852 (0.255)	0.940 (0.332)
I/σI	4.3 (0.6)	3.7 (0.6)	2.3 (0.3)	9.9 (1.0)	6.3 (0.6)	2.8 (0.4)
Multiplicity	5.8 (5.1)	1.8 (1.7)	3.8 (2.3)	3.1 (2.3)	2.5 (2.4)	3.1 (2.3)
Refinement						
Resolution (Å)	61.0-3.2	60.0-3.3	61.5-3.7	60.0-3.0	54.9-3.2	61.0-3.9
No. reflections (test set)	25151 (1302)	20588 (1013)	15748 (799)	31217 (1480)	24826 (1269)	13446 (634)
R_{work}/R_{free} (%)	23.2/26.4	24.5/25.7	28.9/32.5	23.6/25.0	24.2/26.3	25.4/29.4
No. atoms						
Protein	9207	8988	9207	9245	9281	9245
Ligand/Ion	58	119	62	140	30	-
B factors (Å^2)						
Protein	137	124	153	128	140	190
Ligand/Ion	121	119	183	108	90	-
r.m.s.d. bonds (Å)	0.008	0.008	0.015	0.007	0.011	0.018
r.m.s.d. angles (°)	0.90	0.87	1.42	0.86	1.44	1.39
Ramachandran (%)						
Favoured	96.6	96.8	97.3	97.1	96.6	97.2
Allowed	3.2	2.9	2.3	2.6	2.9	2.2
Disallowed	0.2	0.3	0.4	0.3	0.5	0.6

Figure 2.5: Crystallographic Data Table

2.5 Small-angle X-ray scattering (SAXS)

2.5.1 Basic principles of SAXS

Small-angle X-ray scattering (SAXS) is an important complementary technique in structural studies. It can provide information about protein conformation and flexibility in the relatively low-resolution range of 10-50 Å, which for this study, can help identify flexible active states of multi-domain SMO in solution.

Protein molecules that are randomly oriented in solution scatter X-rays to produce a diffuse, radially symmetric scattering pattern. The data in this pattern are generally represented in a plot of $\log I(q)$ versus q, where q is a measure of scattering angle and $I(q)$ is a measure of intensity. Formally, q is "momentum transfer" representing the directional momentum change undergone by the photons. It has units of Å^{-1} and takes the form:

$$q = \frac{4\pi \sin\theta}{\lambda}$$

where 2θ is the scattering angle and λ is the wavelength of the incident X-rays (in Å). The scattering intensity of solute particles, $I(q)$, in a sample is obtained by subtracting solvent scattering (obtained from a buffer-only control) from sample scattering. It is proportional to q. In a $\log I(q)/q$ plot, lower values of q ("low q") provide low resolution information and higher values of q ("high q") provide higher resolution information.

A Fourier transform of the scattering curve $I(q)$, known as the pair-distribution function, $P(r)$, describes the paired-set of all distances between two points within an

object in real-space[292]:

$$P(r) = \frac{r}{2\pi^2} \int_0^\infty I(q) q \sin(qr) dq$$

where r is the distance separating the two points (in Å). The maximum linear dimension of the particle, D_{max}, is the position where $P(r)$ returns to zero at high values of r. If a conformational change alters the longest dimension of the protein then D_{max} will reflect this. More broadly, if a conformational change shifts the relative distances between just a few residues, then this will likely be detectable in the $P(r)$ distribution.

Kratky analysis can be used to quantify the folded- or unfolded-ness of a protein. A Kratky plot is simply $q^2 \cdot I(q)$ versus q. Globular macromolecules will display a bell-shaped curve, whereas unfolded proteins lack this peak and instead show a high plateau. However, the Kratky plot is sensitive to molecular mass and conformational state of the macromolecule. In order to effectively compare scattering profiles on a common scale a Dimensionless Kratky plot is preferred, where the original Kratky plot is normalised for protein concentration, particle volume and radius of gyration (R_g). The plot then becomes $I(q)/I(0) \cdot (qR_g)^2$ versus $q \cdot R_g$. Changes in the Dimensionless Kratky plot reflect the different degrees of flexibility of the particles between each of the samples[295].

2.5.2 SAXS data collection and processing

For size exclusion chromatography coupled small angle X-ray scattering (SEC-SAXS), amphipol-exchanged untreated and 20(S)-OHC-treated (∼5 mM 20(S)-OHC) SMOΔC-BRIL-V329F were loaded onto a 4.8 mL KW-403 column (Shodex), equilibrated in

a no-detergent buffer, on an Agilent 1260 system (B21, Diamond Light Source). Approximately 45 μL of sample at a concentration of 9.6 mg/mL (20(S)-OHC-treated) or 13 mg/mL (apo) was injected at a flow rate of 160 μL per min. Chromatographic elution was directed into a specialised SAXS flow cell, with a 1.6 mm path length, held at 20°C. SAXS measurements were made using a sample-to-detector distance of 4.09 m and X-ray wavelength of 1 Å. SAXS images (frames) were collected as a continuous set of 3-second exposures across the elution peak. The corresponding buffer background measurement for producing the background-subtracted curve was collected at a retention time of greater than 1.5 column volumes.

Frames were corrected for variations in beam current, normalised for exposure time and processed into 1D scattering curves using in-house beamline software (GDA, Diamond Light Source). Buffer subtractions and all other subsequent analysis were performed with the program ScÅtter (http://www.bioisis.net/scatter). Samples were checked for radiation damage by visual inspection of the Guinier region as a function of exposure time. Considerable differences between SAXS curves (at $q > 0.05$ Å $^{-1}$) imply major structural differences between the two states. Large differences observed between the two P(r)-distributions imply a significant structural change in the 20(S)-OHC state.

2.6 Multi-angle light scattering (MALS)

Multi-angle light scattering (MALS) measures the scattering of laser light by a solution of macromolecules across a broad set of angles. The molar mass of a particle is determined from the intensity of reflected light, the known concentration of the particle in a homogeneous solution and the refractive index of the solvent. The intensity of reflected light is proportional to the magnitude of the electromagnetic dipole in-

duced within the macromolecule by the incident light. The larger the macromolecule, the larger the induced dipole and the greater the intensity of reflected light. When connected inline with size exclusion chromatography (SEC) the molecular weights of individual protein peaks can be determined, giving an indication of the oligomeric states or complex species present in solution[117].

In order to avoid background light-scattering due to free detergent in solution, protein samples were exchanged into amphipol (A8-35, Anatrace)[360] at a mass ratio of 3:1 amphipol:protein and rotated at room temperature for 30 min. BioBeads (BIORAD), equilibrated in no-detergent buffer (10 mM HEPES pH7.5, 300 mM NaCl), were added to the protein-detergent-amphipol mixture at 10 mg per 100 μL and incubated overnight at 4°C to remove all detergent molecules. Amphipol-solubilised protein at $\sim$1 mg/mL was then loaded onto a Superose 6 10/300 column (GE Healthcare), equilibrated in no-detergent buffer, on a Shimadzu system with inline MALS detector (Wyatt Technologies). Data were analysed using ASTRA6.1.2 software (Wyatt Technologies). For SMOΔC-BRIL-V329F, the value used for $\frac{\partial n}{\partial c}$ and ε at 280 nm were 0.185 mL/g and 1.541 mL/(mg.cm) respectively. For protein conjugate analysis, $\frac{\partial n}{\partial c}$ used for amphipol A8-35 was 0.150 mL/g[133].

2.7 Generation of nanobodies against Smoothened

2.7.1 Protein for llama immunisation and screening of nanobodies

For initial immunisation of the llama, a combination of the SMOΔC-BRIL-V329F construct and a close variant with a C-terminal Avi tag (GLNDIFEAQKIEWHE) for *in vivo* or *in vitro* biotinylation[113, 85] were expressed and purified as outlined previously (section 2.2.2, page 32). These two constructs contained the extracellular CRD and the intracellular BRIL fusion as solvent-accessible domains. Use of constructs containing the BRIL fusion to immunise the llama was necessitated by the amount of protein required; constructs with BRIL express substantially better than those without. For the screening of nanobodies, a variant of the SMOΔC-BRIL-V329F construct with a C-terminal Avi tag was expressed and purified as outlined previously (section 2.2.2, page 32). This construct contained the extracellular CRD but not the intracellular BRIL fusion. Any nanobodies that had been raised against the BRIL domain derived from the immunisation constructs would therefore not be selected in the screening round. Llama immunisation and initial nanobody screening by phage surface display[322] were carried out in collaboration with Prof. Jan Steyaert, Dr Els Pardon and Eva Beke (Structural Biology Research Center, Vrije Universiteit Brussel, Belgium). They provided us with nine plasmids containing nanobodies that had showed selectivity for SMO binding in their assays.

2.7.2 Periplasmic expression and purification of nanobodies in E. coli

Nanobodies in the pMESy4 vector were expressed in *E. coli* strain WK6*su-* and purified by immobilised metal-affinity chromatography (IMAC)[274]. Bacteria were started in a 14 mL culture of LB supplemented with 100 μg/mL ampicillin, 2.0% glucose and 1 mM $MgCl_2$ and incubated overnight at 37°C, 180 rpm. The culture was then expanded to 2 L of Terrific Broth (TB) with 0.4% glycerol, 0.09% glucose, 2 mM $MgCl_2$, 100 μg/mL ampicillin and incubated at 37°C, 180 rpm to an optical density at 600 nm (OD_{600}) of 0.7-0.8. Expression of nanobodies was induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and bacteria were grown overnight at 22°C, 180 rpm. To harvest, the bacteria were centrifuged at 5000 xg at 15°C for 15 min. The pellet from 1 L culture was resuspended in 23 mL ice-cold TES buffer (0.2 M Tris pH8.0, 0.5 M EDTA, 0.5 M sucrose) and incubated on ice for 4 h at 120 rpm. A total of 40 mL 4-fold diluted TES buffer was added and the samples were incubated on ice overnight at 120 rpm. The samples were centrifuged at 8000 rpm for 30 min at 4°C to remove insoluble material. Nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen) was applied to the supernatant containing nanobodies and gently rotated for 1 h at 4°C. Beads were collected and washed with 50 mM HEPES pH7.5, 500 mM NaCl followed by 50 mM HEPES pH7.5, 500 mM NaCl, 10 mM imidazole and eluted with 20 mM HEPES pH7.5, 150 mM NaCl, 250 mM imidazole. Bead eluate was then loaded onto a Superdex 75 16/60 column (GE Healthcare) equilibrated with 10 mM HEPES pH7.5, 150 mM NaCl. Peak fractions were pooled and concentrated using a 10 kDa MWCO Millipore concentrator. Nanobody purifications were evaluated by reducing SDS-PAGE. Nanobodies were snap-frozen in liquid N_2 and stored at -80°C.

2.8 Binding studies of nanobodies

2.8.1 Bio-layer interferometry (BLI)

Bio-layer interferometry (BLI) is a proprietary technique developed by ForteBio that is an optical biosensor technology similar to surface plasmon resonance (SPR) used to measure binding affinity between two populations of biomolecules[303]. Briefly, white light is transmitted down an optical fibre to a proprietary optical coating with biomolecules attached. The light is partially reflected back through the fibre, producing an interference pattern across the visible spectrum. The thickness of the biological layer affects the spectral shifting of the interference pattern and this can be monitored in real-time[76]. In much the same manner as for SPR, one protein, called the 'ligand', is immobilised onto the surface of the sensing device (in this case the tip of an optical fibre) and a potential binding partner, called the 'analyte', is applied to the immobilised ligand. Unlike SPR, BLI does not require measurement of dielectric constant or refractive index of the solution and as such is claimed to be less sensitive to changes in the solvent. In practice, variations in buffer conditions were detrimental to data quality and comparability.

The binding affinities of purified nanobodies (analytes) for purified SMO (ligand) were assayed as follows. SMO with a C-terminal Avi-tag was biotinylated *in vivo* by co-transfection of BirA biotin ligase[85, 187] and then purified by 1D4-antibody affinity and SEC (see page 32). SMO at ∼0.1 mg/mL in 10 mM HEPES pH7.5, 150 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS, 0.0005% Tween-20 was immobilised onto Dip and Read^{TM}Streptavidin (SA) Biosensors (60 s baseline, 300 s association, 60 s baseline, all at room temperature) using the Octet® RED384 System (Pall ForteBio). As an initial screen, all eight nanobodies raised against SMO (1.0 μM purified nanobody in 10 mM HEPES pH7.5, 150 mM NaCl, 10%

glycerol, 0.03% DDM, 0.006% CHS, 0.0005% Tween-20) were applied to the SMO-conjugated biosensors (60 s baseline, 180 s association, 300 s dissociation). Using fresh biosensors for each experiment, an eight-sample dilution series of each nanobody was run (60 s baseline, 300 s association, 300 s dissociation). The concentration range in the dilution series of each nanobody was estimated based upon the initial screening results. Dissociation constants (K_D) for each binding interaction were determined using the Octet® Analysis software. In order to assess whether the detergent present in buffers had an effect on the binding affinities, SMO was exchanged into amphipols (as for MALS analysis, section 2.6, page 55) and then the BLI procedure was repeated as above but with no-detergent buffers. Results were broadly comparable between detergent and amphipols for all nanobodies tested in both.

2.8.2 Fluorescence microscopy

Fixed-cell fluorescence microscopy was used to detect binding of purified nanobodies to the cell surface of mammalian COS-7 cells expressing SMO. Flag-tagged nanobodies in the pET-22b(+) vector were expressed in Rosetta™(DE3) pLysS competent cells (Novagen ® #70956) and purified by IMAC. Protocol was the same as for nanobodies without the Flag-tag except all media was supplemented with 25 μg/mL chloramphenicol, in addition to ampicillin, glucose and $MgCl_2$. Size exclusion chromatography was not used in this case. To remove excess imidazole, resin eluate was instead dialysed into PBS for 48 h at 4°C. Nanobodies were snap-frozen in liquid N_2 and stored at -80°C.

One day prior to transfection, COS-7 cells (ATCC® CRL-1651™) were plated at 0.25 x 10^6 cells per well of a 6-well plate and incubated overnight at 37°C in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) supplemented with 10%

FCS (Sigma-Aldrich), L-glut (Gibco) and NEAA (Gibco). Cells were transiently transfected using Lipofectamine® 2000 (Thermo Fisher Scientific) in serum-free media (SFM) and then incubated for 4 h at 37°C. Transfected cells were trypsinised, resuspended in complete medium (10% FCS) and aliquoted into a 24-well plate in which each well contained a sterilised coverslip. After 48 h at 37°C, cells were incubated with 5 μM Flag-tagged nanobody in complete medium for 2 h at 37°C. Cells were then washed with phosphate-buffered saline (PBS), fixed with 4% paraformaldehyde (PFA) in PBS and washed again with PBS. Coverslips were blocked with 10% FCS/PBS for 1 h at room temperature before application of the primary antibody: a 1:1000 dilution of rabbit anti-Flag antibody in 0.1% FCS/PBS for 1 h at room temperature. After three brief washes with PBS, the secondary antibody was applied: a 1:1000 dilution of anti-rabbit Alexa Fluor® 488 antibody (Sigma-Aldrich) in 0.1% FCS/PBS, and incubated for 1 h at room temperature in the dark. Coverslips were briefly washed with PBS again and then treated with Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories) before sealing to a microscope slide. All slides were stored at 4°C. Cells were visualised on a wide-field inverted Nikon TE2000U fluorescence microscope (WTCHG Microscopy Core) using NIS-Elements software (Nikon). Cell nuclei were identified by DAPI staining ("mVenus" filter) and nanobody localisation by Alexa Fluor® 488 staining ("GFP" filter).

Chapter 3

SMO is regulated by its extracellular domains

3.1 Abstract

Developmental signals of the Hedgehog (Hh) and Wnt families are transduced across the membrane by Frizzled-class G-protein coupled receptors (GPCRs) composed of both a heptahelical transmembrane domain (TMD) and an extracellular cysteine-rich domain (CRD). How such large extracellular domains of GPCRs regulate signalling by the TMD is unknown. In this chapter, I present the crystal structure of the Hh transducer Smoothened (SMO) including both its TMD and its CRD. The CRD is stacked atop the TMD in the structure, separated by the intervening wedge-like linker domain (LD). Unexpectedly, a cholesterol molecule was bound to the SMO extracellular domain at the interface between the CRD, LD and TMD. Cholesterol binding appears to be important, both for the stability and signalling activity of SMO. Mutations that destabilise the interface between the extracellular and transmembrane domains increase the constitutive signalling activity of SMO. These data suggest that SMO signalling activity is regulated by interactions between its extracellular and transmembrane domains.

Much of the work presented in this chapter has been published by us in two papers:

E. F. X. Byrne, R. Sircar, P. S. Miller, G. Hedger, G. Luchetti, S. Nachtergaele, M. D. Tully, L. Mydock-McGrane, D. F. Covey, R. P. Rambo, M. S. P. Sansom, S. Newstead, R. Rohatgi, and C. Siebold. Structural basis of smoothened regulation by its extracellular domains. *Nature*, 535(7613):517-522, 2016. (ref [36])

G. Luchetti, R. Sircar, J. H. Kong, S. Nachtergaele, A. Sagner, E. F. Byrne, D. F. Covey, C. Siebold, and R. Rohatgi. Cholesterol activates the g-protein coupled receptor smoothened to promote hedgehog signaling. *Elife*, 5, 2016. (ref [226])

Bead-binding and pathway activation experiments reported in this chapter were performed in collaboration with: Ria Sircar, Giovanni Luchetti, Sigrid Nachtergaele and Rajat Rohatgi ('Rohatgi Lab'). Cholesterol- and oxysterol-conjugated beads used in this chapter were produced in collaboration with: Laurel Mydock-McGrane and Douglas Covey ('Covey Lab'). Molecular dynamics simulations reported in this chapter were performed in collaboration with: George Hedger and Mark Sansom ('Sansom Lab'). Specific contributions are noted where relevant. All of these experiments were conducted as a direct result of the crystal structure of SMO CRD-LD-TMD.

3.2 Introduction

3.2.1 Structure of the Smoothened TMD

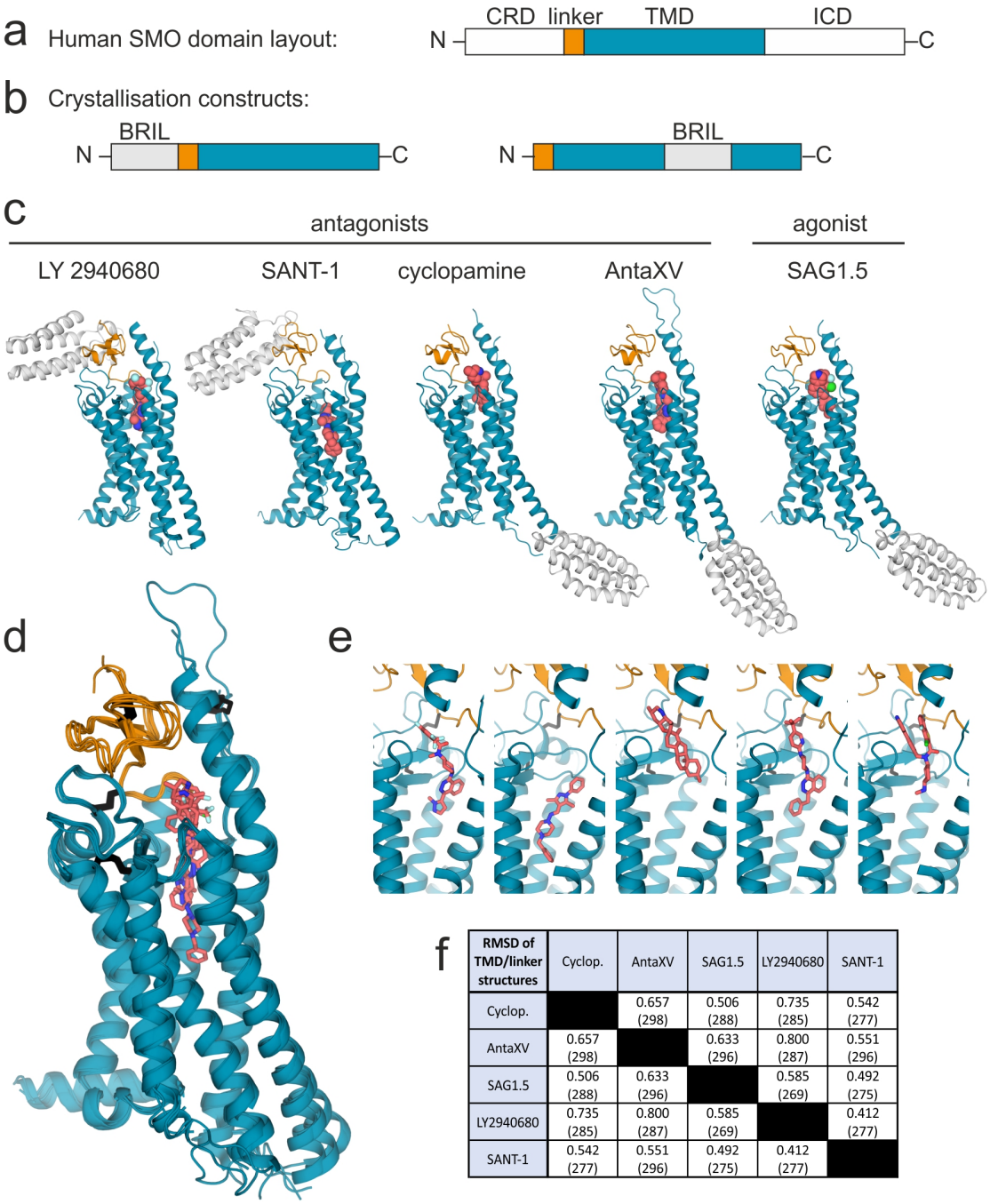
The first crystal structures of the isolated SMO LD-TMD revealed conservation of the canonical GPCR heptahelical scaffold (Fig. 3.1)[379, 378, 388]. The two constructs used for crystallisation had been truncated to remove both the cysteine-rich domain (CRD, residues Met1-Ser189) and the intracellular domain (ICD, residues Gln555-Phe787), leaving just the LD and TMD. Both constructs also contained a thermostabilised apocytochrome b_{562}RIL helical bundle (BRIL) fused either at the N-terminus, replacing the CRD, or in between TM helices V and VI, replacing intracellular loop 3 (ICL3). BRIL fusions with GPCRs had previously been shown to preserve the ligand-binding property of the receptors (for instance, binding of agonist UK432,097 or antagonist ZM241385 to A_{2A}AR) whilst aiding expression levels, improving stability and providing crystal contacts[72, 71]. All the structures were crystallised using the lipidic cubic phase (LCP) method[37].

SMO LD-TMD structures were solved in complex with five different synthetic small molecule ligands. Four were antagonists, LY2940680, SANT1, Anta XV, cyclopamine, and one an agonist, SAG1.5 (refs. [379, 378, 388]). These structures provided a detailed view of the small molecule binding pocket at the extracellular side of the SMO TMD (Fig. 3.1). This pocket is a long and narrow cavity formed mainly by the 7TM helical bundle, with minor contributions from the extracellular LD and the extracellular loops (ECLs). Four of the ligands, LY2940680, Anta XV, cyclopamine and SAG1.5, all bind close to the extracellular surface forming most of their contacts with residues in the LD and ECLs[379, 378, 388]. SANT1, however, binds much deeper within the cavity, forming most of its contacts with residues from the TM helices[378]. The LD and long extracellular loops are stabilised by four disul-

fide bonds, which are essential for proper regulation and subcellular localisation of SMO[41].

Unfortunately, despite the range of small molecules and their various binding modes, these SMO structures did not show the conformational changes typically associated with GPCR signalling (Fig. 3.2)[97, 174]. Comparisons of inactive and active states of various GPCRs show several common features including a substantial outward "swinging" movement of the intracellular tip of TM helix VI in concert with a smaller inward movement of the intracellular tip of helix V. The extent of the movement depends on the receptor involved and its particular active state. For example, in the A_{2a} adenosine receptor (A_{2a}AR) helix VI moves ∼3 Å[404], while its movement in opsin is ∼6 Å[275] and in the β_2 adrenergic receptor (β_2AR) ∼8 Å[298]. In the case of the SMO LD-TMD structure, however, compared to the antagonist-bound structures the agonist-bound SAG1.5 structure showed only minor reorientations of several conserved residues, namely a broken hydrogen bond between Tyr262 and His361 and a shift of Pro263 towards Trp535 (ref. [378]), but no major conformational change as would be expected by an "activated" GPCR conformation.

Although the SMO TMD physically spans the plasma membrane and contains the binding site for many clinically relevant compounds, SMO cannot respond fully to native Hh ligands without the CRD[1]. Several studies have shown an essential role for the CRD in trafficking to the cell membrane (for *Drosophila*) or primary cilium (for vertebrates)[261, 1]. In addition, no endogenous small molecules are known that engage the TMD ligand binding site. While the current structures of the LD-TMD may enhance our molecular understanding of synthetic small molecule recognition by the TMD, they have not revealed the mechanism of action of SMO signal transduction.



f

RMSD of TMD/linker structures	Cyclop.	AntaXV	SAG1.5	LY2940680	SANT-1
Cyclop.		0.657 (298)	0.506 (288)	0.735 (285)	0.542 (277)
AntaXV	0.657 (298)		0.633 (296)	0.800 (287)	0.551 (296)
SAG1.5	0.506 (288)	0.633 (296)		0.585 (269)	0.492 (275)
LY2940680	0.735 (285)	0.800 (287)	0.585 (269)		0.412 (277)
SANT-1	0.542 (277)	0.551 (296)	0.492 (275)	0.412 (277)	

Figure 3.1: Structures of SMO LD-TMD with synthetic small molecule ligands. a & b, Domain schematics of full-length human SMO and two constructs used for crystallisation (orange, LD; blue, TMD; grey, BRIL fusion). **c,** Crystal structures of SMO LD-TMD with five different ligands in cartoon representation from the same view (colours follow domain schematics; red spheres, ligands). **d,** Five structures aligned, with small molecules shown as red sticks and disulfide bonds as black sticks. **e,** Binding site for each ligand from the same view. Order of ligands as in panel c. Part of TM helix VI has been removed for ease of viewing. **f,** RMSD values for pairs of aligned structures with number of aligned Cα atoms in parentheses. (PDB: 4O9R, 4QIM, 4QIN, 4JKV, 4N4W)

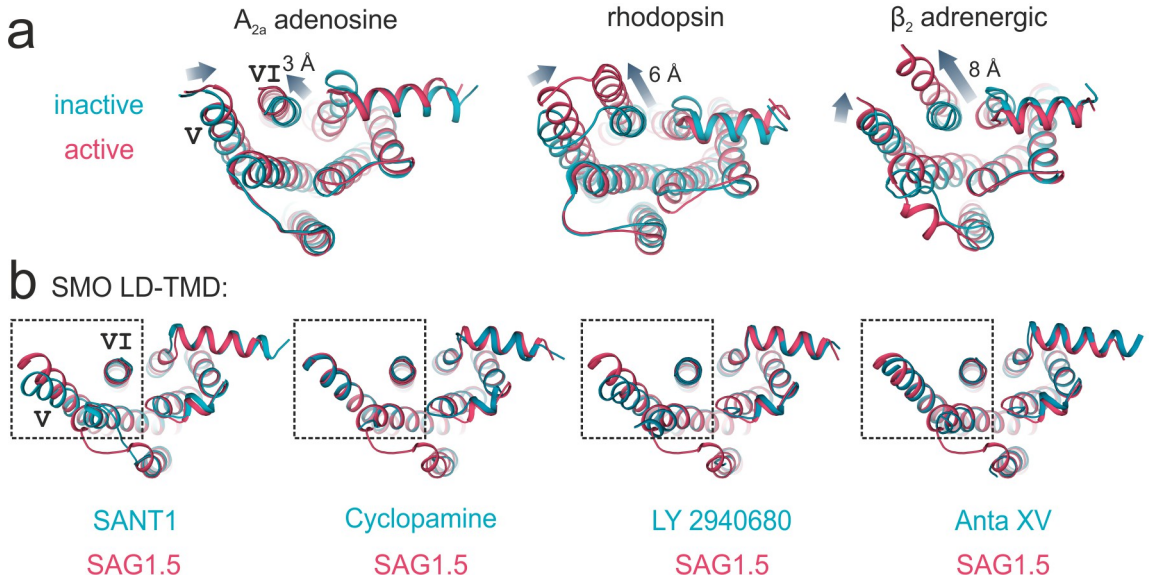


Figure 3.2: Conformational changes of GPCRs associated with activation. a, Intracellular face of A_{2a}AR (PDB: 3QAK, 3EML), rhodopsin (PDB: 3PQR, 1GZM) and β_2AR (PDB: 3P0G, 2RH1) showing active (red) and inactive (blue) states overlaid. Roman numerals indicate TMD helices. Arrows highlight helix movement between states. Adapted from [97]. **b,** SMO from the same view with the agonist SAG1.5 structure (red) overlaid with four antagonist structures (blue). Dotted boxes highlight TMD helices V and VI. Shifts observed in helix V are most likely due to presence/absence of BRIL fused after helix V (see Fig. 3.1, panel c). (PDB: 4O9R, 4QIM, 4QIN, 4JKV, 4N4W)

3.2.2 Structure of the Smoothened CRD

SMO is most closely related to the Frizzled (Fzd) receptors of the Wnt signalling pathway. The Fzd receptors all contain a CRD and it is essential in the binding of Wnt ligands and in receptor dimerisation, which may be required for signal transduction[165, 86, 42]. The SMO CRD has been shown to be indispensable for Hh signalling in both *Drosophila* and vertebrates; SMO lacking the CRD fails to translocate to the plasma membrane (in *Drosophila*) or the primary cilium (in vertebrates) in order to induce maximum signalling[1, 261, 79]. However, the precise role of the SMO CRD is not clear.

Two unliganded structures of the isolated SMO-CRD have been solved, one

by crystallography[258] and the other by NMR spectroscopy[296] (Fig. 3.3). The crystal structure of the zebrafish SMO-CRD (residues Pro41-Glu158) revealed a predominantly α-helical architecture held together by five disulfide bridges common to classical CRD domains[279]. Although this structure did not contain a ligand, the putative oxysterol binding site was identified by a combination of electrostatic and hydrophobic analysis of the solvent-exposed surface, a comparative structural phylogenetic analysis and site-directed mutagenesis paired with a functional assay[263, 256]. The NMR solution structure of the *Drosophila* SMO-CRD (residues Val85-Lys202) revealed a very similar overall structure (2.043 Å RMSD for 78 Cα atoms) including four of the five disulfide bridges. This study also identified a binding site for the small-molecule glucocorticoid, budesonide, that corresponds to the oxysterol binding site suggested by the crystal structure. However, despite the structures of the isolated LD-TMD and CRD, we presently lack structural insights into how the SMO extracellular domains interact with the TMD in order to regulate signalling. More broadly, no structure of a GPCR with a large, functional extracellular domain has yet been solved.

3.2.3 Cholesterol in GPCRs and Hh signalling

Cholesterol performs many roles within the body. It is an essential constituent of healthy animal cell plasma membranes, accounting for 20-25% of the lipid molecules and allowing modulation of membrane fluidity and permeability[154]. It is also the primary sterol synthesised by all vertebrates and acts as a biochemical precursor to many vital steroid hormones and bile acids. Cholesterol is associated with a range of human diseases[237] and has been implicated in morphogen signalling pathways [15, 330].

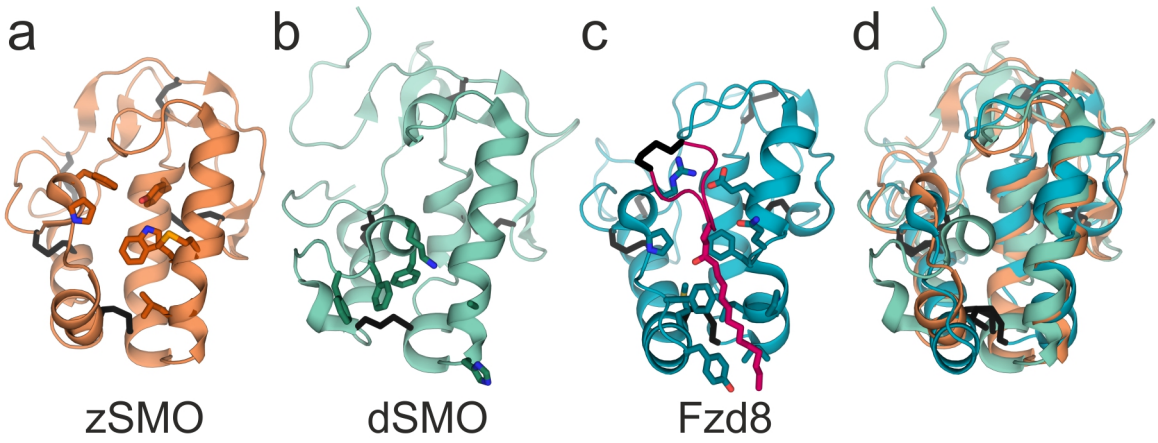


Figure 3.3: Structures of CRDs from SMO and Fzd. a, Crystal structure of zebrafish SMO CRD (orange cartoon, PDB: 4C7A) with residues of predicted oxysterol binding site shown as sticks with atomic colouring. **b,** NMR structure of *Drosophila* SMO CRD (green cartoon, PDB: 2MAH) with residues of predicted budesonide binding site as sticks with atomic colouring. **c,** Crystal structure of mouse Fzd8 CRD (blue cartoon, PDB: 2F0A) with palmitoleic acid lipid moiety from Xenopus Wnt8 (pink sticks) and residues of its binding site as sticks with atomic colouring. **d,** Structural alignment of three CRD structures with colours as in panels a-c. Disulfides shown as black sticks in each panel.

Broadly, cholesterol molecules have been shown to form important intramembrane contacts with many integral membrane proteins, including channels[214] and GPCRs. Several structures of β_2AR have been solved with cholesterol either mediating important intramembrane crystal contacts or bound specifically to a conserved motif on the outside of the 7TM bundle (Fig. 3.4, panel d)[61, 141]. Cholesterol and its derivatives, like cholesteryl hemisuccinate (CHS), can modulate membrane fluidity and curvature and in general have a stabilising effect on receptors both in plasma membranes and in detergent micelles[130, 34, 19, 184, 100]. Cell membrane cholesterol levels can be exogenously altered using cyclodextrin compounds. Cyclodextrin molecules are oligosaccharide rings with a hydrophobic inner surface and a hydrophilic outer surface, allowing inclusion of hydrophobic molecules, such as cholesterol, in aqueous solvents[203]. One such cyclodextrin, methyl-β-cyclodextrin (MβCD), has been used to deplete cell membranes of cholesterol in live-cell culture [406, 185, 414, 191]. MβCD that has been pre-saturated with cholesterol has also been

used to enrich membranes with cholesterol[191, 70]. In addition, cyclodextrin can be used to increase the shuttling of cholesterol between different lipid compartments[69].

Within the Hh signalling pathway, cholesterol and closely related lipids perform a range of important functions[106]. The extracellular Hh ligand must be post-translationally modified with both cholesterol[290] and palmitoyl[280] in order to be fully active. This lipid-modified Hh ligand is spread through tissue by cholesterol-laden lipoprotein particles[273]. The Hh-ligand receptor and essential membrane protein Patched1 (Ptc1) has been implicated in cholesterol transport[22] and its repression of SMO depends on conserved residues that it shares with proteins like NPC1, a cholesterol efflux pump[350]. Homology modelling has also linked the N-terminal domain of NPC1, the structure of which was solved bound to a cholesterol molecule[205], to the CRDs of both SMO and the Fzd receptors[15, 86, 165], suggesting a sterol recognition pocket for SMO (Fig. 3.4, panels a-c).

SMO activity in mammals can be regulated by exogenous small lipophilic molecules, such as cyclopamine[52] and various oxysterols[105, 80] that are structurally related to cholesterol. Crucially, however, no endogenous and physiologically relevant small molecule ligand for SMO has yet been determined. The endogenous sterol derivative, 7-dehydrocholesterol, has also been reported to affect SMO signalling activity[24] but it does not promote trafficking of SMO to the primary cilia[393], making it unlikely to be a physiological ligand. Diseases of distal cholesterol biosynthesis cause developmental disorders, for example Smith-Lemli-Opitz syndrome (SLOS)[78, 179, 288], and compounds that block cholesterol transport are weak Shh antagonists[156]. Additionally, SMO activation has been shown to be dependent on the levels of an important phospholipid, phosphatidylinositol 4-phosphate (PI4P), within specific membrane compartments[408]. These various observations have led to the view that cholesterol is permissive for Hh signalling. However, the specific roles

of cholesterol with respect to SMO are not entirely clear.

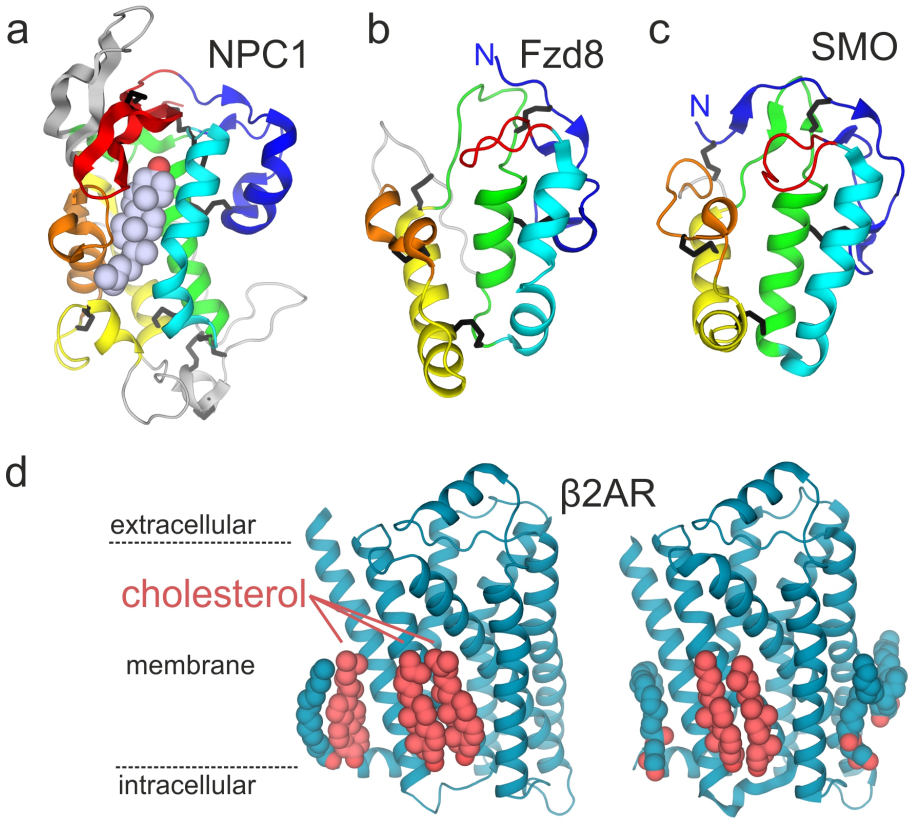


Figure 3.4: Various roles for cholesterol. a, Cholesterol molecule (light purple spheres) bound to the CRD of NPC1 (PDB: 3GKI). **b & c,** CRDs of mouse Fzd8 (PDB: 1IJY) and zebrafish SMO (PDB: 4C79) show structural homology to NPC1. Helices are coloured to demonstrate structural homology. **d,** Structures of $\beta 2$ adrenergic receptor (blue cartoon) with bound lipid molecules (blue spheres), including cholesterol (red spheres) (PDB: 2RH1, 3D4S). Panels a, b, c adapted from [15].

3.2.4 Chapter outline

The aim of this chapter was to characterise the interactions between the CRD, LD and TMD by solving a crystal structure of SMO containing all three domains (Fig. 3.5). Expression constructs were designed and purification strategies for reconstitution in detergent micelles were optimised. The detergent DDM was found to best solubilise and thermostabilise SMO in solution. A single inactivating point mutation,

Val329Phe, was found to substantially and robustly boost SMO expression levels. Crystals were grown in LCP after increasing the protein concentration to 25 mg/mL and extensive optimisation of the crystallisation conditions. By merging data from multiple micro-crystals the structure of SMO-CRD-LD-TMD was solved by molecular replacement to 3.2 Å resolution. The structure revealed an upright arrangement of the three domains, with the CRD stacked atop the LD which separated the CRD from the TMD. The long ECL3 extended away from the TMD to form contacts with both the LD and CRD. A cholesterol molecule occupied the CRD binding site and coordinated interactions between the three domains. Subsequent mutagenesis experiments and MD simulations supported the hypothesis that cholesterol performed a key structural role and that several key residues involved in cholesterol binding were essential for native signalling.

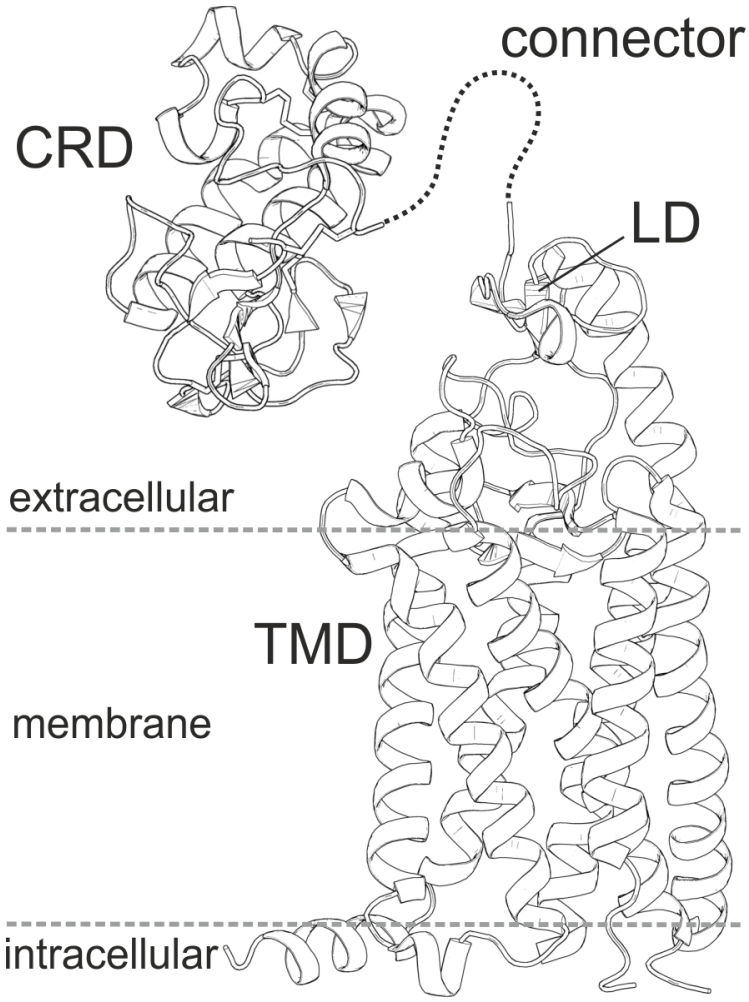


Figure 3.5: Model of SMO CRD-LD-TMD structure. (PDB: 4C79, 4O9R) Figure adapted from [378].

3.3 Results

3.3.1 BRIL fusion increases protein expression

Domain boundaries for constructs were based upon the previously published crystal structures (PDB 4C79, 4JKV), removing the entire unstructured intracellular domain (ICD) (Fig. 3.6). As a benchmark for expression and purification procedures, the first construct, BRIL-ΔCRD-SMOΔC, replicated the construct used to solve the first SMO isolated LD-TMD crystal structure[379], with the CRD replaced by BRIL (Fig. 2.1, page 26). The second construct, SMOΔC, included the CRD, LD and TMD, with no extra fusions. The third construct, SMOΔC-BRIL, also included all three domains and, using an established strategy in GPCR crystallography, the third intracellular loop (ICL3) between TM helices V and VI was replaced by BRIL (Fig. 2.1, page 26)[72]. Constructs were cloned into the pHLSec vector with a C-terminal Rho1D4 antibody epitope tag (1D4-tag) for expression in mammalian cells (described in Materials & Methods, Section 2.1.1, page 24). For initial screening using fluorescence-detection size exclusion chromatograhy (FSEC), a monoVenus fluorescent tag (mVenus) was also added at the C-terminus[175, 259].

Initial screening was performed using transiently transfected adherent HEK293S cells at small scale. Detergent-solubilised SMO constructs were isolated by 1D4-antibody affinity and assayed for monodispersity and yield by FSEC. The three constructs were initially tested for extraction in three detergents: n-octyl-β-D-glucoside (βOG), octyl glucose neopentyl glycol (OGNG) and decyl maltose neopentyl glycol (DMNG), all with cholesteryl hemisuccinate (CHS) at a molar detergent:CHS ratio of 10:1 (for detergent details see Appendix D, page 231). All three constructs expressed relatively well and their FSEC elution profiles revealed relatively sharp and symmetrical peaks, suggestive of monodispersity, at ∼20 min retention time (Fig. 3.7, panel a).

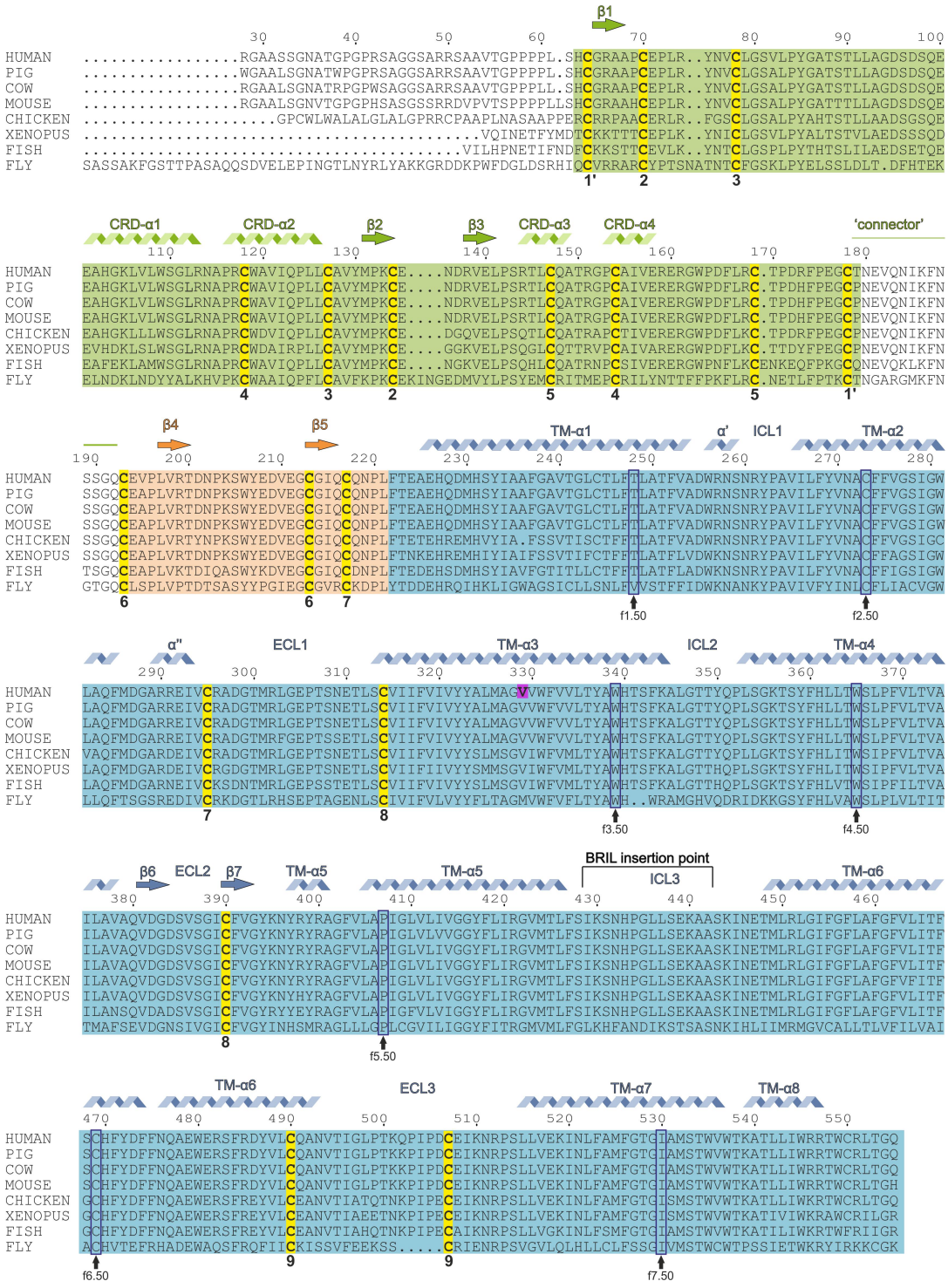


Figure 3.6: Sequence alignment of SMO orthologues. Numbering corresponds to that of human SMO. Coloured blocks show domain boundaries of previous crystal structures. Secondary structure assignments are displayed above the alignment (green, CRD; orange, LD; blue, TMD). Black arrows and numbers (fX.50) below alignment show class F Ballesteros-Weinstein nomenclature for GPCR helices[378]. Disulfide bridges are highlighted in yellow and numbered. Position of the Val329Phe point mutation is highlighted in purple.

Western blotting confirmed that the extracted samples contained substantial amounts of 1D4-tagged protein at the expected sizes for the full-length SMO constructs (86 kDa for BRIL-ΔCRD-SMOΔC-mVenus, 103 kDa for SMOΔC-BRIL-mVenus, 91 kDa for SMOΔC-mVenus) with no significant quantity of degradation products, such as mVenus alone (27 kDa) (Fig. 3.7, panel b). This suggested that the main peak in the FSEC curves corresponded to the full-length constructs. Substantial amounts of higher molecular weight species were apparent in all of the samples, their regular banded appearance suggestive of higher order oligomers or partial aggregation.

The main peaks of the two BRIL-containing constructs, across all three detergent conditions, were substantially higher than those of the construct not containing BRIL, indicating that the BRIL fusions increased the yield of the construct (Fig. 3.7, panel a, left and middle versus right). This may be due to either increased expression in cells, increased receptor stability during purification or both.

3.3.2 Detergent DDM thermostabilises SMO constructs

The three initial detergents all resulted in approximately the same yield for the SMOΔC-BRIL-mVenus construct (Fig. 3.7), indicating that they were all equally efficient in extracting SMO from the cell membrane. However, different detergents may stabilise membrane proteins to differing degrees. In order to ascertain which of the extraction conditions was best able to stabilise SMO, a thermostability assay was used (see Materials & Methods, Section 2.2.3, page 34)[143, 243]. At higher temperatures an increased proportion of protein is denatured, forms aggregates and drops out of solution, which on FSEC corresponds to a lowering of the intensity of the monodisperse peak and/or a shift to higher molecular weight species. The stability of detergents and other extraction conditions were assessed by plotting maximum peak

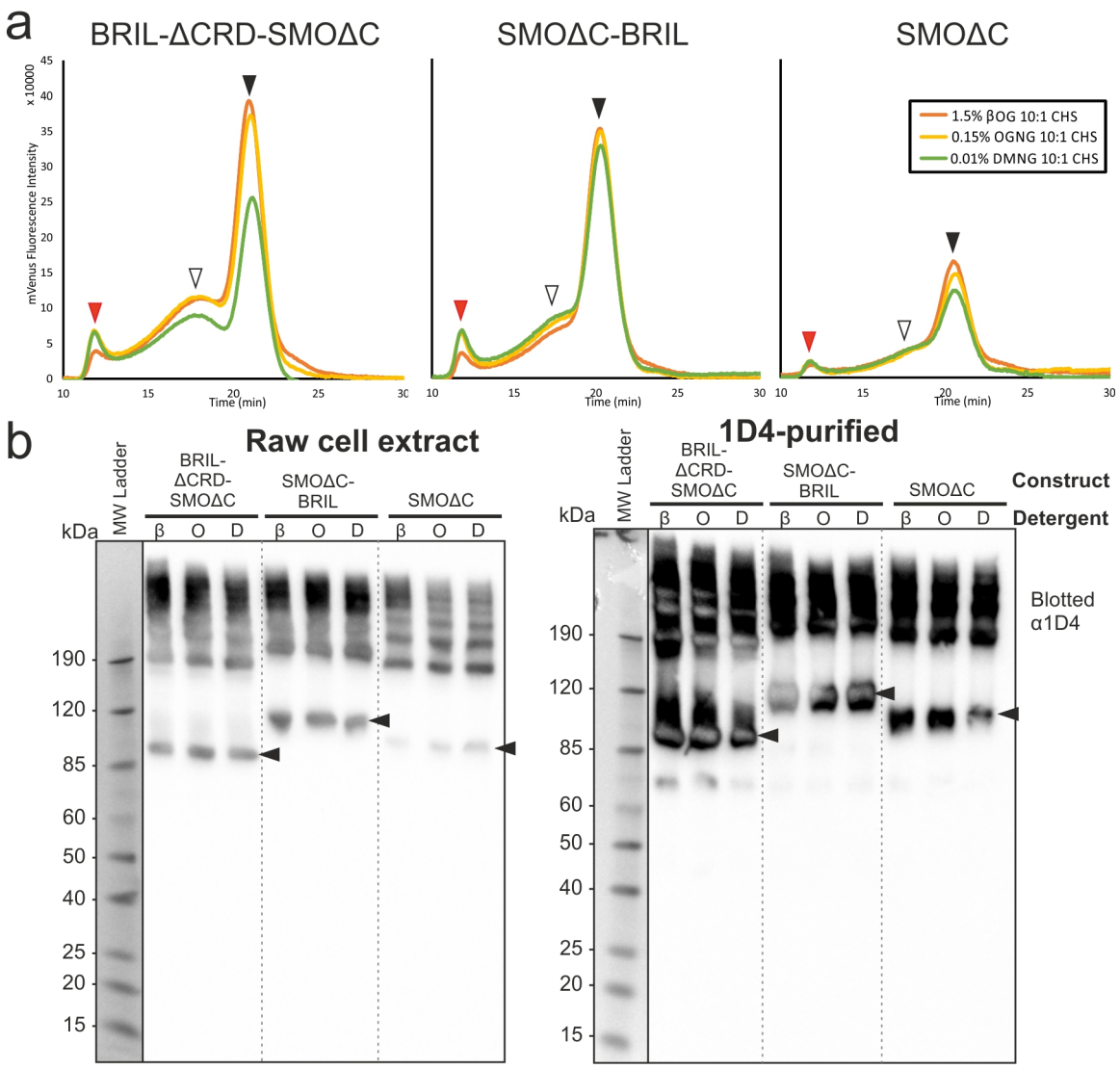


Figure 3.7: Initial screening of three constructs. **a,** FSEC traces for small scale expression tests of YFP-tagged constructs. Constructs were each solubilised in three different detergent conditions. Red arrowheads indicate aggregate peak, black outline arrowheads indicate oligomer peak, black filled arrowheads indicate monomer peak. **b,** Western blots of raw cell extract (left) and affinity purified (right) samples for each of the three constructs extracted in each of the three detergents and detected using anti-1D4 antibody. Black arrowheads indicate monomer SMO bands. (β, 1.5% βOG 10:1 CHS; O, 0.15 % OGNG 10:1 CHS; D, 0.01 % DMNG 10:1 CHS.)

intensity against temperature to produce a 'melting' curve (Fig. 3.8).

The original three detergents plus n-dodecyl-β-D-maltopyranoside (DDM), a detergent commonly used for GPCRs, were screened with different ratios of CHS to determine the best condition with which to proceed (Fig. 3.8). DMNG and DDM, across all CHS concentrations assayed, appeared to stabilise the SMOΔC-BRIL-mVenus construct to a greater extent than OGNG or βOG. The peak intensities for DDM and DMNG did not show the same drop off as those for βOG and OGNG at temperatures higher than 40°C. This effect was observable across all tested ratios of CHS. In fact, the CHS ratio appeared to have little, if any, consistent effect on protein stability. While their responses to high temperatures were similar, ultimately, across temperature ranges and CHS ratios, the peak intensities for DDM-solubilised SMO were marginally higher than those for DMNG. Therefore, DDM was selected as the primary detergent for future work. In addition, the effect of extraction buffer pH on SMO thermostability was tested. Membrane extraction in pH7.8 buffer produced a higher monodisperse peak over the full range of temperatures, compared to pH7.2 or pH6.6 buffers (Fig. 3.8). This higher pH was used for future work.

3.3.3 Inactivating point mutation V329F boosts expression

Due to their inherent conformational heterogeneity, GPCRs generally require the incorporation of stabilising point mutations in order to crystallise[316, 128, 144]. Several point mutations were assayed at small scale for their effect on SMO expression levels. Constructs were generated using site-directed mutagenesis (see Materials & Methods, Section 2.1.3, page 27) and assayed using FSEC. Altogether, four activating and two inactivating point mutations were assayed: the Trp535Leu ("SmoM2"[403]), Leu412Phe, Ala459Val and Asp473Arg[407] mutants were constitu-

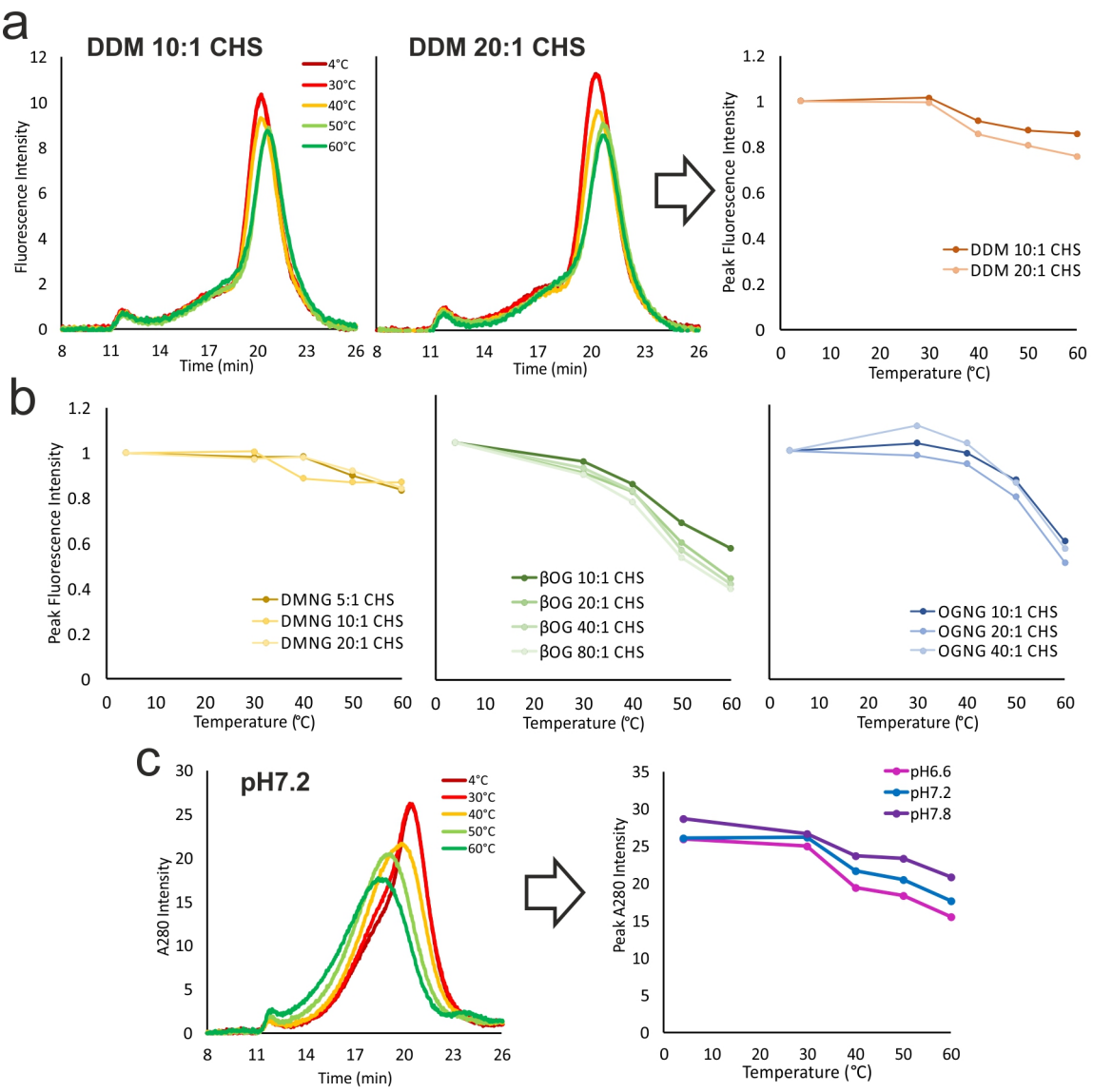


Figure 3.8: Thermostability screening of SMO constructs using FSEC. a, FSEC traces for SMOΔC-BRIL-mVenus extracted in DDM with two CHS ratios after heating samples to five temperatures (left, middle). Normalised, combined peak intensities for the two DDM:CHS ratios plotted against temperature (right). **b,** Normalised, combined peak intensities for SMOΔC-BRIL-mVenus extracted in DMNG (left), βOG (middle), OGNG (right), each at different ratios to CHS. **c,** FSEC trace for SMOΔC-BRIL purified in buffers all at pH7.2 (left). Combined peak intensities for three pH conditions, each at five temperatures, all three extracted in 0.01% DMNG 10:1 CHS (right).

tively active; whereas the Val329Phe and Ile408Phe mutants were inactivating and also blocked SANT1 binding[378]. These mutations were chosen from a wider panel of point mutations that had been pre-screened by our collaborator, Rajat Rohatgi (Stanford University), for their ability to express at small scale and to respond to agonists or antagonists in cells.

Of the six mutations, only Val329Phe displayed substantially increased expression (Fig. 3.9). The other inactivating mutation, Ile408Phe, was expressed to approximately the same extent as the wildtype construct, while the activating mutations all had lower expression levels. A double mutant containing both inactivating mutations (Val329Phe and Ile408Phe) displayed expression levels that were either at or lower than wildtype. These observations were generally consistent with the theoretical paradigm that active states of GPCRs are much more conformationally flexible than inactive states[192]. As a consequence, there exist many more crystal structures of inactive states than active ones.

The human expression construct, SMOΔC-BRIL, containing V329F retained the ability to bind to 20(S)-OHC beads as did the equivalent mutation in full-length mouse SMO, Val333Phe (Fig. 3.10, panels a & b). However, *Smo-/-* mouse fibroblasts stably expressing the Val333Phe mutant SMO were completely unresponsive to activation; neither the native Hh ligand Shh, nor the TMD agonist SAG, nor the CRD agonist 20(S)-OHC was able to activate signalling. This was demonstrated by measuring endogenous *Gli1* mRNA levels through qRT-PCR and through immunoblotting (Fig. 3.10, panels c & d). These data supported the idea that the Val329Phe mutation locked the TMD in an inactive state, reducing flexibility and stabilising the protein during expression and purification.

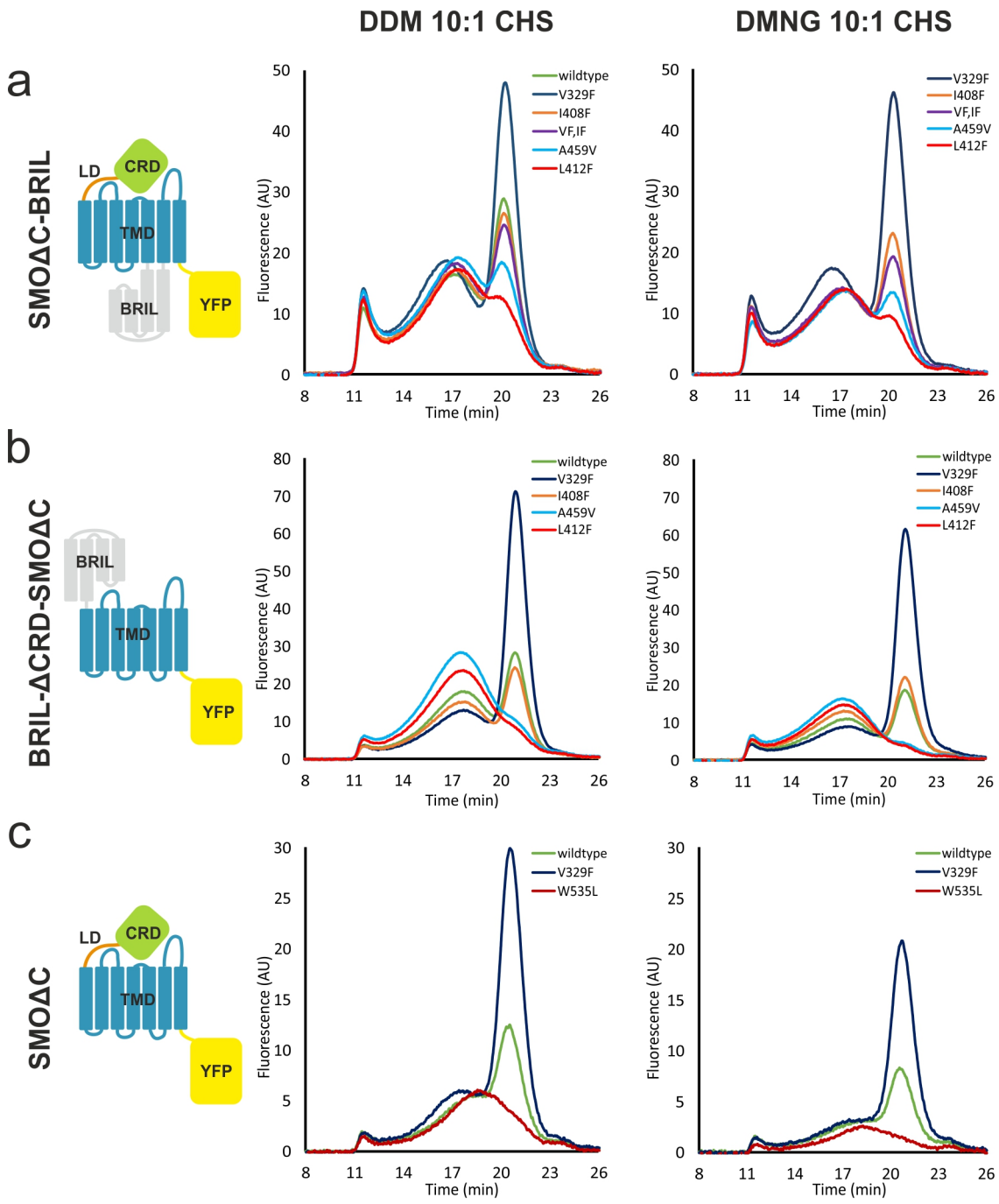


Figure 3.9: Expression screening of point mutations using FSEC. Small scale expression of point mutants using adherent HEK293S cells followed by affinity purification with Rho1D4 beads. Cells were either extracted in DDM 10:1 CHS (left) or DMNG 10:1 CHS (right). The same quantity of cells was used for each sample within each graph. Each experiment was conducted on at least two separate occasions. **a,** Mutations in SMOΔC-BRIL-mVenus background construct. **b,** Mutations in BRIL-ΔCRD-SMOΔC-mVenus background construct. **c,** Mutations in SMOΔC-mVenus background construct. Residue numbers are from human SMO. (VF,IF: double mutant containing V329F and I408F.)

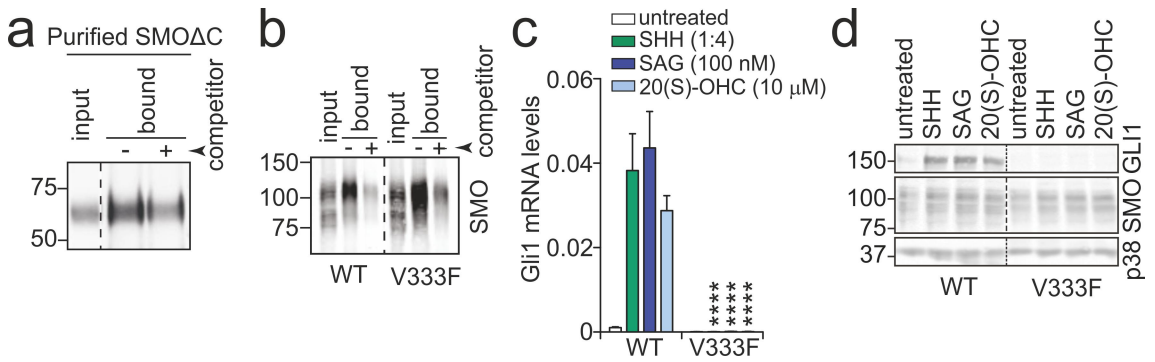


Figure 3.10: Characterisation of the SMO Val329Phe mutation. a, Purified human SMOΔC-BRIL-V329F (the crystallisation construct) binds to 20(S)-OHC beads. Immunoblots, using an anti-SMO antibody directed against the intracellular domain, were used to measure SMO captured on 20(S)-OHC beads. Adding 50 μM free 20(S)-OHC as a competitor reduced binding. **b,** 20(S)-OHC beads can bind both mouse WT SMO and Val333Phe (mouse Val333 corresponds to human Val329). **c,** *Smo-/-* mouse fibroblasts stably expressing SMO-WT or SMO-Val333Phe were exposed to SHH, SAG or 20(S)-OHC. Levels of endogenous *Gli1* mRNA (mean +/- SD, n=4), measured by qRT-PCR, were used as a metric of Hh pathway activity since *Gli1* is a direct Hh target gene. Asterisks indicate statistical significance for the difference in *Gli1* mRNA levels between identically treated SMO-WT and SMO-Val333Phe cells. **d,** Immunoblot shows SMO and GLI1 protein levels in these stable cell lines, with p38 as loading control. These experiments were performed in collaboration with the Rohatgi Lab. Protein in **a** was produced by me.

3.3.4 Optimisation of expression duration and additives

A time course experiment was conducted in order to determine the optimum length of time to allow protein expression following transient transfection of SMOΔC-BRIL-V329F. A total of 3.2 L of HEK293S cells were transfected using our standard protocol (see Materials & Methods, section 2.2.2, page 32) and at regular time points (approximately every 10 h), 50 mL of cells was pelleted, weighed and snap frozen. Samples were solubilised in 1.3% DDM 10:1 CHS and 1D4-affinity purified before analytical SEC. The maximum A280 intensity of the main protein peak and the cell pellet mass were plotted on the same graph against expression time (Fig. 3.11, panel b). The trends in peak intensity and pellet mass were well correlated over time, with an initial rise followed by a plateau around 30-40 hours post transfection (hpt) and then

a further rise to 64 hpt. This result suggested that cells should be left to grow and express protein for as long as possible. Due to equipment and scheduling limitations, the maximum practicable time was 64-72 hours.

In an attempt at further optimisation of the SMO purification procedure and in a manner analogous to crystallisation screening, a 96-condition additive screen was designed and produced in-house for high-throughput stability screening of membrane proteins ("Membrane Protein Additive Screen"; for a summary of conditions, see Materials & Methods, section 2.2.5, page 36, and for a full list of conditions, see Appendix A, page 220). The effects of the additives were tested using a single-point thermostability assay. Briefly, additives were applied to purified protein, heated to 42°C for 1 h and then run on analytical SEC. Potential stabilising additives were further characterised with a multi-point temperature screen. Glycerol was the most promising additive, thermostabilising SMOΔC-BRIL-V329F by several degrees Celsius (Fig. 3.11, panel c). Glycerol at 10% concentration was subsequently included in all wash and size-exclusion buffers.

3.3.5 Large scale purification of SMO for crysallisation

Crystallisation trials require protein concentrations that are very high and may induce aggregation. Therefore, the stability of SMOΔC-BRIL-V329F at high protein concentration was tested; 1D4-purified samples were run on analytical SEC before (∼0.2 mg/mL) and after concentration (∼5.3 mg/mL). Upon concentration the proportion of aggregate to monomer peak intensity did increase somewhat but the sharp monomer peak remained the dominant feature (Fig. 3.12, panel a), suggesting that the protein behaves relatively well at high concentrations. The stability of the protein after a freeze-thaw cycle was also tested. SEC-purified SMOΔC-BRIL-V329F at

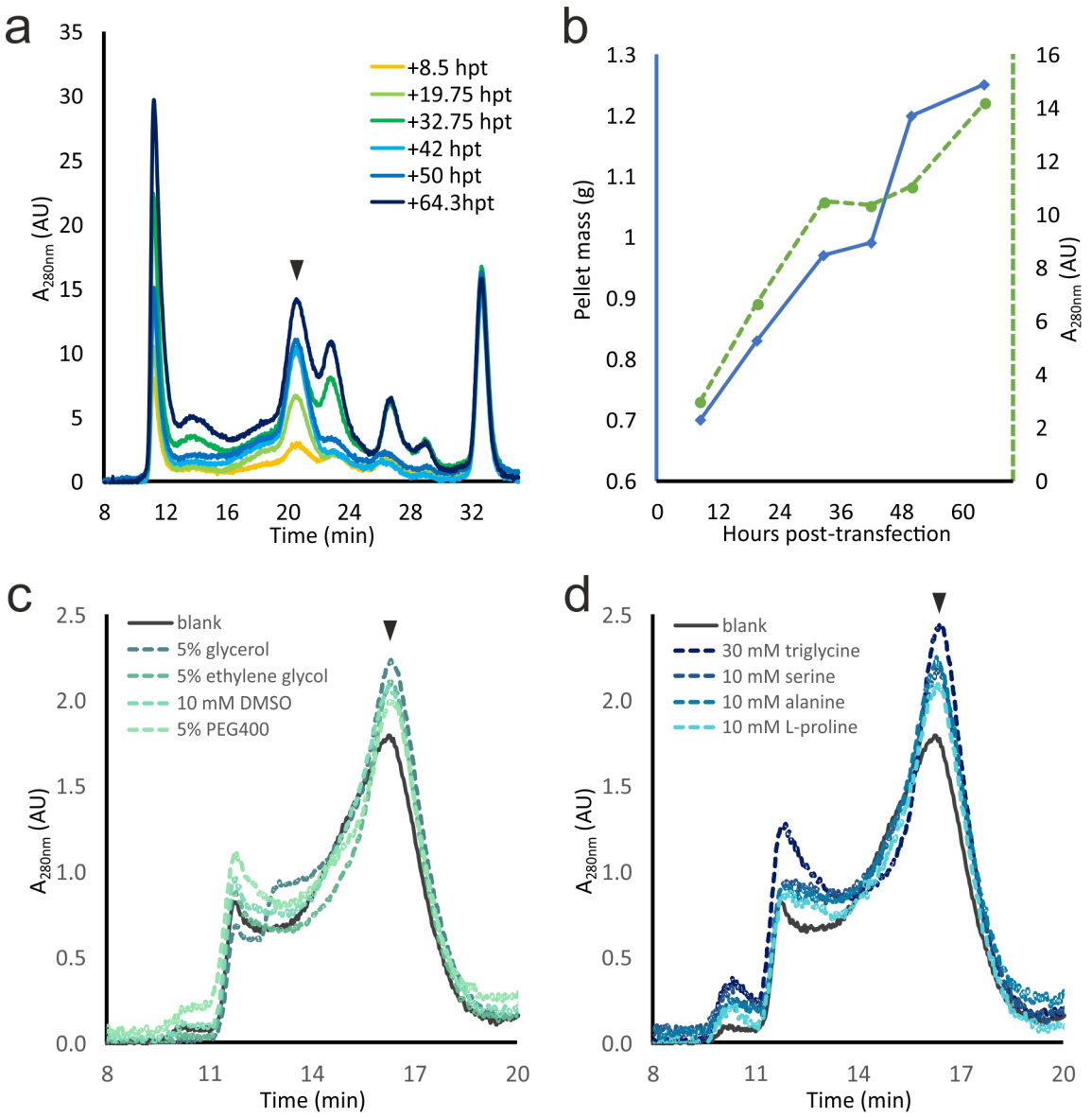


Figure 3.11: Optimisation of expression and stabilisation. a, Analytical SEC of SMO expression time course. Equivalent volumes of HEK293S cells in suspension were harvested at the times indicated (hpt, hours post transfection), affinity-purified and run on Superose 6 3.2/300 column. Black arrowhead indicates main protein peak in a. **b,** Pellet mass (left y-axis, blue) and absorbance at 280 nm (right y-axis, green) over expression time course. Absorbance taken from main protein peak height. **c & d,** Analytical SEC traces of SMO with various stabilising additives from Membrane Protein Additive Screen. Samples run on Superdex 200 3.2/300 column after 1 h heating to 42 °C. Black arrowheads indicate main protein peak.

$\sim$3.1 mg/mL was snap-frozen in liquid N_2, stored for three months at -80°C and then thawed on ice before running on analytical SEC. The main peak retained a sharp and symmetrical appearance and was eluted at the expected retention time for the monomeric protein, with only very minor shoulders to either side of the main peak and a non-protein contaminant at higher retention time (Fig. 3.12, panel c). The lack of aggregates, oligomeric species or protein degradation products indicated that this construct was relatively resilient to at least one freeze-thaw cycle.

The protein was also run on size exclusion chromatography (SEC) coupled to multi-angle light scattering analysis (MALS) to ascertain its molecular weight in solution (Fig. 3.13, panel c). In order to reduce scattering due to detergent in solution, the protein was exchanged into amphipols before running on SEC-MALS (see Materials & Methods, Section 2.6). Amphipols are amphiphilic polymers capable of keeping membrane proteins in solution in the absence of detergent[360]. The total particle size was 147 kDa and, using protein-conjugate analysis (ASTRA 6.1 software, Wyatt), the contribution of the protein was determined to be 84 kDa and that of the amphipol 63 kDa. This result was consistent with monomeric SMO (71 kDa) bound by 14-15 amphipol molecules ($\sim$4.3 kDa each[131]). The difference between the protein MW measured by MALS and the sequence-predicted MW was due to several assumptions made during the protein-conjugate analysis.

Towards setting up crystallisation trials, large scale expression and purification of BRIL-ΔCRD-SMOΔC-V329F and SMOΔC-BRIL-V329F was conducted. Briefly, suspension cultures of HEK293S cells, with a typical batch size of 9.6 L, were transiently transfected with these constructs and harvested after 72 h. Cell pellets were extracted with 1.0% (w/v) DDM, 0.2% (w/v) CHS before binding to 1D4-beads. Beads were washed extensively in buffer containing 10% glycerol and decreasing amounts of DDM and CHS. Eluted protein was further purified by SEC

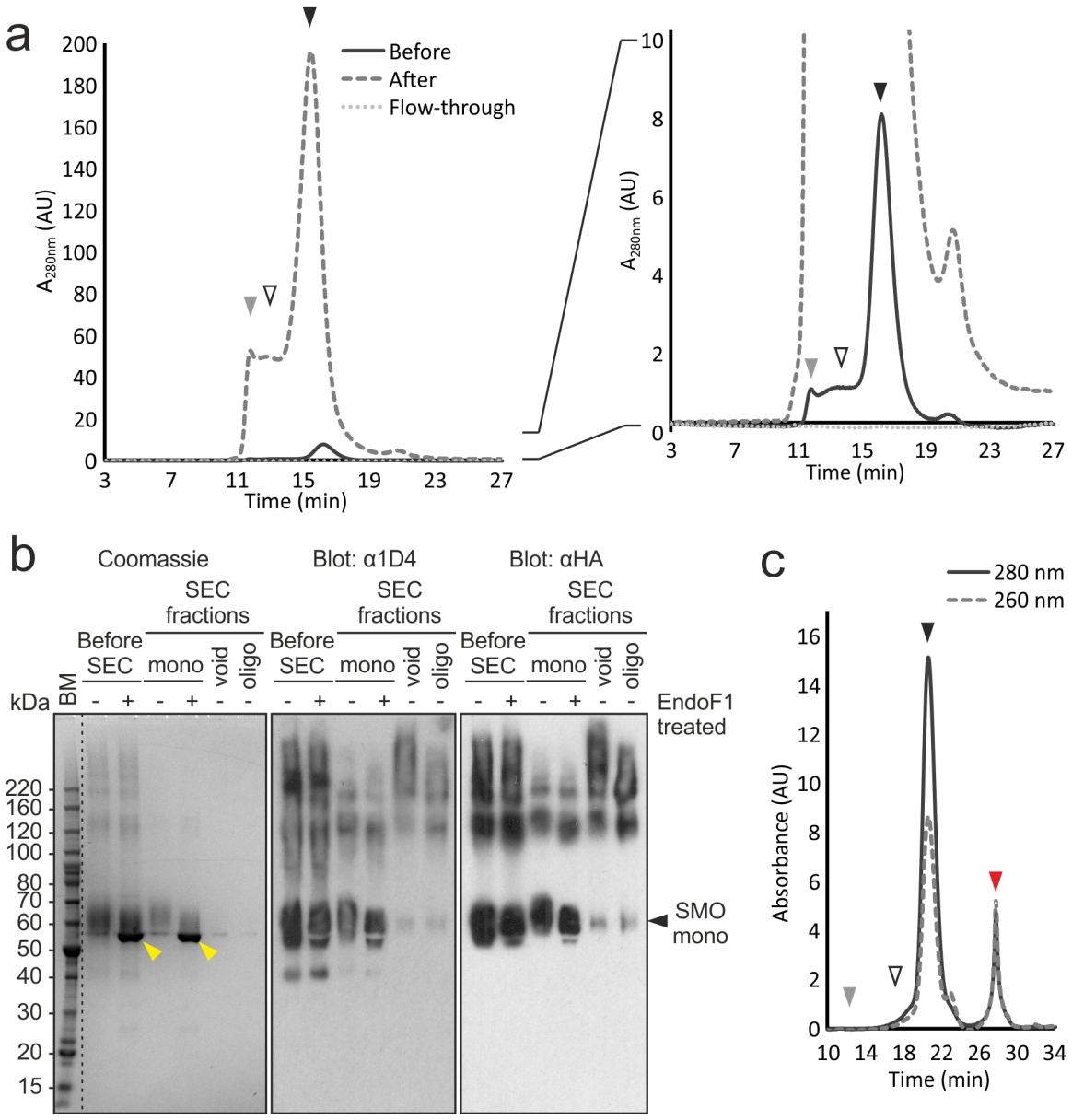


Figure 3.12: Behaviour of SMOΔC-BRIL during and after purification procedure. a, Analytical SEC traces of eluate from Rho1D4 affinity purification before concentration ('Before', 0.2 mg/mL), after concentration ('After', 5.3 mg/mL) and concentrator flow-through ('Flow-through'). Run on Superdex 200 3.2/300 column. Right graph shows expanded y-axis. **b,** Reducing SDS-PAGE analysis of SMOΔC-BRIL purification. Identical loading of three gels followed by Coomassie Blue staining (left), blotting with anti-1D4 antibody (middle) or blotting with anti-HA antibody (right). SMOΔC-BRIL construct tagged with Rho1D4-epitope at C-terminus and HA-epitope at N-terminus. Yellow arrowheads indicate dark EndoF1 bands. **c,** Analytical SEC trace of purified SMOΔC-BRIL after being concentrated to 3.1 mg/mL, snap-frozen, stored at -80 °C for 90 days, thawed on ice and diluted ten-fold. Run on Superose 6 3.2/300 column. Grey arrowheads indicate aggregate peak, black outline arrowheads indicate oligomer peak, black filled arrowheads indicate monomer peak, red arrowhead indicates low MW contamination.

before concentration and crystallisation trials. All buffers were at pH7.5 throughout. (For full details, see Materials & Methods, Section 2.2.2, page 32.) A representative SEC trace and reducing SDS-PAGE analysis from a large-scale purification is shown (Fig. 3.13, panels a & b). Typical cell pellet mass from 9.6 L of suspension culture was 120-180 g; typical purified protein yield from 120-180 g pellet was 0.8-1.6 mg.

3.3.6 SMO crystallises in LCP at high concentration

Crystallisation trials using the vapour diffusion method were set up in Swissci plates (Hampton Research), with three different protein:reservoir solution ratios (1:1, 1:2, 2:1) for each well, with protein initially at ∼3.0 mg/mL. However, they did not produce any crystals despite a substantial proportion of conditions producing precipitate, indicating that the protein concentration was about right. Due to the propensity for GPCRs to crystallise in LCP, I quickly moved on to setting up LCP crystallisation trials. The initial LCP set-ups were focused on trying to crystallise the already-published construct, BRIL-ΔCRD-SMOΔC-V329F [379], in order to validate the expression, purification and LCP procedures. Initial trials were also conducted at the relatively low protein concentration of 10-12 mg/mL due to the difficulty and expense of producing sufficient protein for crystallisation. Purified protein that had been concentrated and treated with EndoF1 was mixed with 9.9 monoacylglycerol (monoolein) in a ratio of 2:3 (v/w) using the mechanical syringe method. Clarified LCP mixture was applied to glass plates in 50 nL boluses and they were covered with 800 nL of precipitant solution before immediately sealing with a coverslip. Plates were incubated at either 20°C or 4°C and monitored by eye. (For full details, see Materials & Methods, Section 2.3.5, page 40.)

A survey of the literature revealed that no GPCR had been crystallised in

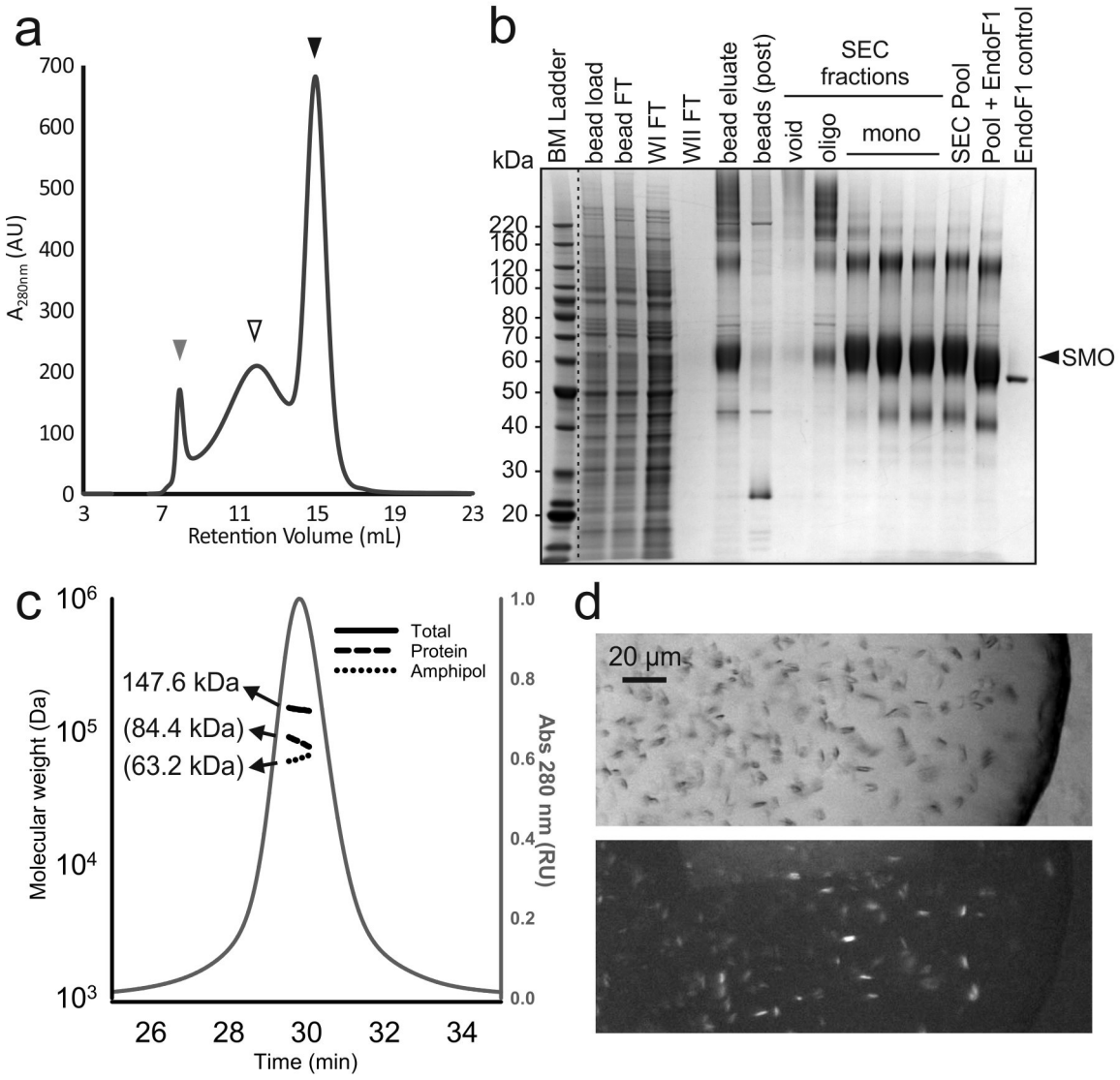


Figure 3.13: Large scale expression and purification of SMOΔ-BRIL-V329F. a, Typical SEC trace from a large scale purification. Grey arrowhead indicates aggregate peak, black outline arrowhead indicates oligomer peak, black filled arrowhead indicates monomer peak. **b,** SDS-PAGE analysis of large scale purification. (BM Ladder, BenchMark Ladder; FT, flow-through; void, aggregate peak; oligo, oligomer peak; mono, monomer peak; EndoF1, Endoglycosidase F1.) **c,** MALS analysis of amphipol-solubilised SMOΔ-BRIL. Protein-conjugate analysis revealed the approximate MW of protein and amphipol within the particle (values shown in parentheses). **d,** Photos of first crystal hit under white light (top) and cross-polarisers (bottom).

LCP using a protein concentration of less than 20 mg/mL, while some had used up to 80 or 100 mg/mL (for example, [244, 102, 397, 147, 337, 379, 378]). In my hands, the first SMO crystals were grown only when the protein concentration was increased to greater than 25 mg/mL. These crystals were of the CRD-containing SMOΔC-BRIL-V329F construct, rather than the already-published BRIL-ΔCRD-SMOΔC-V329F (Fig. 3.13, panel d). Critical to producing enough protein to reach these very high concentrations was increasing the total batch size of a single transfection from 6.4 L to 9.6 L, adding $\sim$50% extra cell mass per purification. Initial crystals were grown in a condition containing 30% (v/v) PEG500 DME, 0.1 M potassium sodium tartrate, 0.1 M sodium acetate pH4.0, 0.5 mM zinc chloride and 0.1 M ammonium fluoride. These crystals were very small (approximately 6 x 2 x 3 μm^3) rods that appeared as a shower throughout the lipid bolus with no crystals outside the bolus (Fig. 3.13, panel d). Crystals were optimised by varying the proportion of PEG500 DME (generally 32-20%) and the salt concentration (generally 200-0 mM). All crystals were grown at 20°C. The morphology of optimised crystals varied between rods, needles, boat-shaped or shards that could be described as bowtie-, lightning- or 'X-Wing Fighter'-shaped. The crystals used for the final structure solution were grown in 0.1 M MES pH6.0, 30% (v/v) PEG500 DME, 0.1 M sodium acetate, 0.5 mM zinc chloride, 0.1 M ammonium fluoride, and were approximately 20-30 x 10 x 5 μm^3 shards.

3.3.7 Structure solution from multiple isomorphous crystals

Prior to data collection crystals were flash-frozen at 105 K (see Materials & Methods, Section 2.3.7, page 42). Data were collected at the microfocus beamline mx I24 at the Diamond Light Source (Harwell, UK). The microfocus capability of this beamline was essential given the small size of the crystals as was the grid-scan software for locating the crystals given the opacity of cryo-cooled LCP[331].

After autoprocessing by Xia2 (ref.[395, 170]) failed, the initial dataset was collated by processing using XDS[170], XSCALE[170] and Aimless[112, 394]. The final dataset required merging of data from seven crystals, with a contribution of between 22.6° and 60° from each. The data collection strategy could not be guided by beamline software (such as EDNA[157] or MOSFLM[213]) as data from individual crystals were often incomplete or of low quality and therefore not conducive to automatic processing. Instead, the data collection strategy was simply to find and shoot as many crystals as possible and then exhaustively test different combinations of the partial datasets ('sweeps'). This was only made possible by the excellent setup and grid-scan option at beamline I24. The structure was solved by molecular replacement, searching for truncated rigid domains using independent search models generated from human SMO TMD (PDB 4QIM[378]), zebrafish CRD (PDB 4C79[258]) and BRIL (PDB 4EIY[222]). Two TMDs, two BRILs and one CRD were found within the asymmetric unit using Phaser[239]. The five independent domains were then re-organised into two chains based upon the location of the N- and C-termini of each domain. One chain contained one of each of the three domains. This three-domain chain was then used as a search model for a new round of molecular replacement, resulting in two three-domain chains within the asymmetric unit (Fig. 3.14, panel a). From this point, missing residues between the domains could be built manually (for instance, the 'connector' region between the CRD and LD shown in Fig. 3.14, panel e).

The SMOΔC-BRIL-V329F structure was first determined 3.95 Å resolution and, after several rounds of optimisation of crystallisation conditions, the resolution was improved to 3.2 Å (Crystallographic Data Table, page 52). Pearson correlation coefficient (CC) analysis was used to relate data quality with model quality and therefore determine the maximum allowable resolution[173]. A CC_{work} and CC_{free} smaller than CC* indicates that the model does not account for all of the signal in the data and is therefore not overfit (Fig. 3.14, panel c). The asymmetric unit,

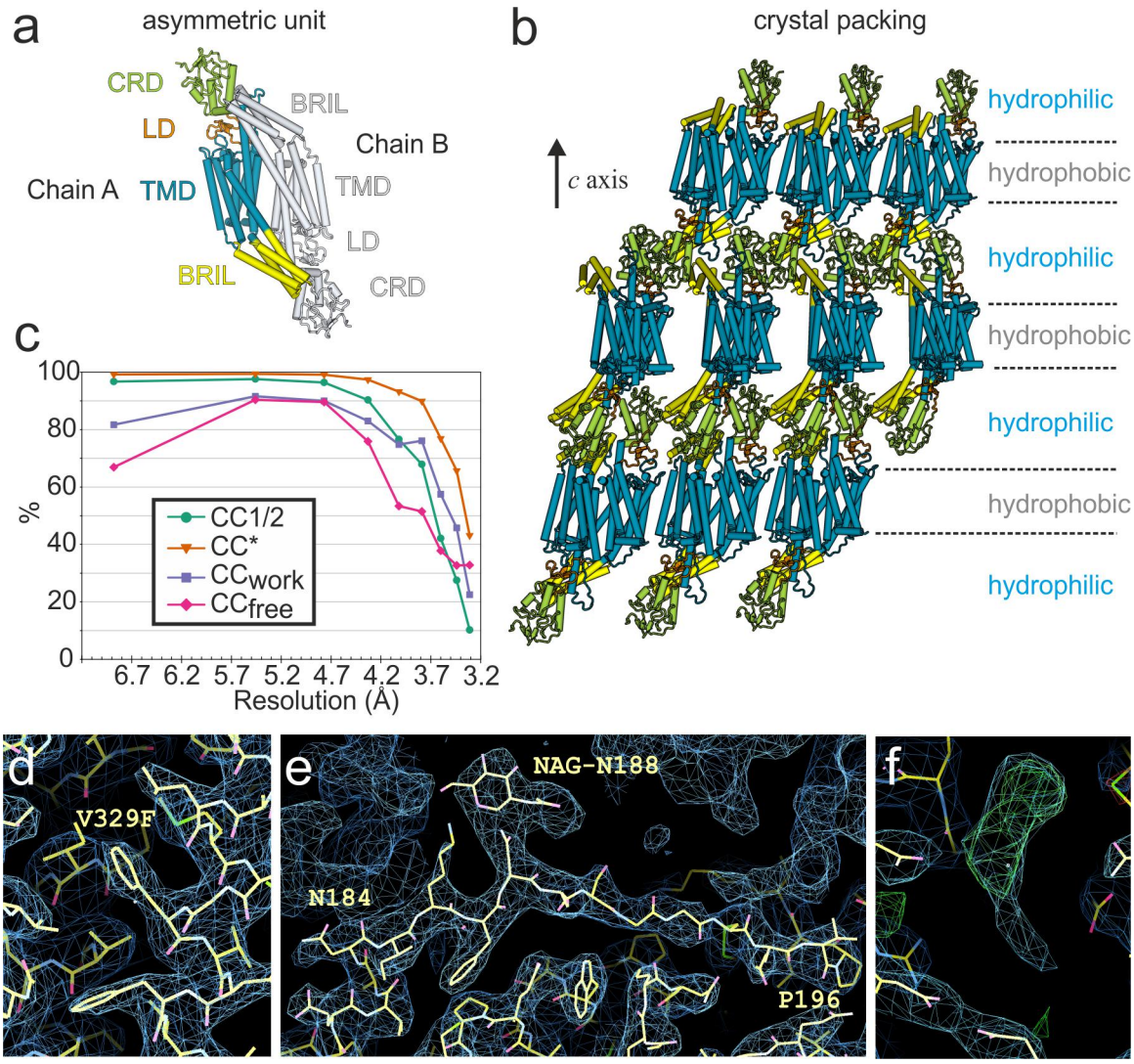


Figure 3.14: Crystallisation and structure solution. a, SMOΔC-BRIL crystal packing. The asymmetric unit consists of two SMOΔC-BRIL chains in an antiparallel arrangement. **b,** LCP crystal packing with alternating hydrophobic and hydrophilic layers perpendicular to the c axis. Molecules coloured as for asymmetric unit Chain A. **c,** Data and model quality. Pearson correlation coefficient (CC) analysis (as described by Karplus and Diederichs[173]) used to relate data quality with model quality. CC1/2: half-set experimental correlation coefficient; CC*: analytical estimate of CC_{true} based on CC1/2 (CC_{true} being the correlation of the averaged dataset with the noise-free true signal); CC_{work}: standard correlation of the experimental intensities with the intensities calculated from the refined model; CC_{free}: cross-validated correlation of the experimental intensities with the intensities calculated from the refined model. **d-f,** SigmaA-weighted 2Fo-Fc electron density maps of the final refinement at a contour level of 1.0 σ. **d,** The region of the Val329Phe mutation. **e,** The 'connector' region linking CRD and LD. Asn188 harbouring an N-linked N-acetyl glycosamine moiety is labelled. **f,** Extra density within the TMD ligand binding pocket with $F_o - F_c$ maps (at a contour level of $+3\sigma$ (green) and -3σ (red)) also shown.

comprising two molecules arranged 'head-to-tail', stacks into alternating hydrophobic and hydrophilic layers along the *c* axis typical for LCP derived crystals (Fig. 3.14, panels a & b). This SMO arrangement within the crystal suggests that SMOΔC-BRIL-V329F is monomeric, in agreement with our SEC-MALS analysis (Fig. 3.13, panel c). We also observed a systematic disorder along this *c* axis, with alternating hydrophilic layers displaying higher and lower ordering, respectively. This disordering was not due to incorrect space-group assignment as re-processing into the lowest symmetry space-group, P1, did not ameliorate the disorder.

3.3.8 Crystal structure of human SMO CRD-LD-TMD

SMO adopts an extended conformation in the structure. The extracellular CRD is perched on top of the LD, which forms a wedge between TMD and CRD. At the apex of this wedge, the CRD contacts the TMD through the elongated TMD extracellular loop 3 (ECL3) (Fig. 3.15). The overall architecture is stabilised by nine disulphide bridges, four of which reveal the canonical disulphide pattern of the CRD fold[15] (numbered 2-5 in Fig. 3.15) and one that is specific for SMO[258] (marked as 1' in Fig. 3.15). This disulphide bridge positions the start of a 'connector' segment that links the CRD to the LD and had not previously been seen in a crystal structure (Fig. 3.16). This connector is tucked along a hydrophobic groove that runs the length of the CRD from top to bottom and shields this groove with three inwardly turned hydrophobic residues (Val182, Ile185, Phe187) in addition to multiple hydrophilic residues and a N-linked glycan facing the solvent (Fig. 3.16, panel a, and Fig. 3.14, panel e). The connector region, combined with mainly hydrophobic interactions between the CRD and both the LD and ECL3 (total buried surface area of 745 Å^2), orients the CRD in an upright conformation with its N-terminus pointing away from the plasma membrane (Figs. 3.15 and 3.16, panels a, c & d). However, we did not observe major

structural changes in the heptahelical TMD bundle compared to previously solved structures of both antagonist- or agonist-bound complexes lacking the CRD[379, 378]. The only exception was a rearrangement of the LD, which in our structure is pushed down towards the TMD by ∼2 Å, perhaps caused by CRD-binding or our inclusion of the inactivating point mutation Val329Phe (Fig. 3.17).

We observed some unexplained extra electron density in the TMD binding site (Fig. 3.14, panel f) that could not be fitted with any known ligands of SMO. This observation is discussed more in the next chapter (see section 4.3.7, page 151).

3.3.9 Structure reveals cholesterol molecule bound to CRD

Unexpectedly, we discovered a cholesterol molecule in our SMO structure (Fig. 3.15). Cholesterol occupies a central position, interacting with all three SMO domains, the CRD, LD and TMD (Fig. 3.16, panel b, and Fig. 3.18, panels a & b). Cholesterol adopts an extended conformation with its tetracyclic sterol ring bound in a shallow groove in the CRD, a site previously shown to bind 20(S)-OHC[258, 263, 256]. The cholesterol iso-octyl tail, located at the interface between the CRD, LD and TMD-ECL3, is buried in the SMO protein core. This arrangement positions cholesterol some 12 Å away from the lipid bilayer of the plasma membrane (Fig. 3.18, panel c) and implies that a cholesterol molecule would have to completely desorb from the plasma membrane surface to access its binding site in SMO.

The sterol-binding site of SMO is predominantly lined by hydrophobic residues from the CRD that stabilise the flat α-face of cholesterol (Fig. 3.18, panels a & b). Mutations in several of these residues (Leu108, Trp109, Pro164 and Phe166) have been noted to prevent SMO binding to 20(S)-OHC and to impair both 20(S)-OHC-driven and SHH-driven signalling[258, 263, 256]. The rough β-face is shielded by the

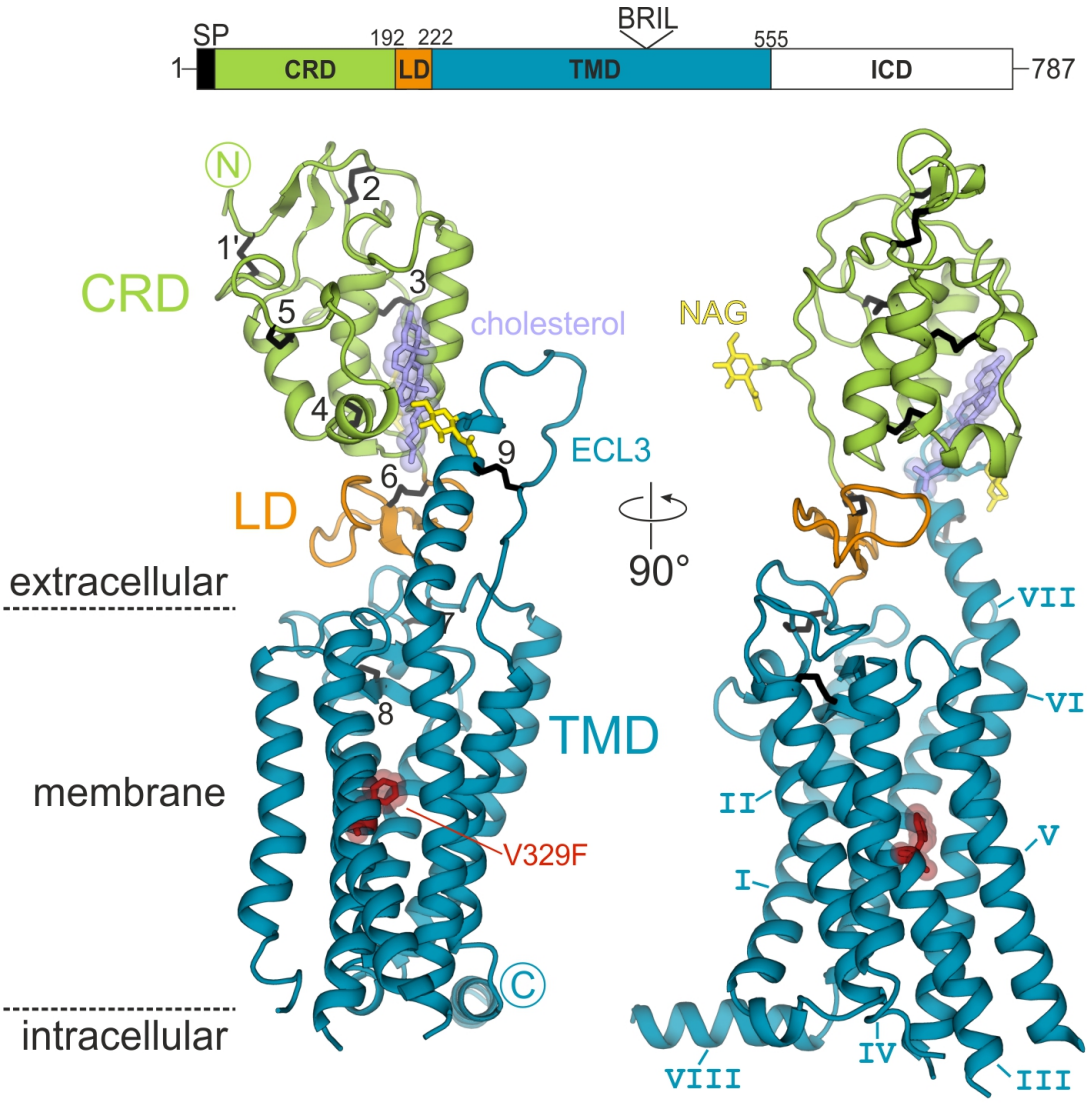


Figure 3.15: Structure of human SMO. Two views of the overall structure showing the full extracellular and transmembrane domains of human SMO in cartoon representation. The cysteine-rich domain (CRD) is highlighted in green, the linker domain (LD) in orange, the transmembrane GPCR domain (TMD) with helices in blue, the inactivating point mutation Val329Phe in red, and cholesterol in violet. The nine disulfide bridges are depicted in black and numbered. Two N-linked glycans (NAG) are shown as yellow sticks. Termini are highlighted. Roman numerals indicate TMD helices. A schematic of SMO is shown above (SP: signal peptide, BRIL indicates the position of the BRIL fusion protein inserted between TMD helices V and VI). (PDB: 5L7D)

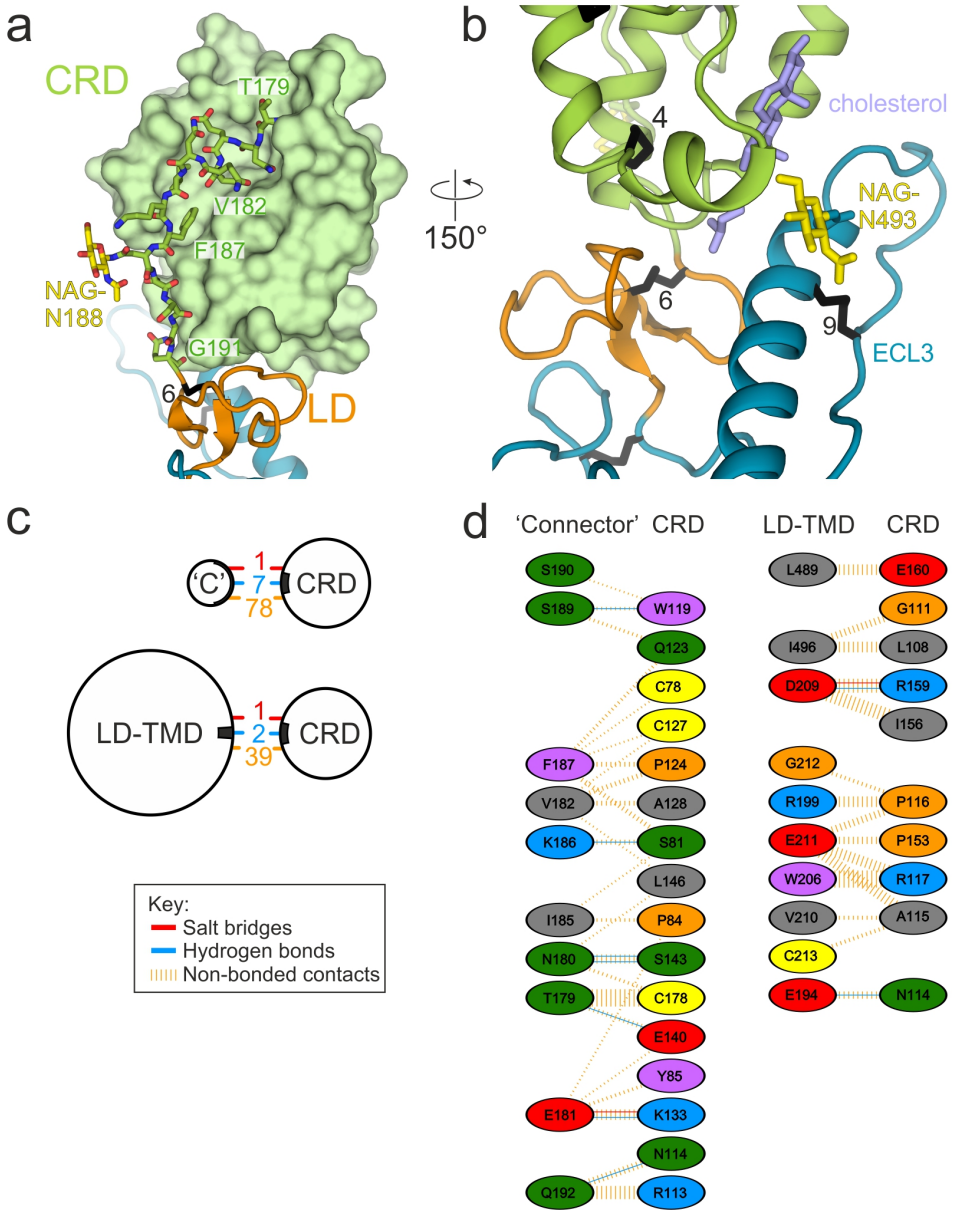


Figure 3.16: Features of human SMO CRD-LD-TMD structure. **a,** The connector region between CRD and LD. CRD is shown as solvent accessible surface with the connector region highlighted in stick representation in atomic colouring. LD and part of the TMD ECL3 loop are shown as cartoons. **b,** Interface between CRD, LD and TMD shown in cartoon representation. The N-linked glycosylation in TMD ECL3 (yellow) and the cholesterol molecule (violet) are shown as sticks (PDB: 5L7D). **c,** Number of interdomain interactions of the SMO CRD with the connector region (top) and the LD-TMD segment (bottom). Bonds coloured as indicated in the key box. **d,** Detail of interdomain interactions. Bond colours as for c. For non-bonded contacts, the width of the striped line is proportional to the number of atomic contacts. Residue colouring is according to amino acid: blue, positive (H,K,R); red, negative (D,E); green, neutral (S,T,N,Q); grey, aliphatic (A,V,L,I,M); purple, aromatic (F,Y,W); orange, proline (P) or glycine (G); yellow, cysteine (C). The figure is adopted from the PDBSUM server.

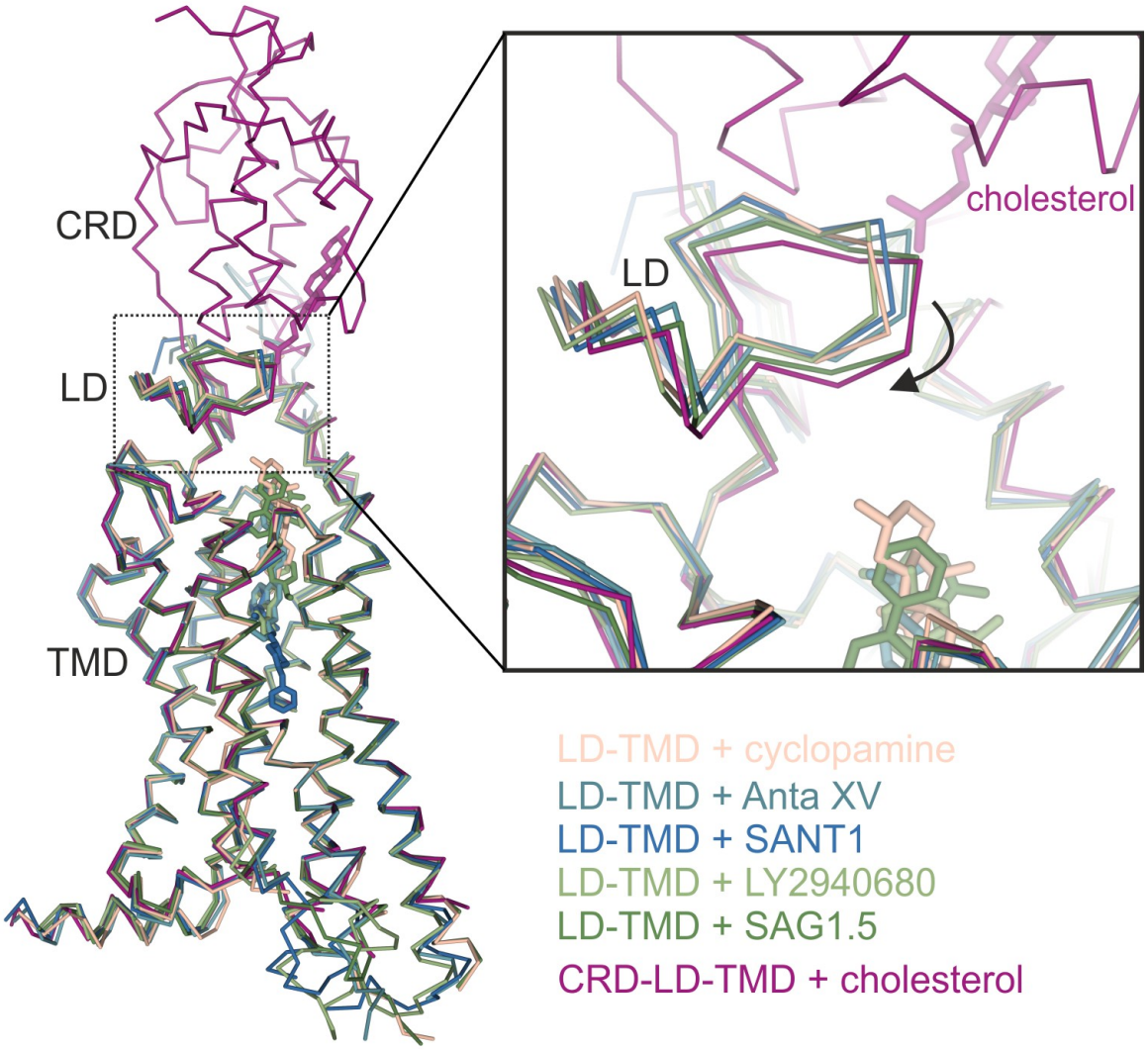


Figure 3.17: Superposition of the SMO CRD-LD-TMD structure with previously determined LD-TMD structures. Structural alignment was performed using the 7TM bundle as template (not including the LD or helix VIII). SMO CRD-LD-TMD with cyclopamine (purple); LD-TMD with cyclopamine (straw, PDB 4O9R, rmsd 0.598 Å for 243 equivalent Cα positions), Anta XV (light blue, PDB 4QIM, rmsd 0.515 Å for 233 equivalent Cα positions), SANT1 (dark blue, PDB 4N4W, rmsd 0.483 Å for 240 equivalent Cα positions), LY2940680 (pale green, PDB 4JKV, rmsd 0.493 Å for 230 equivalent Cα positions) or SAG1.5 (dark green, PDB 4QIN, rmsd 0.623 Å for 262 equivalent Cα positions). The box shows a close-up view of the LD region revealing a structural rearrangement in the CRD-LD-TMD structure compared to the previously determined LD-TMD structures lacking the CRD.

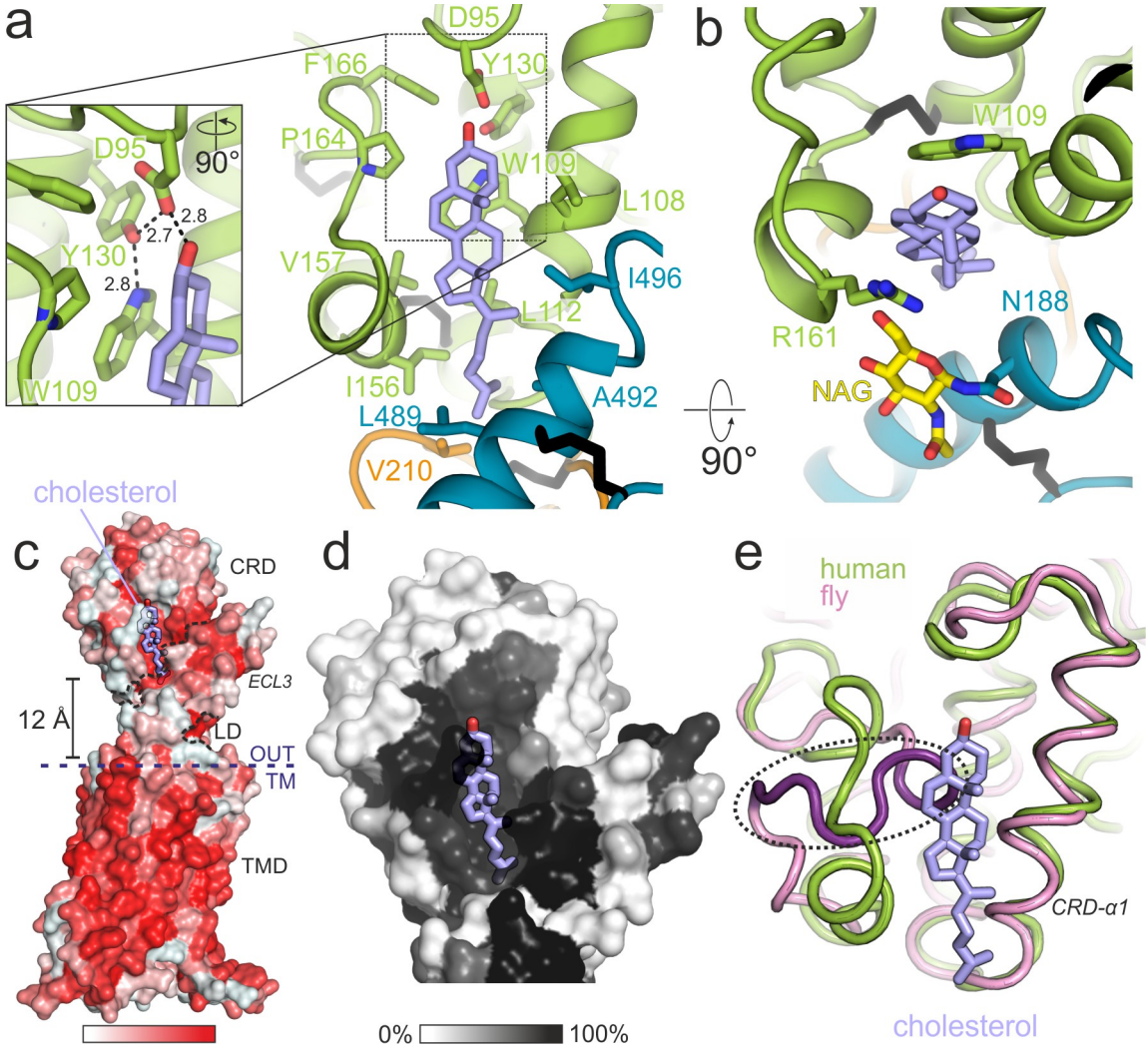


Figure 3.18: The cholesterol binding site in SMO. a, Close-up view of cholesterol and its interacting residues shown in stick representation. Colour coding follows Fig. 3.15. The inset shows the hydrogen bonding network coordinating the cholesterol 3β-hydroxyl group. Black numbers indicate interatomic distances in Å. **b,** Top view of the cholesterol binding site. Orientation is rotated 90° around the x axis relative to a. **c,** Solvent accessible surface representation of SMO colour-coded based on hydrophobicity (red: hydrophobic, white: hydrophilic). **d,** Sequence conservation mapped onto the SMOΔC structure. 55 vertebrate SMO sequences were used for the analysis (black: conserved, white: not conserved). **e,** Superposition of human (green, from this study) and fly (pink/purple, PDB 2MAH) SMO CRD structures. The region of the fly CRD occupying the cholesterol binding site is highlighted.

side-chain of Arg161 and a N-linked glycan (Fig. 3.18, panel b). The cholesterol 3β-hydroxyl group is incorporated in a hydrogen-bonding network with the side chains of Asp95, Trp109 and Tyr130, which optimally position the α-face of the sterol ring system to make a stacking interaction with the indole ring of Trp109 (Fig. 3.18, panel a). TMD-ECL3 residues Leu491, Ala492 and Ile496 and LD residue Val210 orient the iso-octyl tail of cholesterol in an elongated conformation. Residues lining the cholesterol binding site are highly conserved in vertebrates (Fig. 3.18, panel d, and Fig. 3.6).

Cholesterol was not purposefully added to the protein preparations, although it was present at high concentrations in the LCP mix (which contains 6% (w/v) cholesterol). No other small molecule additives (including oxysterols previously shown to bind to the SMO CRD) were included in the preparations either. However, the electron density was immediately discernible in the electron density maps after molecular replacement[239] and initial rounds of structural refinement, indicating high occupancy. The addition of cholesterol to the model improved the R-factor and R_{free} by lowering them by $\sim$1%. However, the limited resolution of the structure prevented completely unambiguous assignment of the extra density. In order to validate both the presence and the apparent binding mode of cholesterol, various functional experiments were embarked upon.

3.3.10 SMO in solution binds cholesterol via CRD binding site

Cholesterol has not before been seen to bind to the SMO CRD. Consistent with the structure, purified SMOΔC-BRIL-V329F was captured by cholesterol-conjugated beads in an assay analogous to the ligand-affinity assay previously developed to mea-

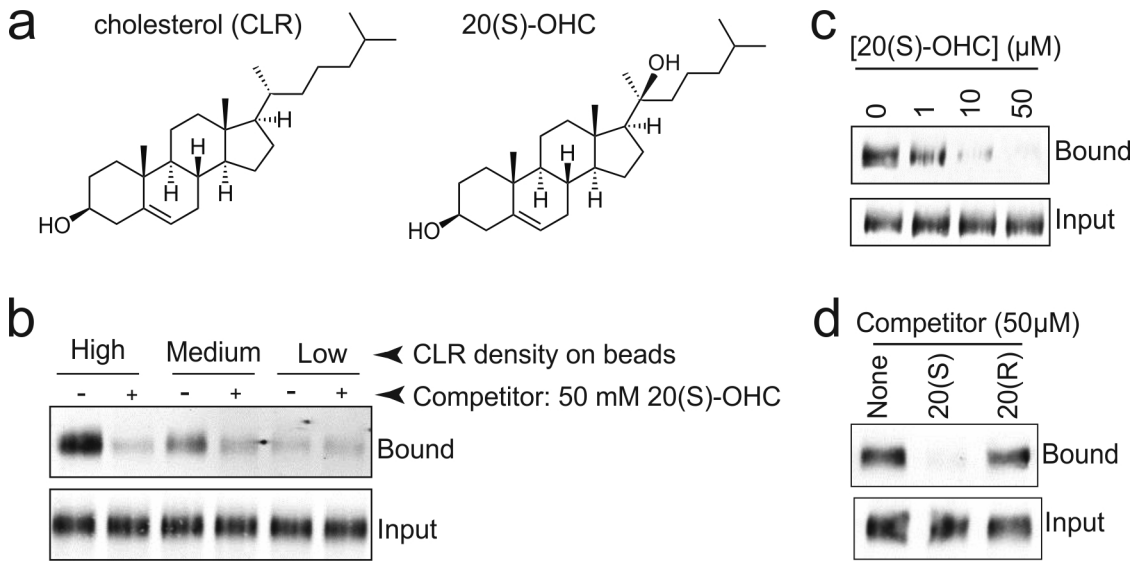


Figure 3.19: SMO binds to cholesterol-conjugated beads. a, Structures of cholesterol and 20(S)-OHC. **b,** Purified SMOΔC-BRIL-V329F captured on beads coupled to increasing concentrations of cholesterol in the presence or absence of free 20(S)-OHC. **c,** SMOΔC-BRIL-V329F captured on cholesterol beads in the presence of increasing concentrations of free 20(S)-OHC. **d,** SMOΔC-BRIL-V329F captured on cholesterol beads in the presence of free 20(S)-OHC or a synthetic epimer 20(R)-OHC. Bead-binding experiments were performed in collaboration with the Rohatgi Lab using beads prepared by the Covey Lab and protein produced by me.

sure the binding of SMO to 20(S)-OHC (Fig. 3.19, panel b)[257]. Binding could be blocked by free 20(S)-OHC added as a competitor at concentrations (1-10 μM) similar to those that activate Hh signalling in cells (Fig. 3.19, panel c)[257], confirming that cholesterol and 20(S)-OHC engaged the same oxysterol-binding groove on the CRD surface. In a stringent specificity control, 20(R)-OHC, an epimer with inverted stereochemistry at a single position (carbon 20) that cannot bind[257] to the CRD and cannot activate Hh signalling[258], failed to block the interaction (Fig. 3.19, panel d).

I next tested the effect of methyl-β-cyclodextrin (MβCD)-mediated cholesterol depletion on the stability of purified SMOΔC-BRIL-V329F using a SEC-based thermostability assay[143], reasoning that removal of the cholesterol from its central position between the three domains may destabilise the purified protein. In the ab-

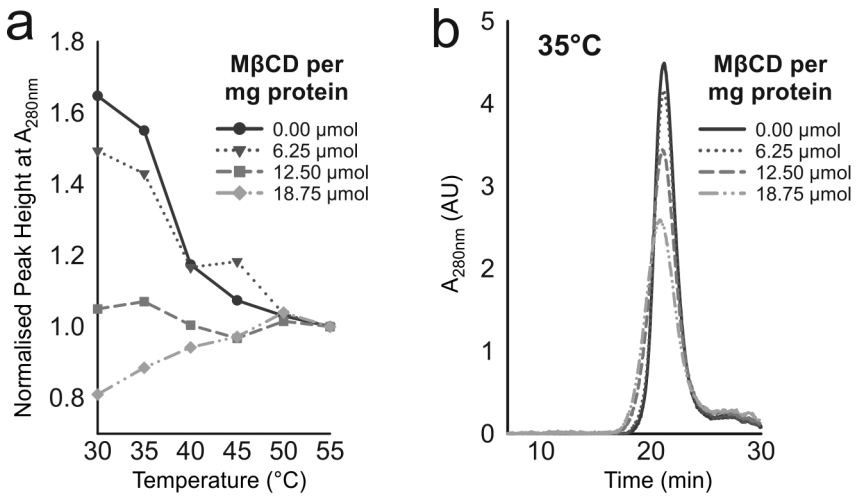


Figure 3.20: Disruption of cholesterol interactions affects SMO stability. a, Compiled peak heights from thermostability SEC analysis of purified SMOΔC-BRIL after treatment with different MβCD concentrations. **b,** Example of raw SEC data used for the analysis in a. Samples were incubated at 35°C for 1 h prior to loading onto the SEC column.

sence of MβCD, the maximum height of the main protein peak showed a decrease over the temperature range 30-55°C (solid line in Fig. 3.20). As more MβCD was added, the maximum peak intensity at the lower temperatures (30-45°C) was reduced in a dose-dependent manner (broken lines in Fig. 3.20), demonstrating that cholesterol depletion is associated with destabilisation of SMOΔC-BRIL-V329F. This result is consistent with the hypothesis that the cholesterol molecule bound to the CRD plays an important role in stabilising isolated SMOΔC-BRIL-V329F at physiologically relevant temperatures. However, I cannot exclude the possibility that the destabilising effect of MβCD is due to extraction of TMD-associated cholesterol molecules, not apparent in our structure but seen in other GPCR structures[321, 61, 397] and recent simulations[194]. Cellular cholesterol levels play a permissive role in Hh signalling[78, 26]; however, this requirement is likely unrelated to the cholesterol binding site seen in our structure since SMO mutants lacking the CRD[256] or carrying mutations in the cholesterol binding groove[26] remain sensitive to cholesterol depletion.

To assess the functional relevance of the cholesterol binding site identified in

our structure, we focused our mutagenesis efforts on the hydrogen-bonding network between the 3β-hydroxyl group of cholesterol and the CRD residues Asp95 and Tyr130 (seen in Fig. 3.18, panel a, and corresponding mouse residues in Fig. 3.21, panel a). The Rohatgi Lab mutated the corresponding residues in mouse SMO, Asp99 and Tyr134, either individually or in combination, to residues that lack hydrogen bond acceptor or donor groups (Asp99Ala and Tyr134Phe). Since cholesterol and oxysterols occupy the same groove in the CRD (Fig. 3.18, panel a), they likely adopt a similar conformation and make at least some of the same interactions. Therefore, 20(S)-OHC binding and signalling was used to evaluate the effect of these mutations. First, the ability of 20(S)-OHC beads to capture SMO from detergent extracts was tested (Fig. 3.21, panels b & c). As previously demonstrated[257], wild-type (WT) SMO can be captured on 20(S)-OHC beads. However, SMO variants carrying the Asp99Ala and Tyr134Phe point mutations failed to bind to 20(S)-OHC beads, highlighting the importance of this hydrogen bonding network for sterol binding (Fig. 3.21, panel c). Next, the Rohatgi Lab stably expressed untagged versions of these SMO mutants in *Smo-/-* fibroblasts (Fig. 3.21, panel b) and assessed their abilities to restore signalling initiated by Shh, the TMD-agonist SAG or the CRD-agonist 20(S)-OHC (Fig. 3.21, panel d) as measured by activation of transcription of *Gli1*, a direct Hh target gene, using quantitative real-time reverse transcription PCR (qRT-PCR). The two mutants did not significantly change the basal activity of SMO in the absence of Hh agonists and, importantly, did not change the signalling response to SAG (Fig. 3.21, panel d). This demonstrates that these mutants retain an intact TMD ligand-binding site and remain competent to transmit signals to cytoplasmic components. However, Asp99Ala and Tyr134Phe did substantially impair the ability of SMO to respond to both 20(S)-OHC and Shh (Fig. 3.21, panel d). These residues, and by implication the cholesterol seen in our structure, appear to be critical for signal-induced SMO activation.

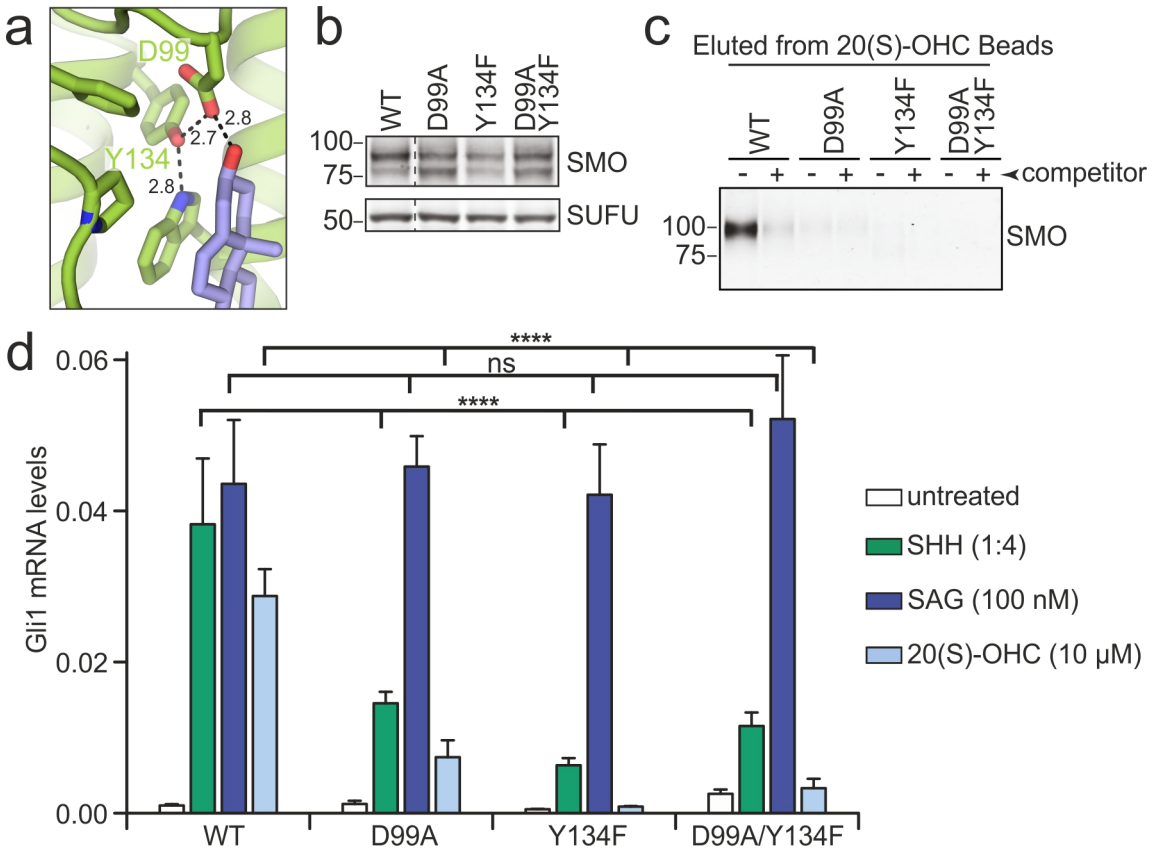


Figure 3.21: Point mutations in the cholesterol binding site disrupt bead-binding and signalling. a, Hydrogen-bonding network within cholesterol binding site (residues numbered according to mouse SMO). Point mutations Asp99Ala and Tyr134Phe are predicted to disrupt hydrogen bonding with cholesterol. **b,** Protein levels of SMO variants stably expressed in *Smo-/-* mouse fibroblasts. **c,** Binding of SMO variants to 20(S)-OHC-coupled beads in the absence or presence of free 20(S)-OHC as competitor (WT, wildtype). **d,** Levels of *Gli1* mRNA (mean +/- SD, n=4) were used as a metric for Hh signalling activity in cell lines shown in b after stimulation with SHH, SAG or 20(S)-OHC. Statistical significance is denoted for the difference in *Gli1* mRNA levels between cells expressing SMO-WT compared to each mutant SMO protein. These experiments were performed in collaboration with the Rohatgi Lab.

3.3.11 Cholesterol is an activator of SMO

While testing the effect of a panel of sterol lipids on Hh signalling in cultured fibroblasts, the Rohatgi Lab observed that cholesterol was able to induce the transcription of Hh target genes. Given the poor solubility of cholesterol in aqueous buffers, it was solubilised in MβCD inclusion complexes. These complexes are able to enrich cell membranes with their hydrophobic cargo and increase shuttling between lipid compartments[191, 70, 69]. Using the same cellular assay detailed above, MβCD:cholesterol was able to activate transcription of *Gli1*, whereas MβCD alone was not (Fig. 3.22, panel a). Furthermore, the transcriptional response to MβCD:cholesterol treatment was on the same time-scale as that of SHH, although it did not reach the same magnitude (Fig. 3.22, panel b).

Next, the Rohatgi Lab tested the ability of the CRD binding site double mutant, Asp99Ala/Tyr134Phe, to respond to various Hh activators. The mutations reduced the ability of MβCD:cholesterol to activate Hh signalling (Fig. 3.22, panel d, leftmost graph). The Asp99Ala/Tyr134Phe SMO mutant was also impaired in its responsiveness to Shh and to 20(S)-OHC (middle graphs), but remained responsive to the TMD-targeting ligand SAG (rightmost graph). A complete deletion of the CRD (SMO-ΔCRD), which increased basal SMO signalling activity like the Asp477Arg mutation, also abolished signalling responses to MβCD:cholesterol (Fig. 3.22, panel c)[256, 263]. This mutational analysis supports the model that the CRD binding-site, rather than the TMD binding-site, mediates the effect of cholesterol on SMO activity and thus on Hh signalling.

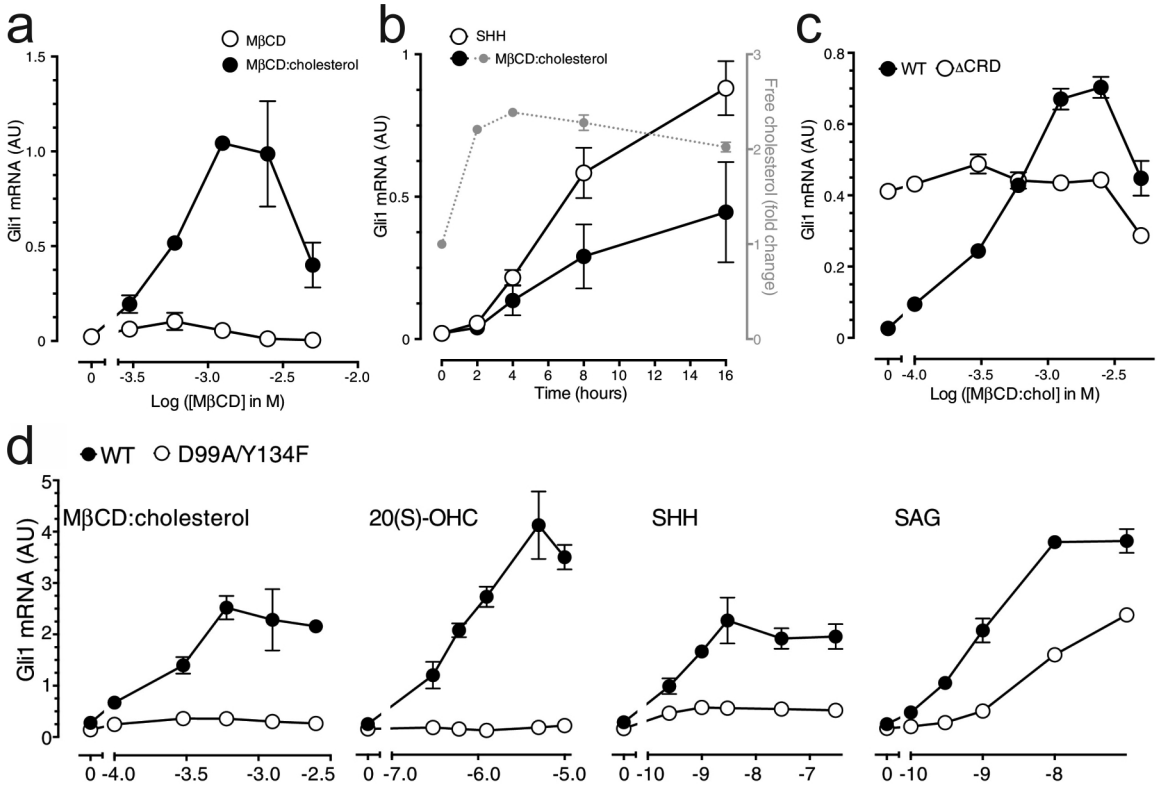


Figure 3.22: Cholesterol activates Hh signalling via SMO CRD binding site. **a,** *Gli1* mRNA was measured by quantitative real-time reverse-transcription PCR (qRT-PCR) and normalised to mRNA levels of the housekeeping gene *GAPDH* after treatment (12 h) with various doses of naked MβCD or a saturated MβCD:cholesterol (8.8:1 molar ratio) complex. In both cases, the concentration of MβCD is plotted on the y-axis. **b,** Time course of *Gli1* induction (left y-axis) after treatment with Shh (265 nM) or the MβCD:cholesterol complex (2.5 mM). The gray circles (right y-axis) show the kinetics of increase in unesterified cholesterol (normalised to total protein) after the addition of MβCD:cholesterol. **c,** Dose-response curves for MβCD:cholesterol in *Smo-/-* cells stably expressing WT SMO (solid black circles) or SMO CRD, an activating N-terminal truncation mutant that lacks the entire CRD.**d,** Dose-response curves for the indicated agonists in *Smo-/-* cells stably expressing WT SMO (always solid black circles) or the Asp99Ala/Tyr134Phe double mutant SMO (open circles). All agonists were applied to cells for 12 h. Values on the y-axis represent Log ([Agonist] in M) and the ordinate for all four graphs is only shown once at the left. In all graphs, circles depict mean values from 3 replicates and error bars show ±SD. These experiments were performed in collaboration with the Rohatgi Lab.

3.3.12 Cholesterol stabilises CRD in MD simulations

In order to assess the structural impact of the CRD in a membrane environment, the Sansom Lab carried out molecular dynamics simulations of SMO in the presence and absence of cholesterol, with our SMO CRD-LD-TMD structure as the starting model. In total 10 independent 100 ns atomistic simulations were performed with SMO embedded in a phosphatidylcholine bilayer (Fig. 3.23, panel a). These simulations revealed that the SMO CRD has significant conformational flexibility in the unliganded state (Fig. 3.23, panel b). In the presence of cholesterol, however, there is a pronounced decrease in this flexibility (Fig. 3.23, panel b, left graph), consistent with a conformationally stabilising effect on the CRD structure. In contrast, cholesterol did not significantly change the conformational stability of the SMO transmembrane region (Fig. 3.23, panel b, right graph). The predominant cholesterol-induced stabilisation was seen in the vicinity of the sterol-binding pocket, with some propagation to more distal CRD regions (Fig. 3.23, panels c, d & e).

The molecular dynamics results prompted us to test the signalling consequences of destabilising interactions within the extracellular region of SMO. To probe the effect of interactions between the CRD and the LD, the Rohatgi Lab generated two mutants, Pro120Ser and Ile160Asn/Glu162Thr, in the context of mouse SMO (Fig. 3.24, panel a). These introduce bulky, N-linked glycosylation sites at positions 156 and 114 of the human CRD that contact the LD (Fig. 3.24, panel a). To disrupt the conformation of the LD itself, two cysteines (Cys197 and Cys217 in mouse SMO) that form a disulfide bond within the LD (numbered "6" in Fig. 3.15 and Fig. 3.24, panel a) were mutated to serine. These mutants were compared to SMO lacking the entire CRD (SMO-ΔCRD) (Fig. 3.24, panels b, c & d). Untagged SMO-ΔCRD stably expressed in *Smo-/-* fibroblasts demonstrated a level of constitutive activity[263, 256] comparable to WT SMO stimulated with saturating Shh, but lower

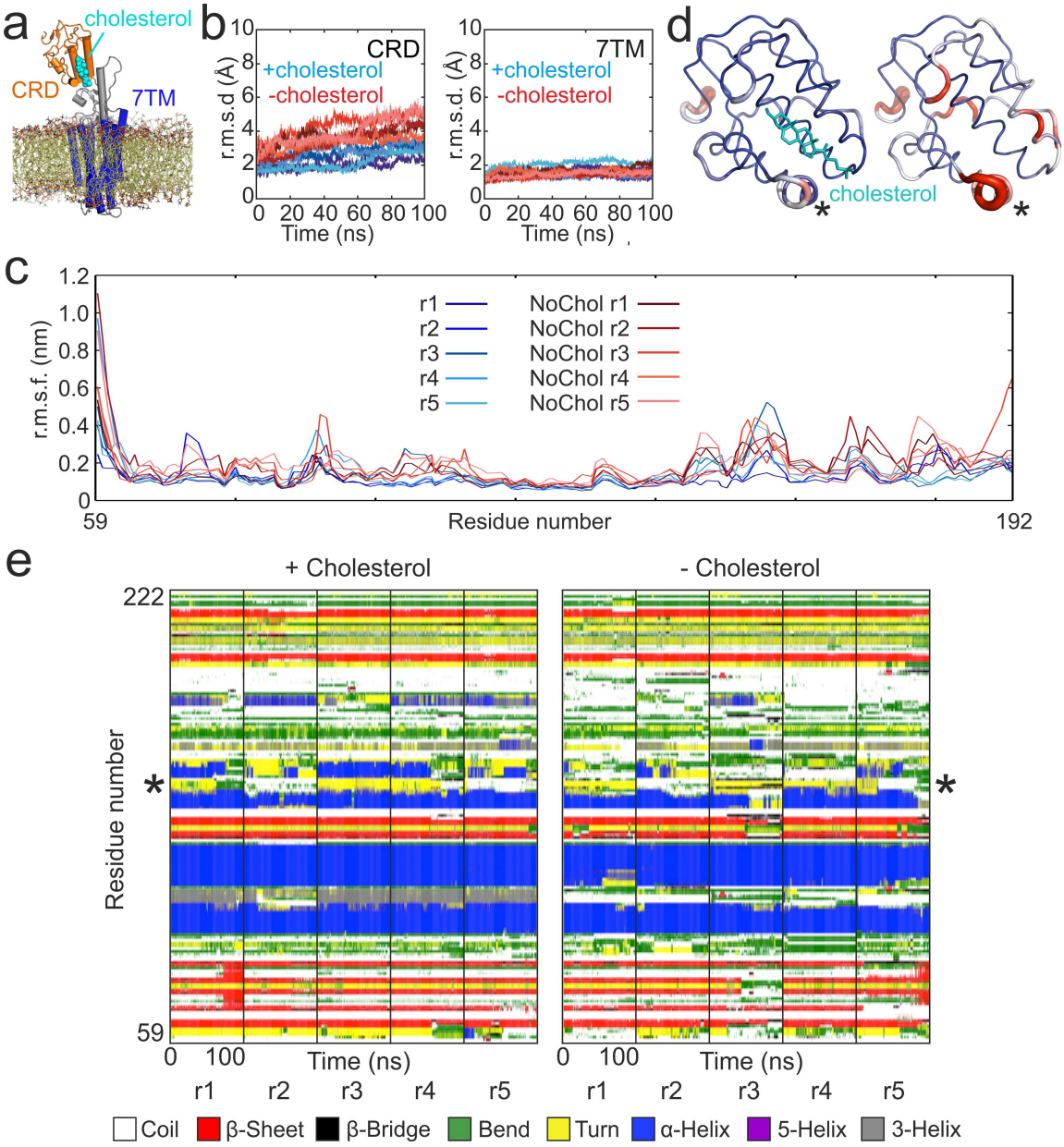


Figure 3.23: Molecular dynamics simulations of SMO in a lipid bilayer. a, SMO embedded in a lipid bilayer with the CRD in orange, the 7-pass transmembrane region excluding intra- and extracellular loops (7TM) in blue and cholesterol in cyan. **b,** Cα r.m.s.d. versus time for CRD (left) and 7TM region (right) with (blues) and without (reds) cholesterol. **c & d,** Relative r.m.s. fluctuations of the Cα atoms over the course of 5 x 100 ns of atomistic molecular dynamics simulation in the presence (blues) and absence (reds) of cholesterol. The structures in d are shown as putty representations coloured from high conformational stability (i.e low r.m.s. fluctuations; blue/thin) to low stability (i.e high r.m.s. fluctuations; red/thick). **e,** Secondary structure DSSP matrices for each of the simulations. The asterisks in d and e all mark the helix spanning from residues 155-160 that is destabilised in the absence of bound cholesterol. These experiments were performed in collaboration with the Sansom Lab.

than that of the oncogenic SMO-M2 mutant[403] (Fig. 3.24, panel d).

The Pro120Ser, Ile160Asn/Glu162Thr and Cys197Ser/Cys217Ser mutations mimicked the effect of a complete CRD deletion - constitutive signalling activity was increased despite the presence of high levels of Ptc1 and there was a complete loss of responsiveness to 20(S)-OHC (Fig. 3.24, panel d). All three mutants remained sensitive to inhibition by SANT1, which binds deep in the TMD and does not contact the LD (Fig. 3.24, panel d, straw-coloured bars)[378]. Cyclopamine and SAG, which both make contacts with the LD[379, 378], could regulate SMO-Pro120Ser and SMO-Ile160Asn/Glu162Thr, but not SMO-Cys197Ser/Cys217Ser. Thus, destabilising either CRD-LD contacts or the LD itself increases the constitutive activity of SMO, implicating these regions in stabilising an inactive state, analogous to the D[E]R$^{3.50}$Y motif in helix III of some Class A GPCRs[370].

To understand the basis for the loss in oxysterol responsiveness, the binding of each of the mutants to 20(S)-OHC beads was measured (Fig. 3.24, panel c). While the CRD mutations Pro120Ser and Ile160Asn/Glu162Thr impaired binding, the Cys197Ser/Cys217Ser mutation in the LD had no effect on binding to 20(S)-OHC. The observation that 20(S)-OHC cannot activate SMO-Cys197Ser/Cys217Ser (Fig. 3.24, panel d) even though it can bind normally (Fig. 3.24, panel c) suggests that the LD may transmit the conformational changes that lead to SMO-TMD activation in response to CRD ligands.

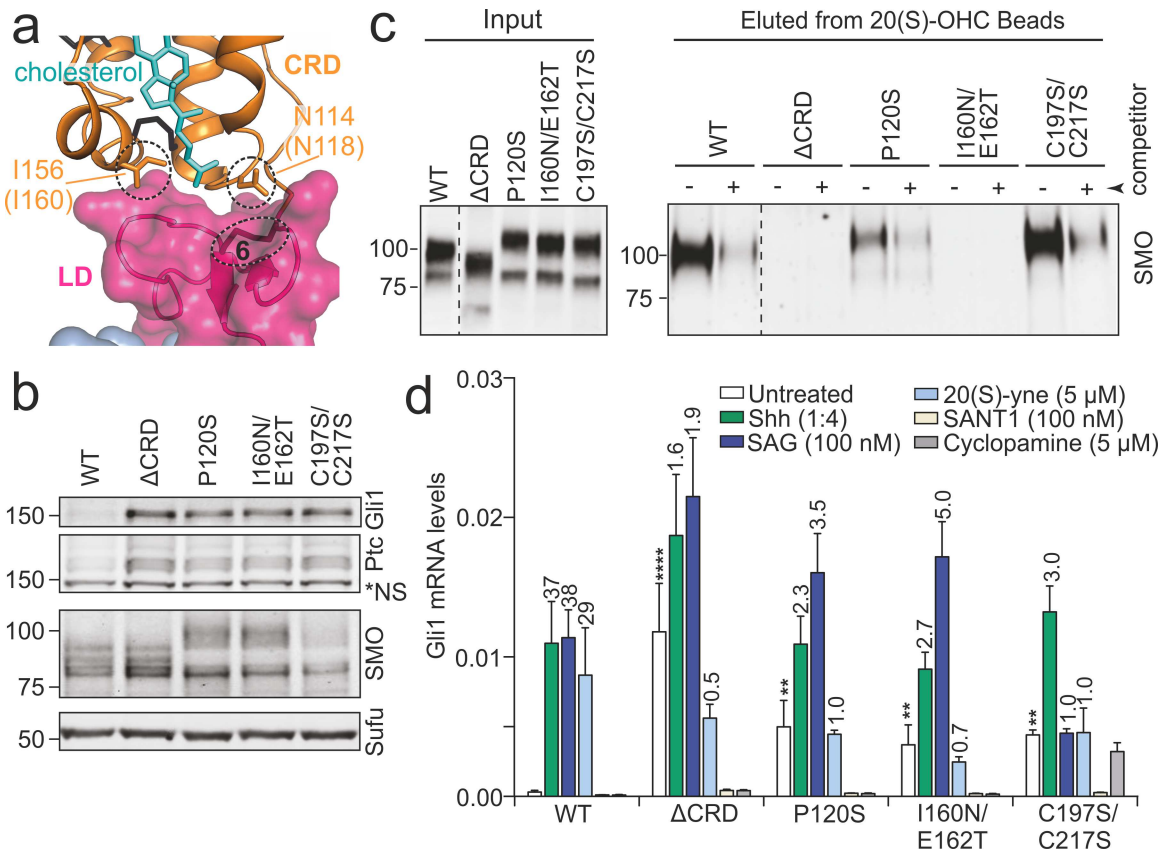


Figure 3.24: Domain interface mutations disrupt signalling ability. **a,** CRD-LD interface with mutated residues highlighted. Corresponding mouse residues are in parentheses. **b,** Protein levels of SMO and also of Ptc1 and Gli1 (each of which is encoded by a direct Hh target gene) measured by immunoblot from *Smo-/-* mouse fibroblasts stably expressing one of five SMO variants: wild-type SMO (WT); SMO lacking the entire CRD (ΔCRD); SMO with two mutations (Pro120Ser or Ile160Asn/Glu162Thr) that introduce glycosylation sites in the LD-CRD interface; and, SMO lacking a conserved disulphide bond (Cys197Ser/Cys217Ser, marked "6" in **a**) in the LD. Elevated levels of Gli1 and Ptc1 reflect high constitutive signalling activity of each mutant. NS: a non-specific band detected by anti-Ptc1 antibody; Sufu: loading control. Different patterns seen in SMO panel is caused by different numbers of N-linked glycosylation sites. **c,** Oxysterol-binding capacity of each SMO variant was determined (right blot) by its ability to bind to 20(S)-OHC beads in the absence or presence of 50 μM free 20(S)-OHC (competitor). Inputs for each binding reaction are shown on the left. **d,** *Gli1* mRNA levels (mean +/- SD, n=4) after treatment with agonists (Shh, SAG, 20(S)-yne) or antagonists (cyclopamine, SANT1). *Gli1* levels in untreated cells reflect constitutive activity of SMO. Numbers above bars indicate *Gli1* fold-increase compared to untreated cells. Asterisks denote statistical significance for comparison with WT SMO. These experiments were performed in collaboration with the Rohatgi Lab.

3.4 Discussion

3.4.1 Domain arrangement suggests signal transduction mechanism between CRD and TMD

The crystal structure solved in this chapter is the first structure solved of a GPCR with a large functional extracellular domain. The arrangement of the CRD held away from the TMD by the LD and ECL3 suggests a mechanism for signal transduction via these two intermediary structural features. Mutations that interfered with LD structure or stability showed an inability to transmit signals from CRD ligands. Furthermore, the ECL3 of SMO contains a helical extension of TM helix VI, which in the active-state structures of Class A GPCRs shows substantial movement compared to their inactive states (Fig. 3.2, page 68)[404, 275, 298]. This suggests that TM helix VI may also mediate activation in SMO, with the ability to integrate structural cues from the CRD, LD and TMD. However, the details of this proposed mechanism remain to be verified and elucidated.

The positioning of the CRD binding site at the intersection of the three domains also points to a key regulatory role for CRD-targeting ligands. It is not yet apparent what features of CRD ligands differentiate agonists from antagonists. However, it seems likely that the precise geometry of the sterols' acyl tails will prove to be important, particularly given that 20(S)-OHC binds to and activates SMO, 20(R)-OHC does not bind SMO, ent-20(S)-OHC does not bind, and 22-NHC binds to and inactivates SMO (Fig. 3.25)[258, 263, 257]. With the new information provided by the crystal structure of cholesterol-bound SMO, we can hypothesise a role for the 3β-hydroxyl group and the tetracyclic sterol ring, as features that are common to all of the sterols that bind to SMO, in facilitating binding to the CRD via the hydrogen-

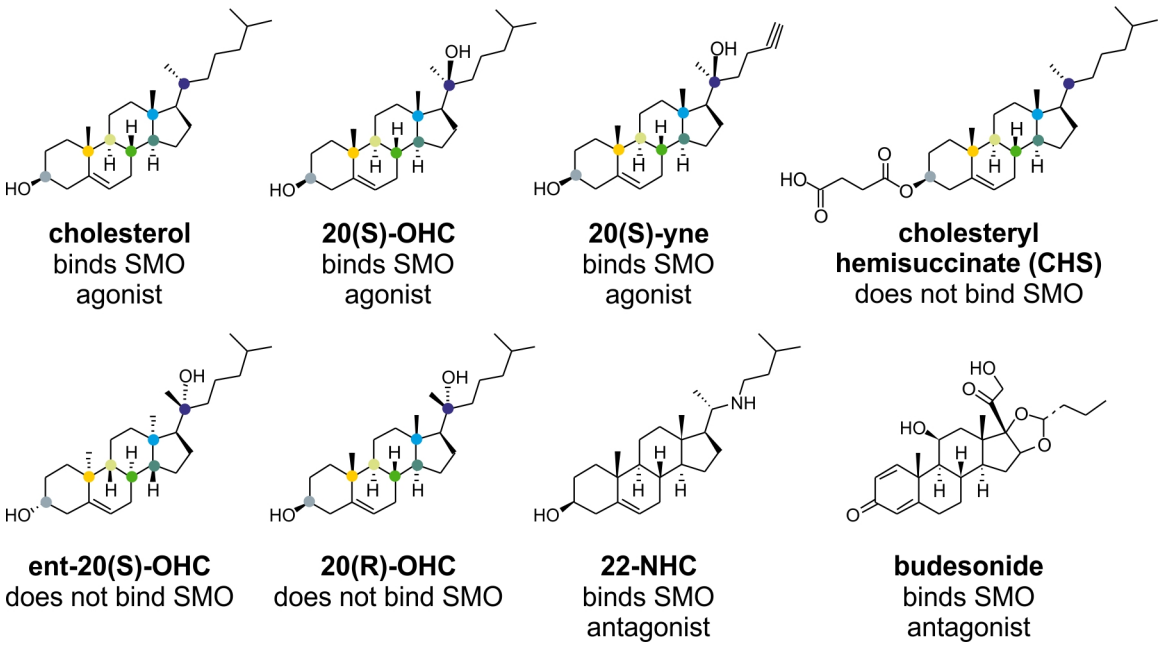


Figure 3.25: Chemical structures of sterols relevant to SMO. Small structural differences in molecules affect binding and activity with respect to SMO. Relevant chiral centres highlighted with coloured dots.

bonding network (made up of Asp95, Trp109, Tyr130) and the associated π-stacking interaction with Trp109 (Fig. 3.18, panels a & b, page 98).

The same CRD ligand binding site has been mapped for interactions with oxysterols[258], budesonide[296] and now cholesterol. The structurally homologous Frizzled 8 CRD has also been solved in this context, bound to a palmitoyl moiety that is conjugated to the Wnt 8 protein ligand (Fig. 3.3, panel c, page 70)[165]. This CRD binding site appears evolutionarily conserved to sense and bind to hydrophobic small molecules. However, while it appears that the general CRD architecture facilitates binding to many different hydrophobic small molecules, the specificity of each CRD is incredibly well-defined, as evidenced by the SMO-oxysterol work detailed above, namely that chiral switching at a single position can abolish binding (Fig. 3.25, 20(S)-OHC versus 20(R)-OHC at dark blue chiral centre). In addition, it has been shown that the Hh pathway activator 20(S)-OHC does not activate the Wnt pathway, demonstrating clear pathway specificity[257].

The SMO CRD-LD-TMD crystal structure also revealed an arrangement of the CRD and TMD that positions their two binding sites some 10 Å distant from one another. This corroborates the previous hypothesis that the two binding sites are completely separable[263, 257], each with its distinct pool of small molecule ligands. Further functional evidence points towards an allosteric interaction between the two sites[257]. However, structural evidence supporting the proposed allosteric interaction, whether synergistic or inhibitory, is yet to be provided.

3.4.2 Cholesterol: a new ligand for SMO

Cholesterol represents the first physiologically relevant small molecule agonist for SMO. Many small molecules are known to bind to and affect Hh signalling through SMO. However, none of them are good candidates for SMO's endogenous ligand. The best candidate, oxysterol 20(S)-OHC, is detectable in vertebrate tissue[220] but likely not at sufficient levels to adequately account for Hh signalling[257, 311, 227]. All other known ligands, whether agonists or antagonists, are exogenous. It is remarkable, given all of the previous work done on identifying and synthesising new SMO ligands, that cholesterol had not previously been identified as a ligand for SMO. We can now see that twin confounding factors came into play. The first was that cholesterol is very difficult to work with in aqueous environments due to its very low solubility, even in detergent solutions. This meant that more soluble alternatives were sought and used in the place of cholesterol itself. Cholesteryl hemisuccinate (CHS), now a commonplace supplement in membrane protein detergent buffers, contains a succinate moiety that makes it more soluble than cholesterol but would sterically hinder its entry into the SMO CRD binding site and, crucially, would prevent the 3β-hydroxyl group from forming a hydrogen-bond with Asp95, a key residue in the SMO-cholesterol interaction (see Fig. 3.25 for structural comparison of cholesterol and CHS, and Fig.

3.18 for the CRD binding site).

The second confounding factor was that cholesterol is ubiquitous in vertebrate cells, performing many essential functions. This means that influencing the bulk supply of cholesterol, for instance by knocking out a key enzyme in the cholesterol biosynthesis pathway within a cell, will critically disrupt a plethora of systems. These kinds of approaches are not sufficiently targeted to distinguish subtle secondary effects. Deconvoluting these overlapping roles for cholesterol within the cellular environment is likely to remain a challenge in the field for some time. Recently Blassberg *et al.* (ref.[26]) were able to differentiate the effects of reduced cholesterol levels on SMO signalling compared to that of precursor DHCR7 levels or derivative oxysterol levels in a study of the developmental disorder Smith-Lemli-Opitz syndrome. While the Rohatgi Lab and I were successfully able to handle cholesterol when it was conjugated to beads or solubilised in MβCD, effective experimental design using these reagents remains difficult.

The cholesterol present in our crystal structure may have originated from the cells in which it was expressed or from the cholesterol-enriched LCP mixture, raising the possibility that the cholesterol was an artefact of crystallisation rather than a *bona fide* ligand. Subsequent binding studies and functional experiments validated cholesterol as an agonist ligand for the SMO CRD. However, a number of questions present themselves. Firstly, assuming that cholesterol does indeed act as an endogenous ligand, by what mechanism does it come to be within the CRD binding site, some 12 Å distant from its normal reservoir of the plasma membrane? Given our evidence that cholesterol helps to stabilise SMO, perhaps cholesterol is acquired by SMO as it folds and helps to stabilise the receptor in lieu of stronger agonists or antagonists. Alternatively, perhaps the already-folded CRD scoops up cholesterol from the membrane or indeed perhaps cholesterol is handed off to SMO by another

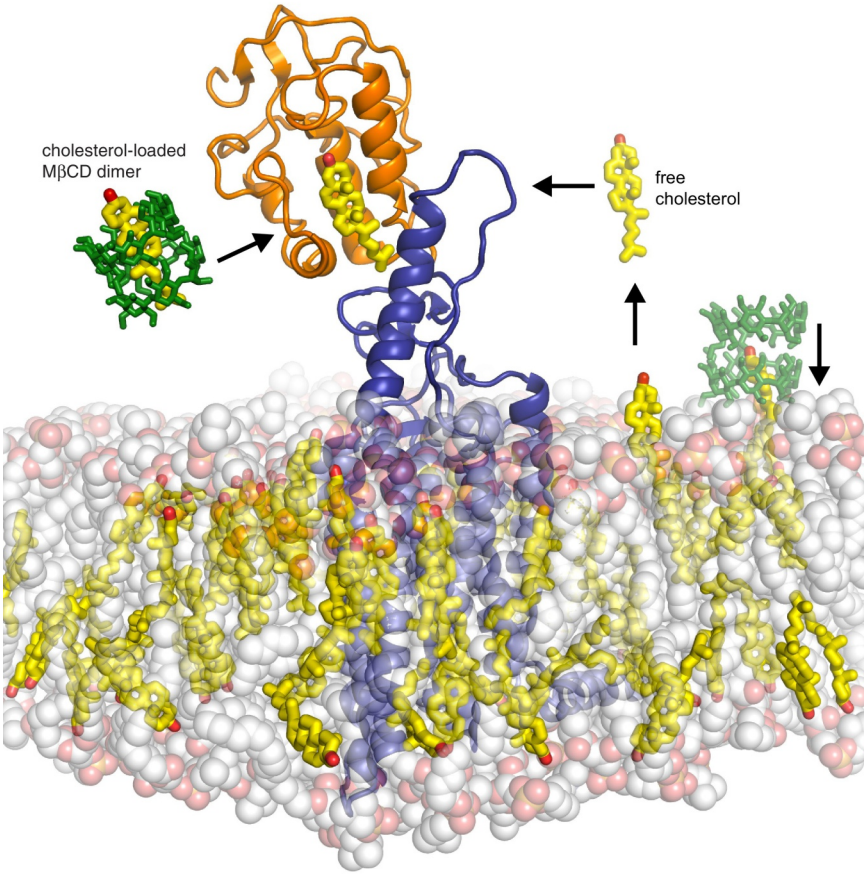


Figure 3.26: Model of cholesterol in membrane with SMO. SMO structure shown in cartoon representation with CRD in orange, LD and TMD in blue, cholesterol shown as yellow sticks with red tip for the 3β-hydroxyl group, MβCD shown as green sticks, lipid bilayer made up of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) shown as translucent spheres in grey and red.

protein, as is the case for NPC2 handing off cholesterol to NPC1 (Fig. 3.26)[205, 405]. Secondly, given the requirement of SMO to translocate to the primary cilium for full Hh signalling in vertebrates, does the lipid environment of the primary cilium display an enrichment or depletion of cholesterol relative to the bulk plasma membrane? Unfortunately, there are conflicting reports on this account depending on the method of detection employed[27, 84, 46, 251]. Further experimentation will be required to ascertain the answer to this question.

3.4.3 Evolutionary divergence between vertebrate and *Drosophila* SMO

There are many important differences between the *Drosophila* and vertebrate Hh signalling pathways[369]. It now seems likely that the role of the SMO CRD may represent yet another difference. It was previously found that neither full-length nor CRD-only dSMO could bind to oxysterols, in contrast to human, mouse and zebrafish SMO[258, 263]. Residues lining the cholesterol binding site are well conserved in vertebrates (Fig. 3.18, panel d, page 98) but less well conserved in *Drosophila* (Fig. 3.27, page 116). The superposition of our human CRD structure with a previously solved NMR solution structure of the *Drosophila* CRD (Fig. 3.18, panel e, page 98, RMSD 1.71 Å for 86 equivalent Cα positions) revealed a major rearrangement of one side of the cholesterol binding site comprising dSMO residues 183-190 (Fig. 3.27, page 116). This segment of dSMO adopts a short helix that protrudes into the CRD groove and consequently may preclude cholesterol and oxysterol binding.

Interestingly, *Drosophila* and all arthropods cannot synthesise sterols, including cholesterol, although they do require them for proper development[43, 75]. Sterol biosynthesis pathways are ancient, being present in yeast for instance, so it is likely that arthropods lost this ability at some point. In the context of SMO this begs several questions: What takes the place of cholesterol in *Drosophila* Hh signalling, if anything? Has the Hh pathway diverged in its evolution due to the loss of endogenous cholesterol biosynthesis capability? Is the utility of cholesterol to vertebrate SMO linked to the formation of primary cilia, which *Drosophila* do not possess? Indeed, the CRD of vertebrate SMO has been suggested to have a role in ciliary localisation[1].

Differences in requirements for N-linked glycosylation between *Drosophila*

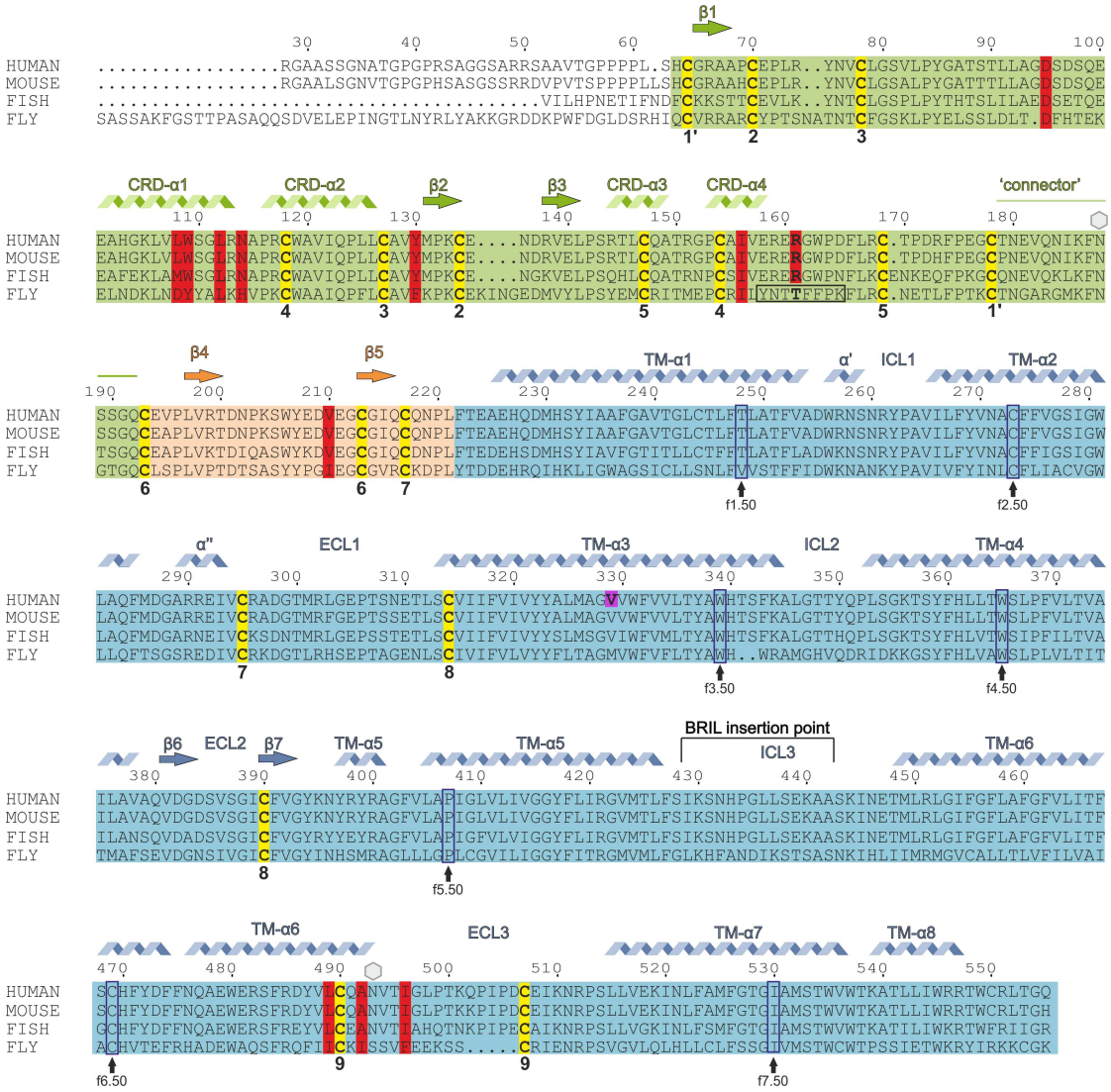


Figure 3.27: Sequence alignment of SMO orthologues with cholesterol binding site highlighted. Colours, secondary structure assignments and numbering as for Fig. 3.6 on page 76. Residues interacting with cholesterol in the CRD-LD-TMD structure are highlighted in red. N-linked glycosylation sites are depicted by a hexagon. Black box indicates residues 183-190 of the dSMO CRD that are rearranged in the NMR structure (ref.[296]) relative to the hSMO CRD-LD-TMD crystal structure.

and mouse SMO have also recently been examined[233]. Of seven N-linked glycosylation sites on dSMO, one was shown to be essential for SMO function. Of the four sites on mSMO, however, all were dispensable for function. Clarification of the differences between vertebrate and *Drosophila* SMO will require further structural work, including a structure of the CRD-LD-TMD of dSMO.

3.4.4 Identical TMD conformation in all SMO structures

Although cholesterol has now been shown to be an agonist for SMO, our crystal structure of SMO CRD-LD-TMD bound to cholesterol appears to be in an inactive conformational state. The LD and TMD in our structure align very closely to those of previously published LD-TMD structures, even when bound to an agonist (Fig. 3.17, page 97). The point mutation that was included to boost expression and stability, Val329Phe, completely abrogates SMO signalling, apparently by locking the TMD into a closed conformation. The stabilising BRIL fusion, in the place of ICL3, also likely prevents native conformational flexibility. These two strategies for stabilising SMO may need to be avoided in future in order to ascertain an active-state structure.

3.5 Summary

Extensive small scale screening of constructs and purification conditions revealed that the BRIL fusion and the inactivating Val329Phe point mutation each boosted expression levels and that SMO was most thermostable when extracted in the detergent DDM. Large scale purifications of SMOΔC-BRIL-V329F yielded sufficient protein for crystallisation trials and micro-crystals were successfully grown in LCP from protein concentrated to >25 mg/mL. Data merged from seven crystals was used to solve the structure to 3.2 Å resolution using molecular replacement. The new structure, containing the full CRD, LD and TMD of SMO, revealed a stacked domain arrangement as well as a cholesterol molecule, a new endogenous ligand for SMO, at the intersection of the three domains. Mutagenesis combined with binding and functional experiments supported the ability of cholesterol to bind to the SMO CRD and also the binding mode of cholesterol observed in the crystal structure. Molecular dynamics simulations suggested a role for the cholesterol molecule in stabilising the CRD. Mutagenesis experiments also confirmed the importance of the LD in conveying signalling information from the CRD to the TMD. Functional experiments further uncovered that cholesterol was not just a new endogenous ligand for SMO but also an agonist for the Hh pathway.

Chapter 4

Small molecule ligands can induce different conformations of SMO

4.1 Abstract

There are many small molecules known to modulate SMO in different ways. The functional state of SMO that these small molecules induce may help to elucidate its activation and inhibition mechanisms. To shed light on the SMO signal transduction mechanism, we determined crystal structures of SMO CRD-LD-TMD in complex with five different small molecule ligands. The first structure of SMO bound to the potent antagonist and anti-cancer drug, vismodegib, revealed a sizeable shift in the position of the CRD with respect to the TMD. In two of the structures, both the CRD and TMD binding sites are occupied by small molecules. In the first case by two cyclopamine molecules and in the second by a cholesterol and a SAG molecule. These are the first structures of SMO in which both binding sites are occupied. Two further structures of SMO were determined in complex with different oxysterol molecules, 20(S)-OHC and 20(S)-yne. These structures provide the structural underpinning for the functional synergy previously observed between the two binding sites. Together, they demonstrate different conformations of SMO that correspond to distinct activation states.

Some of the work presented in this chapter has been published by us in the following paper:

E. F. X. Byrne, R. Sircar, P. S. Miller, G. Hedger, G. Luchetti, S. Nachtergaele, M. D. Tully, L. Mydock-McGrane, D. F. Covey, R. P. Rambo, M. S. P. Sansom, S. Newstead, R. Rohatgi, and C. Siebold. Structural basis of smoothened regulation by its extracellular domains. *Nature*, 535(7613):517-522, 2016. (ref.[36])

Bead-binding experiments reported in this chapter were performed in collaboration with: Ria Sircar, Giovanni Luchetti, Sigrid Nachtergaele and Rajat Rohatgi ('Rohatgi Lab'). Cholesterol-conjugated beads used in this chapter were produced by

our collaborators: Laurel Mydock-McGrane and Douglas Covey ('Covey Lab'). Technical and analytical assistance for SEC-SAXS experiments reported in this chapter was provided by our collaborators: Mark Tully and Robert Rambo ('Rambo Lab'). Specific contributions are noted where relevant.

4.2 Introduction

4.2.1 SMO contains two distinct ligand binding sites

The structural and functional studies carried out in Chapter 3 highlighted the critical regulatory role played by CRD-LD-TMD contacts in controlling the conformation and activity of SMO. The hypothesis that SMO contains two physically-separable ligand binding sites was also borne out in the CRD-LD-TMD structure (Fig. 4.1). The CRD binding site is structurally homologous to the hydrophobic pockets or grooves present in the CRDs of the Frizzled 8 receptor[165], secreted Frizzled-related protein 3 (sFRP3)[86], muscle-specific receptor tyrosine kinase (MuSK)[348], Niemann-Pick C1 protein (NPC1)[205], riboflavin-binding protein (RFBP)[250] and folate receptor α (FRα)[51]. These proteins have roles in signal transduction, cholesterol trafficking and vitamin uptake. The TMD binding site on the other hand is structurally homologous to the orthosteric binding site of Class A GPCRs, such as rhodopsin[272], β_2AR[61] and A_{2a}AR[164], and functions as the primary site of receptor modulation. How these two binding sites of SMO each contribute towards the overall function of the native protein remains unclear. The TMD binding site has thus far proven amenable to therapeutic intervention[318]. Based on functional data, Nachtergaele *et al.* suggested an allosteric interaction between the two sites[257]. However, the structural basis for this interaction remains to be properly elucidated.

Solving a crystal structure of SMO that contains both binding sites provides a unique opportunity to study the structural effects on SMO of small molecules with different interaction characteristics (Fig. 4.1). In the CRD-LD-TMD structure presented in Chapter 3, only the CRD binding site was confirmed to have been occupied by a ligand, cholesterol. Is it possible for both binding sites to be occupied at the same time? Given the allostery previously observed between certain ligands in

functional assays[257], this seems highly likely. If it is possible, how do the binding sites interact and each exert their effect? In addition, given the limited characterisation of cholesterol, being a newly identified ligand for SMO, what conformational state of SMO does the cholesterol structure correspond to? It would be informative to solve CRD-LD-TMD structures with confirmed agonists and also with confirmed antagonists in order to address these questions.

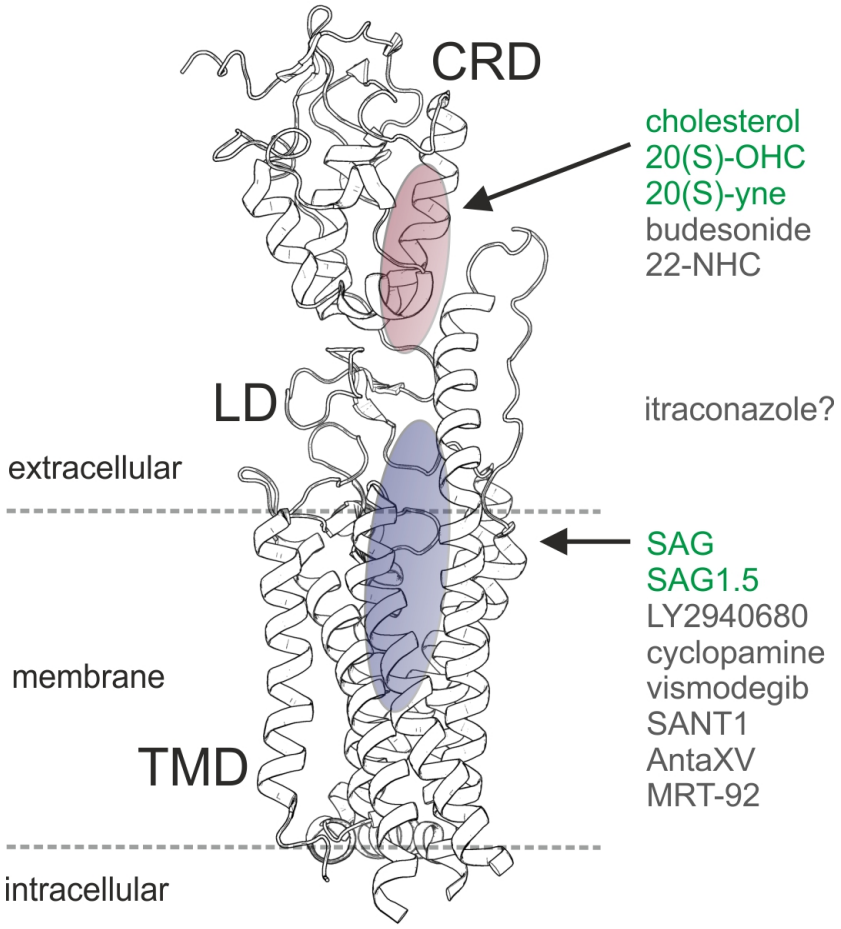


Figure 4.1: SMO CRD-LD-TMD structure with two distinct binding sites. CRD binding site in red, TMD binding site in blue. Ligands for each binding site are listed (not exhaustive) with agonists in green and antagonists in grey. Dotted horizontal lines represent approximate plasma membrane boundaries.

4.2.2 Antagonist-bound SMO

Research into small molecule antagonists of SMO has a long history, beginning even before the discovery of the Hh pathway. The discovery of cyclopamine in 1968 came out of research into the cause of faciocrancial deformations in newborn lambs that were first reported by Basque sheep herders in Idaho in the 1950s. Several studies uncovered that the corn lily (*Veratrum californicum*), when consumed at a particular stage of pregnancy, caused the developmental abnormalities. The causative agent was subsequently isolated from the corn lily and named cyclopamine[178]. It was not until the 1990s, however, that the cyclopic phenotype caused by Shh mutation in the mouse suggested a link between cyclopamine and the Hh pathway[65]. The molecular target of cyclopamine was subsequently identified as SMO[77, 352, 52].

The role of Hh signalling in various cancers has been well established (e.g. [403, 138, 137, 168, 123, 284, 302, 293]). The ability of cyclopamine to reverse the effects of oncogenic mutations in SMO and Ptc1 spurred research into developing further pathway modulators, primarily antagonists like cyclopamine[352]. This research has so far led to two SMO-targeting small molecule inhibitors, vismodegib (GDC-0449, ref.[375]) and sonidegib[44], being approved by the US Food and Drug Administration (FDA) for treatment of advanced basal cell carcinoma (BCC) (Fig. 1.7, page 21). However, spontaneous emergence of drug-resistant tumours has already been observed in a number of cases [320, 407], reinforcing the need for greater understanding of the mode of inhibition of these drugs and for better therapeutics.

To that end, the interaction of cyclopamine with SMO has been thoroughly characterised[155, 52, 313] and a crystal structure solved of the SMO LD-TMD bound to cyclopamine (Fig. 4.2)[388]. The structure of SMO LD-TMD has also been solved in complex with three other TMD-targeting antagonists, namely LY2940680, Anta

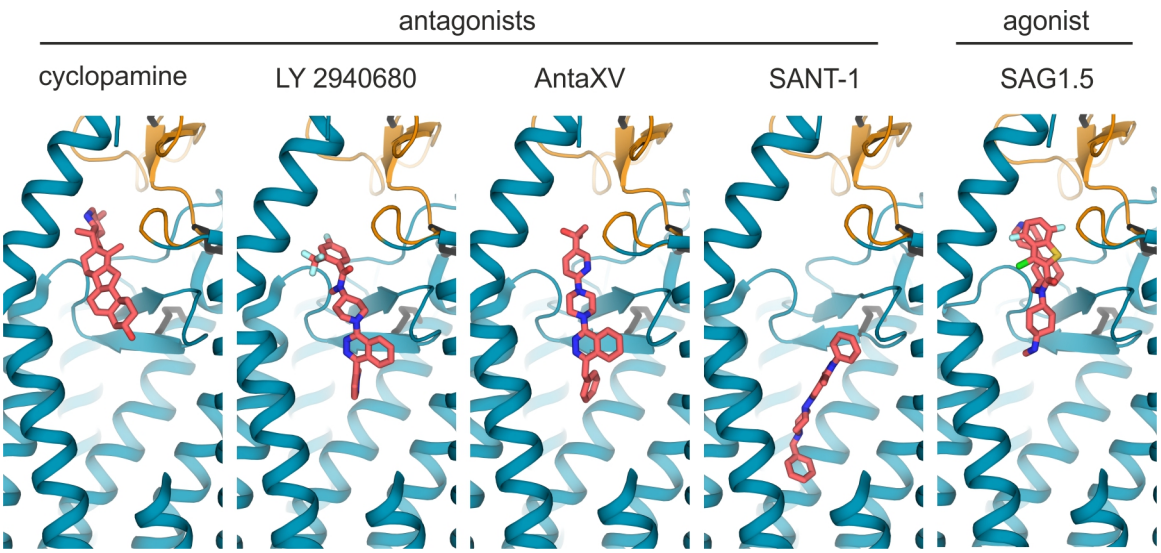


Figure 4.2: Binding modes of SMO ligands from LD-TMD structures. TMD binding site for four SMO antagonists and one SMO agonist. Colours follow Fig. 3.1. Part of TM helix VII has been removed for clarity. Images are all taken from the same orientation. (PDB: 4OR9, 4QIM, 4QIN, 4JKV, 4N4W)

XV and SANT1[379, 378]. Aside from a different binding mode for SANT1, there was very little difference between these four structures (Fig. 4.2 and Fig. 3.17, page 97, for structural alignment). These structures did not include the CRD and as such it is not clear whether these antagonists alter the interactions between the CRD and LD-TMD, precluding an assessment of how the extracellular and transmembrane domains of SMO communicate.

Efforts to find SMO inhibitors that are not sensitive to oncogenic mutations have led to the development of several next-generation small molecule inhibitors, such as compounds MRT-92 (ref.[146]) and A8 (ref.[381]), that combine characteristics of various earlier compounds. Another strategy is to target the CRD binding site instead of or in addition to the TMD binding site. Budesonide[296, 383] and 22-NHC[263] represent two SMO antagonists that target the CRD binding site. The binding position of budesonide has been determined by chemical shift perturbation analysis mapped onto a solution NMR structure of the dSMO CRD[296] (Fig. 3.3 panel b, page 70). In addition, there are a few compounds that bind to SMO but do

not compete for either the CRD or TMD binding site, suggesting yet another small molecule binding site. Itraconazole, a commercially available anti-fungal agent[190], is an example of such a compound (for chemical structures of ligands, see Appendix C, page 228).

4.2.3 Agonist-bound SMO

Since the discovery of cyclopamine, many SMO-targeting small molecules have been identified (e.g. [53, 120, 381, 105, 257, 208, 146]). However, no physiologically-relevant endogenous small-molecule modulators of SMO had been found, until the discovery of cholesterol, a partial agonist, presented in Chapter 3 (section 3.3.9, page 94). Of all the previously-identified molecules that interact with SMO, only one, 20(S)-OHC, also an agonist, is produced endogenously (Fig. 3.25, page 111). Despite its presence in mammalian tissue, however, it seems unlikely that 20(S)-OHC can be an endogenous effector molecule because it is not present in Hh-receiving cells at the levels that would be required for signal activation[257, 328]. This leaves open the crucial question of what small molecule or molecules modulate SMO in its native cellular environment.

There are comparatively few known agonists of SMO. Cholesterol and 20(S)-OHC are part of a group of agonist sterols that target the CRD[263, 258, 105, 326, 257, 80]. Several other agonists, such as SAG[53] and purmorphamine[400, 336], target the TMD. A SMO LD-TMD crystal structure has been solved in complex with SAG1.5, another agonist that is derived from SAG (Fig. 4.2)[378, 120]. However, relative to the antagonist-bound structures, this agonist-bound structure did not reveal any major conformational changes that are normally associated with activation of GPCRs (Fig. 3.2, page 68).

In the past, various strategies have been employed to obtain active state structures of GPCRs. An active state of opsin was obtained without the addition of an agonist[275, 324] and then with an agonist[67]. An active state of $A_{2a}AR$ was solved bound to the potent and selective agonist, UK-432097 (ref.[404]). All other instances of active state structures, of which there have been very few so far, have required the addition of a downstream effector molecule or an effector mimic in order to stabilise the active state sufficiently for structure determination[305, 298, 218, 412, 299, 297, 40]. Active conformational states of GPCRs are more flexible than inactive states and therefore are more difficult to structurally characterise[98]. Given that the CRD is essential for native SMO signal transduction and that many SMO agonists bind to the CRD[258, 257], an agonist-bound CRD-LD-TMD structure may be revealing in terms of the CRD's effect on the TMD.

4.2.4 Chapter outline

The aim of this chapter was to characterise the effect of small molecule binding on the critical interaction between the CRD, LD and TMD, particularly in order to ascertain active and inactive conformational states of SMO. Small molecules that had previously been shown to interact with SMO were screened for any stabilising effect on SMO using a thermostability assay. Cyclopamine showed a substantial thermostabilising effect. Initial crystallisation trials with cyclopamine were unsuccessful, however. After changing the method of small molecule addition, crystals of the TMD antagonist and anti-cancer drug, vismodegib, were obtained in complex with SMOΔC-BRIL-V329F, yielding a conformation that was distinct from the original cholesterol-bound structure. Crystals of SMO in complex with TMD agonist SAG and, separately, CRD agonists 20(S)-OHC and 20(S)-yne were also obtained, yielding structures with conformations very similar to that of the original cholesterol structure.

Remarkably, the SAG structure also retained the cholesterol molecule bound to the CRD. Cyclopamine crystals, when they were successfully grown, produced a structure in which each of the two binding sites contained a cyclopamine molecule. Altogether, these five additional SMO CRD-LD-TMD structures provide the first indication of at least two distinct conformational states for SMO as well as solid evidence for a signal transduction mechanism mediated through ECL3 and the LD.

4.3 Results

4.3.1 Cyclopamine thermostabilises SMO

A panel of small molecules was screened for their ability to thermostabilise the SMOΔC-BRIL-V329F construct. Initially, thermostability was assessed using analytical size exclusion chromatography. Briefly, protein was expressed at large scale and purified in a single batch. After SEC-purification, protein was aliquoted and DMSO-solubilised small molecules were added to a final concentration exceeding their EC_{50} values by at least one order of magnitude (for full details, see Appendix C, page 228). Final DMSO concentration was always 1%. The mixture was then heated to 42°C for 1 h. The samples were run on a Superose 6 3.2/300 column (GE Healthcare) on a Shimadzu system at 20°C and compared to DMSO-only controls at 4°C and 42°C. Only a single temperature for each small molecule was run initially in order to reduce the amount of protein required. The test temperature, 42°C, was selected because it was the temperature at which the intensity of SMO without additives fell to ∼70% of its initial intensity. Any compounds that thermostabilise SMO should produce a less pronounced decease of intensity at this temperature.

Of the small molecules tested, cyclopamine showed substantial stabilisation of SMO at 42°C and SAG showed some stabilisation (Fig. 4.3, panel a). The other small molecules did not show a noticeable difference from the DMSO-only control condition. Curiously, although SAG alone seemed to have an effect, adding SAG plus either of the oxysterols did not result in stabilisation. Cyclopamine and SAG were subsequently assayed against a range of temperatures between 4 and 72°C (Fig. 4.3, panel b). When peak height was plotted against temperature, the cyclopamine sample was stabilised by ∼10°C, whereas SAG did not appear to stabilise the receptor beyond the DMSO-control (Fig. 4.3, panel c). Alternatively, plotting retention time

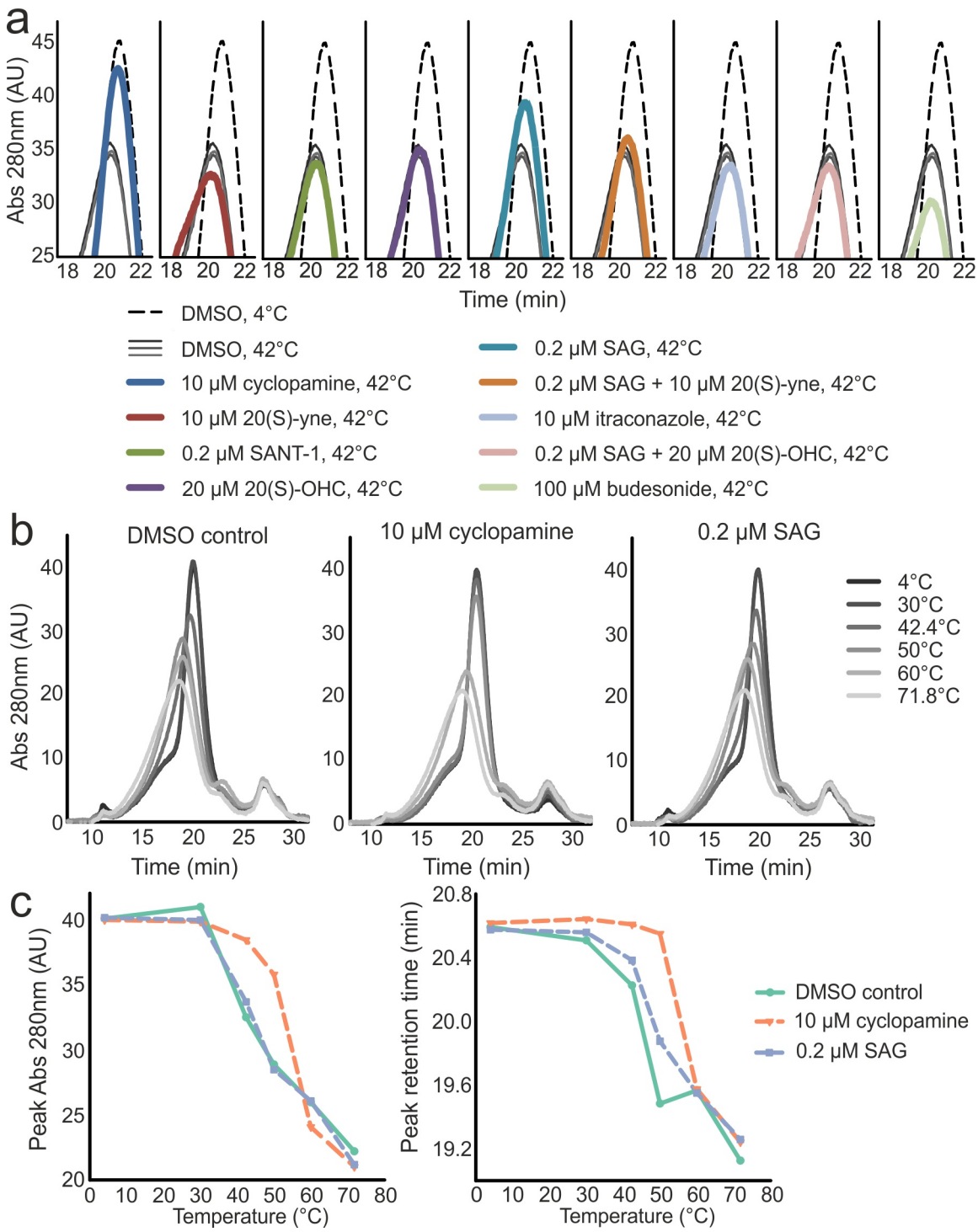


Figure 4.3: Thermostability of SMO with small molecule additives by SEC. **a,** Peaks from SEC traces of SMOΔC-BRIL-V329F with and without small molecule additives after heating to 42°C. Additives that stabilise SMO at 42°C show peaks that are higher than DMSO controls at 42°C. **b,** SEC traces of SMO with promising additives after heating for 1 h at indicated temperatures. **c,** Peak absorbance heights (left) and peak retention times (right) from **b** plotted against temperature for each additive.

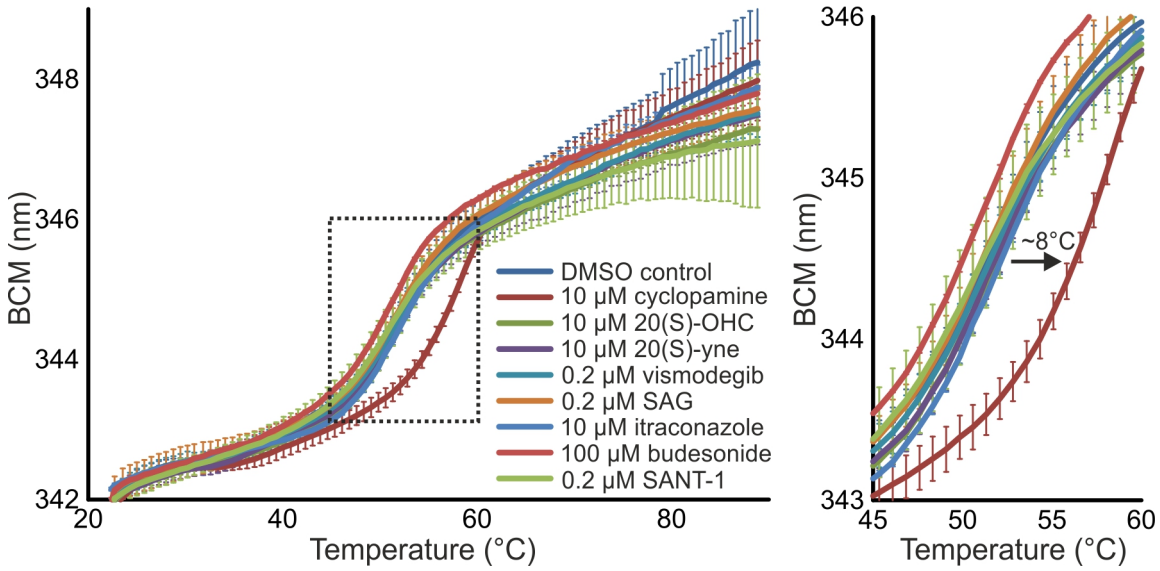


Figure 4.4: Thermostability of SMO with small molecule additives by intrinsic fluorescence (UNCLE system). Barycentric mean (BCM) of fluorescence spectrum plotted at each temperature for each condition. Dotted box indicates portion of graph shown in right-hand plot. Each condition contained 5% DMSO. Solid line drawn through mean value of 2 or 3 technical replicates, error bars represent ±1 standard deviation.

of the peak against temperature revealed that SAG partially stabilised the receptor and cyclopamine again strongly stabilised it (Fig. 4.3, panel c).

The thermostability of the panel of small molecules was also tested by monitoring changes in intrinsic fluorescence over a temperature gradient of 25-90°C using the UNCLE system (Unchained Labs) (Fig. 4.4). Briefly, purified protein from a single batch was aliquoted and DMSO-solubilised small molecules were added before loading the mixtures into the UNCLE cuvettes. The thermostabilising effect of cyclopamine on SMO was reproduced with this assay, showing a stabilisation of ∼8°C relative to the DMSO-only control (Fig. 4.4). The other compounds, including SAG, did not produce a clear effect using this method.

As a result of the promising thermostability results, cyclopamine was included in initial large-scale purifications of SMOΔC-BRIL-V329F and ΔCRD-BRIL-

SMOΔC-V329F for LCP crystallisation (∼100 μM added to bead eluate before SEC then another 20 μM added to pooled SEC fractions before final concentration). However, these initial trials were unsuccessful, producing no crystals. The first crystallisation trials to produce crystals contained no small-molecule additives (see Chapter 3, section 3.3.7, page 90).

4.3.2 Crystallisation and structures of SMO-ligand complexes

To investigate the structural basis for activation and inhibition of SMO, we solved the structure of SMO CRD-LD-TMD in complex with ligands of different properties including the TMD antagonists vismodegib and cyclopamine, the TMD agonist SAG and the CRD agonists 20(S)-OHC and 20(S)-yne (see Crystallographic Data Table, page 52). All compounds were crystallised using the SMOΔC-BRIL-V329F construct detailed in Chapter 3 (section 3.3.1, page 75). Highly concentrated small molecule ligand in DMSO stock was added to purified, EndoF1-treated protein at ∼30 mg/mL, so that the final concentration of DMSO was between 2.5–5.0% (see Materials & Methods, section 2.3.4, page 39). The small molecules generally precipitated out of solution upon addition due to the change into aqueous environment, as indicated by a thick white precipitate, but then dissolved completely again upon mixing, presumably being taken up by the protein and/or the detergent micelle. The protein-ligand solution was then mixed with lipid (10% cholesterol, 90% monoolein) and set up for crystallisation trials (see Materials & Methods, section 2.3.5, page 40).

All of the structures were solved in the C2 space group except for the vismodegib-SMO structure which was solved in P1. The vismodegib-SMO complex was solved to 3.3 Å resolution; cyclopamine-SMO to 3.7 Å; SAG-SMO to 3.0 Å; 20(S)-OHC-SMO to 3.2 Å; and 20(S)-yne-SMO to 3.9 Å (for details, see Crystal-

lographic Data Table, page 52). Pearson correlation coefficient (CC) analysis was conducted on each dataset in order to ascertain the appropriate resolution cut-off point and to validate that the model did not overfit the data (Fig. 4.5). All of the ligand structures showed a similar anti-parallel packing arrangement, with two chains in the asymmetric unit packed side-by-side in the lipid bilayer (Fig. 4.6, panels a and d). The vismodegib-SMO asymmetric unit showed a different arrangement of the two chains, reflective of the different space group for this crystal (Fig. 4.6, panels a and b). This non-physiological dimeric packing is consistent with a monomeric SMO *in vivo*. Similar to the cholesterol structure, each structure displayed a periodic disorder along one axis corresponding to every other hydrophilic layer, evidenced by consistently weaker electron density at one end of the asymmetric unit (Fig. 4.6, panels e and f, and see Materials & Methods, section 2.4.3, page 47). Within each asymmetric unit the CRD from one chain and the BRIL from the other chain were more ordered than the BRIL and CRD at the other end of the asymmetric unit.

4.3.3 Vismodegib induces a conformational change

Vismodegib is a potent TMD antagonist and anti-cancer drug. Unfortunately, resistance has already become a clinically relevant problem[407, 327, 9]. There was no structure previously available for vismodegib in complex with SMO. In the structure, vismodegib is stabilised by a network of hydrogen bonds and hydrophobic interactions formed by the LD, ECLs 2 and 3, as well as the TMD core (Figs. 4.7, panels a and c). As expected, it occupies the TMD binding site previously identified in SMO LD-TMD complexes with other antagonists and agonists, binding high up in the site similar to LY2940680, Anta XV, cyclopamine and SAG1.5 (Fig. 4.2, page 125)[379, 378, 388]. The vismodegib pyridine and chloro-benzyl rings are deeply buried, forming mainly hydrophobic interactions with the 7TM core. The amide linker is stabilised by a

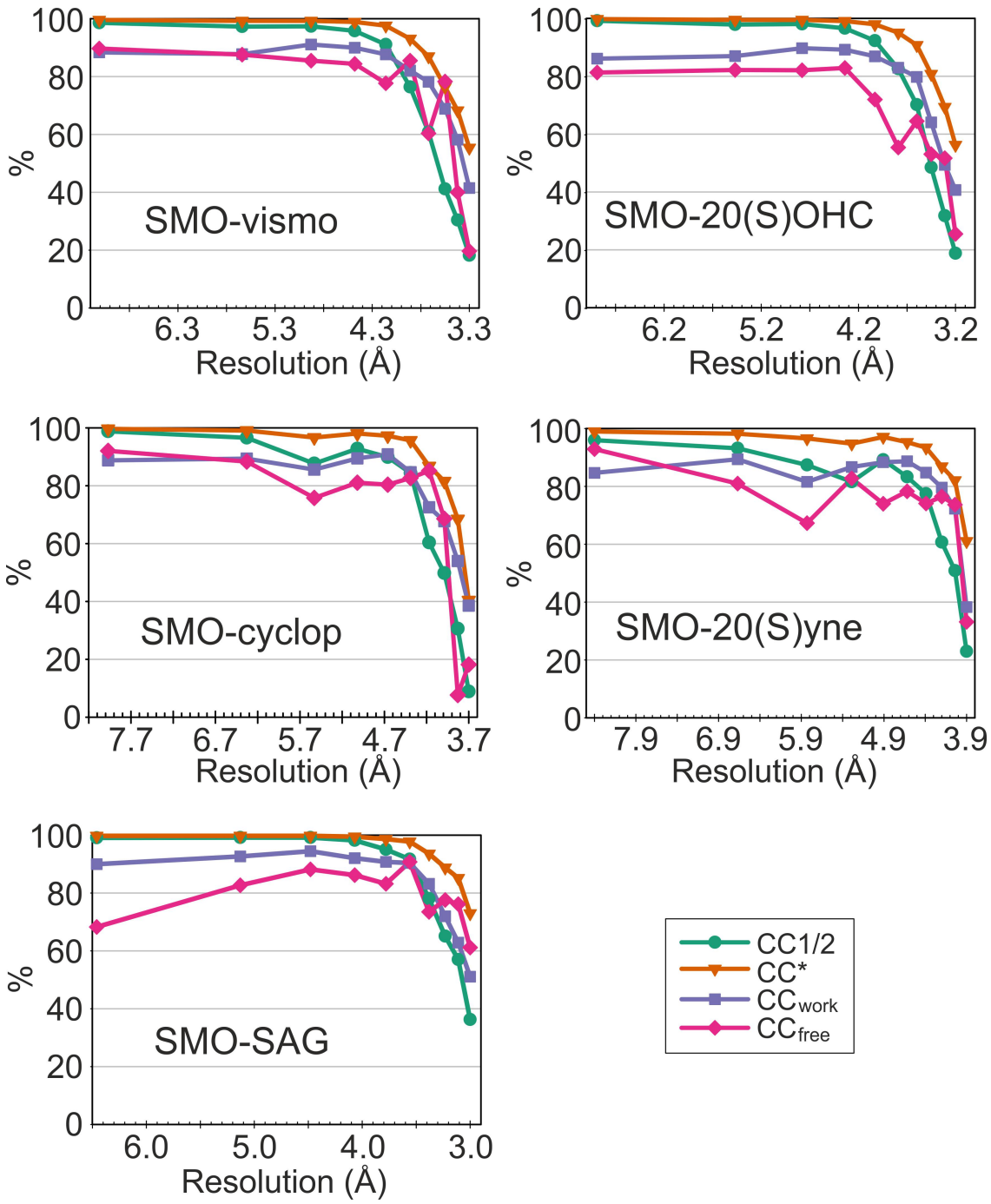


Figure 4.5: Pearson correlation coefficient (CC) analysis of the five SMO-ligand structures. Analysis carried out as in [173]. For more information, see Materials & Methods, section 2.4.4, page 50.

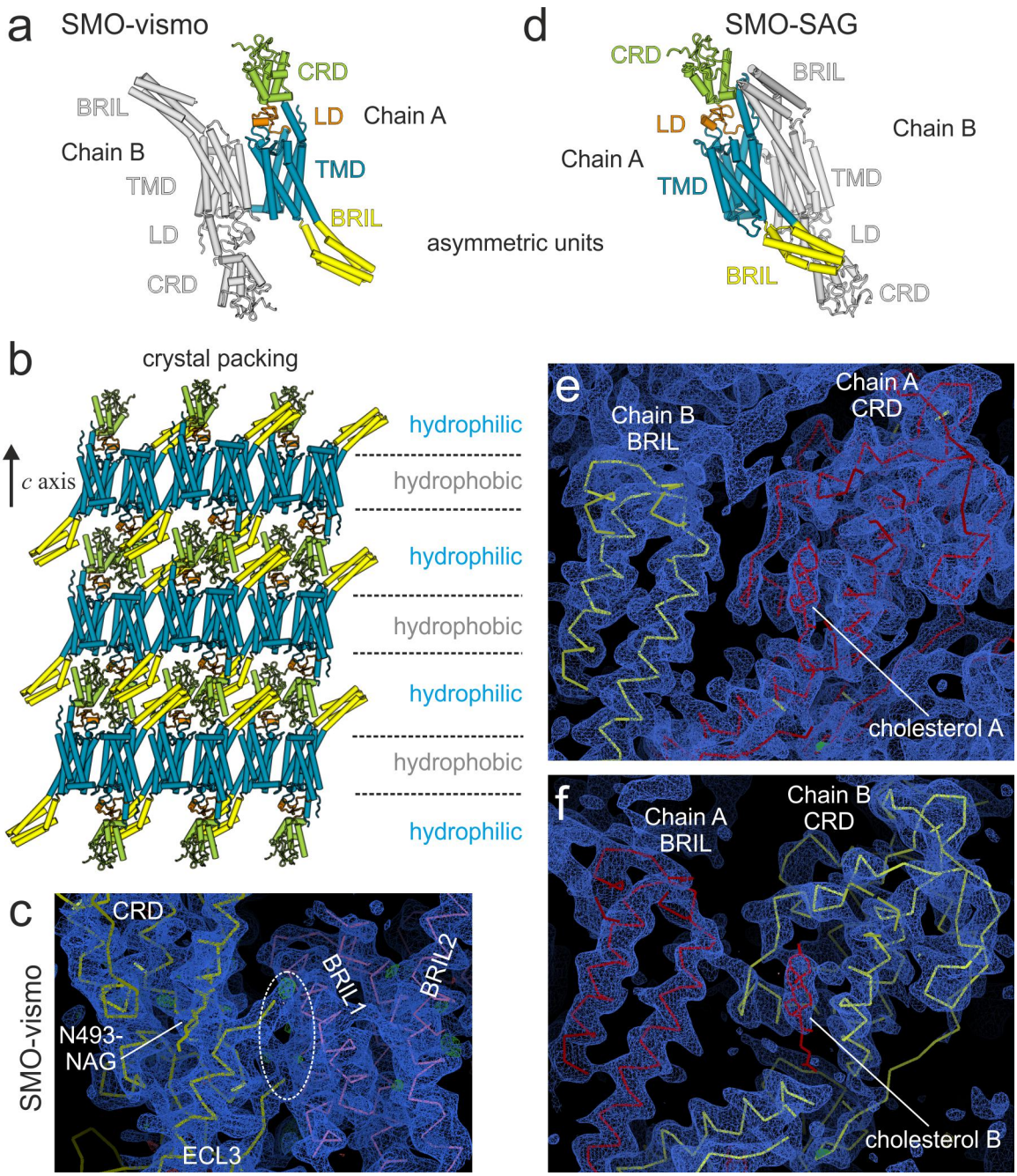


Figure 4.6: Crystal packing of SMO-ligand structures. **a,** Asymmetric unit of SMO-vismo structure. Coloured as in Fig. 3.14. **b,** Crystal packing of SMO-vismo structure. **c,** Crystal contacts near Asn493-NAG in ECL3 of SMO-vismo structure. $2F_o - F_c$ electron density map in blue chicken-wire (1.00 σ contour) and $F_o - F_c$ electron density map in green/red chicken-wire ($\pm$3.00 σ contour). Backbone of main protein chain in yellow sticks, backbone of symmetry-related molecules in pink sticks. White dotted oval indicates beginning of disordered area in ECL3. **d,** Asymmetric unit of SMO-SAG structure. **e & f,** $2F_o - F_c$ electron density map (1.00 σ contour) and backbone model for the SMO-SAG structure showing the CRD of chain A (red in panel **e**) and the CRD of chain B (yellow in panel **f**). These panels correspond to either end of the asymmetric unit. Equivalent views are shown for the sake of comparing electron density.

potential hydrogen bond to the side chain of Asp384 and occupies a central position within the 7TM bundle. The chlorophenyl-methylsulfone moiety is oriented towards the entrance of the TMD pocket and the extracellular domains beyond[378]. This arrangement is stabilised by a hydrophobic stacking interaction of the terminal methylsulfone moiety of vismodegib with the aromatic ring of Phe484 as well as by potential hydrogen bonds of the chlorophenyl to the side chains of Gln477 and Arg400 (Fig. 4.7, panel c). Asp473, a residue mutated in vismodegib-resistant cancers[407, 327, 9], stabilises the potential hydrogen-bonding network around Arg400 (marked with an asterisk in Fig. 4.7, panel c).

Our structure explains the mechanism of vismodegib resistance for several mutations in this binding site that are observed in patients with advanced BCC (Fig. 4.8)[327, 9, 320]. The Asp473His mutation seen in drug-resistant cancers would likely destabilise the intricate vismodegib binding interactions, in agreement with functional experiments showing that SMO-D473H cannot bind this drug[407]. Mutation of Gln477 to Glu leads to a loss of the potential hydrogen bond of the Gln sidechain to the chloride of the vismodegib chlorophenyl-methylsulfone moiety (Fig. 4.8, panel c). Mutation of His231 to Arg would disrupt two potential hydrogen bonds between the imidazole ring of His231 and the carbonyl main-chain atoms of residues Ser385 and Val386, which are located on a loop coordinating the interaction of Asp384 to vismodegibs amide linker (Fig. 4.8, panel d). Residue Trp281 forms a key hydrophobic interaction with the pyrimidine ring of vismodegib that is deeply buried in the SMO 7TM bundle (Fig. 4.8, panel e). Mutation of this residue to Cys would significantly destabilise this interaction while mutation of nearby Val321 to the bulkier Met would likely result in a rearrangement of the Trp281 side chain. These specific structural details of vismodegib-resistance will inform future drug development efforts.

The vismodegib-bound structure of SMO shows a marked conformational

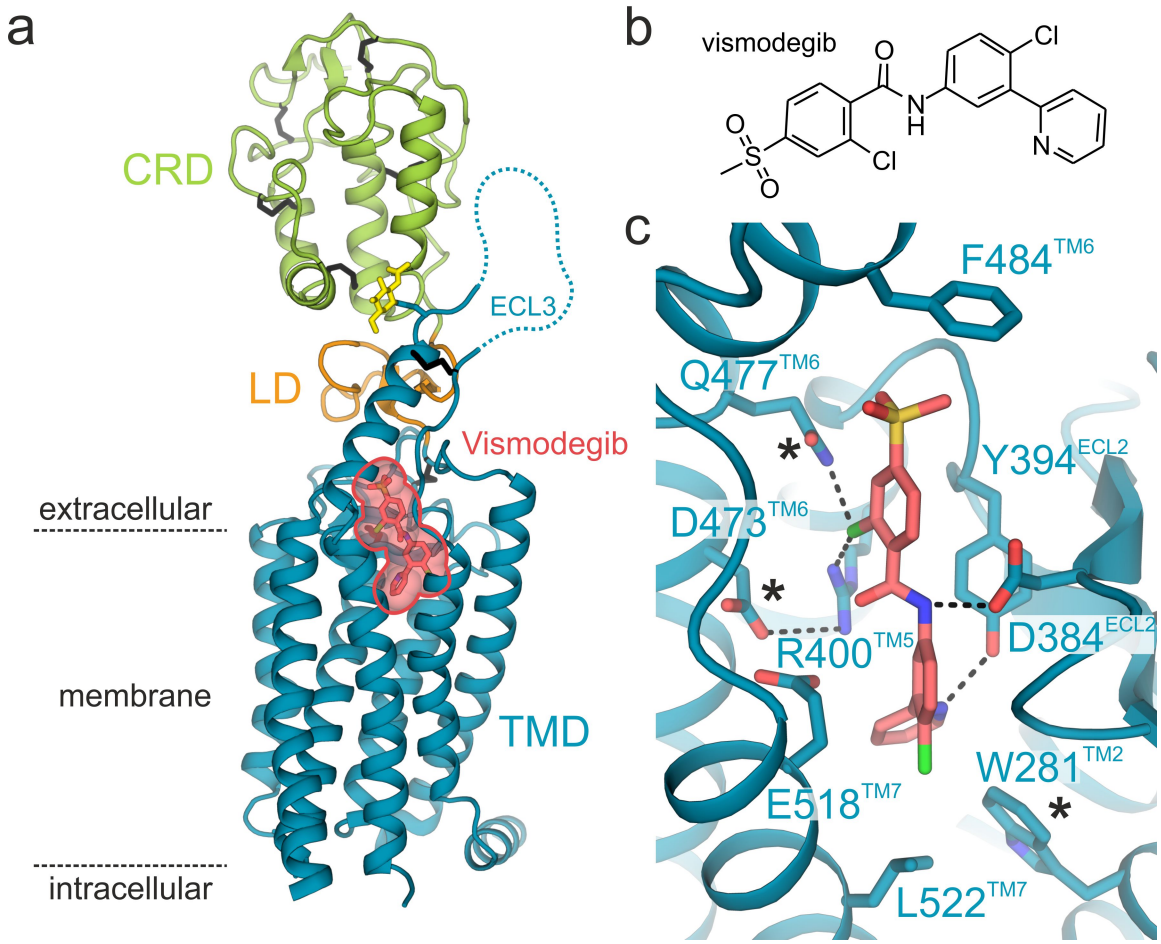


Figure 4.7: Structure of human SMO bound to vismodegib. **a,** Overall structure of vismodegib-bound SMO in cartoon representation with vismodegib in red. Other colours follow Fig. 3.15. Blue dotted line indicates position of flexible loop. **b,** Chemical structure of vismodegib. **c,** Close-up of vismodegib binding site. Important residues are shown as sticks with atomic colouring and labelled. Dotted black lines indicate potential hydrogen bonds. Asterisks denote residues mutated in vismodegib-resistant tumours.

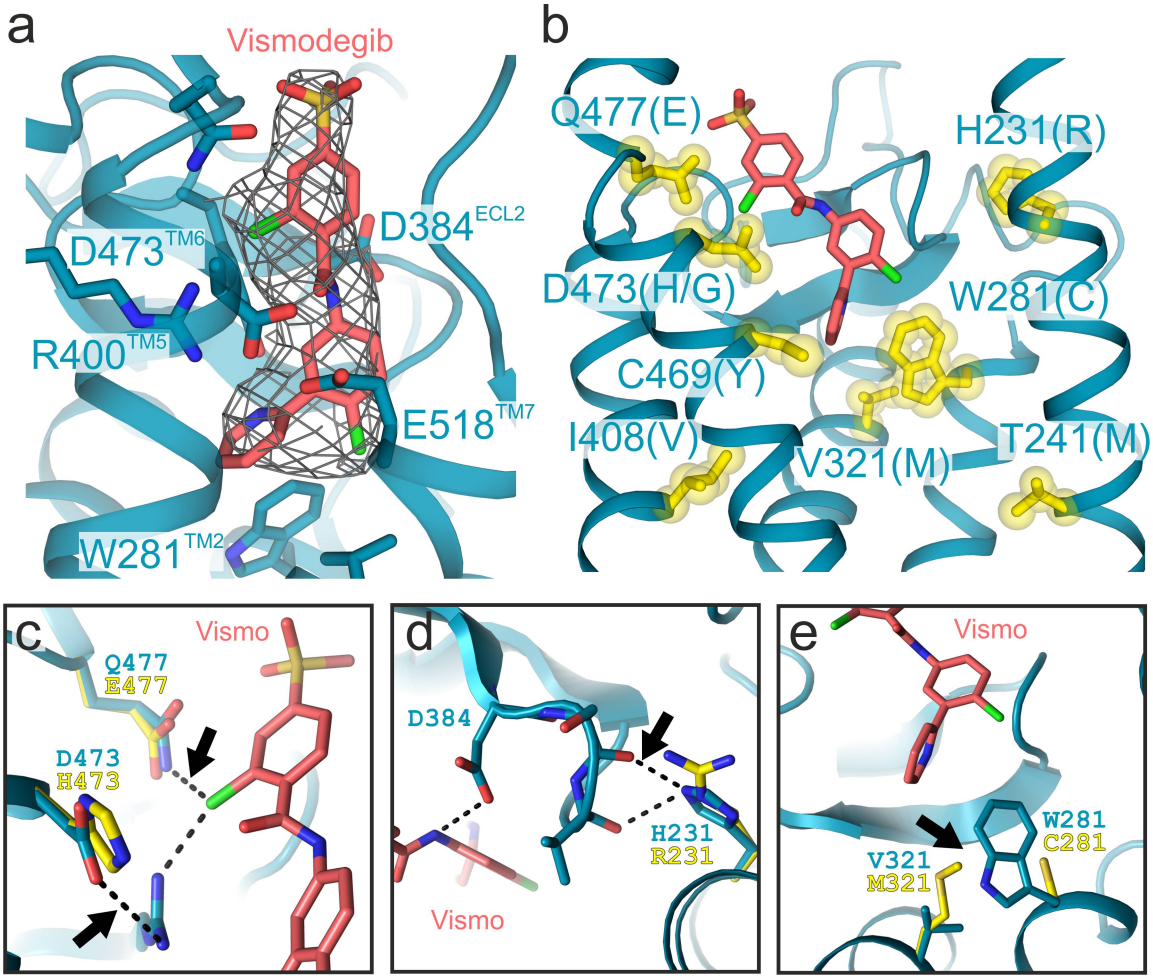


Figure 4.8: Structural analysis of mutations found in vismodegib-resistant tumours. **a,** Close-up view of vismodegib binding site. Colour-coding follows Fig. 4.7. Composite omit map calculated with PHENIX at 1.0σ shown as grey chicken-wire. **b,** Mapping of residues that are mutated in vismodegib-resistant tumours (yellow highlights). Brackets indicate mutant residues. **c-e,** Close-up views of selected interactions. Native residues in blue and mutated residues in yellow. Arrows indicate position of potential clashes that could disrupt vismodegib binding. **c,** Gln477/Asp473 hot spot. The Gln477Glu mutation leads to a loss of the potential hydrogen bond of the glutamine sidechain to the chloride of the vismodegib chlorophenyl-methylsulfone moiety. The Asp473His mutation potentially destabilises the hydrogen-bonding network around Arg400 that coordinates the vismodegib chlorophenyl-methylsulfone moiety. **d,** The imidazole ring of His231 is in hydrogen-bonding distance to two carbonyl main-chain atoms of residues Ser385 and Val386 located on a loop coordinating the interaction of Asp384 to vismodegibs amide linker. **e,** Trp281 forms a key hydrophobic interaction with the vismodegib pyrimidine ring that is deeply buried in the SMO helical bundle core. Mutation to cysteine would significantly destabilise this interaction while mutation of nearby Val321 to the bulkier methionine would likely result in a rearrangement of the Trp281 side chain.

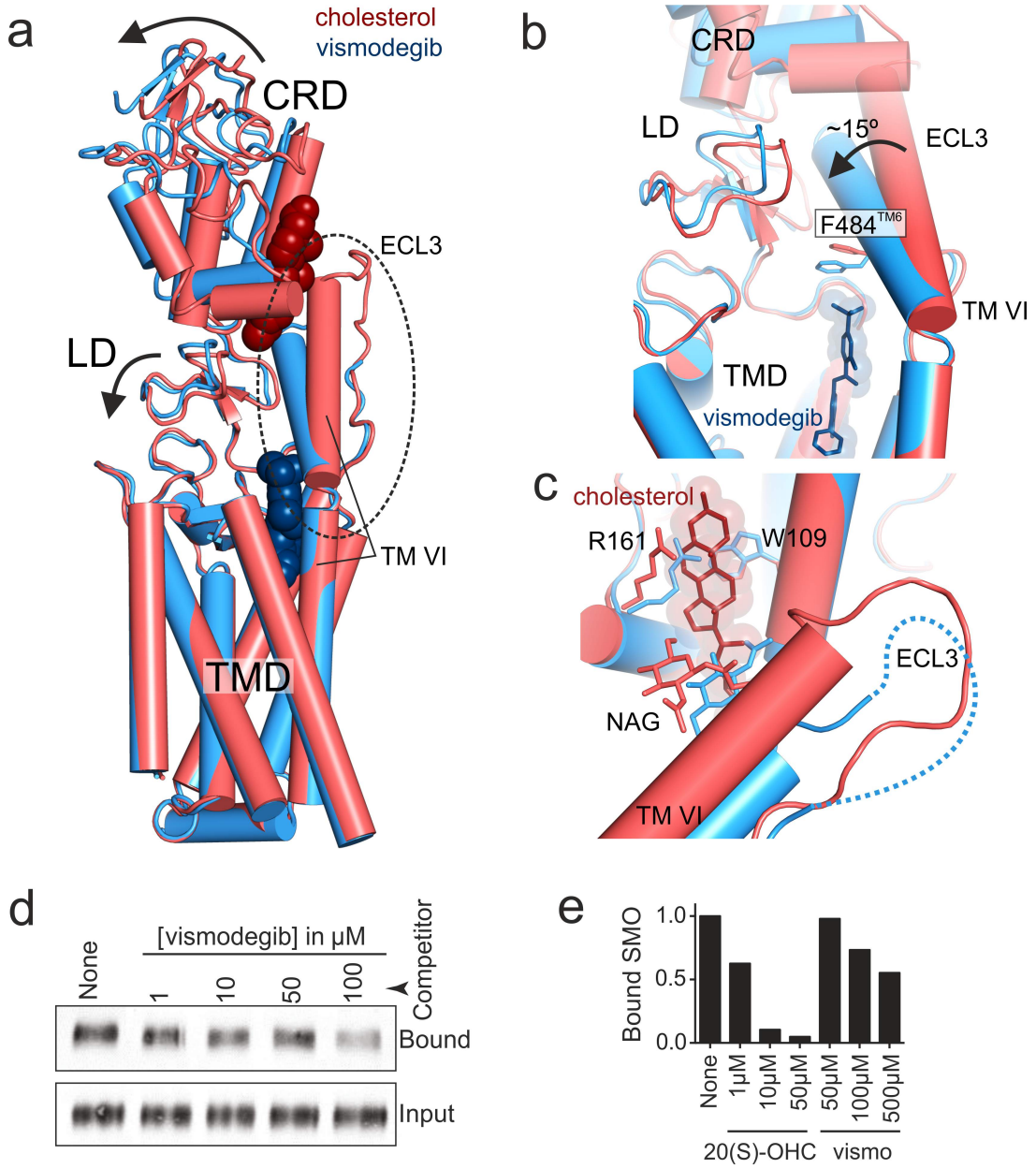


Figure 4.9: Comparison of cholesterol- and vismogedib-bound SMO. a, Superposition of the two SMO structures based on the TMD. Helices shown as cylinders. Arrows indicate domain rotations. Cholesterol (red) and vismodegib (blue) shown as spheres. **b,** Close-up of the TMD/LD/CRD interface. Phe484 contacts the vismodegib methylsulfone group. **c,** Close-up of sterol-binding site. Cholesterol, Arg161 and NAG shown as sticks. The partly-disordered ECL3 loop of vismo-SMO is depicted as dotted line. **d,** SMOΔC-BRIL-V329F captured on cholesterol beads in the presence of increasing concentrations of free vismodegib. **e,** Comparison of SMOΔC-BRIL-V329F captured on cholesterol beads in the presence of increasing concentrations of free 20(S)-OHC or vismodegib. Binding experiments in **d** and **e** were performed in collaboration with the Rohatgi Lab with protein produced by me.

change in the extracellular domains and the ECL3 loop when compared to the cholesterol-SMO structure (Fig. 4.9, panel a). The extracellular extension of TM helix VI, which contacts the CRD and LD, is shifted ∼15° towards the LD and CRD, likely due to an interaction between Phe484 and the terminal methylsulfone moiety of vismodegib (Fig. 4.7, panel c, and Fig. 4.9, panel b). This movement of helix VI results in a rotational movement of the CRD (and to a more minor extent the LD), allowing ECL3 to intrude into the CRD binding site (Fig. 4.9, panel c). As a consequence, the CRD binding site is occluded by the sidechain of CRD residue Arg161, which forms a stacking interaction with Trp106 and occupies the space where cholesterol was located in the cholesterol-SMO structure. Indeed, cholesterol was absent from the vismodegib-SMO structure (Fig. 4.7). Accordingly, vismodegib reduced the binding of purified SMOΔC-BRIL-V329F to cholesterol beads (Fig. 4.9, panels d and e).

In summary, vismodegib binding to the TMD site induces a conformational change that ultimately results in a rearrangement of the CRD binding site. This provides a structural mechanism for communication between the two sites, previously observed to relate allosterically[257] through ligands that bind to either the CRD or TMD sites. More generally, it suggests that a similar conformational change, involving a shift of ECL3 and a pivoting of the CRD on the extracellular end of the TMD bundle, may allow an extracellular signal to be transduced from the CRD to the TMD and ultimately across the membrane.

4.3.4 Cyclopamine molecules occupy both binding sites

Strikingly, cyclopamine, a plant alkaloid and the original SMO antagonist, occupies both the TMD binding site, as shown previously in a LD-TMD structure[388], and the

CRD binding site at the same time (Fig. 4.10, panel a). Despite their very different chemical properties, cyclopamine binds the CRD binding site in almost exactly the same manner as cholesterol (Fig. 4.10, panel b, and Fig. 4.11, panel a). The terminal hydroxyl group of cyclopamine forms a hydrogen bond with the essential residue Asp95, part of the hydrogen-bonding network with Tyr130 and Trp109, the indole ring of which stacks against the flat face of cyclopamine's tetracyclic ring (Fig. 4.10, panel b). The cyclic secondary amine group at the other end of cyclopamine does not extend towards the LD, as the acyl tail of cholesterol does, but instead, due to the chiral geometry of its central carbon, projects outwards from the binding site, slotting between Glu160 of the CRD and Asn493 of ECL3 (Fig. 4.10, panel b). Although no glycosylation was apparent in this structure, it is likely that *in vivo* this outward projection of cyclopamine would be shielded from the bulk solvent by an N-linked glycan attached at position Asn493, as was apparent in the cholesterol structure (Fig. 3.15, page 95). However, the glycan from the cholesterol structure would need to be rearranged in order to accommodate the secondary amine end of cyclopamine (Fig. 4.11, panel a). In this binding mode the cyclopamine bound to the CRD binding site does not directly contact the LD, which may have implications for its effect on signalling.

The TMD-bound cyclopamine molecule sits further out of the central cavity compared to vismodegib (Fig. 4.11, panel b). The terminal hydroxyl group of this cyclopamine molecule is oriented towards the centre of the 7TM bundle and forms part of an extensive hydrogen-bonding network, including residues from the helical bundle ($Arg400^{TM5}$, $Asp473^{TM6}$, $Glu518^{TM7}$) and from the extracellular domains ($Asn219^{LD}$, $Asp384^{ECL2}$, $Ser387^{ECL2}$ and $Tyr394^{ECL2}$) (Fig. 4.10, panel c). This orientation of cyclopamine is also stabilised by many hydrophobic interactions with residues from the extracellular regions of SMO, such as $Leu221^{LD}$, $Phe222^{LD}$, $Lys395^{ECL2}$, $Phe484^{ECL3}$ and $Pro513^{ECL3}$ (Fig. 4.10, panel c). Notably, residue $Phe484^{ECL3}$, the sidechain of

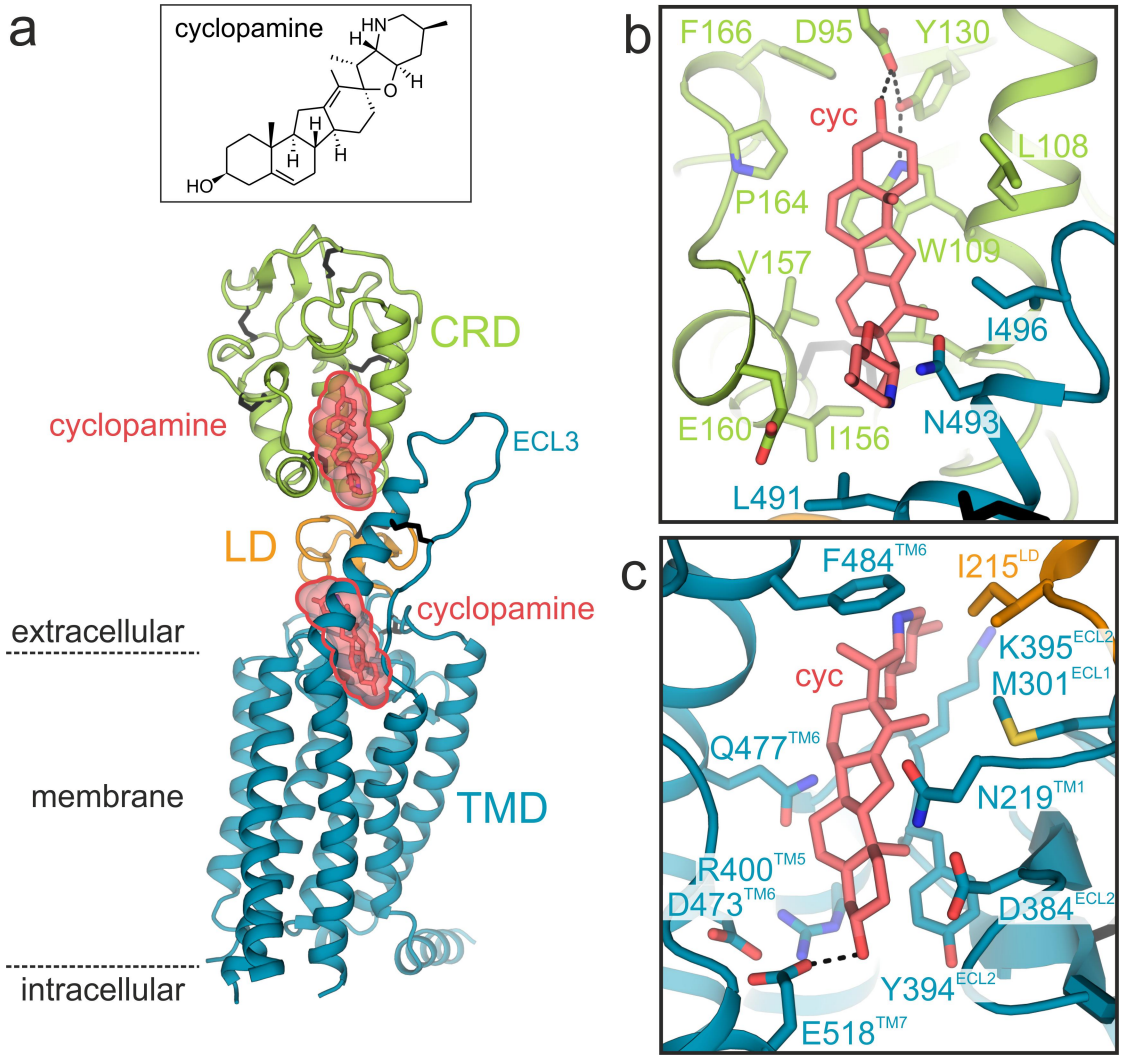


Figure 4.10: Structure of human SMO bound to two cyclopamine molecules. **a,** Overall structure of cyclopamine-bound SMO in cartoon representation with cyclopamine in red. Other colours follow Fig. 3.15. Inset shows chemical structure of cyclopamine. **b & c,** Close-ups of cyclopamine binding sites. Important residues are shown as sticks with atomic colouring and labelled. Dotted black lines indicate potential hydrogen bonds. **b,** CRD binding site. **c,** TMD binding site.

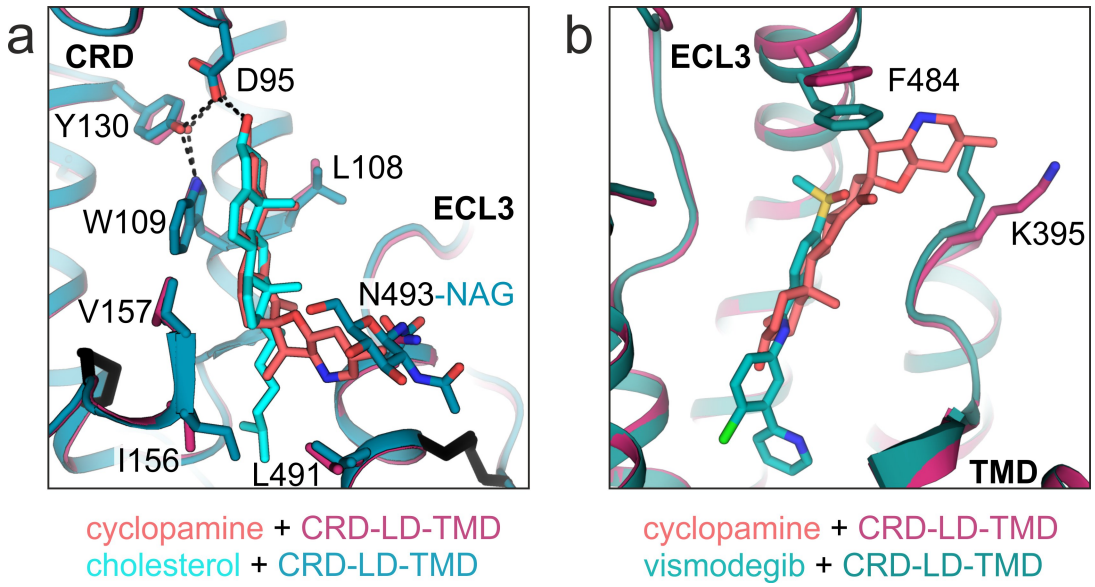


Figure 4.11: Cyclopamine binding compared to other ligands. a, Close-up of CRD binding site with overlay of cyclopamine- and cholesterol-bound SMO. Part of alpha helix 3 has been removed for clarity. **b,** Close-up of TMD binding site with overlay of cyclopamine- and vismodegib-bound SMO. Parts of the LD, ECL1 and ECL2 have been removed for clarity. Important residues are shown as sticks with atomic colouring and labelled. Dotted black lines indicate potential hydrogen bonds. Small molecules and structures coloured as indicated.

which packs down onto the methylsulfone group of vismodegib and may be the cause of a shift in the position of ECL3, is raised in the double-cyclopamine structure, similar to that in the cholesterol structure (Fig. 4.11, panel b). In the double-cyclopamine structure, compared to the vismodegib structure, the sidechain of residue Lys395^{ECL2} adopts a fully extended conformation to avoid a steric clash with the cyclic secondary amine terminus of cyclopamine (Fig. 4.11, panel b). This conformation is similar to the previously published structure of cyclopamine in complex with the SMO LD-TMD[388].

Overall, the double-cyclopamine structure, despite having an apparently antagonist molecule in each binding site, resembles the cholesterol structure far more than it resembles the single antagonist-bound vismodegib structure. The substantial re-orientation of the CRD, the LD and ECL3 observed in the vismodegib-SMO

structure is not seen in the cyclopamine structure, suggesting that vismodegib and cyclopamine may inhibit SMO via different mechanisms.

4.3.5 Cholesterol remains bound in SAG structure

The potent agonist SAG (EC_{50} = ∼5 nM) was one of the first small molecule modulators of SMO to be developed[53, 120, 314]. While a structure of SAG1.5 with SMO LD-TMD has already been solved, it showed only very minor conformational differences relative to the antagonist-bound structures (Fig. 3.2, page 68)[378]. We reasoned that a structure of CRD-LD-TMD SMO with SAG might display an important conformational change that was not apparent from the earlier shorter structure.

In our new CRD-LD-TMD structure, SAG binds the TMD in the same manner as its derivative SAG1.5 in the previously reported structure of SMO LD-TMD (Fig. 4.12, panel a)[378]. Its two aromatic rings are stacked against each other at the extracellular entrance to the binding site with the cyclohexane pointing towards the inside of the 7TM bundle (Fig. 4.12, panel c). In building our structure, when SAG1.5 was placed instead of SAG, the two extra fluorine moieties of SAG1.5 immediately showed up as superfluous electron density on the $F_o - F_c$ map (Fig. 4.12, panel e), demonstrating a well-conserved binding mode between the two agonists and unambiguously identifying the small molecule in our structure as SAG. The binding site of SAG is formed mostly by residues from the LD and extracellular loops (Fig. 4.12, panel c). The footprint of residues that SAG interacts with is much the same as that for cyclopamine, with the methyl amine group of SAG stabilised in the core of the helical bundle by a hydrogen-bonding network with $Asp473^{TM6}$ and $Glu518^{TM7}$. Distinct from cyclopamine, the oxygen from the central carboxamide group of SAG forms a potential hydrogen bond with $Asn219^{LD}$ and the sidechain of $Arg400^{TM5}$ is

re-oriented to avoid a steric clash with the secondary amine tail of SAG (Fig. 4.13, panel a). Notably, Wang *et al.* have chosen a different rotamer for Glu518^{TM7}, positioning the sidechain for hydrophobic packing with the cyclohexane rather than hydrogen-bonding with the secondary amine (Fig. 4.13, panel b)[378].

In this structure, Phe484^{ECL3} is raised and the CRD binding site is not occluded, allowing cholesterol to bind at the same time as SAG. Cholesterol is bound to the CRD binding site in exactly the same manner as in the initial cholesterol-SMO structure (Fig. 4.12, panel b). Given the finding that cholesterol itself is sufficient to activate SMO, this structure represents a double agonist-bound form of SMO. This is consistent with a previous finding that SMO was sensitised to activation by SAG when primed with 20(S)-OHC, and vice versa, demonstrating synergistic activation[257]. Despite two bound agonists, however, the TMD still displayed no change in conformation compared to the earlier structures. The stabilising point mutation Val329Phe and the intracellular BRIL fusion appear to strongly lock the 7TM bundle into the observed conformation, regardless of the bound ligands.

4.3.6 Oxysterol structure unchanged from cholesterol structure

Next we sought a crystal structure of SMO bound to a CRD agonist oxysterol, either 20(S)-OHC or 20(S)-yne. These oxysterols bind tightly to the SMO CRD in order to activate signalling[263, 257]. We reasoned that a structure of SMO activated via the CRD (rather than the TMD) was most likely to present a signal transduction state that extended from the CRD through the TMD and into the intracellular region of SMO. Oxysterols seem to bind more strongly to SMO than cholesterol[226] but the lipid mix contained such a massive quantity of cholesterol that we were concerned

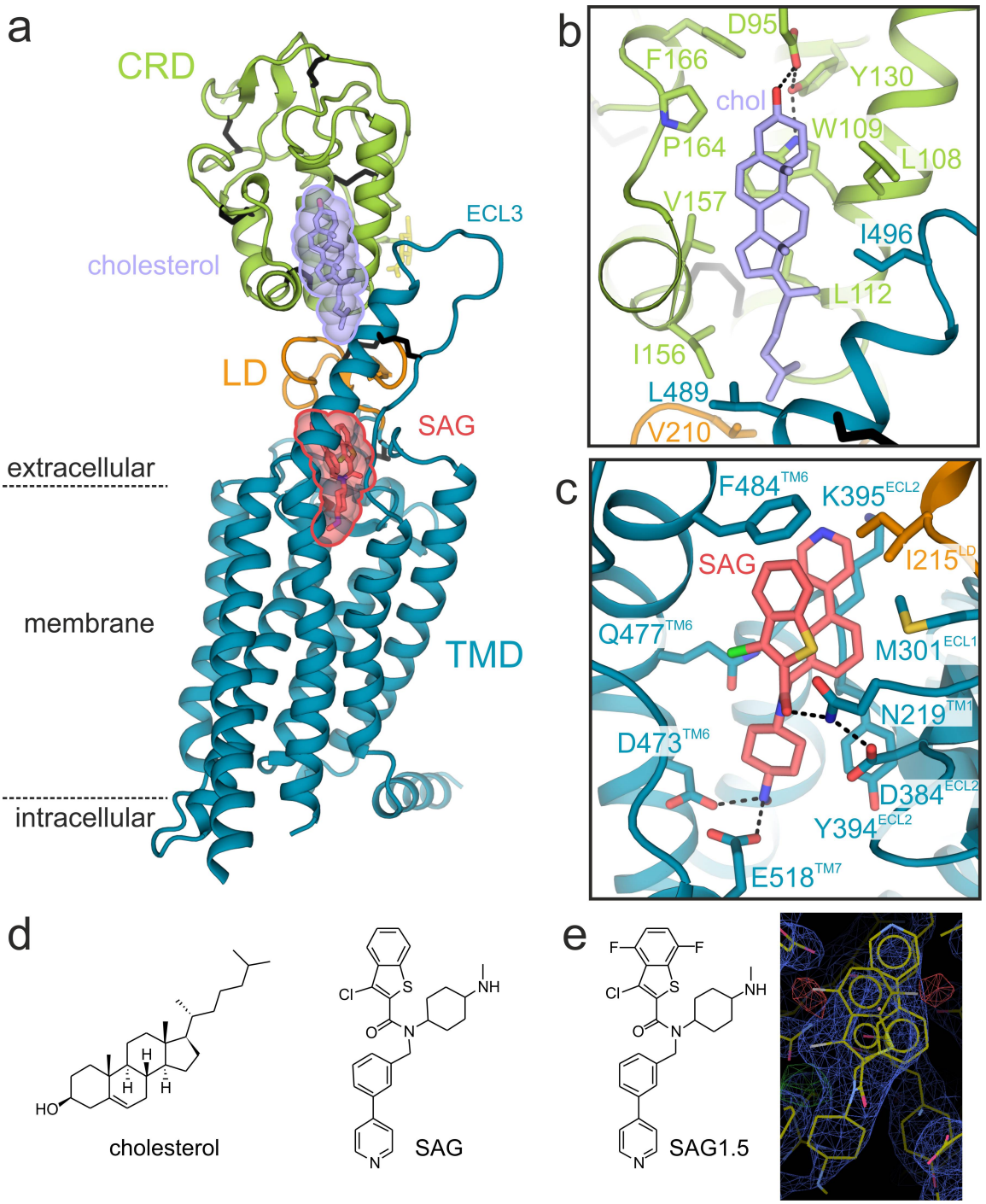


Figure 4.12: Structure of human SMO bound to SAG and cholesterol. **a,** Overall structure of SAG- and cholesterol-bound SMO in cartoon representation. Cholesterol in violet, SAG in red. Other colours follow Fig. 3.15. **b & c** Close-ups of ligand binding sites. Important residues are shown as sticks with atomic colouring and labelled. Dotted black lines indicate potential hydrogen bonds. **b,** Cholesterol binding site. **c,** SAG binding site. **d,** Chemical structures of cholesterol and SAG. **e,** Chemical structure of SAG1.5 (left) and electron density maps of SMO structure with SAG1.5 modelled into TMD binding site (right). $2F_o - F_c$ map as blue net contoured at 1.00 σ, $F_o - F_c$ map as green/red net for positive/negative density contoured at 3.50 σ.

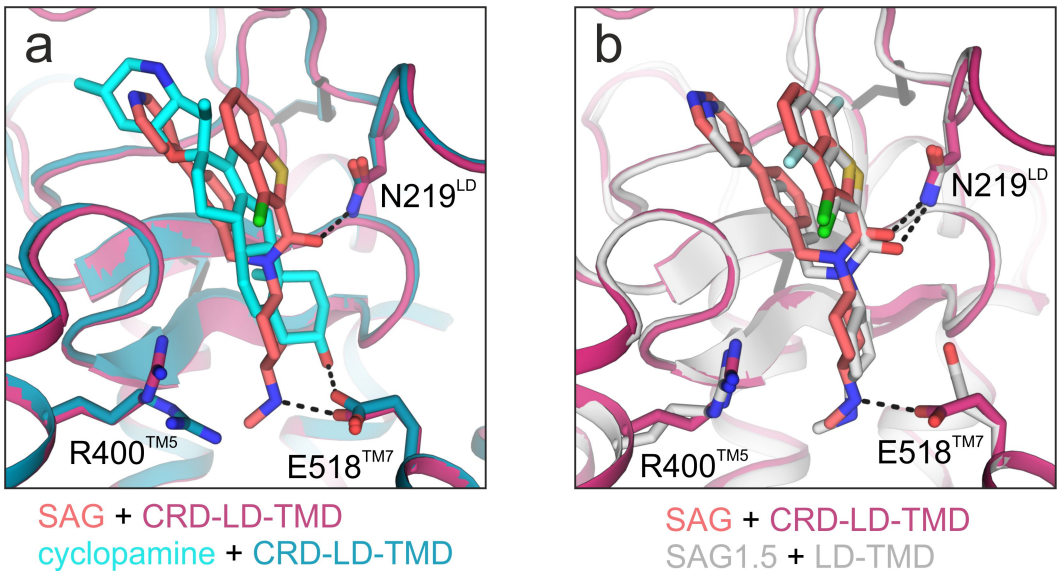


Figure 4.13: SAG binding compared to other ligands. Close-ups of TMD binding site. Parts of TM helix VI and VII have been removed for clarity. Important residues are shown as sticks with atomic colouring and labelled. Dotted black lines indicate potential hydrogen bonds. Small molecules and structures coloured as indicated. **a,** Overlay of SAG- and cyclopamine-bound SMO. **b,** Overlay of SAG- and SAG1.5-bound SMO. (PDB: 4QIN)

it may disrupt the binding of oxysterols. Accordingly, we first sought to replace the cholesterol in the lipid mix with oxysterol, reasoning that this would both remove all exogenous sources of cholesterol and provide a large molar excess of the oxysterol for binding to the protein.

A mixture of 10% 20(S)-OHC and 90% monoolein (20(S)-OHC/MO) behaved as normal but a mixture of 10% 20(S)-yne and 90% monoolein did not melt at all at 42°C. In fact, this mixture hardly melted at 100°C. A mixture of approximately 9% 20(S)-yne, 9% cholesterol and 82% monoolein (20(S)-yne/chol/MO) melted at 42°C but would not clarify, remaining cloudy and white. Crystallisation trials using protein with no added small molecules were set up with both oxysterol-doped lipid mixtures. Crystals grew in the 20(S)-OHC/MO mixture but diffracted to only ~4 Å, even after optimisation of crystallisation conditions. No crystals grew in the 20(S)-yne/chol/MO mixture.

Subsequently, the same approach was taken for the two oxysterols as for the other small molecules; oxysterol dissolved in DMSO at high concentration was added to the purified protein and crystallisation trials set up in the standard 10% cholesterol, 90% monoolein lipid mixture. This strategy produced well-diffracting crystals that allowed datasets for 20(S)-OHC and 20(S)-yne to be collected to 3.2 and 3.9 Å, respectively. Both structures looked exactly the same so only the higher resolution one, 20(S)-OHC-SMO, is reproduced here (Fig. 4.14, panel a). As expected, the oxysterol molecule binds to the CRD at the same place and in the same orientation as cholesterol. Unfortunately, given the nearly identical chemical structures of cholesterol and 20(S)-OHC (Fig. 4.14, panel b) and the relatively low resolution of this structure, it was not possible to unambiguously assign the molecule within the binding site as 20(S)-OHC. Although solved to the same 3.2 Å resolution as the cholesterol structure, the electron density for several regions of the 20(S)-OHC structure looked better than that of the cholesterol, perhaps an indication that the tighter-binding oxysterol was in fact producing a more ordered crystal structure (Fig. 4.15, panel a). The single difference in chemical composition of 20(S)-OHC, a hydroxyl group attached to the position carbon 20, points outwards from the binding site (Fig. 4.14, panels b & c). When aligned by their TMDs, the oxysterol-bound CRD is shifted ∼1 Å from that in the cholesterol structure, in the opposite direction from the vismodegib-bound CRD (Fig. 4.15, panel b).

Given that the crystal structure of oxysterol-SMO did not show a substantial conformational shift compared to cholesterol, we next sought to probe conformational changes in SMO induced by CRD ligands in the solution state. We performed small angle X-ray scattering (SAXS) experiments on amphipol-solubilised SMOΔC-BRIL-V329F in the absence of any exogenous ligands (hereafter called apo-SMOΔC-BRIL-V329F) or SMOΔC-BRIL-V329F loaded with the agonist 20(S)-OHC (Fig. 4.16, panel a). The initial SAXS curves showed a small but significant difference in the

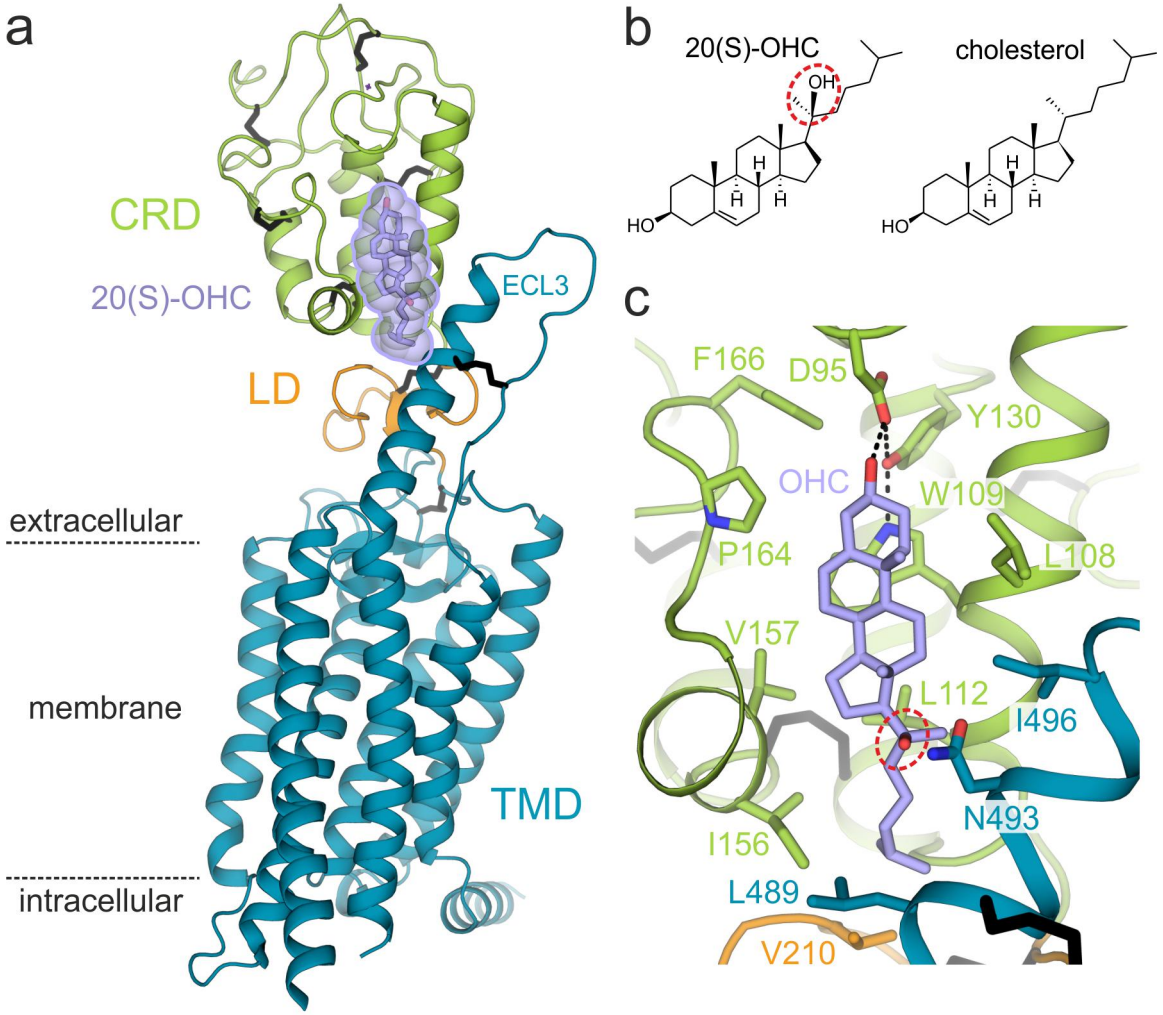


Figure 4.14: Structure of human SMO bound to 20(S)-OHC. a, Overall structure of 20(S)OHC-bound SMO in cartoon representation. 20(S)-OHC in violet, other colours follow Fig. 3.15. **b** Chemical structures of 20(S)-OHC and cholesterol. Red dashed oval highlights the extra hydroxyl group. **c,** Close-up of 20(S)-OHC binding site. Important residues are labelled and shown as sticks with atomic colouring. Black dotted lines indicate potential hydrogen bonds. Red dashed oval highlights the extra hydroxyl group.

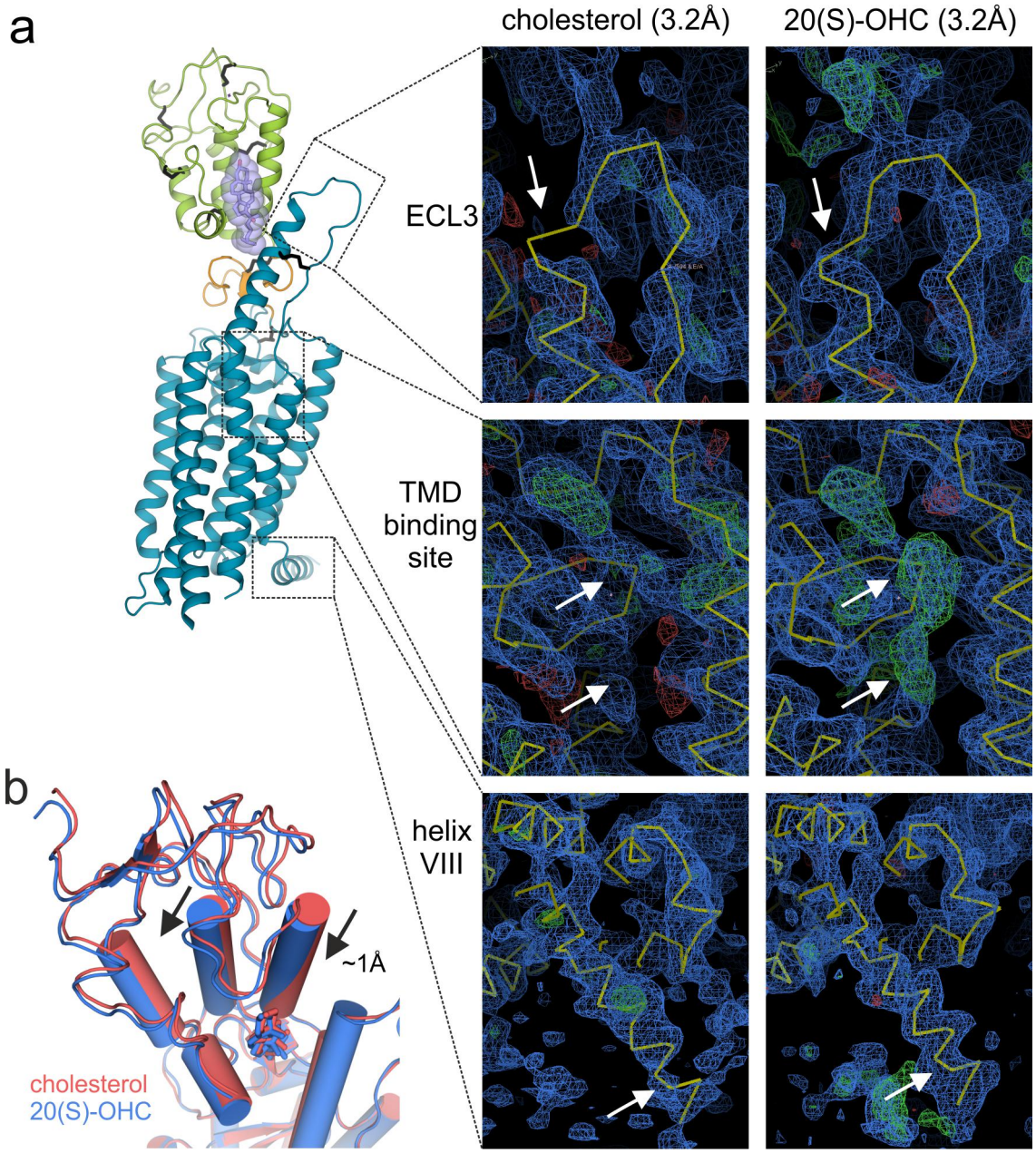


Figure 4.15: Comparison between 20(S)-OHC and cholesterol structures. **a,** Comparison of electron density in three different regions (ECL3, TMD binding site and helix VIII as indicated in overall cartoon representation) between the cholesterol (left) and 20(S)-OHC (right) structures. Blue chicken-wire represents $2F_o - F_c$ map at 1.0 σ contour, red and green chicken-wire represents $F_o - F_c$ map at ± 3.0 σ contour, yellow lines represent the protein chain backbone. White arrows highlight areas of difference between the two structures. **b,** Superposition of cholesterol- and 20(S)-OHC-bound SMO, aligned by the TMD. Helices shown as cylinders. Cholesterol-bound structure in red, 20(S)-OHC in blue.

middle q region (∼0.07-0.21 Å^{-1}) (Fig. 4.16, panel b), while comparison of the $P(r)$ distributions revealed a change in maximum dimension length (D_{max}) from 120 Å for the apo form to 129 Å for the 20(S)-OHC form (Fig. 4.16, panel c)[104, 295]. In addition, a dimensionless Kratky analysis revealed a slower decay of the transformed scattering intensities for 20(S)-OHC, indicating a less spherical particle compared to the apo form (Fig. 4.16, panel d). Taken together, these analyses show that 20(S)-OHC binding induces a conformational change consistent with elongation and reduced globularity of 20(S)-OHC-loaded SMOΔC-BRIL.

4.3.7 TMD binding site contains an unidentified molecule

Just as for the original cholesterol-SMO structure, we again observed extra electron density in the TMD binding site of the 20(S)-OHC-SMO structure (Fig. 4.17, panels a and b). No TMD-binding small molecule additives had been included in the protein preparation so the presence of a molecule in this site was not anticipated. The sigmaA-weighted $F_o - F_c$ electron density map for this binding site taken from the 20(S)-OHC-SMO structure is shown (Fig. 4.17, panels a and b). We have not been able to identify this molecule as of yet. Although the density is located at approximately the same position as all of the TMD-binding small molecules, none of the small molecules previously solved in complex with SMO fit the density (Fig. 4.17, panel c). Attempts at modelling other small molecules that might be present in the cellular context, such as sterols, lipids and vitamins, into this density have so far proved fruitless.

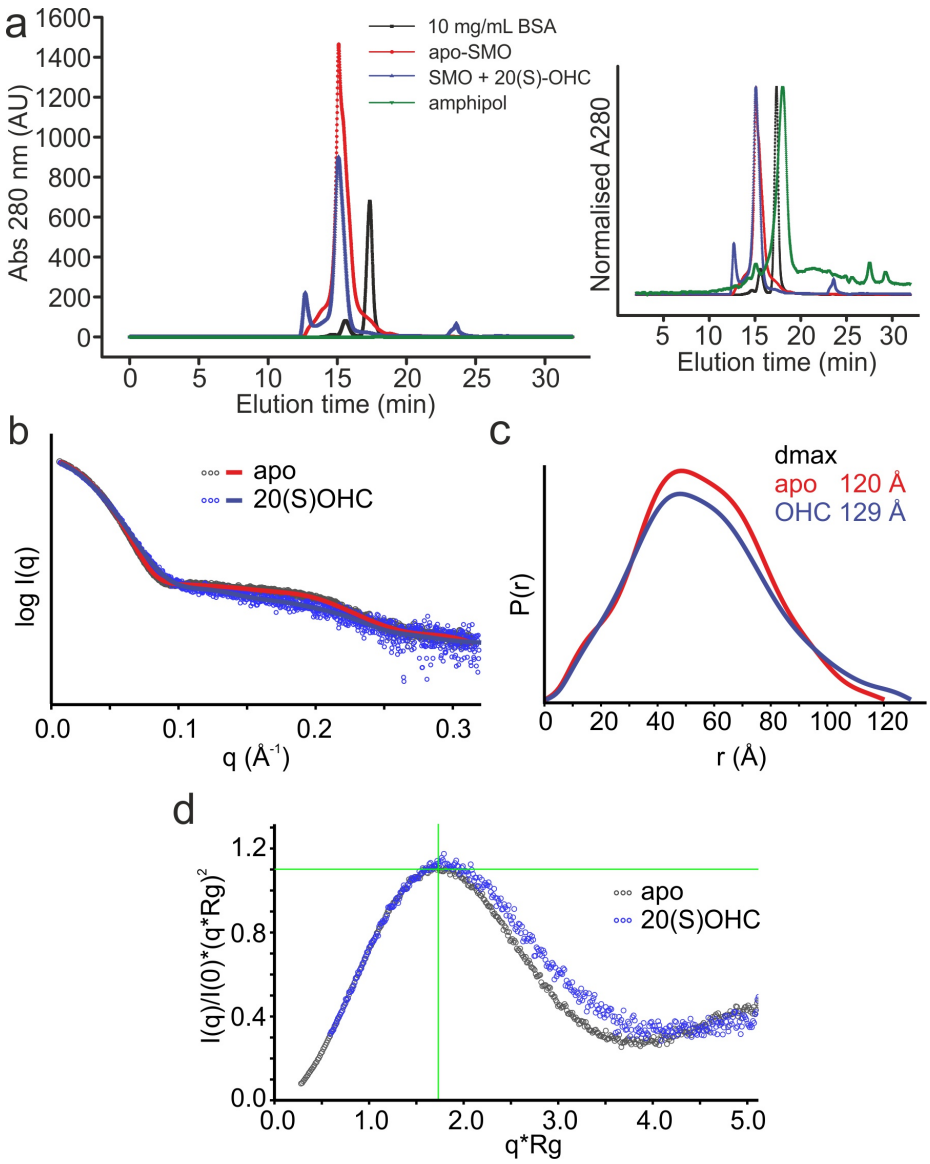


Figure 4.16: SAXS analysis of SMOΔC-BRIL-V329F. a, Overlay of size-exclusion chromatograms monitored at 280 nm (A280) collected during SAXS measurements for apo-SMOΔC (red), 20(S)-OHC + SMOΔC (blue), amphipol (green) and BSA standard (black). Amphipol and BSA were injected at 10 mg/mL. Inset shows curves normalised to peak height. BSA was used as a reference with a radius-of-hydration of 3.7 nm. Absorbance of the free amphipol is negligible and elutes ∼5 min after the amphipol-stabilised SMOΔC samples. **b,** Solution-state SAXS profiles of untreated (grey circles, red line) and 20(S)-OHC-treated (blue circles, navy line) SMOΔC-BRIL-V329F. Lines are fits derived from the indirect Fourier transform of underlying data. **c,** Paired-distance, P(r), distribution functions, normalised to their respective I(0) values, with maximum particle sizes (dmax) indicated. **d,** Dimensionless Kratky plot of apo and 20(S)-OHC-loaded SMOΔC SAXS data. Cross-hairs denote the Guinier-Kratky point ($\sqrt{3}$, 1.1), the peak position for an ideal, globular particle. The slower decay of the transformed scattering intensities for 20(S)-OHC (blue) indicate a comparably less spherical particle.

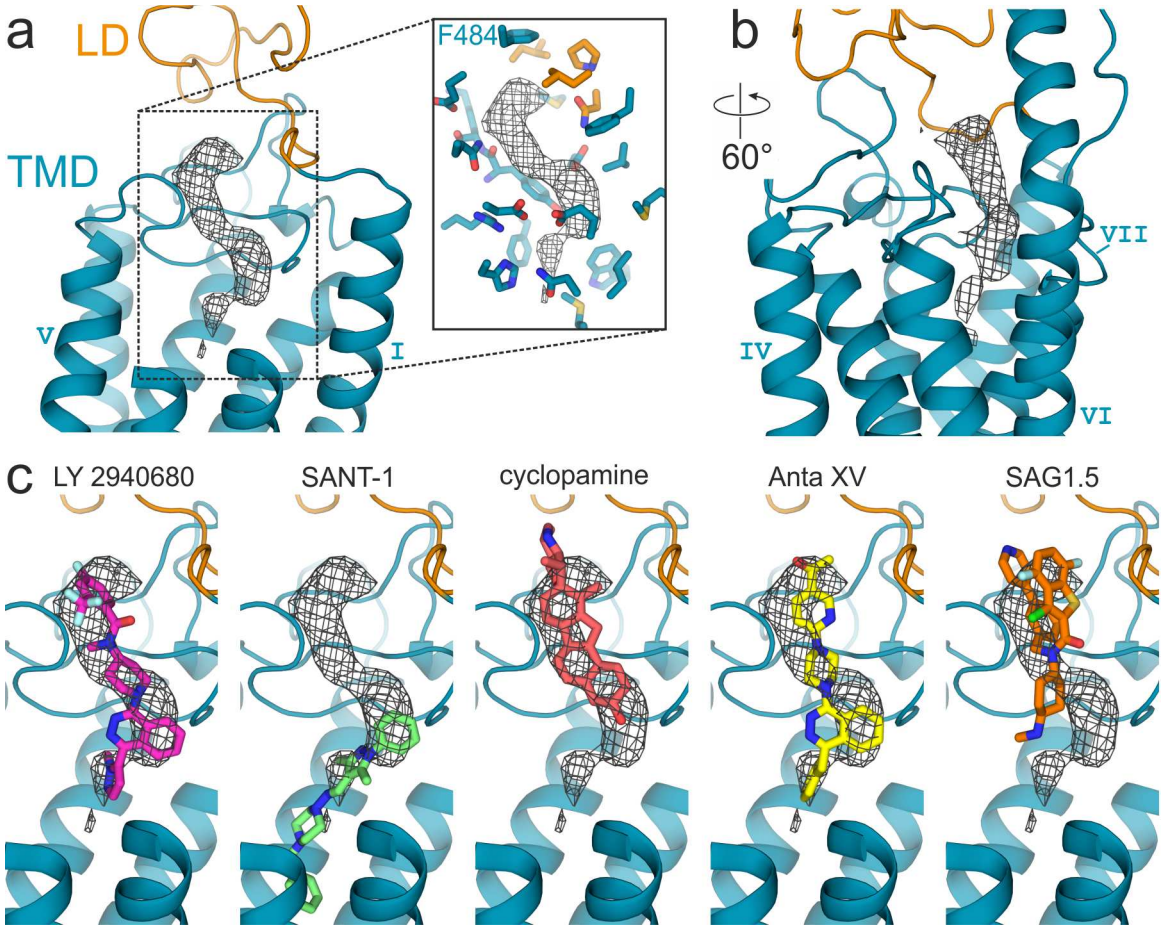


Figure 4.17: Unidentified density in TMD binding site. a, SigmaA-weighted $F_o - F_c$ map (grey chicken-wire) of unidentified electron density within 20(S)-OHC-SMO structure (cartoon representation). Contour level 3.0 σ. Domains coloured as in Figure 3.15. Roman numerals indicate TMD helices. TMD helices VI and VII were removed for clarity. Inset shows potentially interacting residues in stick representation with atomic colouring, with contributions from both the LD (orange) and TMD (blue). **b,** Another view of the unidentified electron density. **c,** Small molecule TMD ligands from previously solved structures overlaid with unidentified electron density. Structures were aligned by their TMDs. All are shown in the same orientation as in **a** with TM helices VI and VII removed for clarity. (PDB: 4O9R, 4QIM, 4QIN, 4JKV, 4N4W)

4.4 Discussion

4.4.1 A conserved binding mode for CRD ligands

A recent paper from Huang *et al.* reported crystal structures of the soluble SMO CRD from *Xenopus laevis* (xSMO) in complex with 20(S)-OHC and cyclopamine at 1.6 and 2.5 Å resolution, respectively[149]. These structures confirmed both the binding position and binding mode of the CRD ligands that we had seen in our various CRD-LD-TMD structures. The 20(S)-OHC-bound xSMO CRD structure (PDB 5KZV) almost perfectly overlays onto our structure of hSMO CRD-LD-TMD bound to 20(S)-OHC (RMSD 0.594 Å for 106 equivalent Cα positions, Fig. 4.18, panel b), despite some amino acid differences between the species, with only minimal loop rearrangements near disulfide bond $1'$ which may be due to the lack of the connector region in the xSMO structure (Fig. 4.18, panel b). A superposition of the cyclopamine-bound xSMO CRD structure (PDB 5KZY) with our hSMO double-cyclopamine CRD-LD-TMD structure revealed a similarly close structural alignment (RMSD 0.587 Å for 112 equivalent Cα positions, Fig. 4.18, panel c). The CRD-bound cyclopamine molecule from the hSMO structure is rotated slightly within the binding site relative to the cyclopamine in the xSMO structure, most likely to prevent a steric clash with the sidechain of Asn493$^{\mathrm{ECL3}}$ (and in the native state likely an attached N-linked glycan too), which is not present in the xSMO structure (Fig. 4.18, panel d). These independently solved structures of the SMO CRD demonstrate both conservation of the CRD binding site architecture between different vertebrates as well as conserved modes of binding between different types of ligands, an oxysterol and an alkaloid.

A third xSMO CRD structure solved by Huang *et al.* did not include any ligands (PDB 5KZZ) and contained a sizeable rearrangement of one edge of the CRD

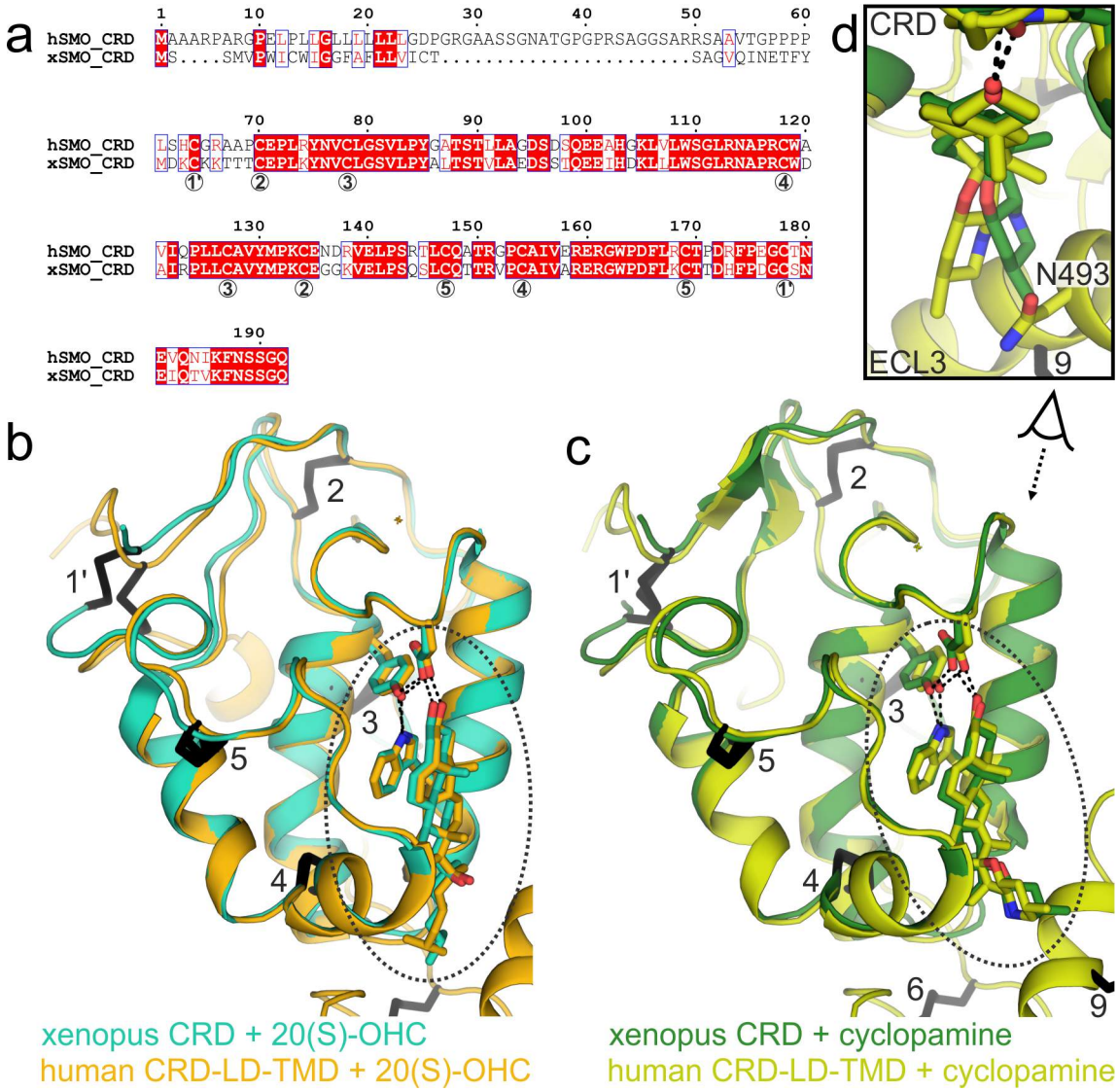


Figure 4.18: Comparison of ligand-bound CRD structures. **a,** Sequence alignment of human and xenopus SMO CRDs. Coloured by % equivalent residues at each position. Encircled black numbers indicate conserved disulfide-forming cysteines. (Produced using Clustal Omega and ESpript3.0, refs [334, 310]). **b,** Superposition of 20(S)-OHC-bound SMO structures, xenopus CRD (PDB 5KZV) in cyan and human CRD-LD-TMD in gold. Cholesterol highlighted with dotted black oval. Disulfides shown as black sticks and numbered as in Fig. 3.15. **c,** Superposition of cyclopamine-bound SMO structures, xenopus CRD (PDB 5KZY) in dark green and human CRD-LD-TMD in yellow-green. Cyclopamine highlighted with dotted black oval. **d,** Top-down close-up of cyclopamine binding position for overlaid structures. Colours as in **c**.

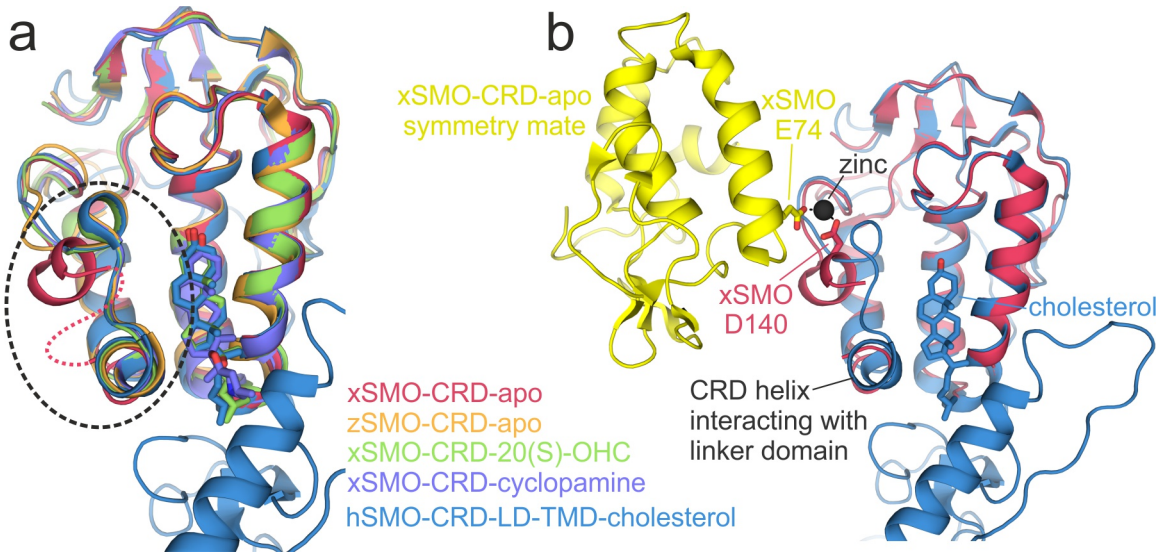


Figure 4.19: Comparison of ligand-bound SMO CRD structures. a, Overlay of five SMO CRD crystal structures, three with ligands and two without. Black dashed ellipse highlights altered region in xSMO-CRD-apo structure. Red dotted line represents loop missing from structure. **b,** Overlay of xSMO-CRD-apo structure (red, symmetry mate in yellow) and hSMOΔC-BRIL structure (blue). A zinc atom mediates a crystal contact between two symmetry-related xSMO-CRD-apo chains via Asp140 in the altered region. This region in the hSMO-CRD-LD-TMD structure interacts with cholesterol and the linker domain. (PDB: 5KZZ, 5KZV, 5KZY, 4C79, 5L7D)

binding site[149]. Excluding the rearrangement, the xSMO apo-CRD structure overlays well with the cholesterol-bound hSMO CRD-LD-TMD structure (RMSD 0.603 Å for 84 equivalent Cα positions), the structure held together by the five conserved disulfide linkages (Fig. 4.19, panel a). The rearrangement appears to consist primarily of the region between disulfide bonds 4 and 5, which includes alpha helix 3 that forms one side of the CRD binding site (Fig. 4.19, panel a). The authors make considerable claims that this rearrangement from apo to ligand-bound form constitutes a conformational change indicative of SMO activation. However, without the other domains of SMO being present, it is very difficult to make a reasonable assessment of how such a rearrangement might affect signal transduction. Indeed, it is not clear from their apo-structure exactly what effect this rearrangement might have on the signalling state of SMO.

The previously solved apo-CRD structure of *Danio rerio* SMO (zSMO, PDB 4C79) did not show this rearrangement but instead showed the same conformation as all of the ligand-bound structures (Fig. 4.19, panel a)[258]. In addition, in the vismodegib structure, where the CRD binding site is not occupied by a ligand, we see a reorientation of the entire CRD but not this rearrangement of alpha helix 3 (Fig. 4.9, page 139). This xSMO apo structure contains a number of zinc atoms which are not present in any of the other structures (the protein was crystallised in a buffer containing 200 mM zinc acetate). One zinc atom facilitates crystal packing by coordinating Asp140 on one polypeptide chain with Glu74 on one of its symmetry mates (Fig. 4.19, panel b). This unusual coordination suggests that the rearrangement may be a crystallisation artefact, perhaps enabled (but not caused) by the lack of a ligand to stabilise the CRD. Indeed, molecular dynamics simulations conducted on the cholesterol-bound hSMO CRD-LD-TMD structure presented in Chapter 3 (section 3.3.12, page 106) demonstrated that inclusion of cholesterol in the CRD binding site helped to stabilise the CRD, particularly in the region of alpha helix 3 (Fig. 3.23, page 107). It seems likely that this unliganded conformation is an artefact of crystallisation and not indicative of the native structure.

4.4.2 Simultaneous binding of ligands at two sites

The double-cyclopamine and SAG/cholesterol crystal structures presented here represent the first structures of SMO in which both ligand binding sites are occupied, in the first case by two of the same ligand and in the second by two different ligands. Previous studies have noted that although the two binding sites are physically separable the synergy observed between 20(S)-OHC and SAG suggests a positive allosteric link between the two domains and also implies that TMD and CRD ligands can bind to SMO simultaneously[257, 258]. Our structure of SMO CRD-LD-TMD with SAG

bound to the TMD and cholesterol bound to the CRD illustrates the structural compatibility of these two ligands and confirms both of these hypotheses (Fig. 4.20, panel a & b).

The prior observation that cyclopamine, previously identified as a TMD antagonist, can inhibit the binding of isolated SMO CRD to 20(S)-OHC beads led to speculation that a single cyclopamine molecule might bridge the two ligand binding sites, combining a higher affinity interaction with the TMD and a lower affinity one with the CRD[258]. In the light of the double-cyclopamine structure, this suggestion now seems highly unlikely. The distance between the binding sites is too great for any overlap of the two - at their closest point the cyclopamine molecules are still ∼15 Å apart (Fig. 4.20, panels c & d). In addition, the opposing orientations of the two cyclopamine molecules, in which the terminal hydroxyl group of each points away from the other, preclude simultaneous binding of a single cyclopamine molecule in both sites. Both hydroxyl groups are also engaged in essential hydrogen-bonding interactions, which could not be formed at the same time by one cyclopamine molecule (Fig. 4.10, panels b & c, page 142).

Another study has demonstrated combinatorial inhibition of SMO by the glucocorticoid budesonide and vismodegib[383]. The binding site of budesonide on the CRD of *Drosophila melanogaster* SMO (dSMO) has been mapped by solution state NMR[296] and corresponds to the same CRD binding site as for the sterols and cyclopamine. However, these data alone are not sufficient to conclude that budesonide and vismodegib bind at the same time. The doubly-bound structures of cyclopamine/cyclopamine and SAG/cholesterol suggest that a vismodegib/budesonide structure is a distinct possibility. However, the binding site in the dSMO CRD structure varies considerably from that of the hSMO CRD-LD-TMD structure (Fig. 3.18, panel e, page 98). If vismodegib and budesonide do bind hSMO at the same time,

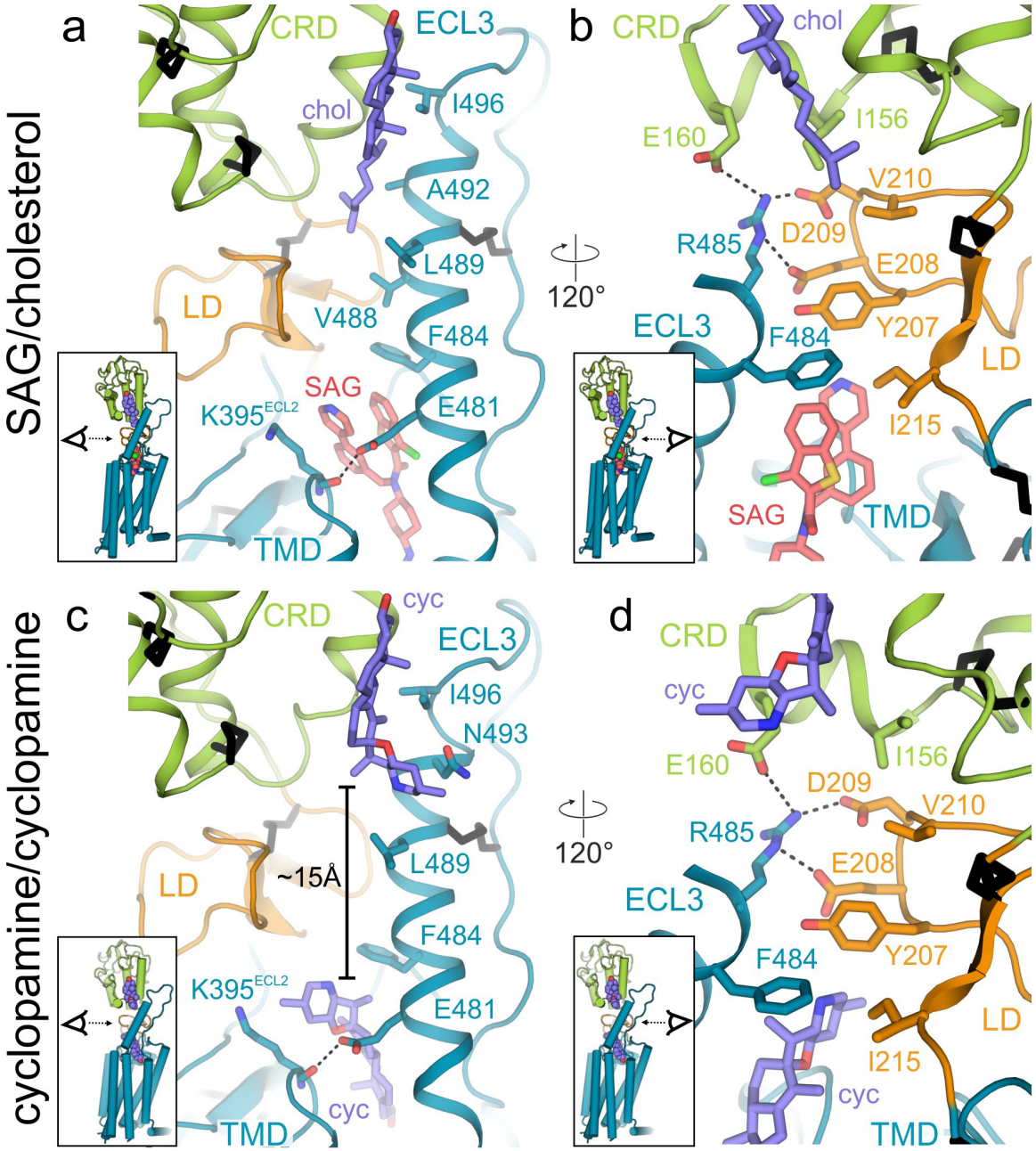


Figure 4.20: Simultaneous binding of ligands at two sites. Potential allosteric interaction networks between the CRD and TMD binding sites. Insets show viewing perspective. Domains coloured as in Fig. 3.15. Important residues labelled and shown as sticks in atomic colouring. **a & b** Cholesterol and SAG structure. **a,** Network along ECL3 helix. **b,** Network through residues from CRD, LD and ECL3. **c & d,** Cyclopamine structure with two cyclopamine molecules bound ∼15 Å distant. **c,** Network along ECL3 helix. **d,** Network through residues from CRD, LD and ECL3. In **a** and **c**, part of the CRD has been removed for clarity. In **b** and **d**, part of the CRD, LD and ECL3 have been removed for clarity.

it is not clear whether this would entail a further conformational adjustment of the CRD binding site, which is occluded by an N-acetylglucosamine from ECL3 in the vismodegib-SMO structure, to accommodate budesonide (Fig. 4.9, panel c, page 139).

4.4.3 Cyclopamine and ciliary localisation

Cyclopamine inhibits the Hh pathway while at the same time inducing ciliary localisation[313, 385, 393, 399], an ability which is special amongst SMO antagonists. However, given the agonist-like binding mode of the CRD-bound cyclopamine molecule, it is intriguing to think that perhaps this molecule would act as an agonist in the absence of another cyclopamine molecule bound at the TMD site. To this end, Huang *et al.* generated a mutant mouse SMO with two point mutations (Asp477Gly/Glu522Lys) in the TMD binding site that prevented the binding of the fluorescent BODIPY-cyclopamine and also prevented activation by SAG[149]. Using this mutant, they found that cyclopamine could indeed activate the Hh pathway via the CRD and it did so to a similar level as cholesterol. This result, in combination with the knowledge from the crystal structure that cyclopamine can bind both sites at once, suggests that cyclopamine simultaneously agonises SMO via its CRD and antagonises SMO via its TMD. In this case, the latter effect outstrips the former due to either superior binding affinity to the TMD or due to the TMD's downstream positioning within SMO's signal transduction network.

This may also help to explain the functional split-personality of cyclopamine; how does cyclopamine both inhibit signalling and promote ciliary localisation? Previous reports have found that the extracellular domain of SMO regulates ciliary localisation[1]. It now seems likely that the two roles can be split between the two binding sites, with the CRD binding site directing SMO to the primary cilium (which

in the absence of an inhibitor may be sufficient to activate signalling), while the TMD binding site governs signal transduction through the 7TM bundle. Interestingly, Huang *et al.* also found using their Asp477Gly/Glu522Lys SMO mutant that Shh, the native pathway activator, was still able to activate the pathway despite the loss of the TMD binding site[149]. This suggests that the TMD binding site may in fact be dispensable for native signalling. Precisely how native Shh-induced activation of SMO occurs is unknown. Identifying the mechanism of interaction between Ptc1 and SMO, including any small molecule second messengers, remains a challenge for studies of Hh signalling.

4.4.4 Multiple CRD positions suggest a continuum of states

The vismodegib-SMO structure revealed a dramatic shift in the positioning of the CRD and LD relative to the TMD, as well as movement of ECL3, compared to the cholesterol structure (Fig. 4.21, panel b). This shift equates to a levering of ECL3 by ∼15° and a pivot of the CRD leading to ∼6.2 Å movement of the Cα of Glu100 at the far end of alpha helix 1 (Fig. 4.21, panel b & d). The cyclopamine structure also appears to produce a subtle shift along the same pivot axis (∼1.0 Å at the same residue) as for vismodegib (Fig. 4.21, panel b & d). The agonist structures, by contrast, cluster very tightly, displaying either no shift or a subtle shift along the same pivot axis but in the opposite direction (Fig. 4.21, panels a & c).

These multiple ligand-bound structures in different signalling states suggest that the pivot position of the CRD-LD may indicate states that enable signal transduction or inhibit it (Fig. 4.21). Alternatively, given that the TMD is locked by the point mutation Val329Phe in all of these structures, it is also possible that this CRD-LD pivot is instead indicative of the ciliary localisation state, with the

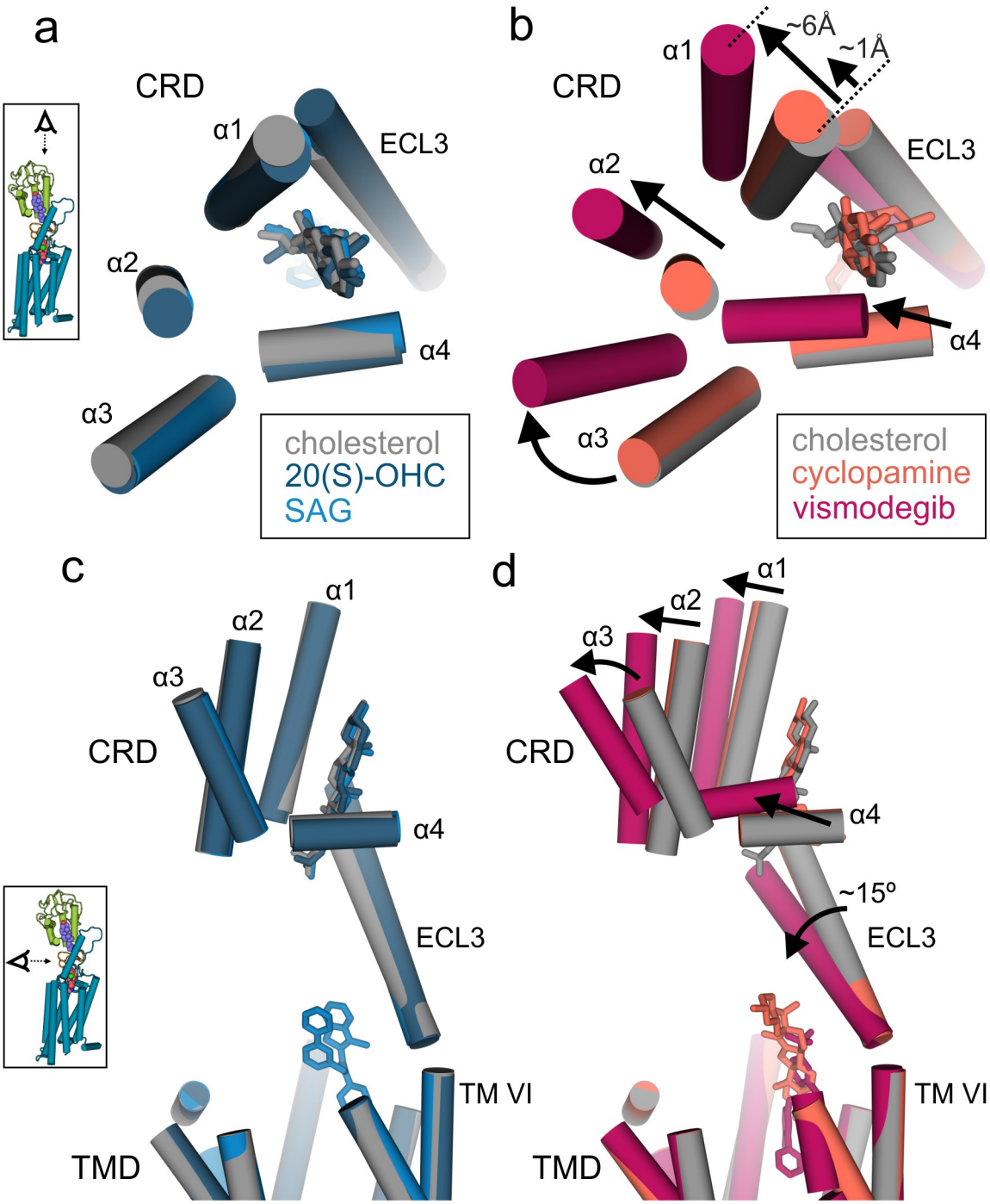


Figure 4.21: Ligand-bound structures reveal continuum of CRD states. Superpositions of ligand-bound structures aligned by their TMDs. Helices shown as cylinders, ligands as sticks. CRD helices labelled as in Fig. 3.6. Insets show viewing perspective. Arrows indicate shifted helices. **a & b** Top-down view of CRD helices for cholesterol plus agonist structures (**a**) and cholesterol plus antagonist structures (**b**). **c & d,** Side-on view of cholesterol plus agonist structures (**c**) and cholesterol plus antagonist structures (**d**).

vismodegib-bound structure as the only structure preventing ciliary localisation. Further structures of antagonist complexes combined with subcellular localisation data will be required to clarify this possibility.

4.5 Summary

Two different thermostability assays revealed cyclopamine to substantially stabilise detergent-solubilised SMO. Despite this, initial attempts to crystallise SMO in the presence of cyclopamine were unsuccessful. Crystallisation trials with other small molecule ligands were subsequently pursued. These efforts resulted in two new antagonist-bound structures, vismodegib and double-cyclopamine, and three new agonist-bound structures, SAG/cholesterol, 20(S)-OHC and 20(S)-yne. Vismodegib induces a distinct conformation of the CRD, LD and ECL3 relative to the TMD, which results in displacement of the extracellular cholesterol ligand. Cyclopamine molecules occupy both binding sites simultaneously, suggesting an explanation for its unique ability to inhibit signalling whilst inducing ciliary accumulation. The SAG/cholesterol structure demonstrates the structural characteristics of the allosteric synergy previously observed between the two binding sites. The domain positions of the SAG/cholesterol structure and the two oxysterol structures are almost identical to that of the original cholesterol structure, reinforcing the hypothesis that the cholesterol structure represents an agonist-bound conformation. From a therapeutic perspective, our results highlight an entirely unanticipated role for the CRD and ECL3 in stabilising a drug-bound inactive state of SMO, providing a structural template for the development of CRD-targeted or bitopic (dual CRD- and TMD-targeted) SMO inhibitors again Hh-driven cancers.

Chapter 5

Nanobodies bind to the extracellular domains of SMO

5.1 Abstract

Nanobodies are recombinant single-domain antigen-binding fragments of camelid-specific heavy-chain only antibodies that have in recent years become important tools for structural biology. Nanobodies have previously been generated against many different protein targets with the purpose of stabilising a specific interaction or conformation for structural and functional analysis. In order to stabilise the conformation of multidomain SMO, we generated nanobodies against the extracellular domains of SMO. One of these nanobodies binds to the extracellular domains of SMO with high affinity and forms a stable complex on size-exclusion chromatography. While crystals of the complex have been produced, they have thus far diffracted poorly. Nevertheless, the identification of a potentially stabilising nanobody that binds to the extracellular domain of SMO with high affinity is a promising step towards a full-length structure of SMO.

Llama immunisation and blood plasma screening reported in this chapter were performed by our collaborators: Els Pardon, Eva Beke and Jan Steyaert ('Steyaert Lab'). Nanobodies were cloned, expressed and purified by Jesper Levring and Rachel Woolley in STRUBI. Fluorescence microscopy experiments were performed by Rebekka Siebold-Schwab in STRUBI.

5.2 Introduction

5.2.1 Brief history of nanobodies

First discovered in 1993, camelids (camels, llamas and vicuñas) can make special heavy-chain only antibodies in addition to the more common heavy- and light-chain antibodies that are produced by the adaptive immune system in all mammals[139]. The variable region of normal antibodies consists of two interacting globular domains, one from the light chain (V_L) and one from the heavy chain (V_H), which together (V_L-V_H) form the antibody's binding surface, its paratope (Fig. 5.1, panel a). In contrast, the variable region of the camelid heavy-chain only antibodies (V_{HH}) only consists of a single domain[254]. Just as the variable domains of normal antibodies can be isolated to produce Fab fragments, the variable domain of heavy-chain only antibodies can be isolated to produce nanobodies[255]. Nanobodies can be cloned and are readily expressed in bacterial or yeast systems to produce milligram quantities of soluble, properly-folded nanobodies[255, 121, 367].

Various crystal structures of nanobodies[96, 93] revealed a consistent concave prolate (football-shaped) architecture that enables them to bind to convex epitopes of antigens, for instance binding pockets and enzyme active-sites[89, 209]. These areas could not be targeted by normal antibodies because of their largely flat and linear paratope surface. A nanobody's paratope is made up of three hypervariable regions, called the complementarity determining regions (CDR), and one of these, CDR3, can extend to produce a substantial interaction surface (Fig. 5.1, panel b)[274].

Methods were subsequently developed to generate libraries of affinity-matured nanobodies against novel and specific antigens. Briefly, this entails immunisation of camelids with purified target antigen, harvesting of 50-100 mL whole blood after suf-

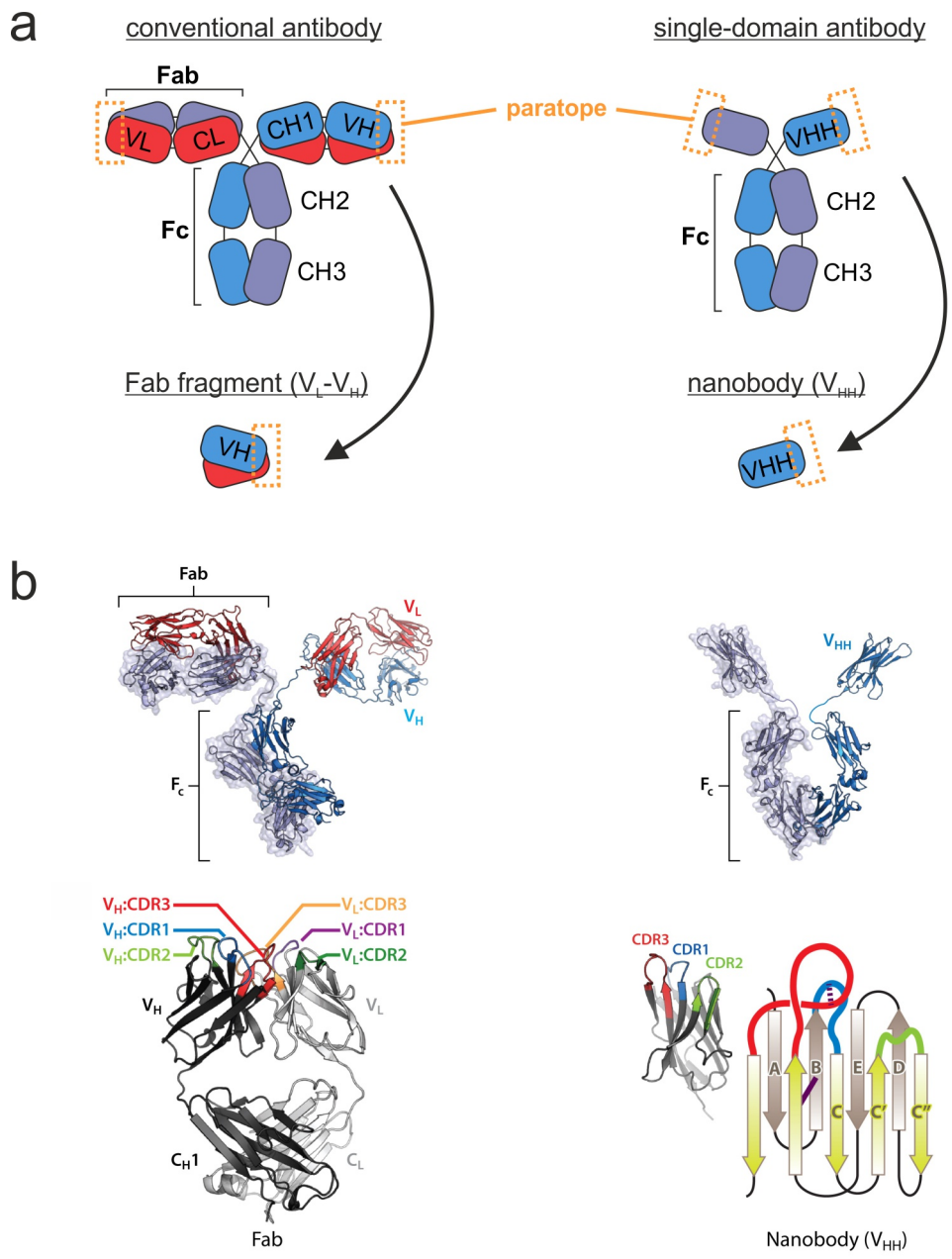


Figure 5.1: Nanobodies are single-chain antigen-binding fragments. **a,** Cartoon domain layout of conventional (left) and single-domain (right) antibodies. Paratope positions indicated with gold dotted rectangles. Adapted from [254]. **b,** Architecture of a conventional (left) and single-domain (right) antibody with epitope-binding domains. Reproduced from [230], figure 2.

ficient immune response, isolation of peripheral blood lymphocytes, generation of an mRNA library through reverse transcription PCR, amplification of nanobody-coding regions by PCR, translation and screening (panning) of nanobodies using phage display or yeast surface display techniques[274, 323, 322].

Nanobodies have been targeted for potential therapeutic use because of several special advantages they have over classic antibody fragments, including their small size, their robustness and their preference to target unique epitopes. Nanobody scaffolds have been humanised for therapeutic use in humans[374]. Nanobodies are currently being investigated for various uses in cancer diagnosis and treatment[364].

5.2.2 Nanobodies as structural biology tools

The special characteristics of nanobodies have made them very useful tools for structural studies. Aside from their small size, high solubility, high thermal and chemical stability, prolate paratope for unique convex epitopes, ease of cloning and recombinant expression, nanobodies can also sample different conformations of flexible proteins without seeming to cause induced fit[323] and they rigidify multidomain proteins[196]. Many protein structures have now been solved in complex with nanobodies (e.g. Fig. 5.2).

For soluble proteins, structures of nanobody complexes have been reported: bound to the active-site of enzymes, either with[270, 271] or without[183] bound substrate; stabilising multidomain proteins[265, 14]; stabilising the interface between two proteins[278]; inhibiting oligomeric states[2, 368, 386, 371, 372, 365]; differentiating between oligomeric states[135, 90, 101]; allosteric enzyme inhibition[340]; or simply as crystallisation chaperones[276, 277, 398, 196, 207].

For characterisation of membrane proteins too, nanobodies have proven very useful. Nanobodies have been used: to obtain crystal structures of active conformational states of GPCRs[230], such as β_2AR[297, 347, 305], muscarinic acetylcholine receptor (M_2R)[200], μ-opioid receptor (μ-OR)[150], and the viral chemokine receptor US28[33]; to study different conformations of GPCRs, such as β_2AR[345, 344, 99, 163] and μ-OR[341], or of membrane transporters, such as LacY[339, 338]; to stabilise binding partners of GPCRs, such as the heterotrimeric G-protein[299] or β-arrestin[38]; as crystallisation chaperones, as for serotonin 5-HT_3 receptor[142], SecY channel[216], Vps34[317], SLC26[127], SLC11[108], archeal mechanosensitive channel[223] and type II secretion system[195]; or as cryoEM aids, such as for the bacterial type IV secretion core complex[308] (Fig. 5.2).

5.2.3 Rationale behind nanobodies for SMO

In the case of SMO, despite having solved the structure of CRD-LD-TMD bound to multiple small molecule ligands, none of the structures showed any change in the conformation of the TMD, likely due to the locking effect of the V329F point mutation or the stabilising effect of the BRIL fusion or both. The extracellular domains, by contrast, did show conformational changes between agonist- and antagonist-bound states. We hypothesise that these different extracellular domain states would affect the TMD states if the TMD was not artificially locked in place. However, if both the extracellular domains and the TMD were allowed to move (by removing the V329F mutation and the BRIL fusion), this would likely result in too much flexibility to successfully crystallise SMO. Generating nanobodies against SMO might help to stabilise the conformation of the extracellular domains in a defined functional state, reducing their flexibility and potentially biasing the TMD toward a particular state in the absence of V329F or BRIL. Such a nanobody may aid in the structural characterisation

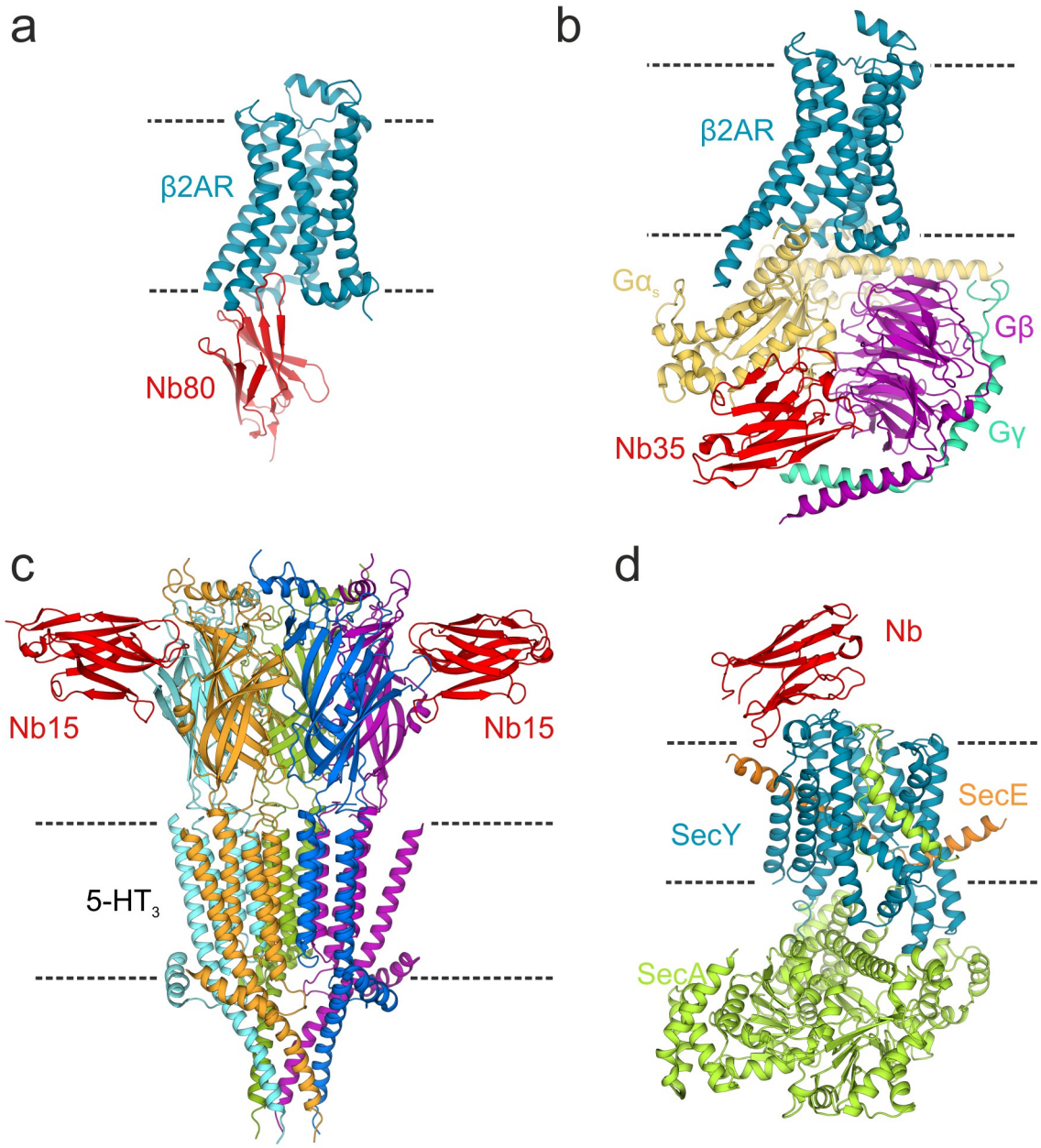


Figure 5.2: Crystal structures of transmembrane proteins solved with the aid of nanobodies. Proteins shown in cartoon representation with each chain in a different colour; all nanobodies shown in red. **a,** β_2AR solved with Nb80 bound to intracellular face. **b,** β_2AR solved in complex with heterotrimeric Gs-protein. The interface between Gα_s and Gβ is stabilised by Nb35. **c,** Pentameric 5-HT$_3$ solved with five Nb15 domains (only two shown). **d,** SecY channel complex solved with a nanobody bound to the periplasmic side. (PDB: 3P0G, 3SN6, 4PIR, 5EUL)

of the active state of SMO and even of full-length SMO.

5.2.4 Chapter outline

The aim of this chapter was to generate nanobodies that could stabilise the interaction between the extracellular regions of SMO (CRD, LD and ECL3) as a step towards solving the structure of a fully active state of SMO. We reasoned that a nanobody directed against the extracellular domains of SMO might help to stabilise the conformation by binding to flexible regions. Purified SMO CRD-LD-TMD was produced with the intracellular BRIL fusion for the immunisation of the llama and then produced again without the BRIL fusion for panning of the nanobody library. This resulted in a panel of nanobodies that were assayed for binding affinity to purified SMO using bio-layer interferometry (BLI) and for cell-surface binding using fixed-cell fluorescence microscopy. Two nanobodies, Nb4 and Nb7, bound strongly to the cell surface of SMO-expressing cells and Nb4 bound to purified detergent-solubilised SMO with a K_D of $\sim$14 nM. Both nanobodies were observed to form a stable complex with SMO by SDS-PAGE analysis after SEC. Multiple crystallisation trials of the Nb-SMO complexes resulted in 5-10 μm crystals that diffracted poorly. The crystallisation conditions require further optimisation to increase crystal size and improve diffraction quality.

5.3 Results

5.3.1 SMO protein production for nanobody generation

Two batches of purified SMO were required for nanobody generation. The first batch was used to immunise the llama and the second was used to screen the phage-displayed nanobody library for SMO-specific nanobodies. Milligram quantities of purified protein was required for llama immunisation so the SMOΔC-BRIL-V329F construct was used because it had the highest expression levels of all the constructs (Fig. 3.9, page 82). The thermostability assays with small molecule ligands showed cyclopamine to substantially thermostabilise SMO (see Fig. 4.3, page 130, and Fig. 4.4, page 131) so cyclopamine was also added to the purification in order to further stabilise the protein. The detergent DMNG was used to solubilise the cell pellet and extract the membrane protein fraction. DMNG has a very low critical micelle concentration (∼0.002% (w/v)) and tightly binds the hydrophobic membrane-spanning domain. It is therefore a robust solubilising agent in the absence of extra detergent molecules in the bulk solvent, such as in the bloodstream of an animal. This DMNG-solubilised purified protein was injected directly into the llama. Monomer peak fractions from two purifications were combined to produce sufficient protein for immunisation (Fig. 5.3, panel b). Protein was not treated with EndoF1 prior to use.

The SMOΔC-BRIL-V329F construct used for immunisation contains the artificial BRIL fusion against which we did not require a nanobody as we ultimately wanted to remove the BRIL. For initial phage display screening of nanobodies, we used the SMOΔC-V329F construct with an Avi-tag and without the BRIL fusion. Using this construct for screening, only nanobodies raised against the solvent-exposed extracellular domains or intracellular loops of SMO would be selected and any nanobodies raised against the BRIL domain would not be selected. The C-terminal Avi-tag was

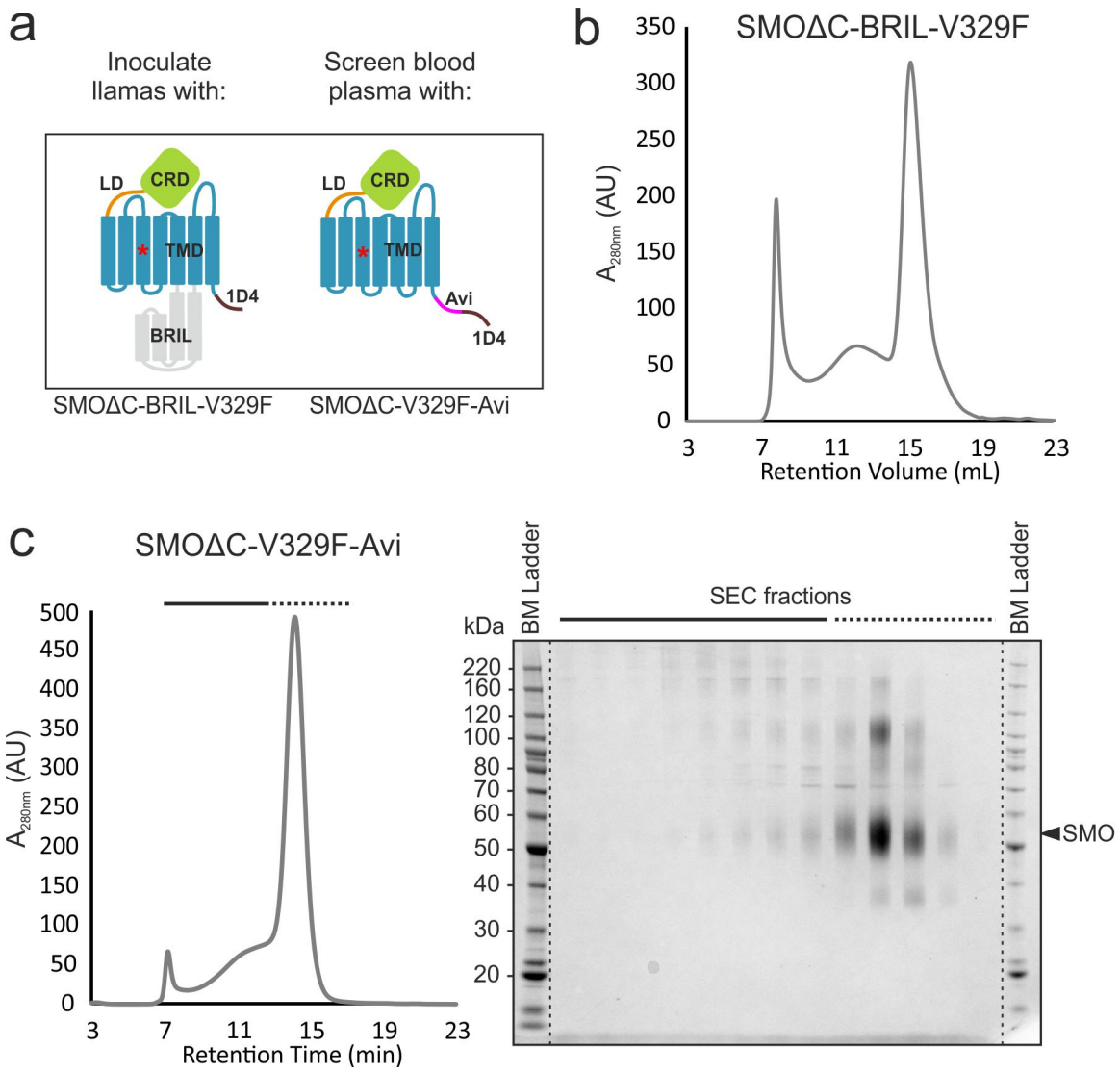


Figure 5.3: Production of SMO for llama immunisation and nanobody screening. **a,** Cartoon schematic representation of constructs used for inoculation of llama and screening of blood plasma, respectively. Coloured as in Fig. 2.1, page 26, with red asterisk indicating V329F mutation. (1D4, anti-Rho1D4 epitope tag; Avi, Avi-tag for biotinylation.) **b,** SEC trace from large-scale SMOΔC-BRIL-V329F purification. **c,** SEC trace (left) and corresponding reducing SDS-PAGE analysis (right) from large-scale SMOΔC-V329F-Avi purification.

used to biotinylate SMO for binding to streptavidin-conjugated ELISA plates. Protein purification produced a high and symmetrical monomer peak on SEC, which on SDS-PAGE was shown to consist of monomer (50-60 kDa) and some dimer (100-120 kDa) (Fig. 5.3, panel c). Protein was not treated with EndoF1 prior to use.

SMO protein was expressed and purified by myself at STRUBI. Llama immunisation, nanobody library generation and phage-display panning were conducted in collaboration with Els Pardon, Eva Beke and Jan Steyaert at the Structural Biology Research Center, Vrije Universiteit Brussel (VUB), Belgium.

5.3.2 Nanobody production

Our collaborators at VUB provided us with nine nanobodies in the pMESy4 plasmid, which included a C-terminal 6xHis-tag and CaptureSelect^{TM}C-tag (see Appendix B, page 222 for plasmid map), and the amino acid sequence for each of the nanobodies[274]. These nine nanobodies were derived from four families. A nanobody family contains nanobodies with identical length and greater than 80% sequence identity in their CDR3 sequence[274]. Nanobodies within the same family derive from the same B-cell lineage and bind to the same antigenic epitope[274]. Each plasmid was sequenced and all nanobodies were expressed in WK6*su E. coli* cells and purified by IMAC followed by SEC (see Materials & Methods, section 2.7.2, page 58). The sequence of the Nb9 plasmid did not match the provided amino acid sequence and an attempt at its purification was also unsuccessful. The other eight nanobodies were sequence-confirmed and successfully purified, with yields between 2.0-6.5 mg from 1 L bacterial cultures. A representative SEC trace and SDS-PAGE analysis from the purification of one of the nanobodies is shown in Fig. 5.4, panel b. The purified nanobodies were used for binding affinity studies and crystallisation trials.

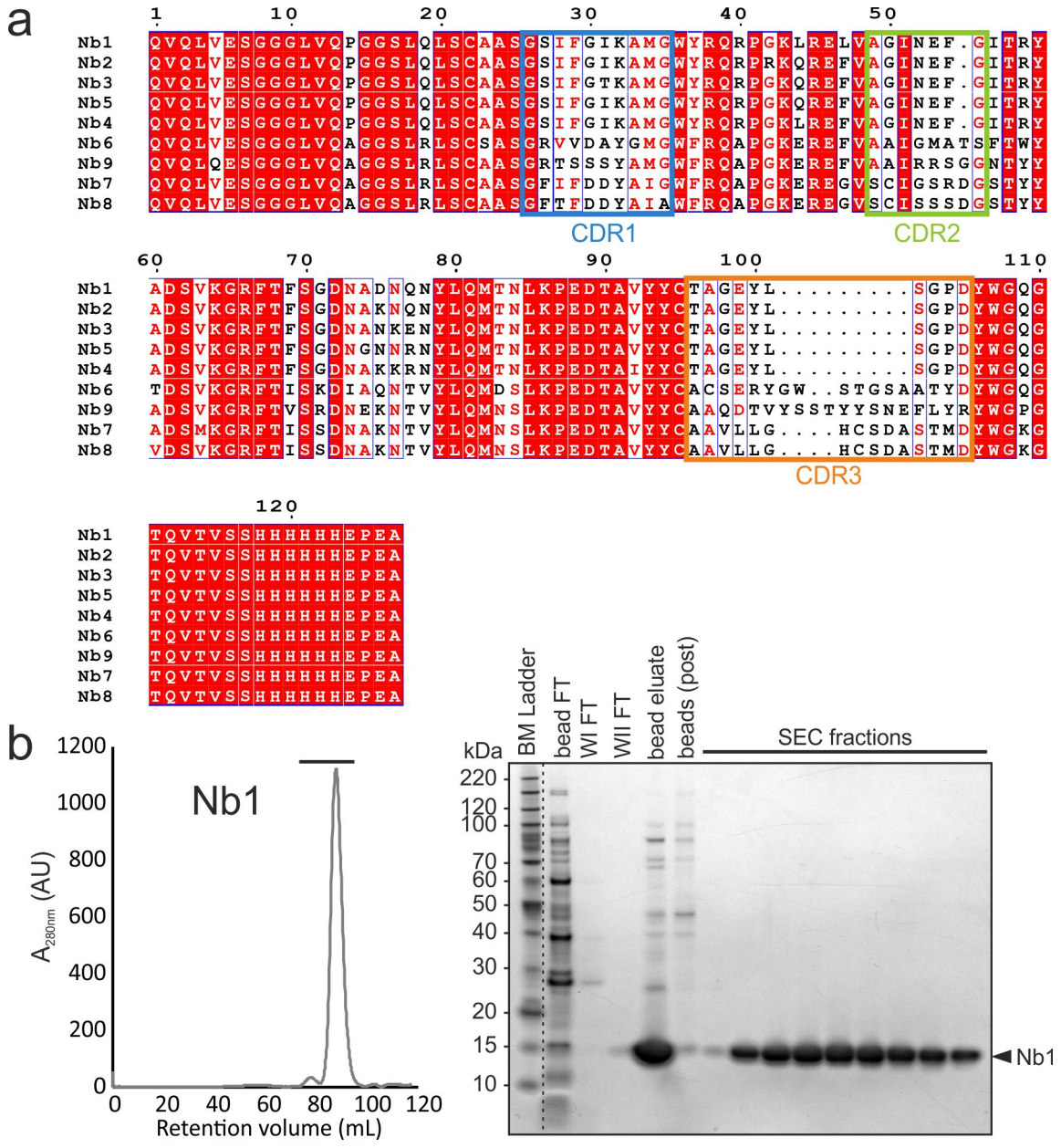


Figure 5.4: Production of nanobodies raised against SMO. a, Sequence alignment of nine nanobodies. Ordered and coloured by % equivalent residues at each position (produced using Clustal Omega and ESpript3.0, refs [334, 310]). The hypervariable CDRs are highlighted with coloured boxes. **b,** SEC trace (left) and corresponding reducing SDS-PAGE analysis (right) from purification of Nb1. Nbs 2-8 showed similar results.

The nanobodies were also cloned into the pET22b(+) vector with an N-terminal Flag-tag (see Appendix B, page 222 for plasmid map). These were expressed in Rosetta^{TM}(DE3) pLysS competent cells and purified by IMAC followed by SEC (see Materials & Methods, section 2.8.2, page 60). Purified nanobodies were used for cell-surface binding experiments using fluorescence microscopy. The majority of the nanobody production and cloning was completed by Jesper Levring as part of his undergraduate summer project in the Siebold Lab under my day-to-day supervision. Several nanobodies were also purified by Rachel Woolley during her first year DPhil rotation in the Siebold Lab.

5.3.3 Nb4 binds SMO with nanomolar affinity

Bio-layer interferometry (BLI) was used to assess the binding affinity of the purified nanobodies for purified SMO. In this technique, one binding partner, the "ligand", is immobilised onto the tip of an optical fibre and the other binding partner, the "analyte", is transiently applied to the tip. Binding of the analyte to the ligand results in a change in the reflected interference pattern of light, which is monitored by the Octet® RED384 system (Pall ForteBio). As is the case for other optical biosensor technology (such as surface plasmon resonance), on- and off-rates of analytes at different concentrations can be used to calculate the binding affinity between ligand and analyte.

SMOΔC-V329F-Avi (no BRIL, not EndoF1-treated) that had been biotinylated *in vivo* by its Avi-tag was immobilised onto streptavidin-conjugated tips, acting as the ligand in this experiment. The various nanobodies were the analytes. Biotinylated SMO was consistently well-immobilised onto the streptavidin tips, with a fast and smooth association response to between 3.0-4.0 nm after 5 min and very little

dissociation (Fig. 5.5, panel a). In the initial screening with nanobodies applied at ~1.0 μM, Nbs 1-6 showed at least some response (Fig. 5.5, panel b). Nb7 and Nb8 showed very limited and apparently irreversible response, at least on the timescale of the experiment. Subsequent experiments using dilution series of the nanobodies resulted in a range of dissociation constant (K_D) values from 12–580 nM. Nb4 bound most strongly to SMO, with a K_D of 14 ± 2 nM (Fig. 5.5, panel c). All of the other nanobodies had binding affinities of higher than 300 nM.

In these initial experiments, biotinylated SMO was solubilised in DDM/CHS micelles and all of the running buffers included 0.03% DDM, 0.006% CHS. However, the effect of these detergent conditions on the nanobodies and their binding affinities to SMO was unknown. In order to reduce this uncertainty, in a subsequent experiment the same biotinylated SMO protein was exchanged into amphipols (see Materials & Methods, section 2.6, page 55) before immobilisation on the Octet tips and the running buffers did not contain DDM or CHS, although they did include a small amount (0.0005%) of Tween-20 in order to reduce non-specific binding. A repeat of the dilution series experiment for Nb4 resulted in a very similar calculated K_D of 25 ± 3 nM, suggesting a robust high affinity binding interaction between SMO and Nb4 (Fig. 5.5, panel d).

Overall, Nb4 was clearly the tightest binding nanobody of the eight tested, displaying nanomolar affinity. However, as both the extracellular and intracellular domains were solvent-exposed in the BLI binding experiments, the question of where on SMO the nanobodies bind remained unanswered.

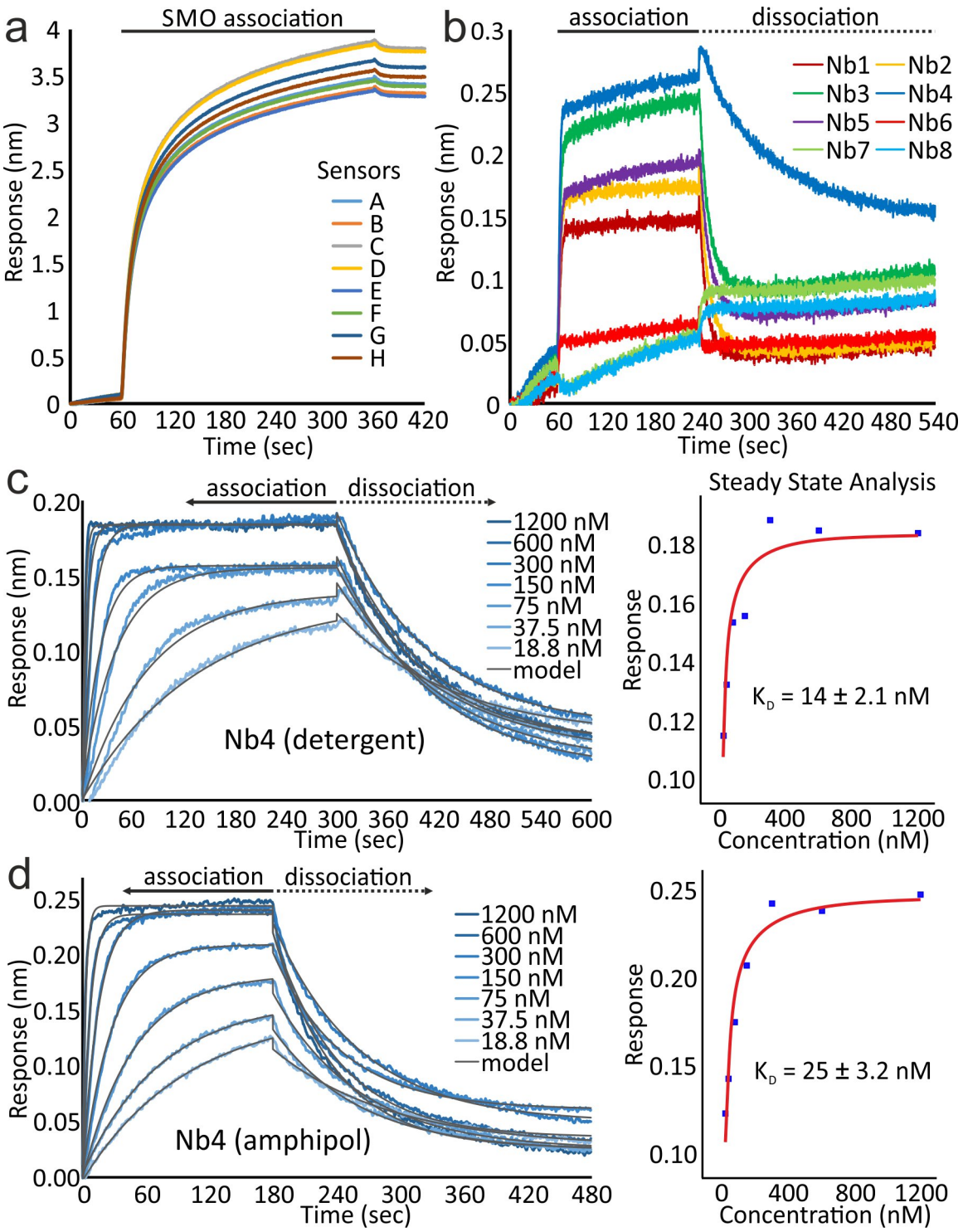


Figure 5.5: Nanobody binding studies using BLI. a, Binding of SMO onto eight identical sensor tips (representative graph). **b,** Initial test of eight nanobodies binding to SMO-conjugated tips. All nanobodies applied at 1 μM. **c** & **d** Response curves for each concentration of Nb4 dilution series with modelled association and dissociation (left), and steady state analysis for near-equilibrium of association step (right) with calculated K_D value shown. **c,** SMO solubilised in DDM detergent and with DDM present in running buffer. **d,** SMO solubilised in amphipol and without DDM in running buffer.

5.3.4 Nb4 and Nb7 bind strongly to the cell surface

Fluorescence microscopy of fixed COS-7 cells was used to assess cell-surface binding capacity of the nanobodies. COS-7 cells were transiently transfected with SMOΔC-BRIL-V329F, plated and allowed to overexpress the construct for 48 h, and then incubated with 5 μM Flag-tagged nanobody, before fixation with 4% PFA and application of rabbit anti-Flag primary antibody followed by anti-rabbit Alexa Fluor® 488 secondary antibody (see Materials & Methods, section 2.8.2, page 60). As specificity controls, cells that had not be transiently transfected with SMO and/or had not been incubated with nanobody were also subjected to the same treatment.

Compared to the no-nanobody control (top row), all eight of the nanobodies displayed at least some low level background binding to the untransfected cells (Fig. 5.6, left column). Nb8 showed substantial background binding to the untransfected cells. In cells overexpressing SMO, Nb4 and Nb7 showed the greatest increase in cell-surface fluorescence. Nb5 and Nb3 also showed noticeable increases in fluorescence. Any increase in cell-surface binding of nanobodies 1, 2 and 6 in the presence of SMO was not readily apparent. The high background fluorescence of Nb8 in untransfected cells prevented an assessment of any increase in binding to SMO-overexpressing cells.

Overall, Nb4 and Nb7 seemed to bind most strongly to the cell surface of SMO-overexpressing cells, indicating that they bind to the solvent-exposed extracellular domains of SMO. Fluorescence microscopy experiments were performed in collaboration with Rebekka Siebold-Schwab in the Siebold Lab.

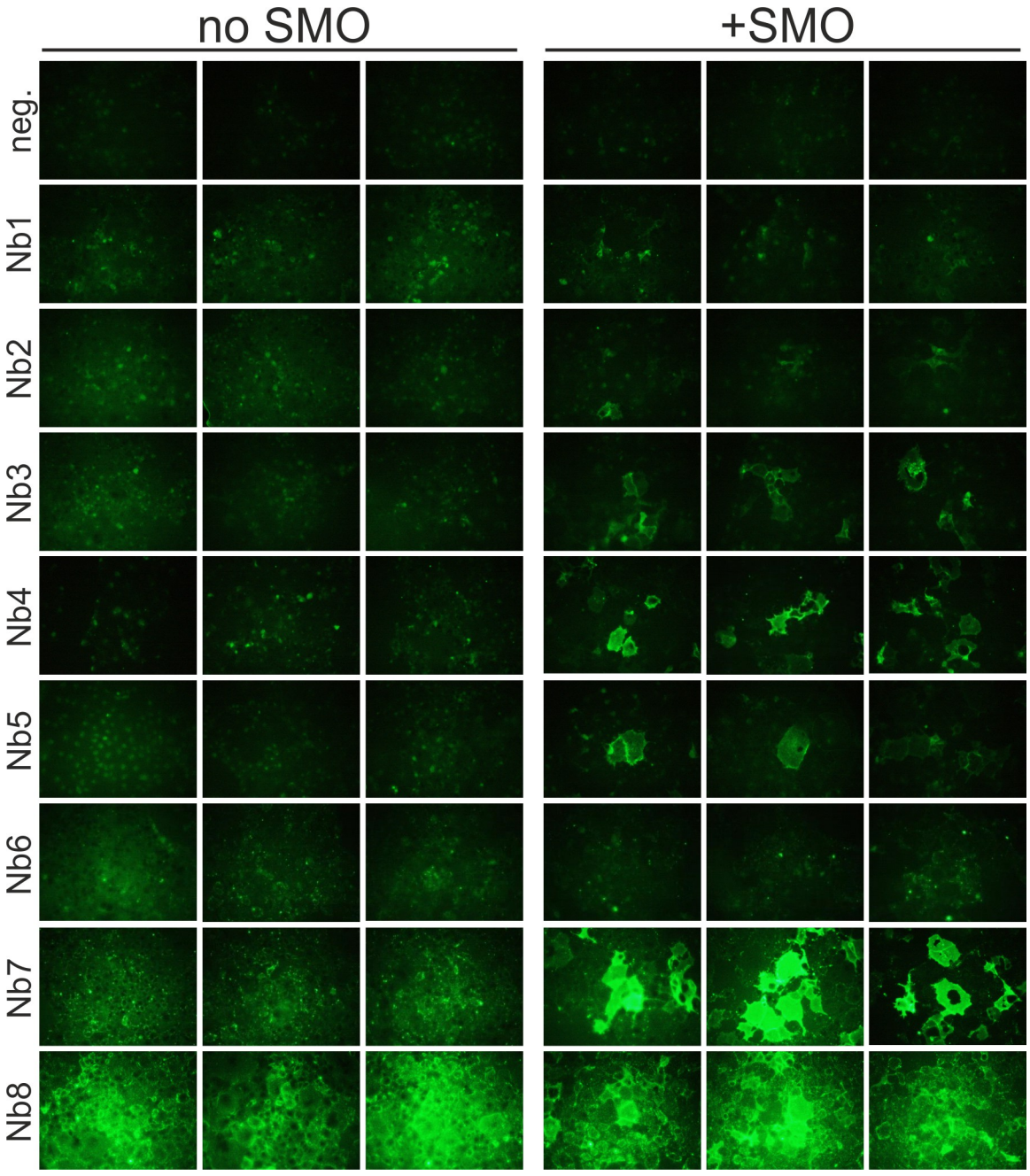


Figure 5.6: Fluorescence microscopy cellular assay of nanobodies binding to SMO. COS-7 cells transiently transfected with SMO (right) or not (left) and then incubated with purified nanobodies. Increased fluorescence in right group compared to left group indicates SMO extracellular domain specific nanobody binding. Three representative images for each sample shown. Fluorescence microscopy experiments were performed by Rebekka Siebold-Schwab.

5.3.5 Towards a SMO-nanobody crystal structure

Nb4 and Nb7 were tested for their ability to form a stable complex with SMOΔC-BRIL-V329F. A 1.5x and 2.0x molar excess of Nb4 and Nb7, respectively, were separately added to purified SMO, incubated briefly on ice, spun down and then run on SEC (Fig. 5.7). Centrifugation resulted in no observable pellets and the SEC traces showed very low aggregate peaks (at retention volume ∼7.5 mL). The two main peaks in each SEC trace were well-separated and were confirmed by SDS-PAGE analysis to be the SMO-nanobody complex and excess nanobody, respectively (Fig. 5.7). There was no SMO-only peak or shoulder in either run, suggesting that SMO was conjugated to saturation by each of the two nanobodies. Furthermore, in the SDS-PAGE analysis, the intensity of nanobody bands in the lanes from the first main peak correlated closely to the intensity of the SMO bands in the same lanes (Fig. 5.7). This suggested that each of the nanobodies was distributed evenly across the respective SMO peak and was therefore bound to SMO in a fixed ratio, rather than simply co-eluting as a result of equivalent size on the column (e.g. as a nanobody oligomer).

SMOΔC-BRIL-V329F was then expressed and purified at large scale, mixed with Nb4 at 1.05x molar excess and set up for LCP crystallisation trials in lipid consisting of 10% cholesterol and 90% monoolein at a final SMO concentration of 23–28 mg/mL. Initial screens produced small crystals that grew in several PEG 500 DME conditions at room temperature. However, when these crystal were shot, they were either salt or very poorly diffracting protein (individual images of ∼7.0 Å at best). Slightly larger crystals were grown after optimisation of crystallisation conditions but these produced similarly poor results when shot. Nb7 was also set up for crystallisation trials in the same manner and produced small crystals in two conditions, which diffracted as protein when mounted and shot. These SMO-Nb7 crystals are yet to be

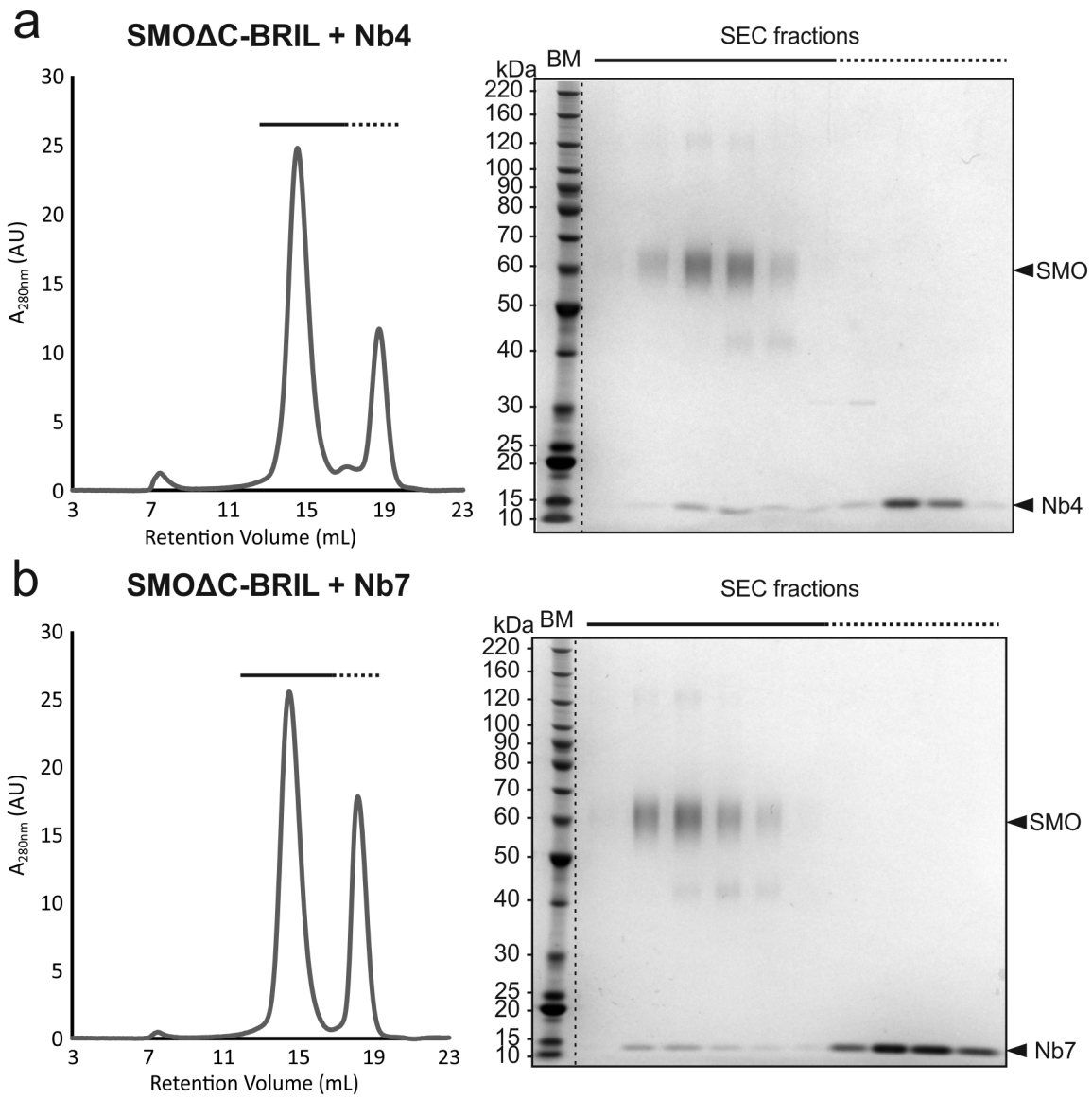


Figure 5.7: Initial mix tests of SMOΔC-BRIL-V329F with nanobodies. a, SEC trace (left) and corresponding reducing SDS-PAGE of SEC fractions (right) of purified SMO after incubation with purified Nb4. **b,** SEC trace (left) and corresponding reducing SDS-PAGE of SEC fractions (right) of purified SMO after incubation with purified Nb7. In each case, the two main peaks correspond to SMO+Nb and Nb-only. (BM, BenchMark Ladder.)

optimised.

A partial dataset of SMO+Nb4 to 7.0 Å was successfully indexed in the spacegroup P222 with unit cell dimensions of 86.3, 103.5, 117.3 Å and angles of 90, 90, 90°. Using a combined molecular mass of 84950 Da (71000 Da for SMO and 13950 Da for Nb4), Matthews Coefficient cell content analysis resulted in 99% probability of one complex per asymmetric unit with 60% solvent content. The spacegroup, unit cell and predicted cell contents are all different from the previously determined structures of SMO without nanobody (see Crystallographic Data Table, page 52). These data suggest that these crystals are not SMO by itself. Unfortunately, attempts at molecular replacement resulted in multiple indefinite solutions.

As an alternative strategy to mixing concentrated SMO with concentrated Nb immediately before setting up LCP, a 2.0x molar excess of Nb4 was added to SEC-purified SMO prior to the final concentration step for LCP crystallisation. Excess Nb4 (at ∼14 kDa) flowed through the 100 kDa MWCO concentrator but Nb4 remained in the concentrated SMO fraction, seen in a clearly discernible band on SDS-PAGE (Fig. 5.8, panel b). This preparation also produced crystals in several PEG 500 DME conditions. However, again, they did not diffract well or at all.

In a further attempt to stabilise the extracellular domains of SMO for nanobody binding, concentrated 20(S)-OHC in DMSO was added to the concentrated pre-mixed SMO-Nb4 complex in the same manner as for the small molecules additions in Chapter 4 (see section 4.3.2, page 132). Curiously, upon addition, the 20(S)-OHC solution turned into white precipitant with the consistency of toothpaste (as per usual) but then did not clarify, even after prolonged mixing. Usually, the white precipitant re-dissolves within a minute with mixing. This observation suggested that the CRD binding site, to which the 20(S)-OHC would bind and therefore dissolve, was blocked in some manner as a result of the nanobody. Upon mixing this SMO-Nb4

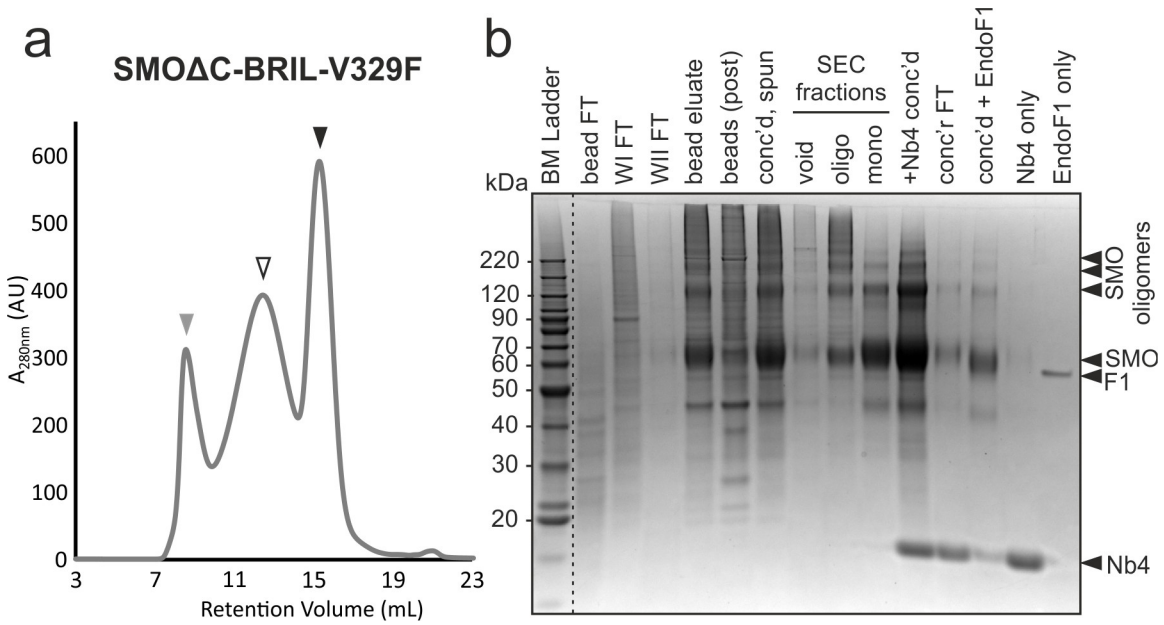


Figure 5.8: Large-scale purification of SMO with addition of Nb4. a, SEC trace from large-scale purification of SMO CRD-LD-TMD-BRIL-V329F before addition of Nb4. Grey arrowhead indicates aggregate peak, black outline arrowhead indicates oligomer peak, black filled arrowhead indicates monomer peak. **b,** Reducing SDS-PAGE analysis of large-scale purification including addition of purified Nb4 prior to final concentration step. (BM Ladder, BenchMark Ladder; FT, flow-through; W, wash; void, aggregate peak; oligo, oligomer peak; mono, monomer peak; EndoF1 or F1, Endoglycosidase F1.)

solution with the LCP lipid, the 20(S)-OHC precipitant dissolved completely, presumably because the large excess of 20(S)-OHC was taken up by the hydrophobic lipid environment. This preparation also produced small crystals in a few conditions but they diffracted poorly if at all. It is possible that adding 20(S)-OHC to SMO before adding Nb4 may prevent this solubilisation issue and allow 20(S)-OHC to successfully bind to the CRD binding site.

5.4 Discussion

5.4.1 Nanobody characteristics

While the high binding affinity and intense cell-surface fluorescence of Nb4 correlate well, the cell-surface binding results for Nb7 do not match with a similar high binding affinity in BLI. Why should this be the case? One possibility is that Nb7 has slow and irreversible binding kinetics. In the fluorescence experiments, nanobodies were applied to cells for a period of four hours before detection; in the BLI assay, nanobodies were applied to the SMO-bound tips for only five minutes (with continuous detection). In order to assess this possibility, nanobodies could be applied for longer periods in the BLI assay or shorter periods in the fluorescence assay. Longer association and dissociation steps may reveal different kinetics for Nb7 compared to the other nanobodies. Nb8 was similar to Nb7 in showing low response in the initial BLI binding experiments. As these two nanobodies are from the same family and therefore potentially bind to the same target epitope, it would not be surprising for them to show similar binding kinetics.

There are several more experiments that could be performed using the BLI assay. For example, SMO could be incubated with a small molecule agonist or antagonist and then assayed against the nanobodies to see if the small molecules induce a change in conformation that affects nanobody binding. This might also give an indication of which nanobody binds to which conformational state. In addition, nanobodies could be applied in combination (e.g. two at a time) in order to assess whether different nanobodies could bind simultaneously, which might suggest that they bind at different epitopes on SMO. In the absence of a crystal structure, nanobody SMO-epitope mapping could also be carried out using the BLI assay by mutating various locations on the solvent-exposed regions of SMO in order to introduce bulky glycosy-

lation moieties. Any introduced glycans that reduced or prevented nanobody binding would indicate the location of an epitope. Furthermore, the CRD-LD-TMD structure revealed two N-glycans in the extracellular domains of SMO. The importance of these glycans could be assessed by deglycosylating SMO before conducting the nanobody BLI assay.

An additional functional assay that could prove informative would be to assess the activation of the Hh pathway target genes (e.g. Gli) upon application of different nanobodies to SMO-expressing cells using a quantitative reverse transcriptase PCR (qRT-PCR) or luciferase assay[253]. This may reveal whether any of the nanobodies are able to activate, inhibit or modulate the activity of SMO. This assay could additionally be performed in the presence of various small molecule agonists and antagonists to assess their interactions. It could also be performed against wild-type SMO and SMO with point mutations causing various functional phenotypes (e.g. Asp473Arg for TMD activation, Asp95Ala for CRD inhibition).

5.4.2 Further steps towards a crystal structure

There are a range of options for moving towards a crystal structure of SMO with a nanobody. The original SMO preparation used for immunisation of the llama contained cyclopamine. Although the BLI assay demonstrated that Nb4 could bind to SMO with nanomolar affinity in the absence of cyclopamine, it is entirely possible that the CRD binding site forms part of the epitope of Nb4, particularly given the observation that Nb4-bound SMO cannot readily uptake (or exchange) a sterol molecule into the CRD binding site. Setting up crystallisation trials with SMO bound to cyclopamine and a nanobody may yield better diffracting crystals.

In addition, both the initial SMO preparation used for llama immunisation

and the follow-up preparation used for screening were not EndoF1-treated. However, all preparations of SMO for crystallisation trials have thus far been EndoF1-treated prior to LCP set-up. If one or both of the N-linked glycans observed in the CRD-LD-TMD structure formed part of the nanobodies' epitope then modifying the glycans may also impede proper binding. This possibility could be assessed quickly using the BLI assay (as outlined above). Interestingly, the SMO preparation that was mixed with Nb4 prior to concentration was still able to be deglycosylated, as indicated by the downward shift in molecular weight on the SDS-PAGE analysis (Fig. 5.8, panel b), suggesting that Nb4 was not binding to and therefore blocking access to the N-glycans. However, this downward shift could still be explained as deglycosylation of just one of the two N-linked glycans on the extracellular domains. As the two glycans are located on opposite sides of the extracellular domains, it is unlikely that both would be involved in the epitope of a single nanobody.

It may also be worth modifying the crystallisation conditions for the SMO/Nb complexes in order to grow larger or better diffracting crystals. MAG7.7 in the place of monoolein (MAG9.9) in the lipid mixture produces larger hydrophilic channels in LCP which may better accommodate the larger soluble domains of the complex[299, 246]. In addition, GPCRs that were solubilised in various maltose neopentyl glycol (MNG) detergents, such as LMNG and DMNG, before reconstitution in LCP have been shown to produce larger crystals than from DDM-solubilised protein[45, 291, 355, 221]. Given the promising early results for solubilisation and stabilisation of SMO with DMNG (Fig. 3.8, page 80), it may be worthwhile to try crystallising the complex from a large-scale purification with DMNG instead of DDM. One early LCP crystallisation trial of SMO-BRIL-V329F was purified in the presence of DMNG rather than DDM but the final concentration of protein was too low (∼12 mg/mL) to produce crystals. Notably, upon initial mixing of the DMNG-solubilised SMO with the monoolein/cholesterol preparation, the mixture took longer to clarify than had been

observed for DDM-solubilised SMO (∼15 min compared to ∼2 min).

When examining crystallisation hits in sandwich plate format, crystals may grow either within the central lipid bolus or within the precipitant solution surrounding the bolus or within both. Crystals within the precipitant solution are unlikely to be protein crystals because the protein is predominantly contained within the lipid phase, having been pre-mixed with the lipid[37]. These other crystals are most likely to be salt. Knowing this in advance, only conditions in which crystals were solely located in the bolus were harvested for data collection. However, many of the SMO-Nb4 crystals harvested from boluses diffracted as salt, suggesting that the concentration of salt in the bolus itself, rather than in the precipitant solution, was too high. Optimisation of the salt that was included in the protein preparation, either lessening the concentration or using a different salt, could result in fewer salt crystals.

5.5 Summary

Two different constructs of SMO (with and without BRIL) were expressed and purified at large scale for llama immunisation and nanobody library screening, respectively. Out of nine nanobodies produced in collaboration with the Steyaert Lab, eight were sequence-confirmed and successfully purified. Using BLI, one of the nanobodies, Nb4, was found to bind to SMO with high affinity. Using a fluorescence microscopy-based cell-surface binding assay, Nb4 and Nb7 were found to bind specifically to SMO-expressing cells. These two nanobodies formed stable complexes with purified SMO as observed by SEC and SDS-PAGE analysis. Crystallisation trials have so far produced crystals that diffract as protein to $\sim$7 Å.

Chapter 6

Discussion

6.1 Activation and inhibition of human SMO

6.1.1 Cholesterol as the endogenous ligand for SMO

The CRD is required for regulation of SMO signalling[258]. There is now strong evidence that cholesterol can function as a direct ligand for SMO through binding to the CRD[36, 226, 149]. Our initial SMO CRD-LD-TMD structure detailed in Chapter 3 contained a cholesterol molecule bound in the CRD binding site, mediating interactions between the CRD, LD and TMD through its ECL3. We showed cholesterol to be important for CRD stability in both MD simulations and for purified protein *in vitro*. Purified SMO binds to cholesterol-conjugated beads and can be competed off the beads by oxysterols. We further showed that the application of MβCD-solubilised cholesterol is sufficient to activate the Hh pathway. Huang *et al.* have independently demonstrated the same result[149]. They also demonstrated that cholesterol and the native pathway activator Shh synergise with each other in activating Hh signalling, a relationship not observed between oxysterols and Shh but one that would be expected for a ligand that is negatively regulated by Ptc1.

Compared to activation by the endogenous pathway activator Shh, cholesterol-based activation appears to be less potent, representing only partial agonism (Fig. 3.22, panel b, page 105). This observation could reflect the method of application of cholesterol (solubilisation in MβCD) being less efficient than the endogenous mechanism or perhaps that cholesterol is not the sole endogenous ligand for SMO. Unlike oxysterol ligands such as 20(S)-OHC, however, cholesterol is present in cells at sufficient concentrations to activate Hh signalling[402], consistent with a role as the native SMO ligand. Importantly, mutations in the CRD binding site that prevent sterol binding also impair signalling by Shh (Fig. 3.21 and ref.[258]), whereas mutations in the TMD binding site that block binding of synthetic ligands, such as vismodegib, do not

affect basal or Shh-stimulated activity of SMO[256, 407, 149]. These findings suggest that the endogenous mechanism of SMO activation relies upon the CRD binding site and not upon the TMD binding site, despite the importance of the TMD as a drug target.

Xiao *et al.* recently published their finding that SMO not only binds to cholesterol but is actually covalently modified by cholesterol[401]. This claim was primarily supported by a whole-cell pull-down assay using a cholesterol probe and tandem mass spectrometry analysis that showed a modification on Asp95 that corresponded to the molecular mass of cholesterol. This is a surprising result as we did not observe a covalent bond between Asp95 and the 3β-hydroxyl group of cholesterol in our crystal structure (Fig. 6.1). Why might they find a covalent bond where we did not? Xiao *et al.* used full-length constructs of SMO with no additional modifications, whereas we used SMO without its intracellular domain and with a BRIL fusion and the Val329Phe stabilising point mutation. These alterations may have somehow prevented formation of the covalent bond in our case, for instance if the intracellular domain of SMO is involved in recruiting an enzyme that catalyses the formation of the covalent bond. Xiao *et al.* did not address the question of how the covalent bond is formed, except to say that autocatalysis is unlikely given that the molecular mass of SMO is unchanged after formation of the bond. The presence of a covalent bond should make the cholesterol very difficult to displace and severely decrease its rate of dissociation. Curiously, no previous studies with small molecule ligands, such as oxysterols that bind the CRD at the same place as cholesterol or TMD antagonists that allosterically occlude the CRD binding site, have reported any difficulties with binding that might be expected for outcompeting a covalently-bound ligand. Perhaps cholesterol-modified SMO is relatively scarce in the cell or turns over relatively quickly and is therefore easy to be overlooked in cellular assays. In the case of short-lived species of SMO, it is not clear what advantage a covalently bound cholesterol

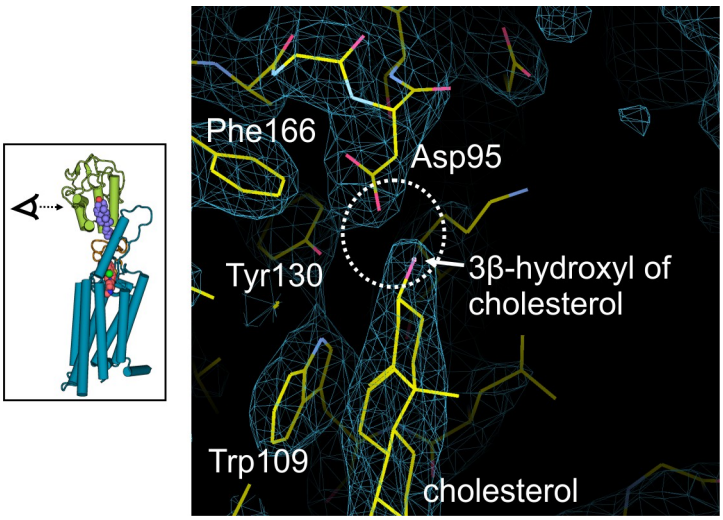


Figure 6.1: No covalent bond between Asp95 and cholesterol in crystal structure. Electron density map (blue net, $2F_o - F_c$ at 1.0 σ contour) with atomic model (yellow lines with atomic colouring) of cholesterol binding site. Dotted white circle indicates position of proposed covalent bond between Asp95 sidechain and the 3β-hydroxyl of cholesterol. Image produced with Coot[111]. Inset shows viewing perspective.

would have over a non-covalently bound cholesterol. In any event, the presence of a covalent bond between cholesterol and Asp95 should not fundamentally alter the action of cholesterol as a SMO ligand and also does not bear upon the question of how cholesterol comes to be in the CRD binding site in the first case.

Given that mature Shh is conjugated with cholesterol and that cholesterol can activate SMO signalling, an obvious question to ask is whether Shh can directly activate SMO in the same manner that Wnt, with its palmitoleic acid conjugate, can activate Fzd[165]. This seems unlikely, however, as extensive functional experimentation has shown that Shh interacts directly with Ptc1 and not with SMO[234, 122, 349, 59]. In addition, the cholesterol moiety of Shh has been shown to bind to the membrane in order to limit the dispersion of Shh[217, 31], rather than operating as a ligand itself. This is supported by the finding that Shh without the cholesterol moiety remains functional, albeit with disturbed tissue distribution[151]. Furthermore, the cholesterol molecule that is conjugated to mature Shh is bound

via its 3β-hydroxyl group, the same hydroxyl group that must be available to bind to Asp95 in the CRD binding site of SMO. This particular chemical characteristic therefore precludes simultaneous binding of one cholesterol molecule to both Shh and SMO.

Ptc1 has been linked to cholesterol in many ways. Ptc1 contains a transmembrane sterol-sensing domain that seems to influence vesicular trafficking[350]. Ptc1 shows sequence similarity to NPC1, an intracellular cholesterol transporter, the N-terminal domain of which was solved in complex with cholesterol[15]. Ptc1 can bind cholesterol directly and affects cholesterol efflux in cells[22]. It is not yet clear, however, how Shh binds to Ptc1, how Ptc1 modulates cholesterol levels or how this modulation then affects SMO. Given the finding that cholesterol is a SMO agonist and that Ptc1 acts catalytically to inhibit SMO, one possibility is that Ptc1 continually removes cholesterol from the SMO CRD, preventing its activation. In this case, binding of Shh to Ptc1 would prevent Ptc1's removal of cholesterol from SMO, allowing SMO to activate downstream signalling. This mechanism of cholesterol shuttling is reminiscent of the handing off of cholesterol from NPC2 to NPC1, two proteins that are distantly related to Ptc1 and SMO. A variation on this hypothesis is that Ptc1, within the plasma membrane compartment of the primary cilium[314], maintains low levels of cholesterol in the membrane, thereby preventing SMO from being activated. Validation of these hypotheses will require more information on a number of factors: What is the concentration of cholesterol within the primary cilium and how does this change with activation of the Hh pathway? Does the SMO CRD obtain cholesterol via Ptc1? If so, how does it do this? If not, how does the SMO CRD obtain cholesterol in an endogenous setting, particularly given the substantial distance between the natural cholesterol reservoir of the plasma membrane and the CRD binding site observed in our structure (Fig. 3.15, page 95)?

6.1.2 Membrane composition as localisation indicator

While the lipid composition of the ciliary membrane is known to be different from the rest of the plasma membrane[252, 210], the relative concentration of cholesterol is not well characterised. Several studies from the 1980's hinted that ciliary membranes may be enriched in cholesterol[46, 251, 342] but more recent work has focused on the distribution of other lipids. In examining the lipid composition of primary cilia with respect to Hh signalling, two groups found that two phosphoinositides, PI(4)P and PI(4,5)P_2, help to define different membrane regions along the length of the primary cilium and that these regions are maintained by the phosphoinositide 5-phosphatase, Inpp5e, which converts PI(4,5)P_2 into PI(4)P (ref.[126, 49]). However, removal of Inpp5e did not affect Ptc1 or SMO trafficking, instead allowing increased ciliary entry of Gpr161, an orphan GPCR that restrains Hh signalling at the Gli stage. Subsequently, Jiang *et al.* showed that application of PI(4)P promotes pathway activation and SMO ciliary accumulation, and correlates with phosphorylation of SMO's intracellular domain[167]. Accordingly, ciliary localisation of Ptc1 was decreased by PI(4)P treatment. Ptc1 was also able to bind PI(4)P-conjugated beads (the bead-binding ability of vertebrate SMO was not mentioned).

Kim *et al.* found that a modified form of Ptc1 that is not able to exit the primary cilium can still respond to Shh to allow SMO activation[188]. While ciliary localisation of Ptc1 is necessary for the suppression of SMO[314], the removal of Ptc1 from the primary cilium upon Shh binding is not essential for releasing its inhibition of SMO. Fan *et al.* demonstrated that ciliary accumulation of SMO is not absolutely necessary for activation of Hh signalling[114]. Furthermore, Bijlsma *et al.* propose that SMO can be activated outside of the primary cilium and that this activity, avoiding the transcriptional response of cilia-localised Gli proteins, induces cytoskeletal reorganisation associated with chemotaxis[23]. Aanstad *et al.* found that

a point mutation in the CRD of zSMO, Cys125Tyr, prevented its ciliary localisation and its activation by the synthetic agonist, purmorphamine[1]. This point mutation corresponds to residue Cys147 in hSMO, one of the conserved cysteines in the CRD and one half of the disulfide bond labelled "5" in Fig. 3.6 on page 76 and Fig. 3.15 on page 95.

Several groups working on the *Drosophila* Hh pathway have implicated lipids in the regulation of dSMO. Shi *et al.* found that clustering of dSMO is necessary for high level Hh signal transduction[332]. Khaliullina *et al.* proposed a model in which dSMO trafficking is regulated by Ptc through its control of a dSMO-inhibitory lipid[182]. Specifically, in endosomes, Ptc could facilitate the transfer of a lipid from lipoprotein particles into the membrane, which then destabilises SMO and leads to its degradation. Ptc with a non-functional sterol sensing domain (SSD) led to the accumulation of dSMO, agreeing with previous work showing that a functional Ptc SSD is essential for repression of dSMO[236, 350]. Interestingly, Khaliullina *et al.*'s data do not support a role for bulk membrane sterol in dSMO regulation because membrane sterol concentration can be reduced sixfold by dietary depletion without affecting Hh signalling or levels of dSMO present on the basolateral membrane. However, dSMO does not bind to oxysterols[258, 263] and the site in the CRD corresponding to the oxysterol binding site is rearranged in dSMO, likely precluding binding of sterol molecules (Fig. 3.18, panel e, page 98).

6.1.3 Synergy between two allosterically-linked binding sites

SMO can be activated or inhibited by small molecules targeted to either of two physically separable binding sites, one at the CRD and another at the TMD[257, 53]. While these can act individually, the two sites can also work in concert. The TMD

agonist SAG and the CRD agonist 20(S)-OHC have been shown to synergistically activate SMO[257]. The SAG/cholesterol structure that I presented in Chapter 4 demonstrates that these two ligands can bind SMO simultaneously, supporting a mechanism of allosteric synergy. Although it was originally believed that cyclopamine bound only to the TMD, it has recently been shown that cyclopamine also has some binding affinity to the CRD[258]. Cyclopamine is the only known ligand that inhibits signal transduction whilst also inducing ciliary localisation. In all other cases, ciliary localisation is paired with pathway activation. The double-cyclopamine structure that I presented in Chapter 4 shows that cyclopamine can occupy both binding sites at the same time. Different roles have been suggested for SMO's different domains (e.g. role for the ECD in ciliary localisation)[1]. The double-cyclopamine structure is consistent with a model in which cyclopamine binding to one domain induces ciliary localisation, while cyclopamine binding to the other domain inhibits signalling. As mentioned in the discussion section of Chapter 4 (page 160), Huang *et al.* found that when SMO was mutated to prevent ligands from binding to the TMD binding site cyclopamine acted as an agonist rather than as an antagonist[149]. These findings, in combination with the double-cyclopamine structure, suggest that cyclopamine is in fact acting as both an agonist and an antagonist at the same time.

Vismodegib and budesonide have been shown to combinatorially antagonise SMO[383]. The vismodegib structure that I presented in Chapter 4 does not contain a second ligand (such as budesonide) bound to the CRD. However, given the other double ligand structures, it may be possible for vismodegib and budesonide to bind at the same time in order to synergistically antagonise SMO through allostery. SANT-1 and SANT-2 have each been shown to interact competitively but allosterically with 20(S)-OHC[257]. From this information, I would predict that a SMO CRD-LD-TMD structure in complex with either of these antagonists would closely resemble that of vismodegib, where the CRD binding site is occluded and allosterically prevents the

binding of a CRD ligand, such as 20(S)-OHC or cholesterol.

6.1.4 Domain conformations and activation states

What conformational differences can we see between these agonised and antagonised structures? The main difference between these structures is a shifting of the entire CRD in relation to the TMD, with the LD acting as a hinge for this movement (Fig. 4.21, page 162). The position of the CRD relative to the TMD seems to reflect the activation state of SMO, with the more upright structures corresponding to the agonist-bound conformations (panels a & c) and the tilted-over structures corresponding to the antagonist-bound conformations (panels b & d).

A recent paper from Zhang *et al.* reported another crystal structure of human SMO CRD-LD-TMD bound to a super-stabilising antagonist, called TC114 (a derivative of the previously reported LY2940680)[411]. As expected, this structure, with a single TMD-bound ligand, demonstrated the same overall architecture for SMO as our structures, with the CRD positioned on top of the LD and braced by the ECL3 helical extension of TM helix VI (Fig. 6.2, panel a). Their TC114 structure more closely resembles our vismodegib structure than our cholesterol structure, with no small molecule ligand in the CRD binding site and instead with several residues of ECL3 intruding into the hydrophobic pocket (Fig. 6.2, panel b). Curiously, while in the cholesterol and vismodegib structures the NAG at Asn493 seems to shield the CRD binding site, in the TC114 structure this residue and the attached NAG point in the opposite direction. The ECL3 helix of the TC114 structure is positioned at the same angle relative to the TMD as in the vismodegib structure, tilted some 13° compared to the cholesterol and other agonist-bound structures (Fig. 6.3, panel a & b). In the same manner as for vismodegib, this arrangement seems to be stabilised by

a hydrophobic interaction between Phe484 on ECL3 and the aromatic ring of TC114 (Fig. 6.3, panel c). The loop region of ECL3 beyond the helix appears very flexible, with a substantial rearrangement in the TC114 structure relative to the agonist structures (Fig. 6.3, panel d). This same region was not able to be assigned in the vismodegib structure due to poor density as a result of flexibility. From the available structures, it is clear that ECL3 can take multiple different conformations and interacts with the CRD in a variety of ways. This is consistent with it playing a role in mediating signal transduction. ECL3 also interacts with both the CRD and TMD binding sites, forming an important allosteric link between them. Furthermore, ECL3 is the extension of TM helix VI, which in other GPCRs propagates the mechanical signal from the orthosteric binding site to the intracellular face (Fig. 3.2, page 68).

Zhang *et al.* solved two structures of the multidomain SMO with TC114, one using an X-ray free electron laser (XFEL) source and one using a standard synchrotron radiation source[411]. The former structure contained a single point mutation in the LD, Glu194Met, that improved protein yield and increased constitutive and Shh-stimulated activity in a Gli-luciferase assay but did not affect stability. This residue mediates contacts between the LD, CRD and ECL3, including with the Asn493-linked NAG (Fig. 6.2, panel c). This sugar moiety was built into the Glu194Met/XFEL structure but not into the wild-type/synchrotron structure, perhaps owing to the slightly higher resolution (2.9 Å versus 3.0 Å) and the higher completeness (100% versus 91.9%) of the Glu194Met/XFEL structure or to the increased stability of the NAG-methionine interaction compared to the NAG-glutamate interaction. In any event, the increased activity of this Glu194Met mutation in the Gli-luciferase assay suggests an important regulatory role for this residue, which is consistent with its position within the structure.

The new TC114 structures, combined with the cholesterol- and vismodegib-

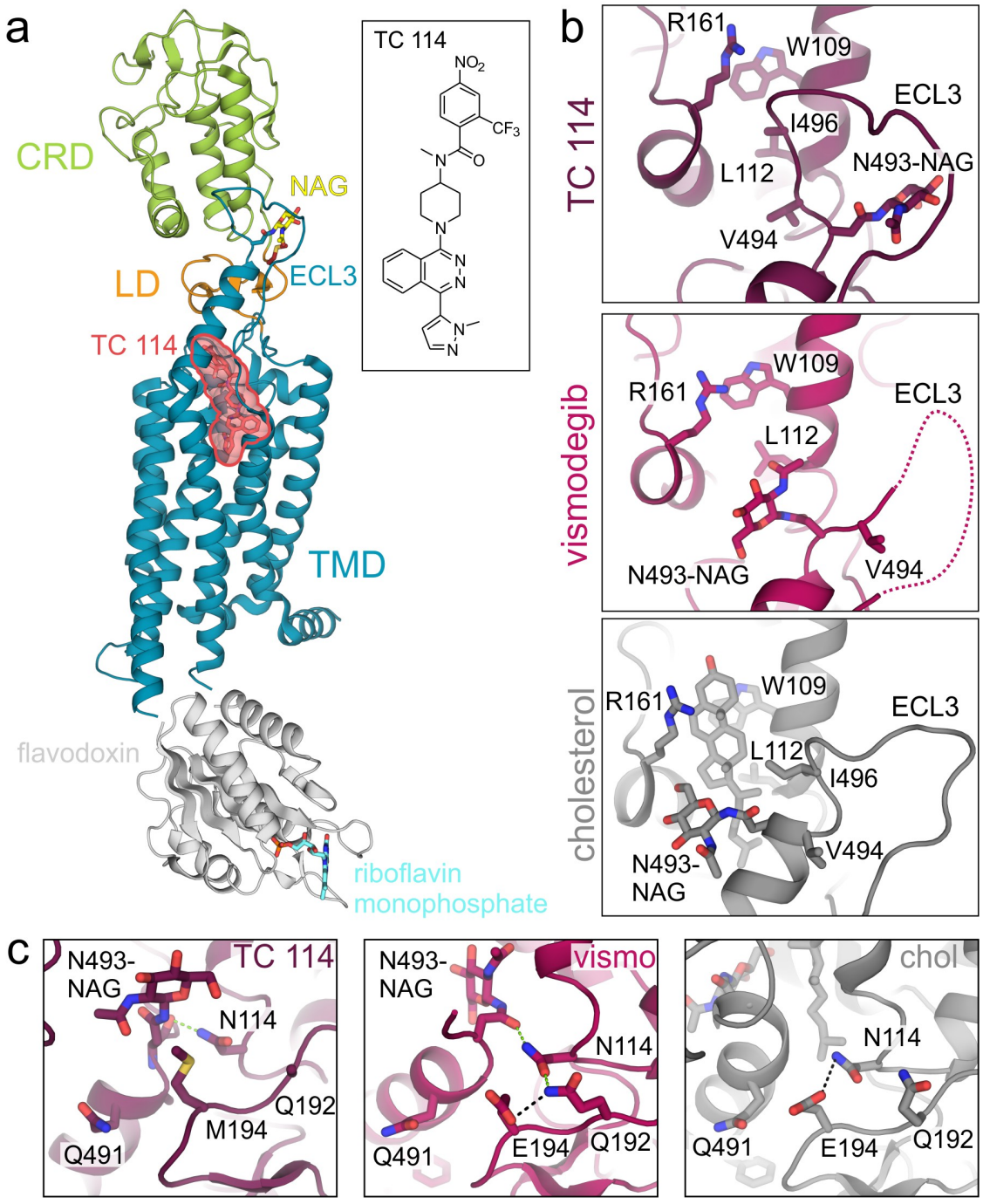


Figure 6.2: Independently solved antagonist-bound structure of SMO closely resembles vismodegib structure. a, Overall structure of TC114-bound SMO in cartoon representation. TC114 in red, flavodoxin fusion in grey, riboflavin monophosphate conjugate in cyan sticks. Other colours follow Fig. 3.15. Inset shows chemical structure of TC114. (PDB: 5V56) **b,** Close-up of ECL3 interaction with sterol binding site for TC114, vismodegib and cholesterol structures. Important residues are shown as sticks with atomic colouring and labelled. **c,** Interactions of LD residue Glu/Met194 with ECL3 and CRD for TC114, vismodegib and cholesterol structures. Dotted green lines indicate potential hydrogen bonds, dotted black lines indicate potential salt bridges.

bound SMO structures previously published by us, led Zhang *et al.* to propose an activation mechanism for SMO based upon the position of the CRD relative to the TMD. This hypothesis is consistent with our findings that agonist-bound structures cluster very closely to the cholesterol structure whereas in antagonist-bound structures the CRD is shifted relative to the TMD (Fig. 4.21, page 162, and Fig. 6.3). However, while the ECL3 helix and the α4 CRD helix overlay very closely between the vismodegib and TC114 structures, the α1-3 helices of the CRD in the TC114 structure are rotated several degrees counterclockwise (when viewed from above) relative to those of the vismodegib structure (Fig. 6.3, panel a). This whole-domain rotation may be the result of the different interaction mode between ECL3 and the CRD, as discussed above, with the Asn493-linked NAG either in front of the CRD binding site in the vismodegib structure or turned back behind the ECL3 in the TC114 structure.

Despite the fact that cyclopamine is an antagonist, the double-cyclopamine structure shows a CRD and ECL3 position much closer to that of the agonist-bound cholesterol structure than to the antagonist-bound vismodegib or TC114 structures. Why should this be the case? Although the double-cyclopamine structure has thus far only been solved to 3.70 Å resolution, such a large domain reorientation, if it occurred, would still be readily observable at this resolution. Cyclopamine is an unusual SMO antagonist in that it can block Hh signalling whilst at the same time inducing SMO accumulation in the primary cilium, an ability that is normally indicative of SMO activation[258]. Given that the CRD has been implicated in SMO trafficking to the primary cilium[1], it is possible that cyclopamine binding to the CRD site induces ciliary accumulation whilst simultaneous cyclopamine binding to the TMD site blocks signal transduction. This hypothesis is consistent with Huang *et al.*'s finding that cyclopamine can act as an agonist for SMO when SMO is mutated such that it cannot bind to TMD ligands[149]. Unfortunately, it is not possible to validate this intriguing hypothesis without a full understanding of how the TMD conformation

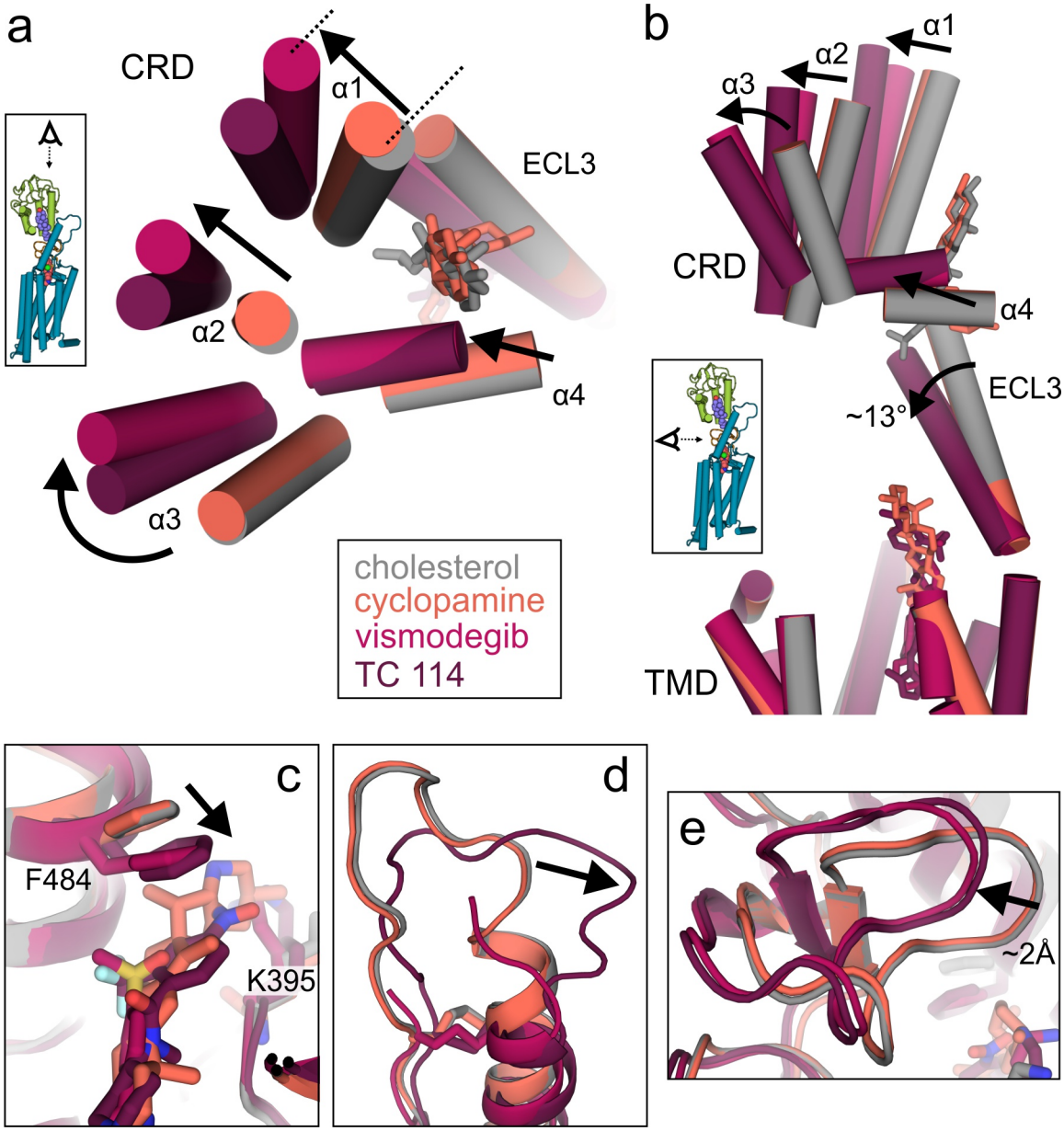


Figure 6.3: Distinct CRD position in antagonist-bound structures. a & b, Superpositions of ligand-bound CRD-LD-TMD structures aligned by their TMDs. Top-down view of CRD (**a**) and side-on view of CRD, LD and extracellular side of TMD (**b**) helices for cholesterol and antagonist structures. Helices shown as cylinders, ligands as sticks. CRD helices labelled as in Fig. 3.6. Insets show viewing perspective. Arrows indicate shifted helices. **c,** Close-up of Phe484 interaction with TMD ligand. **d,** The end of ECL3 is shifted in the TC114 structure. **e,** The LD of TC114 and vismodegib structures is shifted ∼2 Å compared to cholesterol and cyclopamine structures.

shifts in response to activation. Similarly, the effect of CRD position and ECL3 angle relative to the TMD cannot be fully understood without an active-state structure of the SMO CRD-LD-TMD.

An independently-generated multi-domain SMO structure, with a different intracellular fusion (flavodoxin rather than BRIL), no stabilising TMD point mutation and a different antagonist ligand, has reproduced many of the structural observations from our cholesterol and vismodegib structures. Based upon these structures and the double-ligand structures presented in this thesis, the position of the CRD relative to the TMD can be altered by the binding of ligands to either the CRD or the TMD. The effect of agonists and antagonists can be grouped by the relative position of the CRD. Furthermore, cyclopamine, a ligand that shows characteristics of both an agonist and an antagonist, shows a distinct conformation that suggests a tension between the two states. ECL3 appears to play a prominent role in mediating interactions between the three domains, CRD, LD and TMD, as well as between the two binding sites. TM helix VI, as the transmembrane extension of the ECL3 helix, seems most likely to transduce the Hh signal across membrane. However, a structure of SMO with an "active open" TMD will be required to validate this theory.

6.1.5 Towards an active state structure of SMO

While the BRIL fusion and Val329Phe stabilising point mutation facilitated the expression and stabilisation required for generating a crystal structure of human SMO CRD-LD-TMD, these alterations preclude the generation of an "active open" structure. The intracellular BRIL fusion, attached between TM helices V and VI, acts to stabilise the two helices that are most likely, based upon Class A GPCR studies, to be rearranged with activation and thereby to transduce the signal to intracellu-

lar effector molecules. The Val329Phe mutation, located at the base of the TMD binding pocket closest to the intracellular side, causes complete inactivation of SMO signalling[36]. From our structures, the additional steric bulk of the phenylalanine seems to fill the available space at the base of the pocket and reduce conformational flexibility between the TM helices, essentially locking the TMD in an inactive state. Removing both of these alterations is likely to be necessary in order to facilitate the solving of an "active open" structure of SMO. However, without these alterations, low expression levels and low stability due to conformational flexibility are likely to plague any efforts at structural characterisation. Other ways to boost expression and stabilise the "active open" conformation must be sought. For example, SMO could be expressed in the presence of a stabilising small molecule such as cyclopamine or vismodegib (as was done by Zhang *et al.* [411]) or co-expressed with a stabilising nanobody.

Previous work has suggested that GPCRs exist in multiple rapidly interchanging conformational states, most of which are inactive, even when bound to agonists[229, 192]. This means that in order to solve an "active open" structure, the GPCR must be locked into the required state with an intracellular effector molecule (e.g. heterotrimeric G protein) or a mimic of such (e.g. a nanobody or mini-G protein). In past, activated states of Class A GPCRs have been solved with either a nanobody[305], a nanobody-stabilised heterotrimeric G protein[298] or an engineered miniature G protein ("mini-G")[40] bound to the intracellular face. These were all solved with the workhorse GPCRs, $A_{2a}AR$ or β_2AR, neither of which contain an extracellular domain. Producing active-state structures of additional GPCRs using these methods has proven to be very difficult.

SMO has been shown to activate heterotrimeric G proteins[307]. There are several different subtypes of G proteins and the previous active-state studies with

$A_{2a}AR$ were performed with the $G\alpha_s$ subunit or the mini-G_s module. SMO does not bind to $G\alpha_s$, instead binding to $G\alpha_i$ or $G\alpha_o$[307]. Producing stable heterotrimers with these subunits has proven to be more difficult than for $G\alpha_s$ and engineering mini-G proteins out of these subunits has also been difficult[264]. β-arrestin is another intracellular effector molecule that binds directly to Class A GPCRs and has also been implicated in SMO action[56, 197]. However, evidence is limited that β-arrestin binds directly to SMO in the same manner as Class A GPCRs.

Recently, single-particle cryo-electron microscopy (cryo-EM) has been used to determine the active-state structures of two Class B GPCRs, the calcitonin receptor[218] and GLP-1[412], in complex with the heterotrimeric $G\alpha_s\beta\gamma$ protein stabilised by a nanobody. These structures tell us that it is now feasible to pursue structures of GPCRs using cryo-EM methods, albeit in complex with several intracellular proteins in order to increase particle size.

An active-state structure of SMO will require the removal of the intracellular BRIL fusion and the removal of the Val329Phe stabilising point mutation. It will also most likely require the intracellular face of SMO to be bound with an effector molecule or effector mimic. This could perhaps be the heterotrimeric G_i protein, a mini-G_i protein or a nanobody. In order to strongly stabilise the active state of SMO, the TMD should be bound with a high-affinity agonist, such as SAG, and the CRD should similarly be bound with an agonist, such as 20(S)-OHC. If possible, the extracellular domains should also be stabilised by a nanobody, such as Nb4. For crystallisation of such a complex, it may be necessary to use a different lipid mixture in order to accommodate the larger soluble domains. A MAG7.7/cholesterol mixture has been used in past instead of monoolein/cholesterol to generate larger hydrophilic channels in LCP and successfully crystallise a full GPCR-G protein complex[299, 246]. This multi-drug, multi-protein, multi-domain SMO complex could then be set up for LCP

crystallisation or for single-particle cryo-EM.

6.1.6 Models of SMO activation and Ptc interaction

The crystal structures of SMO CRD-LD-TMD that have now been solved in complex with various agonists and antagonists allow us to propose a model for the mechanism of SMO signal transduction. The vismodegib and TC114 structures represent two possible antagonist-bound closed states, both with their ECL3 helix bent down closer to the TM bundle and their CRD-LD pushed backwards (Fig. 6.4, left structure). The cholesterol, SAG/cholesterol, 20(S)-OHC and 20(S)-yne structures represent agonist-bound closed states, all of which show a more upright conformation of ECL3 and the CRD-LD (Fig. 6.4, middle structure). To move from an agonist-bound closed state to an open state will most likely require the binding of an intracellular effector protein, such as a G_i protein (Fig. 6.4, right structure). This model is limited in that it does not propose a mechanism by which SMO is localised to the primary cilium, even though the double-cyclopamine structure offers tantalising clues. However, more information on the structural details of the ICD will be required to properly address this question (for further discussion on this, see section 6.4.1 on page 214).

The position of cholesterol within the binding site of the solvent-exposed CRD in SMO, some 12 Å distant from its natural reservoir of the lipid bilayer (Fig. 3.18, page 98) and requiring a substantial amount of energy to desorb from the membrane[185], implies that some process other than diffusion must mediate its transfer. This may be the point at which Ptc1 interacts with SMO. One possibility is that Ptc1 constitutively prevents cholesterol from accessing the CRD of SMO until binding of Shh deactivates Ptc1, allowing cholesterol into the CRD and leading to SMO activation (Fig. 6.5, panel a). Alternatively, Ptc1 could prevent cholesterol

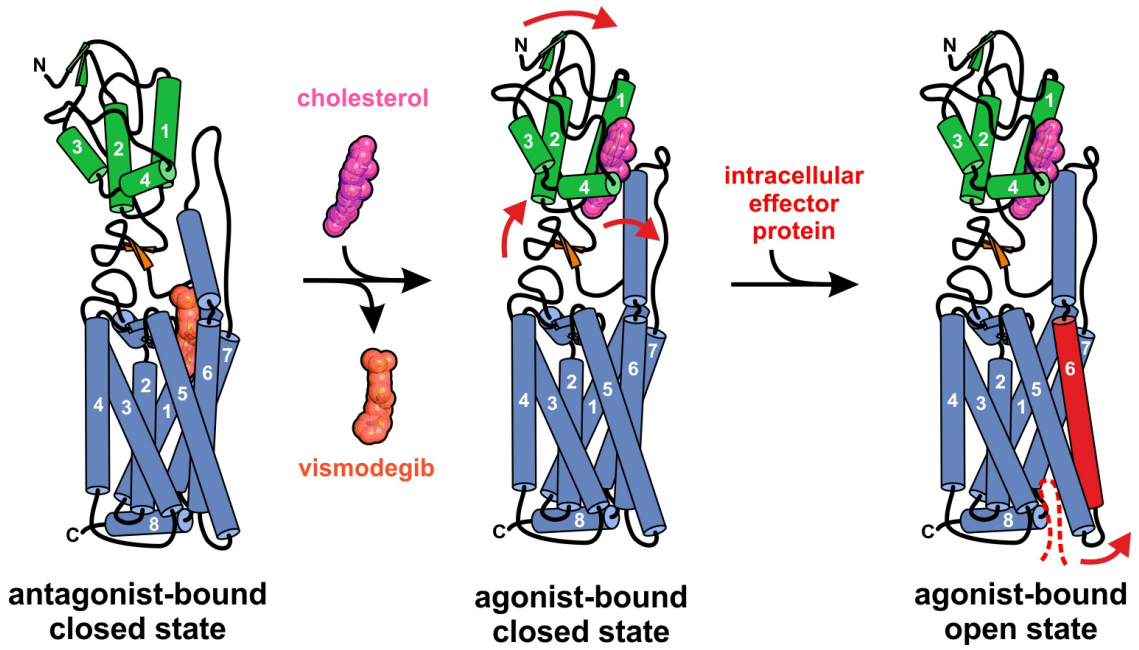


Figure 6.4: Model for SMO activation mechanism. SMO undergoes a conformational change upon binding of an agonist (here represented by cholesterol, pink) and removal of an antagonist (here represented by vismodegib, light red). A further conformational change involving the TM bundle (and most likely TM helix 6, highlighted in red) would be induced by binding of an intracellular effector protein to the intracellular face of SMO. The dashed red line represents a loop from an interacting intracellular effector protein (such as a G protein). Red arrows indicate conformational changes.

access to the SMO TMD until Shh binding (Fig. 6.5, panel b). Another possibility, a variant of the first, is that Ptc1 may influence lipid localisation within the plasma membrane, perhaps functioning as a cholesterol flippase in order to make cholesterol inaccessible to SMO until Shh binding (Fig. 6.5, panel c). This may also help to explain the necessity for SMO to translocate to the primary cilium in order to fully activate downstream signalling[314, 79] - the primary cilium may offer a plasma membrane compartment of limited size in which local lipid concentrations may be readily altered by Ptc1. It may also be the case that a currently unknown lipidic inhibitor may operate through this mechanism (Fig. 6.5, panel d). More information on the mechanism of Ptc1's function as well as on the dynamic lipid composition of ciliary membranes will inform on the crucial interaction between Ptc1 and SMO.

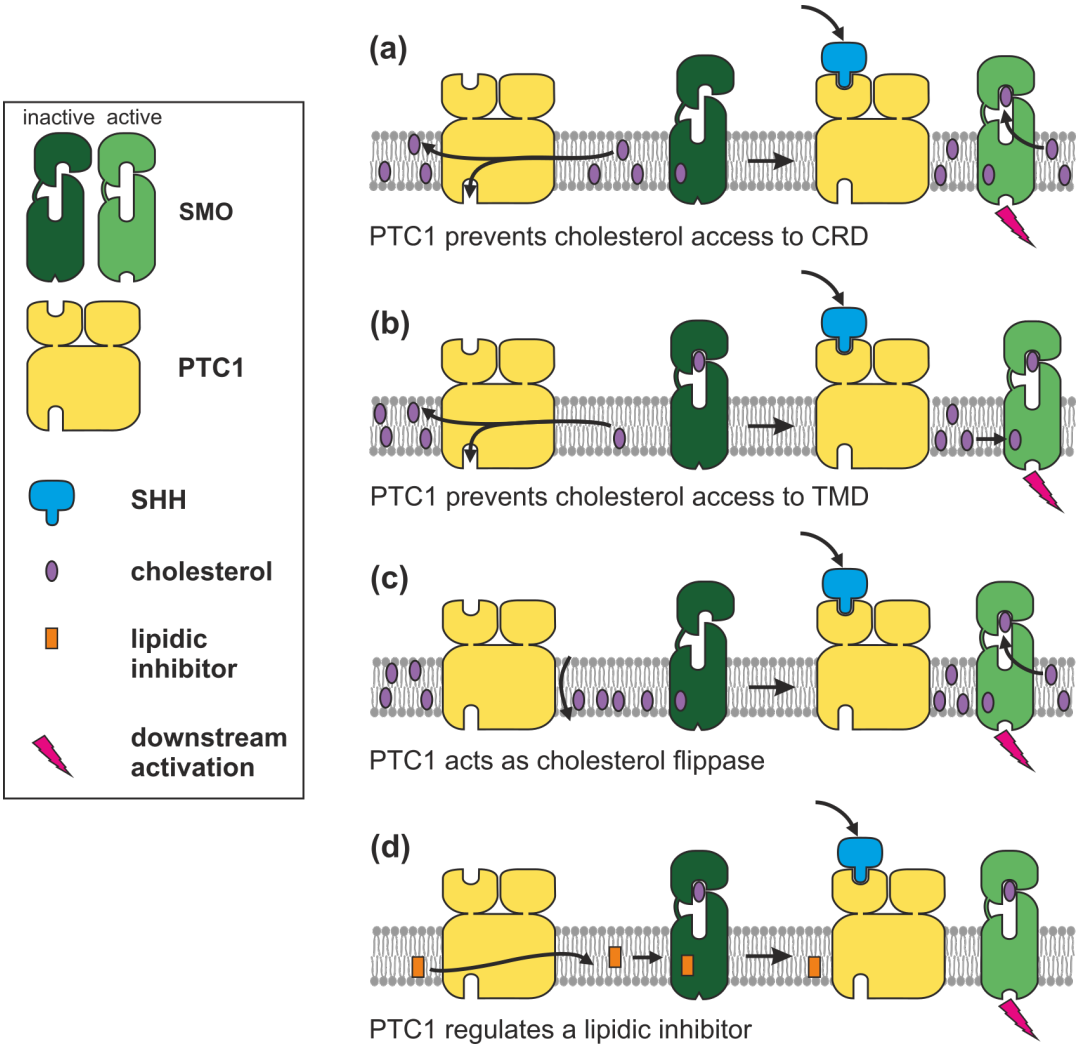


Figure 6.5: Models for Ptc1 function. **a,** In the first model, the SMO TMD is constitutively associated with cholesterol but Ptc1 prevents the SMO CRD from accessing cholesterol, thereby preventing activation. Upon SHH-binding, Ptc1 is inactivated, allowing the CRD to acquire cholesterol and to become activated. **b,** In the second model, the SMO CRD contains cholesterol as a necessary co-factor and Ptc1 prevents the SMO TMD from accessing cholesterol, thereby preventing activation. Upon SHH-binding, Ptc1 is inactivated, allowing the SMO TMD to acquire cholesterol and become activated. **c,** In the third model, a variant of the first, Ptc1 acts as a cholesterol flippase, shifting cholesterol from the outer to the inner leaflet of the lipid bilayer and thereby preventing the SMO CRD from accessing cholesterol. These models are not mutually exclusive and Ptc1 could regulate cholesterol access to both sites. **d,** Cholesterol acts as a co-factor for SMO activation, while Ptc1 regulates a different lipidic ligand (solid square) which could either function as a SMO antagonist as shown here or as a SMO agonist.

6.2 Therapeutic implications of SMO structures

At present, there are two small molecule Hh pathway inhibitors that are approved by the FDA for the treatment of advanced basal cell carcinoma, vismogedib and sonidegib[335, 44]. Both of these drugs act to block uncontrolled Hh signalling by interacting directly with SMO. Specifically, they both bind to the TMD binding site of SMO. There are many clinical trials ongoing to establish whether either of these two compounds could be used in the treatment of other types of cancer or in combination therapies with other anti-cancer drugs[396].

The structural studies carried out in this thesis have focused on understanding the molecular mechanisms of SMO activity. One of the SMO CRD-LD-TMD structures presented in Chapter 4 was bound to the anti-cancer drug, vismodegib (Fig. 4.7, page 137). Vismodegib-resistance mutations have already been reported in patients (Fig. 1.7, panel c, page 21)[320] and this structure now provides a structural explanation for the loss of efficacy of the drug associated with these mutations (Fig. 4.8, page 138).

SMO's CRD can also affect downstream signalling, with the CRD binding site operating either independently from or in synergy with the TMD binding site. All of the structures solved in this thesis have improved our understanding of the structural basis behind the connection between these two functional domains and their two functional binding sites. It may be worthwhile to pursue small molecule inhibitors that target the CRD directly or that affect the availability of endogenous ligands for SMO, for instance by reducing cholesterol availability. These strategies may help to improve therapies that already target the TMD binding site.

6.3 SMO and the related Frizzled-type GPCRs

SMO was designated a Class F or Frizzled-type GPCR because of its sequence similarity to the Frizzled (Fzd) receptors. The Fzd receptors are the principal receptors for the Wnt ligands, performing an essential role in at least three signalling pathways: tissue polarity/PCP signalling, canonical Wnt signalling and Wnt-calcium signalling. There are ten Fzd paralogues in mammals, numbered 1-10. In contrast to SMO, Fzd proteins are direct receptors for the extracellular Wnt ligands. The single-pass transmembrane proteins, LRP5 and LRP6, operate as essential co-receptors for Fzd in canonical Wnt signalling. There is some redundancy in Fzd receptor function but subsets of the receptors seem to be responsible for distinct signalling mechanisms[382]. While there is cross-reactivity between the different Wnt ligands for various Fzd receptors, there is no cross-reactivity of Wnt with SMO, despite the sequence similarity between SMO and the Fzd receptors (∼25-30% identity). Several CRD-only crystal structures of various Fzds have been solved[86], including one in complex with a lipidated Wnt ligand (Fig. 3.3, panel c, page 70)[165]. This enlightening complex structure of *Xenopus laevis* Wnt8 and the CRD of mouse Fzd8 revealed a "thumb and index finger" conformation of Wnt binding to the Fzd CRD. The "thumb" contains a palmitoleic acid lipid group that binds to the Fzd CRD at the equivalent position as cholesterol binds to the SMO CRD. This CRD binding site seems evolutionarily conserved across the Class F GPCRs as a lipid binding site. However, as of yet, there is no structure of a Fzd transmembrane domain.

A sequence alignment of SMO and the Fzds reveals high similarity within the CRD and 7TM helices but striking variation in the length and sequence of the LD, ECL3 and ICD (Fig. 6.6). Eight of the ten disulfide-forming cysteines in SMO's CRD are conserved in all the Fzds, as are the two cysteines in the LD (Fig. 6.7, panel a). The two non-conserved cysteines in SMO's CRD form the disulfide bond

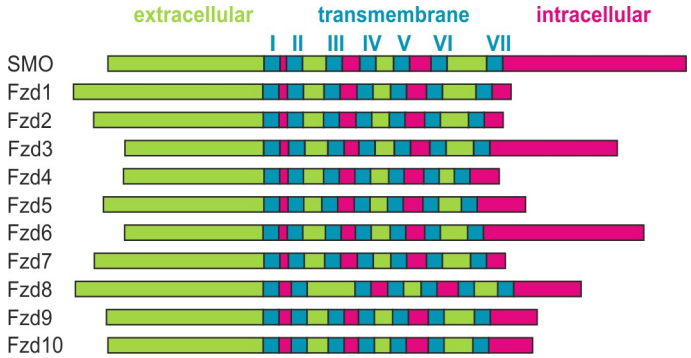


Figure 6.6: Relative lengths of domains for Frizzled family receptors. Linear cartoon schematic of Fzd family topology, aligned to TM helix I. TM helices are numbered with Roman numerals. Colours indicate topology of each region.

that is labelled 1′ in Fig. 3.15 (page 95). Assuming that Fzds retain a similar domain arrangement as for SMO, the sequence variation in the LD, ECL3 and ICD is significant as these are regions in SMO that mediate signal transduction and therefore likely contain the determinants of ligand specificity and signal variation (Fig. 6.7). The ECL3 of Fzd1 is the same length as that of SMO but all other Fzds' ECL3s are shorter (Fig. 6.7, panel b). The ECL3 of Fzd4 is especially short and may mean its signal transduction is regulated in a different manner to that of SMO. In contrast, the LD of SMO is much shorter than many of the Fzd LDs, again suggesting variation in the coordination of CRD positioning relative to the TMD (Fig. 6.7, panel a). Notably, Asp95, one of the key sterol-binding residues in SMO's CRD, is not conserved in Fzds. In fact, the loop containing Asp95 in SMO is two residues shorter in all the Fzds, perhaps to allow the entry of the Wnt-conjugated lipid moiety (Fig. 6.7, panel a). The cholesterol-interacting residues from the SMO ECL3 are not conserved in Fzds, reflecting the overall variation in ECL3 sequences (Fig. 6.7, panel b).

Multi-domain structures of the Fzd receptors will be essential to gaining a full understanding of their signal transduction mechanism. The structural work carried out in this thesis demonstrates a strategy that may also prove useful as a template in pursuing structures for the Fzd receptors.

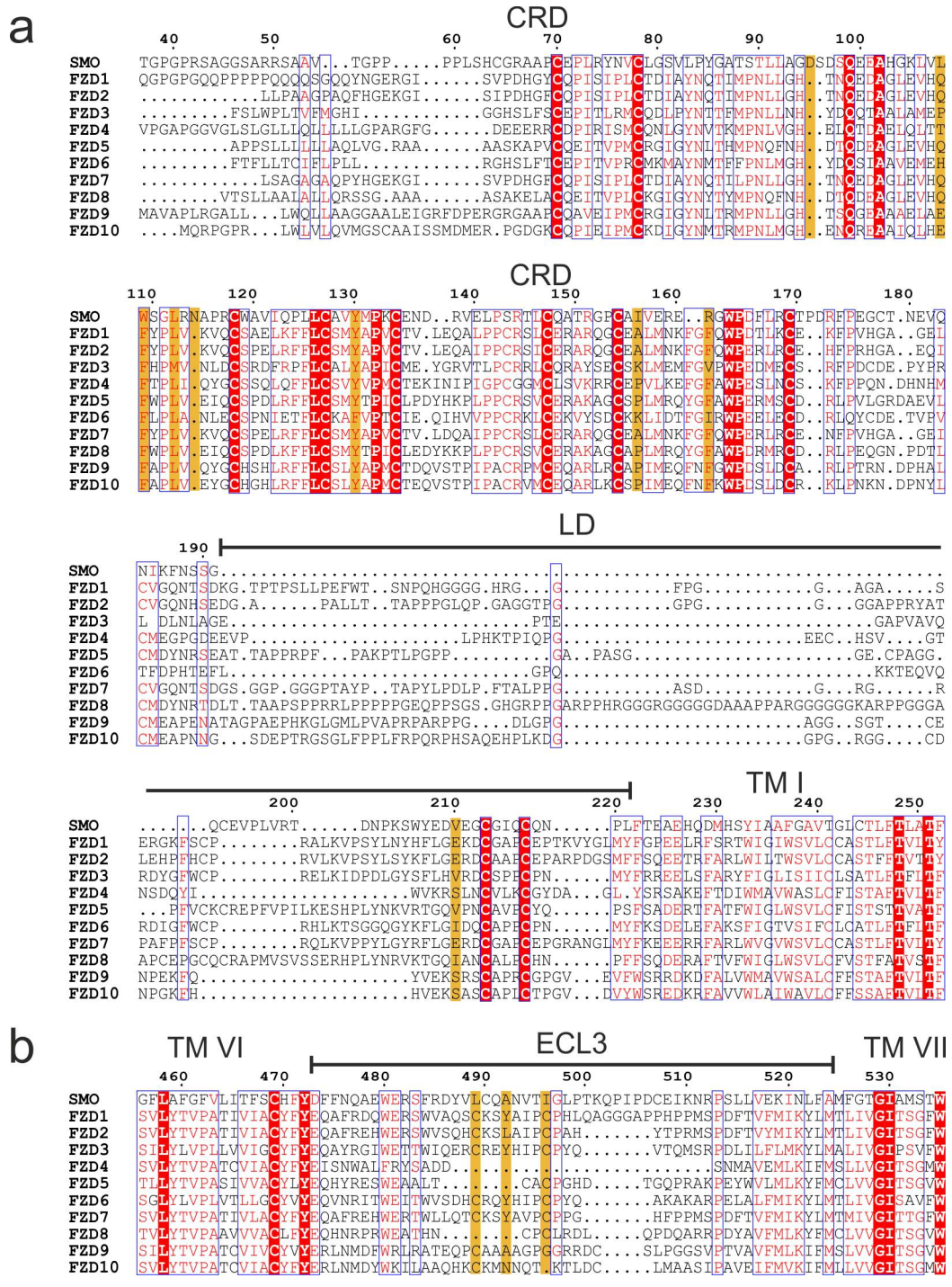


Figure 6.7: Sequence alignment of Frizzled family receptors. Sequence alignment of Fzd family receptors showing CRD (**a**) and LD regions (**b**). Residues of the SMO cholesterol binding site highlighted in gold. Alignment performed using Clustal Omega[334]; visualised using ENDscript[309].

6.4 Hh signalling downstream of SMO

6.4.1 The intracellular domain of SMO

While the crystal structures solved in this thesis contain multiple domains of SMO that are each essential for proper native function, the constructs used are not full-length as they are missing the large and essential ICD. Compared to *Drosophila*, the vertebrate SMO ICD has received relatively little attention, despite the fact that it is required for native Shh signalling[369, 306]. Corbit *et al.* reported that a hydrophobic residue followed by a basic residue, Trp549 and Arg550, in TM helix VII of mouse SMO constituted a conserved ciliary localisation motif[79]. An altered form of mSMO in which this motif had been mutated to alanines was unable to localise to primary cilia and it was also unresponsive to Shh-stimulated signalling. It is not clear, however, whether this motif is truly a ciliary localisation motif or whether it instead causes inactivation of SMO in the same way as most antagonists, such as vismodegib or SANT-1, which also involves prevention of ciliary localisation[399, 313].

Mouse SMO lacking the ICD is unable to reach the primary cilia, even in the presence of Shh or SAG [188, 262]. As part of his doctoral studies, Daniel Nedelcu examined the ICD of vertebrate SMO in some detail[262]. Through truncation analysis, he identified two ciliary localisation sequences in the ICD of mouse SMO, one between residues 615-637 and another between residues 685-793, each of which was sufficient to direct SMO to the primary cilium (Fig. 6.8, panel a)[262]. However, removal of both of these sequences completely prevented SMO ciliary localisation. There are also two basic patches within the ICD that are important for downstream signalling but are not directly involved in ciliary localisation. Mutation of every residue within either of these patches to alanine caused a large decrease in the signalling ability of SMO, although mutating the first basic patch (residues 565-581) showed a much more severe

effect than mutating the second patch (residues 668-680), with downstream signalling in response to SAG almost entirely lost. However, each of these basic patch mutants retained the ability to localise to the primary cilium, consistent with a modular architecture for the ICD. Nedelcu further demonstrated through several chimeras that the ICD of SMO is sufficient to localise other 7TM proteins to the primary cilium but not to activate downstream Hh signalling.

Based on these observations, I designed a number of ICD truncations of hSMO and conducted small scale expression tests. Initial small scale expression tests using a number of SMO CRD-LD-TMD-ICD ('full-length') constructs with progressive truncations of the ICD demonstrated that the addition of just twenty-seven extra residues beyond the conserved helix VIII resulted in extremely low protein yields (Fig. 6.8, panels b & c). The addition of the Val329Phe TMD stabilising mutation did not improve the yield for any of these constructs, indicating that instability of the TMD was not the cause of the huge drop in expression levels between the CRD-LD-TMD and the CRD-LD-TMD-ICD constructs (Fig. 6.8, panel c). Previous efforts within the Siebold Lab to express the SMO ICD alone were unsuccessful (data not shown). The ICD may be natively unstructured or unstructured in the absence of interacting partners, either of which may influence expression levels. The ICD may also include motifs that lead SMO to be rapidly degraded or continually recycled.

There have been several reports of post-translational modification of the ICD of SMO. Chen *et al.* demonstrated that upon Shh stimulation vertebrate SMO is phosphorylated at multiple sites in the ICD and that phosphorylation induces SMO's ciliary localisation[57]. This phosphorylation is carried out by at least two kinases, CK1α and GRK2, and graded increases in Shh stimulation corresponded to increased SMO phosphorylation levels. The authors were able to map five potential phosphorylation sites on the ICD that are each targeted by one or both of the

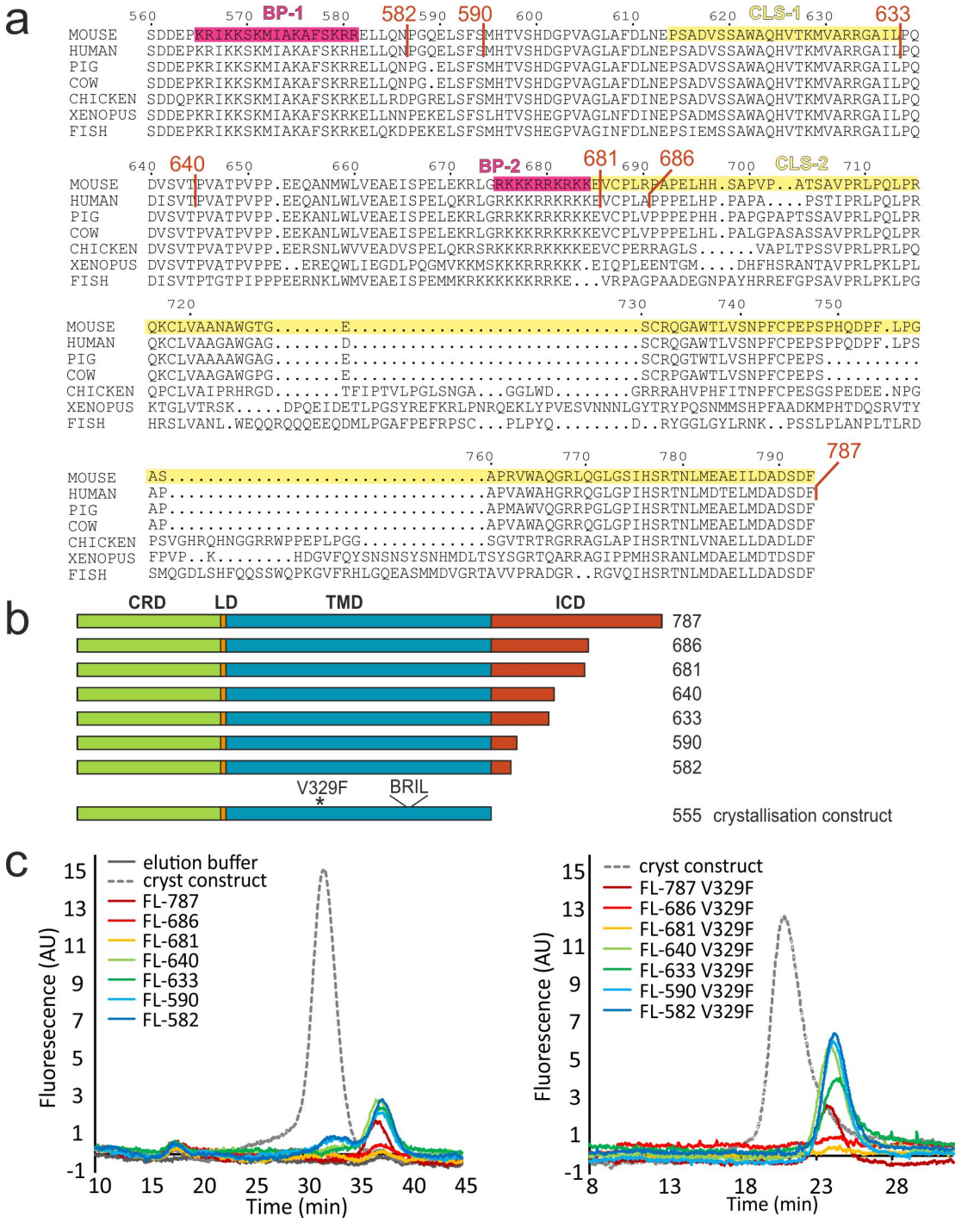


Figure 6.8: SMO intracellular domain. a, Sequence alignment of SMO ICD from seven vertebrates. Sequences follow on from Fig. 3.6. Black numbering corresponds to mouse SMO. Brown numbering corresponds to human SMO. BP, basic patch; CLS, ciliary localisation sequence. Alignment performed using Clustal Omega[334]. **b,** Cartoon schematic of new full-length constructs of human SMO. Truncation points are indicated in a & b. **c,** FSEC traces from small scale expression tests. Full-length constructs without (left) and with Val329Phe mutation (right) compared to the crystallisation construct.

identified kinases. Ma *et al.* showed that vertebrate SMO conjugated to a small ubiquitin-related modifier (SUMO) molecule has higher basal levels of activity and increased ciliary localisation[228]. In addition, they showed that overexpression of a desumoylating enzyme prevented downstream activation of the Hh pathway after stimulation with Shh and also prevented Shh-stimulated ciliary localisation of SMO. The authors were not able to identify a sumoylation consensus site in vertebrate SMO, however. These studies demonstrate a complex modification framework for vertebrate SMO's ICD that includes at least phosphorylation and sumoylation. Control of these processes will likely be necessary for successful high-level production of full-length SMO.

6.4.2 SMO interaction partners

Most GPCRs interact directly with heterotrimeric G proteins and β-arrestins. SMO binds to and activates all members of the G_i family but does not interact with the G_s, G_q or G_{12} families[307, 92]. A SMO mutant lacking the ICD retained its ability to activate $G\alpha_i$ but not its ability to activate downstream Gli transcription, perhaps because this mutant is unable to translocate to the primary cilium[188, 262]. Treatment of NIH-3T3 fibroblasts or C3H10T1/2 cells with pertussis toxin, which modifies $G\alpha_i$ proteins so that they cannot be activated, reduced downstream Gli transcription in response to stimulation by Shh or by the SMO-targeting agonist purmorphamine, suggesting that G_i protein activation is necessary for downstream Hh signalling[307, 329]. A similar dependence of Hh signalling on G_i proteins has been shown in neonatal rat ventricular cardiomyocytes[39]. Furthermore, G_i-dependent but Gli-independent Hh activation has been implicated for cell growth, cell death and chemotaxis in endothelial cells[66], supporting a non-canonical pathway for Hh signalling[23, 287, 286, 373].

SMO also interacts with β-arrestin 2 (ref.[56, 391]). Recruitment of β-arrestin 2 to the cell surface was increased by Shh stimulation in the presence of SMO and also by the presence of both SMO and GRK2. β-arrestins do not affect ciliogenesis but localise to the primary cilium in mouse NIH-3T3 cells[197]. Shh treatment increased the amount of SMO and Kif3A that co-immunoprecipitated with β-arrestin 1 and 2, while cyclopamine reversed this effect. Taken together, these reports suggest that in response to pathway activation a GRK2-dependent SMO-β-arrestin-Kif3A complex is formed, which is required for localisation to the primary cilium. Thus far, only a single GPCR-arrestin complex structure has been solved[171].

Apart from G proteins and β-arrestins, various other intracellular proteins have been implicated in regulation of SMO signalling. Vertebrate SMO interacts with at least two kinases, CK1α and GRK2, as mentioned previously[57, 56, 240]. Activated SMO forms a complex with EVC2 within the primary cilium[103]. Activation of SMO promotes translocation of the anterograde motor protein, Kif7, into the primary cilium and then to its tip[219, 63]. Evidence of a direct interaction between SMO and Kif7 is currently lacking, however. Two proteins, Pitchfork and Gprasp2, have recently been implicated as essential components in a Hh-induced ciliary targeting complex that regulates SMO translocation[169]. Piecing together all of SMO's interactions will require a great deal of future structural characterisation.

6.5 Conclusions

In this thesis, I have detailed my work in pursuing a greater understanding of the molecular workings of SMO. We solved the first crystal structure of SMO CRD-LD-TMD, which was the first structure of a GPCR with a functional extracellular domain. This initial structure revealed a previously unknown small molecule ligand that was also the first reported endogenous ligand for SMO, cholesterol. This initial structure was followed up with five additional crystal structures of SMO bound to different small molecules. One of these was in complex with an FDA-approved drug, vismodegib; another was in complex with not one but two agonists, SAG and cholesterol; yet another was in complex with SMO's classical antagonist, cyclopamine, bound at both the CRD and TMD binding sites. These structures have provided critical insight into the interaction between SMO's extracellular domains and the TMD. In seeking to further stabilise the extracellular domains of SMO, we produced and sought to solve the structure of a nanobody bound to SMO. While characterisation and crystallisation of the nanobodies with SMO has shown promising results, structural determination has thus far been frustrated by poorly diffracting crystals. Nevertheless, such nanobodies may prove useful in future for determining the structure of SMO in an active state.

Appendix A

Membrane Protein Additive Screen

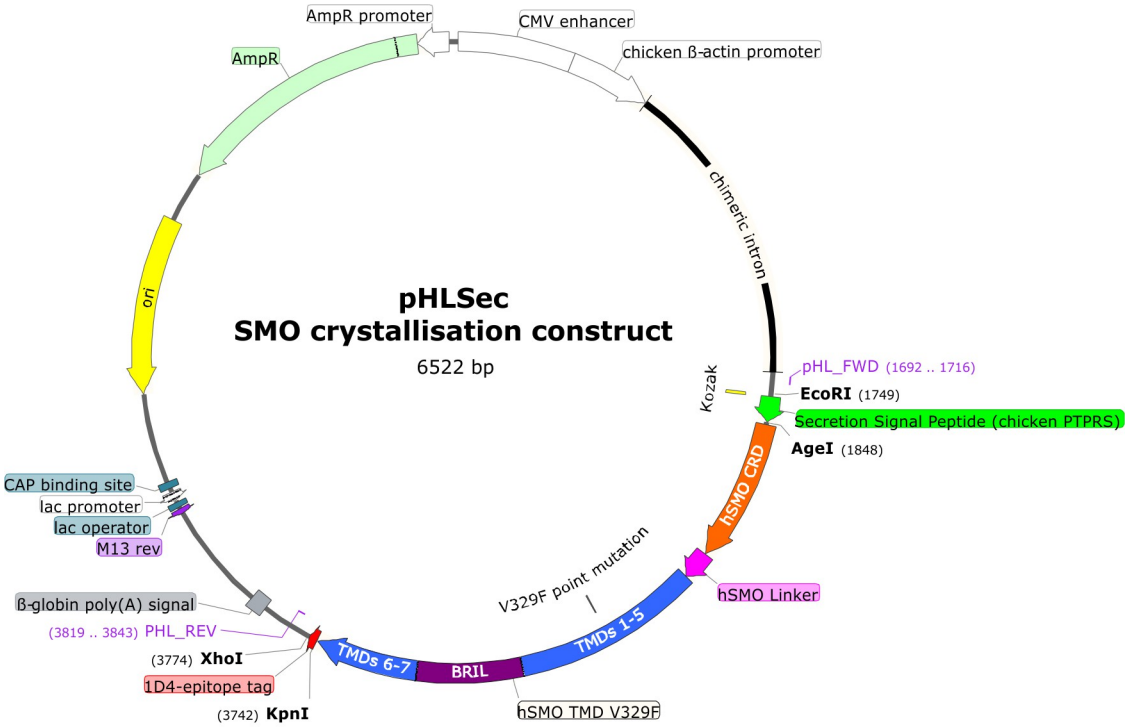
Location	Condition
A1	MilliQ H_2O (blank)
A2	0.50% (w/v) LMNG
A3	0.33% (w/v) digitonin
A4	0.50% (w/v) CYMAL-6 NG
A5	0.50% (w/v) CYMAL-7 NG
A6	100 mM $CaCl_2$
A7	100 mM $MgCl_2$
A8	100 mM $MnCl_2$
A9	100 mM $ZnCl_2$
A10	(100 mM) $FeCl_3$
A11	500 mM KCl
A12	500 mM LiCl
B1	MilliQ H_2O (blank)
B2	100 mM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

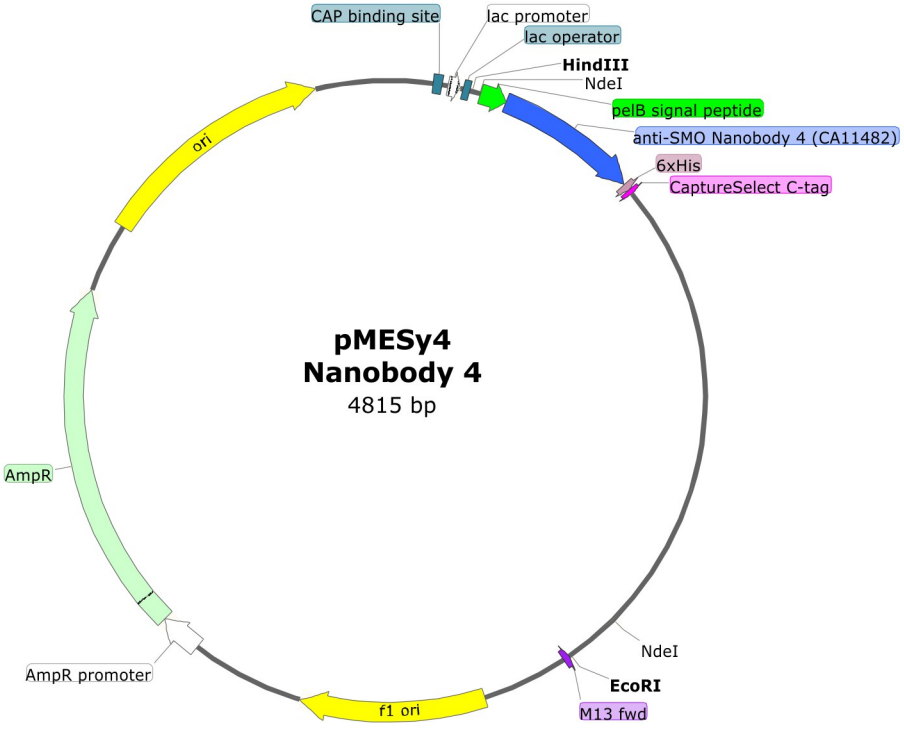
Location	Condition
E1	MilliQ H_2O (blank)
E2	200 mM sodium thiocyanate
E3	50% (v/v) glycerol
E4	50% (v/v) ethylene glycol
E5	10% (v/v) tert-butanol
E6	100 mM DMSO
E7	1M guanidinium hydrochloride
E8	50% (v/v) PEG 4,000
E9	50% (v/v) PEG 400
E10	20% (w/v) D(+)-glucose
E11	20% (w/v) sucrose
E12	100 mM sucrose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

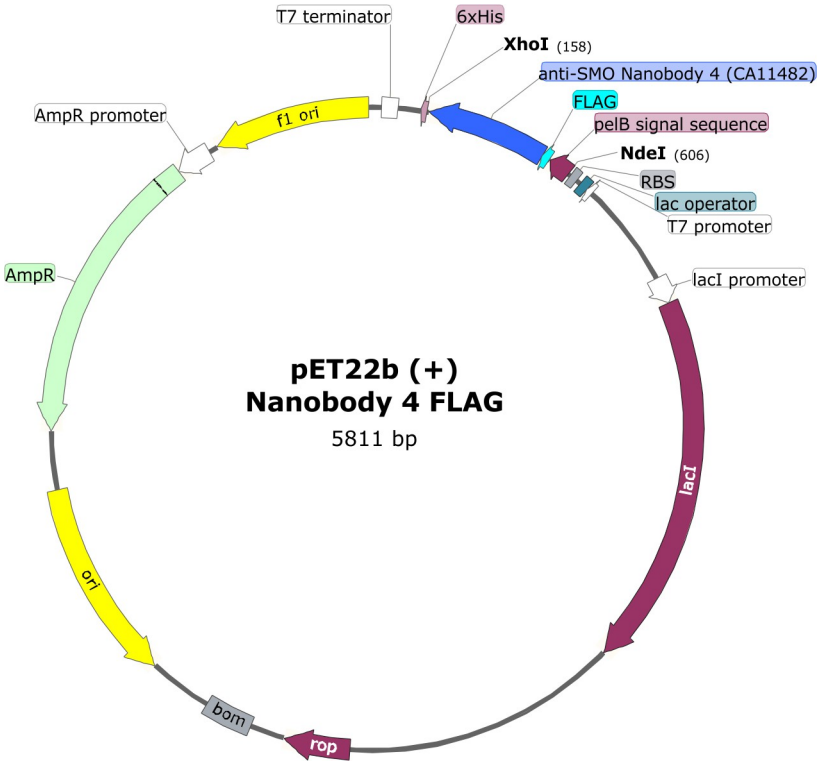
Table A.1: Membrane Protein Additive Screen conditions and locations within 96-well block. This screen was created in order to test the thermostabilising ability of a large number of diverse additives on membrane proteins.

Appendix B

Plasmids & Constructs


pHLSec
SMO crystallisation construct
6522 bp
AmpR promoter
CMV enhancer
chicken ß-actin promoter
AmpR
chimeric intron
ori
Kozak
pHL_FWD (1692 .. 1716)
EcoRI (1749)
Secretion Signal Peptide (chicken PTPRS)
AgeI (1848)
hSMO CRD
hSMO Linker
V329F point mutation
TMDs 1-5
BRIL
TMDs 6-7
hSMO TMD V329F
CAP binding site
lac promoter
lac operator
M13 rev
ß-globin poly(A) signal
(3819 .. 3843) PHL_REV
(3774) XhoI
1D4-epitope tag
(3742) KpnI


pMESy4
Nanobody 4
4815 bp
CAP binding site
lac promoter
lac operator
HindIII
NdeI
pelB signal peptide
anti-SMO Nanobody 4 (CA11482)
6xHis
CaptureSelect C-tag
ori
AmpR
AmpR promoter
f1 ori
NdeI
EcoRI
M13 fwd


T7 terminator
6xHis
XhoI (158)
anti-SMO Nanobody 4 (CA11482)
FLAG
pelB signal sequence
NdeI (606)
RBS
lac operator
T7 promoter
lacI promoter
AmpR promoter
f1 ori
AmpR
pET22b (+)
Nanobody 4 FLAG
5811 bp
lacI
ori
bom
rop

Name	Meaning
"intraBRIL"	ECD+TMD (Q99835 αα32-555) with ICL3 (αα429-440) replaced by BRIL (P0ABE7 αα23-128)
"FzD+linker+7TM"	ECD+TMD (Q99835 αα32-555)
"exBRIL"	N-terminal BRIL (P0ABE7 αα23-128) on TMD (Q99835 αα190-555)
"FL"	full-length (Q99835 αα32-787 or C-terminal truncation at the indicated residue)
"VENUS" or "YFP"	monoVenus fluorescent tag (C-terminal)
"1D4"	Rho1D4 epitope tag (C-terminal)
"HA"	HA tag (N-terminal)

Master Clone #	Name
246	hSmo-fl-bril-GENEART (synthetic gene)
249	pHL_hSmo-intraBRIL-HA-1D4
250	pHL_hSmo-intraBRIL-HA-VENUS-1D4
256	pHL_hSmo-intraBRIL-1D4
257	pHL_hSmo-intraBRIL-VENUS-1D4
267	pHL_hSmo-intraBRIL-HA-CHERRY-1D4
278	pHL_hSmo-intraBRIL-V329F-HA-1D4
279	pHL_hSmo-intraBRIL-V329F-HA-VENUS-1D4
477	pHL_hSmo-intraBRIL-V329F-VENUS-1D4
309	pHL_hSmo-intraBRIL-V329F-1D4 (crystallisation construct)
536	pHL_hSmo-intraBRIL-V329F,G111F-1D4
537	pHL_hSmo-intraBRIL-V329F,G111F-VENUS-1D4
637	pHL_hSmo-intraBRIL-V329F,D95A-1D4
638	pHL_hSmo-intraBRIL-V329F,D95A-VENUS-1D4
639	pHL_hSmo-intraBRIL-V329F,D95A,Y130F-1D4
640	pHL_hSmo-intraBRIL-V329F,D95A,Y130F-VENUS-1D4
500	pHL_hSmo-intraBRIL-ΔCRD-V329F-1D4
280	pHL_hSmo-intraBRIL-I408F-HA-1D4
281	pHL_hSmo-intraBRIL-I408F-HA-VENUS-1D4
282	pHL_hSmo-intraBRIL-V329F,I408F-HA-1D4
283	pHL_hSmo-intraBRIL-V329F,I408F-HA-VENUS-1D4
274	pHL_hSmo-intraBRIL-W535L-HA-1D4 ("SmoM2")
271	pHL_hSmo-intraBRIL-L412F-HA-1D4
272	pHL_hSmo-intraBRIL-L412F-HA-VENUS-1D4
273	pHL_hSmo-intraBRIL-A459V-HA-1D4
298	pHL_hSmo-intraBRIL-A459V-HA-VENUS-1D4
505	pHL_hSmo-intraBRIL-D473R-1D4
506	pHL_hSmo-intraBRIL-D473R-VENUS-1D4
258	pHL_hSmo-FzD+linker+7TM-HA-1D4
259	pHL_hSmo-FzD+linker+7TM-HA-VENUS-1D4
260	pHL_hSmo-FzD+linker+7TM-1D4
261	pHL_hSmo-FzD+linker+7TM-VENUS-1D4
726	pHL_hSmo-FzD+linker+7TM-Avi-1D4
268	pHL_hSmo-FzD+linker+7TM-HA-CHERRY-1D4
276	pHL_hSmo-FzD+linker+7TM-W535L-HA-1D4 ("SmoM2")
277	pHL_hSmo-FzD+linker+7TM-W535L-HA-VENUS-1D4 ("SmoM2")
310	pHL_hSmo-FzD+linker+7TM-V329F-HA-1D4
311	pHL_hSmo-FzD+linker+7TM-V329F-HA-VENUS-1D4
473	pHL_hSmo-FzD+linker+7TM-V329F-HA-Avi-1D4
479	pHL_hSmo-FzD+linker+7TM-V329F-1D4
481	pHL_hSmo-FzD+linker+7TM-I408F-HA-1D4
482	pHL_hSmo-FzD+linker+7TM-I408F-HA-VENUS-1D4
483	pHL_hSmo-FzD+linker+7TM-L412F-HA-1D4

484	pHL_hSmo-FzD+linker+7TM-L412F-HA-VENUS-1D4
485	pHL_hSmo-FzD+linker+7TM-A459V-HA-1D4
507	pHL_hSmo-FzD+linker+7TM-D473R-HA-1D4
508	pHL_hSmo-FzD+linker+7TM-D473R-HA-VENUS-1D4
727	pHL_hSmo-FzD+linker+7TM-D473R-Avi-1D4

245	hSmo-Stephens-GENEART (synthetic gene)
247	pHL_hSmo-exBRIL-HA-1D4
248	pHL_hSmo-exBRIL-HA-VENUS-1D4
254	pHL_hSmo-exBRIL-1D4
255	pHL_hSmo-exBRIL-VENUS-1D4
266	pHL_hSmo-exBRIL-HA-CHERRY-1D4
299	pHL_hSmo-exBRIL-V329F-HA-1D4
300	pHL_hSmo-exBRIL-V329F-HA-VENUS-1D4
348	pHL_hSmo-exBRIL-V329F-1D4
301	pHL_hSmo-exBRIL-I408F-HA-1D4
302	pHL_hSmo-exBRIL-I408F-HA-VENUS-1D4
303	pHL_hSmo-exBRIL-L412F-HA-1D4
304	pHL_hSmo-exBRIL-L412F-HA-VENUS-1D4
305	pHL_hSmo-exBRIL-A459V-HA-1D4
306	pHL_hSmo-exBRIL-A459V-HA-VENUS-1D4
475	pHL_hSmo-exBRIL-W535L-HA-1D4
476	pHL_hSmo-exBRIL-W535L-HA-VENUS-1D4
503	pHL_hSmo-exBRIL-D473R-1D4
504	pHL_hSmo-exBRIL-D473R-VENUS-1D4

516	hSmo-ICD-GENEART (aa556-787)
517	pHL_hSmo-FL-1D4 (aa32-787)
518	pHL_hSmo-FL-YFP-1D4 (aa32-787)
521	pHL_hSmo-FL-V329F-1D4
522	pHL_hSmo-FL-V329F-YFP-1D4
523	pHL_hSmo-FL-686-1D4
524	pHL_hSmo-FL-686-YFP-1D4
687	pHL_hSmo-FL-686-V329F-1D4
688	pHL_hSmo-FL-686-V329F-YFP-1D4
525	pHL_hSmo-FL-681-1D4
526	pHL_hSmo-FL-681-YFP-1D4
689	pHL_hSmo-FL-681-V329F-1D4
690	pHL_hSmo-FL-681-V329F-YFP-1D4
527	pHL_hSmo-FL-640-1D4
528	pHL_hSmo-FL-640-YFP-1D4
691	pHL_hSmo-FL-640-V329F-1D4
692	pHL_hSmo-FL-640-V329F-YFP-1D4
529	pHL_hSmo-FL-633-1D4
531	pHL_hSmo-FL-633-YFP-1D4
693	pHL_hSmo-FL-633-V329F-1D4
694	pHL_hSmo-FL-633-V329F-YFP-1D4
532	pHL_hSmo-FL-590-1D4
533	pHL_hSmo-FL-590-YFP-1D4
695	pHL_hSmo-FL-590-V329F-1D4
696	pHL_hSmo-FL-590-V329F-YFP-1D4
534	pHL_hSmo-FL-582-1D4
535	pHL_hSmo-FL-582-YFP-1D4
697	pHL_hSmo-FL-582-V329F-1D4
698	pHL_hSmo-FL-582-V329F-YFP-1D4

699	pMESy4_hSmoNb1-His (CA11474)
700	pMESy4_hSmoNb2-His (CA11476)
701	pMESy4_hSmoNb3-His (CA11477)
702	pMESy4_hSmoNb4-His (CA11482)
703	pMESy4_hSmoNb5-His (CA11494)
704	pMESy4_hSmoNb6-His (CA11484)
705	pMESy4_hSmoNb7-His (CA11517)
706	pMESy4_hSmoNb8-His (CA11620)
707	pMESy4_hSmoNb9-His (CA11518)
708	pET22b_hSmoNb1-FLAG-His (CA11474)
709	pET22b_hSmoNb2-FLAG-His (CA11476)
710	pET22b_hSmoNb3-FLAG-His (CA11477)
711	pET22b_hSmoNb4-FLAG-His (CA11482)
712	pET22b_hSmoNb5-FLAG-His (CA11494)
713	pET22b_hSmoNb6-FLAG-His (CA11484)
714	pET22b_hSmoNb7-FLAG-His (CA11517)
715	pET22b_hSmoNb8-FLAG-His (CA11620)
716	pET22b_hSmoNb9-FLAG-His (CA11518)

Appendix C

Details of Small Molecule Ligands

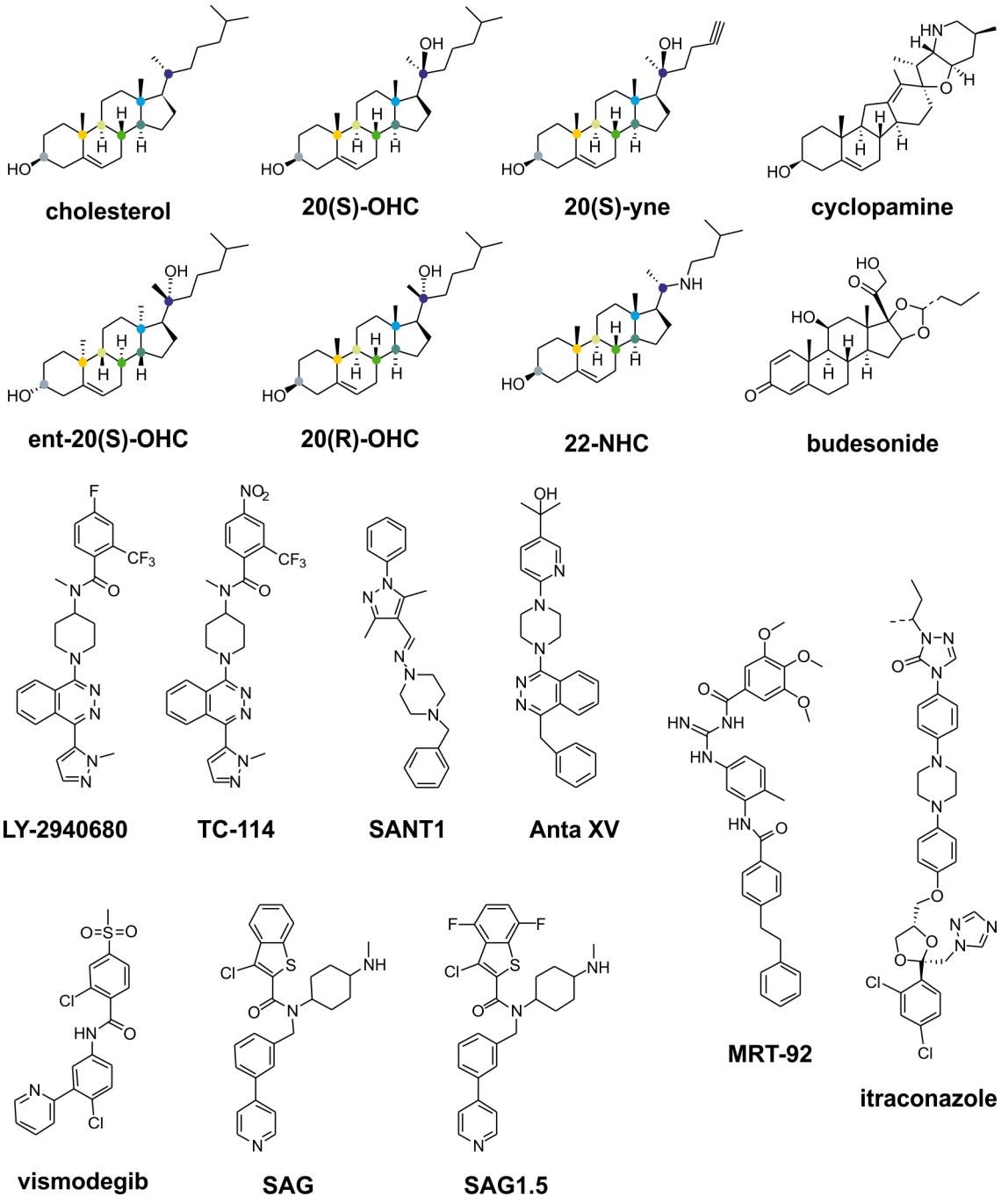


Figure C.1: Chemical structures of small molecules used in this thesis.

Name	**MW (g/mol)**	**Binding site**	**Action on SMO**	**IC50/EC50**	**SMO cilia accumulation**	**Gli-mediated transcription**	**References**
cholesterol	386.7	CRD	agonist	∼100 μM	induces	activates	[226, 149]
20(S)-OHC	402.7	CRD	agonist	3 μM	induces	activates	[257, 258]
20(S)-yne	370.6	CRD	agonist	390 nM	induces	activates	[257, 258]
cyclopamine hydrate	411.6	TMD, CRD	antagonist	300 nM	induces	inhibits	[319, 178, 65, 77, 352]
ent-20(S)-OHC	402.7	(CRD)	-	-	-	-	[257]
20(R)-OHC	402.7	(CRD)	-	-	-	-	[257]
22-NHC	388.4	CRD	antagonist	3 μM	induces	inhibits	[263]
budesonide	430.5	CRD	antagonist	1-10 μM	inhibits	inhibits	[383, 296]
LY2940680	512.5	TMD	antagonist	$<$ 3 μM	(inhibits)	inhibits	[379, 18]
TC-114	540.2	TMD	antagonist	1.6 nM	(inhibits)	inhibits	[411]
SANT-1	373.5	TMD	antagonist	20 nM	inhibits	inhibits	[319]
Anta XV	439.6	TMD	antagonist	3-30 nM	(inhibits)	inhibits	[378, 245]
vismodegib	412.3	TMD	antagonist	150-300 nM	inhibits	inhibits	[134, 136, 375, 399]
SAG	599.0	TMD	agonist	3 nM	induces	activates	[53]
SAG1.5	635.0	TMD	agonist	1 nM	induces	activates	[120, 378]
MRT-92	603.1	TMD	antagonist	0.4 nM	inhibits	inhibits	[146]
itraconazole	705.6	?	antagonist	800 nM	inhibits	inhibits	[190, 186]

Appendix D

Details of Detergents

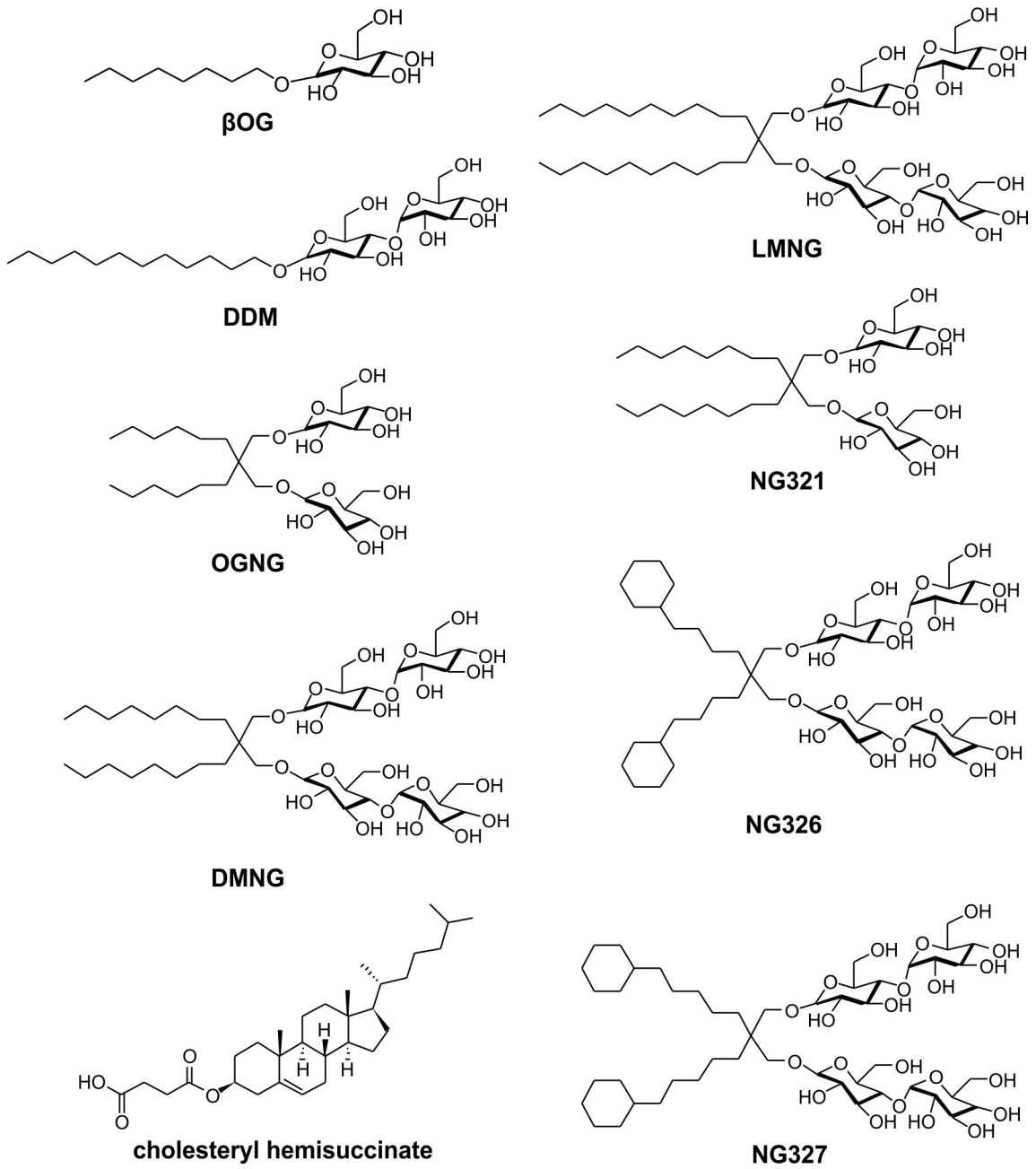


Figure D.1: Chemical structures of detergents used in this thesis. βOG, n-octyl-β-D-glucoside; DDM, n-dodecyl-β-D-maltopyranoside; OGNG, octyl glucose neopentyl glycol; DMNG, decyl maltose neopentyl glycol; CHS, cholesteryl hemisuccinate; LMNG, laurel maltose neopentyl glycol; NG321, decyl glucose neopentyl glycol; NG326, CYMAL-6 neopentyl glycol; NG327, CYMAL-7 neopentyl glycol.

Bibliography

[1] P. Aanstad, N. Santos, K. C. Corbit, P. J. Scherz, A. Trinh le, W. Salvenmoser, J. Huisken, J. F. Reiter, and D. Y. Stainier. The extracellular domain of smoothened regulates ciliary localization and is required for high-level hh signaling. *Curr Biol*, 19(12):1034–9, 2009.

[2] R. N. Abskharon, G. Giachin, A. Wohlkonig, S. H. Soror, E. Pardon, G. Legname, and J. Steyaert. Probing the n-terminal beta-sheet conversion in the crystal structure of the human prion protein bound to a nanobody. *J Am Chem Soc*, 136(3):937–44, 2014.

[3] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. Phenix: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2):213–21, 2010.

[4] J. Alcedo, M. Ayzenzon, T. Von Ohlen, M. Noll, and J. E. Hooper. The drosophila smoothened gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell*, 86(2):221–32, 1996.

[5] R. Alfonsi, B. Botta, S. Cacchi, L. Di Marcotullio, G. Fabrizi, R. Faedda, A. Goggiamani, A. Iazzetti, and M. Mori. Design, palladium-catalyzed synthesis, and biological investigation of 2-substituted 3-aroylquinolin-4(1h)-ones as inhibitors of the hedgehog signaling pathway. *J Med Chem*, 60(4):1469–1477, 2017.

[6] A. M. Arensdorf, M. E. Dillard, J. M. Menke, M. W. Frank, C. O. Rock, and S. K. Ogden. Sonic hedgehog activates phospholipase a2 to enhance smoothened ciliary translocation. *Cell Rep*, 19(10):2074–2087, 2017.

[7] A. R. Aricescu, W. Lu, and E. Y. Jones. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 10):1243–50, 2006.

[8] A. R. Aricescu and R. J. Owens. Expression of recombinant glycoproteins in mammalian cells: towards an integrative approach to structural biology. *Curr Opin Struct Biol*, 23(3):345–56, 2013.

[9] S. X. Atwood, K. Y. Sarin, R. J. Whitson, J. R. Li, G. Kim, M. Rezaee, M. S. Ally, J. Kim, C. Yao, A. L. Chang, A. E. Oro, and J. Y. Tang. Smoothened variants explain the majority of drug resistance in basal cell carcinoma. *Cancer Cell*, 27(3):342–53, 2015.

[10] V. A. Avlani, K. J. Gregory, C. J. Morton, M. W. Parker, P. M. Sexton, and A. Christopoulos. Critical role for the second extracellular loop in the binding of both orthosteric and allosteric g protein-coupled receptor ligands. *J Biol Chem*, 282(35):25677–86, 2007.

[11] K. L. Ayers and P. P. Therond. Evaluating smoothened as a g-protein-coupled receptor for hedgehog signalling. *Trends Cell Biol*, 20(5):287–98, 2010.

[12] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98(18):10037–41, 2001.

[13] B. Barakat, L. Yu, C. Lo, D. Vu, E. De Luca, J. E. Cain, L. G. Martelotto, S. Dedhar, A. J. Sadler, D. Wang, D. N. Watkins, and G. E. Hannigan. Interaction of smoothened with integrin-linked kinase in primary cilia mediates hedgehog signalling. *EMBO Rep*, 14(9):837–44, 2013.

[14] E. Baranova, R. Fronzes, A. Garcia-Pino, N. Van Gerven, D. Papapostolou, G. Pehau-Arnaudet, E. Pardon, J. Steyaert, S. Howorka, and H. Remaut. Sbsb structure and lattice reconstruction unveil ca2+ triggered s-layer assembly. *Nature*, 487(7405):119–22, 2012.

[15] J. F. Bazan and F. J. de Sauvage. Structural ties between cholesterol transport and morphogen signaling. *Cell*, 138(6):1055–6, 2009.

[16] P. A. Beachy, S. G. Hymowitz, R. A. Lazarus, D. J. Leahy, and C. Siebold. Interactions between hedgehog proteins and their binding partners come into view. *Genes Dev*, 24(18):2001–12, 2010.

[17] E. Belloni, M. Muenke, E. Roessler, G. Traverso, J. Siegel-Bartelt, A. Frumkin, H. F. Mitchell, H. Donis-Keller, C. Helms, A. V. Hing, H. H. Heng, B. Koop, D. Martindale, J. M. Rommens, L. C. Tsui, and S. W. Scherer. Identification of sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nat Genet*, 14(3):353–6, 1996.

[18] M.H. Bender, P.A. Hipskind, A.R. Capen, M. Cockman, K.M. Credille, H. Gao, J.A. Bastian, J.M. Clay, K.L. Lobb, D.J. Sall, M.L. Thompson, T. Wilson, G.N. Wishart, and B.K.R. Patel. Identification and characterization of a novel smoothened antagonist for the treatment of cancer with deregulated hedgehol signaling. *Cancer Res*, 71(8):2819, 2011.

[19] M. P. Bennett and D. C. Mitchell. Regulation of membrane proteins by dietary lipids: Effects of cholesterol and docosahexaenoic acid acyl chain-containing phospholipids on rhodopsin stability and function. *Biophysical Journal*, 95(3):1206–1216, 2008.

[20] D. M. Berman, S. S. Karhadkar, A. Maitra, R. Montes De Oca, M. R. Gerstenblith, K. Briggs, A. R. Parker, Y. Shimada, J. R. Eshleman, D. N. Watkins, and P. A. Beachy. Widespread requirement for hedgehog ligand stimulation in growth of digestive tract tumours. *Nature*, 425(6960):846–51, 2003.

[21] D. Bhattarai, J. H. Jung, S. Han, H. Lee, S. J. Oh, H. W. Ko, and K. Lee. Design, synthesis, and biological evaluation of structurally modified isoindolinone and quinazolinone derivatives as hedgehog pathway inhibitors. *Eur J Med Chem*, 125:1036–1050, 2017.

[22] M. Bidet, O. Joubert, B. Lacombe, M. Ciantar, R. Nehme, P. Mollat, L. Bretillon, H. Faure, R. Bittman, M. Ruat, and I. Mus-Veteau. The hedgehog receptor patched is involved in cholesterol transport. *PLoS One*, 6(9):e23834, 2011.

[23] M. F. Bijlsma, H. Damhofer, and H. Roelink. Hedgehog-stimulated chemotaxis is mediated by smoothened located outside the primary cilium. *Sci Signal*, 5(238):ra60, 2012.

[24] M. F. Bijlsma, C. A. Spek, D. Zivkovic, S. van de Water, F. Rezaee, and M. P. Peppelenbosch. Repression of smoothened by patched-dependent (pro-)vitamin d3 secretion. *PLoS Biol*, 4(8):e232, 2006.

[25] R. M. Bill, P. J. Henderson, S. Iwata, E. R. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C. G. Tate, and H. Vogel. Overcoming barriers to membrane protein structure determination. *Nat Biotechnol*, 29(4):335–40, 2011.

[26] R. Blassberg, J. I. Macrae, J. Briscoe, and J. Jacob. Reduced cholesterol levels impair smoothened activation in smith-lemli-opitz syndrome. *Hum Mol Genet*, 2015.

[27] D. K. Breslow, E. F. Koslover, F. Seydel, A. J. Spakowitz, and M. V. Nachury. An in vitro assay for entry into cilia reveals unique properties of the soluble diffusion barrier. *J Cell Biol*, 203(1):129–47, 2013.

[28] G. Bricogne, E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, A. Sharff, O.S. Smart, C. Vonrhein, and T.O. Womack. Buster, 2011.

[29] J. Briscoe, Y. Chen, T. M. Jessell, and G. Struhl. A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol Cell*, 7(6):1279–91, 2001.

[30] J. Briscoe and S. Small. Morphogen rules: design principles of gradient-mediated embryo patterning. *Development*, 142(23):3996–4009, 2015.

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

[32] A. T. Brunger. Free r value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355(6359):472–5, 1992.

[33] J. S. Burg, J. R. Ingram, A. J. Venkatakrishnan, K. M. Jude, A. Dukkipati, E. N. Feinberg, A. Angelini, D. Waghray, R. O. Dror, H. L. Ploegh, and K. C. Garcia. Structural biology. structural basis for chemokine recognition and activation of a viral g protein-coupled receptor. *Science*, 347(6226):1113–7, 2015.

[34] K. Burger, G. Gimpl, and F. Fahrenholz. Regulation of receptor function by cholesterol. *Cell Mol Life Sci*, 57(11):1577–92, 2000.

[35] R. Burke, D. Nellen, M. Bellotto, E. Hafen, K. A. Senti, B. J. Dickson, and K. Basler. Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell*, 99(7):803–15, 1999.

[36] E. F. X. Byrne, R. Sircar, P. S. Miller, G. Hedger, G. Luchetti, S. Nachtergaele, M. D. Tully, L. Mydock-McGrane, D. F. Covey, R. P. Rambo, M. S. P. Sansom, S. Newstead, R. Rohatgi, and C. Siebold. Structural basis of smoothened regulation by its extracellular domains. *Nature*, 535(7613):517–522, 2016.

[37] M. Caffrey and V. Cherezov. Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc*, 4(5):706–31, 2009.

[38] 3rd Cahill, T. J., A. R. Thomsen, J. T. Tarrasch, B. Plouffe, A. H. Nguyen, F. Yang, L. Y. Huang, A. W. Kahsai, D. L. Bassoni, B. J. Gavino, J. E. Lamerdin, S. Triest, A. K. Shukla, B. Berger, J. th Little, A. Antar, A. Blanc, C. X. Qu, X. Chen, K. Kawakami, A. Inoue, J. Aoki, J. Steyaert, J. P. Sun, M. Bouvier, G. Skiniotis, and R. J. Lefkowitz. Distinct conformations of gpcr-beta-arrestin complexes mediate desensitization, signaling, and endocytosis. *Proc Natl Acad Sci U S A*, 114(10):2562–2567, 2017.

[39] C. J. Carbe, L. Cheng, S. Addya, J. I. Gold, E. Gao, W. J. Koch, and N. A. Riobo. Gi proteins mediate activation of the canonical hedgehog pathway in the myocardium. *Am J Physiol Heart Circ Physiol*, 307(1):H66–72, 2014.

[40] B. Carpenter, R. Nehme, T. Warne, A. G. Leslie, and C. G. Tate. Structure of the adenosine a(2a) receptor bound to an engineered g protein. *Nature*, 536(7614):104–7, 2016.

[41] C. E. Carroll, S. Marada, D. P. Stewart, J. X. Ouyang, and S. K. Ogden. The extracellular loops of smoothened play a regulatory role in control of hedgehog pathway activation. *Development*, 139(3):612–21, 2012.

[42] C. Carron, A. Pascal, A. Djiane, J. C. Boucaut, D. L. Shi, and M. Umbhauer. Frizzled receptor dimerization is sufficient to activate the wnt/beta-catenin pathway. *J Cell Sci*, 116(Pt 12):2541–50, 2003.

[43] M. Carvalho, D. Schwudke, J. L. Sampaio, W. Palm, I. Riezman, G. Dey, G. D. Gupta, S. Mayor, H. Riezman, A. Shevchenko, T. V. Kurzchalia, and S. Eaton. Survival strategies of a sterol auxotroph. *Development*, 137(21):3675–85, 2010.

[44] D. Casey, S. Demko, S. Shord, H. Zhao, H. Chen, K. He, A. Putman, W. Helms, P. Keegan, and R. Pazdur. Fda approval summary: Sonidegib for locally advanced basal cell carcinoma. *Clin Cancer Res*, 23(10):2377–2381, 2017.

[45] P. S. Chae, S. G. Rasmussen, R. R. Rana, K. Gotfryd, R. Chandra, M. A. Goren, A. C. Kruse, S. Nurva, C. J. Loland, Y. Pierre, D. Drew, J. L. Popot, D. Picot, B. G. Fox, L. Guan, U. Gether, B. Byrne, B. Kobilka, and S. H. Gellman. Maltose-neopentyl glycol (mng) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat Methods*, 7(12):1003–8, 2010.

[46] B. Chailley and E. Boisvieux-Ulrich. Detection of plasma membrane cholesterol by filipin during microvillogenesis and ciliogenesis in quail oviduct. *J Histochem Cytochem*, 33(1):1–10, 1985.

[47] D. T. Chang, A. Lopez, D. P. von Kessler, C. Chiang, B. K. Simandl, R. Zhao, M. F. Seldin, J. F. Fallon, and P. A. Beachy. Products, genetic linkage and limb patterning activity of a murine hedgehog gene. *Development*, 120(11):3339–53, 1994.

[48] S. D. Chang and M. R. Bruchas. Functional selectivity at gpcrs: new opportunities in psychiatric drug discovery. *Neuropsychopharmacology*, 39(1):248–9, 2014.

[49] M. Chavez, S. Ena, J. Van Sande, A. de Kerchove d'Exaerde, S. Schurmans, and S. N. Schiffmann. Modulation of ciliary phosphoinositide content regulates trafficking and sonic hedgehog signaling output. *Dev Cell*, 34(3):338–50, 2015.

[50] B. Chen, V. Trang, A. Lee, N. S. Williams, A. N. Wilson, Jr. Epstein, E. H., J. Y. Tang, and J. Kim. Posaconazole, a second-generation triazole antifungal drug, inhibits the hedgehog signaling pathway and progression of basal cell carcinoma. *Mol Cancer Ther*, 15(5):866–76, 2016.

[51] C. Chen, J. Ke, X. E. Zhou, W. Yi, J. S. Brunzelle, J. Li, E. L. Yong, H. E. Xu, and K. Melcher. Structural basis for molecular recognition of folic acid by folate receptors. *Nature*, 500(7463):486–9, 2013.

[52] J. K. Chen, J. Taipale, M. K. Cooper, and P. A. Beachy. Inhibition of hedgehog signaling by direct binding of cyclopamine to smoothened. *Genes Dev*, 16(21):2743–8, 2002.

[53] J. K. Chen, J. Taipale, K. E. Young, T. Maiti, and P. A. Beachy. Small molecule modulation of smoothened activity. *Proc Natl Acad Sci U S A*, 99(22):14071–6, 2002.

[54] M. H. Chen, Y. J. Li, T. Kawakami, S. M. Xu, and P. T. Chuang. Palmitoylation is required for the production of a soluble multimeric hedgehog protein complex and long-range signaling in vertebrates. *Genes Dev*, 18(6):641–59, 2004.

[55] V. B. Chen, 3rd Arendall, W. B., J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 1):12–21, 2010.

[56] W. Chen, X. R. Ren, C. D. Nelson, L. S. Barak, J. K. Chen, P. A. Beachy, F. de Sauvage, and R. J. Lefkowitz. Activity-dependent internalization of smoothened mediated by beta-arrestin 2 and grk2. *Science*, 306(5705):2257–60, 2004.

[57] Y. Chen, N. Sasai, G. Ma, T. Yue, J. Jia, J. Briscoe, and J. Jiang. Sonic hedgehog dependent phosphorylation by ck1alpha and grk2 is required for ciliary accumulation and activation of smoothened. *PLoS Biol*, 9(6):e1001083, 2011.

[58] Y. Chen and G. Struhl. Dual roles for patched in sequestering and transducing hedgehog. *Cell*, 87(3):553–63, 1996.

[59] Y. Chen and G. Struhl. In vivo evidence that patched and smoothened constitute distinct binding and transducing components of a hedgehog receptor complex. *Development*, 125(24):4943–8, 1998.

[60] V. Cherezov, M. A. Hanson, M. T. Griffith, M. C. Hilgart, R. Sanishvili, V. Nagarajan, S. Stepanov, R. F. Fischetti, P. Kuhn, and R. C. Stevens. Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 microm size x-ray synchrotron beam. *J R Soc Interface*, 6 Suppl 5:S587–97, 2009.

[61] V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, and R. C. Stevens. High-resolution crystal structure of an engineered human beta2-adrenergic g protein-coupled receptor. *Science*, 318(5854):1258–65, 2007.

[62] A. L. Cherry, C. Finta, M. Karlstrom, Q. Jin, T. Schwend, J. Astorga-Wells, R. A. Zubarev, M. Del Campo, A. R. Criswell, D. de Sanctis, L. Jovine, and R. Toftgard. Structural basis of sufu-gli interaction in human hedgehog signalling regulation. *Acta Crystallogr D Biol Crystallogr*, 69(Pt 12):2563–79, 2013.

[63] H. O. Cheung, X. Zhang, A. Ribeiro, R. Mo, S. Makino, V. Puviindran, K. K. Law, J. Briscoe, and C. C. Hui. The kinesin protein kif7 is a critical regulator of gli transcription factors in mammalian hedgehog signaling. *Sci Signal*, 2(76):ra29, 2009.

[64] C. Chiang, Y. Litingtung, E. Lee, K. E. Young, J. L. Corden, H. Westphal, and P. A. Beachy. Cyclopia and defective axial patterning in mice lacking sonic hedgehog gene function. *Nature*, 383(6599):407–13, 1996.

[65] C. Chiang, Y. Litingtung, E. Lee, K. E. Young, J. L. Corden, H. Westphal, and P. A. Beachy. Cyclopia and defective axial patterning in mice lacking sonic hedgehog gene function. *Nature*, 383(6599):407–13, 1996.

[66] P. Chinchilla, L. Xiao, M. G. Kazanietz, and N. A. Riobo. Hedgehog proteins activate pro-angiogenic responses in endothelial cells through non-canonical signaling pathways. *Cell Cycle*, 9(3):570–79, 2010.

[67] H. W. Choe, Y. J. Kim, J. H. Park, T. Morizumi, E. F. Pai, N. Krauss, K. P. Hofmann, P. Scheerer, and O. P. Ernst. Crystal structure of metarhodopsin ii. *Nature*, 471(7340):651–5, 2011.

[68] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, and J. Damborsky. Caver 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol*, 8(10):e1002708, 2012.

[69] A. E. Christian, H. S. Byun, N. Zhong, M. Wanunu, T. Marti, A. Furer, F. Diederich, R. Bittman, and G. H. Rothblat. Comparison of the capacity of beta-cyclodextrin derivatives and cyclophanes to shuttle cholesterol between cells and serum lipoproteins. *J Lipid Res*, 40(8):1475–82, 1999.

[70] A. E. Christian, M. P. Haynes, M. C. Phillips, and G. H. Rothblat. Use of cyclodextrins for manipulating cellular cholesterol content. *J Lipid Res*, 38(11):2264–72, 1997.

[71] R. Chu, J. Takei, J. R. Knowlton, M. Andrykovitch, W. Pei, A. V. Kajava, P. J. Steinbach, X. Ji, and Y. Bai. Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by nmr and x-ray crystallography. *J Mol Biol*, 323(2):253–62, 2002.

[72] E. Chun, A. A. Thompson, W. Liu, C. B. Roth, M. T. Griffith, V. Katritch, J. Kunken, F. Xu, V. Cherezov, M. A. Hanson, and R. C. Stevens. Fusion partner toolchest for the stabilization and crystallization of g protein-coupled receptors. *Structure*, 20(6):967–76, 2012.

[73] V. Clement, P. Sanchez, N. de Tribolet, I. Radovanovic, and A. Ruiz i Altaba. Hedgehog-gli1 signaling regulates human glioma growth, cancer stem cell self-renewal, and tumorigenicity. *Curr Biol*, 17(2):165–72, 2007.

[74] M. J. Cohn and C. Tickle. Limbs: a model for pattern formation within the vertebrate body plan. *Trends Genet*, 12(7):253–7, 1996.

[75] J. Cooke and J. H. Sang. Utilization of sterols by larvae of drosophila melanogaster. *J Insect Physiol*, 16(5):801–12, 1970.

[76] M. A. Cooper. Current biosensor technologies in drug discovery. *Drug Discovery World*, pages 68–82, 2006 2006.

[77] M. K. Cooper, J. A. Porter, K. E. Young, and P. A. Beachy. Teratogen-mediated inhibition of target tissue response to shh signaling. *Science*, 280(5369):1603–7, 1998.

[78] M. K. Cooper, C. A. Wassif, P. A. Krakowiak, J. Taipale, R. Gong, R. I. Kelley, F. D. Porter, and P. A. Beachy. A defective response to hedgehog signaling in disorders of cholesterol biosynthesis. *Nat Genet*, 33(4):508–13, 2003.

[79] K. C. Corbit, P. Aanstad, V. Singla, A. R. Norman, D. Y. Stainier, and J. F. Reiter. Vertebrate smoothened functions at the primary cilium. *Nature*, 437(7061):1018–21, 2005.

[80] R. B. Corcoran and M. P. Scott. Oxysterols stimulate sonic hedgehog signal transduction and proliferation of medulloblastoma cells. *Proc Natl Acad Sci U S A*, 103(22):8408–13, 2006.

[81] F. Corpet. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*, 16(22):10881–90, 1988.

[82] K. Cowtan. The buccaneer software for automated model building. 1. tracing protein chains. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 9):1002–11, 2006.

[83] K. Cowtan. Recent developments in classical density modification. *Acta Crystallographica Section D-Biological Crystallography*, 66:470–478, 2010.

[84] P. Cuevas and J. A. Gutierrez Diaz. Absence of filipin-sterol complexes from the ciliary necklace of ependymal cells. *Anat Embryol (Berl)*, 172(1):97–9, 1985.

[85] M. G. Cull and P. J. Schatz. Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol*, 326:430–40, 2000.

[86] C. E. Dann, J. C. Hsieh, A. Rattner, D. Sharma, J. Nathans, and D. J. Leahy. Insights into wnt binding and signalling from the structures of two frizzled cysteine-rich domains. *Nature*, 412(6842):86–90, 2001.

[87] J. P. Davies and Y. A. Ioannou. Topological analysis of niemann-pick c1 protein reveals that the membrane orientation of the putative sterol-sensing domain is identical to those of 3-hydroxy-3-methylglutaryl-coa reductase and sterol regulatory element binding protein cleavage-activating protein. *J Biol Chem*, 275(32):24367–74, 2000.

[88] S. J. Davis, M. J. Puklavec, D. A. Ashford, K. Harlos, E. Y. Jones, D. I. Stuart, and A. F. Williams. Expression of soluble recombinant glycoproteins with predefined glycosylation: application to the crystallization of the t-cell glycoprotein cd2. *Protein Eng*, 6(2):229–32, 1993.

[89] E. De Genst, K. Silence, K. Decanniere, K. Conrath, R. Loris, J. Kinne, S. Muyldermans, and L. Wyns. Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc Natl Acad Sci U S A*, 103(12):4586–91, 2006.

[90] E. J. De Genst, T. Guilliams, J. Wellens, E. M. O'Day, C. A. Waudby, S. Meehan, M. Dumoulin, S. T. Hsu, N. Cremades, K. H. Verschueren, E. Pardon, L. Wyns, J. Steyaert, J. Christodoulou, and C. M. Dobson. Structure and properties of a complex of alpha-synuclein and a single-domain camelid antibody. *J Mol Biol*, 402(2):326–43, 2010.

[91] P. G. DeCaen, M. Delling, T. N. Vien, and D. E. Clapham. Direct recording and molecular identification of the calcium channel of primary cilia. *Nature*, 504(7479):315–8, 2013.

[92] D. L. DeCamp, T. M. Thompson, F. J. de Sauvage, and M. R. Lerner. Smoothened activates galphai-mediated signaling in frog melanophores. *J Biol Chem*, 275(34):26322–7, 2000.

[93] K. Decanniere, S. Muyldermans, and L. Wyns. Canonical antigen-binding loop structures in immunoglobulins: more structures, more canonical classes? *J Mol Biol*, 300(1):83–91, 2000.

[94] M. Delling, P. G. DeCaen, J. F. Doerner, S. Febvay, and D. E. Clapham. Primary cilia are specialized calcium signalling organelles. *Nature*, 504(7479):311–4, 2013.

[95] N. Denef, D. Neubuser, L. Perez, and S. M. Cohen. Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothened. *Cell*, 102(4):521–31, 2000.

[96] A. Desmyter, T. R. Transue, M. A. Ghahroudi, M. H. Thi, F. Poortmans, R. Hamers, S. Muyldermans, and L. Wyns. Crystal structure of a camel single-domain vh antibody fragment in complex with lysozyme. *Nat Struct Biol*, 3(9):803–11, 1996.

[97] X. Deupi and J. Standfuss. Structural insights into agonist-induced activation of g-protein-coupled receptors. *Curr Opin Struct Biol*, 21(4):541–51, 2011.

[98] X. Deupi and J. Standfuss. Structural insights into agonist-induced activation of g-protein-coupled receptors. *Curr Opin Struct Biol*, 21(4):541–51, 2011.

[99] B. T. DeVree, J. P. Mahoney, G. A. Velez-Ruiz, S. G. Rasmussen, A. J. Kuszak, E. Edwald, J. J. Fung, A. Manglik, M. Masureel, Y. Du, R. A. Matt, E. Pardon, J. Steyaert, B. K. Kobilka, and R. K. Sunahara. Allosteric coupling from g protein to the agonist-binding pocket in gpcrs. *Nature*, 535(7610):182–6, 2016.

[100] W. X. Ding, X. R. Qi, P. Li, Y. Maitani, and T. Nagai. Cholesteryl hemisuccinate as a membrane stabilizer in dipalmitoylphosphatidylcholine liposomes containing saikosaponin-d. *Int J Pharm*, 300(1-2):38–47, 2005.

[101] K. Domanska, S. Vanderhaegen, V. Srinivasan, E. Pardon, F. Dupeux, J. A. Marquez, S. Giorgetti, M. Stoppini, L. Wyns, V. Bellotti, and J. Steyaert. Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic beta2-microglobulin variant. *Proc Natl Acad Sci U S A*, 108(4):1314–9, 2011.

[102] A. S. Dore, K. Okrasa, J. C. Patel, M. Serrano-Vega, K. Bennett, R. M. Cooke, J. C. Errey, A. Jazayeri, S. Khan, B. Tehan, M. Weir, G. R. Wiggin, and F. H. Marshall. Structure of class c gpcr metabotropic glutamate receptor 5 transmembrane domain. *Nature*, 511(7511):557–62, 2014.

[103] K. V. Dorn, C. E. Hughes, and R. Rohatgi. A smoothened-evc2 complex transduces the hedgehog signal at primary cilia. *Dev Cell*, 23(4):823–35, 2012.

[104] D. Durand, C. Vives, D. Cannella, J. Perez, E. Pebay-Peyroula, P. Vachette, and F. Fieschi. Nadph oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J Struct Biol*, 169(1):45–53, 2010.

[105] J. R. Dwyer, N. Sever, M. Carlson, S. F. Nelson, P. A. Beachy, and F. Parhami. Oxysterols are novel activators of the hedgehog signaling pathway in pluripotent mesenchymal cells. *J Biol Chem*, 282(12):8959–68, 2007.

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

[107] Y. Echelard, D. J. Epstein, B. St-Jacques, L. Shen, J. Mohler, J. A. McMahon, and A. P. McMahon. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of cns polarity. *Cell*, 75(7):1417–30, 1993.

[108] I. A. Ehrnstorfer, E. R. Geertsma, E. Pardon, J. Steyaert, and R. Dutzler. Crystal structure of a slc11 (nramp) transporter reveals the basis for transition-metal ion transport. *Nat Struct Mol Biol*, 21(11):990–6, 2014.

[109] D. Eisenberg, E. Schwarz, M. Komaromy, and R. Wall. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology*, 179(1):125–142, 1984.

[110] B. T. Emmer, D. Maric, and D. M. Engman. Molecular mechanisms of protein and lipid targeting to ciliary membranes. *J Cell Sci*, 123(Pt 4):529–36, 2010.

[111] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 12 Pt 1):2126–32, 2004.

[112] P. R. Evans and G. N. Murshudov. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr*, 69(Pt 7):1204–14, 2013.

[113] M. Fairhead and M. Howarth. Site-specific biotinylation of purified proteins using bira. *Methods Mol Biol*, 1266:171–84, 2015.

[114] C. W. Fan, B. Chen, I. Franco, J. Lu, H. Shi, S. Wei, C. Wang, X. Wu, W. Tang, M. G. Roth, N. S. Williams, E. Hirsch, C. Chen, and L. Lum. The hedgehog pathway effector smoothened exhibits signaling competency in the absence of ciliary accumulation. *Chem Biol*, 2014.

[115] M. J. Fietz, J. P. Concordet, R. Barbosa, R. Johnson, S. Krauss, A. P. McMahon, C. Tabin, and P. W. Ingham. The hedgehog gene family in drosophila and vertebrate development. *Dev Suppl*, pages 43–51, 1994.

[116] M. J. Fietz, A. Jacinto, A. M. Taylor, C. Alexandre, and P. W. Ingham. Secretion of the amino-terminal fragment of the hedgehog protein is necessary and sufficient for hedgehog signalling in drosophila. *Curr Biol*, 5(6):643–50, 1995.

[117] E. Folta-Stogniew. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *Methods Mol Biol*, 328:97–112, 2006.

[118] A. J. Forbes, Y. Nakano, A. M. Taylor, and P. W. Ingham. Genetic analysis of hedgehog signalling in the drosophila embryo. *Dev Suppl*, pages 115–24, 1993.

[119] M. Frank-Kamenetsky, X. M. Zhang, S. Bottega, O. Guicherit, H. Wichterle, H. Dudek, D. Bumcrot, F. Y. Wang, S. Jones, J. Shulok, L. L. Rubin, and J. A. Porter. Small-molecule modulators of hedgehog signaling: identification and characterization of smoothened agonists and antagonists. *J Biol*, 1(2):10, 2002.

[120] M. Frank-Kamenetsky, X. M. Zhang, S. Bottega, O. Guicherit, H. Wichterle, H. Dudek, D. Bumcrot, F. Y. Wang, S. Jones, J. Shulok, L. L. Rubin, and J. A. Porter. Small-molecule modulators of hedgehog signaling: identification and characterization of smoothened agonists and antagonists. *J Biol*, 1(2):10, 2002.

[121] L. G. Frenken, R. H. van der Linden, P. W. Hermans, J. W. Bos, R. C. Ruuls, B. de Geus, and C. T. Verrips. Isolation of antigen specific llama vhh antibody fragments and their high level secretion by saccharomyces cerevisiae. *J Biotechnol*, 78(1):11–21, 2000.

[122] N. Fuse, T. Maiti, B. Wang, J. A. Porter, T. M. Hall, D. J. Leahy, and P. A. Beachy. Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proc Natl Acad Sci U S A*, 96(20):10992–9, 1999.

[123] M. R. Gailani, M. Stahle-Backdahl, D. J. Leffell, M. Glynn, P. G. Zaphiropoulos, C. Pressman, A. B. Unden, M. Dean, D. E. Brash, A. E. Bale, and R. Toftgard. The role of the human homologue of drosophila patched in sporadic basal cell carcinomas. *Nat Genet*, 14(1):78–81, 1996.

[124] A. Gallet, R. Rodriguez, L. Ruel, and P. P. Therond. Cholesterol modification of hedgehog is required for trafficking and movement, revealing an asymmetric cellular response to hedgehog. *Dev Cell*, 4(2):191–204, 2003.

[125] A. Gallet, L. Ruel, L. Staccini-Lavenant, and P. P. Therond. Cholesterol modification is necessary for controlled planar long-range activity of hedgehog in drosophila epithelia. *Development*, 133(3):407–18, 2006.

[126] F. R. Garcia-Gonzalo, S. C. Phua, E. C. Roberson, 3rd Garcia, G., M. Abedin, S. Schurmans, T. Inoue, and J. F. Reiter. Phosphoinositides regulate ciliary protein trafficking to modulate hedgehog signaling. *Dev Cell*, 34(4):400–9, 2015.

[127] E. R. Geertsma, Y. N. Chang, F. R. Shaik, Y. Neldner, E. Pardon, J. Steyaert, and R. Dutzler. Structure of a prokaryotic fumarate transporter reveals the architecture of the slc26 family. *Nat Struct Mol Biol*, 22(10):803–8, 2015.

[128] E. Ghosh, P. Kumari, D. Jaiman, and A. K. Shukla. Methodological advances: the unsung heroes of the gpcr structural revolution. *Nat Rev Mol Cell Biol*, 16(2):69–81, 2015.

[129] G. Gil, J. R. Faust, D. J. Chin, J. L. Goldstein, and M. S. Brown. Membrane-bound domain of hmg coa reductase is required for sterol-enhanced degradation of the enzyme. *Cell*, 41(1):249–58, 1985.

[130] G. Gimpl and F. Fahrenholz. Cholesterol as stabilizer of the oxytocin receptor. *Biochim Biophys Acta*, 1564(2):384–92, 2002.

[131] F. Giusti, J. Rieger, L. J. Catoire, S. Qian, A. N. Calabrese, T. G. Watkinson, M. Casiraghi, S. E. Radford, A. E. Ashcroft, and J. L. Popot. Synthesis, characterization and applications of a perdeuterated amphipol. *J Membr Biol*, 247(9-10):909–24, 2014.

[132] S. C. Goetz and K. V. Anderson. The primary cilium: a signalling centre during vertebrate development. *Nat Rev Genet*, 11(5):331–44, 2010.

[133] Y. Gohon, T. Dahmane, R. W. Ruigrok, P. Schuck, D. Charvolin, F. Rappaport, P. Timmins, D. M. Engelman, C. Tribet, J. L. Popot, and C. Ebel. Bacteriorhodopsin/amphipol complexes: structural and functional properties. *Biophys J*, 94(9):3523–37, 2008.

[134] S. E. Gould, J. A. Low, Jr. Marsters, J. C., K. Robarge, L. L. Rubin, F. J. de Sauvage, D. P. Sutherlin, H. Wong, and R. L. Yauch. Discovery and preclinical development of vismodegib. *Expert Opin Drug Discov*, pages 1–16, 2014.

[135] T. Guilliams, F. El-Turk, A. K. Buell, E. M. O'Day, F. A. Aprile, E. K. Esbjorner, M. Vendruscolo, N. Cremades, E. Pardon, L. Wyns, M. E. Welland, J. Steyaert, J. Christodoulou, C. M. Dobson, and E. De Genst. Nanobodies raised against monomeric alpha-synuclein distinguish between fibrils at different maturation stages. *J Mol Biol*, 425(14):2397–411, 2013.

[136] M. K. Hadden. Hedgehog pathway inhibitors: a patent review (2009–present). *Expert Opin Ther Pat*, 23(3):345–61, 2013.

[137] H. Hahn, C. Wicking, P. G. Zaphiropoulous, M. R. Gailani, S. Shanley, A. Chidambaram, I. Vorechovsky, E. Holmberg, A. B. Unden, S. Gillies, K. Negus, I. Smyth, C. Pressman, D. J. Leffell, B. Gerrard, A. M. Goldstein, M. Dean, R. Toftgard, G. Chenevix-Trench, B. Wainwright, and A. E. Bale. Mutations of the human homolog of drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell*, 85(6):841–51, 1996.

[138] H. Hahn, L. Wojnowski, G. Miller, and A. Zimmer. The patched signaling pathway in tumorigenesis and development: lessons from animal models. *J Mol Med (Berl)*, 77(6):459–68, 1999.

[139] C. Hamers-Casterman, T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman, and R. Hamers. Naturally occurring antibodies devoid of light chains. *Nature*, 363(6428):446–8, 1993.

[140] M. Hammerschmidt and A. P. McMahon. The effect of pertussis toxin on zebrafish development: a possible role for inhibitory g-proteins in hedgehog signaling. *Dev Biol*, 194(2):166–71, 1998.

[141] M. A. Hanson, V. Cherezov, M. T. Griffith, C. B. Roth, V. P. Jaakola, E. Y. Chien, J. Velasquez, P. Kuhn, and R. C. Stevens. A specific cholesterol binding

site is established by the 2.8 a structure of the human beta2-adrenergic receptor. *Structure*, 16(6):897–905, 2008.

[142] G. Hassaine, C. Deluz, L. Grasso, R. Wyss, M. B. Tol, R. Hovius, A. Graff, H. Stahlberg, T. Tomizaki, A. Desmyter, C. Moreau, X. D. Li, F. Poitevin, H. Vogel, and H. Nury. X-ray structure of the mouse serotonin 5-ht3 receptor. *Nature*, 512(7514):276–81, 2014.

[143] M. Hattori, R. E. Hibbs, and E. Gouaux. A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Structure*, 20(8):1293–9, 2012.

[144] F. M. Heydenreich, Z. Vuckovic, M. Matkovic, and D. B. Veprintsev. Stabilization of g protein-coupled receptors by point mutations. *Front Pharmacol*, 6:82, 2015.

[145] A. Hidalgo and P. Ingham. Cell patterning in the drosophila segment: spatial regulation of the segment polarity gene patched. *Development*, 110(1):291–301, 1990.

[146] L. Hoch, H. Faure, H. Roudaut, A. Schoenfelder, A. Mann, N. Girard, L. Bihannic, O. Ayrault, E. Petricci, M. Taddei, D. Rognan, and M. Ruat. Mrt-92 inhibits hedgehog signaling by blocking overlapping binding sites in the transmembrane domain of the smoothened receptor. *FASEB J*, 29(5):1817–29, 2015.

[147] K. Hollenstein, J. Kean, A. Bortolato, R. K. Cheng, A. S. Dore, A. Jazayeri, R. M. Cooke, M. Weir, and F. H. Marshall. Structure of class b gpcr corticotropin-releasing factor receptor 1. *Nature*, 499(7459):438–43, 2013.

[148] J. E. Hooper and M. P. Scott. The drosophila patched gene encodes a putative membrane protein required for segmental patterning. *Cell*, 59(4):751–65, 1989.

[149] P. Huang, D. Nedelcu, M. Watanabe, C. Jao, Y. Kim, J. Liu, and A. Salic. Cellular cholesterol directly activates smoothened in hedgehog signaling. *Cell*, 166(5):1176–1187 e14, 2016.

[150] W. Huang, A. Manglik, A. J. Venkatakrishnan, T. Laeremans, E. N. Feinberg, A. L. Sanborn, H. E. Kato, K. E. Livingston, T. S. Thorsen, R. C. Kling, S. Granier, P. Gmeiner, S. M. Husbands, J. R. Traynor, W. I. Weis, J. Steyaert, R. O. Dror, and B. K. Kobilka. Structural insights into micro-opioid receptor activation. *Nature*, 524(7565):315–21, 2015.

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

[152] D. Huangfu, A. Liu, A. S. Rakeman, N. S. Murcia, L. Niswander, and K. V. Anderson. Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature*, 426(6962):83–7, 2003.

[153] E. W. Humke, K. V. Dorn, L. Milenkovic, M. P. Scott, and R. Rohatgi. The output of hedgehog signaling is controlled by the dynamic association between suppressor of fused and the gli proteins. *Genes Dev*, 24(7):670–82, 2010.

[154] E. Ikonen. Cellular cholesterol trafficking and compartmentalization. *Nat Rev Mol Cell Biol*, 9(2):125–38, 2008.

[155] J. P. Incardona, W. Gaffield, R. P. Kapur, and H. Roelink. The teratogenic veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction. *Development*, 125(18):3553–3562, 1998.

[156] J. P. Incardona, W. Gaffield, Y. Lange, A. Cooney, P. G. Pentchev, S. Liu, J. A. Watson, R. P. Kapur, and H. Roelink. Cyclopamine inhibition of sonic hedgehog signal transduction is not mediated through effects on cholesterol transport. *Dev Biol*, 224(2):440–52, 2000.

[157] M. F. Incardona, G. P. Bourenkov, K. Levik, R. A. Pieritz, A. N. Popov, and O. Svensson. Edna: a framework for plugin-based applications applied to x-ray experiment online data analysis. *J Synchrotron Radiat*, 16(Pt 6):872–9, 2009.

[158] P. W. Ingham. Localized hedgehog activity controls spatial limits of wingless transcription in the drosophila embryo. *Nature*, 366(6455):560–2, 1993.

[159] P. W. Ingham. Signalling by hedgehog family proteins in drosophila and vertebrate development. *Curr Opin Genet Dev*, 5(4):492–8, 1995.

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

[161] P. W. Ingham, S. Nystedt, Y. Nakano, W. Brown, D. Stark, M. van den Heuvel, and A. M. Taylor. Patched represses the hedgehog signalling pathway by promoting modification of the smoothened protein. *Curr Biol*, 10(20):1315–8, 2000.

[162] P. W. Ingham, A. M. Taylor, and Y. Nakano. Role of the drosophila patched gene in positional signalling. *Nature*, 353(6340):184–7, 1991.

[163] R. Irannejad, J. C. Tomshine, J. R. Tomshine, M. Chevalier, J. P. Mahoney, J. Steyaert, S. G. Rasmussen, R. K. Sunahara, H. El-Samad, B. Huang, and M. von Zastrow. Conformational biosensors reveal gpcr signalling from endosomes. *Nature*, 495(7442):534–8, 2013.

[164] V. P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman, and R. C. Stevens. The 2.6 angstrom crystal structure of a human a2a adenosine receptor bound to an antagonist. *Science*, 322(5905):1211–7, 2008.

[165] C. Y. Janda, D. Waghray, A. M. Levin, C. Thomas, and K. C. Garcia. Structural basis of wnt recognition by frizzled. *Science*, 337(6090):59–64, 2012.

[166] T. M. Jessell. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat Rev Genet*, 1(1):20–9, 2000.

[167] K. Jiang, Y. Liu, J. Fan, J. Zhang, X. A. Li, B. M. Evers, H. Zhu, and J. Jia. Pi(4)p promotes phosphorylation and conformational change of smoothened through interaction with its c-terminal tail. *PLoS Biol*, 14(2):e1002375, 2016.

[168] R. L. Johnson, A. L. Rothman, J. Xie, L. V. Goodrich, J. W. Bare, J. M. Bonifas, A. G. Quinn, R. M. Myers, D. R. Cox, Jr. Epstein, E. H., and M. P. Scott. Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science*, 272(5268):1668–71, 1996.

[169] B. Jung, D. Padula, I. Burtscher, C. Landerer, D. Lutter, F. Theis, A. C. Messias, A. Geerlof, M. Sattler, E. Kremmer, K. Boldt, M. Ueffing, and H. Lickert. Pitchfork and gprasp2 target smoothened to the primary cilium for hedgehog pathway activation. *PLoS One*, 11(2):e0149477, 2016.

[170] W. Kabsch. Xds. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2):125–32, 2010.

[171] Y. Kang, X. E. Zhou, X. Gao, Y. He, W. Liu, A. Ishchenko, A. Barty, T. A. White, O. Yefanov, G. W. Han, Q. Xu, P. W. de Waal, J. Ke, M. H. Tan, C. Zhang, A. Moeller, G. M. West, B. D. Pascal, N. Van Eps, L. N. Caro, S. A. Vishnivetskiy, R. J. Lee, K. M. Suino-Powell, X. Gu, K. Pal, J. Ma, X. Zhi, S. Boutet, G. J. Williams, M. Messerschmidt, C. Gati, N. A. Zatsepin,

D. Wang, D. James, S. Basu, S. Roy-Chowdhury, C. E. Conrad, J. Coe, H. Liu, S. Lisova, C. Kupitz, I. Grotjohann, R. Fromme, Y. Jiang, M. Tan, H. Yang, J. Li, M. Wang, Z. Zheng, D. Li, N. Howe, Y. Zhao, J. Standfuss, K. Diederichs, Y. Dong, C. S. Potter, B. Carragher, M. Caffrey, H. Jiang, H. N. Chapman, J. C. Spence, P. Fromme, U. Weierstall, O. P. Ernst, V. Katritch, V. V. Gurevich, P. R. Griffin, W. L. Hubbell, R. C. Stevens, V. Cherezov, K. Melcher, and H. E. Xu. Crystal structure of rhodopsin bound to arrestin by femtosecond x-ray laser. *Nature*, 523(7562):561–7, 2015.

[172] S. S. Karhadkar, G. S. Bova, N. Abdallah, S. Dhara, D. Gardner, A. Maitra, J. T. Isaacs, D. M. Berman, and P. A. Beachy. Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *Nature*, 431(7009):707–12, 2004.

[173] P. A. Karplus and K. Diederichs. Linking crystallographic model and data quality. *Science*, 336(6084):1030–3, 2012.

[174] V. Katritch, V. Cherezov, and R. C. Stevens. Structure-function of the g protein-coupled receptor superfamily. *Annu Rev Pharmacol Toxicol*, 53:531–56, 2013.

[175] T. Kawate and E. Gouaux. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure*, 14(4):673–81, 2006.

[176] B. T. Keady, R. Samtani, K. Tobita, M. Tsuchya, J. T. San Agustin, J. A. Follit, J. A. Jonassen, R. Subramanian, C. W. Lo, and G. J. Pazour. Ift25 links the signal-dependent movement of hedgehog components to intraflagellar transport. *Dev Cell*, 22(5):940–51, 2012.

[177] R. F. Keeler. Teratogenic compounds of veratrum californicum (durand) x. cyclopia in rabbits produced by cyclopamine. *Teratology*, 3(2):175–80, 1970.

[178] R. F. Keeler and W. Binns. Teratogenic compounds of veratrum californicum (durand). v. comparison of cyclopian effects of steroidal alkaloids from the plant and structurally related compounds from other sources. *Teratology*, 1(1):5–10, 1968.

[179] R. I. Kelley and G. E. Herman. Inborn errors of sterol biosynthesis. *Annu Rev Genomics Hum Genet*, 2:299–341, 2001.

[180] T. Kenakin and L. J. Miller. Seven transmembrane receptors as shapeshifting proteins: the impact of allosteric modulation and functional selectivity on new drug discovery. *Pharmacol Rev*, 62(2):265–304, 2010.

[181] P. Keov, P. M. Sexton, and A. Christopoulos. Allosteric modulation of g protein-coupled receptors: a pharmacological perspective. *Neuropharmacology*, 60(1):24–35, 2011.

[182] H. Khaliullina, D. Panakova, C. Eugster, F. Riedel, M. Carvalho, and S. Eaton. Patched regulates smoothened trafficking using lipoprotein-derived lipids. *Development*, 136(24):4111–21, 2009.

[183] S. Khamrui, S. Turley, E. Pardon, J. Steyaert, E. Fan, C. L. Verlinde, L. W. Bergman, and W. G. Hol. The structure of the d3 domain of plasmodium falciparum myosin tail interacting protein mtip in complex with a nanobody. *Mol Biochem Parasitol*, 190(2):87–91, 2013.

[184] G. Khelashvili, A. Grossfield, S. E. Feller, M. C. Pitman, and H. Weinstein. Structural and dynamic effects of cholesterol at preferred sites of interaction with rhodopsin identified from microsecond length molecular dynamics simulations. *Proteins*, 76(2):403–17, 2009.

[185] E. P. Kilsdonk, P. G. Yancey, G. W. Stoudt, F. W. Bangerter, W. J. Johnson, M. C. Phillips, and G. H. Rothblat. Cellular cholesterol efflux mediated by cyclodextrins. *J Biol Chem*, 270(29):17250–6, 1995.

[186] J. Kim, B. T. Aftab, J. Y. Tang, D. Kim, A. H. Lee, M. Rezaee, J. Kim, B. Chen, E. M. King, A. Borodovsky, G. J. Riggins, Jr. Epstein, E. H., P. A. Beachy, and C. M. Rudin. Itraconazole and arsenic trioxide inhibit hedgehog pathway activation and tumor growth associated with acquired resistance to smoothened antagonists. *Cancer Cell*, 23(1):23–34, 2013.

[187] J. Kim, A. B. Cantor, S. H. Orkin, and J. Wang. Use of in vivo biotinylation to study protein-protein and protein-dna interactions in mouse embryonic stem cells. *Nat Protoc*, 4(4):506–17, 2009.

[188] J. Kim, E. Y. Hsia, A. Brigui, A. Plessis, P. A. Beachy, and X. Zheng. The role of ciliary trafficking in hedgehog receptor signaling. *Sci Signal*, 8(379):ra55, 2015.

[189] J. Kim, M. Kato, and P. A. Beachy. Gli2 trafficking links hedgehog-dependent activation of smoothened in the primary cilium to transcriptional activation in the nucleus. *Proc Natl Acad Sci U S A*, 106(51):21666–71, 2009.

[190] J. Kim, J. Y. Tang, R. Gong, J. Kim, J. J. Lee, K. V. Clemons, C. R. Chong, K. S. Chang, M. Fereshteh, D. Gardner, T. Reya, J. O. Liu, E. H. Epstein, D. A. Stevens, and P. A. Beachy. Itraconazole, a commonly used antifungal that inhibits hedgehog pathway activity and cancer growth. *Cancer Cell*, 17(4):388–99, 2010.

[191] U. Klein, G. Gimpl, and F. Fahrenholz. Alteration of the myometrial plasma membrane cholesterol content with beta-cyclodextrin modulates the binding affinity of the oxytocin receptor. *Biochemistry*, 34(42):13784–93, 1995.

[192] B. K. Kobilka and X. Deupi. Conformational complexity of g-protein-coupled receptors. *Trends Pharmacol Sci*, 28(8):397–406, 2007.

[193] C. Koike, T. Mizutani, T. Ito, Y. Shimizu, N. Yamamichi, T. Kameda, E. Michimukai, N. Kitamura, T. Okamoto, and H. Iba. Introduction of wild-type patched gene suppresses the oncogenic potential of human squamous cell carcinoma cell lines including a431. *Oncogene*, 21(17):2670–8, 2002.

[194] H. Koldso and M. S. Sansom. Organization and dynamics of receptor proteins in a plasma membrane. *J Am Chem Soc*, 2015.

[195] K. V. Korotkov, T. L. Johnson, M. G. Jobling, J. Pruneda, E. Pardon, A. Heroux, S. Turley, J. Steyaert, R. K. Holmes, M. Sandkvist, and W. G. Hol. Structural and functional studies on the interaction of gspc and gspd in the type ii secretion system. *PLoS Pathog*, 7(9):e1002228, 2011.

[196] K. V. Korotkov, E. Pardon, J. Steyaert, and W. G. Hol. Crystal structure of the n-terminal domain of the secretin gspd from etec determined with the assistance of a nanobody. *Structure*, 17(2):255–65, 2009.

[197] J. J. Kovacs, E. J. Whalen, R. Liu, K. Xiao, J. Kim, M. Chen, J. Wang, W. Chen, and R. J. Lefkowitz. Beta-arrestin-mediated localization of smoothened to the primary cilium. *Science*, 320(5884):1777–81, 2008.

[198] S. Krauss, J. P. Concordet, and P. W. Ingham. A functionally conserved homolog of the drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. *Cell*, 75(7):1431–44, 1993.

[199] E. Krissinel and K. Henrick. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 372(3):774–97, 2007.

[200] A. C. Kruse, A. M. Ring, A. Manglik, J. Hu, K. Hu, K. Eitel, H. Hubner, E. Pardon, C. Valant, P. M. Sexton, A. Christopoulos, C. C. Felder, P. Gmeiner, J. Steyaert, W. I. Weis, K. C. Garcia, J. Wess, and B. K. Kobilka. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature*, 504(7478):101–6, 2013.

[201] M. Kubo, M. Nakamura, A. Tasaki, N. Yamanaka, H. Nakashima, M. Nomura, S. Kuroki, and M. Katano. Hedgehog signaling pathway is a new therapeutic target for patients with breast cancer. *Cancer Res*, 64(17):6071–4, 2004.

[202] V. Kumar, A. K. Chaudhary, Y. Dong, H. A. Zhong, G. Mondal, F. Lin, V. Kumar, and R. I. Mahato. Design, synthesis and biological evaluation of novel hedgehog inhibitors for treating pancreatic cancer. *Sci Rep*, 7(1):1665, 2017.

[203] S. V. Kurkov and T. Loftsson. Cyclodextrins. *Int J Pharm*, 453(1):167–80, 2013.

[204] P. E. Kuwabara and M. Labouesse. The sterol-sensing domain: multiple families, a unique role? *Trends Genet*, 18(4):193–201, 2002.

[205] H. J. Kwon, L. Abi-Mosleh, M. L. Wang, J. Deisenhofer, J. L. Goldstein, M. S. Brown, and R. E. Infante. Structure of n-terminal domain of npc1 reveals distinct subdomains for binding and transfer of cholesterol. *Cell*, 137(7):1213–24, 2009.

[206] C. Lacroix, I. Fish, H. Torosyan, P. Parathaman, J. J. Irwin, B. K. Shoichet, and S. Angers. Identification of novel smoothened ligands using structure-based docking. *PLoS One*, 11(8):e0160365, 2016.

[207] A. Y. Lam, E. Pardon, K. V. Korotkov, W. G. Hol, and J. Steyaert. Nanobody-aided structure determination of the epsi:epsj pseudopilin heterodimer from vibrio vulnificus. *J Struct Biol*, 166(1):8–15, 2009.

[208] M. Lauth and R. Toftgard. The hedgehog pathway as a drug target in cancer therapy. *Curr Opin Investig Drugs*, 8(6):457–61, 2007.

[209] M. Lauwereys, M. Arbabi Ghahroudi, A. Desmyter, J. Kinne, W. Holzer, E. De Genst, L. Wyns, and S. Muyldermans. Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *EMBO J*, 17(13):3512–20, 1998.

[210] K. F. Lechtreck, J. M. Brown, J. L. Sampaio, J. M. Craft, A. Shevchenko, J. E. Evans, and G. B. Witman. Cycling of the signaling protein phospholipase d through cilia requires the bbsome only for the export phase. *J Cell Biol*, 201(2):249–61, 2013.

[211] J. J. Lee, S. C. Ekker, D. P. von Kessler, J. A. Porter, B. I. Sun, and P. A. Beachy. Autoproteolysis in hedgehog protein biogenesis. *Science*, 266(5190):1528–37, 1994.

[212] J. J. Lee, D. P. von Kessler, S. Parks, and P. A. Beachy. Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. *Cell*, 71(1):33–50, 1992.

[213] A. G. Leslie. Integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr*, 55(Pt 10):1696–702, 1999.

[214] I. Levitan, Y. Fang, A. Rosenhouse-Dantsker, and V. Romanenko. Cholesterol and ion channels. *Subcell Biochem*, 51:509–49, 2010.

[215] P. M. Lewis, M. P. Dunn, J. A. McMahon, M. Logan, J. F. Martin, B. St-Jacques, and A. P. McMahon. Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by ptc1. *Cell*, 105(5):599–612, 2001.

[216] L. Li, E. Park, J. Ling, J. Ingram, H. Ploegh, and T. A. Rapoport. Crystal structure of a substrate-engaged secy protein-translocation channel. *Nature*, 531(7594):395–9, 2016.

[217] Y. Li, H. Zhang, Y. Litingtung, and C. Chiang. Cholesterol modification restricts the spread of shh gradient in the limb bud. *Proc Natl Acad Sci U S A*, 103(17):6548–53, 2006.

[218] Y. L. Liang, M. Khoshouei, M. Radjainia, Y. Zhang, A. Glukhova, J. Tarrasch, D. M. Thal, S. G. B. Furness, G. Christopoulos, T. Coudrat, R. Danev, W. Baumeister, L. J. Miller, A. Christopoulos, B. K. Kobilka, D. Wootten, G. Skiniotis, and P. M. Sexton. Phase-plate cryo-em structure of a class b gpcr-g-protein complex. *Nature*, 2017.

[219] Jr. Liem, K. F., M. He, P. J. Ocbina, and K. V. Anderson. Mouse kif7/costal2 is a cilia-associated protein that regulates sonic hedgehog signaling. *Proc Natl Acad Sci U S A*, 106(32):13377–82, 2009.

[220] Y. Y. Lin, M. Welch, and S. Lieberman. The detection of 20s-hydroxycholesterol in extracts of rat brains and human placenta by a gas chromatograph/mass spectrometry technique. *J Steroid Biochem Mol Biol*, 85(1):57–61, 2003.

[221] D. Linke. Detergents: an overview. *Methods Enzymol*, 463:603–17, 2009.

[222] W. Liu, E. Chun, A. A. Thompson, P. Chubukov, F. Xu, V. Katritch, G. W. Han, C. B. Roth, L. H. Heitman, I. Jzerman AP, V. Cherezov, and R. C. Stevens. Structural basis for allosteric regulation of gpcrs by sodium ions. *Science*, 337(6091):232–6, 2012.

[223] C. Low, Y. H. Yau, E. Pardon, C. Jegerschold, L. Wahlin, E. M. Quistgaard, P. Moberg, S. Geifman-Shochat, J. Steyaert, and P. Nordlund. Nanobody mediated crystallization of an archeal mechanosensitive channel. *PLoS One*, 8(10):e77984, 2013.

[224] W. C. Low, C. Wang, Y. Pan, X. Y. Huang, J. K. Chen, and B. Wang. The decoupling of smoothened from galphai proteins has little effect on gli3 protein processing and hedgehog-regulated chick neural tube patterning. *Dev Biol*, 321(1):188–96, 2008.

[225] W. Lu, Y. Liu, H. Ma, J. Zheng, S. Tian, Z. Sun, L. Luo, J. Li, H. Zhang, Z. J. Yang, and X. Zhang. Design, synthesis, and structure-activity relationship of tetrahydropyrido[4,3-d]pyrimidine derivatives as potent smoothened antagonists with in vivo activity. *ACS Chem Neurosci*, 2017.

[226] G. Luchetti, R. Sircar, J. H. Kong, S. Nachtergaele, A. Sagner, E. F. Byrne, D. F. Covey, C. Siebold, and R. Rohatgi. Cholesterol activates the g-protein coupled receptor smoothened to promote hedgehog signaling. *Elife*, 5, 2016.

[227] D. Lutjohann, O. Breuer, G. Ahlborg, I. Nennesmo, A. Siden, U. Diczfalusy, and I. Bjorkhem. Cholesterol homeostasis in human brain: evidence for an age-dependent flux of 24s-hydroxycholesterol from the brain into the circulation. *Proc Natl Acad Sci U S A*, 93(18):9799–804, 1996.

[228] G. Ma, S. Li, Y. Han, S. Li, T. Yue, B. Wang, and J. Jiang. Regulation of smoothened trafficking and hedgehog signaling by the sumo pathway. *Dev Cell*, 39(4):438–451, 2016.

[229] A. Manglik and B. Kobilka. The role of protein dynamics in gpcr function: insights from the beta2ar and rhodopsin. *Curr Opin Cell Biol*, 27:136–43, 2014.

[230] A. Manglik, B. K. Kobilka, and J. Steyaert. Nanobodies to study g protein-coupled receptor structure and function. *Annu Rev Pharmacol Toxicol*, 57:19–37, 2017.

[231] A. Manglik, H. Lin, D. K. Aryal, J. D. McCorvy, D. Dengler, G. Corder, A. Levit, R. C. Kling, V. Bernat, H. Hubner, X. P. Huang, M. F. Sassano, P. M. Giguere, S. Lober, Duan Da, G. Scherrer, B. K. Kobilka, P. Gmeiner, B. L. Roth, and B. K. Shoichet. Structure-based discovery of opioid analgesics with reduced side effects. *Nature*, 537(7619):185–190, 2016.

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

[233] S. Marada, G. Navarro, A. Truong, D. P. Stewart, A. M. Arensdorf, S. Nachtergaele, E. Angelats, J. T. Opferman, R. Rohatgi, P. J. McCormick, and S. K. Ogden. Functional divergence in the role of n-linked glycosylation in smoothened signaling. *PLoS Genet*, 11(8):e1005473, 2015.

[234] V. Marigo, R. A. Davey, Y. Zuo, J. M. Cunningham, and C. J. Tabin. Biochemical evidence that patched is the hedgehog receptor. *Nature*, 384(6605):176–9, 1996.

[235] E. Marti, D. A. Bumcrot, R. Takada, and A. P. McMahon. Requirement of 19k form of sonic hedgehog for induction of distinct ventral cell types in cns explants. *Nature*, 375(6529):322–5, 1995.

[236] V. Martin, G. Carrillo, C. Torroja, and I. Guerrero. The sterol-sensing domain of patched protein seems to control smoothened activity through patched vesicular trafficking. *Curr Biol*, 11(8):601–7, 2001.

[237] F. R. Maxfield and I. Tabas. Role of cholesterol and lipid organization in disease. *Nature*, 438(7068):612–21, 2005.

[238] C. J. Mayo, J. M. Diprose, T. S. Walter, I. M. Berry, J. Wilson, R. J. Owens, E. Y. Jones, K. Harlos, D. I. Stuart, and R. M. Esnouf. Benefits of automated crystallization plate tracking, imaging, and analysis. *Structure*, 13(2):175–82, 2005.

[239] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. Phaser crystallographic software. *J Appl Crystallogr*, 40(Pt 4):658–674, 2007.

[240] A. R. Meloni, G. B. Fralish, P. Kelly, A. Salahpour, J. K. Chen, R. J. Wechsler-Reya, R. J. Lefkowitz, and M. G. Caron. Smoothened signal transduction is promoted by g protein-coupled receptor kinase 2. *Mol Cell Biol*, 26(20):7550–60, 2006.

[241] L. Milenkovic, M. P. Scott, and R. Rohatgi. Lateral transport of smoothened from the plasma membrane to the membrane of the cilium. *J Cell Biol*, 187(3):365–74, 2009.

[242] L. Milenkovic, L. E. Weiss, J. Yoon, T. L. Roth, Y. S. Su, S. J. Sahl, M. P. Scott, and W. E. Moerner. Single-molecule imaging of hedgehog pathway protein smoothened in primary cilia reveals binding events regulated by patched1. *Proc Natl Acad Sci U S A*, 2015.

[243] P. S. Miller and A. R. Aricescu. Crystal structure of a human gaba receptor. *Nature*, 2014.

[244] J. L. Miller-Gallacher, R. Nehme, T. Warne, P. C. Edwards, G. F. Schertler, A. G. Leslie, and C. G. Tate. The 2.1 a resolution structure of cyanopindolol-bound beta1-adrenoceptor identifies an intramembrane na+ ion that stabilises the ligand-free receptor. *PLoS One*, 9(3):e92727, 2014.

[245] K. Miller-Moslin, S. Peukert, R. K. Jain, M. A. McEwan, R. Karki, L. Llamas, N. Yusuff, F. He, Y. Li, Y. Sun, M. Dai, L. Perez, W. Michael, T. Sheng, H. Lei, R. Zhang, J. Williams, A. Bourret, A. Ramamurthy, J. Yuan, R. Guo, M. Matsumoto, A. Vattay, W. Maniara, A. Amaral, M. Dorsch, and 3rd Kelleher, J. F. 1-amino-4-benzylphthalazines as orally bioavailable smoothened antagonists with antitumor activity. *J Med Chem*, 52(13):3954–68, 2009.

[246] L. V. Misquitta, Y. Misquitta, V. Cherezov, O. Slattery, J. M. Mohan, D. Hart, M. Zhalnina, W. A. Cramer, and M. Caffrey. Membrane protein crystallization in lipidic mesophases with tailored bilayers. *Structure*, 12(12):2113–24, 2004.

[247] J. Mohler. Requirements for hedgehog, a segmental polarity gene, in patterning larval and adult cuticle of drosophila. *Genetics*, 120(4):1061–72, 1988.

[248] J. Mohler and K. Vani. Molecular organization and embryonic expression of the hedgehog gene involved in cell-cell communication in segmental patterning of drosophila. *Development*, 115(4):957–71, 1992.

[249] R. S. Molday and D. MacKenzie. Monoclonal antibodies to rhodopsin: characterization, cross-reactivity, and application as structural probes. *Biochemistry*, 22(3):653–60, 1983.

[250] H. L. Monaco. Crystal structure of chicken riboflavin-binding protein. *EMBO J*, 16(7):1475–83, 1997.

[251] R. Montesano. Inhomogeneous distribution of filipin-sterol complexes in the ciliary membrane of rat tracheal epithelium. *Am J Anat*, 156(1):139–45, 1979.

[252] S. Mukhopadhyay, H. B. Badgandi, S. H. Hwang, B. Somatilaka, I. S. Shimada, and K. Pal. Trafficking to the primary cilium membrane. *Mol Biol Cell*, 28(2):233–239, 2017.

[253] M. Murone, A. Rosenthal, and F. J. de Sauvage. Sonic hedgehog signaling by the patched-smoothened receptor complex. *Curr Biol*, 9(2):76–84, 1999.

[254] S. Muyldermans, T. N. Baral, V. C. Retamozzo, P. De Baetselier, E. De Genst, J. Kinne, H. Leonhardt, S. Magez, V. K. Nguyen, H. Revets, U. Rothbauer, B. Stijlemans, S. Tillib, U. Wernery, L. Wyns, G. Hassanzadeh-Ghassabeh, and D. Saerens. Camelid immunoglobulins and nanobody technology. *Vet Immunol Immunopathol*, 128(1-3):178–83, 2009.

[255] S. Muyldermans and M. Lauwereys. Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. *J Mol Recognit*, 12(2):131–40, 1999.

[256] B. R. Myers, N. Sever, Y. C. Chong, J. Kim, J. D. Belani, S. Rychnovsky, J. F. Bazan, and P. A. Beachy. Hedgehog pathway modulation by multiple lipid binding sites on the smoothened effector of signal response. *Dev Cell*, 26(4):346–57, 2013.

[257] S. Nachtergaele, L. K. Mydock, K. Krishnan, J. Rammohan, P. H. Schlesinger, D. F. Covey, and R. Rohatgi. Oxysterols are allosteric activators of the oncoprotein smoothened. *Nat Chem Biol*, 8(2):211–20, 2012.

[258] S. Nachtergaele, D. M. Whalen, L. K. Mydock, Z. Zhao, T. Malinauskas, K. Krishnan, P. W. Ingham, D. F. Covey, C. Siebold, and R. Rohatgi. Structure and function of the smoothened extracellular domain in vertebrate hedgehog signaling. *Elife*, 2:e01340, 2013.

[259] T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, and A. Miyawaki. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol*, 20(1):87–90, 2002.

[260] Y. Nakano, I. Guerrero, A. Hidalgo, A. Taylor, J. R. Whittle, and P. W. Ingham. A protein with several possible membrane-spanning domains encoded by the drosophila segment polarity gene patched. *Nature*, 341(6242):508–13, 1989.

[261] Y. Nakano, S. Nystedt, A. A. Shivdasani, H. Strutt, C. Thomas, and P. W. Ingham. Functional domains and sub-cellular distribution of the hedgehog transducing protein smoothened in drosophila. *Mech Dev*, 121(6):507–18, 2004.

[262] D. Nedelcu. *Smoothened regulation in the Hedgehog signaling pathway.* Thesis, 2013.

[263] D. Nedelcu, J. Liu, Y. Xu, C. Jao, and A. Salic. Oxysterol binding to the extracellular domain of smoothened in hedgehog signaling. *Nat Chem Biol*, 9(9):557–64, 2013.

[264] R. Nehme, B. Carpenter, A. Singhal, A. Strege, P. C. Edwards, C. F. White, H. Du, R. Grisshammer, and C. G. Tate. Mini-g proteins: Novel tools for studying gpcrs in their active conformation. *PLoS One*, 12(4):e0175642, 2017.

[265] J. A. Newman, P. Savitsky, C. K. Allerston, A. H. Bizard, O. Ozer, K. Sarlos, Y. Liu, E. Pardon, J. Steyaert, I. D. Hickson, and O. Gileadi. Crystal structure of the bloom's syndrome helicase indicates a role for the hrdc domain in conformational changes. *Nucleic Acids Res*, 43(10):5221–35, 2015.

[266] P. Niewiadomski, J. H. Kong, R. Ahrends, Y. Ma, E. W. Humke, S. Khan, M. N. Teruel, B. G. Novitch, and R. Rohatgi. Gli protein activity is controlled by multisite phosphorylation in vertebrate hedgehog signaling. *Cell Rep*, 6(1):168–81, 2014.

[267] C. Nusslein-Volhard and E. Wieschaus. Mutations affecting segment number and polarity in drosophila. *Nature*, 287(5785):795–801, 1980.

[268] S. K. Ogden, D. L. Fei, N. S. Schilling, Y. F. Ahmed, J. Hwa, and D. J. Robbins. G protein galphai functions immediately downstream of smoothened in hedgehog signalling. *Nature*, 456(7224):967–70, 2008.

[269] D. D. Oprian, R. S. Molday, R. J. Kaufman, and H. G. Khorana. Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proc Natl Acad Sci U S A*, 84(24):8874–8, 1987.

[270] D. Oyen, J. Steyaert, and J. N. Barlow. Inhibition of ligand exchange kinetics via active-site trapping with an antibody fragment. *Biochemistry*, 53(12):1879–81, 2014.

[271] D. Oyen, R. Wechselberger, V. Srinivasan, J. Steyaert, and J. N. Barlow. Mechanistic analysis of allosteric and non-allosteric effects arising from nanobody binding to two epitopes of the dihydrofolate reductase of escherichia coli. *Biochim Biophys Acta*, 1834(10):2147–57, 2013.

[272] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano. Crystal structure of rhodopsin: A g protein-coupled receptor. *Science*, 289(5480):739–45, 2000.

[273] D. Panakova, H. Sprong, E. Marois, C. Thiele, and S. Eaton. Lipoprotein particles are required for hedgehog and wingless signalling. *Nature*, 435(7038):58–65, 2005.

[274] E. Pardon, T. Laeremans, S. Triest, S. G. Rasmussen, A. Wohlkonig, A. Ruf, S. Muyldermans, W. G. Hol, B. K. Kobilka, and J. Steyaert. A general protocol for the generation of nanobodies for structural biology. *Nat Protoc*, 9(3):674–93, 2014.

[275] J. H. Park, P. Scheerer, K. P. Hofmann, H. W. Choe, and O. P. Ernst. Crystal structure of the ligand-free g-protein-coupled receptor opsin. *Nature*, 454(7201):183–7, 2008.

[276] Y. J. Park, T. Budiarto, M. Wu, E. Pardon, J. Steyaert, and W. G. Hol. The structure of the c-terminal domain of the largest editosome interaction protein and its role in promoting rna binding by rna-editing ligase l2. *Nucleic Acids Res*, 40(14):6966–77, 2012.

[277] Y. J. Park, E. Pardon, M. Wu, J. Steyaert, and W. G. Hol. Crystal structure of a heterodimer of editosome interaction proteins in complex with two copies of a cross-reacting nanobody. *Nucleic Acids Res*, 40(4):1828–40, 2012.

[278] G. R. Pathare, I. Nagy, P. Sledz, D. J. Anderson, H. J. Zhou, E. Pardon, J. Steyaert, F. Forster, A. Bracher, and W. Baumeister. Crystal structure of the proteasomal deubiquitylation module rpn8-rpn11. *Proc Natl Acad Sci U S A*, 111(8):2984–9, 2014.

[279] J. Pei and N. V. Grishin. Cysteine-rich domains related to frizzled receptors and hedgehog-interacting proteins. *Protein Sci*, 21(8):1172–84, 2012.

[280] R. B. Pepinsky, C. Zeng, D. Wen, P. Rayhorn, D. P. Baker, K. P. Williams, S. A. Bixler, C. M. Ambrose, E. A. Garber, K. Miatkowski, F. R. Taylor, E. A. Wang, and A. Galdes. Identification of a palmitic acid-modified form of human sonic hedgehog. *J Biol Chem*, 273(22):14037–45, 1998.

[281] C. Peters, A. Wolf, M. Wagner, J. Kuhlmann, and H. Waldmann. The cholesterol membrane anchor of the hedgehog protein confers stable membrane association to lipid-modified proteins. *Proc Natl Acad Sci U S A*, 101(23):8531–6, 2004.

[282] R. Petrova and A. L. Joyner. Roles for hedgehog signaling in adult organ homeostasis and repair. *Development*, 141(18):3445–57, 2014.

[283] K. L. Pierce, R. T. Premont, and R. J. Lefkowitz. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol*, 3(9):639–50, 2002.

[284] T. Pietsch, A. Waha, A. Koch, J. Kraus, S. Albrecht, J. Tonn, N. Sorensen, F. Berthold, B. Henk, N. Schmandt, H. K. Wolf, A. von Deimling, B. Wainwright, G. Chenevix-Trench, O. D. Wiestler, and C. Wicking. Medulloblastomas of the desmoplastic variant carry mutations of the human homologue of drosophila patched. *Cancer Res*, 57(11):2085–8, 1997.

[285] O. V. Plotnikova, E. N. Pugacheva, and E. A. Golemis. Primary cilia and the cell cycle. *Methods Cell Biol*, 94:137–60, 2009.

[286] A. H. Polizio, P. Chinchilla, X. Chen, S. Kim, D. R. Manning, and N. A. Riobo. Heterotrimeric gi proteins link hedgehog signaling to activation of rho small gtpases to promote fibroblast migration. *J Biol Chem*, 286(22):19589–96, 2011.

[287] A. H. Polizio, P. Chinchilla, X. Chen, D. R. Manning, and N. A. Riobo. Sonic hedgehog activates the gtpases rac1 and rhoa in a gli-independent manner through coupling of smoothened to gi proteins. *Sci Signal*, 4(200):pt7, 2011.

[288] F. D. Porter. Human malformation syndromes due to inborn errors of cholesterol synthesis. *Curr Opin Pediatr*, 15(6):607–13, 2003.

[289] J. A. Porter, D. P. von Kessler, S. C. Ekker, K. E. Young, J. J. Lee, K. Moses, and P. A. Beachy. The product of hedgehog autoproteolytic cleavage active in local and long-range signalling. *Nature*, 374(6520):363–6, 1995.

[290] J. A. Porter, K. E. Young, and P. A. Beachy. Cholesterol modification of hedgehog signaling proteins in animal development. *Science*, 274(5285):255–9, 1996.

[291] G. G. Prive. Detergents for the stabilization and crystallization of membrane proteins. *Methods*, 41(4):388–97, 2007.

[292] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer. X-ray solution scattering (saxs) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys*, 40(3):191–285, 2007.

[293] C. Raffel, R. B. Jenkins, L. Frederick, D. Hebrink, B. Alderete, D. W. Fults, and C. D. James. Sporadic medulloblastomas contain ptch mutations. *Cancer Res*, 57(5):842–5, 1997.

[294] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J Mol Biol*, 7:95–9, 1963.

[295] R. P. Rambo and J. A. Tainer. Characterizing flexible and intrinsically unstructured biological macromolecules by sas using the porod-debye law. *Biopolymers*, 95(8):559–71, 2011.

[296] R. Rana, C. E. Carroll, H. J. Lee, J. Bao, S. Marada, C. R. Grace, C. D. Guibao, S. K. Ogden, and J. J. Zheng. Structural insights into the role of the smoothened cysteine-rich domain in hedgehog signalling. *Nat Commun*, 4:2965, 2013.

[297] S. G. Rasmussen, H. J. Choi, J. J. Fung, E. Pardon, P. Casarosa, P. S. Chae, B. T. Devree, D. M. Rosenbaum, F. S. Thian, T. S. Kobilka, A. Schnapp, I. Konetzki, R. K. Sunahara, S. H. Gellman, A. Pautsch, J. Steyaert, W. I. Weis, and B. K. Kobilka. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature*, 469(7329):175–80, 2011.

[298] S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, and B. K. Kobilka. Crystal structure of the beta2 adrenergic receptor-gs protein complex. *Nature*, 477(7366):549–55, 2011.

[299] S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, and B. K. Kobilka. Crystal structure of the beta2 adrenergic receptor-gs protein complex. *Nature*, 477(7366):549–55, 2011.

[300] P. J. Reeves, N. Callewaert, R. Contreras, and H. G. Khorana. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous n-glycosylation by a tetracycline-inducible n-acetylglucosaminyltransferase i-negative hek293s stable mammalian cell line. *Proc Natl Acad Sci U S A*, 99(21):13419–24, 2002.

[301] J. Reifenberger, M. Wolter, R. G. Weber, M. Megahed, T. Ruzicka, P. Lichter, and G. Reifenberger. Missense mutations in smoh in sporadic basal cell carcinomas of the skin and primitive neuroectodermal tumors of the central nervous system. *Cancer Res*, 58(9):1798–803, 1998.

[302] J. Reifenberger, M. Wolter, R. G. Weber, M. Megahed, T. Ruzicka, P. Lichter, and G. Reifenberger. Missense mutations in smoh in sporadic basal cell carcinomas of the skin and primitive neuroectodermal tumors of the central nervous system. *Cancer Res*, 58(9):1798–803, 1998.

[303] R. L. Rich and D. G. Myszka. Higher-throughput, label-free, real-time molecular interaction analysis. *Anal Biochem*, 361(1):1–6, 2007.

[304] R. D. Riddle, R. L. Johnson, E. Laufer, and C. Tabin. Sonic hedgehog mediates the polarizing activity of the zpa. *Cell*, 75(7):1401–16, 1993.

[305] A. M. Ring, A. Manglik, A. C. Kruse, M. D. Enos, W. I. Weis, K. C. Garcia, and B. K. Kobilka. Adrenaline-activated structure of beta2-adrenoceptor stabilized by an engineered nanobody. *Nature*, 502(7472):575–9, 2013.

[306] N. A. Riobo and D. R. Manning. Pathways of signal transduction employed by vertebrate hedgehogs. *Biochem J*, 403(3):369–79, 2007.

[307] N. A. Riobo, B. Saucy, C. Dilizio, and D. R. Manning. Activation of heterotrimeric g proteins by smoothened. *Proc Natl Acad Sci U S A*, 103(33):12607–12, 2006.

[308] A. Rivera-Calzada, R. Fronzes, C. G. Savva, V. Chandran, P. W. Lian, T. Laeremans, E. Pardon, J. Steyaert, H. Remaut, G. Waksman, and E. V. Orlova. Structure of a bacterial type iv secretion core complex at subnanometre resolution. *EMBO J*, 32(8):1195–204, 2013.

[309] X. Robert and P. Gouet. Deciphering key features in protein structures with the new endscript server. *Nucleic Acids Res*, 42(Web Server issue):W320–4, 2014.

[310] X. Robert and P. Gouet. Deciphering key features in protein structures with the new endscript server. *Nucleic Acids Res*, 42(Web Server issue):W320–4, 2014.

[311] K. D. Roberts, L. Bandy, and S. Lieberman. The occurrence and metabolism of 20 alpha-hydroxycholesterol in bovine adrenal preparations. *Biochemistry*, 8(3):1259–70, 1969.

[312] E. Roessler, E. Belloni, K. Gaudenz, P. Jay, P. Berta, S. W. Scherer, L. C. Tsui, and M. Muenke. Mutations in the human sonic hedgehog gene cause holoprosencephaly. *Nat Genet*, 14(3):357–60, 1996.

[313] R. Rohatgi, L. Milenkovic, R. B. Corcoran, and M. P. Scott. Hedgehog signal transduction by smoothened: pharmacologic evidence for a 2-step activation process. *Proc Natl Acad Sci U S A*, 106(9):3196–201, 2009.

[314] R. Rohatgi, L. Milenkovic, and M. P. Scott. Patched1 regulates hedgehog signaling at the primary cilium. *Science*, 317(5836):372–6, 2007.

[315] R. Rohatgi and W. J. Snell. The ciliary membrane. *Curr Opin Cell Biol*, 22(4):541–6, 2010.

[316] D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, X. J. Yao, W. I. Weis, R. C. Stevens, and B. K. Kobilka. Gpcr engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science*, 318(5854):1266–73, 2007.

[317] K. Rostislavleva, N. Soler, Y. Ohashi, L. Zhang, E. Pardon, J. E. Burke, G. R. Masson, C. Johnson, J. Steyaert, N. T. Ktistakis, and R. L. Williams. Structure and flexibility of the endosomal vps34 complex reveals the basis of its function on membranes. *Science*, 350(6257):aac7365, 2015.

[318] M. Ruat, L. Hoch, H. Faure, and D. Rognan. Targeting of smoothened for therapeutic gain. *Trends Pharmacol Sci*, 35(5):237–46, 2014.

[319] L. L. Rubin and F. J. de Sauvage. Targeting the hedgehog pathway in cancer. *Nat Rev Drug Discov*, 5(12):1026–33, 2006.

[320] C. M. Rudin, C. L. Hann, J. Laterra, R. L. Yauch, C. A. Callahan, L. Fu, T. Holcomb, J. Stinson, S. E. Gould, B. Coleman, P. M. LoRusso, D. D. Von Hoff, F. J. de Sauvage, and J. A. Low. Treatment of medulloblastoma with hedgehog pathway inhibitor gdc-0449. *N Engl J Med*, 361(12):1173–8, 2009.

[321] J. J. Ruprecht, T. Mielke, R. Vogel, C. Villa, and G. F. Schertler. Electron crystallography reveals the structure of metarhodopsin i. *EMBO J*, 23(18):3609–20, 2004.

[322] S. Ryckaert, E. Pardon, J. Steyaert, and N. Callewaert. Isolation of antigen-binding camelid heavy chain antibody fragments (nanobodies) from an immune library displayed on the surface of pichia pastoris. *J Biotechnol*, 145(2):93–8, 2010.

[323] D. Saerens, J. Kinne, E. Bosmans, U. Wernery, S. Muyldermans, and K. Conrath. Single domain antibodies derived from dromedary lymph node and peripheral blood lymphocytes sensing conformational variants of prostate-specific antigen. *J Biol Chem*, 279(50):51965–72, 2004.

[324] P. Scheerer, J. H. Park, P. W. Hildebrand, Y. J. Kim, N. Krauss, H. W. Choe, K. P. Hofmann, and O. P. Ernst. Crystal structure of opsin in its g-protein-interacting conformation. *Nature*, 455(7212):497–502, 2008.

[325] LLC Schrodinger. The pymol molecular graphics system, version 1.3r1. `https://www.pymol.org/`, 2010.

[326] H. J. Sharpe and F. J. de Sauvage. Signaling: An oxysterol ligand for smoothened. *Nat Chem Biol*, 8(2):139–40, 2012.

[327] H. J. Sharpe, G. Pau, G. J. Dijkgraaf, N. Basset-Seguin, Z. Modrusan, T. Januario, V. Tsui, A. B. Durham, A. A. Dlugosz, P. M. Haverty, R. Bourgon, J. Y. Tang, K. Y. Sarin, L. Dirix, D. C. Fisher, C. M. Rudin, H. Sofen, M. R. Migden, R. L. Yauch, and F. J. de Sauvage. Genomic analysis of smoothened inhibitor resistance in basal cell carcinoma. *Cancer Cell*, 27(3):327–41, 2015.

[328] H. J. Sharpe, W. Wang, R. N. Hannoush, and F. J. de Sauvage. Regulation of the oncoprotein smoothened by small molecules. *Nat Chem Biol*, 11(4):246–255, 2015.

[329] F. Shen, L. Cheng, A. E. Douglas, N. A. Riobo, and D. R. Manning. Smoothened is a fully competent activator of the heterotrimeric g protein g(i). *Mol Pharmacol*, 83(3):691–7, 2013.

[330] R. Sheng, H. Kim, H. Lee, Y. Xin, Y. Chen, W. Tian, Y. Cui, J. C. Choi, J. Doh, J. K. Han, and W. Cho. Cholesterol selectively activates canonical wnt signalling over non-canonical wnt signalling. *Nat Commun*, 5:4393, 2014.

[331] D. A. Sherrell, A. J. Foster, L. Hudson, B. Nutter, J. O'Hea, S. Nelson, O. Pare-Labrosse, S. Oghbaey, R. J. Miller, and R. L. Owen. A modular and compact portable mini-endstation for high-precision, high-speed fixed target serial crystallography at fel and synchrotron sources. *J Synchrotron Radiat*, 22(6):1372–8, 2015.

[332] D. Shi, X. Lv, Z. Zhang, X. Yang, Z. Zhou, L. Zhang, and Y. Zhao. Smoothened oligomerization/higher order clustering in lipid rafts is essential for high hedgehog activity transduction. *J Biol Chem*, 288(18):12605–14, 2013.

[333] J. K. Sicklick, Y. X. Li, A. Jayaraman, R. Kannangai, Y. Qi, P. Vivekanandan, J. W. Ludlow, K. Owzar, W. Chen, M. S. Torbenson, and A. M. Diehl. Dysregulation of the hedgehog pathway in human hepatocarcinogenesis. *Carcinogenesis*, 27(4):748–57, 2006.

[334] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol*, 7:539, 2011.

[335] B. N. Singh, J. Fu, R. K. Srivastava, and S. Shankar. Hedgehog signaling antagonist gdc-0449 (vismodegib) inhibits pancreatic cancer stem cell characteristics: molecular mechanisms. *PLoS One*, 6(11):e27306, 2011.

[336] S. Sinha and J. K. Chen. Purmorphamine activates the hedgehog pathway by targeting smoothened. *Nat Chem Biol*, 2(1):29–30, 2006.

[337] F. Y. Siu, M. He, C. de Graaf, G. W. Han, D. Yang, Z. Zhang, C. Zhou, Q. Xu, D. Wacker, J. S. Joseph, W. Liu, J. Lau, V. Cherezov, V. Katritch, M. W. Wang, and R. C. Stevens. Structure of the human glucagon class b g-protein-coupled receptor. *Nature*, 499(7459):444–9, 2013.

[338] I. Smirnova, V. Kasho, X. Jiang, E. Pardon, J. Steyaert, and H. R. Kaback. Outward-facing conformers of lacy stabilized by nanobodies. *Proc Natl Acad Sci U S A*, 111(52):18548–53, 2014.

[339] I. Smirnova, V. Kasho, X. Jiang, E. Pardon, J. Steyaert, and H. R. Kaback. Transient conformers of lacy are trapped by nanobodies. *Proc Natl Acad Sci U S A*, 112(45):13839–44, 2015.

[340] J. S. Sohier, C. Laurent, A. Chevigne, E. Pardon, V. Srinivasan, U. Wernery, P. Lassaux, J. Steyaert, and M. Galleni. Allosteric inhibition of vim metallo-beta-lactamases by a camelid nanobody. *Biochem J*, 450(3):477–86, 2013.

[341] R. Sounier, C. Mas, J. Steyaert, T. Laeremans, A. Manglik, W. Huang, B. K. Kobilka, H. Demene, and S. Granier. Propagation of conformational changes during mu-opioid receptor activation. *Nature*, 524(7565):375–8, 2015.

[342] T. Souto-Padron and W. de Souza. Freeze-fracture localization of filipin-cholesterol complexes in the plasma membrane of trypanosoma cruzi. *J Parasitol*, 69(1):129–37, 1983.

[343] S. R. Sprang. G protein mechanisms: insights from structural analysis. *Annu Rev Biochem*, 66:639–78, 1997.

[344] D. P. Staus, R. T. Strachan, A. Manglik, B. Pani, A. W. Kahsai, T. H. Kim, L. M. Wingler, S. Ahn, A. Chatterjee, A. Masoudi, A. C. Kruse, E. Pardon, J. Steyaert, W. I. Weis, R. S. Prosser, B. K. Kobilka, T. Costa, and R. J. Lefkowitz. Allosteric nanobodies reveal the dynamic range and diverse mechanisms of g-protein-coupled receptor activation. *Nature*, 535(7612):448–52, 2016.

[345] D. P. Staus, L. M. Wingler, R. T. Strachan, S. G. Rasmussen, E. Pardon, S. Ahn, J. Steyaert, B. K. Kobilka, and R. J. Lefkowitz. Regulation of beta2-adrenergic receptor function by conformationally selective single-domain intrabodies. *Mol Pharmacol*, 85(3):472–81, 2014.

[346] R. C. Stevens, V. Cherezov, V. Katritch, R. Abagyan, P. Kuhn, H. Rosen, and K. Wuthrich. The gpcr network: a large-scale collaboration to determine human gpcr structure and function. *Nat Rev Drug Discov*, 12(1):25–34, 2013.

[347] J. Steyaert and B. K. Kobilka. Nanobody stabilization of g protein-coupled receptor conformational states. *Curr Opin Struct Biol*, 21(4):567–72, 2011.

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

[349] D. M. Stone, M. Hynes, M. Armanini, T. A. Swanson, Q. Gu, R. L. Johnson, M. P. Scott, D. Pennica, A. Goddard, H. Phillips, M. Noll, J. E. Hooper, F. de Sauvage, and A. Rosenthal. The tumour-suppressor gene patched encodes a candidate receptor for sonic hedgehog. *Nature*, 384(6605):129–34, 1996.

[350] H. Strutt, C. Thomas, Y. Nakano, D. Stark, B. Neave, A. M. Taylor, and P. W. Ingham. Mutations in the sterol-sensing domain of patched suggest a role for vesicular trafficking in smoothened regulation. *Curr Biol*, 11(8):608–13, 2001.

[351] T. Tabata and T. B. Kornberg. Hedgehog is a signaling protein with a key role in patterning drosophila imaginal discs. *Cell*, 76(1):89–102, 1994.

[352] J. Taipale, J. K. Chen, M. K. Cooper, B. Wang, R. K. Mann, L. Milenkovic, M. P. Scott, and P. A. Beachy. Effects of oncogenic mutations in smoothened and patched can be reversed by cyclopamine. *Nature*, 406(6799):1005–9, 2000.

[353] J. Taipale, M. K. Cooper, T. Maiti, and P. A. Beachy. Patched acts catalytically to suppress the activity of smoothened. *Nature*, 418(6900):892–7, 2002.

[354] S. Tashiro, T. Michiue, S. Higashijima, S. Zenno, S. Ishimaru, F. Takahashi, M. Orihara, T. Kojima, and K. Saigo. Structure and expression of hedgehog, a drosophila segment-polarity gene required for cell-cell communication. *Gene*, 124(2):183–9, 1993.

[355] C. G. Tate. Practical considerations of membrane protein instability during purification and crystallisation. *Methods Mol Biol*, 601:187–203, 2010.

[356] A. M. Taylor, Y. Nakano, J. Mohler, and P. W. Ingham. Contrasting distributions of patched and hedgehog proteins in the drosophila embryo. *Mech Dev*, 42(1-2):89–96, 1993.

[357] S. P. Thayer, M. P. di Magliano, P. W. Heiser, C. M. Nielsen, D. J. Roberts, G. Y. Lauwers, Y. P. Qi, S. Gysin, C. Fernandez-del Castillo, V. Yajnik, B. Antoniu, M. McMahon, A. L. Warshaw, and M. Hebrok. Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. *Nature*, 425(6960):851–6, 2003.

[358] P. P. Therond. Release and transportation of hedgehog molecules. *Curr Opin Cell Biol*, 24(2):173–80, 2012.

[359] A. B. Tobin. G-protein-coupled receptor phosphorylation: where, when and by whom. *Br J Pharmacol*, 153 Suppl 1:S167–76, 2008.

[360] C. Tribet, R. Audebert, and J. L. Popot. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc Natl Acad Sci U S A*, 93(26):15047–50, 1996.

[361] T. T. Tseng, K. S. Gratwick, J. Kollman, D. Park, D. H. Nies, A. Goffeau, and Jr. Saier, M. H. The rnd permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J Mol Microbiol Biotechnol*, 1(1):107–25, 1999.

[362] H. Tukachinsky, R. P. Kuzmickas, C. Y. Jao, J. Liu, and A. Salic. Dispatched and scube mediate the efficient secretion of the cholesterol-modified hedgehog ligand. *Cell Rep*, 2(2):308–20, 2012.

[363] H. Tukachinsky, L. V. Lopez, and A. Salic. A mechanism for vertebrate hedgehog signaling: recruitment to cilia and dissociation of sufu-gli protein complexes. *J Cell Biol*, 191(2):415–28, 2010.

[364] I. Van Audenhove and J. Gettemans. Nanobodies as versatile tools to understand, diagnose, visualize and treat cancer. *EBioMedicine*, 8:40–8, 2016.

[365] A. Van den Abbeele, S. De Clercq, A. De Ganck, V. De Corte, B. Van Loo, S. H. Soror, V. Srinivasan, J. Steyaert, J. Vandekerckhove, and J. Gettemans.

A llama-derived gelsolin single-domain antibody blocks gelsolin-g-actin interaction. *Cell Mol Life Sci*, 67(9):1519–35, 2010.

[366] M. van den Heuvel and P. W. Ingham. smoothened encodes a receptor-like serpentine protein required for hedgehog signalling. *Nature*, 382(6591):547–51, 1996.

[367] R. H. van der Linden, B. de Geus, G. J. Frenken, H. Peters, and C. T. Verrips. Improved production and function of llama heavy chain antibody fragments by molecular evolution. *J Biotechnol*, 80(3):261–70, 2000.

[368] S. Vanderhaegen, M. Fislage, K. Domanska, W. Versees, E. Pardon, V. Bellotti, and J. Steyaert. Structure of an early native-like intermediate of beta2-microglobulin amyloidogenesis. *Protein Sci*, 22(10):1349–57, 2013.

[369] M. Varjosalo, S. P. Li, and J. Taipale. Divergence of hedgehog signal transduction mechanism between drosophila and mammals. *Dev Cell*, 10(2):177–86, 2006.

[370] A. J. Venkatakrishnan, X. Deupi, G. Lebon, C. G. Tate, G. F. Schertler, and M. M. Babu. Molecular signatures of g-protein-coupled receptors. *Nature*, 494(7436):185–94, 2013.

[371] T. Vercruysse, E. Boons, T. Venken, E. Vanstreels, A. Voet, J. Steyaert, M. De Maeyer, and D. Daelemans. Mapping the binding interface between an hiv-1 inhibiting intrabody and the viral protein rev. *PLoS One*, 8(4):e60259, 2013.

[372] T. Vercruysse, E. Pardon, E. Vanstreels, J. Steyaert, and D. Daelemans. An intrabody based on a llama single-domain antibody targeting the n-terminal alpha-helical multimerization domain of hiv-1 rev prevents viral production. *J Biol Chem*, 285(28):21768–80, 2010.

[373] H. Villanueva, A. P. Visbal, N. F. Obeid, A. Q. Ta, A. A. Faruki, M. F. Wu, S. G. Hilsenbeck, C. A. Shaw, P. Yu, N. W. Plummer, L. Birnbaumer, and M. T. Lewis. An essential role for galpha(i2) in smoothened-stimulated epithelial cell proliferation in the mammary gland. *Sci Signal*, 8(394):ra92, 2015.

[374] C. Vincke, R. Loris, D. Saerens, S. Martinez-Rodriguez, S. Muyldermans, and K. Conrath. General strategy to humanize a camelid single-domain antibody

and identification of a universal humanized nanobody scaffold. *J Biol Chem*, 284(5):3273–84, 2009.

[375] D. D. Von Hoff, P. M. LoRusso, C. M. Rudin, J. C. Reddy, R. L. Yauch, R. Tibes, G. J. Weiss, M. J. Borad, C. L. Hann, J. R. Brahmer, H. M. Mackey, B. L. Lum, W. C. Darbonne, Jr. Marsters, J. C., F. J. de Sauvage, and J. A. Low. Inhibition of the hedgehog pathway in advanced basal-cell carcinoma. *N Engl J Med*, 361(12):1164–72, 2009.

[376] D. Wacker, C. Wang, V. Katritch, G. W. Han, X. P. Huang, E. Vardy, J. D. McCorvy, Y. Jiang, M. Chu, F. Y. Siu, W. Liu, H. E. Xu, V. Cherezov, B. L. Roth, and R. C. Stevens. Structural features for functional selectivity at serotonin receptors. *Science*, 340(6132):615–9, 2013.

[377] T. S. Walter, J. M. Diprose, C. J. Mayo, C. Siebold, M. G. Pickford, L. Carter, G. C. Sutton, N. S. Berrow, J. Brown, I. M. Berry, G. B. Stewart-Jones, J. M. Grimes, D. K. Stammers, R. M. Esnouf, E. Y. Jones, R. J. Owens, D. I. Stuart, and K. Harlos. A procedure for setting up high-throughput nanolitre crystallization experiments. crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr*, 61(Pt 6):651–7, 2005.

[378] C. Wang, H. Wu, T. Evron, E. Vardy, G. W. Han, X. P. Huang, S. J. Hufeisen, T. J. Mangano, D. J. Urban, V. Katritch, V. Cherezov, M. G. Caron, B. L. Roth, and R. C. Stevens. Structural basis for smoothened receptor modulation and chemoresistance to anticancer drugs. *Nat Commun*, 5:4355, 2014.

[379] C. Wang, H. Wu, V. Katritch, G. W. Han, X. P. Huang, W. Liu, F. Y. Siu, B. L. Roth, V. Cherezov, and R. C. Stevens. Structure of the human smoothened receptor bound to an antitumour agent. *Nature*, 497(7449):338–43, 2013.

[380] J. Wang, J. Lu, M. C. Bond, M. Chen, X. R. Ren, H. K. Lyerly, L. S. Barak, and W. Chen. Identification of select glucocorticoids as smoothened agonists: potential utility for regenerative medicine. *Proc Natl Acad Sci U S A*, 107(20):9323–8, 2010.

[381] J. Wang, Jr. Mook, R. A., J. Lu, D. M. Gooden, A. Ribeiro, A. Guo, L. S. Barak, H. K. Lyerly, and W. Chen. Identification of a novel smoothened antagonist that potently suppresses hedgehog signaling. *Bioorg Med Chem*, 20(22):6751–7, 2012.

[382] Y. Wang, H. Chang, A. Rattner, and J. Nathans. Frizzled receptors in development and disease. *Curr Top Dev Biol*, 117:113–39, 2016.

[383] Y. Wang, L. Davidow, A. C. Arvanites, J. Blanchard, K. Lam, K. Xu, V. Oza, J. W. Yoo, J. M. Ng, T. Curran, L. L. Rubin, and A. P. McMahon. Glucocorticoid compounds modify smoothened localization and hedgehog pathway activity. *Chem Biol*, 19(8):972–82, 2012.

[384] Y. Wang, A. P. McMahon, and B. L. Allen. Shifting paradigms in hedgehog signaling. *Curr Opin Cell Biol*, 19(2):159–65, 2007.

[385] Y. Wang, Z. Zhou, C. T. Walsh, and A. P. McMahon. Selective translocation of intracellular smoothened to the primary cilium in response to hedgehog pathway modulation. *Proc Natl Acad Sci U S A*, 106(8):2623–8, 2009.

[386] A. B. Ward, P. Szewczyk, V. Grimard, C. W. Lee, L. Martinez, R. Doshi, A. Caya, M. Villaluz, E. Pardon, C. Cregger, D. J. Swartz, P. G. Falson, I. L. Urbatsch, C. Govaerts, J. Steyaert, and G. Chang. Structures of p-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. *Proc Natl Acad Sci U S A*, 110(33):13386–91, 2013.

[387] D. N. Watkins, D. M. Berman, S. G. Burkholder, B. Wang, P. A. Beachy, and S. B. Baylin. Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature*, 422(6929):313–7, 2003.

[388] U. Weierstall, D. James, C. Wang, T. A. White, D. Wang, W. Liu, J. C. Spence, R. Bruce Doak, G. Nelson, P. Fromme, R. Fromme, I. Grotjohann, C. Kupitz, N. A. Zatsepin, H. Liu, S. Basu, D. Wacker, G. W. Han, V. Katritch, S. Boutet, M. Messerschmidt, G. J. Williams, J. E. Koglin, M. Marvin Seibert, M. Klinker, C. Gati, R. L. Shoeman, A. Barty, H. N. Chapman, R. A. Kirian, K. R. Beyerlein, R. C. Stevens, D. Li, S. T. Shah, N. Howe, M. Caffrey, and V. Cherezov. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat Commun*, 5:3309, 2014.

[389] X. Wen, C. K. Lai, M. Evangelista, J. A. Hongo, F. J. de Sauvage, and S. J. Scales. Kinetics of hedgehog-dependent full-length gli3 accumulation in primary cilia and subsequent degradation. *Mol Cell Biol*, 30(8):1910–22, 2010.

[390] M. Wheatley, D. Wootten, M. T. Conner, J. Simms, R. Kendrick, R. T. Logan, D. R. Poyner, and J. Barwell. Lifting the lid on gpcrs: the role of extracellular loops. *Br J Pharmacol*, 165(6):1688–1703, 2012.

[391] A. M. Wilbanks, G. B. Fralish, M. L. Kirby, L. S. Barak, Y. X. Li, and M. G. Caron. Beta-arrestin 2 regulates zebrafish development through the hedgehog signaling pathway. *Science*, 306(5705):2264–7, 2004.

[392] A. Wilson. Largest likely values for the reliability index. *Acta Crystallographica*, 3(5):397–398, 1950.

[393] C. W. Wilson, M. H. Chen, and P. T. Chuang. Smoothened adopts multiple active and inactive conformations capable of trafficking to the primary cilium. *PLoS One*, 4(4):e5182, 2009.

[394] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson. Overview of the ccp4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*, 67(Pt 4):235–42, 2011.

[395] G. Winter. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, 43:186–190, 2010.

[396] F. Wu, Y. Zhang, B. Sun, A. P. McMahon, and Y. Wang. Hedgehog signaling: From basic biology to cancer therapy. *Cell Chem Biol*, 24(3):252–280, 2017.

[397] H. Wu, C. Wang, K. J. Gregory, G. W. Han, H. P. Cho, Y. Xia, C. M. Niswender, V. Katritch, J. Meiler, V. Cherezov, P. J. Conn, and R. C. Stevens. Structure of a class c gpcr metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science*, 344(6179):58–64, 2014.

[398] M. Wu, Y. J. Park, E. Pardon, S. Turley, A. Hayhurst, J. Deng, J. Steyaert, and W. G. Hol. Structures of a key interaction protein from the trypanosoma brucei editosome in complex with single domain antibodies. *J Struct Biol*, 174(1):124–36, 2011.

[399] V. M. Wu, S. C. Chen, M. R. Arkin, and J. F. Reiter. Small molecule inhibitors of smoothened ciliary localization and ciliogenesis. *Proc Natl Acad Sci U S A*, 109(34):13644–9, 2012.

[400] X. Wu, J. Walker, J. Zhang, S. Ding, and P. G. Schultz. Purmorphamine induces osteogenesis by activation of the hedgehog signaling pathway. *Chem Biol*, 11(9):1229–38, 2004.

[401] X. Xiao, J. J. Tang, C. Peng, Y. Wang, L. Fu, Z. P. Qiu, Y. Xiong, L. F. Yang, H. W. Cui, X. L. He, L. Yin, W. Qi, C. C. Wong, Y. Zhao, B. L. Li, W. W. Qiu, and B. L. Song. Cholesterol modification of smoothened is required for hedgehog signaling. *Mol Cell*, 66(1):154–162 e10, 2017.

[402] C. Xie, S. D. Turley, and J. M. Dietschy. Cholesterol accumulation in tissues of the niemann-pick type c mouse is determined by the rate of lipoprotein-cholesterol uptake through the coated-pit pathway in each organ. *Proc Natl Acad Sci U S A*, 96(21):11992–7, 1999.

[403] J. Xie, M. Murone, S. M. Luoh, A. Ryan, Q. Gu, C. Zhang, J. M. Bonifas, C. W. Lam, M. Hynes, A. Goddard, A. Rosenthal, Jr. Epstein, E. H., and F. J. de Sauvage. Activating smoothened mutations in sporadic basal-cell carcinoma. *Nature*, 391(6662):90–2, 1998.

[404] F. Xu, H. Wu, V. Katritch, G. W. Han, K. A. Jacobson, Z. G. Gao, V. Cherezov, and R. C. Stevens. Structure of an agonist-bound human a2a adenosine receptor. *Science*, 332(6027):322–7, 2011.

[405] S. Xu, B. Benoff, H. L. Liou, P. Lobel, and A. M. Stock. Structural basis of sterol binding by npc2, a lysosomal protein deficient in niemann-pick type c2 disease. *J Biol Chem*, 282(32):23525–31, 2007.

[406] P. G. Yancey, W. V. Rodrigueza, E. P. Kilsdonk, G. W. Stoudt, W. J. Johnson, M. C. Phillips, and G. H. Rothblat. Cellular cholesterol efflux mediated by cyclodextrins. demonstration of kinetic pools and mechanism of efflux. *J Biol Chem*, 271(27):16026–34, 1996.

[407] R. L. Yauch, G. J. Dijkgraaf, B. Alicke, T. Januario, C. P. Ahn, T. Holcomb, K. Pujara, J. Stinson, C. A. Callahan, T. Tang, J. F. Bazan, Z. Kan, S. Seshagiri, C. L. Hann, S. E. Gould, J. A. Low, C. M. Rudin, and F. J. de Sauvage. Smoothened mutation confers resistance to a hedgehog pathway inhibitor in medulloblastoma. *Science*, 326(5952):572–4, 2009.

[408] A. Yavari, R. Nagaraj, E. Owusu-Ansah, A. Folick, K. Ngo, T. Hillman, G. Call, R. Rohatgi, M. P. Scott, and U. Banerjee. Role of lipid metabolism in smoothened derepression in hedgehog signaling. *Dev Cell*, 19(1):54–65, 2010.

[409] D. A. Zacharias, J. D. Violin, A. C. Newton, and R. Y. Tsien. Partitioning of lipid-modified monomeric gfps into membrane microdomains of live cells. *Science*, 296(5569):913–6, 2002.

[410] X. Zhan, J. Wang, Y. Liu, Y. Peng, and W. Tan. Gpcr-like signaling mediated by smoothened contributes to acquired chemoresistance through activating gli. *Mol Cancer*, 13:4, 2014.

[411] X. Zhang, F. Zhao, Y. Wu, J. Yang, G. W. Han, S. Zhao, A. Ishchenko, L. Ye, X. Lin, K. Ding, V. Dharmarajan, P. R. Griffin, C. Gati, G. Nelson, M. S. Hunter, M. A. Hanson, V. Cherezov, R. C. Stevens, W. Tan, H. Tao, and F. Xu. Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand. *Nat Commun*, 8:15383, 2017.

[412] Y. Zhang, B. Sun, D. Feng, H. Hu, M. Chu, Q. Qu, J. T. Tarrasch, S. Li, T. Sun Kobilka, B. K. Kobilka, and G. Skiniotis. Cryo-em structure of the activated glp-1 receptor in complex with a g protein. *Nature*, 546(7657):248–253, 2017.

[413] Y. Zhao, C. Tong, and J. Jiang. Hedgehog regulates smoothened activity by inducing a conformational switch. *Nature*, 450(7167):252–8, 2007.

[414] R. Zidovetzki and I. Levitan. Use of cyclodextrins to manipulate plasma membrane cholesterol content: evidence, misconceptions and control strategies. *Biochim Biophys Acta*, 1768(6):1311–24, 2007.

[415] M. Zocher, C. Zhang, S. G. Rasmussen, B. K. Kobilka, and D. J. Muller. Cholesterol increases kinetic, energetic, and mechanical stability of the human beta2-adrenergic receptor. *Proc Natl Acad Sci U S A*, 109(50):E3463–72, 2012.