# LUND UNIVERSITY

**Crystal structure of N-glycosylated human glypican-1 core protein: Structure of two loops evolutionarily conserved in vertebrate glypican-1.**

Svensson Birkedal, Gabriel; Awad, Wael; Håkansson, Maria; Mani, Katrin; Logan, Derek

Link to publication

Total number of authors:
5

# Crystal Structure of N-Glycosylated Human Glypican-1 Core Protein: Structure of two loops evolutionarily conserved in vertebrate glypican-1

**Gabriel Svensson[a], Wael Awad[b,c], Maria Håkansson[d], Katrin Mani[a], Derek T. Logan[b*]**

[a]Department of Experimental Medical Science, Division of Neuroscience, Glycobiology Group, Lund University, Biomedical Center A13, SE-221 84, Lund, Sweden.
[b]Department of Biochemistry and Structural Biology, Center for Molecular Protein Science, Lund University, Box 124, SE-221 00 Lund, Sweden.
[c]Biophysics Department, Faculty of Science, Cairo University, 12613 Giza, Egypt.
[d]SARomics Biostructures AB, Medicon Village, SE-223 81, Lund, Sweden.

[*]Corresponding author: phone, +4646-2221443; fax.: +4646-2224116; e-mail: derek.logan@biochemistry.lu.se.

Running title: Crystal Structure of Human Glypican-1 Core Protein

**Background:** The glypican family of cell-surface proteoglycans regulates growth factor signaling during development through their core proteins and heparan sulfate chains.
**Results:** The crystal structure of N-glycosylated human glypican-1 is described.
**Conclusions:** The structure reveals the complete disulfide bond arrangement for the conserved Cys residues in glypicans.
**Significance:** Increased structural knowledge of glypicans will help elucidate their important functions in shaping animal development.

## SUMMARY
**Glypicans are a family of cell-surface proteoglycans that regulate Wnt, hedgehog, bone morphogenetic protein, and fibroblast growth factor signaling. Loss-of-function mutations in glypican core proteins and in glycosaminoglycan-synthesizing enzymes have revealed that glypican core proteins and their glycosaminoglycan chains are important in shaping animal development. Glypican core proteins consist of a stable α-helical domain containing 14 conserved Cys residues followed by a glycosaminoglycan attachment domain that becomes exclusively substituted with heparan sulfate (HS) and presumably adopts a random coil conformation. Removal of the α-helical domain results in almost exclusive addition of the glycosaminoglycan chondroitin sulfate, suggesting that factors in the α-helical domain promote assembly of HS. Glypican-1 is involved in brain development and is one of six members of the vertebrate family of glypicans. We expressed and crystallized N-glycosylated human glypican-1 lacking HS and N-glycosylated glypican-1 lacking the HS attachment domain. The crystal structure of glypican-1 was solved using crystals of selenomethionine labelled glypican-1 core protein lacking the HS domain. No additional electron density was observed for crystals of glypican-1 containing the HS attachment domain, and CD spectra of the two protein species were highly similar. The crystal structure of N-glycosylated human glypican-1 core protein at 2.5 Å, the first crystal structure of a vertebrate glypican, reveals the complete disulfide bond arrangement of the conserved Cys residues, and it also extends the structural knowledge of glypicans for one α-helix and two long loops. Importantly, the loops are evolutionarily conserved in vertebrate glypican-1 and one of them is involved in glycosaminoglycan class determination.**

## Introduction

Proteoglycans (PGs) consist of a core protein covalently substituted with glycosaminoglycans (GAGs), long unbranched polysaccharide chains. Several genetic diseases in humans and numerous genetic experiments in fruit flies, nematodes, and mice have revealed that GAGs and their core proteins are important in shaping animal development (1,2). Glypicans are a family of cell-surface PGs substituted with the GAG heparan sulfate (HS). Glypicans regulate growth factor signalling by binding of growth factors to their HS chains, or as has been shown recently, by binding of growth factors to the core proteins. Based on sequence comparisons glypicans fall into two subfamilies: glypicans 1, 2, 4, 6 and glypicans 3 and 5, with approximately 25% amino acid identity between the groups. Glypicans are around 550 amino acids in length and all share an N-terminal secretory signal peptide, 14 evolutionarily conserved Cys residues, a HS attachment domain near the C-terminus and a hydrophobic domain for addition of the glycosylphosphatidylinositol (GPI) anchor at the C-terminus. Aside from the six members of the glypican family in vertebrates, there are also two glypicans in *Drosophila melanogaster* (*D. melanogaster*, Dally and Dally-like protein) and two in *Caenorhabditis elegans* (*C. elegans*, Gpn-1, Lon-2*)*. Dally-like protein (DLP) is divergent from the other glypicans, in particular by the insertion of 100 amino acids N-terminal to the last two conserved Cys residues (Fig. S1). Loss-of-function mutations in *GPC3* cause Simpson-Golabi-Behmel syndrome, an X-linked condition characterized by prenatal and postnatal overgrowth (3). Mutations in *GPC6* cause omodysplasia-1, a congenital disease affecting the skeletal system (4). Clinical features of omodysplasia-1 involve short stature and facial dysmorphism. Also, common variation in *GPC5* is associated with acquired nephrotic syndrome (5). GPC1 is expressed in the developing and adult brain, and *GPC1* knock-out mice have reduced brain size at birth (6).

The structurally related but functionally distinct GAGs HS and chondroitin sulfate (CS) are added to core proteins that contain a GAG attachment motif consisting of the dipeptide Ser-Gly. Synthesis is initiated by the addition of a xylose (Xyl) residue to the hydroxyl group of the Ser residue in the GAG attachment domain, followed by sequential addition of two galactose

(Gal) and one glucuronic acid (GlcA) residue. This forms the linker region common for HS and CS: GlcAβ1-3Galβ1-3Galβ1-4Xylβ1-O-Ser. Synthesis of HS is initiated by the addition of an N-acetylglucosamine (GlcNAc) residue, while CS synthesis begins with the addition of an N-acetylgalactosamine (GalNAc). After the addition of the GlcNAc or GalNAc residues, HS and CS chains are polymerized, forming their respective disaccharide backbones. Exactly what determines whether a GlcNAc residue or a GalNAc is added to the common linker region is not known. It has been suggested that HS assembly is promoted by the repetition of the Ser-Gly dipeptide and by flanking acidic residues (7,8). Glypicans are substituted with HS and contain GAG attachment domains similar to those suggested to promote HS assembly. However, expression of the rat GPC1 GAG attachment domain alone results in substitution with ~90% CS (9), strongly suggesting that factors in the glypican core protein further away from the GAG attachment sites promote the assembly of HS.

Recently it has been shown that functional aspects of glypicans also reside in the protein cores. For example, Hedgehog growth factor signalling is deficient in *D. melanogaster* animals when the *D. melanogaster* glypican *Dlp* has been knocked down, and the Hedgehog response can be rescued by expression of DLP without the HS chains (10,11). Moreover, other reports have shown that glypican core proteins bind directly to growth factors such as Hedgehog (12), bone-morphogenetic protein 4 (13), decapentaplegic, Wnt (14,15), and insulin-like growth factor II (16) with significant biological function independent of the presence of HS chains.

We have previously shown that anchorless recombinant human GPC1 (rhGPC1) is a stable α-helical protein that resists high concentrations of the chemical denaturants urea and guanidine-HCl (17). Unfolding occurrs in a single step and unfolding data can be closely fitted by a two-state model, indicating that GPC1 is a densely packed globular protein. Recently, the core protein structure of DLP was published, revealing an extended α-helical fold for glypican core proteins (18). Here we present the independently solved structure of GPC1 core protein using multiple-wavelength anomalous dispersion (MAD) phasing on selenomethionine (SeMet) labelled rhGPC1, the first crystal structure of a vertebrate glypican. Structural

2

alignment of GPC1 and DLP reveals a common fold for glypican core proteins. The GPC1 structure presented here also expands our knowledge of glypican core proteins, in particular for the disulfide bond arrangement of the conserved Cys residues, one α-helix and two long loops. One of the loops has been shown to be involved in GAG class determination, but lack of structural data has previously hampered further mutagenic investigations.

## Experimental Procedures

**Materials.** Human *GPC1* cDNA clone IMAGE ID 6275649 was obtained from Geneservice (UK). The expression vector pCEP-BM40-HisTEV and 293 human embryonic kidney (HEK293) cells were obtained as described previously (17). All restriction enzymes used were from Fermentas (Canada). Ni-NTA columns were from GE Healthcare Life Sciences (UK). Protein-free medium for CHO cells and Coomassie blue were from Sigma-Aldrich (MO). Lipofectamine, 4–12% Bis-Tris SDS-PAGE gels, L-methionine free Dulbecco's modified Eagle medium, dialyzed fetal bovine serum (FBS), and hygromycin B were from Invitrogen (CA). L-selenomethionine was from Acros Organics (Belgium). Crystallization plates and polyethylene glycol $M_w$ 6000 (PEG 6000) were obtained from Hampton Research (CA). Other materials related to crystallography were purchased from Molecular Dimensions (UK).

   **Cloning and Purification of Human GPC1.** The cDNA sequence coding for the mature form of human GPC1, amino acids 24-529, i.e. without the sequences for the N- and C-terminal signal peptides, was amplified and introduced into the vector pCEP4-BM40-HisTEV as described previously (17). This vector contains the sequence for an N-terminal BM40 secretion peptide followed by a His$_6$ tag and a tobacco etch virus protease cleavage site. Further, the HS attachment sites were disrupted by site-directed mutagenesis in order to create a vector expressing GPC1 ΔHS (17). A vector coding for GPC1 ΔC was created by amplifying the sequence for amino acids 24-479 using the forward primer 5'-TT**AAGCTT**GACCCGGCCAGCAAGAG-3' and the reverse primer 5'-TT**GGATCC**TTAGAAGTCCACGTCGTTGC CGTT-3' and by introducing the PCR-product into the pCEP4-BM40-HisTEV plasmid. Restriction enzyme sites are given in boldface.

All constructs were verified by sequencing at Eurofins MWG Operon (Germany). Stable HEK 293 cell clones expressing GPC1 ΔHS and GPC1 ΔC were generated by limiting dilution after several weeks of growth in selective antibiotics. Stable cells clones expressing large amounts of rhGPC1 were grown to confluence in minimal essential medium with Earle's salts supplemented with 10% (v/v) FBS, 2 mM L-glutamine, penicillin (100 U/ml), streptomycin (100 µg/ml) and hygromycin B (200 µg/ml). After extensive washing with PBS, the rhGPC1 expressing cells were maintained for three or four days in protein-free medium supplemented with 2 mM L-glutamine, penicillin (100 U/ml), streptomycin (100 µg/ml) and hygromycin B (100 µg/ml). Conditioned medium was stored frozen and rhGPC1 was purified from conditioned medium of ~200 mL batches using nickel affinity chromatography. Protein purification was analyzed by SDS-PAGE and purified rhGPC1 was dialyzed to 20 mM Tris-HCl pH 8.0, concentrated by ultrafiltration and stored frozen. SeMet labelling of GPC1 ΔC was performed in L-methionine free Dulbecco's modified Eagle medium supplemented with 2% dialyzed FBS. After a 6 h pre-incubation with L-methionine free medium, the medium was replaced with medium containing 30 mg/l L-selenomethionine. Cells suffered from SeMet toxicity, which made it difficult to grow the cells in medium supplemented with SeMet for more than 24 h. SeMet labelled GPC1 ΔC was purified in the presence of 0.1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) to avoid oxidation of SeMet residues (19).

   **CD Spectroscopy.** CD spectra were recorded using a Jasco J-810 spectropolarimeter equipped with a Peltier thermostated cell holder. The cuvette was thermostated at 20°C. For far-UV CD, a 1-mm quartz cuvette was used, and the measurement range was 190–250 nm. The following parameters were used: sensitivity, 100 millidegrees (mdeg); data pitch, 1 nm; scan rate, 100 nm/min; response time, 4 sec; band width, 1 nm; and accumulation, 2. Near-UV CD was measured in 10 mm quartz cuvette, and the measurement range was 250-340 nm. The following parameters were used: sensitivity, 100 mdeg; data pitch, 0.2 nm; scan rate, 50 nm/min; response time, 4 sec; band width, 0.5 nm; and accumulation, 4. Baseline spectra were recorded for pure buffer and subtracted from the corresponding protein spectra. Stock solutions of rhGPC1 were clarified by centrifugation at

15,000 × *g* for 15 min and protein concentration was determined by measuring the absorbance at 280 nm using a micro-volume spectrophotometer (GE Healthcare Life Sciences) and the molar extinction coefficient for bovine serum albumin. Samples were prepared by diluting the rhGPC1 stock solution to 0.1 mg/ml for far-UV CD and to 1 mg/ml for near-UV CD. Protein was maintained in 20 mM sodium phosphate buffer (pH 7.4). Mean residue ellipticity was calculated using 110.5 and 112.0 as the mean residue molecular weights of GPC1 ΔHS and GPC1 ΔC, respectively.

**Crystallization.** Initial high-throughput crytallization screens performed with GPC1 ΔHS at a protein concentration of 40 mg/ml were set up in MRC plates using a Mosquito nanolitre pipetting robot (TTP Labtech, UK) at the MAX-lab Crystallization Facility, Lund University. The initial screens, PACT premier (Molecular Dimensions), Crystal Screen 1 and 2 (Hampton Research) and JCSG+ (Qiagen GmbH, Germany), were set up using the vapour diffusion technique in sitting drops where 100 nl of protein solution were mixed with 100 nl reservoir solution and equilibrated against 80 μl reservoir solution. After one day needle clusters began to appear under conditions containing polyethylene glycol $M_w = 3350$ (PEG 3350) with anions such as Br⁻, F⁻ and NO₃⁻ (pH 6.5). After one week, crystal plates appeared in drops with 20 % PEG 6000, 0.2 M CaCl₂ in 0.1 M Tris-HCl (pH 8.0). After two weeks, the same appeared in drops with 20 % PEG 6000, 0.2 M NH₄Cl in 0.1 M Tris-HCl (pH 8.0). Crystal-like material also appeared in 20 % PEG 8000, 0.2 M Mg acetate, 0.1 M Na cacodylate trihydrate (pH 6.5) after two weeks. Optimisation screens were produced from stock solutions using a liquid handling pipetting robot (Tecan Group Ltd, Switzerland) and set up in MRC plates using the nanolitre pipetting robot as for the initial screens. The best results were obtained with the crystals grown in CaCl₂ with PEG 6000 as precipitating agent. The optimized conditions were scaled up to microlitre drops in 24-well plates using the sitting drop vapour diffusion method at 20°C with 500 μl reservoir solution. Typically, crystals of rhGPC1 were grown by mixing 2 μl of protein solution with 2 μl of reservoir solution containing 12-14% PEG 6000, 0.1 M Tris-HCl, 0.2 M CaCl₂, pH 8.0. GPC1 ΔC was crystallized at 20 mg/ml. GPC1 ΔHS was crystallized at higher protein concentration, up to 40 mg/ml and

benefited from low concentrations of dithiothreitol in the protein solution. In general, crystallization of GPC1 ΔC was more easily reproduced in comparison to crystallization of GPC1 ΔHS. Crystallization of SeMet labelled GPC1 ΔC was also performed using the same conditions but with lower protein concentration and the protein was maintained in 20 mM Tris-HCl (pH 8.0) supplemented with 0.1 mM TCEP-HCl. Large but thin plate-like crystals with dimensions around 0.6 x 0.3 mm formed within one week for GPC1 ΔHS and GPC1 ΔC. Crystals of SeMet labelled GPC1 ΔC were typically smaller, around 0.2 x 0.2 mm. Crystallization of rhGPC1 was not affected by the presence of the His₆ tag.

**Data Collection and Processing.** Crystals were soaked in 15% ethylene glycol, 18% PEG 6000, 0.1 M Tris-HCl, 0.2 M CaCl₂ (pH 8.0) and flash frozen under a stream of liquid nitrogen. X-ray diffraction data were collected on stations I911-2 and I911-3 of the MAX-II synchrotron radiation source, Lund, Sweden, on 165mm and 225mm CCD detectors from MarResearch (Germany). All data sets were integrated and scaled using the XDS suite (20) and further processed using programs from the CCP4 (21) and Phenix (22) suites. Data collection statistics are given in Table I. Diffraction patterns of GPC1 ΔHS crystals could be indexed in space group $P2_1$ and had unit cell dimensions $a = 47.2$ Å, $b = 169.0$ Å $c = 151.6$ Å, $β = 95.0°$. Variation in cell dimensions was seen between different crystals, in particular for the *c* dimension, with lengths ranging from 148-154 Å. The crystals contain four GPC1 molecules in the asymmetric unit, giving a Matthews volume of 2.7 Å³ / Da and a solvent content of 54%. Crystals of GPC1 ΔC and SeMet GPC1 ΔC also belong to space group $P2_1$ with similar unit cell dimensions. The diffraction anisotropy of the data from GPC1 ΔC crystals was analysed using phenix.xtriage (22) and found to be significant, with a factor of two difference between c* direction and the a* and b* directions. This effectively limited the resolution to around 2.9 Å in the c* direction. However refinement against data truncated using the UCLA anisotropy server (23) did not result in improved maps, so the original data were retained. The data also show a significant translational non-crystallographic symmetry, with a peak at (0.5, 0.021, 0.5) in the native Patterson of 35% of the height of the origin peak. This translation relates chains A and B to C and D in the asymmetric unit respectively.

**Structure Determination and Refinement**. X-ray data were collected to 3.3 Å resolution from SeMet GPC1 ΔC crystals at selenium peak and inflection wavelengths at station I911-3 of MAX-II. The processed data sets were sent to the AutoRickshaw server at the European Molecular Biology Laboratory in Hamburg (24,25) for structure determination using the two-wavelength multiple anomalous dispersion method. In Auto-Rickshaw, the data sets were prepared using the CCP4 programs CAD and SCALEIT (21,26). Further, heavy atoms were found using SHELXD (27) and heavy atom refinement and phase calculation were performed using SHARP (28) and MLPHARE (21). Initial density modification and phase extension were performed using DM (29) and further explored using RESOLVE (30). Finally, the electron density map was interpreted using HELICAP (31) and model building was performed by BUCCANEER (32). This resulted in a more than 90% complete model built at 3.8 Å that was partially refined by CNS (33) and Refmac5 (34) at 3.3 Å ($R_{model}$ = 23% and $R_{free}$ = 32%). The selenium positions were confirmed by calculating an anomalous difference map at peak wavelength using phases from a partially refined model. The initial structure was improved by extensive manual rebuilding in Coot (35) followed by many rounds of refinement using the diffraction data to 2.55 Å from a native GPC1 ΔC crystal. Phenix.refine (22) with Ramachandran restraints was used for the final model refinement and Molprobity (36) was used for validation. Structural alignments were made using the secondary-structure matching algorithm as implemented in Coot. All graphical representations of the GPC1 structure were generated using the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger.

**Structural Conservation Analysis.** Alignments of glypican amino acid sequences were performed using ClustalW2 (37). The amino acid sequences used are given in the figure legend of Fig. 5. DNA sequences were translated using the ExPASy Translate Tool (web.expasy.org/translate). Structural conservation analysis was performed using the ConSurf server (38).

## Results

**Expression and Characterization of Anchorless GPC1.** We expressed His$_6$-tagged full-length rhGPC1 without the GPI-anchor signal peptide in 293 human embryonic kidney cells. This secreted rhGPC1 (protein referred to as GPC1 WT) was purified under native conditions and analysis on SDS-PAGE of the purified protein showed a core protein band of ~60 kDa and a PG smear of ~60-120 kDa (Fig. 1A, lane 1). Degradation with HS lyase and PNGase F showed that this protein was substituted exclusively with HS and also contained N-linked glycans (17,39). In order to produce rhGPC1 for conformational and structural studies, the HS attachment sites were disrupted by site-directed mutagenesis, resulting in expression of core protein containing N-linked glycans with a molecular weight of ~60 kDa (protein referred to as GPC1 ΔHS, Fig. 1A, lane 2). Also, the HS attachment site in GPC1 is regarded, by secondary structure and disorder predictions, as adopting a random coil conformation despite the fact that it contains two Cys residues. Therefore, rhGPC1 was truncated C-terminally before the HS attachment domain, resulting in a core protein band (GPC1 ΔC ) of smaller size (~55 kDa) on SDS-PAGE (Fig. 1A, lane 3). We have previously shown that the PG form of GPC1 WT and GPC1 ΔHS are identical in folding, as judged by far-UV CD and fluorescence emission spectroscopy (17). In order to ensure that GPC1 ΔHS and GPC1 ΔC are also structurally similar, we subjected the two forms of rhGPC1 to far- and near-UV CD spectroscopy. Far-UV CD reports on the peptide bond conformation (i.e. secondary structure content) of the investigated protein and near-UV CD reports on the immediate environment of Trp, Tyr, and Phe residues (i.e. the tertiary structure). Both protein forms exhibited very similar far- and near-UV CD spectra with peaks and troughs located at identical wavelengths, indicating that removal of the HS attachment domain affects neither the secondary nor the tertiary structure of GPC1 (Fig. 1B and C). Also, expression levels were similar for GPC1 ΔHS and GPC1 ΔC. Together, this suggests that the C-terminal HS attachment site in GPC1 does not contribute to the folding and that the two Cys residues in the HS attachment domain either form a disulfide bond together or are free.

**Overall Three-dimensional Structure of GPC1 ΔC.** GPC1 ΔHS and GPC1 ΔC both crystallized in space group $P2_1$ with four molecules in the asymmetric unit. Crystals of GPC1 ΔHS diffracted X-rays to 2.95 Å and crystals of GPC1 ΔC diffracted to 2.5 Å. Aside

from the improved diffraction from GPC1 ΔC crystals, reproducibility of crystallization was greatly improved for GPC1 ΔC compared to GPC1 ΔHS. SeMet derivative crystals of GPC1 ΔC also crystallized in space group $P2_1$ with unit-cell parameters similar to crystals of native GPC1 ΔC and diffracted to 3.3 Å. The structure of GPC1 ΔC was solved by the MAD technique using the peak and inflection wavelengths for selenium. Data collection and processing statistics are presented in Table I. The initial auto-built model was refined against a native data set diffracting to 2.5 Å. The final $R_{model}$ and $R_{free}$ values are 25.3% and 29.2% respectively. These somewhat high values can be explained partly by the significant anisotropy in the diffraction data. Analysis of the dataset for GPC1 ΔC using the UCLA Anisotropy Server (23) revealed strong anisotropy, with a different of approximately 40 $Å^2$ between the B-factors for the $a^*/b^*$ directions and the $c^*$ direction. This effectively restricted diffraction data to 2.9 Å in the $c^*$ direction. We refined the structure against ellipsoidally truncated data from the UCLA server, but this resulted in no improvement in model or map quality, so we reverted to refinement against the full dataset.

The final model encompasses residues Ser27-Asp475 of GPC1 ΔC and the first GlcNAc residues on Asn-79 and Asn-116. A major gap in the model occurs between residues 350 and 361 and a smaller one between 406 and 412. No electron density was observed for the $His_6$ tag or the protease cleavage site. The asymmetric unit of the GPC1 ΔC consists of four monomers apparently organised as two dimers (Fig. 2A). The individual monomers in the dimer pairs are facing in opposite directions. Some variation in the quality of the electron density was seen for the different monomers. The most complete model was obtained with the monomer designated D; the other chains suffer from increased gap lengths. In particular, a major gap between residues 180-200 is seen in the other monomers. The missing amino acids in monomer D are charged residues present in extended loops. The highly extended GPC1 molecules pack in layers with extensive contacts in two dimensions but only limited contact in the third dimension (Supplementary Figure S2). This is consistent with the anisotropy observed in the diffraction datasets.

The structure of GPC1 is similar to the recently published structure of the *D. melanogaster* glypican DLP, with similarly

positioned α helices (Fig. 2B). The GPC1 structure thus reveals a cylindrical (dimensions 120 x 30 x 30 Å) all α-helical fold for glypican core proteins that is structurally conserved between *D. melanogaster* and humans, although GPC1 and DLP only share 25% sequence identity (as determined by FASTA (40)). The overlap between GPC1 and DLP structures is shown as α-carbon traces in Fig. 2B. The r.m.s deviation in α-carbon positions between monomer D of GPC1 ΔC and DLP is 2.6 Å for 331 positions when superimposed by the secondary structure matching algorithm in Coot (35). The α-helical fold is also consistent with the far-UV CD spectra in Fig. 1B, displaying one peak at 195 nm and two troughs at 209 nm and 222 nm, which is typical for α-helical proteins. The GPC1 structure consists of 14 α helices and three major loops, termed α1-α14 and L1-L3 (Fig. 2C). In comparison with the DLP structure, the GPC1 structure extends structural knowledge of glypicans, especially for the N-terminal loops (L1 and L3) and helix α1. The 60-residue, 83 Å long helix α2 (71–130) spans the whole protein, with N-linked glycans near either end (on Asn79 and Asn116). Out of 14 evolutionarily conserved Cys residues in GPC1, 12 are located in a Cys-rich region near the N-terminus. This region was termed the N-lobe (18) based on the observation that it contains the N-terminus of the α helix (α2 in GPC1) that spans the entire protein. We propose that this region should instead be termed the Cys-rich lobe (Fig. 2C), as the lobe is not entirely contained within the N-terminal part of the sequence. The Cys-rich lobe is followed by a middle segment containing five helices: α2, α6, α9, α12 and α14, forming the heart of the structure. This region is stabilized by two evolutionarily conserved hydrophobic centres that connect the α-helices, and also contains the C-terminus of GPC1 ΔC. This lobe was termed the M-lobe in (18). Based on the conserved hydrophobic centers in this lobe, we propose that this part of GPC1 should be called the central lobe. The last part of the GPC1 structure contains a long loop, L2, stretching from Cys-343 to Glu-376. This loop is processed by furin proteases in many of the glypican core proteins (41,42); however, only minor processing of GPC1 occurs in our hands. Due to the disorder of the loop, residues 350 to 361 are not modelled. This lobe was termed the C-lobe in

6

(18); an alternative name could be the protease site lobe.

**Analysis of Crystal Contacts.** The interface between pairs of GPC1 molecules in the crystal is located in the central lobe and involves α9 and α12 in both interacting monomers. Analysis of the crystal packing reveals at least two additional crystal contacts, one involving the end of α2 interacting with α14 in another monomer, the other involving the N-terminal parts of different monomers (Supplementary Fig. 2). The interface involving α2 and α14 is the largest, with a buried surface area of ~500 Å$^2$. However, calculations of the stabilities of the interfaces using PISA software (43) indicate that the interactions are too weak to form oligomeric states of GPC1 in solution. Preliminary dynamic light scattering experiments confirm these calculations, with profiles exhibiting single peaks for both GPC1 ΔHS and GPC1 ΔC (data not shown).

**Evolutionarily Conserved Cysteine Residues in Anchorless GPC1.** All of the 14 Cys residues evolutionarily conserved in the glypican family are clearly defined in the electron density, and all are clearly involved in disulfide bonds. This gives a high degree of confidence to the presented model. The disulfide bonds and their connections are presented in Table II and a detailed view of the model and electron densities for the disulfide bonds is shown in Fig. 3. Six disulfide bonds connect the structure in the Cys-rich lobe. Disulfides Cys268-Cys415 and Cys272-Cys401 connect the structure between the furin-like convertase site, consistent with the two fragments that appear when running reducing SDS-PAGE on glypicans processed by furin-like convertases (41,42). Together with Cys246-Cys279, this forms three disulfide bonds in or closely situated to α8 in the Cys-rich lobe (Fig. 3A). Cys32-Cys68, Cys62-Cys256 and Cys69-Cys259 form three longitudinally placed disulfide bonds in the N-terminal part of the Cys-rich lobe (Fig. 3B). All detailed pictures of the GPC1 structure are presented with the Cys-rich lobe facing upwards. The remaining disulfide bond is located in the protease lobe, being formed by Cys191-Cys343. The structure of DLP (18) lacked three out of seven disulfide bonds, since the electron density was not clear in the Cys-rich region. However the four disulfide bonds present in the DLP structure are consistent with those presented here.

GPC1 also contains two additional Cys residues in the HS attachment region, shared with GPC4 and GPC6. These Cys residues are not part of the presented GPC1 model and can either be free or form a disulfide bond. Crystallization of GPC1 ΔHS is facilitated by the addition of low concentrations of reducing agent, suggesting that the Cys residues are not bound together.

**Other Evolutionarily Conserved Residues in Anchorless GPC1.** A multiple sequence alignment of GPC1 through GPC6, Dally, and DLP is shown in Fig. S1. The alignment reveals that, aside from the 14 conserved Cys residues, other evolutionarily conserved amino acids are mostly of non-polar character. In particular, two hydrophobic centres are formed in the central lobe. Phe107 and Leu227 are identical in human and *D. melanogaster* glypicans (Fig. S1) and form a tight hydrophobic region with the conserved or conservatively substituted Phe111, Phe223 and Phe304 (Fig. 4A). The other conserved hydrophobic region in the central lobe is built from the conserved Trp290 and the conservatively substituted Phe392, Trp393 and Leu96 (Fig. 4B). A Trp to Arg mutation in GPC3 of the amino acid corresponding to Trp290 in Gpc-1 has been found in a Simpson-Golabi-Behmel syndrome patient, and it results in poor substitution of HS chains to the core protein (44). The mutation would result in a buried positive charge, which would either disrupt or prevent proper folding of GPC3.

**Structural conservation of the C-terminal segment in GPC1 orthologs.** Evolutionary conservation of surface exposed amino acids suggests that these exposed amino acids are involved in some interaction with other macromolecules relating to the function of the protein. In order to investigate in detail if surface exposed amino acids are evolutionarily conserved in glypicans, we performed multiple sequence alignment of GPC1 through GPC6, as well as the *D. melanogaster* glypicans and calculated amino acid conservation using the ConSurf Server (consurf.tau.ac.il). Fig. 5A shows the conservation of glypican amino acids as projected onto a space-filling model of GPC1. This model suggests that few surface-exposed amino acids are conserved between the evolutionarily distant glypicans. However, we also performed the same analysis using multiple sequence alignment of GPC1 orthologs, i.e. GPC1 from different vertebrates. Multiple sequence alignment was performed using GPC1

sequences from *Homo sapiens* (human), *Mus musculus* (mouse), *Rattus norvegicus* (rat), Bos taurus (bovine), *Gallus gallus* (chicken), *Danio rerio* (zebrafish)**,** *Xenopus tropicalis* (western clawed frog), *Monodelphis domestica* (short-tailed opossums), *Maccaca mulatta* (rhesus macaque) and *Pan troglodytes* (common chimpanzee). As shown in Fig. 5B, L1 and large parts of L3 are evolutionarily conserved within vertebrate GPC1 sequences, together with the C-terminal α14. Interestingly, L3 has been shown by mutagenesis to be involved in GAG class determination in rat GPC1 (9).

The L1 loop connects α1 and α2 (Fig. 6A). Rather than being extended, the loop is folded upon itself and is stabilized by internal hydrogen bonds. The loop is tethered to the rest of the structure via three disulfide bonds. Cys68 is connected to α1 through Cys32, and Cys62 and Cys69 are connected to α7 through Cys256 and Cys259, respectively. The L3 loop (Fig. 6B) connects α13 and α14 and is formed by amino acids Cys401 to Glu450. Missing in the structure are amino acids 406-412. The L3 loop is stabilised by forming a disulfide bond, Cys268-Cys415, where Cys268 is located in a turn between α7 and α8. The L3 loop is built onto α7, α8 and α9 and forms a number of hydrogen bonds with these α helices (Fig. 6B). Comparison of the temperature factors for L1 and L3 suggest that L3 is far less flexible than L1, consistent with the hydrogen bonds between L3 and α7, α8 and α9 (data not shown).

The structural conservation observed in GPC1 suggests that different structural elements important for specific functions of individual vertebrate glypicans can be identified using similar approaches for the other members in the glypican family. However, our attempts to do so have so far failed, perhaps in part because the sequence of GPC1 is known for many species, contrary to other members of the glypican family.

**HS attachment site in GPC1.** The structure of GPC1 was solved by MAD phasing on crystals of SeMet labelled GPC1 ΔC. In order to build a model of full-length GPC1 in which the HS attachment site was included, we performed model refinement using data from a native GPC1 ΔHS crystal. However, no additional electron density was observed, and $R_{free}$ and $R_{model}$ values were similar to values obtained using GPC1 ΔC (Table 1). Thus it was not possible to obtain information regarding the C-terminal HS

attachment site in GPC1 using the GPC1 ΔHS crystals. This also suggests that the C-terminal HS attachment domain in GPC1 is disordered.

## Discussion

In a previous study, we investigated the in-solution conformation and stability of GPC1 and showed that the GPC1 core protein is a stable α-helical protein (17). Thermal and chemical unfolding of GPC1 ΔHS showed that high temperatures (~70°C) or high concentrations of urea or guanidine-HCl (5.8 M or 2.6 M, respectively) are required to denature the protein. Chemical denaturation occurred in a single step for GPC1 ΔHS for both secondary and tertiary structure, and unfolding data closely fitted a two-state model. The stability of GPC1 ΔHS, $\Delta G_{NU}$, was estimated to be ~60 kJ/mol, which is in the higher range of what is usually seen for globular proteins (45). Together, these data suggested that GPC1 is a densely packed one-domain globular protein. These previous results support the crystal structure of GPC1, which is α-helical, consists of a single domain and, at least as far as the core helices are concerned, is apparently quite rigid. In particular, the disulfide bonds and the long helix α2 possibly contribute to the cooperative unfolding of GPC1.

During refinement of the GPC1 structure, the core protein structure of the *D. melanogaster* glypican DLP was published (18). This structure is similar to the GPC1 core protein structure, with similarly placed α helices (Fig. 2B). The two structures were created independently, giving strength for both models as the correct overall fold for glypican core proteins. It should be noted that the GPC1 structure was determined by using crystals of N-glycosylated GPC1 while DLP was crystallized after mutagenic disruption of both N-glycosylation sites in DLP. N-linked glycans are involved in protein folding, quality control, sorting, degradation and secretion (46). Given the similarities of the structures of N-glycosylated GPC1 and unglycosylated DLP, it is clearly possible to obtain the correct glypican fold in the absence of N-glycosylation. This is also supported by our previous study, investigating the importance the N-glycosylation for HS substitution, folding and stability of GPC1, where we show that the correct folding is obtained for GPC1 even when the two N-glycosylation sites have been disrupted (39).

Also, the N-linked glycans on glypicans are most likely involved in increasing the solubility of the proteins, as N-glycosylated GPC1 was crystallized at a protein concentration of 20 mg/ml in comparison with non-glycosylated DLP at 3.5 mg/ml.

The core proteins of proteoglycans have been regarded as inert, with all function of the PGs related to the GAG chains. This view arose from a wealth of genetic experiments implicating GAGs as regulators of morphogenesis and because GAGs are fairly easy to label and purify. The PG core proteins on the other hand are often large, heavily glycosylated and difficult to express in bacteria and have therefore been more difficult to study. Moreover, although GPC1 is a fairly large protein with longest dimension around 120 Å, this is rather small in comparison to the HS chains that become attached to glypicans. An HS chain can consist of up to 50-100 repeating disaccharide units of GlcNAc and GlcA. From the solution structure of HS, the length of a HS chain consisting of 8 disaccharide units has been estimated to be around 70 Å (47). Thus, a full HS chain can be approximately 500 Å long, as compared to the 100 Å for the folded glypican core protein. However, in recent years many interactions with PG core proteins have been found. For example, the ecto domains of syndecan core proteins have been shown to interact directly with integrins (48,49). Glypican core proteins have been suggested to have functional roles, by interacting with a number of growth factors including hedgehog, bone morphogenetic protein 4, decapentaplegic, Insulin-like growth factor II and Wnt. Recently, a region was identified in DLP as being important for hedgehog signalling, although hedgehog itself does not seem to bind and the interaction partner for this region remains elusive (18). This region is localized to α4 and the first part of α5 (in our numbering α5 and α6, since the first α helix is missing in the DLP structure). This region contains a number of amino acids conserved between GPC1 to GPC6, Dally and DLP that are involved in the folding of the protein. Among the amino acids in this region is the GPC1 Cys191-Cys343 disulfide bond. The structural similarity in this region is confirmed by our GPC1 ΔC structure, with two α-helices connected by a turn. The possibility arises that the interaction is governed by charged residues in this region, and that mutagenesis of GPC1 to mimic the DLP region important for hedgehog signalling may confer

hedgehog signalling effects on GPC1 core protein.

Amino acids in protein loops display higher sequence variation than residues that form the core of a structure, unless the loops are involved in the function of the protein. Also, protein loops can adopt a more unique folding than the common protein secondary structure elements α helix and β sheet. This is perhaps best exemplified by the hypervariable loops on antibodies, which are responsible for the unique binding of antibodies to their antigens. By our analysis of GPC1 vertebrate orthologs, the L1 and L3 loops of GPC1 are evolutionarily conserved, suggesting that they are involved in the function of GPC1. No interaction partner has been found for GPC1 core protein, contrary to some of the other glypican core proteins. *GPC1* knock-out mice are viable and fertile but have reduced brain size at birth (6). This has been suggested to be an effect of a transient reduction in HS dependent fibroblast growth factor signalling during embryogenesis. Moreover, heparan sulfate and chondroitin sulfate proteoglycans (HSPGs and CSPGs, respectively) serve different function during brain development, with CSPGs inhibiting and HSPGs promoting axon extension (50). GPC1 is exclusively substituted with HS, whereas GPC5, for example, can exist as an HS CS/DS (where DS is dermatan sulfate) hybrid. However, expression of the GPC1 GAG attachment domain alone results in addition of ~90% CS (9), strongly suggesting that factors in the GPC1 core protein further away from the GAG attachment sites promote the assembly of HS. Mutagenesis of amino acids in the L3 loop resulted in increased proportion of CS substituted on GPC1 but mutagenesis of the entire loop resulted in no protein expression (9). The structure of the L3 loop presented here should facilitate further mutagenic studies. It should also be noted that the information regarding HS assembly enzymes has increased during recent years. HS polymerization and initiation are performed by the exostosin (EXT) and exostosin-like (EXTL) families, respectively. EXT1 and EXT2 function as a HS co-polymerase and were first described as tumor suppressor genes, as they were found to be the cause of hereditary multiple exostoses, a disease characterized by the formation of multiple cartilaginous tumors (51,52). The EXTL family comprises three members, EXTL1-3, where EXTL2 and EXTL3 harbour HS initiating α1,4-

GlcNAc transferase activity and are ubiquitously expressed in the human body (53). Only EXTL3 has a *D. melanogaster* ortholog, suggesting that this is the functional HS initiation enzyme, whereas structural data is available for EXTL2, the smallest member in the EXTL family. Given the requirement for the glypican core protein to ensure exclusive HS attachment on GPC1, it is reasonable to suggest that EXTL2 or EXTL3 interact not only with the HS attachment site but also with the core protein. Mutagenesis of the L3 loop suggests that this is the interacting surface for EXTL and glypicans. It is also possible that disruption of the L3 loop results in displacement of the HS attachment site. An alternative explanation could also be that the glypican core protein provides steric hindrance for the CS initation enzyme, chondroitin N-acetylgalactosaminyltransferase (54), in order to ensure exclusive HS attachment. Unfortunately, no structural information is available for the CS initiating enzymes. The L1 loop is also evolutionarily conserved as shown by our analysis of structural conservation of GPC1 vertebrate orthologs, but thus far, no function has been suggested for this loop.

The glypican core proteins have long been assumed to adopt a globular fold, but the structures of GPC1 and DLP are cylindrical in shape. The C-terminal $\alpha 14$ is placed on the long side of the protein, making it difficult to predict how the core proteins are situated in the membranes. The first Ser residue in the HS attachment site is located ten amino acids from the C-terminus of the GPC1 $\Delta$C structure, placing the HS attachment close to the folded core protein. We have previously shown that the HS chains on GPC1 do not affect its secondary or tertiary structure (17). In fact, results obtained by differential scanning calorimetry suggest that the HS chains cause the core protein monomers to repel one another, possibly by electrostatic repulsion of the negatively charged HS chains. Others have reported that glypican core protein binds weakly to heparin agarose (18). The C-terminal Ser residue responsible for GPI-attachment is located 50 amino acids downstream of the visible ends of the models of GPC1 and DLP, which is long enough that the C-terminus of the protein could be placed in a wide variety of orientations with respect to the core protein. Thus when the GPI anchor is anchored in the membrane, it may allow great freedom of orientation of GPC1 and DLP with respect to the membrane surface. Our attempts to locate the GPC1 HS attachment domain and C-terminus using crystallography have failed, which together with the presented in-solution data on the GPC1 conformation, indicate that the HS attachment site is a disordered tail in the protein. Other experimental approaches than X-ray crystallography may be needed to finally answer the important question of the placement of C-terminal part in glypicans, how the HS chains are extended from the core protein, and how glypicans are situated in the membranes. Moreover, the presented GPC1 structure provides new information which will be important in determining the mechanism behind exclusive HS attachment on glypicans.

**References**

1.  Bülow, H. E., and Hobert, O. (2006) *Annu. Rev. Cell Dev. Biol.* **22**(1), 375-407
2.  Häcker, U., Nybakken, K., and Perrimon, N. (2005) *Nat. Rev. Mol. Cell Biol.* **6**(7), 530-541
3.  Pilia, G., Hughes-Benzie, R. M., MacKenzie, A., Baybayan, P., Chen, E. Y., Huber, R., Neri, G., Cao, A., Forabosco, A., and Schlessinger, D. (1996) *Nat. Genet.* **12**(3), 241-247
4.  Campos-Xavier, A. B., Martinet, D., Bateman, J., Belluoccio, D., Rowley, L., Tan, T. Y., Baxová, A., Gustavson, K.-H., Borochowitz, Z. U., Innes, A. M., Unger, S., Beckmann, J. S., Mittaz, L., Ballhausen, D., Superti-Furga, A., Savarirayan, R., and Bonafé, L. (2009) *Am. J. Hum. Genet.* **84**(6), 760-770
5.  Okamoto, K., Tokunaga, K., Doi, K., Fujita, T., Suzuki, H., Katoh, T., Watanabe, T., Nishida, N., Mabuchi, A., Takahashi, A., Kubo, M., Maeda, S., Nakamura, Y., and Noiri, E. (2011) *Nat. Genet.* **43**(5), 459-463
6.  Jen, Y.-H., Musacchio, M., and Lander, A. (2009) *Neural Dev.* **4**(1), 33
7.  Zhang, L., and Esko, J. D. (1994) *J. Biol. Chem.* **269**(30), 19295-19299
8.  Zhang, L., David, G., and Esko, J. D. (1995) *J. Biol. Chem.* **270**(45), 27127-27135
9.  Chen, R. L., and Lander, A. D. (2001) *J. Biol. Chem.* **276**(10), 7507-7517
10. Williams, E. H., Pappano, W. N., Saunders, A. M., Kim, M.-S., Leahy, D. J., and Beachy, P. A. (2010) *Proc. Natl. Acad. Sci. U. S. A.* **107**(13), 5869-5874
11. Yan, D., Wu, Y., Yang, Y., Belenkaya, T. Y., Tang, X., and Lin, X. (2010) *Development* **137**(12), 2033-2044
12. Capurro, M. I., Xu, P., Shi, W., Li, F., Jia, A., and Filmus, J. (2008) *Dev. Cell* **14**(5), 700-711
13. Kirkpatrick, C. A., Knox, S. M., Staatz, W. D., Fox, B., Lercher, D. M., and Selleck, S. B. (2006) *Dev. Biol.* **300**(2), 570-582
14. Capurro, M. I., Xiang, Y.-Y., Lobe, C., and Filmus, J. (2005) *Cancer Res.* **65**(14), 6245-6254
15. Ohkawara, B., Yamamoto, T. S., Tada, M., and Ueno, N. (2003) *Development* **130**(10), 2129-2138
16. Cheng, W., Tseng, C.-J., Lin, T. T. C., Cheng, I., Pan, H.-W., Hsu, H.-C., and Lee, Y.-M. (2008) *Carcinogenesis* **29**(7), 1319-1326
17. Svensson, G., Linse, S., and Mani, K. (2009) *Biochemistry* **48**(42), 9994-10004
18. Kim, M.-S., Saunders, A. M., Hamaoka, B. Y., Beachy, P. A., and Leahy, D. J. (2011) *Proceedings of the National Academy of Sciences* **108**(32), 13112-13117
19. Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E., and Hendrickson, W. A. (1994) *Structure (London, England : 1993)* **2**(6), 545-558
20. Kabsch, W. (2010) *Acta Crystallographica Section D* **66**(2), 133-144
21. Collaborative Computational Project, N. (1994) *Acta Crystallographica Section D* **50**(5), 760-763
22. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallographica Section D* **66**(2), 213-221
23. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006) *Proceedings of the National Academy of Sciences* **103**(21), 8060-8065

24. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2005) *Acta Crystallographica Section D* **61**(4), 449-457

25. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2009) *Acta Crystallographica Section D* **65**(10), 1089-1097

26. Howell, P. L., and Smith, G. D. (1992) *Journal of Applied Crystallography* **25**(1), 81-86

27. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallographica Section D* **58**(10-2), 1772-1779

28. de La Fortelle, E., Bricogne, G., and Charles W. Carter, Jr. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In. *Methods Enzymol.*, Academic Press

29. Cowtan, K. (1999) *Acta Crystallographica Section D* **55**(9), 1555-1567

30. Terwilliger, T. C. (2000) *Acta Crystallographica Section D* **56**(8), 965-972

31. Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A., and Lamzin, V. S. (2004) *Journal of Synchrotron Radiation* **11**(1), 56-59

32. Cowtan, K. (2006) *Acta Crystallographica Section D* **62**(9), 1002-1011

33. Brunger, A. T. (2007) *Nat. Protocols* **2**(11), 2728-2733

34. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallographica Section D* **53**(3), 240-255

35. Emsley, P., and Cowtan, K. (2004) *Acta Crystallographica Section D* **60**(12-1), 2126-2132

36. Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) *Acta Crystallographica Section D* **66**(1), 12-21

37. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**(22), 4673-4680

38. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) *Nucleic Acids Res.* **38**(suppl 2), W529-W533

39. Svensson, G., Hyrenius Wittsten, A., Linse, S., and Mani, K. (2011) *Biochemistry* **50**(43), 9377-9387

40. Pearson, W. R., and Lipman, D. J. (1988) *Proceedings of the National Academy of Sciences* **85**(8), 2444-2448

41. De Cat, B., Muyldermans, S.-Y., Coomans, C., Degeest, G., Vanderschueren, B., Creemers, J., Biemar, F., Peers, B., and David, G. (2003) *The Journal of Cell Biology* **163**(3), 625-635

42. Capurro, M. I., Shi, W., Sandal, S., and Filmus, J. (2005) *J. Biol. Chem.* **280**(50), 41201-41206

43. Krissinel, E., and Henrick, K. (2007) *J. Mol. Biol.* **372**(3), 774-797

44. Veugelers, M., Cat, B. D., Muyldermans, S. Y., Reekmans, G., Delande, N., Frints, S., Legius, E., Fryns, J.-P., Schrander-Stumpel, C., Weidle, B., Magdalena, N., and David, G. (2000) *Hum. Mol. Genet.* **9**(9), 1321-1328

45. Makhatadze, G. I., and Privalov, P. L. (1995) Energetics of Protein Structure. In. *Adv. Protein Chem.*, Academic Press

46. Helenius, A., and Aebi, M. (2004) *Annu. Rev. Biochem.* **73**(1), 1019-1049

47. Khan, S., Rodriguez, E., Patel, R., Gor, J., Mulloy, B., and Perkins, S. J. (2011) *J. Biol. Chem.* **286**(28), 24842-24854

48. Whiteford, J. R., and Couchman, J. R. (2006) *J. Biol. Chem.* **281**(43), 32156-32163

49.  Beauvais, D. M., Burbach, B. J., and Rapraeger, A. C. (2004) *Journal Cell Biol* **167**(1), 171-181

50.  Matsumoto, Y., Irie, F., Inatani, M., Tessier-Lavigne, M., and Yamaguchi, Y. (2007) *The Journal of Neuroscience* **27**(16), 4342-4350

51.  Stickens, D., Clines, G., Burbee, D., Ramos, P., Thomas, S., Hogue, D., Hecht, J. T., Lovett, M., and Evans, G. A. (1996) *Nat. Genet.* **14**(1), 25-32

52.  Lind, T., Tufaro, F., McCormick, C., Lindahl, U., and Lidholt, K. (1998) *J. Biol. Chem.* **273**(41), 26265-26268

53.  Kitagawa, H., Shimakawa, H., and Sugahara, K. (1999) *J. Biol. Chem.* **274**(20), 13933-13937

54.  Uyama, T., Kitagawa, H., Tamura, J.-i., and Sugahara, K. (2002) *J. Biol. Chem.* **277**(11), 8841-8846

55.  Read, R. J., and Schierbeek, A. J. (1988) *Journal of Applied Crystallography* **21**(5), 490-495

**Tables**

**Table I. Data collection and refinement statistics.**

|  | Native GPC1 ΔC | SeMet GPC1 ΔC (peak) | SeMet GPC1 ΔC (inflection) | Native GPC1 ΔHS |
|---|---|---|---|---|
| Space group | $P2_1$ | $P2_1$ | $P2_1$ | $P2_1$ |
| Unit-cell parameters (Å, °) | $a = 47.2$, $b = 168.7$, $c = 147.8$, β = 94.6 | $a = 47.1$, $b = 168.5$, $c = 147.0$, β = 94.6 | $a = 47.2$, $b = 168.6$, $c = 147.4$, β = 94.6 | $a = 47.2$, $b = 169.0$ $c = 151.6$ β = 95.0 |
| Beamline | I911-2 | I911-3 | I911-3 | I911-2 |
| Wavelength (Å) | 1.0397 | 0.9789 | 0.9794 | 1.0397 |
| Resolution range (Å) | 30-2.49 (2.64-2.49) | 30-3.3 (3.39-3.30) | 30-3.3 (3.39-3.30) | 30-2.95 (3.03-2.95) |
| R$_{merge}$ (I) (%) | 8.3 (68.3) | 10.7 (33.0) | 12.9 (51.1) | 13.5 (83.2) |
| Total observations | 328111 | 347933 | 381017 | 440496 |
| Unique reflections | 79825 | 66568 | 66846 | 49537 |
| Average redundancy | 4.1 (3.5) | 5.2 (5.3) | 5.7 (5.6) | 8.9 (8.2) |
| Completeness (%) | 98.9 (94.9) | 98.2 (98.0) | 98.0 (95.4) | 99.4 (94.1) |
| $I/\sigma(I)$ | 11.6 (2.37) | 11.3 (4.47) | 10.85 (3.29) | 15.4 (3.4) |
| Refinement |  |  |  |  |
| R$_{work}$/R$_{free}$ (%) | 25.3/29.2 |  |  | 22.1/27.9 |
| Number of atoms |  |  |  |  |
| Protein/carbohydrate | 12572 |  |  | 12630 |
| Water | 242 |  |  | 126 |
| r.m.s deviations from ideal geometry |  |  |  |  |
| Bond lengths, Å | 0.003 |  |  | 0.004 |
| Bond angles, ° | 0.72 |  |  | 0.84 |
| Ramachandran analysis |  |  |  |  |
| In preferred regions (%) | 98.6 |  |  | 98.2 |
| In allowed regions (%) | 1.4 |  |  | 1.8 |
| Outliers (%) | 0 |  |  | 0 |

 The R$_{free}$ values were calculated using 5% of the unique data (3753 reflections for the truncated form and 2529 for the full-length protein).

**Table II. Disulfide bonds in GPC1.**

| Cysteine residues | Connected structural elements |
|---|---|
| Cys32-Cys68 | α1-L1 |
| Cys62-Cys256 | L1-α7 |
| Cys69-Cys259 | L1-α7 |
| Cys191-Cys343 | α5-α11 |
| Cys246-Cys279 | α7-α8 |
| Cys268-Cys415 | Turn between α7 and α8-L3 |
| Cys272-Cys401 | α8-α13 |

**Figure legends**

**Figure 1. SDS-PAGE and spectroscopic analysis of rhGPC1.** (A) SDS-PAGE and Coomassie blue staining of Ni-NTA purified full-length rhGPC1 expressed without the GPI-anchor signal peptide (lane 1). Expression of full-length rhGPC1 with disrupted HS attachment site (GPC1 ΔHS) resulted in

expression of core protein only (lane 2). GPC1 with the C-terminal HS attachment site removed by truncation resulted in expression of core protein of smaller size (lane 3). Three µg of purified protein were added to the well of lane 1 and one µg was loaded in the wells of lane 2 and 3. (B and C) Far and near UV-CD spectra, respectively, of GPC1 ΔHS (black, solid line) and GPC1 ΔC (grey, solid line) in 20 mM sodium phosphate (pH 7.4) were recorded at 20°C.

**Figure 2. Overview of GPC1 ΔC structure and comparison with DLP.** (A) Overview of the asymmetric unit of GPC1 ΔC with two dimer pairs. (B) Ribbon representation of α-carbon alignment of GPC1 ΔC monomer (blue) and DLP (green, PDB ID 3ODN). Alignment was performed in PyMOL, version 1.2r3pre, Schrödinger, LLC. (C) Overview of the GPC1 ΔC structure represented as cartoon in rainbow colours (N-terminus in blue, C-terminus in red) with loops 1-3 (L1-L3) and α-helices (α1-α14). Also indicated are the disulfide bonds (yellow solid lines), the first residues of the N-glycans (GlcNAc) and the assignment of different lobes in the GPC1 ΔC structure.

**Figure 3. Structure model and electron densities of the six disulfide bonds in the N-terminal Cys-rich lobe of GPC1.** *Red*, oxygen; *blue*, nitrogen; *green*, carbon; *yellow*, sulphur. Model of the three disulfide bonds (Cys415-Cys268; Cys272-Cys401; Cys246-Cys279) in α8 in the Cys-rich region (A) and of the three longitudinally placed disulfide bonds in the L1 loop (Cys32-Cys68; Cys69-Cys259 and Cys62-Cys256) (B). Electron densities, contoured at 2.0 σ in A and 1.0 σ in B, were calculated as 2|Fo|-|Fc| maps using the program FFT in CCP4 (21,55). Figures were generated using PyMOL.

**Figure 4. Structure model and electron densities of two hydrophobic centres conserved in glypicans.** *Red*, oxygen; *blue*, nitrogen; *green*, carbon. Phe107 and Leu227 are identical in human and *D. melanogaster* glypicans and form a tight hydrophobic core in the central lobe of GPC1 with the conservatively substituted Phe111, Phe223, Phe304 (A). Another hydrophobic region is also present in the central lobe, built from the conserved Trp290 and the conservatively substituted Phe392, Trp393 and Leu96 (B). Presented in mesh are 2|Fo|-|Fc| electron density maps generated as in Fig. 3. The maps are contoured at 1.0 σ. Figures were generated using PyMOL.

**Figure 5. Structural conservation of surface-exposed amino acids in GPC1 homologs (A) and in GPC1 orthologs (B).** Multiple sequence aligments were performed using ClustalW (37) and conservation scores were calculated with the Bayesian method using the ConSurf server (38). Images of structural conservation of surface exposed amino acids were generated using Consurf output in combination with PyMOL. Amino acid conservation is shown in colors from turquoise (variable) to purple (conserved). The inlaid smaller figures show the structural conservation space fill model in transparency aligned with a ribbon representation of the structure. (A) Multiple sequence alignment and amino acid conservation were calculated using GPC1 homologs. The following amino acid sequences were used (UniProt accession numbers): GPC1 human, P35052; GPC2 human, Q8N158; GPC3 human, P51654; GPC4 human O75487; GPC5 human, P78333; GPC6 human, Q9Y625; Dally *D. melanogaster*, Q24114; DLP *D. melanogaster*, Q9GPL5. (B) The following amino acid sequences were used to visualize structural conservation in GPC1 orthologs (i.e. GPC1 vertebrate homologs): GPC1 human, P35052; GPC1 mouse, Q9QZF2; GPC1 rat, P35052; GPC1 bovine, Q2KJ65; GPC1 chick, P50593; Gpc1 zebrafish, Q1LXM6; gpc1 western clawed frog, Q0V9W0. Also included were the translated mRNA sequences (given by NBCI nucleotide accession numbers) for *GPC1* genes: *Gpc1* short-tailed opossums, XM_001377888; *GPC1* rhesus macaque, XM_001089255; *Gpc1* common chimpanzee, XM_001156197.

**Figure 6. Structure of evolutionarily conserved L1 and L3.** (A) L1 (blue) with start and end marked by arrows is folded upon itself with hydrogen bonds (dotted yellow lines) and is tethered to the rest of the structure by several disulfide bonds (yellow solid lines). (B) L3 (orange and red) with start and end marked by arrows is connected to α7, α8 and α9 with several hydrogen bonds (dotted yellow lines).

# Figure 1

# Figure 2

# Figure 3
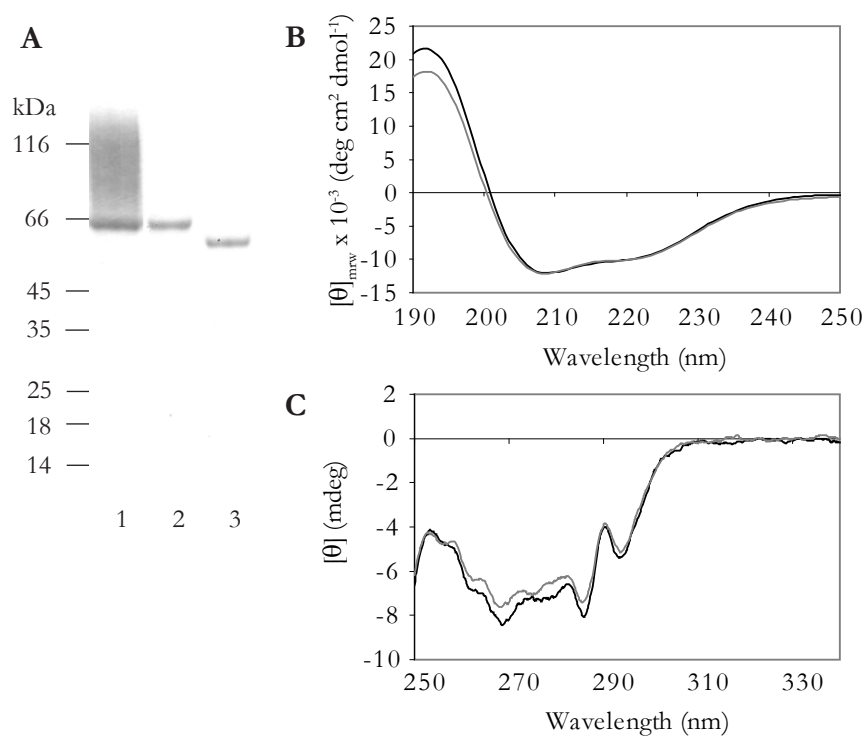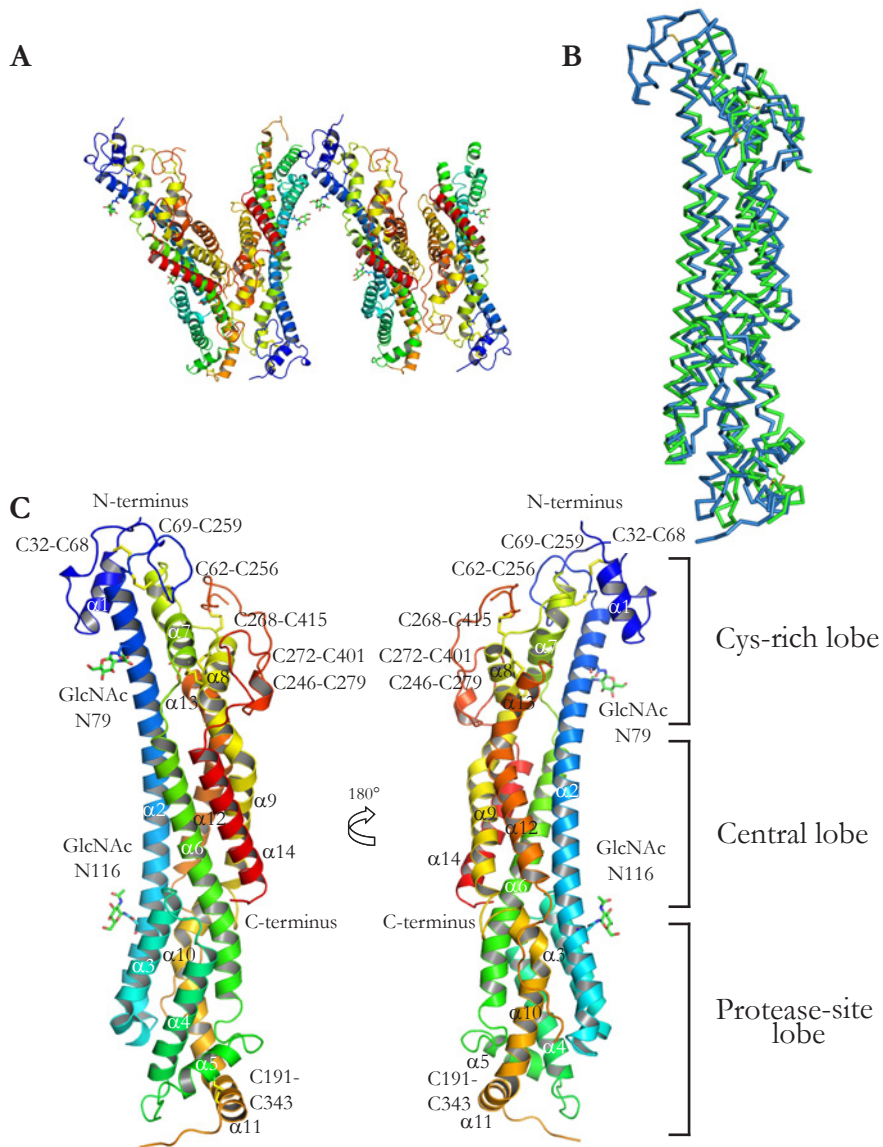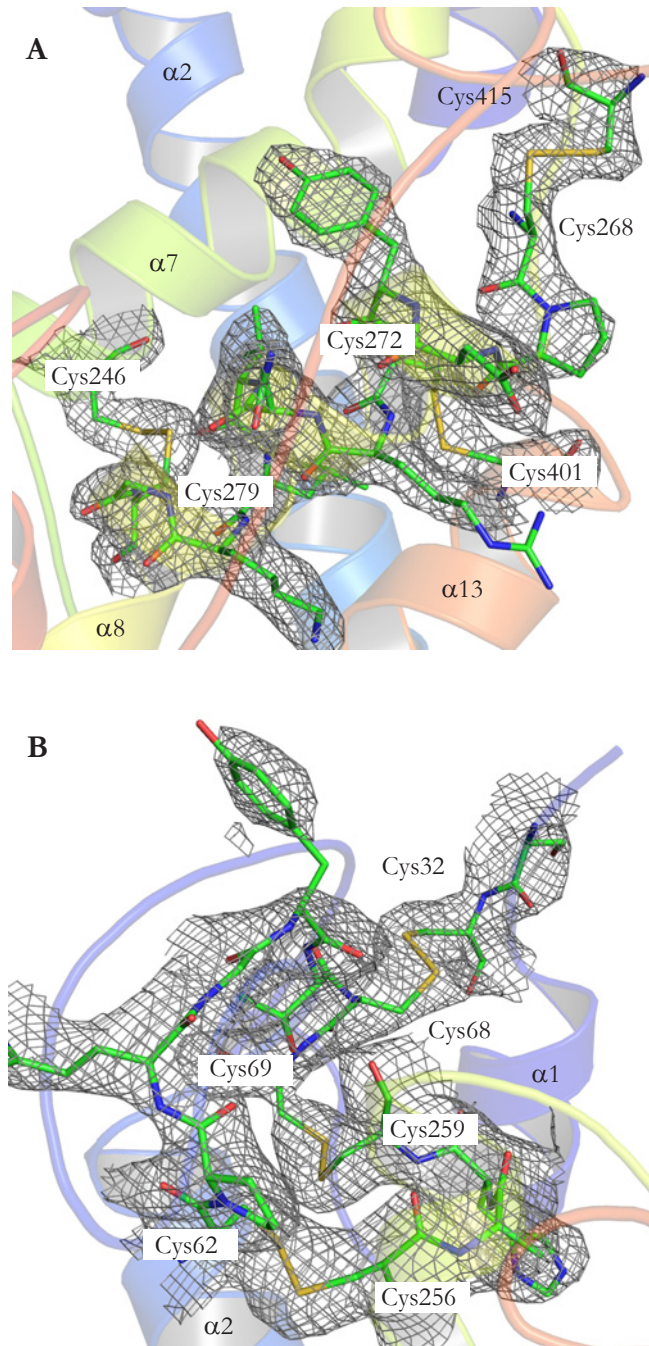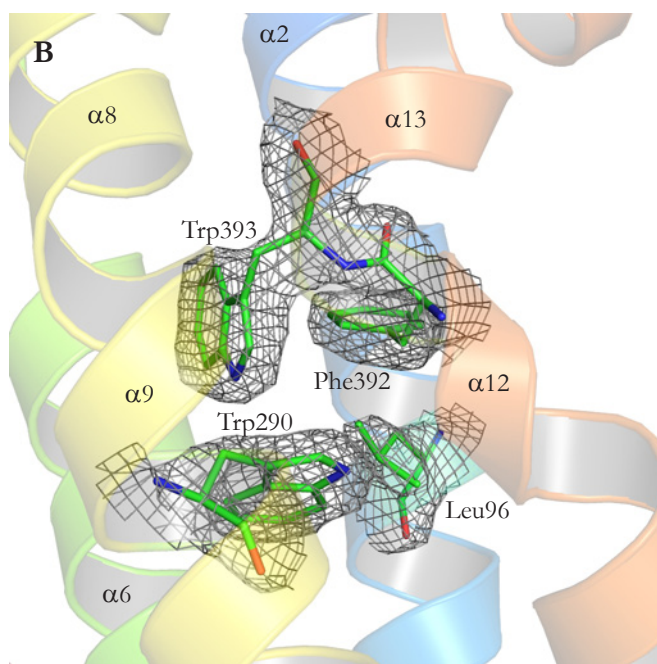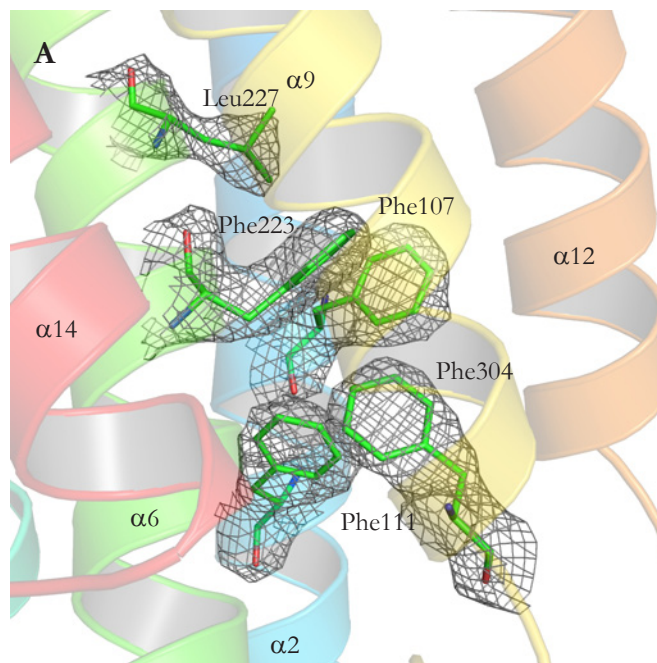
# Figure 4

# Figure 5
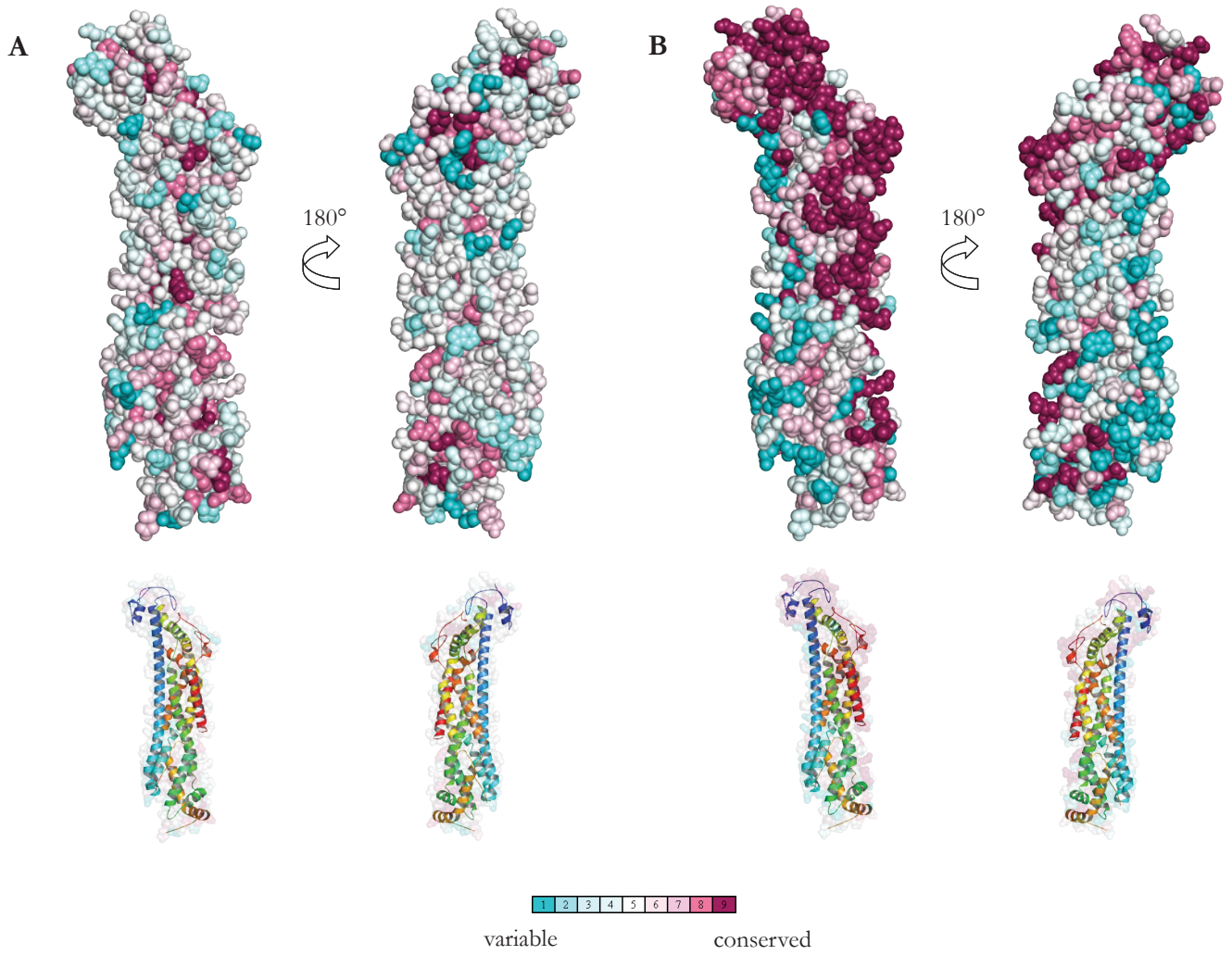
# Figure 6
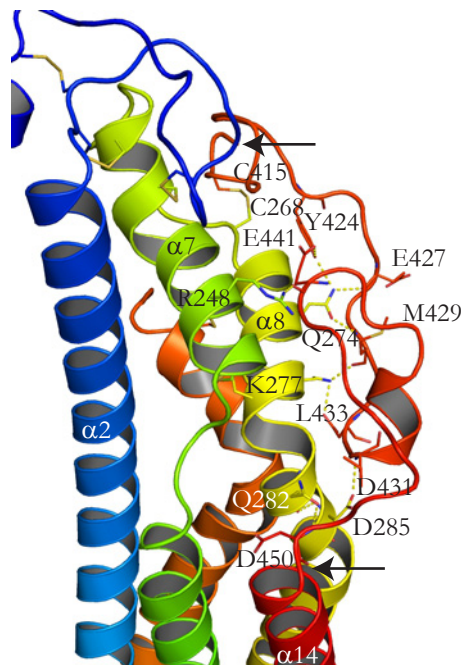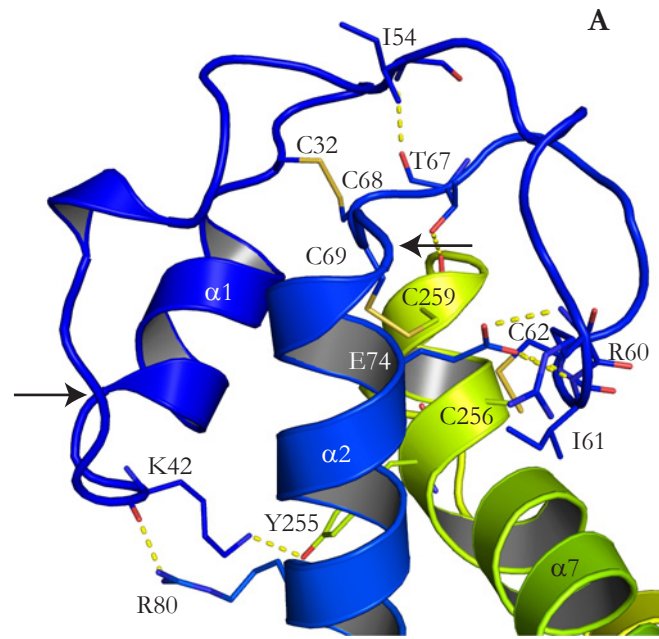
# Supplementary figure 1

```
GPC1_HUMAN    ------------------------MELRARGWWLLCAAAALVACARGDPASK-------  28
GPC2_HUMAN    ------------------------MSALRPLLLLLLPLCPGPGPGPGSEAKV-------  28
GPC3_HUMAN    ---------------------MAGTVRTACLVVAMLLSLDFPGQAQPPPPPP------  31
GPC4_HUMAN    ------------------------MARFGLPALLCTLAVLSAALLAAELK--------  26
GPC5_HUMAN    ----------------------MDAQTWPVGFRCLLLLALVGSARSEG----------  26
GPC6_HUMAN    ------------------------MPSWIGAVILPLLGLLLSLPAGADVK--------  26
DALY_DROME    --------------------MAARSVRLAQLLLFTLLCGFVGLSAAKHLDLDGIHHHQH  39
DLP_DROME     MLHQEQQQQHLHCRRKVTATTTARLVIFSSPLLLLLLTTHLPSTLQADNSPAPQVAALAA  60
                                                                  .
```

**α1**　　　　　　**L1**

```
GPC1_HUMAN    ----------------SRSCGEVR-QIYGAKGFSLSD-VPQAEISGEHLRICPQG--YTC  68
GPC2_HUMAN    ---------------TRSCAETR-QVLGARGYSLNL-IPPALISGEHLRVCPQE--YTC  68
GPC3_HUMAN    ---------------DATCHQVR-SFFQR-LQPGLKWVPETPVPGSDLQVCLPKG-PTC  72
GPC4_HUMAN    ---------------SKSCSEVR-RLYVSKGFNKND-APLHEINGDHLKICPQG--STC  66
GPC5_HUMAN    ---------------VQTCEEVR-KLFQWRLLGAVRGLPDSPRAGPDLQVCISKK-PTC  68
GPC6_HUMAN    ---------------ARSCGEVR-QAYGAKGFSLAD-IPYQEIAGEHLRICPQE--YTC  66
DALY_DROME    HLHSATTHHRRRLQRDSRAKDAVGGSTHQCDAVKSYFESIDIKSSGTYSEKGAICG-GNC  98
DLP_DROME     PNPAGGVAGSSIIDQFSPNCSAVT-HIFQARGIDAIE-IPQKPSNERVLRYCESPSVGTC  118
                            .                      .            .*
```

**α2**

```
GPC1_HUMAN    CTSEMEENLANRSHAELETALRDSSRVLQAMLATQLRSFDDHFQHLLNDSERTLQATFPG  128
GPC2_HUMAN    CSSETEQRLIRETEATFRGLVEDSGSFLVHTLAARHRKFDEFFLEMLSVAQHSLTQLFSH  128
GPC3_HUMAN    CSRKMEEKYQLTARLNMEQLLQSASMELKFLIIQNAAVFQEAFEIVVRHAKNYTNAMFKN  132
GPC4_HUMAN    CSQEMEEKYSLQSKDDFKSVVSEQCNHLQAVFASRYKKFDEFFKELLENAEKSLNDMFVK  126
GPC5_HUMAN    CTRKMEERYQIAARQDMQQFLQTSSSTLKFLISRNAAAFQETLETLIKQAENYTSILFCS  128
GPC6_HUMAN    CTTEMEDKLSQQSKLEFENLVEETSHFVRTTFVSRHKKFDEFFRELLENAEKSLNDMFVR  126
DALY_DROME    CNNATELELRDKAAGMFEQLLHHHTSSLRGVLETNAKQFQSHVLELAQISENMTHSLFSK  158
DLP_DROME     CTYNMETRMAMQSRQQLEGHTKDQISRMSGILGSKATKFKDIFTALLKESRTQFNSMFIR  178
              *.    *.      :    :.        :    :    *.. .   :     :.      *
```

**α3**　　　　　　　　　　　　　　　　　**α4**

```
GPC1_HUMAN    AFGELYTQNARAFRDLYSELRLYYRGAN------------------LHLEETLAEFWARL  170
GPC2_HUMAN    SYGRLYAQHALIFNGLFSRLRDFYGESG------------------EGLDDTLADFWAQL  170
GPC3_HUMAN    NYPSLTPQAFEFVGEFFTDVSLYILGSD------------------INVDDMVNELFDSL  174
GPC4_HUMAN    TYGHLYMQNSELFKDLFVELKRYYVVGN------------------VNLEEMLNDFWARL  168
GPC5_HUMAN    TYRNMALEAAASVQEFFTDVGLYLFGAD------------------VNPEEFVNRFFDSL  170
GPC6_HUMAN    TYGMLYMQNSEVFQDLFTELKRYYTGGN------------------VNLEEMLNDFWARL  168
DALY_DROME    VYTRMVPSSRMMIHQLYTEIMNHLIYTSNYTNSNGQLGRRGIGSVQSNLEEAVRHFFVQL  218
DLP_DROME     TYGVIYERNSYVFSDLFKELETYFANGR------------------VDLLEVMDKFFNTL  220
               :   :       .  :: :   .                     :   :   ::   *
```

**α5**　　　　　**α6**

```
GPC1_HUMAN    LERLFKQLH-----PQLLLPDDYLDCLGKQAEA----LRPFGEAPRELRLRATRAFVAAR  221
GPC2_HUMAN    LERVFPLLH-----PQYSFPPDYLLCLSRLASSTDGSLQPFGDSPRRLRLQITRTLVAAR  225
GPC3_HUMAN    FPVIYTQLMNP---GLPDSALDINECLRGARRD----LKVFGNFPKLIMTQVSKSLQVTR  227
GPC4_HUMAN    LERMFRLVN-----SQYHFTDEYLECVSKYTEQ----LKPFGDVPRKLKLQVTRAFVAAR  219
GPC5_HUMAN    FPLVYNHLINP---GVTDSSLEYSECIRMARRD----VSPFGNIPQRVMGQMGRSLLPSR  223
GPC6_HUMAN    LERMFQLIN-----PQYHFSEDYLECVSKYTDQ----LKPFGDVPRKLKIQVTRAFIAAR  219
DALY_DROME    FPVAYHQMVHLSKNNLGDLHEDYVNCLQHNFDE----MHPFGDIPQQVQSNLGKSVHMSN  274
DLP_DROME     YQKMFTVLN-----TQYTFDENYMRCVSEHMKE----LKPFGDVPDKLSVQIKRSFVATR  271
               :  :            :   *:       :  **: *   :  .  ::.  :.
```

```
                        ┌────────────────┐      ┌──────┐              ┌──────┐
                        │                │      │  α7  │              │  α8  │
                        └────────────────┘      └──────┘              └──────┘
GPC1_HUMAN   SFVQGLGVASDVVRKVAQVP---LGPECSRAVMKLVYCAHCLGVPG-----ARPCPDYCR 273
GPC2_HUMAN   AFVQGLETGRNVVSEALKVP---VSEGCSQALMRLIGCPLCRGVPS-----LMPCQGFCL 277
GPC3_HUMAN   IFLQALNLGIEVINTTDHLK---FSKDCGRMLTRMWYCSYCQGLMM-----VKPCGGYCN 279
GPC4_HUMAN   TFAQGLAVAGDVVSKVSVVN---PTAQCTHALLKMIYCSHCRGLVT-----VKPCYNYCS 271
GPC5_HUMAN   TFLQALNLGIEVINTTDYLH---FSKECSRALLKMQYCPHCQGLAL-----TKPCMGYCL 275
GPC6_HUMAN   TFVQGLTVGREVANRVSKVS---PTPGCIRALMKMLYCPYCRGLPT-----VRPCNNYCL 271
DALY_DROME   VFMNALLQAAEVLSEADALYGEQLTDTCKLHLLKMHYCPNCNGHHSSSRSETKLCYGYCK 334
DLP_DROME    TYGQALTTASEVAKKVLNVR---LNADCTGALTKMQHCGACKGYT------EKPCTNYCV 322
             :  :.*   . :*   . :        *    : ::   *  *        * .:*

             ┌──────┐    ┌──────────────────┐            ┌──────────┐
             │      │    │        α9        │            │   α10    │
             └──────┘    └──────────────────┘            └──────────┘
GPC1_HUMAN   NVLKGC-LANQADLDAEWRNLLDSMV-LITDKFWGTSGVESVIGSVHTWLAEAINALQDN 331
GPC2_HUMAN   NVVRGC-LSSRG-LEPDWGNYLDGLL-ILADKLQGPFSFELTAESIGVKISEGLMYLQEN 334
GPC3_HUMAN   VVMQGC-MAGVVEIDKYWREYILSLE-ELVNGMYRIYDMENVLLGLFSTIHDSIQYVQKN 337
GPC4_HUMAN   NIMRGC-LANQGDLDFEWNNFIDAML-MVAERLEGPFNIESVMDPIDVKISDAIMNMQDN 329
GPC5_HUMAN   NVMRGC-LAHMAELNPHWHAYIRSLE-ELSDAMHGTYDIGHVLLNFHLLVNDAVLQAHLN 333
GPC6_HUMAN   NVMKGC-LANQADLDTEWNLFIDAML-LVAERLEGPFNIESVMDPIDVKISEAIMNMQEN 329
DALY_DROME   NVMRGCSAEYAGLLDSPWSGVVDSLNNLVTTHILSDTGIINVIKHLQTYFSEAIMAAMHN 394
DLP_DROME    NVIKGC-LHYQHEFDSEWENFAMAMD-KVAERLLGSFNIVMVVEPLNIKISEAIMNFQDS 380
             :::**     ::  *        .:  :    :     ..   .    . :.:      .

             ┌──────┐
             │ α11  │
             └──────┘
GPC1_HUMAN   RDTLTAKVIQGCGNPK----VNPQGPGP-------------------------------- 355
GPC2_HUMAN   SAKVSAQVFQECGPPDPVPARNRRAPPPR------------------------------- 363
GPC3_HUMAN   AGKLTTTIGKLCAHSQQRQYRSAYYPEDL------------------------------- 366
GPC4_HUMAN   SVQVSQKVFQGCGPPKPLPAGRISRSISE------------------------------- 358
GPC5_HUMAN   GQKLLEQVNRICGRPVRTPTQSPRCSFDQ------------------------------- 362
GPC6_HUMAN   SMQVSAKVFQGCGQPKPAPALRSARSAPE------------------------------- 358
DALY_DROME   GPELEKKVKKTCGTPSLTPYSSGEPDARP------------------------------- 423
DLP_DROME    GQDITNRVFQGCGRPKLKKMKRSISPKLQGVQILNARSPVEADTLDIDETLDEAIVLRER 440
             :     :  : *.  .


GPC1_HUMAN   --------------------------------------------------EEKRRRGKLAP 366
GPC2_HUMAN   --------------------------------------------------EEAGRLWSMVTE 375
GPC3_HUMAN   ---------------------------------------------------FIDKKV 372
GPC4_HUMAN   --------------------------------------------------SAFSARFRPHHP 370
GPC5_HUMAN   ---------------------------------------------------SKEKHG 368
GPC6_HUMAN   -------------------------------------------------N-FNTRFRPYNP 369
DALY_DROME   ----------------------------------------------------------P 424
DLP_DROME    RAAEPGSQETSSQQSQEQGVGKSGNGGGGGGGGNNRRQQQRRKQQQQRRKQQNNRDDNDDD 500

             ┌──────────┐      ┌────────┐   ┌────────────────┐
             │   α12    │      │  α13   │   │       L3       │
             └──────────┘      └────────┘   └────────────────┘
GPC1_HUMAN   RERPPSGT---LEKLVSEAKAQLRDVQDFWISLPGTLCS-EKMALSTASDDR-CWNGMAR 421
GPC2_HUMAN   EERPTTAAGTNLHRLVWELRERLARMRGFWARLSLTVCGDSRMAADASLEAAPCWTGAGR 435
GPC3_HUMAN   LKVAHVEHEETLSSRRRELIQKLKSFISFYSALPGYICSHSPVAE----NDTLCWNGQEL 428
GPC4_HUMAN   EERPTTAAGTSLDRLVTDVKEKLKQAKKFWSSLPSNVCNDERMAAGNGNEDD-CWNGKGK 429
GPC5_HUMAN   MKTTTRNSEETLANRRKEFINSLRLYRSFYGGLADQLCANELAAA----DGLPCWNGEDI 424
GPC6_HUMAN   EERPTTAAGTSLDRLVTDIKEKLKLSKKVWSALPYTICKDESVTAGTSNEEE-CWNGHSK 428
DALY_DROME   PHKNNVKWATDPDPGMVLFLSTIDKSKEFYTTIVDNFCDEQQHSR----DDHSCWSGDRF 480
DLP_DROME    DNESGGGREPILDRIVRDIRQRVKDYKKFWSNLPHSVCSNEDIAS-SSDVDGMCWNGHTI 559
             .           :     .: :   .*  . :        **.*
```

```
                                              L3                          α14
GPC1_HUMAN    GR-YLPEVMGDGLANQINNPEVEVDITKPDMTIRQQIMQLKIMTNRLRSAYNGNDVDFQD 480
GPC2_HUMAN    GR-YLPPVVGGSPAEQVNNPELKVDASGPDVPTRRRRLQLRAATARMKTAALGHDLDGQD 494
GPC3_HUMAN    VERYSQKAARNGMKNQFNLHELKMKGPEPVVSQIIDKLKHINQLLRTMSMPKGRVLDKNL 488
GPC4_HUMAN    SR-YLFAVTGNGLANQGNNPEVQVDTSKPDILILRQIMALRVMTSKMKNAYNGNDVDFFD 488
GPC5_HUMAN    VKSYTQRVVGNGIKAQSGNPEVKVKGIDPVINQIIDKLKHVVQLLQGRSPKPDKWELLQL 484
GPC6_HUMAN    AR-YLPEIMNDGLTNQINNPEVDVDITRPDTFIRQQIMALRVMTNKLKNAYNGNDVNFQD 487
DALY_DROME    GD-YTQLLINPGTDSQRYNPEVPFNAKAQTGKLNELVDKLFKIRKSIGAAAPSNSIQATH 539
DLP_DROME     DR-YMHSITTEHGSNPEFTG--NPASTKQTAQMASQLSHLKNAIVHLRNAYNGQDVEWSE 616
                 *                                              ..

GPC1_HUMAN    ASDD---GSGSGSGDGCLDDLCSR-------------KVSRKSSSSRTPLTHALPGLSE 523
GPC2_HUMAN    ADED---ASGSGGGQQYADDWMAG-------------AVAPPARPPRPPYPPRRDGSGG 537
GPC3_HUMAN    DEEG----FESGDCGDDEDECIGGSGDGM----IKVKNQLRFLAELAYDLDVDDAPGNSQ 540
GPC4_HUMAN    ISDE---SSGEGSGSGCEYQQCPS-------------EFDYNATDHAGKSANE------ 525
GPC5_HUMAN    GSGGGMVEQVSGDCDDEDG--CGGSGSG------EVKRTLKITDWMPDDMNFSDVKQIHQ 536
GPC6_HUMAN    TSDE---SSGSGSGSGCMDDVCPT-------------EFEFVTTEAPAVDPDR------ 524
DALY_DROME    DIQN-DMGEGSGGGEGQIGDDEEEYGG---------AHGSGDGSGDGPHTPIEESEGTTT 589
DLP_DROME     QEELPYAGSGAGSGSGSEDDEDDDEGSGLGPFEPSHKPDVERPSVDADNDDDEDAGGRGH 676
                     *.

GPC1_HUMAN    Q-------------------------------------------------EGQKTSAASCPQ 536
GPC2_HUMAN    KGGGGSARYN------------------------------------QGRSRSGGASIG 559
GPC3_HUMAN    QATPKDNE-------------------------------------ISTFHNLGN-VH 559
GPC4_HUMAN    -----------------------------------------------KADSAGVRPG 535
GPC5_HUMAN    TDTG-----------------------------------------STLDTTGAGCA 551
GPC6_HUMAN    -------------------------------------------------REVDSSAAQR 534
DALY_DROME    NEVE----------------------------------------SRDSGKTSGSN 604
DLP_DROME     MPTHTSRPTSGVDDKNPLIHTTHFDQDHNDLDEDHRQLDEDEDTDAGHDGANDNRSSDAP 736

GPC1_HUMAN    PPTFLLPLLLFLALTVARPRWR------- 558
GPC2_HUMAN    FHTQTILILSLSALALLGPR--------- 579
GPC3_HUMAN    SPLKLLTSMAISVVCFFFLVH-------- 580
GPC4_HUMAN    AQAYLLTVFCILFLVMQREWR------- 556
GPC5_HUMAN    VATESMTFTLISVVMLLPGIW-------- 572
GPC6_HUMAN    GHSLLSWSLTCIVLALQRLCR------- 555
DALY_DROME    PLEGTATWMLLTLVTMLFSSCS------ 626
DLP_DROME     EKMTLRRALVVYLLPLYMAWFGGVCADLL 765
```

**Supplementary figure 1. Alignment of glypican primary structures.** An alignment of human Gpc-1 through Gpc-6 together with the D. melanogaster glypicans, Dally and Dlp is shown together with a graphical representation of the α helices and major loops in Gpc-1. The alignment was performed using the glypican pro-protein amino acids sequences, i.e. the N- and C-terminal signal peptides are included. Asterisks denote identical amino acid residues in the glypicans while two dots indicate conserved amino acids substitutions (i.e. amino acids belonging to the same class of amino acids) and one dot indicates semiconserved substitutions.

# Supplementary figure 2

**A**



**B**



**C**
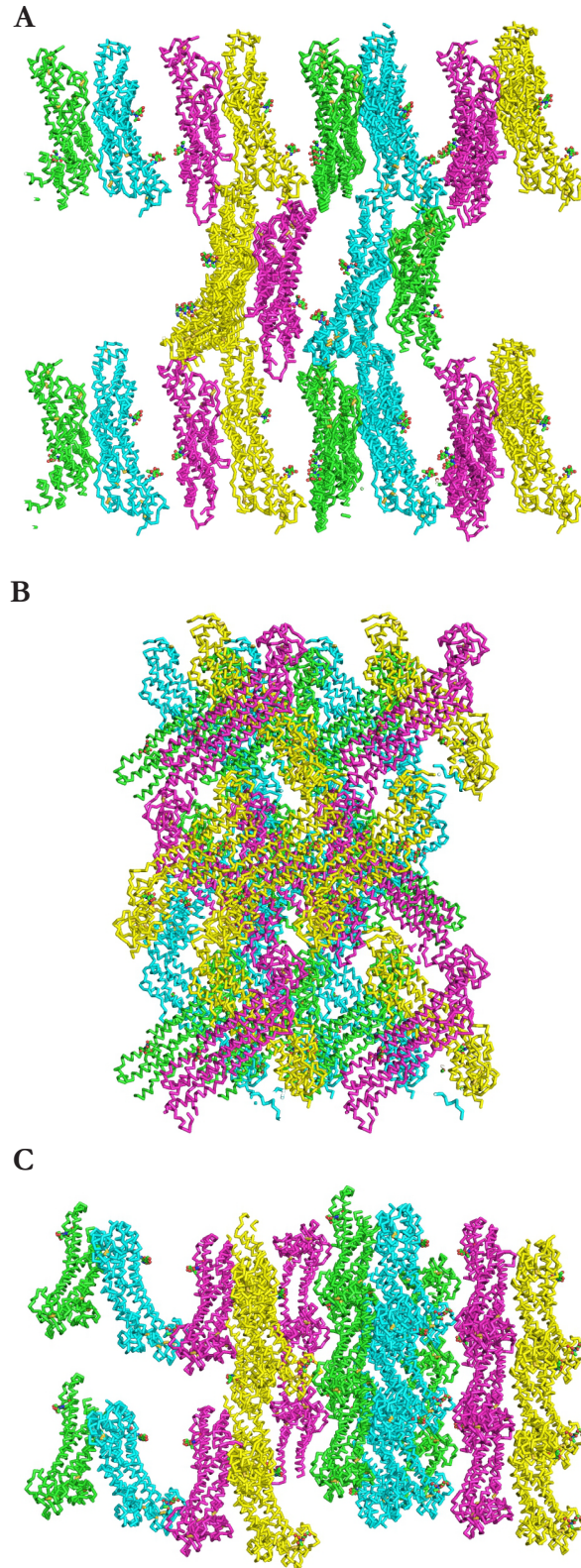


**Supplementary figure 2. Crystal packing of Gpc-1 ΔHS.** Crystal packing of Gpc-1 ΔC and ΔHS in space group $P2_1$ was highly similar. Here shown is the crystal packing of Gpc-1 ΔC represented as α-carbon backbone. The individual monomers in the asymmetric unit are coloured in green, blue, magenta and yellow. Also included are the GlcNAc residues in the N-linked glycans (red and green). In panel B, the molecules are rotated 90° in the horizontal plane compared to A, and in panel C, the molecules are rotated 90° in the vertical plane compared to A.