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

Gabriel Svensson,a Wael Awadb,c, Maria Håkanssond, Katrin Mania, Derek T. Loganc

aDepartment of Experimental Medical Science, Division of Neuroscience, Glycobiology Group, Lund University, Biomedical Center A13, SE-221 84, Lund, Sweden.
bDepartment of Biochemistry and Structural Biology, Center for Molecular Protein Science, Lund University, Getingevägen 60, Box 124, SE-221 00 Lund, Sweden.
cBiophysics Department, Faculty of Science, Cairo University, Cairo, Egypt.
dSARomics Biostructures AB, Box 724, SE-220 07 Lund, Sweden; MAX-lab, Ole Römers väg 1, SE-223 63, Lund, Sweden.

*Corresponding author: phone, +46-46-2221443; fax.: +46-46-2224116; e-mail: derek.logan@biochemistry.lu.se.

Data deposition: The structures have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 4ACR for the C-terminally truncated protein and 4AD7 for the full-length protein)

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

Background: The glypican family of cell-surface proteoglycans regulate growth factor signalling 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 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 Gpc-3 cause Simpson-Golabi-Behmel syndrome, an X-linked condition characterized by prenatal and postnatal overgrowth (3). Mutations in Gpc-6 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 the Gpc-5 gene is associated with acquired nephrotic syndrome (5). Gpc-1 is expressed in the developing and adult brain, and Gpc-1 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-acetylgalactosamine (GlcNAc) residue, while CS synthesis begins with the addition of an N-acetylglucosamine (GlcNAc). 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 Gpc-1 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 Gpc-1 (rhGpc-1) is a stable α-helical protein that resists high concentrations of the chemical denaturants urea and guanidine-HCl (17). Unfolding occurs in a single step and unfolding data can be closely fitted by a two-state model, indicating that Gpc-1 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 Gpc-1 core protein using multiple-wavelength anomalous dispersion (MAD) phasing on selenomethionine (SeMet) labelled rhGpc-1, the first crystal structure of a vertebrate glypican. Structural alignment of Gpc-1 and Dlp reveals a common fold for glypican core proteins. The Gpc-1 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 Gpc-1 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 Mw 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 Gpc-1.**

The cDNA sequence coding for the mature form of human Gpc-1, 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 His6 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 Gpc-1 ΔHS (17). A vector coding for Gpc-1 ΔC was created by amplifying the sequence for amino acids 24-479 using the forward primer 5’-TTAAGCTTGACCCGGCCAGCAAGAG-3’ and the reverse primer 5’-TGGATCCCTTAAAGCTCCAGTGGCATG-3’ and by introducing the PCR-product into the pCEP4-BM40-HisTEV plasmid. Restriction enzyme sites were given in boldface. All constructs were verified by sequencing at Eurofins MWG Operon (Germany). Stable HEK 293 cell clones expressing Gpc-1 ΔHS and Gpc-1 ΔC were generated by limiting dilution after several weeks of growth in selective antibiotics.

Stable cell clones expressing large amounts of rhGpc-1 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 rhGpc-1 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 rhGpc-1 was purified from conditioned medium of ~200 mL batches using nickel affinity chromatography. Protein purification was analyzed by SDS-PAGE and purified rhGpc-1 was dialyzed to 20 mM Tris-HCl pH 8.0, concentrated by ultrafiltration and stored frozen. SeMet labelling of Gpc-1 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 Gpc-1 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 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 rhGpc-1 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 rhGpc-1 stock solution to 0.1 mg/ml for far-UV CD 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 Gpc-1 ΔHS and Gpc-1 ΔC, respectively.

**Crystallization.** Initial high-throughput crystallization screens performed with Gpc-1 ΔHS at a protein concentration of 40 mg/ml were set up in MRC plates using a 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 Mₚ = 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 8000, 0.2 M CaCl², pH 8.0. After two weeks, the same appeared in drops with 20 % PEG 6000, 0.2 M CaCl₂ 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) from stock solutions using a liquid handling robot (TTP Labtech, UK) at the MAXHS at a protein concentration of 40 mg/ml were set up in MRC plates using a nanolitre pipetting robot (Tecan Group Ltd, Switzerland) at the MAX-lab Crystallization Facility, Lund University.

**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 1191-2 and 1911-3 of the MAX-II synchrotron radiation source, Lund, Sweden, on 165nm 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 Gpc-1 ΔHS crystals could be indexed in space group P2₁, 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 Gpc-1 molecules in the asymmetric unit, giving a Matthews volume of 2.7 Å³ / Da and a solvent content of 54%. Crystals of Gpc-1 ΔC and SeMet Gpc-1 ΔC also belong to space group P2₁ with similar unit cell dimensions. The diffraction anisotropy of the data from Gpc-1 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 labelled Gpc-1 ΔC crystals at selenium peak and inflection wavelengths at station 1911-3 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 Gpc-1 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 Gpc-1 structure were generated using the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, www.pymol.org.

**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 Gpc-1.** We expressed His-tagged full-length rhGpc-1 without the GPI-anchor signal peptide in 293 human embryonic kidney cells. This secreted rhGpc-1 (protein referred to as Gpc-1 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 rhGpc-1 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 Gpc-1 ΔHS, Fig. 1A, lane 2). Also, the HS attachment site in Gpc-1 is regarded, by secondary structure and disorder predictions, as adopting a random coil conformation despite the fact that it contains two Cys residues. Therefore, rhGpc-1 was truncated C-terminally before the HS attachment domain, resulting in a core protein band (Gpc-1 ΔC ) of smaller size (~55 kDa) on SDS-PAGE (Fig. 1A, lane 3). We have previously shown that the PG form of Gpc-1 WT and Gpc-1 ΔHS are identical in folding, as judged by far-UV CD and fluorescence emission spectroscopy (17). In order to ensure that Gpc-1 ΔHS and Gpc-1 ΔC are also structurally similar, we subjected the two forms of rhGpc-1 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 Gpc-1 (Fig. 1B and C). Also, expression levels were similar for Gpc-1 ΔHS and Gpc-1 ΔC. Together, this suggests that the C-terminal HS attachment site in Gpc-1 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 Gpc-1 ΔC.** Gpc-1 ΔHS and Gpc-1 ΔC both crystallized in space group P2_1 with four molecules in the asymmetric unit. Crystals of Gpc-1 ΔHS diffracted X-rays to 2.95 Å and crystals of Gpc-1 ΔC diffracted to 2.5 Å. Aside from the improved diffraction from Gpc-1 ΔC crystals, reproducibility of crystallization was greatly improved for Gpc-1 ΔC compared to Gpc-1 ΔHS. SeMet derivative crystals of Gpc-1
ΔC also crystallized in space group P2₁ with unit-cell parameters similar to crystals of native Gpc-1 ΔC and diffracted to 3.3 Å. The structure of Gpc-1 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 Gpc-1 ΔC using the UCLA Anisotropy Server (23) revealed strong anisotropy, with a different of approximately 40 Å² 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 Gpc-1 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 His₆ tag or the protease cleavage site. The asymmetric unit of the Gpc-1 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 Gpc-1 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 Gpc-1 is similar to the recently published structure of the D. melanogaster glypican Dlp, with similarly positioned α helices (Fig. 2B). The Gpc-1 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 Gpc-1 and Dlp only share 25% sequence identity (as determined by FASTA (40)). The overlap between Gpc-1 and Daily-like protein structures is shown as α-carbon traces in Fig. 2B. The r.m.s deviation in α-carbon positions between monomer D of Gpc-1 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 Gpc-1 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 Gpc-1 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 Gpc-1, 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 Gpc-1) 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 Gpc-1 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 Gpc-1 should be called the central lobe. The last part of the Gpc-1 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 Gpc-1 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 (18); an alternative name could be the protease site lobe.

**Analysis of Crystal Contacts.** The interface between pairs of Gpc-1 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 Å². However, calculations of the stabilities of the interfaces using PISA software (43) indicate that the interactions are too weak to form oligomeric states of Gpc-1 in solution. Preliminary dynamic light scattering experiments confirm these calculations, with profiles exhibiting single peaks for both Gpc-1 ΔHS and Gpc ΔC (data not shown).

Evolutionarily Conserved Cysteine Residues in Anchorless Gpc-1. 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 Gpc-1 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.

Gpc-1 also contains two additional Cys residues in the HS attachment region, shared with Gpc-4 and Gpc-6. These Cys residues are not part of the presented Gpc-1 model and can either be free or form a disulfide bond. Crystallization of Gpc-1 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 Gpc-1. A multiple sequence alignment of Gpc-1 through Gpc-6, 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 Gpc-3 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 Gpc-3.

Structural conservation of the C-terminal segment in Gpc-1 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 Gpc-1 through Gpc-6, 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 Gpc-1. 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 Gpc-1 orthologs, i.e. Gpc-1 from different vertebrates. Multiple sequence alignment was performed using Gpc-1 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), Macaca mulatta (rhesus macaque) and Pan troglodytes (common

Evolutionarily Conserved Cysteine Residues in Anchorless Gpc-1. 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 Gpc-1 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. Evolutionarily Conserved Cysteine Residues in Anchorless Gpc-1. 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 Gpc-1 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. Gpc-1 also contains two additional Cys residues in the HS attachment region, shared with Gpc-4 and Gpc-6. These Cys residues are not part of the presented Gpc-1 model and can either be free or form a disulfide bond. Crystallization of Gpc-1 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 Gpc-1. A multiple sequence alignment of Gpc-1 through Gpc-6, 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 Gpc-3 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 Gpc-3. Structural conservation of the C-terminal segment in Gpc-1 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 Gpc-1 through Gpc-6, 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 Gpc-1. 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 Gpc-1 orthologs, i.e. Gpc-1 from different vertebrates. Multiple sequence alignment was performed using Gpc-1 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), Macaca mulatta (rhesus macaque) and Pan troglodytes (common
chimpanzee). As shown in Fig. 5B, L1 and large parts L3 are evolutionarily conserved within vertebrate Gpc-1 sequences, together with the C-terminal α14. Interestingly, L3 has been shown by mutagenesis to be involved in GAG class determination in rat Gpc-1 (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 Gpc-1 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 Gpc-1 is known for many species, contrary to other members of the glypican family.

HS attachment site in Gpc-1. The structure of Gpc-1 was solved by MAD phasing on crystals of SeMet labelled Gpc-1 ΔC. In order to build a model of full-length Gpc-1 in which the HS attachment site was included, we performed model refinement using data from a native Gpc-1 ΔHS crystal. However, no additional electron density was observed, and R_free and R_model values were similar to values obtained using Gpc-1 ΔC (Table 1). Thus it was not possible to obtain information regarding the C-terminal HS attachment site in Gpc-1 using the Gpc-1 ΔHS crystals. This also suggests that the C-terminal HS attachment domain in Gpc-1 is disordered.

Discussion

In a previous study, we investigated the insolution conformation and stability of Gpc-1 and showed that the Gpc-1 core protein is a stable α-helical protein (17). Thermal and chemical unfolding of Gpc-1 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 Gpc-1 ΔHS for both secondary and tertiary structure, and unfolding data closely fitted a two-state model. The stability of Gpc-1 ΔHS, ΔG_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 Gpc-1 is a densely packed one-domain globular protein. These previous results support the crystal structure of Gpc-1, 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 Gpc-1.

During refinement of the Gpc-1 structure, the core protein structure of the D. melanogaster glypican Dlp was published (18). This structure is similar to the Gpc-1 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 Gpc-1 structure was created by using crystals of N-glycosylated Gpc-1 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 Gpc-1 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 Gpc-1, where we show that the correct folding is obtained for Gpc-1 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 Gpc-1 was crystallized at a protein concentration of 20
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 Gpc-1 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 directly interact with integrins (48,49). Glypicans 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 Dally-like structure). This region contains a number of amino acids conserved between Gpc-1 to Gpc-6, Dally and Dlp that are involved in the folding of the protein. Among the amino acids in this region is the Gpc-1 Cys191-Cys343 disulfide bond. The structural similarity in this region is confirmed by our Gpc-1 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 Gpc-1 to mimic the Dlp region important for hedgehog signalling may infer hedgehog signalling effects of Gpc-1 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 Gpc-1 vertebrate orthologs, the L1 and L3 loops of Gpc-1 are evolutionarily conserved, suggesting that they are involved in the function of Gpc-1. No interaction partner has been found for Gpc-1 core protein, contrary to some of the other glycan core proteins. Gpc-1 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). Gpc-1 is exclusively substituted with HS whereas for example Gpc-5 can exist as a HS CS/DS (where DS is dermatan sulfate) hybrid. However, expression of the Gpc-1 GAG attachment domain alone results in addition of ~90% CS (9), strongly suggesting that factors in the Gpc-1 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 Gpc-1 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 EXT family comprises three members, EXT1-3, where EXT1 and EXT3 harbour HS initiating α1,4-GlcNAc transferase activity and are ubiquitously expressed in the human body (53). Only EXT1 has a D. melanogaster ortholog, suggesting that this is the functional HS initiation enzyme, whereas structural data is available for EXT1-2.
the smallest member in the EXTL family. Given the requirement for the glypican core protein to ensure exclusive HS attachment on Gpc-1, 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 initiation 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 Gpc-1 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 Gpc-1 and Dlp are cylindrical in shape. The C-terminal α14 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 Gpc-1 ΔC structure, placing the HS attachment close to the folded core protein. We have previously shown that the HS chains on Gpc-1 do not affect the secondary or tertiary structure of Gpc-1 (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 bind weakly to heparin agarose (18). C-terminal Ser residue responsible for GPI-attachment is located 50 amino acids downstream of the visible ends of the models of Gpc-1 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 Gpc-1 and Dlp with respect to the membrane surface. Our attempts to locate the Gpc-1 HS attachment domain and C-terminus using crystallography have failed, which together with the presented in-solution data on the Gpc-1 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 Gpc-1 structure provides new information which will be important in determining the mechanism behind exclusive HS attachment on glypicans.

Acknowledgments

We thank the beamline staff at MAX-lab for help during data collection and Dr. Mattias Hansson at the MAX-lab Crystallization Facility for help with setting up high-throughput crystallization plates. The technical assistance of Sol Da Rocha is greatly appreciated. We are also grateful to Axel Hyrenius Wittsten for help with CD. The authors also thank the Swedish Research Council (grants 2006-4387 to DL and 2005-6749 and 2010-3914 to KM), the Swedish Cancer Foundation, the Medical Faculty of Lund University, the Royal Physiographic Society in Lund, and the Crafoord, Wiberg, Jeanson, Segerfalk, and Kock Foundations for financial support.
References


Tables

Table I. Data collection and refinement statistics.

<table>
<thead>
<tr>
<th></th>
<th>Native Gpc-1 ΔC</th>
<th>SeMet Gpc-1 ΔC (peak)</th>
<th>SeMet Gpc-1 ΔC (inflection)</th>
<th>Native Gpc-1 ΔHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>(P_2_1)</td>
<td>(P_2_1)</td>
<td>(P_2_1)</td>
<td>(P_2_1)</td>
</tr>
<tr>
<td>Unit-cell parameters (Å, °)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a = 47.2, b = 168.7, c = 147.8, \beta = 94.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a = 47.1, b = 168.5, c = 147.0, \beta = 94.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a = 47.2, b = 168.6, c = 147.4, \beta = 94.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a = 47.2, b = 169.0, c = 151.6, \beta = 95.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>1911-2</td>
<td>1911-3</td>
<td>1911-3</td>
<td>1911-2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.0397</td>
<td>0.9789</td>
<td>0.9794</td>
<td>1.0397</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-2.49 (2.64-2.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-3.3 (3.39-3.30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-3.3 (3.39-3.30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-2.95 (3.03-2.95)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R_{merge}(I)) (%)</td>
<td>8.3 (68.3)</td>
<td>10.7 (33.0)</td>
<td>12.9 (51.1)</td>
<td>13.5 (83.2)</td>
</tr>
<tr>
<td>Total observations</td>
<td>328111</td>
<td>347933</td>
<td>381017</td>
<td>440496</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>79825</td>
<td>66568</td>
<td>66846</td>
<td>49537</td>
</tr>
<tr>
<td>Average redundancy</td>
<td>4.1 (3.5)</td>
<td>5.2 (5.3)</td>
<td>5.7 (5.6)</td>
<td>8.9 (8.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.9 (94.9)</td>
<td>98.2 (98.0)</td>
<td>98.0 (95.4)</td>
<td>99.4 (94.1)</td>
</tr>
<tr>
<td>(I/\sigma(I))</td>
<td>11.6 (2.37)</td>
<td>11.3 (4.47)</td>
<td>10.85 (3.29)</td>
<td>15.4 (3.4)</td>
</tr>
<tr>
<td>Refinement (R_{work}/R_{free}) (%)</td>
<td>25.3/29.2</td>
<td>22.1/27.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of atoms</td>
<td>Protein/carbohydrate 12572</td>
<td>12630</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>242</td>
<td>126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r.m.s. deviations from ideal geometry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths, Å</td>
<td>0.003</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond angles, °</td>
<td>0.72</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramachandran analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In preferred regions (%)</td>
<td>98.6</td>
<td>98.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In allowed regions (%)</td>
<td>1.4</td>
<td>1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 Gpc-1.

<table>
<thead>
<tr>
<th>Cysteine residues</th>
<th>Connected structural elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys32-Cys68</td>
<td>(\alpha_1)-L1</td>
</tr>
<tr>
<td>Cys62-Cys256</td>
<td>(\alpha_1)-(\alpha_7)</td>
</tr>
<tr>
<td>Cys69-Cys259</td>
<td>(\alpha_1)-(\alpha_7)</td>
</tr>
<tr>
<td>Cys191-Cys343</td>
<td>(\alpha_5)-(\alpha_{11})</td>
</tr>
<tr>
<td>Cys246-Cys279</td>
<td>(\alpha_7)-(\alpha_8)</td>
</tr>
<tr>
<td>Cys268-Cys415</td>
<td>Turn between (\alpha_7) and (\alpha_8)-L3</td>
</tr>
<tr>
<td>Cys272-Cys401</td>
<td>(\alpha_8)-(\alpha_{13})</td>
</tr>
</tbody>
</table>

Figure legends

Figure 1. SDS-PAGE and spectroscopic analysis of rhGpc-1. (A) SDS-PAGE and Coomassie blue staining of Ni-NTA purified full-length rhGpc-1 expressed without the GPI-anchor signal peptide (lane 1). Expression of full-length rhGpc-1 with disrupted HS attachment site (Gpc-1 ΔHS) resulted in expression of core protein only (lane 2). Gpc-1 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 Gpc-1 ΔHS (black, solid line) and Gpc-1 ΔC (grey, solid line) in 20 mM sodium phosphate (pH 7.4) were recorded at 20°C.

Figure 2. Overview of Gpc-1 ΔC structure and comparison with Dlp. (A) Overview of the asymmetric unit of Gpc-1 ΔC with two dimer pairs. (B) Ribbon representation of α-carbon alignment of Gpc-1 ΔC monomer (blue) and Dlp (green, PDB ID 3ODN). Alignment was performed in PyMOL, version 1.2r3pre, Schrödinger, LLC (pymol.org). (C) Overview of the Gpc-1 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 Gpc-1 ΔC structure.

Figure 3. Structure model and electron densities of the six disulfide bonds in the N-terminal Cys-rich lobe of Gpc-1. 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; yellow, sulphur. Phe107 and Leu227 are identical in human and D. melanogaster glypicans and form a tight hydrophobic core in the central lobe of Gpc-1 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 Gpc-1 homologs (A) and in Gpc-1 orthologs (B). Multiple sequence alignments 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 Gpc-1 homologs. The following amino acid sequences were used (UniProt accession numbers): Gpc-1 human, P35052; Gpc-2 human, Q8N158; Gpc-3 human, P51654; Gpc-4 human, Q75487; Gpc-5 human, P78333; Gpc-6 human, Q9Y625; Dally D. melanogaster, Q24114; Dlp D. melanogaster, Q9GLP5. (B) The following amino acid sequences were used to visualize structural conservation in Gpc-1 orthologs (i.e. Gpc-1 vertebrate homologs): Gpc-1 human, P35052; Gpc-1 mouse, Q9QZF2; Gpc-1 rat, P35052; Gpc-1 bovine, Q2KJ65; Gpc-1 chick, P50593; Gpc-1 zebrafish, Q1LXM6; Gpc-1 western clawed frog, Q0V9W0. Also included were the translated mRNA sequences (given by NCBI nucleotide accession numbers) for Gpc-1 genes: Gpc-1 short-tailed opossums, XM_001377888; Gpc-1 rhesus macaque, XM_001089255; Gpc-1 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

A

B

C

N-terminus

Cys-rich lobe

Central lobe

Protease-site lobe

GlcNAc N79

GlcNAc N116

α9

α14

C-terminus

C32-C68

C69-C259

C62-C256

C272-C401

C246-C279

C68-C145

C191-C343

C191-C343

α11

α11

180°
Figure 3
Figure 4
Figure 5

A

B

$180^\circ$

variable

conserved
Figure 6
Crystal structure of N-glycosylated human glypican-1 core protein: Structure of two loops evolutionarily conserved in vertebrate glypican-1
Gabriel Svensson, Wael Awad, Maria Hakansson, Katrin Mani and Derek T. Logan

J. Biol. Chem. published online February 20, 2012

Access the most updated version of this article at doi: 10.1074/jbc.M111.322487

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2012/02/20/M111.322487.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2012/02/20/jbc.M111.322487.full.html#ref-list-1