

ARTICLE

Received 24 Jul 2013 | Accepted 20 Nov 2013 | Published 19 Dec 2013

DOI: 10.1038/ncomms3965

# Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling

Rajashree Rana<sup>1,2</sup>, Candace E. Carroll<sup>3,†</sup>, Ho-Jin Lee<sup>1</sup>, Ju Bao<sup>1</sup>, Suresh Marada<sup>3</sup>, Christy R.R. Grace<sup>1</sup>, Cristina D. Guibao<sup>1</sup>, Stacey K. Ogden<sup>2,3</sup> & Jie J. Zheng<sup>1,2</sup>

Smoothened (Smo) is a member of the Frizzled (FzD) class of G-protein-coupled receptors (GPCRs), and functions as the key transducer in the Hedgehog (Hh) signalling pathway. Smo has an extracellular cysteine-rich domain (CRD), indispensable for its function and downstream Hh signalling. Despite its essential role, the functional contribution of the CRD to Smo signalling has not been clearly elucidated. However, given that the FzD CRD binds to the endogenous Wnt ligand, it has been proposed that the Smo CRD may bind its own endogenous ligand. Here we present the NMR solution structure of the *Drosophila* Smo CRD, and describe interactions between the glucocorticoid budesonide (Bud) and the Smo CRDs from both *Drosophila* and human. Our results highlight a function of the Smo CRD, demonstrating its role in binding to small-molecule modulators.

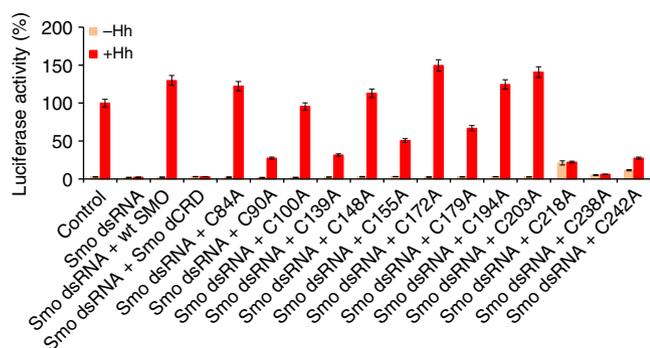
<sup>1</sup>Department of Structural Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. <sup>2</sup>Integrated Program in Biomedical Sciences, University of Tennessee Health Science Center, Memphis, Tennessee 38105, USA. <sup>3</sup>Department of Biochemistry, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. † Present address: Hardin Simmons University, Abilene, Texas, USA. Correspondence and requests for materials should be addressed to J.J.Z. (email: Jie.Zheng@stjude.org) or to S.K.O. (email: Stacey.Ogden@stjude.org).



Clone 8 (C18) cells<sup>27</sup>. Consistent with previous *in vivo* studies, we found that deletion of this domain ablates Smo signalling capacity *in vitro* (Fig. 2). To better dissect this essential functional domain, we studied the structure of the Smo CRD using NMR methods.

**Solution structure of the Smo CRD.** A recombinant *Drosophila* Smo CRD (V85-K202) was expressed and purified using the *Escherichia coli* expression system. For the purpose of NMR studies, <sup>15</sup>N- and <sup>13</sup>C-labelled protein was prepared in 10 mM deuterated acetic acid buffer and 10% D<sub>2</sub>O (volume/volume) at pH 5. Resonance assignments were performed using the standard triple resonance strategy for <sup>13</sup>C-, <sup>15</sup>N-labelled proteins<sup>28</sup>. In the two-dimensional (2D) <sup>1</sup>H-<sup>15</sup>N heteronuclear single quantum coherence (HSQC) spectra, all the amide resonances of the 118 residues of the CRD could be observed except D116. D116 is in the loop region preceding the first helix and hence could be undergoing motions in the intermediate timescale resulting in peak broadening for the backbone amide; nevertheless, all the side-chain resonances of this residue were observed. Using the structural information obtained from the NMR studies, 100 structures were calculated and the 20 structures with the lowest energy are represented as an ensemble (Fig. 3a). The statistical details for the NMR structure calculation are given in Table 1. The structure has no distance violation >0.03 Å and no angle violations >1°. The structure with the lowest energy is used to describe the secondary structure elements (Fig. 3b). The tertiary fold of the Smo CRD is very similar to that of the FzD CRD<sup>16,18</sup> (Fig. 3c, cyan compared with red).

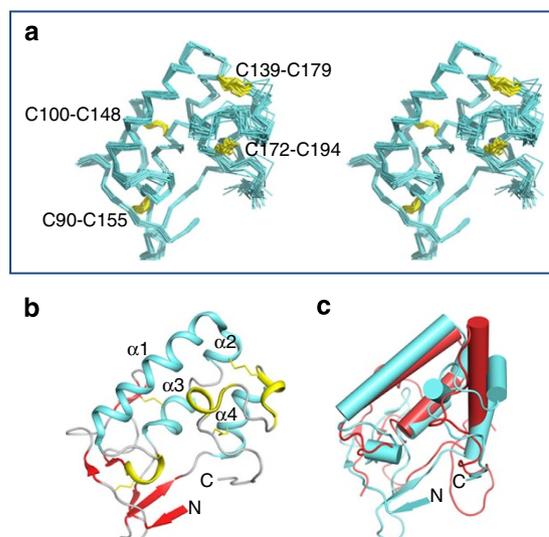
**The cysteines in the Smo CRD.** Structure analysis shows that the eight conserved cysteines in Smo CRD form four disulphide bonds. The chemical shifts observed for the cysteine C-alpha and C-beta carbons are suggestive that all the cysteines are oxidized<sup>29</sup>. The disulphide bond pattern was determined based on the intermolecular nuclear Overhauser effects (NOEs) observed between the β-protons of the cysteine residues and is as follows: C90-C155, C100-C148, C139-C179 and C172-C194 (Fig. 3a). To study the structural and biological relevance of these disulphide bonds, we mutated the cysteines individually to alanine, and checked the ability of the cysteine to alanine



**Figure 2 | Identification of cysteines essential for Smo signalling.** CRD residues C90, C139, C155, C179 and the Smo ECLD residues C218, C238 and C242 are required for proper Hh reporter gene induction. Mutation of C84, 100, 148, 172, 194, 203 to alanine did not compromise the ability of Smo protein to induce the Hh reporter gene activity, and behaved similar to the wild-type protein. Percent activity for each of the experimental assays is shown relative to control. The control level of Hh-induced, *ptcΔ136*-luciferase activity for control dsRNA was set to 100%. For all conditions, luciferase activity is normalized to *pAc-renilla* control. Experiments were performed a minimum of two times in duplicate or triplicate and all data were pooled. Error bars indicate s.e.m.

Myc-Smo mutants to rescue Hh-induced reporter gene in a *smo*-knockdown background. C100A, C148A, C172A and C194A were able to rescue Hh-induced reporter gene induction to wild-type levels (Fig. 2) indicating that the C100-C148 and C172-C194 disulphide bonds are not essential for a functional CRD. Conversely, C90A, C139A, C155A and C179A showed attenuated rescue of the Hh-induced reporter gene induction (Fig. 2), suggesting that the C90-C155 and C139-C179 disulphide play critical roles in maintaining the integrity of the CRD structure. Indeed, additional indirect immunofluorescence revealed that Smo cysteine to alanine mutants that were compromised in their ability to signal were retained in the endoplasmic reticulum (ER), overlapping with the ER marker protein Calreticulin-GFP-KDEL (Fig. 4a and Supplementary Fig. S1a). ER retention is probably because of altered protein folding resulting from the loss of essential disulphide bridges. However, such misfolding did not trigger obvious degradation of Smo protein, as the cysteine to alanine mutants were present at protein levels similar to and greater than that of wild-type Smo (Supplementary Fig. S1b, lane 2 compared with lanes 3–6). Although it is difficult to understand the cause of higher protein levels for cysteine to alanine mutants, we speculate that it may be due to the misfolded Smo mutants failing to exit the ER and cycle to the plasma membrane to be phosphorylated and desensitized like the wild-type protein<sup>30,31</sup>.

**The cysteines in the ECLD of Smo.** In addition to the cysteines of the CRD (V85-K202), the *Drosophila* Smo ECD has five other conserved cysteines, at positions 84, 203, 218, 238 and 242. In the three-dimensional (3D) structure of the *Drosophila* Smo CRD, the amino- and the carboxy termini are in close proximity (Fig. 3b), implying that C84 and C203 may form a disulphide bond. Nevertheless, mutation of C84 and C203 to alanine does



**Figure 3 | Solution structure of the Smo CRD.** (a) Stereo view of the backbone atoms (N, C<sup>α</sup> and C<sup>β</sup>) of the 20 superimposed structures of Smo CRD with the lowest energy. The disulphides are indicated in yellow and are labelled for clarity. (b) Ribbon diagram of the Smo CRD showing the secondary structure elements. The structure with the lowest energy is used to describe the secondary structure elements. The colour scheme is as follows: cyan: alpha helices; red: β strands; yellow: 3<sub>10</sub> helices; grey: random coil. (c) Superimposition of the Smo CRD with FzD CRD. The Smo CRD is represented in cyan and the FzD8 CRD is represented in red. The helices in both proteins are shown as cylinders and the beta-strands as arrowheads. All figures were generated using PyMol.

**Table 1 | NMR and refinement statistics used for calculating the structure of Smo CRD.**

	Protein
<i>NMR distance &amp; dihedral constraints</i>	
Distance constraints	
Total NOE	1,421
Intra-residue	430
Inter-residue	
Sequential ( $ i - j  = 1$ )	422
Medium-range ( $ i - j  < 4$ )	248
Long-range ( $ i - j  > 5$ )	321
Hydrogen bonds	22
Total dihedral angle restraints	139
Phi	71
Psi	68
<i>Structure statistics</i>	
Violations (mean and s.d.)	
Distance constraints (Å)	0.0176 ± 0.0019
Dihedral angle constraints (°)	0.7607 ± 0.0849
Maximum dihedral angle violation (°)	0.9547 ± 0.0849
Maximum distance constraint violation (Å)	0.0238 ± 0.0152
Average pairwise r.m.s.d.* (Å)	
V85-K202	
Heavy	1.28 ± 0.19
Backbone	0.7 ± 0.15
Average pairwise r.m.s.d.* (Å)	
V85-E177	
Heavy	1.08 ± 0.19
Backbone	0.4 ± 0.15

CRD, cysteine-rich domain; NOE, nuclear Overhauser effect; r.m.s.d., root mean squared deviation; Smo, Smoothened.  
\*Average r.m.s.d. from the lowest energy structure was calculated among 20 structures with the lowest energy.

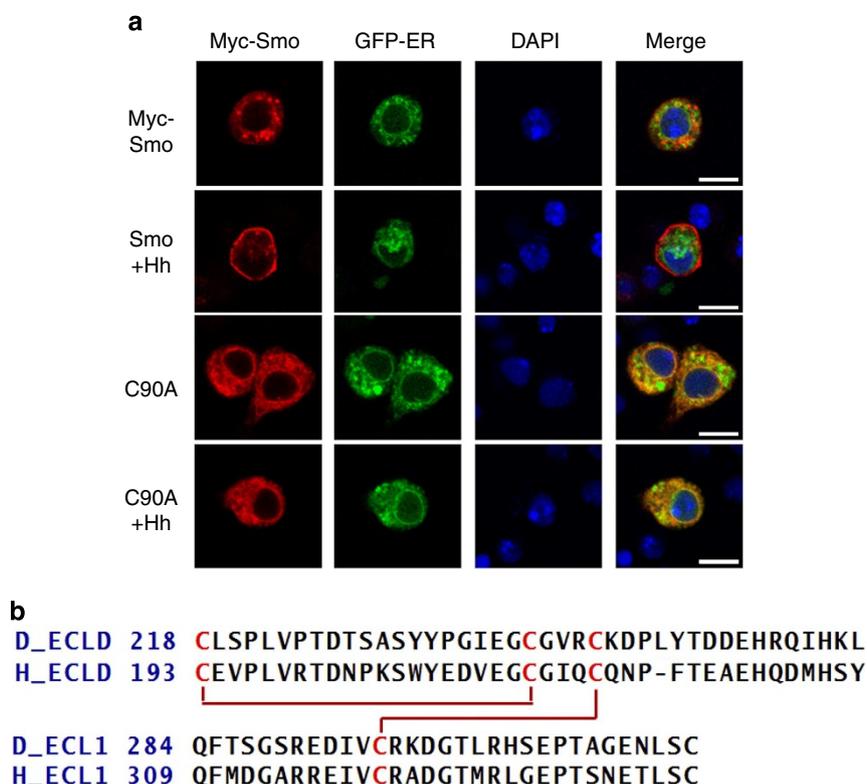
not alter the ability of Smo to rescue reporter gene induction in the *smo*-knockdown background (Fig. 2), suggesting that they are either not engaged in a disulphide bond or the disulphide bond is not crucial in defining the 3D fold of the CRD. However, when C218, C238 and C242 of the ECLD were mutated individually to alanine, they expressed at near-normal levels (Supplementary Fig. S1c), but failed to rescue Hh-induced reporter gene induction following endogenous *smo* knockdown (Fig. 2). C218, C238 and C242 of *Drosophila* Smo are homologous to C193, C213 and C217 of human Smo (Fig. 4b). The crystal structure of the human Smo shows that the ECLD is stabilized towards the ECL by disulphide bonds between C193 and C213 of the ECLD and C217 of the ECLD and C295 ECL1 loop<sup>14</sup> (Fig. 4b). Hence, C218 and C238 in the ECLD of *Drosophila* Smo may form a disulphide bond and C242 in the ECLD is likely to form a disulphide bond with C320 in ECL1 of *Drosophila* Smo. Thus, cysteines both in the Smo CRD and ECD linker play an important role in maintaining the conformation of Smo essential for downstream signalling.

***Drosophila* Smo CRD binds to the glucocorticoid Bud.** Smo activity can be modulated by small-molecule inhibitors such as cyclopamine, which binds in the orthosteric binding site located within a cavity in the 7TM bundle<sup>14,24</sup>. The FzD CRD binds to its endogenous ligand Wnt underscoring the ligand-binding capability of the CRDs<sup>16,18</sup>. Therefore, we decided to test whether the *Drosophila* CRD might bind to 20-OHC or Bud. Because of a solubility issue with 20-OHC, we carried out all experiments with Bud. We used NMR chemical shift

perturbations (CSPs) to identify the residues that are involved in binding<sup>32</sup>, since this method allows us to determine precisely the amino-acid residues involved in the interaction<sup>33</sup>. Despite recent reports stating the inability of 20-OHC to bind to the *Drosophila* Smo protein, we tested whether Bud bound to *Drosophila* Smo CRD<sup>25</sup>. In the 2D <sup>1</sup>H-<sup>15</sup>N HSQC spectra of Smo CRD in the absence and presence of an increasing concentration of the ligand Bud, CSPs induced by the binding of Bud are clearly observed indicating Bud bound to the CRD (Supplementary Figs S2a and S3). The normalized CSP for each residue of Smo in the presence of Bud is shown in Fig. 5a. The residues that have significant CSP are further shown in the backbone 'ribbon' representation of the Smo CRD, wherein the thickness of the ribbon is proportional to CSP values observed on binding of Bud to Smo CRD (Fig. 5b). The results show that the residues of Smo CRD that have the largest CSPs induced by the bound Bud are located between the first helix and the C-terminus <sub>310</sub> helical domain as it folds back between the first and second helix. As a control, we tested the ability of cyclopamine to bind and induce CSPs within the Smo CRD and Smo CRD did not bind to cyclopamine at a ratio of 1:2 of protein to ligand<sup>24</sup> (Supplementary Fig. S2b).

On the basis of NMR studies, we generated the docked structure of Bud bound to Smo CRD using the program HADDOCK (High Ambiguity Driven protein-protein Docking)<sup>34</sup>. The CSP data on the Smo CRD and the bound Bud were used as ambiguous restraints (Supplementary Table S1) for docking and generating 1,000 binding poses. The structure with the lowest binding energy was selected and this agrees with the CSP data on the Smo CRD (Fig. 5b). The modelled structure showed that Bud is buried into a hydrophobic surface formed by the residues A132, H135, F187, F188 and F191 on the Smo CRD (Fig. 5c,d). F191 is conserved in the vertebrate Smo family, whereas F188 is a tryptophan in vertebrate Smo (Fig. 1). The conservation of these residues in the vertebrate family, and the recent demonstration that the domain engages 20-OHC<sup>25,26</sup>, suggests that this pocket may have an important role in binding an as yet unidentified allosteric regulator of *Drosophila* Smo. To determine whether Bud would attenuate signalling by the *Drosophila* Smo protein, we treated Cl8 cells with increasing concentrations of Bud, and observed an ~20–40% reduction in Hh-induced reporter gene activity (Supplementary Fig. S4). Thereby, we speculate that Bud acts as a weak, synthetic mimic of an endogenous ligand for the *Drosophila* Smo CRD, displacing it from this binding pocket.

**The human Smo CRD binds to the glucocorticoid Bud.** Recent reports indicate that 20-OHC binds the vertebrate Smo CRD<sup>25,26</sup>. We therefore tested the binding of Bud to human Smo CRD. The recombinant human Smo CRD (G65-G177) was expressed and purified using the *E. coli* expression system. Similar to the studies of *Drosophila* Smo CRD, <sup>15</sup>N- and <sup>13</sup>C-labelled protein was prepared in 10 mM deuterated acetic acid buffer and 10% D<sub>2</sub>O (volume/volume) at pH 5. In the 2D <sup>1</sup>H-<sup>15</sup>N HSQC spectra of human Smo CRD, all the amide resonances of the 113 residues were assigned except two residues, L73 and R74. NMR titration experiments were performed to determine the specific residues involved in ligand binding. The 2D <sup>1</sup>H-<sup>15</sup>N HSQC spectra of human Smo CRD in the absence and presence of an increasing concentration of Bud showed that Bud bound to human Smo CRD as well (Supplementary Figs S5 and S6). The normalized CSP for each residue of human Smo CRD in the presence of Bud is shown in Fig. 6a. Comparing with *Drosophila* Smo CRD (Fig. 5a), Bud binds to human CRD in a similar manner. However, there are some clear differences. For example, a smaller



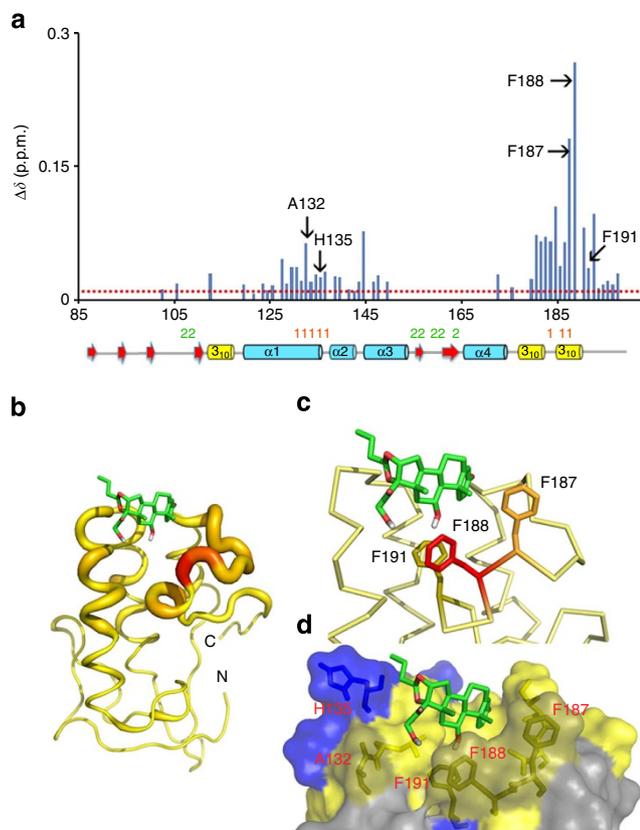
**Figure 4 | Subcellular localization of the cysteines essential for Smo signalling.** (a) CRD mutants with compromised signalling activity have altered subcellular localization. C18 cells expressing wild type or the indicated Myc-Smo-mutant protein, in the presence of Hh (+) or empty vector control, were examined by indirect immunofluorescence. Wild-type Smo translocates to the plasma membrane in response to Hh, whereas Smo CRD cysteine to alanine-mutant C90A that was required for maximal Hh reporter gene induction was largely retained in the ER. Smo was detected using anti-Myc (red), Calreticulin-GFP-KDEL marks the ER (green) and DAPI (blue) marks the nucleus. Scale bar, 10  $\mu$ M. (b) Sequence alignment of the *Drosophila* (D) Smo ECLD and ECL1 with the human (H) Smo ECLD and ECL1. The cysteines engaged in disulphide bond formation in human Smo are conserved in *Drosophila*. The red lines indicate the disulphide bond pattern.

number of residues in human Smo CRD had CSPs induced by the binding of Bud, and the absolute values of CSPs observed in human Smo CRD are smaller. Moreover, the binding affinities measured from different residues in human Smo CRD are clustered around 45  $\mu$ M, whereas those obtained from residues in *Drosophila* Smo CRD range from 200 to 1,000  $\mu$ M indicating that Bud binding induced backbone conformational stabilization in *Drosophila* CRD<sup>35</sup>. Nevertheless, judged by the NMR titration data, it seemed that Bud binds to human Smo CRD with higher affinity (Supplementary Figs S3 and S6). Although NMR is a useful method for determining specific protein ligand interactions in solution, it is not the ideal method for calculating dissociation constants<sup>36</sup>. The associated systemic error in calculating  $K_d$  values from NMR is probably caused by the averaging effect during NMR experiments<sup>37</sup>.  $K_d$  values calculated by this method can only be treated as the upper limit for interaction. Therefore, we validated the binding affinity measurements using additional experiments. Indeed, using Bio-layer interferometry (BLI), we found that Bud bound to *Drosophila* Smo CRD with a  $K_d$  value of  $120 \pm 98 \mu$ M and bound to human Smo CRD with a  $K_d$  value of  $52 \pm 16 \mu$ M (Fig. 6b,c). Both NMR and BLI data suggest that Bud binds to human Smo CRD better than *Drosophila* Smo CRD.

## Discussion

Classical GPCR signalling involves binding of ligands to the extracellular regions of the receptor, which induces a

conformational change in the core of the 7TM domains to modulate signalling events in the cytoplasm<sup>38,39</sup>. Several studies have reported that GPCRs are regulated by natural ligands, as well as exogenous small molecules<sup>38–41</sup>. Furthermore, GPCRs have been reported to possess allosteric binding sites distinct from their orthosteric sites, which when bound by small-molecule modulators, induce distinct signal outputs<sup>42,43</sup>. In the FzD GPCR, the CRD functions to bind to the endogenous ligand Wnt, which interacts with the CRD at two sites<sup>16,18</sup>. The palmitate modification on Wnt engages a groove on the FzD8 CRD termed site 1, whereas the second Wnt-binding site on FzD8 CRD (site 2) is located on the opposite end and involves protein–protein interactions<sup>18</sup>. In the current studies, we showed that Bud interacts with Smo CRD residues analogous to site 1. Within the site 1 of human Smo R161 demonstrated the highest observed CSP. L108, W109, G111, L112, G162 and W163 also demonstrated CSPs. Consistent with these findings, a recent report shows that mutating the homologous murine Smo CRD residues ‘WGL’ (W113, G115 and L116) and R165 to alanine blocked the specific binding of 20-OHC to the Smo CRD and attenuated its signalling ability<sup>26</sup>. Furthermore, mutating murine L112 (human L108) and W113 (human W109) to equivalent *Drosophila* Smo CRD residues blocked 20-OHC binding<sup>25</sup>, suggesting that the site 1 is involved in the binding of hydroxyl-sterols and glucocorticoids. *Drosophila* Smo protein does not bind to 20-OHC *in vitro*<sup>25</sup>; however, that does not preclude the fact that there could be an as yet unidentified



**Figure 5 | *Drosophila* Smo CRD binds to the glucocorticoid Bud.** (a) CSPs of Smo CRD on addition of Bud are plotted versus residue numbers. The red line indicates CSP greater than 0.01 p.p.m. The residues labelled in black form the Bud-binding pocket on Smo CRD as analysed from the HADDOCK-docking experiments. The mouse FzD8-Wnt interacting 'site 1' and 'site 2' residues are shown in orange and green, respectively. The corresponding secondary structure elements of the *Drosophila* Smo CRD are represented below the plot. (b) 'Ribbon' representation of the Smo CRD. The backbone thickness of the ribbon is directly proportional to the weighted sum (in p.p.m.) of the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts on binding to the ligand Bud. (c,d) Results of the HADDOCK docking of Bud on Smo CRD. (c) The aromatic side chains of the Bud contacting Smo CRD residues are shown. (d) Surface representation of the residues that interact with Bud are shown in yellow and the positively charged H135 is shown in blue.

endogenous ligand for *Drosophila* Smo CRD binding to this pocket. Consistent with this hypothesis, genetic experiments suggest that phospholipids could serve as the physiological molecules regulating *Drosophila* Smo activity<sup>44</sup>.

The cache of identified Smo modulators dock on distinct sites on Smo to regulate signalling<sup>14,24,45,46</sup>. The extracellular region of Smo may be flexible (Fig. 7a). Small molecules may bind in the cavity of the 7TM domains (Fig. 7b, right) and modulate signalling irrespective of CRD binding<sup>14,24</sup>. In the present study, we provide structural evidence identifying the specific residues in the Smo CRD that interact with the small-molecule modulator Bud. We speculate that these same residues bind the endogenous allosteric activator, and that activator binding to the CRD may cause a conformational shift that bridges the ECDs and the 7TM domains (Fig. 7b, middle), similar to what is observed for class B GPCRs<sup>40,41,47</sup>. These shifts may enhance communication between the CRD and the 7TM domains to trigger structural alterations in the core and the cytoplasmic domains to regulate downstream signalling (Fig. 7).

Currently, the only Food and Drug Administration-approved therapeutic aimed at targeting Smo is the antagonist Vismodegib, which is likely to bind in the cavity of the 7TM domains and also make contacts with the ECLD and ECL<sup>14,24,48,49</sup>. Despite initial clinical success, resistance to this compound has been observed<sup>48</sup>. Thus, the need exists for developing second-generation drugs that can overcome the acquired drug resistance. On the basis of our data, we speculate that the Bud binding cleft in the Smo CRD could be evaluated for the development of novel antagonists of Smo. Bud is a clinically approved anti-inflammatory glucocorticoid steroid and has the advantage of acting on wild-type as well as Vismodegib-resistant Smo<sup>23,50</sup>. Thus, future investigation and identification of Bud-like molecules that bind to the Smo CRD with high affinity may represent a novel class of Smo regulators.

## Methods

**Vector construction.** The *Drosophila* Smo CRD (V85-K202) was amplified from *pET-SmoN*<sup>3</sup>. The forward and the reverse primers used were 5'-ATA TAT CAT ATG GTC CGA CGT GCC CGT TGC-3' and 5'-ATA TAC TCG AGT TAT TTC GTC GGA AAG AGT G-3'. The PCR-amplified product was cloned in *pET-28a* (+) and validated by nucleotide sequencing. To generate cysteine to alanine mutants, Smo-mutant expression vectors *pAc-myc-smo*<sup>51</sup> was mutated using QuickChange II Site Directed Mutagenesis Kit (Stratagene) using primers generated per manufacturer's instructions. All mutations were verified by sequencing. Smo 5'-UTR double-stranded RNA (dsRNA) was generated using the Megascript T7 kit (Life Technologies)<sup>15</sup>. The human Smo CRD (G65-G177) was amplified from the *smo* complementary DNA purchased from Origene. The forward and the reverse primers used were 5'-TAT ATC ATA TGG GCC GGG CTG CCC CCT GC-3' and 5'-ATA TAT CTC GAG TTA GCC TTC AGG GAA GCG-3', respectively. The PCR-amplified product was cloned in *pET-28a* (+) and validated by nucleotide sequencing.

**Protein expression and purification.** For the expression of both the *Drosophila* and human-recombinant protein *pET-28a-smoCRD* was transformed into BL21 (DE3) cells. The cells were grown in 3-(*N*-morpholino) propane sulphonic acid minimal media supplemented with  $^{15}\text{N}$  ammonium chloride and  $^{13}\text{C}$  glucose as the sole source of nitrogen and carbon, respectively. Cells were grown to an optical density ( $\text{OD}_{600}$ ) of 0.6 and induced with 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside. The protein of interest accumulated in the inclusion bodies and was refolded from the same by the method of rapid dilution. The protein was further purified by reverse-phase HPLC using a C18 column (Hitachi) and gel filtration (AKTA) on a Hiloal 16/60 Superdex 200 preparative grade column (GE Healthcare).

**NMR spectroscopy.** A 200- $\mu\text{M}$  pure monomeric sample of the recombinant *Drosophila* (V85-K202) or human (G65-G177) Smo CRD was prepared in 10 mM deuterated acetic acid buffer and 10%  $\text{D}_2\text{O}$  (volume/volume) for NMR experiments. All spectra were recorded using either  $^1\text{H}$ ,  $^{15}\text{N}$ -labelled or  $^1\text{H}$ ,  $^{15}\text{N}$ - and  $^{13}\text{C}$ -labelled protein on Bruker Avance 600 MHz or 800 MHz NMR spectrometers equipped with  $^1\text{H}/^{15}\text{N}/^{13}\text{C}$  detect, triple resonance cryogenic inverse (TCI) probes at 25 °C. All spectra were processed using Topspin 3.0 NMR software (Bruker Biospin) and analysed using the program, CARA (computer aided resonance assignment)<sup>52</sup>. The backbone chemical shift assignment were obtained using the standard triple-resonance assignment strategy using the 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC and 3D HNCA, HNCACB, CBCA(CO)NH, HN(CA)CO and HNCO NMR experiments. The aliphatic side-chain resonance assignments were obtained using HBHA(CBCACO)NH and (H)CCH-total correlation spectroscopy. Distance constraints for the structure calculation were derived from  $^{15}\text{N}$ / $^{13}\text{C}$  resolved  $^1\text{H}$ - $^1\text{H}$  NOESY with a mixing time of 100 ms. CSP experiments were carried out using  $^{15}\text{N}$ -labelled *Drosophila* or human Smo CRD.  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were recorded in the absence and presence of an increasing concentration of Bud.  $K_d$  values were calculated by a nonlinear least-squares analysis in the program Origin using the equation (1)

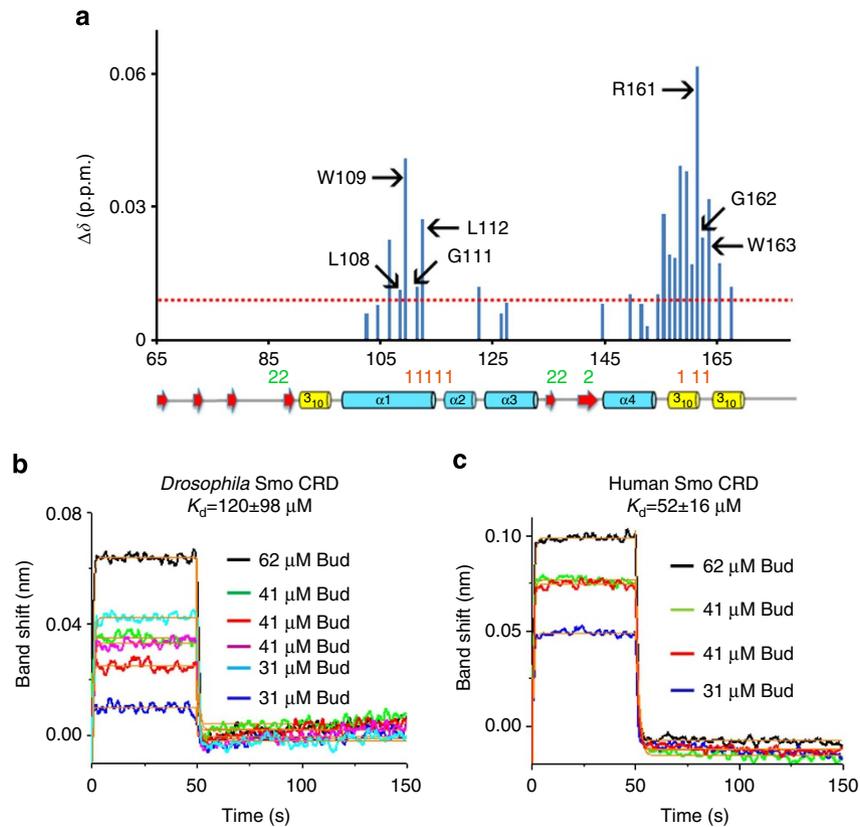
$$\Delta\delta = \Delta\delta_{\text{max}} \left( \frac{([L] + [P] + K_d) - \sqrt{([L] + [P] + K_d)^2 - 4[P][L]}}{2[P]} \right) \quad (1)$$

where [L] is the concentration of the ligand, [P] is the concentration of the protein,  $\Delta\delta$  is the observed chemical shift change and  $\Delta\delta_{\text{max}}$  is the normalized chemical shift change at saturation, calculated using equation (2)

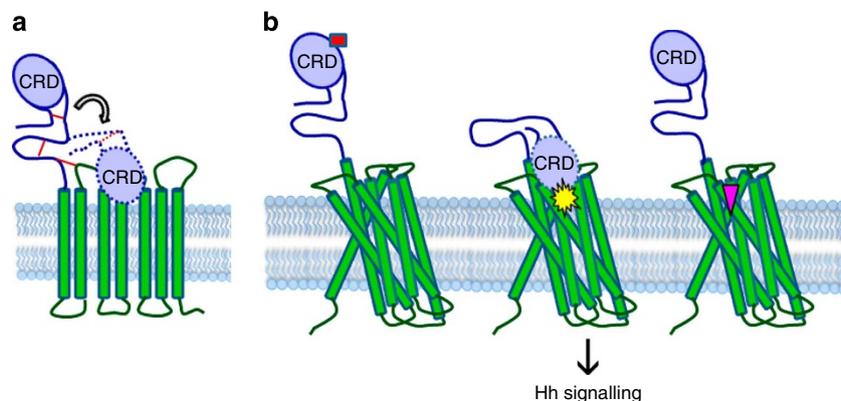
$$\Delta\delta = \sqrt{(\Delta\delta\text{H})^2 + (\Delta\delta\text{N}/5)^2} \quad (2)$$

where  $\Delta\delta$  is the chemical shift in p.p.m.

A stock solution of 50 mM Bud (Sigma) was made in deuterated dimethylsulphoxide (DMSO) for use in NMR titration experiments. The concentration of



**Figure 6 | Analysis of the binding of Bud to human Smo CRD.** (a) CSPs of human Smo CRD on addition of Bud are plotted versus residue numbers. Dotted red line indicates CSP greater than 0.01 p.p.m. L108, W109, G111, L112 and R161 are homologous to the mouse Smo residues that interact with 20-OHC. G162 is homologous to *Drosophila* F187 and W163 is homologous to *Drosophila* Smo F188 that interacts with Bud. All these residues map to the 'site 1' of mouse FzD8–Wnt interaction. 'Site 1' and 'site 2' residues of mouse FzD8–Wnt interaction are shown in orange and green, respectively. The secondary structure elements as in *Drosophila* Smo CRD structure are shown below the plot. (b) BLI-binding assays show that Bud binds to *Drosophila* Smo CRD. The SSA sensors with biotinylated *Drosophila* Smo CRD were exposed to three different concentrations of Bud (62, 41 and 31  $\mu\text{M}$ ). The processed data were fitted locally with the integrated fitting function by a 1:1 binding model (orange line). The respective  $K_d$  values obtained by curve fitting were 89  $\mu\text{M}$  (62  $\mu\text{M}$  Bud, black line), 74  $\mu\text{M}$  (41  $\mu\text{M}$  Bud, green line), 59  $\mu\text{M}$  (41  $\mu\text{M}$  Bud, red line), 318  $\mu\text{M}$  (41  $\mu\text{M}$  Bud, magenta line), 93  $\mu\text{M}$  (31  $\mu\text{M}$  Bud, cyan line) and 85  $\mu\text{M}$  (31  $\mu\text{M}$  Bud, blue line). The average  $K_d$  value of *Drosophila* Smo CRD for Bud was  $120 \pm 98 \mu\text{M}$ . (c) BLI-binding assays show that Bud binds to human Smo CRD. The experimental and data analysis procedure were same as described above. The respective  $K_d$  values obtained by curve fitting were 74  $\mu\text{M}$  (62  $\mu\text{M}$  Bud, black line), 37  $\mu\text{M}$  (41  $\mu\text{M}$  Bud, green line), 44  $\mu\text{M}$  (41  $\mu\text{M}$  Bud, red line) and 54  $\mu\text{M}$  (31  $\mu\text{M}$  Bud, blue line). The average  $K_d$  value of human Smo CRD for Bud was  $52 \pm 16 \mu\text{M}$ .



**Figure 7 | A novel model for Smo allosteric regulation.** (a) Smo CRD may be flexible. The disulphide bonds stabilizing Smo extracellular linker and ECL are shown in red lines. (b) Smo has more than one binding site. We propose that molecules like Bud bind to the Smo CRD (left, red rectangle) to alter its conformation and attenuate its signalling activity. Cyclopamine and vismodegib are known to bind near the orthosteric-binding site located in the cavity of the Smo 7TM domains (right, pink inverted triangle). We speculate that there is the possibility of a class of molecules (middle, yellow star), which would bind to the CRD and cause a conformational change of the Smo extracellular structures to bring it closer to the 7TM domains. This in turn might change the conformation of the transmembrane domains to regulate signalling.

DMSO in the NMR titration experiment was maintained at or below 1%. A control experiment was done by titrating 1% DMSO and no CSPs were observed.

**Structure calculation and refinement.** An automated program UNIO<sup>53</sup> was used for the NOE assignment and all these assignments were manually checked. Integrated NOE peaks were calibrated and converted to distance constraints with the program CALIBA<sup>54</sup>. A total of 1,421 meaningful distance constraints were derived from the NMR data and used as input for structure calculation using the program CYANA 2.1 (ref. 54). The structure was refined using 139 dihedral angle constraints and 22 hydrogen bond constraints. The dihedral angles were generated by the program TALOS+ (ref. 55), which uses six kinds (HN, HA, CA, CB, CO and N) of chemical shift values for a given residue as input to give the phi, psi angles. The Ramachandran statistics are as follows: most favourable 72.5%, additionally allowed 25.7%, generously allowed 1.6% and disallowed 0.2%. Also, four disulphide bonds were added as additional constraints between residues, C90-C155, C100-C148, C139-C179 and C172-C194. A total of 100 conformers were initially generated and 20 conformers with the lowest target function were used to represent the 3D NMR structure.

**Structural modelling of Smo CRD in complex with Bud.** To calculate the structure of the CRD–Bud complex, we used the simulated annealing protocol in the program CNS<sup>34</sup>. All calculations were done using the program HADDOCK 2.1 (ref. 34). Topology and parameter files for Bud were generated by using xplor2d, and the charge parameters of Bud were calculated by using the antechamber module in the AMBER10 software package<sup>56</sup>. We used ambiguous distance restraints resulting from the CSPs of <sup>1</sup>H–<sup>15</sup>N Smo CRD (Supplementary Table S1). To define the active and passive residues, we also calculated a relative accessible surface area by NACCESS<sup>57</sup>. One thousand initial structures of the Smo CRD–Bud complex were generated and the final 10 structures of the complex with the lowest energy were selected manually that were consistent with the CSPs on Smo CRD following Bud binding.

**Lysate preparation and functional assays.** Expression vectors and dsRNA were transfected using Lipofectamine 2000 according to the manufacturer's instructions. For rescue reporter experiments, ~1.5 × 10<sup>6</sup> Cl8 cells were co-transfected with 500 ng Smo 5'-UTR dsRNA, 100 ng *ptcA136-luciferase*<sup>27</sup> 10 ng *pAc-Renilla*<sup>51</sup> and, unless otherwise indicated, 100 ng of *pAc-Hh* (+ Hh) and 50 ng of the indicated *pAc-myc-smo* expression vector. DNA and RNA content were normalized with control dsRNA and/or *pAc5.1A*. Forty-eight hours post transfection, cells were lysed and luciferase activity measured using the Dual Luciferase Assay System (Promega)<sup>15</sup>. For all reporter assays, experiments were performed a minimum of two times in duplicate or triplicate and all data were pooled. Error bars indicate s.e.m.

For the Bud treatment, Cl8 cells were treated with the indicated concentration of Bud (Sigma) or DMSO vehicle. Bud was added in serum-free M3 media (Sigma) 24 h post transfection. Cells were treated with Bud for 24 h before measuring luciferase activity. Experiments were performed at least three times in duplicate and all data were pooled.

For biochemical analysis, ~6 × 10<sup>6</sup> Cl8 cells were transfected with 5 µg of the indicated *pAc-myc-smo* expression vector and 3 µg of *pAc-Hh*, as indicated. DNA content was normalized with empty *pAc5.1A*. Forty-eight hours post transfection, cells were lysed in NP-40 lysis buffer (1% NP-40, 150 mM NaCl, 50 mM Tris, 50 mM NaF, 0.5 mM dithiothreitol and 1 × protease inhibitor cocktail (Roche), pH 8.0), and centrifuged for 10 min at 2,000 × g. The resulting supernatants were analysed by SDS–polyacrylamide gel electrophoresis and western blot using anti-Myc (Roche, catalogue number 11667149001 diluted 1:5,000) and anti-Kinesin (Cytoskeleton, AKIN01-A, diluted 1:10,000). Full scans of the western blots are represented in Supplementary Fig. S7. Antibodies were diluted in Tris-buffered saline with 0.1% Tween and 5% powdered milk.

**Immunofluorescence and microscopy.** For immunofluorescence experiments, ~6 × 10<sup>6</sup> Cl8 cells were transfected with 2 µg of *Calreticulin-GFP-KDEL*<sup>58</sup>, 2 µg *pAc-Hh* and 2 µg of the indicated *pAc-myc-smo* expression vector. DNA content was normalized with *pAc5.1A*. Forty-eight hours post transfection, cells were replated on chamber slides, fixed using 4% formaldehyde and immunostained using anti-Myc (Roche, catalogue number 11667149001)<sup>15</sup>. Calreticulin-GFP-KDEL was used to mark the ER, and DAPI was used to mark the nucleus. Confocal images were collected using a Zeiss LSM 510 NLO Meta microscope.

**BLI assay.** The Octet RED instrument (FortéBio) was used to measure the interaction of *Drosophila* Smo CRD (or human Smo CRD) and Bud. Super streptavidin (SSA) sensors were used to catch the biotinylated target protein. Biotinylation of the target protein was achieved by incubating 200 µl of 1.0 mg ml<sup>-1</sup> EZ-Link NHS-LC-Biotin (Thermo Scientific) with 100 µl of 0.2 mM target protein in 0.1 M potassium phosphate buffer, pH 6.5 for 15 h on a rocking platform at room temperature. Excess unreacted biotin was removed by three successive dialysis under the same buffer conditions. Before the start of the assay, the SSA sensors were pre-wet in 0.1 M potassium phosphate buffer, pH 6.5 with 0.01% Trion-X100 for 10 min. For the binding affinity assay, 3 ~ 5 µM biotinylated

protein was immobilized on the SSA sensor by incubating for 30 min. Subsequently, the free streptavidin sites were quenched by incubating with 0.5 mM biocytin (Thermo Scientific) for 10 min. Excess protein and biocytin were removed by washing with 0.1 M potassium phosphate, pH 6.5 for 10 min. To determine the binding affinity of *Drosophila* Smo CRD (or human Smo CRD) to Bud, three different concentrations of Bud (62, 41 and 31 µM) were used. The association step for 50 s was followed by the dissociation step for 100 s. The assay buffer used was 0.1 M potassium phosphate, pH 6.5 with 0.5% DMSO. Biocytin-loaded SSA sensor without loading biotinylated protein was used as a control to correct the systematic optical artefacts and baseline drifts. All steps were performed at 30 °C with 1,000 r.p.m. rotary shaking. The processed data were fitted locally with the integrated fitting function using the 1:1 binding model in FortéBio analysis software (v6.4). The kinetic constants *k*<sub>on</sub> (on-rate constant), *k*<sub>off</sub> (off-rate constant) and *K*<sub>d</sub> were calculated from curve fitting.

## References

- Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
- van den Heuvel, M. & Ingham, P. W. Smoothed encodes a receptor-like serpentine protein required for hedgehog signalling. *Nature* **382**, 547–551 (1996).
- Alcedo, J., Ayzenzon, M., VonOhlen, T., Noll, M. & Hooper, J. E. The *Drosophila* smoothed gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell* **86**, 221–232 (1996).
- Ogden, S. K. *et al.* G protein G alpha<sub>i</sub> functions immediately downstream of Smoothed in Hedgehog signalling. *Nature* **456**, 967–976 (2008).
- Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
- Jiang, J. & Hui, C. Hedgehog signaling in development and cancer. *Dev. Cell* **15**, 801–812 (2008).
- Cohen, Jr M. M. Hedgehog signaling update. *Am. J. Med. Genet. A* **152A**, 1875–1914 (2010).
- Bale, A. E. Hedgehog signaling and human disease. *Annu. Rev. Genomics Hum. Genet.* **3**, 47–65 (2002).
- Barakat, M. T., Humke, E. W. & Scott, M. P. Learning from Jekyll to control Hyde: Hedgehog signaling in development and cancer. *Trends Mol. Med.* **16**, 337–348 (2010).
- Stone, D. M. *et al.* The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**, 129–134 (1996).
- Hooper, J. E. & Scott, M. P. The *Drosophila* patched gene encodes a putative membrane protein required for segmental patterning. *Cell* **59**, 751–765 (1989).
- Denef, N., Neubuser, D., Perez, L. & Cohen, S. M. Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothed. *Cell* **102**, 521–531 (2000).
- Rohatgi, R. & Scott, M. P. Patching the gaps in Hedgehog signalling. *Nat. Cell Biol.* **9**, 1005–1009 (2007).
- Wang, C. *et al.* Structure of the human smoothed receptor bound to an antitumour agent. *Nature* **497**, 338–343 (2013).
- Carroll, C. E., Marada, S., Stewart, D. P., Ouyang, J. X. & Ogden, S. K. The extracellular loops of Smoothed play a regulatory role in control of Hedgehog pathway activation. *Development* **139**, 612–621 (2012).
- Dann, C. E. *et al.* Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. *Nature* **412**, 86–90 (2001).
- Pei, J. & Grishin, N. V. Cysteine-rich domains related to Frizzled receptors and Hedgehog-interacting proteins. *Protein Sci.* **21**, 1172–1184 (2012).
- Janda, C. Y., Waghay, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science* **337**, 59–64 (2012).
- Nakano, Y. *et al.* Functional domains and sub-cellular distribution of the Hedgehog transducing protein Smoothed in *Drosophila*. *Mech. Dev.* **121**, 507–518 (2004).
- Aanstad, P. *et al.* The extracellular domain of Smoothed regulates ciliary localization and is required for high-level Hh Signaling. *Curr. Biol.* **19**, 1034–1039 (2009).
- Zhao, Y., Tong, C. & Jiang, J. Hedgehog regulates smoothed activity by inducing a conformational switch. *Nature* **450**, 252–258 (2007).
- Nachtergaele, S. *et al.* Oxysterols are allosteric activators of the oncoprotein Smoothed. *Nat. Chem. Biol.* **8**, 211–220 (2012).
- Wang, Y. *et al.* Glucocorticoid compounds modify Smoothed localization and Hedgehog pathway activity. *Chem. Biol.* **19**, 972–982 (2012).
- Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. Small molecule modulation of Smoothed activity. *Proc. Natl Acad. Sci. USA* **99**, 14071–14076 (2002).
- Nedelcu, D., Liu, J., Xu, Y., Jao, C. & Salic, A. Oxysterol binding to the extracellular domain of Smoothed in Hedgehog signaling. *Nat. Chem. Biol.* **9**, 557–564 (2013).
- Myers, B. R. *et al.* Hedgehog pathway modulation by multiple lipid binding sites on the Smoothed effector of signal response. *Dev. Cell* **26**, 346–357 (2013).

27. Chen, C. H., von Kessler, D. P., Park, W. J. & Beachy, P. A. Nuclear trafficking of cubitus interruptus in the transcriptional regulation of Hedgehog target gene expression. *Cell* **98**, 305–316 (1999).
28. Bax, A. & Grzesiek, S. Methodological advances in protein NMR. *Acc. Chem. Res.* **26**, 131–138 (1993).
29. Sharma, D. & Rajarathnam, K. <sup>13</sup>C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR* **18**, 165–171 (2000).
30. Molnar, C. *et al.* Role of the *Drosophila* non-visual beta-arrestin kurtz in Hedgehog signalling. *PLoS Genet.* **7** (2011).
31. Jia, J. H., Tong, C., Wang, B., Luo, L. P. & Jiang, J. Hedgehog signalling activity of Smoothed requires phosphorylation by protein kinase A and casein kinase I. *Nature* **432**, 1045–1050 (2004).
32. Wuthrich, K. Protein recognition by NMR. *Nat. Struct. Biol.* **7**, 188–189 (2000).
33. Zheng, J. *et al.* Identification of the binding site for acidic phospholipids on the pH domain of dynamin: implications for stimulation of GTPase activity. *J. Mol. Biol.* **255**, 14–21 (1996).
34. Dominguez, C., Boelens, R. & Bonvin, A. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
35. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Mag. Res. Sp.* **73**, 1–16 (2013).
36. London, T. B. C., Lee, H. J., Shao, Y. M. & Zheng, J. Interaction between the internal motif KTXXXI of Idax and mDvl PDZ domain. *Biochem. Biophys. Res. Commun.* **322**, 326–332 (2004).
37. Shan, J. F. & Zheng, J. Optimizing Dvl PDZ domain inhibitor by exploring chemical space. *J. Comput. Aided Mol. Des.* **23**, 37–47 (2009).
38. Lagerstrom, M. C. & Schioth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* **7**, 339–357 (2008).
39. Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
40. Grace, C. R. R. *et al.* NMR structure and peptide hormone binding site of the first extracellular domain of a type B1 G protein-coupled receptor. *Proc. Natl Acad. Sci. USA* **101**, 12836–12841 (2004).
41. Koth, C. M. *et al.* Molecular basis for negative regulation of the glucagon receptor. *Proc. Natl Acad. Sci. USA* **109**, 14393–14398 (2012).
42. Christopoulos, A. & Kenakin, T. G protein-coupled receptor allostery and complexing. *Pharmacol. Rev.* **54**, 323–374 (2002).
43. Hudson, B. D., Ulven, T. & Milligan, G. The therapeutic potential of allosteric ligands for free fatty acid sensitive GPCRs. *Curr. Top. Med. Chem.* **13**, 14–25 (2013).
44. Yavari, A. *et al.* Role of lipid metabolism in Smoothed derepression in hedgehog signaling. *Dev. Cell* **19**, 54–65 (2010).
45. Rohatgi, R., Milenkovic, L., Corcoran, R. B. & Scott, M. P. Hedgehog signal transduction by Smoothed: pharmacologic evidence for a 2-step activation process. *Proc. Natl Acad. Sci. USA* **106**, 3196–3201 (2009).
46. Wu, V. M., Chen, S. C., Arkin, M. R. & Reiter, J. F. Small molecule inhibitors of Smoothed ciliary localization and ciliogenesis. *Proc. Natl Acad. Sci. USA* **109**, 13644–13649 (2012).
47. Dong, M. Q., Li, Z. J., Pinon, D. I., Lybrand, T. P. & Miller, L. J. Spatial approximation between the amino terminus of a peptide agonist and the top of the sixth transmembrane segment of the secretin receptor. *J. Biol. Chem.* **279**, 2894–2903 (2004).
48. Von Hoff, D. D. *et al.* Inhibition of the hedgehog pathway in advanced basal-cell carcinoma. *N. Engl. J. Med.* **361**, 1164–1172 (2009).
49. Robarge, K. D. *et al.* GDC-0449-a potent inhibitor of the hedgehog pathway. *Bioorg. Med. Chem. Lett.* **19**, 5576–5581 (2009).
50. Balanag, V. M., Yunus, F., Yang, P. C. & Jorup, C. Efficacy and safety of budesonide/formoterol compared with salbutamol in the treatment of acute asthma. *Pulm. Pharmacol. Ther.* **19**, 139–147 (2006).
51. Ogden, S. K. *et al.* Smoothed regulates activator and repressor functions of Hedgehog signaling via two distinct mechanisms. *J. Biol. Chem.* **281**, 7237–7243 (2006).
52. Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial* (CANTINA Verlag, 2004).
53. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171–189 (2002).
54. Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 (2004).
55. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
56. Case, D. A. *et al.* AMBER 10 (University of California, 2008).
57. Hubbard, S. J. & Thornton, J. M. NACCESS (Department of Biochemistry and Molecular Biology, University College London, 1993).
58. Casso, D. J., Tanda, S., Biehs, B., Martoglio, B. & Kornberg, T. B. *Drosophila* signal peptide peptidase is an essential protease for larval development. *Genetics* **170**, 139–148 (2005).

### Acknowledgements

This work was supported by NIH research Grants R01 GM100909 (J.J.Z.), R01 GM081492 (J.J.Z.), R01GM101087 (S.K.O.), March of Dimes MOD5-FY10-6 (S.K.O.), NIH Cancer Center grant P30 CA021765 and ALSAC. C.E.C. was supported by a UNCF/Merck Postdoctoral Science Research Fellowship.

### Author contributions

J.J.Z. and S.K.O. conceived the study, designed the experiments and served as corresponding authors. R.R. performed protein purification, NMR data acquisition and analysis, structure determination and refinement. R.R. cloned, purified and assigned the human Smo CRD and performed both *Drosophila* and human NMR titration experiments. H.-J.L. determined the  $K_d$  from the NMR experiments and performed the BLI assays. H.-J.L. and J.B. performed HADDOCK docking experiments. C.E.C. and S.M. performed all cell-based assays. C.R.R.G. helped with NMR data acquisition and structure calculation. C.D.G. made the clone of the *Drosophila* protein for NMR studies. R.R., S.K.O. and J.J.Z. coordinated the experimental work and wrote the manuscript with contribution from all authors.

### Additional information

**Accession code:** Chemical shifts of *Drosophila* Smo CRD have been deposited at the BioMagResBank under accession code 19354. The coordinates for the structure are deposited in the Protein Data Bank under the accession code 2MAH.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npublishing.nature.com/reprintsandpermissions/>

**How to cite this article:** Rana, R. *et al.* Structural insights into the role of the Smoothed cytosine-rich domain in Hedgehog signalling. *Nat. Commun.* **4**:2965 doi: 10.1038/ncomms3965 (2013).