LUND UNIVERSITY

FROM SERUM TO TISSUE: GALECTIN-BINDING GLYCOFORMS OF SERUM GLYCOPROTEINS AS FUNCTIONAL BIOMARKERS

Carlsson, Michael

2012

From the department of Laboratory Medicine, Section of Microbiology,
Immunology and Glycobiology, Lund University, Sweden

FROM SERUM TO TISSUE:

GALECTIN-BINDING GLYCOFORMS OF SERUM GLYCOPROTEINS AS FUNCTIONAL BIOMARKERS

LUND UNIVERSITY
Faculty of Medicine

AKADEMISK AVHANDLING

Som med vederbörligt tillstånd av Medicinska Fakulteten vid Lunds universitet för
avläggande av doktorsexamen i medicinsk vetenskap
kommer att offentligen försvaras i Belfragesalen BMC D15, Lund.
fredagen den 4 maj 2012, kl. 9.00.

Handledare: Professor Hakon Leffler

Fakultetsopponent: Docent Hans Wandall
Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine,
University of Copenhagen, Denmark

DOKUMENTDATABLAD enl SIS 61 41 21

Organization LUND UNIVERSITY Department of Laboratory Medicine Section of Microbiology, Immunology and Glycobiology (MIG)	Document name DOCTORAL DISSERTATION Date of issue May 4, 2012 Sponsoring organization
Author(s) Michael Carlsson	

Title and subtitle

From serum to tissue: Galectin-binding glycoforms of serum glycoproteins as functional biomarkers

Abstract

Galectins are small soluble proteins defined by a conserved sequence motif that forms a defining β-galactoside binding site shared among galectins; adjacent binding sites give each galectin a unique fine specificity for a specific subset of galactose containing glycans. A wealth of data now suggests important roles of galectins in both cancer and inflammation. The aim of this project is to get a clearer picture how these simple proteins can play such a wide variety of roles. Our findings suggest a basic cellular mechanism– galectins use their unique fine specificities to bind and regulate the cellular trafficking of glycoproteins, which could lead to functional effects, such as targeting a glycoprotein to either side of a cell, changing the time it stays at the cell surface, and organizing it relative to other glycoproteins. Coupled to the well-known changes of glycans occurring during pathological conditions, this mechanism predicts a rich and complex array of regulatory functions of the galectins. This will depend on the match between the binding specificity and cellular expression of the galectin and the structure and cellular expression of the glycan.

Key words: Galectin, specificity, disease, serum glycoproteins, cancer, intracellular sorting

Classification system and/or index termes (if any):

Supplementary bibliographical information:	Language English
ISSN and key title: 1652-8220 From serum to tissue: Galectin-binding glycoforms of serum glycoproteins as functional biomarkers	ISBN 978-91-86871-92-5

Recipient's notes	Number of pages	Price
	Security classification	

FROM SERUM TO TISSUE:

GALECTIN-BINDING GLYCOFORMS OF SERUM GLYCOPROTEINS AS FUNCTIONAL BIOMARKERS

Michael Carlsson
Section of Microbiology, Immunology and Glycobiology

LUND UNIVERSITY
Faculty of Medicine

2012

ISSN 1652-8220
ISBN 978-91-86871-92-5
Lund University, Faculty of Medicine Doctoral Dissertation Series 2012:30
Printed in Sweden by Tryckeriet i E-huset, LTH, Lund 2012

Till pappa

TABLE OF CONTENTS

LIST OF PAPERS

I. Emma Salomonsson*, **Michael C. Carlsson***, Veronica Osla, Ruth Hendus-Altenburger, Barbro Kahl Knutson, Christopher T. Öberg, Anders Sundin, Rickard Nilsson, Eva Nordberg-Karlsson, Ulf J. Nilsson, Anna Karlsson, James M. Rini, and Hakon Leffler. Mutational Tuning of Galectin-3 Specificity and Biological Function. *The Journal of Biological Chemistry* (2010), 285(45):35079-35091

*Equal contribution

II. **Michael C. Carlsson**, Cecilia Cederfur, Viveka Schaar, Crina I. A. Balog, Adriana Lepur, Franck Touret, Emma Salomonsson, André M. Deelder, Mårten Fernö, Håkan Olsson, Manfred Wuhrer and Hakon Leffler. Galectin-1-binding glycoforms of haptoglobin with altered intracellular trafficking, and increase in metastatic breast cancer patients. *PLoS One* (2011) *6*,e26560. Epub 2011 Oct 18.

III. **Michael C. Carlsson**, Omran Bakoush, Lotta Tengroth, Ola Kilsgård, Johan Malmström, Thomas Hellmark, Mårten Segelmark, and Hakon Leffler. Galectin-8 in IgA Nephritis: Decreased Binding of IgA by Galectin-8 Affinity Chromatography and Associated Increased Binding in Non-IgA Serum Glycoproteins. *Journal of Clinincal Immunology* (2012) 32:246–255

IV. **Michael C. Carlsson**, Crina I. A. Balog, Ola Kilsgård, Thomas Hellmark, Omran Bakoush, Mårten Segelmark, Mårten Fernö, Håkan Olsson, Johan Malmström, Manfred Wuhrer and Hakon Leffler. Different fractions of human serum glycoproteins bind galectin-1 or galectin-8, and their ratio may provide a refined biomarker for pathophysiological conditions in cancer and inflammatory disease. *Biochimica et Biophysica Acta* (2012). Epub 2012 Jan 17.

V. **Michael C. Carlsson**, Per Bengtson and Hakon Leffler. A key role for galectin-3 as regulator of transferrin trafficking after endocytosis. *Submitted*

LIST OF ABBREVIATIONS

AUC	Area under the curve
CA	Cancer antigen
CDG	Congenital disorders of glycosylation
CDT	Carbohydrate deficient transferrin
CE	Capillary electrophoresis
CEA	Carcinoembryonic antigen
CRD	Carbohydrate recognition domain
CRP	C reactive protein
ECM	Extracellular matrix
EGF	Epidermal growth factor
ER	Endoplasmic reticulum
Fuc	Fucose
Gal	Galactose
GalNAc	N-acetylgalactosamine
GalNAcT	N-acetylgalactosaminyltransferase
Glc	Glucuronic acid
GlcNAc	N-acetylglucosamine
GlcNAcT	N-acetylglucosaminyltransferase
HPLC	High-performance liquid chromatography
IgAN	Immunoglobulin A Nephritis
K_d	Dissociation constant
Lac	Lactose
LacNAc	N-acetyllactosamine
Man	Mannose
MS	Mass spectrometry
Neu5Ac	N-acetylneuraminic acid
Neu5Gc	N-glycolylneuraminic acid
PSA	Prostate specific antigen
ROC	Receiver operating characteristic
Siglecs	Sialic acid-binding, immunoglobulin like lectin
TCR	T-cell receptor
Xyl	Xylose

One and three letter codes for amino acids

A=Ala=alanine, C=Cys=cysteine, D=Asp=aspartate, E=Glu=glutamate, F=Phe=phenylalanine, G=Gly=glycine, H=His=histidine, I=Ile=isoleucine, K=Lys=lysine, L=Leu=leucine, M=Met=methionine, N=Asn=asparagine, P=Pro=proline, Q=Glu=glutamine, R=Arg=arginine, S=Ser=serine, T=Thr=threonine, V=Val=valine, W=Trp=tryptophane and Y=Tyr=tyrosine

Summary

Galectins are small soluble proteins defined by a conserved sequence motif that forms a defining β-galactoside binding site shared among galectins; adjacent binding sites give each galectin a unique fine specificity for a specific subset of galactose containing glycans. A wealth of data now suggests important roles of galectins in both cancer and inflammation. The aim of this project is to get a clearer picture of how these simple proteins can play such a wide variety of roles. Our findings suggest a basic cellular mechanism– galectins use their unique fine specificities to bind and regulate the cellular trafficking of glycoproteins, which could lead to functional effects, such as targeting a glycoprotein to either side of a cell, changing the time it stays at the cell surface, and organizing it relative to other glycoproteins. Coupled to the well-known changes of glycans occurring during pathological conditions, this mechanism predicts a rich and complex array of regulatory functions of the galectins. This will depend on the match between the binding specificity and cellular expression of the galectin and the structure and cellular expression of the glycan.

Populärvetenskaplig sammanfattning

Cellerna i kroppen är uppbyggda av fyra olika byggstenar: DNA, proteiner, fetter samt kolhydrater. Kolhydrater förknippas framförallt med energi och det pågår en intensiv debatt om kolhydraternas vara eller icke vara i vår kost. Kolhydraterna har dock också en helt annan funktion, en som är obestridligt oumbärlig för alla levande organismer. I denna roll förekommer de bland annat som en extra tillsatt del på proteiner och fetter och bildar på så vis så kallade glykoproteiner och glykolipider. Dessa finns på ytan av kroppens alla celler och utgör en stor del av de molekyler som finns i vårt blod. Här har kolhydraterna, eller glykanerna som de kallas, en reglerande funktion genom att forma en sorts nedskriven kod på dessa molekyler som informerar omgivningen om vad som skall hända med dem. Avläsningen av denna kod görs av kolhydratsbindande proteiner som kallas lektiner, som finns i vår kropp men även hos infekterande bakterier och virus. Olika lektiner känner igen olika kolhydratsstrukturer och kan på så vis sortera ut specifika glykoproteiner eller glykolipider. Bland annat använder sig flera virus av lektiner för att ta sig in i kroppen och det faktum att en femtedel av befolkningen är immuna mot vinterkräksjukan beror på att dessa saknar den kolhydratsstruktur som viruset använder sig av för att ta sig in i vävnaden. Samspelet mellan glykoproteiner och lektiner är även av stor vikt för flera cellulära processer i kroppen. En cell kan ändra vilka kolhydrater den sätter på sina proteiner beroende på hur den mår och vad den vill utföra. Detta är väldigt uppenbart i cancer då tumörceller börjar producera andra kolhydrater än vad de normalt gör, t. ex för att växa snabbare.

En specifik lektinfamilj är galektinerna som binder till glykoproteiner som bär kolhydrater innehållande galaktos. Genom att binda till dessa kolhydrater, får galektiner en reglerande roll i immunförsvaret, inflammation och cancer. Galektiners bindningsstyrka för ett visst glykoprotein, och därpå följande biologiska effekt, styrs inte bara av förekomsten av galaktos, utan även av angränsande kolhydratsstrukturer

och proteindelar, det vi kallar galektinernas finspecificitet. Denna avhandling ställer tre frågor baserade på detta: vad är mekanismen för denna finspecificitet (arbete I och II)? Hur är den kopplad till biologisk funktion (arbete I,II och V)? Hur är den kopplad till sjukdom (arbete II och III)?

För att få en djupare förståelse för hur galektinernas finspecificitet fungerar har vi först bytt ut (muterat) vissa aminosyror (byggstenarna i proteiner) som bygger upp den kolhydratbindande domänen, och bildat så kallade mutanter av det ursprungliga galektinet. Sedan har vi mätt hur dessa mutationer påverkar bindingen till olika kolhydratstrukturer och glykoproteiner. Genom att mutera galektin-3, kunde vi förbättra bindingen till en del kolhydratstrukturer och försämra den till andra. Genom att sedan testa dessa i ett biologiskt system kunde vi fastställa vilka delar av finspecificiteten som är viktiga för galektinets funktion, vilket i framtiden kan vara en av startpunkterna för utvecklandet av läkemedel som har galektiner som måltavla för sin effekt.

För att förstå vad som händer i celler när ett galektin binder ett glykoprotein separerade vi först två av de vanligast förekommande glykoproteinerna i blod, transferrin och haptoglobin, baserat på om de binder till galektiner eller inte. En mindre del (5%) av transferrinet band till galektin-3 och genom att märka in denna fraktion med fluorescerande färg, och studera dess upptag i celler med mikroskop, kunde vi se att denna galektinbindande fraktion styrdes till andra delar inne i cellen jämfört med den som inte band. I celler som saknar galektin-3 transporteras de båda fraktionerna likadant, men om galektin-3 tillsätts utifrån får man skillnad i transporten. Med andra ord styr galektin-3 hur transferrin ska transporteras i cellen. Transferrin försörjer alla celler i kroppen med järn, en livsviktig funktion, och hur transferrinet rör sig inne i cellen är grundligt studerat. Likväl har ingen studerat hur kolhydraterna på transferrinet påverkar denna process, vilket gör vår upptäckt mycket intressant. Vidare visar det på en biologisk funktion för galektinernas finspecificitet: att välja ut en undergrupp av ett glykoprotein och styra det annorlunda inne i cellen.

En tredjedel av allt haptoglobin som en frisk person har i blodet binder till galektin-1. För att förstå vad detta har för biologisk konsekvens, använde vi oss av samma modell som ovan. Precis som för transferrin styrdes det haptoglobin som band till galektin-1

annorlunda jämfört med det som inte band. Det förefaller som att galektin-1-bindande haptoglobin undviker de delar av cellen som sköter nedbrytning och istället återcirkulerar ut ur cellen igen. Som tidigare nämnts brukar sjukdomar, som exempelvis cancer, medföra att celler producerar glykoproteiner med annorlunda kolhydrater. När vi istället studerade haptoglobin hos patienter med långt framskriden bröstcancer såg vi att mer än hälften band till galektin-1, d.v.s. vi såg en ökning med närmare tjugo procentenheter jämfört med friska individer. Denna förändring skulle betyda att en mindre del haptoglobin bryts ner i cancer än hos friska, vilket i sin tur kan vara fördelaktigt för cancern då haptoglobin har visat sig vara viktig för att tumörer skall kunna växa.

Av den stora mängd studier som gjorts på galektiner har de flesta en koppling till sjukdomstillstånd som inflammation och cancer. Rollen av galektinernas finspecificitet i detta är relativt okänd, men troligen är den kopplad till de förändringar av kolhydratsstrukturer som förknippas med sjukdom. Ett sätt att studera detta är att analysera vilka glykoproteiner, och hur mycket av dessa, som binder till galektiner i blod från sjuka patienter och jämföra med friska. Bland annat så har vi studerat galektin-8s bindning till glykoproteiner i blodet hos en stor mängd patienter (100 st) med en inflammatorisk njursjukdom kallad IgA nefrit. Vi fann att en kolhydratförändring hos en antikropp resulterade i minskad bindning till galektin-8. Denna förändring är tidigare känd men har aldrig kopplats till ett galektin. Dessutom fann vi en förändrad bindning till andra proteiner, vilket innebär att det måste ske en förändring av kolhydratstrukturerna även på dessa, vilket är ett helt nytt fynd.

Varje år dör uppskattningsvis sju miljoner människor i cancer och antalet ökar för varje år även om medicineringen har förbättrats. Tidig upptäckt är kanske det bästa redskapet för att bekämpa cancer men för detta krävs enkla test som kan separera frisk från sjuk långt innan tumören har hunnit spridas. Enklast vore om man i ett simpelt blodprov kunde hitta markörer som kunde tala om detta. Dessa så kallade biomarkörer är kraftfulla verktyg för screening, diagnos och prognos samt utvärdering av olika behandlingar. Svårigheten att få fram en förstklassig biomarkör ligger i att hitta en som är känslig nog att fånga upp alla som är sjuka, utan, att av misstag, plocka ut några friska i denna process. Detta är ett stort problem för PSA-provet som

idag används för att testa män för prostatacancer, och anses därför inte vara tillräcklig bra för screening i vissa länder. Att utnyttja förändringar av kolhydratstrukturer för tidig upptäckt av cancer har alla förutsättningar för att bli en ny generation biomarkörer.

Som tidigare nämnt ser vi en betydande ökning av galektin-1-bindande haptoglobin hos patienter med bröstcancer. Trots att detta är baserat på begränsat antal patienter, är ökningen i galektin-1- bindande haptoglobin statistiskt säkerställd, och är en lika bra biomarkör som många av de som används idag. Även galektin-8 band haptoglobin, men intressant nog inte samma fraktion som galektin-1. Dessutom band galektin-8 mindre mängd haptoglobin hos cancerpatienterna än hos friska. Troligen känner dessa två galektiner igen helt olika kolhydratsstrukturer på haptoglobin. Ökningen av galektin-1 bindande haptoglobin är en bra biomarkör i sig, men genom att kombinera den med minskningen av galektin-8-bindande haptoglobin, d.v.s. ta kvoten av dessa, får man en biomarkör som är överlägsen de som används idag. Även om detta är preliminära resultat, och fler studier behövs, så finns det en god chans att galektiner kan användas för att hitta biomarkörer för tidig upptäckt av cancer. Förutom att de verkar vara *känsliga* nog för att urskilja sjuka individer i en större population, *specifika* nog för att med stor säkerhet kunna skilja sjuk från frisk och med en provtagning som innebär så *lite lidande* för patienten som möjligt har vi visat att de dessutom är kopplade till en biologisk funktion. Dessa, så kallade funktionella biomarkörer, kan förutom att upptäcka och följa sjukdomen, även användas som måltavlor för behandling av sjukdomen.

Introduction to Glycobiology

The sugar code

The four major classes of organic molecules in living systems are proteins, nucleic acids, lipids and carbohydrates. Of these, the carbohydrates, is the far most abundant and makes up a major portion of the biomass on earth; mostly as cellulose or chitin. The name derives historically from the fact that glucose, the first simple carbohydrate purified, has the molecular formula $C_6H_{12}O_6$ and was thought to be a "hydrate of carbon". This term persists even though a carbohydrate is definitely not a hydrated carbon atom and is used to refer loosely to the broad group of polyhydroxylated aldehydes and ketones commonly called sugars[1]. Even if most people recognize carbohydrates as energy source and simply call it food, this is probably its least intriguing role. Dietary intake of carbohydrates is not essential[2], but their biological roles inside the body are[3]. We actually mention these roles in everyday life, but do not reflect that there are carbohydrates behind these processes. They are the D in DNA, the determining antigens in the ABO blood group system, the homing signal for many microbes and the slipperiness of secretions we cough up.

Every life form contains free or covalently attached carbohydrates, known as glycans. The biological functions of glycans are diverse, with different roles in different tissue or at different times. Hence, studying glycans in their natural setting might be necessary to fully elucidate their functions. Indeed, the term glycobiology was coined by Raymond Dwek in 1988 to emphasize the importance of relating glycans back to basic biology instead of analyzing them after isolation and thereby removing them from their biological context[4].

The spectrum of glycan functions span from nonessential activities to key processes that are crucial for the development, function and survival of an organism. Some of

the functions of carbohydrates can be directly related to their structural and physiological properties, such as stabilizing proteins or increasing solubility. A more captivating role for glycans is that they function as a code for conveying biological information, just as DNA encodes protein sequences[5]. This sugar code, similar to any language, contains an alphabet with specific letters (monosaccharides) that can be assembled into a sentence or biological code (glycan structure) by specific writers (glycan processing enzymes) and attached to a protein or lipid to form glycoconjugates. The code (sugar moiety) can be recognized by a reader (lectin) that is familiar with the language and the area, and accordingly interpret and deliver the message (biological consequence). Almost all major biological events including several human disease conditions involve acquired (noninherited) glycosylation changes. Characteristic alterations are observed in several inflammatory disorders such as rheumatoid arthritis[6], cystic fibrosis[3] and IgA nephritis[7]. Moreover, altered glycosylation is also considered to be a universal feature of cancer cells. Code rewriting can occur at different stages of the glycosylation processes (red arrows in figure 1), most of them arise from changes in the expression of glycosyltransferases in the diseased cell (e.g. cancer cell) but can also occur in "normal cells" as a response to the disease[8]. These disease-related modifications are acquired (beside a rare group of diseases caused by a genetic disorder of glycosylation) and are likely to be the effect and not the primary cause of the disease. Nevertheless, several of these aberrant glycan structures have a functional significance in the disease process, related to an altered recognition by lectins. Accordingly, this recognition can be exploited in assays to detect functional diagnostic and prognostic markers of the disease.

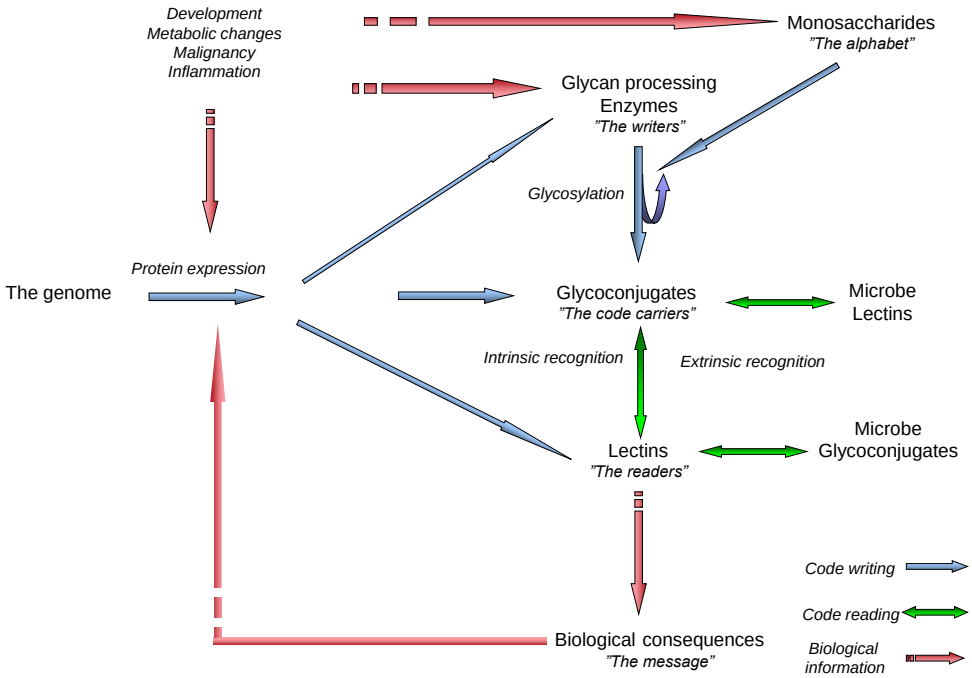

Figure 1. Glycans amplify the genomic information content of cells: a schematic simplified overview of "the sugar code". The assembly of amino acids to form proteins is encoded in the genome and a large portion is destined to the ER and Golgi departments of the cell for the attachment of glycan chains providing dynamic structural and functional diversity. The process of glycosylation (code writing) is driven by the concerted actions of specific glycan processing enzymes such as glycosyltransferases (occasionally glycosylated themselves) which in turn are dependent upon the concentrations and localization of the monosaccharide building blocks (the alphabet). The glycan moieties can be identified by specific glycan binding proteins, referred to as lectins, a recognition that can be either intrinsic (same organism) or extrinsic (different organism, i.e. microbe). The biological consequences of intrinsic recognition include regulation of cell signaling, intracellular trafficking of glycoproteins and cell-cell/matrix interactions[9-11], whereas extrinsic recognition mediates attachment of invading pathogens or their toxins[12]. It should be noted that host lectins also can recognize microbe glycans[13, 14]. Since both lectins and the glycan processing enzymes are encoded in the genome, glycosylation depend upon many factors directly tied to both gene expression and cellular metabolism. Hence, there can be a considerable difference in glycosylation of a glycoconjugate not only between cell types but also within the glycosylation sites in the polypeptide chain. Physiological events such as development, metabolic changes, malignancy and inflammation can significantly influence the glycosylation machinery resulting in altered binding to lectins modifying the specific role of the glycoconjugate. Additional regulation of glycosylation can occur as a consequence of host-pathogen arms races, the so called "red queen effect" discussed further in the coming chapter.

Writing the code

The expanding language

The complexity of the glycome reflects the biodiversity at the cell surface. All cells are covered with a dense coating of glycans named the glycocalyx[15] and at this "outermost" location glycans are in a position to modulate and mediate numerous biological events which are important for the development, growth and function of an organism. As the first entities encountered by incoming pathogens, glycans can also mediate interactions between different organisms i.e. act as specific host cell binding sites for a variety of microbes. Invading microorganisms exploit host cell glycans as recognition targets and a large number of pathogens depend on binding of these to initiate infection. As a consequence the long lived hosts must evade the more rapidly evolving pathogens that infect them by changing their glycans and consequently glycans may be trapped in endless cycles of evolutionary "Red Queen" effects[16, 17]. The Red Queen metaphor was suggested by Van Valen and recalls the comment to Alice by the Red Queen in Lewis Carroll's *Through the Looking-Glass*:

"Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that"[18].

The Red Queen hypothesis captures the idea that an evolutionary advance by one species produces a net negative effect of the same magnitude across all other coexisting species[16, 19]. Thus, the remarkable structural variations of glycans, which play an essential role in biological diversity and specialization, might be attributed to the Red Queen effect. In addition a secondary Red Queen may arise from many pathogens masquerading as "self," by disguising themselves with cell-surface glycans similar to those of their host[20].

Indeed, this is a very complex phenomenon, since the advantage gained under certain circumstances might be weakness under others. For instance, the inactivation of a glycosylation gene in 20% of the human population (called non-secretors) has been associated with decreased susceptibility to infection by caliciviruses causing the winter

vomiting disease[21] but these individuals show increased susceptibility to Streptococcus infections[22]. Even if a significant portion of the overall glycan diversity most likely reflects a pathogen-mediated selection process, a crucial functional role of a glycan would override the Red Queen effect as well as a positive selection of cell-type-specific glycans beneficial for the organism. Thus, the most plausible explanation of the diversification of glycan expression is that it represents a compromise between all these events[17]. In all, the diverse nature of glycans might contribute to formation of new species but the cost of this flexibility might be the incidence of aberrant glycosylation in pathological conditions such as cancer and inflammation.

The main strategies taken by the scientific world to understand glycan functions include targeting the biosynthesis of glycans (chemically or genetically) to prevent or alter the formation of glycan structures and the study of naturally occurring variants of mutants in glycosylation[23]. Pathophysiological roles of glycans in both genetic disorders of glycosylation and in acquired human diseases have also contributed to assign some basic functions to glycans[24]. As Goethe remarked: "It is in her moments of abnormalities that nature reveals her secrets". A major breakthrough was the discovery that the absence of a certain glycan structure resulted in failure to sort specific glycoproteins to the lysosomes, causing the rare Inclusion cell disease ("I-cell" disease)[25]. This particular sorting relies on lectin recognition of the glycan structure and was the first definite observation of a biological role for glycans on mammalian glycoproteins as well as a link between glycan biosynthesis and disease. For a fluent understanding of the complex glycan language we need a deeper knowledge of all its elements, from the structure of the letters, words and sentences to the recognition of these by lectins.

The Letters: Monosaccharides

Carbohydrates are commonly classified into two main groups, simple and complex. Simple sugars, or monosaccharides, such as glucose, cannot be hydrolyzed into smaller molecules; while complex sugars or polysaccharides, composed of two or more monosaccharides linked together can be broken down to their constituent monosaccharide units by hydrolysis[3]. Unlike proteins and nucleotides, which are linear since they contain only one form of linkage, monosaccharides can theoretically

generate either an α or a β linkage to any of several positions on another monosaccharide. This is the primary source of glycan diversity, just the combination of three different monosaccharides could generate up to 27648 unique trisaccharides.

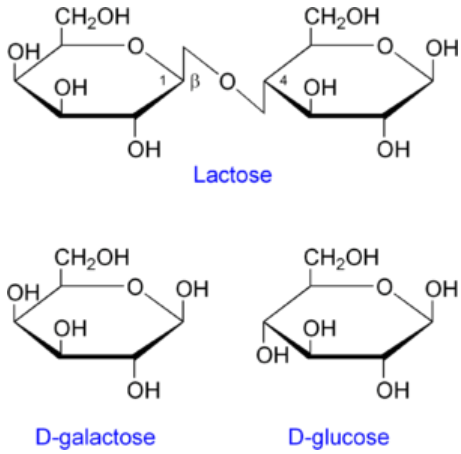

Figure 2. Structure of lactose and the products of its hydrolysis.

There are nine monosaccharides used in the enzymatic process of glycosylation in mammals; glucose (Glc), glucuronic acid (GlcA), xylose (Xyl), mannose (Man), fucose (Fuc), galactose (Gal), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc) and sialic acids (N-acetylneuraminic acid and N-glycolylneuraminic acid, Neu5Ac and Neu5Gc respectively). The monosaccharide iduronic acid is not directly synthesized from a sugar donor but created from glucuronic acid after incorporation in a glycan chain. Together with protein and lipid acceptors they produce different kinds of glycans determined by the assembling glycosyltransferases[3].

The code writers: Glycan processing enzymes

Rudely, glycosylation is often put as the fifth member of posttranslational modifications next to phosphorylation, acetylation, ubiquitination, and methylation. However, due to their vast diversity, the number of glycan modifications by far exceeds the other posttranslational alterations. The attachment of glycan chains is one of the most common and complicated modifications that proteins undergo after

synthesis and are controlled by factors that differ significantly among cell types and species[26, 27]. The human genome consists of at least 250 glycan building enzymes, the glycosyltransferases and their location and substrate competition primarily control the glycosylation process[28]. Despite the complex nature of glycosylation, glycan expression patterns can be strikingly regulated and clear-cut, for instance T-cells go from peanut agglutinin negative to positive during thymic development[29].

In eukaryotic cells the main events of this assembly occur in the endoplasmic reticulum- Golgi pathway, while other machineries are responsible for the assembly in prokaryotes. The building blocks in this process are monosaccharides e.g. glucose which is taken up in the cell by specific transporters in the cell membrane. These monosaccharides have to be activated into a high-energy donor form such as Uridine diphosphate (UDP) e.g. UDP-Glc. The monosaccharide moiety of these high-energy donor sugars are transferred to the acceptor substrate such as an oligosaccharide or a polypeptide by the action of glycosyltransferases[3, 30]. Each glycosyltransferase is essentially responsible to form a specific glycosidic linkage, the so called "one enzyme-one linkage" concept[31], a hypothesis that is not completely accurate since there are circumstances in which more than one enzyme can catalyze the same linkage from the same donor substrate[32].

Despite that glycosylation genes are highly conserved among species in the same taxonomic group, intraspecies variations in glycosylation can be found, e.g. the ABO blood group system in humans. Interspecies variations are more frequent and can be quite extreme, such as the loss of the common mammalian sialic acid Neu5Gc among humans: one of the few true genetic differences between us and the great apes[33].

The Words: Glycans

Glycans can be associated with proteins or lipid to form glycoproteins, glycolipids or glycosaminoglycans in a process called glycosylation. These glycans consist of polysaccharides that vary in length from a few sugars to several hundreds. The glycans possess a diversity far beyond that of polymers of amino and nucleic acids through their capacity for stereochemical linkage (α or β at any of several positions on another monosaccharide) and becomes even more intricate when the number of

monosaccharide building blocks increases. Thus, these monosaccharides can be assembled into an almost unlimited number of combinations[34] and this unique feature of carbohydrates makes them highly suitable as units of hardware for storing immense amounts of biological information. Storing and receiving information from this high-density coding system is however not as strait forward as for nucleic acids and proteins. Glycan structures cannot be predicted from the genome – instead they are the result of the concerted activity of synthesizing and modifying enzymes, which in turn are regulated by both gene expression and cell metabolism[28].

Given the limited number of genes in the mammalian genome[35], post translational modifications such as glycosylation have a more important role in cell phenotype than was previously expected. The "glycome" (the totality of glycan structures) has the capacity to far exceed that of the genome and proteome[3, 36] and while the genome is essentially the same in all living organisms, the glycome show considerable diversity in its structure and expression. The central dogma of biology that states that DNA leads to RNA leads to protein[37], fails to explain several biological facts such as phenotypic differences between genetically identical cells[38]. In the now revised version glycans serve to bridge the gap between genotype and phenotypic expression[39].

The Sentences: Glycoconjugates

Glycosylation of proteins and lipids, to form glycoproteins and glycolipids, is a key factor in modulating their structures and functions. Glycoconjugate is the term used to describe polymeric substances consisting of carbohydrates covalently linked to a non-carbohydrate moiety. There are five major classes of glycoconjugates, primarily defined by the linkage to the protein or lipid (or lack thereof, see hyaluronan). Two of these are attached to a core protein and therefore defined as glycoproteins (*O*-glycans and *N*-glycans). The third member are the glycolipids, in which the glycans are linked to the lipid moiety ceramide, while glycosylphosphatidylinositol anchors (GPI anchors) are defined as a distinct group of glycolipids that are covalently attached to proteins that anchor them to the cell membrane. The fifth and final class, hyaluronan, appears to exist primarily as a free sugar chain[3]. In this thesis, focus will be on the glycoproteins.

Approximately 50% of the proteins found in nature have been proposed to be glycoproteins[40]. The carbohydrate content of these ranges from less than 1% to over 80%, but for the more common glycoproteins, such as the heavy chain of immunoglobulins and the transport glycoprotein transferrin, the range is 2 -15%[41]. Glycans can be present in various forms of the same glycoprotein, making each glycoprotein a collection of "glycoforms". These glycoforms are highly reproducible in a given physiological state but can alter dramatically in disease[42] and thereby change their molecular properties. For example, the occurrence of a particular glycoform of a specific glycoprotein can significantly affect their serum half-life and potency *in vivo*[43].

The carbohydrate moiety is covalently linked to a given amino acid in the polypeptide backbone, either by an *O*-glycosidic bond to a serine or threonine or by *N*-glycosidic bond to an asparagine. These structures are assembled in a step-wise manner by glycosyltransferases in the ER and the Golgi[23], a process initiated by the uptake of monosaccharides. It has been estimated that 75% of all glycoproteins should be *N*-linked, about 10% both *N*- and *O*-linked and about 15% only *O*-linked[40].

The structure of a sentence: *O*-glycans

The formation of *O*-glycans usually starts with the linkage of N-acetylgalactosamine (GalNAc) or mannose to the OH-group of serine or threonine residues forming O-GalNAc or O-mannose structures. However, some other *O*-linked glycans also exist; O-xylose, O-fucose and O-GlcNAc.

If the first added sugar to the core protein is a mannose, this linkage occurs in the ER and is followed by addition of GlcNAc, galactose and sialic acid in the Golgi, forming complex *O*-mannose structures. Disruption of this pathway gives rise to a distinct group of clinical disorders, which all are forms of congenital muscular dystrophy [24]. *O*-linked xylose initiates the assembly of glycosaminoglycans which are found in heparan sulphate and heparin. Commonly, these are classified as an own group of glycoconjugates but since most of them actually are initiated with *O*-linked xylose[44], this definition may be more accurate. *O*-linked fucose is found on the epidermal growth factor (EGF)[45] as well as the Notch receptor where it is necessary for proper

signaling[46]. The discovery that nuclear and cytosolic proteins could be dynamically glycosylated with a single *O*-GlcNAc moiety[47], demolished the dogma that glycosylation was only found on secreted molecules, on the exterior layer of the plasma membrane or on intracellular organelle membranes. GlcNAc is found on nuclear and cystolic proteins and are now considered as one of the most common posttranslational modifications, with important functions in regulation of cell signaling and protein expression[48, 49]. The intracellular and extracellular pools of glycoproteins are subjected to different evolutionary forces- and are therefore also diverse regarding glycan structure and function[16].

The more classical *O*-glycans start with the addition of a GalNAc initiated by members of the polypeptide N-acetylgalactosaminyltransferases (ppGalNAcTs) family of enzymes in the Golgi[24]. This is followed by the addition of any of galactose, GlcNAc, fucose or sialic acid to generate linear or multi-branched chains. The four most common *O*-GalNAc core structures are designated cores 1 to 4 and differ in their additional extensions (figure 3). Core 1 structures (T-antigen) are most abundant and are formed when the peptide linked GalNAc is substituted by galactose in a β1-3 linkage, whereas core 3 glycans contain a GlcNAc in the 3 position of GalNAc. While the two former are linear, core 2 and core 4 glycans contain a branching GlcNAc added to the 6 position of the core 1 or core 3 glycan respectively[3, 50]. In addition, several less common core structures exist; sialic acid can terminate both the Tn and T-antigen to form sialyl-Tn-antigen or sialyl-T-antigen.

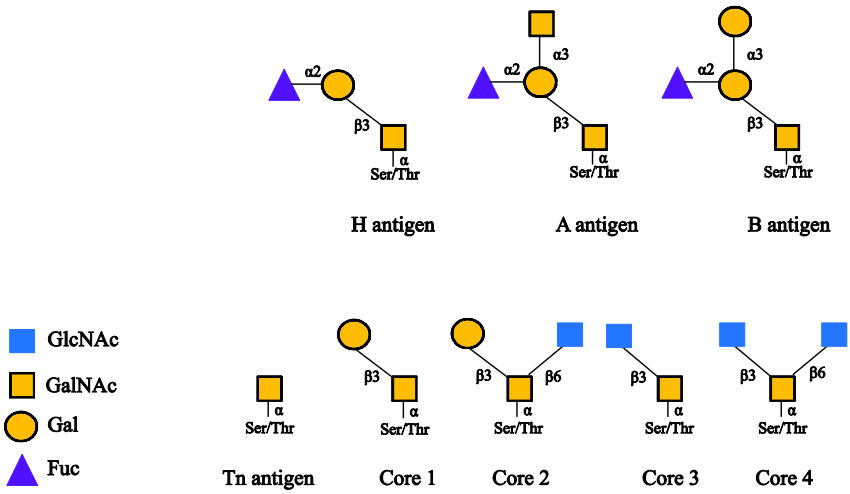

Figure 3. Structures of the major types of *O*-GalNAc cores (1-4) and some common antigenic epitopes. Monosaccharides are symbolized according to the Nomenclature Committee for Functional Glycomics.

Homologues of the ppGalNAcT enzymes driving the *O*-glycosylation machinery is found throughout the animal kingdom but is lacking in plants and fungi[51]. Within eumetazoans there is both a high sequence and functional conservation among ppGalNAcTs suggesting unique biological roles[50]. In mammals at least 15 ppGalNAcTs have been found (1-15) (20 genes have been identified in humans and 15 have been confirmed to be expressed and active) and each tissue or cell type express more than one of these enzymes[50, 52]. They possess a significant degree of overlap in substrate specificities which likely accounts for the lack of a detectable phenotype after deletion of ppGalNAcT-8 in murine T-cells[53]. Further upstream modifications to form a core 1 glycan is synthesized by core 1 β1-3 galactosyltransferase (C1GalT-1) a process that requires the specific molecular chaperone *Cosmc*[54]. Branching of core 1 to form the abundant core 2 structure, important for lymphocyte activation, involves core 2 β1-6 N-acetylglucosaminyltransferase (C2 GalNAcT-1)[29]. Generally, early acting enzymes such as ppGalNAcTs are enriched in *cis*- and medial compartments of the Golgi whereas late acting enzymes such as sialyltransferases are enriched in the *trans*-Golgi

network[52]. The specificities of glycosyltransferases limit the theoretical number of possible *O*-glycans to a few hundred. In addition there are only a limited number of possible assembly pathways. For example, overexpression of α2-6 sialyltransferase (ST6GalNAc), adds sialic acid to the Tn-antigen at an early stage and prevents further addition to this structure. This is rarely seen in normal tissue but is common in cancer[52].

Several early studies suggest that *O*-glycans have important biological roles. For example, *O*-glycans on glycoproteins in the mouse egg's zona pellucida serve as receptors for sperm[55] and increased expression of core 2 *O*-glycans are associated with acquisition of invasive and metastatic potential of mammary carcinoma cells *in vitro*[56] and of bladder tumour cells *in vivo*[57]. The regulation of specific enzyme activities is primarily responsible for these changes and blocking these pathways has provided vast information of functional roles of these glycans[23].

The *O*-GalNAc glycosylation is most likely an essential process since all mammals studied express ppGalNAcTs. However, as mentioned above, the partly overlapping substrate specificity makes it difficult to detect any obvious phenotype after inactivation of different ppGalNAcTs. In this regard it has been speculated that it may be necessary to knockout two or more ppGalNAcTs for any phenotypic effect as seen for *O*-mannosyltransferases[29]. However, at least one member of this family has been shown to be essential for viability and development[31]. Most studies focus on the C2 GalNAcTs due to their involvement in the immune response[58, 59]. It has been demonstrated that a switch in core 2 *O*-glycan expression is associated with T-cell activation and development[59] and mice deficient in C2 GalNAcT displayed defective neutrophil rolling and recruitment to sites of inflammation. The core 1 glycans tend to be crucial for survival since depletion of C1 GalNAcT caused vasculature deformation and impaired angiogenesis and as a result the mice died from fatal brain hemorrhages[60].

The most common defects in the *O*-glycan machinery that lead to human disease occur in the *O*-mannose and *O*-xylose pathways. At least five types of congenital muscular dystrophy (CMD), which can result in severe muscular abnormalities, are caused by mutations in genes responsible for the synthesis of *O*-mannose glycans.

Mutations in the *O*-xylose pathway affect various stages in GAG synthesis and cause distinct disease phenotypes. Defects in the formation of heparan sulphate cause hereditary multiple exostosis (HME) leading to bone and cartilage abnormalities[24].

Defects in *O*-GalNAc glycans can cause familiar tumoral calcinosis, a severe autosomal recessive metabolic disorder that shows phosphataemia and massive calcium deposits in the skin and subcutaneous tissues. Moreover, a rare autoimmune disease, the Tn syndrome, is caused by a mutation in the *Cosmc* gene affecting folding and activity of C1GalT-1 which results in a loss of core 1 glycans on blood cells, exposing the immunoreactive Tn-antigen. These are recognized by anti-Tn antibodies that are present in almost all humans, triggering an autoimmune response leading to a range of disease states such as anemia[54]. IgA nephropathy (IgAN), the most common glomerulonephritis worldwide, is thought to be caused by a similar defect in the *Cosmc* gene. Considerable evidence point towards that the accumulation of IgA in the mesanglial deposits of the kidney originates from circulating aberrantly glycosylated IgA1 complexes in which the hinge-region *O*-linked glycans are deficient in galactose. It is speculated that IgA producing B-cells have an *O*-glycosyltransferase imbalance which results in decreased formation of core 1 structures and instead increased addition of NeuAcα2-6 to the GalNAc[7]. The absence of galactose prevents further addition of NeuAcα2-3 to form the most common *O*-glycan structure found on healthy human IgA[61]. It has been implied that IgG antibodies specific for the Gal-deficient *O*-glycans of IgA1 form complexes that accumulated in the glomerular mesanglial area inducing glomerular injury[62], but this has not been further evaluated.

Functional information can also be found by studying their carriers, i.e. the *O*-GalNAc glycoproteins such as mucins. The mucins are heavily *O*-glycosylated proteins (*O*-glycans may comprise 80% of its molecular weight) that are assembled into homo-oligomers which give mucus its viscous property that forms a physical barrier between lumen and epithelium. Consistent with the general role of glycosylations, *O*-glycosylation of mucins also offers a practically complete protection from protease degradation[63].

The structure of a sentence: *N*-Glycans

The most common protein-attached oligosaccharides are *N*-linked and no mammals are able to survive without them[64]. *N*-glycan synthesis starts with the formation of an amide linkage between GlcNAc and asparagine with a minimal amino acid sequence of Asn-X-Ser/Thr (X represents any amino acid except proline). Of all these sequons it has been estimated that 90% are glycosylated[65]. The biosynthesis of a *N*-glycan is complicated and the pathway begins in the ER with assembly of a precursor oligosaccharide initially synthesized on a lipid like molecule, dolichol phosphate (Dol-P) to form a lipid-linked oligosaccharide containing 2 GlcNAc, 9 mannose and 3 glucose units. This is followed by an "en bloc" transfer in the ER lumen of the entire glycan to a newly synthesized protein with an available asparagine in a suitable context as mentioned above[65, 66]. Indeed, all *N*-glycans share a common sugar sequence: Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn-X-Ser/Thr. *N*-glycans are classified into three main groups: high mannose (or oligomannose), with only mannose residues linked to the pentasaccharide core; complex, in which both terminal GlcNAc residues are attached to both of the terminal mannoses in the core structure; and hybrid, which is a combination of high mannose and complex type with only mannose residues attached to the Manα1-6 and a GlcNAc on the Manα1-3[3].

The early processing after the attachment of the 14 sugar oligomannose glycan to the protein starts with trimming of the glycan by glucosidases and a mannosidase to remove three glucose residues and one mannose residue in the ER. This is also an important check point in protein folding, in which specialized chaperones act by binding to monoglucosylated proteins and prevent them from exciting the ER until properly folded. Upon proper folding the last glucose is removed and the complex dissociates and the glycoprotein is transported to the Golgi for additional processing[67]. Here additional mannose residues are removed, to give $Man_5GlcNAc_2$ which either can escape further processing to form high mannose structures or be subjected to additional modification forming complex or hybrid *N*-glycans.

Biosynthesis of complex and hybrid *N*-glycans are initiated by the addition of GlcNAc to the α1-3 mannose by GlcNAcT-I[65] and mice lacking this enzyme

(disruption of the Mgat1 gene that encodes GlcNAcT-I) die before birth, showing that these carbohydrate structures are essential for proper development[68]. GlcNAc addition is usually followed by the action of α-mannosidase II which removes the two terminal mannoses to form a $GlcNAcMan_3GlcNAc$ which will be further processed to form complex *N*-glycans. If these are not removed, hybrid *N*-glycans will be formed as seen in mutant cells lacking α-mannosidase II activity or cells treated with a α-mannosidase inhibitor which results in the accumulation of hybrid *N*-glycans and inhibition of complex oligosaccharide synthesis[69]. If the two mannoses are removed, a second GlcNAc will be added to the α1-6 Man by GlcNAcT-II which will be a precursor to all complex *N*-glycans. From this biantennary structure more highly branched *N*-glycans (glycans with more than two antennae) can be generated by addition of GlcNAc to the terminal mannoses. Six different types of antennae have been characterized based on to which mannose the GlcNAc is attached (α1-3 or α1-6) and the type of linkage (β1-2, β1-4, β1-6)[70]. These are generated by the actions of GlcNAcTs (IV, V, VI and IX) and form tri-, tetra-, penta- and hexa- antennary structures, see figure 4. Hepta-antennary structures exist, but have so far only been found in birds and fish[71, 72]. Moreover, GlcNAcT-IX expression is restricted to the brain and testis and catalyzes the transfer of a GlcNAc with a β1-6 linkage to the α1-6 mannose[73, 74]. Branching is also dependent on previous glycosylation events. An additional bisecting GlcNAc can be transferred to the "central" or β-linked mannose by GlcNAcT-III. If a bisecting GlcNAc is inserted, this prevents further branching of complex *N*-glycans as this residue inhibits the action of α-mannosidase, resulting in generation of hybrid *N*-glycans[70].

Highly branched complex *N*-glycans are required for normal function of multicellular animals. Incomplete *N*-glycan branching due to absence of α-mannosidase II lead to an autoimmune disease in mice similar to human systemic lupus erythematosus (SLE), probably triggered by innate immune self-recognition[75]. Disruptions in both the β1-4 and β1-6 GlcNAc branching have been shown to cause apparent phenotypic abnormalities, most likely as a consequence of altered residence time of cell surface receptors. Reduction of GlcNAcT-IV expression in mice reduces the residence time of the glucose receptor, resulting in metabolic dysfunction diagnostic of type 2 diabetes[76]. In contrast, an increase in GlcNAcT-V expression as seen in oncogenic transformation[77], modifies the *N*-glycans on cell surface growth receptors delaying

their removal by constitutive endocytosis and consequently, enhancing cell migration and tumour metastasis[9].

Complex *N*-glycans are further processed in the *trans*-Golgi to produce a broad array of mature complex *N*-glycans. This maturation modification can be divided into three main groups: (1) sugar additions to the core, such as the addition of α1-6 fucose to the innermost GlcNAc moiety by fucosyltransferase VIII in vertebrates[78]; (2) elongation of GlcNAc residues such as the addition of Gal to a GlcNAc in the core to form a Galβ1-4GlcNAc (LacNAc) which can be further elongated with LacNAc repeats to form poly(LacNAc)[79] and (3) capping of elongated branches with galactose, fucose, GalNAc, sulphate or sialic acid[3].

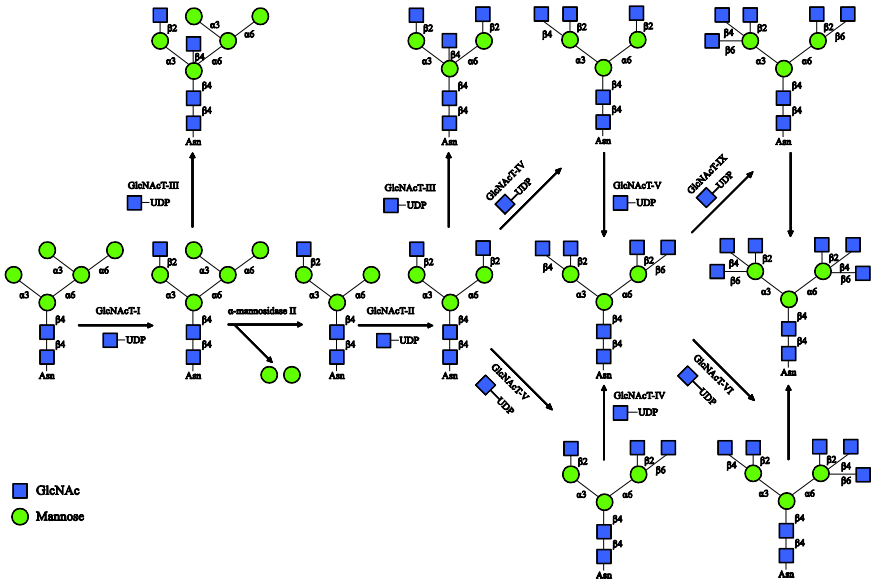

Figure 4. Biosynthetic scheme for hybrid and complex N-glycans. After initial processing in the ER and cis-Golgi, glycoproteins transit to the medial-Golgi to generate hybrid or complex N-glycans due to the action of a number of GlcNAcTs. Trimming of the $Man_5GlcNAc_2$ starts with the action of GlcNAcT-I which adds a GlcNAc to the C-2 of α3Man. It should be noted that no GlcNAcT can act prior to the action of GlcNAcT-I. After the addition of the first GlcNAc, the glycoprotein can be subjected to the actions of either GlcNAcT-III which transfers a GlcNAc to the β-linked Man to generate hybrid N-glycans or α-mannosidase which removes the two Man on the core α6Man in a reaction that will allow addition of GlcNAc to yield a biantennary complex *N*-glycan. Their cell-type specific localization within the Golgi determines which enzyme that will encounter the *N*-glycan first. This will limit the future fate of the *N*-glycan since action of one will prevent additional processing from the other[70]. More highly branched *N*-glycans can be generated by the actions of GlcNAcT (IV, V, VI and XI). The β1-6 GlcNAc branching structures catalyzed by GlcNAcT-V are associated with tumour metastasis[80, 81] and the gene expression of this enzyme is significantly increased during oncogenic transformation[77].

The end of a sentence: Sialic acids

Sialylation on the distal end of glycan chains is a typical characteristic of most intestinal epithelium in mammals. While sialic acids, unlike other monosaccharides, are generally not found in plants, prokaryotes (except bacterial capsules) and invertebrates, they appear to be critical in the development of vertebrates[82]. Sialic acid is defined as a nine-carbon sugar containing an α-ketoacid group and is a large family with over 50 members. The two most common sialic acid structures in mammalian cells are Neu5Ac and Neu5Gc, the latter being absent from normal adult human tissue[83]. The capping with sialic acid (Neu5Ac) has been shown to be a remarkably significant modification on both *N*- and *O*-linked glycans. It is formed by a α-linkage between the C-2 of Neu5Ac and the C-3 or C-6 positions of Gal or the C-6 position of GalNAc (*O*-glycans) and these linkages are directed by distinct sialyltransferases[84]. Sialic acids have been shown to be important contributors to both physiological as well as pathological interactions. Just like their structural variations in nature, the physiological roles of sialic acids are equally diverse, ranging from fertilization and development[85] to the sense of fear[86]. Moreover, sialic acids are central in T-cell homeostasis[87] as well as regulation of IgG properties - loss of sialylation switched IgG from having anti-inflammatory effects at steady state to having pro-inflammatory activity after antigen challenge[88]. These roles are mediated through numerous of cellular and molecular processes. Beside general functions such as stabilization of molecules and membranes and protection from proteases and glycosidases, sialic acids can modulate cell signaling by regulating the affinity of receptors[85]. Here the role of sialic acids are dual: the presence of sialic acid can act either as a glycan mask shielding recognition sites so that they are *not* recognized by lectins[89] or they can themselves serve as ligands for a variety of lectins[90]. Such regulation has been shown to have a fundamental role in immunity by skewing the Th1/Th2 balance by a galectin mediated recognition mechanism[87].

The most well-known role of sialic acids may be as targets for invading microbes such as influenza viruses. Only a subtle difference in structure, i.e. the linkage to the core structure, can have a considerable impact on how viruses infect and spread. For example, the differences in distribution of α2-3 and α2-6-linked sialic acids between

birds and humans account for the rare human-to-human transmission of the avian flu virus[91].

Since sialic acids are involved in so many cellular mechanisms, disturbances in their biosynthesis can cause or be a result of several diseases. Especially, distinct changes in sialylation are associated with malignant transformation, progression and prognosis of tumours. Predominantly these changes are caused by increased activities of specific sialyltransferases and subsequent overall increase of sialic acid content. This increase is often manifested as an increase in α2-6-linked sialic acids[92] but increased amounts of α2-3-linked has also been reported[93]. Examples of specific cancer-associated sialic acid epitopes are sialyl Lewis x (NeuAcα2-3Galβ1-4(Fucα1-3)GlcNAc, sialyl-Tn and polysialic acid[8, 85]. Finally, changes in sialic acid expression is also linked to other diseases, such as previously mentioned IgA nephritis[7] and loss of sialylation on serum transferrin has been found to be an effect of both chronic alcohol consumption[94] and congenital disorders of glycosylation (CDG)[95].

Altered glycosylation in disease

The glycome is influenced by the cell's genome, transcriptome and proteome as well as by environmental factors such as nutrients. Individual glycosylation sites on the same polypeptide may contain different glycan structures that reflect both the type and status of the cell in which they are synthesized. For example, the glycans on one protein can differ significantly depending on if it is expressed in the brain or in immune cells[96]. Hence, determining the glycome would allow us to get a clearer insight on the physiological state of the cell and enable us to expand our knowledge of biological systems and subsequently provide us with new clinical tools for future diagnosis and treatment[24].

Altered glycosylation is a hallmark of disease. This is however not simply the random consequence of a disease associated disturbance, but may be a result of a selective process during disease progression. While glycans are essential for the host cells to function properly they also act as specific ligands for invading microbes[12] and metastasizing tumour cells[80, 97]. Just as host cells are constantly trying to evade the

evolving pathogens, cancer cells can evade the attacking immune cells by altering their glycans[57], (figure 5).

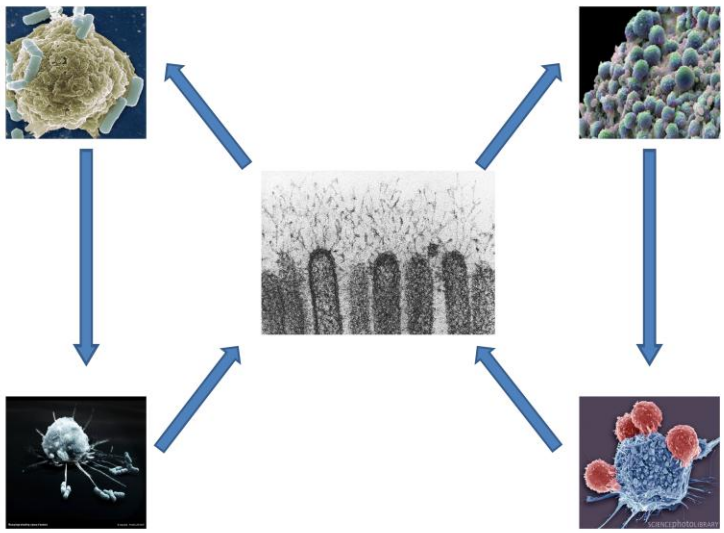

Figure 5. Illustration of the possible selectivness of cancer-associated glycosylation changes. Similar to pathogens evolving to bind host glycans or avoid immune recognition, cancer specific glycan changes shown to mediate tumour growth, metastatic potential and immune evasion, may be driven by a related vicious cycle. Figure showing colour-enhanced scanning electron microscopy pictures of bacteria invading a host cell (top left) macrophage engulfing bacteria (low left) glycocalyx of epithelial cells (center), prostate cancer cells (blue-green) attached to epithelial cells (top right) and T-cells attacking a tumour cell (low left). (Courtesy of Science photo library).

Since just minor alterations in the physiological state of the cell can have a significant impact on the glycosylation machinery, aberrant glycosylation can potentially affect nearly every glycoprotein produced in the diseased cell. Hence, glycans might reflect pathologies in instances when reliable changes in protein profiles cannot be identified. Besides the previously mentioned inherited CDGs, cancer is the best documented disease accompanied by glycosylation changes.

Cancer-associated altered glycosylation patterns was observed over 4 decades ago by the demonstration that healthy fibroblasts have smaller membrane glycoproteins than

their virus-transformed counterparts[98]. Since then, a comprehensive characterization of these cancer-associated changes using lectins, specific monoclonal antibodies and mass spectrometry has provided a sizable array of potential glycan biomarkers[99]. Cancer-associated glycosylation changes take a variety of forms, which have to be distinguished from glycosylation changes originated from non-neoplastic and inflammatory diseases[100] as well as other factors such as age and gender[101]. Still, some striking correlations between certain alteration of glycosylation and malignancy are discussed below.

Increased branching in cancer

Alterations of glycosylation during cancer progression are often a consequence of changed expression of glycosyltransferases in the Golgi leading to modifications of the core structures of both *N*- and *O*-linked glycans. The early observation of an increased size of tumour glycans by Meezan *et al* was later explained by an increase in GlcNAcT-V expression resulting in an increase in β1-6 branching of *N*-glycans[81]. This was followed up by strong functional evidence; first Demetriou *et al* showed that GlcNAcT-V-transfected cells formed solid tumours in mice due to loss of contact inhibition of growth and increased motility[102]. Consistent with this hypothesis, Granovsky *et al* found a striking reduction of tumour growth and metastasis in mice lacking GlcNAcT-V expression[80]. Amplified branching due to increased GlcNAcT-V expression is now considered as major hallmark of cancer progression[103, 104] with several links to a functional, rather than just simply correlative role in tumour growth and metastasis[9, 80, 105, 106]. Also GlcNAcT-IV activity is remarkably increased (up to 66 times) upon oncogenic activation, resulting in formation of abnormal biantennary and triantennary sugar chains[107]. In addition, the increased branching also provides additional sites for terminal sialic acids which ultimately leads to the overall increase in sialylation observed in cancer[8]. It is important to stress that these features not only are found on cell surface glycoproteins but also on secreted glycoproteins. Indeed, increased branching of the glycans on the acute phase serum glycoproteins haptoglobin and α1 acid glycoprotein has been observed in cancer, mostly carrying fucosylated di-, tri- and tetra-branched glycans[108, 109].

In addition, increased branching of *O*-linked glycans, as the previously mentioned additional branching of core 1 to form core 2 *O*-glycans, may be an important feature of cancer. Increased C2GalNAcT expression in bladder tumour cells has been shown to promote a NK cell escape mechanism by silencing NK cell activation, resulting in longer survival of the tumour cells[57].

Incomplete *O*-linked glycans in cancer

In addition to changes in the glycan core structure, alterations in the terminal structures are also a universal feature of cancer. Altered activity of different sialyltransferases has been observed in cancer[110] with the ability to generate a more malignant phenotype, i.e. a more aggressive cell behavior[111]. One study has even reported a mutation in the chaperone protein *Cosmc* in cervical cancer patients as the underlying mechanism for this aberrant glycosylation[112]. One of the most common cancer-associated changes is the expression of truncated *O*-glycans due to "poor glycosylation" on the cell surface. Increased number of Tn-, sialyl-Tn- and T-antigens have been reported in several type of cancer and are associated with cancer malignancy[113, 114]. In conjunction, overexpression of mucins carrying aberrant glycosylations has been described in cancer for many years. Intriguingly, autoantibodies directed against these *O*-glycopeptide epitopes have been found in breast, ovarian, and prostate cancer patients, but are rare in normal adult tissues. The clinical or physiological role of these autoantibodies is unknown but might provide a novel specific form of early biomarkers[115, 116], which will be discussed later.

Sialylated Lewis structures in cancer

The Lewis blood group antigens (such as sialyl-Le^x), are overexpressed in human carcinomas (pancreatic, breast, lung and colon)[8, 117] and correlates with prognosis[118]. Similar to several other cancer-associated changes, this is the result of altered activities of sialyltransferases and fucosyltransferases[119]. These sialylated and fucosylated structures serve as key targets for cell surface lectins, such as selectins, and thus increased expression of sialyl Le^x can promote metastasis by mediating attachment of colon tumour cells to selectins *in vitro*[97]. Acute phase glycoproteins such as

haptoglobin and α1-acid glycoprotein show increased numbers of sialyl-Le^x epitopes in ovarian cancer, but the functional consequences of these glycoforms are unknown[109]. In contrast, no increase of sialyl-Le^x haptoglobin glycoforms was observed in sera from colon cancer patients, only a general increase of fucosylated haptoglobin[120].

Interestingly, for some acute phase proteins (transferrin and α1-antitrypsin) glycosylation is essentially unaffected in cancer with no increase in branching or sialyl-Le^x epitopes, suggesting that particular glycan modifications are relevant for the function of individual glycoproteins in malignancy[109].

Inflammation associated glycans

In contrast to the puzzling functional roles of the cancer-associated glycosylation changes discussed above, chronic inflammation associated glycans are well understood. Upon activation, the immune cell glycome is altered to mediate critical biological functions such as cell adhesion and trafficking, as well as receptor binding and activation. For example, leukocyte migration into lymphoid organs or inflamed tissues during an immune response requires glycan recognition of selectins. The selectins bind sialylated and fucosylated epitopes such as sialyl-Le^x and form a cell–cell adhesion system that operates mainly between leukocytes and endothelial cells[121, 122]. Thus, a general feature observed in any inflammatory response is an increased expression of both selectins, glycoproteins that "carry" sialyl-Le^x and the enzymes that contribute to the synthesis of sialyl-Le^x[123]. Indeed, sialylation plays a crucial role in the inflammatory response, e.g. mice deficient in the α2,3 sialyltransferase show significantly impaired leukocyte adhesion and extravasation[124]. It should also be noted that certain glycans are highly inflammation specific, such as 6-sulpho-sLe^x, which are absent from non-inflamed endothelial tissues, but are generally expressed on the endothelium of chronically inflamed tissues[8].

Altered glycosylation of acute phase proteins are common in response to an inflammatory stimulus. These changes are induced by specific inflammation-derived cytokines that independently of their effects on the production of acute phase

proteins alter the expression of glycosyltransferases in the hepatocytes[125]. Similar to cancer increased branching of glycoproteins has been observed, as well as the inflammation characteristic increase of sialyl-Le^{x}[109].

Specific glycoforms of glycoproteins have been linked to different inflammatory conditions; heavily fucosylated α1 acid glycoprotein have been reported in patients with rheumatoid arthritis and type 1 diabetes[126], poorly galactosylated IgG is associated with rheumatoid arthritis[6] and undergalactosylated IgA has been suggested as the cause of IgA nephritis[7]. Transferrin with decreased sialylation, called carbohydrate deficient transferrin (CDT) is a biomarker for excessive alcohol intake, routinely used to diagnose chronic alcoholism[127]. In contrast to this and the reported conserved glycosylations in cancer, increased sialylation of transferrin has been observed in several chronic inflammatory conditions, such as rheumatoid arthritis and ulcerative colitis[126]. This shows that glycosylation can change in opposite directions depending on the disease. Inflammation related glycosylation changes of haptoglobin have only been found in a few studies, mainly reporting a minor increase of fucosylation, galactosylation and branching of this acute phase protein[108, 120, 128]. Finally, high fat diet induced type 2 diabetes (today considered to be an inflammatory condition) is associated with reduced GlcNAcT-IV expression and subsequent reduced branching of *N*-linked glycans[76].

Alterations in glycosylation as cancer biomarkers

It has been estimated that more than seven million people died of cancer in 2008 worldwide, making cancer the leading cause of death in economically developed countries[129]. Population aging and growth, in combination with an unhealthier lifestyle (physical inactivity and diet), this number will most likely increase over time. To counter this, biomarkers for earlier and more specific diagnosis as well as monitoring the effect of medication are required to increase the survival rate of those afflicted with the disease[129]. Cancer biomarkers are valued according to their specificity and sensitivity, suiting them for different clinical roles. Sensitivity measures the proportion of actual cancer positives which are correctly identified as such, while specificity measures the proportion of cancer negatives which are correctly identified. The most popular measure of biomarker discrimination is the receiver operating

characteristic (ROC) curve, a plot of sensitivity vs. 1 – specificity. Accuracy is measured as area under the curve (AUC) and can range from 0.5 (no predictive ability) to 1 (perfect discrimination) where values being >0.90 are considered excellent in the field[130]. The desire of the cancer biomarker field is to develop simple non-invasive tests that allow early cancer detection, are able to classify tumours (to be able to select the appropriate therapy) and monitor disease progression, regression and recurrence. However, none of the available biomarkers have been able to provide the sensitivity and specificity needed for sufficient clinical impact[131, 132].

Exploiting differences in glycosylation between malignant and healthy tissues will most likely provide excellent opportunities to identify sensitive and specific cancer biomarkers. Abnormal glycosylation in cancer has been observed from early stages (proliferation), to late stages, i.e. invasion and metastasation. For example, increased branching has been connected to prolonged cell surface residence time of growth receptors[9], promotion of cell detachment and increased motility[102]. Glycosylation changes on glycoconjugates are mainly present on the surface of cancer cells, but are also often expressed in the circulation, either on secreted products or by shedding from cell surfaces (and therefore identifiable as serum biomarkers)[133]. Roughly, the circulation is considered to reflect the glycome of all tissues[134].

As discussed earlier these serum biomarkers can be produced in some non-neoplastic tissues and inflammatory conditions such as liver cirrhosis, nephrotic syndrome and pancreatitis[100]. Consequently the glycosylation changes observed in cancer can be less specific for screening purposes due to “clouding” inflammatory glyco-biomarkers. Hence, a cancer-derived glycosylation change unbiased from inflammation would yield a more specific and sensitive biomarker.

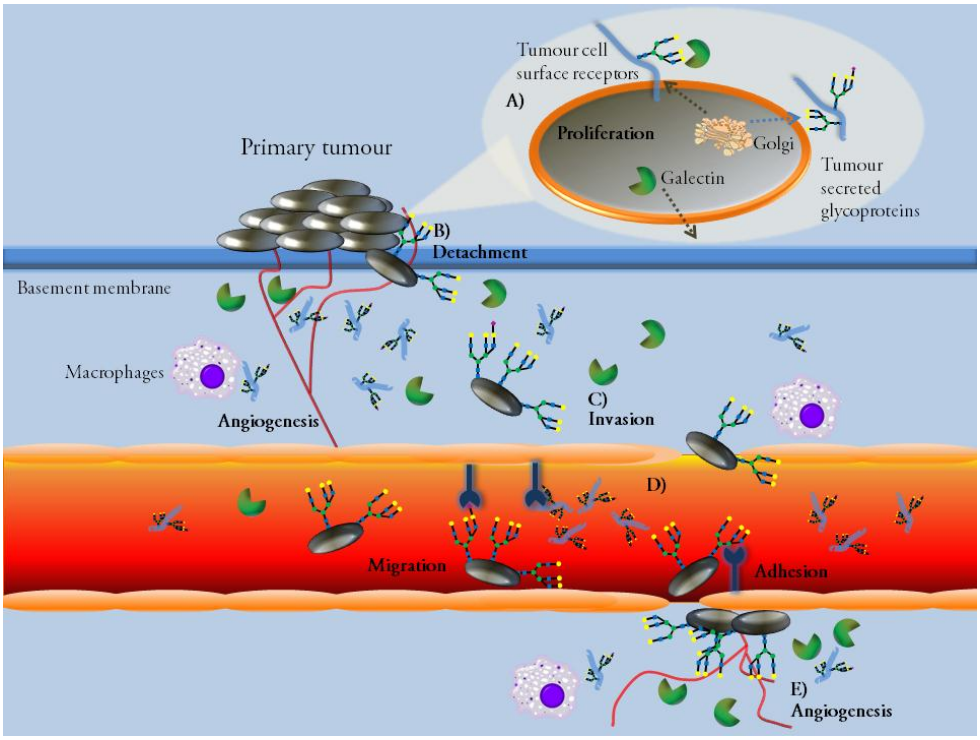

Figure 6. Glycans in all stages of tumour progression. The progression from primary to metastatic tumours is a multigenic and multistep process that involves proliferation (A), detachment (B), cell invasion (C), migration, cell adhesion (D), and angiogenesis (E). These steps are often associated with altered glycosylation of tumour-derived glycoproteins and increased expression of selectins and galectins[97, 135]. (A) Changes in the expression levels of glycosyltransferases in the Golgi compartment of cancerous cells lead to altered glycan structures of cell surface proteins and secreted glycoproteins. One of the most common cancer- associated changes is an increased branching, represented here by a triantennary *N*-glycan. Increased branching has been shown to cause growth receptor hyperactivation due to an increased cell surface residence time resulting in increased cell proliferation[9]. Constitutive endocytosis appears to be prevented by lattice formation following the binding of galectins to these branched *N*-glycans, effectively trapping the glycoprotein receptors at the cell surface. In addition, malignant cells release glycoproteins carrying disease-related glycan epitopes into the extracellular matrix (ECM), where they can reach the circulation. (B) Increased expression of sialylated glycans has been shown to promote dissociation of tumour cells. (C) Tumour cells secrete glycosidases to degrade ECM components as well as galectins, known to attract macophages[136], which contribute further to the ECM breakdown. (D) Sialylated Lewis structures form critical components of most natural ligands for selectins on activated epithelium promoting tumour cell migration. (E) Several evidence support a pro-angiogenetic role for heparan sulphate[137],galectin-1 and galectin-3[138, 139].

Today many of the most widely used clinical cancer biomarkers, such as carcinoembryonic antigen (CEA), cancer antigen 15-3 (CA 15-3) and prostate-specific antigen (PSA) are glycoproteins with increased serum levels in colon, ovarian and prostate cancer, respectively[119]. All of these, however, lack specificity and sensitivity (AUC of about 0.75), two of the main criteria for a good biomarker. In fact, for many cancers there are no serum markers available at all. Despite that the glycan moieties of CEA, CA 15-3 and PSA change during oncogenesis they are monitored solely based on protein levels. Capitalizing on the glycan alterations may boost the clinical performance of these biomarkers significantly, increasing diagnostic sensitivity and specificity[8, 133, 140]. Based on the numerous distinct cancer-derived glycan structures discussed in the preceding chapter, one could imagine that these coupled with a protein may serve as attractive biomarker candidates.

In general, two separate approaches have been applied to detect glycan biomarkers; characterization of the entire genome of a tissue or body fluid or characterization of glycans on a specific glycoprotein. Due to the dynamic nature and pleiotropic effects of glycosylation, the broad "glycomic" overview might be beneficial. However, studying glycans from a specific glycoprotein may provide functional information important for the understanding of the pathogenesis of the disease[28]. We are still far from having the technologies to completely characterize the glycome of even a simple cell. Hence, choosing the best suited approach to address the problem should be based on which level of information is required, whether quantitative or qualitative data are needed and how much material is available[141].

Several technologies offer the possibility of analyzing tumours at the molecular level without subjecting the patients to clinical interventions to obtain tumour tissue (biopsy). The dynamic nature of the circulatory system and its constituents (serum proteins) reflects diverse physiological or pathological states, and the ease with which the blood can be sampled perhaps makes it the most logical choice for biomarker purposes[131]. The main challenge of analyzing serum proteins is their wide abundance range, which spans nine orders of magnitude. The tumour derived proteins are most likely in the lower range of this spectrum, under the detection limit of the mass-spectrometry detectors. Numerous tactics have been used to overcome these issues to achieve an in depth analysis of low abundance plasma proteins as potential

biomarkers[131]. Essentially, these can be divided into three main approaches, of which two involve glycan-lectin recognition, figure 7.

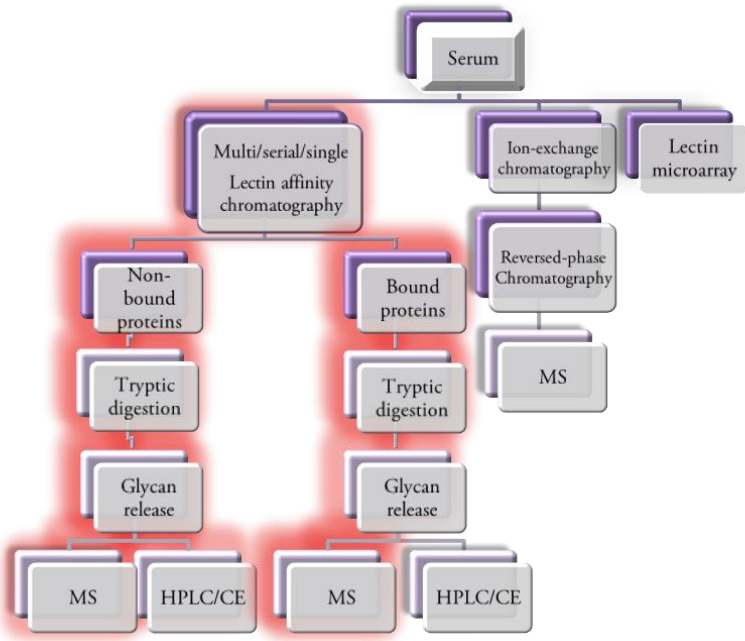

Figure 7. Main approaches for in depth analysis of serum proteins. (1) A proteomic approach is to subject serum proteins to extensive fractionation (by ion-exchange chromatography and reversed-phase chromatography) before MS analysis of individual fractions. (2) Serum samples labeled with fluorescent probes can be directly analyzed with lectin or glycan specific antibody microarrays to provide global information about the types of glycan epitopes present. (3) The third approach (highlighted in red) first selects a subgroup of glycoproteins for detailed analysis using lectin affinity chromatography. These can be used singly, in serial combinations, or as mixtures of lectins depending on which information is required. Fractionated glycoproteins are commonly denaturated and subjected to trypsin digestion before enzymatic release of glycans using an amidase. Profiling of released glycans can be performed either by high-performance liquid chromatography (HPLC) or capillary electrophoresis (CE) after labeling with fluorescent probes, or by different MS methodologies[141].

All these methods have weaknesses; the proteomic approach completely disregards the glycans and subsequently fails to provide biomarkers with high sensitivity and specificity. The high-throughput lectin/antibody microarray neither delivers any detailed structural information nor provides information about the core protein[142].

Lectin enrichment followed by analysis of released glycans probably suffers the least drawbacks, and these are mostly due to the multiple steps involved. Given the wide concentration range of serum proteins, lectin affinity chromatography is usually used prior to analysis for depletion of abundant proteins (or concentration of low abundant proteins). Recently, Zeilinska *et al* showed that using lectin enrichment prior to MS analysis increased the proportion of detected glycans with a factor of about 100[134]. As a result of advances in determining the specificities of many lectins[143], this step may also provide additional information about glycan structure, e.g. linkage. Indeed, by using a combination of plant lectins that selectively enrich sialic-acid-containing glycoproteins from cancer patients, Zhao *et al* gained information about the abundance of specific glycoproteins and the glycosylation state of these, including the linkage of the sialic acids[144]. Because of easy isolation and commercial availability, plant-derived lectins have been the basic choice for glycomic analysis. These, however, suffer from multiple disadvantages including varying activity and availability as well as autoreactivity since most plant lectins are glycoproteins themselves[3]. In addition, despite that plant lectins recognize animal glycans with high specificity (e.g. ricin as discussed in the next chapter) these interactions are not physiologically relevant since they only occur under particular circumstances (e.g. poisoning)[145]. Hence, the use of plant lectin specificity to find endogenous glycoprotein ligands for human lectins is an indirect method and involves several assumptions[146]. Only a few studies have used recombinant human lectin affinity columns to detect natural ligands in serum followed by glycan analysis[106, 147] and only one has been used to identify a cancer biomarker linked to a cellular function[106].

After glycoprotein enrichment, different approaches are required for the release of *N*- and *O*-glycans. Intact *N*-glycans can be released from glycoproteins with peptide N-glycosidase F (PNGase F) or endoglycosidase D (endo D), while endo H selectively cleaves high mannose and hybrid type structures. Up to now, only one specific *O*-glycanase, active against the Core 1 glycan structure has been described[148], which made chemical methods such as alkaline β elimination (despite striking limitations due to the high pH required), the preferred choice for *O*-glycan cleavage. Another chemical method, hydrazinolysis, is currently the best method for generating *O*-glycans with high yield and has therefore been widely adopted.

The sequential analysis of released glycans with HPLC, CE or MS all have limitations. HPLC suffers from unstable baselines, loss of sensitivity and high salt concentrations. EC requires chemical modifications of glycans, while MS primarily lack the sensitivity and quantitative power of the others[28, 141]. Despite its problems, MS is considered to be the key technique for future glycomic studies and recently a number of improvements have been made to overcome some of the issues[149]. The use of a separation technique such as HPLC prior to MS analysis significantly improves sensitivity[134, 150] and other strategies have been developed to yield improved quantitative data[151].

Based on the methods described in the previous paragraph, several promising glycan cancer biomarkers have been detected. To date most of the focus is on the mucins, owing to their high expression in carcinomas and for their proposed contribution to the biological properties of tumours. They also function as scaffolds for most of the above-listed cancer-associated epitopes. Mucins are overexpressed and underglycosylated in most cancers and several diagnostic target glycan epitopes are present on the core protein such as CA 19-9 and CA 125 as well as the previously mentioned CA 15-3[152]. Beside their role as diagnostic biomarkers they appear to have important pathophysiological roles in cancer progression. Detection of these "functional biomarkers" may therefore not only be valuable for the diagnosis but may also be used to develop glycan-based therapies in future cancer treatment. Likewise, one cancer-associated underglycosylated mucin has been used for noninvasive imaging of breast cancer, monitoring the course of the disease[153].

Using the immunogenic response towards aberrantly glycosylated mucin epitopes has great potential as biomarkers for early detection of cancer. Autoantibodies to aberrant *O*-glycopeptide mucin epitopes have been detected in breast, ovarian and prostate patients, while having very low prevalence in healthy controls[115]. As previously noted, cancer marker levels may be increased due to various secondary factors, such as inflammation. For example, CA 15-3 levels increase in chronic active hepatitis and liver cirrhosis[132]. Detecting autoantibodies directed against cancer-associated glycopeptide epitopes may circumvent such effects, since these are essentially absent in sera from inflammatory bowel disease (IBD) patients[154].

Beside the mucins, several other glycoproteins carrying aberrant glycosylations have been suggested as promising biomarkers. A change of haptoglobin glycosylation was first discovered 20 years ago in ovarian cancer and has been shown to be a potential biomarker for other cancer types[155]. Fucosylated haptoglobin has been suggested as a novel marker for pancreatic cancer[156] and an enhanced branching and fucosylation of haptoglobin has been associated with prostate[157] and lung cancer[158]. Just measuring total serum haptoglobin has proven to be a slightly better prognostic marker for hepatocellular carcinoma (HCC), than Alpha-fetoprotein (AFP), the currently best available serum marker[159]. The main disadvantage with using haptoglobin as a cancer biomarker, is that in addition to various malignant diseases, elevated serum levels of this protein can be observed in infection and acute inflammation[125, 160]. In addition, many of the known cancer-derived haptoglobin glycoforms have been detected in chronic inflammatory conditions[109]. Nevertheless, distinct haptoglobin glycoforms unbiased of benign causes have recently been detected in serum from metastatic breast cancer patients and thus provide an interesting novel cancer biomarker candidate[161]. PSA has been used for the diagnosis of prostate cancer and monitoring patient response to treatment for approximately 20 years. However, elevated levels can also be caused by benign prostate diseases and therefore PSA detection can be unspecific and unreliable resulting in false positives. Indeed, in ROC curves, the AUC for the PSA test scored only 0.76[162]. Recently, a number of studies have successfully focused on PSA glycosylation changes as an indicator of prostate cancer to improve diagnostics[163].

Analyzing *N*-glycans of the total serum glycoproteins have also been highly fruitful. The cancer-associated increase in sialyl-Le^x has been shown to be a better indicator of metastasis and breast cancer progression in individual patients than CA 15-3, the most commonly used marker for breast cancer[140, 164]. By analyzing the entire serum glycomes from 100 lung cancer patients and comparing it to 84 age-matched controls, Arnold *et al* found several potential disease markers. The combination of all these glyco-biomarkers gave the best AUC, a remarkable 0.938, unbiased of smoking[158].

Reading the Code

Owing to their structural richness, glycans could be as versatile as the protein backbone that carries them. Nature seems to have taken full advantage of its enormous selection of glycan structures expressed in organisms by evolving an array of specialized proteins that recognize discrete glycan structures deciphering their physiological or pathological message. Specific biological roles of glycans are mediated via recognition of glycan binding proteins. Lectins (from Latin *lectus,* the past tense of *legere*, to select or choose) are glycan binding proteins other than enzymes or antibodies[165]. Nevertheless, interaction of lectins with particular carbohydrates can be as specific as the interaction between those of antigen and antibody or substrate and enzyme[166] and is generated by their carbohydrate recognition domain (CRD). The lectin binding affinities for small carbohydrates are often relatively low, with a K_d in the range of 1000-100 μM, which is too weak to be biologically relevant. Biology may compensate for this low avidity, by multivalent interactions[167]. However, there are some notable exceptions in which most of the binding affinity is provided by a single CRD and the multimerization has more of a functional role[168].

From their discovery in plants in the end of the 19th century, as proteins with the ability to agglutinate erythrocytes, lectins have been found in all living organisms studied[169]. There is no single universal accepted classification of these, but using structural and evolutionary sequence similarities some generally agreed upon names have emerged[3]. Simplified, microbial lectins can be classified into a separate family which includes the well-studied influenza virus hemagglutinin (The "H" in the virus nomenclature) which binds to sialic acid-containing glycans on target epithelial cells of the host. The specificity of the interaction with the sialic acid capped glycan structure varies significantly between different subtypes of the virus, e.g. human or avian strains[170], hence determining the host tropism of the virus as discussed in the previous section.

Eukaryote lectin classification is not straightforward, mostly due to the fact that they have been classified based either on what they bind; ions (Ca^{++})[171], phosphorylated residues[172] or glycan structures (β-galactoside)[173]; where they are primarily found (leguminous plants or *R. communis* seeds)[174]; or if they are related to other protein families (immunoglobulin superfamily)[175]. From this classification six major families of animal lectins emerge, i.e. C-type lectins, I-type lectins, P-type lectins, R-type lectins, L-type lectins and galectins (first classified as S-type lectins).

The classification of *C-type lectins* is based on its CRD architecture and its Ca^{++} dependent glycan binding. The unique compact globular structure of the C-type lectin CRD is also found in proteins that do not bind carbohydrates. There are at least 17 subgroups of C-type lectins, distinguished by their CRD structures[171]. Selectins are most likely the best characterized of these due to their well-studied role in tethering and subsequent rolling of leukocyte from the blood stream into sites of inflammation and lymphatic tissues, probably by high affinity binding to fucosylated core 2 *O*-glycans containing a sulphated sialyl Lewis X determinant on leukocyte glycoproteins[122]. Other members of the family also have well established functions within the immune system such as DC-SIGN, collectins and NK receptors[171]. In addition, a novel class of C-type lectins has displayed antibacterial activity that targets primarily Gram positive bacteria, revealing a primitive mechanism of lectin-mediated innate immunity[176].

L-type lectins were first discovered in the seeds of leguminous plants in which they are primarily found. Some structural motifs of these have also been found in eukaryotic organisms such as pentraxins, a superfamily of plasma proteins in innate immunity, such as the frequently used biomarker C-reactive protein (CRP)[177].

There are only two members of the *P-type lectin* family; cation-dependent mannose 6-phosphate receptor (CD-MPR) and the insulin-like growth factor II/mannose 6-phosphate receptor (IGF-II/MPR). They are distinguished by their ability to recognize phosphorylated mannose residues mainly found in a family of mannose-6-phosphate-bearing *N*-glycans. The predominant role of P-type lectins are as receptors for lysosomal enzyme trafficking[172]. Disruption of this pathway, either by a deficiency in mannose-6-phosphate signal generation or a lack of the two P-type lectin receptors,

results in a failure of intracellular degradation as seen in patients with the I-cell disease[25].

The first *R-type lectin*, ricin, was found as a hemagglutinating protein in the seed extracts of the poisonous plant *Ricinus communis* (castor bean)[145]. Beside its agglutinating properties it is widely known as an extremely potent toxin; less than 0.3 mg is considered a lethal dose for humans depending on the mode of exposure. Ricin preferentially binds β-linked galactose and enters cells via binding to cell surface glycans terminating in the sequence Galβ1–4GlcNAc-R (terminal LacNAc). Inside the cytoplasm, ricin effectively blocks ribosome function and subsequent protein synthesis[145]. Membership in the R-type lectin family requires structural and sequential similarities to the ricin-type CRD. This lectin fold is found in several vertebrate and invertebrate proteins, such as glycosidases and hydrolases. The fact that the ricin-type CRD is the only one conserved between animal and bacterial lectins suggests an early origin for these domains[178]. In animals, the R-type lectin domain is mainly found in the mannose receptor family important for phagocytosis of mannose rich bacteria, leukocyte trafficking to the germinal center and antigen presentation, and in ppGalNAcTs involved in the *O*-glycans biosynthesis discussed earlier. The slightly different acceptor specificity of the lectin domains found between ppGalNAcTs most likely contributes to the large variety of *O*-linked glycoconjugates[179].

All *I-type lectins* contain an immunoglobulin-like fold with non-antibody glycan-binding properties. The family can be divided into two main groups, sometimes referred to as Siglecs (sialic acid-binding, immunoglobulin like lectin) and "others"[180]. Of these the Siglecs are undoubtedly the best characterized I-type lectins, all being integral membrane proteins with an amino terminal V-set domain that binds sialic acid. In contrast to other sialic acid binding lectins, such as the selectins, sialic acid glycan moieties are an absolute requirement for Siglec binding. The Siglecs can be further divided into two subgroups, based on sequence similarity and evolutionary conservation between mammalian species: the conserved Siglecs and human CD33-related Siglecs[181]. The conserved Siglecs includes sialoadhesin, CD22; myelin-associated protein (MAG) and Siglec-15 and have clear orthologues in all mammals. Of these CD22 is the most well studied and has been found to be a strong negative regulator of B-cell signaling. This is due to a complex interplay between CD22, its

sialylated ligands and the B-cell receptor[182]. In addition, CD22 is evolutionary interesting due to its equally strong preference for Neu5Acα2–6Gal and Neu5Gcα2–6Gal structures in humans and in apes, while murine CD22 prefers Neu5Gcα2–6Gal suggesting that this adaption occurred before the human specific loss of Neu5Gcα[83].

In contrast to the conserved Siglecs, the human CD33 related Siglecs have evolved rapidly making it difficult to define orthologues between species. Predominantly expressed on immune cells, the CD33 related Siglecs regulate the immune response, e.g. by down-regulating leukocyte activity. Notably, human T-cells lack CD33 related Siglecs on their surfaces, while our closest ancestors, the chimpanzees express considerable numbers of CD33 related Siglecs on their T-cell surfaces. These Siglecs were particularly abundant on chimpanzee $CD4^+$ T-cells. These T-cells are involved in the pathology of many human diseases, including AIDS, inflammatory bowel disease, rheumatoid arthritis and type 1 diabetes[183]. Remarkably, neither of these diseases have been reported in chimpanzees. Hence, lack of CD33-related Siglec expression in humans may contribute to CD4 T-cell hyperactivity in these disorders[183].

Galectins

In 1852 Charcot and Robins observed small crystals in the blood and spleen of a leukemia patient and 19 years later Ernst Viktor von Leyden found similar crystals in the mucus secretion from a patient suffering from bronchial asthma. Leyden described them as follows: "Amidst this just described graft we found a large number of very delicate soft crystals with a colorless gloss and with uninterrupted point shaped ends creating an octahedron" (translated from Leyden[184]). Similar crystals, termed Charcot-Leyden crystals, were later discovered in both malignant and inflammatory conditions and their origin was tracked to the cytoplasm of eosinophils and basophils[185]. Cloning of the Charcot-Leyden crystal revealed a polypeptide of 16.5 kDa with 142 amino acids similar to members of a superfamily of S-type lactose and IgE-binding animal lectins[186].

The first member of this family was a lactose-inhibited hemagglutinin termed electrolectin, isolated from the extracts of the electronic organ of the eel[187]. The following years, numerous of structurally related lectins were discovered in vertebrates, forming a distinct group of animal lectins. All members in this group share some essential properties: affinity for β-galactosides, one or more free cysteines, extra- and intracellular localization (despite lacking a classical signal peptide for secretion) and soluble in the absence of detergents. These were referred to as S-type lectins to stress their sulphohydryl dependency, since they all required the presence of β-mercaptoethanol to maintain activity (reducing conditions)[188]. When a number of structurally related lectins with affinity for β-galactosides that did not require reduced conditions were identified, such as a ~30 kDa IgE binding protein[189] the term S-type lectin seemed obsolete and confusing. To simplify classification the main co-workers in the field agreed on the general term galectin. They stated that: "Membership in the galectin family requires fulfillment of two criteria: affinity for β-galactosides and significant sequence similarity in the carbohydrate binding site, the relevant amino

acid residues of which have been identified by X-ray crystallography (Lobsanov *et al.,* 1993)"[173].

Galectins form a moderate sized protein family of 15 mammalian members (named galectin-1 to -15) not counting a few galectin like proteins that share the sequence similarity but lacks the β-galactoside binding activity. They are an evolutionary conserved and ubiquitous group of lectins represented in most animal taxa examined so far, even in plants and fungi[190, 191]. While the majority of galectins have orthologues in most mammalian species, galectin-5 has only been found in rat, galectin-6 only in mouse and galectin-15 only in sheep[191]. The galectin family can further be classified based on structural features in the presentation of their CRD, figure 8. A majority of galectins contain one CRD (mono-CRD type or prototype), while some others contain two distinct CRDs connected by a short linker region (bi-CRD type or tandem repeat type). Galectin-3 is the sole mammalian chimera type galectin, with two separate domains consisting of one CRD and one non-CRD N-terminal peptide. This classification was established a year before the term "galectin" was introduced and therefore based on very few proteins[192] and may be passé as discussed in Leffler *et al*[191]. Hence, for clarity, galectins will further on be described as having one or two CRDs (mono-CRD or bi-CRD). Since most galectins act as multivalent lectins, the mono-CRD galectins exist in a dimerized form. This makes them very well suited for cross-linking of ligands and a number of the functional roles of galectins have been attributed to their ability to cross-link glycoprotein receptors at the cell surface, forming lattices that enhance residence time at the cell surface[193].

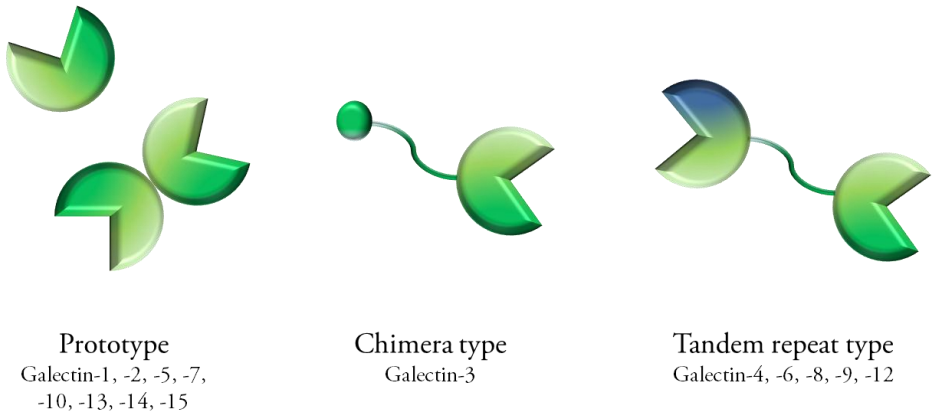

Figure 8. Classification of galectins.

As illustrated by Leffler *et al*, using the term "galectin" to search the PubMed gave over 1150 hits in 2004[191]. Today the same search gives over 3500 hits, so it is a rapidly growing field of research. In these publications, galectins have been assigned important regulatory functions in immunity[194, 195], inflammation[196] and cancer[11]. Galectins are found intracellularly in the nucleus and cytoplasm[197] but are also secreted through a nonclassical mechanism that is not well understood[198]. Given this widespread localization, multiple functions have been assigned to this protein family from RNA splicing[197] to anti-bacterial activity[13]. Some galectins are widely expressed and are distributed in a variety of tissues (e.g. galectin-1, -3 and -8)[199-201], while others are more specific (e.g. galectin-7 in the skin and galectin-12 in adipose tissue)[190]. Malignant transformation is often accompanied with altered galectin expression, usually overexpression[202], but in some cases when normal cells do express high levels of selected galectins, these are downregulated when those cells become cancerous[135]. Increased concentrations of circulating galectins have also been found in several types of cancer. It is however unclear if this is functionally relevant since the concentrations seem relatively low, <1nM[203, 204] (except for galectin-1 in one study[205]).

The crystal structure of galectin-2 solved by Lobsanov *et al* showed that the galectin conserved CRDs consist of about 130 amino acids, arranged in two anti-parallel β-sheets (F1–5 and S1–6)[206]. The two sheets are slightly bent with the F-sheet forming

the convex side and the S-sheet the concave side. The concave side forms the binding groove which is long enough to accommodate a linear tetrasaccharide and may therefore be divided into four subsites A-D[191, 207]. Site C is the defining β-galactoside binding site made up of approximately six conserved amino acids (with some exceptions, e.g. human galectin-4C has an arginine replaced by a lysine not found in other galectins or species). Elongation with a Glc (to form a LacNAc) into site D may enhance affinity hundred-fold[208, 209]. Site D is the second most conserved site but nonetheless a source to the variation in binding specificity between galectins. For example, galectin-1 and -3 prefer LacNAc (Galβ1-4GlcNAc) over lactose (Galβ1-4Glc)[208-211] while galectin-8 displays the opposite profile[212].

In addition to interactions in site D, further elongation (site B and A) with substitutions at positions 2 and 3 of the Gal are tolerated or favored in different ways by different galectins, giving each its unique *fine specificity* for longer saccharides[207, 210, 213, 214]. For example, substitution at the 3 position of Gal with a GalNAcα(Fucα1–2) significantly improves binding by galectin-3, while binding to galectin-1 is reduced[213], and galectin-8 demonstrate particularly high affinity if the β-galactoside is 2–3-sialylated as found in the common *O*-glycan structure NeuAcα2-3Galβ1-4GalNAc[212, 214]. Several biological functions of galectins seem to be determined or tuned by their fine specificity[210, 215], while others seem to depend on a broader binding specificity[212]. Equally important is that substitutions at the 4 and 6 position, such as α2–6-linked sialic acid of Gal, prevent binding to all galectins[214] and has been suggested as an "off switch" for galectin function[89]. Indeed, α2-6-sialylation protected human and mouse T_H2 cells from galectin-1 induced apoptosis resulting in hyper-T_H1 and hyper-T_H17 responses[87].

Given the broad expression and unique specificities of this lectin family, the spectrum of potential binding partners (e.g. glycoproteins) presumably is equally extensive and specific. This suggests different preferences for certain natural ligands of galectins, not only of individual glycoproteins, but also different glycoforms of these. This selective recognition may be a way of regulating glycoproteins, as proposed in several studies[9, 76, 105, 106], and could explain why galectins display such divergent functions. The next section will aim to summarize the link between the specificity, natural ligands and biological function of galectin-1, -3 and -8.

Galectin-1

Human galectin-1 (14kDa) exists as a homodimer in solution and most of the biological activities of galectin-1 require the dimeric form[216]. The dimerization is however not required for enhancing its binding affinity, but more likely permits cross-linking important for biological function[168].

The galectin-1 CRD prefers glycans with terminal LacNAc residues with or without sialic acid at the 3-position[213], but this binding is relatively weak compared to galectin-3[214]. For higher affinity, divalent binding to two glycans at the same time has been proposed[216]. However, a recent study argued against this by showing that galectin-1 interaction with ligands are mainly monovalent[168] and therefore, additional interactions with adjacent glycans and/or neighboring parts of the protein are presumably required for high-affinity galectin-1 binding. Modifications of the glycan itself, e.g. capping with 2-3 linked sialic acid, could be an explanation, but the presence of terminal α2-3NeuAc seems only to be tolerated by galectin-1 and not required for binding to serum glycoproteins[106] or cell surfaces[213, 217].

As mentioned in the previous section, galectin-1 and -3 share some binding features, but differ in others (such as preference for extensions of Gal with GalNAcα or GlcNAcβ in site B). This is reflected in their binding to natural ligands, e.g. galectin-3 binds a considerable fraction and wide range of serum glycoproteins, while galectin-1 (despite being the most abundant of all galectins in serum[205]) binds a smaller fraction of a more limited set of glycoproteins[218]. In addition, galectin-1 shows limited affinity towards blood group determinants[213] signified by that it fails to recognize blood group antigens expressed on *E.coli*[13]. While early studies suggested that galectin-1 mainly binds glycoproteins carrying LacNAc repeats (poly(LacNAc))[219] it is apparent that most of the binding of galectin-1 to serum glycoproteins is not mediated by poly(LacNAc). Instead, galectin-1 binding of serum glycoproteins requires increased proportion of branched *N*-glycans (as found on e.g. tumour-derived glycoproteins) lacking terminal 2-6 linked sialic acid[106].

The fine specificity of galectin-1 shapes its biological functions giving it some unique features in regulation of immunity and tumour progression[11, 194, 199]. In general, galectin-1 is established as a negative regulator of both innate and adaptive immune responses. Early studies showed that galectin-1 is able to induce T-cell death, with CD45 as a potential ligand for this effect[220]. This was later found to be a characteristic of several galectins, including galectin-3[221] and -8[222]. At first, it was suggested that galectin-3 and galectin-1 kill T-cells by binding to the same cell surface glycoproteins, since both of them bind CD29, CD43 and CD45. However, the subtle differences in specificity were later found to dictate how the different galectins can interact with glycosylated receptors on the cell surface. Whereas galectin-1 induced T-cell death requires CD7, galectin-3 does not bind CD7 nor require CD7 to trigger T-cell death. Moreover, CD45 is not required for galectin-1-induced death, while it appears to be required for galectin-3-induced death[223]. In an elegant study, Toscano *et al* revealed that galectin-1, but not galectin-3 selectively induces apoptosis in proinflammatory T_H1 and T_H17 cell subsets but not in T_H2 cells. This "hyper-T_H1" response in galectin-1 null mice was associated with greater susceptibility to autoimmune disease[87]. Accordingly, galectin-1 has demonstrated a protective role by shifting the T_H-cell balance in both type 1 diabetes[224] and autoimmune retinal disease[225] as well as in fetomaternal tolerance[226]. In contrast to the well-studied role for galectin-1 in the adaptive immune response, the potential effects on the innate immune response have only been moderately examined. In general, galectin-1 has been suggested as an anti-inflammatory regulator of macrophage activation, either by promoting alternative activation of macrophages (M2) or by "deactivating" inflammatory macrophages (M1)[227].

There is a distinctive high abundance of galectin-1 in privileged immune sites, such as the brain, placenta and the reproductive organs and growing evidence shows that galectin-1 expression is correlated with aggressive phenotypes in many types of tumours[135, 228, 229]. This suggests that galectin-1's role as negative regulator of immunity may provide an immune escape mechanism for cancer cells, in which tumour-derived galectin-1 might trigger the death of infiltrating tumour targeting T-cells and protect these sites from antitumour immune responses. In fact, blocking of galectin-1 promotes tumour rejection, presumably by shifting the balance in favor of an efficient T-cell-mediated antitumor immune response[230]. In addition, galectin-1

participate in tumour transformation by binding oncogenic H-Ras resulting in Ras membrane anchorage[231] and in cancer cell invasion and metastasis by interactions with a number of ECM components including laminin, fibronectin, vitronectin and integrins[232]. Galectin-1 has also been shown to facilitate tumour metastasis by mediating cell-cell interactions such as the interaction between cancer and endothelial cells[233]. Another important role of galectin-1 in cancer progression is in the tumour vasculature during cancer progression. Galectin-1 is upregulated in angiogenically active endothelial cells and galectin-1–null mice reveal inhibited tumour growth due to decreased angiogenic activity and do not respond to galectin-1 targeted anti-angiogenesis therapy[138, 234].

Galectin-3

The structure of galectin-3 (29kDa) seems to be unique among all vertebrate galectins (but common in invertebrates), consisting of a single polypeptide chain that forms two structurally distinct domains: one uncharacteristic N-terminal domain lacking carbohydrate-binding activity and a C-terminal CRD[191].

Galectin-3 has been revealed to have a remarkable multifaceted fine specificity. Similar to galectin-1 it prefers Galβ1-4GlcNAc over Galβ1-4Glc and does not bind Galβ1-3GalNAc (T-antigen) but in contrast to galectin-1 it also accepts Galβ1-3GlcNAc in sites C-D[210, 213]. Several additions on the 3 position of the galactose are tolerated or may enhance its affinity. Extension with GlcNAcβ and 2-3αNeuAc are tolerated while extensions with GalNAcα and Galα (as found in blood group antigens) increase binding affinities significantly[210]. For example, galectin-3 showed specificity for human blood group A and B antigens expressed on *E.coli* but in contrast to galectin-4 and -8 it did not affect the viability of the bacteria[13]. Unlike galectin-1, which requires the terminal LacNAc for poly(LacNAc) recognition, galectin-3 shows a preference for internal poly(LacNAc), probably due to its tolerance for GlcNAcβ extensions[213]. This may also explain the protective role of terminal α2-6-sialylation on T_H2 cells from galectin-1 but not from galectin-3 induced cell death[87]. Commonly, binding to poly(LacNAc) has been proposed to be involved in galectin-3-

mediated modulation of intracellular trafficking[10], surface retention of growth receptors[9, 104], immune recognition[235] and neutrophil activation[210]. However, as seen for several serum glycoproteins, galectin-3 does not require poly(LacNAc) for high affinity binding (K_d<5μM)[218].

Numerous of natural ligands for galectin-3 have been identified. These are structurally and functionally very diverse and include cytoplasmic proteins, cell surface glycoproteins, glycosylated components of extracellular matrix, and serum glycoproteins[200, 218]. Binding of cell surface glycoproteins commonly involves recognition of poly(LacNAc) structures[105, 210]. For example, galectin-3 binds cell surface LAMP-1 and -2, perhaps the most densely N-glycosylated cellular protein and major carriers of poly(LacNAc) antennae[236]. These are, however, unlikely binding sites for serum glycoproteins since poly(LacNAc) has not been reported among human serum *N*-glycans[237]. Among mature serum glycoproteins, complex, highly branched *N*-glycans are most abundant[237] and these are potentially suitable galectin-3 ligands with the reservation that they carry at least one available LacNAc residue (free or 2–3 sialylated)[213, 218]. Statistically, at least one such structure is present on α-2-macroglobulin and accordingly all bound galectin-3 when subjected to affinity chromatography. Galectin-3 also binds less glycosylated proteins e.g. a small fraction (~5%) of transferrin carrying partially sialylated bi- and triantennary glycans[218]. While this galectin predominantly binds serum glycoproteins known to carry *N*-glycans, binding of *O*-glycans can't be ruled out[218]. Most *O*-linked glycans found on serum proteins are poor galectin-3 ligands, such as the common core 1 structures Galβ1-3GalNAc (T-antigen) and NeuAcα2-3Galβ1-3GalNAc[213]. For example, galectin-3 binds only traces of IgA known to carry *O*-glycans predominantly containing these epitopes. Despite this, galectin-3 has been shown to bind the heavily *O*-glycosylated mucin MUC1, an interaction mediated largely via binding to the T-antigen[238]. In fact, for galectin-3-dependent apical sorting, *O*- and not *N*-glycans seem to play a key role and it is speculated that the sterical arrangement of the *O*-glycan is critical for the binding affinity of galectin-3[239].

The multifaceted specificity of galectin-3 may reflect its different properties for different cell types and/or cell compartments. The preference for LacNAc over lactose seems to be a universal galectin-3 binding feature since it is required for binding to

serum glycoproteins[218], activation of neutrophils, intracellular targeting into vesicles[210] and apical targeting of glycoproteins[10, 239]. In addition, extra numbers and levels of branching increase the probability for galectin-3 binding (similar to galectin-1)[105]. Additional branching is as discussed above considered to be a trademark of cancer progression. Lau *et al* showed that the metabolic flux of the donor substrate, UDP-GlcNAc elicits increased branching of EGFR and TCR *N*-glycans and can therefore serve as regulator of the strength of the association with the galectin-3 lattice and thus a modulator of the choice between growth and arrest[105]. Similarly, shortage in branched N-glycans (Mgat5$^{-/-}$) reduces galectin-3 cross-linking which in turn cause receptor internalization and reduces downstream EGFR and TCR signaling[9, 240].

In addition to the growth promoting properties described above, galectin-3 has important roles during different steps of tumorigenic processes. Similar to galectin-1, galectin-3 interacts with oncogenic activated Ras and promotes strong K-Ras activation that downstream regulates the gene expression at the transcriptional level. Activated Ras is capable of preventing apoptosis and this may explain the anti-apoptotic activity of galectin-3. Indeed, both Ras activation and anti-apoptotic ability was abrogated by a point mutation in the galectin-3 CRD[241, 242]. Further support for a tumour promoting role for galectin-3 is that breast cancer cells transfected with antisense galectin-3 cDNA show significantly decreased cell proliferation[243] and treatment with truncated galectin-3 (galectin-3C) inhibits tumour growth and metastasis *in vivo*[244]. Moreover, galectin-3 has been found to bind and regulate β-catenin in a carbohydrate dependent manner subsequently affecting downstream oncogenic signaling events[245].

In vitro, galectin-3 controls the inflammatory responses by modulating cell adhesion, migration and activation of various innate immune cells. Galectin-3 has been shown to promote neutrophil adhesion to laminin, a mechanism which might have implications in pathological processes such as metastasis[246]. Additionally, galectin-3 promotes neutrophil activation most likely through binding of receptors containing LacNAc repeats (poly(LacNAc))[210]. Studies of galectin-3 null mice suggest a pro-inflammatory role for this galectin *in vivo*. Upon immune challenge, these mice revealed fewer and less activated recruited immune cells[247]. Furthermore, in a mouse model of asthma, the null mice also developed less severe allergic airway inflammation

and airway hyper-responsiveness compared to wild-type mice[248]. Contradictory, intranasal administration of plasmids carrying galectin-3 in a murine model of chronic allergic asthma, led to an early complete blockade of antibody responses, airway hyper-responsiveness, and remodeling[249].

It has been known for 30 years that macrophages express galectin-3[250] and today it is acknowledged as being one of the top five most abundant proteins in macrophages (FO. Martinez, personal communication). Consequently, there are large amounts of information on the function of galectin-3 in macrophages e.g. in phagocytosis, chemotaxis and protection against apoptosis[247]. Galectin-3 expression is significantly higher in alternatively activated macrophages (M2) than in the classically activated macrophages (M1)[251] and may have an important role in regulating this polarization. Evidently galectin-3 drives M2-polarization, since macrophages upon gene disruption of galectin-3 fail to become alternatively activated, both *in vitro* and *in vivo*[252]. Consistently, galectin-3 null mice have a reduced fibrotic phenotype probably as a consequence of reduced numbers of pro-fibrotic M2 macrophages[253].

Galectin-8

Galectin-8 (34kDa) belongs to the bi-CRD galectins (or tandem repeat type) and was first isolated from a rat liver cDNA library screened with an antibody against insulin receptor substrate-1 (IRS-1)[254]. Since galectin-8 bears no sequence similarity either to IRS-1 or to the peptide used as immunogen, this interaction was probably caused by an undefined recognition; e.g. binding of galectin-8 to the IgG, as rabbit IgG contains *O*-glycans to which galectin-8 may bind with high affinity[255] (author's own speculation). The two galectin-8 CRDs show 38% sequence similarity to each other and hence are as different from each other as they are from other members of the lectin family[191]. Undeniably, this sequence dissimilarity is reflected in their specificity. The N-terminal CRD of galectin-8 (galectin-8N) has a particularly high affinity if the β-galactoside is 2–3-sialylated, which is a unique feature among galectins. In contrast to galectin-1 and -3, the preferred sialylated β-galactosides are Galβ1-4Glc, Galβ1-3GlcNAc or Galβ1-3GalNAc and not Galβ1-4GlcNAc. The strong affinity for 2-3-sialylated galactosides is not shared by the C-terminal of

galectin-8 (galectin-8C) which generally displays weaker affinities towards most glycans, but with a slight preference for Galβ1-3(Fucα1-2)Galβ1-3/4GlcNAc (blood group A determinant)[212]. This comparatively weak affinity for blood group antigens is however relevant when it comes to immune recognition. Whereas galectin-8C independently recognized blood group antigens and killed *E. coli*, galectin-8N failed to bind these blood group antigens and was consequently also unsuccessful in killing the bacteria[13].

The unique specificity of the N-terminal appears necessary for binding of serum glycoproteins[256] and integrins[257] as well as for proper intracellular targeting[215], but not for binding to cell surfaces or activation of neutrophils[212]. The galectin-8N preference for Galβ1-3GalNAc over Galβ1-4GlcNAc suggests that it favors *O*- over *N*-linked glycans. Accordingly galectin-8 binds approximately 50 percent of all IgA (carrying mainly *O*-glycans) in serum from healthy individuals[256]. In addition, galectin-8 specifically interacts with CD44 (not known to bind any other galectin) which carries a number of *O*-glycans[258]. Undeniably, galectin-8 binds a slightly different spectra of serum glycoproteins compared to galectin-1 and -3 and BAL-fluid glycoproteins compared to galectin-3[218, 259]. Unlike galectin-3, galectin-8 does not bind significant amounts of serum transferrin and in contrast to galectin-1, binds IgA instead of IgM. Similar to galectin-1 and -3, galectin-8 binds haptoglobin containing mostly *N*-glycans but also suggested to carry at least one *O*-glycan[260]. The presence of an *O*-glycan would explain the observed variation of the galectin-1/-8 haptoglobin binding patterns in different directions with different diseases[161]. Nevertheless, galectin-8 also binds the highly N-glycosylated α-2-macroglobulin not known to carry any *O*-glycans. This could be explained by binding of 2-3-sialylated galactosides on one of the antennae in a fraction of tri- or tetra-antennary N-glycans in serum (a known but less common feature in normal human serum *N*-glycans)[237].

Like other galectins, galectin-8 exhibit immunoregulatory functions such as modulation of immune cell adhesion, migration[261], growth[262], apoptosis[258] and pathogen recognition[13]. Galectin-8 has been shown to bind several members of the main cell adhesion molecules, the integrins, with both inhibitory and stimulatory effects on cell adhesion. Galectin-8 seems to selectively interact with subgroups of β1 integrins while it interacts to a very limited extent with β3 integrins. Addition of

galectin-8 in the presence of serum was shown to inhibit adhesion of human carcinoma cells to plates coated with integrin ligands. Galectin-8 also induced apoptosis in these cells, but in contrast to inhibition of cell adhesion, this required the absence of serum[222]. A year later, another study showed that coating plats with galectin-8 *promotes* cell adhesion and migration and that this was partially inhibited by serum proteins[263]. These findings suggest that complex formation between galectin-8 and serum components (serum glycoproteins) generates a matrix that attenuates cell adhesion. Moreover, galectin-8 has been shown to promote cell adhesion by binding to β1 integrin in activated T-cells and galectin-8 autoantibodies produced in patients with systemic lupus erythematosus could block this process[261].

The galectin-8 binding of integrins, must at least in part, be due to the unique affinity of the N-CRD for 2-3-sialylated galactosides. Indeed, one study showed that β1 integrins in the eye trabecular meshwork (TM), primarily carrying 2-3-linked sialic acid, served as galectin-8N ligands (but not galectin-1 and -3) and promoted cell adhesion[257]. The pro- and anti-adhesive roles of galectin-8 might seem unorganized and confusing. Generally, soluble galectin-8 negatively regulates cell-matrix interactions while immobilized galectin-8 promotes cell adhesion by ligation of cell surface integrin receptors. However, this is further complicated by the suggested complex formation between galectin-8 and serum components.

Galectin-8 is the most frequently expressed galectin in cancer, found in 59 out of 61 tested human tumour cell lines of different origin[202]. It is therefore feasible that the anti-adhesive mechanism operates in tumours, and that tumour derived galectin-8 will augment the detachment of tumour cells and promote metastasis. Paradoxically, overexpression of galectin-8 inhibits cell growth by promoting the accumulation of the cyclin-dependent kinase inhibitor p21[262]. In view of these cytostatic effects, it is an interesting question how the overexpression of galectin-8 observed in cancer could benefit its progression.

Recently, galectin-8 (and galectin-3 and -9) was found to accumulate on damaged *S. Typhimurium* bacteria-containing vesicles, by binding to exposed luminal glycans[264]. By using siRNA to deplete galectin expression, Thurston *et al.* found an enhanced proliferation of bacteria in cells lacking galectin-8, but not -3 and -9. This

was caused by disruption of a galectin-8 mediated sorting mechanism that target bacteria-containing vesicles to lysosomes for degradation. The exact mechanism is unclear, but it seems like galectin-8N senses cytosolic host glycans exposed on damaged vesicles and the C-terminal binds the autophagy receptor NDP52, that directs the complex for lysosomal degradation. Interestingly, this finding may shed some light on the observation that the intracellular targeting of galectin-8 is determined by the unique fine specificity of the N-terminal CRD to 2-3 sialylated galactosides[215].

Present investigation

Aim

It is without doubt that galectins play a key role in both health and disease. This is primarily based on a variety of intriguing observations *in vitro* and *in vivo*, in a context where a particular galectin is absent, e.g. null mutant mice. These studies however fail to provide a link between the galectin's carbohydrate ligands and the resulting biological consequence of the interaction with them. It is like flipping to the last page of an Agatha Christie novel - to try to find out the ending. You will perhaps find out who the murderer is but you will miss out on the rest. In this setting "the rest" may conceal critical elements, such as potential drug targets and/or first-rate biomarkers.

All the observed biological effects of galectins must depend on a match between the structural context of the glycan in the particular glycoprotein and the *fine specificity* of the particular galectin. Alterations in the glycosylation machinery during disease will undoubtedly disturb this pairing and accordingly also modify the corresponding biological effects. In this thesis we are trying to answer three questions based on this:

1. What is the molecular mechanism for the galectin fine specificity?
2. How is this related to biological function?
3. How is this related to disease?

Molecular mechanisms of the galectin fine specificity (paper I and II)

Several approaches have been used to elucidate the mechanism of galectin fine specificity, including frontal affinity chromatography[214], glycan array[213] and fluorescence polarization[265]. These methods differ considerably in some aspects, e.g. how the glycans or galectins are presented (immobilized or in solution) and hence may also generate slightly different binding constants. Moreover, while the first two are powerful high-throughput tools that give an overview of the glycan specificities of a lectin, they fail to provide detailed structural information of the interactions.

Fluorescence polarization is a more sensitive method for studying molecular interactions in solution that does not require physical separation between the galectin and its ligand. The method measures the movement of a fluorescently tagged saccharide (probe) that is excited with plane polarized light. The following emitted light is largely depolarized because the molecule rotates rapidly in solution during the fluorescence event (the time between excitation and emission). However, if the fluorescently tagged saccharide is bound to a much larger molecule (e.g. a galectin) its effective total volume is increased, the rotation reduced and the emitted light becomes less depolarized. In fluorescence polarization or anisotropy assays (polarization and anisotropy is closely related and anisotropy is usually used to simplify calculations) this relationship can be used to estimate the concentrations of bound and free probe and in turn the affinity of the interaction. In detail, a dilution series of galectin is added to a fixed concentration of the fluorescent saccharide probe followed by excitation with plane polarized light. A binding curve can be generated by plotting the anisotropy against the galectin concentration; the anisotropy will increase with increased galectin concentration (if binding occurs) to finally reach saturation. The position and shape of the curve will provide a good approximation of affinity. The method can also be used for indirect binding studies, in which a non-fluorescent ligand (e.g. glycoprotein) competes with the fluorescent probe for galectin binding. The potency of the competing ligand to inhibit the galectin probe interaction (reflected in the % bound probe) can be used to estimate the binding constants[265].

We have previously exploited a combination of both glycan array and fluorescence anisotropy (FA) to characterize the fine specificity of galectin-8 and its separate CRDs. By combining the glycan array analysis, which simultaneously tests the binding to a panel of surface bound glycans, with the solution based FA assay that delivers affinity constants and binding mode, we achieved an in depth analysis of galectin-8's binding specificities[212]. Here (paper I), we use a similar approach for a comprehensive analysis of the galectin-3 fine specificity. From this we could confirm that galectin-3 displays a broad fine specificity with a clear binding to representatives of three structural groups (Galβ1–4GlcNAc, Galβ1–3GlcNAc, Galβ1–4Glc but not Galβ1–3GalNAc) in site C-D as well as evident binding to all representative groups extending into site B (sulfate-3, NeuAcα2–3, GalNAcα1–3, Galα1–3, GlcNAcβ1–3). I an attempt to gain an even more detailed analysis of the galectin-3 fine specificity, we produced 10 mutants of human galectin-3, with changes in or near site B-D and tested their binding. In this way we managed to fine tune the binding specificity of galectin-3, i.e. alter carbohydrate-binding fine specificity while retaining the basic β-galactoside binding activity.

The most striking consequence was the specific loss of affinity for the galectin-3 preferred saccharide LacNAc (Galβ1–4GlcNAc), when substituting Arg186 to Ser or Ile (R186S and R186I). Notably, their affinity for lactose was only slightly reduced, thus giving these mutants a more galectin-8N like binding profile (prefering lactose over LacNAc). This was quite reasonable since galectin-8N has an Ile instead of an Arg at this position. Correspondingly, a mutation in site B, shifting the Ala146 to a Glu (A146Q) as found in galectin-8N, retained the affinity for the 2-3-sialylated galactosides while it decreased binding to all other tested glycan structures.

Replacing the Lys176 with a Leu (K176L), also in site B, indeed fine-tuned the specificity; while it lost affinity for NeuAc substituted galactosides, it *gained* affinity for GlcNAcβ1–3-substituted galactosides, commonly found in poly(LacNAc) structures. The fine specificity of the K176L mutant resembles that of *Xenopus* galectin-3, which has the Lys replaced with a Met, giving some insight in the species specific sequence differences most likely caused by an adaptation to recognize encountered host or pathogen glycans.

From this it is apparent that subtle variations in the CRD significantly alter the specificities of galectins for glycan epitopes, but how does this reflect their binding to endogenous ligands?

A simple and powerful method for purifying binding partners is affinity chromatography. In this technique, the protein of interest, e.g. a galectin is immobilized on an affinity support, e.g. NHS-activated Sepharose. Ligands from a mixture of biomolecules, such as serum, will non-covalently interact with the stationary phase and be retained. These can then be eluted with a competitive ligand (lactose, in the case of galectin) and be used for further analysis, e.g. MS for proteomics. Affinity chromatography has proven to be a highly suitable method to evaluate galectin specificity for natural ligands, since it selectively separates and collects the bound and unbound glycoproteins. In combination with a method like MS, a vast number of ligands can be identified, but equally important, it provides the possibility to study differences between binding and non-binding glycoforms in biological systems, as discussed later.

We earlier used affinity chromatography to identify endogenous galectin ligands in serum[218]. From this we could conclude that the multifaceted specificity of galectin-3 also is evident when it comes to recognition of serum glycoproteins. Galectin-3 binds approximately 50% of all glycoproteins in serum, whereas galectin-8 and especially galectin-1 binds a smaller fraction of a more restricted set. The identities of these imply that galectin-3 binds primarily *N*-linked glycans, as it binds all of the heavily N-glycosylated α-2-macroglobulin, but very little of the *O*-glycosylated IgA. Most likely, galectin-3 selects glycoproteins in serum based on at least one available LacNAc residue (free or 2–3-sialylated), probably on the third antennae of triantennary glycans, as suggested for galectin-3 binding of transferrin (paper V). As discussed earlier, a larger number and more branching of *N*-glycans were shown to increase galectin-3 binding to cell surface receptors[105], further supporting this theory. Generally, recognition of LacNAc seems to be required for galectin-3 binding of endogenous ligands, since the R186 mutant which has lost this affinity, does not bind any serum glycoproteins[218]. In contrast, the presence of terminal 2-3 sialic acids seem to be irrelevant for galectin-3 binding, as removal of sialylations generated only an additional 4% of galectin-3 binding transferrin (paper II). Together this suggests that

the multifaceted specificity of galectin-3 enables it to recognize most *N*-glycosylated glycoproteins, stipulating that they carry at least one triantennary *N*-glycan with an available LacNAc.

It is obvious that galectin-1 and galectin-3 differ significantly in their recognition of glycans, as demonstrated in their binding to serum glycoproteins. Both recognize poly(LacNAc), but while galectin-3 is capable of binding internal poly(LacNAc), galectin-1 only recognizes the terminal LacNAc. Substituting Val32 in site B of galectin-1 with an Ala (as found in galectin-3) opens the ability to bind the GlcNAcβ1-3 substituted galactosides found in internal poly(LacNAc) glycans. However, this mutation did not alter the binding to serum glycoproteins. Hence, most of the binding of galectin-1 to serum glycoproteins is not mediated by poly(LacNAc), which was as expected, since as discussed above, poly(LacNAc)s have not been reported among human serum *N*-glycans. This signifies that the binding of galectin-3 to its unique ligands, such as transferrin, must be due to other binding features. Sialylation could be an explanation, since galectin-3 tends to bind these structures better (galectin-3 but not galectin-1 binds the considerably sialylated glycoprotein fetuin). Still, similar to galectin-3, galectin-1 tolerates, but does not require terminal 2-3 sialylations for binding of serum glycoproteins (paper II), leaving the mechanisms of the different recognition of glycoproteins between the two galectins unanswered.

Despite its generally low affinity for most serum glycoproteins, galectin-1 binds almost all α-2-macroglobulin and approximately one-third of all serum IgM and haptoglobin. As pointed out above, an α-2-macroglobulin molecule should statistically display several promising galectin binding sites and hence also bind galectin-1. The same applies for IgM and haptoglobin which both tend to form higher-order multimers, giving rise to a substantial number of *N*-glycosylation sites. Glycan analysis of galectin-1 bound haptoglobin revealed that the major recognition site for galectin-1 is a triantennary *N*-glycan directed away from the suggested binding sites for haemoglobin. As the interaction with galectin-1 is mainly monovalent, interaction of one galectin-1 CRD with this haptoglobin *N*-glycan is probably sufficient for binding, leaving the other open for interaction with cellular glycoproteins involved in the function of haptoglobin.

The major discrepancy between serum ligands of galectin-1 and galectin-8 is that galectin-8 binds IgA instead of IgM. IgA carries mainly *O*-glycans and approximately one-third of these contain the galectin-8N favored NeuAcα2-3Galβ1-3GalNAc and an equal amount of IgA also binds galectin-8N. An additional one-third of the IgA *O*-glycans includes uncapped Galβ1-3GalNAc, not sufficient for galectin-8N binding, since the binding of galectin-8N to serum glycoproteins depends on the presence of 2–3-sialylated galactosides (paper III). Galectin-8 binds a significant but variable (unbiased of disease) fraction of haptoglobin, ranging from 20-50% depending on the source (pooled purified haptoglobin or haptoglobin analyzed directly from serum). Notably, this is a largely different fraction of haptoglobin glycoforms compared to galectin-1, suggesting that these two galectins recognize separate glycosylation features of this glycoprotein (paper IV). However, the *N*-glycans on galectin-8N binding haptoglobin does not appear to contain any identifiable characteristics that could account for this discrepancy. One plausible explanation for this would be that galectin-8 recognizes *O*-glycans, but nearly all studies have neglected or ruled out the existence of these on haptoglobin. Nevertheless, at least two recent studies have reported the presence of *O*-glycans on haptoglobin[266, 267] which might clarify this issue.

From the above it is apparent that galectins are able to bind a large and diverse set of glycoforms and it is highly conceivable that these interactions must have functional consequences. Even if these interactions must occur in serum itself, it is doubtful that these are physiologically relevant since the concentration of galectin is relatively low compared to its ligands (more than a thousand-fold, for either of the main bound glycoproteins). However, as serum glycoproteins encounter circulating immune cells and erythrocytes, or surrounding endothelial cells, all with considerable amounts of galectins, these interactions may well be functionally relevant. In fact, the estimated steady-state concentration of most serum proteins in tissue is as high as half of that in serum itself[268], providing suitable binding partners for the tissue abundant galectins. Indeed, several of the serum glycoproteins, such as haptoglobin, also have functions outside the circulation by interaction with tissue cells[269].

Biological function of the galectin fine specificity (paper I, II and V).

Given that galectins are present both inside and outside cells, their biological functions are a result of intracellular and/or extracellular effects. Extracellularly, galectins can bind to cell-surface and ECM glycoconjugates to modulate cell adhesion, proliferation or apoptosis (discussed repeatedly in the previous section). They are also capable of activating NADPH-oxidase in primed neutrophils, causing the production of reactive oxygen species (oxidative burst) critical for innate immunity[212, 270]. While the fine specificity appeared to be of minor importance for galectin-8 activation of neutrophils, the galectin-3 fine specificity seems decisive in this context. Both galectin-3 mutations in site D, R186S and R186I, were incapable of cell surface binding and activation of neutrophils (paper I). Similarly, another study showed that the R186S mutant failed to attract neutrophils as well as promote macrophage clearance of apoptotic cells[271], suggesting that galectin-3 requires its preference for LacNAc for its innate immune functions. The most potent neutrophil activators were mutants with the K176L substitution, suggesting that binding of GlcNAcβ in site B is decisive, while sialylations seemed to be of minor importance. Overall the ability to accommodate GlcNAcβ in site B correlates best with the ability to activate neutrophils, suggesting binding to cell surface poly(LacNAc) structures (paper I).

Several of the regulatory roles of galectins involve their role in guiding intracellular traffic of glycoproteins[9, 10, 76]. We have previously shown that the fine specificity of galectin-8 determines its intracellular trafficking after endocytosis[215] and galectin-3 recognition of LacNAc is required for apical sorting of glycoproteins[239]. Correspondingly, we found that binding to LacNAc was essential for galectin-3 entry into vesicles and that a more subtle modification in fine specificity could significantly alter the intracellular trafficking of galectin-3 (paper I).

Intracellular sorting of a number of glycoproteins is a critical process and disruption of this could have severe consequences, e.g. the I-cell disease[25]. Several of the serum glycoproteins recognized by galectins have important functions after uptake into

tissue cells, such as iron delivery by transferrin[272] or haemoglobin scavenging by haptoglobin[273]. Considering that transferrin is a highly glycosylated protein and its endocytosis has been extensively studied, it is surprising that no effort has been made to find a function of its glycans in this process. Here (paper V) we show that galectin-3 binding and non-binding glycoforms of transferrin are sorted to different endosomal compartments and with diverse kinetics. The galectin-3 binding transferrin utilizes a shorter and faster route of recycling and shows a distinct overlap with endogenous galectin-3, verifying an important role for this lectin in the sorting machinery. Indeed, in cells lacking galectin-3 no difference in intracellular localization could be observed. It has previously been shown that in the absence of endogenous galectin-3, exogenously added galectin-3, can "rescue" its function both *in vitro*[274] and *in vivo*[226]. Likewise, in our study galectin-3 deficient cells could acquire the ability for diverse intracellular sorting of these glycoforms when treated with exogenously added wild type galectin-3, but not with the R186S mutant. To our knowledge this is the first report of a functional role for transferrin glycans in intracellular trafficking after uptake.

A relatively large fraction of haptoglobin binds galectin-1 (30 % compared to the 5% of galectin-3 bound transferrin) and as discussed in the next section this goes up in cancer. In order to study a possible biological consequence of this interaction, we studied the uptake of haptoglobin into alternatively activated macrophages (M2 macrophages) in complex with haemoglobin (paper II). The major function of haptoglobin is to capture free haemoglobin (Hb) released in the circulation during haemolysis, resulting in clearance of potentially oxidative and toxic iron groups that haemoglobin contains. This complex can subsequently be taken up by M2 macrophages through haptoglobin binding to CD163[273]. We show that the interaction of haptoglobin with galectin-1 becomes decisive in this process, as galectin-1 binding appears to divert the complex from lysosomal degradation, and instead targets it to galectin-1 positive vesicles with a so far unknown function. To reveal the identity of these vesicles, further studies are needed, but it is liable to suspect that they carry out some sort of protein recycling. Nevertheless, it is clear that, under healthy conditions one-third of all haptoglobin involved in haemoglobin scavenging carry *N*-glycans recognized by galectin-1 and this recognition may determine its intracellular faith after uptake into tissue cells. Since galectin-1 is highly

abundant in macrophages, especially the M2 phenotype[251] and one of the main functions of these cells is to remove free haemoglobin, it is feasible that this interaction actually occurs *in vivo*.

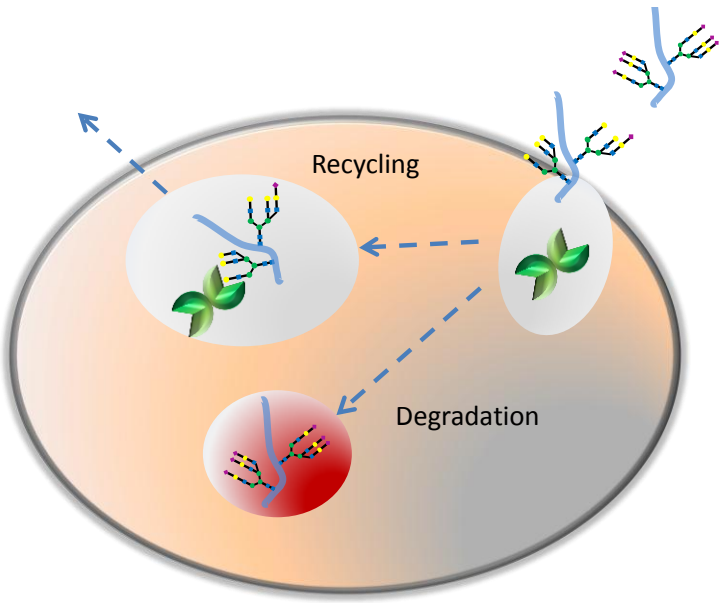

Figure 9. Schematic illustration of selective intracellular targeting of glycoproteins based on galectin interactions. Intracellular sorting of certain glycoproteins after uptake appears to be determined by galectin interactions. Endocytosed glycoproteins most likely enter galectin containing vesicles accumulated in approximity to the cell surface. Here they appear to be subjected to a galectin selection process in which glycoforms recognised by galectins are targeted to different intracellular compartments than unrecognized glycoforms.

Galectin fine specificity in relation to disease (paper II, III, IV).

There is ample evidence linking galectins to inflammatory and malignant conditions. Numerous of these are based on addition or removal of the galectin in question, resulting in a disease phenotype. Others are based on novel observations in "naturally" developed diseases, commonly signified by an increased expression of one or more galectins. Together this suggests that the increased amounts of galectins must

recognize and alter the function of aberrantly glycosylated glycoproteins important for the disease process. Indeed, it has been shown that glycosylation changes of specific glycoproteins are advantageous to the disease; as a result of altered galectin recognition (e.g. prolonged residence time of growth receptors). This may explain some of the diverse effects observed for galectins, but it does not provide a direct link to disease. Serum contains an enormous pool of glycoproteins, which are known to be altered in disease. A number of these disease associated changes may increase galectin binding (e.g. increased branching in cancer) or decrease binding (e.g. degalactosylated IgA in IgA nephritis). Based on the previous section these may provide the basis for a new type of functional biomarkers.

Galectin-1 has previously been proven to bind a small selective group of serum glycoproteins (α-2-macroglobulin, IgM and haptoglobin). This makes it highly suitable to detect disturbed glycosylation of any of these as a result of disease. The characteristic increased glycan branching seen in cancer may very well give rise to increased galectin-1 binding of these glycoproteins. Indeed, by subjecting sera to affinity chromatography using immobilized galectin-1 we found that serum from 25 metastatic breast cancer patients contained approximately twice as much galectin-1 binding serum glycoproteins compared with healthy individuals (paper II). This was mainly due to an increase of galectin-1 binding haptoglobin, which increased from on average 30% in healthy sera to about 50% in cancer sera, probably due to increased proportion of additional antennae with less terminal sialic acids. In contrast, galectin-1 binding to IgM actually decreased in cancer (from 37% in healthy to 28 % in cancer patients). Despite that further studies are needed to fully evaluate the potential of galectin-1 binding glycoforms as a cancer biomarker, it is at present clear that the method has the capacity to discriminate breast cancer patients from healthy controls. Individually, the galectin-1 binding concentrations of haptoglobin and IgM gave an AUC value of 0.90, and using the ratio of these improved the AUC to 0.95, which by far exceeds several of the currently used biomarkers for breast cancer. Certainly, this study compares late stage metastatic cancer patients with healthy controls which are at the far ends of the disease spectrum. For the limited number of patients studied here no correlation with treatment or disease severeness (i.e. number of metastatic lymph nodes) was found. Moreover, there is a trend that patients with larger primary tumours have more galectin-1 binding glycoproteins, suggesting that small primary

tumours at the early stages of cancer development may not alter the serum glycome significantly enough for detection with this method. To evaluate this, future studies need to include different stages of cancer development, especially early, and also patients with benign tumours.

Several pieces of evidence suggest that abnormally O-glycosylated IgA plays a key role in the pathogenesis of the relatively common disease IgA nephritis. This is probably due to a galactosyltransferase imbalance in the IgA producing B cells, leading to decreased formation of Galβ1-3GalNAc and the consequent lack of galactose in turn prevents further addition of NeuAcα2-3. This should in theory reduce the binding of IgA to galectin-8N. To address whether altered glycosylation of IgA would lead to an altered binding to galectin-8N, we subjected 100 sera from IgA nephritis patients, 92 sera from patients with other histological patterns of glomerulonephritis and 20 sera from healthy controls to affinity chromatography with immobilized galectin-8N (paper III). As expected we found that the IgA in patients with IgAN have reduced binding to galectin-8N. However, since IgA nephritis patient's sera contain significantly more total IgA, this is not a result of decreased amounts of galectin-8N binding IgA but rather an increase of galectin-8N non-binding IgA. This decrease in galectin-8N binding was also associated with disease severity; a low ratio of galectin-8N bound/unbound IgA leads to faster disease progress. Intriguingly, the galectin-8N binding fraction of serum glycoproteins other than IgA was significantly higher in the IgAN cases. These glycoproteins are derived from different sources than IgA, e.g. haptoglobin is mainly secreted from hepatocytes, suggesting that the disease have cell specific effects on the glycosylation machinery. Notably, this is not a general feature of chronic inflammation, since this increase in galectin-8N binding is not seen in other forms of glomerulonephritis. Hence, it appears as if a substantial proportion of plasma cells, but not all, overproduces galactose deficient IgA molecules in IgA nephritis and that this is a result of a cell specific alteration of glycosylation.

As discussed above, galectin-1 and galectin-8 recognize not only different glycoproteins, but also different glycoforms of the same glycoprotein. Consequently, it is feasible that this may be reflected in the detection of disease associated glycoforms. To evaluate this, a number of breast cancer sera from paper II were subjected to affinity chromatography with immobilized galectin-8N and a number of

IgA nephritis sera from paper III were subjected to affinity chromatography with immobilized galectin-1 (paper IV). We found that the levels of the different glycoforms recognized by galectin-1 and -8 change in opposite direction with disease. Most striking was that cancer appears to significantly decrease binding to galectin-8N, partially due to a reduced yield of haptoglobin, but also other glycoproteins must contribute. Using the ratio of the galectin-1 bound/galectin-8N bound glycoforms showed a remarkable correlation (increase) with cancer and displayed an AUC of 0.98, and hence equal to combined data from multiple (>6) biomarkers reported[275]. Moreover, this appears to be unbiased of inflammation since this ratio moves in the other direction, i.e. decrease, in IgA nephritis patients.

Concluding remarks

Because of their wide range of important biological activities, galectins are enigmatic in terms of identifying a central physiological function, if only one exists. Their role in intracellular guidance of glycoproteins shown here emerges as a feasible explanation for these multifaceted activities. Several human diseases are due to defects of intracellular trafficking[276], hence altered sorting of some of the major serum glycoproteins found here almost certainly will have a physiological impact. Even a seemingly trivial increase in galectin binding to one of these, e.g. 20 percentage points, renders a substantial amount of protein (0.1-1 mg/ml), far beyond that of the common cancer-associated and functionally relevant mucins (5-100ng/ml). Moreover, this effect will be boosted even further since increased numbers of galectin-1 ligands found in cancer sera will encounter increased numbers of galectin-1 in the tissue cells, since the latter is a feature of most cancers. Such increase in both ligand and receptor is most likely not circumstantial, but instead coordinated to benefit the tumour. At first glance, it is puzzling how these concerted actions could promote tumour progression, but evasion of lysosomal degradation by a considerable fraction of a pro-angiogenic protein, such as haptoglobin emerges as a conceivable hypothesis.

Acknowledgments

Min tid som doktorand har varit oerhört rolig och stimulerande. För detta vill jag tacka följande:

Min handledare **Hakon**, som är hjärnan bakom allt detta arbete. Stort tack för att du delat med dig av din enorma kunskap och dina fantastiska idéer! Du har inte bara gett mig möjligheten att få ta del av spännande forskning utan även släppt mig fri och låtit mig styra projekten, delvis på egen hand, vilket har varit otroligt stimulerande. Tack även för utförsåkning i San Francisco, åskoväder i Tonga room och fantastiska gåsmiddagar.

Barbro, labbets hjärta, utan vars engagemang och kunskap detta arbete inte hade blivit av. Tack för alla trevliga stunder och skratt på labbet och i fikarummet.

Emma, för delat kontor, författarskap och tankar. Vi har klarat av allt ifrån hopplösa FACS försök (tappade rör), feltryckta avhandlingar, 3000 meter hinder på Charles de Gaulle, konstiga mutanter, och obegripliga kolhydratsstrukturer.

Adriana, thanks for macrophages, wigs and inspiration. But especially all discussions and laughs we had throughout the years.

Susanne och **Cecilia**, tidigare medlemmar av gruppen som har lagt grunden för detta arbete. Tack för all hjälp under mina första år som doktorand.

Viveka, **Lotta**, och **Annelie**, som gjort examensarbeten och sommarprojekt i gruppen och bidragit till arbeten och god stämning.

All international students that have joined the group for a short period of time, especially **Franck** (we will beat you in Kiev this summer) and **Edel** (want some coffee?) for both laughs and science.

Ett stort tack till **Ulf,** för kemi, trevligt samarbete och utflykter!

Tack till alla övriga på MIG och i korridorerna på andra våningen för hjälp och inspiration. Ett speciellt tack till **Rickard** för alla ”kalla” stunder i kylrummet. Lovar att komma och besöka dig så ofta jag kan (om jag får låna din mössa).

Manfred and **Crina**, for your great contributions to the papers!

Johan och **Ola**, för er värdefulla hjälp med MS analyser av mina prover.

Per, för serum och transferrinanalyser.

Ett stort tack till alla övriga medförfattare, utan ert bidrag hade inget av detta varit möjligt.

Mina tre amigos för alla dessa år, utan er hade jag definitivt inte varit här idag:

Viktor ”el rocho” Walldén som sedan vår debut på vita duken 86 har varit en pålitlig vän. Jag kommer aldrig glömma när vi jagade nyckeln till FilmTajm eller när du fick kramp i alperna.
Erik ”el migraino” Oleskog för alla krigs”lekar”, Bondfilmer, härliga samtal, skratt (tur att vi förstår vår fenomenala humor, tror nämligen ingen annan gör det) och fantastisk vänskap.
Anders ”el indefinido” Roos för allt vi upplevt tillsammans, från att måla soffan med tandkräm som 3 åringar, första skoldagen, Lindau och Oxford, skidresor, flytt till Lund och ”vandrarhemmet tåget”, grillfester på MH, jag har haft tur som haft dig med på resan.

Rikard ”Lex Luthor” Hjelm, **Jones** ”Tight T” Sköld, **Ola** ”Lidl” Morin, **Håkan** ”Big Daddy” Erlandsson, **Mårten** ”Ghetto Gospel” Hansson, **Olaf** ”Tysken” Wulf och sist

men absolut inte minst **Cliff**, för att varje morgon klockan 06:00 under flera år visat att råstyrka är viktigare än intelligens. Ni har bidragit stort till att jag har haft fullt med energi för att varje dag åka till Lund (och ibland till en hangar i Dalby) för att få ihop innehållet till denna bok.

Sasha, för all cola, skinka, grillfester och playstation. Jag hoppas att vi får mer tid till allt detta nu i sommar. Du är en fantastisk ”styvbror” men en ännu bättre vän!

Marina, din glädje och energi smittar av sig. Tack för alla fantastiska fester, konstiga fötter, och så naturligtvis för valborgsoppan!

Milan och **Mira**, för att ni tagit hand om mig och alltid ställt upp och hjälp till när det behövts. Jag hade tur som träffade Helena, men haft lika mycket tur som har fått lära känna hennes underbara föräldrar!

Mina syskon, **Lena** och **Björn**, för att ni ALLTID ställt upp, inspirerat och hjälpt till från att jag var liten tuss till stor klump.

Även ett stort tack till resten av familjen: **Felix**, **Alexander, Gabriel, Eva** och **Jan** för alla roliga dagar och kvällar när jag är på besök i Göteborg. Ni får komma hit och hälsa på snart!

Min **mormor** som tagit hand om mig som liten, alltid funnits där och alltid ställt upp. Jag har haft en oerhörd tur som haft dig.

Mina fantastiska föräldrar för en underbar uppväxt. Ni har alltid trott på mig och låtit mig fatta mina egna beslut. **Mamma**: du har alltid ställt upp, stöttat och tagit hand om mig, och för detta är jag är evigt tacksam. **Pappa**: min stora idol. Du är den mest omtänksamma person jag någonsin träffat. Du kommer alltid att finnas med mig.

Slutligen vill jag tacka min egna lilla doktor **Helena**. Hur ska jag kunna beskriva hur mycket du betyder för mig, denna bok måste sluta någon gång? Allt jag kan säga är att jag älskar dig, men det är bara en viskning av vad jag verkligen känner.

References

1. Stone, W.E. THE USE OF THE TERM "CARBOHYDRATES.". *Science* **ns-21**, 149-150 (1893).
2. Bier, D.M. et al. Report of the IDECG Working Group on lower and upper limits of carbohydrate and fat intake. International Dietary Energy Consultative Group. *Eur J Clin Nutr* **53 Suppl 1**, S177-8 (1999).
3. Varki, A. Essentials of glycobiology (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2009).
4. Rademacher, T.W., Parekh, R.B. & Dwek, R.A. Glycobiology. *Annual Review of Biochemistry* **57**, 785-838 (1988).
5. Feizi, T. & Chai, W. Oligosaccharide microarrays to decipher the glyco code. *Nat Rev Mol Cell Biol* **5**, 582-588 (2004).
6. Parekh, R.B. et al. Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316**, 452-7 (1985).
7. Suzuki, H. et al. IgA1-secreting cell lines from patients with IgA nephropathy produce aberrantly glycosylated IgA1. *J Clin Invest* **118**, 629-39 (2008).
8. Dube, D.H. & Bertozzi, C.R. Glycans in cancer and inflammation potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **4**, 477-488 (2005).
9. Partridge, E.A. et al. Regulation of cytokine receptors by Golgi N-glycan processing and endocytosis. *Science* **306**, 120-4 (2004).
10. Delacour, D., Koch, A. & Jacob, R. The Role of Galectins in Protein Trafficking. *Traffic* **10**, 1405-1413 (2009).
11. Liu, F.-T. & Rabinovich, G.A. Galectins as modulators of tumour progression. *Nat Rev Cancer* **5**, 29-41 (2005).
12. Karl-Anders, K. Microbial recognition of target-cell glycoconjugates. *Current Opinion in Structural Biology* **5**, 622-635 (1995).
13. Stowell, S.R. et al. Innate immune lectins kill bacteria expressing blood group antigen. *Nat Med* **16**, 295-301 (2010).
14. Paz, I. et al. Galectin-3, a marker for vacuole lysis by invasive pathogens. *Cellular Microbiology* **12**, 530-544 (2010).
15. Rambourg, A. & Leblond, C.P. ELECTRON MICROSCOPE OBSERVATIONS ON THE CARBOHYDRATE-RICH CELL COAT PRESENT AT THE SURFACE OF CELLS IN THE RAT. *The Journal of Cell Biology* **32**, 27-53 (1967).

16. Varki, A. Nothing in Glycobiology Makes Sense, except in the Light of Evolution. *Cell* **126**, 841-845 (2006).
17. Gagneux, P. & Varki, A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* **9**, 747-755 (1999).
18. Carroll, L. Through the Looking-Glass (pp. 224. Macmillan & Co.: London, 1872).
19. Van Valen, L. *A New Evolutionary Law, Evolutionary Theory 1*, 1-30 (1973).
20. Varki, A. & Angata, T. Siglecs—the major subfamily of I-type lectins. *Glycobiology* **16**, 1R-27R (2006).
21. Lindesmith, L. et al. Human susceptibility and resistance to Norwalk virus infection. *Nat Med* **9**, 548-553 (2003).
22. Haverkorn, M.J. & Goslings, W.R. Streptococci, ABO blood groups, and secretor status. *Am J Hum Genet* **21**, 360-75 (1969).
23. Bertozzi, C.R. & Kiessling, L.L. Chemical glycobiology. *Science* **291**, 2357-64 (2001).
24. Freeze, H.H. Genetic defects in the human glycome. *Nat Rev Genet* **7**, 537-551 (2006).
25. Reitman, M.L., Varki, A. & Kornfeld, S. Fibroblasts from patients with I-cell disease and pseudo-Hurler polydystrophy are deficient in uridine 5'-diphosphate-N-acetylglucosamine: glycoprotein N-acetylglucosaminylphosphotransferase activity. *The Journal of Clinical Investigation* **67**, 1574-1579 (1981).
26. Spiro, R.G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R (2002).
27. Spiro, R.G. Glycoproteins. *Adv Protein Chem* **27**, 349-467 (1973).
28. Hart, G.W. & Copeland, R.J. Glycomics Hits the Big Time. *Cell* **143**, 672-676 (2010).
29. Lowe, J.B. & Marth, J.D. A GENETIC APPROACH TO MAMMALIAN GLYCAN FUNCTION. *Annual Review of Biochemistry* **72**, 643-691 (2003).
30. Opat, A.S., van Vliet, C. & Gleeson, P.A. Trafficking and localisation of resident Golgi glycosylation enzymes. *Biochimie* **83**, 763-773 (2001).
31. Ten Hagen, K.G., Fritz, T.A. & Tabak, L.A. All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology* **13**, 1R-16R (2003).
32. Grabenhorst, E., Nimtz, M., Costa, J. & Conradt, H.S. In Vivo Specificity of Human α1,3/4-Fucosyltransferases III-VII in the Biosynthesis of LewisX and Sialyl LewisX Motifs on Complex-type N-Glycans. *Journal of Biological Chemistry* **273**, 30985-30994 (1998).
33. Alper, J. Sugar Separates Humans From Apes. *Science* **291**, 2340 (2001).
34. Laine, R.A. Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 × 1012 structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* **4**, 759-767 (1994).
35. Clamp, M. et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences* **104**, 19428-19433 (2007).

36. Cummings, R.D. The repertoire of glycan determinants in the human glycome. *Molecular BioSystems* **5**, 1087-1104 (2009).
37. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561-563 (1970).
38. Shapiro, J.A. Revisiting the Central Dogma in the 21st Century. *Annals of the New York Academy of Sciences* **1178**, 6-28 (2009).
39. Nagai, Y. Glycobiology in the 21ST Century: Coming developments in glycobiology. *Glycoconjugate Journal* **19**, 161-163 (2002).
40. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1473**, 4-8 (1999).
41. Arnold, J.N., Wormald, M.R., Sim, R.B., Rudd, P.M. & Dwek, R.A. The Impact of Glycosylation on the Biological Function and Structure of Human Immunoglobulins. *Annual Review of Immunology* **25**, 21-50 (2007).
42. Ohtsubo, K. & Marth, J.D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855-867 (2006).
43. Fiete, D., Srivastava, V., Hindsgaul, O. & Baenziger, J.U. A hepatic reticuloendothelial cell receptor specific for SO4-4GalNAc^{2}1, 4GlcNAc^{2}1,2Man± that mediates rapid clearance of lutropin. *Cell* **67**, 1103-1110 (1991).
44. Götting, C., Kuhn, J. & Kleesiek, K. Human xylosyltransferases in health and disease. *Cellular and Molecular Life Sciences* **64**, 1498-1517 (2007).
45. Harris, R.J. & Spellman, M.W. O-Linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **3**, 219-224 (1993).
46. Sasamura, T. et al. neurotic, a novel maternal neurogenic gene, encodes an O-fucosyltransferase that is essential for Notch-Delta interactions. *Development* **130**, 4785-4795 (2003).
47. Torres, C.R. & Hart, G.W. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *Journal of Biological Chemistry* **259**, 3308-3317 (1984).
48. Hart, G.W., Housley, M.P. & Slawson, C. Cycling of O-linked [beta]-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-1022 (2007).
49. Hart, G.W., Haltiwanger, R.S., Holt, G.D. & Kelly, W.G. Glycosylation in The Nucleus and Cytoplasm. *Annual Review of Biochemistry* **58**, 841-874 (1989).
50. Tian, E. & Ten Hagen, K.G. Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconj J* **26**, 325-34 (2009).
51. Hashimoto, K. et al. Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans. *Carbohydrate Research* **344**, 881-887 (2009).
52. Gill, D.J., Clausen, H. & Bard, F. Location, location, location: new insights into O-GalNAc protein glycosylation. *Trends in Cell Biology* **21**, 149-158 (2011).
53. Hennet, T., Hagen, F.K., Tabak, L.A. & Marth, J.D. T-cell-specific deletion of a polypeptide N-acetylgalactosaminyl-transferase gene by site-directed recombination. *Proceedings of the National Academy of Sciences* **92**, 12070-12074 (1995).

54. Ju, T. & Cummings, R.D. Protein glycosylation: Chaperone mutation in Tn syndrome. *Nature* **437**, 1252-1252 (2005).
55. Florman, H.M. & Wassarman, P.M. O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. *Cell* **41**, 313-324 (1985).
56. Yousefi, S. et al. Increased UDP-GlcNAc:Gal beta 1-3GaLNAc-R (GlcNAc to GaLNAc) beta-1, 6-N-acetylglucosaminyltransferase activity in metastatic murine tumor cell lines. Control of polylactosamine synthesis. *J Biol Chem* **266**, 1772-82 (1991).
57. Tsuboi, S. et al. A novel strategy for evasion of NK cell immunity by tumours expressing core2 O-glycans. *EMBO J* **30**, 3173-3185 (2011).
58. Lowe, J.B. Glycosylation in the control of selectin counter-receptor structure and function. *Immunological Reviews* **186**, 19-36 (2002).
59. Tsuboi, S. & Fukuda, M. Roles of O-linked oligosaccharides in immune responses. *BioEssays* **23**, 46-53 (2001).
60. Xia, L. et al. Defective angiogenesis and fatal embryonic hemorrhage in mice lacking core 1–derived O-glycans. *The Journal of Cell Biology* **164**, 451-459 (2004).
61. Mattu, T.S. et al. The Glycosylation and Structure of Human Serum IgA1, Fab, and Fc Regions and the Role of N-Glycosylation on Fcα Receptor Interactions. *Journal of Biological Chemistry* **273**, 2260-2272 (1998).
62. Suzuki, H. et al. Aberrantly glycosylated IgA1 in IgA nephropathy patients is recognized by IgG antibodies with restricted heterogeneity. *The Journal of Clinical Investigation* **119**, 1668-1677 (2009).
63. Thornton, D.J., Rousseau, K. & McGuckin, M.A. Structure and Function of the Polymeric Mucins in Airways Mucus. *Annual Review of Physiology* **70**, 459-486 (2008).
64. Azim, M. & Surani, H. Glycoprotein synthesis and inhibition of glycosylation by tunicamycin in preimplantation mouse embryos: Compaction and trophoblast adhesion. *Cell* **18**, 217-227 (1979).
65. Kornfeld, R. & Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annual Review of Biochemistry* **54**, 631-664 (1985).
66. Chavan, M. & Lennarz, W. The molecular basis of coupling of translocation and N-glycosylation. *Trends in Biochemical Sciences* **31**, 17-20 (2006).
67. Helenius, A. How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol Biol Cell* **5**, 253-65 (1994).
68. Ioffe, E. & Stanley, P. Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. *Proc Natl Acad Sci U S A* **91**, 728-32 (1994).
69. Annette, H. Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1473**, 96-107 (1999).
70. Schachter, H. The 'yellow brick road' to branched complex N-glycans. *Glycobiology* **1**, 453-61 (1991).
71. Taguchi, T. et al. Activity of UDP-GlcNAc:GlcNAcβ1→6(GlcNAcβ1→2) Manα1→R[GlcNAc to Man]β1→4N- Acetylglucosaminyltransferase VI (GnT VI)

from the Ovaries of Oryzias latipes(Medaka Fish). *Biochemical and Biophysical Research Communications* **230**, 533-536 (1997).
72. Taguchi, T. et al. Purification and Characterization of UDP-GlcNAc: GlcNAcβ1–6(GlcNAcβ1–2)Manα1-R [GlcNAc to Man]-β1, 4-N-acetylglucosaminyltransferase VI from Hen Oviduct. *Journal of Biological Chemistry* **275**, 32598-32602 (2000).
73. Inamori, K.-i. et al. Molecular Cloning and Characterization of Human GnT-IX, a Novel β1,6-N-Acetylglucosaminyltransferase That Is Specifically Expressed in the Brain. *Journal of Biological Chemistry* **278**, 43102-43109 (2003).
74. Kaneko, M. et al. A novel β(1,6)-N-acetylglucosaminyltransferase V (GnT-VB). *FEBS Letters* **554**, 515-519 (2003).
75. Green, R.S. et al. Mammalian N-Glycan Branching Protects against Innate Immune Self-Recognition and Inflammation in Autoimmune Disease Pathogenesis. *Immunity* **27**, 308-320 (2007).
76. Ohtsubo, K. et al. Dietary and Genetic Control of Glucose Transporter 2 Glycosylation Promotes Insulin Secretion in Suppressing Diabetes. *Cell* **123**, 1307-1321 (2005).
77. Buckhaults, P., Chen, L., Fregien, N. & Pierce, M. Transcriptional Regulation ofN-Acetylglucosaminyltransferase V by the srcOncogene. *Journal of Biological Chemistry* **272**, 19575-19581 (1997).
78. Ma, B., Simala-Grant, J.L. & Taylor, D.E. Fucosylation in prokaryotes and eukaryotes. *Glycobiology* **16**, 158R-184R (2006).
79. Sasaki, K. et al. Expression cloning of cDNA encoding a human beta-1,3-N-acetylglucosaminyltransferase that is essential for poly-N-acetyllactosamine synthesis. *Proc Natl Acad Sci U S A* **94**, 14294-9 (1997).
80. Granovsky, M. et al. Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nat Med* **6**, 306-12 (2000).
81. Dennis, J.W., Laferte, S., Waghorne, C., Breitman, M.L. & Kerbel, R.S. Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science* **236**, 582-5 (1987).
82. Schwarzkopf, M. et al. Sialylation is essential for early development in mice. *Proceedings of the National Academy of Sciences* **99**, 5267-5270 (2002).
83. Brinkman-Van der Linden, E.C.M. et al. Loss of N-Glycolylneuraminic Acid in Human Evolution. *Journal of Biological Chemistry* **275**, 8633-8640 (2000).
84. Roland, S. Sialic acids as regulators of molecular and cellular interactions. *Current Opinion in Structural Biology* **19**, 507-514 (2009).
85. Varki, A. Sialic acids in human health and disease. *Trends Mol Med* **14**, 351-60 (2008).
86. Angata, K. et al. Sialyltransferase ST8Sia-II Assembles a Subset of Polysialic Acid That Directs Hippocampal Axonal Targeting and Promotes Fear Behavior. *Journal of Biological Chemistry* **279**, 32603-32613 (2004).
87. Toscano, M.A. et al. Differential glycosylation of TH1, TH2 and TH-17 effector cells selectively regulates susceptibility to cell death. *Nat Immunol* **8**, 825-834 (2007).

88. Kaneko, Y., Nimmerjahn, F. & Ravetch, J.V. Anti-Inflammatory Activity of Immunoglobulin G Resulting from Fc Sialylation. *Science* **313**, 670-673 (2006).
89. Zhuo, Y. & Bellis, S.L. Emerging Role of α2,6-Sialic Acid as a Negative Regulator of Galectin Binding and Function. *Journal of Biological Chemistry* **286**, 5935-5941 (2011).
90. Varki, A. Sialic acids as ligands in recognition phenomena. *The FASEB Journal* **11**, 248-255 (1997).
91. Shinya, K. et al. Avian flu: Influenza virus receptors in the human airway. *Nature* **440**, 435-436 (2006).
92. Hedlund, M., Ng, E., Varki, A. & Varki, N.M. α2-6–Linked Sialic Acids on N-Glycans Modulate Carcinoma Differentiation In vivo. *Cancer Research* **68**, 388-394 (2008).
93. Babál, P., Janega, P., Černá, A., Kholová, I. & Brabencová, E. Neoplastic transformation of the thyroid gland is accompanied by changes in cellular sialylation. *Acta Histochemica* **108**, 133-140 (2006).
94. Romppanen, J. et al. Serum Sialic Acid as a Marker of Alcohol Consumption: Effect of Liver Disease and Heavy Drinking. *Alcoholism: Clinical and Experimental Research* **26**, 1234-1238 (2002).
95. Freeze, H.H. & Aebi, M. Altered glycan structures: the molecular basis of congenital disorders of glycosylation. *Current Opinion in Structural Biology* **15**, 490-498 (2005).
96. Parekh, R.B., Tse, A.G., Dwek, R.A., Williams, A.F. & Rademacher, T.W. Tissue-specific N-glycosylation, site-specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. *EMBO J* **6**, 1233-44 (1987).
97. Biancone, L., Araki, M., Araki, K., Vassalli, P. & Stamenkovic, I. Redirection of tumor metastasis by expression of E-selectin in vivo. *J Exp Med* **183**, 581-7 (1996).
98. Meezan, E., Wu, H.C., Black, P.H. & Robbins, P.W. Comparative studies on the carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. II. Separation of glycoproteins and glycopeptides by sephadex chromatography. *Biochemistry* **8**, 2518-24 (1969).
99. Shriver, Z., Raguram, S. & Sasisekharan, R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat Rev Drug Discov* **3**, 863-873 (2004).
100. Sevinc, A., Adli, M., Kalender, M.E. & Camci, C. Benign causes of increased serum CA-125 concentration. *The Lancet Oncology* **8**, 1054-1055 (2007).
101. Knežević, A. et al. Variability, Heritability and Environmental Determinants of Human Plasma N-Glycome. *Journal of Proteome Research* **8**, 694-701 (2008).
102. Demetriou, M., Nabi, I.R., Coppolino, M., Dedhar, S. & Dennis, J.W. Reduced contact-inhibition and substratum adhesion in epithelial cells expressing GlcNAc-transferase V. *J Cell Biol* **130**, 383-92 (1995).
103. Dennis, J.W., Granovsky, M. & Warren, C.E. Glycoprotein glycosylation and cancer progression. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1473**, 21-34 (1999).
104. Lau, K.S. & Dennis, J.W. N-Glycans in cancer progression. *Glycobiology* **18**, 750-760 (2008).

105. Lau, K.S. et al. Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation. *Cell* **129**, 123-134 (2007).
106. Carlsson, M.C. et al. Galectin-1-binding glycoforms of haptoglobin with altered intracellular trafficking, and increase in metastatic breast cancer patients. *PLoS One* **6**, e26560 (2011).
107. Takamatsu, S. et al. Unusually High Expression of N-Acetylglucosaminyltransferase-IVa in Human Choriocarcinoma Cell Lines: A Possible Enzymatic Basis of the Formation of Abnormal Biantennary Sugar Chain. *Cancer Research* **59**, 3949-3953 (1999).
108. Nakano, M. et al. Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: A novel approach for the development of tumor markers. *International Journal of Cancer* **122**, 2301-2309 (2008).
109. Arnold, J.N., Saldova, R., Hamid, U.M. & Rudd, P.M. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **8**, 3284-93 (2008).
110. Sewell, R. et al. The ST6GalNAc-I Sialyltransferase Localizes throughout the Golgi and Is Responsible for the Synthesis of the Tumor-associated Sialyl-Tn O-Glycan in Human Breast Cancer. *Journal of Biological Chemistry* **281**, 3586-3594 (2006).
111. Pinho, S. et al. Biological significance of cancer-associated sialyl-Tn antigen: Modulation of malignant phenotype in gastric carcinoma cells. *Cancer Letters* **249**, 157-170 (2007).
112. Ju, T. et al. Human Tumor Antigens Tn and Sialyl Tn Arise from Mutations in Cosmc. *Cancer Research* **68**, 1636-1646 (2008).
113. Springer, G. T and Tn, general carcinoma autoantigens. *Science* **224**, 1198-1206 (1984).
114. Leivonen, M., Nordling, S., Lundin, J., von Boguslawski, K. & Haglund, C. STn and Prognosis in Breast Cancer. *Oncology* **61**, 299-305 (2001).
115. Wandall, H.H. et al. Cancer Biomarkers Defined by Autoantibody Signatures to Aberrant O-Glycopeptide Epitopes. *Cancer Research* **70**, 1306-1313 (2010).
116. Tarp, M.A. et al. Identification of a novel cancer-specific immunodominant glycopeptide epitope in the MUC1 tandem repeat. *Glycobiology* **17**, 197-209 (2007).
117. Itzkowitz, S.H. et al. Lewisx- and Sialylated Lewisx-related Antigen Expression in Human Malignant and Nonmalignant Colonic Tissues. *Cancer Research* **46**, 2627-2632 (1986).
118. Hoff, S.D. et al. Increased Expression of Sialyl-Dimeric Lex Antigen in Liver Metastases of Human Colorectal Carcinoma. *Cancer Research* **49**, 6883-6888 (1989).
119. Reis, C.A., Osorio, H., Silva, L., Gomes, C. & David, L. Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol* **63**, 322-329 (2010).
120. Park, S.-Y. et al. N-glycosylation status of β-haptoglobin in sera of patients with colon cancer, chronic inflammatory diseases and normal subjects. *International Journal of Cancer* **126**, 142-155 (2010).
121. Rosen, S.D. Ligands for L-Selectin: Homing, Inflammation, and Beyond. *Annual Review of Immunology* **22**, 129-156 (2004).

122. Somers, W.S., Tang, J., Shaw, G.D. & Camphausen, R.T. Insights into the Molecular Basis of Leukocyte Tethering and Rolling Revealed by Structures of P- and E-Selectin Bound to SLeX and PSGL-1. *Cell* **103**, 467-479 (2000).
123. Marth, J.D. & Grewal, P.K. Mammalian glycosylation in immunity. *Nat Rev Immunol* **8**, 874-887 (2008).
124. Frommhold, D. et al. Sialyltransferase ST3Gal-IV controls CXCR2-mediated firm leukocyte arrest during inflammation. *J Exp Med* **205**, 1435-1446 (2008).
125. Gabay, C. & Kushner, I. Acute-Phase Proteins and Other Systemic Responses to Inflammation. *New England Journal of Medicine* **340**, 448-454 (1999).
126. Gornik, O. & Lauc, G. Glycosylation of serum proteins in inflammatory diseases. *Dis Markers* **25**, 267-78 (2008).
127. Flahaut, C., Michalski, J.C., Danel, T., Humbert, M.H. & Klein, A. The effects of ethanol on the glycosylation of human transferrin. *Glycobiology* **13**, 191-198 (2003).
128. Saroha, A., Biswas, S., Chatterjee, B.P. & Das, H.R. Altered glycosylation and expression of plasma alpha-1-acid glycoprotein and haptoglobin in rheumatoid arthritis. *Journal of Chromatography B* **879**, 1839-1843 (2011).
129. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer Statistics, 2010. *CA: A Cancer Journal for Clinicians* **60**, 277-300 (2010).
130. Cook, N.R. Statistical Evaluation of Prognostic versus Diagnostic Models: Beyond the ROC Curve. *Clinical Chemistry* **54**, 17-23 (2008).
131. Hanash, S.M., Pitteri, S.J. & Faca, V.M. Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571-579 (2008).
132. Jesneck, J. et al. Do serum biomarkers really measure breast cancer? *BMC Cancer* **9**, 164 (2009).
133. Drake, P.M. et al. Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation. *Clinical Chemistry* **56**, 223-236 (2010).
134. Zielinska, D.F., Gnad, F., Wiśniewski, J.R. & Mann, M. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **141**, 897-907 (2010).
135. van den Brûle, F., Califice, S. & Castronovo, V. Expression of galectins in cancer: A critical review. *Glycoconjugate Journal* **19**, 537-542 (2002).
136. Sano, H. et al. Human Galectin-3 Is a Novel Chemoattractant for Monocytes and Macrophages. *The Journal of Immunology* **165**, 2156-2164 (2000).
137. Fuster, M.M. & Esko, J.D. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat Rev Cancer* **5**, 526-542 (2005).
138. Thijssen, V.L.J.L. et al. Galectin-1 is essential in tumor angiogenesis and is a target for antiangiogenesis therapy. *Proceedings of the National Academy of Sciences* **103**, 15975-15980 (2006).
139. Nangia-Makker, P. et al. Galectin-3 Induces Endothelial Cell Morphogenesis and Angiogenesis. *The American Journal of Pathology* **156**, 899-909 (2000).
140. Abd Hamid, U.M. et al. A strategy to reveal potential glycan markers from serum glycoproteins associated with breast cancer progression. *Glycobiology* **18**, 1105-1118 (2008).

141. Marino, K., Bones, J., Kattla, J.J. & Rudd, P.M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* **6**, 713-723 (2010).
142. Rillahan, C.D. & Paulson, J.C. Glycan Microarrays for Decoding the Glycome. *Annual Review of Biochemistry* **80**, 797-823 (2011).
143. Gabius, H.-J., André, S., Jiménez-Barbero, J., Romero, A. & Solís, D. From lectin structure to functional glycomics: principles of the sugar code. *Trends in Biochemical Sciences* **36**, 298-313 (2011).
144. Zhao, J., Simeone, D.M., Heidt, D., Anderson, M.A. & Lubman, D.M. Comparative Serum Glycoproteomics Using Lectin Selected Sialic Acid Glycoproteins with Mass Spectrometric Analysis: Application to Pancreatic Cancer Serum. *Journal of Proteome Research* **5**, 1792-1802 (2006).
145. Sjur, O. The history of ricin, abrin and related toxins. *Toxicon* **44**, 361-370 (2004).
146. Clark, G.F. et al. Tumor Biomarker Glycoproteins in the Seminal Plasma of Healthy Human Males Are Endogenous Ligands for DC-SIGN. *Molecular & Cellular Proteomics* **11** (2012).
147. Malhotra, R., Ward, M., Sim, R.B. & Bird, M.I. Identification of human complement Factor H as a ligand for L-selectin. *Biochem J* **341 (Pt 1)**, 61-9 (1999).
148. Bhavanandan, V.P., Umemoto, J. & Davidson, E.A. Characterization of an endo-α-N-acetyl galactosaminidase from Diplococcus pneumoniae. *Biochemical and Biophysical Research Communications* **70**, 738-745 (1976).
149. Wada, Y. et al. Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **17**, 411-422 (2007).
150. Chu, C.S. et al. Profile of native N-linked glycan structures from human serum using high performance liquid chromatography on a microfluidic chip and time-of-flight mass spectrometry. *Proteomics* **9**, 1939-1951 (2009).
151. Malmstrom, J. et al. Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460**, 762-765 (2009).
152. Hollingsworth, M.A. & Swanson, B.J. Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer* **4**, 45-60 (2004).
153. Moore, A., Medarova, Z., Potthast, A. & Dai, G. In Vivo Targeting of Underglycosylated MUC-1 Tumor Antigen Using a Multimodal Imaging Probe. *Cancer Research* **64**, 1821-1827 (2004).
154. Pedersen, J.W. et al. Seromic profiling of colorectal cancer patients with novel glycopeptide microarray. *International Journal of Cancer* **128**, 1860-1871 (2011).
155. Kam, R. & Poon, T. The Potentials of Glycomics in Biomarker Discovery. *Clinical Proteomics* **4**, 67-79 (2008).
156. Okuyama, N. et al. Fucosylated haptoglobin is a novel marker for pancreatic cancer: A detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *International Journal of Cancer* **118**, 2803-2808 (2006).
157. Fujimura, T. et al. Glycosylation status of haptoglobin in sera of patients with prostate cancer vs. benign prostate disease or normal subjects. *International Journal of Cancer* **122**, 39-49 (2008).

158. Arnold, J.N. et al. Novel Glycan Biomarkers for the Detection of Lung Cancer. *Journal of Proteome Research* **10**, 1755-1764 (2011).
159. Ang, I.L. et al. Study of Serum Haptoglobin and Its Glycoforms in the Diagnosis of Hepatocellular Carcinoma: A Glycoproteomic Approach. *Journal of Proteome Research* **5**, 2691-2700 (2006).
160. Zhao, C. et al. Circulating haptoglobin is an independent prognostic factor in the sera of patients with epithelial ovarian cancer. *Neoplasia* **9**, 1-7 (2007).
161. Carlsson, M.C. et al. Different fractions of human serum glycoproteins bind galectin-1 or galectin-8, and their ratio may provide a refined biomarker for pathophysiological conditions in cancer and inflammatory disease. *Biochimica et Biophysica Acta (BBA) - General Subjects* (2012).
162. Lilja, H., Ulmert, D. & Vickers, A.J. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat Rev Cancer* **8**, 268-278 (2008).
163. Adamczyk, B., Tharmalingam, T. & Rudd, P.M. Glycans as cancer biomarkers. *Biochimica et Biophysica Acta (BBA) - General Subjects*.
164. Saldova, R., Reuben, J.M., Abd Hamid, U.M., Rudd, P.M. & Cristofanilli, M. Levels of specific serum N-glycans identify breast cancer patients with higher circulating tumor cell counts. *Ann Oncol* **22**, 1113-9 (2011).
165. Barondes, S.H. Bifunctional properties of lectins: lectins redefined. *Trends in Biochemical Sciences* **13**, 480-482 (1988).
166. Tamara, M. Drug targeting to the colon with lectins and neoglycoconjugates. *Advanced Drug Delivery Reviews* **56**, 491-509 (2004).
167. Lee, R.T. & Lee, Y.C. Affinity enhancement by multivalent lectin–carbohydrate interaction. *Glycoconjugate Journal* **17**, 543-551 (2000).
168. Salomonsson, E. et al. Monovalent Interactions of Galectin-1. *Biochemistry* **49**, 9518-9532 (2010).
169. Sharon, N. Lectins: past, present and future. *Biochem Soc Trans* **36**, 1457-60 (2008).
170. Matrosovich, M.N. et al. Avian Influenza A Viruses Differ from Human Viruses by Recognition of Sialyloligosaccharides and Gangliosides and by a Higher Conservation of the HA Receptor-Binding Site. *Virology* **233**, 224-234 (1997).
171. Zelensky, A.N. & Gready, J.E. The C-type lectin-like domain superfamily. *FEBS Journal* **272**, 6179-6217 (2005).
172. Dahms, N.M. & Hancock, M.K. P-type lectins. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1572**, 317-340 (2002).
173. Barondes, S.H. et al. Galectins: A family of animal β-galactoside-binding lectins. *Cell* **76**, 597-598 (1994).
174. Remy, L. Principles of structures of animal and plant lectins. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1572**, 198-208 (2002).
175. Powell, L.D. & Varki, A. I-type Lectins. *Journal of Biological Chemistry* **270**, 14243-14246 (1995).
176. Cash, H.L., Whitham, C.V., Behrendt, C.L. & Hooper, L.V. Symbiotic Bacteria Direct Expression of an Intestinal Bactericidal Lectin. *Science* **313**, 1126-1130 (2006).

177. Pepys, M.B. & Hirschfield, G.M. C-reactive protein: a critical update. *The Journal of Clinical Investigation* **111**, 1805-1812 (2003).
178. Dodd, R.B. & Drickamer, K. Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology* **11**, 71R-79R (2001).
179. Raman, J. et al. The Catalytic and Lectin Domains of UDP-GalNAc:Polypeptide α-N-Acetylgalactosaminyltransferase Function in Concert to Direct Glycosylation Site Selection. *Journal of Biological Chemistry* **283**, 22942-22951 (2008).
180. David C, K. Animal lectins: a historical introduction and overview. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1572**, 187-197 (2002).
181. Crocker, P.R., Paulson, J.C. & Varki, A. Siglecs and their roles in the immune system. *Nat Rev Immunol* **7**, 255-266 (2007).
182. Collins, B.E., Smith, B.A., Bengtson, P. & Paulson, J.C. Ablation of CD22 in ligand-deficient mice restores B cell receptor signaling. *Nat Immunol* **7**, 199-206 (2006).
183. Nguyen, D.H., Hurtado-Ziola, N., Gagneux, P. & Varki, A. Loss of Siglec expression on T lymphocytes during human evolution. *Proceedings of the National Academy of Sciences* **103**, 7765-7770 (2006).
184. Leyden, E. Zur Kenntniss des Bronchial-Asthma. *Virchows Arch. Path. Anat* **54**, 324-344 (1872).
185. Archer, G.T. & Blackwood, A. FORMATION OF CHARCOT-LEYDEN CRYSTALS IN HUMAN EOSINOPHILS AND BASOPHILS AND STUDY OF THE COMPOSITION OF ISOLATED CRYSTALS. *J Exp Med* **122**, 173-180 (1965).
186. Ackerman, S. et al. Molecular cloning and characterization of human eosinophil Charcot- Leyden crystal protein (lysophospholipase). Similarities to IgE binding proteins and the S-type animal lectin superfamily. *The Journal of Immunology* **150**, 456-468 (1993).
187. Teichberg, V.I., Silman, I., Beitsch, D.D. & Resheff, G. A beta-D-galactoside binding protein from electric organ tissue of Electrophorus electricus. *Proc Natl Acad Sci U S A* **72**, 1383-7 (1975).
188. Drickamer, K. Two distinct classes of carbohydrate-recognition domains in animal lectins. *J Biol Chem* **263**, 9557-60 (1988).
189. Laing, J.G., Robertson, M.W., Gritzmacher, C.A., Wang, J.L. & Liu, F.T. Biochemical and immunological comparisons of carbohydrate-binding protein 35 and an IgE-binding protein. *J Biol Chem* **264**, 1097-10 (1989).
190. Di Lella, S. et al. When Galectins Recognize Glycans: From Biochemistry to Physiology and Back Again. *Biochemistry* **50**, 7842-7857 (2011).
191. Leffler, H., Carlsson, S., Hedlund, M., Qian, Y. & Poirier, F. Introduction to galectins. *Glycoconj J* **19**, 433-40 (2004).
192. Hirabayashi, J. & Kasai, K. The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution. *Glycobiology* **3**, 297-304 (1993).

193. Brewer, C.F., Miceli, M.C. & Baum, L.G. Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. *Curr Opin Struct Biol* **12**, 616-23 (2002).
194. van Kooyk, Y. & Rabinovich, G.A. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol* **9**, 593-601 (2008).
195. Rabinovich, G.A. & Toscano, M.A. Turning 'sweet' on immunity: galectin-glycan interactions in immune tolerance and inflammation. *Nat Rev Immunol* **9**, 338-352 (2009).
196. Vasta, G.R. Roles of galectins in infection. *Nat Rev Micro* **7**, 424-438 (2009).
197. Liu, F.T., Patterson, R.J. & Wang, J.L. Intracellular functions of galectins. *Biochim Biophys Acta* **1572**, 263-73 (2002).
198. Hughes, R.C. Secretion of the galectin family of mammalian carbohydrate-binding proteins. *Biochim Biophys Acta* **1473**, 172-85 (1999).
199. Camby, I., Le Mercier, M., Lefranc, F. & Kiss, R. Galectin-1: a small protein with major functions. *Glycobiology* **16**, 137R-157R (2006).
200. Dumic, J., Dabelic, S. & Flögel, M. Galectin-3: An open-ended story. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1760**, 616-635 (2006).
201. Bidon-Wagner, N. & Le Pennec, J.P. Human galectin-8 isoforms and cancer. *Glycoconj J* **19**, 557-63 (2004).
202. Lahm, H. et al. Comprehensive galectin fingerprinting in a panel of 61 human tumor cell lines by RT-PCR and its implications for diagnostic and therapeutic procedures. *Journal of Cancer Research and Clinical Oncology* **127**, 375-386 (2001).
203. Saussez, S. et al. The determination of the levels of circulating galectin-1 and -3 in HNSCC patients could be used to monitor tumor progression and/or responses to therapy. *Oral Oncology* **44**, 86-93 (2008).
204. Vereecken, P. et al. Evaluation of the prognostic significance of serum galectin-3 in American Joint Committee on Cancer stage III and stage IV melanoma patients. *Melanoma Research* **19**, 316-320 10.1097/CMR.0b013e32832ec001 (2009).
205. Barrow, H. et al. Serum Galectin-2, -4, and -8 Are Greatly Increased in Colon and Breast Cancer Patients and Promote Cancer Cell Adhesion to Blood Vascular Endothelium. *Clinical Cancer Research* **17**, 7035-7046 (2011).
206. Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H. & Rini, J.M. X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-A resolution. *J Biol Chem* **268**, 27034-8 (1993).
207. Seetharaman, J. et al. X-ray Crystal Structure of the Human Galectin-3 Carbohydrate Recognition Domain at 2.1-Å Resolution. *Journal of Biological Chemistry* **273**, 13047-13052 (1998).
208. Leffler, H. & Barondes, S.H. Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian beta-galactosides. *J Biol Chem* **261**, 10119-26 (1986).
209. Sparrow, C.P., Leffler, H. & Barondes, S.H. Multiple soluble beta-galactoside-binding lectins from human lung. *J Biol Chem* **262**, 7383-90 (1987).

210. Salomonsson, E. et al. Mutational Tuning of Galectin-3 Specificity and Biological Function. *Journal of Biological Chemistry* **285**, 35079-35091 (2010).
211. Abbott, W.M., Hounsell, E.F. & Feizi, T. Further studies of oligosaccharide recognition by the soluble 13 kDa lectin of bovine heart muscle. Ability to accommodate the blood-group-H and -B-related sequences. *Biochem J* **252**, 283-7 (1988).
212. Carlsson, S. et al. Affinity of galectin-8 and its carbohydrate recognition domains for ligands in solution and at the cell surface. *Glycobiology* **17**, 663-76 (2007).
213. Stowell, S.R. et al. Galectin-1, -2, and -3 Exhibit Differential Recognition of Sialylated Glycans and Blood Group Antigens. *Journal of Biological Chemistry* **283**, 10109-10123 (2008).
214. Hirabayashi, J. et al. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1572**, 232-254 (2002).
215. Carlsson, S., Carlsson, M.C. & Leffler, H. Intracellular sorting of galectin-8 based on carbohydrate fine specificity. *Glycobiology* **17**, 906-12 (2007).
216. Leppanen, A., Stowell, S., Blixt, O. & Cummings, R.D. Dimeric galectin-1 binds with high affinity to alpha 2,3-sialylated and nonsialylated terminal N-acetyllactosamine units on surface bound extended glycans. *J. Biol. Chem.*, M412019200 (2004).
217. Patnaik, S.K. et al. Complex N-glycans are the major ligands for galectin-1, -3, and -8 on Chinese hamster ovary cells. *Glycobiology* **16**, 305-317 (2006).
218. Cederfur, C. et al. Different affinity of galectins for human serum glycoproteins: galectin-3 binds many protease inhibitors and acute phase proteins. *Glycobiology* **18**, 384-94 (2008).
219. Merkle, R.K. & Cummings, R.D. Asparagine-linked oligosaccharides containing poly-N-acetyllactosamine chains are preferentially bound by immobilized calf heart agglutinin. *Journal of Biological Chemistry* **263**, 16143-9 (1988).
220. Perillo, N.L., Pace, K.E., Seilhamer, J.J. & Baum, L.G. Apoptosis of T cells mediated by galectin-1. *Nature* **378**, 736-9 (1995).
221. Fukumori, T. et al. CD29 and CD7 Mediate Galectin-3-Induced Type II T-Cell Apoptosis. *Cancer Research* **63**, 8302-8311 (2003).
222. Hadari, Y.R. et al. Galectin-8 binding to integrins inhibits cell adhesion and induces apoptosis. *Journal of Cell Science* **113**, 2385-2397 (2000).
223. Stillman, B.N. et al. Galectin-3 and Galectin-1 Bind Distinct Cell Surface Glycoprotein Receptors to Induce T Cell Death. *The Journal of Immunology* **176**, 778-789 (2006).
224. Perone, M.J. et al. Suppression of autoimmune diabetes by soluble galectin-1. *J Immunol* **182**, 2641-53 (2009).
225. Toscano, M.A. et al. Galectin-1 Suppresses Autoimmune Retinal Disease by Promoting Concomitant Th2- and T Regulatory-Mediated Anti-Inflammatory Responses. *The Journal of Immunology* **176**, 6323-6332 (2006).

226. Blois, S.M. et al. A pivotal role for galectin-1 in fetomaternal tolerance. *Nat Med* **13**, 1450-7 (2007).
227. Barrionuevo, P. et al. A Novel Function for Galectin-1 at the Crossroad of Innate and Adaptive Immunity: Galectin-1 Regulates Monocyte/Macrophage Physiology through a Nonapoptotic ERK-Dependent Pathway. *The Journal of Immunology* **178**, 436-445 (2007).
228. Jung, E.-J. et al. Galectin-1 expression in cancer-associated stromal cells correlates tumor invasiveness and tumor progression in breast cancer. *International Journal of Cancer* **120**, 2331-2338 (2007).
229. van den Brule, F. et al. Galectin-1 accumulation in the ovary carcinoma peritumoral stroma is induced by ovary carcinoma cells and affects both cancer cell proliferation and adhesion to laminin-1 and fibronectin. *Lab Invest* **83**, 377-86 (2003).
230. Rubinstein, N. et al. Targeted inhibition of galectin-1 gene expression in tumor cells results in heightened T cell-mediated rejection: A potential mechanism of tumor-immune privilege. *Cancer Cell* **5**, 241-251 (2004).
231. Paz, A., Haklai, R., Elad-Sfadia, G., Ballan, E. & Kloog, Y. Galectin-1 binds oncogenic H-Ras to mediate Ras membrane anchorage and cell transformation. *Oncogene* **20**, 7486-93 (2001).
232. Moiseeva, E.P., Williams, B. & Samani, N.J. Galectin 1 inhibits incorporation of vitronectin and chondroitin sulfate B into the extracellular matrix of human vascular smooth muscle cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1619**, 125-132 (2003).
233. Glinsky, V.V., Huflejt, M.E., Glinsky, G.V., Deutscher, S.L. & Quinn, T.P. Effects of Thomsen-Friedenreich Antigen-specific Peptide P-30 on β-Galactoside-mediated Homotypic Aggregation and Adhesion to the Endothelium of MDA-MB-435 Human Breast Carcinoma Cells. *Cancer Research* **60**, 2584-2588 (2000).
234. Thijssen, V.L. et al. Tumor Cells Secrete Galectin-1 to Enhance Endothelial Cell Activity. *Cancer Research* **70**, 6216-6224 (2010).
235. van den Berg, T.K. et al. LacdiNAc-Glycans Constitute a Parasite Pattern for Galectin-3-Mediated Immune Recognition. *The Journal of Immunology* **173**, 1902-1907 (2004).
236. Sarafian, V. et al. Expression of Lamp-1 and Lamp-2 and their interactions with galectin-3 in human tumor cells. *Int J Cancer* **75**, 105-11 (1998).
237. Kita, Y. et al. Quantitative Glycomics of Human Whole Serum Glycoproteins Based on the Standardized Protocol for Liberating N-Glycans. *Molecular & Cellular Proteomics* **6**, 1437-1445 (2007).
238. Yu, L.-G. et al. Galectin-3 Interaction with Thomsen-Friedenreich Disaccharide on Cancer-associated MUC1 Causes Increased Cancer Cell Endothelial Adhesion. *Journal of Biological Chemistry* **282**, 773-781 (2007).
239. Delacour, D. et al. Apical Sorting by Galectin-3-Dependent Glycoprotein Clustering. *Traffic* **8**, 379-388 (2007).

240. Demetriou, M., Granovsky, M., Quaggin, S. & Dennis, J.W. Negative regulation of T-cell activation and autoimmunity by Mgat5N-glycosylation. *Nature* **409**, 733-739 (2001).
241. Akahani, S., Nangia-Makker, P., Inohara, H., Kim, H.-R.C. & Raz, A. Galectin-3: A Novel Antiapoptotic Molecule with A Functional BH1 (NWGR) Domain of Bcl-2 Family. *Cancer Research* **57**, 5272-5276 (1997).
242. Shalom-Feuerstein, R., Cooks, T., Raz, A. & Kloog, Y. Galectin-3 Regulates a Molecular Switch from N-Ras to K-Ras Usage in Human Breast Carcinoma Cells. *Cancer Research* **65**, 7292-7300 (2005).
243. Vandenbrule, F. et al. Antisense galectin-3 alters thymidine incorporation in human MDA-MB435 breast cancer cells. *Int J Oncol* **11**, 261-4 (1997).
244. John, C.M., Leffler, H., Kahl-Knutsson, B., Svensson, I. & Jarvis, G.A. Truncated Galectin-3 Inhibits Tumor Growth and Metastasis in Orthotopic Nude Mouse Model of Human Breast Cancer. *Clinical Cancer Research* **9**, 2374-2383 (2003).
245. Shimura, T. et al. Galectin-3, a novel binding partner of beta-catenin. *Cancer Res* **64**, 6363-7 (2004).
246. Kuwabara, I. & Liu, F.T. Galectin-3 promotes adhesion of human neutrophils to laminin. *J Immunol* **156**, 3939-44 (1996).
247. Liu, F.-T. & Rabinovich, G.A. Galectins: regulators of acute and chronic inflammation. *Annals of the New York Academy of Sciences* **1183**, 158-182 (2010).
248. Zuberi, R.I. et al. Critical Role for Galectin-3 in Airway Inflammation and Bronchial Hyperresponsiveness in a Murine Model of Asthma. *The American Journal of Pathology* **165**, 2045-2053 (2004).
249. López, E. et al. Inhibition of Chronic Airway Inflammation and Remodeling by Galectin-3 Gene Therapy in a Murine Model. *The Journal of Immunology* **176**, 1943-1950 (2006).
250. Ho, M.K. & Springer, T.A. Mac-2, a novel 32,000 Mr mouse macrophage subpopulation-specific antigen defined by monoclonal antibodies. *J Immunol* **128**, 1221-8 (1982).
251. Novak, R., Dabelic, S. & Dumic, J. Galectin-1 and galectin-3 expression profiles in classically and alternatively activated human macrophages. *Biochimica et Biophysica Acta (BBA) - General Subjects*.
252. MacKinnon, A.C. et al. Regulation of Alternative Macrophage Activation by Galectin-3. *The Journal of Immunology* **180**, 2650-2658 (2008).
253. Henderson, N.C. & Sethi, T. The regulation of inflammation by galectin-3. *Immunological Reviews* **230**, 160-171 (2009).
254. Hadari, Y.R. et al. Galectin-8. *Journal of Biological Chemistry* **270**, 3447-3453 (1995).
255. Kabir, S. & Gerwig, G.J. The structural analysis of the O-glycans of the jacalin-bound rabbit immunoglobulin G. *Biochemistry and Molecular Biology International* **42**, 769-778 (1997).

256. Carlsson, M.C. et al. Galectin-8 in IgA Nephritis: Decreased Binding of IgA by Galectin-8 Affinity Chromatography and Associated Increased Binding in Non-IgA Serum Glycoproteins. *J Clin Immunol* (2011).
257. Diskin, S., Cao, Z., Leffler, H. & Panjwani, N. The role of integrin glycosylation in galectin-8-mediated trabecular meshwork cell adhesion and spreading. *Glycobiology* **19**, 29-37 (2009).
258. Eshkar Sebban, L. et al. The Involvement of CD44 and Its Novel Ligand Galectin-8 in Apoptotic Regulation of Autoimmune Inflammation. *The Journal of Immunology* **179**, 1225-1235 (2007).
259. Cederfur, C. et al. Glycoproteomic identification of galectin-3 and -8 ligands in bronchoalveolar lavage of mild asthmatics and healthy subjects. *Biochimica et Biophysica Acta (BBA) - General Subjects*.
260. Chandrasekaran, E.V. et al. Mammalian sialyltransferase ST3Gal-II: its exchange sialylation catalytic properties allow labeling of sialyl residues in mucin-type sialylated glycoproteins and specific gangliosides. *Biochemistry* **50**, 9475-87 (2011).
261. Cárcamo, C. et al. Galectin-8 binds specific β1 integrins and induces polarized spreading highlighted by asymmetric lamellipodia in Jurkat T cells. *Exp Cell Res* **312**, 374-386 (2006).
262. Arbel-Goren, R., Levy, Y., Ronen, D. & Zick, Y. Cyclin-dependent Kinase Inhibitors and JNK Act as Molecular Switches, Regulating the Choice between Growth Arrest and Apoptosis Induced by Galectin-8. *Journal of Biological Chemistry* **280**, 19105-19114 (2005).
263. Levy, Y. et al. Galectin-8 Functions as a Matricellular Modulator of Cell Adhesion. *Journal of Biological Chemistry* **276**, 31285-31295 (2001).
264. Thurston, T.L.M., Wandel, M.P., von Muhlinen, N., Foeglein, A. & Randow, F. Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. *Nature* **482**, 414-418 (2012).
265. Sorme, P., Kahl-Knutsson, B., Huflejt, M., Nilsson, U.J. & Leffler, H. Fluorescence polarization as an analytical tool to evaluate galectin-ligand interactions. *Anal Biochem* **334**, 36-47 (2004).
266. Fujimura, T. et al. Glycosylation status of haptoglobin in sera of patients with prostate cancer vs. benign prostate disease or normal subjects. *Int J Cancer* **122**, 39-49 (2008).
267. Chandrasekaran, E.V. et al. Mammalian Sialyltransferase ST3Gal-II: Its Exchange Sialylation Catalytic Properties Allow Labeling of Sialyl Residues in Mucin-Type Sialylated Glycoproteins and Specific Gangliosides. *Biochemistry* **50**, 9475-9487 (2011).
268. Rippe, B. & Haraldsson, B. Transport of macromolecules across microvascular walls: the two-pore theory. *Physiological Reviews* **74**, 163-219 (1994).
269. Quaye, I.K. Haptoglobin, inflammation and disease. *Trans R Soc Trop Med Hyg* **102**, 735-42 (2008).

270. Feuk-Lagerstedt, E., Jordan, E.T., Leffler, H., Dahlgren, C. & Karlsson, A. Identification of CD66a and CD66b as the Major Galectin-3 Receptor Candidates in Human Neutrophils. *The Journal of Immunology* **163**, 5592-5598 (1999).
271. Karlsson, A. et al. Galectin-3 functions as an opsonin and enhances the macrophage clearance of apoptotic neutrophils. *Glycobiology* **19**, 16-20 (2009).
272. Richardson, D.R. & Ponka, P. The molecular mechanisms of the metabolism and transport of iron in normal and neoplastic cells. *Biochim Biophys Acta* **1331**, 1-40 (1997).
273. Madsen, M., Graversen, J.H. & Moestrup, S.K. Haptoglobin and CD163: captor and receptor gating hemoglobin to macrophage lysosomes. *Redox Rep* **6**, 386-8 (2001).
274. Stechly, L. et al. Galectin-4-regulated delivery of glycoproteins to the brush border membrane of enterocyte-like cells. *Traffic* **10**, 438-50 (2009).
275. Gonzalez, R.M., Daly, D.S., Tan, R., Marks, J.R. & Zangar, R.C. Plasma Biomarker Profiles Differ Depending on Breast Cancer Subtype but RANTES Is Consistently Increased. *Cancer Epidemiology Biomarkers & Prevention* **20**, 1543-1551 (2011).
276. Aridor, M. & Hannan, L.A. Traffic Jam: A Compendium of Human Diseases that Affect Intracellular Transport Processes. *Traffic* **1**, 836-851 (2000).