ARTICLE

Structural basis of Smoothened regulation by its extracellular domains

Eamon F. X. Byrne[1]*, Ria Sircar[2]*, Paul S. Miller[1], George Hedger[3], Giovanni Luchetti[2], Sigrid Nachtergaele[2], Mark D. Tully[4], Laurel Mydock-McGrane[5], Douglas F. Covey[5], Robert P. Rambo[4], Mark S. P. Sansom[3], Simon Newstead[3]*, Rajat Rohatgi[2] & Christian Siebold[1]

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 the large extracellular domains of GPCRs regulate signalling by the TMD is unknown. We present crystal structures of the Hh signal transducer and oncoprotein Smoothened, a GPCR that contains two distinct ligand-binding sites: one in its TMD and one in the CRD. The CRD is stacked atop the TMD, separated by an intervening wedge-like linker domain. Structure-guided mutations show that the interface between the CRD, linker domain and TMD stabilizes the inactive state of Smoothened. Unexpectedly, we find a cholesterol molecule bound to Smoothened in the CRD binding site. Mutations predicted to prevent cholesterol binding impair the ability of Smoothened to transmit native Hh signals. Binding of a clinically used antagonist, vismodegib, to the TMD induces a conformational change that is propagated to the CRD, resulting in loss of cholesterol from the CRD–linker domain–TMD interface. Our results clarify the structural mechanism by which the activity of a GPCR is controlled by ligand-regulated interactions between its extracellular and transmembrane domains.

The extracellular region of Smoothened (SMO) is composed of an N-terminal CRD followed by a small linker domain, which then connects to the TMD and a C-terminal intracellular domain (ICD; Fig. 1a). Studies using small-molecule agonists and antagonists of SMO have defined two separable ligand-binding sites, one in the TMD and one in the CRD[1]. The TMD binding site binds the plant-derived inhibitor cyclopamine[2,3], the synthetic agonist SAG[4,5], and the anti-cancer drug vismodegib[6], which is used clinically to treat advanced basal cell cancer. Side-chain oxysterols such as 20(S)-hydroxycholesterol (20(S)-OHC) represent a distinct class of SMO ligands[7–9] that activate signalling by engaging a hydrophobic groove on the surface of the SMO CRD[10–12]. The native morphogen Sonic Hedgehog (SHH) functions by binding and inactivating Patched 1 (PTCH1), the major receptor for Hh ligands, which suppresses SMO activity[13]. Despite the discovery of numerous exogenous SMO ligands, no *bona fide* endogenous SMO ligand that regulates Hh signalling has been identified. 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]. By contrast, several mutations in the TMD site that blocked the binding and activity of synthetic ligands failed to have any effect on the basal or SHH-stimulated activity of SMO[12,14]. These data suggest that an endogenous SMO ligand that can regulate Hh signalling engages the CRD groove on SMO.

Crystal structures of the isolated SMO linker domain–TMD in complex with both agonist and antagonist ligands[15–17] have shown that the GPCR heptahelical scaffold is conserved and provided a detailed view of a small-molecule binding pocket, but did not show the 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, structural insights into how the extracellular domains and TMD interact to regulate signalling in SMO (or in any other GPCR) are lacking.

Overall structure of SMO

We determined the crystal structure of human SMO containing both the CRD and the TMD, connected by the juxta-membrane linker domain (SMO∆C, Fig. 1a and Extended Data Fig. 1). To study the SMO TMD in a defined functional state and to reduce conformational flexibility, we included a single amino acid mutation, Val329Phe[16], 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 transmembrane helices 5 and 6 was replaced by thermostabilized apocytochrome *b*562RIL (BRIL)[21]. The SMO∆C structure was determined to 3.2 Å resolution (Extended Data Table 1 and Extended Data Fig. 3). The asymmetric unit, comprising two molecules arranged 'head-to-tail', stacks into alternating hydrophobic and hydrophilic layers along one axis, as is 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 linker domain, which forms a wedge between the 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 stabilized by nine disulfide bridges, four of which (numbered 2–5 in Fig. 1a) reveal the canonical disulfide pattern of the CRD fold[22] and one that is specific

[1]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [2]Departments of Biochemistry and Medicine, Stanford University School of Medicine, Stanford, California, 94305, USA. [3]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK. [4]Diamond Light Source Ltd, Harwell Science & Innovation Campus, Didcot OX11 0DE, UK. [5]Department of Developmental Biology, Washington University School of Medicine, St. Louis, Missouri, 63110, USA.
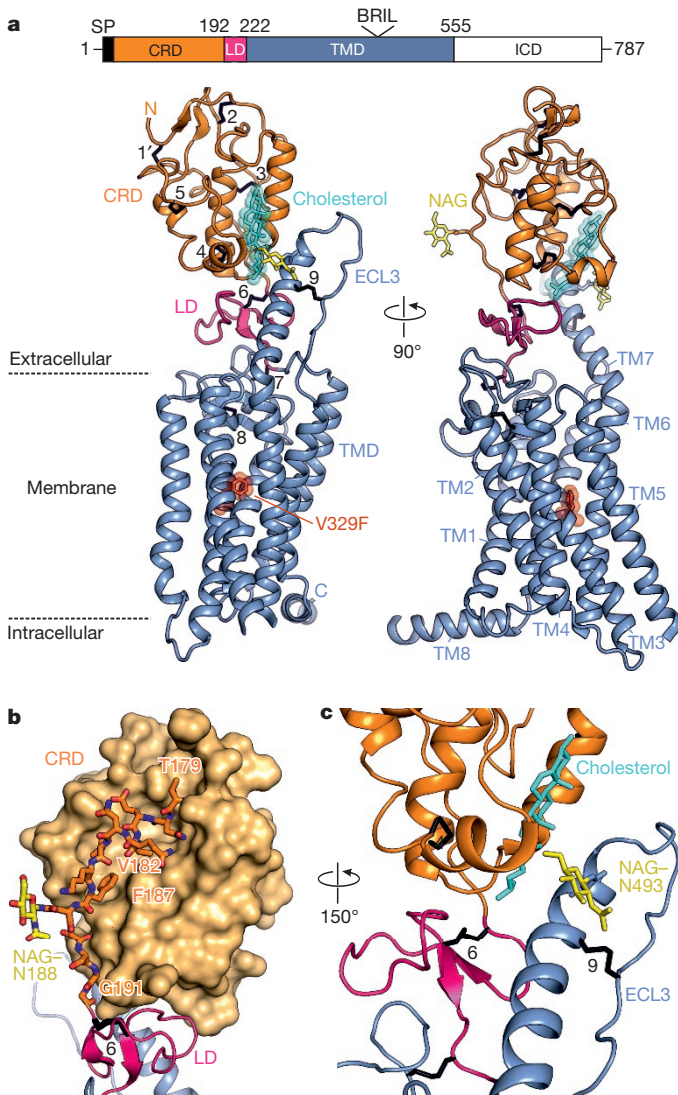*These authors contributed equally to this work.

Figure 1 | Structure of human SMO. a, Two views of the overall structure showing the extracellular and transmembrane domains of human SMO in cartoon representation. Orange, CRD; pink, linker domain (LD); blue, TMD; red, inactivating point mutation Val329Phe; cyan, cholesterol; black, nine numbered disulfide bridges; yellow sticks, two N-linked glycans (NAG). 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 the CRD and linker domain highlighted as sticks in atomic colouring, with the CRD shown as a solvent-accessible surface and the linker domain and part of the TMD ECL3 loop as cartoons. **c**, Interface between the CRD, linker domain and TMD shown in cartoon representation. Yellow sticks, ECL3-NAG; cyan sticks, cholesterol.

for SMO[10] (marked as 1′ in Fig. 1a). This disulfide bridge positions the start of a 'connector' segment that links the CRD and linker domain (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 and Phe187), multiple hydrophilic residues and an N-linked glycan facing the solvent (Fig. 1b and Extended Data Fig. 4a). The connector region, combined with mainly hydrophobic interactions between the CRD and both the linker domain 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 (Fig. 1a, b and Extended Data Fig. 4b). However, we did not observe major structural changes in the heptahelical TMD bundle when we compared it to previously solved structures of either antagonist- or agonist-bound complexes lacking the CRD[15–17]. The

only exception was a rearrangement of the linker domain, 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 probably require co-crystallization with its (still unknown) downstream effector or the use of an active-state stabilizing 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, linker domain and TMD; Figs 1c and 2a, b), and adopts an extended conformation with its tetracyclic sterol ring bound in a shallow groove in the CRD, a site previously shown[10–12] to bind 20(S)-OHC in SMO and the palmitoleyl group of Wnt ligands in Frizzled receptors[23]. The cholesterol iso-octyl tail, located at the interface between the CRD, linker domain 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 indicates that a cholesterol molecule would have to completely desorb from the plasma membrane surface to access its binding site in SMO. Cellular cholesterol levels are permissive for Hh signalling[24,25]; however, this requirement is likely to be unrelated to the cholesterol binding site seen in our structure because SMO mutants lacking the CRD[12] or carrying mutations in the cholesterol-binding groove[25] remain sensitive to cholesterol depletion (Supplementary Discussion).

The sterol-binding site of SMO is predominantly lined with hydrophobic residues from the CRD, which stabilize 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 signalling driven by either 20(S)-OHC or SHH[10–12]. The β-face is shielded by the side-chain of Arg161 and an N-linked glycan (Fig. 2b). The cholesterol 3β-hydroxyl group is
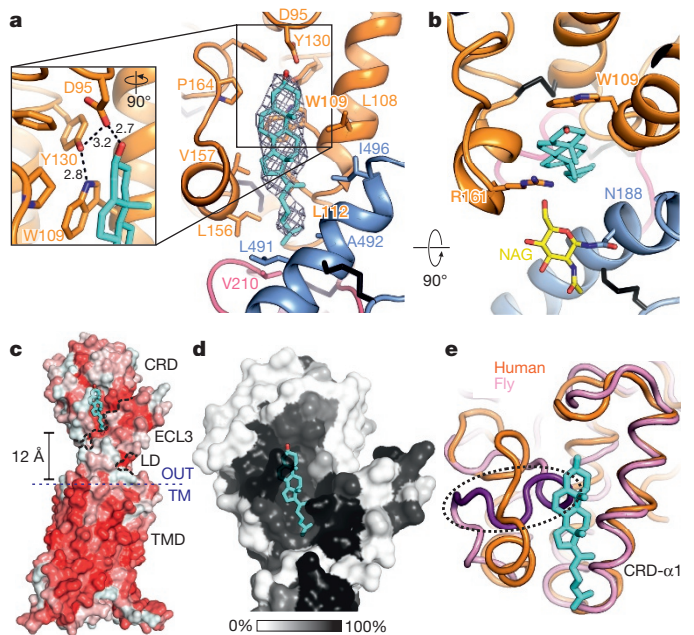

Figure 2 | The cholesterol binding site. a, b, Close-up of cholesterol with interacting residues as sticks. Initial $2F_o–F_c$ map at 1.0σ before inclusion of cholesterol shown as chicken-wire. Colour-coding follows Fig. 1. Inset shows the potential hydrogen-bonding network coordinating the cholesterol 3β-hydroxyl group. Interatomic distances (Å) 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 SMOΔC (black, conserved; white, not conserved). **e**, Superposition of human (orange, this study) and fly (purple, PDB 2MAH[20]) SMO CRD structures. The fly CRD region occupying the cholesterol-binding site is highlighted by the dashed line.

incorporated in a hydrogen-bonding network with the side chains of Asp95, Trp109 and Tyr130; this optimally positions the α-face of the sterol ring system to make a stacking interaction with the indole ring of Trp109 (Fig. 2a, b). The Leu491, Ala492 and Ile496 residues of TMD ECL3 and the Val210 residue of the linker domain 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 and Extended Data Fig. 1). The superposition of our human CRD structure with a previously solved solution structure of the *Drosophila* CRD[20] (r.m.s.d. 1.71 Å for 86 equivalent Cα positions) revealed a major rearrangement of one edge of the cholesterol-binding site comprising fly SMO residues 183–190 (Fig. 2e and Extended Data Fig. 1). This segment of fly SMO forms a short helix that protrudes into the CRD groove and consequently may preclude cholesterol binding. In fact, fly SMO does not bind oxysterols[10,11].

Consistent with the structure, purified SMOΔC bound to cholesterol in a ligand-affinity assay (Fig. 3a–c), analogous to the assay we
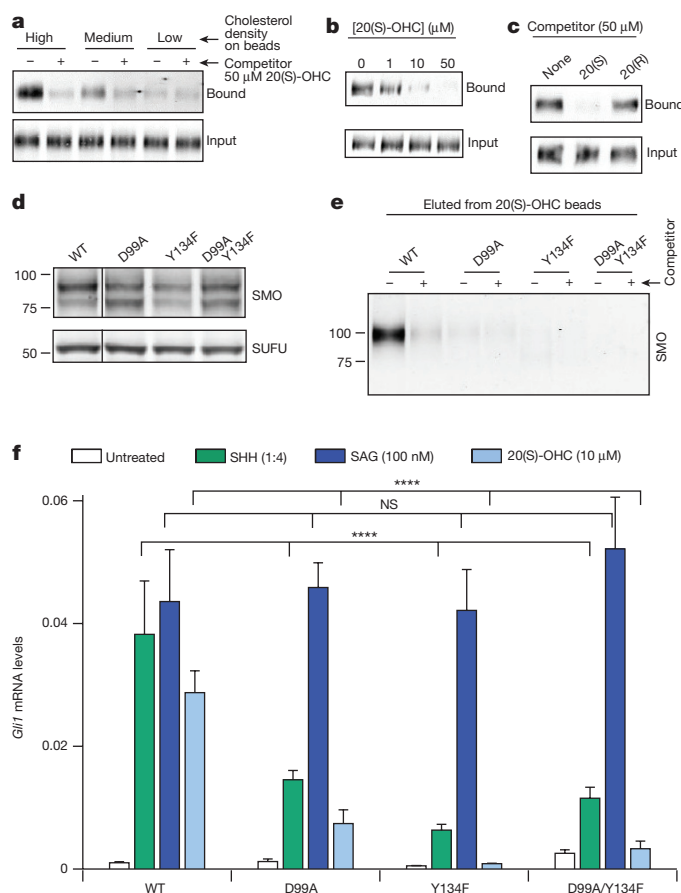

Figure 3 | The cholesterol-binding site regulates SMO signalling activity. a, Purified SMOΔC captured on beads coupled to increasing concentrations of cholesterol in the presence or absence of free 20(S)-OHC. **b, c**, SMOΔC captured on cholesterol beads in the presence of increasing concentrations of free 20(S)-OHC (**b**) or in the presence of a synthetic epimer 20(R)-OHC (**c**). **d**, Protein levels of SMO variants stably expressed in *Smo*[−/−] mouse fibroblasts. WT, wild-type. The protein Suppressor of Fused (SUFU) served as a loading control. **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 arbitrary units ± s.d., $n = 4$) 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 based on one-way ANOVA is denoted for the difference in *Gli1* mRNA levels between cells expressing wild-type SMO and cells expressing each mutant SMO protein. n.s., $P > 0.05$; ****$P \leq 0.0001$. Each experiment was repeated 3 or more times.

previously developed to measure the binding of SMO to 20(S)-OHC[7]. Beads covalently coupled to cholesterol captured purified SMOΔC (Fig. 3a). Binding could be blocked by free 20(S)-OHC added as a competitor at concentrations that activate Hh signalling in cells[7] (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 CRD[7] and cannot activate Hh signalling[10], 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 3β-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. As cholesterol and oxysterols occupy the same groove in the CRD (Fig. 3b), they probably adopt a similar conformation and participate in 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 (Fig. 3d, e). As previously demonstrated[7], wild-type 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*[−/−] mouse 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 increase the basal activity of SMO in the absence of Hh agonists and did not significantly 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 significantly 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 stabilization by the SMO CRD

To assess the structural influence 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 (Fig. 4a, b and Extended Data Fig. 6). These simulations revealed that the SMO CRD has substantial conformational flexibility when not bound to a ligand. In the presence of cholesterol, however, there was a pronounced decrease in this flexibility (Fig. 4a), consistent with the idea that cholesterol stabilizes the CRD structure. By contrast, cholesterol did not substantially change the conformational stability of the SMO TMD (Fig. 4b). The predominant cholesterol-induced stabilization 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 SMOΔC was reduced when cholesterol was depleted with methyl-β-cyclodextrin (Extended Data Fig. 6f, g and Supplementary Discussion).

The molecular dynamics results prompted us to test the signalling consequences of destabilizing interactions within the extracellular region of SMO. To probe the effect of interactions between the CRD and the linker domain, we generated two mutants (Pro120Ser and Ile160Asn/Glu162Thr) in the context of mouse SMO. These introduce bulky, *N*-linked glycosylation sites at positions 114 and 156, respectively, of the human CRD that contact the linker domain
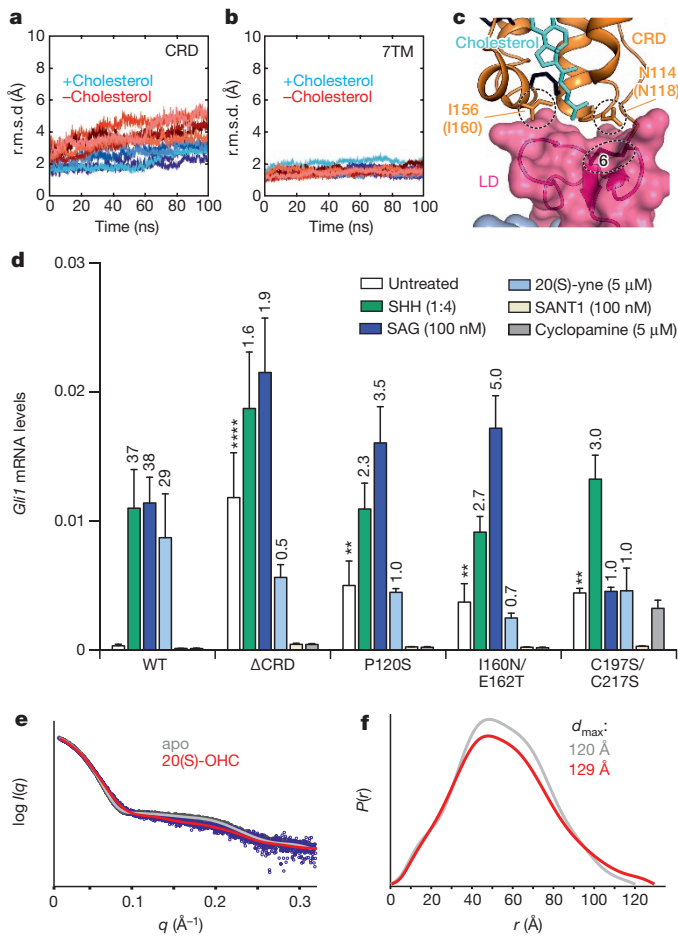
Figure 4 | SMO activity is regulated by the stability of its extracellular domain. a, b, Molecular dynamics (MD) simulations of SMO. Cα r.m.s.d. versus time for CRD (**a**) and 7TM region (**b**) with (blues) and without (reds) cholesterol. **c, d,** Mutations in the extracellular region increase constitutive signalling activity of SMO. **c,** CRD–linker domain interface with mutated residues highlighted. Corresponding mouse residues are in brackets. **d,** *Gli1* mRNA levels (mean arbitrary units ± s.d., $n = 4$, ≥3 independent repeats) after treatment with agonists (SHH, SAG, 20(S)-yne) or antagonists (cyclopamine, SANT-1). *Gli1* mRNA levels in untreated cells reflect constitutive activity of SMO. Numbers above bars indicate *Gli1* fold-increase compared to untreated cells. Asterisks denote statistical significance based on one-way ANOVA for comparison with wild-type SMO. **$P \leq 0.01$, ****$P \leq 0.0001$. **e,** Solution-state SAXS profiles of untreated (grey) and 20(S)-OHC-treated (red) SMOΔC. Lines are fits derived from the indirect Fourier transform of the shown data points. **f,** Paired-distance ($P(r)$) distribution functions, normalized to their respective $I(0)$ values, with maximum particle sizes (d_{max}) of 120 Å for untreated (grey) and 129 Å for 20(S)-OHC-treated (red) SMOΔC.

(Fig. 4c). To disrupt the conformation of the linker domain itself, we mutated two cysteines (Cys197 and Cys217 in mouse SMO) that form a disulfide bond within the linker domain (numbered 6 in Figs 1a and 4c) to serines. We compared these mutants to SMO lacking the entire CRD (SMO-ΔCRD) (Fig. 4d and Extended Data Fig. 7a). Untagged SMO-ΔCRD stably expressed in $Smo^{-/-}$ fibroblasts demonstrated a similar level of constitutive activity[11,12] to that of wild-type SMO stimulated with saturating SHH (Fig. 4d and Extended Data Fig. 7b).

The Pro120Ser, Ile160Asn/Glu162Thr, and Cys197Ser/Cys217Ser mutations mimicked the effect of 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)-yne (Fig. 4d). All three mutants remained sensitive to inhibition by SANT-1, which binds deep in the TMD and does not contact the linker domain[16]. Cyclopamine and SAG, both of

which make contacts with the linker domain[16,17], could regulate SMO-Pro120Ser and SMO-Ile160Asn/Glu162Thr, but not SMO-Cys197Ser/Cys217Ser. Thus, destabilization of either CRD–linker domain contacts or the linker domain itself increases the constitutive activity of SMO, implicating these regions in stabilizing an inactive state, analogous to the D(E)R$^{3.50}$Y motif in helix III of some Class A GPCRs[26].

To understand the basis for the loss of oxysterol responsiveness, we measured binding of each of the mutants to 20(S)-OHC beads (Extended Data Fig. 7c). Although the CRD mutations Pro120Ser and Ile160Asn/Glu162Thr impaired binding, the Cys197Ser/Cys217Ser mutation in the linker domain 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 linker domain may transmit the conformational changes that lead to SMO TMD activation in response to CRD ligands. To investigate conformational changes in SMO induced by CRD ligands in the solution state, we performed small angle X-ray scattering (SAXS) experiments on SMOΔC in the absence of any exogenous ligands (*apo*-SMOΔC) or SMOΔC loaded with the agonist 20(S)-OHC (Extended Data Fig. 8a). The initial SAXS curves (Fig. 4e) and further analyses[27,28] (Fig. 4f and Extended Data Fig. 8b) show that 20(S)-OHC binding induces a conformational change consistent with elongation and reduced globularity of SMOΔC. As 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–linker domain–TMD contacts in controlling the conformation and activity of SMO. It is unclear whether these interactions are altered by TMD-targeted small molecules used to treat patients with Hh-driven cancers, 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*–SMOΔC; Fig. 5a and Extended Data Fig. 9a, b). 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 stabilized by a network of hydrogen bonds and hydrophobic interactions formed by the SMO TMD core (Fig. 5b). As expected, vismodegib occupies the TMD binding site previously identified in SMO complexes with other antagonists and agonists[15–17] (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, which is stabilized by a potential hydrogen bond to the side chain of Asp384, occupies a central position within the seven transmembrane domain (7TM) bundle and connects to the chlorophenyl–methylsulfone moiety, which is oriented towards the extracellular domains and the entrance of the TMD pocket[16]. This arrangement is stabilized by 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 that is mutated in vismodegib-resistant cancers[14,29,30], stabilizes the potential hydrogen-bonding network around Arg400 (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 drug[14]. Our structure also explains the vismodegib resistance mechanism for many other mutations in the TMD binding site observed in patients with advanced basal cell cancer[29,30] (Extended Data Fig. 9c–f).
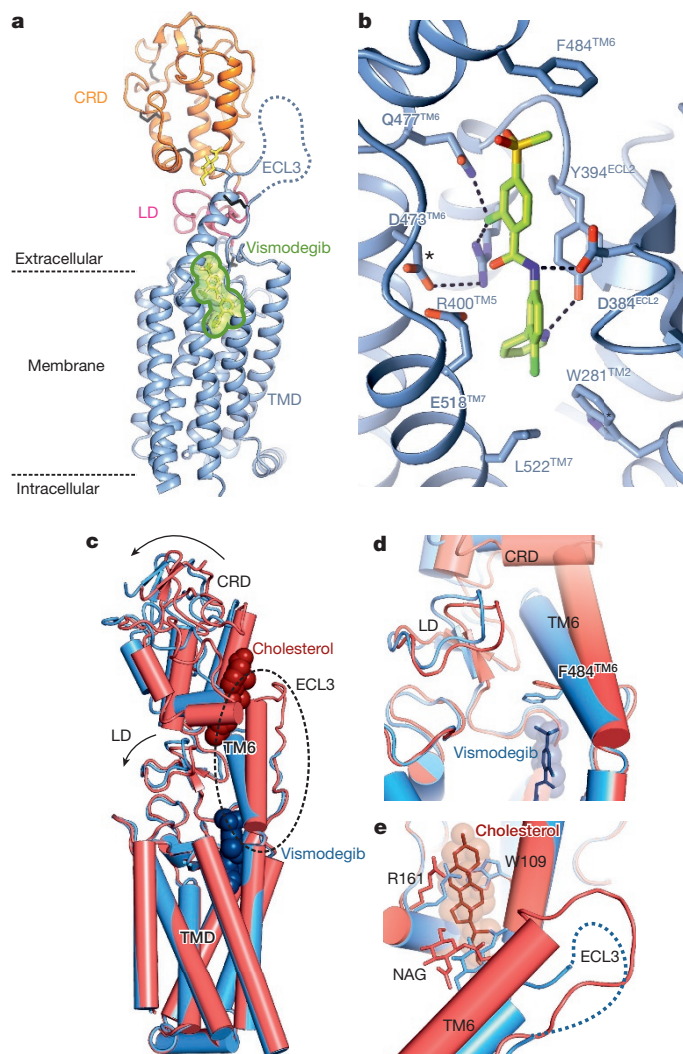
Figure 5 | Structure of SMO in complex with the antagonist vismodegib.
a, Overall structure showing full extracellular and transmembrane domains of human SMO (cartoon) in complex with vismodegib (green). Colour-coding follows Fig. 1. **b**, Close-up of the vismodegib-binding site. The asterisk denotes 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. The dashed oval highlights the conformational change of the TM6 and ECL3. **d**, Close-up of the TMD–linker domain–CRD interface. Phe484 contacts the vismodegib methylsulfone group. **e**, Close-up of the sterol-binding site. Cholesterol, Arg161 and NAG shown as sticks. The partly disordered ECL3 loop of *vismo*–SMO is depicted as dotted line.

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 transmembrane helix 6, which contacts the CRD and linker domain, is moved ~15° towards the linker domain and CRD, probably owing to an interaction between Phe484 and the methylsulfone moiety of vismodegib (Fig. 5b, d). This movement of helix 6 results in a rotational movement of the CRD (and to a lesser extent the linker domain), 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 reorganization 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 SMOΔC with cholesterol beads (Extended Data

Fig. 9g, h). In summary, binding of vismodegib to the TMD site induces a conformational change that ultimately results in rearrangement of the sterol-binding site, providing a structural communication mechanism that explains the previously observed allosteric interaction[7] between CRD and TMD ligands. More generally, this finding 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 unanticipated role for the CRD and ECL3, including displacement of an extracellular cholesterol ligand, in stabilizing 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 stabilizes 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 such as vismodegib that act allosterically to occlude the cholesterol-binding site. SAXS data suggest that CRD ligands such as 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, linker domain, and TMD that allow the TMD to adopt an active signaling state.

1. Sharpe, H. J., Wang, W., Hannoush, R. N. & de Sauvage, F. J. Regulation of the oncoprotein Smoothened by small molecules. *Nat. Chem. Biol.* **11,** 246–255 (2015).
2. Taipale, J. *et al.* Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* **406,** 1005–1009 (2000).
3. Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* **16,** 2743–2748 (2002).
4. Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. Small molecule modulation of Smoothened activity. *Proc. Natl Acad. Sci. USA* **99,** 14071–14076 (2002).
5. Frank-Kamenetsky, M. *et al.* Small-molecule modulators of Hedgehog signaling: identification and characterization of Smoothened agonists and antagonists. *J. Biol.* **1,** 10 (2002).
6. Robarge, K. D. *et al.* GDC-0449-a potent inhibitor of the hedgehog pathway. *Bioorg. Med. Chem. Lett.* **19,** 5576–5581 (2009).
7. Nachtergaele, S. *et al.* Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nat. Chem. Biol.* **8,** 211–220 (2012).
8. Corcoran, R. B. & Scott, M. P. Oxysterols stimulate Sonic hedgehog signal transduction and proliferation of medulloblastoma cells. *Proc. Natl Acad. Sci. USA* **103,** 8408–8413 (2006).
9. Dwyer, J. R. *et al.* Oxysterols are novel activators of the hedgehog signaling pathway in pluripotent mesenchymal cells. *J. Biol. Chem.* **282,** 8959–8968 (2007).
10. Nachtergaele, S. *et al.* Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. *eLife* **2,** e01340 (2013).
11. Nedelcu, D., Liu, J., Xu, Y., Jao, C. & Salic, A. Oxysterol binding to the extracellular domain of Smoothened in Hedgehog signaling. *Nat. Chem. Biol.* **9,** 557–564 (2013).
12. Myers, B. R. *et al.* Hedgehog pathway modulation by multiple lipid binding sites on the smoothened effector of signal response. *Dev. Cell* **26,** 346–357 (2013).
13. Briscoe, J. & Thérond, P. P. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* **14,** 416–429 (2013).
14. Yauch, R. L. *et al.* Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science* **326,** 572–574 (2009).

15. Wang, C. *et al.* Structure of the human smoothened receptor bound to an antitumour agent. *Nature* **497,** 338–343 (2013).
16. Wang, C. *et al.* Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nat. Commun.* **5,** 4355 (2014).
17. Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5,** 3309 (2014).
18. Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* **21,** 541–551 (2011).
19. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53,** 531–556 (2013).
20. Rana, R. *et al.* Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling. *Nat. Commun.* **4,** 2965 (2013).
21. Chun, E. *et al.* Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20,** 967–976 (2012).
22. Bazan, J. F. & de Sauvage, F. J. Structural ties between cholesterol transport and morphogen signaling. *Cell* **138,** 1055–1056 (2009).
23. Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science* **337,** 59–64 (2012).
24. Cooper, M. K. *et al.* A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis. *Nat. Genet.* **33,** 508–513 (2003).
25. Blassberg, R., Macrae, J. I., Briscoe, J. & Jacob, J. Reduced cholesterol levels impair Smoothened activation in Smith-Lemli-Opitz syndrome. *Hum. Mol. Genet.* **25,** 693–705 (2016).
26. Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494,** 185–194 (2013).
27. Durand, D. *et al.* NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169,** 45–53 (2010).
28. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95,** 559–571 (2011).
29. Sharpe, H. J. *et al.* Genomic analysis of smoothened inhibitor resistance in basal cell carcinoma. *Cancer Cell* **27,** 327–341 (2015).
30. Atwood, S. X. *et al.* Smoothened variants explain the majority of drug resistance in basal cell carcinoma. *Cancer Cell* **27,** 342–353 (2015).

Author Contributions E.F.X.B. produced the protein with P.S.M, and carried out crystallization with S.Ne. C.S. and E.F.X.B. determined the crystal structures. R.S., S.Na., and G.L. performed Hh signalling and biochemical assays. L.M.-M. and D.F.C. synthesized sterol analogues. 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*-SMOΔC and *vismo*–SMOΔC crystal structures have been deposited in the Protein Data Bank (PDB) under accession numbers 5L7D and 5L7I. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.R. (rrohatgi@stanford.edu) or C.S. (christian@strubi.ox.ac.uk).

Reviewer Information *Nature* thanks J. Briscoe, R. Dror, F. de Sauvage and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Reagents. NIH 3T3 and 293T cells were certified stocks obtained directly from ATCC. $Smo^{-/-}$ fibroblasts (which were used to express all the SMO mutants) have been described previously[31] and were originally obtained from J. Chen and P. Beachy. The authenticity of $Smo^{-/-}$ cells was established by immunoblotting to ensure lack of endogenous SMO protein expression. Incoming cell lines were confirmed to be negative for mycoplasma contamination. SAG was obtained from Enzo Life Sciences; cyclopamine was obtained from Toronto Research Chemicals; SANT-1 was obtained from EMD Millipore; 20(S)-OHC was obtained from Steraloids. The synthesis of 20(S)-yne, 20(R)-OHC[7] and 20(S)-amine-coupled beads for SMO binding assays has been described in detail previously[10]. Rabbit polyclonal antibodies against SMO, PTCH1 and SUFU have been described previously[32,33], and the mouse monoclonal antibody against GLI1 was obtained from Cell Signaling Technologies (L42B10). SHH-containing conditioned medium 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 (SMOΔC) 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 SMOΔC was obtained from Geneart (Grand Island, NY) and cloned into the pHLSec vector[34] in frame to either a C-terminal Rho1D4 antibody epitope tag[35,36] or a monoVenus tag[37,38] followed by a Rho1D4 antibody epitope tag.

All mouse SMO mutants were made using Quikchange methods using a previously described construct[39] encoding mouse SMO (pCS2+:YFP-mSmo), after removal of the N-terminal YFP tag by XhoI digestion. The mouse SMO-ΔCRD construct lacks residues 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. SMOΔC 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% CO_2, 130 r.p.m. to densities of 2–3×10^6 cells ml⁻¹ in protein expression medium (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 r.p.m., 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 r.p.m. Culture media were subsequently topped up to 0.8 l with PEM and returned to 130 r.p.m. 48–72 h after transfection, cell pellets were collected by centrifugation, snap-frozen in liquid N_2 and stored at –80 °C, resulting in a total of ~150 g of cell mass per 9.6 l suspension medium[40].

Frozen cell pellets were thawed, re-suspended in 10 mM HEPES pH 7.5, 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 r.p.m., 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 pH 7.5, 300 mM NaCl, 10% (v/v) glycerol, 0.1% DDM, 0.02% CHS buffer and then with 50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 0.05% DDM, 0.01% CHS buffer and eluted overnight in 50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS, 500 μM TETSQVAPA peptide (Genscript). Eluate was concentrated to ~500 μl 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 pH 7.5, 150 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS. Peak fractions were pooled and concentrated to ~30 mg ml⁻¹ 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[37,38], 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)[41].

For thermostability experiments, SMOΔC 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 pH 7.5, 0.03% DDM and elution with 300 mM NaCl, 50 mM Hepes, pH 7.5, 0.03% DDM, 500 μM TETSQVAPA peptide. The thermostability of the different samples was assessed by heating aliquots of their eluates to the indicated temperatures for 1 h before loading them onto a Superose 6 3.2/300 column on a Shimadzu system and following 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 pMSCVpuro[39]. Retroviral supernatants were produced after transient transfection of Bosc23 helper cells with the pMSCV constructs[42,43]. Virus-containing media from these transfections was directly used to infect $Smo^{-/-}$ fibroblasts, and stable integrants were selected with puromycin (2 μg ml⁻¹).

We had previously[10] 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 a fluorescent protein, or transient overexpression of tagged or untagged SMO, could impact the assessment of its signalling activity. For example, we previously measured a lower level of constitutive signalling activity for SMO-ΔCRD compared to the present study, probably due to the presence of an N-terminal YFP tag[10]. Moreover, previous reports from three groups (including our own) reached somewhat divergent conclusions regarding the role of the CRD in basal and ligand-stimulated SMO activity[10–12]. 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 wild-type 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 (20,000g, 30 min, 4 °C), the protein concentration of each lysate was measured using the bicinchoninic acid assay (BCA, Pierce/Thermo Scientific). Lysate aliquots containing equal amounts of total protein were fractionated on SDS–PAGE gels (either an 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.

Gli1 mRNA is a commonly used metric for Hh signalling activity, because *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′-AGCCCGCTTCTTTGTTAATTTGA-3′) and *Gapdh* (forward primer: 5′-AGTGGCAAAGTGGAGATT-3′ and reverse primer: 5′-GTGGAGTCATACTGGAACA-3′). Transcript levels relative to Gapdh were calculated using the ΔC_t method and reported in arbitrary units. 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, $P > 0.05$; *$P \le 0.05$, **$P \le 0.01$, ***$P \le 0.001$, ****$P \le 0.0001$.

Ligand affinity chromatography. Ligand affinity chromatography to assess the interaction between SMO protein in detergent extracts and beads covalently coupled to 20(S)-amine has been described previously[10]. 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 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 antibody[32] and infrared imaging (Li-Cor Odyssey). Each experiment was repeated twice with similar results.

Preparation of cholesterol–PEG3–sepharose using the azide-alkyne Huisgen cycloaddition reaction from the Click Chemistry toolbox. 100 µl packed PEG3-azide sepharose resin (22 µmole per ml, Click Chemistry Tools) was washed 3 × 1 ml with 20% ethanol (v/v aq.) and re-suspended in 500 µl 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 µmoles (high) of LKM-26, a previously described[44,45] 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 1:1,000. 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 µl. Each reaction contained 1 µg purified SMOΔC, the same protein used for crystallization studies. Competitors were added to the binding reaction and incubated for 1 h at room temperature before the addition of 20 µl packed cholesterol–PEG3–sepharose. Reactions were incubated for 16 h at 4 °C for affinity capture. The binding reactions were washed 3 × 1 ml with binding buffer and eluted using 2× Laemmli buffer for 30 min at room temperature. SMOΔC 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 1:2,000). Each experiment was repeated twice with similar results.

Crystallization and data collection. Protein samples were reconstituted into lipidic cubic phase (LCP) by mixing with molten lipid in a mechanical syringe mixer[46]. Molten lipid, consisting of 10% cholesterol (Sigma-Aldrich) and 90% 9.9 monoacylglycerol (monoolein, Sigma-Aldrich), was mixed with detergent-solubilized protein (either *apo* or with 10 mM vismodegib in DMSO) at ~30 mg ml⁻¹ 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 µl of precipitant solution onto each of 96 positions on a siliconized glass plate, which were 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-polarizers, and subsequently imaged using a UV-imaging system (Rigaku Minstrel). *apo*-SMOΔC 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*-SMOΔC 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 (refs 47, 48), scaled using XSCALE[48] and merged using Aimless[49,50]. The final data set used for structure solution and refinement was merged from data from nine crystals for *apo*-SMOΔC and two crystals for *vismo*-SMOΔC. Data collection statistics are shown in Extended Data Table 1.

Structure determination, refinement and analysis. The *apo*-SMOΔC structure was solved by molecular replacement in PHASER[51] using the structure of human SMO TMD (PDB 4QIM[16]), zebrafish SMO CRD (PDB 4C79[10]) and BRIL (PDB 4EIY[52]) as search models. Extra electron density accounting for the region between CRD and LD, BRIL and TMD was immediately discernible after density modification in PARROT[53] (Extended Data Fig. 3e). We also observed 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*-SMOΔC polypeptide chain was traced using iterative rounds of BUCCANEER[54], manual building in COOT[55] and refinement in autoBUSTER[56] and PHENIX[57]. This resulted in a well-defined model for the *apo*-SMOΔC structure that included two molecules of SMO (residues 59–549) with a BRIL protein segment inserted between SMO TMD helices 5 and 6, two *N*-linked glycans and a cholesterol molecule. 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 by crystal non-isomorphy or pseudo-symmetry, because reducing the space group from *C*2 to *P*1 had no effect on the disorder. The *vismo*–SMOΔC structure was solved by molecular replacement using the *apo*-SMOΔC structure. Extra electron density accounting for vismodegib and two well-ordered monoolein molecules was immediately apparent. The structure was refined using autoBUSTER[56] and PHENIX[57] with non crystallographic and secondary structure restraints. Crystallographic and Ramachandran statistics are given in Extended Data Table 1. Stereochemical properties were assessed by MOLPROBITY[58]. Superpositions were calculated using the program COOT[55], electrostatic potentials were calculated using APBS[59] and hydrophobicity was calculated according to the Eisenberg hydrophobicity scale[60], as implemented in PyMOL[61]. Buried surface areas of protein–protein interfaces were calculated using the PISA webserver[62] with a probe radius of 1.4 Å. Sequence alignment was performed using MULTALIN[63] and formatted with ESPRIPT[64]. Program Caver was used with default settings to visualize the SMO TMD ligand binding pocket[65].

Amphipol exchange and MALS. In order to avoid background light-scattering due to free detergent in solution, protein samples were exchanged into amphipol[66](A8-35, Anatrace) at a mass ratio of 3:1 amphipol:protein and rotated at room temperature for 30 min. BioBeads (BIORAD), equilibrated in detergent-free buffer (10 mM HEPES pH 7.5, 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-solubilized protein at 1 mg ml⁻¹ 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 SMOΔC, the values used for d*n*/d*c* and ε at 280 nm were 0.185 ml g⁻¹ and 1.541 ml (mg.cm)⁻¹, respectively. For protein conjugate analysis, the d*n*/d*c* used for amphipol A8-35 was 0.15 ml g⁻¹ (ref. 67).

Molecular dynamics system setup. Simulations were performed using the GROMACS v4.6.3 simulation package[68]. Side chain ionization states were modelled using pdb2gmx (Histidine) and PropK_a (all other residues)[69,70]. 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 4N4W (ref. 16). The protein structure was then energy-minimized using the steepest descents algorithm implemented in GROMACS, before being converted to a coarse-grained representation using the MARTINI 2.2 force field[71]. The energy-minimized coarse-grained 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 coarse-grained simulation was applied to permit the self-assembly[72] of a POPC lipid bilayer around the GPCR. During the coarse-grained simulation, the structure of the protein was maintained by an elastic network, allowing local conformational flexibility of the protein. Thus, the protein was able to adopt its optimal orientation within the lipid bilayer[73]. The endpoint of the coarse-grained bilayer self-assembly simulation was converted back to atomic detail using a fragment-based protocol for the lipid conformations[74], while 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 through 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 field[75] and its extension to proteins[71,76] was used to describe all system components. During the coarse-grained self-assembly simulation an ELNEDYN network[77] was applied to the protein using force constant of 500 kJ mol⁻¹ nm⁻² and a cutoff of 1.5 nm. Temperature was maintained at 310 K using a Berendsen thermostat[78] with a coupling constant of τ_t = 1 ps, and pressure was controlled at 1 bar using a Berendsen barostat[78] with a coupling constant of τ_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 protocol[75]. An integration time step of 20 fs was applied. Covalent bonds were constrained to their equilibrium values using the LINCS

algorithm[79]. All simulations were run in the presence of standard MARTINI water particles[75], and ions added to an approximate concentration of 0.15 M NaCl.

Atomistic simulations. Atomistic simulations were run using the GROMOS53a6 force field[80], and its extension to glycans[81]. 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 fs. Temperature was maintained at 310 K using a V-rescale thermostat[82] with a coupling constant of 0.1 ps, while pressure was controlled at 1 bar through coupling to a Parrinello–Rahman barostat[83] with a coupling constant of 1 ps. Particle Mesh Ewald (PME)[84] was applied to model long-range electrostatics. The LINCS algorithm was used to constrain covalent bond lengths[79]. The g_dist, g_rmsf and g_rms tools implemented in the GROMACS v4.6.3 software package[68] were applied to analyse the simulations, with VMD[85] and PyMOL[61] used for visualization. Cα r.m.s.d. calculations for the CRD were performed after a least-squares fit of the trajectory to Cα particles of the initial (crystal structure) CRD coordinates. Cα r.m.s.d. calculations for the TMD were performed after a least squares fit to Cα 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 GROMACS[68].

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) SMOΔC 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 was injected at 9.6 (20(S)-OHC-treated) or 13 (apo) mg ml^{-1} using a flow rate of 160 μl per min. 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 at a wavelength of 1 Å. SAXS images (frames) were collected as a continuous set of 3 s exposures across the elution peak. The corresponding buffer background frames for producing the background-subtracted SAXS curve was collected at greater than 1.5 column volumes.

Images were corrected for variations in beam current, normalized 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 the SAXS curves ($q > 0.05$ Å^{-1}) of the treated and untreated samples 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.

31. Varjosalo, M., Li, S. P. & Taipale, J. Divergence of hedgehog signal transduction mechanism between *Drosophila* and mammals. *Dev. Cell* **10,** 177–186 (2006).
32. Rohatgi, R., Milenkovic, L. & Scott, M. P. Patched1 regulates hedgehog signaling at the primary cilium. *Science* **317,** 372–376 (2007).
33. Humke, E. W., Dorn, K. V., Milenkovic, L., Scott, M. P. & Rohatgi, R. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes Dev.* **24,** 670–682 (2010). 10.1101/gad.1902910
34. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1243–1250 (2006).
35. Molday, R. S. & MacKenzie, D. Monoclonal antibodies to rhodopsin: characterization, cross-reactivity, and application as structural probes. *Biochemistry* **22,** 653–660 (1983).
36. Oprian, D. D., Molday, R. S., Kaufman, R. J. & Khorana, H. G. Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proc. Natl Acad. Sci. USA* **84,** 8874–8878 (1987).
37. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296,** 913–916 (2002).
38. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20,** 87–90 (2002).
39. Rohatgi, R., Milenkovic, L., Corcoran, R. B. & Scott, M. P. Hedgehog signal transduction by Smoothened: pharmacologic evidence for a 2-step activation process. *Proc. Natl Acad. Sci. USA* **106,** 3196–3201 (2009).
40. Miller, P. S. & Aricescu, A. R. Crystal structure of a human GABA$_A$ receptor. *Nature* **512,** 270–275 (2014).
41. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14,** 673–681 (2006).
42. Bailey, E. C., Milenkovic, L., Scott, M. P., Collawn, J. F. & Johnson, R. L. Several PATCHED1 missense mutations display activity in patched1-deficient fibroblasts. *J. Biol. Chem.* **277,** 33632–33640 (2002).
43. Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. USA* **90,** 8392–8396 (1993).
44. Ciepla, P. *et al.* New chemical probes targeting cholesterylation of Sonic Hedgehog in human cells and zebrafish. *Chem. Sci. (Camb.)* **5,** 4249–4259 (2014).
45. Chung, S.-K., Shim, J.-Y., Kang, M. G., Lee, K. W. & Kang, H. I. Studies of steroids as potential antifungal agent 2. Side chain modified cholesterols. *Korean J. Med. Chem.* **8,** 14–17 (1998).
46. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protocols* **4,** 706–731 (2009).
47. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43,** 186–190 (2010).
48. Kabsch, W. Xds. *Acta Crystallogr. D* **66,** 125–132 (2010).
49. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69,** 1204–1214 (2013).
50. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67,** 235–242 (2011).
51. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
52. Liu, W. *et al.* Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337,** 232–236 (2012).
53. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 470–478 (2010).
54. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62,** 1002–1011 (2006).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
56. BUSTER v. 2.10.2 (Global Phasing Ltd., Cambridge, United Kingdom, 2011).
57. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
58. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
59. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98,** 10037–10041 (2001).
60. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179,** 125–142 (1984).
61. Schrodinger, L.L.C. The PyMOL Molecular Graphics System, Version 1.3r1 (2010).
62. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
63. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16,** 10881–10890 (1988).
64. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–W324 (2014).
65. Chovancova, E. *et al.* CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLOS Comput. Biol.* **8,** e1002708 (2012).
66. Tribet, C., Audebert, R. & Popot, J. L. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl Acad. Sci. USA* **93,** 15047–15050 (1996).
67. Gohon, Y. *et al.* Bacteriorhodopsin/amphipol complexes: structural and functional properties. *Biophys. J.* **94,** 3523–3537 (2008).
68. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4,** 435–447 (2008).
69. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).
70. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7,** 2284–2295 (2011).
71. de Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9,** 687–697 (2013).
72. Scott, K. A. *et al.* Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* **16,** 621–630 (2008).
73. Stansfeld, P. J. *et al.* MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **23,** 1350–1361 (2015).
74. Stansfeld, P. J. & Sansom, M. S. Molecular simulation approaches to membrane proteins. *Structure* **19,** 1562–1572 (2011).
75. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111,** 7812–7824 (2007).
76. Monticelli, L., Sorin, E. J., Tieleman, D. P., Pande, V. S. & Colombo, G. Molecular simulation of multistate peptide dynamics: a comparison between microsecond timescale sampling and multiple shorter trajectories. *J. Comput. Chem.* **29,** 1740–1752 (2008).
77. Periole, X., Cavalli, M., Marrink, S. J. & Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **5,** 2531–2543 (2009).

78. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).
79. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).
80. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25,** 1656–1676 (2004).
81. Pol-Fachin, L., Verli, H. & Lins, R. D. Extension and validation of the GROMOS 53A6(GLYC) parameter set for glycoproteins. *J. Comput. Chem.* **35,** 2087–2095 (2014).
82. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).
83. Parrinello, M. & Rahman, A. Polymorphic transitions in single-crystals — a new molecular-dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).
84. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).
85. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
86. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).
87. Zoonens, M. & Popot, J. L. Amphipols for each season. *J. Membr. Biol.* **247,** 759–796 (2014).
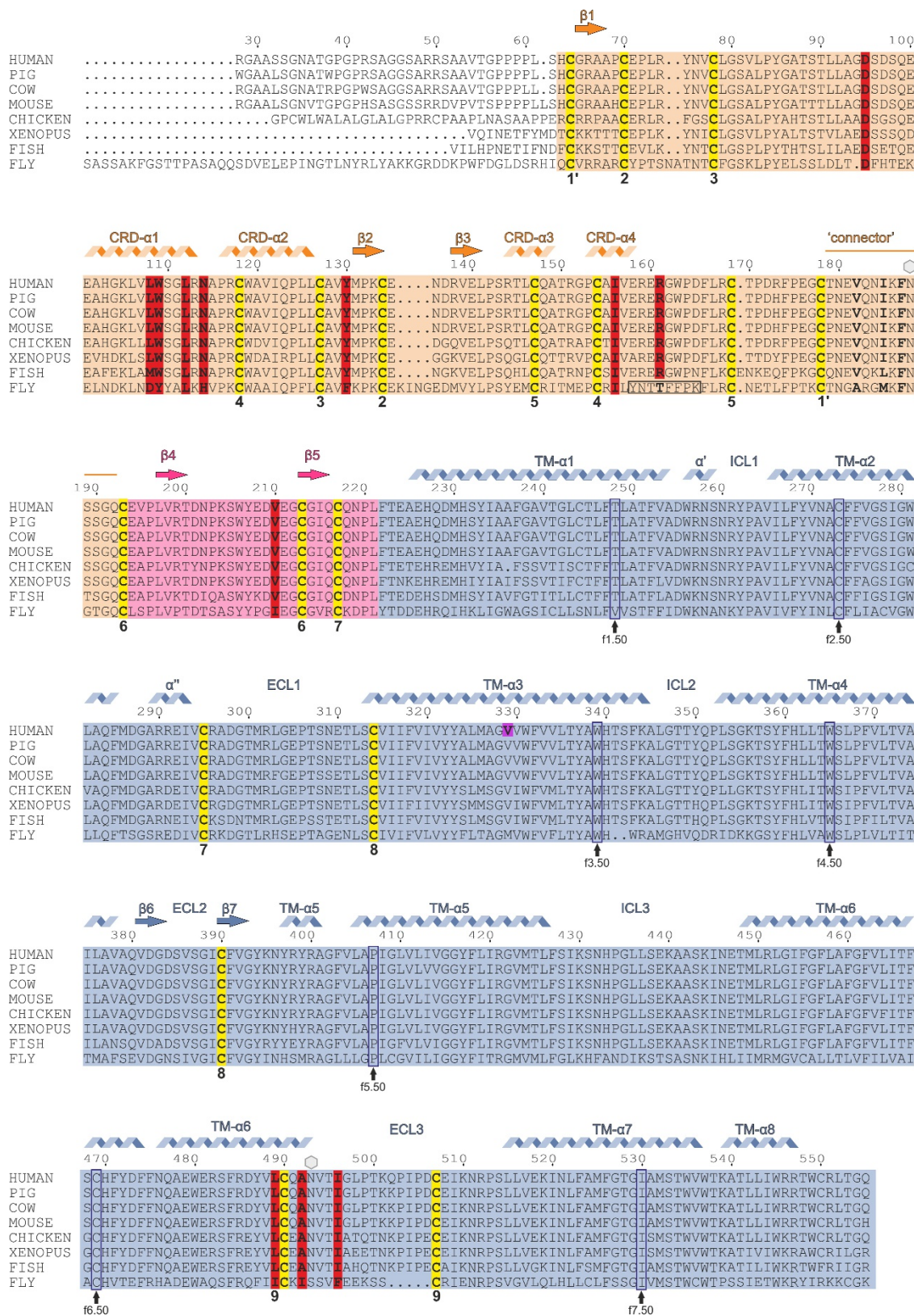
Extended Data Figure 1 | Sequence alignment of SMO orthologues.
Numbering corresponds to that of human SMO. Secondary structure
assignments are displayed above the alignment and colour-coded as in
Fig. 1. Black arrows and numbers (fX.50) below alignment show class
F Ballesteros–Weinstein nomenclature for GPCR helices[16]. Residues
interacting with cholesterol are highlighted in red. Disulfide bridges are
highlighted in yellow and numbered. N-linked glycosylation sites are
depicted by a hexagon. The position of the Val329Phe point mutation is
highlighted in purple.
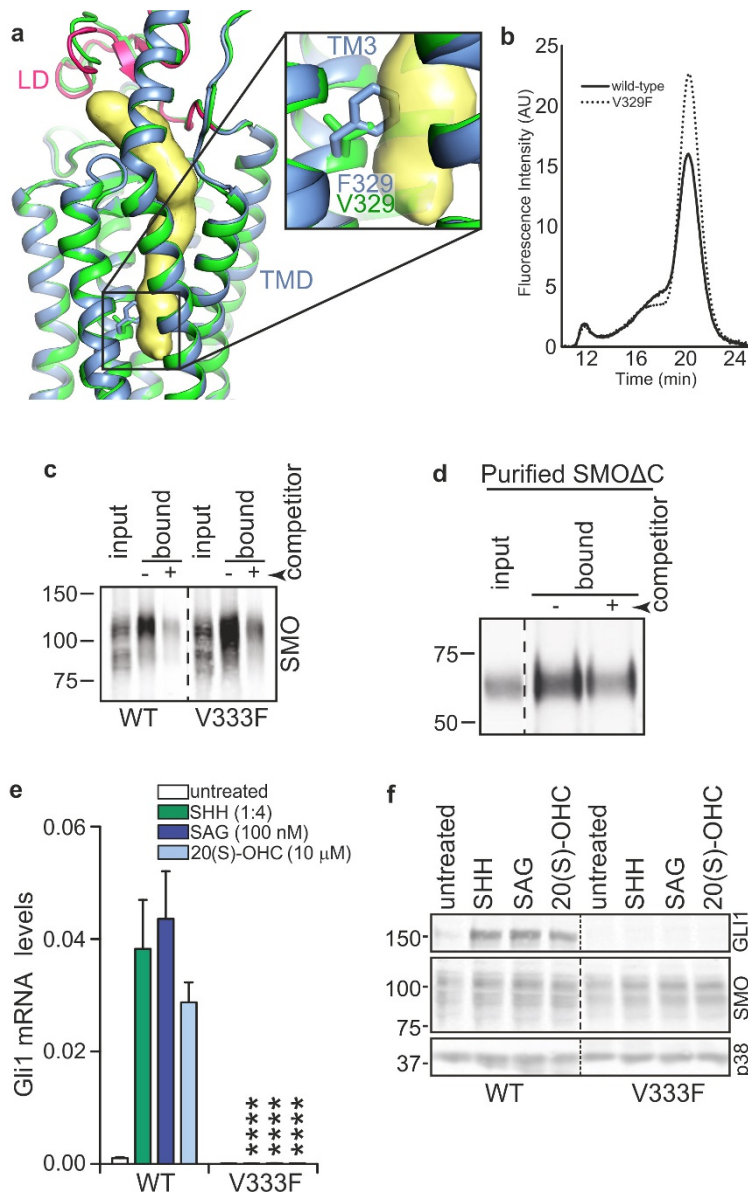
Extended Data Figure 2 | Characterization of the SMO Val329Phe mutation. a, Superposition of SMOΔC structure (blue) with the SMO-SANT-1 complex structure, which lacks the CRD (green, PDB 4N4W (ref. 16)), showing the TMD ligand-binding pocket as a yellow surface. Inset shows Val329, mutated to Phe in our structure. **b**, SEC analysis of fluorescently labelled SMOΔC 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 wild-type 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 µM free 20(S)-OHC as a competitor reduced

binding. **d**, Purified human SMOΔC (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 arbitrary units ± s.d., $n = 4$), measured by qRT–PCR, were used as a metric of Hh pathway activity because $Gli1$ is a direct Hh target. Asterisks indicate statistical significance (****$P ≤ 0.0001$) based on one-way ANOVA 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. Each experiment was replicated ≥2 times with similar results.
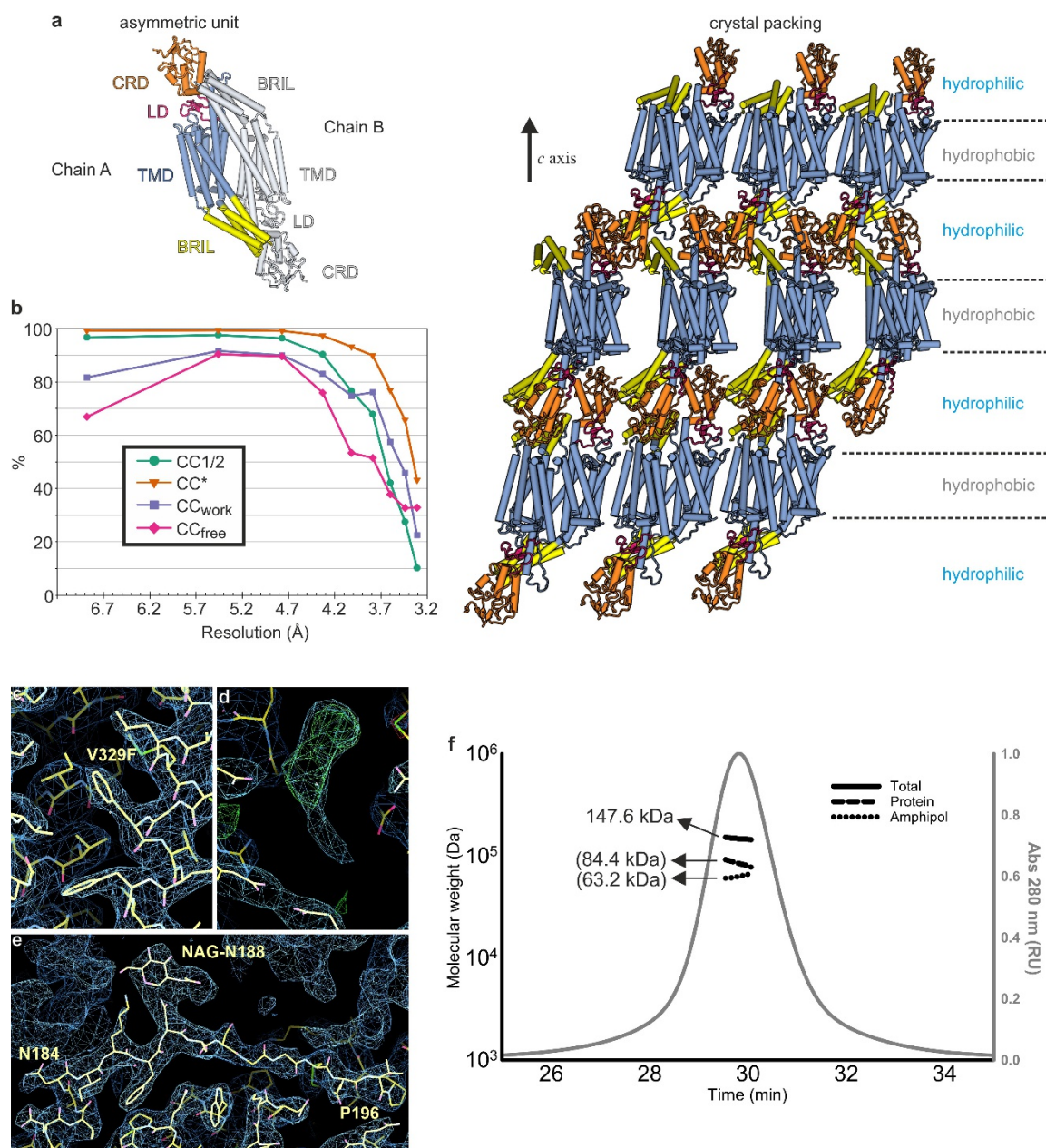
Extended Data Figure 3 | Crystallization, structure solution and oligomeric state of SMOΔC. a, SMOΔC crystal packing. Asymmetric unit consists of two antiparallel SMOΔC chains. Chain A coloured as in Fig. 1 with BRIL fusion in yellow; Chain B in grey. LCP crystal packing with alternating hydrophobic and hydrophilic layers perpendicular to the c axis. Molecules coloured as for Chain A. **b**, Pearson correlation coefficient (CC) analysis[86] 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 $2F_o–F_c$ electron density maps of final refinement at 1.0σ contour level. **c**, Val329Phe mutation. **d**, Extra density within TMD ligand-binding pocket ($F_o–F_c$ maps shown at contour level of $+3\sigma$ (green) and $–3\sigma$ (red)) (This density could not be confidently assigned, probably because of low occupancy within the crystal.). **e**, 'Connector' region linking the CRD and linker domain, with Asn188 and linked N-acetyl glycosamine moiety. **f**, SEC–MALS analysis of amphipol-solubilized SMOΔC. 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 shown only around main absorption peak. Theoretical M_W of SMOΔC based on sequence is 71 kDa with the extra mass observed in MALS (~13 kDa) probably due to three N-linked glycosylation sites. Expected mass of protein-bound amphipol (A8–35) was previously determined to be 40–75 kDa (ref. 87), in agreement with our data. This analysis suggests that SMOΔC is a monomer under our purification conditions.
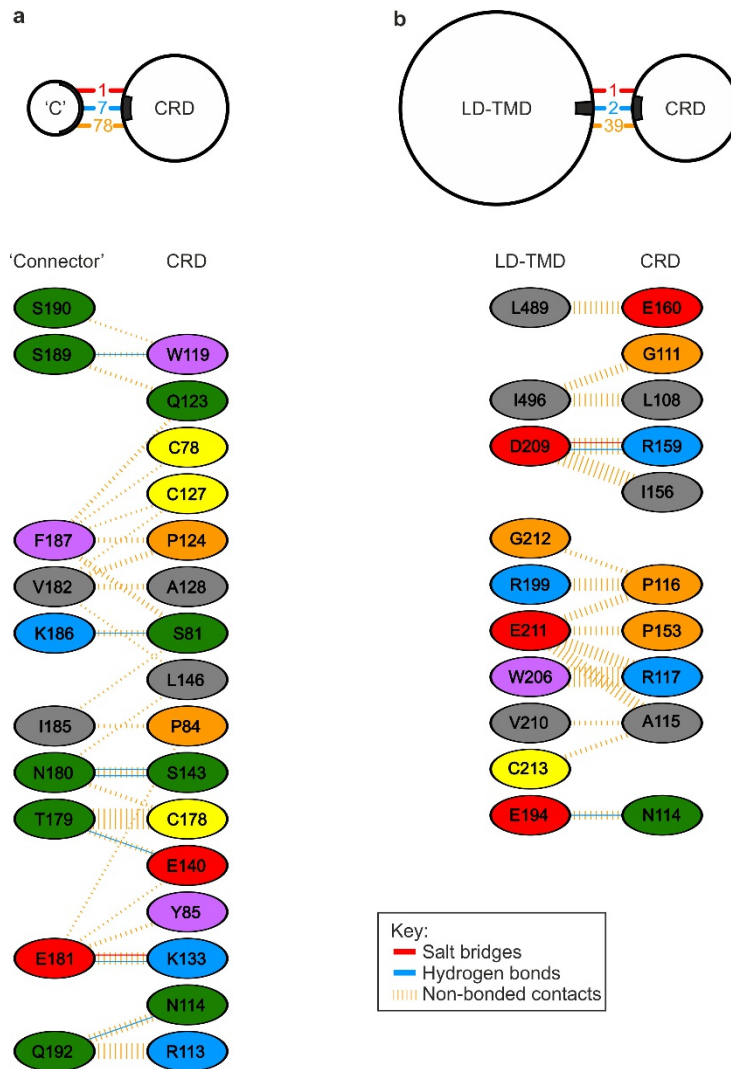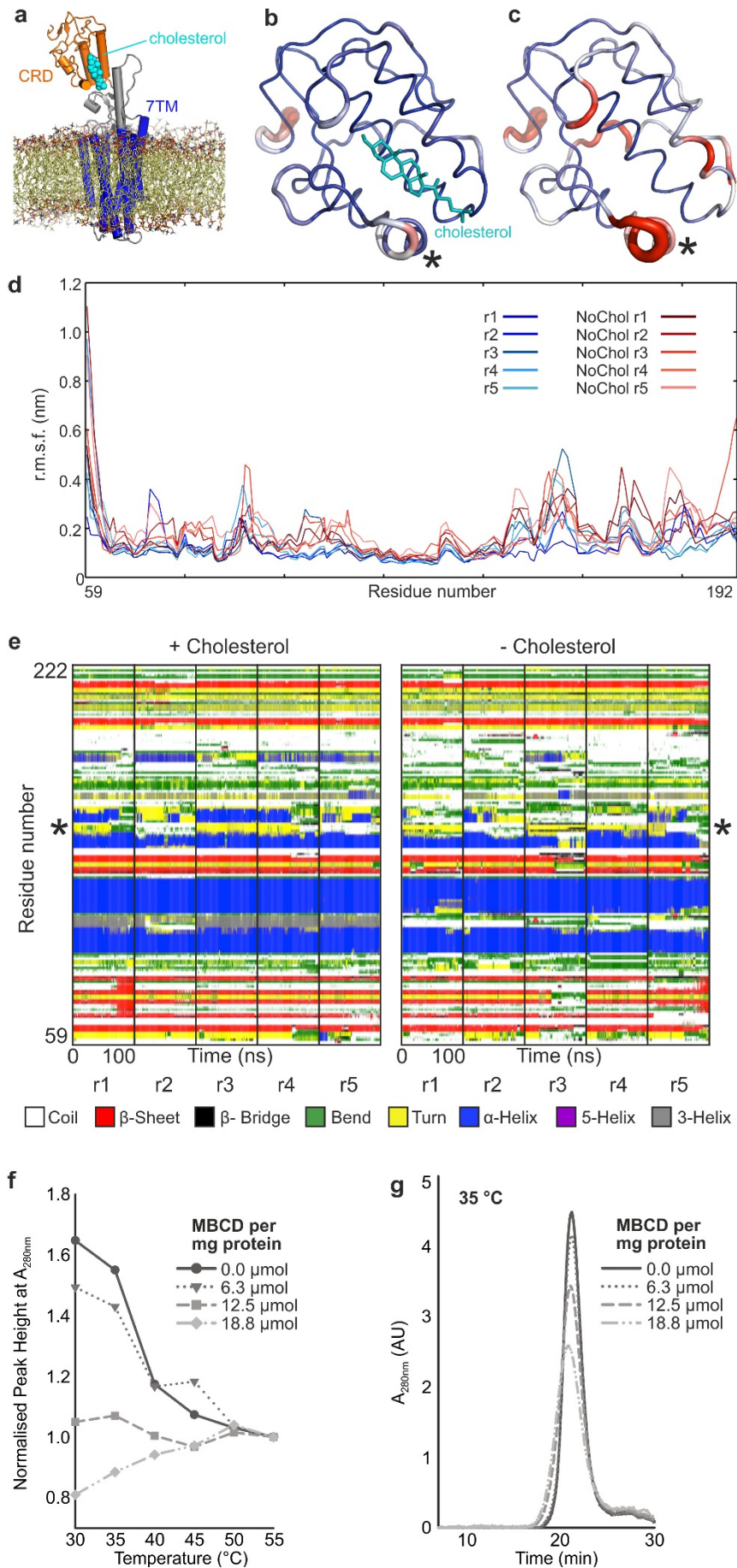
Extended Data Figure 4 | Interfaces of the SMO CRD. a, b, Detailed interactions of the CRD with the connector region (**a**) and with the linker domain–TMD segment (**b**). The number of interactions are indicated in the top panel and coloured as indicated in the key box. For non-bonded contacts, the 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 (http://www.ebi.ac.uk/pdbsum/).
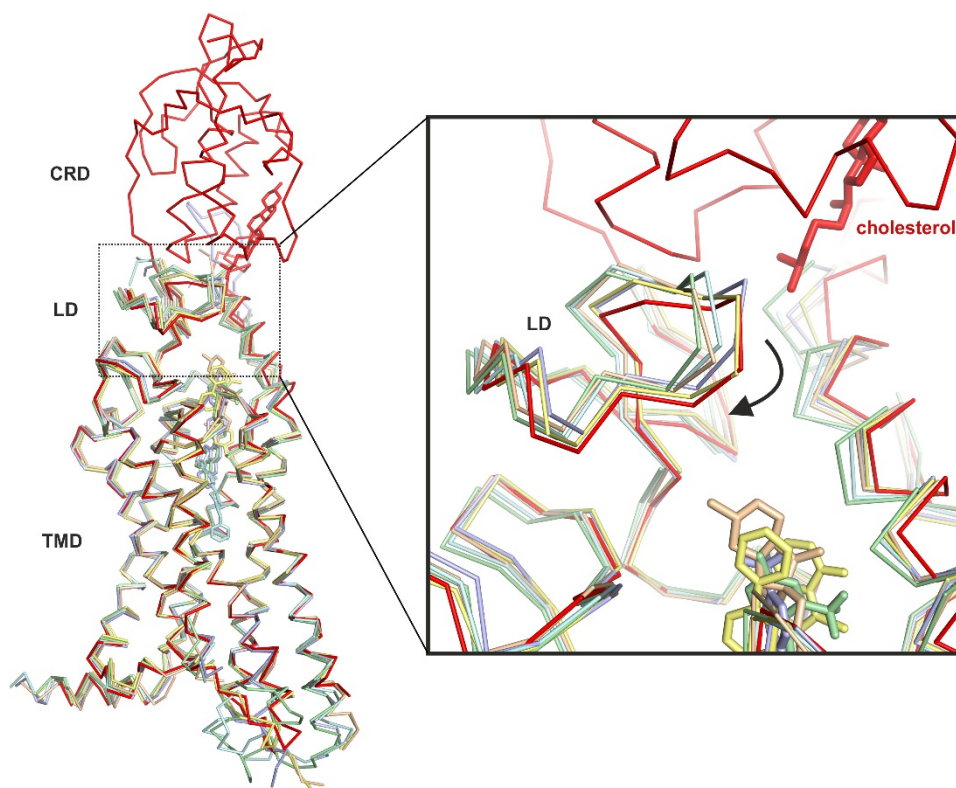
Extended Data Figure 5 | Comparison of SMOΔC with previously determined SMO TMD structures. Superposition of the SMOΔC structure with SMO TMD structures. Structural alignment was performed using the 7TM bundle as template (not including the linker domain or TMD helix 8). SMOΔC (red), SMO TMD complexed with cyclopamine (light orange, PDB 4O9R, r.m.s.d. 0.598 Å for 243 equivalent Cα positions), antaXV (light blue, PDB 4QIM, r.m.s.d. 0.515 Å for 233 equivalent Cα positions), SANT1 (pale cyan, PDB 4N4W, r.m.s.d. 0.483 Å for 240 equivalent Cα positions), LY2940680 (pale green, PDB 4JKV, r.m.s.d. 0.493 Å for 230 equivalent Cα positions), SAG1.5 (pale yellow, PDB 4QIN, r.m.s.d. 0.623 Å for 262 equivalent Cα positions). The box shows a close-up view of the linker domain region revealing a structural rearrangement in the SMOΔC structure compared to the previously determined SMO TMD structures lacking the native extracellular domain.

a CRD, cholesterol, 7TM

b cholesterol

c

d

r1, r2, r3, r4, r5, NoChol r1, NoChol r2, NoChol r3, NoChol r4, NoChol r5

r.m.s.f. (nm) vs Residue number (59 to 192)

e

+ Cholesterol   - Cholesterol

Residue number (59 to 222) vs Time (ns)

r1 r2 r3 r4 r5   r1 r2 r3 r4 r5

☐ Coil ■ β-Sheet ■ β-Bridge ■ Bend ■ Turn ■ α-Helix ■ 5-Helix ■ 3-Helix

f

Normalised Peak Height at A_{280nm} vs Temperature (°C)

MBCD per mg protein
— 0.0 μmol
⋯ 6.3 μmol
- - 12.5 μmol
-·- 18.8 μmol

g 35 °C

A_{280nm} (AU) vs Time (min)

MBCD per mg protein
— 0.0 μmol
⋯ 6.3 μmol
- - 12.5 μmol
-·- 18.8 μmol

Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Cholesterol stabilizes SMO. a–e, MD simulations of SMO in a lipid bilayer. **a**, SMO embedded in a lipid bilayer with the CRD in orange, the seven-pass transmembrane region excluding intra- and extracellular loops (7TM) in blue and cholesterol in cyan. **b–d**, Relative r.m.s. fluctuations of the Cα atoms over the course of 5 × 100 ns of atomistic MD simulation in the presence and absence of cholesterol. The structures in **b** and **c** are shown as putty representations coloured from high conformational stability (that is, low r.m.s. fluctuations; blue/thin) to low stability (that is, 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 residues 155–160, which is destabilized in the absence of bound cholesterol. **f**, **g**, Thermostability of purified SMOΔC. See Supplementary Discussion for details. **f**, Compiled peak heights from thermostability SEC analysis of purified SMOΔC 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 before loading onto the SEC column.
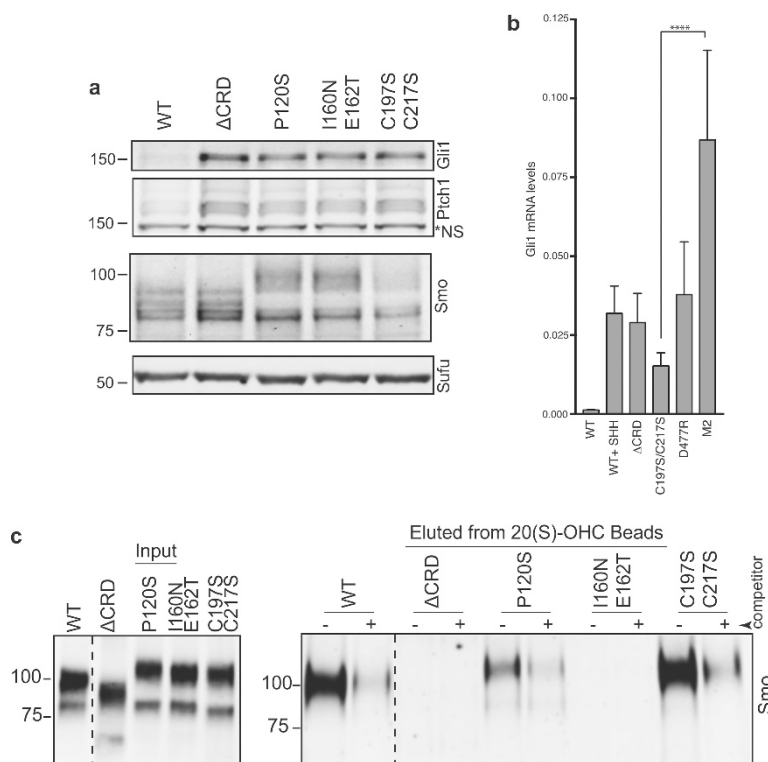
Extended Data Figure 7 | Effect of domain interface mutations on expression levels and 20(S)-OHC binding. a, Protein levels of SMO and also of PTCH1 and GLI1 (each of which is encoded by a direct Hh target gene) measured by immunoblot from $Smo^{-/-}$ mouse fibroblasts stably expressing one of five SMO 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 linker domain–CRD interface; and SMO lacking a conserved disulfide bond (Cys197Ser/Cys217Ser, marked 6 in Fig. 4c) in the linker domain. 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 are caused by different numbers of N-linked glycosylation sites. **b**, *Gli1* mRNA levels (mean arbitrary units ± s.d., $n = 3$) were used to assess Hh signalling activity in $Smo^{-/-}$ cells stably expressing the indicated mouse SMO variants. One-way ANOVA was used to assess statistical significance (****$P \leq 0.0001$). 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. Each experiment was repeated 2 or more times with similar results.
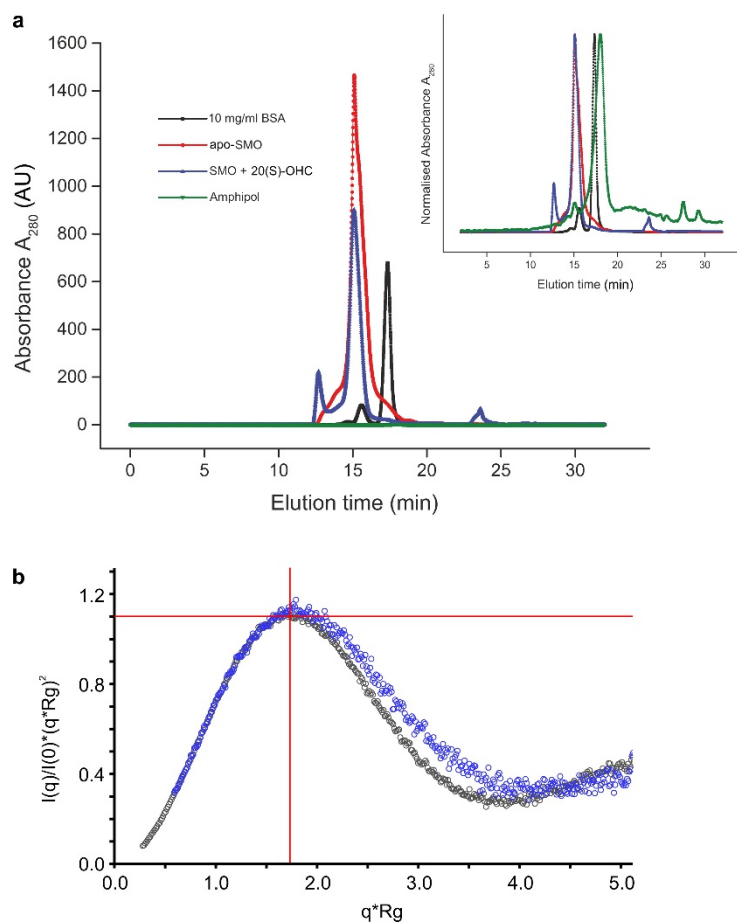
Extended Data Figure 8 | SAXS analysis of SMOΔC. a, Overlay of size-exclusion chromatograms monitored at 280 nm (A_{280}) collected during SAXS measurements for *apo*-SMOΔC (red), (+)20(S)-OHC SMOΔC (blue), amphipol (green) and BSA standard (black). Amphipol and BSA were injected at 10 mg ml^{-1}. Inset 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 min after the amphipol-stabilized SMOΔC samples. **b**, 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.

Extended Data Table 1 | Crystallographic data collection and refinement statistics

	apo-SMOΔC	*vismo*-SMOΔC
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		
$\quad$ a, b, c (Å)	122.9, 63.0, 208.6	52.8, 71.3, 107.0
$\quad$ 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/σI	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_{work}/R_{free} (%)	23.2/26.4	24.5/25.7
No. atoms		
$\quad$ Protein	9207	8988
$\quad$ Ligand/Ion	58	119
B factors (Å^2)		
$\quad$ Protein	137	124
$\quad$ Ligand/Ion	121	119
r.m.s.d. bonds (Å)	0.008	0.008
r.m.s.d. angles (°)	0.90	0.87
Ramachandran (%)		
$\quad$ Favoured	96.6	96.8
$\quad$ Allowed	3.2	2.9
$\quad$ Disallowed	0.2	0.3

Numbers in parentheses refer to the highest resolution shell. R_{free}, 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.