Structural Aspects of N-Glycosylations and the C-terminal Region in Human Glypican-1*§

Received for publication, April 29, 2015, and in revised form, July 10, 2015 Published, JBC Papers in Press, July 22, 2015, DOI 10.1074/jbc.M115.660878

Wael Awaad†*, Barbara Adamczyk†, Jessica Örnros‡, Niclas G. Karlsson§, Katrin Mani¶, and Derek T. Logan†*,‡*

From the †Department of Biochemistry and Structural Biology, Centre for Molecular Protein Science, Lund University, Box 124, SE-221 00 Lund, the ‡Department of Biochemistry and Cell Biology, University of Gothenburg, Box 440, SE-40530 Gothenburg, and the §Department of Experimental Medical Science, Lund University, SE-221 84, Lund, Sweden

Background: Glypicans are a family of cell surface proteoglycans implicated in diverse cell signaling pathways.

Results: We provide a description of the structure and functions of the N-glycans and C terminus of human glypican-1.

Conclusion: The studies revealed the structural topology of glypicans with respect to the membrane and the protective roles of their N-glycans.

Significance: Improved structural knowledge of glypican-1 helps to elucidate the functional roles of the glypicans.

Glypicans are multifunctional cell surface proteoglycans involved in several important cellular signaling pathways. Glypican-1 (Gpc1) is the predominant heparan sulfate proteoglycan in the developing and adult human brain. The two N-linked glycans and the C-terminal domain that attach the core protein to the cell membrane are not resolved in the Gpc1 crystal structure. Therefore, we have studied Gpc1 using crystallography, small angle x-ray scattering, and chromatographic approaches to elucidate the composition, structure, and function of the N-glycans and the C terminus and also the topology of Gpc1 with respect to the membrane. The C terminus is shown to be highly flexible in solution, but it orients the core protein transverse to the membrane, directing a surface evolutionarily conserved in Gpc1 orthologs toward the membrane, where it may interact with signaling molecules and/or membrane receptors on the cell surface, or even the enzymes involved in heparan sulfate substitution in the Golgi apparatus. Furthermore, the N-glycans are shown to extend the protein stability and lifetime by protection against proteolysis and aggregation.

Glypilated proteins are anchored to the extracellular surface of the eukaryotic cell membrane by covalent linkage of their C termini to glycosylphosphatidylinositol (GPI)3 (1). Glypicans (Gpcs) are a family of glypilated extracellular proteoglycans that mainly work as co-regulators of several signaling pathways, and they are thereby involved in the control of many biological processes such as cellular division, differentiation, and morphogenesis. To date, six different Gpcs have been identified in vertebrates, Gpc1 to Gpc6, with ~25% amino acid identity within the family, two in Caenorhabditis elegans (Gpc1 and Lon-2), two in Drosophila melanogaster (Dally and Dally-like protein), and one in zebrafish (knypek). The mature forms of human Gpcs have a core protein of ~60–70 kDa in size and share a pattern of 14 conserved cysteine residues. In the C-terminal region, Gpcs share attachment sites for glycosaminoglycan (GAG) chains and a hydrophobic sequence for the addition of the GPI anchor (2).

GAGs are O-linked linear anionic hetero-polysaccharides found in all mammalian tissues in extracellular and/or intracellular environments. Their sequences are composed of repeating disaccharide units and are divided into classes depending on the disaccharide building blocks. The most widespread categories are as follows: 1) heparan sulfate (HS), which consists of repeating units of glucuronic acid (GlcUA) or iduronic acid (IdoUA) followed by N-acetylgalactosamine (GlcNAc), i.e. (GlcUA/IdoUA-GlcNAc)n, 2) chondroitin sulfate (CS) and dermatan sulfate, which consist of GlcNAc or IdoUA followed by N-acetylgalactosamine (GalNAc), i.e. (GlcUA/IdoUA-GalNAc)n. The biosynthesis of GAG chains is performed by membrane-bound glycosyltransferases in the Golgi apparatus and starts with addition of xylose (Xyl) to the GAG attachment serine residue by a xylosyltransferase. Then two consecutive galactose units and a GlcUA unit are transferred, forming the linkage tetrasaccharide ([GlcUAβ1–3Galβ1–3Galβ1–4Xylβ1–O-Ser]). Addition of the next residue is the critical factor in determining which type of GAG will be formed as follows: incorporation of GalNAc promotes CS/dermatan sulfate assembly, and addition of GlcNAc initiates HS synthesis. The GAG chain then becomes elongated by addition of the corresponding repeating disaccharides and undergoes serial modification-competition, leading to the formation of a variety of different GAGs that mediate specific biological processes.

*This work was supported by the Swedish Research Council, European Union FP7 GastricGlycoExplorer ITN Grant 316929, the Swedish Cancer Society, Alfred Österlund, and the Ahlen and Olle Engkvist Foundation. The authors declare that they have no conflicts of interest with the contents of this article.

†This article contains supplemental Tables S1–S3.

The atomic coordinates and structure factors (code 4YWT) have been deposited in the Protein Data Bank (http://wwpdb.org/).

1 Present address: Dept. of Biophysics, Faculty of Science, Cairo University, 12316 Cairo, Egypt.

2 To whom correspondence should be addressed. Tel.: 4646-222-1443; Fax: 4646-222-4116; E-mail: derek.logan@biochemistry.lu.se.

3 The abbreviations used are: GPI, glycosylphosphatidylinositol; Gpc, glypcican; HS, heparan sulfate; Ni-NTA, nickel-nitriotic acid; CS, chondroitin sulfate; PDB, Protein Data Bank; EndoH, endoglycosidase H; 2-AB, 2-aminobenzamide; GU, glucose unit; SEC, size-exclusion chromatography; SAXS, small angle x-ray scattering; DLS, dynamic light scattering; GAG, glycosaminoglycan; MES, minimal ensemble search; HLIC-FLD-UPLC, hydrophilic interaction liquid chromatography with fluorescence detection-competition-liquid performance liquid chromatography; EXT, exostosin; PI, polydispersity index; Gpcs, glypicans; EOM, ensemble optimization method.
Solution Structure of Human Glypican-1 Core Protein

cations, including N-deacetylation, N- and O-sulfation, and epimerization, generating GAG chains with heterogeneous structures (3).

Human Gpc1 is mainly expressed in the neural and skeletal systems during development, as well as many other tissues in the adult. Gpc1 is involved in the uptake of different macromolecules such as growth factors (FGF2), viral proteins, cytokines, and polyamines (4). Gpc1 knock-out mice have significant reduction in brain size at birth (about 30%), indicating a role for Gpc1 in brain development (5). Many studies have revealed involvement of Gpc1 in the pathogenesis of neurodegenerative diseases, including Alzheimer disease and transmissible spongiform encephalopathies (6, 7). Moreover, it has been shown that Gpc1 is up-regulated in many human cancer types such as glioma, pancreatic cancer, and breast cancer (8, 9).

The Gpc1 protein is composed of an N-terminal core (residues 24–474) and a C-terminal GAG attachment region (residues 475–530), ending with a sequence of hydrophobic residues that link it to the GPI anchor (Fig. 1A). The Gpc1 protein is decorated with two N-linked glycans at positions Asn-79 and Asn-116 and is also substituted with three chains of HS at Ser-486, Ser-488, and Ser-490 (Fig. 1A). Previously, we determined the crystal structure of the N-glycosylated, C-terminally truncated Gpc1 core protein at 2.6 Å (PDB entry 4ACR) and showed that it has an elongated cylindrical form with a minimally truncated Gpc1 core protein at 2.6 Å (PDB entry 4ACR). A

Previously, we achieved crystallization of Gpc1-dHS protein was performed

Expression and Purification of Human Gpc1—The design of different Gpc1 variants is summarized in Table 1. All Gpc1 constructs were cloned and expressed in stable HEK293 cells as described previously (13). The conditioned medium was collected, and the proteins were purified using Ni-NTA affinity chromatography. The proteins were then dialyzed into 20 mM NaCl, 20 mM Tris, pH 8, for the C-terminally truncated protein (Gpc1-dC) and the same buffer with additional 2 mM DTT for the others before any further experiments.

Enzymatic Treatments—Deglycosylation of Gpc1-dC and the full-length protein without the HS chains (Gpc1-dHS) was carried out by growing the Gpc1-producing cells in protein-free medium containing 10 µM plant alkaloid kifunensine, which forces the cells to express proteins with N-glycans of the high mannose type that are sensitive to endoglycosidase H (EndoH) (14). Afterward, EndoH treatment (New England Biolabs) of 1 mg of native high mannose Gpc1 was carried out in 25 mM sodium phosphate buffer, pH 7.0, by overnight incubation at 37 °C with 30 milliunits of the enzyme. The deglycosylation efficiency was tested by SDS-PAGE, which showed a reduction in the size of the Gpc1-dC and Gpc1-dHS core protein bands by 5 kDa after deglycosylation, producing deglycosylated forms of the Gpc1 core proteins (Gpc1-dC-dN and Gpc1-dHS-dN), respectively. EndoH was removed from the sample by repeating the Ni-NTA purification.

Digestion of the HS chains was carried out as follows: the purified proteins were dialyzed into 10 mM HEPES and 3 mM Ca(OAc)₂ buffer, pH 7.0, overnight, and then 150 milliunits of HS lyase (Seikagaku, Japan) were added to 1 mg of protein and incubated overnight at 37 °C. The protein buffer was exchanged to 0.3 M NaCl, 50 mM sodium phosphate, pH 8.0, by ultrafiltration, and then incubated with DE-53 DEAE-cellulose (for anion exchange) on a rocker at 8 °C for 1 h. The unbound proteins (without heparan sulfate) were released by washing the cellulose twice with 0.3 M NaCl, 50 mM sodium phosphate, pH 8, and finally the HS lyase was removed by repeating the Ni-NTA purification. HS lyase digests the HS polysaccharide chains, leaving only the tetrasaccharide linkers (GlcUAβ1–3Galβ1–3Galβ1–4Xylβ1) on the consensus serine residues of the HS attachment sites. Removal of HS chains from the wild type (Gpc1-WT) and the proteins with disrupted N-glycosites, Gpc1-N79Q and Gpc1-N116Q, produced the Gpc1-WT-dHS, Gpc1-N79Q-dHS, and Gpc1-N116Q-dHS proteins, respectively.

Protein Crystallization, Dehydration, and Structure Determination—Crystalization of Gpc1-dHS protein was performed using sitting drop vapor diffusion by mixing 2 µl of protein at 25 mg/ml with 2 µl of reservoir solution containing 14–16% PEG 6000, 0.1 M Tris-HCl, pH 8.0, and 0.2 M CaCl₂, equilibrated over 0.5 ml of reservoir solution. Thin plate-like crystals of dimen-
sions around 0.9 × 0.2 × 0.05 mm grew in 2 weeks within a lot of precipitated protein. Gpc1-dHS crystals were soaked in 16% PEG 6000, 0.2 m CaCl₂, 0.1 m Tris, pH 8.0, and 15% ethylene glycol and then mounted on a mesh LithoLoop (Molecular Dimensions, UK) in the HC1b machine at beamline I1911-3 of the MAX IV Laboratory, Lund, Sweden. The Gpc1-dHS crystals were dehydrated for a total incubation time of 50 – 60 min to a final relative humidity of 86 – 88%. The dehydrated crystals were then flash-frozen and stored in liquid nitrogen using the CATS sample changer (IRELEC, Saint-Martin-d’Hères, France), and subsequently complete diffraction data sets were collected at 100 K. Diffraction images were indexed and scaled using XDS (15) and were further processed using programs from CCP4 package (16). The initial model was obtained by rigid body refinement of the dehydrated Gpc1-dC model (PDB code 4BWE) using REFMAC5 (17), followed by manual building in Coot (18) and rounds of refinement using phenix.refine (19). Finally, model validation was performed using Molprobity (20). PyMOL Version 1.6 (Schrödinger, LLC) was used for molecular rendering.

-N-Glycan Characterization by HILIC-FLD-UPLC—The Gpc1 proteins (150 μg) were immobilized on a 10-kDa spin filter ( Pall, Port Washington, NY), treated with denaturation buffer (50 μl of 50 mM DTT, 20 mM NaHCO₃), and incubated at 65 °C for 15 min to denature the proteins to allow for efficient cleavage of the N-glycans. After cooling to room temperature for 10 min, an iodacetamide solution (50 mM, 50 μl per well) was added and incubated at room temperature for 30 min. Samples were spun at 12,000 rpm for 5 min, and the flow-through was discarded, followed by two washings with MilliQ water (50 μl) to remove residual amounts of denaturation reagents. N-Glycans were enzymatically released by digestion with recombinant N-glycosidase F (50 μl, 0.5 milliunits in 20 mM NaHCO₃, pH 7.0) (ProZyme, Hayward, CA) and incubated at 37 °C overnight. After extraction, glycans were derivatized with 2-aminobenzamide (2-AB), sodium cyanoborohydride, 30% v/v ammonia acetate in 5% acetonitrile. The elution was repeated from 30 to 47% over a 25-min period. An injection volume of 25 μl of sample prepared in 70% v/v acetonitrile was used through- out the SAXS measurements with the known absolute scattering of water (I₀,abs(water) = 1.632.10⁻² cm⁻¹, at 25 °C).

Size exclusion directly in-line with SAXS (SEC-SAXS) was used to obtain scattering data from highly pure monomers of Gpc1-dHS and Gpc1-dC using the HPLC system (Viscotek GPCmax, Malvern Instruments) attached directly to the sample inlet valve of the sample changer at BM29. 100 μl of ~6 mg/ml clarified sample was loaded onto a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with 2 column volumes of the protein standard buffer. The protein was eluted at a flow rate of 0.5 ml/min and passed through the capillary cell, and a scattering frame was collected every 2 s with a total of 1200 frames. The EDNA pipeline (25) provided a one-dimen- sional profile for each frame. The individual frame was subtracted and processed using tools from the ATSAS suite (26) calculating the forward scattering I(0) and radius of gyration Rg. Frames with stable Rg values from the monomer peak scattering intensity (the second half of the peak) were merged to provide a single averaged frame corresponding to the scattering of an individual SEC-purified monomer.

The sequence and linkage specificities of the glycans were analyzed by exoglycosidase enzymes purchased from ProZyme (Hayward, CA). 2-AB-labeled glycans were digested in 10 μl of supplied buffers at 37 °C overnight, using arrays of the following enzymes: ABS, Arthrobacter ureafaciens sialidase (EC 3.2.1.18, releases α2–3,6,8-linked nonreducing terminal sialic acid); NANA, recombinant sialidase (EC 3.2.1.18, releases α2–3-linked nonreducing terminal sialic acids); bovine kidney α-fucosidase (EC3.2.1.51, releases α1–2,6-linked nonreducing terminal fucose residues more efficiently than α1–3,4-linked fucose, digests core α1–6 fucose); bovine testes β-galactosidase (EC 3.2.1.23, hydrolyzes nonreducing terminal β1–4- and β1–3-linked galactose); hexosaminidase cloned from Streptococcus pneumoniae expressed in Escherichia coli (EC 3.2.1.30), releases GlcNAc residues but not a bisecting GlcNAc-linked to Man); and jack bean α-mannosidase (EC 3.2.1.24, hydrolyzes α(1–2,3,6)-linked mannose from oligosaccharides). After incubation, enzymes were removed by filtration using Pall spin filters of 10 kDa. N-Glycans were then analyzed by HILIC-FLD-UPLC as described above.

Analytical SEC and DLS—SEC analysis was performed at room temperature using a pre-equilibrated Superdex 200 10/300 GL column (GE Healthcare) in the standard buffers. Sample homogeneity was assayed by DLS using a Zetasizer APS DLS system (Malvern Instruments Ltd., Malvern, UK) under the same conditions as the SAXS measurements were done, and the data were analyzed using Zetasizer software version 7.03.

SAXS Data Collection and Processing—Synchrotron radiation SAXS data were collected on the ID14-3 and BM29 beamlines, ESRF, Grenoble, France (24). Water measurements (empty capillary and water) were employed as a reference for further measurements and to give preliminary estimates for the sample molecular weight using the known absolute scattering of water (I₀,abs(water) = 1.632.10⁻² cm⁻¹, at 25 °C).

A representative HILIC-FLD-UPLC profile of Gpc1 was annotated with glucose unit (GU) values by comparison with a dextran hydrolysate ladder (21). Initial structural assignment of the glycans present in the peaks was performed by comparison of experimental data with GlycoBase (22). The Consortium for Functional Glycomics (CFG) glycan notation is used throughout (23).

The sequence and linkage specificities of the glycans were analyzed by exoglycosidase enzymes purchased from ProZyme (Hayward, CA). 2-AB-labeled glycans were digested in 10 μl of supplied buffers at 37 °C overnight, using arrays of the following enzymes: ABS, Arthrobacter ureafaciens sialidase (EC 3.2.1.18, releases α2–3,6,8-linked nonreducing terminal sialic acid); NANA, recombinant sialidase (EC 3.2.1.18, releases α2–3-linked nonreducing terminal sialic acids); bovine kidney α-fucosidase (EC3.2.1.51, releases α1–2,6-linked nonreducing terminal fucose residues more efficiently than α1–3,4-linked fucose, digests core α1–6 fucose); bovine testes β-galactosidase (EC 3.2.1.23, hydrolyzes nonreducing terminal β1–4- and β1–3-linked galactose); hexosaminidase cloned from Streptococcus pneumoniae expressed in Escherichia coli (EC 3.2.1.30), releases GlcNAc residues but not a bisecting GlcNAc-linked to Man); and jack bean α-mannosidase (EC 3.2.1.24, hydrolyzes α(1–2,3,6)-linked mannose from oligosaccharides). After incubation, enzymes were removed by filtration using Pall spin filters of 10 kDa. N-Glycans were then analyzed by HILIC-FLD-UPLC as described above.

Analytical SEC and DLS—SEC analysis was performed at room temperature using a pre-equilibrated Superdex 200 10/300 GL column (GE Healthcare) in the standard buffers. Sample homogeneity was assayed by DLS using a Zetasizer APS DLS system (Malvern Instruments Ltd., Malvern, UK) under the same conditions as the SAXS measurements were done, and the data were analyzed using Zetasizer software version 7.03.

SAXS Data Collection and Processing—Synchrotron radiation SAXS data were collected on the ID14-3 and BM29 beamlines, ESRF, Grenoble, France (24). Water measurements (empty capillary and water) were employed as a reference for further measurements and to give preliminary estimates for the sample molecular weight using the known absolute scattering of water (I₀,abs(water) = 1.632.10⁻² cm⁻¹, at 25 °C).

Size exclusion directly in-line with SAXS (SEC-SAXS) was used to obtain scattering data from highly pure monomers of Gpc1-dHS and Gpc1-dC using the HPLC system (Viscotek GPCmax, Malvern Instruments) attached directly to the sample inlet valve of the sample changer at BM29. 100 μl of ~6 mg/ml clarified sample was loaded onto a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with 2 column volumes of the protein standard buffer. The protein was eluted at a flow rate of 0.5 ml/min and passed through the capillary cell, and a scattering frame was collected every 2 s with a total of 1200 frames. The EDNA pipeline (25) provided a one-dimen- sional profile for each frame. The individual frame was subtracted and processed using tools from the ATSAS suite (26) calculating the forward scattering I(0) and radius of gyration Rg. Frames with stable Rg values from the monomer peak scattering intensity (the second half of the peak) were merged to provide a single averaged frame corresponding to the scattering of an individual SEC-purified monomer.
Solution Structure of Human Glypican-1 Core Protein

For concentration series measurements, the samples were dialyzed overnight against the standard buffers, and concentration series were prepared. A sample changer robot was used to load 35 μl of the clarified sample into the measurement capillary, and SAXS data were collected in flow mode. Scattering profiles of the filtered dialysis buffers were subtracted from the corresponding sample scattering profiles. Real space $R_e$ excluded particle (Porod) volume, and the maximum particle dimension ($D_{max}$) were estimated from the pair distance distribution function $P(r)$ using the program GNOM. Molecular mass was determined using the program SAXSMoW. Porod-Debye plots were employed to analyze the protein flexibility according to Rambo and Tainer (27).

Ab Initio Modeling—Ab initio shape reconstruction was performed using the simulated annealing method as implemented in the DAMMIN program (28). 20 different models were aligned and averaged, and the most typical model was generated using DAMAVER program suite (29). To test the dependence of the SAXS shape reconstruction on methods, the program GASBOR (30) was also used to compute a set of 20 different ab initio shape envelopes using GNOM output up to 4.5 nm$^{-1}$ and the same $D_{max}$ and $R_g$ values used with DAMMIN.

All-atom Modeling, MES, and EOM—All-atom models (AllosMod) of Gpc1-dC and Gpc1-dHS were generated using tools on the SaliLab web server (54). ModLoop was used for modeling of loops missing in the Gpc1 crystal structure (PDB code 4YWT) (31). No building was performed for the missing residues at the N and C termini of the core proteins. Next, static models of Gpc1-dC and Gpc1-dHS with flexible glycans were generated in MODELLER utilizing the ModLoop output. Monosialylated digalactosylated biantennary glycans (at position Asn-79) and core-fucosylated monogalactosylated biantennary complex glycans with bisecting GlcNAc (at position Asn-116), corresponding to the predominant glycoforms identified in the MS studies, were added with ideal geometries, followed by a 1 Å randomization of the all-atom coordinates, where the motions were restricted to loops and surface side chains. The generated core models were then used for further modeling jobs. Finally, a 330 K AllosMod simulation was used to generate 2000 protein conformations with rigid sugars consistent with the input structure. The ensembles of glycosylated models were fitted to the raw SAXS data using FoXS (32). The $\chi^2$ value was used to evaluate the goodness of fit and to select the AllosMod models of best/poorest consistency with the SAXS data. Minimal ensemble searches (MES) with up to four conformations were tried to minimize the $\chi^2$ values (33).

Furthermore, an ensemble containing a number of different conformations of the flexible parts was obtained by the EOM (34). The static Gpc1 model generated by MODELLER (residues 26–476) was used as a fixed core in RANCH to generate a large pool of 40,000 independent conformers of the N- and C-terminal residues of Gpc1-dHS with the four best orientations of $N$-glycans (obtained from Gpc1-dC MES results). GAJOE was used for ensemble selection by minimizing the discrepancy $\chi^2$.

Rigid Body Modeling—Molecular modeling of the N and C termini was conducted using rigid body modeling as implemented in the program CORAL (26), based on the static model generated by MODELLER. $N$-Glycan chains with structures identical to the ones used in Allosmod runs were treated as rigid bodies with contact restraints of 7 Å to the appropriate asparagine residues Asn-79 and Asn-116. Fifty independent CORAL runs were performed by minimizing the discrepancy $\chi^2$ between the theoretical scattering curve calculated from the model and the experimental data. For Gpc1-dC dimer generation, 2-fold symmetry was applied. All CORAL models were aligned in Supcomb13 (35) and analyzed in PyMOL to identify the most typical structures. The theoretical scattering amplitude of the generated models together with the their discrepancies $\chi^2$ to the experimental data were calculated using CRYSOl (36).

Peptide Mass Spectrometry—In-gel trypsin digestion of Gpc1 SDS-PAGE bands followed by LC-MS/MS analysis was performed to identify protein sequences as described previously (37). Data-dependent mass spectrometry experiments were performed with an EASY LC Nano Flow high performance liquid chromatography (HPLC) system (Proxeon Biotools, Odense, Denmark) connected to an LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific, Waltham, WA) equipped with a nano-Easy spray ion source (Proxeon Biosystems, Odense, Denmark). The chromatographic separation was performed at 40 °C on a 15-cm (75-μm inner diameter) EASY-Spray column packed with 3 μm of resin (Proxeon Biosystems, Odense, Denmark). The nano-HPLC intelligent flow control gradient was 5–20% solvent B (0.1% (v/v) formic acid, 100% (v/v) acetonitrile in water) in solvent A (0.1% (v/v) formic acid in water) for 120 min and then 20–40% for 60 min followed by an increase to 90% for 5 min. A flow rate of 300 nl/min was used through the whole gradient. An MS scan (400–1400 m/z) was recorded in the Orbitrap mass analyzer set at a resolution of 60,000 at 400 m/z, 1 × 10^6 automatic gain control target, and 500-ms maximum ion injection time. The MS was followed by data-dependent collision-induced dissociation MS/MS scans on the eight most intense multiply charged ions in the LTQ at a 500 signal threshold, 3 m/z isolation width, 10-ms activation time at 35 normalized collision energy and dynamic exclusion enabled for 60 s. The general mass spectrometric conditions were as follows: spray voltage 2.0 kV; no sheath or auxiliary gas flow; S-lens 60%; ion transfer tube temperature 275 °C. Raw data were processed by Mascot Distiller searching the SwissProt database (release, December 11, 2013, containing 541,954 entries) with an in-house Mascot database. The search parameters for the Mascot searches were as follows: taxonomy, *Homo sapiens*; enzyme, trypsin or chymotrypsin; variable modifications, oxidation (methylation); precursor tolerance, 20 ppm; and MS/MS fragment tolerance, 0.1 Da.

Results

We expressed and purified various His$_{6}$-tagged constructs of GPI anchorless Gpc1 (summarized in Table 1) as described previously (13, 38). Briefly, we produced anchorless full-length Gpc1 with and without HS chains (by site-directed mutagenesis of HS attachment sites S486A, S488A, and S490A) (Gpc1 WT(24–529) and Gpc1-dHS(24–529)), as well as C-terminally truncated Gpc1 lacking the C-terminal region (Gpc1-dC(24–
Gpc1-dHS Crystal Dehydration and Structure Determination—The full-length Gpc1-dHS protein crystallized in space group P2₁, containing four monomers in the asymmetric unit, with unit cell dimensions of \(a = 47.2, b = 169.0, c = 151.6 \text{ Å}, \beta = 95.0^\circ\), and crystals diffracted to \(\sim 3\) Å resolution (10). Similarly to crystals of Gpc1-dC, the Gpc1-dHS crystals were not isomorphous (with \(c\) dimension varying between 148 and 155 Å) and diffracted anisotropically, as revealed by a significantly higher \(B\) factor in the \(c^*\) dimension than in the \(a^*\) and \(b^*\) directions (\(\Delta B = 56 \text{ Å}^2\)) (39). Some parts of the structure, including the long C terminus, were disordered and not visible in the electron density map.

We have previously shown that controlled dehydration using the HC1b machine greatly improved the diffraction properties and amount of visible structure in Gpc1-dC crystals (12). Here, we investigated whether we could improve the diffraction quality of full-length Gpc1-dHS crystals using the same method, aiming to resolve at least some of the structure of the C-terminal region. We optimized the dehydration protocol for Gpc1-dHS crystals and succeeded in reproducing isomorphous dehydrated crystals with unit cell dimensions of \(46.8, b = 166.6, c = 137.7 \text{ Å}, \beta = 90.4^\circ\), which diffracted to 2.1 Å in the best orientation. This represents a reduction in the \(c\) axis length by 13.9 Å. Complete 2.3 Å data were collected from a crystal dehydrated to a final relative humidity of 87% with dehydration rate of 0.5% per 200 s and total incubation time of 55 min in the humidified air stream of the HC1b machine (Table 2). The anisotropy \(\Delta B (22 \text{ Å}^2)\) and Wilson \(B\) factor (38.7 Å) of the new data were reduced by 61 and 27%, respectively, compared with nondehydrated crystals. These revealed significant improvement in the lattice order and packing after dehydration, generated much better and less noisy electron density maps, allowing the building of more complete monomers in the asymmetric unit (5% more than in 4AD7), and displayed a better defined side-chain density. The final Gpc1-dHS model included residues Pro-25–Asp-475 with only a few missing residues (Pro-350 to Arg-360 and Ser-408 to Asp-412). The overall model \(B\) factor fell from 74.8 to 59.1 Å² after dehydration. The backbone flexibility of Gpc1 as measured by the \(C_\beta\) \(B\) factors was significantly higher than average for residues close to the Asn-79 glycosylation site, the N-terminal region (including elements \(\alpha_1\), \(L_1\), and part of \(\alpha_2\)), and parts of the protease site lobe (involving \(\alpha_4\), \(\alpha_5\), \(\alpha_{11}\), and \(L_2\)). The L3 loop was less flexible than the others, being stabilized by two disulfide bonds (Fig. 1B). Unfortunately, no additional electron density was observed for the C-terminal domain after dehydration, which confirms that the HS attachment region is highly flexible and has no unique structure in the Gpc1-dHS crystals. This raises the question whether the C terminus is intrinsically disordered or contains any locally ordered structure.

### Solution Structure of Human Glypican-1 Core Protein

| Gpc1 proteins as cited in text | Characteristics |
|-------------------------------|-----------------|
| Gpc1-dC                        | C-terminally truncated Gpc1 carrying two N-glycans |
| Gpc1-dC-dN                     | EndoH-deglycosylated C-terminally truncated Gpc1 |
| Gpc1-dHS                       | Full-length Gpc1 carrying two N-glycans but no HS substitution (by mutagenesis of S486A, S488A, and S490A) |
| Gpc1-dHS-dN                    | EndoH-deglycosylated full-length Gpc1 |
| Gpc1-WT                        | Wild-type Gpc1 substituted with three HS chains and two N-glycans |
| Gpc1-WT-dHS                    | Gpc1-WT treated with HS lyase enzyme |
| Gpc1-N79Q                      | Gpc1 substituted with three HS chains and one glycan at Asn-116 (by mutagenesis of N79Q) |
| Gpc1-N79Q-dHS                  | Gpc1-N79Q treated with HS lyase enzyme |
| Gpc1-N116Q                     | Gpc1 substituted with three HS chains and one glycan at Asn-79 (by mutagenesis of N116Q) |
| Gpc1-N116Q-dHS                 | Gpc1-N116Q treated with HS lyase enzyme |

| TABLE 2 |
| Dehydrated Gpc1-dHS (4YWT) |
|---------------------------|
| Wavelength (Å)            | 1.000 |
| Resolution range (Å)      | 43.22–2.38 (2.47–2.38) |
| Space group               | P2₁ |
| Unit cell                 | 46.8 166.6 137.7 90.4 90.4 |
| Total reflections         | 299,725 (30301) |
| Unique reflections        | 82,475 (8185) |
| Multiplicity              | 3.6 (3.7) |
| Completeness (%)          | 97.9 (97.1) |
| Mean I/σ(I)               | 7.44 (1.62) |
| Wilson B factor           | 38.7 |
| Rmerge (F)                | 0.108 (0.695) |
| CCC(2)/2 (F)              | 0.996 (0.631) |
| CC(1) (F)                 | 0.999 (0.880) |
| Rwork (F)                 | 0.239 (0.340) |
| Rfree (F)                 | 0.273 (0.348) |
| No. of non-hydrogen atoms | 13,110 |
| Macromolecules            | 12,761 |
| Ligands                   | 65 |
| Water molecules           | 285 |
| Protein residues          | 1679 |
| Root mean square deviation from ideal geometry bonds (Å) | 0.004 |
| Root mean square deviation, angles (°) | 0.97 |
| Ramachandran favored (%)  | 98 |
| Ramachandran outliers (%) | 0.1 |
| Average B-factor          | 59.00 |
| Macromolecules            | 59.87 |
| Ligands                   | 78.74 |
| Solvent                   | 44.76 |

| TABLE 1 |
| Description of all forms of Gpc1 proteins used in this study |
|-------------------|
| Gpc1-WT            | Wild-type Gpc1 substituted with three HS chains and two glycans |
| Gpc1-dC            | C-terminally truncated Gpc1 carrying two N-glycans |
| Gpc1-dC-dN         | EndoH-deglycosylated C-terminally truncated Gpc1 |
| Gpc1-dHS           | Full-length Gpc1 carrying two N-glycans but no HS substitution (by mutagenesis of S486A, S488A, and S490A) |
| Gpc1-dHS-dN        | EndoH-deglycosylated full-length Gpc1 |
| Gpc1-WT            | Wild-type Gpc1 substituted with three HS chains and two N-glycans |
| Gpc1-WT-dHS        | Gpc1-WT treated with HS lyase enzyme |
| Gpc1-N79Q          | Gpc1 substituted with three HS chains and one glycan at Asn-116 (by mutagenesis of N79Q) |
| Gpc1-N79Q-dHS      | Gpc1-N79Q treated with HS lyase enzyme |
| Gpc1-N116Q         | Gpc1 substituted with three HS chains and one glycan at Asn-79 (by mutagenesis of N116Q) |
| Gpc1-N116Q-dHS     | Gpc1-N116Q treated with HS lyase enzyme |

479)). To investigate the structural roles of the individual N-glycans, two different Gpc1 constructs with disrupted N-glycosylation sites but intact HS attachment sites were also produced (Gpc1-N79Q(24–529) and Gpc1-N116Q(24–529)).
The structures of the N-glycans decorating the Gpc1 core protein were investigated by chromatographic approaches. The total N-glycan pools of Gpc1-WT, Gpc1-N79Q, and Gpc1-N116Q were assigned using exoglycosidase digestions and analysis on the 1.7-/H9262 m HILIC phase (Fig. 2). The resulting data confirmed the structural diversity of the glycans in all the samples. Emphasis was placed on the identification of the most abundant peaks to use in SAXS modeling (see below) rather than a comprehensive analysis of the structures of all N-glycans. The Gpc1-WT profile consisted of 44 chromatographic peaks (supplemental Table S1). The largest proportions of oligosaccharides on Gpc1-WT have core-fucosylated biantennary structures with bisecting GlcNAc carrying one galactose (FA2BG1) and N-glycan core structure with 4 N-acetylhexosamine residues (HexNAc), which represent 9.7% of the total glycan pool. Core fucosylated biantennary structures with bisecting GlcNAc carrying one galactose (FA2BG1) and N-glycan core structure with 4 N-acetylhexosamine residues (HexNAc), which represent 9.7% of the total glycan pool. Core fucosylated biantennary structures with bisecting GlcNAc carrying two galactoses (FA2BG2), mannose-5 (M5), biantennary structure with bisecting GlcNAc carrying two galactoses and one sialic acid (A2BG2S1(6)), and finally biantennary glycan with bisecting GlcNAc and one galactose (A2BG1) accounted for 5.8, 4.6, 4.4, and 4% of the total glycan pool, respectively (supplemental Table S2).

A series of exoglycosidase digestions was performed to identify and assign N-glycan structures to particular chromatographic peaks. The undigested profile of Gpc1-WT consisted of 44 chromatographic peaks, but the majority of the peaks had less than 2% abundance (supplemental Table S1). A panel of digestions for Gpc1 N-glycans that includes the most commonly used exoglycosidase enzymes (α-sialidase, α-fucosidase, β-galactosidase, β-hexosaminidase, and α-mannosidase) is shown in Fig. 3. The digested glycans separated by HILIC have a logical movement in GU value, whereby each oligosaccharide residue can be accounted for by a constant value, depending on linkage. The GU shifts followed the consecutive removal of terminal sugar residues. After sequential digestion with exoglycosidases, we were able to assign the most abundant peaks.

After sialidase treatment (ABS digestion), biantennary structures with bisecting GlcNAc carrying two galactoses and one sialic acid (α2,6-linkage) with GU 8.30 digest back to biantennary structures with bisecting GlcNAc carrying two galactoses with GU 7.15. The digestion with NAN1 enzyme (a recombinant sialidase that removes α2,3-linked nonreducing terminal sialic acids) confirms that sialic acid is linked via an α2,6-linkage, as the addition of this enzyme does not move the peak. The peak with GU 7.65 does not move upon addition of sialidase enzymes, but upon addition of α-fucosidase, the enzyme moves back by 0.5 GU, suggesting that is core-fucosylated. In addition treatment with β-galactosidase resulted in peak movement by 1.6 GU, corresponding to two galactose residues. The most abundant peak with GU 6.86 appeared to contain two co-eluting structures. Digestion with β-galactosidase confirmed the presence of one galactose residue at the terminal end, but some proportion (30%) of the peak remained after addition of siali-
FIGURE 2. Overall analyses of N-glycans from purified Gpc1. The N-glycans present on Gpc1-WT (A), Gpc1-N79Q (B), and Gpc1-N116Q (C) as determined using exoglycosidase digestions and analysis on the 1.7-μm HILIC phase are shown. For each sample, the most abundant N-glycans are labeled, and a percentage area of the total profile, glucose unit (GU), and retention time (RT) are shown in supplemental Table S2. Nomenclature was from Ref. 23.
dase, fucosidase, galactosidase, and mannosidase, suggesting the presence of terminal HexNAc residues. Following enzymatic treatment with β-N-acetyhexosaminidase the peak changed elution position. Unfortunately, because of the specificity of the enzyme, it is not possible to confirm whether the terminal residue represents a GlcNAc or a GalNAc residue. Based on database matching, the structure could correspond to a tetra-antennary structure but also to the structure containing LacdiNAc residue (GalNAcβ1–4GlcNAc), which seems more likely because HEK cells were previously reported to express this structural feature (49, 50).

The identification of GU 6.5 peak was primarily based on α-fucosidase digestion, as the peak did not change either its elution or migration position following enzymatic treatment, but its relative area increased due to digestion of the core-fucosylated analogue. The presence of a high mannose structure in the peak with GU 6.17 was confirmed by database matching and also with α-mannosidase digestion.
To identify glycans originating from each glycosylation site, the Gpc1 mutants of Gpc1-N79Q and Gpc1-N116Q were analyzed by HILIC-FLD-UPLC, allowing for identification of the most abundant structures from the Asn-116 and Asn-79 sites, respectively. Most of the structures overlapped between the two glycosylation sites; however, there were some trends specific for each one. Less complex glycans were more abundant at Asn-116 (Gpc1-N116Q), whereas more complex sialylated glycans were more prominent at Asn-79 (Gpc1-N116Q) (Fig. 2, B and C).

**Biophysical Characterization of Purified Gpc1—** Analytical SEC of Gpc1-dC and Gpc1-dHS displayed a main elution peak representing a Gpc1 monomer (Fig. 4A). The molecular mass of the Gpc1-dC monomer calculated from SEC (~62 kDa) was significantly smaller than that of the Gpc1-dHS monomer (~71 kDa), confirming the presence of the C-terminal domain in the dHS protein (verified by SDS-polyacrylamide gel). The homogeneity of the eluted protein fractions was assessed by DLS, without performing further concentration (Fig. 4B). The volume distribution plots of the fractions from the second half of the SEC peak of Gpc1-dC and Gpc1-dHS showed single, monodisperse peaks with a polydispersity index of 11–14% and similar patterns with $R_g = 4.1$ nm for Gpc1-dC and 4.25 nm for Gpc1-dHS. An estimation of the molecular mass assuming a globular protein suggested a mass of 89.4 ± 12.6 kDa for Gpc1-dC and 99.6 ± 27 kDa for Gpc1-dHS. The reason for such molecular mass overestimation can be the highly nonglobular shape of Gpc1 and/or the flexibility of the $N$-glycan chains decorating the core protein.

**Structural Studies Using SAXS—** To elucidate the structure of glycosylated Gpc1 in solution, SAXS data were collected for both the glycosylated monomeric Gpc1-dC and the Gpc1-dHS proteins using in-line SEC-SAXS (Fig. 4 and Table 3). No sign of protein aggregation was observed. The SAXS curves of Gpc1-dC and Gpc1-dHS were distinctly similar over the whole $q$-range collected, except in the region between 0.8 and 1.5 nm$^{-1}$, producing similar $R_g$ (~3.6 nm), but with a larger Porod volume for the full-length protein (about ~14% higher). The most accurate method for the estimation of the mass of the glycoproteins was provided by SAXSMOW calculation (51), with an expected uncertainty of 10%. In our hands, the SAXSMOW estimated masses of 67.0 and 74.9 kDa for Gpc1-dC and Gpc1-dHS, respectively, are consistent with monomeric proteins. Thus, the solution scattering data meet the essential requirements for extracting accurate monomer shape information.

To estimate the distribution of masses within the particle and its shape, the pair-distance distribution $P(r)$ function was calculated from the scattering data using GNOM (52). The $P(r)$ profiles of Gpc1-dC and Gpc1-dHS showed main peaks around 2.75 and 2.87 nm, respectively, with inclined distributions of vector lengths (Fig. 4C). This indicates that both proteins have an extended structure in solution, with maximum dimensions of 118.0 and 119.5 Å, respectively, consistent with the crystal structures. However, the $P(r)$ plot of Gpc1-dHS exhibited an asymmetric fall-off accompanied by an extra shoulder around 5.8 nm that may be correlated to the C-terminal extension. The characteristic shape factor $\sigma$ (the $R_g/R_n$ ratio) for a globular protein is ~0.774; however, when molecules deviate from globular shape to ellipsoidal, $\sigma$ increases, as the $R_n$ becomes larger than $R_g$ (53). Gpc1-dC has $\sigma = 0.9$, whereas Gpc1-dHS has $\sigma = 0.87$, which reveals a more elongated structure for the truncated version of Gpc1 than for the full-length one. This is consistent with the idea that the C-terminal region may extend perpendicular to the protein surface (see below). To gain insight into Gpc1 flexibility, Porod Debye plots were calculated and displayed a loss of the plateau in the full-length protein compared with the truncated version (Fig. 4D). This implies that the scattering contrast had become more diffuse, suggesting increased flexibility in the presence of the C-terminal domain.

The DAMMIN ab initio shape reconstruction program (28) was employed to generate 20 models of Gpc1-dC and Gpc1-dHS (Fig. 4C, insets, and Table 3) with good structural convergence, as reflected in the low normalized spatial discrepancy values following structural alignment (normalized spatial discrepancy 0.71 ± 0.02 and 0.64 ± 0.02 for Gpc1-dHS and Gpc1-dC respectively; see Table 3). The resulting models revealed sizes and shapes consistent with the core protein substructures, in good agreement with the SAXS patterns (Fig. 4E). Despite the fact that the N and C termini and two $N$-glycans should contribute to the SAXS patterns, they were not resolved as features in the averaged low resolution envelopes. These would presumably provide extra protrusions to the envelopes close to their attachment sites in the individual bead models, which are averaged out during the process of averaging and filtering to generate the final conserved model. The Gpc1-dHS ab initio model showed an additional small conserved bulge that protruded ~10 Å from the middle of the protein, proximal to the last C-terminal residue visible in the crystal structure (Fig. 4C). To test the dependence of the ab initio envelopes on the program used to determine them, a different algorithm,
namely GASBOR, was also tested. Both GASBOR and DAMMIN produced truncated and full-length models with similar overall shapes (data not shown).

Structural Reconstruction of the N-Glycans Decorating the Gpc1 Core Protein—To explore the N-linked glycan assembly on the Gpc1 core protein, we employed the Gpc1 crystal structure determined in this study to reconstruct the N-glycan structure using the SAXS data from GPC1-dC (i.e. lacking the C-terminal extension). First, the structures of missing loops in Gpc1 were built and optimized using the ModLoop protocols (31). Subsequently, the predominant N-linked glycan structures and the N terminus were reconstructed using all atom modeling (AllosMod, as described in the “Methods” section). Modeling with the most abundant glycan structures has been demonstrated to give good agreement with SAXS data (54). The predominant N-glycan structures vary between different batches of purified protein, so we assigned the structures of the predominant glycoforms from the same protein batch that was used for SAXS data collection, and these structures were used for further modeling. Monosialylated digalactosylated bi-antennary complex glycans (at position Asn-79) and core-fucosylated monogalactosylated bi-antennary complex glycans with bisecting GlcNAc (at position Asn-116) were the predominant glycoforms derived from the chromatographic data of purified Gpc1 without heparan sulfate (data not shown).

The simulated Gpc1-dC model without N-glycans gave a poor agreement with the observed SAXS data, even when the N terminus was reconstituted by modeling (Fig. 5A). In contrast, good agreement with the SAXS data were achieved with the glycosylated models; thus, the sampling of glycan conformations (accounting for ∼9% of the total scattering mass) was crucial for generating accurate models. A comparison of the 10 best and poorest fitting AllosMod models of Gpc1-dC indicated an obvious difference in the glycan orientation (Fig. 5, A and B). In the 10 best models ($\chi^2 = 1.0 \pm 0.05$), the two glycan chains protrude outward from helix α2, and the N terminus spreads out close to the protein surface, whereas in the 10 worst fitting models ($\chi^2 = 3.99 \pm 0.08$), the glycan chains were localized close to the protein surface and the N terminus protruded away from the core.

Fifty CORAL runs (26) were carried out to check the Gpc1-dC model reproducibility using a different algorithm than AllosMod. The proposed CORAL models were consistent throughout different modeling runs. The five CORAL models having the best agreement with the SAXS data ($\chi^2 = 0.9 \pm 0.002$) had the glycan chains consistently pointing outward in similar orientations to the best models from AllosMod (data not shown). Nonetheless, AllosMod has the advantage of an all-atom modeling approach to obtain stereochemically sound models.
Because of the inherent flexibility of the N-glycan chains, MES was attempted to improve the agreement with the SAXS data by including up to four models from each ensemble (33). MES provided a modest improvement in the consistency with the experimental SAXS data over the single model ($\chi^2 = 0.87$ versus 0.94 for the best single model) (Fig. 5, A and C). Both of the N-glycans were modeled in proximity to high B factor regions of the core protein crystal structure (Fig. 1B) and oriented alongside helix $\alpha_2$ with higher locational variability for Asn-79 glycans than for those on Asn-116. The Asn-79 glycan chains diverged widely over an area, including the N-terminal helix $\alpha_1$, the hydrophobic part of L1 (GFSLSDVPQA), and the beginning of the L3 loop, suggesting that they are not involved in specific interactions with the core protein. In contrast, the Asn-116 glycan chains were oriented in the vicinity of the end of L2 with smaller divergence (Fig. 5C).

Spatial Occupancy of the Gpc1 C Terminus in Solution—To elucidate the structural orientation of the C-terminal residues and to allow an estimation of the distance between Gpc1 and the cell membrane, similar modeling strategies were used to build models that agree with the Gpc1-dHS SAXS data. The best N-glycosylated truncated Gpc1 structure lacking the C-terminal residues had poor agreement with the Gpc1-dHS SAXS data ($\chi^2 = 1.73$, Fig. 5D). Therefore, generating Gpc1-dHS SAXS-consistent models require the capturing of the correct N and C termini and glycan orientations. The C terminus is not folded around the core protein in the best fitting models ($\chi^2 = 0.960 \pm 0.007$; Fig. 5E) but is rather extended toward the periphery, stretching for $\sim$40 Å from the last helix in the crystal structure ($\alpha_{14}$). In the poorest fitting models ($\chi^2 = 4.6 \pm 0.1$), the C terminus was even more highly extended ($>70$ Å). Furthermore, the modeled positions of the N terminus and N-glycan chains in the best and worst models generated by AllosMod were highly similar in both Gpc1-dC and Gpc1-dHS, which strengthens our confidence in the interpretation of these structural features. The CORAL approach reproduced the characteristic extension of the C terminus obtained by the AllosMod method with comparable agreement with the experimental SAXS data ($\chi^2 = 0.92 \pm 0.008$; data not shown).

Because of the high number of variables in Gpc1-dHS modeling (N and C termini and two glycan chains), the probability of capturing a single conformation that most correctly explains the SAXS pattern is quite limited. MES was successful in improving the fit over a single model ($\chi^2$ of 0.9 and 0.95, respectively). The best fitting C-terminal conformations form a compact cluster, spreading between 35 and 40 Å in a direction perpendicular to the core protein and mainly localized between the central and the protease site lobes (Fig. 5F). Furthermore, EOM modeling (34) was used to check the consistency of the MES modeling processes. The EOM ensemble of models of Gpc1-dHS showed the C terminus in the same orientation as in the MES results but with a shorter extension (between 30 and 35 Å; data not shown). The Gpc1 membrane-proximal surface, defined by this orientation of the C terminus, is composed of the conserved $\alpha_{14}$, L1, and L3 in the Cys-rich lobe and $\alpha_4$ and $\alpha_5$ in the protease-site lobe (Fig. 10). The Asn-79 glycan chains are directed perpendicular to the C terminus, whereas the Asn-116 glycans are located on the opposite surface of the protein.

FIGURE 6. Comparative studies of Gpc1-dC and Gpc1-dHS at different protein concentrations. A, analytical SEC profiles showing the concentration effect on the elution volume. One hundred $\mu$l of 0.2 mg/ml (light blue), 0.5 mg/ml (blue), 1.4 mg/ml (orange), 3.4 mg/ml (green), and 6 mg/ml (red) solutions of Gpc1-dC and 1 mg/ml (violet) and 6 mg/ml (brown) Gpc1-dHS solution were loaded onto the Superdex 200 column. SAXS curves of Gpc1-dC (B) and Gpc1-dHS (C) at protein concentrations of 0.75 mg/ml (orange), 3 mg/ml (green), and 6 mg/ml (red) in a comparison with the SAXS pattern of the equivalent monomer (blue) collected from SEC-SAXS setup. The related P(r) plots are shown in insets B and C.

Does Gpc1 Self-associate in Solution?—In the SEC experiments, when we increased the loading concentration of Gpc1-dC to $\geq$1.5 mg/ml, the position of the main peak moved toward a higher mass, and a small shoulder was eluted at 12.4 ml (Fig. 6A), suggesting a protein dimer. This shift in the eluted peaks might originate from various monomer/dimer mixtures in the solution. SAXS measurements of Gpc1-dC at different concentrations (0.75–6 mg/ml) showed apparent changes in the low-{$q$} region indicative of concentration-dependent self-
association or oligomerization (Fig. 6B and Table 4). A small change in the SAXS curves (particularly at 6 mg/ml) beyond the Guinier region was observed when compared with Gpc1-dC SEC-SAXS data, which support the emergence of dimer species at high concentrations. Furthermore, the measured SEC peak shift or change in the SAXS curves (particularly at 6 mg/ml) beyond the Guinier region was observed when compared with Gpc1-dC at 6 mg/ml suggested an elongated shape for the dimer in which the N terminus was pre-dimerized in the vicinity of the protease site lobe, particularly dominantly extended away from the protein (Fig. 7). Gpc1-dC forms a highly extended dimer with a shape factor of 1.2. Moreover, the processing parameters are for the highest protein concentration sample.

The DLS data show that at 6 mg/ml concentration, Gpc1-dC was directly proportional to the protein concentration range 0.75–6 mg/ml and suggests that the presence of the C-terminal tail prevents unwanted interactions of the core proteins. N-Glycosylation Protects Aggregation- and Degradation-prone Regions of Gpc1—EndoH enzymatic removal of the N-glycans from Gpc1-dHS was achieved as described under “Experimental Procedures,” producing the deglycosylated Gpc1-dHS-dN (Fig. 6C). Taken together, this indicates that self-association does not occur for Gpc1-dHS in the tested concentration range of 0.75 to 6 mg/ml and suggests that the presence of the C-terminal tail prevents unwanted interactions of the core proteins.

N-Glycosylation Protects Aggregation- and Degradation-prone Regions of Gpc1—EndoH enzymatic removal of the N-glycans from Gpc1-dHS was achieved as described under “Experimental Procedures,” producing the deglycosylated Gpc1-dHS-dN with a mass ~5 kDa lower than that of the glycosylated protein, as shown on an SDS-polyacrylamide gel (Fig. 8A). Importantly, SAXS measurements of Gpc1-dHS-dN showed little concentration-dependent behavior, displaying invariability in the q-range beyond the Guinier region, which speaks against variation in the oligomeric structure in the concentration range 0.7–2.5 mg/ml. However, models generated by DAMMIN and CORAL using the merged SAXS data of dHS-dN revealed an elongated dimeric structure rather than a monomer (Fig. 8, B and C). Gpc1-dHS-dN dimerized through parts of the Cys-rich lobe that are covered by the Asn-79 gly-
cans in the glycosylated Gpc1. Removing the N-glycans from Gpc1-dC resulted in complete aggregation during purification. Furthermore, the mutant N79Q showed a high propensity for aggregation after removing the HS chains (Gpc1-N79Q-dHS; data not shown). Taken together, these data suggest a vital role for Asn-79 glycans in preventing protein aggregation, at least in vitro. The solubility difference between Gpc1-N79Q-dHS and deglycosylated Gpc1-dHS might originate from the first GlcNAc residue linked to Asn-79, which is not removed by EndoH. This GlcNAc confers most of the thermal stabilization effect of the N-glycans on Gpc1, as reported previously (13), and it seems likely that it also confers solubility enhancement.

In addition to aggregation, the Gpc1-N79Q samples showed degradation on SDS-polyacrylamide gels, with a strong extra band at ~47 kDa (Fig. 9A), which indicates that the lack of the Asn-79-glycosylation has rendered N79Q susceptible to intracellular and/or extracellular proteolytic degradation. This proteolysis site was identified by a mass spectrometric analysis and shown to be close to the beginning of the L3 loop (between Lys-404–Arg-414). This evidence reveals a role for Asn-79 glycans in protecting against degradation of Gpc1.

We used highly monodisperse preparations of purified Gpc1-WT-dHS and Gpc1-N116Q-dHS (after removal of the HS by HS lyase) for SAXS data collection at different protein concentrations (up to 3 mg/ml) (Fig. 9, A and B). The HS lyase enzyme degraded the HS chains but did not remove the tetrasaccharide linker attached to the serine. The Gpc1-WT-dHS and Gpc1-N116Q-dHS scattering profiles were smoother than for the proteins without the HS linkers and had almost identical concave patterns accompanied by negligible differences in \( R_g \) and \( D_{\text{max}} \) (Table 4), without concentration-dependent self-association or aggregation at the tested concentrations. Their molecular masses and Porod volumes were highly overestimated, presumably because of the flexibility of the partially glycosylated C-terminal tails (bound to three tetrasaccharide linkers) and/or the presence of some oligomers. The pair distribution functions of Gpc1-WT-dHS and Gpc1-N116Q-dHS suggested similar shapes, with Gpc1-WT-dHS being slightly larger due to its extra Asn-116-linked glycan chain (Fig. 9C). The generated ab initio envelopes featured similar overall expanded shapes, with a larger volume for the WT samples. Attempts to model Gpc1-WT-dHS and Gpc1-N116Q-dHS using AllosMod and CORAL failed due to the complexity at the
C terminus, with three flexible tetrasaccharide chains in a very confined region; thus the programs were unable to model them correctly. We concluded from comparison of Gpc1-WT-dHS and Gpc1-N116Q-dHS that no structural changes in Gpc1 (at the resolution level of SAXS) could be noticed after disrupting the Asn-116 site, and therefore that this glycan might have other functional roles.

Discussion

Glypicans are multifunctional GAG-substituted proteoglycans involved in the regulation of several cellular signaling pathways. Defects in their function lead to developmental distortions (2, 55). The regulatory activity of Gpcs is based on their ability to either inhibit or stimulate the interaction of many growth factors with their signaling receptors. GAG chains are responsible for many of the biological functions of Gpcs. However, recent studies suggest regulatory roles for Gpc core proteins, for example in mediating cell signaling by direct binding to e.g. BMP4, FGF2, Wnt, and Hedgehog (56–58).

The crystal structure of the Gpc1 core protein reveals a quite rigid, elongated single-domain α-helical fold, although flexibility is higher toward the ends. Almost identical disulfide connectivity patterns and structural similarity between the Cys-rich lobe of Gpc1 and the Cys-rich domain of the Frizzled receptors (functioning in Wnt signaling) have been reported (59), but whether this has functional relevance is still unknown.

Structural Features of the Gpc1 N-Glycans—The long α2 helix of the Gpc1 structure traverses the entire length of the protein and carries two N-glycan chains, one at each end (10). The Gpc1 N-glycan chains are not fully resolved in the crystal structure, due to intrinsic heterogeneity and flexibility. Glycoproteins are usually micro-heterogeneous, presenting a range of different N-glycans at each glycosylation site. The expression system and/or the folded protein itself can influence the N-glycan diversity, probably by affecting the substrate availability and the proximity of the N-glycosylation sites to the Golgi glycosyltransferases and glycosidases (60).

The N-glycan analyses presented here have revealed highly heterogeneous complex-type glycoforms for Gpc1, but it is not clear whether this heterogeneity has a functional impact. Characteristics of the Asn-79-linked glycans include a high incidence of core fucosylation and sialylation, presenting relatively strong electronegative charges, which we discovered, based on the SAXS modeling, cover the hydrophobic patches on the Cys-rich lobe, including parts of the L1 loop and the α1
and α2 helices, that would otherwise promote aggregation. Furthermore, these Asn-79 oligosaccharides seem to limit intra- and/or extracellular proteolysis of the Gpc1 core protein, specifically at the beginning of the L3 loop, extending the protein stability and lifetime, whereas removal of Asn-116 glycans does not affect the overall structural characteristics of Gpc1.

Topology of the Gpc1 Core Protein, N- and O-Glycans with Respect to the Cell Surface—The Gpc1 C-terminal region lacks significant secondary or tertiary structure and is disordered in Gpc1-dHS crystals, even after controlled crystal dehydration, which otherwise significantly improves crystalline order, resolution, and diffraction anisotropy. In view of this, we coupled our crystallographic knowledge on Gpc1 with SAXS to obtain information on the localization of the C terminus and thus the potential spatial orientations of Gpc1 relative to the cell surface. Our results show that the C-terminal domain is highly flexible and extends ~35–40 Å from the core protein. Molecular dynamics simulations have shown that the GPI anchor of many glypiated proteins is highly flexible but that it nevertheless can retain the proteins at distances between 9 and 13 Å from the cell surface, with minor impact on the proteins’ degrees of freedom for movement and orientation relative to the membrane (61, 62). Accordingly, by adding the length of the GPI anchor to that of the C-terminal tail, we suggest that the Gpc1 core protein is located approximately ~44–53 Å from the cell membrane. This distance is likely sufficient for HS assembly enzymes or even a membrane receptor candidate to interact with the membrane-proximal surface of Gpc1. It seems likely that Gpc1 “lies down” in a transverse orientation to the membrane, with the Gpc1 orthologs’ evolutionarily conserved surface (L1 and L3 loops, α4, α5, and α14 helices) facing the cell surface. No evolutionary conservation has been detected for the other surfaces of Gpc1 (Fig. 10). Evolutionary conservation of these surface-exposed residues suggests that they are implicated in interaction with other macromolecules related to Gpc1 function. In the Drosophila glypican Dally-like core protein (lacking the HS chains), structure-guided mutational evidence suggested that helices α4 and α5 are important for mediating Hedgehog signaling (63), consistent with these being oriented toward the membrane. The flexibility of the C-terminal region presumably allows great freedom for Gpc1 and other glypicans to reorient to accommodate binding to receptors and other signaling molecules, generally with the participation of the HS chains.

Bioinformatics tools predict a highly disordered structure for the anionic linker (10 residues) between the folded core protein and the GAG attachment site (at least 10–15 Å). Consequently, we can hypothesize that the HS attachment sites are located more than 30 Å from the membrane, and therefore they could mediate the interaction of glypicans with other cell surface proteins. Previous work has reported that the HS chains do not stabilize, and probably do not interact with, the Gpc1 core protein (38), which is consistent with the extended C-terminal structure that we observe. However, the HS chains on Gpc1 protect the protein against irreversible aggregation, probably by providing the protein additional negative charge, resulting in electrostatic repulsion.

Hence, we predict that the anionic HS chains extend from the C-terminal region in a plane initially approximately parallel to the membrane, then more divergent from it, to avoid
repulsive interactions with the negatively charged phospholipid bilayer (Fig. 10). Glypicans like Gpc5 possess both HS and CS (66). Chen et al. (67) concluded that the core protein plays a vital role in directing the assembly of HS rather than CS on rat Gpc1, as expression of the GAG attachment domain without the core protein results in substitution with ~90% CS. Moreover, mutational analysis shows that the L3 loop, α14, and other nearby parts of the core protein are required for preferential HS assembly (67). Previous work and the present data both confirm that the N-glycosylations of Gpc1 are not involved in GAG class determination but do affect the amount of HS substitution and also chain elongation (13). Accordingly, parts or all of the surface-conserved elements (L1, L3, α4, α5, and α14) would be predicted to be involved in GAG class determination and synthesis by interacting with some components of the GAG biosynthetic pathway, e.g. glycosyltransferases, necessary cofactors for the enzyme activity or some components required for trafficking the protein within the Golgi apparatus to regulate the HS substitution on Gpc1. HS biosynthesis is mediated by Golgi apparatus transmembrane glycosyltransferases of the exostosin (EXT) family, which initiate, elongate, and terminate HS backbone formation (68). Five members have been identified in mammals, including EXT1, EXT2, EXT1L, EXT2L, and EXT3L. In general, the published results so far suggest that EXT3L works as an initiator of HS chain biosynthesis, as no HS was detected in 9-day-old mouse embryos lacking EXT3L (69). Future work would be of importance to determine precisely how the Gpc1 core protein regulates HS class determination and assembly. This could be by searching for the Gpc1 interacting partner from the exostosis family and in particular by studying its affinity and interaction with EXT3L as an HS initiator enzyme. Furthermore, systematic mutagenesis studies of the evolutionarily conserved surface structural elements would be helpful to map more distinctly the residues of the Gpc1 core protein that might be involved in preferential HS assembly.

**Solution Structure of Human Glypican-1 Core Protein**

![Diagram](image-url)

**FIGURE 10. Predicted topology of Gpc1 on the membrane.** A, a schematic overview of the Gpc1 structure represented in rainbow colors (blue, N terminus; red, C terminus) with N-glycans as red spheres. The model is aligned with a transparent surface showing the sequence conservation of surface residues, colored as in the inlay figures. The inset smaller figures show the structural conservation of surface-exposed residues in Gpc1 orthologs (Gpc1 human, P35052; Gpc1 mouse, Q9QZF2; Gpc1 rat, P35052; Gpc1 bovine; Q2KJ65; Gpc1 chick, P50593; Gpc1 zebrafish, Q1LXM6; Turkey, G1MQ88; Gpc1 chimpanzee, K7BM18) as generated using ConSurf server (70) and presented by PyMOL as spheres colored from turquoise (variable) to purple (conserved). B, supposed spatial orientation of Gpc1 on the cell membrane. The membrane is shown as a gray lipid bilayer with orange GPI anchor connected to the Gpc1 C terminus and carried three chains of HS (blue).
Author Contributions—W. A. purified all glypic-an constructs, crys-
stallized them, performed crystal dehydration, collected data and
refined the crystal structure, conceived, and carried out SAXS exper-
iments. B. A., J. Ö., and N. G. K. carried out the analysis of the glyco-
sylation patterns on Gpc1. K. M. expressed the glypic-an constructs.
D. T. L. and K. M. conceived and supervised the project. All authors
contributed to writing the paper.

Acknowledgments—We thank the staff at beamline 1911-3 of the MAX
IV Laboratory for help with data collection, in particular Mar-
jolein Thunnissen for assistance with the HClb machine. We thank
staff at the ESRF SAXS beamlines (Adam Round and Martha Bren-
nich) for help with SAXS data collection. We also appreciate the assis-
Received from http://www.bioge.org/Downloaded on December 4, 2019

References
1. Brown, D., and Waneck, G. L. (1992) Glycosyl-phosphatidylinositol-anchored
membrane proteins. J. Am. Soc. Nephrol. 3, 895–906
2. Fico, A., Maina, F., and Dono, R. (2011) Fine-tuning of cell signaling by
glypicans. Cell. Mol. Life Sci. 68, 923–929
3. Whitelock, J. M., and Iozzo, R. V. (2005) Heparan sulfate: a complex poly-
cation charged with biological activity.
4. Awad, W., Logan, D. T., and Mani, K. (2014) GPC1 (glypican 1).
5. Jen, Y. H., Musacchio, M., and Lander, A. D. (2009) Glypican-1 controls
6. Hooper, N. M. (2011) Glypican-1 facilitates prion conversion in lipid rafts.
7. Watanabe, N., Araki, W., Chui, D. H., Makifuchi, T., Ihara, Y., and Tabira,
8. Matsuda, K., Maruyama, H., Guo, F., Keffe, J., Itakura, J., Matsumoto, Y.,
9. Su, G., Meyer, K., Nandini, C. D., Qiao, D., Salamat, S., and Friedl, A.
10. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and
development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501
11. Bowler, M. W., Mueller, U., Weiss, M. S., Sanchez-Weatherby, J., So-
nichols, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5
for the refinement of macromolecular crystal structures. Acta Crystallogr.
12. Royle, L., Radcliffe, C. M., Dwek, R. A., and Rudd, P. M. (2006) Detailed
structural analysis of N-glycans released from glycoproteins in SDS-poly-
crylamide gel bands using HPLC combined with exoglycosidase digest
methods. Methods Mol. Biol. 347, 125–143
13. Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. A., and Rudd, P. M.
(2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis.
Bioinformatics 24, 1214–1216
14. Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Marth,
15. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty,
16. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans,
17. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A.,
18. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and
development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501
19. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty,
20. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino,
21. Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. A., and Rudd, P. M.
(2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis.
Bioinformatics 24, 1214–1216
22. Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. A., and Rudd, P. M.
(2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis.
Bioinformatics 24, 1214–1216
23. Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Marth,
24. Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon,
25. Su, G., Meyer, K., Nandini, C. D., Qiao, D., Salamat, S., and Friedl, A.
(2006) Glypican-1 is frequently overexpressed in human breast cancer
and modulates the mitogenic effects of multiple heparin
26. Su, G., Meyer, K., Nandini, C. D., Qiao, D., Salamat, S., and Friedl, A.
(2006) Glypican-1 is frequently overexpressed in human gliomas and en-
27. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsi-
cally unstructured biological macromolecules by SAS using the Porod-
28. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsi-
cally unstructured biological macromolecules by SAS using the Porod-
29. Incardona, M. F., Bourenkov, G. P., Revik, K., Pierzitz, R. A., Popov, A. N.,
30. Perrot, P., Round, A., Vian, G., Antibolinos, A., Gobbo, A., Gordon, E.,
31. Svergun, D. I. (1999) Restoring low resolution structure of biological mac-
32. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) FoXS: a web
33. Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility
34. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) FoXS: a web
35. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and
36. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans,
37. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A.,
38. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A.,
39. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A.,
40. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A.,
41. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A.,
evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. 28, 768–773
37. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. 1, 2856–2860
38. Svensson, G., Linse, S., and Mani, K. (2009) Chemical and thermal unfolding of glypican-1: protective effect of heparan sulfate against heat-induced irreversible aggregation. Biochemistry 48, 9994–10004
39. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006) Toward the structural genomics of complexes: Crystal structure of a PE/PP protein complex from Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U.S.A. 103, 8060–8065
40. Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21, 3433–3434
41. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim. Biophys. Acta 1804, 996–1010
42. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003) Protein disorder prediction: implications for structural proteomics. Structure 11, 1453–1459
43. Hirose, S., Shimizu, K., Kanai, S., Kuroda, Y., and Noguchi, T. (2007) POODLE-L: a two-level SVM prediction system for reliably predicting long disordered regions. Bioinformatics 23, 2046–2053
44. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PISpred Protein Analysis Workbench. Nucleic Acids Res. 41, W349–W357
45. Cole, C., Barber, J. D., and Barton, G. J. (2008) The Ipred 3 secondary structure prediction server. Nucleic Acids Res. 36, W197–W201
46. Ouali, M., and King, R. D. (2000) Cascaded multiple classifiers for secondary structure prediction. Protein Sci. 9, 1162–1176
47. Linding, R., Russell, R. B., Neduva, V., and Gibson, T. J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 31, 3701–3708
48. Cheng, J., Randall, A. Z., Sweredoski, M. J., and Baldi, P. (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res. 33, W72–W76
49. Hamouda, H., Kaup, M., Ullah, M., Berger, M., Sandig, V., Tauber, R., and Blanchard, V. (2014) Rapid alignment of cell surface N-glycosylation from living cells using mass spectrometry. J. Proteome Res. 13, 6144–6151
50. Kim Hie, Saldova, R., Park, J. H., Harvey, D. J., Wormald, M. R., Wynne, K., Elia, G., Kim, H.-J., Rudd, P. M., and Lee, S.-T. (2013) The presence of outer arm fucose residues on the N-glycans of tissue inhibitor of metalloproteinases-1 reduces its activity. J. Proteome Res. 12, 3547–3560
51. Fischer, H., de Oliveira Neto, M., Napolitano, H. B., Polikarpov, I., and Craievich, A. F. (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J. Appl. Crystallogr. 43, 101–109
52. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. 25, 495–503
53. Brewer, A. K., and Striegel, A. M. (2011) Characterizing the size, shape, and compactness of a polydisperse prolate ellipsoidal particle via quadrupole-detector hydrodynamic chromatography. Analyt. Chem. 136, 515–519
54. Gutman, M., Weinкам, P., Salì, A., and Lee, K. K. (2013) All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosylated proteins. Structure 21, 321–331
55. Filmus, J., Capurro, M., and Rast, J. (2008) Glypicans. Genome Biol. 9, 224
56. Kirkpatrick, C. A., Knox, S. M., Staatz, W. D., Fox, B., Lercher, D. M., and Selleck, S. B. (2006) The function of a Drosophila glypican does not depend entirely on heparan sulfate modification. Dev. Biol. 300, 570–582
57. Cheng, W., Tseng, C.-J., Lin, T. T., Cheng, I., Pan, H.-W., Hsu, H.-C., and Lee, Y.-M. (2008) Glypican-3-mediated onco genesis involves the insulin-like growth factor-signaling pathway. Carcinogenesis 29, 1319–1326
58. Williams, E. H., Pappano, W. N., Saunders, A. M., Kim, M.-S., Leahy, D. J., and Beachy, P. A. (2010) Dally-like core protein and its mammalian homologues mediate stimulatory and inhibitory effects on Hedgehog signal response. Proc. Natl. Acad. Sci. U.S.A. 107, 5869–5874
59. Pei, J., and Grishin, N. V. (2012) Cysteine-rich domains related to Frizzled receptors and Hedgehog-interacting proteins. Proteins 21, 1172–1184
60. Varki, A., Freeze, H. H., and Gagneux, P. (2009) in Essentials of Glycobiology (Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E., eds) 2nd Ed., Chap. 19, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
61. Zuegg, J., and Gready, I. E. (2000) Molecular dynamics simulation of human prion protein including both N-linked oligosaccharides and the GPI anchor. Glycobiology 10, 959–974
62. Rudd, P. M., Morgan, B. P., Wormald, M. R., Harvey, D. J., van den Berg, C. W., Davis, S. J., Ferguson, M. A., and Dwek, R. A. (1997) The glycosylation of the complement regulatory protein, human erythrocyte CD59. J. Biol. Chem. 272, 7229–7244
63. Kim, M.-S., Saunders, A. M., Hamaoka, B. Y., Beachy, P. A., and Leahy, D. J. (2011) Structure of the protein core of the glypican Dally-like and localization of a region important for hedgehog signaling. Proc. Natl. Acad. Sci. U.S.A. 108, 13112–13117
64. Khan, S., Fung, K. W., Rodriguez, E., Patel, R., Gor, J., Mullboy, B., and Perkins, S. J. (2013) The solution structure of heparan sulfate differs from that of heparin: implications for function. J. Biol. Chem. 288, 27737–27751
65. Pang, S., Urquhart, P., and Hooper, N. M. (2004) N-Glycans, not the GPI anchor, mediate the apical targeting of a naturally glycosylated, GPI-anchored protein in polarised epithelial cells. J. Cell Sci. 117, 5079–5086
66. Li, F., Shi, W., Capurro, M., and Filmus, J. (2011) Glypican-5 stimulates rhabdomyosarcoma cell proliferation by activating Hedgehog signaling. J. Cell Biol. 192, 691–704
67. Chen, R. L., and Lander, A. D. (2001) Mechanisms underlying preferential assembly of heparan sulfate on glypicans-1. J. Biol. Chem. 276, 7507–7517
68. Busse-Wicher, M., Wicher, K. B., and Kusche-Gullberg, M. (2014) The extostosin family: proteins with many functions. Matrix Biol. 35, 25–33
69. Takahashi, I., Noguchi, N., Naka, K., Yamada, S., Kaneiwa, T., Mizumoto, S., Ikeda, T., Sugihara, K., Asano, M., Yoshikawa, T., Yamachi, A., Shervani, N. J., Uruno, A., Kato, I., Unno, M., et al. (2009) Important role of heparan sulfate in postnatal islet growth and insulin secretion. Biochem. Biophys. Res. Commun. 383, 113–118
70. Ashkenazi, H., Erez, E., Martz, E., Pepko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38, W529–W533