Supporting Information

Lin et al. 10.1073/pnas.1010689108

SI Materials and Methods

Data Collection and Structure Determination. Screening of betaglycan zona pellucida (ZP)-C crystals was performed on beamline 8.3.1 at the Advanced Light Source (for native crystals) and beamlines 9-2 and 11-1 at the Stanford Synchrotron Radiation Laboratory (for heavy atom derivative crystals). All data were indexed, integrated, and scaled using the programs Denzo and Scalepack in the HKL2000 suite (1). Phases were calculated from K_2PtCl_4, K_2OsO_4, and NaBr derivatives by multiple isomorphous replacement with anomalous scattering using the program Sharp (2). Density-modified maps, derived from DM in the CCP4 suite (3), were calculated with the assumption of 48.3% solvent content and two ZP-C molecules per asymmetric unit. Histogram matching, solvent flattening, and twofold noncrystallographic symmetry averaging were used to obtain the initial density-modified maps.

Initial model building was done using ARP/warp (4). Subsequently, iterative rounds of coordinate and *B*-factor refinement were done with Refmac (5) or Phenix (6), interspersed with manual model building in Coot (7). From the 173 betaglycan ZP-C residues (591–763), 167 residues (591–757) in chain A and 163 residues (591–732, 737–757) in chain B were modeled. Electron density for N-linked carbohydrate was observed and modeled on N591 on both chains A and B. The final model was refined to 2.0 Å with R_{work} and R_{free} values of 18.6 and 24.5%, respectively. The geometry of the carbohydrate moieties was assessed using pdb-care (8), and the geometry of the protein model was assessed using MolProbity (9), PROCHECK (10), and WHAT_CHECK (11).

Sequence and Structure Analysis. Structure-based multiple sequence alignment was assembled from results obtained from STRAP (12), Expresso (13), and PROMALS3D (14). In the low-homology AB, CD, and FG loop regions, additional multiple sequence alignments with COBALT (15) and CLUSTAL W (16) as well as pairwise alignments using National Center for Biotechnology Information Blast2Seq (17) were used to help guide the alignment process. Published external hydrophobic patch (EHP) sequences for ZP1-3 (18, 19) and uromodulin (20) were used to help align the ZP-C G strand region. ZP-N and ZP-C domain structure overlay was performed using secondary-structure matching superposition (21) in Coot. Poisson–Boltzmann electrostatic calculations were generated using APBS (22) using input generated from PDB2PQR (23) and colored according to potential on solvent accessible surface. All structural figures were generated using PyMol (Schrödinger, LLC).

Betaglycan ZP-C Binding Assays. Affinity measurements between betaglycan ZP-C domain, full ZP region, and full-length ectodomain to biotinylated inhibin were measured in an ELISA. Inhibin-A was obtained from stable CHO cell lines (24) and biotinylated prior to the assay using EZ-Link NHS-chromogenic biotinylation kit (Pierce/Thermo Fisher) as instructed by the manufacturer. The molar ratio of biotin per inhibin used in this assay ranges from 1.0–1.5 as determined by absorbance at 354 nm for chromogenic biotin and 280 nm for inhibin. The ZP-C domain, the full ZP region, and the full-length ectodomain were immobilized on a 96-well plate at 100, 150, and 350 ng/well, respectively, overnight at 4 °C. The wells were blocked using 0.5% bovine serum albumin in phosphate-buffered saline (PBS) for 4 h at 37 °C. Subsequently, the wells were sequentially incubated with 100 μL/well of biotinylated inhibin at the indicated concentrations for overnight at 4 °C and 100 μL/well of horseradish peroxidase (HRP)-conjugated streptavidine (R&D Biosystems) for 1 h at room temperature, with washing steps using PBS containing 0.05% Tween-20. Binding was visualized by using the HRP substrate, 3,3′,5,5′-tetramethylbenzidine (Invitrogen), at 650 nm as recommended by the manufacturer. The Kd values were determined by performing a nonlinear curve fitting assuming one-site binding on Prism 5 (GraphPad Software, Inc.).

1. Otwinowski Z, Minor W (1997) *Processing of X-ray Diffraction Data Collected in Oscillation Mode* (Academic, New York) pp 307–326.
2. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364:215–230.
3. CCP4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D*50:760–763.
4. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3:1171–1179.
5. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
6. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66:213–221.
7. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. Acta Crystallogr D 60:2126–2132.
8. Lutteke T, von der Lieth CW (2004) pdb-care (PDB carbohydrate residue check): A program to support annotation of complex carbohydrate structures in PDB files. *BMC Bioinformatics* 5:69.
9. Davis IW, Murray LW, Richardson JS, Richardson DC (2004) MOLPROBITY: Structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 32:W615–W619.
10. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231:1049–1067.
11. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272.
12. Gille C, Frommel C (2001) STRAP: Editor for structural alignments of proteins. *Bioinformatics* 17:377–378.
13. Armougom F, et al. (2006) Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* 34:W604–W608.
14. Pei J, Grishin NV (2007) PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* 23:802–808.
15. Papadopoulos JS, Agarwala R (2007) COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23:1073–1079.
16. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
18. Zhao M, et al. (2003) Mutation of a conserved hydrophobic patch prevents incorporation of ZP3 into the zona pellucida surrounding mouse eggs. *Mol Cell Biol* 23:8982–8991.
19. Jovine L, Qi H, Williams Z, Litscher ES, Wassarman PM (2004) A duplicated motif controls assembly of zona pellucida domain proteins. *Proc Natl Acad Sci USA* 101:5922–5927.
20. Schaeffer C, Santambrogio S, Perucca S, Casari G, Rampoldi L (2009) Analysis of uromodulin polymerization provides new insights into the mechanisms regulating ZP domain-mediated protein assembly. *Mol Biol Cell* 20:589–599.
21. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D* 60:2256–2268.
22. 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* 98:10037–10041.
23. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res* 32:W665–W667.
24. Pangas SA, Woodruff TK (2002) Production and purification of recombinant human inhibin and activin. *J Endocrinol* 172:199–210.

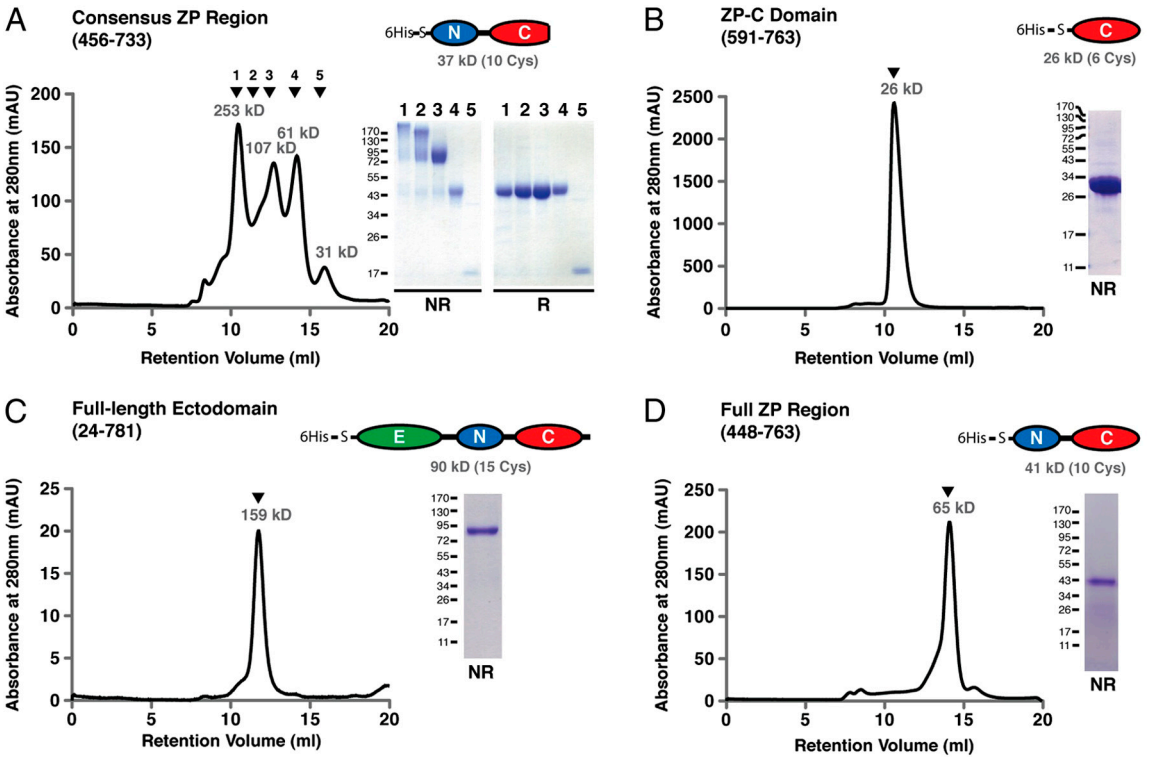

Fig. S1. Expression and purification of betaglycan. Size exclusion chromatography results of purified betaglycan extracellular regions incorporating (*A*) the consensus ZP region (456–733), (*B*) the ZP-C domain (591 –763), (*C*) the full-length ectodomain (24–781), and (*D*) the full ZP region (448–763). The molecular weights corresponding to the major peaks are based on calibrated molecular weight standards run on Superdex 200 (*A*,*C*, and *D*) and Superdex 75 (*B*) columns. The schematic diagrams on the upper right corner of each panel indicate the individual betaglycan extracellular domains: green (E) complete oval represents the endoglin-homology domain, blue (N) complete oval represents the ZP-N domain, red (C) incomplete oval represents the ZP-C region without the EHP, and red (C) complete oval represents the full ZP-C domain. Below the schematic, the expected molecular weight of each protein, taking into account the N-terminal 6His- and S-tags but not the carbohydrates, is shown along with the total number of cysteines. The SDS-PAGE gels on the bottom right corner of each panel are analyses of peak fraction(s) indicated by the inverted black triangle(s) on the size exclusion chromatogram. NR, nonreducing condition; R, reducing condition.

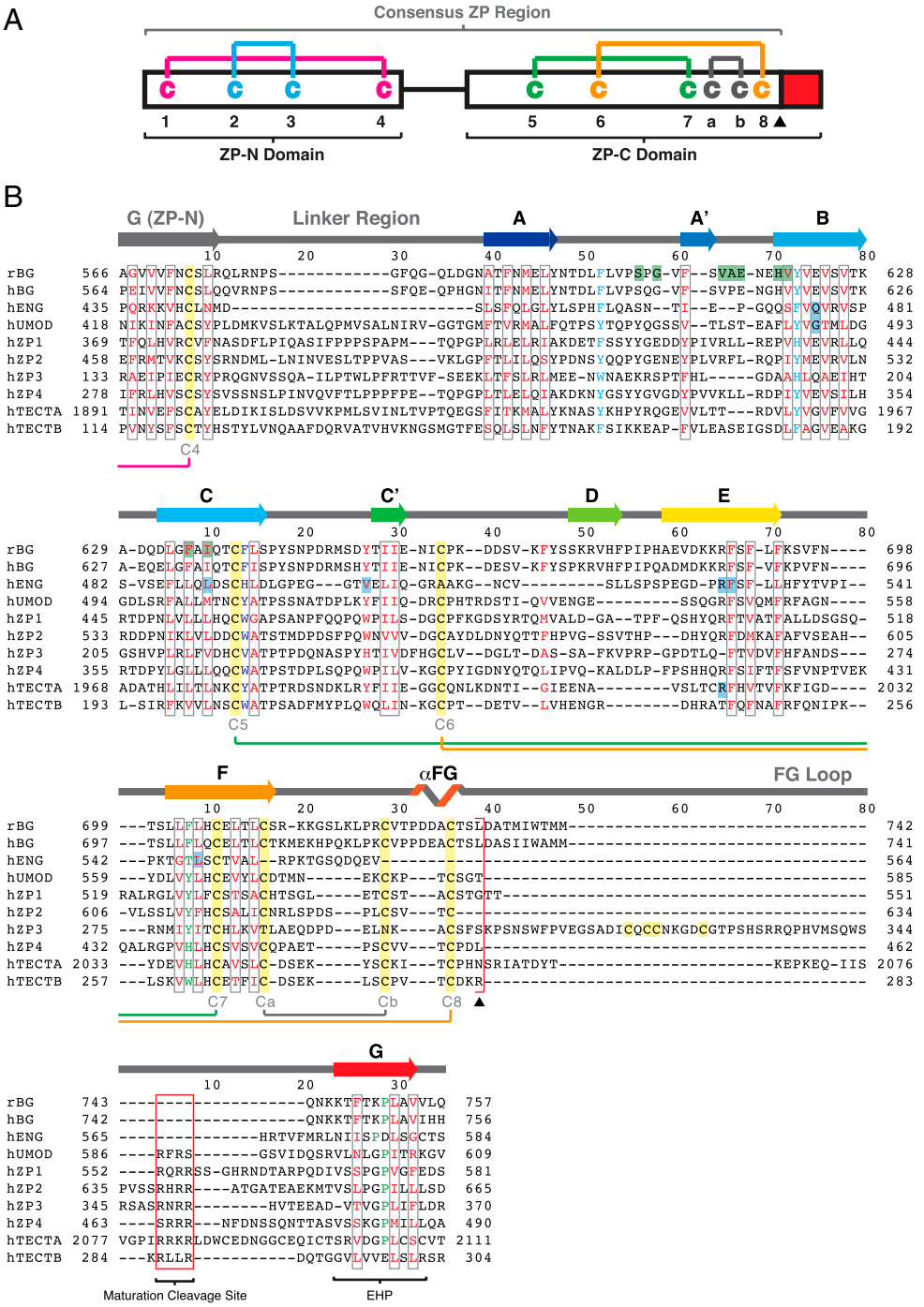

Fig. S2. Structure-based alignment of ZP proteins. (*A*) Schematic representation of the disulfide linkage pattern in betaglycan ZP-N and ZP-C domains. The numbers 1–8 and letters a/b represent the position of the conserved cysteine residues, the black triangle denotes the end of the consensus ZP region, and the red rectangle denotes the EHP region. (*B*) Structure-based alignment of ZP-C domains from several well-characterized ZP proteins: rBG and hBG, rat and human betaglycan; hENG, human endoglin; hUMOD, human uromodulin; hZP1-4, human ZP proteins; hTECTA and hTECTB, human α- and β-tectorins. Sequences from the ZP-N G strand and the interdomain linker region are shown as reference. Open gray boxes indicate inward-facing residues on β-strands, open red box indicates the maturation cleavage site, and open red bracket and black triangle indicate the C-terminus of the consensus ZP region. Solid yellow boxes indicate the conserved cysteine residues. Solid green boxes indicate residues implicated in TGF-β ligand binding (see Fig. S4). Solid blue boxes indicate residues that have naturally occurring mutations leading to pathological phenotypes (see Fig. S5). Amino acids colored in red indicate residues with buried side chains. Amino acids colored in cyan denote conserved hydrophobic contact residues between the AB loop and the B strand. Amino acids colored in blue on the C strand indicate outward-facing aromatic residues that are masked by the CD loop. Amino acids colored in green indicate conserved hydrophobic contact residues between the F and G (EHP) strands.

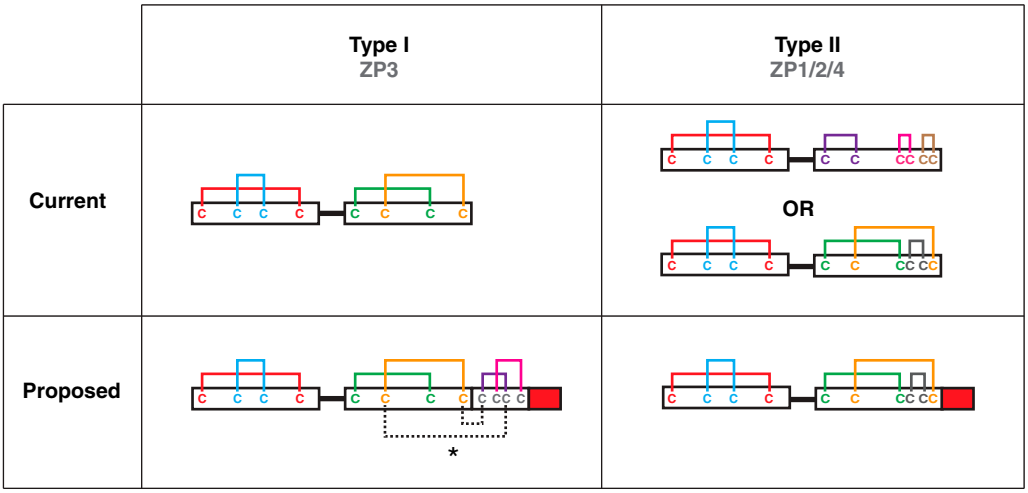

Fig. S3. Domain boundaries and disulfide linkage arrangements in ZP proteins. Schematic representation of the current and the proposed type I and II ZP protein domain boundaries and disulfide linkages. The type II ZP proteins, including ZP1, ZP2, ZP4, uromodulin, α-/β-tectorins, and betaglycan, have been predicted to have two different disulfide arrangements. The betaglycan ZP-C domain structure is consistent with only one of these. We propose a revision to the ZP-C domain boundaries that includes the EHP region (red rectangle) corresponding to the betaglycan ZP-C G strand. The four extra cysteines of type I ZP proteins, which is represented by ZP3, are the four additional conserved cysteine residues found in the ZP3 FG loop insertion. The asterisk denotes an alternative disulfide linkage pattern observed in a recent crystal structure of chicken ZP3 (1), with distinctive C6–C11 and C8–C9 disulfide bonds that are shown as dashed lines.

1 Han L, et al. (2010) Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell* 143:404–415.

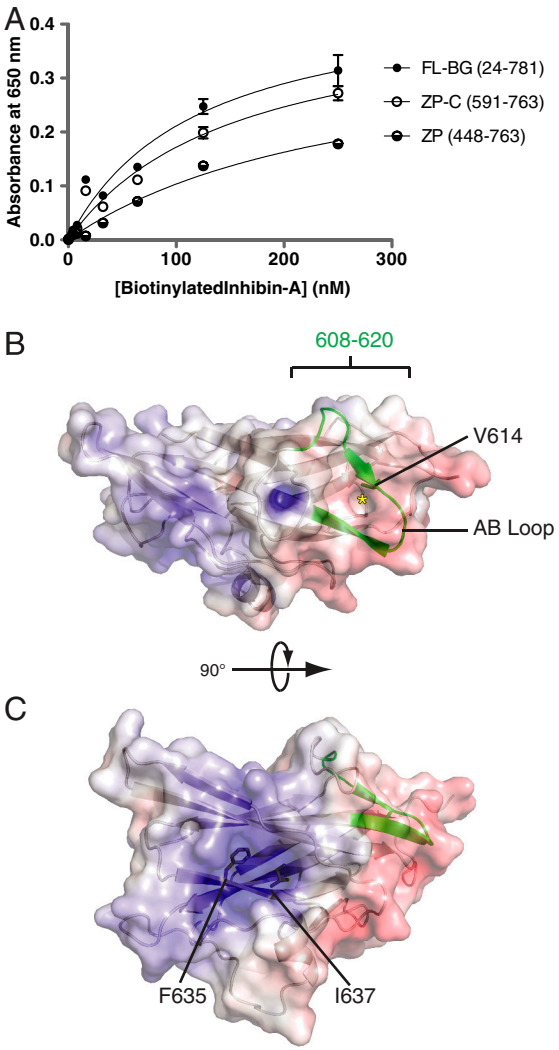

Fig. S4. Betaglycan ZP-C Domain Interaction with Inhibin. (*A*) Equilibrium binding of soluble biotinylated inhibin to equal molar amounts of immobilized betaglycan full-length ectodomain (BG), full ZP region (ZP), and ZP-C domain (ZP-C); Kd for BG $= 119 \pm 30$ nM, ZP $= 266 \pm 59$ nM, and ZP-C $= 154 \pm 37$ nM. The experiment was performed in triplicates and the error bars represent the standard error mean. (*B* and *C*) Amino acids implicated in inhibin and TGF-β1 ligand binding include (*B*) residues 608–620 (green), which map mostly onto the ZP-C AB loop and form a small hydrophobic pocket beside V614 (denoted by the yellow asterisk) and (*C*) residues F635 and I637, which map onto the basic pocket on the ZP-C convex surface. The surface representation is colored by Poisson–Boltzmann electrostatic calculations contoured from −10 (red) to 10 (blue) kT/e.

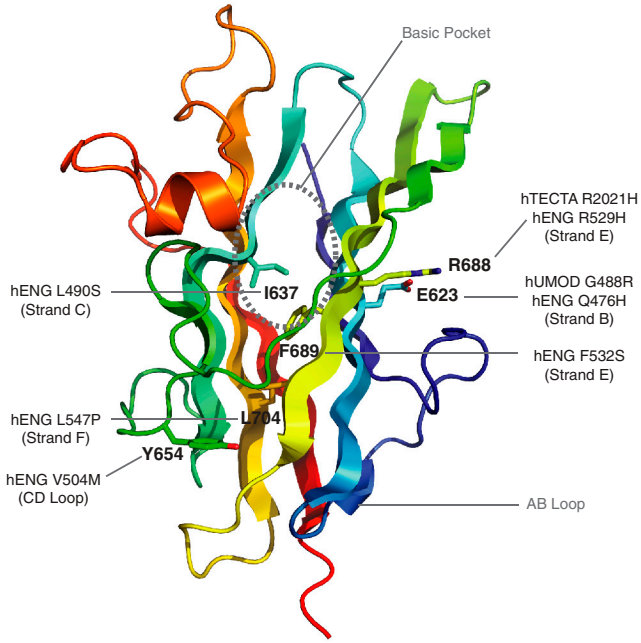

Fig. S5. Three-dimensional mapping of ZP-C domain mutations. Known missense ZP-C mutations that lead to pathological phenotypes are shown and mapped onto the current betaglycan ZP-C structure. hENG, human endoglin; hUMOD, human uromodulin; hTECTA and hTECTB, human α- and β-tectorins. With exception of the hENG V504M mutation, which maps onto a conserved inward-facing residue on the CD loop that contacts the hydrophobic core, all other mutations map onto conserved β-strand regions as indicated within the parenthesis. The gray dashed line defines the location of the basic pocket on the betaglycan ZP-C domain convex surface.

Table S1. Data collection and refinement statistics

	Native	K_2PtCl_4	K_2OsO_4	NaBr
Data collection				
Source	ALS BL 8.3.1	SSRL BL 11-1	SSRL BL 9-2	SSRL BL 11-1
Wavelength, Å	1.11587	1.06698	1.13993	0.91889
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimension				
A, Å	53.53	53.02	53.07	53.84
B, Å	63.57	63.65	63.56	62.68
C, Å	107.22	107.47	108.27	106.42
Resolution (Å)	48–2.00	48–3.10	48–2.63	48–2.60
(Last shell)*	(2.07–2.00)	(3.21–3.10)	(2.72–2.63)	(2.69–2.60)
Mosaicity, °	0.27–0.41	0.46–0.75	0.34–0.62	0.40–0.80
R_{sym}, %	7.6 (38.4)	14.1 (43.6)	9.4 (49.4)	10.2 (76.7)
I/σ	13.1 (2.8)	19.6 (6.7)	19.0 (2.9)	9.7 (2.0)
Completeness, %	98.6 (97.6)	99.8 (99.3)	99.9 (99.2)	100.0 (100.0)
Redundancy	4.0 (4.0)	14.3 (14.5)	7.8 (7.5)	5.1 (5.0)
Phasing from MIRAS† data				
No. of heavy atom sites	—	15	8	16
Phasing $power_{isomorphous}$ (acentric/centric)	—	0.743/0.675	0.365/0.374	0.700/0.642
Phasing $power_{anomalous}$ (acentric)	—	0.820	1.186	0.602
Refinement				
Resolution, Å	48–2.0			
R_{work}/R_{free}, %	18.6/24.5			
No. of atoms				
Protein	2674			
Carbohydrate	140			
Water	173			
Rms deviations				
Bond length, Å	0.0195			
Bond angle, °	1.982			
Mean B value, $Å^2$	27.8			
Ramachandran plot‡				
Favored, %	96.4			
Allowed, %	3.6			
Outlier, %	0.0			

*Values in parentheses indicate the highest resolution shell.
†MIRAS, multiple isomorphous replacement with anomalous scattering.
‡Performed in MolProbity.

Table S2. Summary of ZP-C domain missense mutations shown in Fig. S5

Protein*	Mutation	Equivalent rat betaglycan ZP-C position	Predicted effect	Refs.
h*ENG*	L490S	I637	destabilizes hydrophobic core	1
			potentially affects TGF-β signaling because I637V on rat betaglycan reduces TGF-β ligand binding	2
h*ENG*	F532S	F689	destabilizes hydrophobic core	3
h*ENG*	V504M	Y654	destabilizes the CD loop because Y654 on rat betaglycan anchors the CD loop to the hydrophobic core	4, 5
h*ENG*	L547P	L704	disrupts continuity of strand F and thus destabilizes interaction with the EHP region on strand G	1
h*ENG*	Q476H	E623	Mutations on these exposed residues would be expected to affect the	6
h*UMOD*	G488R		surface charge around the important basic pocket on the convex surface.	7
h*ENG*	R529H	R688		1, 2
h*TECTA*	R2021H			8

*The protein names are defined as follows: h*ENG*, human endoglin (UniProtKB/Swiss-Prot P17813); h*UMOD*, human uromodulin (UniProtKB/Swiss-Prot P07911); h*TECTA*: human α-tectorin (UniProtKB/Swiss-Prot O75443).

1 Bossler AD, Richards J, George C, Godmilow L, Ganguly A (2006) Novel mutations in ENG and ACVRL1 identified in a series of 200 individuals undergoing clinical genetic testing for hereditary hemorrhagic telangiectasia (HHT): Correlation of genotype with phenotype. *Hum Mutat* 27:667–675.
2 Wiater E, Harrison CA, Lewis KA, Gray PC, Vale WW (2006) Identification of distinct inhibin and transforming growth factor beta-binding sites on betaglycan: functional separation of betaglycan co-receptor actions. *J Biol Chem* 281:17011–17022.
3 Gedge F, et al. (2007) Clinical and analytical sensitivities in hereditary hemorrhagic telangiectasia testing and a report of de novo mutations. *J Mol Diagn* 9:258–265.
4 Lesca G, et al. (2004) Molecular screening of ALK1/ACVRL1 and ENG genes in hereditary hemorrhagic telangiectasia in France. *Hum Mutat* 23:289–299.
5 Brusgaard K, et al. (2004) Mutations in endoglin and in activin receptor-like kinase 1 among Danish patients with hereditary haemorrhagic telangiectasia. *Clin Genet* 66:556–561.
6 Wehner LE, et al. (2006) Mutation analysis in hereditary haemorrhagic telangiectasia in Germany reveals 11 novel ENG and 12 novel ACVRL1/ALK1 mutations. *Clin Genet* 69:239–245.
7 Williams SE, et al. (2009) Uromodulin mutations causing familial juvenile hyperuricaemic nephropathy lead to protein maturation defects and retention in the endoplasmic reticulum. *Hum Mol Genet* 18:2963–2974.
8 Iwasaki S, et al. (2002) Association of clinical features with mutation of TECTA in a family with autosomal dominant hearing loss. *Arch Otolaryngol* 128:913–917.