



Published in final edited form as:

Nature. 2016 July 28; 535(7613): 517–522. doi:10.1038/nature18934.

Structural basis for Smoothened regulation by its extracellular domains

Eamon F.X. Byrne^{#1}, Ria Sircar^{#2}, Paul S. Miller¹, George Hedger³, Giovanni Luchetti², Sigrid Nachtergaele², Mark D. Tully⁴, Laurel Mydock-McGrane⁵, Douglas F. Covey⁵, Robert P. Rambo⁴, Mark S. P. Sansom³, Simon Newstead^{#3}, Rajat Rohatgi^{2,#}, and Christian Siebold^{1,#}

¹Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

²Departments of Biochemistry and Medicine, Stanford University School of Medicine, Stanford, California, United States of America

³Department of Biochemistry, University of Oxford, Oxford, UK

⁴Diamond Light Source Ltd, Harwell Science & Innovation Campus, Didcot, UK

⁵Department of Developmental Biology, Washington University School of Medicine, St. Louis, Missouri, United States of America

These authors contributed equally to this work.

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. We present crystal structures of the Hh signal transducer and oncoprotein Smoothened (SMO), which contains two distinct ligand-binding sites in its TMD and CRD. The CRD is stacked atop the TMD, separated by an intervening wedge-like linker domain (LD). Structure-guided mutations show that the interface between the CRD, LD and TMD stabilises the inactive state of SMO. Unexpectedly, we find a cholesterol molecule bound to SMO in the CRD-binding site. Mutations predicted to prevent cholesterol binding impair the ability of SMO to transmit native Hh signals. Binding of a clinically used antagonist, vismodegib, to the TMD induces a conformational change

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence to R.R. (rrohagti@stanford.edu) or C.S. (christian@strubi.strubi.ox.ac.uk).

Author Contributions

E.F.X.B. produced the protein with P.S.M. and carried out crystallization with Simon N. C.S. and E.F.X.B. determined the crystal structures. R.S., Sigrid N., and G.L. performed Hh signalling and biochemical assays. L.M.-M. and D.F.C. synthesized sterol analogs. G.H. and M.S.P.S performed MD analysis. M.D.T and R.P.R. carried out SAXS analysis. C.S. and R.R. supervised the project. E.F.X.B., R.R. and C.S. wrote the paper, with input from all authors.

Author Information

Atomic coordinates and structure factors for the *apo*-SMO C and *vismo*-SMO C crystal structures have been deposited in the Protein Data Bank (PDB) under accession numbers 5L7D and 5L7I. The authors declare no competing financial interests.

that is propagated to the CRD, resulting in loss of cholesterol from the CRD-LD-TMD interface. Our work elucidates the structural mechanism by which the activity of a GPCR is controlled by ligand-regulated interactions between its extracellular and transmembrane domains.

The SMO extracellular region is composed of a N-terminal CRD followed by a small LD, which then connects to the TMD and a C-terminal intracellular domain (ICD) (Fig. 1a). Small-molecule agonists and antagonists of SMO have defined two separable ligand-binding sites in the TMD and CRD¹. The TMD-site binds the plant-derived inhibitor cyclopamine^{2,3}, the synthetic agonist SAG^{4,5}, and the anti-cancer drug vismodegib⁶ used to treat advanced basal cell cancer (BCC) in the clinic. Side-chain oxysterols such as 20(S)-hydroxycholesterol (20(S)-OHC) represent a distinct class of SMO ligands⁷⁻⁹ that activate signalling by engaging a hydrophobic groove on the surface of the SMO-CRD¹⁰⁻¹². The native morphogen Sonic Hedgehog (SHH) functions by binding and inactivating Patched 1 (PTCH1), the major receptor for Hh ligands that restrains SMO activity¹³. Despite the discovery of numerous exogenous SMO ligands, a *bona fide* endogenous SMO ligand that regulates Hh signalling remains unknown. Structure-guided mutations that disrupt 20(S)-OHC binding to the CRD groove or sterol-based inhibitors that occlude this groove impair signalling by SHH^{10,11}. In contrast, several mutations in the TMD-site that blocked the binding and activity of synthetic ligands failed to have any effect on either the basal or the SHH-stimulated activity of SMO^{12,14}. These data suggest that an endogenous SMO ligand capable of regulating Hh signalling engages the CRD groove on SMO.

Crystal structures of the isolated SMO-LD-TMD in complex with both agonist and antagonist ligands¹⁵⁻¹⁷ revealed conservation of the GPCR heptahelical scaffold and provided a detailed view of a small molecule binding pocket, but did not show conformational changes typically associated with GPCR signalling^{18,19}. In addition, two unliganded structures of the isolated SMO-CRD have been solved^{10,20}. However, we presently lack structural insights of how the extracellular domains and TMD interact to regulate signalling in SMO (or in any other GPCR).

Structure of the extracellular and transmembrane domains of SMO

We determined the crystal structure of human SMO containing both the CRD and the TMD, connected by the juxta-membrane LD (SMO_C, Fig. 1a and Extended Data Fig. 1). To study the SMO TMD in a defined functional state and reduce conformational flexibility, we included a single amino acid mutation, Val329Phe¹⁶, in TMD helix 3 that locked SMO in an inactive state and substantially improved expression levels (Extended Data Fig. 2 and supplementary discussion). Using an established strategy in GPCR crystallography, the third intracellular loop (ICL3) between TM helices 5 and 6 was replaced by thermostabilised apocytochrome b562RIL (BRIL)²¹. The SMO_C structure was determined to 3.2 Å resolution (Extended Data Table 1, Extended Data Fig. 3). The asymmetric unit, comprising two molecules arranged ‘head-to-tail’, stacks into alternating hydrophobic and hydrophilic layers along one axis typical for lipidic cubic phase (LCP) derived crystals (Extended Data Fig. 3a). This SMO arrangement within the crystal suggests that SMO_C is monomeric, in

agreement with size exclusion chromatography (SEC) coupled to multi-angle light scattering analysis (MALS) (Extended Data Fig.3f).

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. 1a). The overall architecture is stabilised by nine disulphide bridges, four of which reveal the canonical disulphide pattern of the CRD fold²² (numbered 2-5 in Fig. 1a) and one that is specific for SMO¹⁰ (marked as 1' in Fig. 1a). This disulphide bridge positions the start of a 'connector' segment that links the CRD and LD (Fig.1b). The connector is tucked along a hydrophobic groove that runs the length of the CRD and shields this groove with three inwardly turned hydrophobic residues (Val182, Ile185, Phe187), multiple hydrophilic residues and a N-linked glycan facing the solvent (Fig. 1b, Extended Data Fig. 4a). The connector region, combined with mainly hydrophobic interactions between the CRD and both the LD and ECL3 (total buried surface area of 745 Å²), orients the CRD in an upright conformation with its N-terminus pointing away from the plasma membrane (Figs. 1a and 1b, Extended Data Fig. 4b). However, we did not observe major structural changes in the heptahelical TMD bundle compared to previously solved structures of either antagonist- or agonist-bound complexes lacking the CRD^{15,16}. The only exception was a rearrangement of the LD, which in our structure is pushed down towards the TMD, perhaps by CRD-binding (Extended Data Fig. 5). A structure of the activated state of the SMO TMD will likely require co-crystallization with its still-unknown downstream effector or the use of an active-state stabilising antibody.

Cholesterol is a ligand for SMO

Unexpectedly, we discovered a cholesterol molecule in our SMO structure (Fig. 1 and Extended Data Methods). Cholesterol occupies a central position, interacting with all three SMO domains, the CRD, LD and TMD (Figs. 1c, 2a and 2b), and 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)-OHC¹⁰⁻¹² in SMO and the palmitoleyl group of Wnt ligands in Frizzled receptors²³. 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. 2c) and implies that a cholesterol molecule would have to completely desorb from the plasma membrane surface to access its binding site in SMO. Cellular cholesterol levels play a permissive role in Hh signalling^{24,25}; however, this requirement is likely unrelated to the cholesterol binding site seen in our structure since SMO mutants lacking the CRD¹² or carrying mutations in the cholesterol binding groove²⁵ remain sensitive to cholesterol depletion (see supplementary discussion).

The sterol-binding site of SMO is predominantly lined by hydrophobic residues from the CRD that stabilise the flat α -face of cholesterol (Fig. 2a). 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 signalling¹⁰⁻¹². The β -face is shielded by the side-chain of Arg161 and a N-linked glycan (Fig. 2b). The cholesterol

3 β -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 (Figs. 2a and 2b). 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 but are less conserved in *Drosophila* (Fig.2d, Extended Data Fig.1). The superposition of our human CRD structure with a previously solved solution structure of the *Drosophila* CRD²⁰ (r.m.s.d. 1.71 Å for 86 equivalent C α positions) revealed a major rearrangement of one edge of the cholesterol binding site comprising fly SMO residues 183-190 (Fig. 2e, Extended Data Fig. 1). This segment of fly SMO adopts a short helix that protrudes into the CRD groove and consequently may preclude cholesterol binding. In fact, fly SMO fails to bind oxysterols^{10,11}.

Consistent with the structure, purified SMO C bound to cholesterol in a ligand-affinity assay (Figs. 3a-3c), analogous to the assay we previously developed to measure the binding of SMO to 20(S)-OHC⁷. Beads covalently coupled to cholesterol captured purified SMO C (Fig. 3a). Binding could be blocked by free 20(S)-OHC added as a competitor at concentrations that activate Hh signalling in cells⁷ (Fig. 3b), confirming that cholesterol and 20(S)-OHC engaged the same binding groove on the CRD surface. In a stringent specificity control, 20(R)-OHC, an epimer with inverted stereochemistry at a single position that cannot bind to the CRD⁷ and cannot activate Hh signalling¹⁰, failed to block this interaction (Fig. 3c).

Cholesterol promotes SMO signalling

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 cholesterol 3 β -hydroxyl group and Asp95 and Tyr130 (Fig. 2a). The corresponding residues in mouse SMO, Asp99 and Tyr134, were mutated either individually or in combination to residues (Asp99Ala and Tyr134Phe) that lack hydrogen bond acceptor or donor groups. Since cholesterol and oxysterols occupy the same groove in the CRD (Fig. 3b), they likely adopt a similar conformation and make at least some of the same interactions. Therefore, we used 20(S)-OHC binding and signalling to evaluate the effect of these mutations. First, we tested the ability of 20(S)-OHC beads to capture SMO from detergent extracts (Figs. 3d and 3e). As previously demonstrated⁷, 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. 3e). Next, we stably expressed untagged versions of these SMO mutants in *Smo*^{-/-} fibroblasts (Fig. 3d) and assessed their abilities to restore signalling initiated by SHH, the TMD agonist SAG or the CRD agonist 20(S)-OHC (Fig. 3f). SMO-Asp99Ala and SMO-Tyr134Phe 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. 3f). This demonstrates that these mutants retained an intact TMD ligand-binding site and remained competent to transmit signals to cytoplasmic components. However, Asp99Ala and Tyr134Phe substantially impaired the ability of SMO to respond to both

20(S)-OHC and SHH (Fig.3f). We conclude that these residues, and by implication the cholesterol seen in our structure, are important for signal-induced SMO activation.

Inactive-state stabilisation by the SMO extracellular domain

In order to assess the structural impact of the CRD in a membrane environment, we carried out molecular dynamics simulations of SMO in the presence and absence of cholesterol. Ten independent 100 ns atomistic simulations were performed with SMO embedded in a phosphatidylcholine bilayer (Figs. 4a and 4b, Extended Data Fig. 6). These simulations revealed that the SMO CRD has significant conformational flexibility in the un-liganded state. In the presence of cholesterol, however, there was a pronounced decrease in this flexibility (Fig. 4a), consistent with a conformationally stabilising effect on the CRD structure. In contrast, cholesterol did not significantly change the conformational stability of the SMO TMD (Fig. 4b). The predominant cholesterol-induced stabilisation was seen in the vicinity of the sterol-binding pocket, with some propagation to more distal CRD regions (Extended Data Fig. 6). Consistent with these simulations, the thermostability of purified SMO C was reduced when cholesterol was depleted with methyl- β -cyclodextrin (Extended Data Figs. 6f and 6g, supplementary discussion).

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, we generated two mutants, Pro120Ser and Ile160Asn/Glu162Thr, in the context of mouse SMO. These introduce bulky, N-linked glycosylation sites at positions 156 and 114 of the human CRD that contact the LD (Fig. 4c). To disrupt the conformation of the LD itself, we mutated two cysteines (Cys197 and Cys217 in mouse SMO) that form a disulfide bond within the LD (numbered “6” in Figs. 1a and 4c) to serine. We compared these mutants to SMO lacking the entire CRD (SMO- CRD) (Fig. 4d, Extended Data Fig. 7a). Untagged SMO- CRD stably expressed in *Smo*^{-/-} fibroblasts demonstrated a level of constitutive activity^{11,12} comparable to WT SMO stimulated with saturating SHH (Fig. 4d, Extended Data Fig. 7b).

The Pro120Ser, Ile160Asn/Glu162Thr, and Cys197Ser/Cys217Ser mutations mimicked the effect of a complete CRD deletion - constitutive signalling activity was increased (Fig. 4d) despite the presence of high levels of PTCH1 (Extended Data Fig. 7a) and there was complete loss of responsiveness to 20(S)-OHC (Fig. 4d). All three mutants remained sensitive to inhibition by SANT-1, which binds deep in the TMD and does not contact the LD¹⁶. Cyclopamine and SAG, which both make contacts with the LD^{15,16}, 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^{3.50Y} motif in helix III of some Class A GPCRs²⁶.

To understand the basis for the loss in oxysterol responsiveness, we measured binding of each of the mutants to 20(S)-OHC beads (Extended Data Fig. 7c). 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. 4d) even though it can bind normally (Extended Data Fig. 7c) suggests that the LD may transmit the conformational changes that lead to SMO-TMD activation in response to CRD ligands. To probe conformational changes in SMO induced by CRD ligands in the solution state, we performed small angle X-ray scattering (SAXS) experiments on SMO C in the absence of any exogenous ligands (hereafter called *apo*-SMO C) or SMO C loaded with the agonist 20(S)-OHC (Extended Data Fig. 8a). The initial SAXS curves (Fig. 4e) and further analyses^{27,28} (Fig. 4f, Extended Data Fig. 8b) show that 20(S)-OHC binding induces a conformational change consistent with elongation and reduced globularity of SMO C. Since 20(S)-OHC and cholesterol bind to the same groove on the CRD (Fig. 3b), we conclude that replacement of cholesterol by 20(S)-OHC (a molecule that carries only a single additional hydroxyl) produces a conformational change that leads to SMO activation.

The structure of SMO bound to vismodegib

Our structural and functional studies highlighted the critical regulatory role played by CRD-LD-TMD contacts in controlling the conformation and activity of SMO. Are these interactions altered by TMD-targeted small molecules used to treat patients with Hh-driven cancers? This question has remained unanswered because prior structures of SMO bound to small molecules have not included the CRD, precluding an assessment of how the extracellular and transmembrane domains of SMO communicate. We determined the crystal structure of SMO in complex with vismodegib, a potent TMD antagonist, to 3.3 Å resolution (*vismo*-SMO C, Fig. 5a and Extended Data Fig. 9a and 9b). Despite the fact that vismodegib is the most commonly used Hh pathway inhibitor in patients and resistance has already become a clinically relevant problem^{14,29,30}, there are no structures available for vismodegib in complex with SMO.

Vismodegib is stabilised by a network of hydrogen bonds and hydrophobic interactions formed by the SMO TMD core (Fig. 5b). As expected, it occupies the TMD binding-site previously identified in SMO complexes with other antagonists and agonists¹⁵⁻¹⁷ (Extended Data Fig. 5). The vismodegib pyrimidine and chloro-benzyl rings are deeply buried, forming mainly hydrophobic interactions with the TMD core. The amide linker, stabilised by a potential hydrogen bond to the side chain of Asp384, occupies a central position within the 7TM bundle and connects to the chlorophenyl-methylsulfone moiety, which is oriented towards the extracellular domains and the entrance of the TMD pocket¹⁶. This arrangement is stabilised via potential hydrogen bonds to the side chains of Gln477 and Arg400 and a hydrophobic stacking interaction of the vismodegib methylsulfone moiety and the aromatic ring of Phe484. Asp473, a residue mutated in vismodegib-resistant cancers^{14,29,30} stabilises the potential hydrogen-bonding network around Arg400 (marked with an asterisk in Fig. 5b). The SMO-Asp473His mutation seen in drug-resistant cancers is ideally positioned to disrupt the vismodegib binding site, in agreement with functional experiments showing that SMO-Asp473His cannot bind this drug¹⁴. Our structure also explains the vismodegib resistance mechanism for a whole swath of other mutations in the TMD binding site observed in patients with advanced BCC^{29,30} (Extended Data Figs. 9c-9f).

The vismodegib-bound structure of SMO in its inactive state shows a marked conformational change in the extracellular domains and the ECL3 loop when compared to the *apo*-SMO structure (Fig. 5c). The upper part of TM helix 6, which contacts the CRD and LD, is moved $\sim 15^\circ$ towards the LD and CRD, likely due to an interaction between Phe484 and the methylsulfone moiety of vismodegib (Figs. 5b and 5d). This movement of helix 6 results in a rotational movement of the CRD (and to a more minor extent the LD), allowing ECL3 to intrude into the sterol-binding groove of the CRD. As a consequence, the side chain of CRD residue Arg161 forms a stacking interaction with Trp109 and occupies the space where cholesterol was located in the *apo*-SMO structure (Fig. 5e). This reorganisation is predicted to occlude the cholesterol binding site. Indeed, cholesterol was absent from the *vismo*-SMO structure (Fig.5a), and the addition of vismodegib reduced the interaction of purified SMO C with cholesterol beads (Extended Data Figs. 9g and 9h). In summary, vismodegib binding to the TMD site induces a conformational change that ultimately results in a rearrangement of the sterol-binding site, providing a structural communication mechanism that explains the previously observed allosteric interaction⁷ between CRD and TMD ligands. More generally, it suggests that a similar conformational change, involving a shift of the 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. From a therapeutic perspective, our results highlight an entirely unanticipated role for the CRD and ECL3, including displacement of an extracellular cholesterol ligand, in stabilising a drug-bound inactive state of SMO, providing a structural template for the development of the next generation of SMO inhibitors against Hh-driven cancers.

Conclusion

The structure of SMO provides insights into the mechanism by which a large extracellular region and two allosterically-linked ligand binding sites may regulate the activity of a GPCR. We propose that cholesterol functions as an endogenous SMO ligand that occupies the CRD groove and stabilises a resting or *apo* conformation poised to respond to Hh signals. SMO signalling activity is compromised by mutations that prevent cholesterol binding or by antagonists like vismodegib that act allosterically to occlude the cholesterol binding site. SAXS data suggest that CRD ligands like 20(S)-OHC, which displace cholesterol, produce an additional conformational change that leads to SMO activation. Identification of the mechanism by which PTCH1 inhibits SMO will be necessary to understand how cholesterol - bound SMO is activated in response to endogenous Hh morphogens. We predict that SMO activation will involve alterations in the interactions between the CRD, ECL3 and LD-TMD domains that allow the TMD to adopt an active signalling state.

METHODS - Byrne et al

Reagents

NIH 3T3 and 293T cells were obtained from ATCC, *Smo*^{-/-} fibroblasts have been described previously³³ and were originally obtained from Drs. James Chen and Philip Beachy, SAG

was from Enzo Life Sciences, Cyclopamine from Toronto Research Chemicals, SANT-1 from EMD Millipore, 20(S)-OHC from Steraloids. The synthesis of 20(R)-OHC⁷ and 20(S)-amine coupled beads for SMO binding assays has been described in detail previously¹⁰. Rabbit polyclonal antibodies against SMO, PTCH1 and SUFU have been described previously^{34,35}, and the mouse monoclonal antibody against GLI1 was obtained from Cell Signaling Technologies (Cat#L42B10). SHH-containing conditioned media was made from 293T cells transfected with the N-terminal signalling domain of SHH and used at saturating concentrations (dilution of 1:4)².

Constructs

For large-scale expression and crystallization, a SMO construct (SMO C) was designed by truncating the N- and C-termini of human SMO, to leave the extracellular and transmembrane domains (UniProt Id. Q99835; residues 32-555), and replacing intracellular loop 3 (Q99835; residues 429-440) with BRIL (UniProt Id. P0ABE7; residues 23-128). The synthetic gene encoding SMO C was obtained from Genent (Grand Island, NY) and cloned into the pHLSec vector³⁶ in frame to either a C-terminal Rho1D4 antibody epitope tag^{37,38} or a monoVenus tag^{39,40} followed by a Rho1D4 antibody epitope tag.

All mouse SMO mutants were made using Quikchange methods using a previously described construct⁴¹ encoding mouse SMO (pCS2+:YFP-mSmo), after removal of the N-terminal YFP tag by XhoI digestion. The mouse SMO- CRD construct lacks a.a. 68-184. For stable-line construction, the mouse SMO coding sequence was transferred from pCS2+ to pMSCVpuro using Gibson cloning.

Expression and purification of SMO

SMO C was 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% CO₂, 130 rpm to densities of $2-3 \times 10^6$ cells mL⁻¹ in protein expression media (PEM, Invitrogen) supplemented with L-glutamine, non-essential amino-acids (NEAA, Gibco) and 1% fetal calf serum (FCS, Sigma-Aldrich). Cells from 0.8 L cultures were collected by centrifugation (1,100 rpm, 7 min) and re-suspended in 120 mL Freestyle293 medium (Invitrogen) containing 1.2 mg PEI Max (Polysciences), 0.4 mg plasmid DNA and 5 mM valproic acid (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, snap-frozen in liquid N₂ and stored at -80°C, resulting in a total of ~150 g of cell mass per 9.6 L suspension medium⁴².

Frozen cell pellets were thawed, re-suspended 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 1.3% (w/v) n-dodecyl-β-D-maltopyranoside (DDM, Anatrace) and 0.26% (w/v) cholesteryl hemisuccinate (CHS, Anatrace), then rotated for 1.5 h at 4°C. Insoluble material was removed by centrifugation (10,000 rpm, 12 min, 4°C) and the supernatant incubated for 2 h at 4°C with purified Rho-1D4 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 μ M “TETSQVAPA” peptide (Genscript). Eluate was concentrated to ~500 μ L using a Vivaspin Turbo 4 PES 100 kDa MWCO concentrator and loaded onto a Superose 6 10/300 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 to ~30 mg/mL using a Vivaspin 500 PES 100 kDa MWCO concentrator. Samples were deglycosylated with Endoglycosidase F1 and incubated at room temperature for 1 h. For the vismodegib complex, vismodegib (GDC-0449, Selleck Chem) dissolved at high concentration in DMSO, was added to the protein sample to a final concentration of 10 mM.

For small-scale screening, cells in adherent format were transiently transfected using Lipofectamine2000 (Invitrogen) and the expressed protein, tagged with YFP^{39,40}, was prepared in the same manner as above with quantities adjusted appropriately. For analysis, samples were loaded onto a Superose 6 3.2/300 column (GE Healthcare) attached to a high-performance liquid chromatography system with automated micro-volume loader and in-line fluorescence detection (Shimadzu)⁴³.

For thermostability experiments, SMO C was expressed and purified as described above. After SEC purification, the pooled peak fractions were re-applied to purified Rho-1D4 antibody coupled to CNBr-activated sepharose beads. Equal amounts 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 300 mM NaCl, 50 mM Hepes pH7.5, 0.03% DDM and elution with 300 mM NaCl, 50 mM Hepes, pH7.5, 0.03% DDM, 500 μ M “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 followed using 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.

Stable Cell Lines

Stable cell lines expressing untagged SMO mutants were made by infecting *Smo*^{-/-} fibroblasts with a retrovirus carrying these constructs cloned into pMSCVpuro⁴¹. Retroviral supernatants were produced after transient transfection of Bosc23 helper cells with the pMSCV constructs^{44,45}. Virus-containing media from these transfections was directly used to infect *Smo*^{-/-} fibroblasts, and stable integrants were selected with puromycin (2 μ g/mL). We had previously¹⁰ constructed stable lines using an identical strategy with SMO constructs carrying an N-terminal Fluorescent Protein (FP) tag; however, we found that epitope tagging of SMO with an FP, or transient overexpression of tagged or untagged SMO can impact the assessment of its signalling activity. For example, we previously measured a lower level of constitutive signalling activity for SMO- CRD compared to the present study, likely due to the presence of an N-terminal YFP tag¹⁰. Moreover, prior reports from three groups (including our own) reached somewhat divergent conclusions regarding the role of

the CRD in basal and ligand-stimulated SMO activity¹⁰⁻¹². These differences may have been related to the use of different epitope tags and expression systems.

Hence, all Hh signalling assays used in this study were performed with untagged SMO and SMO mutants stably expressed in *Smo*^{-/-} fibroblasts, with assessment of SMO protein levels in stable cell lines by immunoblotting.

Hedgehog signalling assays

Stable cell lines expressing either WT SMO or SMO mutants were grown to confluence in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% Fetal Bovine Serum (FBS, Optima Grade, Atlanta Biologicals) and then exposed to DMEM containing 0.5% FBS for 24 h to induce primary cilia assembly. These ciliated cells were then treated with saturating concentrations of various Hh agonists and antagonists in DMEM containing 0.5% FBS for 12 h.

For detection of proteins by immunoblotting, cells were washed in ice-cold PBS and lysed (30 min, 4°C) by agitation in modified RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM sodium chloride, 2% NP-40, 0.25% deoxycholate, 0.1% sodium-dodecyl sulfate [SDS], 1 mM dithiothreitol, 1 mM sodium fluoride and the SigmaFast Protease inhibitor cocktail). After clarification (20,000.g, 30 min, 4°C), the protein concentration of each lysate was measured using the bicinchoninic acid assay (BCA, Pierce/Thermo Scientific, Rockford, IL). Lysate aliquots containing equal amounts of total protein were fractionated on SDS-PAGE gels (either a 8% Tris-Glycine gel or a 4-12% Bis-Tris Gel), and transferred to a nitrocellulose membrane. Quantitative immunoblotting with the various antibodies was performed using the Li-Cor Odyssey infrared imaging system. In all immunoblots, vertical dashed black lines represent non-contiguous lanes from the same immunoblot juxtaposed for clarity. Each immunoblot was repeated 2-3 times for all experiments shown in the paper.

Gli1 mRNA is a commonly used metric for Hh signaling activity, since *Gli1* is a direct Hh target gene. *Gli1* and *Gapdh* mRNA levels were measured by quantitative, reverse-transcription PCR (qRT-PCR) using the *Power SYBR Green Cells-To-CT* kit from Thermo Fisher Scientific and custom primers for *Gli1* (forward primer: 5'-ccaagccaactttatgtcaggg-3' and reverse primer: 5'-agcccgcctcttggtaattga-3') and *Gapdh* (forward primer: 5'-agtggcaaatggagatt-3' and reverse primer: 5'-gtggagtcatactggaaca-3'). Transcript levels relative to *Gapdh* were calculated using the C_t method. Each qRT-PCR experiment, which was repeated 2-4 times, included two biological replicates, each with two technical replicates.

Statistical analysis of *Gli1* mRNA levels across samples was performed using an ordinary one-way ANOVA test with a Holm-Sidak post-test to correct for multiple comparisons using the GraphPad Prism suite. Statistical significance in the figures is denoted as follows: ns: $p > 0.05$, *: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Ligand affinity chromatography

Ligand affinity chromatography to assess the interaction between SMO protein in detergent extracts and beads covalently coupled to 20(S)-amine have been described previously¹⁰.

Membranes from cells transiently or stably expressing constructs encoding mouse SMO variants were lysed in a DDM extraction buffer (50 mM Tris pH 7.4, 150 mM NaCl, 10% v/v glycerol, 0.5% w/v DDM and the SigmaFast EDTA-Free protease inhibitor cocktail) for 2 h at 4°C, followed by removal of insoluble material by ultracentrifugation (100,000.g, 30 min). This DDM extract was incubated with 20(S)-OHC beads for overnight at 4°C to allow binding to equilibrium. After extensive washing, proteins captured on the beads were eluted with reducing LDS sample buffer (Life Technologies) and 100 mM dithiothreitol. The presence of SMO in these eluates was determined by quantitative immunoblotting with an anti-SMO antibody³⁴ and infrared imaging (Li-Cor Odyssey).

Preparation of cholesterol-PEG3-sepharose using the azide-alkyne Huisgen cycloaddition reaction from the click chemistry toolbox

100 μ L of packed PEG3-azide sepharose resin (22 μ mole / mL, Click Chemistry Tools) was washed 3 \times 1 mL with 20% ethanol (v/v aq.) and re-suspended in 500 μ L of 20% ethanol (v/v). The bead slurry was supplemented with 1 mM CuSO₄, 5 mM Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), and 15 mM sodium ascorbate. The cycloaddition coupling reaction was initiated by adding 0.044 μ moles (high) of **LKM-26**, a previously-described^{46,47} alkynyl cholesterol derivative synthesized in-house, to achieve an approximate ligand density of 1:50 (moles coupled azide functional groups : moles uncoupled azide functional groups). High, medium and low ligand densities (Fig. 2f) represent coupling ratios of 1:50, 1:200 and 1:1000. The reaction was protected from light and allowed to proceed at room temperature for 20 h with end-over-end rotation. The supernatant was removed from the resin, and the reaction quenched with 1 mM EDTA in 20% ethanol (v/v aq.). The supernatant was extracted with diethyl ether and loaded on a normal-phase TLC plate next to an alkynyl cholesterol standard to assess the efficiency of coupling. The Thin Layer Chromatography (TLC) plate was developed using 2% methanol in chloroform (v/v) and stained using 10% CuSO₄/10% H₃PO₄ followed by charring at 200°C for visualization.

SMO pull-down assays using cholesterol-PEG3-sepharose

Binding reactions were carried out in buffer containing 20 mM HEPES pH 7.4, 150 mM NaCl, and 0.03% DDM in a final volume of 100 μ L. Each reaction contained 1 μ g of purified SMO C, the same protein used for crystallization studies. Competitors were added to the binding reaction and incubated for 1 hr at room temperature prior to the addition of 20 μ L of packed cholesterol-PEG3-sepharose. Reactions were incubated for 16 hr at 4°C for affinity capture. The binding reactions were washed 3 \times 1 mL with binding buffer and eluted using 2X Laemmli buffer for 30 minutes at room temperature. SMO C levels in the input, flow-through and captured on the beads were determined by quantitative immunoblotting (Li-Cor Odyssey) using the 1D4 primary antibody (mouse 1:2000).

Crystallization and data collection

Protein samples were reconstituted into lipidic cubic phase (LCP) by mixing with molten lipid in a mechanical syringe mixer⁴⁸. Molten lipid, consisting of 10% cholesterol (cholesterol, Sigma-Aldrich) and 90% 9.9 monoacylglycerol (monoolein) (Sigma-Aldrich),

was mixed with detergent-solubilised protein (either apo or with 10 mM vismodegib in DMSO) at ~30 mg/mL in a ratio of 60:40. A Gryphon robot (Art Robbins Instruments) was used to dispense 50 nL boluses of protein laden mesophase followed by 0.8 μ L of precipitant solution onto each of 96 positions on a siliconized glass plate and then covered with a coverslip in a “sandwich-plate” format. 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). *apo*-SMO C crystallized in 0.1 M MES pH 6, 30% (v/v) PEG500 DME, 0.1 M sodium acetate, 0.5 mM zinc chloride, 0.1 M ammonium fluoride. *vismo*-SMO C crystallized in 0.09 M sodium acetate pH4, 0.09 M sodium malonate, 27% (v/v) PEG500 DME, 0.1 M sodium acetate, 0.5 mM zinc chloride, 0.1 M ammonium fluoride.

X-ray data collection was conducted at MX beamline I24 at the Diamond Light Source (Harwell, UK). Prior to data collection crystals were flash-frozen at 105 K. X-ray data were processed using Xia2^{49,50}, scaled using XSCALE⁵⁰ and merged using Aimless^{51,52}. The final dataset used for structure solution and refinement was merged from data from 9 crystals for *apo*-SMO C and two crystals for *vismo*-SMO C. Data collection statistics are shown in Extended Data Table 1.

Structure determination, refinement and analysis

The *apo*-SMO C structure was solved by molecular replacement in PHASER⁵³ using the structure of human SMO TMD (PDB 4QIM¹⁶), zebrafish SMO CRD (PDB 4C79¹⁰) and BRIL (PDB 4EIIY⁵⁴) as search models. Extra electron density accounting for the region between CRD and LD, BRIL and TMD were immediately discernible after density modification in PARROT⁵⁵ (Extended Data Fig. 3e). Surprisingly, we also observed very strong extra density within the CRD ligand binding pocket (Fig.2a) 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 derived from the cells (mM concentrations within the cell membrane) or from the LCP crystallization mix (that contained 10% (w/v) cholesterol). The *apo*-SMO C polypeptide chain was traced using iterative rounds of BUCCANEER⁵⁶, manual building in COOT⁵⁷ and refinement in autoBUSTER⁵⁸ and PHENIX⁵⁹. This resulted in a well-defined model for the *apo*-SMO C structure that included two molecules of SMO (residues 59-549) inserted with a BRIL protein segment between SMO TMD helices 5 and 6, two N-linked glycans and a cholesterol molecule. Interestingly, we observed a systematic disorder along the c axis (Extended Data Fig. 2a), 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. The *vismo*-SMO C structure was solved by molecular replacement using the *apo*-SMO C structure. Extra electron density accounting for vismodegib and two well-ordered monoolein molecules was immediately apparent. The structure was refined using autoBUSTER⁵⁸ and PHENIX⁵⁹ using non crystallographic and secondary structure restraints. Crystallographic and Ramachandran statistics are given in Extended Data Table 1. Stereochemical properties were assessed by

MOLPROBITY⁶⁰. Superpositions were calculated using the program COOT⁵⁷ and electrostatic potentials were generated using APBS⁶¹ and hydrophobicity was calculated according to the Eisenberg hydrophobicity scale⁶², as implemented in PyMOL⁶³. Buried surface areas of protein-protein interfaces were calculated using the PISA webserver⁶⁴ for a probe radius of 1.4 Å. Sequence alignment were performed using MULTALIN⁶⁵ and formatted with ESPRIPT⁶⁶. Program Caver was used with default settings to visualize the SMO TMD ligand binding pocket⁶⁷.

Amphipol exchange and MALS

In order to avoid background light-scattering due to free detergent in solution, protein samples were exchanged into amphipol⁶⁸(A8-35, Anatrace) at a mass ratio of 3:1 amphipol:protein and rotated at RT for 30 min. BioBeads (BIORAD), equilibrated in detergent-free 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. For multi-angle light scattering (MALS) experiments, amphipol-solubilised protein at 1 mg/mL was loaded onto a Superose 6 10/300 column (GE Healthcare), equilibrated in detergent-free buffer, on a Shimadzu system with inline MALS detector (Wyatt). Data were analysed using ASTRA6.1.2 software (Wyatt). For SMO C, the value used for dn/dc and ϵ at 280 nm were 0.185 mL/g and 1.541 mL/(mg.cm) respectively. For protein conjugate analysis, dn/dc used for amphipol A8-35 was 0.15 mL/g⁶⁹.

Molecular dynamics system setup

Simulations were performed using the GROMACS v4.6.3 simulation package⁷⁰. Side chain ionization states were modelled using pdb2gmX (Histidine) and PropKa (All other residues)^{71,72}. The N and C-termini were treated with neutral charge. Intracellular loop 3 (occupied by the BRIL fusion in our crystal structure) was modelled using coordinates from the PDB entry 4N4W¹⁶. The protein structure was then energy minimised using the steepest descents algorithm implemented in GROMACS, before being converted to a coarse-grained (CG) representation using the MARTINI 2.2 force field⁷³. The energy minimised CG structure was centred in a simulation box with dimensions 100 × 100 × 180 Å³. 270 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids were randomly placed around the protein and the system solvated and neutralised to a concentration of 0.15 M NaCl. An initial 1 µs of CG simulation was applied to permit the self-assembly⁷⁴ of a POPC lipid bilayer around the GPCR. During the CG simulation the structure of the protein was maintained via an elastic network, allowing local conformational flexibility of the protein. Thus the protein was able to adopt its optimal orientation within the lipid bilayer⁷⁵. The endpoint of the CG bilayer self-assembly simulation was converted back to atomic detail using a fragment-based protocol for the lipid conformations⁷⁶, whilst retaining the original crystal structure of the protein, now located in its optimal orientation and position within the lipid bilayer. Equilibration of the atomic system was achieved *via* 1 ns of NPT simulation with the protein coordinates restrained, before the system was subjected to 100 ns of unrestrained atomistic molecular dynamics. Simulations were performed both in the presence and absence of the cholesterol ligand. Five repeat simulations were run for each case.

Coarse-grained simulations

The standard MARTINI force field⁷⁷ and its extension to proteins^{73,78} was used to describe all system components. During the CG self-assembly simulation an ELNEDYN network⁷⁹ was applied to the protein using force constant of 500 kJ/mol/nm² and a cutoff of 1.5 nm. Temperature was maintained at 310 K using a Berendsen thermostat⁸⁰ with a coupling constant of $\tau_t = 1$ ps, and pressure was controlled at 1 bar using a Berendsen barostat⁸⁰ with a coupling constant of $\tau_p = 1$ ps and a compressibility of 5×10^{-6} bar⁻¹. Electrostatics and van der Waals interactions in the CG simulations were shifted between 0 and 1.2 nm, and 0.9 and 1.2 nm respectively, using the standard MARTINI protocol⁷⁷. An integration time step of 20 femtoseconds was applied. Covalent bonds were constrained to their equilibrium values using the LINCS algorithm⁸¹. All simulations were run in the presence of standard MARTINI water particles⁷⁷, and ions added to an approximate concentration of 0.15 M NaCl.

Atomistic simulations

Atomistic simulations were run using the GROMOS53a6 force field⁸², and its extension to glycans⁸³. The system was solvated using the SPC water model, and ions added to yield an electrically neutral system with a NaCl concentration of approximately 0.15 M. Systems contained approximately 140,000 atoms including 270 POPC molecules, ~41,000 water molecules, 149 sodium ions, and 154 chloride ions. Periodic boundary conditions were applied, with a simulation time step of 2 femtoseconds. Temperature was maintained at 310 K using a V-rescale thermostat⁸⁴ with a coupling constant of 0.1 ps, whilst pressure was controlled at 1 bar through coupling to a Parrinello-Rahman barostat⁸⁵ with a coupling constant of 1 ps. Particle Mesh Ewald (PME)⁸⁶ was applied to model long-range electrostatics. The LINCS algorithm was used to constrain covalent bond lengths⁸¹. The `g_dist`, `g_rmsf` and `g_rms` tools implemented in the GROMACS v4.6.3 software package⁷⁰ were applied to analyse the simulations, with VMD⁸⁷ and PyMOL⁶³ used for visualisation. C α RMSD calculations for the CRD were performed after a least squares fit of the trajectory to C α particles of the initial (crystal structure) CRD coordinates. C α RMSD calculations for the TMD were performed after a least squares fit to C α particles of the initial (crystal structure) transmembrane helix coordinates, excluding the inter-helix loop regions of the TMD from this calculation. DSSP matrices were produced using the `do_dssp` tool implemented in GROMACS⁷⁰.

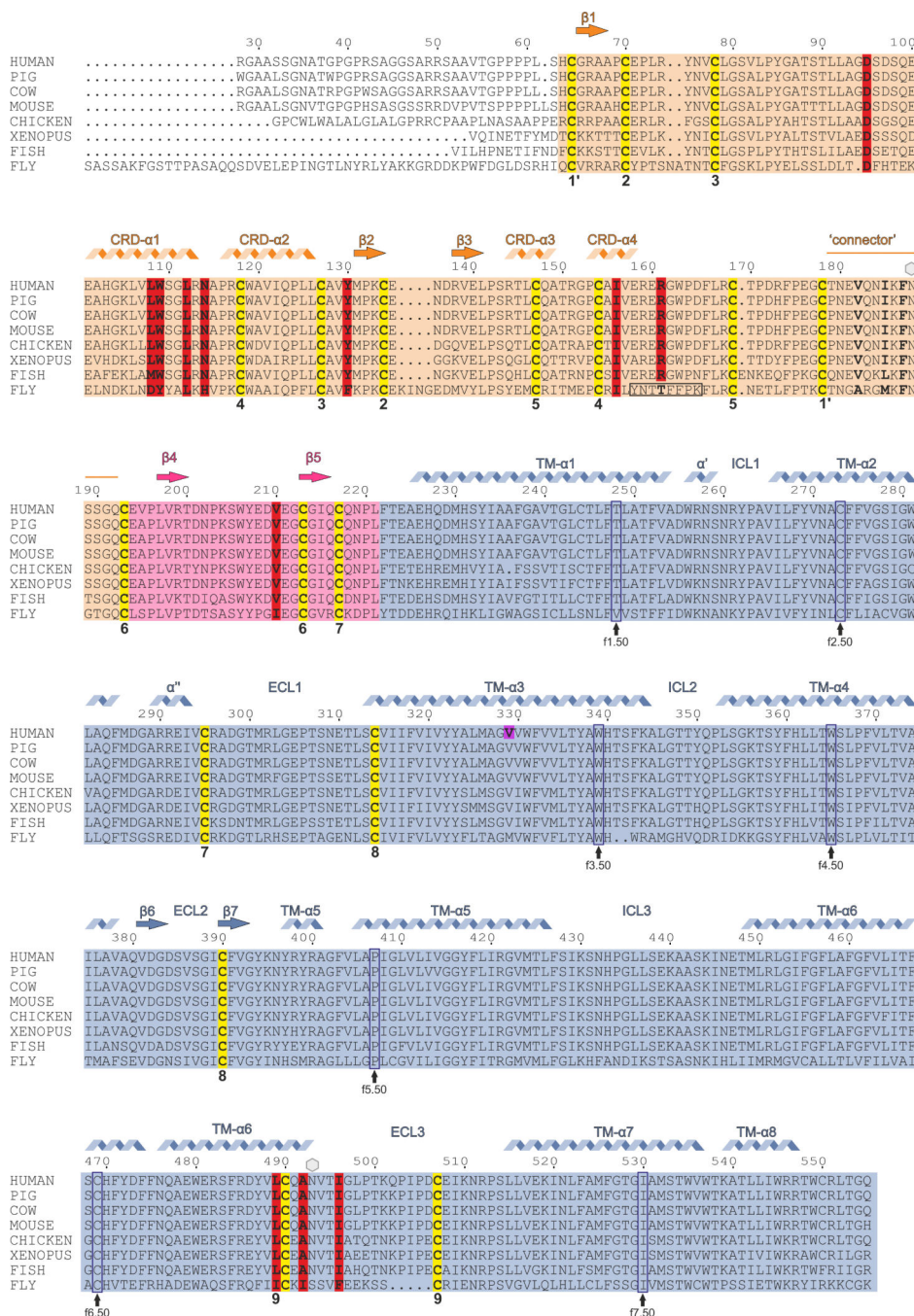
SAXS experiments

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) SMO C 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 μ L of sample was injected at 9.6 (20(S)-OHC-treated) or 13 (apo) mg/mL using a flow rate of 160 μ L per minute. Chromatographic elution was directed into a specialized 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

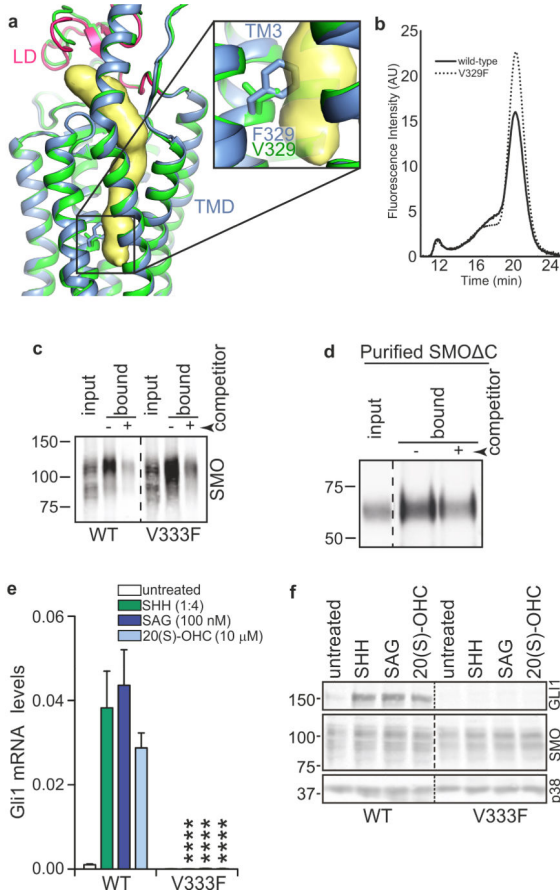
background measurement for producing the background-subtracted curve was collected at greater than 1.5 column volumes.

Images were corrected for variations in beam current, normalised for exposure time and processed into 1D scattering curves using in-house beamline software (GDA). 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 ($q > 0.05 \text{ \AA}^{-1}$) imply major structural differences between the two states. Large differences between the two $P(r)$ -distributions imply a significant structural change in the 20(S)-OHC state.

Extended Data

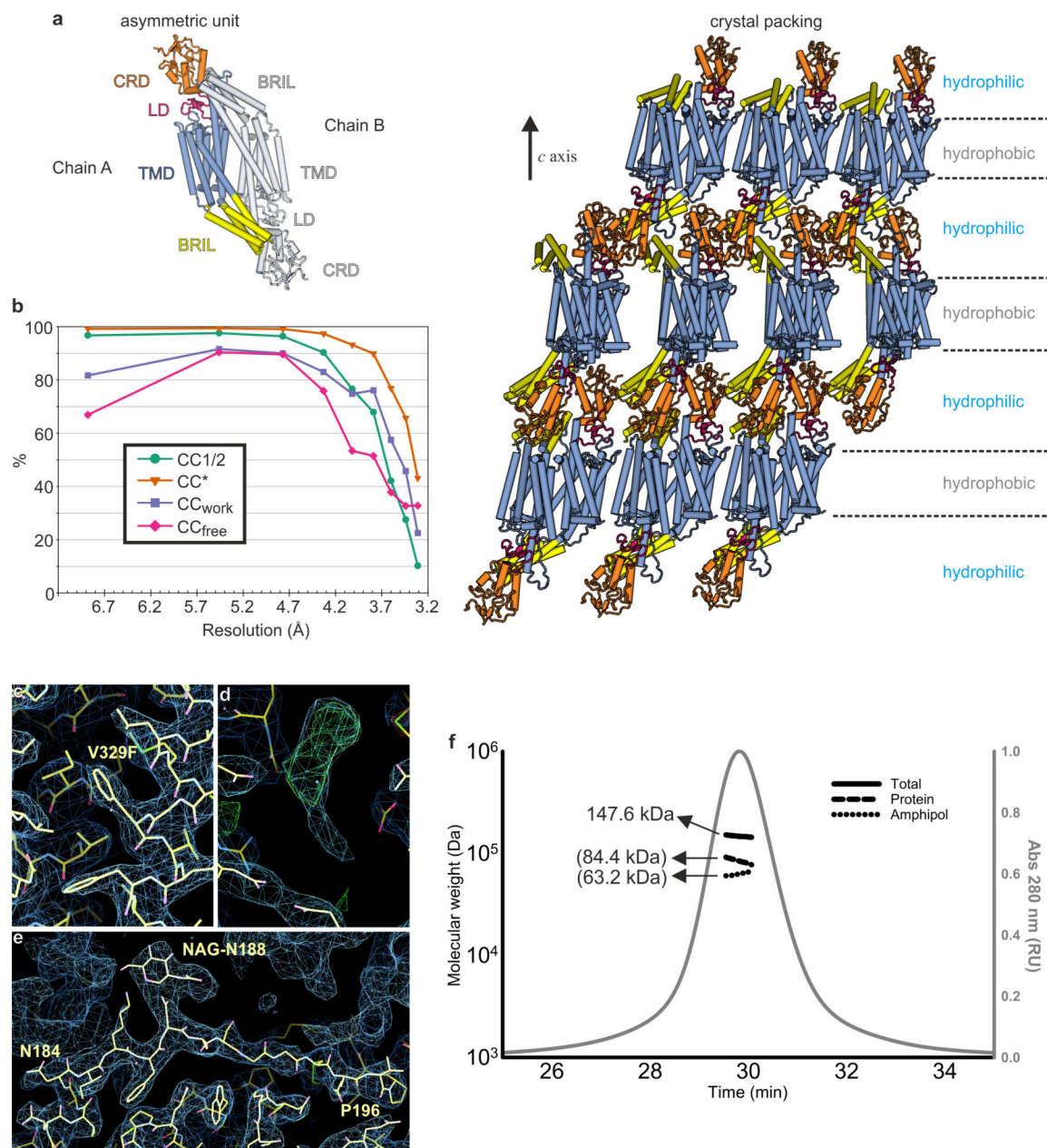


Extended Data Figure 1. Sequence alignment of SMO orthologues
 Numbering corresponds to that of human SMO. Secondary structure assignments are displayed above the alignment and colour-coded as in Fig. 1. Black arrows and numbers (fX.50) below alignment show class F Ballesteros-Weinstein nomenclature for GPCR helices¹⁶. Residues interacting with cholesterol are highlighted in red. Disulphide bridges are highlighted in yellow and numbered. N-linked glycosylation sites are depicted by a hexagon. Position of the Val329Phe point mutation is highlighted in purple.



Extended Data Figure 2. Characterization of the SMO Val329Phe mutation

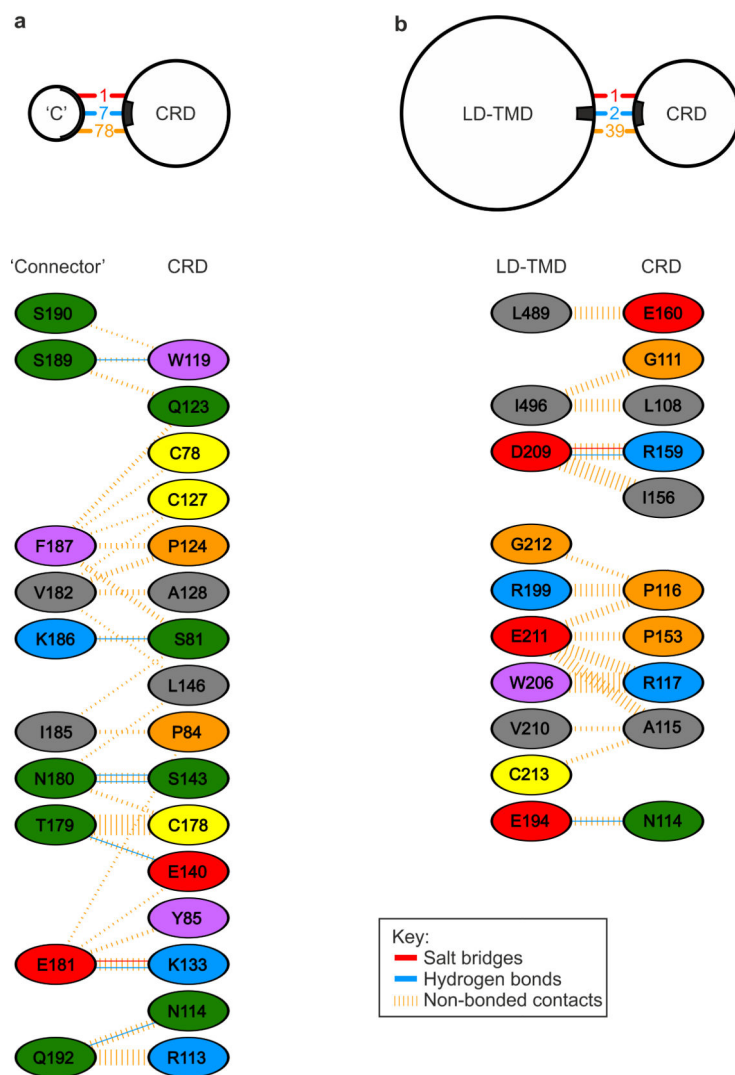
a, Superposition of SMO C structure (blue) with the SMO-SANT-1 complex structure, which lacks the CRD (green, PDB 4N4W¹⁶), showing the TMD ligand-binding pocket as a yellow surface. Inset shows Val329, mutated to Phe in our structure. **b**, SEC analysis of fluorescently-labelled SMO C showing difference in expression levels of wild-type and Val329Phe variant (main protein peak ~20 min). **c**, 20(S)-OHC beads can bind both mouse WT SMO and Val333Phe (mouse Val333 corresponds to human Val329). Immunoblots, using an anti-SMO antibody directed against the ICD, were used to measure SMO captured on 20(S)-OHC beads. Adding 50 μ M free 20(S)-OHC as a competitor reduced binding. **d**, Purified human SMO C (the crystallization construct) binds to 20(S)-OHC beads. **e**, *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 \pm 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. **f**, Immunoblot shows SMO and GLI1 protein levels in these stable cell lines, with p38 as loading control.



Extended Data Figure 3. Crystallization, structure solution and oligomeric state of SMO C

a, SMO C crystal packing. Asymmetric unit consists of two antiparallel SMO C chains. Chain A coloured as in Fig. 1 with BRIL-fusion in yellow; Chain B in grey. LCP crystal packing with alternating hydrophobic and hydrophilic layers perpendicular to the *c* axis. Molecules coloured as for Chain A. **b**, Pearson correlation coefficient (CC) analysis³¹ used to relate data quality with model quality. 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). **c-e**, SigmaA-weighted 2Fo-Fc electron density maps of final refinement at 1.0 σ contour level. **c**, Val329Phe mutation; **d**, extra density within TMD ligand-binding pocket (Fo-Fc maps shown at contour level of $+3\sigma$ (green) and -3σ (red)) (N.B. This density could not be confidently assigned, likely because of low occupancy within the crystal.); **e**, ‘connector’

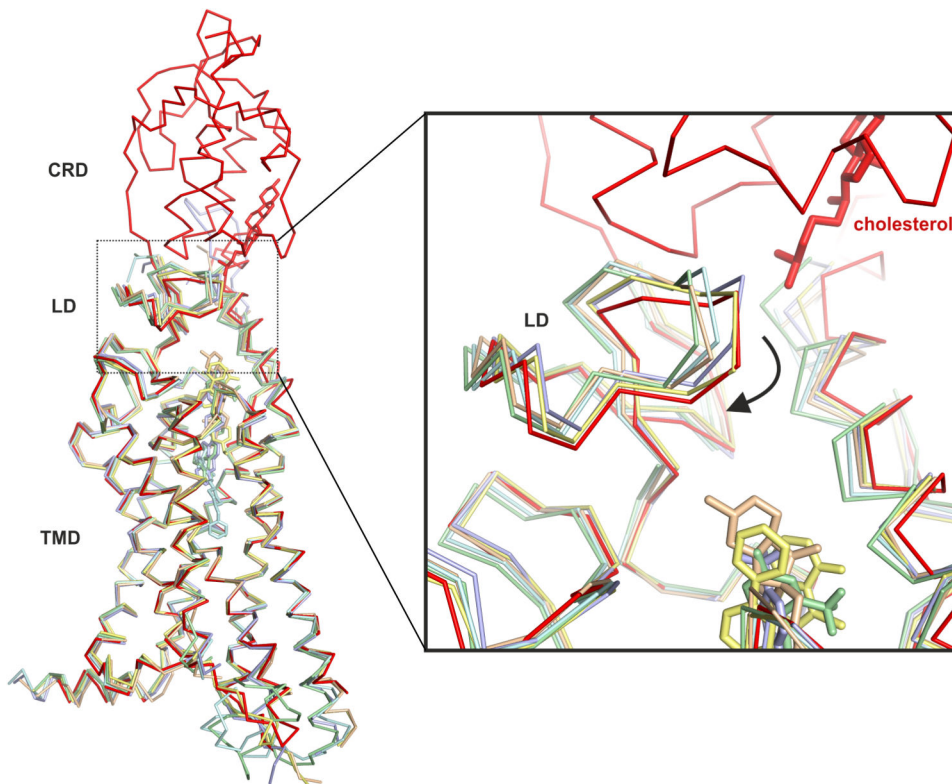
region linking CRD and LD, with Asn188 and linked N-acetyl glycosamine moiety. **f**, SEC-MALS analysis of amphipol-solubilised SMO C. Molar masses (M_W , black lines) and 280 nm absorption (grey line) plotted against elution time. M_W derived from protein-conjugate analysis indicated in parentheses. For clarity, graphs of M_W are only shown around main absorption peak. Theoretical M_W of SMO C based on sequence is 71 kDa with the extra mass observed in MALS (~13 kDa) likely due to three N-linked glycosylation sites. Expected mass of protein-bound amphipol (A8-35) was previously determined to be 40-75 kDa³², in agreement with our data. This analysis suggest that SMO C is a monomer under our purification conditions.



Extended Data Figure 4. Interfaces of the SMO CRD

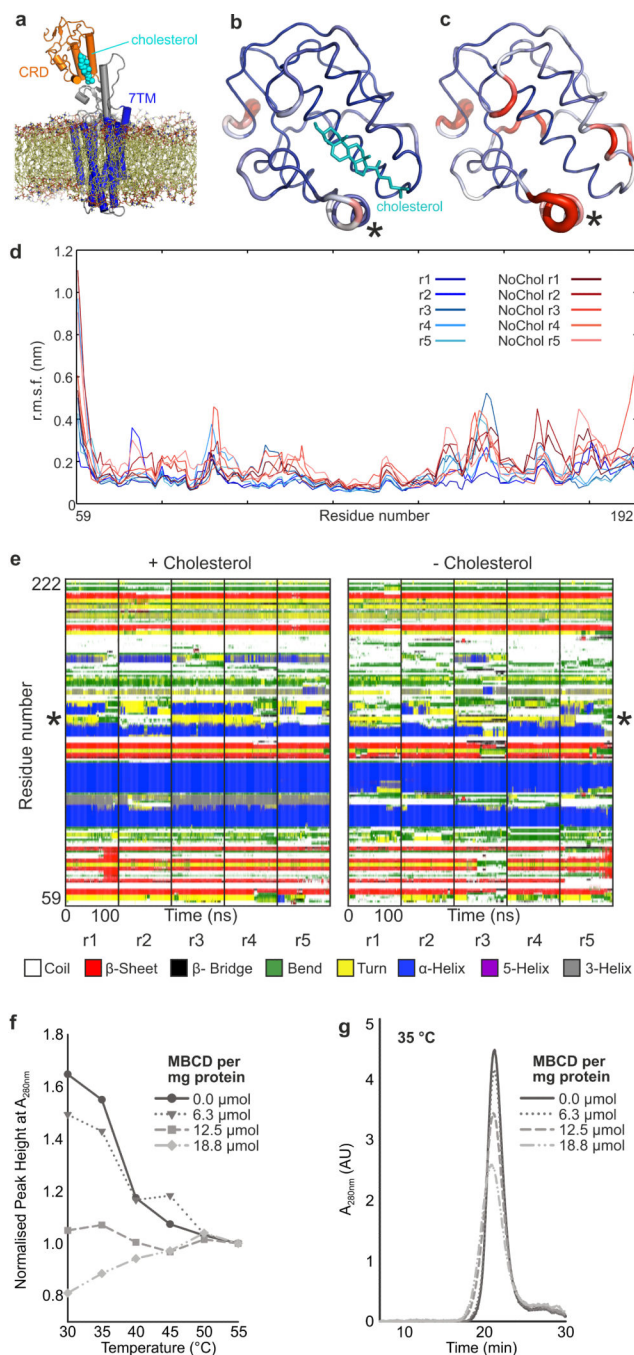
Detailed interactions of the CRD with the connector region (**a**) and with the LD-TMD segment (**b**). The number of interactions are indicated in the top panel and coloured as indicated in the key box. For non-bonded contacts, the width of the striped line is proportional to the number of atomic contacts. 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.



Extended Data Figure 5. Comparison of SMO C with previously determined SMO TMD structures

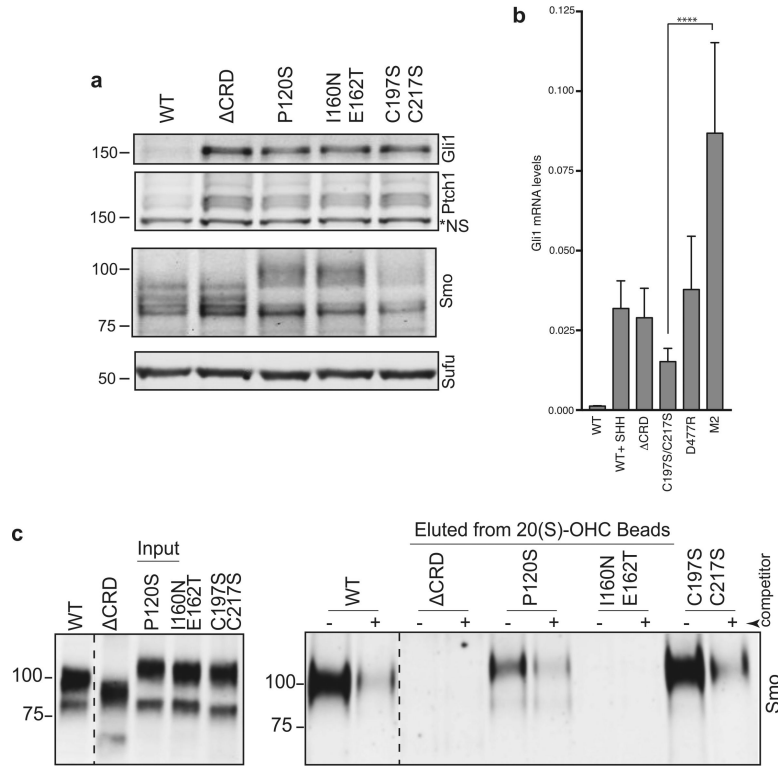
Superposition of the SMO C structure with SMO TMD structures. Structural alignment was performed using the 7TM bundle as template (not including the LD or TMD helix 8). SMO C (red), SMO TMD complexed with cyclopamine (light orange, PDB 4O9R, r.m.s.d. 0.598 Å for 243 equivalent Ca positions), antaXV (light blue, PDB 4QIM, r.m.s.d. 0.515 Å for 233 equivalent Ca positions), SANT1 (pale cyan, PDB 4N4W, r.m.s.d. 0.483 Å for 240 equivalent Ca positions), LY2940680 (pale green, PDB 4JKV, r.m.s.d. 0.493 Å for 230 equivalent Ca positions), SAG1.5 (pale yellow, PDB 4QIN, r.m.s.d. 0.623 Å for 262 equivalent Ca positions). The box shows a close-up view of the LD region revealing a structural rearrangement in the SMO C structure compared to the previously determined SMO TMD structures lacking the native extracellular domain.



Extended Data Figure 6. Cholesterol stabilises SMO. a-e. Molecular dynamics simulations of SMO in a lipid bilayer

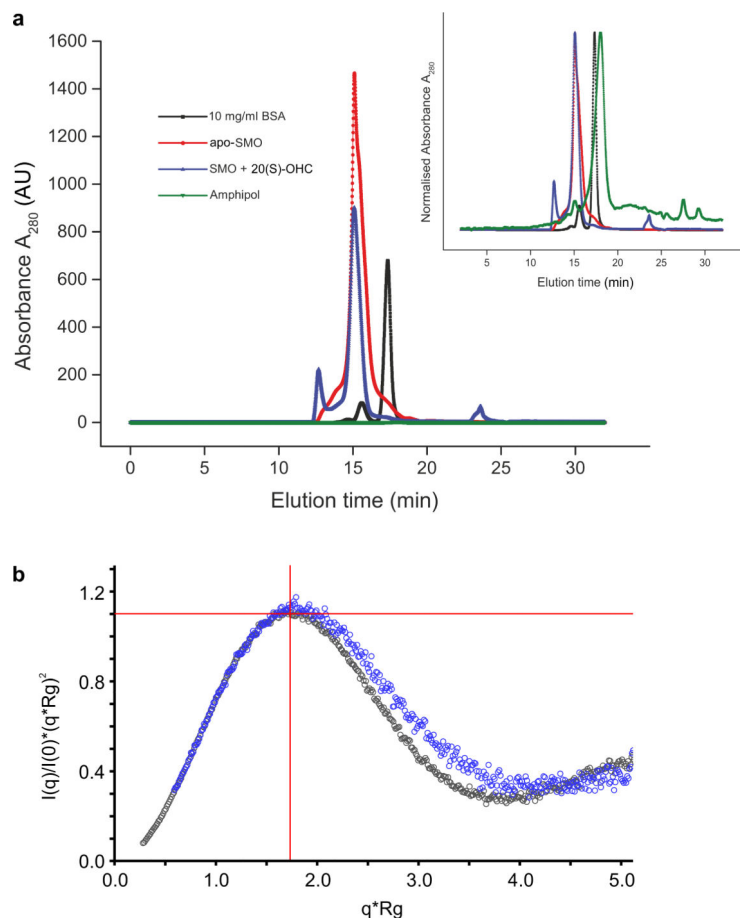
a, SMO embedded in a lipid bilayer with the CRD in orange, the 7-pass transmembrane region excluding intra- and extracellular loops (7TM) in blue and cholesterol in cyan. **b-d**, Relative r.m.s. fluctuations of the C α atoms over the course of 5×100 ns of atomistic molecular dynamics simulation in the presence and absence of the cholesterol. The structures in **b** and **c** 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 **b**, **c** and **e** all mark the helix spanning from residues 155-160, that is destabilised in the absence of bound cholesterol. **f-g**, Thermostability of purified SMO. **c**, See supplementary discussion for details. **f**, Compiled peak heights from thermostability SEC analysis of purified SMO. **c** after treatment with different MBCD concentrations. **g**, Example of raw SEC data used for the analysis in **f**. Samples were incubated at 35°C for 1 h prior to loading onto the SEC column.



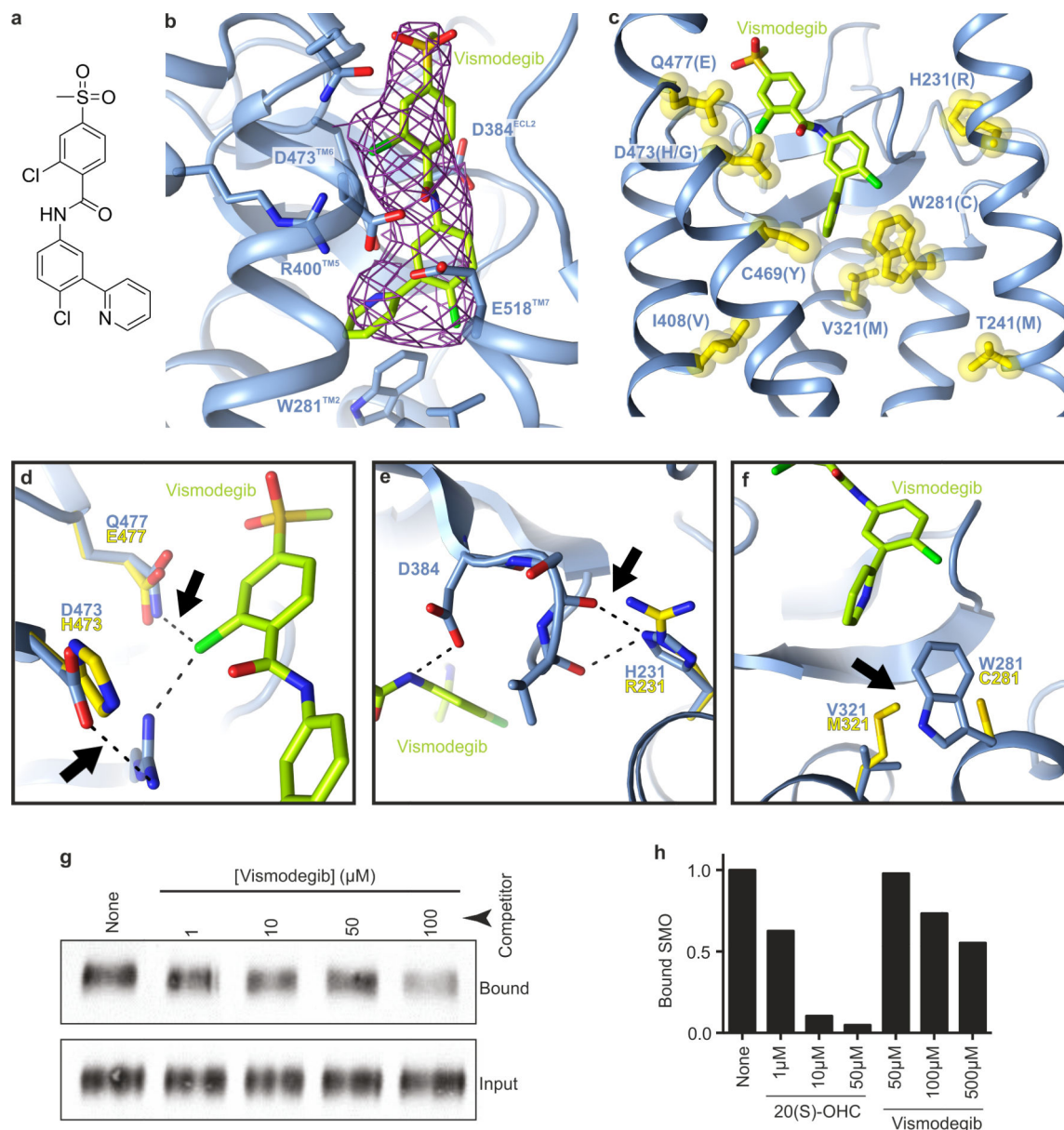
Extended Data Figure 7. Effect of domain interface mutations on expression levels and 20(S)-OHC binding

a, Protein levels of SMO and also of PTCH1 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 Fig. 4c) in the LD. Elevated levels of GLI1 and PTCH1 reflect high constitutive signalling activity of each mutant. NS: a non-specific band detected by anti-PTCH1 antibody; SUFU: loading control. Different patterns seen in SMO panel is caused by different numbers of N-linked glycosylation sites. **b**, *Gli1* mRNA levels were used to assess Hh signalling activity in *Smo*^{-/-} cells stably expressing the indicated mouse SMO variants. D477R and M2 (Trp539Leu) are two previously described mutations in the TMD that increase constitutive signalling. **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. Inputs for each binding reaction are shown on the left.



Extended Data Fig. 8. SAXS analysis of SMO C

a, Overlay of size-exclusion chromatograms monitored at 280 nm (A_{280}) collected during SAXS measurements for apo-SMO C (red), (+) 20(S)-OHC SMO C (blue), amphipol (green) and BSA standard (black). Amphipol and BSA were injected at 10 mg/ml. Insert shows curves normalized 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 minutes after the amphipol-stabilised SMO C samples. **b**, Dimensionless Kratky plot of *apo* and (+) 20(S)-OHC-loaded SMO C SAXS data. Cross-hairs denote the Guinier–Kratky point (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.



Extended Data Fig. 9. Crystal structure of the SMO C-vismodegib complex and structural analysis of mutations found in vismodegib-resistant cancers

a, Chemical structure of vismodegib. **b**, Close-up view of vismodegib binding site. Colour-coding follows Fig. 5b. Composite omit map calculated with PHENIX at 1.0σ shown as magenta chicken-wire. **c**, Mapping of residues that are mutated in vismodegib-resistant tumours (yellow highlights). Brackets indicate mutant residues. **d-f**, 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. **d**, 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. **e**, 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 vismodegib's amide linker. **f**, 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. **g-h**, SMO C captured on cholesterol beads in the presence of increasing concentrations of free vismodegib or 20(S)-OHC (**h**).

Extended Data Table 1
Crystallographic data collection and refinement
statistics

Numbers in parentheses refer to the highest resolution shell. R_{free} equals the R-factor against 5% of the data. r.m.s.d.: root mean square deviation from ideal geometry. R_{pim} : precision indicating merging R-factor. $CC_{1/2}$: correlation coefficient between random half data sets. CC^* : analytical estimate of CC_{true} based on $CC_{1/2}$. 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. Ramachandran statistics were calculated using MolProbity.

	<i>apo</i> -SMO C	<i>vismo</i> -SMO C
Data collection		
Beamline	DIAMOND-I24	DIAMOND-I24
Wavelength (Å)	0.9686	0.9686
No. crystals	9	2
Resolution (Å)	61.0-3.2 (3.41-3.20)	68.4-3.3 (3.39-3.30)
Space group	C2	P1
Cell dimensions		
a, b, c (Å)	122.9, 63.0, 208.6	52.8, 71.3, 107.0
α, β, γ (°)	90.0, 96.6, 90.0	91.5, 98.2, 105.9
Unique reflections	25214 (3815)	21097 (2380)
Completeness (%)	93.8 (78.8)	94.7 (85.8)
R_{merge} (%)	30.5 (>100)	15.2 (>100)
R_{pim} (%)	13.8 (>100)	9.9 (>100)
$CC_{1/2}$	0.980 (0.237)	0.953 (0.357)
CC^*	0.995 (0.580)	0.988 (0.460)
CC_{work}	0.923 (0.354)	0.703 (0.340)
CC_{free}	0.941 (0.333)	0.941 (0.315)
I/ σ I	4.3 (0.6)	3.7 (0.6)
Multiplicity	5.8 (5.1)	1.8 (1.7)
Refinement		
Resolution (Å)	61.0-3.2	60.0-3.3
No. reflections (test set)	25151 (1302)	20588 (1013)
$R_{\text{work}}/R_{\text{free}}$ (%)	23.2/26.4	24.5/25.7

	<i>apo-SMO C</i>	<i>vismo-SMO C</i>
No. atoms		
Protein	9207	8988
Ligand/Ion	58	119
B factors (Å ²)		
Protein	137	124
Ligand/Ion	121	119
r.m.s.d. bonds (Å)	0.008	0.008
r.m.s.d. angles (°)	0.90	0.87
Ramachandran (%)		
Favoured	96.6	96.8
Allowed	3.2	2.9
Disallowed	0.2	0.3

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the staff of beamlines I24 and B21 (MX10627) at the DLS, UK for assistance. We thank S. Masiulis, D. Staunton, K. Jungnickel and A.R. Aricescu for helpful discussions. The work was supported by Cancer Research UK (C20724/A14414), the US National Institutes of Health (GM106078) and Wellcome Trust (102890/Z/13/Z, 092970/Z/10/Z and 090532/Z/09/Z). Further support by NDM Oxford (E.F.X.B.), Medical Research Council UK (G.H.), Ford Foundation (G.L.) and National Science Foundation (Sigrid N.) is acknowledged.

REFERENCES

- Sharpe HJ, Wang W, Hannoush RN, de Sauvage FJ. Regulation of the oncoprotein Smoothed by small molecules. *Nature chemical biology*. 2015; 11:246–255. doi:10.1038/nchembio.1776. [PubMed: 25785427]
- Taipale J, et al. Effects of oncogenic mutations in Smoothed and Patched can be reversed by cyclopamine. *Nature*. 2000; 406:1005–1009. doi:10.1038/35023008. [PubMed: 10984056]
- Chen, J.f K.; Taipale, J.; Cooper, MK.; Beachy, PA. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothed. *Genes & development*. 2002; 16:2743–2748. doi:10.1101/gad.1025302. [PubMed: 12414725]
- Chen JK, Taipale J, Young KE, Maiti T, Beachy PA. Small molecule modulation of Smoothed activity. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99:14071–14076. doi:10.1073/pnas.182542899. [PubMed: 12391318]
- Frank-Kamenetsky M, et al. Small-molecule modulators of Hedgehog signaling: identification and characterization of Smoothed agonists and antagonists. *Journal of biology*. 2002; 1:10. [PubMed: 12437772]
- Robarge KD, et al. GDC-0449-a potent inhibitor of the hedgehog pathway. *Bioorganic & medicinal chemistry letters*. 2009; 19:5576–5581. doi:10.1016/j.bmcl.2009.08.049. [PubMed: 19716296]
- Nachtergaele S, et al. Oxysterols are allosteric activators of the oncoprotein Smoothed. *Nature chemical biology*. 2012; 8:211–220. doi:10.1038/nchembio.765. [PubMed: 22231273]
- Corcoran RB, Scott MP. Oxysterols stimulate Sonic hedgehog signal transduction and proliferation of medulloblastoma cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:8408–8413. doi:10.1073/pnas.0602852103. [PubMed: 16707575]

9. Dwyer JR, et al. Oxysterols are novel activators of the hedgehog signaling pathway in pluripotent mesenchymal cells. *The Journal of biological chemistry*. 2007; 282:8959–8968. doi:10.1074/jbc.M611741200. [PubMed: 17200122]
10. Nachtergaele S, et al. Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. *eLife*. 2013; 2:e01340. doi:10.7554/eLife.01340. [PubMed: 24171105]
11. Nedelcu D, Liu J, Xu Y, Jao C, Salic A. Oxysterol binding to the extracellular domain of Smoothened in Hedgehog signaling. *Nature chemical biology*. 2013; 9:557–564. doi:10.1038/nchembio.1290. [PubMed: 23831757]
12. Myers BR, et al. Hedgehog pathway modulation by multiple lipid binding sites on the smoothened effector of signal response. *Developmental cell*. 2013; 26:346–357. doi:10.1016/j.devcel.2013.07.015. [PubMed: 23954590]
13. Briscoe J, Therond PP. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nature reviews. Molecular cell biology*. 2013; 14:416–429. doi:10.1038/nrm3598. [PubMed: 23719536]
14. Yauch RL, et al. Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science*. 2009; 326:572–574. doi:10.1126/science.1179386. [PubMed: 19726788]
15. Wang C, et al. Structure of the human smoothened receptor bound to an antitumour agent. *Nature*. 2013; 497:338–343. doi:10.1038/nature12167. [PubMed: 23636324]
16. Wang C, et al. Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nature communications*. 2014; 5:4355. doi:10.1038/ncomms5355.
17. Weierstall U, et al. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nature communications*. 2014; 5:3309. doi:10.1038/ncomms4309.
18. Deupi X, Standfuss J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Current opinion in structural biology*. 2011; 21:541–551. doi:10.1016/j.sbi.2011.06.002. [PubMed: 21723721]
19. Katritch V, Cherezov V, Stevens RC. Structure-function of the G protein-coupled receptor superfamily. *Annual review of pharmacology and toxicology*. 2013; 53:531–556. doi:10.1146/annurev-pharmtox-032112-135923.
20. Rana R, et al. Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling. *Nature communications*. 2013; 4:2965. doi:10.1038/ncomms3965.
21. Chun E, et al. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure*. 2012; 20:967–976. doi:10.1016/j.str.2012.04.010. [PubMed: 22681902]
22. Bazan JF, de Sauvage FJ. Structural ties between cholesterol transport and morphogen signaling. *Cell*. 2009; 138:1055–1056. doi:10.1016/j.cell.2009.09.006. [PubMed: 19766557]
23. Janda CY, Waghray D, Levin AM, Thomas C, Garcia KC. Structural basis of Wnt recognition by Frizzled. *Science*. 2012; 337:59–64. doi:10.1126/science.1222879. [PubMed: 22653731]
24. Cooper MK, et al. A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis. *Nature genetics*. 2003; 33:508–513. doi:10.1038/ng1134. [PubMed: 12652302]
25. Blassberg R, Macrae JI, Briscoe J, Jacob J. Reduced cholesterol levels impair Smoothened activation in Smith-Lemli-Opitz syndrome. *Human molecular genetics*. 2015 doi:10.1093/hmg/ddv507.
26. Venkatakrishnan AJ, et al. Molecular signatures of G-protein-coupled receptors. *Nature*. 2013; 494:185–194. doi:10.1038/nature11896. [PubMed: 23407534]
27. Durand D, et al. NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of structural biology*. 2010; 169:45–53. doi:10.1016/j.jsb.2009.08.009. [PubMed: 19723583]
28. Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*. 2011; 95:559–571. doi:10.1002/bip.21638. [PubMed: 21509745]
29. Sharpe HJ, et al. Genomic analysis of smoothened inhibitor resistance in Basal cell carcinoma. *Cancer cell*. 2015; 27:327–341. doi:10.1016/j.ccell.2015.02.001. [PubMed: 25759019]
30. Atwood SX, et al. Smoothened variants explain the majority of drug resistance in Basal cell carcinoma. *Cancer cell*. 2015; 27:342–353. doi:10.1016/j.ccell.2015.02.002. [PubMed: 25759020]

31. Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science*. 2012; 336:1030–1033. doi:10.1126/science.1218231. [PubMed: 22628654]
32. Zoonens M, Popot JL. Amphipols for each season. *The Journal of membrane biology*. 2014; 247:759–796. doi:10.1007/s00232-014-9666-8. [PubMed: 24969706]
33. Varjosalo M, Li SP, Taipale J. Divergence of hedgehog signal transduction mechanism between *Drosophila* and mammals. *Developmental cell*. 2006; 10:177–186. doi:10.1016/j.devcel.2005.12.014. [PubMed: 16459297]
34. Rohatgi R, Milenkovic L, Scott MP. Patched1 regulates hedgehog signaling at the primary cilium. *Science*. 2007; 317:372–376. doi:10.1126/science.1139740. [PubMed: 17641202]
35. Humke EW, Dorn KV, Milenkovic L, Scott MP, Rohatgi R. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes & development*. 2010; 24:670–682. doi:10.1101/gad.1902910. [PubMed: 20360384]
36. Aricescu AR, Lu W, Jones EY. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta crystallographica. Section D, Biological crystallography*. 2006; 62:1243–1250. doi:10.1107/S0907444906029799. [PubMed: 17001101]
37. Molday RS, MacKenzie D. Monoclonal antibodies to rhodopsin: characterization, cross-reactivity, and application as structural probes. *Biochemistry*. 1983; 22:653–660. [PubMed: 6188482]
38. Oprian DD, Molday RS, Kaufman RJ, Khorana HG. Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proceedings of the National Academy of Sciences of the United States of America*. 1987; 84:8874–8878. [PubMed: 2962193]
39. Zacharias DA, Violin JD, Newton AC, Tsien RY. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science*. 2002; 296:913–916. doi:10.1126/science.1068539. [PubMed: 11988576]
40. Nagai T, et al. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature biotechnology*. 2002; 20:87–90. doi:10.1038/nbt0102-87.
41. Rohatgi R, Milenkovic L, Corcoran RB, Scott MP. Hedgehog signal transduction by Smoothed: pharmacologic evidence for a 2-step activation process. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:3196–3201. doi:10.1073/pnas.0813373106. [PubMed: 19218434]
42. Miller PS, Aricescu AR. Crystal structure of a human GABA receptor. *Nature*. 2014 doi:10.1038/nature13293.
43. Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure*. 2006; 14:673–681. doi:10.1016/j.str.2006.01.013. [PubMed: 16615909]
44. Bailey EC, Milenkovic L, Scott MP, Collawn JF, Johnson RL. Several PATCHED1 missense mutations display activity in patched1-deficient fibroblasts. *The Journal of biological chemistry*. 2002; 277:33632–33640. doi:10.1074/jbc.M202203200. [PubMed: 12072433]
45. Pear WS, Nolan GP, Scott ML, Baltimore D. Production of high-titer helper-free retroviruses by transient transfection. *Proceedings of the National Academy of Sciences of the United States of America*. 1993; 90:8392–8396. [PubMed: 7690960]
46. Ciepla P, et al. New chemical probes targeting cholesterylation of Sonic Hedgehog in human cells and zebrafish. *Chem Sci*. 2014; 5:4249–4259. doi:10.1039/c4sc01600a. [PubMed: 25574372]
47. Chung S-K, Shim J-Y, Kang MG, Lee KW, Kang HI. Studies of steroids as potential antifungal agent 2. Side chain modified cholesterol. *Korean Journal of Medicinal Chemistry*. 1998; 8:14–17.
48. Caffrey M, Cherezov V. Crystallizing membrane proteins using lipidic mesophases. *Nature protocols*. 2009; 4:706–731. doi:10.1038/nprot.2009.31.
49. Winter G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography*. 2010; 43:186–190. doi:10.1107/S0021889809045701.
50. Kabsch W. Xds. *Acta crystallographica. Section D, Biological crystallography*. 2010; 66:125–132. doi:10.1107/S0907444909047337.
51. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography*. 2013; 69:1204–1214. doi:10.1107/S0907444913000061.

52. Winn MD, et al. Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography*. 2011; 67:235–242. doi:10.1107/S0907444910045749.
53. McCoy AJ, et al. Phaser crystallographic software. *Journal of applied crystallography*. 2007; 40:658–674. doi:10.1107/S0021889807021206. [PubMed: 19461840]
54. Liu W, et al. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science*. 2012; 337:232–236. doi:10.1126/science.1219218. [PubMed: 22798613]
55. Cowtan K. Recent developments in classical density modification. *Acta Crystallogr D*. 2010; 66:470–478. doi:10.1107/S090744490903947x. [PubMed: 20383000]
56. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica. Section D, Biological crystallography*. 2006; 62:1002–1011. doi:10.1107/S0907444906022116.
57. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*. 2004; 60:2126–2132. doi:10.1107/S0907444904019158.
58. BUSTER v. 2.10.2. Global Phasing Ltd.; Cambridge, United Kingdom: 2011.
59. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography*. 2010; 66:213–221. doi:10.1107/S0907444909052925.
60. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*. 2010; 66:12–21. doi:10.1107/S0907444909042073.
61. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98:10037–10041. doi:10.1073/pnas.181342398. [PubMed: 11517324]
62. Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *Journal of molecular biology*. 1984; 179:125–142. doi:10.1016/0022-2836(84)90309-7. [PubMed: 6502707]
63. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010
64. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*. 2007; 372:774–797. doi:10.1016/j.jmb.2007.05.022. [PubMed: 17681537]
65. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic acids research*. 1988; 16:10881–10890. [PubMed: 2849754]
66. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*. 2014; 42:W320–324. doi:10.1093/nar/gku316. [PubMed: 24753421]
67. Chovancova E, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS computational biology*. 2012; 8:e1002708. doi:10.1371/journal.pcbi.1002708. [PubMed: 23093919]
68. Tribet C, Audebert R, Popot JL. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93:15047–15050. [PubMed: 8986761]
69. Gohon Y, et al. Bacteriorhodopsin/amphipol complexes: structural and functional properties. *Biophysical journal*. 2008; 94:3523–3537. doi:10.1529/biophysj.107.121848. [PubMed: 18192360]
70. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*. 2008; 4:435–447. doi:10.1021/ct700301q. [PubMed: 26620784]
71. Olsson MHM, Sondergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *Journal of chemical theory and computation*. 2011; 7:525–537. doi:10.1021/ct100578z. [PubMed: 26596171]
72. Sondergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. *Journal of chemical theory and computation*. 2011; 7:2284–2295. doi:10.1021/ct200133y. [PubMed: 26606496]

73. de Jong DH, et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of chemical theory and computation*. 2013; 9:687–697. doi:10.1021/ct300646g. [PubMed: 26589065]
74. Scott KA, et al. Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure*. 2008; 16:621–630. doi:10.1016/j.str.2008.01.014. [PubMed: 18400182]
75. Stansfeld PJ, et al. MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure*. 2015; 23:1350–1361. doi:10.1016/j.str.2015.05.006. [PubMed: 26073602]
76. Stansfeld PJ, Sansom MS. Molecular simulation approaches to membrane proteins. *Structure*. 2011; 19:1562–1572. doi:10.1016/j.str.2011.10.002. [PubMed: 22078556]
77. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry. B*. 2007; 111:7812–7824. doi:10.1021/jp071097f. [PubMed: 17569554]
78. Monticelli L, Sorin EJ, Tieleman DP, Pande VS, Colombo G. Molecular simulation of multistate peptide dynamics: a comparison between microsecond timescale sampling and multiple shorter trajectories. *Journal of computational chemistry*. 2008; 29:1740–1752. doi:10.1002/jcc.20935. [PubMed: 18307167]
79. Periolo X, Cavalli M, Marrink SJ, Ceruso MA. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *Journal of chemical theory and computation*. 2009; 5:2531–2543. doi:10.1021/ct9002114. [PubMed: 26616630]
80. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys*. 1984; 81:3684–3690. doi:10.1063/1.448118.
81. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry*. 1997; 18:1463–1472. doi:10.1002/(Sici)1096-987x(199709)18:12<1463::Aid-Jcc4>3.0.Co;2-H.
82. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry*. 2004; 25:1656–1676. doi:10.1002/jcc.20090. [PubMed: 15264259]
83. Pol-Fachin L, Verli H, Lins RD. Extension and Validation of the GROMOS 53A6(GLYC) Parameter Set for Glycoproteins. *Journal of computational chemistry*. 2014; 35:2087–2095. doi:10.1002/jcc.23721. [PubMed: 25196137]
84. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007; 126 doi:10.1063/1.2408420.
85. Parrinello M, Rahman A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J Appl Phys*. 1981; 52:7182–7190. doi:10.1063/1.328693.
86. Essmann U, et al. A Smooth Particle Mesh Ewald Method. *J Chem Phys*. 1995; 103:8577–8593. doi:10.1063/1.470117.
87. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph Model*. 1996; 14:33–38. doi:10.1016/0263-7855(96)00018-5.

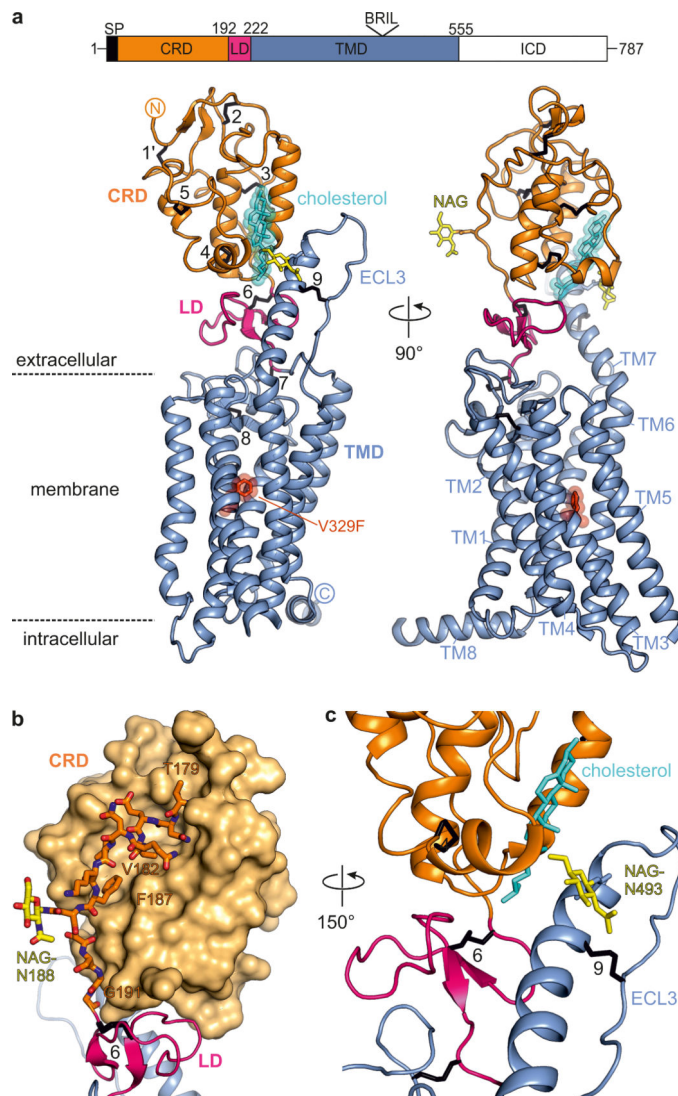


Figure 1. Structure of human SMO

a, Two views of the overall structure showing extracellular and transmembrane domains of human SMO in cartoon representation with the CRD in orange, LD in pink, TMD in blue. The inactivating point mutation Val329Phe is depicted in red, cholesterol in cyan, nine numbered disulphide bridges in black, and two N-linked glycans (NAG) as yellow sticks. A schematic of SMO is shown above (SP: signal peptide, BRIL: position of the BRIL fusion protein inserted between TMD helices 5 and 6). **b**, The ‘connector’ region between CRD and LD highlighted as sticks in atomic colouring, with the CRD shown as a solvent accessible surface and the LD and part of TMD ECL3 loop as cartoons. **c**, Interface between CRD, LD and TMD shown in cartoon representation with ECL3-NAG and cholesterol as yellow and cyan sticks, respectively.

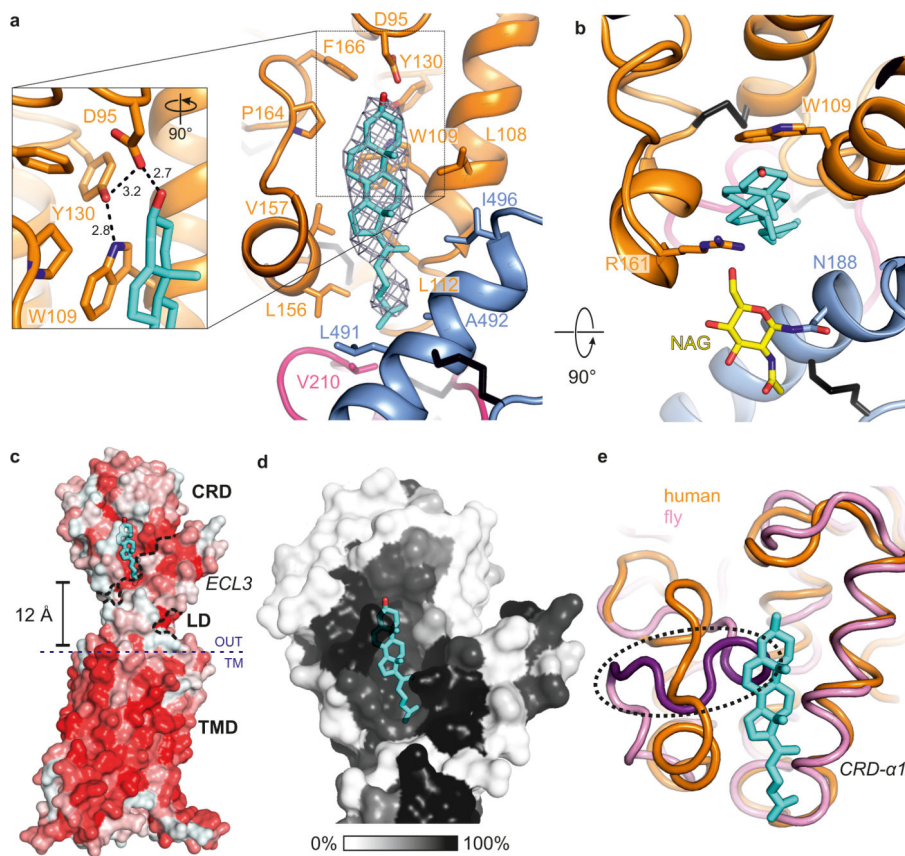


Figure 2. The cholesterol binding site

a-b, Close-up of cholesterol with interacting residues as sticks. Initial 2Fo-Fc map at 1.0 σ before inclusion of cholesterol shown as chicken-wire. Colour-coding follows Fig.1. Inset shows the potential hydrogen-bonding network coordinating the cholesterol 3 β -hydroxyl group. Interatomic distances (\AA) are shown in black. **c**, Solvent accessible surface colour-coded by hydrophobicity (red: hydrophobic, white: hydrophilic). **d**, Sequence conservation (based on 55 vertebrate SMO sequences) mapped onto SMO C (black: conserved, white: not conserved). **e**, Superposition of human (orange, this study) and fly (purple, PDB 2MAH²⁰) SMO CRD structures. The fly CRD region occupying the cholesterol binding site is highlighted.

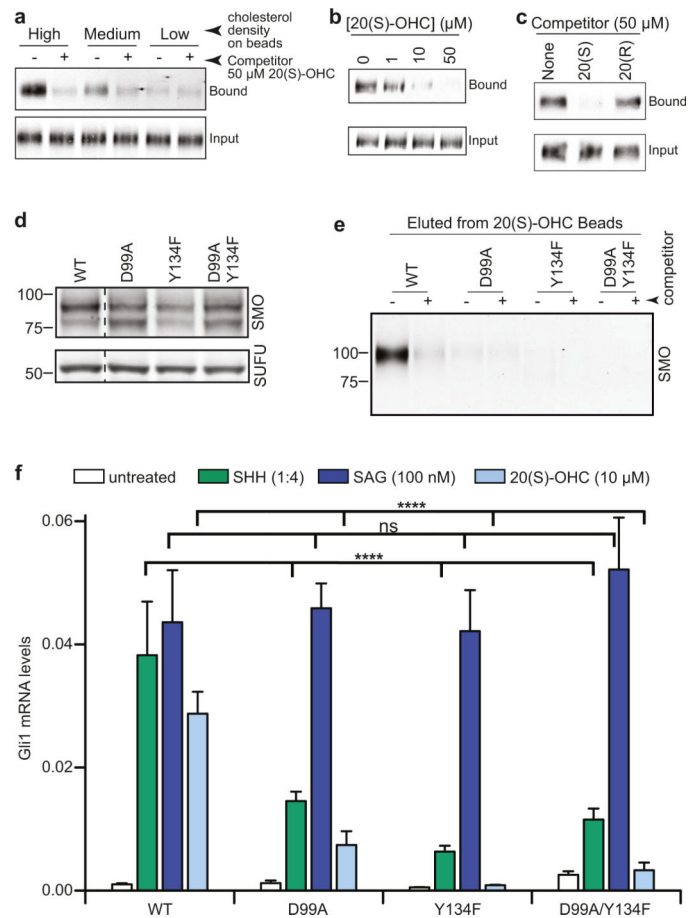


Figure 3. The cholesterol-binding site regulates SMO signalling activity

a, Purified SMO C captured on beads coupled to increasing concentrations of cholesterol in the presence or absence of free 20(S)-OHC. **b**, SMO C captured on cholesterol beads in the presence of increasing concentrations of free 20(S)-OHC or **c**, in the presence of a synthetic epimer 20(R)-OHC. **d**, Protein levels of SMO variants stably expressed in *Smo*^{-/-} mouse fibroblasts. **e**, Binding of SMO variants to 20(S)-OHC-coupled beads. Asp99Ala and Tyr134Phe are predicted to disrupt hydrogen bonding with cholesterol (Fig.2a). **f**, Levels of *Gli1* mRNA (mean \pm SD, $n=4$) were used as a metric for Hh signalling activity in cell lines shown in **d** 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.

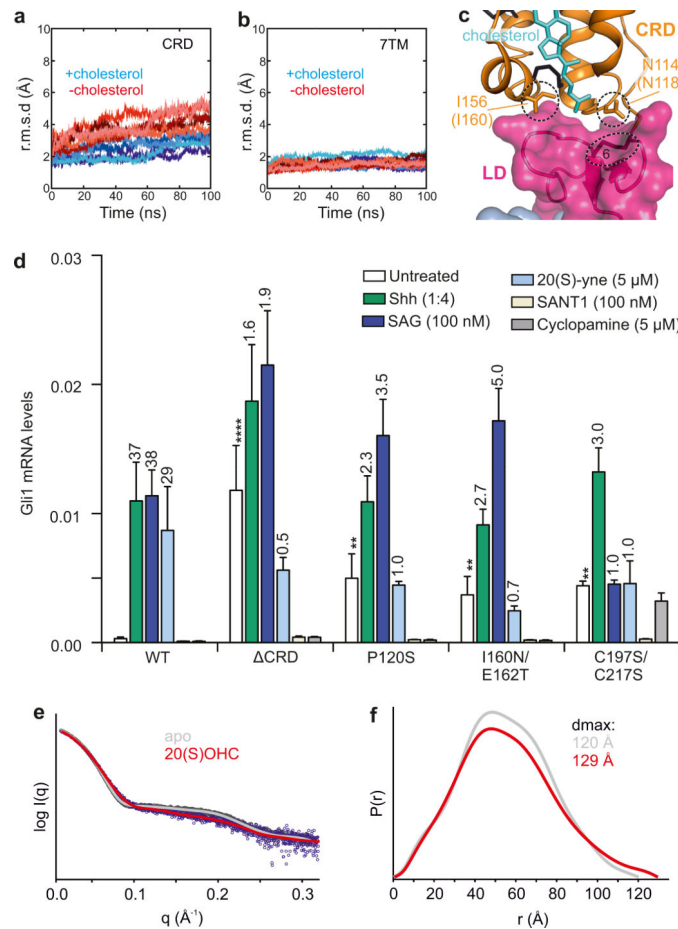


Figure 4. SMO activity is regulated by the stability of its extracellular domain
a-b, MD simulations of SMO. **a**, r.m.s.d. versus time for CRD (**a**) and 7TM region (**b**) with (blues) and without (reds) cholesterol. **c-d**, Mutations in the extracellular region increase constitutive signalling activity of SMO. **c**, CRD-LD interface with mutated residues highlighted. Corresponding mouse residues are in brackets. **d**, *Gli1* mRNA levels (mean \pm SD, $n=4$) after treatment with agonists (SHH, SAG, 20(S)-yne) or antagonists (cyclopamine, SANT-1). *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. **e**, Solution-state SAXS profiles of untreated (grey) and 20(S)-OHC-treated (red) SMO. Lines are fits derived from the indirect Fourier transform of **e**. **f**, Paired-distance, $P(r)$, distribution functions, normalised to their respective $I(0)$ values, with maximum particle sizes (d_{\max}) indicated.

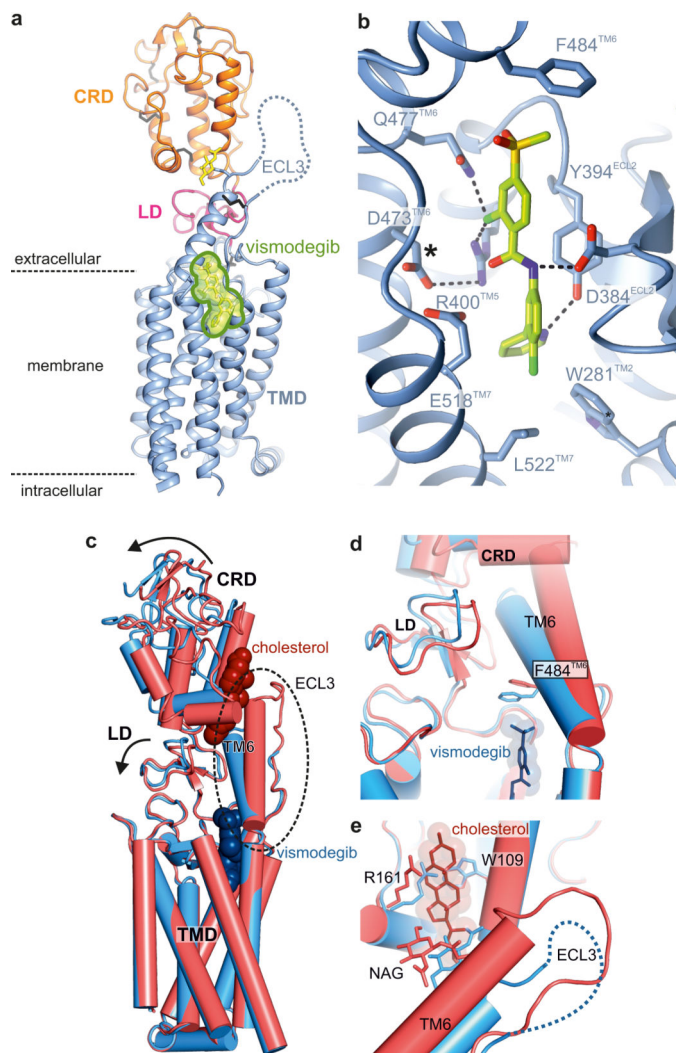


Figure 5. Structure of SMO in complex with the antagonist Vismodegib

a, Overall structure showing full extracellular and transmembrane domains of human SMO (cartoon) in complex with vismodegib (green). Colour-coding follows Fig.1. **b**, Close-up of the vismodegib-binding site. The asterisk denotes a residue mutated in vismodegib-resistant tumours. **c-e** Comparison of *apo*-SMO (red) and *vismo*-SMO (blue). **c**, Superposition of the two SMO structures based on the TMD. Arrows indicate domain rotations. Cholesterol (red) and vismodegib (blue) shown as spheres. **d**, Close-up of the TMD/LD/CRD interface. Phe484 contacts the vismodegib methylsulfone group. **e**, 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.