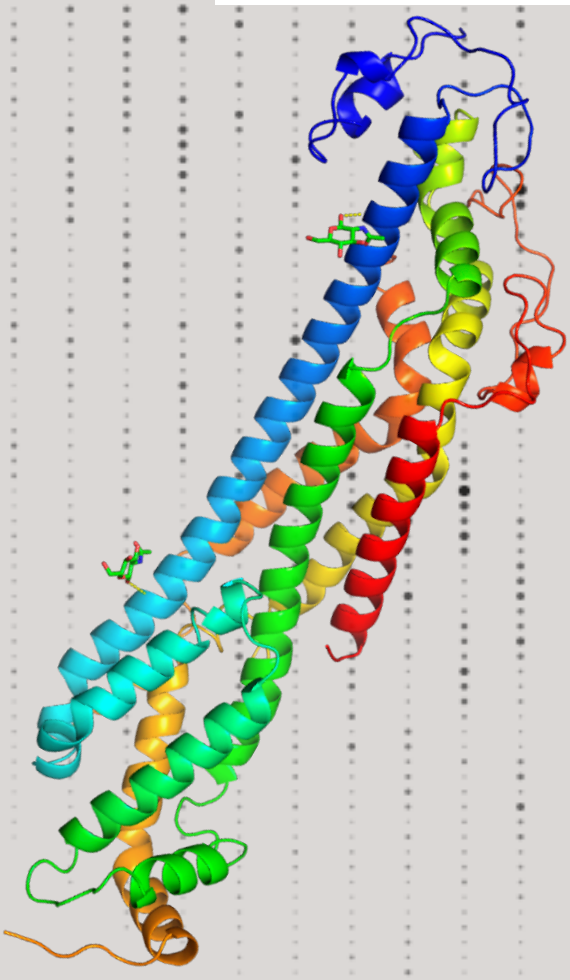
Glypican-1: Structural and functional analysis of the N-glycosylated human protein

Abdelhady, Wael Awad

2015

Glypican-1: Structural and functional analysis of the N-glycosylated human protein

WAEL AWAD MOHAMED
BIOCHEMISTRY AND STRUCTURAL BIOLOGY | LUND UNIVERSITY



Glypican-1: Structural and functional analysis of the N-glycosylated human protein

Wael Awad Mohamed

LUND
UNIVERSITY

by due permission of the Faculty Science, Lund University, Sweden.

To be publicly defended on Thursday May 28th 2015 at 9:30 a.m. in Kemicentrum, Lecture Hall C

Faculty opponent

Prof. Dr. Savvas N. Savvides
Department of Biochemistry and Microbiology,
Ghent University, Ghent, Belgium.

Organization	Document name
LUND UNIVERSITY Department of biochemistry and Structural Biology Center for Molecular Protein Science	DOCTORAL DISSERATION
	Date of issue 28-5-2015
Author(s) Wael Awad Mohamed	Sponsoring organization

Title and subtitle

Glypican-1: Structural and functional analysis of the N-glycosylated human protein

Abstract

Glypicans are multifunctional cell surface heparan sulphate proteoglycans co-regulating numerous signalling pathways, and are thereby involved in the control of cellular division, differentiation, and morphogenesis. The heparan sulphate (HS) chains are responsible for many of those biological functions; nevertheless recent studies suggest functional roles for the glypican core proteins in mediating the signalling of various growth factors. Glypican-1 (GPC1) is the predominant HS proteoglycan in the developing and adult human brain. In addition, GPC1 is involved in Alzheimer's disease and scrapie, among others. There is a shortage of detailed structural knowledge regarding the GPC1 core protein and accordingly, we proposed in this thesis to structurally and functionally characterize the human GPC1 core protein and to elucidate its overall topology with respect to the membrane.

First, we determined the crystal structure of the human N-glycosylated GPC1 core protein by the two-wavelength MAD method on a SeMet-substituted protein crystal. The GPC1 structure revealed a quite rigid, cylindrical single-domain all -helical fold with three substantial loops. Shortly afterwards, we achieved improvements of GPC1 crystal diffraction properties by controlled crystal dehydration using a humidity control device (HC1b) and generated better electron density for crystals of GPC1, allowing the building of previously disordered parts of the structure. Using small angle X-ray scattering and other biophysical approaches, we found that the GPC1 core protein lies on the membrane in a transverse orientation, directing a surface evolutionarily conserved in GPC1 orthologues towards the membrane, where it can interact with enzymes involved in HS substitution in the Golgi apparatus. Furthermore, the N-linked glycans are shown to extend the protein stability and lifetime by protection against proteolysis and aggregation.

The EXTL3 protein, a member of the exostosin family, functions mainly as an initiator for HS assembly on the glypicans. We have investigated the spectroscopic and structural characteristics of the catalytic region of EXTL3, which exhibits a quite stable extended monomeric structure with two functional domains containing a majority of β sheets. Additionally, it was found that catalytic EXTL3 is occupied with N-glycans at least at two sites and these N-glycans seem critical for proper EXTL3 biosynthesis. To precisely determine how the GPC1 core protein regulates HS assembly through interactions with EXTL3, investigations of the GPC1-EXTL3 complexes are ongoing, and some preliminary results are presented here.

Key words: Proteoglycans; glypicans; glypican-1; heparan sulphate; N-glycosylation; exostosin-proteins; X-ray crystallography; crystal dehydration; diffraction anisotropy; SAXS, protein spectroscopy

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language: English
ISSN and key title	ISBN: 978-91-7422-399-6
Recipient's notes	Number of pages 149 / Price
	Security classification

Glypican-1: Structural and functional analysis of the N-glycosylated human protein

Doctoral Dissertation

Wael Awad Mohamed

LUND
UNIVERSITY

Department of Biochemistry and Structural Biology
2015

Supervisors: Derek Logan
Katrin Mani

Faculty of Science
Department of Biochemistry and Structural Biology
Center of Molecular Protein Science, Lund, Sweden

ISBN 978-91-7422-399-6

Printed in Sweden by Media-Tryck, Lund University
Lund 2015

Dedicated to

My parents, brother and sister
My wife, Sanaa
My lovely kids: NourEldin, Leen & Jana

Table of contents

Abstract

Glypicans are multifunctional cell surface heparan sulphate proteoglycans co-regulating numerous signalling pathways, and are thereby involved in the control of cellular division, differentiation, and morphogenesis. The heparan sulphate (HS) chains are responsible for many of those biological functions; nevertheless recent studies suggest functional roles for the glypican core proteins in mediating the signalling of various growth factors. Glypican-1 (GPC1) is the predominant HS proteoglycan in the developing and adult human brain. In addition, GPC1 is involved in Alzheimer's disease and scrapie, among others. There is a shortage of detailed structural knowledge regarding the GPC1 core protein and accordingly, we proposed in this thesis to structurally and functionally characterize the human GPC1 core protein and to elucidate its overall topology with respect to the membrane.

First, we determined the crystal structure of the human N-glycosylated GPC1 core protein by the two-wavelength MAD method on a SeMet-substituted protein crystal. The GPC1 structure revealed a quite rigid, cylindrical single-domain all α-helical fold with three substantial loops. Shortly afterwards, we achieved improvements of GPC1 crystal diffraction properties by controlled crystal dehydration using a humidity control device (HC1b) and generated better electron density for crystals of GPC1, allowing the building of previously disordered parts of the structure. Using small angle X-ray scattering and other biophysical approaches, we found that the GPC1 core protein lies on the membrane in a transverse orientation, directing a surface evolutionarily conserved in GPC1 orthologues towards the membrane, where it can interact with enzymes involved in HS substitution in the Golgi apparatus. Furthermore, the N-linked glycans are shown to extend the protein stability and lifetime by protection against proteolysis and aggregation.

The EXTL3 protein, a member of the exostosin family, functions mainly as an initiator for HS assembly on the glypicans. We have investigated the spectroscopic and structural characteristics of the catalytic region of EXTL3, which exhibits a quite stable extended monomeric structure with two functional domains containing a majority of

β sheets. Additionally, it was found that catalytic EXTL3 is occupied with N-glycans at least at two sites and these N-glycans seem critical for proper EXTL3 biosynthesis. To precisely determine how the GPC1 core protein regulates HS assembly through interactions with EXTL3, investigations of the GPC1-EXTL3 complexes are ongoing, and some preliminary results are presented here.

List of papers

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

Svensson G, Awad W, Håkansson M, Mani K & Logan DT (2012) *J. Biol. Chem.* 287, 14040-14051.

II. *Improvements of the order, isotropy and electron density of glypican-1 crystals by controlled dehydration.*

Awad W, Svensson Birkedal G, Thunnissen MMGM, Mani K & Logan DT (2013) Acta Crystallogr. D Biol. Crystallogr. 69 (12), 2524-2533.

III. *Structure-function analysis of N-glycosylation and C-terminus in human glypican-1.*

Awad W, Adamczyk B, Örnros J, Karlsson NG, Mani K & Logan DT
Submitted manuscript

IV. *Expression, purification and biophysical characterization of N-glycosylated human EXTL3.*

Awad W, Svensson Birkedal G, Mani K & Logan DT.
Manuscript

My contribution to the papers

I. Within a team, I took part in crystal manipulation and diffraction data collection. I made a major contribution to solving the crystal structures and carried out all model building, refinement and validation. I contributed to data analysis and revision of the manuscript.

II. I participated in designing the study. I purified and crystallized the protein, performed crystal dehydration and data collection. I carried out all model building and refinement. I performed data analysis and drafted the manuscript.

III. I took the major role in designing the study. I purified the proteins and fulfilled the crystallography, SAXS and the other biophysical investigations except for N-glycan mass spectrometry. I performed data analysis and drafted the manuscript.

IV. I took the major role in designing the study. I purified the protein and carried out all the biochemical and biophysical experiments. I performed data analysis and drafted the manuscript.

Additional papers not included in the thesis

I. *Global motions from the strain of a single hydrogen bond.*

Danielsson J, Awad W, Saraboji K, Kurnik M, Lang L, Leinartaité L, Marklund SL, Logan DT & Oliveberg M (2013) *Proc. Natl. Acad. Sci. USA* 110, 3829–34.

II. *GPC1 (glypican 1).*

Awad W, DT Logan, K Mani (2014) Atlas Genet Cytogenet Oncol Haematol – 18(7).

Abbreviations

C. elegans	*Caenorhabditis elegans*
CD	circular dichroism
CS	chondroitin sulphate
D. melanogaster	*Drosophila melanogaster*
DLS	dynamic light scattering
DSF	differential scanning fluorimetry
ECM	extracellular matrix
EndoH	endoglycosidase
EXT	exostosin
GAG	glycosaminoglycan
GalNAc	*N*-acetylgalactosamine
Gal	galactose
GlcNAc	*N*-acetylglucosamine
GlcA	glucuronic acid
GPC	glypican
GPI	phosphatidylinositol
HC1b	humidity-controlled device
HS	heparan sulphate
HSPG	Heparan sulphate proteoglycan
MS	mass spectrometry
PG	proteoglycan
SAXS	small angle X-ray scattering
SEC	size exclusion chromatography
RH	relative humidity
T_{inc}	total incubation dehydration time

Preface

All living things somehow communicate with each other. Cell-to-cell communication, or signalling, occurs on the molecular level, regulating the body's activities and coordinating various cell actions and is thereby valuable for realizing cell as well as system functions. Errors in the cellular transferred information may induce diseases such as cancer and autoimmune diseases, among others. In fact, most diseases enclose at least one malfunction in cell communication pathways. Understanding the cellular signalling pathways and the components involved would be significant for effective treatment of those diseases. Many components of the extracellular environment, in particular the cell surface receptors, enable the cells to recall the extracellular signals and then trigger intracellular chains of biochemical events creating the response. These receptors often use heparan sulphate proteoglycans (HSPG) to promote and control ligand binding and activation, due to the interactions of HSPG core proteins and/or the heparan sulphate (HS) chains with the ligands.

Glypican (GPC) is a family of HSPG proteins that are anchored to the external leaflet of the cell membrane where they interact with several extracellular ligands and receptors and therefore act as mandatory co-receptors. GPCs are involved in the regulation of many biological processes such as cellular adhesion, division, differentiation and morphogenesis. The HS chains are responsible for many of these biological functions, but recent studies suggest functional roles for the GPC core proteins in mediating various morphogen and growth factor signalling.

Glypican-1 (GPC1) is one of the six members of the vertebrate GPC family that is mainly expressed in the neural and skeletal systems during development and ubiquitously in the adult. GPC1 is involved in the uptake of different macromolecules such as growth factors, viral proteins, polyamines and cytokines. Many reports concluded that GPC1 is important for brain development and function, and further revealed its involvement in the pathogenesis of several neurodegenerative diseases and glioma, pancreatic and breast cancers. Unfortunately,

there is a shortage in structural knowledge about the GPC core proteins. The overall objective of this thesis is to structurally characterize the GPC1 core protein and its overall topology with respect to the cell surface. This will be of great assistance to gain insights into the functional roles of GPC1 and the mechanism behind HS assembly on their core proteins.

I hope that I have written this dissertation at a level at which readers with scientifically diverse backgrounds can understand and appreciate it. First there is a general introduction about the proteins and their post-transcriptional modifications focusing on the HS chains biosynthesis via the exostosin family enzymes. In the second chapter there is a brief, but sufficiently detailed description of the HSPG, in particular the GPC protein family and their roles in modulating various signalling processes. Afterwards, I will try to summarize, in chapter three, the available functional and biochemical knowledge regarding the GPC1 proteoglycan. Chapter four introduces an investigation of the current study followed by concise description of the methods that were used. Finally, chapter five pinpoints the main findings of the papers included in this dissertation and discusses further future directions.

Chapter 1

Background

Cells are able to communicate and interact with their surroundings through proteins, in particular membrane proteins. Membrane proteins are structurally and functionally highly diverse, and they are involved in cell-cell and cell-matrix interactions, signal transduction, internalization, and intercellular connections. Moreover, they connect the cytoskeleton to the extracellular matrix (ECM). Membrane proteins are classified into several categories including the integral membrane proteins, which have one or more segments permanently embedded into the membrane, and peripheral proteins that do not penetrate into the hydrophobic lipid core but are usually temporarily adhered either to other integral proteins or directly to the lipid bilayer by a combination of hydrophobic, electrostatic or other non-covalent interactions.

The majority of membrane proteins have a signal recognition sequence, which targets them to the endoplasmic reticulum (ER) for translation. Then they are transported to the Golgi apparatus and finally to the cell membrane. The oxidizing surroundings in the ER enable the formation of disulphide bridges, which are fundamental for the accurate folding of many eukaryotic proteins (1). Furthermore, the ER and Golgi apparatus contain specific enzymes that mediate protein folding and post-translational modifications. Post-translational modifications involve changing the chemical nature of the amino acids (deamidation, carbamylation, etc.), proteolytic cleavage of proteins, or addition of functional groups, such as phosphate, acetate, lipids or carbohydrates (glycosylation). In this chapter, I will describe the most common types of protein modifications in eukaryotes.

1.1 Protein glycosylation

Glycosylation is the process involving the covalent linking of carbohydrates to a protein partner, and it is a form of site-specific co- and post-translational modification. More than 50 % of all proteins are glycosylated, with varying carbohydrate contents (1% to 90%). Different types of glycosylation can occur, including 1) addition of a glycosylphosphatidylinositol (GPI)-anchor to the protein

hydrophobic C-terminus, 2) N-glycosylation, where the glycans are attached to the amino group of asparagine residues and 3) O-glycosylation with attachment of sugar molecules to the oxygen atoms of serine or threonine (2).

1.1.1 Glycosylphosphatidylinositol anchorage

Glypiated proteins are peripheral proteins that lack a transmembrane domain but are instead anchored to the eukaryotic cell membrane by a covalent linkage to a GPI anchor (3). To date, more than 250 glypiated proteins have been found that play vital roles in various biological processes such as cell-cell interactions, signalling, complement regulation, antigen activation, differentiation, development and also have other miscellaneous functions (4). Interestingly, GPI-anchored proteins have been shown to play an essential role for viability, as defects in the biosynthesis of GPI-anchor are embryonic lethal in mammals (5). The GPI-anchor has a complex structure comprising a phosphoethanolamine linker, a glycan core, and a phospholipid tail, and it is positioned at the protein C-terminus during posttranslational modification in the endoplasmic reticulum (ER). The glypiated protein is then transferred via vesicles to the Golgi apparatus and finally to the external leaflet of the cell membrane.

Glypiated proteins are routinely associated into highly dynamic ordered microdomains of the membrane, called lipid rafts, which are heterogeneous in lipid content, enriched in cholesterol molecules, glycosphingolipids and certain types of lipidated proteins, making them detergent resistant (6). The majority of the saturated hydrocarbon chains of sphingolipids are tightly packed with cholesterol molecules, which makes the rafts more ordered than the surrounding lipids (7). Lipid rafts participate in the sorting of linked proteins, serve as supply sites for assembling cytoplasmic signalling molecules and also function as membrane domains involved in vesicular trafficking. They are further stabilized through protein-protein and protein-lipid interactions. GPI-anchored proteins in the lipid rafts are susceptible to phospholipase enzymes that cleave the GPI anchor from its associated protein in a rapid process, which may be used by the cells for selective regulation of signal transduction and sometimes to disrupt cell-cell adhesion (4).

1.1.2 N-glycosylation

Asparagine-linked (N-linked) glycosylation is a common, diverse and essential post-translational modification in all domains of life (Bacteria, Archaea, and Eukaryotes), where the glycans are attached to the nitrogen of an Asn side chain in the consensus sequon of Asn-Xaa-Ser/Thr (where Xaa may be any amino acid except proline) during their passage through the ER (8). All eukaryotic N-glycans share a common core of pentasaccharide structure, Manα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAcβ1-Asn-X-Ser/Thr, and are classified into three groups: high-mannose type

(oligomannose), hybrid type and complex type. Representative chemical structures of the three classes of N-glycans are shown in Figure 1.1.

During the first step of N-glycosylation, pre-assembled precursors of oligosaccharides with a defined structure (*N*-acetylglucosamine$_2$-mannose$_9$-glucose$_3$) are transferred *en bloc* from a lipid carrier onto target polypeptide chains by the oligosaccharyltransferase (OST) complex in the ER lumen. The attached N-glycan is further processed in the ER and Golgi apparatus by a complex series of reactions catalysed by various membrane-bound glycosidases and glycotransferases. Each glycosyltransferase enzyme has a unique substrate specificity, which determines the final N-glycan structure. Several N-glycan maturation processes occur for the hybrid and complex glycans during the glycosylation process, including e.g. fucosylation, galactosylation and sialylation.

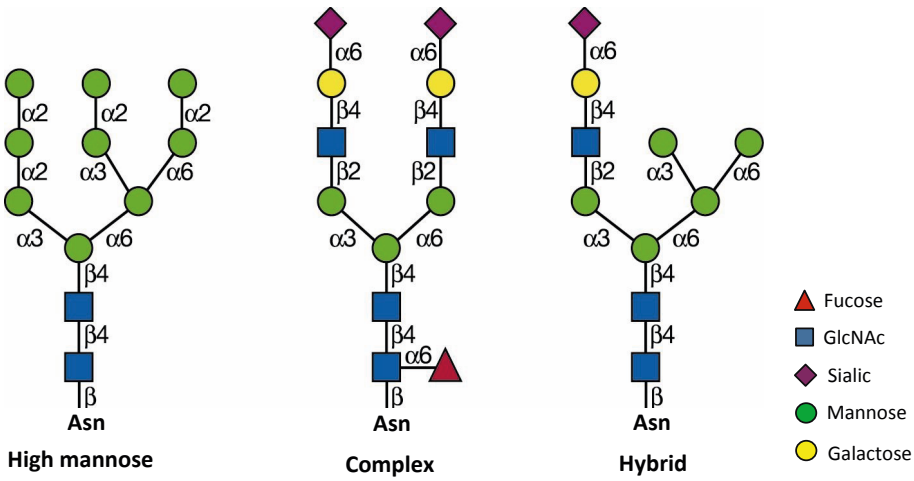


Figure 1.1 Types of N-glycans in mature glycoproteins: high mannose, complex, and hybrid. Each N-glycan tree contains the common $Man_3GlcNAc_2Asn$ core.

Interestingly, the N-glycosylation sites can be variably occupied, which contributes to N-glycan macro-heterogeneity. Moreover, the N-glycans are often chemically and structurally micro-heterogeneous and each glycosylation site may carry a range of different N-glycans. Glycan micro-heterogeneity can also be a result of tissue and cell-type specific pathways and leads to further diversity of the N-glycan structures. The protein conformation may also affect N-glycan heterogeneity, probably by affecting substrate availability or proximity to the glycan modification enzymes.

Many studies suggest a chaperone-like activity of the N-glycans during protein folding, where the protein does not fold to the active state in the absence of glycans. Moreover, the N-glycans may affect different properties of the glycoproteins

including their conformation, oligomerization, solubility, stability, quality control and identification by glycan-binding proteins. Therefore, N-glycosylations are important for various cellular processes including signalling, protein secretion, intracellular sorting and trafficking. The N-glycan population is strictly regulated during development and differentiation and altered in diseases (for reviews, see (8, 9)).

1.1.3 O-glycosylation and proteoglycans

O-glycosylation is a form of glycosylation that occurs mainly in eukaryotes, where the sugar is linked to the oxygen atom of the consensus glycosylation residue (serine, or threonine) and occurs predominantly in the Golgi apparatus. Proteins modified by covalent attachment of glycosaminoglycan polysaccharides (GAGs) to the oxygen of specific serine residues are called proteoglycans (PGs). All mammalian cells produce PGs and incorporate them into the cell membrane, secrete them into the ECM or sort them in secretory granules (intracellular PGs) (Figure 1.2). PGs influence various cellular processes, such as cell adhesion, migration, proliferation and differentiation. Moreover, PGs may interact with growth factors, cytokines and protein receptors involved in cell signalling and communication (for a review see (10)).

GAGs are linear, negatively charged polysaccharide chains, composed of repeated disaccharide building blocks containing amino sugars; (1) *N*-acetylglucosamine (GlcNAc) or *N*-acetylgalactosamine (GalNAc), (2) the uronic acids (glucuronic acid (GlcA) or iduronic acid (IdoA)), or (3) galactose (Gal). The GAG content varies between different PGs, where some contain only one GAG chain (e.g. decorin), whereas others can carry more than 100 GAG chains (e.g. aggrecan). Based on their disaccharide building blocks GAGs have been divided into four classes: dermatan sulphate (DS), keratan sulphate (KS), chondroitin sulphate (CS), heparin and heparan sulphate (HS). Moreover, hyaluronan is considered as a GAG, which is not covalently linked to a PG, but instead may interact non-covalently with PGs via its binding motifs (for review see (11)).

HS and CS/DS are the most common categories of GAGs. HS chains consist of repeating disaccharide units of GlcNAc and GlcA/IdoA, whereas GlcNAc is replaced by GalNAc in CS/DS. HS & CS/DS synthesis is established via membrane-bound glycosyltransferases in the Golgi apparatus. Xylosyltransferases initiate the process by the addition of xylose (Xyl) to the consensus serine residue of the attachment site. A glycine residue is invariably found after the serine attachment site, but a precise consensus sequence for xylosylation has not been identified yet. After Xyl addition, a linkage tetrasaccharide assembles on the serine amino acid (GlcA-Gal-Gal-Xyl-Ser). The addition of the next residue is the critical factor in determining which GAG will be formed: the addition of GalNAc results in initiation of CS/DS while addition of GlcNAc results in formation of HS.

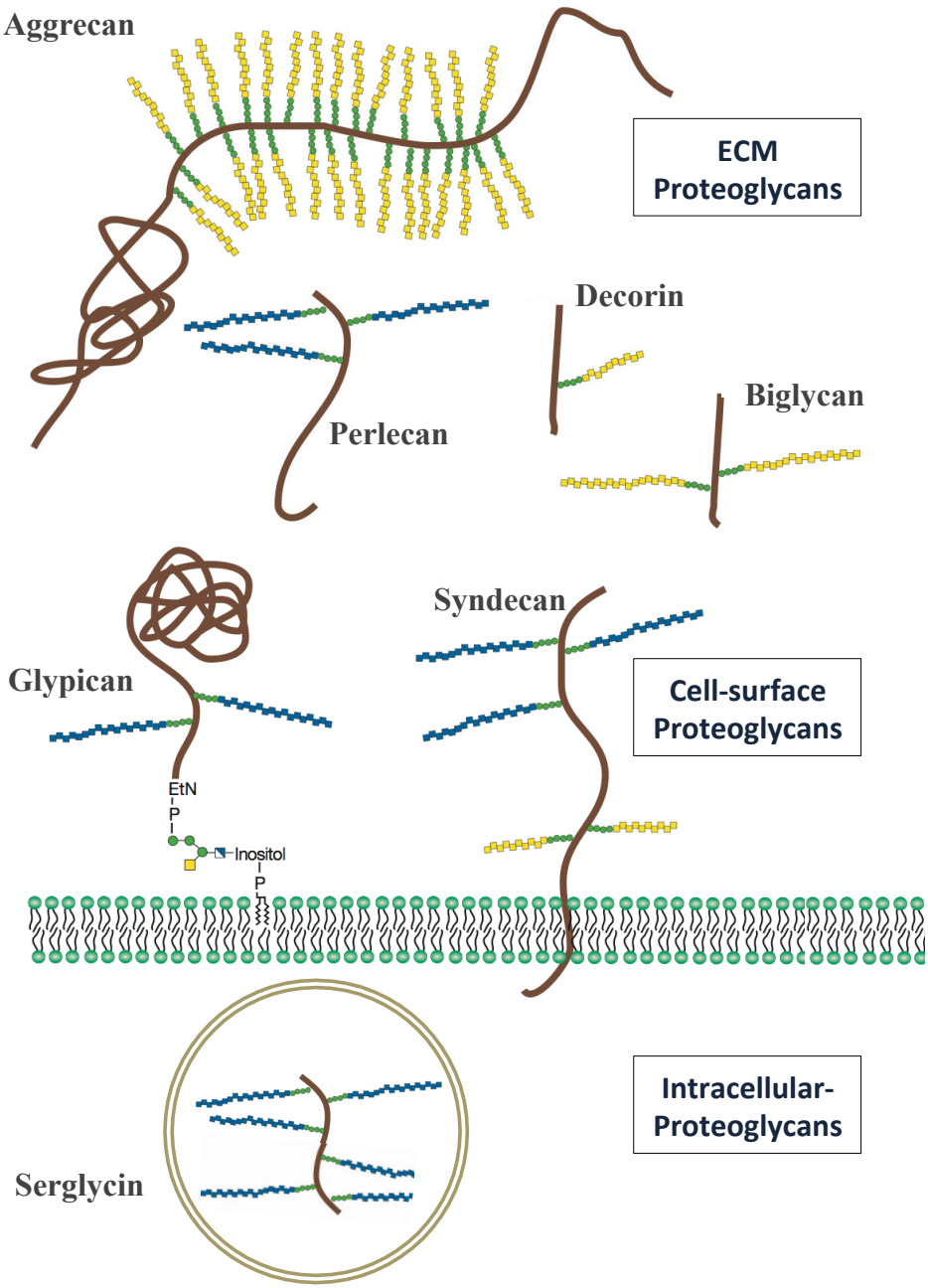


Figure 1.2 PGs consist of a protein core (brown) and one or more attached GAG chains: [blue] HS and [yellow] CS/DS. Different categories of PGs as displayed: ECM, cell-surface, and intracellular PGs. ''The figure is modified with permission from (11)''.

Afterwards, the GAG chains are elongated by the sequential addition of corresponding sugar residues of the repeating disaccharides (11, 12). The exact factors that decide the type of GAGs are still unknown, however, it seems likely that the core protein has a vital regulatory role in this process (13, 14). For example, most PGs substituted with HS contain the GAG attachment dipeptide Ser-Gly flanked by clusters of aspartic and glutamic acids and adjacent tryptophan. Mutations in crucial acidic residues within this region give rise to more CS than HS (15). The substitution of HS chains on the PGs is a complicated process and will be described below.

1.2 Heparan sulphate biosynthesis and exostosin family

HS biosynthesis is performed by glycosyltransferases of the exostosin family (EXTs), which initiate, elongate and terminate the HS backbone formation. Five members have been identified in mammals, including EXT1, EXT2, EXTL1, EXTL2 and EXTL3. The EXT proteins are well conserved, particularly in their C-terminal parts (Figure 1.3). Many cysteine residues are conserved in all EXTs, suggesting that they share a common fold, at least in the C-terminal domain. The exostosin proteins contain one or more conserved Asp-Xxx-Asp (DXD) motif, except for EXTL1 (16, 17). The DXD motifs are typical for glycotransferases utilizing nucleotide-activated sugars as donor substrates and they are most likely involved in either substrate recognition and/or catalysis (18). EXTL2, which comprises ~330 amino acids, is apparently half the size of the other family members, whereas EXTL3 (~900 amino acids) is the largest one and contains an additional N-terminal fragment with no homology to the other EXT family members.

EXT1 and EXT2 genes were first identified as responsible for hereditary multiple exostoses (HME), an autosomal inherited human disorder characterized by formation of cartilage-capped bony outgrowths at the epiphyseal growth plates (19). The EXT1 and EXT2 polymerases show sequence similarities and are responsible for HS chain elongation and polymerization by alternating transfer of GlcA and GlcNAc residues to the growing polymer, where the levels of the individual proteins affect the polymerization process. Furthermore, neither of the two EXT proteins can substitute for the other one (20, 21). Co-expression experiments show that two EXT enzymes form a hetero-oligomeric complex in vivo that possesses significantly higher glycosyltransferase and polymerase activities than the individual enzymes, suggesting that this complex represents a biologically functional polymerization unit involved in HS chain synthesis (20, 22, 23). The other members of the EXT family are known as the EXT-like proteins and include EXTL1, EXTL2 and EXTL3. They work in a complex way to initiate and control the HS synthesis process. EXTL1 has been identified as possessing a GlcNAc transferase (GlcNAc-TII) activity, catalysing the addition of GlcNAc to the growing HS chain (24). Few reports have been published regarding EXTL1, and its precise function is still unclear.

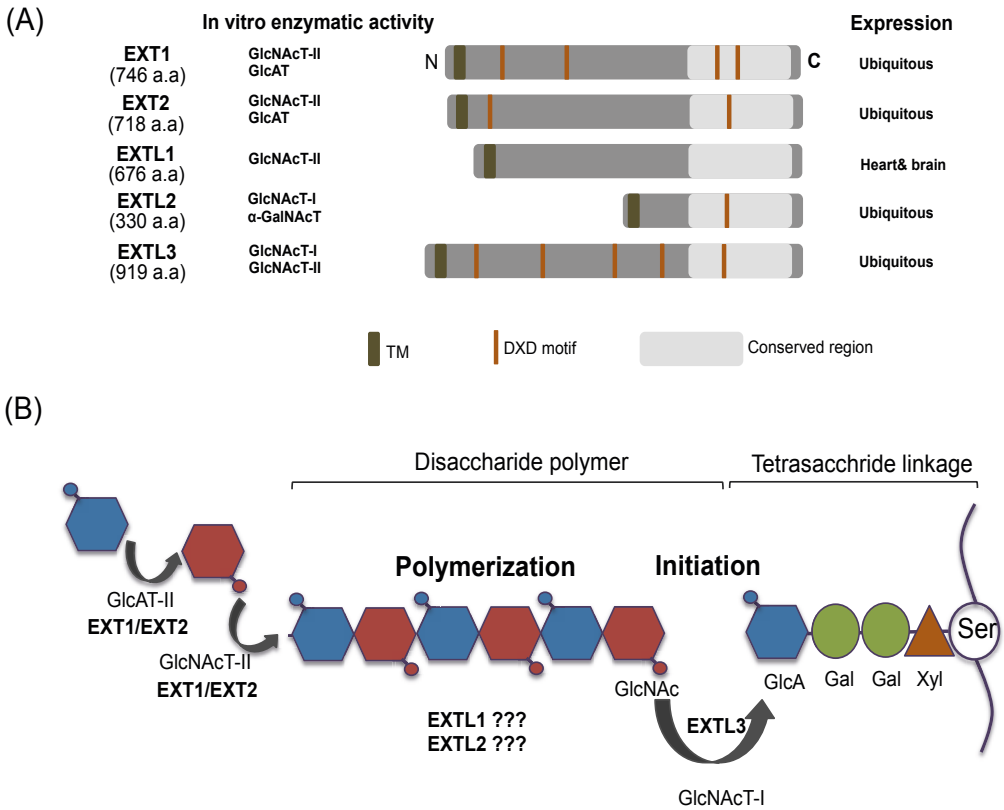


Figure 1.3 EXT family and HS biosynthesis. (A) Comparison of the human EXT members. The membrane spanning domains and the DXD motifs are indicated by brown bars and yellow bars respectively, while the C-terminal conserved regions are shown as grey shadows. The common expression patterns and the in vitro activities of each protein are also indicated. (B) Schematic representation of the HS assembly pathways with illustration of the different glycosyltransferase activities involved in HS initiation and elongation.

EXTL2 has a weak in vitro GlcNAc-TI activity, meaning that it transfers GlcNAc to the linkage tetrasaccharide. However, recent studies demonstrate that EXTL2-knockout mice produce significantly higher HS than wild-type ones (25). Other studies have shown that EXTL2 can transfer a GlcNAc residue to the tetrasaccharide linkage that is phosphorylated by a xylose kinase-1 and subsequently terminates chain elongation (25, 26). Interestingly, *Caenorhabditis elegans* (*C. elegans)* and *Drosophila melanogaster (D. melanogaster)* lack the orthologues of the mammalian EXTL1 and EXTL2, suggesting that they are perhaps not essential for HS biosynthesis. Clearly, additional studies are required to sort out the precise functions of EXTL1 and EXTL2 in HS assembly.

EXTL3 is evidently a bifunctional enzyme with GlcNAc-TI and GlcNAc-TII activities, which suggests its participation in both HS initiation and elongation (24). Published studies show that reduction of EXTL3 expression levels results in synthesis of longer HS chains whereas the EXTL3 overexpression has no clear effect on the HS chain lengths (20). Furthermore, no HS was detected in 9-day old mouse embryos lacking EXTL3 (27). Mutations in corresponding enzymes in *D. melanogaster*, *C. elegans*, and zebrafish *(Danio rerio)* resulted in reduced HS synthesis (28-30). In general, the published results so far suggest that EXTL3 works as an initiator of HS chain biosynthesis. The significance of EXT family for HS synthesis and HS-reliant signalling has been shown in many animal models. Loss of function of these genes in mouse, zebrafish, *D. melanogaster, C. elegans* and human affects the cellular signalling pathways and causes severe developmental abnormalities as well as serious pathologies (Table 1.1).

1.3 HS modifications

Concomitantly with the chain elongation, extensive modifications of the HS polysaccharide are carried out by different sulfotransferases and epimerases. The HS modifications are initiated through N-deacetylation and N-sulphation of the GlcNAc by N-deacetylase/N-sulfotransferase enzymes (NDST1-4). This reaction has been regarded as a key control step, as the subsequent modifications only occur in the vicinity of N-sulphated glucosamines. Subsequent modifications include C-5-epimerization, 2-O-sulphation, 6-O-sulphation and 3-O-sulphation (31). The variable length of HS chains together with detailed structural modifications result in extensive diversity of HS chains of PGs and thereby various biological functions (for a review see (32)).

Table 1.1 Loss of function and diseases associated with EXT and their orthologue gene mutants.

Enzyme	Model	Mutant phenotype/syndrome/disease	Ref.
EXT1	Mouse	-Null allele: embryonic lethality & lack of HS chains.	(33)
	D. melanogaster	-Segment polarity defects, impaired morphogen distribution and reduced HS.	(28)
	C. elegans	-Embryonic lethality and impaired HS synthesis.	(34)
	Human	-Hereditary multiple exostoses. -Epigenetic inactivation: leukaemia and non-melanoma cancer.	(35, 36) (37)
EXT2	Mouse	-Null allele: embryonic lethality and lack of HS chains.	(26)
	D. melanogaster	-Segment polarity defects, impaired morphogen distribution and reduced HS.	(21, 28, 38)
	Danio rerio	-Retinal ganglion cell axons missorting in the optic tract, fin defect and low HS level	(39)
	Human	-Hereditary multiple exostoses. -Breast carcinoma.	(36, 40) (41)
EXTL1	No mutant reported		
EXTL2	Mouse	-Viable, impaired liver regeneration and high HS level.	(25, 42)
EXTL3	Mouse	-Null allele: embryonic lethality and lack of HS.	(27)
	D. melanogaster	-Segment polarity defects, impaired morphogen distribution and reduced HS.	(28, 38)
	Danio rerio	-Retinal ganglion cell axons missorting in the optic tract, fin defect and low HS level.	(39)
	C. elegans	-Embryonic lethality and mpaired HS synthesis.	(29)
	Human	-Colorectal cancer.	(43, 44)

Chapter 2

Heparan sulphate proteoglycans & glypicans

2.1 Heparan sulphate proteoglycans (HSPG)

One of the most widespread types of PG is heparan sulphate proteoglycan (HSPG), with the characteristic of containing one or more HS chains. The HS chains, due to their broad structural diversity, are able to interact with a wide variety of proteins, such as chemokines, growth factors, morphogens, extracellular matrix components and enzymes, among others. The specificity of the interaction between the HS chains and their targets is affected by the fine structure of the polysaccharide chain, the overall organization of HS domains and the precise protein motifs. A relatively small set of HSPGs (~20) has been identified, and HSPGs are classified into three groups according to their localization: [1] membrane associated HSPGs that may either span the lipid bilayer through a short hydrophobic domain, such as syndecans (45), or that are linked to the cell membrane by a GPI anchor, like glypicans (46, 47); [2] secreted HSPGs in the ECM (e.g. perlecan, agrin, type XVIII collagen), [3] the secretory vesicle HSPGs, e.g. serglycin (10, 32) (Table 2.1).

A key function of the HSPG on the cell surface is to promote receptor-ligand binding and thereby signalling through high-affinity receptors. An example is the fibroblast growth factors (FGF), which form FGF/HS complex that interact with the FGF receptor (FGFR) kinases (48). Furthermore, HS can also control morphogen gradients required for tissue differentiation (10, 49). Even so, HSPG can trigger cell response through signal transduction pathways, along with translocation to intracellular compartments, due to the interactions of the core protein and/or HS chains with specific ligands (50). Here, a family of cell-surface HSPGs, glypicans, will be described in more details.

2.2 Glypicans

Glypicans (GPCs) are an HS-substituted PG family attached to the external leaflet of the cell membrane by GPI anchorage, where they may interact with extracellular signals and receptors. Genetic and biochemical studies have shown that GPCs co-

Table 2.1: Human heparan sulphate proteoglycans summary

HSPG	Human chromosome localization	GAG type	Protein core (kDa)	Ref.
GPI-anchored				
Glypican 1	Chromosome: 2 location: 2q35-q37	HS	56	(51)
Glypican 2	Chromosome: 7 location: 7q22.1	HS	59	(52)
Glypican 3	Chromosome: X location: Xq26.1	HS	59	(53)
Glypican 4	Chromosome: X location: Xq26.1	HS	58	(54)
Glypican 5	Chromosome: 13 location: 13q32	HS, CS	59	(55)
Glypican 6	Chromosome: 13 location: 13q32	HS	58	(56)
Transmembrane				
Syndecan 1	Chromosome: 2 location: 2p24.1	HS, CS/DS	33	(57)
Syndecan 2	Chromosome: 8 location: 8q22-23	HS	23	(58)
Syndecan 3	Chromosome: 1 location: 1pter-p22.3	HS, CS/DS	43	(59)
Syndecan 4	Chromosome: 20 location: 20q12	HS	22	(60)
CD44	Chromosome: 11 location: 11p13	HS/CS	37–81	(61)
Betaglycan	Chromosome: 1 location: 1p33-p32	HS/CS	110	(62)
Neuropilin-1	Chromosome: 10 location: 10p12	HS/CS	130	(63)
Extracellular matrix				
Perlecan	Chromosome: 2 location: 1p36.1-p34	HS	400–450	(64)
Agrin	Chromosome: 1 location: 1p36.33	HS	212	(65)
Type XVIII collagen	Chromosome: 21 location: 21q22.3	HS	180–200	(66)
Testican 1	Chromosome: 5 location: 5q31.2	HS/CS	48	(67)
Testican 2	Chromosome: 10 location: 10pter-q25.3	HS/CS	45	(68)
Testican 3	Chromosome: 4 location 4q32.3	HS/CS	47	(69)
Secretory vesicles				
Serglycin	Chromosome: 10 location: 10q22.1	HS/CS	10–19	(70)

regulate and modulate various cellular signals, namely FGFs, the wingless-integrated (Wnt), Hedgehog (Hh), bone morphogenic protein (BMP), Slit and insulin-like growth factors. GPCs are involved in various cellular and biological processes such as cell adhesion, cell division and proliferation, cell differentiation, homeostasis and development. In addition, GPCs may work as carriers for cellular uptake of compounds and complexes with positive net charges such as polyamines (For reviews on GPCS function, see (46, 47, 71-75)).

2.2.1 Characteristic features and localization of glypicans

GPCs have only been found throughout the Eumetazoa, where the first GPC was identified in 1990 (76). Six GPC members have been identified in vertebrates (GPC1-GPC6) (56, 76-79) and five members in invertebrates; two homologues in *Drosophila* (Dally and Dally-like (Dlp)) (80, 81), two in *C. elegans* (gpn-1 and lon-2) (82, 83) and one in zebrafish (Knypek) (84). From an evolutionary perspective, vertebrate GPCs are classified into two subfamilies: GPC1, -2, -4 & -6 and GPC3 & -5, with approximately 25% sequence identity between the two groups. For the first GPC group, GPC-4 and -6 are closely related (65% identity), whereas GPC1 and -2 are relatively divergent. Dally is more similar to the GPC-3 & -5 subfamily, and Dlp is more similar to the other subfamily. The mammalian GPCs contain between 550 and 580 amino acids, whereas Dally and Dlp have longer insertion sequences and are composed of 626 and 765 residues respectively. The mammalian GPCs have a characteristic pattern of expression during development and adolescence. GPC1 is ubiquitously expressed at various levels in different tissues in the adult, whereas it is mainly expressed in the CNS and skeletal system during embryonic development. GPC3, -4, -5 and -6 are widely expressed in many tissues and organs in the embryo but are more limited in the adult. GPC3, -4 & -5 are expressed extensively in the adult CNS and GPC6 is found in the heart, liver, kidney, intestines, and ovaries. GPC2 is located in the CNS only during embryonic development without any detected expression in the adult.

The mature human GPC core proteins are ~60-70 kDa in size and their sequences share a characteristic pattern involving an N-terminal secretory signal and 14 evolutionarily conserved cysteine residues, indicating a conserved tertiary structure for all GPC core proteins (Figure 2.1). Furthermore, all GPC core proteins share, in their carboxyl terminal regions, attachment sites for HS GAG chains and a hydrophobic sequence for GPI anchorage. The location of the GAG attachment sites close to the GPC C-termini will place the GAG chains close to the cell membrane, suggesting their interaction with other cell-surface receptors and other molecules. GPCs are predicted to contain variously: no N-glycosylation sites (GPC2 & -6), one site (GPC4), two sites (GPC1), or three sites (GPC3 & -5). The N-glycans are probably involved in protein quality control and stabilization processes. Generally, GPCs carry HS chains, but GPC5 also exhibits CS chains (55, 85). The GAG polysaccharides of

the GPC proteins can be altered in a various ways, having variable length (50-150 disaccharides), sulphation and epimerization patterns. All these factors together result in a rich structural diversity in the glycanated GPCs that may differ from tissue to tissue and even from cell to cell. Furthermore, the GAG chains' structural variabilities are most likely core protein-specific, as it has been found that GPC3 and GPC5, expressed in the same cell type, exhibit different degrees of sulphation (85).

Like other GPI-anchored proteins, GPCs are mostly found at the exterior leaflet of the cell membrane, specifically in lipid raft domains. However, GPCs have also been detected outside the lipid rafts and at the basolateral surface of polarized cells (86). Moreover, GPCs have been detected intracellularly. For example, GPC3 has been detected in the cytoplasm of liver cancer cells (87), although whether cytoplasmic GPC3 plays an specific role is still unknown. Further, GPC1 has been shown to undergo a copper and nitric oxide-dependent recycling through the endosomal pathway and has been detected in the nuclei of many cells, as described later in chapter 3.

GPCs can be cleaved off at their GPI anchors by extracellular lipases, allowing the release of PG into the extracellular environment, in a process called protein shedding. The Notum enzyme was first identified to cleave Dlp but not Dally (88). In vertebrates, Notum has so far been shown to cleave GPC3, -4, -5, and -6 (89). Several studies have shown that the shedding process results in secretion of functionally active GPCs. For example, it has been shown that shed GPCs have a role in transport of Wnt, Hh and BMP and for regulation of morphogen gradient formation in *Drosophila* (90-92). Proteolytic cleavage of the core proteins is another process that can contribute to generation of various GPC forms. Protease cleavage sites have been identified in GPC1, -3, & -4 (76, 93, 94). It has been shown that the GPC3 core protein is processed by a furin-like convertase generating a ~40 kDa N-terminal subunit and a ~30 kDa HS-carrying C-terminal subunit. These subunits remain connected by disulphide bonds. This internal cleavage by convertase is required for GPC3 modulation of Wnt signalling and gastrulation movements (93), and is furthermore essential for GPC3-induced inhibition of Hh signalling (95).

2.2.2 Biological activities of glypicans and mutation effects

GPCs modulate various intracellular signalling events by acting as mandatory co-receptors. The functions of GPCs in a specific cellular process rely on their structural features and on the set of growth factors and growth factor receptors. GPCs may introduce either a stimulatory or an inhibitory effect on the cell signalling. For example, it has been shown that GPC3 binds to Wnt and the Frizzled receptor and stimulates Wnt signalling by facilitating and/or stabilizing the interaction between the Wnt and Frizzled proteins (96). Other studies show that GPC3 inhibits Hh signalling during development by competing with the Hh Patched receptor (46).

Figure 2.1 Multiple sequence alignment of the human glypicans (GPC1-6), including the signal peptide sequences. The positions of the GPC1 secondary structure elements are indicated above the alignment. The 14 evolutionarily conserved cysteines are indicated with pink arrows. N-glycosylation sites are in bold face and green boxes.

The levels of the GPC core proteins on the cell surface may regulate the concentration of morphogens. As described earlier, GPCs can be shed from the cell surface by Notum or protease cleavage. The shedding may then play a vital role to reduce the cell surface concentration of morphogens. On the other hand, it has been shown that shedding of cell surface PGs and thereby release of cell surface-bound growth factors induces long-term growth factor signalling (97). Recent studies suggest functional roles for GPC core proteins in mediating growth factor signalling pathways by direct binding to BMP4, FGF2, Wnt and Hedgehog signals (98-100).

A powerful genetic approach to investigate protein function is to perform gene knockouts or mutations. Of the six GPCs, only gene knockouts for GPC1 (101), GPC3 (102, 103), GPC4 (104, 105) and GPC6 (105) have been published (Table 3). GPC1 knockout in mice results in reduction of brain size by 30% in birth, indicating a significant role of GPC1 in brain development. GPC3 and GPC4 gene mutations and knockout studies indicate that they play a role in cartilage and bone development. GPC4 knockout also results in defective synapse formation. Further studies demonstrate that mice lacking GPC6 die shortly after birth from breathing difficulties, suggesting neural dysfunction (105). Moreover, several mutational disruption studies of the GPC core proteins point towards the functional importance of the core proteins in mediating cellular signalling (Table 2.2).

Table 2.2 Mutants altered in glypicans.

Protein	Model	Mutant phenotype/syndrome	Ref.
GPC1	Mouse	- GPC1 knockout: reduced brain size - Athymic mice (lacking GPC1): show decreased tumour angiogenesis and metastasis.	(101) (106)
GPC2	No mutant reported		
GPC3	Mouse	- GPC3 knockout: Simpson–Golabi–Behmel syndrome - Alterations in Wnt signalling - Increased Hedgehog signalling - Alterations in BMP- and FGF-signalling	(103) (107) (108) (109)
	Human	- Simpson–Golabi–Behmel syndrome	(110)
GPC4	Zebrafish (Knypek)	- GPC4 (knypek) deficient: craniofacial skeletal defect - Impaired cell polarity during convergent extension	(104) (111)
	Xenopus	- Gastrulation defects (Wnt disruption) - Dorsal forebrain defect (impaired FGF signalling)	(112) (113)
	Mouse	- Gpc4- knockout mice: defective synapse formation	(105)
GPC5	No mutant reported		
GPC6	Human	- Impaired endochondral ossification and omodysplasia	(114)

Chapter 3

Glypican-1: What do we know so far?

Glypican-1 (GPC1) is mainly expressed in the neural and skeletal systems during development and ubiquitously in the adult. The human GPC1 gene codes for a protein of 558 amino acids with a predicted molecular weight of 62 kDa. The GPC1 protein is composed of a N-terminal core (residues: 24-474) and a small C-terminal domain (residues: 475-530) containing three GAG attachment sites at S486, S488 and S490, and it ends with hydrophobic residues to link the protein to a GPI-anchor. The full-length GPC1 core protein is O-glycanated exclusively with three HS chains at Ser-486, Ser-488 and Ser-490 and further decorated with two N-linked glycans at positions Asn-79, and Asn-116 (115).

3.1 Function and Recycling

Many functions of GPC1 are related to the HS chains, which are capable of binding and/or transporting and/or activating a variety of proteins and signalling molecules such as growth factors (FGF2, VEGF), several types of cytokines, chemokines, polyamines and viral proteins. Thereby, GPC1 is involved in various aspects of cell biology including cell proliferation, differentiation, division, vascular and brain development and angiogenesis (reviewed in (51, 73)). It is known that both the core protein and the HS chains of GPC1 are important for brain function, as knockout of the GPC1 gene results in reduction of brain size by 30% probably due to the transient reduction in FGF signalling during embryogenesis (101).

Furthermore, athymic mice that lack GPC1 exhibit decreased tumour angiogenesis and metastasis following implantation of either human pancreatic cancer or murine melanoma cells, indicating that GPC1 has vital role in malignancy (106). A role for GPC1 in axonal guidance and regeneration via Slit has also been proposed (116, 117). The shedding of GPC1 from the satellite cell surface sequester FGF2 from its tyrosine kinase receptor which allows differentiation to occur as FGF2 inhibits differentiation (118). Furthermore, errors in HS metabolism, as in mucopolysaccharidosis type III or Sanfilippo syndromes, result in neurodegeneration and mental retardation (119).

GPC1 is a cell surface HSPG, which can be internalized via a non-classical, caveolin-1 associated pathway. It undergoes a recycling from cell surface to endosomes and back to the cell surface via Golgi. Before or during endocytosis, the non-conserved cysteine residues in the C-terminal region of the GPC1 become *S*-nitrosylated by nitric oxide in a copper-dependent reaction (120, 121), a reaction that can be catalysed for example by prion protein, the brain-specific splice variant of ceruloplasmin or amyloid precursor protein (122). During recycling, the HS chains of GPC1 are degraded by heparanase and further on by a novel copper, nitric oxide and vitamin C-dependent deaminative cleavage. Subsequently, new HS chains are synthesized on the stubs remaining on the core protein (122, 123). GPC1 recycling can internalize and transport basic compounds that are bound to their negatively charged HS chains such as polyamines and net cationic DNA-basic peptide complexes. The HS cleavage may be essential for the intracellular release of the cargo.

The GPC1 core protein contains a basic nuclear localization signal (NLS) KRRRGK (358-366) peptide and GPC1 has been detected in the nuclei of neurons, glia cells and Chinese hamster ovary cells (124, 125). The functional role of GPC1 inside the nucleus remains to be elucidated. Recent reports propose involvement of GPC1 in regulation of the cell cycle in endothelial cells, glioma cells and astrocytes (126, 127). However, it is still unclear whether these effects arise from extracellular or nuclear GPC1.

3.2 GPC1 in human diseases

Numerous studies have shown involvement of GPC1 in the pathogenesis of neurodegenerative diseases. GPC1 co-localizes with the cellular prion protein (PrP^{C}) on the cell surface and enhances its association with lipid rafts. Furthermore, it has been shown that GPC1 triggers and accelerates conversion of cellular PrP^{C} to the infectious form (PrP^{Sc}), which is the critical step in the pathogenesis of prion diseases such as Creutzfeldt-Jakob disease in humans and scrapie in sheep (128, 129).

Several studies have demonstrated the involvement of GPC1 in Alzheimer's disease, where GPC1 has been localized to the brain amyloid plaques (130). Both NO- and heparanase-induced degraded GPC1 HS have been found to be associated with the toxic amyloid β deposits in the brains of human Alzheimer's patients and transgenic Alzheimer's mice (131-133). Further, the HS oligosaccharides released from GPC1 by Cu/NO/vitamin C-induced cleavage form conjugates with amyloid β peptides and modulate the formation of toxic amyloid β oligomers in hippocampal slices from Alzheimer's mice (131). Other studies have shown that amyloid β toxicity is attenuated in cells overexpressing heparanase, suggesting protective effects of HS oligosaccharides generated by cleavage with heparanase (132, 134) (for a review see (135).

GPC1 is upregulated in human cancer cells such as glioma, pancreatic and breast cancers and supports and maintains the mitogenic effect of several growth factors (136, 137). Downregulation of GPC1 results in prolonged doubling times and decreased growth of cancer cells in vitro, as well as attenuation of tumour growth, angiogenesis, and metastasis in vivo. Accordingly, targeting GPC1 might provide new approaches for the prevention of cancer metastasis and for cancer therapy (106)

3.3 Biochemical characterization of GPC1

Based on amino acid sequences alignments, GPCs have been predicted to consist of a globular domain and a GAG attachment domain. Studies from our laboratory have previously shown that the human GPC1 core protein, expressed by HEK293 cells, has a predominantly α-helical single domain fold, with identical CD spectra for the core protein with and without the GAG chains, indicating that the HS chains do not influence the folding and the structure of the core protein (138). Furthermore, the HS chains have been shown to protect GPC1 against irreversible heat-induced aggregation, probably by supplying further negative charges to the core protein and thereby inducing electrostatic repulsion between exposed acidic residues of the core protein and the polyanionic HS chains.

The GPC1 amino acid sequence contains two potential attachment sites for N-glycosylation, Asn-79 & Asn-116. Svensson et al. (2011) investigated the N-glycosylation of the anchorless GPC1 and showed that the N-glycan attachment sites of GPC1 are invariably occupied (115). The N-linked glycans at these sites affected GPC1 protein expression levels and improved the thermal stability of the protein. Nevertheless, the protein was folded correctly even in the absence of N-linked glycans. In addition, the N-glycans seemed to influence the length, and also possibly the fine structure of the HS chains, but they were not involved in the GAG class determination.

GPC1 is exclusively substituted with HS, whereas other GPCs like GPC5 may possess both HS and CS (85). Robert et al. (2001) suggested that the core protein plays a vital role in the preferential assembly of HS on rat GPC1 by showing that exclusive expression of the GAG attachment domain results in substitution with ~90% CS. Mutational analysis shows that sequences at least 70 amino acids away from the first GAG attachment site and other nearby parts of the core protein are required for HS assembly (13). These data strongly suggest that the correctly folded GPC1 core protein is involved in GAG class determination and is required for the preferential assembly of HS. Systematic mutagenesis and GPC1 structural determination would elucidate which part(s) of the GPC1 core protein are involved in influencing the HS substitution. Unfortunately, there is a shortage of detailed structural knowledge about the GPC core proteins. So far, attempts to express

truncated or mutated forms of GPC1 core domain have resulted in expression failure in many cases, probably due to misfolding. Structural studies would also reveal the functional role of the GPC1 core protein in more detail.

Chapter 4

The present investigation

4.1 Aim of the work

The principal objective of this thesis was to characterize structurally the GPC1 core protein and its overall topology on the cell membrane. This information would be of great assistance in elucidating the functional roles of GPC1 and the mechanism behind HS assembly on their core proteins. This main objective can be characterized by four specific questions (aims) as shown in Figure 4.1.

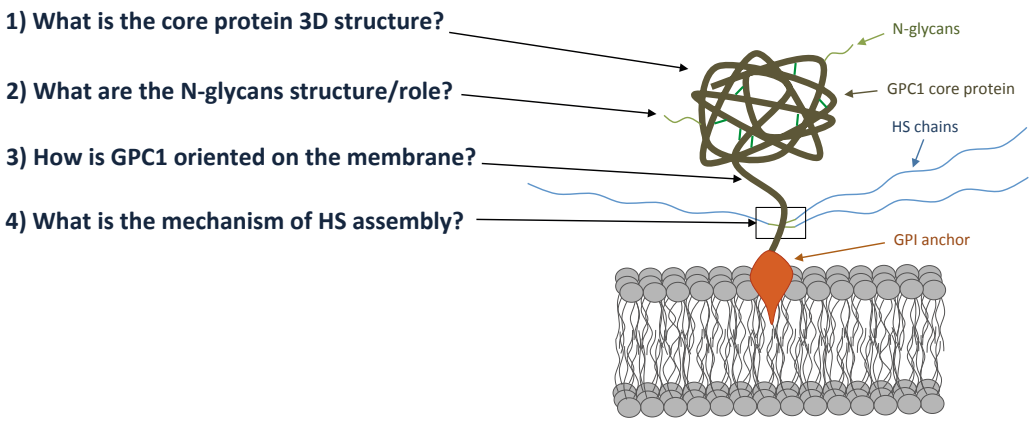


Figure 4. 1 The overall objectives of this thesis.

In this thesis, I have employed structural biology techniques to investigate these questions. The first three aims were accomplished mainly by the GPC1 protein crystallography and SAXS studies in papers I, II and III. To further enquire into the mechanism of HS assembly (aim 4), we expressed and characterized the EXTL3 protein (paper IV), which is supposed to initiate the HS assembly on the GPC1, and carried out preliminary explorations of its affinity and binding to the GPC1 core protein.

4.2 Methodology

I will briefly describe in this section the techniques that were used in this thesis to structurally characterize GPC1, EXTL3 and their complex. As this is a structural biology project, large amounts of highly purified proteins (~purity >95%) are required. Both GPC1 and EXTL3 are large (>500 residues) glycosylated proteins with a high cysteine content, which make them difficult to express using bacterial systems. Accordingly, we have used mammalian systems for the expression of our target proteins in order to maintain the correct glycosylation, folding, and other post-translational modifications. The main drawbacks of working with mammalian cells are the large time investment (taking months), high costs and the low expression yields in comparison with the protein productivity in *E. coli* but all these factors can be compensated by the high quality of the produced proteins (139).

In this study, the cDNA of the desired protein was introduced into a plasmid vector pCEP-BM40-HisTEV coding for an N-terminal secretion signal (BM40), a 6His-tag and protease cleavage site (TEV). The resulting plasmid was used to transfect 293 human embryonic kidney (HEK293) cells. Stable HEK293 clones expressing the desired protein construct were generated by growing the cells in selective antibiotic (hygromycin B) for several weeks (limiting dilution technique). The conditioned medium was collected and the proteins were purified using immobilized metal affinity chromatography on Ni-NTA columns using a linear imidazole gradient, followed by size exclusion chromatography in some cases.

4.2.1 Protein Characterization

The purified proteins were extensively characterized using various biochemical and biophysical methods to inspect the protein identity, purity, folding, and conformational stability in solution, as listed in Figure 4.2. The expressed proteins were further optimized for the structural studies by mutagenesis and enzymatic removal of polysaccharide chains. Further, the protein buffer content was optimized to increase the protein solubility and stability, in order to facilitate the structural analysis by crystallography and SAXS techniques. Moreover, the N-glycan chains of the GPC1 core protein were structurally characterized using mass spectrometry (MS) techniques.

A summary of the different recombinant GPC1 proteins used during this thesis work is listed in Figure 1 of paper III. Only one EXTL3 construct has been used during the course of this study (paper IV).

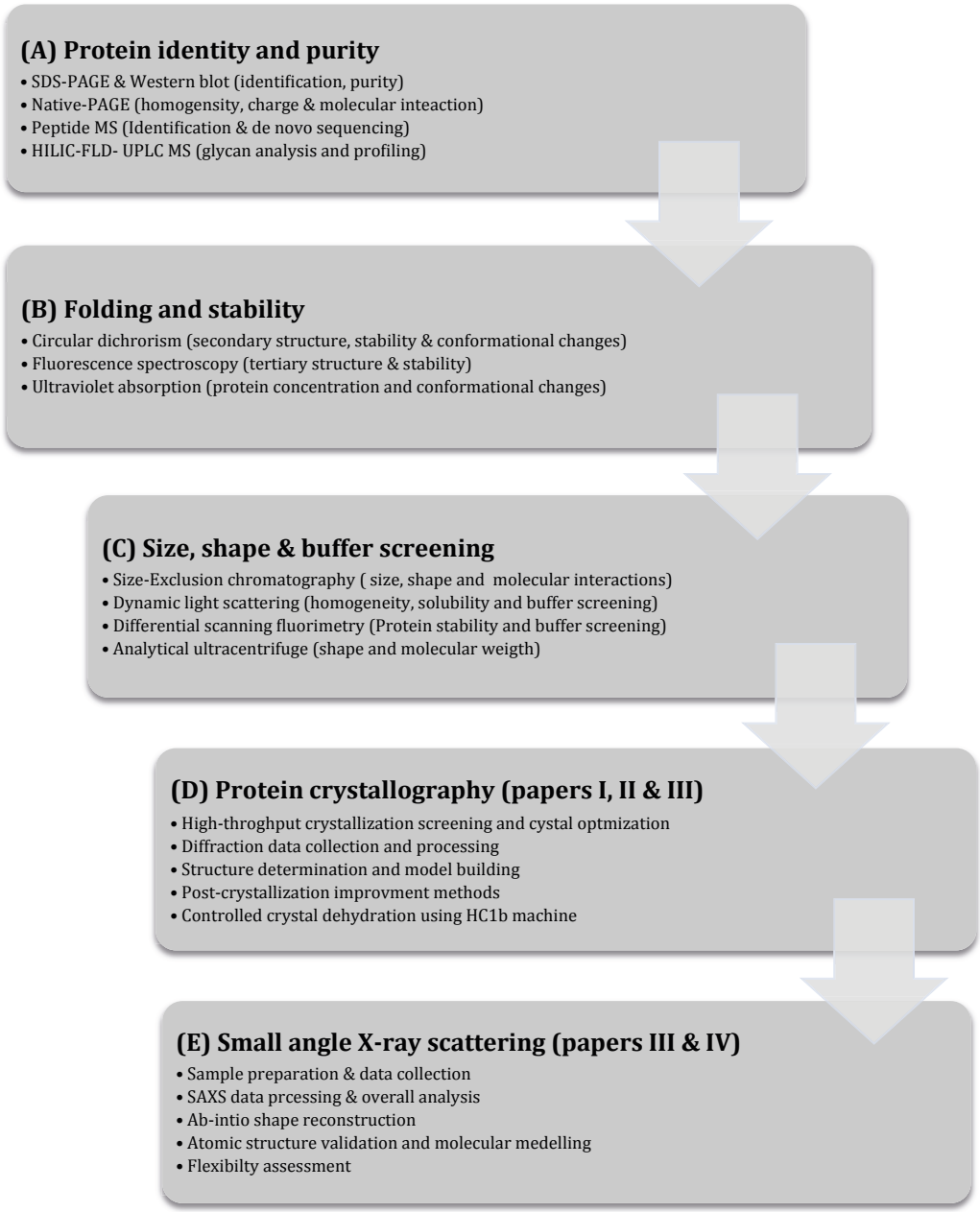


Figure 4.2 Biochemical characterizations of recombinant proteins. The first step is to evaluate the expressed protein identity and purity by various methods (A). These are followed by checking the protein folding and stability (B), and by characterizing the protein shape and size (C). Lastly, the protein buffer content is optimized (C) for further structural studies (D&E).

4.2.2 Protein crystallography (papers I, II, III & IV)

Before starting this project, there was no structural information available regarding the glypican core protein, therefore we planned to use protein crystallography to investigate the GPC1 core protein structurally at a high-resolution level. X-ray crystallography is a powerful form of very high-resolution microscopy that allows us to visualize protein structures at the atomic level to enhance our understanding of their function. Crystallography relies on the diffraction of X-rays by the electrons in the molecules constituting the investigated crystal and consequently the result of an X-ray diffraction experiments is a three-dimensional map showing the distribution of the electrons inside the unit cell, the electron density map. Subsequently, the relevant protein structure can be built in this electron density and the model is further evaluated (Figure 4.3) (For reviews about protein crystallography, see (140, 141)).

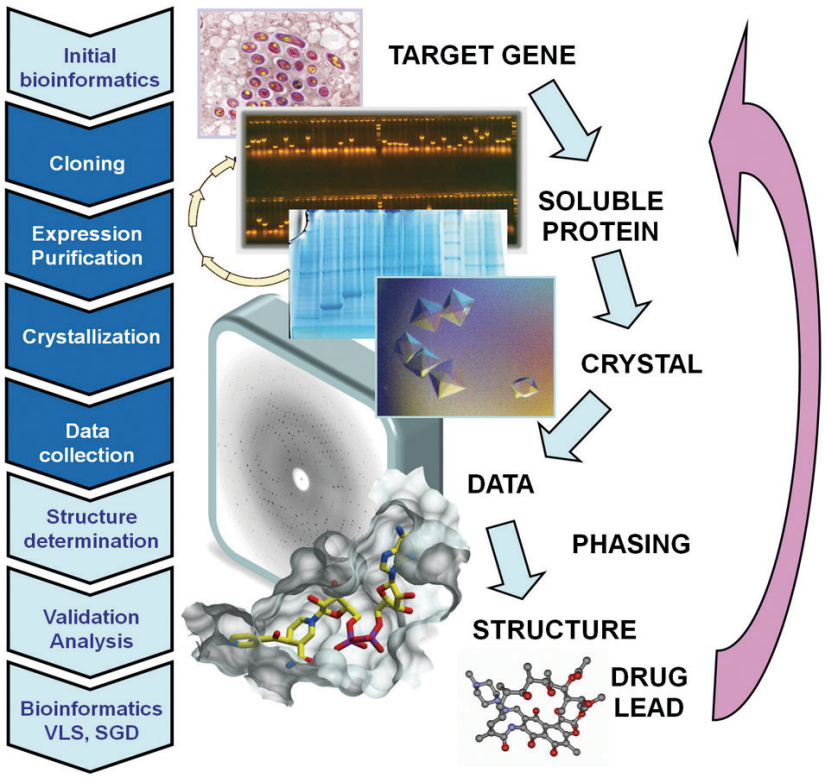


Figure 4.3 Key stages in the crystal structure determination process. The blue shaded bars indicate experimental procedures while the light shading bars indicate computer in-Silico works. The figure is reproduced with permission from "Biomolecular Crystallography" by Bernhard Rupp (140).

4.2.2.1 Protein crystallization

Protein crystallization is the critical step during the course of a protein structure determination. The general crystallization procedure is to reduce the protein solubility, with the goal that the protein molecules are separated from the solution and periodically self–assembled into a diffracting crystal, where these molecules are held together by a network of few but specific weak intermolecular interactions. The intermolecular packing may conformationally change the surface residues or the flexible loops of the target molecule, whereas generally no change occurs for the core of the protein in the crystalline state. In fact protein enzymes usually remain active inside the crystal. There are large voids between the packed protein molecules, which are filled with solvent, so-called solvent channels. These solvent channels allow the crystal to exchange liquid with the environment, which could be employed in ligand or heavy atom soaking.

The most common crystallization technique is vapour diffusion, where water vapour equilibrates from a drop containing protein and precipitant into a larger reservoir solution with higher precipitant concentration. This allows for a gentle and gradual increase in the protein and precipitant concentrations inside the drop, which could result in the supersaturation of the protein solution, and if the appropriate crystallization solutions (pH, ionic strength, type of buffer and precipitants) are used for a given protein, a well ordered protein crystal will be formed. Vapour diffusion is usually performed in sitting drop or hanging drop setups. Other protein crystallization methods are sometimes employed to obtain and/or improve the protein crystal, including free-interface diffusion, batch crystallization under oil and dialysis methods. The actual setup of the crystallization experiment is simple and easy to automate, which facilitates the initial crystallization screening using nano-volumes of the proteins (for reviews, see (142, 143)).

The challenges in crystallography are to produce highly purified, homogenous proteins that can be crystallized. Various characteristics of the protein affect its crystallizability, such as the presence of affinity tags, disordered regions, transmembrane parts, flexible domains and/or glycosylation, where all of these factors affect the sample conformational homogeneity. Accordingly, we expressed, purified various versions of GPC1 in large quantities, and then high throughput crystallization screening was performed.

More than 400 crystallization conditions were tested (at the MAX IV Laboratory crystallization facility, Lund University) for the full-length, truncated, deglycosylated and non-glycosylated GPC1 proteins (paper I). The crystallization conditions were further optimized and scaled up to microliter drops in order to improve the reproducibility of the crystallization and crystal quality. Various crystallization methods were also tested including; vapour diffusion (either in hanging and sitting drop formats), micro-batch and free-interface diffusion (Paper I).

Large, thin plates-like GPC1 (dC & dHS) crystals with dimensions around 0.8 x 0.3 x 0.05 mm grew within two weeks using sitting drop vapour diffusion (Figure 4.4).

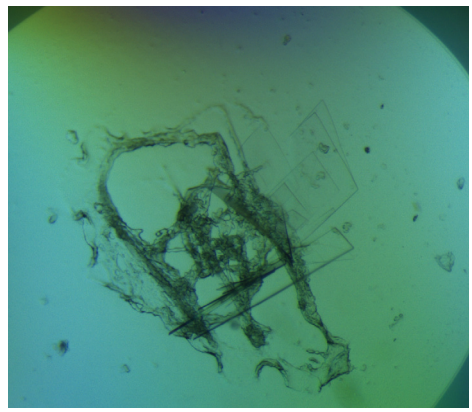


Figure 4.4 GPC1-dHS crystals grow against a reservoir solution of 14% PEG 6000, 0.1M Tris-HCL, 0.2 M $CaCl_2$ pH 8, in the presence of 2 mM dTT.

4.2.2.2 Diffraction data collection

The protein crystal can be treated as 3D diffraction grating, where the highly similar structural motifs are repeated in a periodic fashion inside the individual unit cells throughout the entire volume of a crystal. Accordingly, the diffraction signals will be magnified in some directions (constructive interference) and extinguished completely in all others (destructive interference). The diffraction appears on the detector as series of discrete spots, which are known as reflections. Each reflection (*hkl*) includes information on all atoms in the structure and conversely each atom contributes to the intensity of each reflection. If the crystal is well ordered, then the diffraction will be measurable at high angles or high resolution and a detailed structure should result.

X-ray electromagnetic radiation exhibits wave properties, in other words it has both amplitude and phase. In order to calculate the electron density we need to know the structure factor of each reflection F_{hkl}, which is a vector quantity of the amplitude ($|F_{hkl}|$) and phase (α), ($F_{hkl}=|F_{hkl}|.\ e^{i\alpha}$). A Fourier transform is used to calculate the electron density function $\rho(xyz)$ at each point within the unit cell given by the coordinates (*x, y, z*), referring to the unit cell axes

$$\rho(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} |F_{hkl}|\ e^{i\alpha_{hkl}}\ e^{-2\pi i(hx+ky+lz)} \qquad (4.1)$$

where V is the volume of the unit cell. We calculate $\rho(xyz)$ at closely spaced intervals of x, y, and z to make a three-dimensional function that can be contoured and displayed on a computer screen. Only the structure factor amplitude can be calculated from the square root of the corresponding reflection intensity, whereas all the phase information is absent. This is known as "the phase problem".

4.2.2.3 Structure determination and model building

Obtaining the missing phases is the experimentally most challenging step of the protein crystallography project. There are two major approaches to obtain the phase angles, called molecular replacement and experimental phasing. Phases can be calculated indirectly using molecular replacement, which employs a known reference structure (homologue model) to determine the crystal structure of a target molecule of similar structure. This task is achieved by finding the correct position of the search molecule in the crystal using the "rotation function" to find the orientation of the reference molecule in the target unit cell, and subsequent "translation function" to find its position. These yield the initial phases suggesting how the target molecule is posed in the unit cell, which can then be refined. In the absence of a suitable reference homologue model, phases should be determined experimentally by phasing experiments such as Multiple Wavelength Anomalous Dispersion (MAD), Single Wavelength Anomalous Dispersion (SAD), Multiple Isomorphous Replacement (MIR) and Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS).

The structure of the GPC1 core protein, the first crystal structure of a vertebrate glypican, was determined *de novo* by the two-wavelength MAD method on SeMet-substituted GPC1-dC crystals. MAD phasing relies on the occurrence of an anomalous scatterer (heavy atoms such as selenium) in the unit cell. Anomalous diffraction is collected from the same crystal at two or three different wavelengths using a synchrotron radiation source, typically in order to maximize the absorption and anomalous dispersion effect (an induced phase shift in all of the reflections collected at difference wavelengths). Using dispersive and anomalous signals, one can solve the substructure of the anomalously diffracting atoms and hence produce an electron density map for the whole molecule. This initial structure phase information is further enhanced by numerous density modification techniques providing substantially improved electron density maps, which are then interpreted by fitting the initial protein atomic model to it. This model can be used to refine the phases, leading to an improved model, and so on.

Manual model building is subsequently carried out using computer graphics programs that display the electron density (real space) and allow the placement and manipulation of the protein backbone chain and residues, matching the protein sequence and containing appropriate bond lengths and angles. The atomic positions and their respective B-factors (accounting for the thermal motion of the atoms) can be further refined to fit the experimental diffraction data (reciprocal space refinement), resulting in a better set of phases. Several rounds of rebuilding and refinement are carried out until the correlation between the experimental observed (F_{obs}) and model calculated (F_{cal}) structure factors is maximized. This agreement is measured by the R-factor:

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{cal}||}{\sum_{hkl} |F_{obs}|} \quad (4.2)$$

where k is a scaling factor to put $|F_{obs}|$ and $|F_{cal}|$ on the same scale. A similar quality gauge is R_{free}, which is calculated from a subset of couple of thousand, up to 10% of the dataset reflections that are not included in the structure refinement, where the R/R_{free} measure the overall model-data matching. Various evaluation criteria are applied to an entire crystal structure, most notably the resolution, isotropy and the completeness and multiplicity of the diffraction data. Furthermore, assessment of both geometry and electron density give an accurate picture of the reliability of the model.

4.2.2.4 Post-crystallization improvement methods

The GPC1 crystals diffracted anisotropically, reflected in a Wilson B factor (144) that was twice as large in the c* direction than in the a* and b* directions. This anisotropy may be a result of intrinsic disorder in the lattice packing, where there are more lattice contacts in the *a* and *b* directions than in the *c* direction (paper I). Various strategies to overcome such diffraction-quality problems have been described in the literature, using post-crystallization treatments such as crystal annealing, tempering, soaking, chemical cross-linking and dehydration (145, 146). GPC1 crystal annealing, cross-linking and chemical dehydration did not produce reproducible diffraction improvements, but controlled crystal dehydration was successful in our hands, as shown in the study in paper II.

Crystal dehydration is likely of great general interest for crystals with high solvent content and/or poor order, where a change in the solvent content might yield critical improvement in the diffraction quality. This crystal improvement is mostly driven by the possibility of inducing a change to more ordered crystalline state (with a higher number of crystal contacts) and is not dictated simply by a change in the solvent content of the sample. Furthermore, the mechanical forces induced by water loss in the crystal may rearrange local areas of disorder, thereby improving the diffraction profiles (145, 147). Several classic protocols have been developed for protein crystal dehydration, including soaking with dehydration compounds, air dehydration, and addition of a dehydrating solution to the crystallization drop. However, all of these methods are time- and crystal consuming. In order to dehydrate crystals more reproducibly, a number of devices have been designed to control the relative humidity (RH) surrounding the crystal, such as the humidity-control device (HC1b) (Figure 4.5) (148).

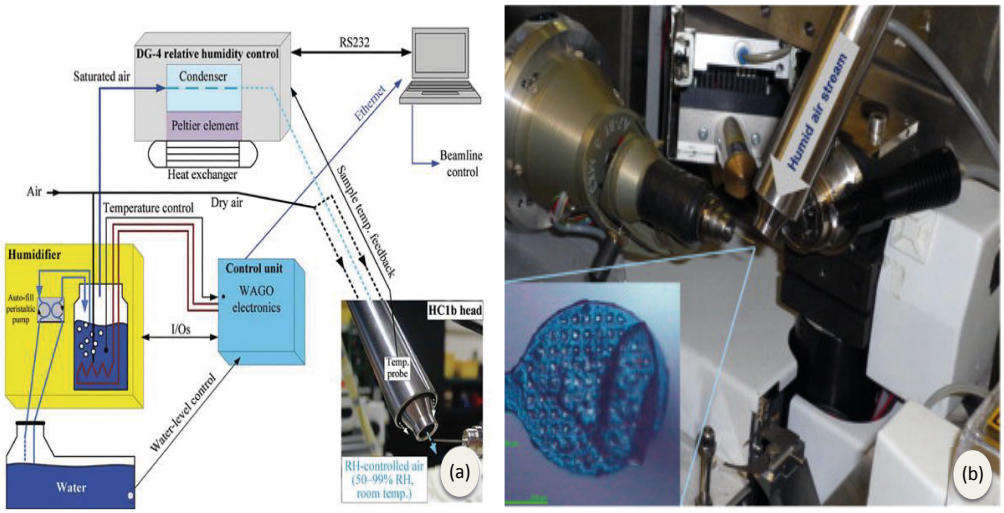


Figure 4. 5. The HC1b humidity control device. (a) Schematic of the HC1b design. (b) Detail of the sample environment. The crystal is mounted on a mesh loop and is kept under a humid air stream during the dehydration (reproduced with permission from (148)).

4.2.2.5 Controlled crystal dehydration using the HC1b machine

The HC1b machine is a user–friendly apparatus for crystal dehydration experiments that is capable of generating a controlled humidified air stream and is designed to allow full functionality at a synchrotron beamline without any disruption to the experimental environment. The HC1b delivers a precisely controlled humidified air stream (via its air-dispensing nozzle) of a precise RH, which can be employed to affect the crystal solvent content. The HC1b control software is used to monitor and control humidity changes and display a live image of the crystal during the experiment. The components and operations of HC1b have been described in detail, with a few examples (148, 149). The simplicity and the efficiency of HC1b operation make the crystal dehydration experiment reproducible and achievable within a reasonable time.

GPC1 crystals were mounted on standard mesh-loops and put in the HC1b air stream (at beamline I911-3, MAX IV Laboratory) for controlled dehydration (papers II & III). Various parameters of the dehydration protocol were optimized to induce a reproducible rearrangement of the molecules in the crystal. Such variables included: the RH step size, dehydration rate, equilibration time, annealing, number of steps and total time for the protocol as shown in Figure (4.6).

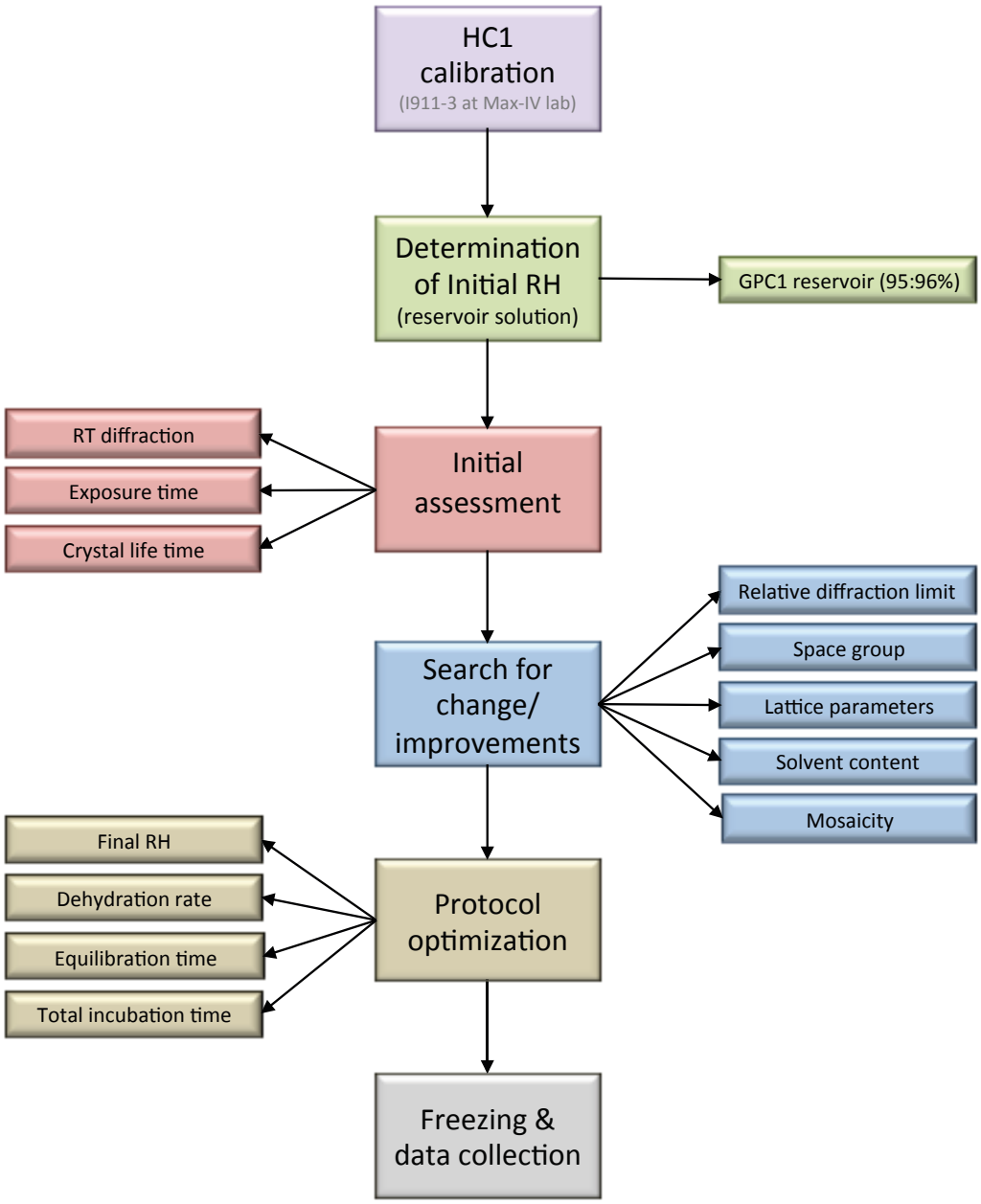


Figure 4. 6. The HC1b controlled dehydration experiment workflow (paper II).

Once the optimum hydration level is obtained, cryo-cooling of the treated crystals is easy to perform by hand or using a CATS sample changer (IRELEC, Saint-Martin-d'Hères, France) and the diffraction data are collected from the frozen crystals immediately or later on. More details are available in papers II & III.

4.2.3 Small angle x-ray scattering (papers III & IV)

Small angle x-ray scattering (SAXS) has become a versatile tool for the structural investigation of macromolecules in the solution state, in particular to characterize structural and conformational changes under physiologically relevant conditions that are difficult to achieve by other methods. It can be used to investigate functionally important conformational changes of various biomolecules, including complex formation, macromolecular folding–unfolding, movement of flexible domains and characterization of allosteric regulation (for reviews on SAXS, see (150-154)). Whilst limited by resolution (~10–40 Å), SAXS is not limited by the size of the biomolecules, unlike electron microscopy (dominant for large macromolecules) or NMR (size limit of ~50 kDa). The versatility of SAXS (using various solution conditions) makes it a powerful complementary technique to the high-resolution methods of x-ray crystallography and NMR. This combinatorial approach has been aided and developed in the recent years through the advances in computing power and software to provide much detailed analysis. A publicly available program suite for use in SAXS data processing and analysis is the ATSAS package (155), which was developed at the EMBL Hamburg, by the group of Dmitri Svergun. ATSAS is the most comprehensive collection of tools covering the major manipulation and interpretation tasks of SAXS data. The overall scheme of SAXS experiments is shown in Figure 4.7.

4.2.3.1 SAXS sample preparation and data collection

SAXS data should be collected from highly pure monodisperse samples. The protein purity can be assessed by SDS-PAGE gel. Any aggregation needs to be removed directly by spinning down the sample and/or filtration before the measurements. The sample monodispersity can be checked by dynamic light scattering (DLS) and improved by size-exclusion chromatography (SEC). There is a need for a perfectly matched solvent (equivalent to the one the sample is in) to be used as reference in the processing, which can be obtained by overnight dialysis of the sample against the appropriate buffer before the measurement.

There are two ways to collect SAXS data on proteins, either by collecting the scattering from samples at different protein concentrations (concentration series), or from an inline size exclusion chromatography with SAXS configuration (SEC-SAXS) (paper III & IV). SEC-SAXS can be used to obtain scattering data from highly pure monodisperse conformations of the macromolecules using an HPLC system that is

attached directly to the sample-measuring capillary cell. Subsequently, scattering frames are collected every short period (one or two seconds) and lastly, the scattering images with stable radius of gyration (R_g) (from similar species) are merged and scattering of the buffer subtracted to provide a single averaged scattering profile (Figure 4.8). The SAXS profiles are generally described as [log I (Q)] versus [Q (nm^{-1})], where I is the scattering intensity as function of the momentum transfer (Q), which is a function of scattering angle (2θ); Q= ($4\pi \sin \theta/\lambda$).

4.2.3.2 SAXS data processing and analysis

All the data manipulations and processing are carried out using the PRIMUS software (156). The radius of gyration R_g, which is the root mean square of the distances of all electrons from the centre of gravity of the particle, can be estimated by the Guinier approximation(157). The Guinier plot is most sensitive to the longest distances present in the sample and therefore it can be used to identify protein aggregation or interparticle interference in the sample and to estimate the extrapolated intensity at zero scattering angle I(0) value (Figure 4.7).

To estimate the distribution of masses within the particle and its shape, the pair distribution P(r) function is calculated from the scattering data using GNOM (158). The real space R_g and D_{max} (maximum dimension of scattering particle) are estimated from the P(r) and then used for the calculation of the Porod volume (V_p) (Figure 4-8). Sample molecular weight (MW) can be estimated from the SAXS data by various methods, particularly using the SAXSMoW program (159) that estimates the molecular mass through a calculation of volume times the average protein density of 0.83 x 10^{-3} kDa A^{-3}.

4.2.3.3 Atomic structure validation

One of the most popular uses for SAXS data is structure validation. The program CRYSOL is used to calculate the theoretical scattering amplitudes of proteins from their atomic coordinates (obtained from crystallography, NMR or homology modelling), taking into account the hydration shell and excluded solvent volume (160). The similarity between the model and experimental SAXS data can be quantified using the χ^2 values between the theoretical and experimental patterns;

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^{N} \left[\frac{I_{exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)} \right]^2 \qquad (4.3)$$

where N is the number of points, $I_{exp}(s_j)$ and $I_{calc}(s_j)$ are the experimental and calculated intensities, c is the scaling factor, and σ denotes the experimental errors.

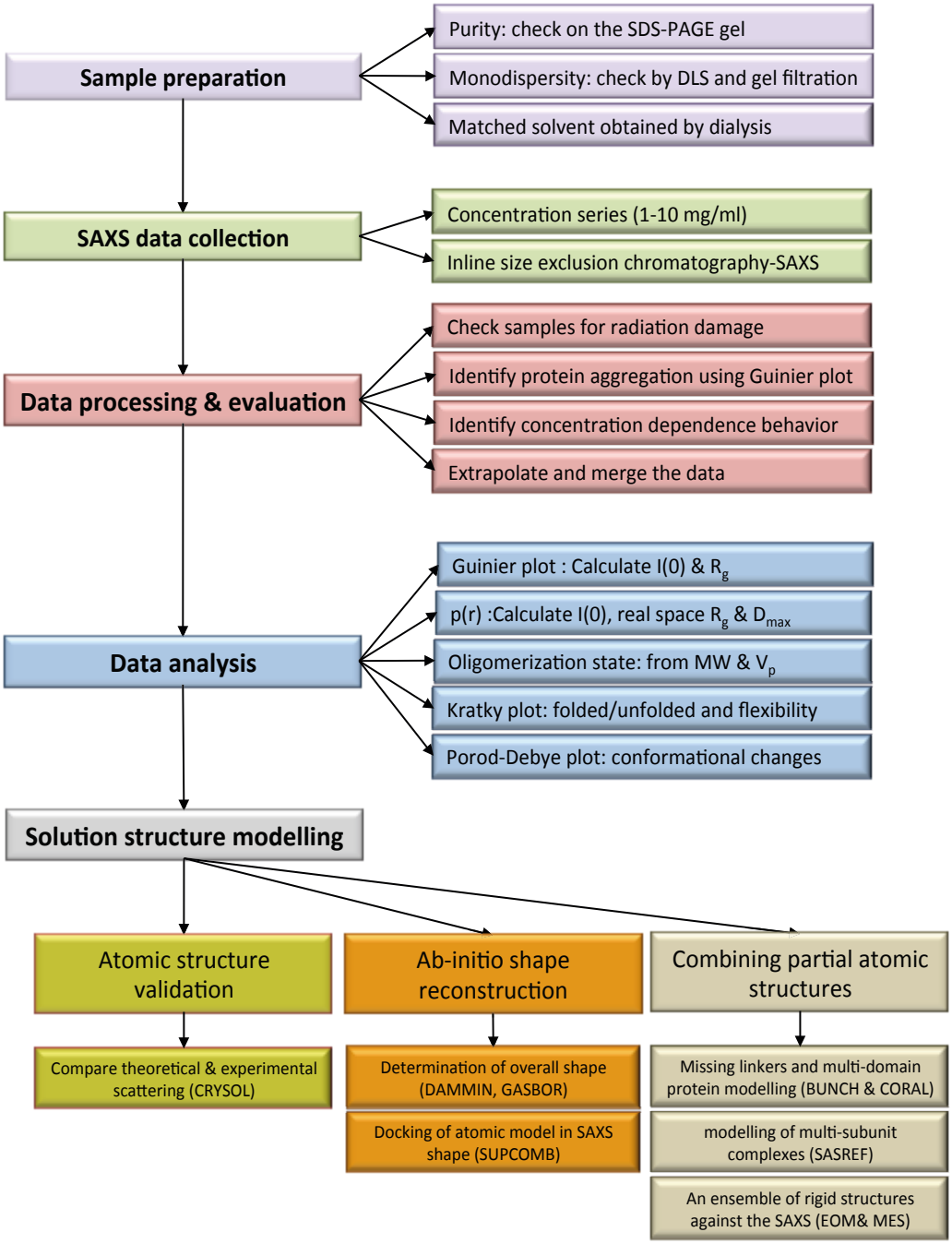


Figure 4.7 Schematic for SAXS sample preparation, data collection, evaluation, analysis and modelling.

4.2.3.4 Ab-initio shape reconstruction

One of the major advantages of SAXS is the ability to reconstruct molecular shape from the SAXS data alone, without any prior information. *Ab initio* shape reconstruction can be performed using simulated annealing optimization of a dummy atom set, as implemented in the programs DAMMIF and DAMMIN (161, 162). DAMMIF represents the particle as random configuration starting from a spherical assembly of densely packed beads (dummy atoms), of volume compatible with the experimentally determined R_g, and theoretical scattering amplitudes are calculated. Simulated annealing is employed to find a ''physically sound'' model (i.e. compact and interconnected) that best fit the experimental pattern while minimizing the discrepancy χ^2. The results of independent DAMMIF runs can be compared using the structural alignment program SUPCOMB (163) which minimizes the spatial discrepancy (NSD) to find the best alignment between the models.

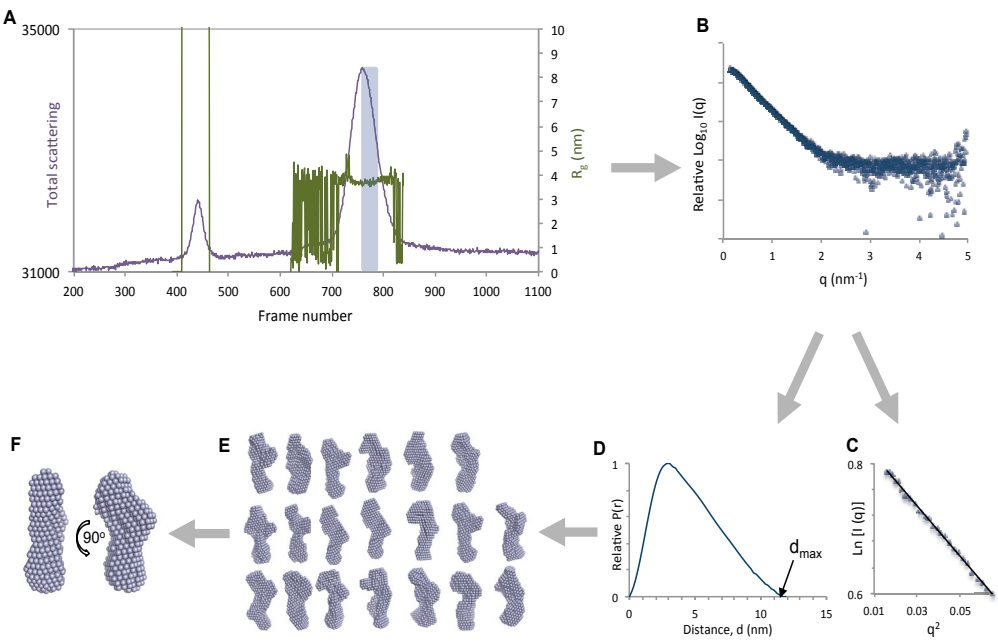


Figure 4.8 SEC-SAXS run of GPC1-dHS. (A) Total scattering intensity (violet) and the calculated R_g (green) of the frames collected from the eluted GPC1-dHS protein from Superdex-200 column. The scattering images of stable R_g (shown as grey shadow) were merged together, buffer subtracted to produce final SAXS pattern showed at (B). The SAXS derived Guinier plot (C) and normalized P(r) (D) are subsequently calculated. (E) Gallery of the individual DAMMIN reconstruction of GPC1-dHS. (F) The average *ab-initio* model of GPC1-dHS calculated by DAMVER from the 20 individual models.

The mean NSD between the independent DAMMIF reconstructions provides a useful estimate of the reliability of the models. Subsequently, the different DAMMIF models can be averaged using the program suite DAMAVER (164), by aligning the *ab initio* models, selecting the most probable ones and building the averaged model. DAMFILT can be used as a part of the DAMAVER suite to filtered out the low bead occupancy positions and loosely connected atoms from the averaged model, generating the most representative/typical model that agrees with the particle excluded volume. In some cases in our work, the damstart model, generated from the averaged model by DAMMAVER suite, was used as initial model for another slow mode DAMMIN run (162) (Figure 4.8).

To test the dependence of the SAXS shape reconstruction on the used algorithms, the program GASBOR (165) can also be used to compute a set 20 different ab initio shape envelopes using the GNOM output (having the same D_{max} and R_g), which have been used with DAMMIF & DAMMIN. The GASBOR algorithm uses a simulated annealing method to search for a chain-compatible spatial distribution of dummy residues (number of residues C_α corresponding to the protein molecular weight) that fit the experimental scattering data without constraining them to sit in a fixed lattice.

4.2.3.5 SAXS molecular modelling

SAXS-based molecular modelling of protein complexes or multidomain proteins that have component(s) lacking some fragment(s) (e.g. interdomain linkers or termini regions are missing), is conducted using rigid body modelling as implemented in the program CORAL (COmplexes with RAndom Loops) (155), an advanced version of the rigid body-modelling program BUNCH (166). CORAL represents the linker/missing loops as interconnected dummy residue chains attached to the appropriate C_α atoms in the rigid domain, where symmetry/contacts could be taken into account (155). A simulated annealing protocol is employed to locate the most favourable positions and orientations of available domain atomic models and the approximate conformations of the missing portions of polypeptide chain(s). Again, confirmation of theses missed regions was conducted by minimizing the discrepancy χ^2 between the scattering curves calculated from the model and the experimental data.

4.2.3.6 Characterization of flexible systems

As SAXS data are time- and ensemble-averaged for all the protein conformations in the sample, it is possible to measure data for flexible systems. Flexible structures can be presented as an ensemble containing a number of different conformations of the flexible parts, obtained by the Ensemble Optimization Method (EOM) (167). A large pool of random independent conformers that theoretically covers all possible configurational space is generated by the RANCH program. CRYSOL is subsequently used to calculate the theoretical scattering patterns and then a genetic

algorithm (GAJOE program) is employed to select the ensemble having the average scattering pattern that best matches the experimental data. GAJOE was used for ensemble selection data by minimizing the discrepancy χ^2. A flexibility estimate is achieved by comparing the R_g distribution of the random pool to the R_g distribution obtained from the ensemble possible conformation s selected by EOM.

Chapter 5

Results and general discussion

We aimed in this thesis to investigate the three dimensional structure of the GPC1 core protein and its spatial relationship to the cell membrane and to determine the mechanism behind the HS assembly. In paper I we have expressed and purified the human N-glycosylated GPC1 core protein and succeeded in crystallizing and solving its crystal structure. In paper II we piloted a follow-up study in which we thoroughly <explored all the parameters necessary for improvement of the GPC1 crystal diffraction by controlled dehydration. In Paper III we have coupled the knowledge from X-ray crystallography with SAXS and other biophysical techniques to provide valuable insights into the characteristics of GPC1 protein in solution. In paper IV we expressed, purified, and preliminarily characterized the EXTL3 protein. Lastly, we carried out preliminary characterization of the GPC1-EXTL3 complexes using biochemical methods.

5.1 Paper I

Based on amino acid sequence alignment, GPC1 has been predicted to consist of a globular domain and a GAG attachment domain. Previous studies have demonstrated that the human GPC1 core protein has a predominantly α-helical single domain fold, with identical CD spectra for the core protein with and without the GAG chains, indicating that the HS chains do not influence the folding and the structure of the core protein (138). The GPC1 core protein contains two N-glycosylation sites at Asn-79 & Asn-116, which are found to be invariably occupied (115). The N-linked glycans at these sites affect GPC1 protein expression and heparan sulphate substitution. Nevertheless the protein is folded correctly even in the absence of N-linked glycans (115). There was no GPC core protein structure available when we started this study, although one was published during our work.

In this work, anchorless His6-tagged full-length human N-glycosylated GPC1 without HS chains (GPC1-dHS) and C-terminally truncated GPC1 (GPC1-dC) were expressed in human embryonic kidney cells, purified and crystallized. GPC1-dC

crystallized in space group $P2_1$, with typical unit cell dimensions of a = 47.2, b = 168.7, c= 147.8 Å, β= 94.6°. Variation in cell dimensions was noticed between different crystals. In particular the length of the c axis varied between 147–154 Å. The GPC1 crystals were delicate, highly fragile plates of typical dimensions around 0.8 mm x 0.3 mm x 0.05 mm, with a solvent content of 54%. The asymmetric unit of GPC1 crystals contained four chains in two pairs, A/B and C/D, related by translational pseudosymmetry.

The structure of the GPC1 core protein, the first crystal structure of a vertebrate glypican, was determined by the two-wavelength MAD method on a selenomethionine-substituted protein crystal (168). Automatic phasing and model building was done via the AutoRickshaw pipeline (169) and the model refinement was done using Refmac5 (170) and phenix.refine (171) against the 2.6 Å native data set. All data sets were collected at station I911-3 of the MAX IV Laboratory, Lund, Sweden. All suffered from significant anisotropy, with much higher Wilson B-factors in the c* direction than in the other two.

The GPC1 structure revealed a highly extended cylindrical, all α-helical fold with dimensions of 120 Å x 30 Å x 30 Å (Figure 5.1). Its structural similarity to the Dally-like protein from *Drosophila* (172) confirmed a conserved overall fold for the glypican family. The GPC1 structure consists of 14 α-helices (α1- α14) and three major loops (L1-L3). Sequence analysis showed that the major loops are evolutionarily conserved in vertebrate GPC1, and at least one of them (L3) may be involved in glycosaminoglycan class determination. The extended helix α2 (83 Å) traverses the whole protein, carrying two N-linked glycans close to its ends. The GPC1 structure revealed the complete arrangement of the 14 Cys residues conserved across the GPC family, in 7 disulfide bonds, 6 of them located close to the molecule's N terminus in a region termed the "Cys-rich lobe'". This lobe is followed by a region that forms the heart of the structure, called the "central lobe". This lobe is stabilized by evolutionarily conserved hydrophobic centres. The last region of the GPC1 molecule is termed the "protease site lobe" because of the presence of a protease site in this part.

No additional electron density was observed from the crystals of GPC1 containing the HS attachment region at the C-terminus when compared to the truncated protein, which demonstrates the flexibility and/or disorder of this part. Thereby, this long, flexible C-terminus (53 residues) could give the core protein a freedom in its orientation when GPC1 is anchored to the cell membrane. Other experimental approaches are required to answer the important questions about the placement of the HS attachment region relative to the glypican core protein and how these proteoglycans are situated on the cell membrane.

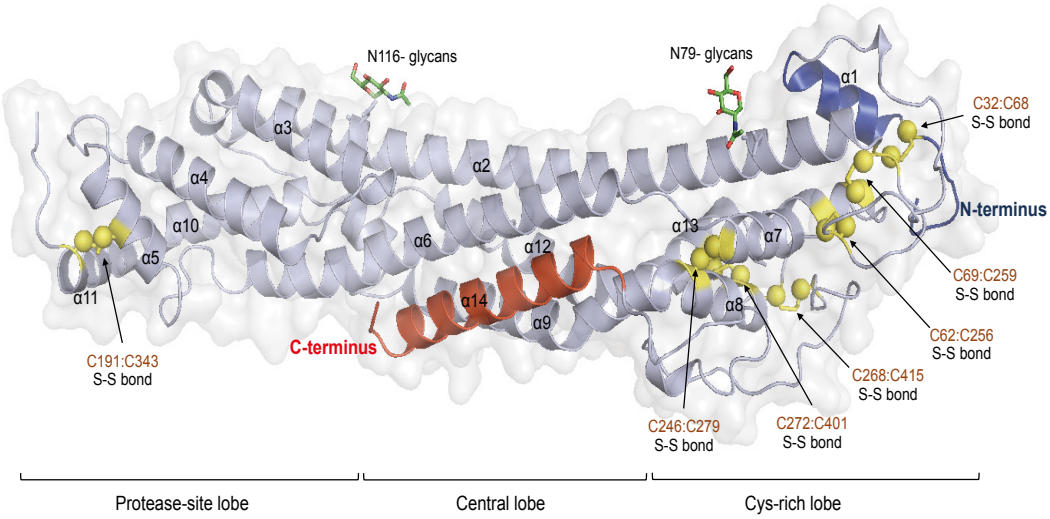


Figure 5.1. Crystal structure of the N-glycosylated human GPC1 core protein (PDB entry: 4ACR); cartoon diagram of GPC1 in which the whole body of the structure is coloured light blue, the N-terminal helix (dark blue) and the C-terminal helix (red) with the loops (L1:L3) and alpha-helices (α1:α14) are exhibited. The seven disulfide bonds are indicated in yellow and the assignment of different lobes in the GPC1 structure is displayed.

Paper I: Conclusions
➢ The GPC1 structure reveals a cylindrical (dimensions 120 x 30 x 30 Å), all α-helical fold for the glypican core protein, with three major loops. ➢ The loops are evolutionarily conserved in vertebrate GPC1 and at least one of them may be involved in glycosaminoglycan class determination. ➢ No additional electron density was observed from crystals of GPC1 containing the HS attachment region at the C-terminus. ➢ Other experimental approaches than X-ray crystallography are needed to answer the important questions about the placement of the C-terminal HS attachment region, and how GPC1 is situated on the membrane.

5.2 Paper II

GPC1 crystals were delicate highly fragile plates that displayed poor isomorphism, with varying cell dimensions between different crystals of the same type; in particular the length of the c axis varied between 147–155Å (115). Furthermore, they diffracted anisotropically, reflected in a Wilson B factor that was twice as large in the c* direction as in the a* and b* directions, which limited the effective resolution to 2.9 Å in the c* direction for GPC1-dC crystals. Refinement against ellipsoidally truncated data produced from the UCLA Anisotropy server (http://services.mbi.ucla.edu/anisoscale) (173) did not improve either the electron density map or the model.

The anisotropy observed in GPC1 crystals is largely a result of intrinsic disorder in the lattice packing, where there are more lattice contacts in a and b directions than in the c direction. In the a and b directions strongly packed layers are formed through heterotypic contacts between the chains in the A/B and C/D pairs and homotypic interactions between identical chains in adjacent cells related by translation along the a axis (total area 2121 $Å^2$). In contrast, the layers are held together by a single interaction in the c direction between chains B and C related by a pure x=1 translation, of which the buried surface area is only 275 $Å^2$. Moreover, the diffraction data show significant translational non-crystallographic symmetry, with a peak at (0.5, 0.021, 0.5) in the native Patterson of about 35% of the height of the original peak. This relates chains A and B to C and D in the asymmetric unit, respectively.

In this work we have shown GPC1 crystals to be a successful case for improvement in diffraction properties by controlled crystal dehydration using the HC1b device, which delivers an air stream of a precise relative humidity (RH) that can be used to alter the solvent content inside the crystal. The optimal dehydration protocol was developed by investigation of the following parameters: the final relative humidity (RH_f), the dehydration rate and the total incubation time (T_{inc}) as shown in Figure 4.6. Through the investigation of the optimal final humidity, the diffraction quality of the GPC1 crystals was clearly enhanced, with significant and reproducible improvements in the Wilson B factor and the diffraction isotropy. Whilst it is very difficult to deconvolute the effects of incubation time and dehydration rate, as they are to a certain extent correlated, we have demonstrated that the total incubation time is at least a factor to be considered. Notably, too extended incubation times turned out to be detrimental to crystal quality.

The marked improvement in the GPC1 diffraction after the dehydration generates better, less noisy electron density maps. The new density maps allowed the building of more complete models for all the GPC1 monomers in the asymmetric unit, where 60 residues from the protease site lobe become ordered after the dehydration, forming two helices and one loop. Furthermore, the new maps displayed well-defined side

chain density, allowing more reliable side chain placement and facilitated the location of more water and solvent molecules (Figure 5.2). The HC1b controlled dehydration method could be generally useful for structurally anisotropic proteins such as membrane proteins and protein-protein complexes, which have deficient crystal contacts in some directions.

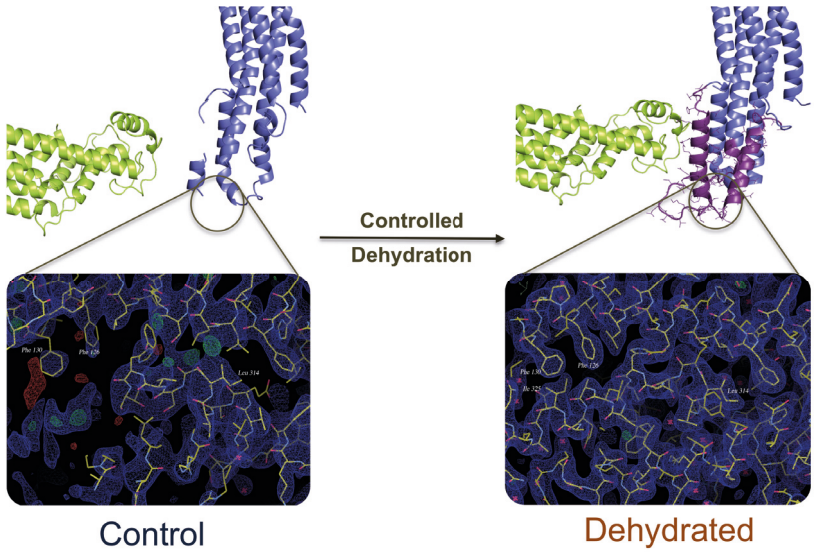


Figure 5.2. Crystal packing of chains A (slate blue) & C (lemon green) of control (left panel) and dehydrated models (right panel). The protease-site lobe (purple-violet) in chain A (60 residues) was ordered after dehydration, with a better, less noisy electron density map (contoured at 1σ).

Paper II: Conclusions
➢ The GPC1 unit cell shrank after HC1b controlled dehydration, which was reflected in the crystal packing and the model quality. ➢ The diffraction quality of the GPC1 crystals was clearly improved after dehydration, with significant reduction in both Wilson B-factor & anisotropy. ➢ This generated improved electron density maps, which allowed the building of previously disordered parts of the model and displayed better-defined side chains. ➢ The total incubation time is an important parameter to be considered in the optimization of HC1b protocol. ➢ HC1 crystal controlled dehydration may be useful for structurally anisotropic proteins such as membrane proteins and protein-protein complexes.

5.3 Paper III

In Papers I & II, we have determined the N-glycosylated GPC1 core protein crystal structure and have shown the truncated GPC1 crystals to be a successful case for quality improvement by controlled dehydration (168, 174). As is normal for glycosylated proteins, the GPC1 N-glycan chains are not fully resolved in the crystal structure, due to innate heterogeneity and flexibility. Our MS analyses have revealed highly heterogeneous complex-type glycoforms for GPC1, but it is not clear whether this heterogeneity has a functional impact. The prominent structural roles of the Asn-79-glycans were proposed to extend the protein stability and lifetime on the membrane by the protection against proteolysis and aggregation, whereas the Asn-116-glycans seem likely to have more functional impacts. Overall, the N-linked glycans assumed to influence the geometry of the GPC1 protein packing and subsequently impact its distribution on the cell surface.

Unfortunately, as noted above, the 3 Å full-length GPC1 (GPC1-dHS) crystal structure (PDB entry 4AD7) (168) contains no electron density for the C-terminal domain (approximately 53 residues) that attaches the core protein to the cell membrane. Because of the disorder of the C-terminus in the crystal structure, it was difficult to predict how distant the core protein and the HS sites might be from the cell membrane surface. Furthermore, the extended structure of the GPC1 core protein increase the difficulty of predicting how the glypicans are oriented near the membrane and which part(s) of the core protein face the cell surface.

Here, we have shown that the GPC1 C-terminal domain still lacks significant tertiary structure and is disordered in the GPC1-dHS crystals, even after controlled crystal dehydration, which otherwise significantly improved the crystalline order, resolution (2.3 Å), Wilson B-factor and diffraction isotropy. Furthermore, bioinformatics tools predict a highly disordered structure for the C-terminus. In view of this, we coupled our crystallographic knowledge on GPC1 with SAXS to elucidate the spatial relationships between the cell membrane, glypican, and their N- and O-linked glycans. It seems likely that the C-terminus is highly flexible in the solution, placing the core protein at an approximate distance of ~44–53 Å from the cell membrane, including the GPI anchorage. We propose that GPC1 “lies down” in a transverse orientation to the cell surface, where the GPC1-orthologues’ evolutionarily conserved plane is exposed to the membrane, serving both in targeting & linking the GPC1 protein to the consensus lipid raft and in GAG substitution. In this orientation the protease site lobe is in closer proximity to the lipid bilayer than the other lobes (Figure 5.3).

In future work it could be of importance to precisely determine how the GPC1 core protein regulates the HS class determination and assembly. This could be accomplished by searching for the GPC1 interacting partner from the EXT-family

and in particular to study its affinity and interaction with the EXTL3 as HS-initiator enzyme. Furthermore, systematic mutagenesis studies of the evolutionarily conserved structural elements facing the membrane would be important to map more distinctly the residues of the GPC1 core protein that might be involved in the preferential HS assembly.

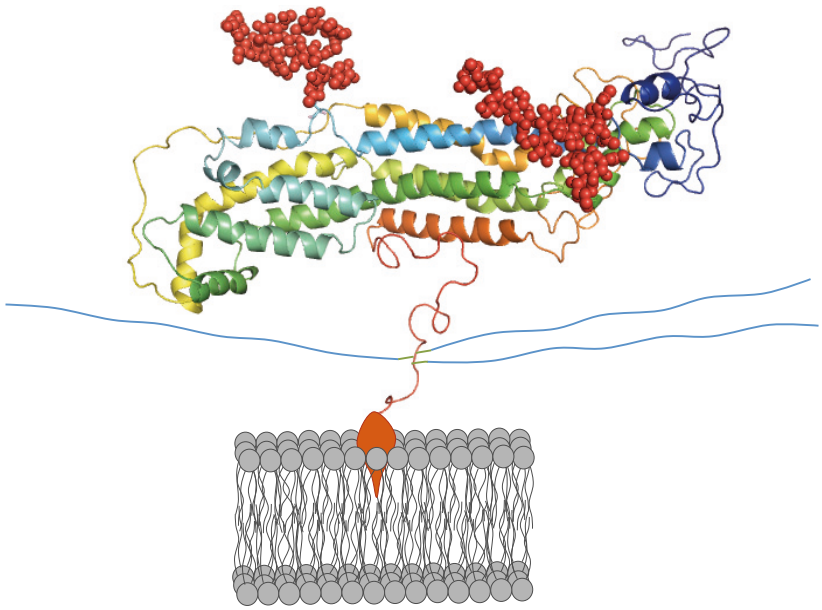


Figure 5.3. Topology of the GPC1 on the membrane. The GPC1 structure is represented in rainbow colors (blue N-terminus, red C-terminus) with N-glycans as red spheres and the HS chains as blue long chains. The membrane is shown as a grey lipid bilayer with orange GPI anchor connected to the GPC1 core protein.

Paper III: Conclusions
➢ We report the highest resolution crystal structure of a GPC family member to date. ➢ GPC1 carries complex type N-glycans that extend the protein stability and lifetime on the membrane by protection against proteolysis and aggregation. ➢ The GPC1 core protein is placed at an approximate distance of ~44–53 Å, including the GPI anchor, from the membrane and the evolutionary conserved GPC1 surface is oriented towards the cell surface. ➢ A systematic mutagenesis study could be important to map more precisely the GPC1 surface exposed residues that regulate the GAG class determination and assembly.

5.4 Paper IV

Heparan sulphate assembly on the HSPG, like glypicans, is performed at the Golgi apparatus by glycotransferases of EXT family (section 1.2). EXTL3 is supposed to initiate the process by transferring the GlcNAc to the tetrasaccharide common core linked to the target serine (GlcNAc-TI activity). Then the HS chains are further elongated by the addition of the repeating disaccharides via EXT1 / EXT2 polymerases enzyme (GlcNAc-TII activity)(16). Recent studies suggested that EXTL2 worked as a terminator of the HS elongation process (25), whereas the significance of EXTL1 remains unclear. EXTl3 is distinctly bifunctional enzyme with both GlcNAc-TI and GlcNAc-TII activities and accordingly, EXTL3 provides an excellent model for obtaining a complete view of the structure–based mechanism of glycotransferase activities catalysed by the EXT protein family. In this investigation, the human lumenal N-glycosylated EXTL3 protein (EXTL3ΔN) was cloned and expressed in human embryonic kidney cells and further purified using Ni-NTA affinity and gel filtration. Correct protein expression was confirmed by Western plot and MS analysis and subsequently the purified proteins were investigated using various biophysical-biochemical approaches to clarify the N-glycosylation sites and function, and the conformation and stability of the enzyme in the solution.

EXTL3 sequence reveals four possible N-glycosylation sites at Asn-277, Asn-290, Asn-592 and Asn-790. LC-MS/MS data showed that Asn-277 is non-glycosylated, whereas Asn-290 and Asn-592 are variably and invariably occupied with N-glycans respectively (Figure 5.4A). No data confirms the absence or presence of N-glycosylation on Asn-790. There is severely impaired secretion of EXTL3ΔN devoid of N-glycosylation via tunicamycin treatment, which prevent N-glycosylation, of the EXTL3ΔN producing cells. These indicate a quality control effect of the N-glycans chains on the EXTL3ΔN biosynthesis in the cells.

The thermal stability of EXTL3ΔN was examined by CD, DLS and DSF techniques, which revealed a quite stable protein fold, as high temperatures (~59 ºC) are required for denaturation. Further, the results from the heat denaturation of EXTL3ΔN in the presence of reducing agents (dTT) suggesting destabilization effect that indicate the presence of some disulphide bonds, as the lumenal EXTL3ΔN contain orthologous evolutionary conserved 16 Cys residues, 7 of them are conserved between all the EXT proteins. Careful analyses of the EXTL3ΔN far-UV CD spectrum (Figure 5.4B) indicated substantial fraction of β sheets structures (25%), with 14%, 23% and 35% for α-helical conformations, turns and unordered structures respectively.

Bioinformatics investigation showed that the EXTL3ΔN contain at least two functional domains performing the enzyme activities: GlcNAc-TII activity by large GT47 domain and GlcNAc-TI activity by smaller GT64 domain. Solution structural studies using SAXS and DLS revealed extended monomeric structure of two

distinctive regions, a narrow and a broad one (Figure 5.4C). The homology model of the EXTL3 small GT64 domain has similar dimensions to the narrow region suggesting its localization there, whereas the broad region seems likely to contain the GT47 domain. To get insight the detailed structure of the catalytic EXTL3ΔN, crystallization screening was performed, but unfortunately no mountable crystals have been obtained yet. Further experiments, such as His-tag cleavage and N-glycans removal are suggested to get well-diffracted EXTL3ΔN crystals.

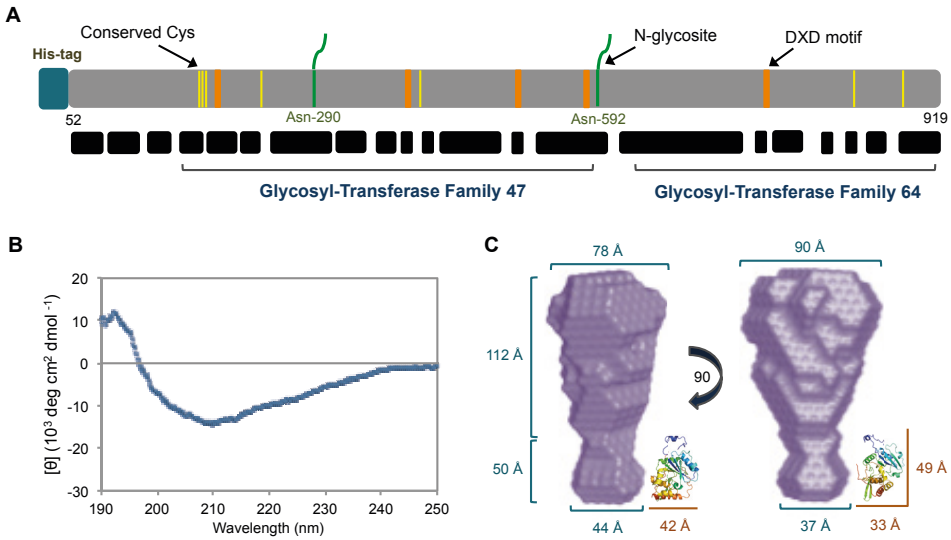


Figure 5.4. Biophysical characterization of EXTL3ΔN in the solution. (A) Schematic presentation of EXTL3ΔN sequence with annotation of functional domains, DXD motifs, conserved cysteines and the two found N-glycosites. In addition, the MS peptide mapping is shown as black boxes. (B) The far-UV CD spectrum collected at 25 °C is displayed as blue curve. (C) SAXS based ab-initio model of EXTL3ΔN is shown as surface volume with the GT64 domain homology model as an insert.

In Paper IV: Conclusions
➢ We cloned, expressed and purified the EXTL3ΔN for the first time.
➢ EXTL3ΔN protein is occupied with N-glycans at least at two positions, Asn-290 and Asn-592, which seem critical for proper protein expression.
➢ EXTL3ΔN is stable protein, which displayed substantial fraction of β sheets structures.
➢ Structural studies of EXTL3ΔN revealed extended monomeric structure of narrow and broad regions.
➢ Initial crystallization screening was performed and additional experiments are suggested to get well-diffracted EXTL3ΔN crystals

5.5 Closing remarks and future directions

In this book, I have introduced my humble contribution to the biochemical society on glypican proteins and their related enzymes, like EXTL3. We aimed to characterize structurally and functionally the GPC1 core protein and its overall topology with respect to the cell surface and to gain insights into the mechanism behind HS assembly on the core proteins. Indeed, we succeeded to report the first and highest resolution (2.3 Å) crystal structures of human GPC family members to date. The story was initialized with the successful determination of the N-glycosylated GPC1 core protein crystal structure by the two-wavelength MAD method on a SeMet-substituted protein crystal, which uncovers a quite rigid, elongated single-domain all α-helical fold (α1- α14) with three major loops (L1, L2 and L3) structure (Paper I). The long α2 helix of the GPC1 structure traversing the entire length of the protein carries two N-glycan chains, one at each end. The complete disulphide pattern of the 14 conserved Cys residues across the GPC family was also revealed in the structure.

Afterwards, we achieved improvements of GPC1 crystal diffraction properties by controlled crystal dehydration using the HC1b machine that is installed at the MAX IV Laboratory. This enhanced diffraction generates better electron density for the GPC1 crystals, allowing the building of previously disordered parts of the structure (Paper II). The optimized dehydration protocol was developed by exploring all the protocol parameters and we found that total incubation time is an important parameter to be considered during the dehydration experiments.

Unfortunately, despite dehydration the full-length GPC1 structure contains no electron density for the C-terminal region, ~53 residues attach the core protein to the cell membrane. In view of this, we employed SAXS and other biophysical methods to elucidate the solution structural characteristics of the GPC1 protein (Paper III). The C-terminus was highly flexible in the solution, laying the core protein transverse to the cell surface at approximate distance of ~44-53 Å, taking into consideration the GPI anchor. This orientation assemble the GPC1-orthologues' evolutionarily conserved plane directed to the membrane, where it might interact with the EXT enzymes that are involved in HS substitution in the Golgi apparatus. The GPC1 N-glycan chains are not fully resolved in the crystal structure, due to innate heterogeneity and flexibility. The N-glycan chromatographic analyses revealed highly heterogeneous complex-type glycoforms for GPC1, but it is not clear whether this heterogeneity has a functional impact. Moreover, we have shown that these N-glycan oligosaccharides extend the GPC1 protein stability by protecting both of aggregation- and degradation-prone regions of the protein molecule.

The HS chains are the responsible for many of the biological functions of the GPC proteins, as they are able to interact with a wide variety of biological molecules due to

their structural diversity. To explore the mechanism of HS assembly on GPC, among other HSPG proteins families, it is very important to start working on the EXT-family proteins that inaugurate and polymerize the HS biosynthesis. All published reports so far suggested EXTL3 to be the initiator for the HS synthesis. Therefore, we expressed, purified and further structurally characterized the human lumenal EXTL3ΔN catalytic protein using various biophysical techniques (Paper IV). Our data show that EXTL3ΔN is occupied with N-glycans at least at two positions, Asn-290 and Asn-592 that seem to be vital for appropriate EXTL3ΔN biosynthesis. EXTL3ΔN was shown to be a quite stable protein with a majority of β sheets. SAXS low-resolution analysis revealed extended monomeric structure with two distinctive regions, narrow and broad respectively.

Detailed EXTL3ΔN structural information would be valuable in elucidating the structural based mechanism of glycotransferase activities (GlcNAc-TI and -TII) catalysed by the EXT protein family. Therefore, crystallization screening for the purified EXTL3ΔN proteins was performed, but sadly no diffracting crystals have been produced right now. To obtain well-diffracting EXTL3ΔN crystals; N-glycans enzymatic removal, His-tag cleavage, and/or further crystallization methods (such as micro-batch and free-interface diffusion) are suggested to improve the initial crystallization conditions. These experiments are already ongoing. Afterwards, we are compelled to solve the EXTL3ΔN structure by experimental phasing, as there is no known structural homologue for the whole EXTL3ΔN protein. For that experiments, heavy metals soaking and co-crystallization could be inspected, otherwise selenomethionine should be incorporated into the proteins as we carried out for the GPC1 protein. EXTL3ΔN contains 19 methionine residues, which should give a sufficient signal for MAD phasing and thereby structure determination.

A crucial question would be how the GPC core proteins regulate the GAG class determination and assembly. To provide some insights into that issue, we performed a preliminary examination for the affinity of the EXTL3ΔN with both of the C-terminally truncated GPC1-dC and the full-length GPC1-dHS proteins using solid phase assays (175). In this assay, the EXTL3ΔN protein was absorbed to the wells of ELISA plate, then the plate was blocked with bovine serum albumin and various concentrations of GPC1 mutants were then added. After careful washing, anti-rabbit GPC1 antibody was added to the wells followed by calorimetric detection steps. These preliminary experiments showed that GPC1-dC and GPC1-dHS proteins have similar affinity to the EXTL3ΔN protein, which indicates a significant effect of the N-terminal core protein on the binding rather than the C-terminus HS-attached region. More experiments are planned to further investigate the thermodynamics and kinetics of the GPC1-EXTL3ΔN complexes by different techniques like surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) (176, 177). Structural analysis of the GPC1-EXTL3ΔN complexes using X-ray crystallography or even by low-resolution structural approaches of mass spectrometry cross-linking and

SAXS would be of great assistance to specify the interacting interfaces of the two proteins. This will give insights into the mechanism of action of glycosyl transferases on the GPCs proteins. In addition, systematic mutagenesis of these structural elements might be significant to map more accurately the residues of the GPC1 core protein that are involved in the preferential HS assembly.

Mammalian GPC family contained six members, which all share 14 evolutionary conserved cysteines, and hence expected to have similar fold. From an evolutionary perspective, vertebrate GPCs are classified into two subfamilies: GPC1, -2, -4 & -6 and GPC3 & 5, with approximately 25% sequence identity between the two groups. We have solved the crystal structure of the GPC1, which belongs to the first subfamily then we tried to express the GPC3 protein, as a member of another subfamily, for further structural investigation to inspect whether all GPCs have identical fold. Furthermore, GPC3 is an emerging therapeutic target for tumour antigen-specific immunotherapy against liver cancer (178), so its structure will be invaluable for the anti-GPC3 antibody design studies(179). Initially we had aggregation problems for GPC3 after expression in human embryonic kidney cells, but recently we achieved solubility improvement by incorporating solubility enhancement peptide tags at the protein termini and now protein characterization experiments are in progress.

Acknowledgments

First and foremost, all praise and gratefulness are to ALLAH for the kindness, strengths and inspiration in lending me to accomplish this thesis and many blessings are upon our prophet Muhammad, who said ''A person who is not thankful to his benefactor is not thankful to ALLAH''. I have learnt more things than I could ever imagine through out these four and half years I stayed in Lund. Many people helped me and it would be impossible to mention all of them, I would like to express my sincere gratitude to all of you. It is pleasure to express special thanks to the following people:

Derek, my supervisor, for accepting me as PhD student in your lab, introducing me to the world of structural biology and to give me the opportunity to work on different projects. You are a truly inspiring and generous person. I am appreciating your creativity and the scientific freedom you give me in the lab. You never say 'no' for any request regarding the work. Thanks for always replying to my endless questions about crystallography, carefully revising my thesis, and for all the opportunities of travelling around the world; it has been a true privilege.

Katrin, my co-supervisor, for a fruitful collaboration, for your continuous support, and for your availability most of the time to talk about the project suggesting pretty ideas. You are a true kind and helpful scientist. Thank you for proofreading of my thesis. This project was impossible without your assistance in tissue culture and taking care for the cells.

Gabriel, for guiding me around the lab in the beginning of my studies, for being great collaborator on the glypican project and for long hours spent at the synchrotrons. Without your hard work on the GPC1 crystal optimization, the first structure would not have appeared. *Saraboji*, former postdoc in the lab, for advising me on the protein crystallization and for fabulous collaboration on the SOD project.

Oskar, for all the chats about our projects & science in general and for accompany in the KEMM15 teaching. I will not forget our fantastic trips around the world; England, Italy, France, Egypt, Denmark and Canada enjoying the science and travel companionship. I will be in contact with my best Swedish awesome friend and I am sure we will meet again there ☺.

Mostafa, my colleague at Cairo and Lund Universities. You made a difference in my life after you came to Lund, as I found the guy who I could gossip with about everything like Egyptian news, family and work, among others. I wish you all the best with your thesis.

*Marjolein***,** for nice collaboration and assistance with the HC1 machine and to always help to allocate beam times at MAX-lab. *Thomas* and *Roberto* for fixing problems at the beam. I would like to thank all the people at the crystallization facility, *Maria*, *Mattias and Raymond*, for help

in setting up crystallization screens. *Salam* and *Anders* for the fantastic courses you taught me and for accepting the invitation to visit Cairo University arranging the structural bioinformatics course there. Indeed, it was a fantastic time.

I further thank all of former and current group members for having group meeting and talking about science weekly. *Renzo*, thank you for the doda file, it was very beneficial in the data processing. *Francesco* and *Rohit***,** I hope you have lactose out of the crystals soon and good luck with your sweet project ☺. *Tejas*, Good luck in your new career. I would like to thank *Mahmudul* and all the people from Salam's group; *Eva-Christina, Janina, Abdo, Abdelbaset, Ali* and *Wessen* for sharing stuff and for giving company at the department and during synchrotron visits. Many thanks also to the people at Katrin's lab at BMC; *Sol* for great assistance in the cloning work; *Fang* for help in the Western blotting and *Andrea* & *Lars-Åke* for fabulous discussion about the glypicans.

I enjoyed all the retreats, breakfast seminars and mini-symposia organized by the CMPS. It is a fantastic place to work at and all the people here are very helpful, thank you all. A special thanks to: *Sven* for assistance with the mass spectrometry; *Robert* for great help with the AUC runs; *Tommy* for introducing the CD to me; *Henry* for borrow me the His-tag antibody as it was very useful; *Yusra* for nice collaboration and discussions; *Susanna, Adine, Gert* & *Magnus* for administrative and technical aids.

I very much appreciate the collaboration with *Barbara* and *Niclas* (Gothenburg University) on the N-glycans analyses. Many thanks for *Mikael Oliveberg* and *Jens Danielsson* (Stockholm University) for our *PNAS* article. I am grateful for the collaboration with the LP3 facility, for insect expression of GPC4 by *Wolfgang* and cloning of GPC3 by *Annika*. Thank you *Claes* for nice collaboration on the REX proteins. I thank the staff at ESRF SAXS beamlines (*Adam Round* and *M. Brennich*) for help with SAXS data collection.

I am very glad to have many Egyptian friends in Lund and appreciating their support, social gathering and memorably barbecue in the gardens. I will not forget *T. Selim, M Abdella, A Fawzy, Abdalla, Yasser, Abdelrazek, Wael, T. Deshisha, M. Shamekh, M Essam, Mahmod, M Ismail, Haitham, A. Sofi, M Ibrahim, M. Mabrouk, Ibrahim* and all other friends who are not mentioned here. I wish you all the best.

My former supervisor in Egypt, Prof *M. Mady*, thank you for continuous encouragement and inspiration before and during my PhD studies, I still remember your advice about work and life management. I also acknowledge the support from Prof *A. Abdelazim* and Prof *W. Elshemy* during the short visits to my home.

And last but definitely not least; I dedicated this thesis to my father and mother for sincerely raising me with their care, love, encouragement and prayers before and during my studies. Many thanks to my sister, *Aml* and my brother, *Mahmod*, for taking care of my parents during my absence from Egypt and for always supporting me. Special thanks to my mother-in-law for sincere prayers and support. I would like to thank my dear wife *Sanaa*. I am glad to have awesome inspiring partner like you. Thank you for always smiling ☺, for sacrifice & taking care of our active kids during the course of my study and for accompanying me all the time. I dedicate this thesis to you. My lovely kids, *NourEldin*; *Leen* and *Jana,* you make my life happy. Please forgive me for being quite busy the last few days. I wish for you a great future ahead, full of happiness and achievements.

Wael Awad,
Lund, 20150420

Bibliography

1. Csala M, Kereszturi É, Mandl J, & Bánhegyi G (2011) The Endoplasmic Reticulum As the Extracellular Space Inside the Cell: Role in Protein Folding and Glycosylation. *Antioxidants & Redox Signaling* 16(10):1100-1108.
2. Varki A*, et al.* (2009) *Essentials of Glycobiology.* 2nd Ed.
3. Brown D & Waneck GL (1992) Glycosyl-phosphatidylinositol-anchored membrane proteins. *Journal of the American Society of Nephrology* 3(4):895-906.
4. Paulick MG & Bertozzi CR (2008) The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry* 47(27):6991-7000.
5. Kawagoe K, *et al.* (1996) Glycosylphosphatidylinositol-anchor-deficient mice: implications for clonal dominance of mutant cells in paroxysmal nocturnal hemoglobinuria. *Blood* 87(9):3600-3606.
6. Pike LJ (2006) Rafts defined: a report on the Keystone symposium on lipid rafts and cell function. *Journal of Lipid Research* 47(7):1597-1598.
7. García-Sáez AJ & Schwille P (2010) Stability of lipid domains. *FEBS letters* 584(9):1653-1658.
8. Tannous A, Pisoni GB, Hebert DN, & Molinari M (N-linked sugar-regulated protein folding and quality control in the ER. *Seminars in cell & developmental biology* (0).
9. Stanley P, Schachter H, & Taniguchi N (2009) N-Glycans. *Essentials of Glycobiology*, eds Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, & Etzler ME (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)), 2nd Ed.
10. Couchman JR & Pataki CA (2012) An Introduction to Proteoglycans and Their Localization. *Journal of Histochemistry & Cytochemistry* 60(12):885-897.
11. JD E, K K, & Lindahl U (2009) Proteoglycans and Sulfated Glycosaminoglycans. *Essentials of Glycobiology*, eds Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, & Etzler ME (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)), 2nd Ed.
12. Whitelock JM & Iozzo RV (2005) Heparan Sulfate: A Complex Polymer Charged with Biological Activity. *Chemical Reviews* 105(7):2745-2764.
13. Chen RL & Lander AD (2001) Mechanisms underlying preferential assembly of heparan sulfate on glypican-1. *The Journal of biological chemistry* 276(10):7507-7517.
14. Zhang L, David G, & Esko JD (1995) Repetitive Ser-Gly Sequences Enhance Heparan Sulfate Assembly in Proteoglycans. *Journal of Biological Chemistry* 270(45):27127-27135.

15. Zhang L & Esko JD (1994) Amino acid determinants that drive heparan sulfate assembly in a proteoglycan. *Journal of Biological Chemistry* 269(30):19295-19299.
16. Busse-Wicher M, Wicher KB, & Kusche-Gullberg M (2014) The extostosin family: Proteins with many functions. *Matrix Biology* 35(0):25-33.
17. Zak BM, Crawford BE, & Esko JD (2002) Hereditary multiple exostoses and heparan sulfate polymerization. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1573(3):346-355.
18. Breton C, Bettler E, Joziasse DH, Geremia RA, & Imberty A (1998) Sequence-Function Relationships of Prokaryotic and Eukaryotic Galactosyltransferases. *The Journal of Biochemistry* 123(6):1000-1009.
19. Wuyts W*, et al.* (1998) Mutations in the EXT1 and EXT2 genes in hereditary multiple exostoses. *American Journal of Human Genetics* 62(2):346-354.
20. Busse M*, et al.* (2007) Contribution of EXT1, EXT2, and EXTL3 to heparan sulfate chain elongation. *The Journal of biological chemistry* 282(45):32802-32810.
21. Han C*, et al.* (2004) Distinct and collaborative roles of Drosophila EXT family proteins in morphogen signalling and gradient formation. *Development* 131(7):1563-1575.
22. McCormick C, Duncan G, Goutsos KT, & Tufaro F (2000) The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. *Proceedings of the National Academy of Sciences* 97(2):668-673.
23. Busse M & Kusche-Gullberg M (2003) In Vitro Polymerization of Heparan Sulfate Backbone by the EXT Proteins. *Journal of Biological Chemistry* 278(42):41333-41337.
24. Kim B-T*, et al.* (2001) Human tumor suppressor EXT gene family members EXTL1 and EXTL3 encode α1,4- N-acetylglucosaminyltransferases that likely are involved in heparan sulfate/ heparin biosynthesis. *Proceedings of the National Academy of Sciences* 98(13):7176-7181.
25. Nadanaka S*, et al.* (2013) EXTL2, a Member of the EXT Family of Tumor Suppressors, Controls Glycosaminoglycan Biosynthesis in a Xylose Kinase-dependent Manner. *The Journal of biological chemistry* 288(13):9321-9333.
26. Nadanaka S & Kitagawa H (2013) EXTL2 controls liver regeneration and aortic calcification through xylose kinase-dependent regulation of glycosaminoglycan biosynthesis. *Matrix biology : journal of the International Society for Matrix Biology*.
27. Takahashi I*, et al.* (2009) Important role of heparan sulfate in postnatal islet growth and insulin secretion. *Biochemical and biophysical research communications* 383(1):113-118.
28. Bornemann DJ, Duncan JE, Staatz W, Selleck S, & Warrior R (2004) Abrogation of heparan sulfate synthesis in Drosophila disrupts the Wingless, Hedgehog and Decapentaplegic signaling pathways. *Development* 131(9):1927-1938.
29. Morio H*, et al.* (2003) EXT gene family member rib-2 is essential for embryonic development and heparan sulfate biosynthesis in Caenorhabditis elegans. *Biochemical and biophysical research communications* 301(2):317-323.
30. Lee JS*, et al.* (2004) Axon sorting in the optic tract requires HSPG synthesis by ext2 (dackel) and extl3 (boxer). *Neuron* 44(6):947-960.
31. Kreuger J & Kjellen L (2012) Heparan sulfate biosynthesis: regulation and variability. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 60(12):898-907.
32. Sarrazin S, Lamanna WC, & Esko JD (2011) Heparan Sulfate Proteoglycans. *Cold Spring Harbor Perspectives in Biology* 3(7).
33. Lin X*, et al.* (2000) Disruption of Gastrulation and Heparan Sulfate Biosynthesis in EXT1-Deficient Mice. *Developmental biology* 224(2):299-311.

34. Kitagawa H, *et al.* (2007) Expression of rib-1, a Caenorhabditis elegans Homolog of the Human Tumor Suppressor EXT Genes, Is Indispensable for Heparan Sulfate Synthesis and Embryonic Morphogenesis. *Journal of Biological Chemistry* 282(11):8533-8544.
35. Ahn J, *et al.* (1995) Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1). *Nature genetics* 11(2):137-143.
36. Wuyts W & Van Hul W (2000) Molecular basis of multiple exostoses: mutations in the EXT1 and EXT2 genes. *Human Mutation* 15(3):220-227.
37. Ropero S, *et al.* (2004) Epigenetic loss of the familial tumor-suppressor gene exostosin-1 (EXT1) disrupts heparan sulfate synthesis in cancer cells. *Human molecular genetics* 13(22):2753-2765.
38. Takei Y, Ozawa Y, Sato M, Watanabe A, & Tabata T (2004) Three Drosophila EXT genes shape morphogen gradients through synthesis of heparan sulfate proteoglycans. *Development* 131(1):73-82.
39. Lee J-S, *et al.* (Axon Sorting in the Optic Tract Requires HSPG Synthesis by ext2 (dackel) and extl3 (boxer). *Neuron* 44(6):947-960.
40. Stickens D, *et al.* (1996) The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes. *Nature genetics* 14(1):25-32.
41. Yoneda A, Lendorf ME, Couchman JR, & Multhaupt HAB (2012) Breast and Ovarian Cancers: A Survey and Possible Roles for the Cell Surface Heparan Sulfate Proteoglycans. *Journal of Histochemistry & Cytochemistry* 60(1):9-21.
42. Nadanaka S, Kagiyama S, & Kitagawa H (2013) Roles of EXTL2, a member of the EXT family of tumour suppressors, in liver injury and regeneration processes. *Biochemical Journal* 454(1):133-145.
43. Arai T, *et al.* (1999) EXTL3/EXTR1 alterations in colorectal cancer cell lines. *International journal of oncology* 15(5):915-919.
44. Karibe T, Fukui H, Sekikawa A, Shiratori K, & Fujimori T (2008) EXTL3 promoter methylation down-regulates EXTL3 and heparan sulphate expression in mucinous colorectal cancers. *The Journal of pathology* 216(1):32-42.
45. Couchman JR, Gopal S, Lim HC, Nørgaard S, & Multhaupt HAB (2014) Syndecans: from peripheral coreceptors to mainstream regulators of cell behaviour. *International Journal of Experimental Pathology*:n/a-n/a.
46. Filmus J, Capurro M, & Rast J (2008) Glypicans. *Genome biology* 9(5):224.
47. Fransson LA (2003) Glypicans. *The international journal of biochemistry & cell biology* 35(2):125-129.
48. Schlessinger J, *et al.* (2000) Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Molecular cell* 6(3):743-750.
49. Nahmad M & Lander AD (2011) Spatiotemporal mechanisms of morphogen gradient interpretation. *Curr Opin Genet Dev* 21(6):726-731.
50. Dreyfuss JL, *et al.* (2009) Heparan sulfate proteoglycans: structure, protein interactions and cell signaling. *Anais da Academia Brasileira de Ciências* 81:409-429.
51. Awad W, Logan DT, & Mani K (2014) GPC1 (glypican 1). *Atlas Genet Cytogenet Oncol Haematol* 18(7):461-464.
52. Filmus J (2002) The contribution of in vivo manipulation of gene expression to the understanding of the function of glypicans. *Glycoconjugate journal* 19(4-5):319-323.
53. Capurro M, Martin T, Shi W, & Filmus J (2014) Glypican-3 binds to frizzled and plays a direct role in the stimulation of canonical Wnt signaling. *Journal of Cell Science*.

54. Sakane H, Yamamoto H, Matsumoto S, Sato A, & Kikuchi A (2012) Localization of glypican-4 in different membrane microdomains is involved in the regulation of Wnt signaling. *Journal of Cell Science* 125(2):449-460.
55. Saunders S, Paine-Saunders S, & Lander AD (1997) Expression of the Cell Surface Proteoglycan Glypican-5 Is Developmentally Regulated in Kidney, Limb, and Brain. *Developmental biology* 190(1):78-93.
56. Veugelers M*, et al.* (1999) Glypican-6, a New Member of the Glypican Family of Cell Surface Heparan Sulfate Proteoglycans. *Journal of Biological Chemistry* 274(38):26968-26977.
57. Eriksson AS & Spillmann D (2012) The Mutual Impact of Syndecan-1 and Its Glycosaminoglycan Chains—A Multivariable Puzzle. *Journal of Histochemistry & Cytochemistry* 60(12):936-942.
58. Essner JJ, Chen E, & Ekker SC (2006) Syndecan-2. *The international journal of biochemistry & cell biology* 38(2):152-156.
59. Pacifici M*, et al.* (2005) Syndecan-3: a cell-surface heparan sulfate proteoglycan important for chondrocyte proliferation and function during limb skeletogenesis. *J Bone Miner Metab* 23(3):191-199.
60. Elfenbein A & Simons M (2013) Syndecan-4 signaling at a glance. *Journal of Cell Science* 126(17):3799-3804.
61. Naor D, Wallach-Dayan SB, Zahalka MA, & Sionov RV (2008) Involvement of CD44, a molecule with a thousand faces, in cancer dissemination. *Seminars in Cancer Biology* 18(4):260-267.
62. Bilandzic M & Stenvers KL (2011) Betaglycan: A multifunctional accessory. *Molecular and Cellular Endocrinology* 339(1–2):180-189.
63. Shintani Y*, et al.* (2006) Glycosaminoglycan modification of neuropilin-1 modulates VEGFR2 signaling. *The EMBO Journal* 25(13):3045-3055.
64. Farach-Carson MC, Warren CR, Harrington DA, & Carson DD (2014) Border patrol: Insights into the unique role of perlecan/heparan sulfate proteoglycan 2 at cell and tissue borders. *Matrix Biology* 34(0):64-79.
65. Ngo ST, Noakes PG, & Phillips WD (2007) Neural agrin: A synaptic stabiliser. *The international journal of biochemistry & cell biology* 39(5):863-867.
66. Seppinen L & Pihlajaniemi T (2011) The multiple functions of collagen XVIII in development and disease. *Matrix Biology* 30(2):83-92.
67. Edgell C-JS, BaSalamah MA, & Marr HS (2004) Testican-1: A Differentially Expressed Proteoglycan with Protease Inhibiting Activities. *International Review of Cytology*, (Academic Press), Vol Volume 236, pp 101-122.
68. Schnepp A*, et al.* (2005) Mouse Testican-2: EXPRESSION, GLYCOSYLATION, AND EFFECTS ON NEURITE OUTGROWTH. *Journal of Biological Chemistry* 280(12):11274-11280.
69. Hartmann U*, et al.* (2013) Testican-3: a brain-specific proteoglycan member of the BM-40/SPARC/osteonectin family. *Journal of neurochemistry* 125(3):399-409.
70. Kolset SO & Tveit H (2008) Serglycin – Structure and biology. *Cell. Mol. Life Sci.* 65(7-8):1073-1085.
71. Filmus J (2001) Glypicans in growth control and cancer. *Glycobiology* 11(3):19R-23R.
72. Fico A, Maina F, & Dono R (2011) Fine-tuning of cell signaling by glypicans. *Cellular and molecular life sciences : CMLS* 68(6):923-929.
73. Fransson LÅ*, et al.* (2004) Novel aspects of glypican glycobiology. *CMLS, Cell. Mol. Life Sci.* 61(9):1016-1024.

74. Song HH & Filmus J (2002) The role of glypicans in mammalian development. *Biochimica et biophysica acta* 1573(3):241-246.
75. Filmus J & Selleck SB (2001) Glypicans: proteoglycans with a surprise. *The Journal of clinical investigation* 108(4):497-501.
76. David G, *et al.* (1990) Molecular cloning of a phosphatidylinositol-anchored membrane heparan sulfate proteoglycan from human lung fibroblasts. *The Journal of cell biology* 111(6):3165-3176.
77. Filmus J, Shi W, Wong ZM, & Wong MJ (1995) Identification of a new membrane-bound heparan sulphate proteoglycan. *Biochemical Journal* 311(Pt 2):561-565.
78. Stipp C, Litwack E, & Lander A (1994) Cerebroglycan: an integral membrane heparan sulfate proteoglycan that is unique to the developing nervous system and expressed specifically during neuronal differentiation. *The Journal of cell biology* 124(1):149-160.
79. Veugelers M, *et al.* (1997) Characterization of Glypican-5 and Chromosomal Localization of HumanGPC5,a New Member of the Glypican Gene Family. *Genomics* 40(1):24-30.
80. Nakato H, Futch TA, & Selleck SB (1995) The division abnormally delayed (dally) gene: a putative integral membrane proteoglycan required for cell division patterning during postembryonic development of the nervous system in Drosophila. *Development* 121(11):3687-3702.
81. Baeg GH, Lin X, Khare N, Baumgartner S, & Perrimon N (2001) Heparan sulfate proteoglycans are critical for the organization of the extracellular distribution of Wingless. *Development* 128(1):87-94.
82. Hudson ML, Kinnunen T, Cinar HN, & Chisholm AD (2006) C. elegans Kallmann syndrome protein KAL-1 interacts with syndecan and glypican to regulate neuronal cell migrations. *Developmental biology* 294(2):352-365.
83. Gumienny TL, *et al.* (2007) Glypican LON-2 Is a Conserved Negative Regulator of BMP-like Signaling in Caenorhabditis elegans. *Current Biology* 17(2):159-164.
84. Marlow F, *et al.* (1998) Functional Interactions of Genes Mediating Convergent Extension,knypekandtrilobite,during the Partitioning of the Eye Primordium in Zebrafish. *Developmental biology* 203(2):382-399.
85. Li F, Shi W, Capurro M, & Filmus J (2011) Glypican-5 stimulates rhabdomyosarcoma cell proliferation by activating Hedgehog signaling. *The Journal of cell biology* 192(4):691-704.
86. Mertens G, Van der Schueren B, van den Berghe H, & David G (1996) Heparan sulfate expression in polarized epithelial cells: the apical sorting of glypican (GPI-anchored proteoglycan) is inversely related to its heparan sulfate content. *The Journal of cell biology* 132(3):487-497.
87. Capurro M, *et al.* (2003) Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology* 125(1):89-97.
88. Kreuger J, Perez L, Giraldez AJ, & Cohen SM (2004) Opposing activities of Dally-like glypican at high and low levels of Wingless morphogen activity. *Dev Cell* 7(4):503-512.
89. Traister A, Shi W, & Filmus J (2008) Mammalian Notum induces the release of glypicans and other GPI-anchored proteins from the cell surface. *The Biochemical journal* 410(3):503-511.
90. Fujise M, *et al.* (2003) Dally regulates Dpp morphogen gradient formation in the Drosophila wing. *Development* 130(8):1515-1522.
91. Han C, Belenkaya TY, Wang B, & Lin X (2004) Drosophila glypicans control the cell-to-cell movement of Hedgehog by a dynamin-independent process. *Development* 131(3):601-611.

92. Akiyama T, *et al.* (2008) Dally regulates Dpp morphogen gradient formation by stabilizing Dpp on the cell surface. *Developmental biology* 313(1):408-419.
93. De Cat B, *et al.* (2003) Processing by proprotein convertases is required for glypican-3 modulation of cell survival, Wnt signaling, and gastrulation movements. *The Journal of cell biology* 163(3):625-635.
94. Watanabe K, Yamada H, & Yamaguchi Y (1995) K-glypican: a novel GPI-anchored heparan sulfate proteoglycan that is highly expressed in developing brain and kidney. *The Journal of cell biology* 130(5):1207-1218.
95. Capurro M, Shi W, Izumikawa T, & Filmus J (2015) Processing by convertases is required for Glypican-3-induced inhibition of Hedgehog signaling. *The Journal of biological chemistry*.
96. Capurro MI, Xiang Y-Y, Lobe C, & Filmus J (2005) Glypican-3 Promotes the Growth of Hepatocellular Carcinoma by Stimulating Canonical Wnt Signaling. *Cancer research* 65(14):6245-6254.
97. Hou S, Maccarana M, Min TH, Strate I, & Pera EM (2007) The Secreted Serine Protease xHtrA1 Stimulates Long-Range FGF Signaling in the Early Xenopus Embryo. *Developmental Cell* 13(2):226-241.
98. Kirkpatrick CA, *et al.* (2006) The function of a Drosophila glypican does not depend entirely on heparan sulfate modification. *Developmental biology* 300(2):570-582.
99. Cheng W, *et al.* (2008) Glypican-3-mediated oncogenesis involves the Insulin-like growth factor-signaling pathway. *Carcinogenesis* 29(7):1319-1326.
100. Williams EH, *et al.* (2010) Dally-like core protein and its mammalian homologues mediate stimulatory and inhibitory effects on Hedgehog signal response. *Proceedings of the National Academy of Sciences* 107(13):5869-5874.
101. Jen YH, Musacchio M, & Lander AD (2009) Glypican-1 controls brain size through regulation of fibroblast growth factor signaling in early neurogenesis. *Neural development* 4:33.
102. Viviano BL, *et al.* (2005) Altered hematopoiesis in glypican-3-deficient mice results in decreased osteoclast differentiation and a delay in endochondral ossification. *Developmental biology* 282(1):152-162.
103. Chiao E, *et al.* (2002) Overgrowth of a Mouse Model of the Simpson–Golabi–Behmel Syndrome Is Independent of IGF Signaling. *Developmental biology* 243(1):185-206.
104. LeClair EE, Mui SR, Huang A, Topczewska JM, & Topczewski J (2009) Craniofacial skeletal defects of adult zebrafish Glypican 4 (knypek) mutants. *Developmental Dynamics* 238(10):2550-2563.
105. Allen NJ, *et al.* (2012) Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors. *Nature* 486(7403):410-414.
106. Aikawa T, *et al.* (2008) Glypican-1 modulates the angiogenic and metastatic potential of human and mouse cancer cells. *The Journal of clinical investigation* 118(1):89-99.
107. Song HH, Shi W, Xiang Y-Y, & Filmus J (2005) The Loss of Glypican-3 Induces Alterations in Wnt Signaling. *Journal of Biological Chemistry* 280(3):2116-2125.
108. Capurro MI, *et al.* (2008) Glypican-3 Inhibits Hedgehog Signaling during Development by Competing with Patched for Hedgehog Binding. *Developmental Cell* 14(5):700-711.
109. Grisaru S, Cano-Gauci D, Tee J, Filmus J, & Rosenblum ND (2001) Glypican-3 Modulates BMP- and FGF-Mediated Effects during Renal Branching Morphogenesis. *Developmental biology* 231(1):31-46.
110. Pilia G, *et al.* (1996) Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nature genetics* 12(3):241-247.

111. Topczewski J, *et al.* (2001) The Zebrafish Glypican Knypek Controls Cell Polarity during Gastrulation Movements of Convergent Extension. *Developmental Cell* 1(2):251-264.

112. Ohkawara B, Yamamoto TS, Tada M, & Ueno N (2003) Role of glypican 4 in the regulation of convergent extension movements during gastrulation in Xenopus laevis. *Development* 130(10):2129-2138.

113. Galli A, Roure A, Zeller R, & Dono R (2003) Glypican 4 modulates FGF signalling and regulates dorsoventral forebrain patterning in Xenopus embryos. *Development* 130(20):4919-4929.

114. Campos-Xavier AB, *et al.* (2009) Mutations in the Heparan-Sulfate Proteoglycan Glypican 6 (GPC6) Impair Endochondral Ossification and Cause Recessive Omodysplasia. *The American Journal of Human Genetics* 84(6):760-770.

115. Svensson G, Hyrenius Wittsten A, Linse S, & Mani K (2011) The structural role of N-linked glycans on human glypican-1. *Biochemistry* 50(43):9377-9387.

116. Bloechlinger S, Karchewski LA, & Woolf CJ (2004) Dynamic changes in glypican-1 expression in dorsal root ganglion neurons after peripheral and central axonal injury. *The European journal of neuroscience* 19(5):1119-1132.

117. Lau E & Margolis RU (2010) Inhibitors of slit protein interactions with the heparan sulphate proteoglycan glypican-1: potential agents for the treatment of spinal cord injury. *Clinical and experimental pharmacology & physiology* 37(4):417-421.

118. Velleman SG, Song Y, Shin J, & McFarland DC (2013) Modulation of turkey myogenic satellite cell differentiation through the shedding of glypican-1. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology* 164(1):36-43.

119. Verhoeven WMA, *et al.* (2010) Sanfilippo B in an elderly female psychiatric patient: a rare but relevant diagnosis in presenile dementia. *Acta Psychiatrica Scandinavica* 122(2):162-165.

120. Ding K, Mani K, Cheng F, Belting M, & Fransson LA (2002) Copper-dependent autocleavage of glypican-1 heparan sulfate by nitric oxide derived from intrinsic nitrosothiols. *The Journal of biological chemistry* 277(36):33353-33360.

121. Cheng F, Svensson G, Fransson LA, & Mani K (2012) Non-conserved, S-nitrosylated cysteines in glypican-1 react with N-unsubstituted glucosamines in heparan sulfate and catalyze deaminative cleavage. *Glycobiology* 22(11):1480-1486.

122. Fransson LA & Mani K (2007) Novel aspects of vitamin C: how important is glypican-1 recycling? *Trends in molecular medicine* 13(4):143-149.

123. Cheng F, *et al.* (2002) Nitric oxide-dependent processing of heparan sulfate in recycling S-nitrosylated glypican-1 takes place in caveolin-1-containing endosomes. *The Journal of biological chemistry* 277(46):44431-44439.

124. Liang Y, Haring M, Roughley PJ, Margolis RK, & Margolis RU (1997) Glypican and biglycan in the nuclei of neurons and glioma cells: presence of functional nuclear localization signals and dynamic changes in glypican during the cell cycle. *The Journal of cell biology* 139(4):851-864.

125. Fransson LA, *et al.* (2004) Novel aspects of glypican glycobiology. *Cellular and molecular life sciences : CMLS* 61(9):1016-1024.

126. Qiao D, Meyer K, & Friedl A (2013) Glypican 1 Stimulates S Phase Entry and DNA Replication in Human Glioma Cells and Normal Astrocytes. *Molecular and Cellular Biology* 33(22):4408-4421.

127. Qiao D, Meyer K, & Friedl A (2012) Glypican-1 stimulates Skp2 autoinduction loop and G1/S transition in endothelial cells. *The Journal of biological chemistry* 287(8):5898-5909.

128. Taylor DR, Whitehouse IJ, & Hooper NM (2009) Glypican-1 mediates both prion protein lipid raft association and disease isoform formation. *PLoS pathogens* 5(11):e1000666.

129. Hooper NM (2011) Glypican-1 facilitates prion conversion in lipid rafts. *Journal of neurochemistry* 116(5):721-725.

130. Watanabe N*, et al.* (2004) Glypican-1 as an A beta binding HSPG in the human brain: Its localization in DIG domains and possible roles in the pathogenesis of Alzheimer's disease. *Faseb Journal* 18(6):1013-+.

131. Cheng F*, et al.* (2011) Suppression of amyloid beta A11 antibody immunoreactivity by vitamin C: possible role of heparan sulfate oligosaccharides derived from glypican-1 by ascorbate-induced, nitric oxide (NO)-catalyzed degradation. *The Journal of biological chemistry* 286(31):27559-27572.

132. Sandwall E*, et al.* (2010) Heparan sulfate mediates amyloid-beta internalization and cytotoxicity. *Glycobiology* 20(5):533-541.

133. Ohmi K, Zhao HZ, & Neufeld EF (2011) Defects in the medial entorhinal cortex and dentate gyrus in the mouse model of Sanfilippo syndrome type B. *PloS one* 6(11):e27461.

134. Zhang X*, et al.* (2012) Heparanase overexpression impairs inflammatory response and macrophage-mediated clearance of amyloid-β in murine brain. *Acta Neuropathol* 124(4):465-478.

135. Zhang G-l, Zhang X, Wang X-m, & Li J-P (2014) Towards Understanding the Roles of Heparan Sulfate Proteoglycans in Alzheimer's Disease. *BioMed Research International* 2014:516028.

136. Matsuda K*, et al.* (2001) Glypican-1 is overexpressed in human breast cancer and modulates the mitogenic effects of multiple heparin-binding growth factors in breast cancer cells. *Cancer research* 61(14):5562-5569.

137. Su G*, et al.* (2006) Glypican-1 is frequently overexpressed in human gliomas and enhances FGF-2 signaling in glioma cells. *The American journal of pathology* 168(6):2014-2026.

138. Svensson G, Linse S, & Mani K (2009) Chemical and thermal unfolding of glypican-1: protective effect of heparan sulfate against heat-induced irreversible aggregation. *Biochemistry* 48(42):9994-10004.

139. Gray D & Subramanian S (2001) Choice of cellular protein expression system. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* Chapter 5:Unit5 16.

140. Rupp B (2010) *Biomolecular crystallography : principles, practice, and application to structural biology* (Garland Science, New York) pp xxi, 809 p.

141. Anonymous (2008) Protein crystallography: a concise guide. *Choice: Current Reviews for Academic Libraries* 46(1):129-129.

142. Bergfors TM (2009) *Protein crystallization* (International University Line, La Jolla, Calif.) pp xxvii, 474 p.

143. Bergfors TM (1999) *Protein crystallization : techniques, strategies, and tips : a laboratory manual* (International University Line, La Jolla, Calif.) pp xix, 306 p.

144. Wilson A (1949) The probability distribution of X-ray intensities. *Acta Crystallographica* 2(5):318-321.

145. Heras B & Martin JL (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D* 61(9):1173-1180.

146. Newman J (2006) A review of techniques for maximizing diffraction from a protein crystal in stilla. *Acta Crystallographica Section D* 62(1):27-31.

147. Heras B*, et al.* (2003) Dehydration Converts DsbG Crystal Diffraction from Low to High Resolution. *Structure* 11(2):139-145.

148. Sanchez-Weatherby J*, et al.* (2009) Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallographica Section D* 65(12):1237-1246.

149. Russi S*, et al.* (2011) Inducing phase changes in crystals of macromolecules: Status and perspectives for controlled crystal dehydration. *Journal of Structural Biology* 175(2):236-243.

150. Skou S, Gillilan RE, & Ando N (2014) Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat. Protocols* 9(7):1727-1739.

151. Rambo RP & Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* 95(8):559-571.

152. Putnam CD, Hammel M, Hura GL, & Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics* 40(3):191-285.

153. Bernado P & Svergun DI (2012) Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Molecular BioSystems* 8(1):151-167.

154. Svergun DI, Koch MHJ, Timmins PA, & May RP (*Small angle x-ray and neutron scattering from solutions of biological macromolecules* First Edition. Ed pp ix, 358 pages.

155. Petoukhov MV*, et al.* (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography* 45:342-350.

156. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, & Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* 36(5):1277-1282.

157. Guinier A (1939) La diffraction des rayons X aux tr`es petits angles: application `a l'etude de phenom`enes ultramicroscopiques. Th`ese (Univ. de Paris., Paris,).

158. Svergun DI (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography* 25:495-503.

159. Fischer H, de Oliveira Neto M, Napolitano HB, Polikarpov I, & Craievich AF (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography* 43(1):101-109.

160. Svergun D, Barberato C, & Koch MHJ (1995) CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* 28:768-773.

161. Franke D & Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography* 42:342-346.

162. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999). *Biophysical Journal* 77(5):2896-2896.

163. Kozin MB & Svergun DI (2001) Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography* 34:33-41.

164. Volkov VV & Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* 36:860-864.

165. Svergun DI, Petoukhov MV, & Koch MHJ (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* 80(6):2946-2953.

166. Petoukhov MV & Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical Journal* 89(2):1237-1250.

167. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, & Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society* 129(17):5656-5664.

168. Svensson G, Awad W, Hakansson M, Mani K, & Logan DT (2012) Crystal structure of N-glycosylated human glypican-1 core protein: structure of two loops evolutionarily conserved in vertebrate glypican-1. *The Journal of biological chemistry* 287(17):14040-14051.

169. Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, & Tucker PA (2005) Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallographica Section D* 61(4):449-457.

170. Murshudov GN*, et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* 67(Pt 4):355-367.

171. Afonine PV*, et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* 68(Pt 4):352-367.

172. Kim M-S, Saunders AM, Hamaoka BY, Beachy PA, & Leahy DJ (2011) Structure of the protein core of the glypican Dally-like and localization of a region important for hedgehog signaling. *Proceedings of the National Academy of Sciences* 108(32):13112-13117.

173. Strong M*, et al.* (2006) Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences* 103(21):8060-8065.

174. Awad W, Svensson Birkedal G, Thunnissen MM, Mani K, & Logan DT (2013) Improvements in the order, isotropy and electron density of glypican-1 crystals by controlled dehydration. *Acta crystallographica. Section D, Biological crystallography* 69(Pt 12):2524-2533.

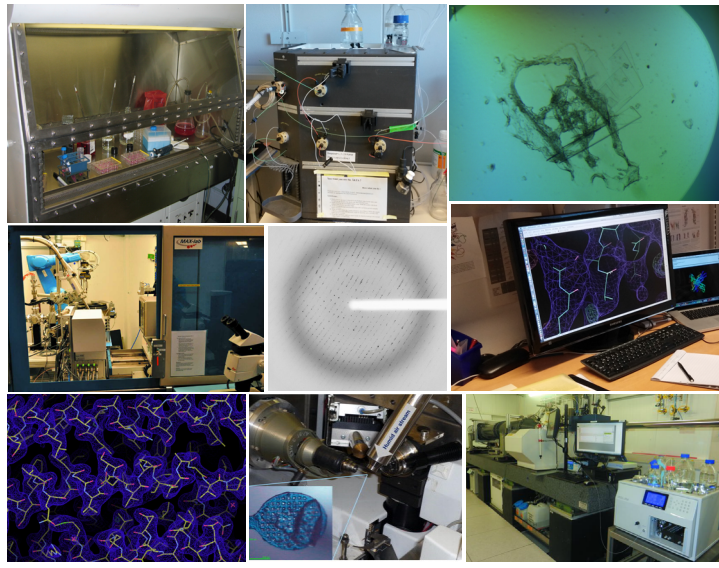
175. Mould PA (2009) Solid Phase Assays for Studying ECM Protein–Protein Interactions. *Extracellular Matrix Protocols,* Methods in Molecular Biology, eds Even-Ram S & Artym V (Humana Press), Vol 522, pp 195-200.

176. Velazquez-Campoy A, Leavitt SA, & Freire E (2015) Characterization of Protein-Protein Interactions by Isothermal Titration Calorimetry. *Methods in molecular biology* 1278:183-204.

177. Lipschultz CA, Li Y, & Smith-Gill S (2000) Experimental Design for Analysis of Complex Kinetics Using Surface Plasmon Resonance. *Methods* 20(3):310-318.

178. Ho M & Kim H (2011) Glypican-3: A new target for cancer immunotherapy. *European Journal of Cancer* 47(3):333-338.

179. Ho MFM (2014) Glypican-3 antibodies: a new therapeutic target for liver cancer. *FEBS letters* 588(2):377-382.



In this book, I have introduced to the biochemical community my modest contribution on glypican co-receptors. We succeeded in reporting the first and highest resolution crystal structures of human glypicans to date. To achieve these research objectives, I truly enjoyed learning various techniques during the course of the study, as shown in these pictures, starting from mammalian protein expression, through to protein crystallography and SAXS measurements. It was a long but extremely exciting and entertaining story, as these experiments were accomplished in many cities including Lund, Grenoble and Hamburg.

WAEL AWAD MOHAMED

Glypican-1: Structural and functional analysis of the N-glycosylated human protein

LUND UNIVERSITY

Lund University
Biochemistry and Structural Biology
ISBN 978-91-7422-399-6

2015