Supporting Information

Whalen et al. 10.1073/pnas.1310097110

SI Materials and Methods

Expression and Purification of Wild-Type Sonic Hedgehog Constructs. Two constructs of the murine Sonic Hedgehog N-terminal signaling domain (ShhN) (GenBank entry AK077688, UniProtKB/Swiss-Prot Q62226) were cloned into the pET22b vector (Novagen), introducing a C-terminal hexa-histidine tag: $ShhN_{\Delta 24}$ (residues Cys25–Ala194), containing the Cardin–Weintraub (CW) sequence and $ShhN_{\Delta 39}$ (residues Leu40–Ala194), lacking the CW sequence. Protein constructs were expressed by isopropyl-β-D-thiogalactopyranoside (IPTG) induction in *Escherichia coli* Rosetta (DE3)pLysS cells (Invitrogen). Cultures were grown at 37 °C to an optical density (A_{600}) of 0.6–0.8, cooled to room temperature (~20 °C), induced by the addition of 250 μM IPTG, and grown overnight at 18 °C. Proteins were purified from bacterial lysate by immobilized cobalt- (TALON beads; Clontech Laboratories) or nickel-affinity chromatography, followed by size exclusion chromatography (Superdex resin; GE Healthcare) in 10 mM Hepes pH 7.5, 150 mM NaCl.

Site-Directed Mutagenesis. Two mutant forms of ShhN to validate the Shh core glycosaminoglycan (GAG)-binding site ($ShhN_{\Delta 24}$–K88E/R124E/R154E/R156E and $ShhN_{\Delta 39}$–K88E/R124E/R154E/R156E) were generated by a two-step overlapping PCR experiment using Pyrobest Polymerase (Takara). PCR products were cloned into the pET22b vector, introducing a C-terminal hexa-histidine tag. The constructs were then verified by DNA sequencing. The two mutant ShhN proteins were produced at similar levels as corresponding wild-type proteins.

Heparin Affinity Chromatography. Four purified ShhN constructs were loaded (1 mg/mL, 1 mL each) onto a 5-mL HiTrap heparin HP column (GE Healthcare Life Sciences) equilibrated in 40 mM NaCl, 10 mM Hepes pH 7.5. ShhN elution (flow rate 2 mL/min) was followed by absorption at 280 nm. Proteins were eluted using a linear NaCl gradient, from 40 mM NaCl, 10 mM Hepes pH 7.5 to 1 M NaCl, 10 mM Hepes pH 7.5 over 10 column volumes.

Crystallization and Data Collection. Before crystallization, purified $ShhN_{\Delta 39}$ was concentrated to 9 mg/mL. The $ShhN_{\Delta 39}$–heparin complex was crystallized in the presence of 2 mM 18-mer heparin (Iduron; molecular weight ~5,200 Da) and the $ShhN_{\Delta 39}$–chondroitin-4-sulfate (C4S) complex in the presence of 10 mM chondroitin 4-sulfate (Sigma) and 5 mM calcium chloride. Crystallization trials, using 100 nL protein solution plus 100 nL reservoir solution in sitting drop vapor diffusion format were set up in 96-well Greiner plates using a Cartesian Technologies robot (1). Crystallization plates were maintained at 20.5 °C in a TAP Homebase storage vault and imaged via a Veeco visualization system (2). $ShhN_{\Delta 39}$–heparin complex crystals were obtained in 200 mM sodium acetate, 100 mM Tris·HCl pH 8.5, 30% (wt/vol) PEG 4000, 8% (vol/vol) 1,1,1,3,3,3-hexafluoro-2-propanol. $ShhN_{\Delta 39}$–CS complex crystals were obtained in 25% (wt/vol) PEG 3350, 100 mM Hepes pH 7.5. Before flash freezing in liquid nitrogen, $ShhN_{\Delta 39}$–heparin complex crystals were soaked with 50 mM calcium chloride in mother liquor for 20 min to ensure the full occupancy of the Shh–Ca binding sites. Twenty percent (vol/vol) glycerol in mother liquor was used as cryoprotectant for all crystals. Data were collected at 100 K at beamline I03 at the Diamond Light Source. All X-ray data were processed and scaled with the HKL suite (3) and XIA2 (4). Data collection statistics are shown in Table S1.

Structure Determination and Refinement of the $ShhN_{\Delta 39}$–Heparin Complex. The structure was solved by the molecular replacement method with PHASER (5) in space group $P2_12_12$ (Table S1) using the structure of the ShhN apo form as search model [Protein Data Bank (PDB) ID code 1VHH]. Additional electron density for the heparin chain and the zinc and calcium ions were immediately discernible (Fig. S2). The heparin molecule was built in two distinct conformations, which are related by the 2_1 symmetry operation of the crystallographic b axis resulting in two heparin chains (each with an occupancy of 50%) running in opposite directions. Together the symmetry-related sugar chains account for the observed density (Fig. S2 *B* and *C*). For heparin refinement, the pyranose rings were modeled in the lowest energy chair $^4C_1$ conformation, using REFMAC (6) dictionaries edited to contain more stringent bond length/angle restraints as well as the addition of planarity restraints to maintain the correct ring pucker. This was necessary to prevent overfitting of the pyranose rings to the electron density at the expense of correct geometry during refinement. The Shh sidechains of residues K179 and R154 (which are part of the Shh core GAG-binding site) showed higher flexibility compared with the rest of the molecule. This observed disorder appears to be a common feature for GAG–protein complex structures and is likely due to the multiple interaction modes between the positively charged residues of Shh and the negatively charged interaction sites on heparin (7). Iterative rounds of refinement in REFMAC (6) and manual building in COOT (8) resulted in a model that includes a Shh dimer (each monomer consisting of residues 40–190, one zinc and two calcium ions) and two 12-residue-long heparin molecules (each with an occupancy of 50%) (Fig. S2*C*).

Structure Determination and Refinement of the $ShhN_{\Delta 39}$–C4S complex. The structure was solved by molecular replacement as described for the $ShhN_{\Delta 39}$–heparin complex (Table S1). Additional electron density for the C4S chain and zinc and calcium ions were also immediately discernible (Fig. S7). In contrast to the Shh–heparin complex, only *N*-acetylgalactosamine-4-sulfate (ASG) residue I, which directly interacts with the Shh core GAG-binding site, was well resolved after refinement. The remaining three saccharides present in the asymmetric unit [ASG-II and glucuronic acid (BSG) residues I and II] were only partially ordered (Fig. S6), which is reflected in the higher B factors of C4S compared with those of protein atoms (Table S1). The structure was refined using PHENIX (9) applying occupancy refinement for the C4S molecule (occupancy converged to 0.93) and stringent geometric restraints to maintain correct saccharide ring puckering. The final model resulted in an R_{work} of 16.0% (R_{free}: 19.3%) with one Shh molecule, a zinc, and two calcium ions, and a C4S tetrasaccharide per asymmetric unit.

Structural Analysis. Stereochemical properties were assessed in COOT (8) and MOLPROBITY (10). Superpositions were calculated using COOT (8) and electrostatic potentials were generated using APBS software (11). Buried surface areas of protein–protein interactions were calculated using the PISA webserver (12) for a probe radius of 1.4 Å and figures were prepared using PyMOL (13). The ConSurf server (14) was used to calculate the evolutionary conservation of Shh residues and map these results onto the $ShhN_{\Delta 39}$ crystal structure. Ligplot (15) was used to analyze $ShhN_{\Delta 39}$–GAG interactions.

Surface Plasmon Resonance-Based Binding Studies. Surface plasmon resonance (SPR) experiments were performed using a Biacore T100 machine (GE Healthcare) at 25 °C in 10 mM Hepes pH 7.5, 120 mM NaCl, 3 mM $CaCl_2$, 0.05% (vol/vol) Tween 20. Protein concentrations were determined from the absorbance at 280 nm using calculated molar extinction coefficients. Heparin (Iduron; average molecular weight >9,000 Da) and heparan sulfate (HS) from porcine mucosa (Iduron) were biotinylated using EZ-link Biotin-LC-Hydrazide (Thermo Fisher Scientific) as described previously (16). Biotinylated GAGs were immobilized on the streptavidin-coated SPR CM5 chip (Biacore) surface (less than 50 response units). For measurements of Shh-binding to immobilized GAGs, samples were injected at 5 μL/min. The signal from experimental flow cells was corrected by subtraction of a buffer and reference signal from a control protein (streptavidin without biotinylated GAG)-coupled flow cell. Shh association with GAGs was disrupted and surface regeneration was performed in 1.5 M NaCl, 10 mM Hepes pH 7.5, 0.05% Tween 20 (100 μL/min, 1 min). K_d values were obtained by nonlinear curve fitting of the Langmuir binding isotherm (bound = $(C \times B_{max})/(K_d + C)$, where C is analyte concentration and B_{max} is the maximum analyte binding).

Analytical Ultracentrifugation. Sedimentation velocity experiments were carried out at 20 °C in a Beckman Optima XL-I analytical ultracentrifuge (Beckman Instruments) using a scanning absorbance of 280 nm and interference optics. Sample 1, 1.33 mg/mL $ShhN_{\Delta 39}$; sample 2, 1.33 mg/mL $ShhN_{\Delta 39}$ and 2 mM 6-mer heparin (Sigma, average chain length 6 saccharide units); sample 3, 1.33 mg/mL $ShhN_{\Delta 39}$ and 10 μM 30-mer heparin (Iduron; average chain length 30 saccharide units); sample 4, 2 mM 6-mer heparin; and sample 5, 10 μM 30-mer heparin (all in 10 mM Hepes pH 7.5, 120 mM NaCl) were held in 12-mm Epon sector-shaped two-channel centerpieces and were spun at 40,000 rpm (An60Ti rotor, Beckman Coulter Inc., CA), with 80 sample distribution scans being taken in increments of 6 min apart using a scanning absorbance of 280 nm, and interference optics. Interference data were used to monitor the sugar, which does not absorb at 280 nm, but were not required for further analysis. Absorbance data of samples 6–50 were analyzed using Sedfit (17), for size-and-shape distribution (c(*s.fr*) where *fr* is the frictional ratio and a sphere has a frictional ratio of 1 and other species, >1). This allows the plotting of contour plots of c(s,M), where M is the weight of the protein. For the protein-only cells, a value of 0.73 mL/g was taken for the partial specific volume. A carbohydrate has a partial specific volume of 0.53. Therefore, for the cell containing the 6-mer heparin, a vbar was calculated from the proportion of sugar:protein ratio expected from the predicted complexes (6-mer heparin/$ShhN_{\Delta 39}$, 1:1 complex, vbar = 0.71). As the number, size, and type of complexes in the experimental conditions containing the 30-mer heparin were not initially known, it was not possible to ab initio calculate a vbar to confidently describe all possible complexes. Analysis of these conditions was therefore performed using a range of vbar values (0.50–1.00), the use of which did not materially affect the results. In performing the presented analysis, a value of 0.73 was used, because it describes the largest and therefore most populous c(s) peak.

1. Walter TS, et al. (2005) A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* 61(Pt 6): 651–657.
2. Mayo CJ, et al. (2005) Benefits of automated crystallization plate tracking, imaging, and analysis. *Structure* 13(2):175–182.
3. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography Pt A* 276:307–326.
4. Winter G (2010) xia2: An expert system for macromolecular crystallography data reduction. *J Appl Cryst* 43:186–190.
5. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(Pt 4):658–674.
6. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240–255.
7. Fukuhara N, Howitt JA, Hussain SA, Hohenester E (2008) Structural and functional analysis of slit and heparin binding to immunoglobulin-like domains 1 and 2 of *Drosophila* Robo. *J Biol Chem* 283(23):16226–16234.
8. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
9. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
10. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375–W383.
11. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98(18):10037–10041.
12. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774–797.
13. Schrödinger LLC (2012) The PyMOL Molecular Graphics System, version 1.5.0.4.
14. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38(Web Server issue):W529–W533.
15. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8(2):127–134.
16. Malinauskas T, Aricescu AR, Lu W, Siebold C, Jones EY (2011) Modular mechanism of Wnt signaling inhibition by Wnt inhibitory factor 1. *Nat Struct Mol Biol* 18(8):886–893.
17. Brown PH, Schuck P (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys J* 90(12):4651–4661.

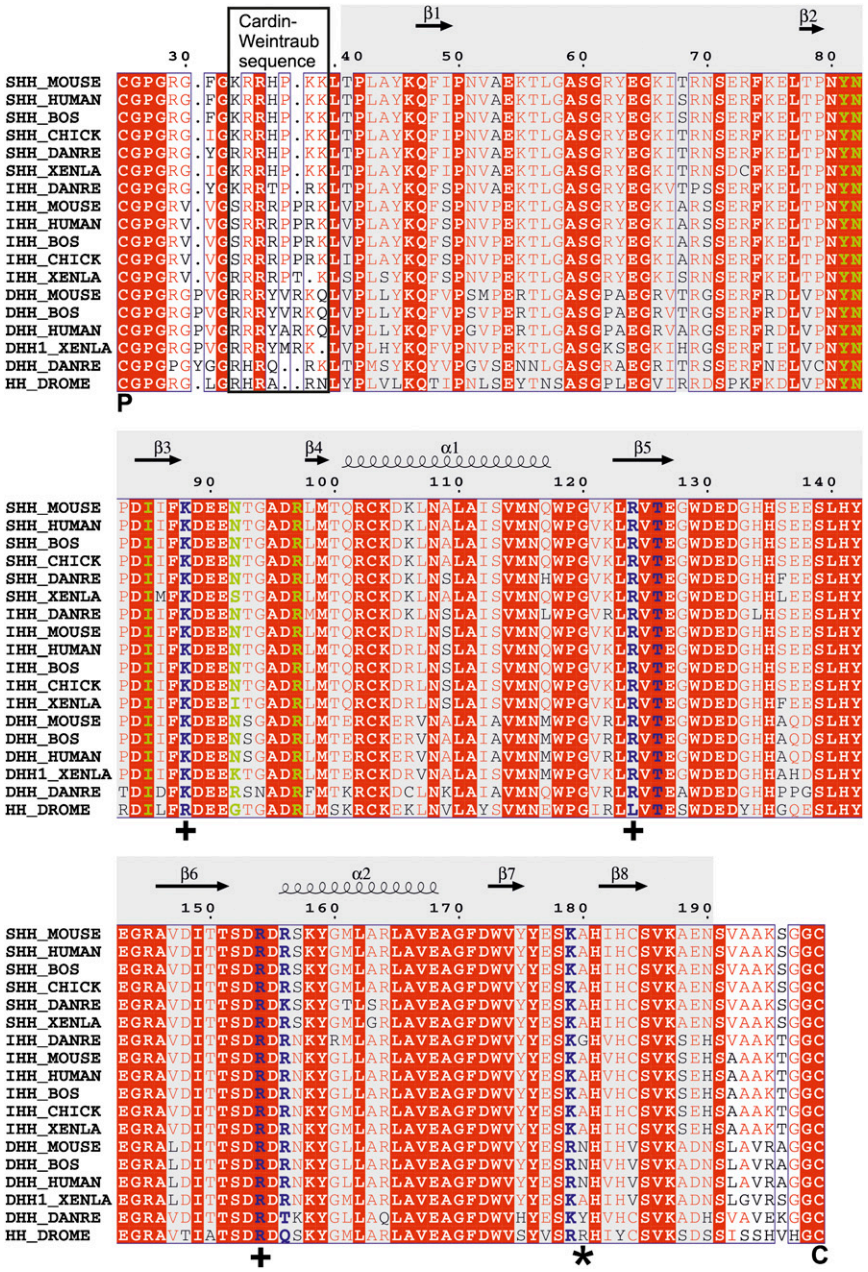

Fig. S1. Amino acid sequence alignment of Hh family members. Sequences were aligned using MultAlin (http://bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPript (http://espript.ibcp.fr/ESPript/ESPript/). Numbering corresponds to mouse Shh. The CW is boxed with a dotted line. Shh residues interacting with heparin are colored in blue; residues contributing to the Shh dimer interface are shown in green. Shh–C4S interacting residues are marked with a plus sign (+). Palmitate and cholesterol attachment sites are annotated with P and C, respectively, below the sequence. *Drosophila* Hedgehog (fHh)-R239 is marked with an asterisk (*); this residue occupies the same side-chain position as mouse Shh-R156 (which is not conserved in fly Hh). Secondary structure elements of the mouse $ShhN_{\Delta N39}$ structure are shown above the sequence. The gray box in the background indicates the amino acid residues present in the structure.

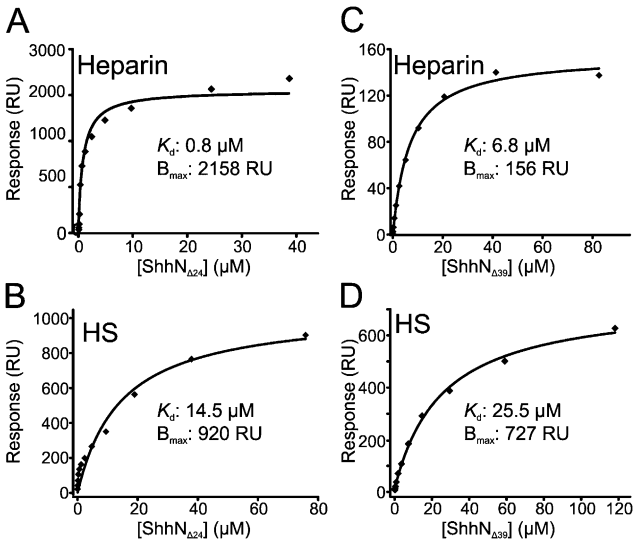

Fig. S2. GAG-binding properties of different ShhN constructs. Representative SPR equilibrium-binding experiments of the GAG–Shh interactions. Different concentrations of ShhN$_{\Delta 24}$ and ShhN$_{\Delta 39}$ were injected over surfaces coupled with GAGs: (*A*) ShhN$_{\Delta 24}$–heparin; (*B*) ShhN$_{\Delta 24}$–HS; (*C*) ShhN$_{\Delta 39}$–heparin; and (*D*) ShhN$_{\Delta 24}$–HS.

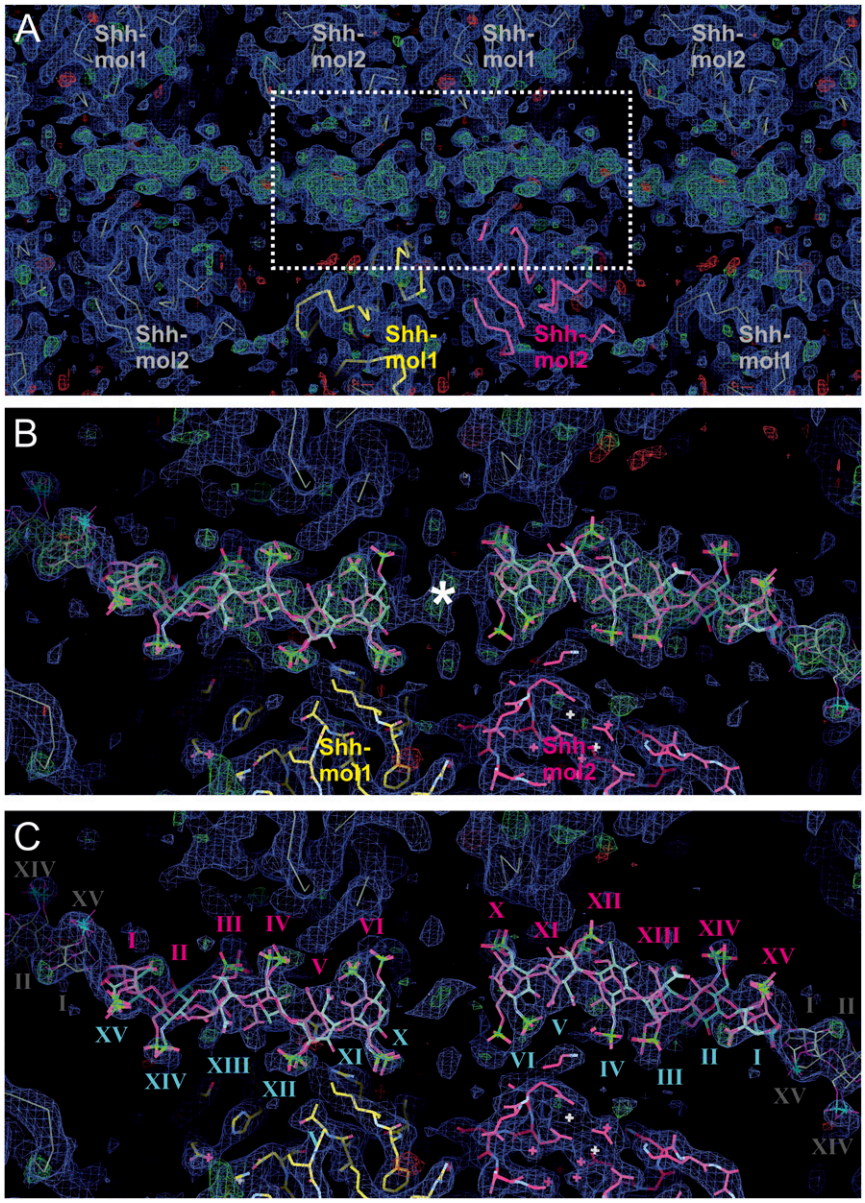

Fig. S3. Structure determination of the $ShhN_{\Delta 39}$–heparin complex. (*A*) View of the crystal packing showing the two Shh molecules (red and yellow ribbons) in the asymmetric unit. The $2F_o–F_c$ (0.9 σ, blue) and $F_o–F_c$ (±3.0 σ, green and red) maps omitting the heparin chain are shown. The symmetry of the crystal results in a continuous stretch of heparin, resulting in a head-to-tail arrangement of the sugar chains through the whole crystal. (*B*) Close-up view of the box in *A* with the final model of the Shh–heparin complex. The asterisk indicates a crystallographic twofold axis. (*C*) $2F_o–F_c$ (0.9 σ, blue) and $F_o–F_c$ (±3.0 σ, green and red) maps after final refinement in REFMAC (6). View is as in *B*. The heparin molecule (magenta sticks and roman numerals) has a symmetry mate (cyan sticks and roman numerals) related by a 2_1 screw axis lying parallel to the sugar. Even numerals correspond to N,O6-disulfoglucosamine, odd numerals to O2-sulfoiduronic acid. The two heparin chains together account for the observed density. The regular positions of the sulfate groups allow the antiparallel heparin strands to place a sulfate group in regular matching positions irrespective of the directionality. The saccharides of the symmetry-related heparin molecules of the neighboring asymmetric units are labeled in gray.

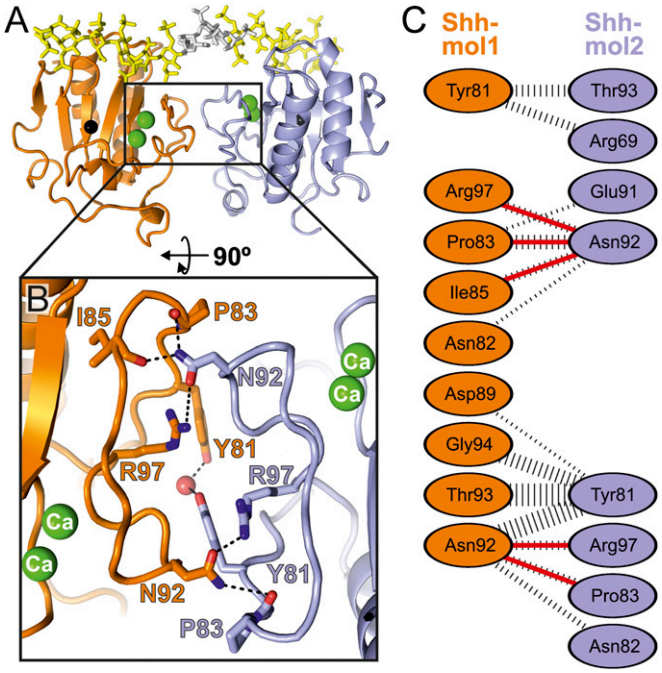

Fig. S4. Dimer formation in the $ShhN_{\Delta 39}$–heparin complex structure. (*A*) Cartoon representation of the $ShhN_{\Delta 39}$–heparin complex. The two Shh molecules are shown in orange and slate. Heparin is depicted in yellow sticks. Orientation is similar to Fig. 2*A*. (*B*) Close-up view of the dimer interface boxed in *A*. Residues forming hydrogen bonds are shown in stick representation (carbon, orange or slate; oxygen, red; nitrogen, blue). A water molecule is shown as a red sphere, calcium ions as green spheres. Hydrogen bonds are marked with black dotted lines. (*C*) Table of Shh dimer interactions adapted from PDBsum (www.ebi.ac.uk/pdbsum/). Red lines, hydrogen bonds; black striped lines, van der Waals contacts.

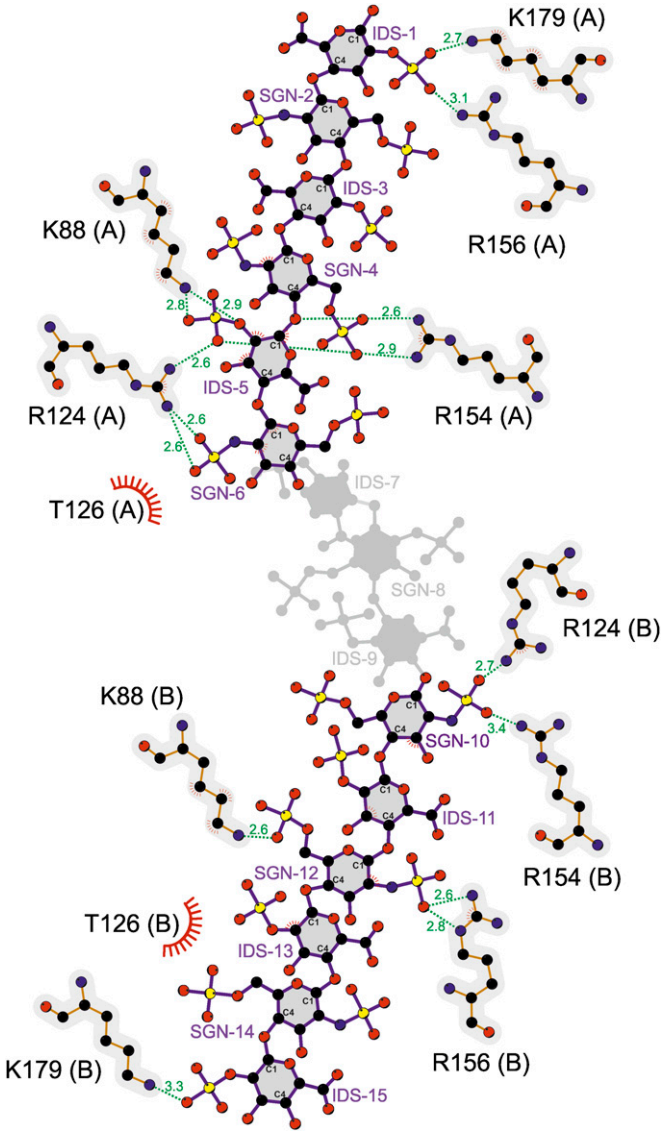

Fig. S5. Schematic representation of the $Shh_{\Delta 39}$–heparin interactions. Atoms are colored as follows: N, blue; O, red; S, yellow; C, black. Bonds are colored violet (heparin) and orange (Shh). Brackets after amino acids correspond to the respective Shh chains. Hydrogen bonds are shown in green with lengths in angstroms. Red "eyelashes" indicate hydrophobic interactions. SGN, N,O6-disulfoglucosamine; IDS, O2-sulfoiduronic acid. IDS-7, SGN-8, and IDS-9 that were omitted in the structure refinement are shown in gray.

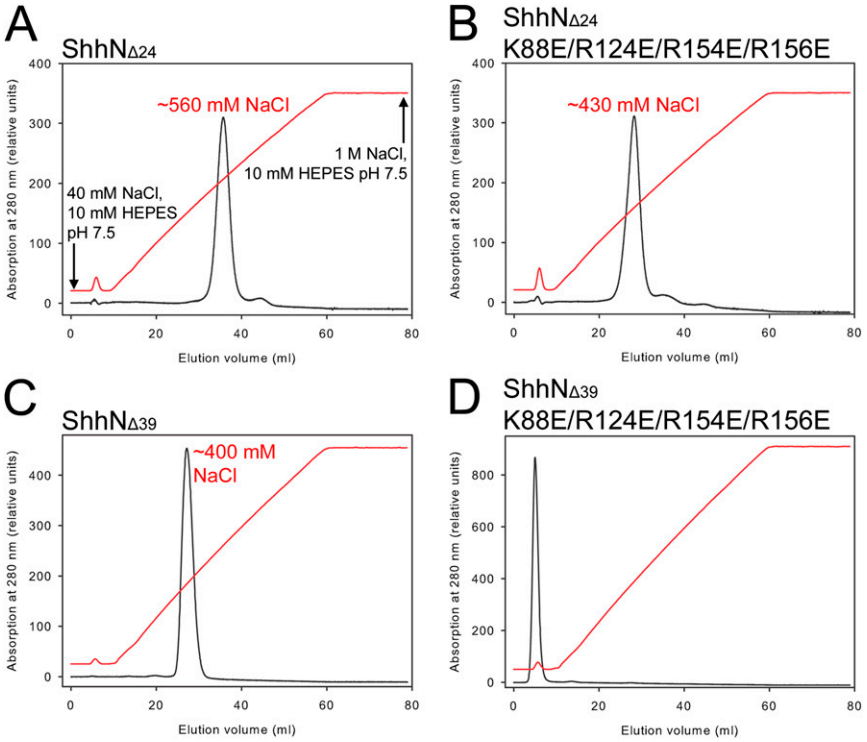

Fig. S6. Heparin affinity chromatography of mouse ShhN. Wild-type $ShhN_{\Delta 24}$ (*A*), mutant $ShhN_{\Delta 24}$–K88E/R124E/R154E/R156E (*B*), wild-type $ShhN_{\Delta 39}$ (*C*), and mutant $ShhN_{\Delta 39}$–K88E/R124E/R154E/R156E (*D*) were purified and loaded onto a 5-mL heparin column (HiTrap Heparin HP; GE Healthcare Life Sciences) equilibrated in 40 mM NaCl, 10 mM Hepes pH 7.5. ShhN elution was followed by absorption at 280 nm. The elution gradient is represented by the red trace showing the conductivity from around 5 (start of run) to around 91 (end of run) mS/cm. Wild-type $ShhN_{\Delta 24}$ (*A*), mutant $ShhN_{\Delta 24}$–K88E/R124E/R154E/R156E (*B*), and wild-type $ShhN_{\Delta 39}$ (*C*) bound to the column and were eluted with ∼560 mM NaCl, 10 mM Hepes pH 7.5 (conductivity 52 mS/cm at the peak), ∼430 mM NaCl, 10 mM Hepes pH 7.5 (40 mS/cm), and ∼400 mM NaCl, 10 mM Hepes pH 7.5 (37 mS/cm), respectively. $ShhN_{\Delta 39}$–K88E/R124E/R154E/R156E (*D*) did not bind to the column and was eluted with the binding buffer (elution volume 3–8 mL, ∼5 mS/cm).

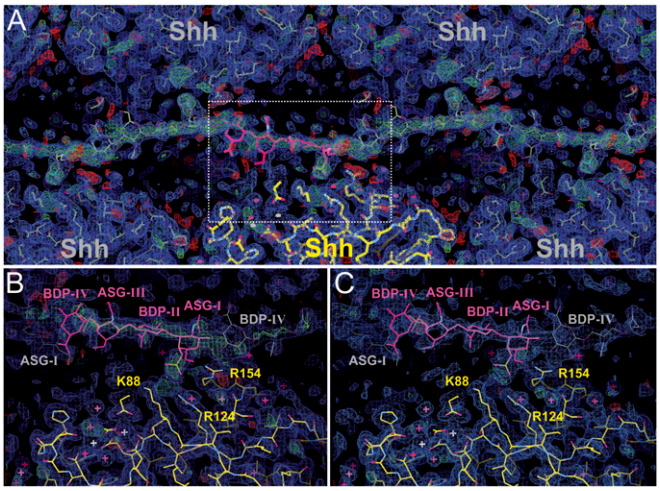

Fig. S7. Structure determination of the $ShhN_{\Delta 39}$–C4S complex. (*A*) View of the crystal packing showing the Shh molecule of the asymmetric unit in yellow. The $2F_o–F_c$ (0.6 σ, blue) and $F_o–F_c$ (±2.7 σ, green and red, respectively) electron density maps calculated with the C4S chain omitted are shown. As observed for the $ShhN_{\Delta 39}$–heparin structure, a large patch of continuous density (in this case corresponding to the C4S molecule) is evident. (*B* and *C*) Close-up view of the box in *A*, showing the Shh-binding region for the C4S. A C4S tetrasaccharide unit accounts for the observed electron density. ASG, *N*-acetylgalactosamine-4-sulfate; BDP, glucuronic acid. Saccharide molecules, which are part of the asymmetric unit are colored in magenta and their symmetry mates in gray. In *B*, the electron density maps are calculated as in *A*. In *C*, the $2F_o–F_c$ (0.6 σ, blue) and $F_o–F_c$ (±2.7 σ, green and red, respectively) maps after the final round of refinement with included C4S tetrasaccharide are shown. Mouse Shh residues K88, R124, and R154, which form part of the Shh core GAG-binding site, are marked.

Table S1. Data collection and refinement statistics

	$ShhN_{\Delta 39}$–heparin complex	$ShhN_{\Delta 39}$–C4S complex
Data collection		
Space group	$P2_12_12$	$P2_12_12_1$
Cell dimensions, Å	a = 95.7	a = 40.4
	b = 61.6	b = 55.7
	c = 62.3	c = 71.0
Resolution	22.00–2.35 (2.42–2.35)	43.80–1.74 (1.79–1.74)
Completeness, %	99.8 (99.9)	99.9 (100)
R_{merge}, %	6.5 (62.7)	8.2 (70.0)
I/σ(I)	18.8 (2.7)	15.3 (2.7)
Multiplicity	6.4 (5.5)	6.9 (7.0)
Refinement		
Resolution range, Å	22.00–2.35 (2.42–2.35)	43.80–1.74 (1.84–1.74)
No. reflections	14,900 (1,070)	14,617 (2,608)
R_{work}, %	20.2 (30.6)	16.0 (20.1)
R_{free}, %	23.2 (34.8)	19.3 (27.8)
No. atoms (protein/Zn/Ca/GAG/water)	2,434/2/4/216/64	1,228/1/2/60/129
B factors, $Å^2$ (protein/Zn/Ca/GAG/water)	69/45/50/125/51	18/8/15/32/78
rmsd		
Bond lengths, Å	0.010	0.006
Bond angles, °	1.413	1.118
Ramachandran statistics*		
Favored, %	96.4	98.0
Disallowed, %	0	0
MolProbity score*	1.38 (100th percentile)	1.11 (99th percentile)
MolProbity clashscore*	3.72 (99th percentile)	1.21 (99th percentile)

Each structure was determined from one crystal. Numbers in parentheses refer to the highest resolution shell. R_{free} equals the R factor against 5% of the data.
*Calculated with MolProbity.