Mass spectrometric analysis of proteoglycans –

Novel tools for studying prohormones in insulin-producing cells

Mahnaz Nikpour

Department of Laboratory Medicine
Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2020

Cover illustration: Electron microscopy image of the B4galt7 knockdown INS-1 832/13 insulinoma cell, MS^2 fragmentation spectrum of a CS-substituted glycopeptide corresponding to chromogranin-A and derived from rat INS-1 832/13 cells, and schematic representation of a secretory granule containing granule (glyco)proteins.

By: Mahnaz Nikpour

Mass spectrometric analysis of proteoglycans –
Novel tools for studying prohormones in insulin-producing cells

Mahnaz.nikpour@gu.se

ISBN 978-91-8009-124-4 (PRINT)
ISBN 978-91-8009-125-1 (PDF)

Printed in Borås, Sweden 2020
Printed by Stema Specialtryck AB, Borås

Dedicated to my father

“The beginning of knowledge is the discovery of something we do not understand.”

Frank Herbert

Mass spectrometric analysis of proteoglycans –

Novel tools for studying prohormones in insulin-producing cells

Mahnaz Nikpour

Department of Laboratory Medicine, Institute of Biomedicine
Sahlgrenska Academy at the University of Gothenburg
Sweden, 2020

ABSTRACT

Proteoglycans (PGs) are proteins that carry one or more negatively charged glycosaminoglycan (GAG) chains. Proteoglycans have been identified in essentially all multicellular organisms, and are implicated in a wide range of biological and pathological processes. To further decipher the influence of GAG glycosylation on biological and pathological events in humans, detailed structural characterization of PGs/GAGs is needed. However, studying the PGs/GAGs is very challenging and has been hampered by the limited number of analytical tools available.

In this thesis work, we have developed and applied glycoproteomics and glycomics methods to characterize PGs and their GAG structures and investigated their possible influence on the cellular characteristics of insulin-producing cells. The workflow included isolation, enrichment, and enzymatic depolymerization of PG/GAG structures followed by structural analysis using liquid chromatography tandem mass spectrometry. We identified several different chondroitin/dermatan sulfate (CS/DS) and heparan sulfate (HS) PGs, some of which are novel PGs. Several of the identified PGs, such as chromogranin-A (CgA) belong to the granin family of secretory granules which are typically co-stored, co-processed and co-released with the insulin hormone. Considering that several of the PGs identified belong to the granin family, and that the granins are important for the biogenesis of dense-core secretory granules, we started to explore the cellular effects of blocking the GAG glycosylation in rat INS-1 832/13 cells by using the CRISPR/Cas9 technique. Our data showed that B4galt7-KO clones had a major, but not always complete, block of the GAG glycosylation of the CgA protein. Furthermore, the cellular localization of CgA as well as its proteolytic processing was different in KO cells compared to WT cells. Further studies of

the effects of downregulation of the GAG biosynthesis in these clones are ongoing.

In summary, our structural findings may assist in elucidating the influence of GAG modifications on the storage, processing, and secretion of peptide hormones of endocrine cells, with particular relevance to insulin-secreting beta cells. Given the paramount importance of insulin on glucose homeostasis, these novel aspects of GAG glycosylation presented herein may provide new insights into diabetes research and future treatment strategies.

Keywords:
CRISPR/Cas9

ISBN 978-91-8009-124-4 (PRINT)
ISBN 978-91-8009-125-1 (PDF)

SAMMANFATTNING PÅ SVENSKA

Proteoglykaner (PG) är en grupp proteiner vilka till vilka en eller flera negativt laddade långa sockerkedjor s.k. glykosaminoglykaner (GAG) är fast kopplade till särskilda aminosyror via typiska s.k. länkregioner. PG har identifierats i väsentligen alla multicellulära organismer och har visats delta i en mängd olika biologiska och patologiska processer. För att kunna kartlägga hur olika PG och deras GAG kedjor påverkar dessa processer måste man i detalj kunna kartlägga sockerkedjornas kemiska strukturer. Att göra detta är emellertid väldigt utmanande, inte minst på grund av PG storlek och varierande GAG strukturer, och lämpliga metoder har tidigare inte funnits till hands.

Vi har i denna avhandling utarbetat och tillämpat nya känsliga metoder för att kunna identifiera PG, kartlägga deras GAG-kedjor samt ange var i proteinet dessa kedjor sitter förankrade. Särskilt har vi, efter att vi upptäckt att flera (pro)hormoner faktiskt är PG, intresserat oss för att studera PG/GAG i insulin-producerande celler från människa, mus och råtta. Vår metodik inkluderar extraktion av PG, degradation med olika proteinspjälkande enzymer, anrikning av GAG-peptider med jonbytarkromatografi, de-polymerisering av GAG med olika sockerspjälkande enzymer och slutligen analys av erhållna sockerföreningar med vätskekromatografi och masspektrometri. Vi har med denna metodik kunnat bekräfta tidigare fynd om PG men också kunnat identifiera flera nya PG bärande på 2 olika typer av GAG (kondroitinsulfat respektive heparansulfat). Fler av de PG vi identifierat t.ex. kromogranin-A och kromogranin-B (secretogranin-1) tillhör den s.k. granin familjen som är typisk för sekretoriska granulae, vilka återfinns i (neuro)endokrina celler och i vilka graninerna lagras, spjälkas och utsöndras tillsammans med andra proteinhormoner som produceras i dessa celler. Graninerna, och de peptider som spjälkas av från dem, är centrala för själva bildandet av sekretoriska granulae. Vi ville därför specifikt undersöka den funktionella betydelsen av PG/GAG i en av de insulin-producerande cell linjer, INS-1 832/13, som många forskare tidigare använt för att förstå graninernas roll vid fysiologisk och patologisk insulin-sekretion (såsom vid diabetes), och vilken vi i våra första arbeten hade studerat i större detalj. Med hjälp av den s.k. gensax tekniken (CRISPR/Cas9) har vi därför etablerat flera cellkloner från denna cellinje där vi genetiskt blockerat tillverkningen av GAG genom att slå ut en av nyckelenzymerna i deras biosyntes. Våra resultat visar att vi i flera kloner fått en mer eller mindre fullständig hämning av GAG biosyntesen. I dessa celler processas och distribueras kromogranin-A på ett annorlunda sätt än i ursprungscellerna och de sekretoriska granulae ser också annorlunda ut. Fortsatta studier är dock nödvändiga för att kartlägga betydelsen av dessa fynd.

Sammanfattningsvis kan vi slå fast att de metoder vi nu utvecklat och tillämpat har gett oss ny kunskap om den strukturella variationsrikedomen av PG och GAG, och gett oss nya möjligheter att kartlägga om och hur GAG kan påverka lagring, spjälkning och utsöndring av peptidhormoner t.ex. i insulin-producerande celler. Med tanke på insulinets centrala roll i upprätthållande av en normal glukosbalans i kroppen, är det troligt att de metoder och resultat kring PG struktur och funktion, som presenteras i denna avhandling, kan ge nya insikter och uppslag för vidare forskning kring diabetes sjukdomarnas orsaker, komplikationer och behandlingsstrategier.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Noborn F, **Nikpour M**, Persson A, Sihlbom C, Nilsson J, and Larson G. A glycoproteomic approach to identify novel proteoglycans. Accepted for publication in Methods in Molecular Biology, Special volume on Glycosaminoglycans: Chemistry & Biology, 2020.

II. **Nikpour M**, Nilsson J, Persson A, Noborn F, Vorontsov E, and Larson G. Proteoglycan profiling of human, rat and mouse insulin-secreting cells. Manuscript

III. Persson A, **Nikpour M***, Vorontsov E*, Nilsson J, and Larson G. Structural domain mapping of proteoglycan-derived glycosaminoglycans from rat insulinoma cells. Manuscript
* Equal contribution

IV. **Nikpour M**, Madsen TD, Satir DM, Gomez Toledo A, Nilsson A, Persson A, Noborn F, Schjoldager KT, and Larson G. Establishing B4galt7 knock-down clones of the rat INS-1 832/13 insulinoma cell line for studying biological effects of downregulation of GAG biosynthesis. Manuscript

.

Publications not included in this thesis:

Madsen TD, Dworkin LA, Hintze J, Hansen L, **Nikpour M**, Larson G, Joshi HJ, Vakhrushev SY, Schjoldager KT. Wide occurrence of O-glycosylation implicate diverse roles in the extended granin-family and controls dense-core granule size and cargo in the SHSY5Y cell-line. Manuscript

El Masri R, Seffouh A, Roelants C, Seffouh I, Gout E, Pérard J, Dalonneau F, Nishitsuji K, Noborn F, **Nikpour M**, Larson G, Crétinon Y, Uchimura K, Daniel R, Lortat-Jacob H, Filhol O and Vivès RR. Extracellular endosulfatase Sulf-2 harbours a chondroitin/dermatan sulfate chain that modulates its enzyme activity. Submitted

CONTENT

ABBREVIATIONS

β3GalT6	β1,3-Galactosyltransferase 6
β4GalT7	β1,4-Galactosyltransferase 7
CgA	Chromogranin-A
ChABC	Chondroitinase ABC
ChAC	Chondroitinase AC
ChB	Chondroitinase B
CS	Chondroitin sulfate
CSF	Cerebrospinal fluid
DBA	Dibutylamine
DCG	Dense-core secretory granule
dHex	Deoxyhexose
DMMB	1,9-Dimethyl-methylene blue
DS	Dermatan sulfate
ECM	Extracellular matrix
ESI	Electrospray ionization
GAG	Glycosaminoglycan
Gal	Galactose
GalNAc	*N*-Acetylgalactosamine
Glc	Glucose

GlcA	Glucuronic acid
GlcAT-I	β1,3-Glucuronyltransferase 3
GlcNAc	*N*-Acetylglucosamine
GT	Glycosyltransferase
HA	Hyaluronic acid
HCD	Higher-energy collisional dissociation
Hex	Hexose
HexA	Hexuronic acid
HexNAc	*N*-Acetylhexosamine
HS	Heparan sulfate
IdoA	Iduronic acid
KS	Keratan sulfate
LC-MS/MS	Liquid chromatography-tandem MS
MS	Mass spectrometry
NCE	Normalized collision energy
Neu5Ac	*N*-Acetylneuraminic acid
NRE	Non-reducing end
PG	Proteoglycan
PTM	Post-translational modification
RER	Rough endoplasmic reticulum

SAX	Strong anion exchange chromatography
SG	Secretory granule
TGN	Trans-Golgi network
Xyl	Xylose
XylT	Xylosyltransferase

> "We live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance."
>
> John Archibald Wheeler

1 INTRODUCTION

Glycosylation refers to the covalent attachment of carbohydrates to proteins and lipids, which leads to the formation of glycoproteins and glycolipids, respectively. Glycosylation is the most frequent post-translational modification of proteins, which provides a vast structural diversity that is vital for their various functions (Khoury, Baliban, and Floudas 2011). Glycans exist in different types, including *N*- and *O*-linked glycans attached to glycoproteins, proteoglycans, glycosphingolipids, and GPI-anchored conjugates (Stanley 2011). Glycans are involved in an enormous number of physiological and pathological processes of cells, and changes in their expressions and structures are linked to many diseases, notably cancer and inflammatory diseases (Reily et al. 2019). The development of new technologies for deciphering the structure and function relationships of glycans has opened up the growing field of glycobiology. Glycobiology encompasses the study of various aspects of glycans such as structure, biosynthesis, evolution, and function (Varki 2017). Studying the various aspects of glycans not only benefits the glycobiology field, but also the biomedical sciences in general.

In the studies presented in this thesis, I have focused on two types of proteoglycan-linked glycans, namely the chondroitin/ dermatan sulfate and the heparan sulfate glycosaminoglycans, to understand the structure and function of these glycans in insulin-producing cells.

1.1 Proteoglycans

Proteoglycans (PGs) are composed of a core protein to which linear polysaccharides, glycosaminoglycans (GAGs), are bound. One or more of GAG chains may be attached to each core protein (Lindahl et al. 2015). GAGs are categorized into four distinct groups, based on the composition of the GAG chain: chondroitin/dermatan sulfate (CS/DS), heparin/heparan sulfate (HS), hyaluronic acid (HA) and keratan sulfate (KS) (Kjellen and Lindahl 1991; Lindahl et al. 2015). Core proteins that carry HS and/or CS GAG chains are called HSPG and CSPG, respectively. Less than hundred mammalian PG core proteins are known to date, which are traditionally classified based on their cellular and subcellular localization into four groups including: intracellular, cell surface, pericellular and extracellular matrix (ECM) PGs (Figure 1) (Iozzo and Schaefer 2015; Toledo et al. 2020).

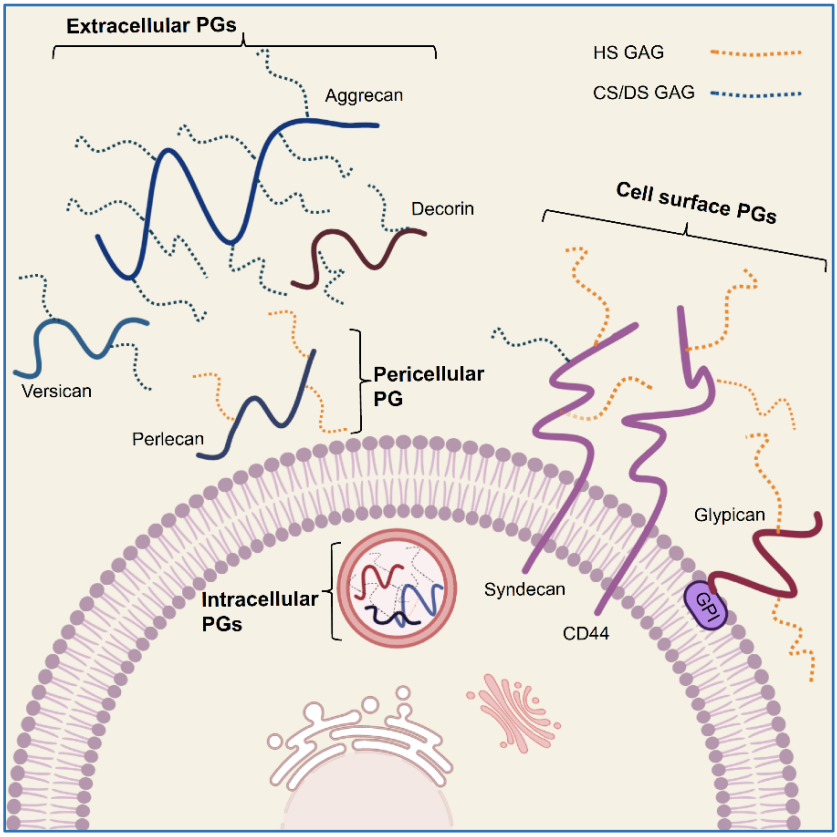

Figure 1. Cellular and subcellular localization of PGs.

The PG serglycin, that can carry heparin GAG chains, was for a long time considered the only intracellular PG, packed inside secretory granules (SGs) of cells, specifically mast cells. Serglycin has an important role in mast cell function since it affects the storage of numerous secretory proteases inside the cells (Abrink, Grujic, and Pejler 2004). Recently, several secreted prohormones such as chromogranin-A (CgA), neuropeptide W, secretogranin-1 and secretogranin-3 were identified to be decorated with GAG chains (Gomez Toledo et al. 2015). These are typically found in SGs and therefore qualify as intracellular PGs. Cell surface PGs are divided into transmembrane and glycosyl-phosphatidyl-inositol (GPI) anchored subgroups. Syndecans (1-4) and glypicans (1-6) which mostly carry HS GAG chains are the best known cell surface PGs. Pericellular PGs are associated with the cell surface either through attachment to the cell surface receptors such as integrins, or associated to the basement membrane. Perlecan, agrin and collagens XV and XVIII are pericellular PGs and typically HSPGs. The hyalectan PGs, small leucine-rich PGs (SLRPs) and the SPOCK family (testicans 1-3) are components typical for the ECM. There are four members of the hyalectan family: aggrecan, brevican, neurocan, and versican. SLRPs, harboring multiple leucine-rich repeats, comprise the largest family of the ECM PGs and exist in a wide range of connective tissues together with collagens, elastin and fibronectin (Iozzo and Schaefer 2015).

1.1.1 Glycosaminoglycan structures

GAGs constitute a very heterogeneous group of linear, negatively charged polysaccharide chains that are bound to core proteins. HS and CS/DS GAGs are made up of repeating building blocks composed of hexuronic acid (HexA) [either glucuronic acid (GlcA) or iduronic acid (IdoA)] and *N*-acetylhexosamine (HexNAc) [either *N*-acetylgalactosamine (GalNAc) or *N*-acetylglucosamine (GlcNAc)]. (Esko, Kimata, and Lindahl 2009; Kjellen and Lindahl 1991).

CS comprises repetitive (-4GlcAβ1-3GalNAcβ1-) disaccharides that can be modified by addition of sulfate to the hydroxyl groups at position C2 of GlcA and at positions C4 and C6 of GalNAc (Lauder 2009) (Figure 2). Positions and extent of CS sulfation varies depending on the source, for instance articular

cartilage in adult human is 6-*O*-sulfated whereas tracheal cartilage is mostly 4-*O*-sulfated (Lauder, Huckerby, and Nieduszynski 2000). The non-sulfated disaccharides may be present in the CS backbone at low percentages (Li et al. 2015), however, the presence of tri-sulfated building blocks are very rare (Nandini and Sugahara 2006).

DS, also known as CS-B, is structurally formed from the same building blocks as CS (Figure 2). However, due to the activity of DS-epimerase enzymes (DS-epi1 and DS-epi2) some of the GlcA are epimerized to IdoA (Pacheco, Malmstrom, and Maccarana 2009). The extent of this modification ranges from one IdoA per GAG chain to almost 100% IdoA in other chains, which leads to the generation of hybrid CS/DS structures (Malmstrom et al. 2012). DS can be modified by sulfation at the same positions as CS. However, 4-*O*-sulfation appears to be of particular importance as it tends to closely accompany epimerization (Tykesson et al. 2018). Since IdoA residues are more often sulfated than GlcA residues, they generate highly modified domains (higher negative charge content) within a co-polymeric CS/DS GAG chain. Co-polymeric CS/DS GAGs have been shown to have a greater conformational flexibility compare to the GAG chains composed of only CS. Structural heterogeneity of these hybrid GAGs enable them to be involved in different and more specialized processes (Ferro et al. 1990).

Heparin and HS are both composed of the same building blocks of (-4GlcAβ-4GlcNAcα-) (Figure 2). Heparin is more sulfated than HS, which makes it a highly negatively charged macromolecule. In addition, it has been shown that the sulfation pattern of heparin and HS is also different, HS has some highly sulfated regions (*N*-sulfated or NS regions) and some unmodified regions (*N*-acetylated or NA regions) while the sulfation pattern of heparin is more evenly distributed (Turnbull and Gallagher 1991). Heparin and HS GAGs contain IdoA residues at different positions along the chains. The conversion into IdoA typically occurs adjacent to GlcNS residues and the chains are then further modified by sulfation (Capila and Linhardt 2002; Kjellen and Lindahl 1991).

HA is the simplest GAG which is composed of (-4GlcAβ-3GlcNAcβ-) repetitive disaccharides (Figure 2). In contrast to the other GAGs, HA is a free polysaccharide chain, not attached to any core protein. Moreover, HA is not modified by sulfation since it is synthetized at the cell surface and does not pass through the Golgi apparatus (Weigel and DeAngelis 2007).

KS consists of repeated (-3Galβ4-GlcNAcβ-) disaccharides and does thus not contain any hexuronic acid (Figure 2). Both monosaccharide residues can be further modified by 6-*O*-sulfation. The KS chain is linked to the core PG through either *N*- or *O*-glycan core structures. Based on the linkage type that connects KS to the core protein, KS chains are divided into three types. KS type I (KS-I) is attached to the core protein via an *N*-glycan, type II (KS-II) is attached through an *O*-glycan and type III (KS-III) is attached through an *O*-linked 2-*O*-mannose structure (Caterson and Melrose 2018).

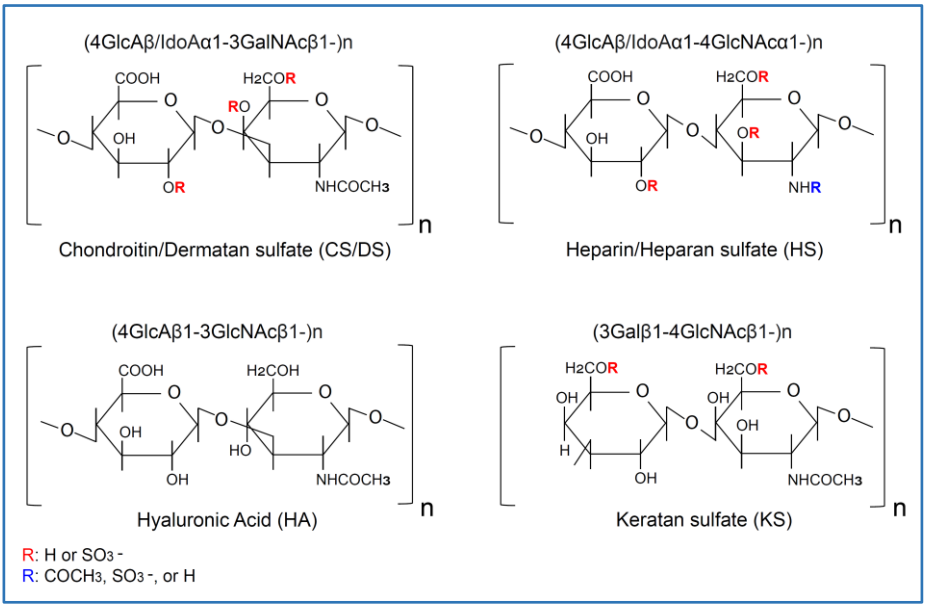

Figure 2. GAG disaccharide units.

1.1.2 Biosynthesis of the GAG tetrasaccharide linkage region

The assembly of CS/DS and HS GAGs, is initiated by the biosynthesis of a common tetrasaccharide linkage region (GlcAβ1-3Galβ1-3Galβ1-4Xylβ1-*O*-Ser) at a selected serine residue of the core protein. Following the biosynthesis of the core proteins in the rough endoplasmic reticulum (RER), the biosynthesis of the GAG chains takes place in the lumen of the Golgi. The building blocks of the GAG chains, monosaccharides and sulfates, are activated in the cell cytosol to form nucleotide monosaccharides and 3′-phosphoadenosine 5′-phosphosulphate (PAPS). These activated building blocks are translocated by nucleotide sugar transporters and the PAPS transporters into the Golgi apparatus, where the CS/DS and HS/heparin are then synthesized (Prydz and Dalen 2000).

Biosynthesis of the linkage region (Figure 3) is initiated by adding a xylose (Xyl) residue from UDP-Xyl to the core protein at a specific serine residue by the action of a xylosyltransferase (XylTs). This common step of HS and CS/DS biosynthesis is a crucial step for the allocation of the GAG attachment site(s) onto a specific core protein. The biosynthesis of GAGs on a specific acceptor serine is favored when serine is flanked by glycine(s), and there is an acidic cluster in close proximity to the acceptor serine (Esko and Zhang 1996). However, the underlying mechanisms of how the XylT enzyme selects a particular serine residue of a particular core protein to catalyze the reaction, are still unclear (Briggs and Hohenester 2018).

It has been shown that vertebrates have two isoforms of xylosyltransferases, XylT-1 and XylT-2. A study, which included 33 human cell lines, showed that XylT-2 mRNA level expression is often higher than that of XylT-1. However, the xylosylation activity of XylT-1 enzyme was, in the same study, reported to be higher than that of XylT-2. Besides, both enzymes acted in the same manner towards the acceptor substrates (Roch et al. 2010).

The action of XylT is followed by stepwise addition of two galactose (Gal) residues and one glucuronic acid (GlcA) residue to the growing glycan chain. These enzymatic reactions are catalyzed by sequential actions of β1,4-galactosyltransferase 7 (β4GalT7/GalT-I), β1,3-galactosyltransferase 6 (β3GalT6/GalT-II) and of β1,3-glucuronyltransferase 3 (GlcAT-I) enzymes (Bai et al. 2001; Kreuger and Kjellen 2012; Mikami and Kitagawa 2013).

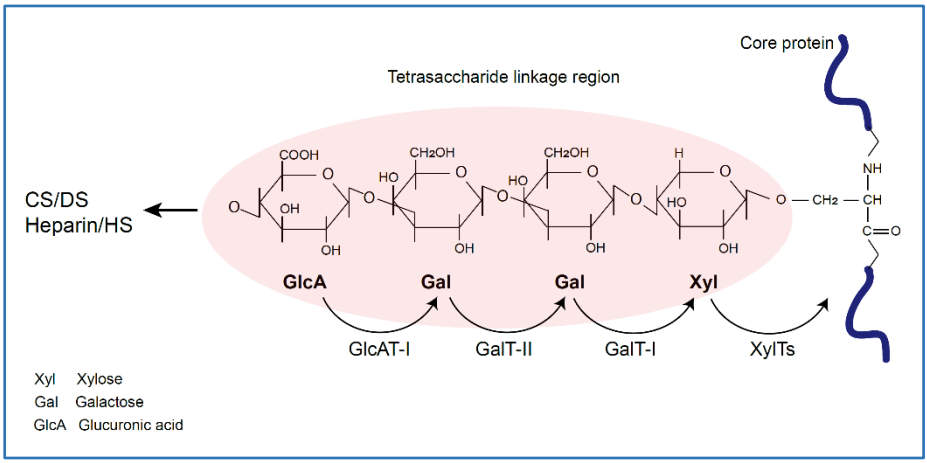

Figure 3. Biosynthesis of tetrasaccharide linkage region of CS/DS and Heparin/HS PGs.

The result of all above enzymatic reactions is the formation of the tetrasaccharide linkage region, which can further undergo several modifications such as phosphorylation, sulfation, sialylation and fucosylation (Gomez Toledo et al. 2015; Kitagawa et al. 2008; Prydz 2015). Phosphorylation of Xyl at the C2 hydroxyl group can occur in both CS/DS and HS biosynthesis and is catalyzed by the secretory pathway kinase, Fam20b (family with sequence similarity 20, member B) (Koike et al. 2009; Zhang, Zhu, et al. 2018). The presence of the Gal-Xyl disaccharide appears necessary for the kinase activity of Fam20b, which in turn enhances the activity of GalT-II enzyme adding the second Gal residue (Kayton et al. 2015). It has been reported that, in the human U2OS osteosarcoma cell line, the absence of Fam20b activity abrogates the formation and later on the elongation of tetrasaccharide linkage regions, which finally leads to the formation of truncated GAG chains capped with an *N*-Acetylneuraminic acid (Neu5Ac) (Kayton et al. 2015). Both Fam20b-dependent Xyl phosphorylation and the

enhanced activity of GalT-II, which lead to the GAG chain elongation, are considered as events that occur in the same time frame. Xyl phosphorylation does not last long and is removed by 2-phosphoxylose phosphatase I (XylP) before the addition of GlcA by GlcAT-I. This step is considered important for the GAG chain elongation, since the GAG polymerization of the tetrasaccharide linkage region may not occur while the Xyl is phosphorylated (Kayton et al. 2015; Koike et al. 2014). Considering that Xyl phosphorylation take place in both CS/DS and HS GAG chains, contrary to Gal sulfation, it does not seem to be a general factor for the selection of GAG type biosynthesis (Prydz 2015). In this work, we identified several linkage regions with Xyl phosphorylation. This was in the agreement with previous findings that the GAG chains could elongate in the presence of Xyl-phosphorylation (Gomez Toledo et al. 2015; Izumikawa et al. 2015).

In contrast to Xyl phosphorylation, *O*-sulfation of Gal residues in the linkage regions have so far been detected only in CS/DS chains. Both Gal residues may be modified by 6-*O*-sulfation, while only the outer Gal may be modified by 4-*O*-sulfation. Since these modifications have been identified only in CS/DS chains and not in HS chains, it may suggest that Gal sulfation serves as a regulatory signal to direct the biosynthesis towards CS/DS (Prydz 2015; Sugahara and Kitagawa 2002). In agreement with these observations, it has been shown that sulfation of Gal residues at C4 or C6 positions influences both the activity and the specificity of the CS GalNAcT-1. This enzyme is responsible for transferring a GalNAc onto a growing linkage region and therefore strictly directing the biosynthesis towards CS/DS chains (Gulberti et al. 2012).

Recently, novel modifications of the CS linkage region including sialylation and fucosylation have been reported (Gomez Toledo et al. 2015), although the impact of these modifications on GAG biosynthesis is not yet clear. Additionally, a non-canonical CS linkage region trisaccharide was recently defined as a minor constituent of the bikunin PG, found in human urine (Persson et al. 2019), which further illustrates that there may be additional variations in the linkage region structures of PGs, at least of CSPGs.

1.1.3 Biosynthesis of the CS/DS backbone

Biosynthesis of the CS/DS GAG chain continues by adding GalNAc and GlcA monosaccharides to the common tetrasaccharide linkage region. CSGalNAcT-1 and CSGalNAcT-2 are the two glycosyltransferases capable of adding the first GalNAc to the linkage region (Uyama et al. 2002). These enzymes are located in the Golgi apparatus and are involved in both the initiation and elongation processes of the CS/DS GAG chain. The CSGalNAcT-1 has been shown to have a higher activity towards the tetrasaccharide linkage than CSGalNAcT-2. On the other hand, CSGalNAcT-2 has been reported to have a higher activity towards specific elongated and sulfated GlcA-GalNAc substrates (Sato et al. 2003).

In addition to CSGalNAcTs, specific polymerizing enzymes are involved in the CS/DS chain elongation. These enzymes are classified into two pairs according to the similarity of their amino acid sequences. Chondroitin sulfate synthase 1 (CSS1) (Kitagawa, Uyama, and Sugahara 2001), also known as chondroitin synthase 1 (ChSy1), and chondroitin sulfate synthase 3 (CSS3) (Yada, Sato, et al. 2003), also known as chondroitin synthase 2 (ChSy2), comprise one pair which exhibit both β3GlcAT and β4GalNAcT activities. The other pair of enzymes is comprised of chondroitin sulfate synthase 2 (CSS2), also known as chondroitin-polymerizing factor (ChPF) (Yada, Sato, et al. 2003), and chondroitin sulfate glucuronyltransferase (CSGlcA-T) (Izumikawa et al. 2008a), also known as chondroitin polymerizing factor 2 (ChPF-2). Similar to CSS1 and CSS3, CSS2 also express both β3GlcAT and β4GalNAcT activities, while CSGlcAT has only β3-GlcAT activity toward GalNAc (Yada, Gotoh, et al. 2003). The *in vitro* results suggest that despite CSS1, CSS2 and CSS3 are all bifunctional enzymes, the co-expression of any two of them is the minimum requirement for the CS polymerizing machinery to work, implying that hetero-oligomers of two glycosyltransferases are needed (Izumikawa et al. 2008b). Furthermore, the length of the CS chain polymerized by different combinations of polymerizing enzymes appears to differ; for instance, the combination of CSS1 and CSS2 has been shown to produce longer CS chain in vivo, in comparison to the combination of CSS3 and CSS2, or CSS3 and CSS1. However, out of all these combinations, the CSS1 and CSS2 hetero-dimer complex has been shown to have the highest in vitro activity and is considered as the most important combination for the CS backbone formation. Nevertheless, these results are not always applicable to the in vivo conditions,

and the exact underlying mechanisms of CS polymerization in vivo remain to be clarified (Izumikawa et al. 2008b; Izumikawa et al. 2007).

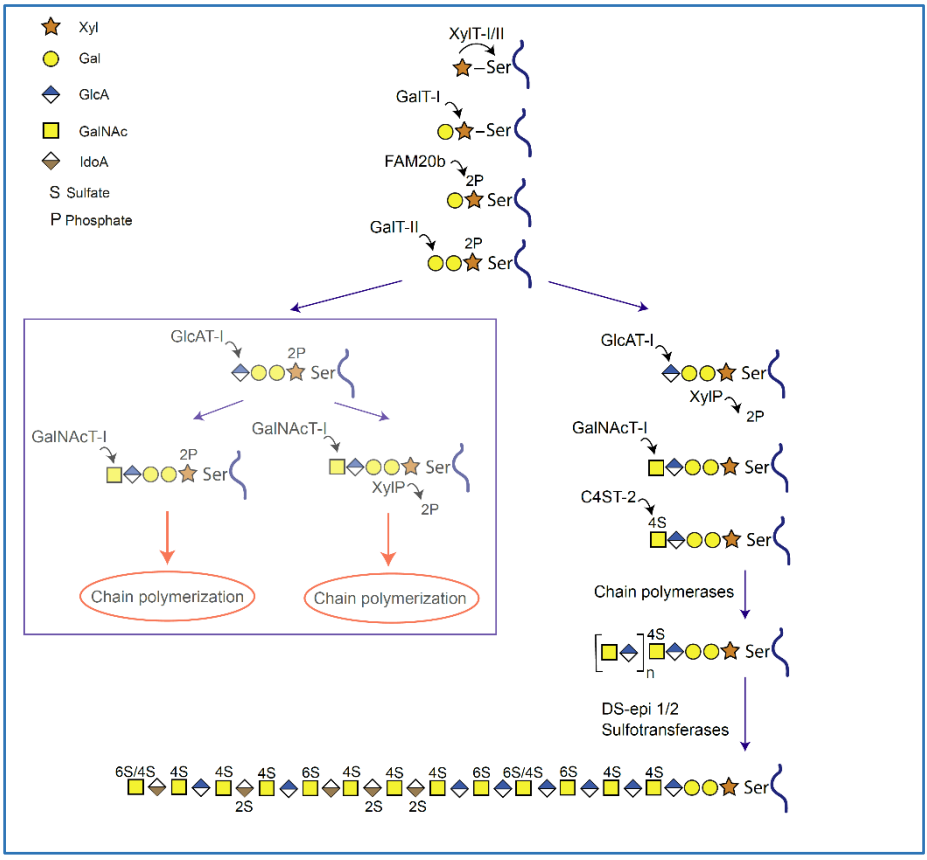

Figure 4. Biosynthesis of CS/DS GAGs. The right panel presents a biosynthesis pathway, which is suggested by most studies. The left pathway, in boxed and translucent, shows the alternative pathways of CS/DS biosynthesis.

1.1.4 Modification of the CS/DS backbone

Two major modifications occur at the CS/DS GAG backbone; epimerization of GlcA to IdoA and *O*-sulfation (Figure 4). These modifications generate a high structural diversity in the GAG chains that seems to be cell- and tissue-dependent (Kusche-Gullberg and Kjellen 2003). The exact order of the different steps of CS/DS backbone modifications is still largely unknown, but

it seems that the chain modifications take place parallel to the chain polymerization.

Epimerization of GlcA to IdoA, which is catalyzed by DS-epimerase enzymes (DS-epi1 and DS-epi2) (earlier described in the paragraph *Glycosaminoglycan structures*), occurs within the same time frame as the *O*-sulfation (Malmstrom 1984). Sulfate groups can be attached to the hydroxyl groups at positions C4 and C6 of GalNAc and C2, C3 of GlcA/IdoA of the growing CS/DS chain (Kusche-Gullberg and Kjellen 2003). Chondroitin 6-*O*-sulfotransferase (C6ST) and 4-*O*-sulfotransferase (C4ST) modifies the GalNAc residue at C6 and C4 positions, respectively (Kitagawa et al. 2000; Nolan et al. 1995; Yamauchi et al. 2000). Chondroitin 4-*O*-sulfotransferase has three isomers, C4ST-1, C4ST-2 (Hiraoka et al. 2000), and C4ST-3 (Kang et al. 2002) that are capable of adding sulfate groups to GalNAc residues adjacent to GlcA or IdoA. Addition of 4-*O*-sulfate groups to the GalNAc residues adjacent to IdoA is catalyzed by the enzymatic activity of dermatan 4-*O*-sulfotransferase (D4ST) (Evers et al. 2001). D4ST-1 has been shown to transfer the 4-*O*-sulfate group to the GalNAc residue, when flanked by two IdoA residues, thereby it is essential for the formation of iduronic acid blocks in the DS chains (Pacheco, Maccarana, and Malmstrom 2009). The 2-*O*-sulfation of GlcA and IdoA are catalyzed by uronyl-2-*O*-sulfotransferase (CS/DS2ST) (Kobayashi et al. 1999). The different *O*-sulfotransferases can also work in combination and generate disulfated units of the CS/DS chains. Along with the *O*-sulfation at different CS/DS chain positions, other modifications such as fucosylation (Vieira and Mourao 1988; Gomez Toledo et al. 2015) can occur and increase the structural complexity of CS/DS chains. Altogether, various modifications of the CS/DS GAG chains generate the structural heterogeneity, which is linked to the biological functions of these molecules (Sugahara et al. 2003).

1.1.5 Biosynthesis of the HS/heparin backbone

The HS formation begins with the addition of a GlcNAc residue in α1,4 position to the GlcA of the common tetrasaccharide linkage region by the enzymatic action of exostosin-like 3 (EXTL3). Polymerization of the HS chain continues by the action of HS polymerases with dual GlcNAcT-II and GlcAT-II activities. The two enzymes responsible for polymerization belong to the

exostosin (EXT) family and are called EXT1 and EXT2 (Lind et al. 1998; Presto et al. 2008).

1.1.6 Modification of the HS/heparin backbone

Along with the chain elongation, HS chains are subsequently modified through several enzymatic steps. Sulfation of GlcNAc residues takes place by the action of a family of GlcNAc *N*-deacetylase/*N*-sulfotransferase (NDST) enzymes, which substitute *N*-acetyl groups from selected GlcNAc residues with *N*-sulfate groups (Carlsson et al. 2008). HS chains are further modified by C5-epimerase enzyme activity which catalyzes the epimerization of GlcA into IdoA (Li et al. 2001) and also by 2-*O*-, 6-*O*- and 3-*O*-sulfotransferases activities which generate HS *O*-sulfations of the monosaccharide residues along the length of the HS chain (Figure 5) (Sugahara and Kitagawa 2002). Since the subsequent HS chain variations are mostly confined to the *N*-sulfated regions, the NDST enzymes (NDST1-4) are considered to play a major role in the overall HS structural design (Presto et al. 2008). Structural diversity in HS chain length and sulfation pattern generates important variability regarding biological functions that can be unique to each HS chain and affect the interactions with other molecules, e.g. the heparin-binding proteins.

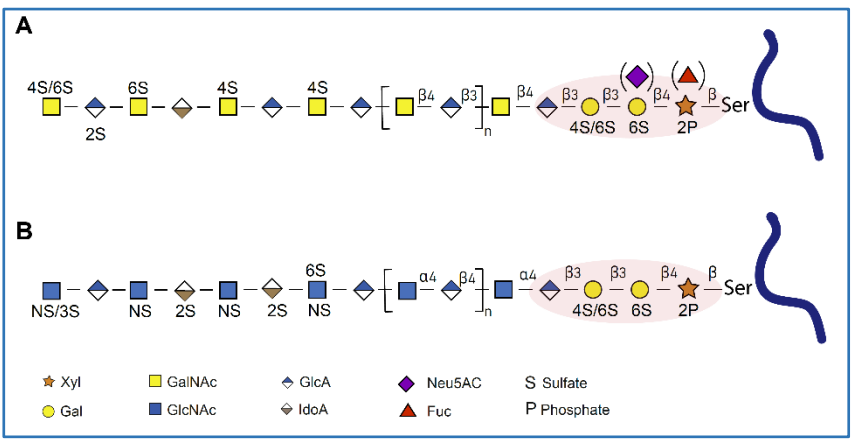

Figure 5. Schematic presentation of CS/DS (A) and HS (B) GAGs with a common tetrasaccharide linkage region (shaded) attached to a serine residue of the core protein.

1.2 Secretory granule biogenesis

Specialized secretory cells such as endocrine, neuroendocrine, exocrine, and some hematopoietic cells are known to form dense-core secretory granules (DCGs), which store and release cargo molecules upon stimulation (Kim et al. 2006). The morphologic term “dense-core granule” refers to the typical electron-dense appearance of these vesicles when viewed by transmission electron microscopy (TEM). Secretory granules (SGs) are loaded with high concentrations of peptide hormones, neuropeptides, processing enzymes, and neurotransmitters. Moreover, SGs are harboring the processing of stored proteins and maintaining the communication with the extra granular environment via regulating the ionic composition of the granule lumen by ion channels and ATP-dependent proton pumps (Day and Gorr 2003). SG peptides are synthesized as larger precursor proteins by ribosomes at the surface of the RER, then transported into the lumen of this organelle by their hydrophobic signal peptide, and subsequently trafficked through the Golgi complex to the trans-Golgi network (TGN). The precursor proteins, along with other regulatory secretory pathway proteins, are packaged into immature SGs. Subsequently, propeptides are partially cleaved to active peptide hormones, which results in the increasing condensation of SGs and their conversion into mature SGs, which are finally secreted in a regulated manner in response to certain stimuli (Borgonovo, Ouwendijk, and Solimena 2006). This controlled secretion of hormones and neuropeptides constitutes the regulated (stimulus-inducible) secretory pathway. All cells, including endocrine, exocrine, and neuronal cells, also have another secretory pathway called the constitutive secretory pathway. In this pathway, continuous secretion is limited by the availability of the products, and the SGs release their contents without any external stimulation. These two pathways are segregated from each other at the TGN, where immature SGs are formed due to the milieu-induced selective aggregation of secretory proteins, and constitutive SGs are formed from soluble secretory proteins (Tooze 1998)

1.2.1 Protein sorting into the regulated secretory pathway

The mechanism(s) by which some proteins are sorted into the regulated secretory pathway at the TGN have been studied intensively. Based on

numerous studies, two distinct and possibly parallel sorting mechanisms have been proposed: receptor-mediated (active sorting) and aggregation-mediated (passive sorting) (Loh, Snell, and Cool 1997). According to the receptor-mediated model, the binding of prohormone sorting signal to the corresponding sorting receptor in the lumen of the TGN leads the prohormone targeting to the regulated secretory pathway. In the aggregation-mediated model, prohormones are packaged into SGs upon selective aggregation in the lumen of the TGN (Loh, Snell, and Cool 1997). Identification of the sorting signals in the sequence of several prohormones, including CgA, CgB, and pro-opiomelanocortin (POMC), provides evidence to support the active sorting model (Chanat et al. 1993; Parmer et al. 1993; Tam, Andreasson, and Loh 1993). Moreover, identification of the common receptor for sorting signals of many prohormones provided evidence for the validity of this model (Zhang, Snell, and Loh 1999). This identified sorting receptor is a membrane-bound form of carboxypeptidase E (CPE), which acts both as a sorting receptor and a processing enzyme within the SGs (Cool et al. 1997; Loh, Snell, and Cool 1997). However, the passive sorting mechanism, triggered by acidic pH and high Ca^{2+} concentration in the lumen of the TGN, is also confirmed as a valid model of prohormones sorting into a regulated secretory pathway (Chanat and Huttner 1991; Colomer, Kicska, and Rindler 1996). In terms of the suggested models, it seems that both aggregation- and receptor-mediated mechanisms work together for sorting (pro)hormones into SGs. According to this, (pro)hormones first aggregate in the lumen of the TGN and then these aggregated (pro)hormones bind to their sorting receptors at the TGN via their sorting signals; and finally accumulate into SGs (Loh, Snell, and Cool 1997). Precise packaging of (pro)hormones within SGs is essential since (pro)hormones are not only packed but also processed to active peptides inside the granules. Thus, any failure in this process would result in the secretion of unprocessed or partly processed prohormones, which in turn could cause (neuro)endocrine dysfunctions (Cool et al. 1997).

1.2.2 Granins in granule biogenesis

Several studies have proposed that the granin family of glycoproteins, particularly chromogranin-A (CgA) and chromogranin-B (CgB/secretogranin 1), play key roles in the granule biogenesis due to their pH- and Ca^{2+}-dependent

aggregation properties and their ability to interact with other SG components (Chanat and Huttner 1991; Colomer, Kicska, and Rindler 1996; Gerdes et al. 1989; Gorr et al. 1988). The granin family is a group of acidic, soluble secretory proteins that have similar structural and biochemical properties and constitute the major components in DCGs (Bartolomucci et al. 2011). The granin family consists of several genetically distinct proteins, including CgA, CgB, chromogranin-C (CgC/secretogranin 2), and other secretogranins (3-7) (Helle 2004). From studying the granin family mRNA distribution it has been shown that they are expressed in various neuroendocrine tissues with distinctly regulated secretory pathways (Taupenot, Harper, and O'Connor 2003).

The aggregation of CgA and CgB, the two major proteins of neuroendocrine SGs, is triggered by a low pH (5.5-6.5) and high Ca^{2+} (10-15 mM) environment in the TGN, and is critical for SG formation, as already discussed (Laguerre, Anouar, and Montero-Hadjadje 2020). Aggregation of vesicle matrix proteins and chromogranins, particularly CgA and CgB, in the acidic milieu of the TGN appears to be a result of pH and Ca^{2+} -induced conformational changes of these proteins (Yoo 1995). In the regulated secretory pathway, aggregation of chromogranins produces a physical force that induces the TGN membrane budding and finally the formation of DCGs. However, this Ca^{2+}-induced aggregation has been suggested to function in a cell-specific manner and is neither necessary nor sufficient for sorting to the regulated secretory pathway in all cell types (Cowley et al. 2000).

Moreover, similar to several other intravesicular matrix proteins, chromogranins have been shown to interact with SG membrane proteins at acidic pH, and at higher concentrations of Ca^{2+} ions (~10 mM) these pH-dependent interactions are stabilized. The interactions dissociate upon increasing pH to near physiological pH of 7.5 and thus seems to be biologically important for CgA targeting to SGs (Chanat and Huttner 1991; Yoo 2010).

1.2.3 Chromogranin-A and the biogenesis of secretory granules

In the search for a master molecule responsible for driving the DCG biogenesis, Kim. et al proposed that in endocrine cells, CgA act as an "on/off switch" that regulates DCG biogenesis and hormone secretion (Kim et al. 2001). In their study, downregulation of CgA using antisense RNAs in PC12 neuroendocrine cells led to impaired granule formation and prohormone secretion as well as a reduction of other SG proteins, such as CgB and CPE, due to an enhanced posttranslational degradation. This phenomenon was rescued upon the restoration of CgA biosynthesis in CgA-depleted PC12 cells (Kim et al. 2001). In the same study, the expression of bovine CgA restored the regulated secretory pathway in the 6T3 mouse cell line, which lacks endogenous CgA expression and DCGs, thus, supporting the critical role of CgA in DCG biogenesis (Kim et al. 2001). Further, data from $Chga^{-/-}$ mice confirmed the putative functions of CgA in regulating DCG biogenesis *in vivo*. In accordance with the in vitro findings, the knock-out mice had less granules of smaller size, and less amounts of other granule proteins than the wild-type ($Chga^{+/+}$) mice (Mahapatra et al. 2005). In contrast to these findings, Huh et.al reported that CgB plays a major and more crucial role in inducing SG formation than CgA, both in PC12 cells and in non-neuroendocrine cells (Huh, Jeon, and Yoo 2003). To further elucidate the underlying mechanism of DCG biogenesis, gene expression profiling was performed on endocrine 6T3 wild-type cells (6T3-WT) and stably transfected cells over-expressing bovine CgA (6T3-bCgA). The comparison of WT and transfected cells revealed that the protease nexin-1 (PN-1), a serine protease inhibitor located in the Golgi apparatus, was up-regulated in the 6T3-bCgA cells. Up-regulation of PN-1 in the Golgi protected granule proteins such as CgA, CgB, and CPE from degradation and consequently enhanced DCG biogenesis (Kim and Loh 2006). Further investigation on how CgA increases PN-1 expression led to the identification of a 2.9-kDa CgA-derived peptide, serpinin that can induce PN-1 mRNA transcription via affecting the translocation of Sp1, a transcription factor that has been shown to be able to induce the expression of PN-1 in the nucleus. Serpinin-mediated PN-1 up-regulation via cAMP, protein kinase A (PKA), and Sp1 signaling pathway

indicate a central role for serpinin in DCG biogenesis (Figure 6) (Koshimizu et al. 2011).

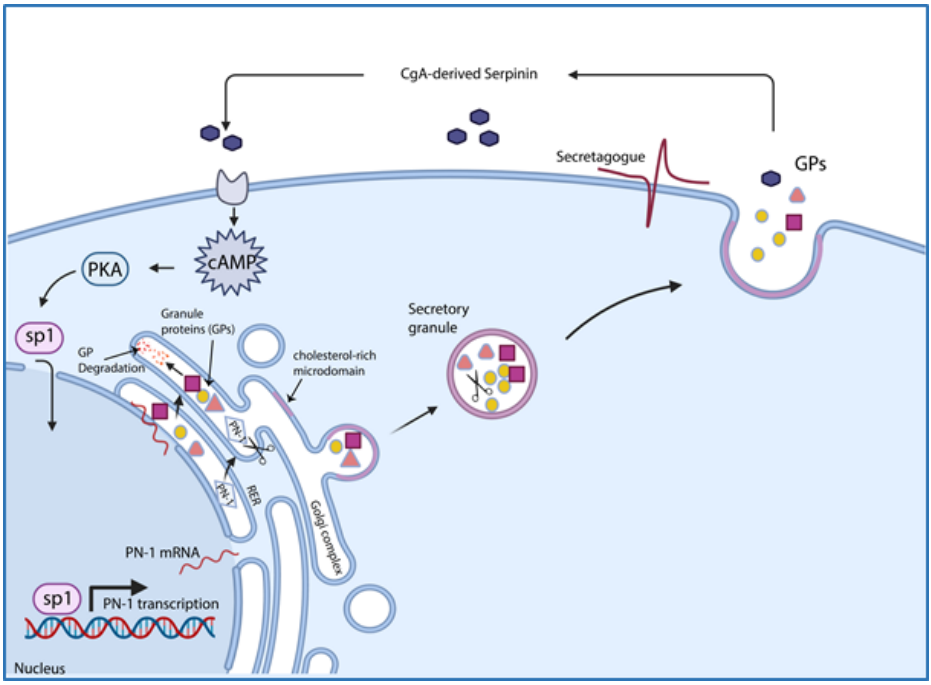

Figure 6. Schematic model of CgA-dependent regulated secretory pathway in neuro(endocrine) cells. CgA and other granule proteins (GPs) are synthesized in the RER, trafficked to the Golgi complex, aggregated and sorted into SGs at cholesterol-rich membrane microdomains. CgA and other prohormones, are cleaved by proteases in the SGs to generate biologically active peptides including serpinin. Upon stimulation of cells by secretagogues, SGs are released into the extracellular space. Serpinin binds to a cognate receptor and subsequently up-regulates PN-1 transcription in the nucleus. PN-1 protein stabilizes and increases the level of GPs at the Golgi complex, which finally promotes biogenesis of SGs. The figure was modified from (Salton. SR Endocrine Reviews, 2011 Dec; 32(6):755-97).

1.2.4 Chromogranin-A as a prohormone

Chromogranin-A has an 18 amino acid long signal peptide and several paired basic amino acids that are potential sites for processing by prohormone convertases (Figure 7). Endoproteases (proprotein convertases) PC1/3 (also termed PC1 or PC3), PC2 and cathepsin L, may thus cleave CgA at these paired basic residues, leaving one or two basic residues at the C-terminal ends of the cleaved peptides. Processing of CgA continues with CPE that removes the remaining basic amino acids generated after the cleavages by endopeptidases. CPE, also known as carboxypeptidase H (CPH), is a proteolytic enzyme located in Golgi and DCGs of (neuro)endocrine cells that exists in both soluble and membrane-bound forms. Soluble CPE is involved in the processing of several prohormones via releasing the C-terminal basic amino acids (Lys and/or Arg). Membrane-bound CPE functions as a sorting receptor for targeting prohormones to the regulated secretory pathway, as discussed earlier (Ji et al. 2017).

Following the initial proteolytic cleavage of CgA prohormone in the TGN, the major processing of CgA takes place within the SGs; however circulating enzymes such as plasmin can further process released CgA extracellularly (Hutton, Davidson, and Peshavaria 1987). CgA is known as a prohormone since upon proteolytic cleavages, it generates several biologically active peptides with different functions including vasostatins (I-II), pancreastatin, catestatin, WE-14 and serpinin (Figure 6). The degree of CgA processing in SGs differs among different tissues, but processing is, in general, less complete compared to the other prohormones. Thus, also intact CgA is released after granule exocytosis (Eskeland et al. 1996; Koshimizu et al. 2010).

The N-terminal domain of CgA is well-conserved over several species and contains a hydrophobic disulfide bridge between cysteine residues 35 and 56 (bovine CgA), that has been shown to be a requirement for CgA sorting into regulated secretory pathway granules in PC12 neuroendocrine cells but not in endocrine GH4C1 cells. Contrarily, the conserved C-terminal domain of CgA (90 aa), that seems to be necessary for CgA sorting in GH4C1 cells, is not critical for CgA sorting in PC12 cells. These results propose that the highly-conserved N- and C-terminal domains of CgA protein may act in a cell-type specific manner for sorting CgA prohormone into the SGs (Cowley et al. 2000). Additionally, Hosaka et al. have shown that in PC12 and pituitary AtT-

20 cells, as well as in pancreatic beta cells, binding of the N-terminal domain of CgA to secretogranin 3 (SgIII) is involved in the routing of CgA prohormone into the regulated secretory pathway (Han et al. 2008; Hosaka et al. 2002). In agreement with the passive aggregation model of sorting prohormones, it has been reported that adrenomedullin is co-localized with CgA and SgIII inside SGs and co-aggregates with CgA in PC12 cells. Taken together, CgA prohormone is sorted into the regulated secretory pathway, through binding to the SgIII (functional domain 214-373) as well as aggregating upon coupling to adrenomedullin in these cells (Han et al. 2008; Sun et al. 2013).

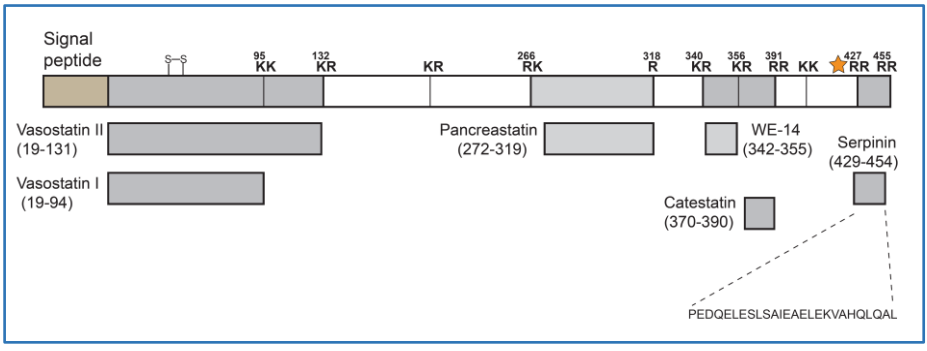

Figure 7. Schematic presentation of human CgA and some of its bioactive peptide fragments. The potential dibasic cleavages sites of CgA are indicated bold. The identified GAG attachment site at Ser-424, is indicated with the star symbol for Xyl.

1.2.5 Chromogranin-A and insulin secretion

Chromogranin-A is co-stored with the insulin prohormone within the SGs of islet beta cells (Lukinius et al. 1992). As discussed earlier, CgA undergoes proteolytic cleavages within the SGs and generates several biologically active peptides/hormones. Upon stimulation by external secretagogues, CgA and CgA derived fragments, as well as the other resident hormones of SGs are released. Chromogranin-A has been shown to be processed in a cell and tissue-specific manner. Thus, the type and the amount of CgA and its derived peptides are varying between tissues (Bartolomucci et al. 2011). The first biologically-active CgA derived peptide was isolated from the porcine pancreas in 1986 and named pancreastatin (PST) (Tatemoto et al. 1986). After the isolation of PST, several

independent studies have shown that the PST is indeed a proteolytic product of CgA (Huttner and Benedum 1987; Konecki et al. 1987). In the study that was performed on endocrine cells from the bovine gut, pancreas and adrenal medulla, PST was a major product of CgA processing in bovine pancreas (Watkinson et al. 1991). Further investigation on PST function in several in vitro studies revealed that PST is involved in glucose homeostasis and functions as an anti-insulin peptide (Valicherla et al. 2013). In humans, PST (amino acids 272-319 of CgA) has been shown to inhibit glucose uptake by 50% in hepatocytes and adipocytes, thereby increasing blood glucose concentration (O'Connor et al. 2005).

In addition to PST, WE-14 (amino acids 342-355 of human CgA), a CgA derived peptide, has been suggested to function as an antigen, which can stimulate the response from highly autoreactive CD^{4+} cells in a non-obese diabetic (NOD) mouse model of type 1 diabetes mellitus (T1DM) (Stadinski et al. 2010). The same function has been proposed for WE-14 in humans, since the T cells of new-onset diabetic patients could recognize and reacts towards the WE-14 as an autoantigen (Gottlieb et al. 2014). Despite the correlation between CgA processing products and diabetes pathogenesis, the impact of GAG glycosylation on the CgA processing is not clear yet.

1.2.6 HS core proteins and insulin secretion

Along with the granin family and other secretory proteins, SGs in endocrine beta cells selectively store insulin, which is secreted in a regulated manner upon appropriate stimulation by glucose or other secretagogues. The immature form of insulin, preproinsulin (110 aa), synthesized in RER, is converted to proinsulin (86 aa) by removal of the first 24 aa (signal peptide). While proinsulin is folded in RER, and before it is sorted into the SGs, three disulfide bridges are formed, two inter-chain (A7- B7 and A20-B19) and one intra-chain (A6-A11) (Chang et al. 2003; Liu, Weiss, et al. 2018). In the TGN, newly synthesized proinsulin is sorted into the immature SGs. Along with the acidification of the immature SGs, due to the activation of proton pumps, further maturation of proinsulin takes place and mature insulin (A and B polypeptide-chains) and C-peptide are formed. A- and B- polypeptide-chains of insulin are connected together through di-sulfide bridges (Figure 8). Like

many other prohormones, processing of preproinsulin is catalyzed by proteolytic enzymes including PC1/3, PC2 and, CPE endo- and exopeptidases. PC2 and PC1/3 that are optimally active at pH 5.5, cleave between the A polypeptide-chain and C-peptide and B polypeptide-chain and C-peptide, respectively. Products of proinsulin cleavage by PC2 and PC1/3 enzymes, further trimmed by CPE to yield mature insulin (Figure 8) (Weiss, Steiner, and Philipson 2000).

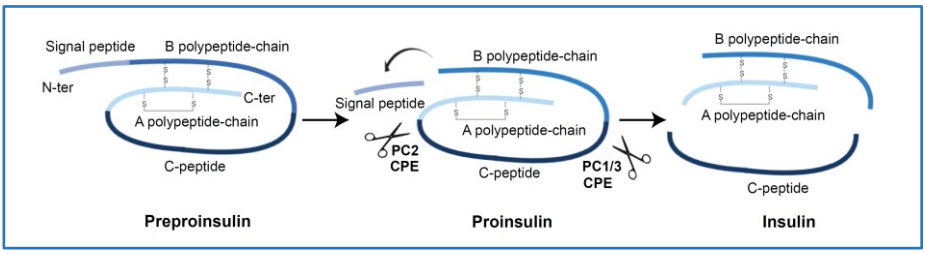

Figure 8. Schematic presentation of preproinsulin to mature insulin processing.

Inside the mature SGs, insulin at a concentration of 100 mM is crystalized together with Zn^{2+} and Ca^{2+} ions to form the cores of the DCGs (diameter of 300-350 nm) (Fu, Gilbert, and Liu 2013; Hou, Min, and Pessin 2009; Hutton 1994). Inside the islets of Langerhans, insulin-producing beta cells (65-90% of the total cell number) are randomly surrounded by other components of the pancreatic islet including glucagon-producing alpha-cells (15-20%), somatostatin-producing delta-cells (3-10%), pancreatic polypeptide-producing PP-cells (1%), and ghrelin-producing epsilon cells (Da Silva Xavier 2018). Beta cells of pancreas possesses the unique ability to secrete insulin, the only hormone that can lower the concentration of blood glucose. Any failure in insulin secretion and/or action can result in a cluster of metabolic dysfunctions characterized by chronic hyperglycemia known as DM (American Diabetes 2009).

Inside the SGs, insulin is packed together with several other proteins. In the search for the co-localizing proteins, several proteomics analyses have been performed and proposed that immature SGs contain more than 140 different proteins. Identified proteins of immature SGs are divided into membrane proteins and intravesicular proteins, in which a large proportion of

intravesicular proteins comprises insulin and proteins of the granin family (Brunner et al. 2007; Li et al. 2018; Stutzer, Esterhazy, and Stoffel 2012). Immunostaining of pancreatic islets revealed that HS structures are intensively present in these areas. Gene expression profiling of adult mice beta cells showed that several members of the EXT family, *N*-deacetylase/*N*-sulfotransferases, C5-epimerase, and HS sulfotransferases are expressed in these cells, indicating that some of the proteins are most likely decorated with HS structures. Further investigation has revealed that HS structures play a role in the maturation of beta cells and are necessary for normal insulin secretion (Takahashi et al. 2009). Takahashi et al. has shown that sulfated HS structures are important for normal insulin secretion in the MIN6 insulin-secreting cell line by suppressing the expression of 3-*O*-sulfotransferase isoform 1, Hs3st1. This finding indicates that sulfate groups of HS GAGs may interact with intra- and extracellular molecules involved in insulin secretion; and indeed Hs3st1 silencing results in impaired glucose-induced insulin secretion (GIIS) (Takahashi, Ohashi, and Nata 2012). This conclusion is further supported by the fact that HS sulfation pattern has been shown to be important for fibroblast growth factor (FGF) signaling (Theodoraki et al. 2015; Turnbull et al. 1992). Indeed, FGF binds to its receptor, fibroblast growth factor receptor (FGFR), and HS GAGs function as a co-receptor for this interaction. The signaling of FGF1 and FGF2 ligands through binding to FGFR3c has been shown to be favored in the presence of high IdoA and sulfate content of HS GAGs (Sterner et al. 2013). FGFR1 has been reported to express in beta cells of adult mouse and be involved in several important features of these cells function, including glucose homeostasis, insulin content, insulin processing and beta cells survival (Hart et al. 2000).

In addition to the vital role of HS structures in insulin secretion, it has been suggested that HS is involved in beta cells survival by protecting the beta cells from oxidative damage which can induce cell death. This line of research also proposed that preventing the degradation of HS structures localized in between insulin secreting beta cells, can protect these cells from destructive autoimmunity and development of autoimmune T1DM (Ziolkowski et al. 2012). In line with the findings from rodent models, in patients with functional mutations in either EXT1 or EXT2 HS polymerases, the insulin secretion response to glucose stimulation through the oral glucose tolerance test (OGTT) has been reported to be impaired (Moens et al. 2014).

In the search for possible HS-carrying core proteins in pancreatic beta cells, syndecan-4 (sdc4) has been identified in MIN6 insulin-secreting cells and been reported to be involved in maintaining normal insulin secretion in these cells (Cheng, Whitelock, and Poole-Warren 2012; Takahashi, Yamada, and Nata 2018). Despite that the involvement of HS structures in several important features of beta cells function has been shown, the HS-associated core proteins, except for Sdc4, have not yet been defined. Two independent studies have been reported that the mRNA of HS-associated core proteins including glypicans and collagen XVIII are expressed in murine islets (Takahashi et al. 2009; Zertal-Zidani, Bounacer, and Scharfmann 2007). However, there are no protein expression data that could confirm the presence of these core proteins in HS positive areas of the pancreatic islets.

1.2.7 Diabetes and insulin secretion

Diabetes is a major chronic disease characterized by deranged insulin secretion in response to elevated blood glucose concentrations. Diabetes has a global prevalence of almost 500 million people, which is yearly increasing with around 10%, and has an expected prevalence of 700 million people in 2045 (Saeedi et al. 2019). Diabetes is more common in urban areas and in high income countries and is the seventh leading cause of death in the US. Almost half of all diabetic patients are unaware of their diagnosis, more so in the low- and middle-income countries. Yearly, more than 2 million people die from elevated blood glucose concentrations and its complications such as kidney failure, heart attacks, stroke, gangrene and limb amputations.

The deranged insulin secretion is pathognomonic for the two classical types of diabetes, usually known T1DM and T2DM, although for completely different reasons. For T1DM, which typically starts at younger ages, an autoimmune response directed towards the pancreatic beta cells, kills these cells and thus selectively eliminates the only source for insulin, physiologically secreted in response to elevated blood glucose concentrations. For T2DM, which typically affects middle-aged people, there is a general resistance to insulin in peripheral tissues which demands an excessive amount of insulin production finally exhausting the capacity of the beta cells. Although these are the two major types of DM there is actually a panorama of different but often rare subtypes

of DM, related to both genetic and environmental factors, as was recently reviewed (Flannick, Johansson, and Njolstad 2016; Ilonen, Lempainen, and Veijola 2019).

Although DM can be treated and complications avoided by substitution therapy with insulin, with anti-diabetic drugs and with well-designed diets, there is no actual cure for the disease. In my work I have aimed to reveal basic mechanisms on the function of GAGs and PGs in the biology of SGs, such as those in which insulin is processed, aggregated and finally secreted. If my findings can have some bearing on the understanding, diagnosis or treatment of DM my efforts will have been highly rewarded.

2 AIMS

The general aim of this thesis was to develop advanced glycoproteomics and glycomics methods to obtain structural information of PGs and GAGs from biological samples and consequently use this information to decipher some of the biological functions of GAG glycosylation. The specific aims are:

- Validate a nano-scale liquid chromatography tandem mass spectrometry (nLC-MS/MS) based glycoproteomic approach that facilitates identification and characterization of CS/DS and HS PGs in biological tissues, organisms and cell cultures (Paper I)
- Structurally characterize PGs of human, mouse and rat insulin-secreting cells to investigate their variability or conservation in different species and cell lines (Paper II)
- Develop of a novel LC-MS/MS method for structural analysis of linkage regions, internal oligosaccharides and non-reducing ends of PG-derived CS/DS GAGs (Paper III)
- Establish a B4galt7 knock-down cell line for studies of the biological roles of GAGs in SG formation and prohormone processing (Paper IV)

"If I have seen further it is by standing on the shoulders of giants."
Isaac Newton

3 METHODS

3.1 Ethical permit

In Paper I, human de-identified cerebrospinal fluid (CSF) samples were collected from individuals undergoing a diagnostic lumbar puncture in the clinic but showing no biochemical alterations. The collection and use of these samples were done according to Swedish law and ethical permissions.

In Paper II, we used human pancreatic islets from deceased donors. These human cell preparations were provided by Prof. Olle Korsgren (Uppsala University, Sweden), under the ethical permission of the Swedish Ethical Review Authority, Uppsala University, 2009/371 updated 20170712. General laboratory and ethical guidelines were applied while working with the human material.

3.2 Cell models

In this thesis I used several (neuro)endocrine cell lines from different species (human, rat and mouse). We intentionally used these cell lines to model the GAG glycosylation status of pancreatic beta cells. Using cell lines is generally cost-effective, they are easier to handle and to manipulate compared to primary cells, and they are also easier to maintain and expand in large quantities. Moreover, reproducing the results is more feasible when using cell lines as the study model. The major disadvantage of using cell lines is that they may not resemble the molecular and functional phenotypes of the primary cells, which makes it difficult to translate findings from cell models to whole organisms and from other species to humans. In addition, genotypic and phenotypic variations may occur over an extended period of time after serial passages of cell lines. Altogether, rodent-derived beta cells lines are widely used as a model for

investigating beta cell development, differentiation, and function. Below, I am discussing all the cell lines that I have used in my work.

3.2.1 Rat INS-1 832/13 insulinoma cell line

The INS-1 832/13 cell line is a sub-clone of a well-established rat insulinoma cell line, INS-1 (Hohmeier et al. 2000). The INS-1 cell line was established by Asfari and colleagues in 1992, by diffusion of a radiation-induced rat insulinoma into the cell culture medium containing 2-mercaptoethanol (2-ME) reducing agent (Asfari et al. 1992). 2-ME maintains the levels of glutathione, which is involved in the neutralization of free radicals, and therefore enhances the defense against oxidative stress. As a consequence of adding 2-ME to the cell culture, the continuous growth of INS-1 and its derivative clones (including INS-1 832/13) is dependent on the routine addition of this component to the cell culture media. The doubling time of these cells is quite long (100 h), which makes them difficult to study. More importantly, their insulin secretory response to glucose is low (2- to 4-fold compare to 15-fold in freshly isolated primary islets) (Asfari et al. 1992). Since these cells were not isolated as a single clone, the cell line consists of a mixture of cells with diverse insulin response capacities, which are also not stable over a long-term culture period. To bypass these problems, the parental INS-1 cells were stably transfected with the plasmid containing human proinsulin gene under the control of cytomegalovirus (CMV) promoter, and followed by the selection of the most glucose-responsive clone, INS-1 832/13, out of 58 different independent colonies (Hohmeier et al. 2000). In contrary to the parental INS-1 cells, glucose responsiveness (almost 10-fold increase upon increasing the glucose concentration from 3 to 15 mM) of INS-1 832/13 cells was sustained in long-term culturing (for at least 7.5 months of culture). In contrast to the parental INS-1 cells, which are heterogeneous, the INS-1 832/13 cells consist of a pure population of insulin-secreting cells, which makes them a good model for studying beta cells. The short doubling time of these cells (48 h) makes them very efficient to work with. The INS-1 832/13 cells are producing both rat (endogenous) and to a lesser extent human (exogenous) insulin, which makes it difficult to assess the differentiated state of the cells based only on their insulin content (Hohmeier et al. 2000).

3.2.2 Mouse MIN6 insulinoma cell line

MIN6 cells originate from a mouse pancreatic beta cells insulinoma. The insulin promoter-simian virus 40 large tumor antigen (SV40-LT) was injected into C57BL/6 mice, and the MIN6 cell line was established from the insulinoma tumors further developed in transgenic mice (Miyazaki et al. 1990). Glucose-stimulated insulin secretion (GSIS) of MIN6 has been shown to be similar to the response from isolated pancreatic cells. Gene expression profiling of these cells shows that they are primarily insulin secretory cells but also able to produce glucagon, somatostatin, and ghrelin (Nakashima et al. 2009).

3.2.3 Mouse NIT-1 insulinoma cell line

The NIT-1 cell line was established from non-obese diabetic (NOD/Lt) mice in which the beta cells had been transformed by SV40-LT under the control of the rat insulin-promoter. These mice developed beta cells adenomas, and from the tumors, the NIT-1 insulin-producing insulinoma cells were established. NIT-1 cells show features of mouse beta cells and show an insulin response to glucose stimulation, and are therefore known as a validated cell model to study beta cells (Hamaguchi, Gaskins, and Leiter 1991).

3.2.4 Human SH-SY5Y neuroblastoma cell line

The human SH-SY5Y neuroblastoma cell line originates from a parental line, the SK-N-HS cell line, which was established from a human bone tumor biopsy in 1970. SH-SY5Y cells can be differentiated into more neuron-like cells that are, expressing neuronal markers. This cell line is widely used to study neuronal differentiation, metabolism, and function as well as neurodegenerative diseases (Kovalevich and Langford 2013).

3.2.5 Human GOT1 neuroendocrine cell line

The GOT1 cell line was established from a human liver metastasis by Nilsson and colleagues in 2001 (Kolby et al. 2001). Primary cell cultures from the liver tumor material were prepared and subsequently grafted to nude mice and then the tumor regeneration was repeated. After five grafted tumor generations, cell cultures were re-established and characterized as having a well-preserved phenotype and a long doubling time (≥ 18 days) (Grozinsky-Glasberg, Shimon, and Rubinfeld 2012). GOT1 cells express endocrine markers, such as CgA and synaptophysin, but also midgut carcinoid markers, e.g., serotonin (5-HT). In addition, GOT1 cells respond to changes in Ca^{2+} levels similar to other neuroendocrine cells such as pancreatic islets and PC12 cells (Kolby et al. 2001). This GOT1 cell line has been used for drug testing and functional studies in xenograft mice. Here, we used them as a neuroendocrine cell model to characterize their PGs.

3.2.6 Comparison of cell models

The generation and use of rodent pancreatic beta cells lines have been continued over the past 30 years, and a series of insulin-secreting cell lines with different characteristics have been established. Although each cell model has some limitations, combining different cell models provides a valuable tool to study rodent beta cells function in detail. One of the advantages of using rodent cell lines is that they offer an unlimited, homogenous, and reproducible supply of beta cells.

Rodent beta cells share many similarities with human beta cells; however, these cells are not identical due to several notable differences between them (Scharfmann, Rachdi, and Ravassard 2013). For instance, while there is one unique insulin coding gene in humans, insulin is coded for by two distinct genes in rodents (rat and mice) (Melloul, Marshak, and Cerasi 2002). Moreover, in rodents, beta cells are located in the islet centers where they are surrounded by alpha (glucagon-producing) and delta (somatostatin-producing) cells. In human islets, beta, alpha, and delta cells are dispersed throughout the islet (Brissova et al. 2005). There are also other differences between rodent and

human islets, such as differences in transcription factors, functionality, and susceptibility to beta cells injuries.

Development of novel tools to freshly isolate, prepare and study human beta cells from human islets from deceased donors have shown that the human beta cells tend to be more challenging to study than the established rodent cells (Kayton et al. 2015). One reason for this is that human materials are prepared from donors with various characteristics such as gender, age, body mass index, co-morbidity and cause of death (Scharfmann, Staels, and Albagli 2019). Thus, to circumvent the limited availability and practicality of human primary beta cells for research purposes, a stable and immortal human beta cells line with comparable physiological characteristics to primary beta cells has been needed. In that regard, a lot of efforts have been ongoing to generate a reliable human beta cell line that reflects our understanding of human islet cell biology and pathology.

3.2.7 Human insulin-secreting cell lines

The first human-derived immortalized beta cells line, BetaLox-5, was established from human adult beta cells in 1999 and turned out to express low levels of insulin, which was also unstable and lost with increasing the number of passages (de la Tour et al. 2001). Scientists in the field took advantage of the experience from the establishment of functional rodent cell lines and applied this knowledge to generate a functional human beta cells line. In 2011, a new human beta cells line, EndoC-βH1, generated by Ravassard and colleagues, opened up a possibility to study human beta cells biology in vitro (Ravassard et al. 2011). To generate this cell line, fetal pancreatic buds were transduced with the SV40LT-lentiviral vector under the control of rat insulin promoter. The resulting tissue, was grafted into severe combined immune-deficient (SCID) mice where it was allowed to develop into insulinoma cells. The insulinoma cells were then removed and transduced with a second lentivirus expressing human telomerase reverse transcriptase (hTERT) and grafted into other SCID mice. The resulting cells were expanded in culture as insulin-secreting cells with the ability to maintain insulin expression over more than 80 passages (Ravassard et al. 2011). EndoC-βH1 cells have since then been shown to be able to secret insulin in response to the glucose stimuli and

also express very low levels of other islet cell types than beta cells. In conclusion, EndoC-βH1 cell have been proven as an alternative to rodent beta cell lines since the physiological characteristics are closer to human primary beta cells than to any other beta cell of rodent origin (Gurgul-Convey, Kaminski, and Lenzen 2015). However, the EndoC-βH1 cells have some limitations such as slow growth (doubling time is 7 days), a requirement of expertise for propagation and handling, and is not suitable for cell therapy due to their tumoral properties (Scharfmann, Staels, and Albagli 2019).

More recently human pancreatic organoids have evolved as models to study human beta cell functionality in vitro. These organoids are developed in a three-dimensional culture, which allows the pancreatic beta cells to mature and differentiate in an environment close to their inherent cellular environment (Georgakopoulos et al. 2020; Grapin-Botton 2016; Loomans et al. 2018).

3.2.8 Limitations of rodents and human beta cell lines

Rodent and human islet cell lines provide valuable tools to study the beta cell physiology in the context of diabetes research. However, it is important to remember that these cells, both human and rodent beta cells, are not equal to primary islet cells and that there are several points that one need to consider when working with beta cell lines. First, even under the best of conditions, the insulin expression in beta cells is much less than in normal primary beta cells, especially in human beta cell line, EndoC-βH1, the insulin gene expression is 10–20 times lower than in primary human beta cells (Scharfmann, Rachdi, and Ravassard 2013). Second, beta cells may lose their beta cell characteristics over a period of continuous growth in culture. Thus, it is important to closely monitor the cells and be careful that other islet cell types do not appear in the culture (Weir and Bonner-Weir 2011). Third, the unlimited growth of many engineered beta cells is linked with their tumor origin and therefore they may acquire changes in chromosomal content, protein expression, metabolism as well as genetic mutations (Skelin, Rupnik, and Cencic 2010). Additionally, the cell cultures are isolated from the influences of other organ systems, such as the liver and the GI-tract, the musculoskeletal and adipose tissues as well as the immune system, which is important for understanding the pathogenesis of various diabetic subtypes.

3.3 CRISPR/Cas9 genome editing

Different genome editing technologies (GETs), such as zinc finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN), have provided valuable tools for researchers to manipulate genomic sequences (Gaj et al. 2016). The same concept of using "sequence-specific DNA-binding domains" that is coupled to a "DNA cleavage nucleases" is behind all these site-specific genome targeting technologies (Gaj, Gersbach, and Barbas 2013). All these technologies have enabled researchers to manipulate the targeted sequence by adding, removing and/or replacing the genetic material at specific locations within the genome. However, most of these earlier GETs are costly, time-consuming and more importantly difficult to reproduce. The genome editing revolution really started when researchers introduced the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated protein 9 (Cas9) system to target and edit double-strand DNA (Figure 9) (Doudna and Charpentier 2014; Zhang et al. 2017).

The CRISPR system is part of the adaptive immune system in bacteria and archaea that enables them to resist repeated viral and bacterial infections. These repetitive DNA sequences or CRISPRs, consisting of 30-40 bp, are separated by sections of DNA called "spacer DNA sequences" that exactly match the viral DNA sequences that the bacteria has already encountered. Upon viral infection, bacteria transcribe these DNA sequences to RNA and this RNA, guides a nuclease, a protein that cleaves DNA, to the viral DNA to cut it apart and disable the virus propagation, which thus finally provides protection against the virus.

Researchers demonstrated that RNAs could be designed to bind to specific DNA sequences, guide a programmable endonuclease Cas9 to the site and subsequently induce DNA double-strand breaks (DSBs) on the target DNA sequences in a site-specific manner (Figure 9). Once the double-strand DNA is cut, the host cell repair machinery is used to repair the genome at the site of Cas9-generated DSB (Barrangou and Doudna 2016). Two distinct mechanisms could be involved in the DSB repair in both prokaryotic and eukaryotic organisms: homology-directed repair (HDR) or non-homologous end-joining (NHEJ) (Barrangou and Doudna 2016). HDR uses a double-stranded DNA donor template that has homology with DNA sequences around the break which needs to be repaired. Although this method is very accurate when

repairing the DNA break it has extremely low efficiency. Quick-fix NHEJ utilizes no or limited homology, is susceptible to errors and frequently results in small insertions and deletions (indels) in the genome. NHEJ is often the pathway of choice to create targeted mutations when utilizing CRISPR/Cas9 system because of its efficiency and flexibility compared to the HDR process (Liu, Rehman, et al. 2018).Testing of this technique lead to a revolution in the genome editing field and revealed that this system can work in almost all types of cells, including human cells.

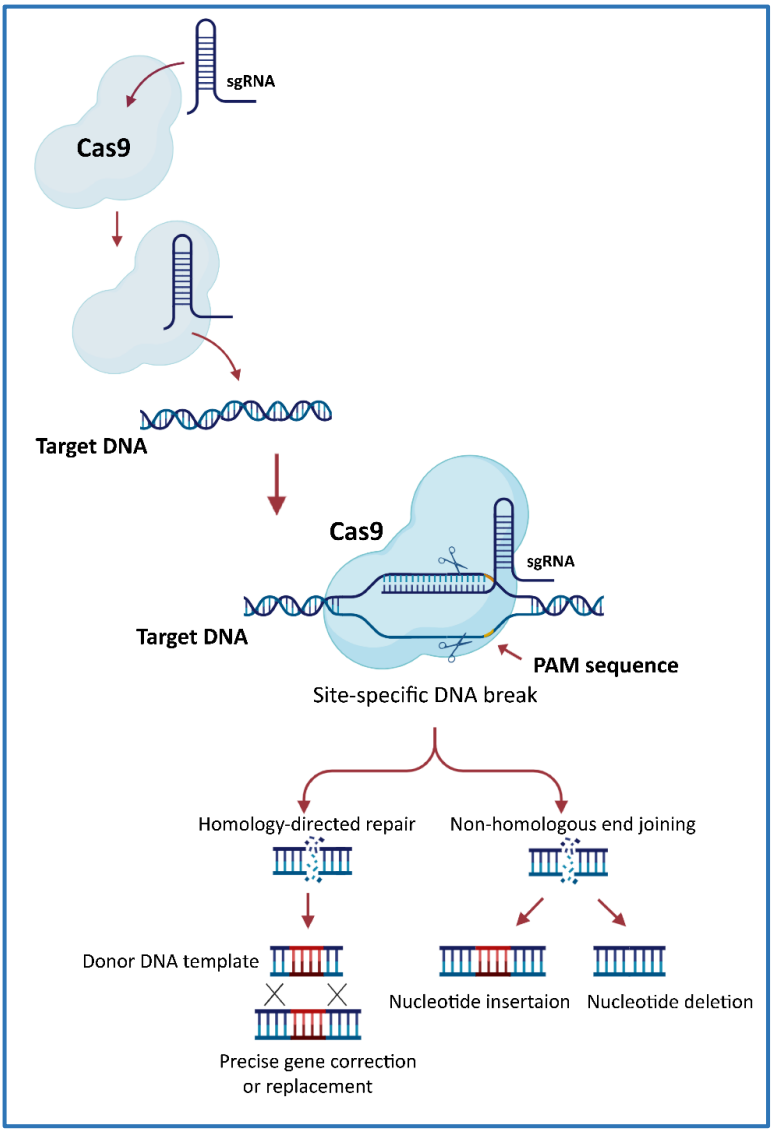

Figure 9. CRISPR/Cas9 genome editing method. Single guide RNA (sgRNA) loaded into the Cas9 endonuclease and subsequently binds to the targeted DNA sequence in close proximity to protospacer adjacent motif (PAM). Upon conformational change of the Cas9 enzyme, which leads to its activation, a DSB takes place in a DNA sequence three nucleotides upstream of the PAM motif. The host cell repairs the DNA break by either NHEJ or HDR endogenous repair machineries. In HDR, an external DNA donor is provided that works as a template for precise repair of the DNA break.

3.3.1 Limitations of CRISPR/Cas9 system

CRISPR/Cas9 system is an extremely powerful tool to target and edit genomic DNA because of its precision, accuracy and speed compared to the other existing genome editing methods. However, it may yet have important limitations that one needs to consider. First, the delivery of CRISPR/Cas9 material to mature cells is challenging and one needs to test different delivery systems to overcome this problem. However, viral vectors seems to be the most commonly used CRISPR/Cas9 delivery method. Second, CRISPR/Cas9 system in not 100% efficient, which means that even the cells that take up CRISPR/Cas9 may not have genome editing activity. Third, the CRISPR/Cas9 system is not 100% accurate, which means that uncontrollable “off-target” mutations, may happen and lead to adverse consequences to the target cell or organism (Hsu, Lander, and Zhang 2014; Manghwar et al. 2019).

3.4 Enrichment of proteoglycan and glycosaminoglycan derivatives

The first step in the characterization of PGs and their attached GAG chains, is to extract them from their native biological sources. There are many different protocols published, depending on what sources and what PGs you are interested in to study (Whitelock and Iozzo 2002; Woods and Couchman 2018). The GAG extraction methods from diverse sources should be appropriate for eliminating and/or decreasing contaminants that may produce noise background in mass spectrometric analysis (Staples and Zaia 2011). In this thesis, the focus has been directed towards biological fluids and cell media and the soluble PGs identified in those.

The enrichment of PGs with intact GAG chains is facilitated by their highly negatively charged nature due to the presence of carboxylic acid-containing sugars and/or sulfate groups (Prydz 2015). The negative charges of GAG chains thus allow for their separation from neutral or positively charged molecules by anion-exchange chromatography. In this technique, positively charged functional groups, which are covalently bound to a solid matrix in the chromatography column, are binding to the negatively charged target molecules. Initially, the molecules that are not or weakly bound to the stationary phase are washed away and then the molecules that are more strongly bound to the stationary phase are eluted by increasing the pH and the salt concentrations of the buffer, which enables either a gradient elution or an "en-bloc" elution.

Chromatography columns are filled with either strong or weak functional groups that reflect the ability of the exchange matrix to maintain its charge with changes in pH conditions. Strong exchangers that are used in strong-anion exchange chromatography (SAX), have functional groups that remain ionized regardless of the buffer pH and a wide range of pH buffers (pH 0-14) can be used. Quaternary (Q) ammonium cation resins are the most common functional group that are used in SAX methods. Weak ion exchangers that are used in weak-anion exchange chromatography (WAX), can be neutralized by changing the pH and deliver optimal performance only over a small pH range (pH 5-9). Diethylaminoethyl (DEAE), is one of the weak anion exchangers.

Anion-exchange chromatography allows for enrichment of PGs and GAGs from different sources and. However, in this method, the molecules are separated based on their charge only, thus, other molecules with negative charge will be enriched together with the PGs and GAGs. This is the case for acidic proteins and peptides as well as for the negatively charged nucleic acids.

3.4.1 Protease digestion

In order to increase PGs sequence coverage in the down-stream LC-MS/MS analyses, the core proteins need to be proteolytically digested with one single protease or with a combination of proteases. Different proteases are able to hydrolyze the peptide bonds of core proteins at different sites depending on the specificity of the proteases, thus giving rise to a mixture of naked peptides and glycopeptides that carry glycan modifications, e.g. the GAG chain(s). Trypsin is the most-frequently used protease for MS-based (glyco)proteomics sample preparation, mostly due to its high specificity, efficiency and availability (Giansanti et al. 2016). Trypsin cleaves peptide bonds at the C-terminal side of Arg and Lys residues, giving rise to the mixture of positively charged peptides with different lengths (Olsen, Ong, and Mann 2004). To promote the proteolytic digestion of proteins, a reducing agent such as dithiothreitol (DTT) or 2-ME is commonly applied to break the disulfide bridges followed by alkylation with iodoacetamide (IAA) to prevent the re-formation (Shajahan et al. 2017).

One drawback of using trypsin as a single protease is that a high percentage of tryptic peptides are too short (56% ≤ 6 residues) and some too long to become unambiguously identified by MS. This means that the protein sequence coverage becomes limited and different protein isoforms cannot be clearly distinguished (Tsiatsiani and Heck 2015). Digesting the same proteome with multiple proteases can overcome these restrictions and lead to an increased number of identified peptides and proteins, as well as the identification of new PTM sites (Choudhary et al. 2003; Dau, Bartolomucci, and Rappsilber 2020; Giansanti et al. 2016; Swaney, Wenger, and Coon 2010). Besides trypsin, other serine proteases such as lysyl endopeptidase (Lys-C), chymotrypsin, glutamyl peptidase I (Glu-C) and pepsin are the proteases most often used to obtain non-tryptic data sets giving complementary information (Tsiatsiani and Heck

2015). In Paper II, in addition to the sole use of trypsin we also used chymotrypsin and a combination of both enzymes. Chymotrypsin preferably cleaves amino acid sequences at the aromatic residues; tyrosine, phenylalanine and tryptophan and to a lesser extent at leucine and methionine.

In addition, and particularly in cases where glycoproteins are known to undergo substantial protease digestion by the action of endogenous proteases, LC-MS/MS experiments may be conducted without the usage of any proteases. In this way, native protease cleavage sites of CgA peptides and glycopeptides were investigated (Paper II). Experiments were conducted both with and without the use of trypsin to provide complementary coverage of the sequences.

Trypsin digestion of glycoproteins results in the generation of both glycosylated peptides as well as non-glycosylated peptides that may suppress the glycosylated peptides signals. In addition, the tryptic-generated glycosylated peptides are sometimes too long for MS analysis, since glycan modifications per se may interfere with the protease ability of trypsin to hydrolyze adjacent peptides bonds. Another approach to hydrolyze the (glyco)proteins is then to use an unspecific proteolytic enzyme such as pronase. Pronase is a mixture of several unspecific endo- and exoproteases that are able to completely hydrolyze the proteins and leave glycopeptides with a very short peptide portion. The benefits of using pronase digestion include reduced sample complexity due to substantial protein digestion and thus facilitated data interpretation of the glycan part. On the other hand, the core protein identification of glycopeptides becomes limited since the generated peptide parts of the glycopeptides are relatively short (Dodds et al. 2009). In Paper III, we used pronase to more or less completely digest the PG core proteins, to facilitate the simultaneous analysis of the released GAG conjugates and the short glycopeptides using negative mode LC-MS/MS.

3.4.2 Oligonucleotide digestion

In Paper III, we used benzonase, a commercially available endonuclease, to eliminate both DNA and RNA (single and double stranded) from sample and also to improve the GAG depolymerization with chondroitinase ABC

(ChABC) (Further described in *Depolymerization of glycosaminoglycans*). Low specificity activity of benzonase, results in the degradation of all types of nucleic acids without any base preference. This enzyme has no proteolytic activity and is commonly used in different studies to eliminate nucleic acids contamination (Ledin et al. 2004; Shao et al. 2013).

3.4.3 Depolymerization of glycosaminoglycans

Characterization of GAG chains is challenging due to their heterogeneous nature and large size. To release the GAG chains from their core proteins, alkaline β-elimination can be performed leading to the liberation of the GAG chains and formation of a xylitol (Xyt) reducing end (Ly, Laremore, and Linhardt 2010). Following release, the GAG chains can be further enriched by anion-exchange chromatography. Partial enzymatic depolymerization of GAG chains, is a way to reduce their sizes and make the analysis of their structures more feasible. Specific bacterial enzymes that degrade the GAG chains at expected sequences are commonly used for GAG structural analysis. Depolymerization of GAG chains with such enzymes results in the creation of oligosaccharides of different lengths depending on the GAG structure and the specificities of the enzymes used. There are several chondroitinases and heparinases that cleave CS/DS and HS GAGs into different lengths. These enzymes are lyases, which means that they catalyze the cleavage of GAG chains via elimination reactions. For dissecting the CS/DS structures different chondroitinases are available; ChABC is the most commonly used enzyme to digest CS/DS structures at both GlcA and IdoA residues, chondroitinase AC (ChAC-I and ChAC-II) digests CS/DS structures at GlcA and chondroitinase B (ChB) digests CS/DS structures at IdoA residues (Figure 10) (Linhardt et al. 2006). ChABC cleaves the β1,4-linkage between GalNAc and GlcA/IdoA residues and generates 4,5-unsaturated hexuronic acids (HexA) at the non-reducing ends of disaccharides and hexasaccharide linkage region products. GlcA and IdoA residues are not distinguishable after degradation with ChABC (Ernst et al. 1995). The combined use of ChAC-I (endolyase) and ChAC-II (exolyase) results in the generation of products with different lengths (Linhardt et al. 2006). When there is no IdoA presents, degradation with ChAC generates

disaccharides and if there are IdoA-containing domains (CS/DS chain), oligosaccharides are generated (Zhao et al. 2013).

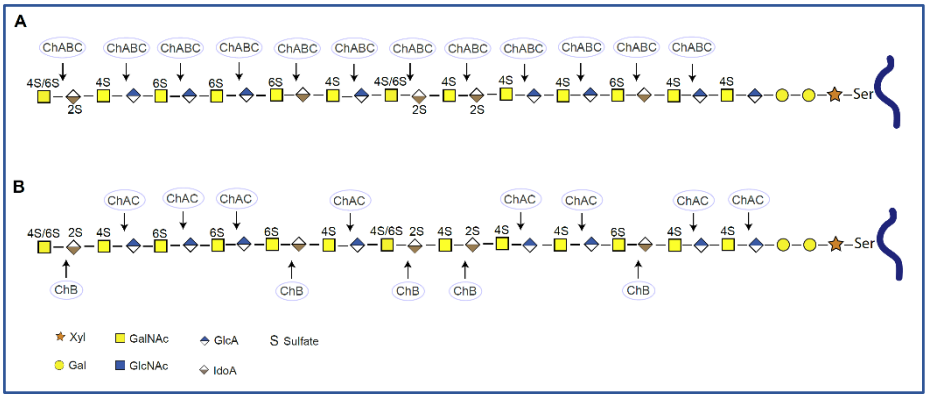

Figure 10. Depolymerization of GAG chains using bacterial lyases ChABC (A), ChAC and ChB (B) indicating their substrate specificities along the CS/DS chains.

A recent study comparing the different chondroitinases (ChABC, ChAC and ChB) from different vendors, showed that there were variations in enzymatic activities, substrate specificities and thus in the GAG product profiles when using different enzymes (Toledo et al. 2020). This underscores the importance of exploring the enzyme specificities and activities by using GAG standards and biological controls (Toledo et al. 2020).

To elucidate the heparin and heparan sulfate structures various microbial heparinases with known substrate specificity are available. Heparinase I cleaves the glycosidic bond between GlcNS±6S and 2-*O*-sulfated IdoA residues within the highly sulfated sequences (NS domains) of HS and H. Heparinase II has a broader specificity than heparinase I and is able to degrade highly sulfated (NS), unmodified *N*-acetylated (NA) and NA/NS domains. Heparinase III, acts in non-sulfated NA and NA/NS domains of HS and heparin (Desai, Wang, and Linhardt 1993; Dong et al. 2012).

In this thesis we used bacterial lyases for depolymerization of CS/DS and HS GAG chains. ChABC was used in Paper I and IV, ChABC and heparinases (II/III) in Paper II and in Paper III we used ChABC, ChAC (-I and –II) and ChB.

3.5 Measuring glycosaminoglycan quantities

Several GAG quantification methods have been developed to measure both intact GAGs and disaccharides generated after GAGs depolymerization (Kubaski et al. 2017). Dye-based spectrophotometric methods such as alcian blue (Bjornsson 1998) and 1,9-dimethyl-methylene blue (DMMB) (Farndale, Sayers, and Barrett 1982), have been used to measure total GAG quantities from different biological samples. More specific and sensitive methods such as enzyme-linked immunosorbent assay (ELISA) assays for detecting GAGs in blood and urine samples and HPLC to detect each specific GAG has been developed. The drawback of such methods is that they are in general time and money consuming (Kubaski et al. 2017). Tandem mass spectrometry (MS/MS) analysis of GAG-derived disaccharides is a very sensitive and specific method that has been developed and applied for different biological samples, which will be further described in a specific paragraph.

The DMMB method was used to provide an approximate estimation of the GAG concentration in Papers II, III and IV. DMMB is a positively charged dye that binds to negatively charged molecules including sulfated GAGs, which results in an absorbance shift. The shift in absorbance is in the visible spectrum and thus matches a shift in color, from 595 nm for blue color to 510-530 nm for purple color, which is measured by spectrophotometry. These methods have been modified by different research groups and is one of the most commonly used methods to quantify GAGs. In general, the DMMB method is fast and simple but has some limitations. The DMMB dye can bind also to other negatively charged molecules than GAGs, such as DNA and acidic proteins that are present in the sample. Moreover, other factors like the purity of the dye itself and the instability of DMMB-GAG complex, may affect the test results (Kubaski et al. 2017). As mentioned above, in our work, the DMMB method was used mainly to appreciate of the amounts of GAGs in various preparations, especially during the optimization of our GAG purification methodology used in Paper III.

3.6 Liquid chromatography tandem mass spectrometry

Mass spectrometry has provided a powerful analytical tool for the identification and quantification of a diverse set of molecules by converting them to the ionized forms followed by the separation, fragmentation and detection of gas-phase ions by their mass-to-charge ratio (*m/z*) (Pitt 2009). Upon introducing the sample into the MS ion source, multiple gas-phase ions are generated from the same sample and then separated based on their *m/z* values and finally the relative abundance of each ion is measured with the detection system (Ho et al. 2003). Several different technologies have been developed for ionization, and analyzing the gas-phase ions resulting in the generation of different mass spectrometry instruments. However, all MS instruments are composed of three major parts: an ion source, mass analyzer, and detector system (Han, Aslanian, and Yates 2008).

Matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) are two advanced soft ionization methods (without or with low level of fragmentation of the primary analyte) (Figure 11). In the MALDI technique, samples are co-crystallized with a matrix, and the ions, mostly singly charged, are generated from large and small biomolecules by irradiating a matrix with a pulsed laser (Zabet-Moghaddam et al. 2004). In recent years, ESI based LC-MS/MS has become the dominating technique used in proteomics and glycoproteomics. In the ESI method, liquid samples containing the analytes are pumped through a capillary and sprayed into the ionization chamber. A high voltage (generally 2-5 kV) is applied to the capillary tip and produces charged droplets via solvent evaporation. These charged droplets are further evaporated and dispersed into often highly charged ions (Banerjee and Mazumdar 2012; Pitt 2009). In this ionization method, there is very little in-source fragmentation; thus the analyte ions remain intact. Samples may be run in either positive (protonation of the analyte) or negative (deprotonation of the analyte) ion modes. In positive mode, positively charged ions are formed by protonation at low pH; and for negative mode, molecules are deprotonated at a pH higher than the isoelectric point of the analytes. The amount of gas-phase ions that are generated from a specific analyte, the ionization efficiency, is different when different ionization modes (positive or negative mode) are used. In addition to the ionization mode, the composition of the solvents affects the

ionization efficiency, and thus the sensitivity and detection limit of the ESI method (Liigand et al. 2017). The ESI method is now widely used for the identification of polar molecules such as (glyco)proteins and (glyco)peptides. The ESI ion source is often coupled to a reversed-phase LC system employed to fractionate the analytes prior to the MS analysis, and to improve the sensitivity of the ESI step through the pre-concentration and desalting of the sample (Gundry et al. 2009). The highly charged ions that are generated in the ionization chamber are accelerated into the mass analyzer where they are filtered according to their *m/z* values in electric and/or magnetic fields. These primary ions (precursor ions) are typically also fragmented in parallel or sequentially prior to the mass analysis, which is referred as a tandem mass spectrometry (MS/MS or MS^n indicating that that the fragmentation can be repeated several (n) times) (Banerjee and Mazumdar 2012).

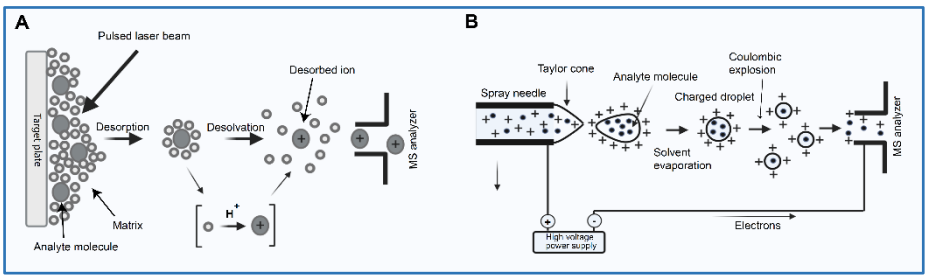

Figure 11. Principles for the most commonly used soft ionization methods MALDI (A) and ESI (B).

Different mass analyzers are available for the mass measurements and filtering of ions include time-of-flight (TOF), the quadrupole (Q), and the ion trap (IT). The TOF analyzer basically measures the time that it takes for a specific ion to travel a particular distance with a specific kinetic energy. The quadrupole analyzer is composed of four cylindrical metal rods located in parallel to each other, and ionized ions are selectively transmitted through the electrodynamic field between these four rods. The ion trap analyzer is a three-dimensional quadrupole which can both measure and filter ions for further fragmentation events (Banerjee and Mazumdar 2012).

The Orbitrap, from Thermo Scientific, is a unique pulsed ion trap mass analyzer with central spindle-like electrodes and outer barrel-like electrodes, which has a very high accuracy and resolving power, which makes it a common

choice for analysis of very complex samples (Perry, Cooks, and Noll 2008). In particular, Orbitrap instruments are assembled into "hybrid" instruments where two or more types of analyzers are combined in order to obtain the best feature of each analyzer and improve the performance and resolution of the MS/MS analysis (Glish and Burinsky 2008). The final part of the MS instrument is a detector, which converts the current of ions into electric signals and further into intensity values to be displayed as the mass spectrum (Banerjee and Mazumdar 2012).

In this thesis work, LC-MS/MS based (glyco)proteomics and glycomics has been implemented for the analysis of PG-derived glycopeptides, glycans and peptides. In Papers I, II and IV glycoproteomics samples were run on an Orbitrap Fusion Tribrid mass spectrometer coupled to an Easy-nLC system. The Orbitrap Fusion Tribrid mass spectrometer is a tri-hybrid instrument that has a combination of quadrupole, linear ion trap and Orbitrap mass analyzers. The combination of mass analyzers, high mass accuracy and speed offers a possibility to identify low-abundance and complex glycoconjugates. Nanospray ionization, which is a developed downscaled ESI, has been commonly used as an ion source, which requires very low amounts of sample (in concentration as well as in volume) compared to the standard ESI process (Banerjee and Mazumdar 2012). In Paper III, glycomics analysis of samples was performed on a linear ion-trap quadrupole (LTQ) Orbitrap Elite mass spectrometer in the negative ESI mode. This instrument is a linear ion trap Orbitrap hybrid mass spectrometer with a high resolving power that has been used in both bottom-up and top-down experiments (Michalski et al. 2012).

Using MS instruments equipped with an Orbitrap analyzer for both glycoproteomics and glycomics approaches allowed us to use higher-energy collisional dissociation (HCD) for the MS/MS fragmentation of the sample molecules. The combination of the quadrupole mass filter, the HCD collision cell and the Orbitrap mass analyzer result in a high-resolution fragmentation technique that is suitable for the analysis of protein modifications, including GAG glycosylation (Yu et al. 2018). HCD also benefits from the possibility to produce MS/MS spectra at the full *m/z* 100-2000 region enabling the simultaneous detection of diagnostic oxonium ions in the low mass region and glycopeptide fragments at the higher *m/z* values. Moreover, HCD fragmentation is well suited for selective fragmentation of the glycan and peptide cores using different levels of energy specified as normalized collision

energies (NCEs). When applying stepped collision energy HCD, or the alternate use of both low and high NCEs on the same precursor, for the analysis of glycopeptides, the higher HCD energies are employed for peptide fragmentation and the lower HCD energies for glycan fragmentations (Cao et al. 2014; Gomez Toledo et al. 2015; Jedrychowski et al. 2011; Riley et al. 2020; Yang, Yang, and Sun 2018). HCD was the default dissociation technique applied in all Papers (I-IV).

3.7 Proteomics

Proteomics is the science of studying the "proteome", the complete set of proteins that are expressed under a particular condition from a cell, tissue or organism. Several different modifying agents such as apoptosis, differentiation and different pathological conditions can affect the proteome. The final goal of proteomics is to correlate the identified proteins (including protein isoforms and splice variants), their PTMs, biological functions and interactions with and without the presence of the modifying agents in a certain cell at a given time (Altelaar, Munoz, and Heck 2013). The development of new MS instrument technologies with higher resolution and the lower detection limit, has significantly affected the proteomics field. However the proteomics analyses are still challenging due to the considerable heterogeneity and complexity of proteins/peptides in biological samples and the low abundance of some proteins in such samples (Gulcicek et al. 2005).

Mass spectrometry is the method of choice to identify intact proteins or a mixture of peptides from different biological sources (Aebersold and Goodlett 2001; Aebersold and Mann 2003; Bantscheff et al. 2007). Two types of approaches are commonly applied for proteomics analysis; "bottom-up" and "top-down" approaches. In bottom-up approach, MS-MS techniques are used to identify peptides ($0.7 < Mw < 3$ kDa) that are obtained from enzymatic or chemical digestion of intact proteins. In contrast, in top-down approach, intact small proteins ($10 < Mw < 30$ kDa) are directly analyzed by MS or MS/MS techniques. More recently, a "middle-down" approach has been introduced to identify large peptides by combining the benefits of both top-down and bottom-up approaches (Wu et al. 2012). In this method, generated peptides after enzymatic digestion are larger (≥ 3 kDa) and fewer in numbers compared

to the bottom-up method, which increases the proteome coverage (Cristobal et al. 2017).

3.8 Glycoproteomics

A glycoproteomics approach combines the "proteomics" knowledge of core proteins with the "glycomics" knowledge of the attached glycan(s). Such approaches resolve the structural identities and sometimes the relative quantities of the glycan structures that are bound to specific amino acid residues, the glycosites, of any particular glycoprotein appearing under particular conditions of time, location, and environment (Rudd et al. 2015). The total composition of different glycans that are expressed within a particular cell at a certain time is often called the "glycome" or when linked to proteins the "glycoproteome". In analogy with this nomenclature the global analysis of PGs at any given location and time has been named "proteoglycomics"(Ly, Laremore, and Linhardt 2010). In humans, over 214 distinct genes are encoding the glycosyltransferases that are involved in 15 different glycosylation pathways known for human cells. In addition to glycosyltransferases, other enzymes such as sulfotransferases, epimerases and transporters are involved in the biosynthesis contributing to the structural diversity of glycans, thus increasing the complexity of the glycome (Joshi et al. 2018; Mizumoto, Yamada, and Sugahara 2014). The necessity of exploring the glycome, as well as single glycans and glycoproteins, has become even more evident when a vast number of studies have shown that the glycan repertoire change during both physiological and pathological processes (Rudd et al. 2015).

The complexity of the glycoproteome has made it impossible to use a single technique for the identification and quantification of all glycoproteins. Thus, to make it an easier task, several approaches may be applied in parallel (Rudd et al. 2015). One approach is to release the glycans from proteins and then analyze the proteins and the released glycans separately, preferentially using different LC-MS/MS analyses. Although this approach provides information on any glycan structure, the identity of the core proteins and also the attachment site of each glycan cannot be obtained from such an approach. Information on the exact attachment site of glycans to selected core proteins is

important to further understand and explore the glycan-mediated functions of glycoproteins (Varki 2017). Thus, to obtain a more comprehensive picture of the glycan structure, assign the glycosite, and to identify the corresponding glycoprotein, the glycosidic link between the glycan and the protein must be preserved. Different ionization patterns of GAGs and core proteins, as well as GAGs size heterogeneity and complexity has been considered as major obstacles for the analysis of PGs for a long time (Ly, Laremore, and Linhardt 2010). However, many different approaches have been developed and customized according to the nature of the biological samples, the composition of the glycans and the information required. These MS-based methods have been developed to identify glycoproteins, quantify and characterize glycans at their specific glycosite as well as determine the level of the glycosite occupancy in a single analysis (Chi et al. 2008; Gomez Toledo et al. 2015; Olson et al. 2006; Zaia 2010). The development of such methods is not only of value for the characterization of PGs but has also been implemented for the analysis of *N*- and *O*-linked glycans of other glycoproteins (Halim et al. 2013; Li et al. 2019; Nilsson et al. 2009; Thaysen-Andersen and Packer 2014; Thaysen-Andersen, Packer, and Schulz 2016; Zhang et al. 2003). Such information is needed to build a structural platform for further unravelling the biological importance of glycans and relate the structure of glycans to their functions.

Glycoproteomics analysis utilizes the bottom-up and top-down approaches, similar to those that are used for proteomics analysis (earlier described in the paragraph *Proteomics*). In the bottom-up approach, glycoproteins are digested and generated glycopeptides are characterized by LC-MS/MS, whereas in the top-down glycoproteomics approaches, intact glycoproteins are subjected to further analysis by LC-MS/MS (Yu et al. 2018). There are presently several levels of limitation in glycoproteomics analysis, including the need for manual interpretation and verification of data due to the lack of profound bioinformatics tools suitable to efficiently handle the vast amount of data obtained from the high-throughput LC-MS/MS analyses (Further described below in the paragraph *Bioinformatics*).In addition, MS/MS fragmentation cannot generally provide detailed linkage and stereochemistry information. Complementary analytical techniques such as gas chromatography-mass spectrometry (GC-MS) and nuclear magnetic resonance (NMR) spectroscopy may enable researchers to obtain this information (Rudd et al. 2015; Shajahan et al. 2017) but usually demands much higher amounts of sample.

3.9 Glycomics

The term “glycomics” describes the study of all glycans that are synthesized by a cell or a tissue under particular conditions, the “glycome” (Bertozzi and Sasisekharan 2009). Glycans are attached to the core protein either through *N*-linked (attached to the amino group of the Asn) or *O*-linked (attached to the hydroxyl group of Ser/Thr/Tyr) glycosylation. Glycans have a high heterogeneity both regarding their distribution on a core protein (microhetergeneity) and their occupancy at a specific glycosite (macroheterogeneity), which makes the glycan analysis challenging (Cummings and Pierce 2014). Despite the analytical challenges, several MS-based glycomics methods have been developed and applied successfully to study *N*- and *O*-linked glycans (Han and Costello 2013; North et al. 2009).

Structural analysis of GAG glycans is extremely challenging due to the heterogeneous nature and large size of GAGs. To overcome this complexity, one approach is to release the GAGs from core proteins followed by partial or complete lyase digestion to produce a less complex range of oligosaccharides which are subsequently analyzed by bottom-up MS analysis (described in the paragraph *Proteomics*). MS/MS approaches have emerged over the past few years to decipher the GAG structures (Staples and Zaia 2011; Volpi et al. 2014; Zaia 2010, 2008). A recent development of GAG domain mapping, GAGDoMa approach, facilitated the structural analysis of linkage regions, internal and terminal non-reducing ends (NREs) of GAG chains (Persson et al. 2020). In this method GAG chains are depolymerized by bacterial lyases, then separated by reversed-phase dibutylamine (DBA) ion-paring chromatography prior to introducing the separated molecules to the MS/MS analysis (Persson et al. 2020). In addition to sequencing GAG chains by bottom-up approaches, a top-down glycomics approach (described in *Proteomics*) has recently been successfully implemented to characterize GAGs of the PGs bikunin (Ly et al. 2011) and decorin (Yu et al. 2017).

Although these glycomics methods provide valuable information on the structural composition of the attached glycans, they don’t allow for the characterization of glycosidic linkage positions and the carbohydrate configurations. This limitation has recently been overcome by the introduction of the ion mobility mass spectrometry (IMMS) method as a powerful analytical technique for the separation and structural characterization of glycan isomers (Lanucara et al. 2014; Mookherjee and Guttman 2018). IMMS is able to

distinguish ions with the same masses and provides information on the isomeric carbohydrates, their configurations and glycosidic bounds (Gray et al. 2016; Gray et al. 2017; Miller et al. 2020).

In Paper III, we applied a glycomics approach similar to the one that has been used to characterize and semi-quantify xyloside-primed GAGs from human breast cancer cells and fibroblasts (Persson et al. 2018). Some modification in the previously published method, has now been introduced for the analysis of PGs, including benzonase and hyaluronidase digestions. Samples were run on Orbitrap Elite mass spectrometer in the negative mode with HCD fragmentation. The LC step was performed in the presence of dibutylamine (DBA) as ion-pairing agent which significantly improves the chromatographic separation and also reduces the formation of alkali adducts (Kuberan et al. 2002).

3.10 Bioinformatics

Another limiting factor when studying the MS-based "omics" techniques, is processing and analyzing the large amount of data obtained in the MS/MS analysis (Chandramouli and Qian 2009). This becomes even more complicated when it comes to the glycopeptide analysis and there is a need for more effective database and search algorithms to overcome the present limitations (Darula and Medzihradszky 2018; Li et al. 2020).

In bottom-up proteomics, data are commonly obtained by either data-dependent acquisition (DDA) or data-independent acquisition (DIA) methods, depending on the purpose of the study. In the DDA methods that have been used extensively for proteomics analyses, the most intense precursor ions at a given retention time at the MS1 stage are fragmented in the MS2 step. In DIA methods, all precursor ions that are in the same mass window (isolation window) are fragmented at MS2 stage, which means that all detected precursor ions are MS2 fragmented regardless of their intensities. Obtaining proteomics data by the DIA method, makes the analysis even more complex since the resulting DIA MS/MS data are very complicated and need to be analyzed by very robust analysis tools (Doerr 2015; Guan et al. 2020). Thus, the MS-generated data must be converted to an appropriate format used in the search

protocols. There are several database search programs such as SEQUEST, Mascot and X!Tandem that facilitate the accurate and sensitive peptide identification from MS/MS spectra (Huang et al. 2012). In Papers I, II and IV, the Mascot search engine was implemented (Perkins et al. 1999). Mascot is used to identify proteins from peptide sequences by using a probability-based scoring algorithm which allows it to score experimental peptides against the database of known peptides. The mascot score is -10Log10 (P), where the P stands for probability; thus the best scored peptides in mascot are the ones with lowest probability to occur by chance (Perkins et al. 1999).

For the identification of glycopeptides using the Mascot search engine, the masses of the anticipated glycans are added as variable modifications in the Mascot database searches. For HCD fragmentation, the used modifications must also include subtraction of the same masses from the peptide fragment ions including the glycosite (Nasir et al. 2016). In Papers I and II, the masses of the hexasaccharide including 0-2 sulfate/phosphate groups were used as the allowed modifications of Ser residues for the CS/DS core proteins. Also, in Paper II the masses of the tetrasaccharide including 0-1 sulfate/phosphate groups were used as the allowed modifications of Ser residues for the HS core proteins. Bioinformatics tools may be used to verify glycopeptide hits (Nasir et al. 2016), but for Papers I and II, manual verifications of the fragmentation spectra from the Mascot hits were performed. In order to be considered a true hit, the presence of the fully de-glycosylated but otherwise intact peptide ion had to be present in the fragment spectrum. Also, the presence of an oxonium ion at *m/z* 362.10 significant for the hexasaccharide structure had to be present for verification of the CS/DS glycopeptide hits (Gomez Toledo et al. 2015).

"We are not lost. We're locationally challenged."

John M. Ford

4 RESULTS

As discussed in the introduction, PGs are very complex structures, and studying them is hindered by several limiting factors. In this thesis work, we aimed to validate a glycoproteomics method that allows us to characterize the linkage region composition and identify the core protein while the linkage region is still attached to the Ser glycosite. This novel method, couples the enrichment and enzymatic depolymerization of GAG chains with a high-resolution nLC-MS/MS analysis. In this method, discussed in a Paper I, the complexity of the CSPG containing sample, in this case CSF, was reduced by trypsin digestion, which tends to generates peptides with optimal sizes for MS/MS analysis. The glycopeptides that carry CS chains were enriched by SAX chromatography and then ChABC was used to depolymerize the CS polysaccharides into the much less complex hexasaccharide structures. The hexasaccharide linkage region generated after ChABC depolymerization is composed of the tetrasaccharide linkage region plus a dehydrated disaccharide. The samples were analyzed on an Orbitrap Fusion mass spectrometer coupled to nLC for the better separation of peptides. The raw data was processed by proteomics software and analyzed with the Mascot database search algorithm including the allowed fixed and variable modifications for glycopeptides identification. All glycopeptide results were further verified by manual interpretation to avoid false positives. The positioning of phosphate, sulfate, and sialic acid modifications was also done manually. This approach was previously applied for several biological complex samples, including CSF, Caenorhabditis elegans, urine, cells and cell culture media from different human and rodent origins. Applying this method for identifying CSPGs in the CSF sample has resulted in the global site-specific identification of several novel core proteins including CgA, CgB, and neuropeptide W.

Considering that the understanding of the GAG glycosylation pattern of SGs may benefit the understanding of the biological role(s) of PGs or their GAG chains in endocrine cell physiology, we designed the work that is presented as

Paper II in this thesis. In Paper II, we aimed to investigate if the SGs proteins in neuro(endocrine) cells, known to have specialized SGs, carry CS and/or HS GAG chains. We applied the method that we developed in Paper I, and optimized it according to the type of samples used in Paper II. Using our method, we identified core proteins, their CS and/or HS GAG chains and the sites that these are attached to in the core proteins. We used endocrine insulin-secreting cells from human islets, rat (INS-1 832/13) and mouse (MIN6 and NIT-1) insulinoma cell lines as models for GAG isolation and identification. We identified CgA with CS modification in all the cell lines and in human islets. In rat insulin-secreting cells, CgA carried either CS or HS GAG chains, thus this is a so called hybrid site. Besides CgA prohormone, we identified other core proteins that carry CS GAGs, including islet amyloid polypeptide (IAPP), secretogranin-1 (SCG1), secretogranin-2 (SCG2), and immunoglobulin superfamily member 10 (IGSF10). Depolymerization of GAG chains with heparinases II and III, enabled us to identify two known HSPGs, i.e. syndecan-1 (SDC1) and syndecan-4 (SDC4), and three novels HSPGs including CgA, SCG1, and neurexin-2 (Nrxn2). Together, we could identify several different CSPGs and HSPGs in human, rat and mouse insulin-secreting cells (Figure 12). However, their biological roles remain to be determined.

In Paper III, we set up a novel LC-MS/MS method for structural analysis of linkage regions, internal oligosaccharides and NREs of CS/DS GAGs from PGs secreted from rat INS-1 832/13 cells. This method was adapted and modified from the glycosaminoglycan domain mapping, GAGDoMa, method of structurally analyzing complex mixtures of CS/DS primed on xylosides. To study the GAG chains derived from PGs, we isolated CS/DS GAG glycopeptides from their corresponding core proteins, enriched from culture media of INS-1 832/13 cells, by extensive digestion of the PGs using a non-specific protease, pronase. The released CS/DS GAG conjugates were then isolated by SAX chromatography, followed by benzonase and hyaluronidase digestions. The optimization of the sample preparation step by including benzonase and hyaluronidase digestions resulted in more complete digestions of GAG chains by ChABC and ChAC enzymes and in cleaner, more enriched, CS/DS preparations. Isolated GAG conjugates were further orthogonally depolymerized by bacterial lyases, analyzed by nLC-MS/MS in negative acquisition mode, and the data was processed by Proteome Discoverer software. In addition to applying Proteome Discoverer software for LC-MS

peak detection and intensity measurements that enabled us to replicate samples to provide a more reliable semi-quantification, all data were analyzed manually. In summary, this method facilitated the detailed structural description and semi-quantification of the terminal, internal, and linkage region domains of the CS/DS on CgA, the major PG produced by the INS-1 832/13 cells. The data showed that in rat INS-1 832/13 cells, the secreted GAGs were heterogeneous because of the presence of both co-polymeric CS/DS GAGs and CS GAGs of secreted PGs. Different modifications of the linkage region such as sulfation, phosphorylation, and sialylation increased the heterogeneity of the CS/DS linkage regions. However, the CS/DS GAG chains from CgA, appeared to have essentially only one 4-*O*-sulfate modification per GalNAc residue and thus, they showed rather simple repetitive internal structures.

By developing two advanced glycoproteomics and glycomics approaches that enabled us to evaluate the GAG glycosylation status of PGs in insulin-secreting cells, we further aimed to combine our novel methodologies with other techniques to better understand the biological role(s) of GAG-glycosylation in rat INS-1 832/13 insulin-secreting cells. Thus, in Paper IV, using the CRISPR/Cas9 genome-editing system, we generated stable beta-1,4-galactosyltransferase 7 (B4galt7) knock-down (KO) INS-1832/13 cells. The idea was to block an early step in the GAG biosynthesis in the cell machinery; thus, CgA and other GAG-glycosylated core proteins would remain without GAG chains. Several of the identified CSPGs and HSPGs in Paper II, are prohormones. Prohormones are typically stored and also processed within DCGs before they are released from neuroendocrine cells upon stimulation. Among all the prohormones stored in DCGs, CgA has been shown to have a critical role in DCG formation through its C-terminal derived peptide, serpinin. Notably, in Paper II, we showed that the GAG glycosite is close to the serpinin cleavage site. However, the involvement of the CgA GAG structures in such an effect as the processing of CgA itself is not yet studied. In summary, the CRISPR/Cas9 technique was used to establish multiple clones with a genetic block in the GAG biosynthesis, followed by phenotypic characterization of the cells by assessing possible changes in cell morphology, CgA expression, CgA localization in SGs, and in PG structures. Our data showed that B4galt7-KO cell lines had a significant, but not complete, block of the GAG glycosylation of the CgA protein. Inspection of the knock-down cells showed that these cells are more round and smaller than the WT cells. Immunostaining of CgA showed

an altered distribution of this protein in the knock-down clones when compared to WT cells.

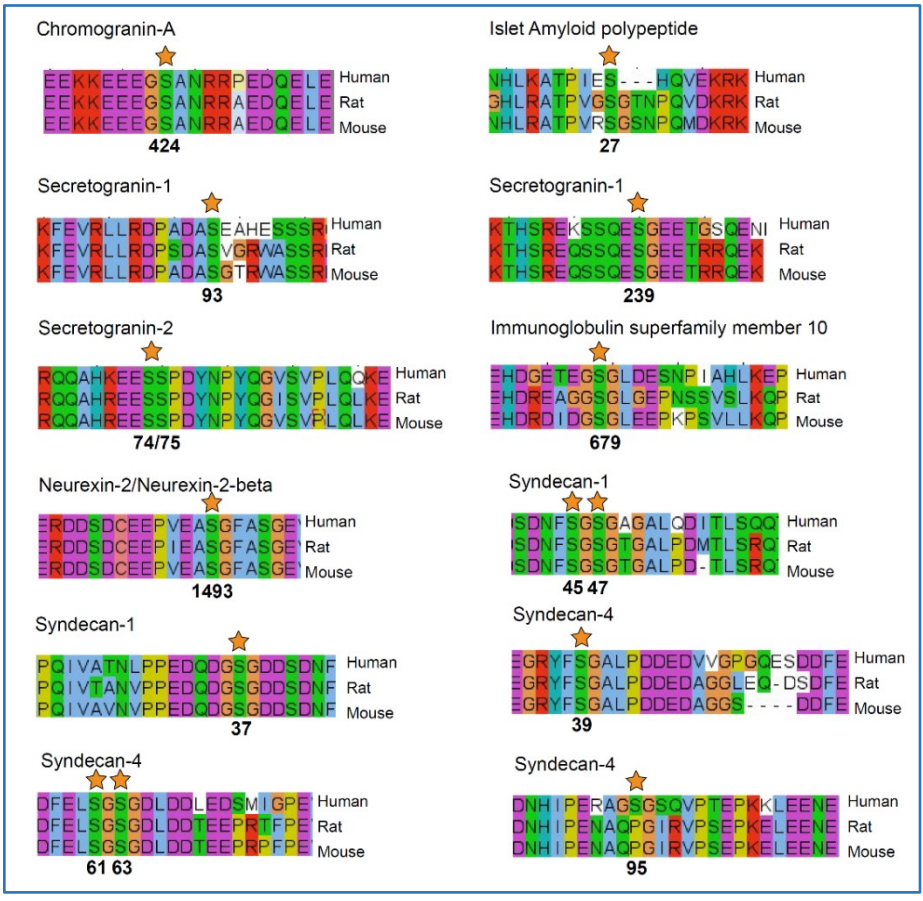

Figure 12. Alignment of GAG glycosites from human, rat and mouse insulin producing cell PGs identified in this thesis. The exact site of Xyl attachment is shown by the star symbol and the corresponding serine residues are numbered.

"Science may set limits to knowledge, but should not set limits to imagination."

Bertrand Russell

5 DISCUSSION

PGs are composed of a core protein to which one or more GAG chain(s) are covalently attached. The emergence of mass spectrometric glycomic techniques has enabled studies of the basic units, the disaccharides, of PGs and in parallel cell and molecular biology techniques have opened up great possibilities to study also the functional aspects of GAGs. However, due to the structural complexity of these molecules, there is still a need to develop novel tools to identify PGs and study their complete structures in more detail, as has been done for only very few of them (Ly et al. 2011; Yu et al. 2017).

With this aim, we have developed a glycoproteomics approach, which is based on the enrichment of GAG-substituted glycopeptides, followed by enzymatic depolymerization of GAG chain and nLC-MS/MS analysis in positive mode MS/MS. In this method, the precursor ions were fragmented by applying a range of different HCD energy levels, which provided the complementary sequential glycosidic and peptide backbone fragmentations. This sequential fragmentation provided information on both GAG linkage region composition and modifications and identification of the core protein as well as of the serine glycosite to which the GAG chain is attached (Paper I). The modifications of the GAG linkage region included phosphorylation, sulfation, and sialylation. Sulfate groups are fragile during the fragmentation in positive mode; however; the negative mode ionization and the use of ion-pairing agents are useful to prevent losses of sulfate groups. We applied this later approach in the glycomics method, where we analyzed samples in the negative mode combined with DBA ion-pairing agent, which gave us additional information on the internal oligosaccharides and the non-reducing ends of the PGs (Paper III).

Sulfation of linkage regions, and of the extended GAG chains in general, is the most typical modification of GAGs. The different sulfation patterns of GAGs, generate a high variety of GAG chains and domains with a wide range of biological actions (Schjoldager et al. 2020). Considering the limited arsenal of methods for structural characterization of biologically active PGs, their GAG domains and their sulfated motifs, the development of the glycoproteomics (Paper I) and glycomics (Paper III) methods described in this thesis and which enabled us to characterize such structures, will probably be valuable.

In this thesis work, we applied the glycoproteomics approach to analyze the CSPGs/DSPGs and HSPGs from human pancreatic islets and rat and mouse insulin-secreting cells. The rat INS-1 832/13 and two mouse cell lines, MIN6 and NIT-1, have been widely used as models to explore various aspects of beta cells pathology in diabetes research. A total of six of core proteins carrying CS/DS GAG chains were identified in this study, including two novel CS/DS-modified core proteins (IAPP and IGSF10). Besides, we identified two known (SDC1, and SDC4) and three novel HSPGs (CgA, SCG1, and Nrxn2). Identification of some PGs, including SDC4, SDC1 and Nrxn2, was possible upon using the combination of trypsin and chymotrypsin enzymes. This indicates the importance of the protease selection based on the origin of the sample. All CSPGs and HSPGs identified in this study showed a cell-type-specific expression pattern, except for CgA that had a broad expression pattern and was identified in human islets and in media from rat and mouse insulinoma cells. Identification of CgA as a PG in the three species investigated, as well as in human CSF and urine (Gomez Toledo et al. 2015), may indicate some importance of the GAG glycosylation for CgA function.

Among all PGs identified from the three different species, only CgA contained either CS/DS or HS chains, thereby CgA may be called a hybrid PG. This hybrid GAG character at the same glycosite is conserved between species, and was identified in rat INS-1 832/13 cells cell lines but not in human beta cells. This could be explained by the fact that the HS structures were less abundant than the CS structures, as in the insulinoma cells, and thus fell below the level of detection in the human islets. Interestingly, HSPGs and CSPGs often function differently in e.g. neural extensions (Shen 2014); thus, one could speculate that this hybrid site of CgA may be of importance for the function and processing of CgA in these cells.

Several of the PGs identified in this work are known as prohormones that are typically stored and processed in SGs. Prohormones have not been reported to be decorated with HS GAGs before; thus the identification of two prohormones carrying HS GAG was surprising. CS/DS and HS GAG-modifications of several prohormones led us to hypothesize that the highly acidic GAG chains may be involved in the storage and/or processing of granule proteins within the SGs. Future research is however, needed to unravel the biological importance of prohormone GAG-modifications.

Prohormones are known to undergo several proteolytic cleavages, which lead to the generation of smaller active peptide fragments from larger inactive prohormones (Laslop, Doblinger, and Weiss 2000). Several studies have shown that glycosylation plays a major role in protein processing, thereby contributing to several pathophysiological processes (Nakagawa et al. 2017; Schjoldager et al. 2011; Schjoldager and Clausen 2012; Topaz et al. 2004). It has been suggested that *O*-glycosylation in or immediately close to the cleavage site of the prohormone convertase (PC) endopeptidases assist these enzymes in identifying and cleaving their target proteins (Schjoldager et al. 2011). To our knowledge, there is no study on the effects of GAG glycosylation on the core protein processing and future investigations are necessary to explore this issue. In this thesis work, the GAG site identified in CgA is in close proximity to the serpinin cleavage site (figure 7). Whether occupying this site with a CS/DS or an HS GAG, may affects the release of serpinin from CgA is of importance, as serpinin has been shown to be the main regulator of DCG biogenesis in neuroendocrine AtT-20 cells (Koshimizu et al. 2011).

In humans, less than one hundred PGs, including CS/DSPG, HSPG and KSPGs, have been characterized (Toledo et al. 2020). The number of known HSPGs is limited, and until recently, only 18 HSPGs have been identified (Xu and Esko 2014; Zhang, Lu, et al. 2018). Thus, by identifying novel HSPGs in this thesis work, we have expanded the number of HSPGs, up to 21. Identifying novel PGs, especially three novel HSPGs, indicates that this glycoproteomic method may be used routinely to identify HSPGs in biologically relevant samples. Besides, the identification of novel PGs, many of which seem to be part-time PGs (Bekku et al. 2003; Yamada et al. 1994; Mann et al. 1990), indicates that PGs may have a larger structural and functional complexity than earlier conceived.

Applying our glycoproteomics method enabled us to determine the exact GAG attachment sites and the surrounding amino acids within the core proteins (Figure 12). This allowed us to investigate whether the selected GAG attachment sites are consensus CS/DS and/or HS sites. Although there is no definite consensus motif to predict the potential GAG sites, certain features have been reported to favor the attachment of CS or HS GAG chains into the selected core protein. For instance, repeated serine-glycine (Ser-Gly)-residues in a short sequence and in the proximity of acidic amino acid clusters and of tryptophan residues seems to favor the synthesis of the HS GAG chain, whereas a single Ser-Gly motif in close proximity of acidic amino acid clusters mostly promotes the CS/DS GAG chain synthesis (Winzen, Cole, and Halfter 2003; Yu and Linhardt 2018; Zhang, David, and Esko 1995). Most of the GAG sites identified in this work are consistent with the above criteria, e.g. the HS GAG sites in SDC1, and SDC4 which contain a repetitive Ser-Gly motif adjacent to acidic clusters (Figure 12). However, not all GAG attachment sites conform to this consensus motif as some display a Ser-Ala sequence (Toledo et al. 2020; Noborn et al. 2015) as illustrated with the GAG attachment site of CgA.

Not all the potential GAG sites are occupied in the identified core proteins; for instance, in human CgA, just one CS/HS site (Ser-424) was identified out of three potential Ser-Gly and three Ser-Ala GAG sites (Paper II). This could be explained by the fact that the degree of Ser xylosylation is different in different cells and has been considered to be an incomplete procedure in some PGs (Esko, Kimata, and Lindahl 2009). The initial attachment of Xyl to the core protein is regulated via the expression of distinct XylT isoenzymes (XylT-1 and XylT-2) that are possibly expressed differently in different cells; thus, the cell of origin may or may not produce particular PGs depending on its needs (Schjoldager et al. 2020). Even for one specific GAG glycosite, the degree of xylosylation, i.e. the occupancy, may differ. This phenomenon, early coined as part-time PGs (Fransson 1987), indicates the dynamic nature of glycosylation in living organisms (Ohtsubo and Marth 2006). Our data showed that level of occupancy at Ser-424 of CgA in human beta islets was around 33% (Paper II).

Our data showed that all but one identified GAG attachment site, Ser-95 of SDC4, were conserved within the three species of this study. The conservation of GAG sites among SGs prohormones, CgA, SCG1, and SCG2, may indicate their biological importance in SGs. However, this speculation needs to be

evaluated further on. Interestingly, the HS GAG site in at the Ser-95 residue, 93-AGSGSQVPTEPK-104, was reported to be decorated with CS GAG chain in human BeWo choriocarcinoma cells (Toledo et al. 2020). This finding may indicate that different cells may decorate the same GAG site with either HS or CS chains, possibly depending on their needs. This observation further indicates that it is important to be careful when choosing the experimental cell or animal model, to be able to translate the obtained information from such models to humans.

Detailed analysis of the glycan structures of the CS and HS linkage regions of identified glycopeptides in this study showed a variety of linkage regions with different patterns of Xyl phosphorylation and/or sulfation of the GalNAc/Gal residues. Considering the fact that the GAGs have a very heterogeneous nature in terms of size, disaccharide composition, and sulfation pattern, and many of biological functions arise from the heterogeneity of structures, we aimed to further unravel the GAGs structures from rat INS-1 832/13 cells (Paper III). We expanded and optimized the previously published method to decipher the GAGs disaccharide composition of linkage region, internal and NREs of CS/DS PG-derived GAGs (Persson et al. 2020). The semi-quantification of different GAG structures, was possible to perform using this approach. Quantifying the GAG glycosylation heterogeneity may help us to follow alterations induced during various physiological and pathological conditions (Hackett and Zaia 2020). A recent study has shown that the strict size and sulfation pattern of HS GAGs are required for the *in vitro* Tau protein cellular uptake (Stopschinski et al. 2018). This finding indicates the importance of developing glycomics methods as the one described in this thesis work, to correlate the structure of GAGs to their biological functions. This analytical method can assist us in expanding our knowledge of the glycome and to decipher the complexity and diversity of GAG structures.

In continuation, we established stable B4galt7 knock-down INS-1 832/13 cells (Paper IV). This particular gene codes for the β1,4-galactosyltransferase 7, responsible for addition of the inner Gal in the linkage region common to all CS/DS and HS GAGs. By knocking out this gene, which does not have any known homologues, we expected to block the normal extension of all CSPGs and HSPGs in INS-1 832/13 cells, including those on CgA. The major reduction of CgA GAG glycosylation, observed with western blot and glycoproteomics analysis in KO cells, was, somewhat unexpectedly,

accompanied with a reduction of CgA at the mRNA and protein levels. This may indicate that there is a relation between CgA glycosylation and its expression. The altered distribution of SGs and processing of CgA prohormone in KO compared to WT cells, may suggest that the GAG-modification influences the processing of prohormones and in particular CgA, which may affect the DCG formation. Further studies of the effect of GAG biosynthesis downregulation in these cells is ongoing.

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

Marie Curie

6 CONCLUSIONS

This thesis has made a number of significant contributions to the field of glycobiology, by the:

- development of two LC-MS/MS based glycoproteomics and glycomics approaches, that are possible to use for the global structural analysis of PGs and GAGs
- identification of several known and novel CS/DSPGs and HSPGs from human, rat and mouse insulin-producing cells
- establishment of CRISPR/Cas9 B4galt7 KO clones of the insulin-secreting rat cell line INS-1 832/13 to better understand the importance of GAGs in biological processes e.g. in DCG biogenesis

"Important thing in science is not so much to obtain new facts as to discover new ways of thinking about them."

Sir William Bragg

7 FUTURE PERSPECTIVES

Understanding the biology of islet cells, especially beta cells, is essential for encoding the molecular events that lead to diabetes and to its long-term complications. The goal of diabetes research is to prevent the development of this disease and to provide novel therapeutic options for the diabetic patients. The basic research in the field is very dependent on the availability to appropriate models of human pancreatic islets. As we have discussed in this thesis, rodent and in less extent human beta cell lines have been widely used in diabetes research. In future studies, it is important to connect the findings from cell models to human cells, tissues, organoids and individuals. One way is using the models that better mimic the *in vivo* environment of pancreatic cells, e.g. pancreatic organoids. Working on a model which can simulate the *in vivo* biology, one could also try to further optimize the methods to be able to detect possibly more PGs and study the connection between the structural composition of PGs and their biological functions. Applying the CRISPR/Cas9 method enables further studies of the function of PGs by generating KO models of specific core proteins, sequences or serine residues of PGs. Additionally, it will be very valuable to generate CRISPR/Cas9 libraries with different KOs of specific glycosyltransferases and of sulfotransferases, epimerases and transporters in order to better understand the GAGs biosynthesis and to understand how the environmental cues could effect this process. Deciphering the structural variabilities of PGs and understanding the effect of biological and pathological conditions on GAG biosynthesis may further open up the use of GAG glycopeptides/proteins as biomarkers of human diseases.

The emergence of very high resolution MS methodologies, has enabled us to study some GAG-related diseases, i.e. the linkeropathies. Linkeropathy refers to a group of severe congenital diseases with enzymatic defects in the biosynthesis of the common GAG tetrasaccharide linkage regions attached to the core proteins of PGs (Mizumoto, Yamada, and Sugahara 2015). Applying our glycoproteomics method we have been able to detect variants of the GAG linkage regions, similarly to the method that led to the identification of a non-canonical CS linkage region in human urine (Persson et al. 2019). According to other studies of human linkeropathies, the clinical symptoms are not always in agreement with glycosyltransferase enzymatic defects of GAG linkage region biosynthesis (Job et al. 2016; Jones et al. 2015; Mizumoto, Yamada, and Sugahara 2015). This may be due to the activation of partially compensational mechanisms by homologous enzymes, which makes it difficult to diagnose the disease just based on the clinical symptoms and enzymatic activities of the crippled enzymes. Whole exome sequencing is now used as the method of choice for correct diagnose for these patients. However, exploring the heterogeneity of GAG chains in different linkeropathies using the methods described in this thesis, may open up an opportunity to use these as diagnostic tools for human linkeropathies and also as possible analysis for follow-up on future therapies.

ACKNOWLEDGEMENT

This work was performed at the Department of Laboratory Medicine, University of Gothenburg, Sweden. The work was supported by grants from the Swedish Research Council and Swedish state, the ALF agreement. I would like to take the opportunity and to thank all the people who have been with me and helped me to accomplish my PhD project and especially to:

My main supervisor, Professor **Göran Larson**, Thank you for giving me an opportunity to join your group and introducing me to the fascinating world of glycobiology. I would like to express my gratitude for your excellent academic support, patience, encouragement, optimistic look, and care throughout my research period. It has been a great pleasure for me to be a part of your group. Thank you again for being such a wonderful supervisor!

Thanks to my previous co-supervisor, Professor **Susanna Cardell**; it was my pleasure to get to know you and discuss my project with you.

The **Glycobiology** group, which without their help, support, and knowledge, I would not have been able to complete this research:

Andrea Persson, for your practical advice and always being supportive, thank you for all your feedback on my work and your understanding. Your help has been invaluable!

Jonas Nilsson, thank you for being the first guide into the mass spectrometry field and always patient and kind to me.

Fredrik Noborn, thanks for your contribution to my work and your intellectual discussions.

I would like to thank the former members of the Glycobiology group: **Inga Rimkute** and **Alejandro Gomez**, special thanks to you for all the nice time that we shared, all the lunch breaks, and for all your support and care. Thanks to **Lisa Gilborne**, for the practical help in the lab and lunch breaks.

I express my great sense of gratitude to all my past and present colleagues in the clinical chemistry department; **Cecillia**, **Maria**, **Tina**, **Camilla**, **Petra**, **Lotta**, **Alexandra**, **Hamdy**, **Josefine**, **Marion**, **Daniel**, **Mukesh**, **Pradeep**. Thank you very much for being always cooperative in the lab.

Thanks to all Cell-3 members that I shared a lot of time with. It was a pleasure to work with all of you: **Fredrik**, **Berta**, **Kristina**, **Sanhita**, **Anna-Karin** and, **Debbi**.

I would like to say a special thanks to all the people that were contributed to my research project:

Katrine Ter-Borch Gram Schjoldager for hosting me at Copenhagen Center for Glycomics (CCG) and many thanks to **Thomas Daugbjreg Madsen** for practical help in the CRISPR project and for being always positive and helpful. It was a great experience to work with you!

Egor Vorontsov and **Carina Sihlbom** for running all the mass spectrometry analyses in this thesis work. This work was impossible to be done without your expertise and knowledge!

Massimo Micaroni and **Julia Fernandez**, thank you for the great help in electron microscopy. It was really fascinating to see the cells that I was working with for four years so closely!

Many thanks to all my lovely friends, whom my PhD journey would not have been possible without their help, care, and love:

Thanks to all the lovely persons that I share my office with **Aditi**, **Sandra**, **Mary**, **Katti,** for being always nice, understanding and supportive! Especial thanks to **Aditi** and **Sandra** for the help and support during my writing time. You are wonderful office mates!

Aida, Thank you very much for all your support, kind words, and caring about me. I am very thankful for our friendship!

Tugce, I really don't know how to say thank you for being always with me, in good and bad moments and always being supportive and inspiring.

Narmin, my stay in Gothenburg would be totally different without knowing you! Thank you so much for your big heart and a big smile.

Jasmin and **Steffie**, thanks for being such wonderful and supporting friends. I will always keep the memory of my first camping in Sweden with you.

Laia, I was very lucky to get you to know, you are an amazing person!

Guisy and **Paula**, there is no way to thank you for everything you have done for me. Simply, thanks for being always supportive, inspiring and caring. I love both of you, and I hope that to see you soon.

Fathi and **Najme**, one of the best things that happened to me in Gothenburg was knowing you! Good luck **Fathi,** with your dissertation, I would love to listen to you, but since we have our dissertations on the same day and even at the same time, it is obviously impossible.

Thanks to all my Iranian friends at Sahlgrenska hospital: **Parisa**, **Mona**, **Sara(s)**, **Maryam**, **Azadeh**, **Pegah**, **Laleh**, **Gitti**, **Maddi**, **Shohreh**, **Farnaz**. Thanks for all your kindness, support, and help.

Thanks to lovely after work girls: **Asal**, **Emel**, and **Roya**

Thank you, **Björn,** for the technical help you offered.

I also thank my friends in Switzerland, for providing support and friendship that I needed, especially to **Ornella**, **Sara, Rocco**, and a lovely couple **Mohsen** and **Sara,** for their always great hospitality.

Finally, thanks to all my family members for believing my decisions. Without your support, all these accomplishments would have been impossible. I love you all ♥ ♥ ♥

REFERENCES

Abrink, M., M. Grujic, and G. Pejler. 2004. 'Serglycin is essential for maturation of mast cell secretory granule', *J Biol Chem*, 279: 40897-905.

Aebersold, R., and D. R. Goodlett. 2001. 'Mass spectrometry in proteomics', *Chem Rev*, 101: 269-95.

Aebersold, R., and M. Mann. 2003. 'Mass spectrometry-based proteomics', *Nature*, 422: 198-207.

Altelaar, A. F., J. Munoz, and A. J. Heck. 2013. 'Next-generation proteomics: towards an integrative view of proteome dynamics', *Nat Rev Genet*, 14: 35-48.

American Diabetes, Association. 2009. 'Diagnosis and classification of diabetes mellitus', *Diabetes Care*, 32 Suppl 1: S62-7.

Asfari, M., D. Janjic, P. Meda, G. Li, P. A. Halban, and C. B. Wollheim. 1992. 'Establishment of 2-mercaptoethanol-dependent differentiated insulin-secreting cell lines', *Endocrinology*, 130: 167-78.

Bai, X., D. Zhou, J. R. Brown, B. E. Crawford, T. Hennet, and J. D. Esko. 2001. 'Biosynthesis of the linkage region of glycosaminoglycans: cloning and activity of galactosyltransferase II, the sixth member of the beta 1,3-galactosyltransferase family (beta 3GalT6)', *J Biol Chem*, 276: 48189-95.

Banerjee, S., and S. Mazumdar. 2012. 'Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte', *International Journal of Analytical Chemistry*.

Bantscheff, M., M. Schirle, G. Sweetman, J. Rick, and B. Kuster. 2007. 'Quantitative mass spectrometry in proteomics: a critical review', *Analytical and Bioanalytical Chemistry*, 389: 1017-31.

Barrangou, R., and J. A. Doudna. 2016. 'Applications of CRISPR technologies in research and beyond', *Nat Biotechnol*, 34: 933-41.

Bartolomucci, A., R. Possenti, S. K. Mahata, R. Fischer-Colbrie, Y. P. Loh, and S. R. Salton. 2011. 'The extended granin family: structure, function, and biomedical implications', *Endocr Rev*, 32: 755-97.

Bekku, Y., W. D. Su, S. Hirakawa, R. Fassler, A. Ohtsuka, J. S. Kang, J. Sanders, T. Murakami, Y. Ninomiya, and T. Oohashi. 2003. 'Molecular cloning of Bral2, a novel brain-specific link protein, and immunohistochemical colocalization with brevican in perineuronal nets', *Molecular and Cellular Neuroscience*, 24: 148-59.

Bertozzi, C. R., and R. Sasisekharan. 2009. 'Glycomics.' in nd, A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler (eds.), *Essentials of Glycobiology* (Cold Spring Harbor (NY)).

Bjornsson, S. 1998. 'Quantitation of proteoglycans as glycosaminoglycans in biological fluids using an alcian blue dot blot analysis', *Anal Biochem*, 256: 229-37.

Borgonovo, B., J. Ouwendijk, and M. Solimena. 2006. 'Biogenesis of secretory granules', *Curr Opin Cell Biol*, 18: 365-70.

Briggs, D. C., and E. Hohenester. 2018. 'Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1', *Structure*, 26: 801-09 e3.

Brissova, M., M. J. Fowler, W. E. Nicholson, A. Chu, B. Hirshberg, D. M. Harlan, and A. C. Powers. 2005. 'Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy', *J Histochem Cytochem*, 53: 1087-97.

Brunner, Y., Y. Coute, M. Iezzi, M. Foti, M. Fukuda, D. F. Hochstrasser, C. B. Wollheim, and J. C. Sanchez. 2007. 'Proteomics analysis of insulin secretory granules', *Mol Cell Proteomics*, 6: 1007-17.

Cao, L., N. Tolic, Y. Qu, D. Meng, R. Zhao, Q. B. Zhang, R. J. Moore, E. M. Zink, M. S. Lipton, L. Paga-Tolic, and S. Wu. 2014. 'Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation', *Analytical Biochemistry*, 452: 96-102.

Capila, I., and R. J. Linhardt. 2002. 'Heparin - Protein interactions', *Angewandte Chemie-International Edition*, 41: 391-412.

Carlsson, P., J. Presto, D. Spillmann, U. Lindahl, and L. Kjellen. 2008. 'Heparin/heparan sulfate biosynthesis: processive formation of N-sulfated domains', *J Biol Chem*, 283: 20008-14.

Caterson, B., and J. Melrose. 2018. 'Keratan sulfate, a complex glycosaminoglycan with unique functional capability', *Glycobiology*, 28: 182-206.

Chanat, E., and W. B. Huttner. 1991. 'Milieu-induced, selective aggregation of regulated secretory proteins in the trans-Golgi network', *J Cell Biol*, 115: 1505-19.

Chanat, E., U. Weiss, W. B. Huttner, and S. A. Tooze. 1993. 'Reduction of the disulfide bond of chromogranin B (secretogranin I) in the trans-Golgi network causes its missorting to the constitutive secretory pathways', *EMBO J*, 12: 2159-68.

Chandramouli, K., and P. Y. Qian. 2009. 'Proteomics: challenges, techniques and possibilities to overcome biological sample complexity', *Hum Genomics Proteomics*, 2009.

Chang, S. G., K. D. Choi, S. H. Jang, and H. C. Shin. 2003. 'Role of disulfide bonds in the structure and activity of human insulin', *Mol Cells*, 16: 323-30.

Cheng, J. Y. C., J. Whitelock, and L. Poole-Warren. 2012. 'Syndecan-4 is associated with beta-cells in the pancreas and the MIN6 beta-cell line', *Histochemistry and Cell Biology*, 138: 933-44.

Chi, L., J. J. Wolff, T. N. Laremore, O. F. Restaino, J. Xie, C. Schiraldi, T. Toida, I. J. Amster, and R. J. Linhardt. 2008. 'Structural analysis of bikunin glycosaminoglycan', *J Am Chem Soc*, 130: 2617-25.

Choudhary, G., S. L. Wu, P. Shieh, and W. S. Hancock. 2003. 'Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS', *Journal of Proteome Research*, 2: 59-67.

Colomer, V., G. A. Kicska, and M. J. Rindler. 1996. 'Secretory granule content proteins and the luminal domains of granule membrane proteins aggregate in vitro at mildly acidic pH', *Journal of Biological Chemistry*, 271: 48-55.

Cool, D. R., E. Normant, F. S. Shen, H. C. Chen, L. Pannell, Y. Zhang, and Y. P. Loh. 1997. 'Carboxypeptidase E is a regulated secretory pathway sorting receptor: Genetic obliteration leads to endocrine disorders in Cpe(fat) mice', *Cell*, 88: 73-83.

Cowley, D. J., Y. R. Moore, D. S. Darling, P. B. Joyce, and S. U. Gorr. 2000. 'N- and C-terminal domains direct cell type-specific sorting of chromogranin A to secretory granules', *J Biol Chem*, 275: 7743-8.

Cristobal, A., F. Marino, H. Post, H. W. van den Toorn, S. Mohammed, and A. J. Heck. 2017. 'Toward an Optimized Workflow for Middle-Down Proteomics', *Anal Chem*, 89: 3318-25.

Cummings, R. D., and J. M. Pierce. 2014. 'The challenge and promise of glycomics', *Chem Biol*, 21: 1-15.

Da Silva Xavier, G. 2018. 'The Cells of the Islets of Langerhans', *J Clin Med*, 7.

Darula, Z., and K. F. Medzihradszky. 2018. 'Analysis of Mammalian O-Glycopeptides-We Have Made a Good Start, but There is a Long Way to Go', *Mol Cell Proteomics*, 17: 2-17.

Dau, T., G. Bartolomucci, and J. Rappsilber. 2020. 'Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin', *Analytical Chemistry*, 92: 9523-27.

Day, R., and S. U. Gorr. 2003. 'Secretory granule biogenesis and chromogranin A: master gene, on/off switch or assembly factor?', *Trends Endocrinol Metab*, 14: 10-3.

de la Tour, D., T. Halvorsen, C. Demeterco, B. Tyrberg, P. Itkin-Ansari, M. Loy, S. J. Yoo, E. Hao, S. Bossie, and F. Levine. 2001. 'Beta-cell differentiation from a human pancreatic cell line in vitro and in vivo', *Mol Endocrinol*, 15: 476-83.

Desai, U. R., H. M. Wang, and R. J. Linhardt. 1993. 'Specificity studies on the heparin lyases from Flavobacterium heparinum', *Biochemistry*, 32: 8140-5.

Doerr, A. 2015. 'DIA mass spectrometry', *Nature Methods*, 12: 35-35.

Dong, W., W. Lu, W. L. McKeehan, Y. Luo, and S. Ye. 2012. 'Structural basis of heparan sulfate-specific degradation by heparinase III', *Protein Cell*, 3: 950-61.

Doudna, J. A., and E. Charpentier. 2014. 'Genome editing. The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346: 1258096.

Ernst, S., R. Langer, C. L. Cooney, and R. Sasisekharan. 1995. 'Enzymatic Degradation of Glycosaminoglycans', *Critical Reviews in Biochemistry and Molecular Biology*, 30: 387-444.

Eskeland, N. L., A. Zhou, T. Q. Dinh, H. Wu, R. J. Parmer, R. E. Mains, and D. T. O'Connor. 1996. 'Chromogranin A processing and secretion: specific role of endogenous and exogenous prohormone convertases in the regulated secretory pathway', *J Clin Invest*, 98: 148-56.

Esko, J. D., K. Kimata, and U. Lindahl. 2009. 'Proteoglycans and Sulfated Glycosaminoglycans.' in nd, A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler (eds.), *Essentials of Glycobiology* (Cold Spring Harbor (NY)).

Esko, J. D., and L. Zhang. 1996. 'Influence of core protein sequence on glycosaminoglycan assembly', *Curr Opin Struct Biol*, 6: 663-70.

Evers, M. R., G. Xia, H. G. Kang, M. Schachner, and J. U. Baenziger. 2001. 'Molecular cloning and characterization of a dermatan-specific N-acetylgalactosamine 4-O-sulfotransferase', *J Biol Chem*, 276: 36344-53.

Farndale, R. W., C. A. Sayers, and A. J. Barrett. 1982. 'A direct spectrophotometric microassay for sulfated glycosaminoglycans in cartilage cultures', *Connect Tissue Res*, 9: 247-8.

Ferro, D. R., A. Provasoli, M. Ragazzi, B. Casu, G. Torri, V. Bossennec, B. Perly, P. Sinay, M. Petitou, and J. Choay. 1990. 'Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences', *Carbohydr Res*, 195: 157-67.

Flannick, J., S. Johansson, and P. R. Njolstad. 2016. 'Common and rare forms of diabetes mellitus: towards a continuum of diabetes subtypes', *Nat Rev Endocrinol*, 12: 394-406.

Fransson, L. A. 1987. 'Structure and Function of Cell-Associated Proteoglycans', *Trends in Biochemical Sciences*, 12: 406-11.

Fu, Z., E. R. Gilbert, and D. Liu. 2013. 'Regulation of insulin synthesis and secretion and pancreatic Beta-cell dysfunction in diabetes', *Curr Diabetes Rev*, 9: 25-53.

Gaj, T., C. A. Gersbach, and C. F. Barbas, 3rd. 2013. 'ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering', *Trends Biotechnol*, 31: 397-405.

Gaj, T., S. J. Sirk, S. L. Shui, and J. Liu. 2016. 'Genome-Editing Technologies: Principles and Applications', *Cold Spring Harb Perspect Biol*, 8.

Georgakopoulos, N., N. Prior, B. Angres, G. Mastrogiovanni, A. Cagan, D. Harrison, C. J. Hindley, R. Arnes-Benito, S. S. Liau, A. Curd, N. Ivory, B. D. Simons, I. Martincorena, H. Wurst, K. Saeb-Parsy, and M. Huch. 2020. 'Long-term expansion, genomic stability and in vivo safety of adult human pancreas organoids', *BMC Dev Biol*, 20: 4.

Gerdes, H. H., P. Rosa, E. Phillips, P. A. Baeuerle, R. Frank, P. Argos, and W. B. Huttner. 1989. 'The Primary Structure of Human Secretogranin-Ii, a Widespread Tyrosine-Sulfated Secretory Granule Protein That Exhibits Low Ph-Induced and Calcium-Induced Aggregation', *Journal of Biological Chemistry*, 264: 12009-15.

Giansanti, P., L. Tsiatsiani, T. Y. Low, and A. J. Heck. 2016. 'Six alternative proteases for mass spectrometry-based proteomics beyond trypsin', *Nat Protoc*, 11: 993-1006.

Glish, G. L., and D. J. Burinsky. 2008. 'Hybrid mass spectrometers for tandem mass Spectrometry', *Journal of the American Society for Mass Spectrometry*, 19: 161-72.

Gomez Toledo, A., J. Nilsson, F. Noborn, C. Sihlbom, and G. Larson. 2015. 'Positive Mode LC-MS/MS Analysis of Chondroitin Sulfate Modified Glycopeptides Derived from Light and Heavy Chains of The Human Inter-alpha-Trypsin Inhibitor Complex', *Mol Cell Proteomics*, 14: 3118-31.

Gorr, S. U., W. L. Dean, T. L. Radley, and D. V. Cohn. 1988. 'Calcium-binding and aggregation properties of parathyroid secretory protein-I (chromogranin A)', *Bone Miner*, 4: 17-25.

Gottlieb, P. A., T. Delong, R. L. Baker, L. Fitzgerald-Miller, R. Wagner, G. Cook, M. R. Rewers, A. Michels, and K. Haskins. 2014. 'Chromogranin A is a T cell antigen in human type 1 diabetes', *Journal of Autoimmunity*, 50: 38-41.

Grapin-Botton, A. 2016. 'Three-dimensional pancreas organogenesis models', *Diabetes Obes Metab*, 18 Suppl 1: 33-40.

Gray, C. J., B. Schindler, L. G. Migas, M. Picmanova, A. R. Allouche, A. P. Green, S. Mandal, M. S. Motawia, R. Sanchez-Perez, N. Bjarnholt, B. L. Moller, A. M. Rijs, P. E. Barran, I. Compagnon, C. E. Eyers, and S. L. Flitsch. 2017. 'Bottom-Up Elucidation of Glycosidic Bond Stereochemistry', *Analytical Chemistry*, 89: 4540-49.

Gray, C. J., B. Thomas, R. Upton, L. G. Migas, C. E. Eyers, P. E. Barran, and S. L. Flitsch. 2016. 'Applications of ion mobility mass spectrometry for high throughput, high resolution glycan analysis', *Biochimica Et Biophysica Acta-General Subjects*, 1860: 1688-709.

Grozinsky-Glasberg, S., I. Shimon, and H. Rubinfeld. 2012. 'The role of cell lines in the study of neuroendocrine tumors', *Neuroendocrinology*, 96: 173-87.

Guan, S., P. P. Taylor, Z. Han, M. F. Moran, and B. Ma. 2020. 'Data Dependent-Independent Acquisition (DDIA) Proteomics', *J Proteome Res*, 19: 3230-37.

Gulberti, S., J. C. Jacquinet, M. Chabel, N. Ramalanjaona, J. Magdalou, P. Netter, M. W. Coughtrie, M. Ouzzine, and S. Fournel-Gigleux. 2012. 'Chondroitin sulfate N-acetylgalactosaminyltransferase-1 (CSGalNAcT-1) involved in chondroitin sulfate initiation: Impact of sulfation on activity and specificity', *Glycobiology*, 22: 561-71.

Gulcicek, E. E., C. M. Colangelo, W. McMurray, K. Stone, K. Williams, T. Wu, H. Zhao, H. Spratt, A. Kurosky, and B. Wu. 2005. 'Proteomics and the analysis of proteomic data: an overview of current protein-profiling technologies', *Curr Protoc Bioinformatics*, Chapter 13: Unit 13 1.

Gundry, R. L., M. Y. White, C. I. Murray, L. A. Kane, Q. Fu, B. A. Stanley, and J. E. Van Eyk. 2009. 'Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow', *Curr Protoc Mol Biol*, Chapter 10: Unit10 25.

Gurgul-Convey, E., M. T. Kaminski, and S. Lenzen. 2015. 'Physiological characterization of the human EndoC-betaH1 beta-cell line', *Biochem Biophys Res Commun*, 464: 13-9.

Hackett, W. E., and J. Zaia. 2020. 'Calculating glycoprotein similarities from mass spectrometric data', *Mol Cell Proteomics*.

Halim, A., U. Ruetschi, G. Larson, and J. Nilsson. 2013. 'LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins', *J Proteome Res*, 12: 573-84.

Hamaguchi, K., H. R. Gaskins, and E. H. Leiter. 1991. 'NIT-1, a pancreatic beta-cell line established from a transgenic NOD/Lt mouse', *Diabetes*, 40: 842-9.

Han, L., and C. E. Costello. 2013. 'Mass spectrometry of glycans', *Biochemistry-Moscow*, 78: 710-20.

Han, L., M. Suda, K. Tsuzuki, R. Wang, Y. Ohe, H. Hirai, T. Watanabe, T. Takeuchi, and M. Hosaka. 2008. 'A large form of secretogranin III functions as a sorting receptor for chromogranin a aggregates in PC12 cells', *Molecular Endocrinology*, 22: 1935-49.

Han, X. M., A. Aslanian, and J. R. Yates. 2008. 'Mass spectrometry for proteomics', *Current Opinion in Chemical Biology*, 12: 483-90.

Hart, A. W., N. Baeza, A. Apelqvist, and H. Edlund. 2000. 'Attenuation of FGF signalling in mouse beta-cells leads to diabetes', *Nature*, 408: 864-68.

Helle, K. B. 2004. 'The granin family of uniquely acidic proteins of the diffuse neuroendocrine system: comparative and functional aspects', *Biol Rev Camb Philos Soc*, 79: 769-94.

Hiraoka, N., H. Nakagawa, E. Ong, T. O. Akama, M. N. Fukuda, and M. Fukuda. 2000. 'Molecular cloning and expression of two distinct human chondroitin 4-O-sulfotransferases that belong to the HNK-1 sulfotransferase gene family', *J Biol Chem*, 275: 20188-96.

Ho, C. S., C. W. Lam, M. H. Chan, R. C. Cheung, L. K. Law, L. C. Lit, K. F. Ng, M. W. Suen, and H. L. Tai. 2003. 'Electrospray ionisation mass spectrometry: principles and clinical applications', *Clin Biochem Rev*, 24: 3-12.

Hohmeier, H. E., H. Mulder, G. Chen, R. Henkel-Rieger, M. Prentki, and C. B. Newgard. 2000. 'Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion', *Diabetes*, 49: 424-30.

Hosaka, M., T. Watanabe, Y. Sakai, Y. Uchiyama, and T. Takeuchi. 2002. 'Identification of a chromogranin A domain that mediates binding to

secretogranin III and targeting to secretory granules in pituitary cells and pancreatic beta-cells', *Mol Biol Cell*, 13: 3388-99.

Hou, J. C., L. Min, and J. E. Pessin. 2009. 'Insulin granule biogenesis, trafficking and exocytosis', *Vitam Horm*, 80: 473-506.

Hsu, P. D., E. S. Lander, and F. Zhang. 2014. 'Development and applications of CRISPR-Cas9 for genome engineering', *Cell*, 157: 1262-78.

Huang, T., J. Wang, W. Yu, and Z. He. 2012. 'Protein inference: a review', *Brief Bioinform*, 13: 586-614.

Huh, Y. H., S. H. Jeon, and S. H. Yoo. 2003. 'Chromogranin B-induced secretory granule biogenesis: comparison with the similar role of chromogranin A', *J Biol Chem*, 278: 40581-9.

Huttner, W. B., and U. M. Benedum. 1987. 'Chromogranin A and pancreastatin', *Nature*, 325: 305.

Hutton, J. C. 1994. 'Insulin secretory granule biogenesis and the proinsulin-processing endopeptidases', *Diabetologia*, 37 Suppl 2: S48-56.

Hutton, J. C., H. W. Davidson, and M. Peshavaria. 1987. 'The mechanism of chromogranin A processing', *Nature*, 325: 766.

Ilonen, J., J. Lempainen, and R. Veijola. 2019. 'The heterogeneous pathogenesis of type 1 diabetes mellitus', *Nat Rev Endocrinol*, 15: 635-50.

Iozzo, R. V., and L. Schaefer. 2015. 'Proteoglycan form and function: A comprehensive nomenclature of proteoglycans', *Matrix Biol*, 42: 11-55.

Izumikawa, T., T. Koike, S. Shiozawa, K. Sugahara, J. I. Tamura, and H. Kitagawa. 2008a. 'Identification of chondroitin sulfate glucuronyltransferase as chondroitin synthase-3 involved in chondroitin polymerization - Chondroitin polymerization is achieved by multiple enzyme complexes consisting of chondroitin synthase family members', *Journal of Biological Chemistry*, 283: 11396-406.

Izumikawa, T., T. Koike, S. Shiozawa, K. Sugahara, J. Tamura, and H. Kitagawa. 2008b. 'Identification of chondroitin sulfate glucuronyltransferase as chondroitin synthase-3 involved in chondroitin polymerization: chondroitin polymerization is achieved by multiple enzyme complexes consisting of chondroitin synthase family members', *J Biol Chem*, 283: 11396-406.

Izumikawa, T., B. Sato, T. Mikami, J. Tamura, M. Igarashi, and H. Kitagawa. 2015. 'GlcUAbeta1-3Galbeta1-3Galbeta1-4Xyl(2-O-phosphate) is the preferred substrate for chondroitin N-acetylgalactosaminyltransferase-1', *J Biol Chem*, 290: 5438-48.

Izumikawa, T., T. Uyama, Y. Okuura, K. Sugahara, and H. Kitagawa. 2007. 'Involvement of chondroitin sulfate synthase-3 (chondroitin synthase-2) in chondroitin polymerization through its interaction with chondroitin synthase-1 or chondroitin-polymerizing factor', *Biochem J*, 403: 545-52.

Jedrychowski, M. P., E. L. Huttlin, W. Haas, M. E. Sowa, R. Rad, and S. P. Gygi. 2011. 'Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics', *Molecular & Cellular Proteomics*, 10.

Ji, L., H. T. Wu, X. Y. Qin, and R. Lan. 2017. 'Dissecting carboxypeptidase E: properties, functions and pathophysiological roles in disease', *Endocr Connect*, 6: R18-R38.

Job, F., S. Mizumoto, L. Smith, N. Couser, A. Brazil, H. Saal, M. Patterson, M. I. Gibson, S. Soden, N. Miller, I. Thiffault, C. Saunders, S. Yamada, K. Hoffmann, K. Sugahara, and E. Farrow. 2016. 'Functional validation of novel compound heterozygous variants in B3GAT3 resulting in severe osteopenia and fractures: expanding the disease phenotype', *BMC Med Genet*, 17: 86.

Jones, K. L., U. Schwarze, M. P. Adam, P. H. Byers, and H. C. Mefford. 2015. 'A homozygous B3GAT3 mutation causes a severe syndrome with multiple fractures, expanding the phenotype of linkeropathy syndromes', *Am J Med Genet A*, 167A: 2691-6.

Joshi, H. J., L. Hansen, Y. Narimatsu, H. H. Freeze, B. Henrissat, E. Bennett, H. H. Wandall, H. Clausen, and K. T. Schjoldager. 2018. 'Glycosyltransferase genes that cause monogenic congenital disorders of glycosylation are distinct from glycosyltransferase genes associated with complex diseases', *Glycobiology*, 28: 284-94.

Kang, H. G., M. R. Evers, G. Xia, J. U. Baenziger, and M. Schachner. 2002. 'Molecular cloning and characterization of chondroitin-4-O-sulfotransferase-3. A novel member of the HNK-1 family of sulfotransferases', *J Biol Chem*, 277: 34766-72.

Kayton, N. S., G. Poffenberger, J. Henske, C. Dai, C. Thompson, R. Aramandla, A. Shostak, W. Nicholson, M. Brissova, W. S. Bush, and A. C. Powers. 2015. 'Human islet preparations distributed for research exhibit a variety of insulin-secretory profiles', *Am J Physiol Endocrinol Metab*, 308: E592-602.

Khoury, G. A., R. C. Baliban, and C. A. Floudas. 2011. 'Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database', *Sci Rep*, 1.

Kim, T., M. C. Gondre-Lewis, I. Arnaoutova, and Y. P. Loh. 2006. 'Dense-core secretory granule biogenesis', *Physiology (Bethesda)*, 21: 124-33.

Kim, T., and Y. P. Loh. 2006. 'Protease nexin-1 promotes secretory granule biogenesis by preventing granule protein degradation', *Mol Biol Cell*, 17: 789-98.

Kim, T., J. H. Tao-Cheng, L. E. Eiden, and Y. P. Loh. 2001. 'Chromogranin A, an "on/off" switch controlling dense-core secretory granule biogenesis', *Cell*, 106: 499-509.

Kitagawa, H., M. Fujita, N. Ito, and K. Sugahara. 2000. 'Molecular cloning and expression of a novel chondroitin 6-O-sulfotransferase', *J Biol Chem*, 275: 21075-80.

Kitagawa, H., K. Tsutsumi, A. Ikegami-Kuzuhara, S. Nadanaka, F. Goto, T. Ogawa, and K. Sugahara. 2008. 'Sulfation of the galactose residues in the glycosaminoglycan-protein linkage region by recombinant human chondroitin 6-O-sulfotransferase-1', *J Biol Chem*, 283: 27438-43.

Kitagawa, H., T. Uyama, and K. Sugahara. 2001. 'Molecular cloning and expression of a human chondroitin synthase', *J Biol Chem*, 276: 38721-6.

Kjellen, L., and U. Lindahl. 1991. 'Proteoglycans: structures and interactions', *Annu Rev Biochem*, 60: 443-75.

Kobayashi, M., G. Sugumaran, J. A. Liu, N. W. Shworak, J. E. Silbert, and R. D. Rosenberg. 1999. 'Molecular cloning and characterization of a human uronyl 2-sulfotransferase that sulfates iduronyl and glucuronyl residues in dermatan chondroitin sulfate', *Journal of Biological Chemistry*, 274: 10474-80.

Koike, T., T. Izumikawa, B. Sato, and H. Kitagawa. 2014. 'Identification of phosphatase that dephosphorylates xylose in the glycosaminoglycan-protein linkage region of proteoglycans', *J Biol Chem*, 289: 6695-708.

Koike, T., T. Izumikawa, J. I. Tamura, and H. Kitagawa. 2009. 'FAM20B is a kinase that phosphorylates xylose in the glycosaminoglycan-protein linkage region', *Biochemical Journal*, 421: 157-62.

Kolby, L., P. Bernhardt, H. Ahlman, B. Wangberg, V. Johanson, A. Wigander, E. Forssell-Aronsson, S. Karlsson, B. Ahren, G. Stenman, and O. Nilsson. 2001. 'A transplantable human carcinoid as model for somatostatin receptor-mediated and amine transporter-mediated radionuclide uptake', *Am J Pathol*, 158: 745-55.

Konecki, D. S., U. M. Benedum, H. H. Gerdes, and W. B. Huttner. 1987. 'The primary structure of human chromogranin A and pancreastatin', *J Biol Chem*, 262: 17026-30.

Koshimizu, H., N. X. Cawley, T. Kim, A. L. Yergey, and Y. P. Loh. 2011. 'Serpinin: a novel chromogranin A-derived, secreted peptide up-

regulates protease nexin-1 expression and granule biogenesis in endocrine cells', *Mol Endocrinol*, 25: 732-44.

Koshimizu, H., T. Kim, N. X. Cawley, and Y. P. Loh. 2010. 'Chromogranin A: a new proposal for trafficking, processing and induction of granule biogenesis', *Regul Pept*, 160: 153-9.

Kovalevich, J., and D. Langford. 2013. 'Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology', *Methods Mol Biol*, 1078: 9-21.

Kreuger, J., and L. Kjellen. 2012. 'Heparan sulfate biosynthesis: regulation and variability', *J Histochem Cytochem*, 60: 898-907.

Kubaski, F., H. Osago, R. W. Mason, S. Yamaguchi, H. Kobayashi, M. Tsuchiya, T. Orii, and S. Tomatsu. 2017. 'Glycosaminoglycans detection methods: Applications of mass spectrometry', *Mol Genet Metab*, 120: 67-77.

Kuberan, B., M. Lech, L. Zhang, Z. L. Wu, D. L. Beeler, and R. D. Rosenberg. 2002. 'Analysis of heparan sulfate oligosaccharides with ion pair-reverse phase capillary high performance liquid chromatography-microelectrospray ionization time-of-flight mass spectrometry', *J Am Chem Soc*, 124: 8707-18.

Kusche-Gullberg, M., and L. Kjellen. 2003. 'Sulfotransferases in glycosaminoglycan biosynthesis', *Curr Opin Struct Biol*, 13: 605-11.

Laguerre, F., Y. Anouar, and M. Montero-Hadjadje. 2020. 'Chromogranin A in the early steps of the neurosecretory pathway', *IUBMB Life*, 72: 524-32.

Lanucara, F., S. W. Holman, C. J. Gray, and C. E. Eyers. 2014. 'The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics', *Nature Chemistry*, 6: 281-94.

Laslop, A., A. Doblinger, and U. Weiss. 2000. 'Proteolytic processing of chromogranins', *Adv Exp Med Biol*, 482: 155-66.

Lauder, R. M. 2009. 'Chondroitin sulphate: a complex molecule with potential impacts on a wide range of biological systems', *Complement Ther Med*, 17: 56-62.

Lauder, R. M., T. N. Huckerby, and I. A. Nieduszynski. 2000. 'A fingerprinting method for chondroitin/dermatan sulfate and hyaluronan oligosaccharides', *Glycobiology*, 10: 393-401.

Li, G. Y., L. Y. Li, F. Tian, L. X. Zhang, C. H. Xue, and R. J. Linhardt. 2015. 'Glycosaminoglycanomics of Cultured Cells Using a Rapid and Sensitive LC-MS/MS Approach', *Acs Chemical Biology*, 10: 1303-10.

Li, J. P., F. Gong, K. El Darwish, M. Jalkanen, and U. Lindahl. 2001. 'Characterization of the D-glucuronyl C5-epimerase involved in the biosynthesis of heparin and heparan sulfate', *J Biol Chem*, 276: 20069-77.

Li, M., W. Du, M. Zhou, L. Zheng, E. Song, and J. Hou. 2018. 'Proteomic analysis of insulin secretory granules in INS-1 cells by protein correlation profiling', *Biophys Rep*, 4: 329-38.

Li, Q., Y. Xie, M. Wong, and C. B. Lebrilla. 2019. 'Characterization of Cell Glycocalyx with Mass Spectrometry Methods', *Cells*, 8.

Li, X., Z. Xu, X. Hong, Y. Zhang, and X. Zou. 2020. 'Databases and Bioinformatic Tools for Glycobiology and Glycoproteomics', *Int J Mol Sci*, 21.

Liigand, P., K. Kaupmees, K. Haav, J. Liigand, I. Leito, M. Girod, R. Antoine, and A. Kruve. 2017. 'Think Negative: Finding the Best Electrospray Ionization/MS Mode for Your Analyte', *Anal Chem*, 89: 5665-68.

Lind, T., F. Tufaro, C. McCormick, U. Lindahl, and K. Lidholt. 1998. 'The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate', *J Biol Chem*, 273: 26265-8.

Lindahl, U., J. Couchman, K. Kimata, and J. D. Esko. 2015. 'Proteoglycans and Sulfated Glycosaminoglycans.' in rd, A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger (eds.), *Essentials of Glycobiology* (Cold Spring Harbor (NY)).

Linhardt, R. J., F. Y. Avci, T. Toida, Y. S. Kim, and M. Cygler. 2006. 'CS lyases: structure, activity, and applications in analysis and the treatment of diseases', *Adv Pharmacol*, 53: 187-215.

Liu, M., S. Rehman, X. Tang, K. Gu, Q. Fan, D. Chen, and W. Ma. 2018. 'Methodologies for Improving HDR Efficiency', *Front Genet*, 9: 691.

Liu, M., M. A. Weiss, A. Arunagiri, J. Yong, N. Rege, J. Sun, L. Haataja, R. J. Kaufman, and P. Arvan. 2018. 'Biosynthesis, structure, and folding of the insulin precursor protein', *Diabetes Obes Metab*, 20 Suppl 2: 28-50.

Loh, Y. P., C. R. Snell, and D. R. Cool. 1997. 'Receptor-mediated targeting of hormones to secretory granules: role of carboxypeptidase E', *Trends Endocrinol Metab*, 8: 130-7.

Loomans, C. J. M., N. Williams Giuliani, J. Balak, F. Ringnalda, L. van Gurp, M. Huch, S. F. Boj, T. Sato, L. Kester, S. M. C. de Sousa Lopes, M. S. Roost, S. Bonner-Weir, M. A. Engelse, T. J. Rabelink, H. Heimberg,

R. G. J. Vries, A. van Oudenaarden, F. Carlotti, H. Clevers, and E. J. P. de Koning. 2018. 'Expansion of Adult Human Pancreatic Tissue Yields Organoids Harboring Progenitor Cells with Endocrine Differentiation Potential', *Stem Cell Reports*, 10: 712-24.

Lukinius, A., E. Wilander, B. Eriksson, and K. Oberg. 1992. 'A chromogranin peptide is co-stored with insulin in the human pancreatic islet B-cell granules', *Histochem J*, 24: 679-84.

Ly, M., T. N. Laremore, and R. J. Linhardt. 2010. 'Proteoglycomics: recent progress and future challenges', *OMICS*, 14: 389-99.

Ly, M., F. E. Leach, T. N. Laremore, T. Toida, I. J. Amster, and R. J. Linhardt. 2011. 'The proteoglycan bikunin has a defined sequence', *Nature Chemical Biology*, 7: 827-33.

Mahapatra, N. R., D. T. O'Connor, S. M. Vaingankar, A. P. Hikim, M. Mahata, S. Ray, E. Staite, H. Wu, Y. Gu, N. Dalton, B. P. Kennedy, M. G. Ziegler, J. Ross, and S. K. Mahata. 2005. 'Hypertension from targeted ablation of chromogranin A can be rescued by the human ortholog', *J Clin Invest*, 115: 1942-52.

Malmstrom, A. 1984. 'Biosynthesis of dermatan sulfate. II. Substrate specificity of the C-5 uronosyl epimerase', *J Biol Chem*, 259: 161-5.

Malmstrom, A., B. Bartolini, M. A. Thelin, B. Pacheco, and M. Maccarana. 2012. 'Iduronic acid in chondroitin/dermatan sulfate: biosynthesis and biological function', *J Histochem Cytochem*, 60: 916-25.

Manghwar, H., K. Lindsey, X. Zhang, and S. Jin. 2019. 'CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing', *Trends Plant Sci*, 24: 1102-25.

Mann, D. M., Y. Yamaguchi, M. A. Bourdon, and E. Ruoslahti. 1990. 'Analysis of Glycosaminoglycan Substitution in Decorin by Site-Directed Mutagenesis', *Journal of Biological Chemistry*, 265: 5317-23.

Melloul, D., S. Marshak, and E. Cerasi. 2002. 'Regulation of insulin gene transcription', *Diabetologia*, 45: 309-26.

Michalski, A., E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Muller, R. Viner, J. Schwartz, P. Remes, M. Belford, J. J. Dunyach, J. Cox, S. Horning, M. Mann, and A. Makarov. 2012. 'Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes', *Molecular & Cellular Proteomics*, 11.

Mikami, T., and H. Kitagawa. 2013. 'Biosynthesis and function of chondroitin sulfate', *Biochimica Et Biophysica Acta-General Subjects*, 1830: 4719-33.

Miller, R. L., S. E. Guimond, R. Schworer, O. V. Zubkova, P. C. Tyler, Y. Xu, J. Liu, P. Chopra, G. J. Boons, M. Grabarics, C. Manz, J. Hofmann, N. G. Karlsson, J. E. Turnbull, W. B. Struwe, and K. Pagel. 2020. 'Shotgun ion mobility mass spectrometry sequencing of heparan sulfate saccharides', *Nat Commun*, 11: 1481.

Miyazaki, J., K. Araki, E. Yamato, H. Ikegami, T. Asano, Y. Shibasaki, Y. Oka, and K. Yamamura. 1990. 'Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms', *Endocrinology*, 127: 126-32.

Mizumoto, S., S. Yamada, and K. Sugahara. 2014. 'Human Genetic Disorders and Knockout Mice Deficient in Glycosaminoglycan', *Biomed Research International*, 2014.

Mizumoto, S. 2015. 'Mutations in Biosynthetic Enzymes for the Protein Linker Region of Chondroitin/Dermatan/Heparan Sulfate Cause Skeletal and Skin Dysplasias', *Biomed Research International*, 2015: 861752.

Moens, S. J. B., H. L. Mooij, H. C. Hassing, J. K. Kruit, J. J. Witjes, M. A. J. V. de Sande, A. J. Nederveen, D. Xu, G. M. Dallinga-Thie, J. D. Esko, E. S. G. Stroes, and M. Nieuwdorp. 2014. 'Carriers of Loss-of-Function Mutations in EXT Display Impaired Pancreatic Beta-Cell Reserve Due to Smaller Pancreas Volume', *Plos One*, 9.

Mookherjee, A., and M. Guttman. 2018. 'Bridging the structural gap of glycoproteomics with ion mobility spectrometry', *Current Opinion in Chemical Biology*, 42: 86-92.

Nakagawa, Y., T. Nishikimi, K. Kuwahara, A. Fujishima, S. Oka, T. Tsutamoto, H. Kinoshita, K. Nakao, K. Cho, H. Inazumi, H. Okamoto, M. Nishida, T. Kato, H. Fukushima, J. K. Yamashita, W. J. Wijnen, E. E. Creemers, K. Kangawa, N. Minamino, K. Nakao, and T. Kimura. 2017. 'MiR30-GALNT1/2 Axis-Mediated Glycosylation Contributes to the Increased Secretion of Inactive Human Prohormone for Brain Natriuretic Peptide (proBNP) From Failing Hearts', *Journal of the American Heart Association*, 6.

Nakashima, K., Y. Kanda, Y. Hirokawa, F. Kawasaki, M. Matsuki, and K. Kaku. 2009. 'MIN6 is not a pure beta cell line but a mixed cell line with other pancreatic endocrine hormones', *Endocr J*, 56: 45-53.

Nandini, C. D., and K. Sugahara. 2006. 'Role of the sulfation pattern of chondroitin sulfate in its biological activities and in the binding of growth factors', *Adv Pharmacol*, 53: 253-79.

Nasir, W., A. G. Toledo, F. Noborn, J. Nilsson, M. Wang, N. Bandeira, and G. Larson. 2016. 'SweetNET: A Bioinformatics Workflow for

Glycopeptide MS/MS Spectral Analysis', *J Proteome Res*, 15: 2826-40.

Nilsson, J., U. Ruetschi, A. Halim, C. Hesse, E. Carlsohn, G. Brinkmalm, and G. Larson. 2009. 'Enrichment of glycopeptides for glycan structure and attachment site identification', *Nat Methods*, 6: 809-11.

Noborn, F., A. Gomez Toledo, C. Sihlbom, J. Lengqvist, E. Fries, L. Kjellen, J. Nilsson, and G. Larson. 2015. 'Identification of chondroitin sulfate linkage region glycopeptides reveals prohormones as a novel class of proteoglycans', *Mol Cell Proteomics*, 14: 41-9.

Nolan, E. M., T. C. Cheung, D. W. Burton, and L. J. Deftos. 1995. 'Identification and characterization of a neuroendocrine-specific 5'-regulatory region of the human chromogranin A gene', *Endocrinology*, 136: 5632-8.

North, S. J., P. G. Hitchen, S. M. Haslam, and A. Dell. 2009. 'Mass spectrometry in the analysis of N-linked and O-linked glycans', *Curr Opin Struct Biol*, 19: 498-506.

O'Connor, D. T., P. E. Cadman, C. Smiley, R. M. Salem, F. W. Rao, J. Smith, S. D. Funk, S. K. Mahata, M. Mahata, G. Wen, L. Taupenot, C. Gonzalez-Yanes, K. L. Harper, R. R. Henry, and V. Sanchez-Margalet. 2005. 'Pancreastatin: Multiple actions on human intermediary metabolism in vivo, variation in disease, and naturally occurring functional genetic polymorphism', *Journal of Clinical Endocrinology & Metabolism*, 90: 5414-25.

Ohtsubo, K., and J. D. Marth. 2006. 'Glycosylation in cellular mechanisms of health and disease', *Cell*, 126: 855-67.

Olsen, J. V., S. E. Ong, and M. Mann. 2004. 'Trypsin cleaves exclusively C-terminal to arginine and lysine residues', *Molecular & Cellular Proteomics*, 3: 608-14.

Olson, S. K., J. R. Bishop, J. R. Yates, K. Oegema, and J. D. Esko. 2006. 'Identification of novel chondroitin proteoglycans in Caenorhabditis elegans: embryonic cell division depends on CPG-1 and CPG-2', *Journal of Cell Biology*, 173: 985-94.

Pacheco, B., M. Maccarana, and A. Malmstrom. 2009. 'Dermatan 4-O-sulfotransferase 1 is pivotal in the formation of iduronic acid blocks in dermatan sulfate', *Glycobiology*, 19: 1197-203.

Pacheco, B., A. Malmstrom, and M. Maccarana. 2009. 'Two dermatan sulfate epimerases form iduronic acid domains in dermatan sulfate', *J Biol Chem*, 284: 9788-95.

Parmer, R. J., X. P. Xi, H. J. Wu, L. J. Helman, and L. N. Petz. 1993. 'Secretory protein traffic. Chromogranin A contains a dominant targeting signal for the regulated pathway', *J Clin Invest*, 92: 1042-54.

Perkins, D. N., D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell. 1999. 'Probability-based protein identification by searching sequence databases using mass spectrometry data', *Electrophoresis*, 20: 3551-67.

Perry, R. H., R. G. Cooks, and R. J. Noll. 2008. 'Orbitrap mass spectrometry: instrumentation, ion motion and applications', *Mass Spectrom Rev*, 27: 661-99.

Persson, A., A. Gomez Toledo, E. Vorontsov, W. Nasir, D. Willen, F. Noborn, U. Ellervik, K. Mani, J. Nilsson, and G. Larson. 2018. 'LC-MS/MS characterization of xyloside-primed glycosaminoglycans with cytotoxic properties reveals structural diversity and novel glycan modifications', *J Biol Chem*, 293: 10202-19.

Persson, A., J. Nilsson, E. Vorontsov, F. Noborn, and G. Larson. 2019. 'Identification of a non-canonical chondroitin sulfate linkage region trisaccharide', *Glycobiology*, 29: 366-71.

Persson, A., E. Vorontsov, G. Larson, and J. Nilsson. 2020. 'Glycosaminoglycan Domain Mapping of Cellular Chondroitin/Dermatan Sulfates', *Sci Rep*, 10: 3506.

Pitt, J. J. 2009. 'Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry', *Clin Biochem Rev*, 30: 19-34.

Presto, J., M. Thuveson, P. Carlsson, M. Busse, M. Wilen, I. Eriksson, M. Kusche-Gullberg, and L. Kjellen. 2008. 'Heparan sulfate biosynthesis enzymes EXT1 and EXT2 affect NDST1 expression and heparan sulfate sulfation', *Proc Natl Acad Sci U S A*, 105: 4751-6.

Prydz, K. 2015. 'Determinants of Glycosaminoglycan (GAG) Structure', *Biomolecules*, 5: 2003-22.

Prydz, K., and K. T. Dalen. 2000. 'Synthesis and sorting of proteoglycans', *J Cell Sci*, 113 Pt 2: 193-205.

Ravassard, P., Y. Hazhouz, S. Pechberty, E. Bricout-Neveu, M. Armanet, P. Czernichow, and R. Scharfmann. 2011. 'A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion', *J Clin Invest*, 121: 3589-97.

Reily, C., T. J. Stewart, M. B. Renfrow, and J. Novak. 2019. 'Glycosylation in health and disease', *Nat Rev Nephrol*, 15: 346-66.

Riley, N. M., S. A. Malaker, M. D. Driessen, and C. R. Bertozzi. 2020. 'Optimal Dissociation Methods Differ for N- and O-Glycopeptides', *Journal of Proteome Research*, 19: 3286-301.

Roch, C., J. Kuhn, K. Kleesiek, and C. Gotting. 2010. 'Differences in gene expression of human xylosyltransferases and determination of acceptor specificities for various proteoglycans', *Biochem Biophys Res Commun*, 391: 685-91.

Rudd, P., N. G. Karlsson, K. H. Khoo, and N. H. Packer. 2015. 'Glycomics and Glycoproteomics.' in rd, A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger (eds.), *Essentials of Glycobiology* (Cold Spring Harbor (NY)).

Saeedi, P., I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala, K. Ogurtsova, J. E. Shaw, D. Bright, R. Williams, R. Almutairi, P. A. Montoya, A. Basit, S. Besanccon, C. Bommer, W. Borgnakke, E. Boyko, J. L. Chan, H. Divakar, A. Esteghamati, N. Forouhi, L. Franco, E. Gregg, M. Hassanein, C. Ke, D. Levitt, L. L. Lim, G. D. Ogle, D. Owens, M. Pavkov, J. Pearson-Stuttard, A. Ramachandran, W. Rathmann, M. Riaz, D. Simmons, A. Sinclair, E. Sobngwi, R. Thomas, H. Ward, S. Wild, X. L. Yang, L. L. Yuen, P. Zhang, and IDF Diabet Atlas Comm. 2019. 'Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition', *Diabetes Research and Clinical Practice*, 157.

Sato, T., M. Gotoh, K. Kiyohara, T. Akashima, H. Iwasaki, A. Kameyama, H. Mochizuki, T. Yada, N. Inaba, A. Togayachi, T. Kudo, M. Asada, H. Watanabe, T. Imamura, K. Kimata, and H. Narimatsu. 2003. 'Differential roles of two N-acetylgalactosaminyltransferases, CSGalNAcT-1, and a novel enzyme, CSGalNAcT-2. Initiation and elongation in synthesis of chondroitin sulfate', *J Biol Chem*, 278: 3063-71.

Scharfmann, R., L. Rachdi, and P. Ravassard. 2013. 'Concise review: in search of unlimited sources of functional human pancreatic beta cells', *Stem Cells Transl Med*, 2: 61-7.

Scharfmann, R., W. Staels, and O. Albagli. 2019. 'The supply chain of human pancreatic beta cell lines', *J Clin Invest*, 129: 3511-20.

Schjoldager, K. T., and H. Clausen. 2012. 'Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family', *Biochim Biophys Acta*, 1820: 2079-94.

Schjoldager, K. T., Y. Narimatsu, H. J. Joshi, and H. Clausen. 2020. 'Global view of human protein glycosylation pathways and functions', *Nat Rev Mol Cell Biol*.

Schjoldager, K. T., M. B. Vester-Christensen, C. K. Goth, T. N. Petersen, S. Brunak, E. P. Bennett, S. B. Levery, and H. Clausen. 2011. 'A systematic study of site-specific GalNAc-type O-glycosylation modulating proprotein convertase processing', *J Biol Chem*, 286: 40122-32.

Shajahan, A., C. Heiss, M. Ishihara, and P. Azadi. 2017. 'Glycomic and glycoproteomic analysis of glycoproteins-a tutorial', *Anal Bioanal Chem*, 409: 4483-505.

Shen, Y. J. 2014. 'Traffic lights for axon growth: proteoglycans and their neuronal receptors', *Neural Regeneration Research*, 9: 356-61.

Skelin, M., M. Rupnik, and A. Cencic. 2010. 'Pancreatic beta cell lines and their applications in diabetes mellitus research', *ALTEX*, 27: 105-13.

Stadinski, B. D., T. Delong, N. Reisdorph, R. Reisdorph, R. L. Powell, M. Armstrong, J. D. Piganelli, G. Barbour, B. Bradley, F. Crawford, P. Marrack, S. K. Mahata, J. W. Kappler, and K. Haskins. 2010. 'Chromogranin A is an autoantigen in type 1 diabetes', *Nature Immunology*, 11: 225-U5.

Stanley, P. 2011. 'Golgi glycosylation', *Cold Spring Harb Perspect Biol*, 3.

Staples, G. O., and J. Zaia. 2011. 'Analysis of Glycosaminoglycans Using Mass Spectrometry', *Current Proteomics*, 8: 325-36.

Sterner, E., L. Meli, S. J. Kwon, J. S. Dordick, and R. J. Linhardt. 2013. 'FGF-FGFR Signaling Mediated through Glycosaminoglycans in Microtiter Plate and Cell-Based Microarray Platforms', *Biochemistry*, 52: 9009-19.

Stopschinski, B. E., B. B. Holmes, G. M. Miller, V. A. Manon, J. Vaquer-Alicea, W. L. Prueitt, L. C. Hsieh-Wilson, and M. I. Diamond. 2018. 'Specific glycosaminoglycan chain length and sulfation patterns are required for cell uptake of tau versus alpha-synuclein and beta-amyloid aggregates', *J Biol Chem*, 293: 10826-40.

Stutzer, I., D. Esterhazy, and M. Stoffel. 2012. 'The pancreatic beta cell surface proteome', *Diabetologia*, 55: 1877-89.

Sugahara, K., and H. Kitagawa. 2002. 'Heparin and heparan sulfate biosynthesis', *IUBMB Life*, 54: 163-75.

Sugahara, K., T. Mikami, T. Uyama, S. Mizuguchi, K. Nomura, and H. Kitagawa. 2003. 'Recent advances in the structural biology of

chondroitin sulfate and dermatan sulfate', *Curr Opin Struct Biol*, 13: 612-20.

Sun, M., T. Watanabe, H. Bochimoto, Y. Sakai, S. Torii, T. Takeuchi, and M. Hosaka. 2013. 'Multiple sorting systems for secretory granules ensure the regulated secretion of peptide hormones', *Traffic*, 14: 205-18.

Swaney, D. L., C. D. Wenger, and J. J. Coon. 2010. 'Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics', *Journal of Proteome Research*, 9: 1323-29.

Takahashi, I., N. Noguchi, K. Nata, S. Yamada, T. Kaneiwa, S. Mizumoto, T. Ikeda, K. Sugihara, M. Asano, T. Yoshikawa, A. Yamauchi, N. J. Shervani, A. Uruno, I. Kato, M. Unno, K. Sugahara, S. Takasawa, H. Okamoto, and A. Sugawara. 2009. 'Important role of heparan sulfate in postnatal islet growth and insulin secretion', *Biochem Biophys Res Commun*, 383: 113-8.

Takahashi, I., K. Ohashi, and K. Nata. 2012. 'Involvement of heparan sulfate 3-O-sulfotransferase isoform-1 in the insulin secretion pathway', *J Diabetes Investig*, 3: 362-70.

Takahashi, I., S. Yamada, and K. Nata. 2018. 'Effects of heparan sulfate proteoglycan syndecan-4 on the insulin secretory response in a mouse pancreatic beta-cell line, MIN6', *Mol Cell Endocrinol*, 470: 142-50.

Tam, W. W., K. I. Andreasson, and Y. P. Loh. 1993. 'The amino-terminal sequence of pro-opiomelanocortin directs intracellular targeting to the regulated secretory pathway', *Eur J Cell Biol*, 62: 294-306.

Tatemoto, K., S. Efendic, V. Mutt, G. Makk, G. J. Feistner, and J. D. Barchas. 1986. 'Pancreastatin, a Novel Pancreatic Peptide That Inhibits Insulin-Secretion', *Nature*, 324: 476-78.

Taupenot, L., K. L. Harper, and D. T. O'Connor. 2003. 'The chromogranin-secretogranin family', *N Engl J Med*, 348: 1134-49.

Thaysen-Andersen, M., and N. H. Packer. 2014. 'Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome', *Biochim Biophys Acta*, 1844: 1437-52.

Thaysen-Andersen, M., N. H. Packer, and B. L. Schulz. 2016. 'Maturing Glycoproteomics Technologies Provide Unique Structural Insights into the N-glycoproteome and Its Regulation in Health and Disease', *Molecular & Cellular Proteomics*, 15: 1773-90.

Theodoraki, A., Y. Hu, S. Poopalasundaram, A. Oosterhof, S. E. Guimond, P. Disterer, B. Khoo, A. C. Hauge-Evans, P. M. Jones, J. E. Turnbull, T. H. van Kuppevelt, and P. M. Bouloux. 2015. 'Distinct patterns of

heparan sulphate in pancreatic islets suggest novel roles in paracrine islet regulation', *Mol Cell Endocrinol*, 399: 296-310.

Toledo, A. G., J. Pihl, C. B. Spliid, A. Persson, J. Nilsson, M. A. Pereira, T. Gustavsson, S. Choudhary, H. Z. Oo, P. C. Black, M. Daugaard, J. D. Esko, G. Larson, A. Salanti, and T. M. Clausen. 2020. 'An Affinity-Chromatography and Glycoproteomics Workflow to Profile the Chondroitin Sulfate Proteoglycans that interact with malarial VAR2CSA in the Placenta and in Cancer', *Glycobiology*.

Tooze, S. A. 1998. 'Biogenesis of secretory granules in the trans-Golgi network of neuroendocrine and endocrine cells', *Biochim Biophys Acta*, 1404: 231-44.

Topaz, O., D. L. Shurman, R. Bergman, M. Indelman, P. Ratajczak, M. Mizrachi, Z. Khamaysi, D. Behar, D. Petronius, V. Friedman, I. Zelikovic, S. Raimer, A. Metzker, G. Richard, and E. Sprecher. 2004. 'Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis', *Nat Genet*, 36: 579-81.

Tsiatsiani, L., and A. J. Heck. 2015. 'Proteomics beyond trypsin', *FEBS J*, 282: 2612-26.

Turnbull, J. E., D. G. Fernig, Y. Ke, M. C. Wilkinson, and J. T. Gallagher. 1992. 'Identification of the basic fibroblast growth factor binding sequence in fibroblast heparan sulfate', *J Biol Chem*, 267: 10337-41.

Turnbull, J. E., and J. T. Gallagher. 1991. 'Distribution of iduronate 2-sulphate residues in heparan sulphate. Evidence for an ordered polymeric structure', *Biochem J*, 273 (Pt 3): 553-9.

Tykesson, E., A. Hassinen, K. Zielinska, M. A. Thelin, G. Frati, U. Ellervik, G. Westergren-Thorsson, A. Malmstrom, S. Kellokumpu, and M. Maccarana. 2018. 'Dermatan sulfate epimerase 1 and dermatan 4-O-sulfotransferase 1 form complexes that generate long epimerized 4-O-sulfated blocks', *J Biol Chem*, 293: 13725-35.

Uyama, T., H. Kitagawa, J. Tamura Ji, and K. Sugahara. 2002. 'Molecular cloning and expression of human chondroitin N-acetylgalactosaminyltransferase: the key enzyme for chain initiation and elongation of chondroitin/dermatan sulfate on the protein linkage region tetrasaccharide shared by heparin/heparan sulfate', *J Biol Chem*, 277: 8841-6.

Valicherla, G. R., Z. Hossain, S. K. Mahata, and J. R. Gayen. 2013. 'Pancreastatin is an endogenous peptide that regulates glucose homeostasis', *Physiological Genomics*, 45: 1060-71.

Varki, A. 2017. 'Biological roles of glycans', *Glycobiology*, 27: 3-49.

Vieira, R. P., and P. A. S. Mourao. 1988. 'Occurrence of a Unique Fucose-Branched Chondroitin Sulfate in the Body Wall of a Sea-Cucumber', *Journal of Biological Chemistry*, 263: 18176-83.

Volpi, N., F. Galeotti, B. Yang, and R. J. Linhardt. 2014. 'Analysis of glycosaminoglycan-derived, precolumn, 2-aminoacridone-labeled disaccharides with LC-fluorescence and LC-MS detection', *Nature Protocols*, 9: 541-58.

Watkinson, A., A. C. Jonsson, M. Davison, J. Young, C. M. Lee, S. Moore, and G. J. Dockray. 1991. 'Heterogeneity of chromogranin A-derived peptides in bovine gut, pancreas and adrenal medulla', *Biochem J*, 276 (Pt 2): 471-9.

Weigel, P. H., and P. L. DeAngelis. 2007. 'Hyaluronan synthases: a decade-plus of novel glycosyltransferases', *J Biol Chem*, 282: 36777-81.

Weir, G. C., and S. Bonner-Weir. 2011. 'Finally! A human pancreatic beta cell line', *J Clin Invest*, 121: 3395-7.

Weiss, M., D. F. Steiner, and L. H. Philipson. 2000. 'Insulin Biosynthesis, Secretion, Structure, and Structure-Activity Relationships.' in K. R. Feingold, B. Anawalt, A. Boyce, G. Chrousos, W. W. de Herder, K. Dungan, A. Grossman, J. M. Hershman, H. J. Hofland, G. Kaltsas, C. Koch, P. Kopp, M. Korbonits, R. McLachlan, J. E. Morley, M. New, J. Purnell, F. Singer, C. A. Stratakis, D. L. Trence and D. P. Wilson (eds.), *Endotext* (South Dartmouth (MA)).

Whitelock, J. M., and R. V. Iozzo. 2002. 'Isolation and purification of proteoglycans', *Methods Cell Biol*, 69: 53-67.

Winzen, U., G. J. Cole, and W. Halfter. 2003. 'Agrin is a chimeric proteoglycan with the attachment sites for heparan sulfate/chondroitin sulfate located in two multiple serine-glycine clusters', *J Biol Chem*, 278: 30106-14.

Woods, A., and J. R. Couchman. 2018. 'Proteoglycan Isolation and Analysis', *Curr Protoc Cell Biol*, 80: e59.

Wu, C., J. C. Tran, L. Zamdborg, K. R. Durbin, M. X. Li, D. R. Ahlf, B. P. Early, P. M. Thomas, J. V. Sweedler, and N. L. Kelleher. 2012. 'A protease for 'middle-down' proteomics', *Nature Methods*, 9: 822-+.

Xu, D., and J. D. Esko. 2014. 'Demystifying heparan sulfate-protein interactions', *Annu Rev Biochem*, 83: 129-57.

Yada, T., M. Gotoh, T. Sato, M. Shionyu, M. Go, H. Kaseyama, H. Iwasaki, N. Kikuchi, Y. D. Kwon, A. Togayachi, T. Kudo, H. Watanabe, H. Narimatsu, and K. Kimata. 2003. 'Chondroitin sulfate synthase-2. Molecular cloning and characterization of a novel human glycosyltransferase homologous to chondroitin sulfate

glucuronyltransferase, which has dual enzymatic activities', *J Biol Chem*, 278: 30235-47.

Yada, T., T. Sato, H. Kaseyama, M. Gotoh, H. Iwasaki, N. Kikuchi, Y. D. Kwon, A. Togayachi, T. Kudo, H. Watanabe, H. Narimatsu, and K. Kimata. 2003. 'Chondroitin sulfate synthase-3. Molecular cloning and characterization', *J Biol Chem*, 278: 39711-25.

Yamada, H., K. Watanabe, M. Shimonaka, and Y. Yamaguchi. 1994. 'Molecular-Cloning of Brevican, a Novel Brain Proteoglycan of the Aggrecan Versican Family', *Journal of Biological Chemistry*, 269: 10119-26.

Yamauchi, S., S. Mita, T. Matsubara, M. Fukuta, H. Habuchi, K. Kimata, and O. Habuchi. 2000. 'Molecular cloning and expression of chondroitin 4-sulfotransferase', *J Biol Chem*, 275: 8975-81.

Yang, H., C. X. Yang, and T. L. Sun. 2018. 'Characterization of glycopeptides using a stepped higher-energy C-trap dissociation approach on a hybrid quadrupole orbitrap', *Rapid Communications in Mass Spectrometry*, 32: 1353-62.

Yoo, S. H. 1995. 'pH- and Ca(2+)-induced conformational change and aggregation of chromogranin B. Comparison with chromogranin A and implication in secretory vesicle biogenesis', *J Biol Chem*, 270: 12578-83.

Yoo, S. H. 2010. 'Secretory granules in inositol 1,4,5-trisphosphate-dependent Ca2+ signaling in the cytoplasm of neuroendocrine cells', *FASEB J*, 24: 653-64.

Yu, A., J. Zhao, W. Peng, A. Banazadeh, S. D. Williamson, M. Goli, Y. Huang, and Y. Mechref. 2018. 'Advances in mass spectrometry-based glycoproteomics', *Electrophoresis*, 39: 3104-22.

Yu, Y., J. Duan, F. E. Leach, 3rd, T. Toida, K. Higashi, H. Zhang, F. Zhang, I. J. Amster, and R. J. Linhardt. 2017. 'Sequencing the Dermatan Sulfate Chain of Decorin', *J Am Chem Soc*, 139: 16986-95.

Yu, Y., and R. J. Linhardt. 2018. 'Xylosyltransferase 1 and the GAG Attachment Site', *Structure*, 26: 797-99.

Zabet-Moghaddam, M., R. Kruger, E. Heinzle, and A. Tholey. 2004. 'Matrix-assisted laser desorption/ionization mass spectrometry for the characterization of ionic liquids and the analysis of amino acids, peptides and proteins in ionic liquids', *Journal of Mass Spectrometry*, 39: 1494-505.

Zaia, J. 2008. 'Mass spectrometry and the emerging field of glycomics', *Chem Biol*, 15: 881-92.

Zaia, J. 2010. 'Mass Spectrometry and Glycomics', *Omics-a Journal of Integrative Biology*, 14: 401-18.

Zertal-Zidani, S., A. Bounacer, and R. Scharfmann. 2007. 'Regulation of pancreatic endocrine cell differentiation by sulphated proteoglycans', *Diabetologia*, 50: 585-95.

Zhang, C. F., C. R. Snell, and Y. P. Loh. 1999. 'Identification of a novel prohormone sorting signal-binding site on carboxypeptidase E, a regulated secretory pathway-sorting receptor', *Molecular Endocrinology*, 13: 527-36.

Zhang, H., X. J. Li, D. B. Martin, and R. Aebersold. 2003. 'Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry', *Nature Biotechnology*, 21: 660-66.

Zhang, H., Q. Zhu, J. Cui, Y. Wang, M. J. Chen, X. Guo, V. S. Tagliabracci, J. E. Dixon, and J. Xiao. 2018. 'Structure and evolution of the Fam20 kinases', *Nat Commun*, 9: 1218.

Zhang, K., N. Raboanatahiry, B. Zhu, and M. Li. 2017. 'Progress in Genome Editing Technology and Its Application in Plants', *Front Plant Sci*, 8: 177.

Zhang, L. J., G. David, and J. D. Esko. 1995. 'Repetitive Ser-Gly Sequences Enhance Heparan-Sulfate Assembly in Proteoglycans', *Journal of Biological Chemistry*, 270: 27127-35.

Zhang, P., H. Lu, R. T. Peixoto, M. K. Pines, Y. Ge, S. Oku, T. J. Siddiqui, Y. Xie, W. Wu, S. Archer-Hartmann, K. Yoshida, K. F. Tanaka, A. R. Aricescu, P. Azadi, M. D. Gordon, B. L. Sabatini, R. O. L. Wong, and A. M. Craig. 2018. 'Heparan Sulfate Organizes Neuronal Synapses through Neurexin Partnerships', *Cell*, 174: 1450-64 e23.

Zhao, X., B. Yang, K. Solakylidirim, E. J. Joo, T. Toida, K. Higashi, R. J. Linhardt, and L. Y. Li. 2013. 'Sequence Analysis and Domain Motifs in the Porcine Skin Decorin Glycosaminoglycan Chain', *Journal of Biological Chemistry*, 288: 9226-37.

Ziolkowski, A. F., S. K. Popp, C. Freeman, C. R. Parish, and C. J. Simeonovic. 2012. 'Heparan sulfate and heparanase play key roles in mouse beta cell survival and autoimmune diabetes', *J Clin Invest*, 122: 132-41.