

# **Development and application of affinity- mass spectrometry to identify protein- carbohydrate interactions**

Inauguraldissertation zur

Erlangung des akademischen Grades

Doktor der Medizinwissenschaften

der Universitätsmedizin Rostock

Gutachter:

Pr. Dr. Michael O. Glocker, Universitätsmedizin Rostock, Institut für Immunologie

Pr. Dr. Dr.h.c. Michael Przybylski, Steinbeis center AffyMSLifeChem

Pr. Dr. Christopher Gerner, Universität Wien, Institute für Analytische Chemie

Jahr der Einreichung: 2022

Jahr der Verteidigung: 2022

Vorgelegt von

Yannick Baschung, geb. am 04.05.1988 in Mulhouse (FR)

Aus 17 rue du 19 Novembre, 68870 Brinckheim (FR)

Brinckheim, 29.11.2021

*For my parents who always supported me*

---

## ACKNOWLEDGEMENTS

The present work has been performed from October 2015 to December 2018 at the Steinbeis Center for Biopolymer Analysis and Biomedical Mass Spectrometry, under the supervision of Prof. Dr. Dr. h. c. Michael Przybylski.

I would like to express my deep gratitude and appreciation to Professor Michael Przybylski for the opportunity to perform my PhD work in his laboratory, for his guidance and kind support throughout the years. His expert knowledge, enthusiasm and dedication to research were an inspiration to me.

I would like also to thank my supervisor from Rostock Medical University, Prof. Dr. Michael Glocker, who allowed me to apply to structured curriculum for MD and PhD candidates "Molecular Mechanisms of Regenerative Processes" and helped me through these years with great scientific advices and interesting talks.

I would also like to thank my colleagues with whom I shared enlightening scientific discussion and very nice moments:

Loredana Lupu, for her support, kindness and her deep passion for medical sciences.

Pascal Wiegand and Hendrik Rusche, who both helped me a lot in the search of solutions, and for the nice time spent abroad during scientific conferences.

Delia Mihoc, Nico Huttman, Zdenek Kuckaka, Fabio Borri, Stefan Maeser and many other for their constant positive mind and the nice work atmosphere that they provided.

I am also grateful to all the participants of the semester's meeting of our Ph.D. curriculum for all the interesting talks and research they shared.

---

Special thanks to Stephan Rawer for his precious guidance and help for the synthesis of the peptides at thermofisher in Darmstadt, and for the long discussion we had about our shared passion for scuba diving.

I would like to also thank Dr. Alexander Lazarev from Pressure Biosciences Inc., Boston, for cooperation and support in high-pressure proteolytic digestion.

Last but not least, my deepest gratitude is dedicated to my parents, who provided help, encouragement and support for all these years, and to my wife, who helped me to stay constantly positive with her support and smile.

---

## PUBLICATION

**Y. Baschung**, L. Lupu, A. Moise, M. Glocker, S. Rawer, A. Lazarev, and M. Przybylski. Epitope Ligand Binding Sites of Blood Group Oligosaccharides in Lectins Revealed by Pressure-Assisted Proteolytic Excision Affinity Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 29, 1881-1891.

## Conference presentations

1. **Yannick Baschung**, Michael Przybylski, Michael Glocker (2015) Characterization of carbohydrate and chaperone binding sites in human  $\alpha$ -galactosidase A by proteolytic affinity - mass spectrometry.  
[Poster] 1<sup>st</sup> MS-peptide day, 10-12 February 2016, Florence, Italy
2. **Yannick Baschung**, Adrian Moise, Loredana Lupu, Michael Przybylski (2016) Characterization of carbohydrate binding sites in galectins by proteolytic affinity – mass spectrometry.  
[Poster] DGMS, 29 February 2016, Hamburg, Germany
3. **Yannick Baschung** (2016) Epitope Identification and Affinity Determination of A $\beta$ - specific Antibodies by online SPR- MS.  
[Oral presentation] Ph. D. curriculum participant camp, 29 June 2016, Rostock, Germany
4. **Yannick Baschung** (2017) Characterization of Chaperone Binding Sites to Lysosomal Enzymes by Affinity- Mass Spectrometry.  
[Oral presentation] “Affinity - Mass spectrometry – New Methods for Biochemical and Clinical Applications”, 2-3 November 2017, Russelsheim, Germany

- 
5. **Yannick Baschung** (2017) Aspartic Proteases from carnivorous plant *Nepenthes* as a tool for proteolytic affinity- mass spectrometry.  
[Oral presentation] Ph. D. curriculum participant camp, 28 June 2017, Rostock, Germany
  
  6. **Yannick Baschung**, Loredana Lupu, Michael Glocker, Michael Przybylski (2018) Structure identification and affinity quantification of soybean agglutinin – carbohydrate interaction epitopes using proteolytic affinity – mass spectrometry  
[Poster] ASMS 2018, 4-7 June 2018, San Diego, CA, USA.
  
  7. **Yannick Baschung** (2018) Epitope ligand sites of blood group oligosaccharides in lectins revealed by high pressure proteolytic extraction affinity mass spectrometry.  
[Oral presentation] International Symposium & Summer School: “Mass spectrometry in Medical Technology and Biotechnology”, 8-9 November 2018, Russelsheim, Germany.

---

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	3
PUBLICATION.....	5
1. INTRODUCTION.....	10
1.1. Lectins: History, structure, function and applications .....	10
1.2. Structure and functions of glycans .....	16
1.3. Glycoproteins and glycolipids .....	18
1.4. Development of carbohydrates as therapeutic agents.....	20
1.4.1. Glycomimetic drugs .....	20
1.4.2. Carbohydrate-based vaccines.....	21
1.5. SPR biosensor analysis for kinetic determination of protein-carbohydrate interaction ....	24
1.6. High-pressure proteolytic methods as a new tool for enzyme hydrolysis .....	26
1.7. Mass-spectrometric methods for protein interactions analysis .....	30
1.8. Affinity-Mass-Spectrometry for protein-ligand structure and interaction analysis .....	33
1.9. Scientific goal of the dissertation.....	38
2. RESULTS AND DISCUSSION .....	41
2.1. Structural characterization of lectins .....	41
2.1.1. Characterization of Glycine Max isolectins.....	41
2.1.2. Development of an efficient proteolytic procedure for Soybean Agglutinin .....	43
2.1.3. Structural characterization of lectins by proteolytic peptide mapping - mass spectrometry 49	
2.1.4. Synthesis and purification of a peptide spacer for glycan immobilization .....	53
2.2. Identification of carbohydrate recognition epitopes in lectins.....	56
2.2.1. Pressure enhanced proteolytic excision and extraction of lectins .....	56

---

2.2.2.	Development of buffer and elution system for lectin affinity interactions .....	58
2.2.3.	Identification of saccharide and disaccharide binding sites in lectins .....	60
2.2.4.	Identification of binding sites of blood group oligosaccharides in SBA .....	62
<b>2.3.</b>	<b>Determination of dissociation constants of lectin - carbohydrates interactions.....</b>	<b>66</b>
<b>2.4.</b>	<b>Surface immobilization procedure of carbohydrate for the identification of protein binding sites .....</b>	<b>72</b>
<b>3.</b>	<b>EXPERIMENTAL PART .....</b>	<b>77</b>
<b>3.1.</b>	<b>Materials and Reagents .....</b>	<b>77</b>
<b>3.2.</b>	<b>Chemical modification and hydrolysis of proteins .....</b>	<b>78</b>
3.2.1.	Reduction and alkylation .....	78
3.2.2.	Proteolytic digestion .....	79
3.2.3.	Pressure-enhanced digestion .....	80
<b>3.3.</b>	<b>Separation methods .....</b>	<b>81</b>
3.3.1.	Electrophoresis .....	81
3.3.2.	Liquid chromatography.....	82
3.3.3.	Sample cleaning and desalting .....	83
<b>3.4.</b>	<b>Affinity methods.....</b>	<b>84</b>
3.4.1.	Preparation of immobilized carbohydrate Sepharose columns .....	84
3.4.2.	Proteolytic extraction and excision .....	85
3.4.3.	Pressure enhanced proteolytic excision .....	85
3.4.4.	Surface acoustic wave biosensor .....	86
3.4.5.	Carbohydrate-glycophage conjugation.....	90
<b>3.5.</b>	<b>Mass spectrometric methods.....</b>	<b>90</b>
3.5.1.	ESI-Ion trap MS .....	90
3.5.2.	MALDI -MS.....	91
3.5.3.1.	ESI-Quadrupole mass spectrometry.....	91
3.5.3.2.	High-performance liquid chromatography mass spectrometry.....	92
<b>3.6.</b>	<b>Solid phase peptide synthesis .....</b>	<b>93</b>
<b>3.7.</b>	<b>Offline coupling of SAW with MALDI-MS .....</b>	<b>95</b>



---

4. ZUSAMMENFASSUNG .....	97
5. REFERENCES .....	100
6. APPENDIX .....	114
ABBREVIATIONS.....	118
DECLARATION OF INDEPENDENCE .....	120

## 1. INTRODUCTION

### 1.1. Lectins: History, structure, function and applications

Lectins are a large family of proteins able to bind both free carbohydrates and carbohydrates attached to glycoconjugates, such as glycoproteins (proteins which contain oligosaccharide chains covalently attached to amino acid side-chains), intact cells or viruses (1).

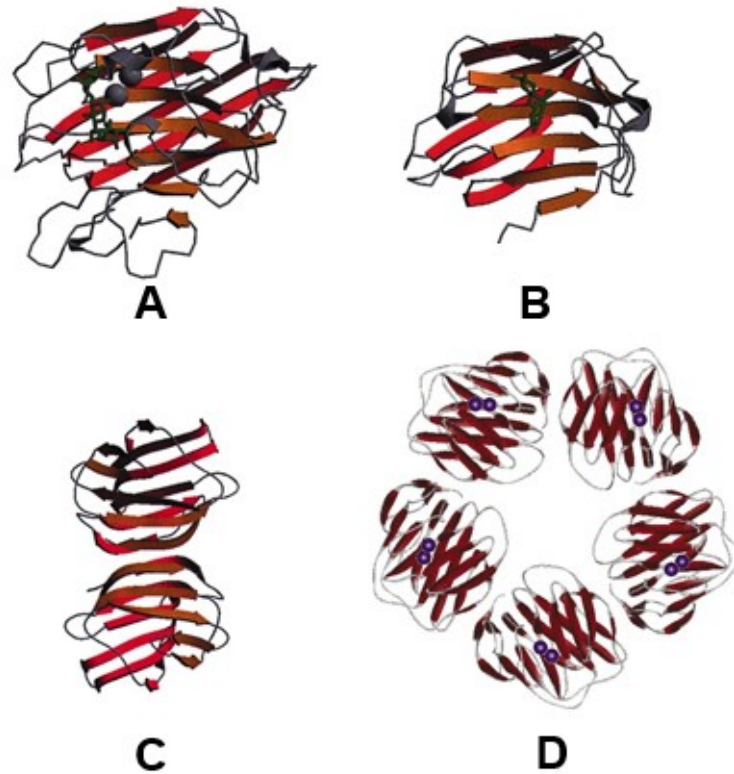
The earliest description of lectins is believed to have been given by Peter Hermann Stillmark in his doctoral thesis presented in 1888 to the University of Dorpat, when extracts of castor bean seeds were found to agglutinate animal red blood cells. Subsequently, lectins were found in all living organisms and classified into evolutionarily-related families identified by carbohydrate-recognition domains (CRDs) based on primary and/or three-dimensional structural similarities. Lectins are particularly common in the seeds of leguminous plants and the first hemagglutinin, Concanavalin A (figure 1A), was isolated from jack-bean by James B. Sumner in 1919 and later on reported to agglutinate cells like erythrocytes and yeast as well as glycogen from solution (2). The specific glycan-binding activities of plant lectins is an extremely useful scientific tools for cell separation and analysis, bacterial strain typing (characterization of bacterial strains in order to ascertain whether they are derived from a single parental organism), detection and targeting of aberrant glycans for cancer diagnosis and therapy, and purification of glycoconjugates. However, although lectins can comprise up to 30% of the total protein in legume seeds, their biological functions remain speculative (3). Several reviews or books have summarized the possible functions of plant lectins (4) (5). Generally, plant lectins have both internal and external activities. The former refers to the functions acting within plants during various developmental processes such as interactions with storage proteins or enzymes (6). External activities include the roles of lectins in response to various biotic and abiotic stresses. For example, some of the lectins may be associated with the binding of symbiotic rhizobia (Gram-negative soil bacteria) to form root nodules.

The term lectin was introduced by W.C. Boyd in 1954 (7), after the discovery of the human blood-group (or blood type) binding specificity of the hemagglutinins by K.

Landsteiner, for which he was awarded the Nobel Prize for Physiology or Medicine in 1930. W.C. Boyd and Karl O. Renkonen observed that crude extracts of the lima bean, *Phaseolus limensis*, and the tufted vetch, *Vicia cracca*, agglutinated blood type A erythrocytes but not blood type B or O cells, whereas an extract of the asparagus pea, *Lotus tetragonolobus*, agglutinated specifically blood type O erythrocytes (8). This discovery played an important role in the early investigation on the structural basis of the specificity of antigens associated with the ABO blood group system. Walter J. T. Morgan and Winifred M. Watkins discovered in the 1950s that the agglutination of type A red cells was best inhibited by  $\alpha$ -linked *N*-acetyl-D-galactosamine and that of type O cells was best inhibited by  $\alpha$ -linked L-fucose, leading to the conclusion that  $\alpha$ -*N*-acetyl-D-galactosamine and  $\alpha$ -L-fucose are the sugar determinants conferring A and O blood group specificity, respectively.

The first animal lectin, asialoglycoprotein receptor (ASGPR), was discovered a few years later in late 1960 in the liver cells of rabbits by A. Morell and Gilbert Ashwell (9). The ASGPR are lectins which bind asialoglycoprotein and glycoproteins from which a sialic acid has been removed to expose galactose residues.

Studies of the amino-acid sequence of lectins and elucidation of their 3D structure became more intense in the 1970s, with the development of affinity chromatography for the isolation of lectins. Concanavalin A was the first lectin whose amino-acid sequence has been established (10) and its 3D structure was solved by high resolution X-ray crystallography (11). Today, the number of lectin primary and 3D structures has increased dramatically. Despite the lack of amino-acid sequence similarities, tertiary structures observed in lectins (figure 1), and referred to as the lectin fold, show high similarities consisting characteristically of an elaborate jelly roll, derived from antiparallel  $\beta$ -strands, arranged as two  $\beta$ -sheets (12). This fold has been found in numerous animals and plants lectins, such as Galectins, Concanavalin A, and pentraxins (13).



**Figure 1.** Structures of concanavalin A (A), human galectin-7 (B), Lencil lectin (C) and Serum amyloid protein (D) represented as ribbon diagrams, showing high similarities in their antiparallel  $\beta$ -strands.

Despite a wide diversity in three dimensional structures and in the carbohydrate binding sites among lectins, there are some features of the interaction of proteins with carbohydrate that appear to be common to all lectins. Both hydrogen bonds and van der Waals interactions are involved in stabilizing lectin-carbohydrate interactions. Lectins also often require metal ions such as  $\text{Ca}^{2+}$  and  $\text{Mn}^{2+}$  for the maintenance of their conformation and their ability to bind to the target carbohydrate.

Lectins are separated into eight general types on the basis of their carbohydrate-binding specificity: complex, fucose, galactose, N-acetylglucosamine, mannose, mannose/glucose, mannose/maltose, and sialic acid. Lectins are classified into different families (table 1) according to their carbohydrate recognition domains (CRDs). C-type lectins, which are lectins containing  $\text{Ca}^{2+}$ -dependent CRDs, are the most abundant lectins in human genome (14), and provide the sugar recognition

activity in a variety of cell surface and extracellular lectins. Other modules in these proteins initiate biological processes such as adhesion, endocytosis and complement fixation (15). R-type, I-type and galectins are other extracellular lectin families that are either secreted into the extracellular matrix or body fluids, or localized to the plasma membrane.

**Table 1.** Classification of lectin families and their subcellular location for vertebrates lectins

Lectin family	Saccharide ligand	Subcellular location	Example of functions
Calnexin	Glc <sub>1</sub> Man <sub>9</sub>	Endoplasmic reticulum	Protein sorting in the endoplasmic reticulum.
M-type	Man <sub>8</sub>	Endoplasmic reticulum	ER-associated degradation of glycoproteins.
L-type	Various	Endoplasmic reticulum, Golgi	Protein sorting in the endoplasmic reticulum.
P-type	Man 6-phosphate, others	Secretory pathway	Various (Protein sorting post-Golgi, glycoprotein trafficking, enzyme targeting).
C-type	Various	Cell membrane, extracellular	Various (glycoprotein clearance, innate immunity).
Galectins	b-Galactosides	Cytoplasm, extracellular	Glycan crosslinking in the extracellular matrix.
I-type	Sialic acid	Cell membrane	Cell adhesion.
R-type	Various	Golgi, Cell membrane	Enzyme targeting, glycoprotein hormone turnover.
F-box	GlcNAc <sub>2</sub>	Cytoplasm	Degradation of misfolded glycoproteins.
Ficolins	GlcNAc, GalNAc	Cell membrane, extracellular	Innate immunity.
Chitinase-like	Chito-oligosaccharides	Extracellular	Collagen metabolism (YKL-40).
F-type	Fuc-terminating oligosaccharides	Extracellular	Innate immunity.
Intelectins	Gal, galactofuranose, pentoses	Extracellular/cell membrane	Innate immunity. Fertilization and embryogenesis.

L-type lectins on the other hand, are intracellular lectins found in abundance in the seeds of leguminous plants, such as concanavalin A from jack beans. The L-like lectin domain has an overall globular shape composed of a beta-sandwich of two major twisted antiparallel beta-sheets. The ligand binding site is in a negatively-charged cleft of conserved residues. Most L-type CRDs require metal ions for ligand binding: in

concanavalin A for example, a transition metal is bound at one site and  $\text{Ca}^{2+}$  at the second site. M-type, P-type and calnexins are other intracellular lectins located in luminal compartments of the secretory pathway and playing a role in the trafficking, sorting and targeting of maturing glycoproteins.

In addition to these lectin families, F-box lectins, ficolins, chitinase-like lectins, F-type lectins and intelectins are recently identified new groups of animal lectins (16) (17) (18) (19) (20), some of which have roles complementary to those of the well-established lectin families.

The interaction between lectins and carbohydrates regulates a wide variety of biological processes, such enzymatic synthesis and degradation, intracellular sorting of glycoconjugates and tumor metastasis. These interactions also play a key role in the attachment of parasite, fungi, bacteria or viruses to host cells, setting the first stage of infections (21) (22). Such interaction requires high selectivity which is provided by the specific stereochemical fit of the target carbohydrate to carbohydrate recognition domain (CRD) of the lectin. The ability of lectins to recognize and differentiate other carbohydrates in a distinct way makes its use relevant for research involving structural analysis, purification, biotechnological use in different areas such as molecular and cell biology, immunology, pharmacology, medicine, clinical analysis, nanotechnology as well as in systems for drug release (23).

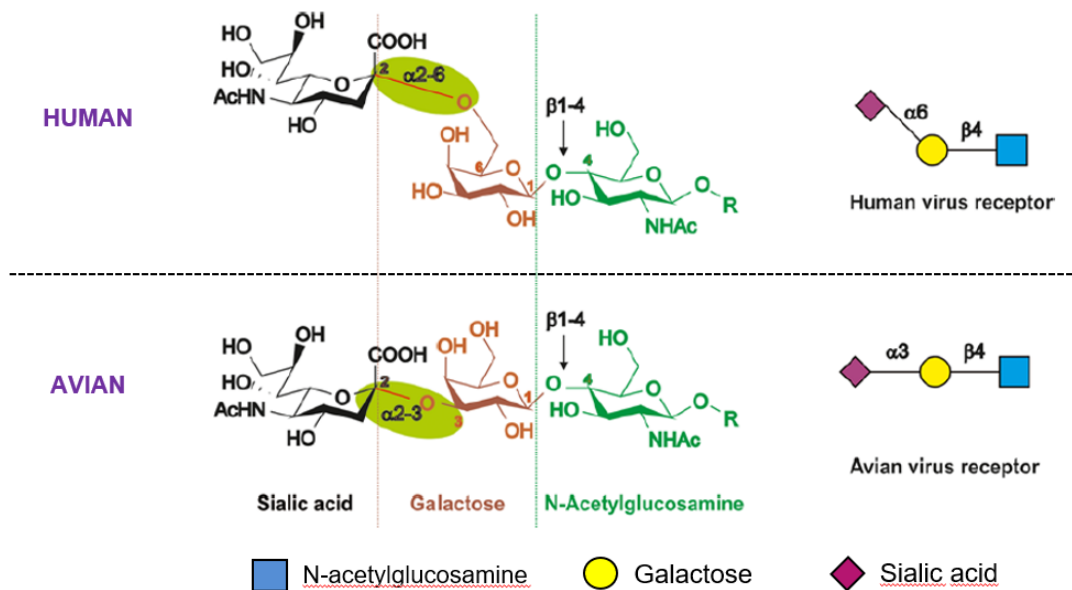
Since the end of the XX<sup>th</sup> century and following the development of powerful instrumentations and advancement in synthetic and carbohydrate chemistry as well as X-ray crystallography, great progress has been made in characterization of protein-carbohydrate interactions, leading to an increased comprehensiveness of their role in infectious disease, as cell surfaces are covered with complex carbohydrates in the form of glycoprotein or glycolipids that are of key importance in the antigen recognition process.

The importance of lectin-carbohydrate interactions is highlighted by the considerable number of diseases which result from deficient biosynthesis or processing of glycans. For example, inherited defects of lysosomal alpha-galactosidase A, responsible for the degradation of globotriaosylceramide, lead to pathological levels of substrate

accumulation in patients suffering from Fabry disease, a lysosomal storage disease (LSD) (24) (25).

Lectins can not only differentiate between different monosaccharides within a building block of glycans, but also between different glycan linkages positions.

A good example of this specificity is the capacity of the lectins *Sambucus nigra* agglutinin (SNA) and *Maackia amurensis* agglutinin (MAA) to recognize a terminal sialic acid (SA) linked to a galactose via respectively an  $\alpha$ 2-6 linkage or an  $\alpha$ 2-3 linkage (figure 2), thus differentiating between the human and the avian influenza virus (26).



**Figure 2.** Difference between the  $\alpha$ 2-6 sialic acid determinant of the Human (top) and the  $\alpha$ 2-3 sialic acid determinant of Avian influenza virus receptor (bottom).

Lectin-carbohydrate interactions have however considerably lower dissociation constants ( $K_D$ ) than antigen-antibody interactions. Dissociation values of lectins or glycopeptides binding to monosaccharides are usually in the  $\mu$ M to mM range while binding strength to complex carbohydrates is in the  $\mu$ M range (15) (4).

Lectins may have multiple CRDs with different carbohydrate specificities. They are also classified according to their overall tertiary and quaternary structure. The major

secondary structure component in most lectins is  $\beta$ -sheet. This secondary structure can participate in the formation of multiple tertiary structures such as  $\beta$ -sandwich,  $\beta$ -trefoil,  $\beta$ -propeller,  $\beta$ -prism I and II (27) (28).

Lectins form complexes with carbohydrates with the help of a network of hydrogen bonds, depending on the hydroxyl groups of the carbohydrate and the amino-acid side chains of the protein (29), and through hydrophobic interactions between the hydrophobic surface of the carbohydrate, created by the steric disposition of the hydroxyl groups, and the hydrophobic side chains of the protein (30) (31). Less frequently, bonds are formed through electrostatic interaction and coordination with metal ions (32) (33).

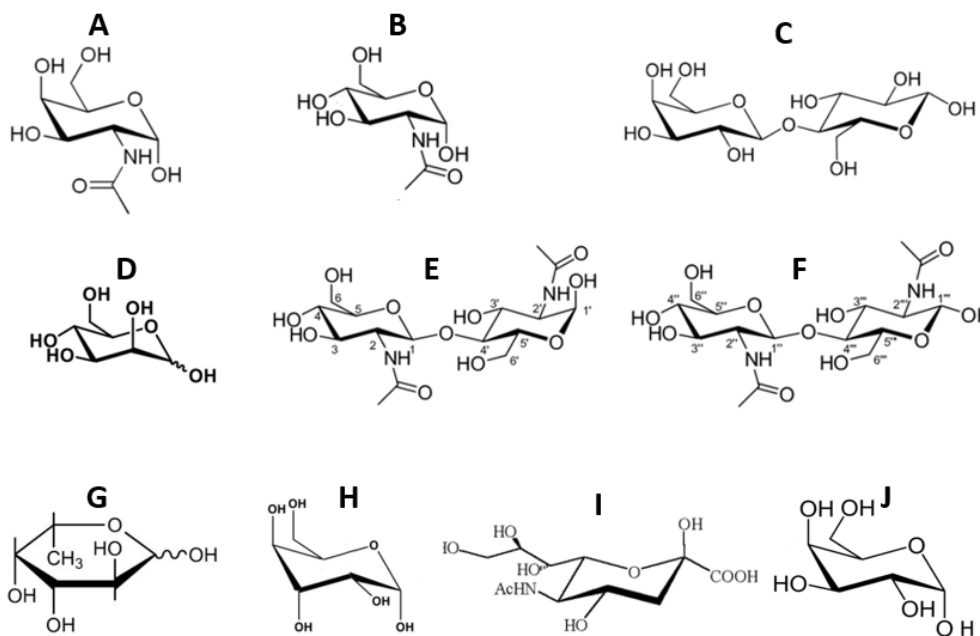
## **1.2. Structure and functions of glycans**

As with the glycoproteins and lectins, the biological functions of glycans span a wide spectrum of functions, from structural and modulatory properties such as nutrient storage, to molecular mimicry of host glycans and specific recognition by other molecules such as the lectins. Glycans also have many protective, stabilizing, organizational, and barrier functions. Indeed, the external location of glycans on most glycoproteins can provide a general shield, protecting the underlying polypeptide from recognition by proteases, blocking antibody binding, and even (in the case of mucins) protecting entire tissue surfaces from microbial attachment (34) (35).

The biosynthesis of glycans takes place in the endoplasmic reticulum (ER) and the Golgi apparatus and is catalyzed by glycosyl transferases and glycosyl hydrolases. The assembly of glycan blocks is a non-template-driven approach (36) and the final glycan structures depend on the carbohydrate specificity of the enzymes which are expressed in each organism. An illustrative example of the effect of expression of different glycosyltransferases is the polymorphism of complex carbohydrate structures of glycoproteins and glycolipids expressed at the surface of red blood cells. This process leads to the different ABO blood group phenotypes, which are of high importance for blood transfusion and organ transplantation.



The structure of the most abundant saccharides forming building blocks and found to interact with carbohydrate-binding proteins is illustrated in figure 3.



**Figure 3.** Structure of common mono and bi- saccharides ligand found to bind lectins and discussed in this thesis. (A) N-Acetylgalactosamine (GalNAc), (B) N-Acetylglucosamine (GlcNAc), (C) Lactose (Lac), (D) Mannose (Man), (E) (GlcNAc)<sub>2</sub> β-α form, (F) (GlcNAc)<sub>2</sub> β-β form, (G) Fucose (Fuc), (H) Glucose (Glu), (I) Sialic acid (Sia), (J) Galactose (Gal).

The extend of compositional and structural variability of glycan oligomers is unsurpassed in nature due to the unique property of independently combining parameters such anomeric status, linkage position, addition of branches and ring size with the carbohydrate sequence (37).

These building blocks of glycans surpass by far amino acids and nucleotides in information-storing capacity (38). Glycans can also differentiate between various post-translational modifications (PTM) such as sulfation, methylation, acetylation and phosphorylation (39).

The study of the functions of glycans is challenging and functions are sometimes discovered accidentally. In some instances, the investigator who has elucidated complete details of the structure and biosynthesis of a specific glycan is left without knowing its functions (35). Experiments need to be designed in order to differentiate

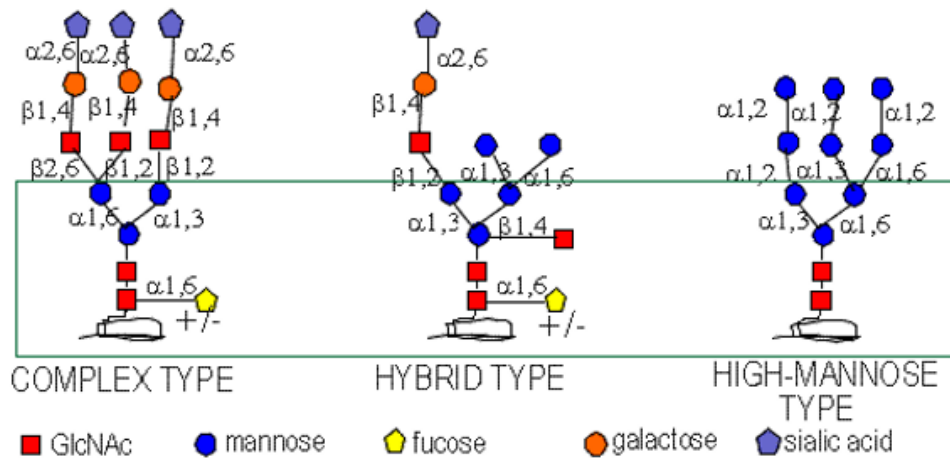
between the multiple functions mediated by each glycan. Various approaches to study the glycans binding sites and strengths currently exist, such as the use of Glycan-Recognizing Probes, the metabolic inhibition or alteration of glycosylation, and techniques such as hemagglutination, flow cytometry, surface plasmon resonance, and affinity chromatography. Strategies to investigate the carbohydrate-recognition domains of glycan-recognizing proteins will be explored in this thesis.

### 1.3. Glycoproteins and glycolipids

Glycoproteins and glycolipids are formed through covalent attachment of a glycan to the functional group of peptides and lipids in an enzymatic process called glycosylation, one of the most abundant post-translational modification (PTM). It is estimated that up to 70% of human proteins are glycosylated (40), and this modification serves also other purposes than carbohydrate-lectin interactions, such as protein stabilization with linkage to the amine nitrogen of asparagine residues and protein folding (41).

Depending on the linkage between the sugar and the amino acid, the sugar chains are mainly classified as either N-glycans, O-glycans, Glycosaminoglycans, or Glycosphingolipids. All N-glycans possess a common trimannosyl core (Man- $\alpha$ 1,6-(Man- $\alpha$ 1,3-)Man- $\beta$ 1,4-GlcNAc- $\beta$ 1,4-GlcNAc) linked to the amide nitrogen of Asn in the protein, to which different monosaccharides are added conforming the three subtypes of N-glycosylation : high-mannose, hybrid or complex (figure 4). High-mannose oligosaccharides can have from 3 mannose residues to as many as 60 in protozoans and yeast. Hybrid and complex N-glycans can construct various outer chains, constructed from a series of core portions and modifications (fucosylation, sialylation, galactosylation, glucuronylation, and/or sulfation) of their Gal- $\beta$ 1,4-GlcNAc units, on their GlcNAc residues.

### N-linked oligosaccharides

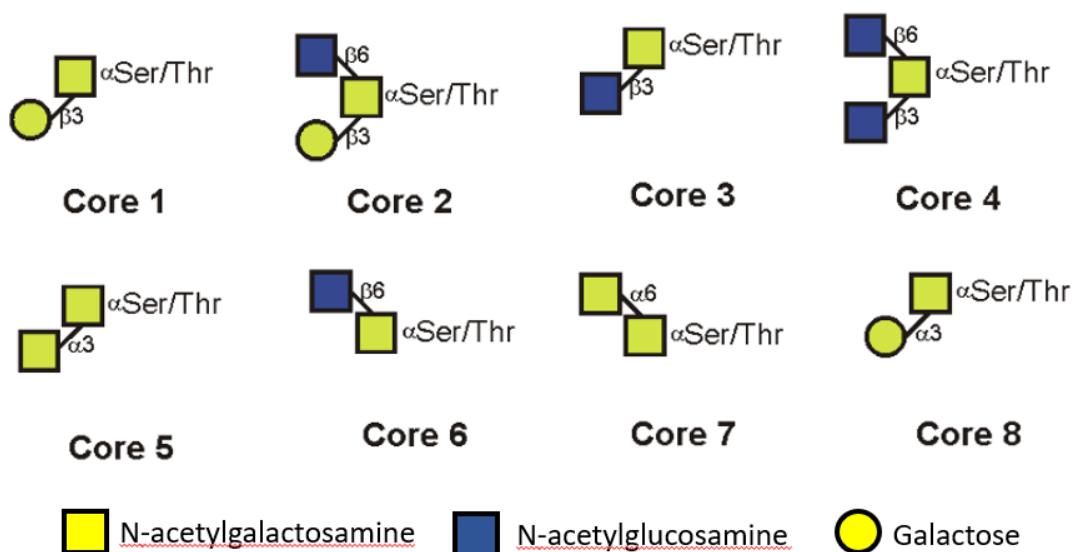


**Figure 4.** Illustration of differences between complex, hybrid and high-mannose N-linked oligosaccharides. The common pentasaccharide core (Man $\alpha$ 1-3(Man $\alpha$ 1-6)Man $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc $\beta$ 1-Asn-X-Ser/Thr) is highlighted in green. Complex N-glycans can have up to six branches initiated by GlcNAc and each can be elongated with Gal $\beta$ 1-4GlcNAc (LacNAc) repeats.

As an indication of the structural diversity found in N-glycans, nowadays the CarbBank database lists more than 1000 distinct N-glycan structures (42).

O-glycans oligosaccharides consist of a residue (usually N-Acetylgalactosamine, less commonly, galactose, mannose or xylose form O-glycosidic), attached to a hydroxyl group of Serine or threonine. O-linked oligosaccharides are generally short (1-4 sugar residues), although some oligosaccharides such as o-glycans of ABO blood group antigens can be longer. Currently, the O-linked glycans are grouped into seven core subtypes, with the cores 1 and 2 (figure 5) as the most commonly found in mucins and glycoproteins.

Glycosaminoglycans are long linear polysaccharides consisting of repeating disaccharide units (either GlcNAc or GalNAc) covalently attached to a serine residue. Glycosphingolipids oligosaccharide usually attached via Glc or Gal to the terminal primary hydroxyl group of the lipid moiety ceramide. This basic unit can be further extended by the linear addition of further monosaccharides (Gal, GlcNAc, and GalNAc).



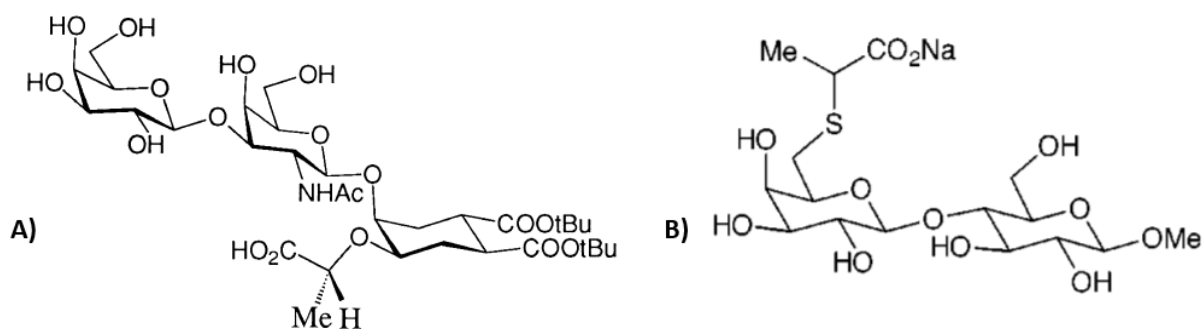
**Figure 5.** Structure of O-linked core glycans. Core 1 and Core 2 O-glycan subtype are the most commonly found formation in glycoproteins and is controlled by the activity of Core 1 Gal-T and Core 2 GlcNAc-T enzymes.

## 1.4. Development of carbohydrates as therapeutic agents

### 1.4.1. Glycomimetic drugs

The key role played by protein-carbohydrate interactions in various diseases processes and the technological advances in glycobiology have increased the interest in the development of chemical entities that mimics the biological essence of carbohydrate (glycomimetics) as pharmaceutical agents. In general, native carbohydrates lack the properties necessary for efficacious drugs and historically have not been successful candidates to capitalize on these applications (43). However, recent advances in analytical techniques and in the understanding of the bioactive conformation and molecular interactions of functional carbohydrates allowed the rational design of small molecule glycomimetics that exhibit improved drug-like properties such as increased affinity, serum half-life, stability, and bioavailability, opening the door to a new class of therapeutic drugs to target molecular mechanisms that can address many of the current unmet needs in the treatment of disease.

For example, the sialylmimetic disaccharide illustrated in figure 6a exhibit micromolar affinities toward the cholera toxin (44) and the lactose-based sialylmimetic disaccharide illustrated in figure 6b has been found to inhibit the human rotavirus (45).



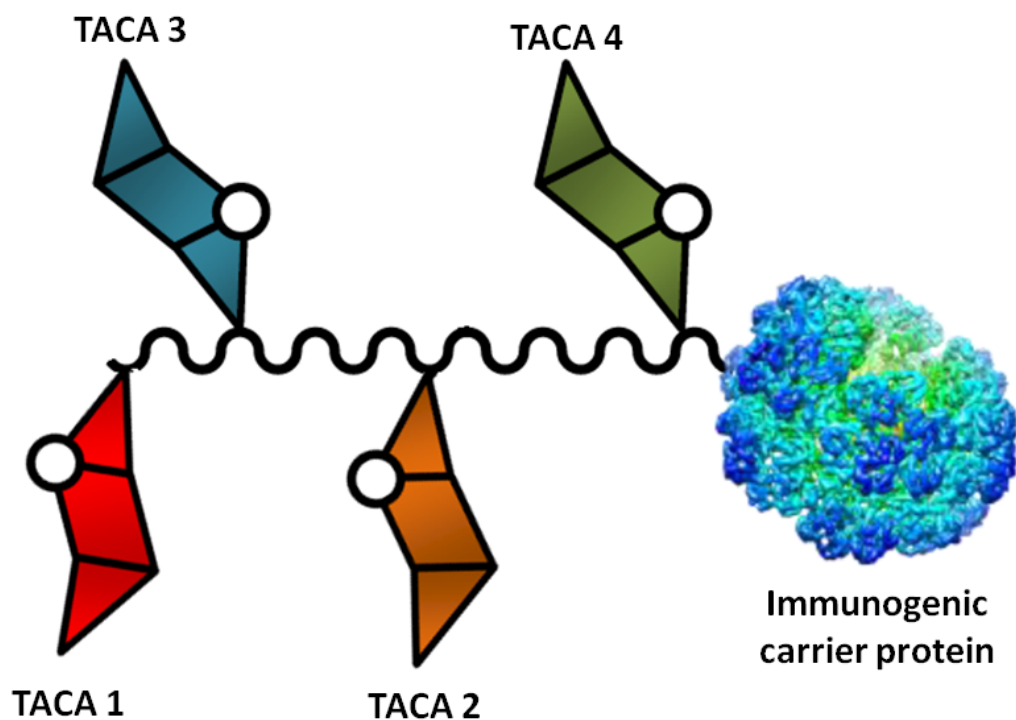
**Figure 6.** Example of glycomimetics with inhibitory effects towards cholera (A) and human rotavirus (B).

Another example of the importance of glycomimetics is the growing importance of pharmaceutical chaperone therapy for lysosomal storage disorders. Chaperones are active-site-specific molecules that act as a folding template in the endoplasmic reticulum to facilitate folding of mutant proteins, thereby accelerating their smooth escape from the endoplasmic reticulum associated degradation to maintain a higher level of residual enzyme activity. This approach has been successfully used to treat Fabry disease patients, by using the iminosugar 1-Deoxygalactonojirimycin (DGJ) to increase residual  $\alpha$ -galactosidase A level in patients suffering from the disease (46).

#### 1.4.2. Carbohydrate-based vaccines

Vaccines are the most powerful and cost-efficient way to prevent and eliminate human infectious diseases. However, vaccines against most human parasites including malaria, leishmaniasis, and schistosomiasis are either still inexistent or of limited efficacy (47) (48). It is therefore important to improve the understanding of the detailed molecular interactions between the host cells and the parasites to develop new vaccines candidates and improve the control of infectious diseases.

Oligosaccharides are playing a growing role in the development of carbohydrate-based vaccines. For example, human tumor cells are known to have an overexpression of unique oligosaccharides on their surface, named “tumor associated carbohydrates antigens” (TACA). TACAs are barely expressed on normal tissues due to their tumor-specificity (49), and have therefore the potential to act as biomarkers of specific tumor cells and are considered as promising targets for the design of anticancer vaccines (50) (51). Unfortunately, the poor ability of carbohydrates to provoke an immune response in human bodies in the human body are challenging to overcome. The main design of TACA-based antitumor vaccine is therefore to conjugate TACAs with a protein or peptide carrier containing established T-cell epitopes (52) (53) and incorporate multiple TACA domains to address the their heterogenicity (figure 7).



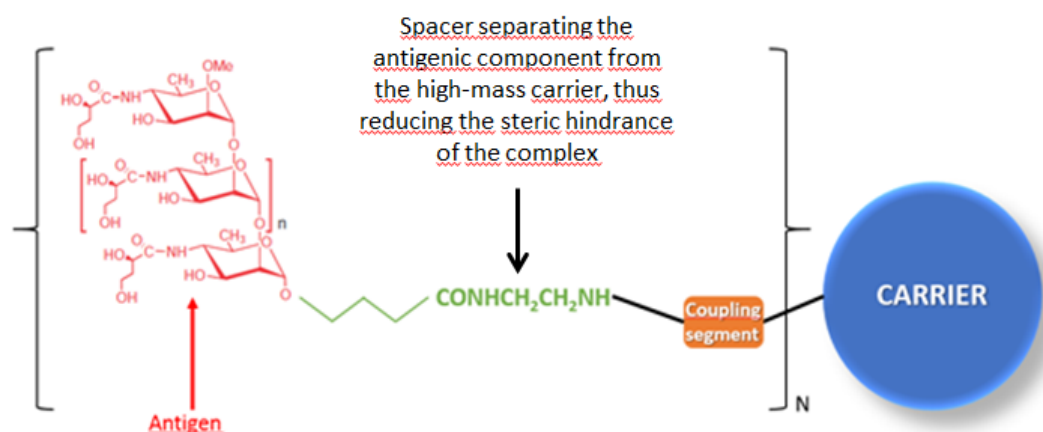
**Figure 7.** Design consideration of tumor-associated carbohydrate antigen (TACA) based antitumor vaccines. The branched TACA-based construct is coupled to immunogenic elements, such as carrier proteins. Covalent linkage of B-cell antigens (TACAs) to carrier proteins promotes initiation of T-cell-dependent pathways and the formation of IgG antibodies against the B-cell epitopes.

For example, Globo H antigen, an hexasaccharide expressed on tumor cells, has been conjugated to monophosphoryl lipid A (MPLA) and showed significant beneficial results for breast and prostate cancer patients (54).

Gram negative bacteria such as Cholera are another good example of target for carbohydrate-based vaccines. Indeed, these bacteria are presenting lipopolysaccharides (LPS) on their outer membranes. LPS consist of three parts: The lipid A, the core, and O-specific polysaccharide (O-SP). The O-SP provide the serotype specificity to the molecule and interacts with the receptor/binding sites of elements of the immune system when the organism is challenged by bacterial pathogen (55) (56). Therefore, immunity is often provided by antibodies that recognize the O-SP part of the LPS.

Vaccines exploiting the key immunogenic role of the O-SP part in the Cholera's LPS have been developed, by combining the antigen with a high mass protein carrier as described above. A schematic representation of the general structure of carbohydrate–protein constructs from oligosaccharide fragments of the O-SP of cholera O1 is shown in Figure 8 (57).

Such vaccine has successfully been shown to induce memory response and protection against Cholera (58) (59).



**Figure 8.** Schematic representation of the structure of the carbohydrate-based vaccine against *V. cholerae* O1, serotype Ogawa. The coupling segment varies depending on the conjugation chemistry. The carrier is usually a medically acceptable immunogenic protein.

X-ray crystallography is commonly used to determine the structure of these protein-antigen complexes, an effective but complex and time-consuming method. However, there is a need for methods enabling the rapid molecular determination of glycomimetics and chaperone binding sites to their corresponding targets, with high sensitivity and low requirements in sample purity (60).

### **1.5. SPR biosensor analysis for kinetic determination of protein-carbohydrate interaction**

Kinetics measurements enable the determination of specific thermodynamic and energetic parameters of protein-carbohydrate interactions (e.g. equilibrium constants, kinetic constants and binding energies). Such determination is often called *affinity measurements*. Most of the analytical methods for affinity measurements rely on a surface at which one of the binding partners is immobilized. Several types of biosensors, such as Surface Plasmon Resonance (SPR), Surface Acoustic Wave (SAW), Quartz Crystal Microbalance (QCM) and Bio-Layer Interferometry (BLI), have been applied in the analysis of protein-carbohydrate, and more generally protein-ligand, interactions.

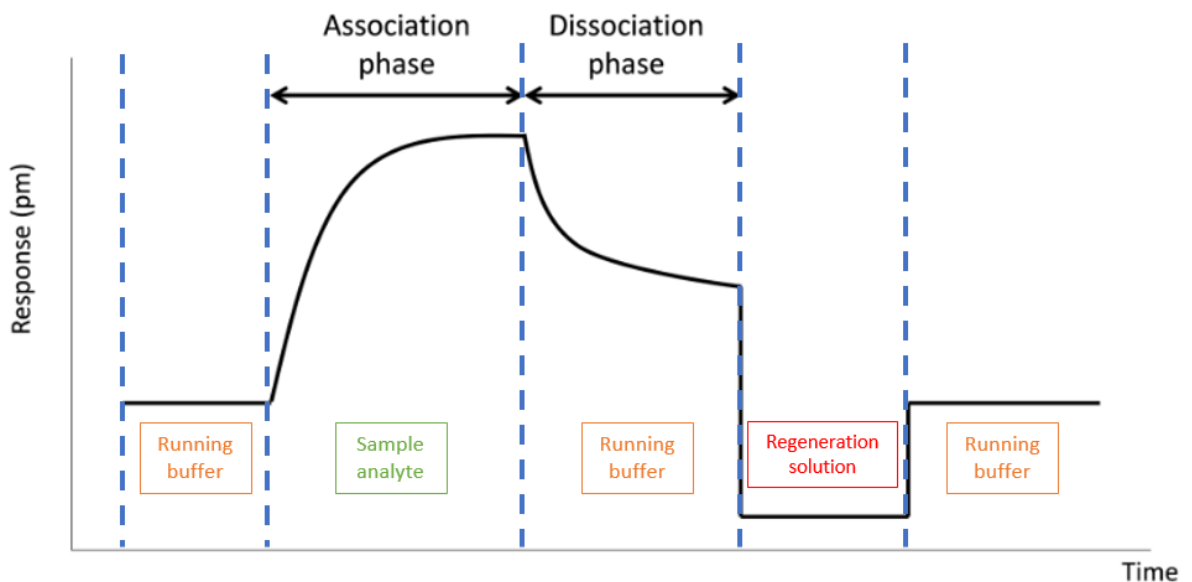
The basic feature of a biosensor is the interaction of an analyte with a biorecognition element (i.e. lectin or glycan), which is in direct contact with a physico-chemical transducer. The main role of the transducer is to change a biorecognition event into a measurable signal, proportional to the analyte concentration (61).

Each of these biosensor techniques has its strengths and weaknesses with regard to the detection approaches. However, SPR and SAW has proven particularly suitable for the study of protein-carbohydrate bindings in numerous studies (62) (63), providing



both quantitative and qualitative interaction data, in real time and under conditions closely mimicking physiological ones.

These instruments rely on the same type of association/dissociation procedure to obtain a binding curve. Typically, a target component (analyte) is captured by a capturing molecule (ligand) previously immobilized on the surface, giving rise on a measurable signal. As shown in figure 9, the measurement can be decomposed in several different step: The sensor surface is first equilibrated and stabilized with a suitable buffer solution. Then, the free analyte is injected over the surface in an aqueous solution, and is accumulating on the surface under continuous flow, until it reaches an equilibration state, where the amount of analytes associating to the surface is equal to the ones dissociating from it. Next, the buffer is flown over the surface and the non-specifically bound components are flushed off, and dissociation of the analyte from the complex takes place and leads to a decrease of the signal response. Finally, in order to prepare the surface for the next injection, an acidic and/or organic regeneration solution is injected, which leads to complete dissociation of the complex.



**Figure 9.** Association and dissociation curve. A protein binds to the covalently immobilized ligand during sample injection, resulting in an increase in signal (response). At the end of the injection, the sample is replaced by a continuous flow of running buffer, and the decrease in signal reflects dissociation of the protein. The surface is then regenerated with a strong eluting solution.

Appropriate interpretation of the association and dissociation data displayed by the sensorgram can then provide quantitative information on active concentration of molecule in the sample, as well as on the affinity kinetics of the interaction.

Most sugar-protein kinetics studies rely on lectin immobilization, in order to compare easily the affinity of a series of different glycoconjugates (64) (26), and because sugar immobilization requires demanding procedures and least to accessibility issues of the lectin to the immobilized sugar (65).

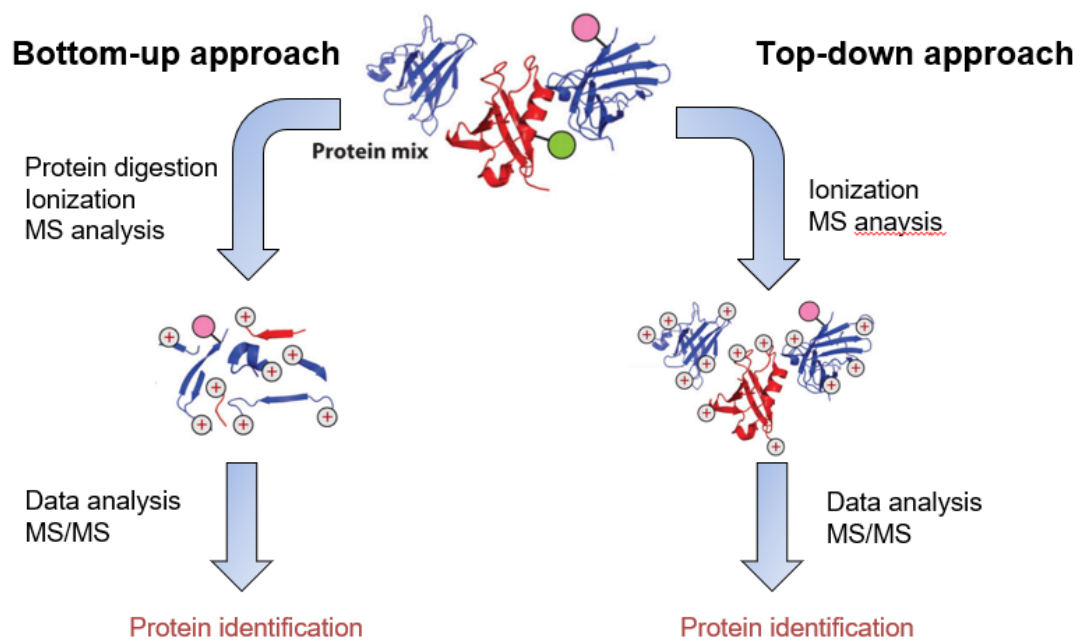
Electrochemical lectin biosensors have been applied for the analysis of viral glycoproteins, P-glycoprotein (responsible for multidrug resistance of cancerous cells), carcinoembryonic antigen (CEA, colorectal cancer),  $\alpha$ -fetoprotein (different types of cancer) and intact cancerous cells (66) (67) (68).

An effective sugar immobilization would allow the identification of novel carbohydrate-binding proteins, and would facilitate profiling studies in array type settings, for the study on complex carbohydrate interactions with bacterial adhesins, viral agglutinins or human lectins (69). Vila-Perello et Al (70) investigated the use of a short 6 amino-acid peptide tag as a linker between the glycan and the gold surface to facilitate the orientation and the accessibility of the sugar moiety with great results, providing alternative to the preparation of oligosaccharide microarrays based on neoglycolipid technology (71). Such peptide-linker has been used in the present work for the investigation of the protein-carbohydrate thermodynamics.

## **1.6. High-pressure proteolytic methods as a new tool for enzyme hydrolysis**

The strategies used to prepare proteins samples for structure characterization by MS involve many steps. The two main approaches to identifying and characterizing proteins using MS are the “bottom-up” and “top-down” methodologies (Figure 10).

In top-down proteomics, the proteins are generally isolated in an ion trapping mass spectrometer ion for mass measurement and then further studied using tandem mass spectrometry (MS/MS) analysis or other protein purification methods such as two-dimensional gel electrophoresis in conjunction with MS/MS. However, top-down proteomics faces many technical challenges for large proteins and virtually all membrane proteins, since they require a detergent such as sodium dodecyl sulfate (SDS) for their solubilization, which is not compatible with the use of ESI and it can therefore not be applied in top-down proteomics. In addition, a protein's tertiary structure becomes more difficult to disrupt with the increase of the protein's molecular weight, meaning most top-down applications are generally limited to proteins of less than 50 kDa.



**Figure 10.** Schematic illustration of the difference between top-down and bottom-up proteomics. In bottom-up proteomics (left), the proteins are subjected to proteolytic digestion prior analysis of the resulting peptide mixture by MS and MS/MS. In top-down proteomics, proteins are directly analyzed by MS. Subsequently, a specific proteoform can be isolated and fragmented by MS/MS to obtain sequence information, which can be used to identify the protein via database searching.

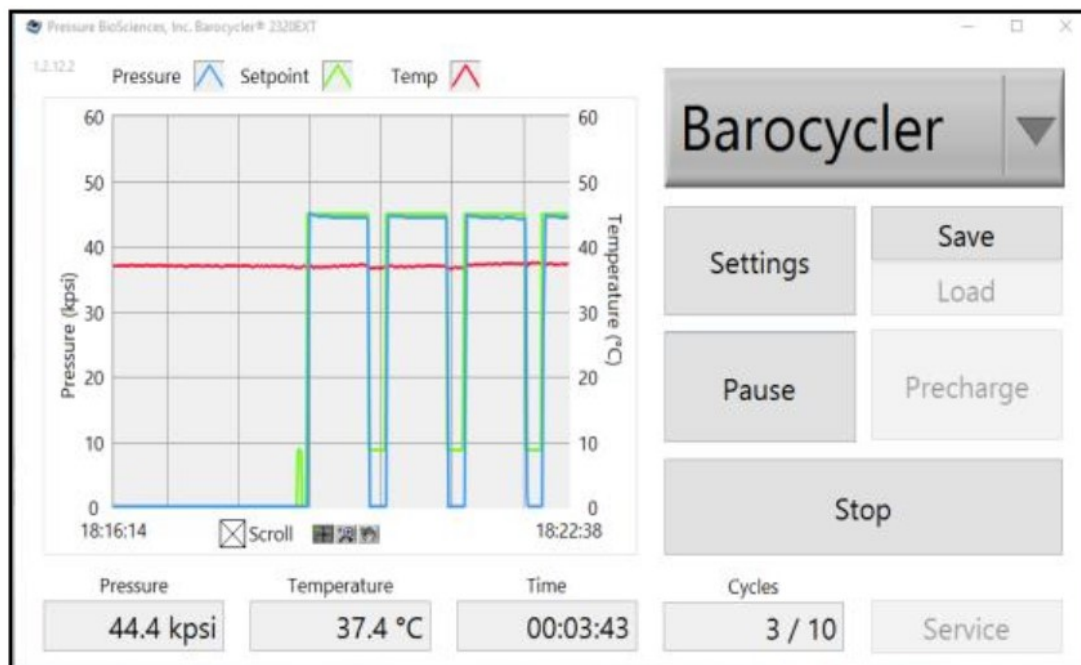
In bottom-up proteomics, the protein is first broken up into peptides, either by chemical or enzymatic digestion, prior to MS analysis. Chemical processes, including alkaline or acid hydrolysis, tend to be difficult to control, and yield products with modified amino acids. Enzymatic hydrolysis however, can be performed under mild conditions, and could avoid the extreme environments required by chemical treatments. Additionally, enzymes present substrate specificity which permits the development of protein hydrolysates with better defined chemical characteristics. On the other hand, total protein hydrolysis is sometime challenging, especially for proteins with compact three-dimensional structures, effectively inhibiting the access of the enzyme to the cleavage site.

Optimization of the proteolysis step is of key importance in proteomics, and often rely on time-consuming procedures. Protein digestion has traditionally been performed using serine proteases trypsin or chymotrypsin in a buffered medium over a defined length of time, generally overnight (~12 h), in order to have a complete and accurate proteolysis and for the enzyme to access to all cleavage sites. The rate of digestion strongly relies on several environmental factors such as temperature, solvents used, pH range, and enzyme-to-substrate ratio. Higher hydrolysis temperature generally increases digestion yields, but also destabilize the enzyme and increase autolysis. Modified and thermo-stable trypsin are constantly studied in order to overcome this issue (72). Higher temperatures also cause protein to aggregate in solution and, when denaturing chaotropes solvents such as urea are used, can lead to the formation of chemically modified peptides such as carbamylated peptides. The use of mixed solvent buffers containing organic solvents has also been proven to reduce hydrolysis time and sequence coverage (73). Increasing the enzyme-to-substrate ratio also lead to faster reaction time, but also lead to higher autolysis peptide and higher hydrolysis cost as quality sequence-grade enzyme trypsin are not inexpensive.

More recently, an alternative parameter has been investigated to increase the performance of the established protein hydrolysis procedures. High pressure has been well documented to increase enzymatic activity in food products between 100 to 400 MPa (74).

High pressure can weaken hydrophobic interactions but can enhance electrostatic ones (75), and acts synergistically with chaotropes and detergents leading to protein denaturation. However, pressure-perturbed proteins have been shown to assume conformational forms drastically different from those resulting from thermal or chemical treatment, as cycles of hydrostatic pressure can disrupt protein secondary, tertiary and quaternary structures and alter their conformation, thereby exposing otherwise hidden peptide sequences to proteolytic cleavage (76). Previous experiments performed by applying high hydrostatic pressure during enzymatic digestion of  $\beta$ -lactoglobulin has shown to increase the hydrolysis rate without significant qualitative changes in the peptic peptide profiles produced (77).

This premise has been applied for pressure cycling technology (PCT). PCT uses repeated cycles of hydrostatic pressure between ambient and ultra-high levels (up to 45,000psi) during proteolytic digestion, in order to facilitate the thermodynamic perturbation of molecular interactions. The pressure, time hydrolysis, temperature, and number of cycles can all be parameter through its user interface (figure 11).



**Figure 11.** Example of the Barocyler user interface. The ongoing hydrolysis parameter are shown in the graph. In red: The hydrolysis temperature. In green: The pressure requested in the parameters. In blue: The actual pressure measured.

PCT has been shown to drastically increase hydrolysis speed and recovery and improve sequence coverage, even for hydrolysis-resistant proteins such as highly folded proteins and lectins (78) (79).

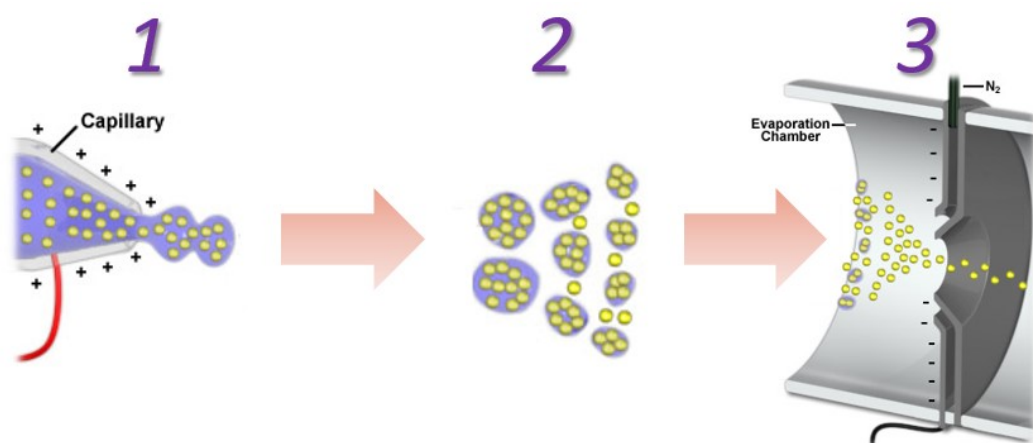
### **1.7. Mass-spectrometric methods for protein interactions analysis**

Most mass-spectrometry based methods for the identification of peptides and protein rely of soft-ionization methods such as electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI) (80) (81) (82). MADLI and ESI methods are considered “soft” ionization methods because the low energy applied causes minimal or no fragmentation and allows the molecular ions of analytes to be identified, even in complex mixtures of biopolymers such as cell lysates. These methods allow fast, sensitive and accurate measurements and requires minimal amount of material.

The gentleness of the ESI process allows intact protein complexes to be directly detected by mass spectrometry, effectively preserving the non-covalent interactions present in solution (83). The ionization technique produces gas-phase ions from analytes originally present in solution. These ions are sprayed from a small tube into a strong electric field in the presence of a flow of warm nitrogen to assist desolvation. The generated droplet size can vary depending the viscosity and the conductivity of the sample. Typically, addition of formic acid in the liquid sample will increase the conductivity and therefore reduce the droplet size.

The charged droplets are then sampled into the first vacuum stage of a mass spectrometer through a capillary carrying a potential difference of approximately 3000 to 6000V. By applying an elevated source temperature and/or another stream of drying

nitrogen gas, the solvent is evaporated, thus resulting in continuously reduced size of the charged droplets, leading to an increase of surface charge density and a decrease of the droplet radius. When the electric field strength around the droplet becomes greater than the surface tension, the droplet undergoes Coulomb fission, whereby the original droplet splits into smaller, more stable droplets (figure 12) (84) (85).



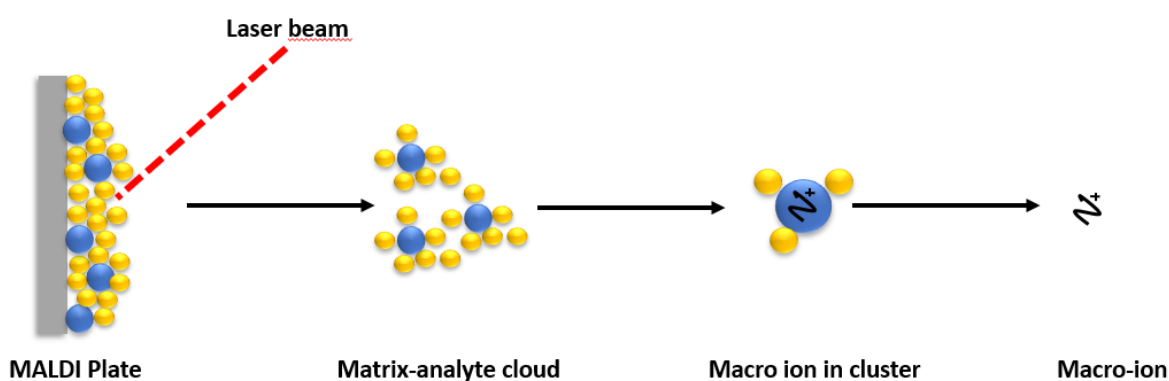
**Figure 12.** Illustration of the electrospray ionization. (1) Formation of liquid drops spray under high voltage in the capillary (2) Due to the large surface-area / volume ratio of the charged sample droplets, and the nitrogen gas pumped into the chamber, the solvent evaporates from the droplet and the surface tension is increasing (3) When the electrostatic repulsion exceed the surface tension, the droplet dissociates, leaving a stream of charged ions into the analyzer.

ESI generates mainly protonated ( $[M+nH^+]^{n+}$ ) or deprotonated ( $[M-nH^+]^{n-}$ ) molecules, and the number of charges depend on the molecular size and characteristic of the biomolecule measured. In positive-ion mode, the charges come mainly from protons, due to the acidic pH of the solvent and various electrochemical reactions, and from other ions naturally present in the sample, such as  $Na^+$ ,  $K^+$ , etc. The number of charges is also influenced by the number of solvent-exposed basic amino acid residues of peptides and proteins and their unmodified N-terminus (86). In negative-mode however, the ionization occurs by deprotonation of acidic amino acids and of the free C-terminus.

MALDI ionization involves mixing and co-crystallizing the sample with a large amount of non-volatile matrix, such as  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) or 2,5-dihydroxybenzoic acid (DHB). The preparation is generally done in a mixture of organic solvent and water in order to allow both hydrophilic and hydrophobic molecules to dissolve. A counter-ion, such as trifluoroacetic acid (TFA) is generally added to help the protonation of the analyte. The mixture is then spotted on a metal plate and the solvent is left to dry, leaving the crystallized matrix embed with the analyte.

Short laser pulses at 337 nm wavelength are then shot on the analyte-matrix mix. The matrix absorb energy at the wavelength of a laser, causing desorption-ionization of the sample mixture (figure 13), and accelerating the analyte ions into the mass analyzer by applying a high potential electric field between the sample and the orifice of the analyzer (81).

Like ESI, attachment of alkali metal ions such as  $\text{Na}^+$  or  $\text{K}^+$  can be observed during positive ionization of the analyte molecules (87). However, ions produced by MALDI are most of the time singly charged (88). Multiple charges can be observed for large biomolecules (89), yet not to the same extent as ESI.



**Figure 13.** Schematic representation of the MALDI process. Laser irradiation causes the Desorption-ionization of the matrix (in yellow) and analyte (in blue) mixture from the stainless steel MALDI target which transfers protons to the analyte. Analyte ionization occurs in the resulting gas cloud.



Time-of-flight analyzer, such as the one used for this work, are the most commonly used analyzer for MALDI, mainly due to its large mass-range, allowing to measure the singly charged high mass biomolecules. Ions are accelerated by an electric field, entering the analyzer with a define kinetic energy. They are then separated according to their mass-to-charge ratio with the help of a reflection, which uses a constant electrostatic field to reflect the ion beam toward the detector, allowing more energetic ions to penetrate deeper into the reflectron, and take a slightly longer path to the detector.

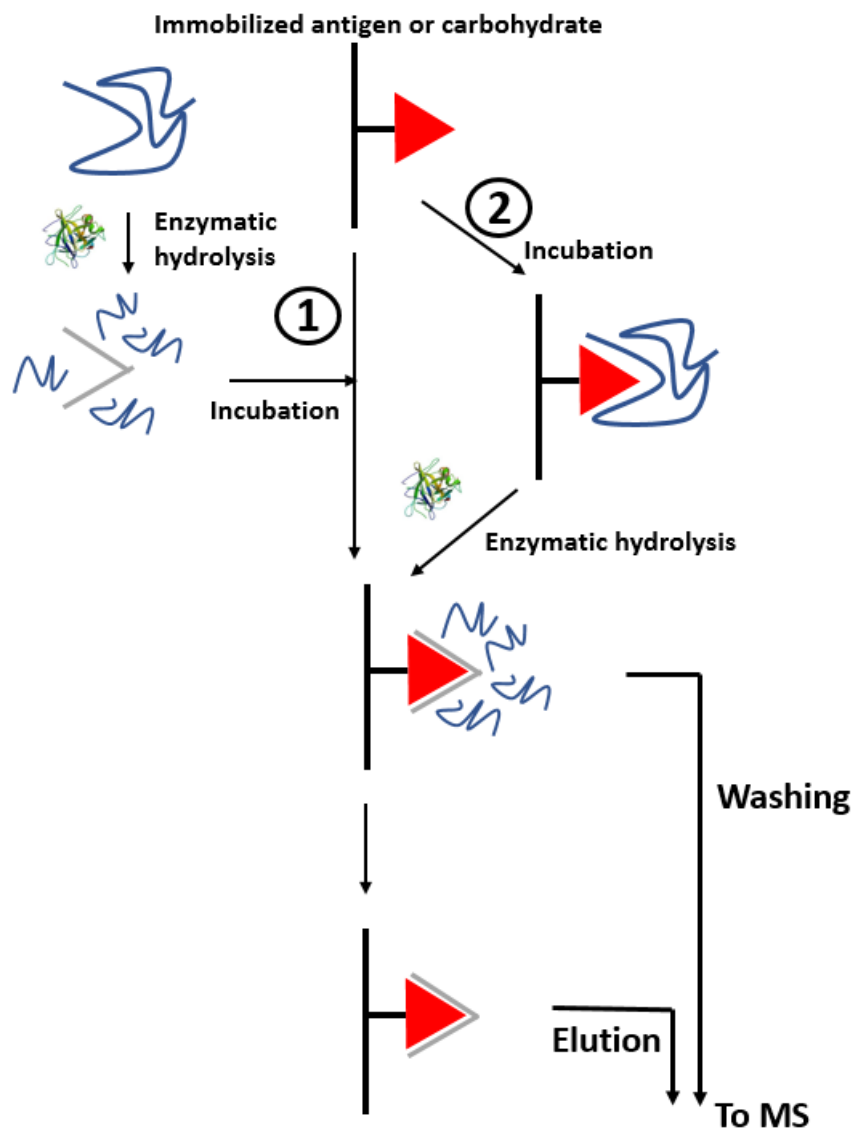
Mass-spectrometry not only allow an accurate measurement of molecular weight, but can also deliver structural information for a wide range of biomolecules. Moreover, complete amino-sequence information and location of post-translational modification can be obtained with various fragmentation techniques through tandem mass spectrometry (MS/MS). When using MS/MS, ions of a selected mass-to-charge ratio are selected after the first stage of MS and fragmented in the gas phase by colliding with neutral molecules (generally helium, argon or nitrogen). This fragmentation technique, named collision-induced dissociation, leads to breakage of peptide bonds and allow the characterization of the exact sequence of a peptide (90).

### **1.8. Affinity-Mass-Spectrometry for protein-ligand structure and interaction analysis**

A growing interest in bioaffinity-mass spectrometry has been observed in recent years. Such combination of methods has been shown to be a powerful tool for the identification of epitopes for protein-carbohydrates interactions as well as for paratopes in protein-protein and protein-antigen complexes.

The proteolytic excision and extraction methods (PROTEX) combined to mass spectrometry has been first developed by our laboratory (91) and is now a standard method for the identification of biomolecular recognition structure for many groups with successful application such as the identification of the antibody epitope of human  $\alpha$ -galactosidase or carbohydrate-recognizing epitopes (92) (93) (94) (95) (96) (97). This approach combines the advantages of the proteolytic stability of antibodies and antigen epitopes by shielding with affinity methods with the powerful identification tool that is mass spectrometry. The general approach of such experiment is illustrated in figure 14, and has been adapted in this thesis the recognition of carbohydrate binding sites. In contrast to antigen-antibody, only one interacting partner, the lectin, can be proteolytically digested during carbohydrate-protein interactions, simplifying the identification of the immobilized peptide (epitope), thus highlighting the combination of proteolytic digestion of affinity bound proteins with mass spectrometry as an efficient tool for identification of epitope and paratope structures.

As a first step of the PROTEX experiment, an affinity column is created by immobilizing a carbohydrate on an affinity matrix such as Sepharose microcolumn and incubated with a carbohydrate-recognizing lectin. Sepharose is an agarose-based, bead-formed matrix, containing 2, 4, or 6 % agarose, hence designated Sepharoses 2B, 4B, or 6B. In addition to sepharose, other microspheres matrices can also be used to immobilize carbohydrate and are commercially available (Mini-leak agarose, Synsorb, Seralose, and Spheron besides others), either in unmodified form, CNBR, epoxy or CH-activated form or with pre-immobilized carbohydrate (98; 99; 100). For immobilization of small carbohydrates, it is usually necessary to introduce a spacer between the ligand and the matrix, to allow accessibility to the lectin's binding site. The spacer may be linked to the matrix before immobilizing the carbohydrate or the carbohydrate may be converted into a glycoside by derivatizing it with the spacer. After the carbohydrate immobilization, blocking of any remaining active groups on the matrix is performed with the use of small organic molecules (e.g. ethanolamine).



**Figure 14.** Scheme of proteolytic extraction (1) and proteolytic excision (2) methods for epitope identification, exemplified on an immobilized antigen or carbohydrate. First, the antigen or carbohydrate is bound to an affinity matrix. In the proteolytic extraction method, the sample is then digested with various proteases prior incubation on the immobilized carbohydrate/antigen. In the secondary method, proteolytic excision, the sample is incubated with the immobilized carbohydrate/antigen before enzymatic hydrolysis. Unbound fragments are subsequently removed by washing and analyzed by MS, until no signal can be observed. Affinity-bound epitope peptides are then dissociated from the immobilized carbohydrate/antigen with an adequate solvent and the elution fractions are analyzed by MS and compared.

A selective enzyme hydrolysis is then carried out on the lectin (proteolytic excision approach). The epitope fragment recognizing the immobilized carbohydrate will remain bound to the carbohydrate, while the other non-bound peptides can be washed away and collected for mass-spectrometric analysis. Finally, the carbohydrate-lectin epitope complex is dissociated with the use of an acidic or organic solvent, eluted, and identified by mass-spectrometry.

Alternatively, the lectin can be proteolytically digested prior to immobilization on the carbohydrate-Sepharose column, and the resulting peptide mixture is then incubated on the micro-column (proteolytic extraction). Due to the specificity of the carbohydrate-lectin interaction, only lectin epitopes would bind to the Sepharose-carbohydrate column. The non-bound peptides are then washed and epitopes are then eluted similarly as the proteolytic excision experiment, using either organic solvents or a competitive carbohydrate solution.

These approaches, when applied to lectin-carbohydrate interaction, have been named “carbohydrate recognition domain excision/extraction” (CREDEX) (97). One of the main advantages of the CREDEX methods of other methods used for interaction recognition such as hydrogen-deuterium exchange is the low sample consumption and the straightforward analysis procedure by MS, allowing a lower analysis time. Moreover, this method brings the possibility to isolate the epitopes-peptide sequences that are relevant to the carbohydrate-recognition domain of the lectin. These peptides can then easily be chemically synthesized and can be used as model system for biomedical application.

One of the major concerns however, when using the CREDEX method for the identification of lectin epitope is the lower affinity between lectins and carbohydrate compared to protein-protein or protein-antibody interactions. Epitope peptide of lectin would typically have interaction strengths in the millimolar range (15) (4). However, there are records of synthesized and natural occurring peptides that can bind to carbohydrates. Yamamoto et Al. (101) synthesized an 9 amino-acid peptide from *Bauhinia purpurea* lectin (BPA) that have shown to exhibit lactose binding affinity in the presence of calcium. Defensins, a naturally occurring cysteine-rich peptides that participate in both vertebrate and invertebrate host defense, are another example of

host defense peptide that can bind to complex glycans present on the surface of bacteria and viruses with affinities comparable with those of lectins (102).

Protein-ligand interactions may also be studied by other techniques such as chemical crosslinking coupled to mass spectrometry (CXMS) (103) (104) or hydrogen-deuterium exchange of backbone hydrogens with mass spectrometry (HDX-MS) (105) (106). CXMX works through covalently binding a linker between two amino acids in proximity to each other. These amino acids are constrained by the length of the linker used. Digestion of the protein followed by LC-MS/MS gives information relating to the distal restraints set by the linker. This allows for further insight into protein folding and how proteins interact with each other, which in turn can be used to refine or create *in silico* models of the protein of interest. The main drawback of this method is the difficulty of data analysis of the many types of chemical structures that form, such as unmodified peptides, mono-linked peptides (only one end of the cross-linker will react with the protein while the other is deactivated), intra-peptide crosslinks, regular crosslinked peptides and higher order crosslinked peptides (107). Other drawbacks include the modification of lysine residues by amine-reactive crosslinking reagents, which results in the blocking of tryptic cleavage sites and the possible creation of very large crosslinked peptides which may be difficult to analyze (108).

HDX-MS on the other hand, a protein is exposed to a solvent containing D<sub>2</sub>O, and regions of the protein that are not involved in hydrogen bonding exhibit a burst phase exchange kinetic, whereas regions that are in the interior of the protein or involved in hydrogen bonds, exchange at much slower rates. The differences in this exchange process can occur over seconds to hour timeframe and is an indicator of the local structure. The Deuterium is then frozen or "quenched" at low temperature and pH and the location of the Deuterium uptake can be detected accurately using high-resolution mass spectrometry. To minimize the inevitable back-exchange, both proteolysis and MS must be carried out quickly and at low temperature. However, the major drawback of HDX-MS is that it requires significant time and effort for sample preparation, management of back-exchange and assignment of exchange rates to single amino acid residues, based on the data obtained for the peptic fragments.

## 1.9. Scientific goal of the dissertation

Carbohydrates belong to the most abundant natural products, and are fundamental constituents of cell surfaces and involved in vital cellular recognition processes through protein-carbohydrate interactions.

Such interactions are ubiquitous in nature, and set carbohydrates as relatively untapped source of new drugs, offering therefore exciting new therapeutic opportunities. Indeed, advances in the functional understanding of carbohydrate–protein interactions have enabled the development of a new class of small-molecule drugs, known as glycomimetics.

The characterization of these interactions is also a crucial step in understanding the mechanisms by which lectins, anti-carbohydrate antibodies and enzymes exert their functions.

Identification of the binding sites of oligosaccharides to proteins – the epitope - offers new insights to disease diagnostics by using their capacity to recognize carbohydrate moieties on antigens, and bring leads to the developments of therapeutics such as chaperones molecules for the treatment of lysosomal storage disorders.

Analytical techniques such as X-ray crystallography and NMR are the standard methods used to identify carbohydrate-recognition domains, but these approaches are limited by their high requirements in sample purity and amount, are time-consuming and obtaining satisfactory resolution is challenging. Other methods such as HDX-MS and CX-MS are of growing relevance but demand high data treatment, and are time consuming or sample consuming procedures.

The major goal of this dissertation is the identification and affinity characterization of carbohydrate interaction epitopes. To this goal, techniques for immobilization of carbohydrates and for high pressure protein digestion were developed. Within these goals, the following major objectives were pursued:

- 1. Development of a new immobilization procedure for carbohydrates on surface acoustic wave biosensor chips for the determination of dissociation constants.**

Using both divinyl sulphone activation and a peptide carrier, immobilization of carbohydrate and oligosaccharides moieties for characterization of their binding strengths with carbohydrate-recognizing proteins.

- 2. Development of a high-pressure proteolysis method to fasten lectin hydrolysis.**

Development of a pressure-enhanced proteolysis method for a short-time hydrolysis of Glycine max lectin, an enzyme-resistant lectin, with comparison of the rate and efficiency against a standard in-solution enzymatic hydrolysis at atmospheric pressure.

- 3. Identification of glycine max lectin epitopes to bound GalNAc and blood group trisaccharides.**

Analysis of the primary structure of glycine max lectin by pressure-enhanced proteolytic peptide mapping – mass spectrometry. Identification of the discontinued carbohydrates binding sites using the carbohydrate recognition domain extraction method and the to the newly developed pressure-enhanced proteolytic excision method.

- 4. Development of a new affinity column for epitope peptide identification using gold chip immobilized carbohydrate.**

---

Immobilization of carbohydrates on a SAW gold chip via the perfected immobilization procedure, injection of glycine max lectin tryptic peptides obtained with pressure-enhanced proteolysis and capture of the epitope peptides. Following incubation of the chip and extensive washing of the gold chip surface until no peptide can be detected, elution of the epitope peptide following identification by MALDI-MS.



## 2. RESULTS AND DISCUSSION

### 2.1. Structural characterization of lectins

#### 2.1.1. Characterization of Glycine Max isolectins

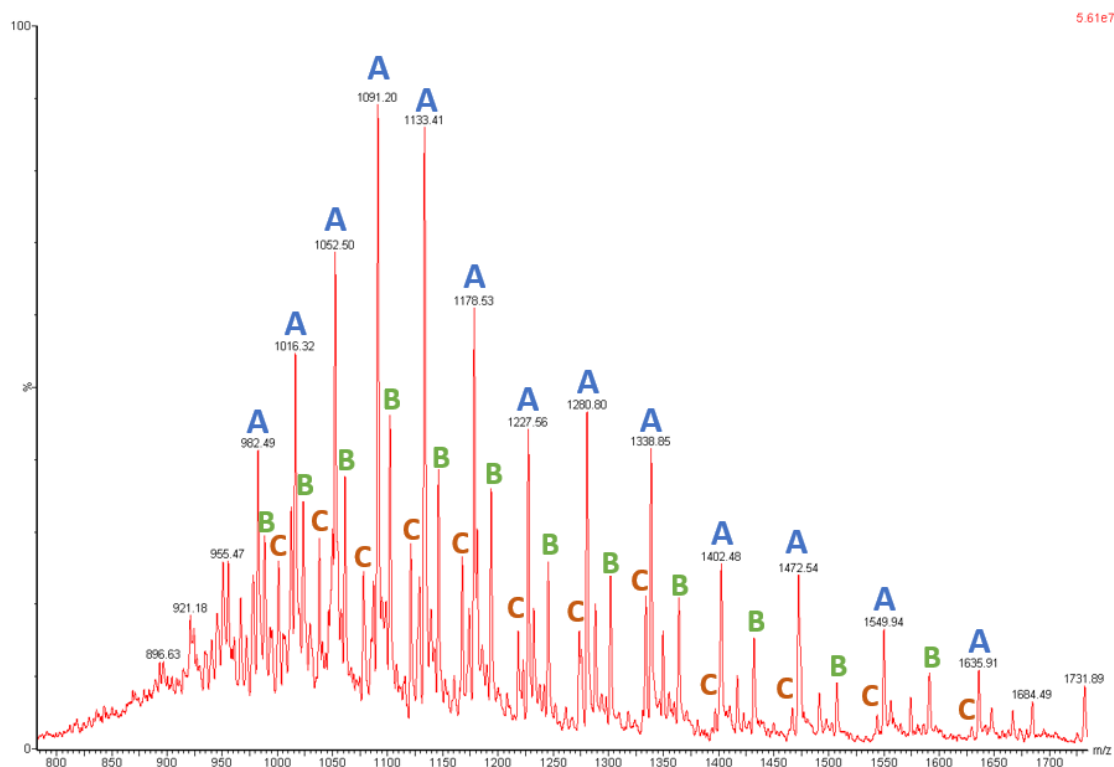
Glycine Max Lectin, also known as Soybean Agglutinin (SBA), is a 253 amino-acid tetrameric lectin with a molecular mass of 27571 Da. The lectin also possesses one glycosylation site consisting of mannose and N -acetylglucosamine units,  $\text{Man}_9(\text{GlcNAc})_2$ , attached to Asn-75 and with a molecular mass of 1866 Da (109). Native SBA also exists as multiple isolectins caused by C-terminal truncation. These isolectins have similar binding and immunochemical properties (110). The complete sequence of SBA and the known C-terminal truncation position is resumed in figure 15. The lectin provided by the company Medicago was therefore analyzed by electrospray mass-spectrometry to identify the isolectins present in the sample and to facilitate the later steps of proteolytic peptide mapping.

1	AETVSFSWNK	FVPKQPNMIL	QGDAIVTSSG	KLQLNKVDEN	GTPKPSSLGR
			$\text{Man}_9(\text{GlcNAc})_2$		
51	ALYSTPIHIW	DKETGSVASF	AASFNFTFYA	PDKRLADGL	AFFLAPIDTK
101	PQTHAGYLGL	FNENESGDQV	VAVEFDTFRN	SWDPPNPHIG	INVNSIRSIK
151	TTSWDLANNK	VAKVLITYDA	STSLLVASLV	YPSQRTSNI	LSDVVDLKTS
201	LPEWVRIGFS	AATGLDIPGE	SHDVLSWSFA	SNLPHASSNI	DPLDLTSFLH
251	EAI				

**Figure 15.** Sequence of the [1-253] SBA protein. The position of the glycosylation site  $\text{Man}_9(\text{GlcNAc})_2$  is indicated with a yellow arrow. Known position of C-terminal truncation of the complete [1-253] sequence is indicated with the lightning symbol.

100  $\mu\text{g}$  SBA was diluted with 200  $\mu\text{L}$  of 40% methanol with 0.1% formic acid and the solution was analyzed by ESI-QMS. The resulting mass spectrum is shown in figure 16.

As the isolectins are issued from the c-terminal truncation of the lectin, identification of the different subunit is possible by analysis of the  $m/z$  observed for the different isolectins detected.



**Figure 16.** Intact SBA measured by ESI-QMS. The different subunits that can be observed with different protonation are labeled A, B and C.

The mass of the subunit A corresponds to the entire [1-253] SBA protein together with the glycosylation mass. The mass of the subunits B and C correspond respectively to the subunits [1-246] and [1-240] with their oligomannose chain. Details of the calculation are shown in table 2. Both isolectins has also been identified from native SBA by Mandal et Al (110). Other subunits can be observed with lower intensities.

The present findings thus demonstrate that isolectins of SBA arise from different combinations of intact and C-terminal fragmented subunits. Moreover, the sequence [241-253] PLDLTSFVLHEAI, absent from the shorter isolectin, do not contain any arginine or lysine residue, thus assuring it will have no influence in the case of a c-terminal epitope.

**Table 2.** Identification of the isolectins from the ESI-QMS mass spectra. The mass of each highest peak has been used for the calculation of the measured mass. The formula  $M=z(m-1)$  has been applied, with M being the mass of the protein, z the charge and m the measured m/z signal.

<b>Man(9) mass (Da)</b>	1866			
<b>Sequence length of the Isolectin</b>	<b>Theoretical average mass (Da)</b>	<b>Theoretical average mass including Man<sub>9</sub>(GlcNAc)<sub>2</sub>(Da)</b>	<b>Measured mass (Da)</b>	<b>Difference (Da)</b>
[1-253]	27572	29438	29436	2
[1-246]	26763	28629	28628	1
[1-240]	26135	28001	28000	1

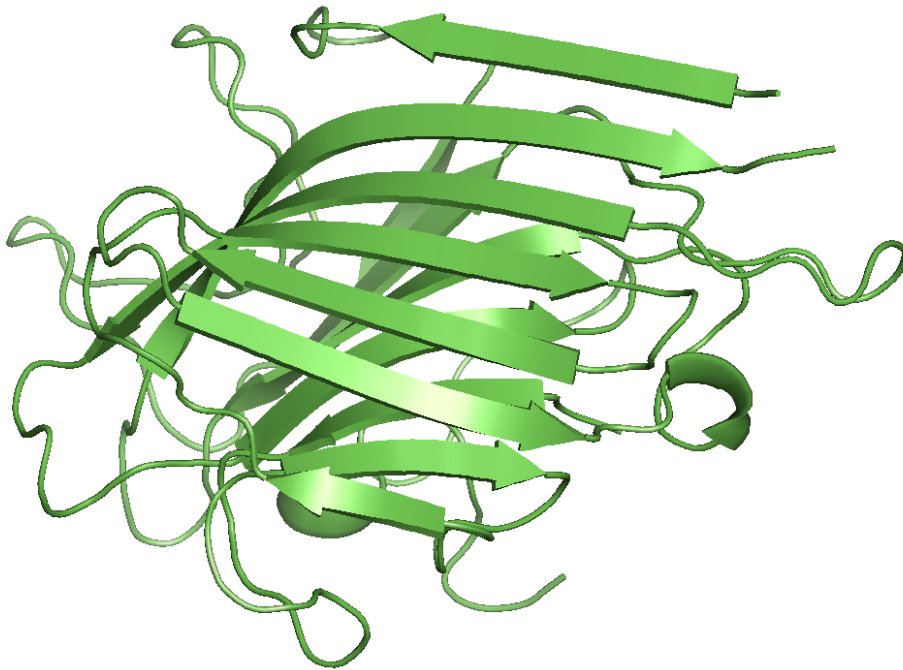
### 2.1.2. Development of an efficient proteolytic procedure for Soybean Agglutinin

Native SBA is resistant to digestive enzymes in the gastrointestinal (GI) tract and binds to the small intestinal brush boarder, causing increased weight of the small intestine and pancreatic hypertrophy (111). In addition, SBA can inhibit the disaccharidases and proteases in the intestines, interfere with the absorption of nonheme iron and lipid from the diet, lower the circulating insulin level and cause degenerative changes in the liver and kidneys (112).

SBA shows high affinities towards carbohydrate-containing molecules, with the highest affinity towards N-acetyl-D-galactosamine (GalNAc) moieties. However, the presence of Ca<sup>2+</sup> and Mn<sup>2+</sup> ions are required for carbohydrate binding activity (113). The structure of the SBA monomer is shown in figure 17.

Different methods have been used to deactivate SBA, with heat treatment being the most commonly used. Dry heat treatment is not as effective as moist heat in deactivating SBA (114).

Here, we investigate an enzymatic hydrolysis method for the digestion of SBA. Several different conditions have been tested in order to maximize the efficiency of the hydrolysis, by testing different buffer conditions, different enzymes and enzyme to substrate ratio, different denaturing conditions and hydrolysis duration. The different hydrolysis conditions tested have been summarized in table 3.



**Figure 17.** Schematic ribbon diagram of the SBA structure.

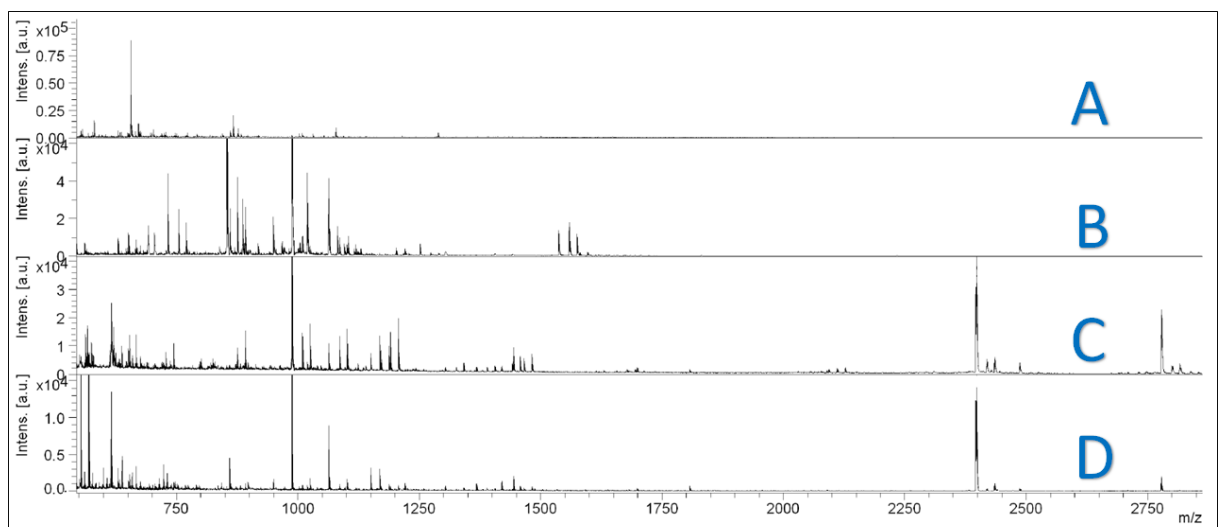
The procedures 1 and 2 yielded close to no tryptic peptides (see appendix 2). The digest yield was controlled by SDS Page over 24 hours. However, results were observed to be slightly better with the use of tris buffer in the procedure 2. Ma Y. et Al observed in their work that single enzymes such as trypsin did not deactivate SBA, and suggesting the use of multiple enzymes and denaturation procedures to increase the efficiency of the lectin deactivation.

The use of multiple enzymes in the procedure 3 confirmed the observations made in the previously mentioned publication, as about 60% of sequence coverage was recovered by proteolytic peptide mapping using a mixture of trypsin and chymotrypsin.

**Table 3.** Procedure details for the development of an effective hydrolysis method on SBA. T: Trypsin. C: Chymotrypsin. E:S : Enzyme to substrate ratio.

Procedure	1	2	3	4	5	6
<b>Denaturation</b>	-	-	-	Urea	Boiling	Boiling
<b>Buffer</b>	100 mM NH <sub>4</sub> HCO <sub>3</sub> , pH 7,8	100 mM Tris, pH 7,8	100 mM Tris, pH 7,8	100 mM Tris, 10 mM CaCl <sub>2</sub> , pH 7,8	100 mM Tris, 10 mM CaCl <sub>2</sub> , pH 7,8	100 mM Tris, 10 mM CaCl <sub>2</sub> , pH 7,8
<b>Enzyme</b>	T	T	T / C	T / C	T / C	T / C
<b>E : S</b>	1:40	1:40	1:40 / 1:20	1:40 / 1:20	1:40 / 1:20	1:40 / 1:20
<b>Incubation</b>	37°C	37°C	37°C	37°C	37°C	50°C, 40kpsi

The impact of different denaturation procedures on the lectin's digestion yield was measured by using the procedure 3 after denaturing the lectin by heating (95°C, 20min. Procedure 5) or by chemical denaturation using urea. Both procedures have shown a complete lectin hydrolysis by SDS-Page, with sequence coverage of about 90%. All enzymatic procedures 1 to 6 were stopped after 12 hours of incubation by adding 1-5  $\mu$ L TFA. They were then cleaned and concentrated using Ziptip, and measured by MALDI-MS (figure 18).



**Figure 18.** MALDI-MS measurement of aliquots of the 12 hours SBA digests using the hydrolysis methods described in table 2. A: Method 2 (100 mM Tris, pH 7,8, trypsin ration

1;40, 37°C). B: Method 3 (100 mM Tris, pH 7,8, trypsin ratio 1;40, Chymotrypsin ratio 1:20, 37°C). C: Method 4 (urea denaturation, 100 mM Tris, 10 mM CaCl<sub>2</sub>, pH 7,8, trypsin ratio 1;40, Chymotrypsin ratio 1:20, 37°C). D: Method 5 (high temperature denaturation, 100 mM Tris, 10 mM CaCl<sub>2</sub>, pH 7,8, trypsin ratio 1;40, Chymotrypsin ratio 1:20, 37°C).

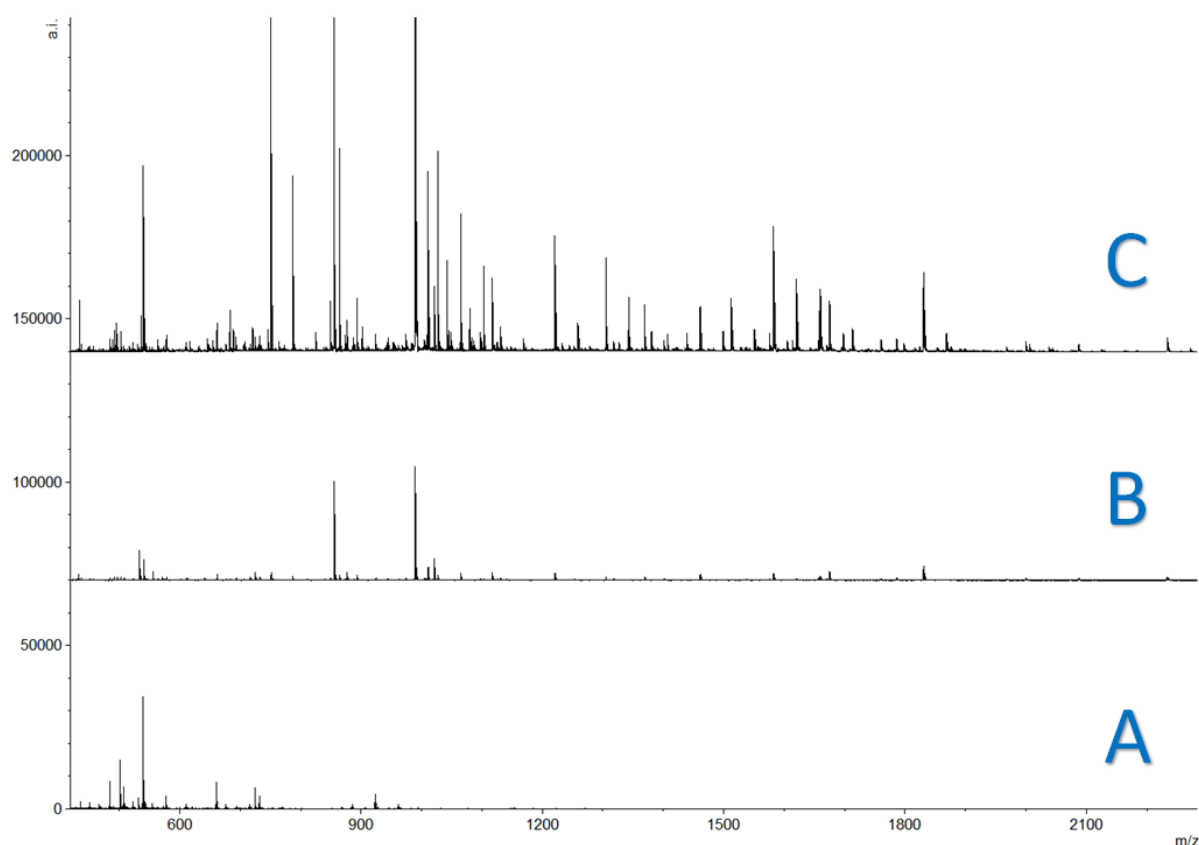
As we could observe carbamylations at the N-terminal of peptides and at the side chain amino groups of lysine and arginine residues caused by the use of urea in procedure 4, the boiling denaturation procedure 6 was judged more satisfactory and less time consuming for further experiments.

Since the incubation time needed to obtain a complete lectin digest is still high with the procedure described below, a barocycler was used in order to diminish the total procedure time. Indeed, alternating high hydrostatic pressure has been shown to facilitate thermodynamic perturbation of molecular interactions and increase hydrolysis speed, recovery and improve sequence coverage (75) (79). Several pressure conditions (25, 30 and 40 kpsi) as well as temperature (30°C and 50°C) have been tested with an incubation time of 2 hours.

The pressure cycles have been of 50 seconds at high pressure, and 10 seconds at atmospheric pressure.

The most efficient hydrolysis conditions were observed for 40 kpsi cycles at 50°C. In order to control the necessary incubation time, aliquots of digest have been sampled over the incubation time and were cleaned and concentrated as above and measured by MALDI-MS.

The results, shown in figure 19, shows the peptide mapping results after 5 minutes (A), 30 minutes (B) and 60 minutes (C). It can be seen that 60 minutes under digesting conditions 6 in the barocycler is enough to completely hydrolyze SBA and to obtain sequence coverage close to what is obtained for 12 hours of the same digestion procedure at room temperature.



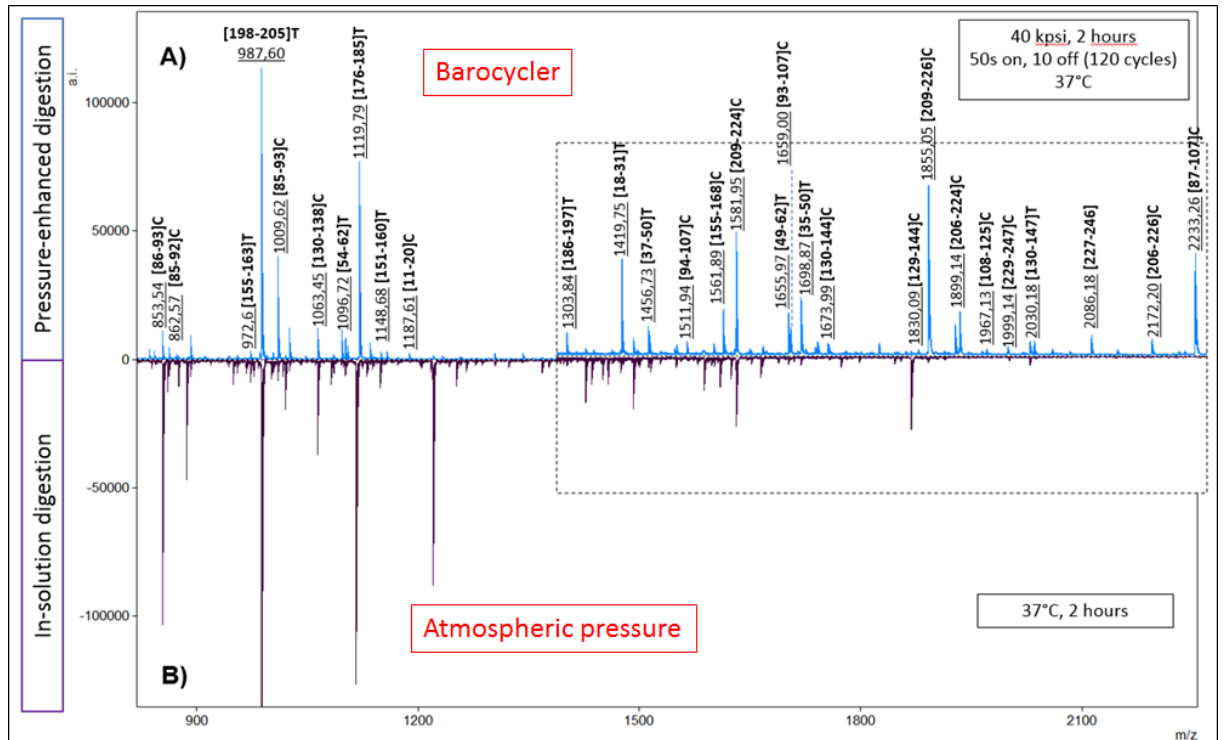
**Figure 19.** MALDI-MS peptide mapping of the SBA hydrolysis sample following the procedure 6 described in table 2. Results of enzymatic hydrolysis are shown after 5 minutes (A), 30 minutes (B) and 60 minutes (C). An offset of 70k a.i. and 140k a.i. intensity has been applied respectively to (B) and (C).

Comparison needed therefore to be made between the barocycler and standard atmospheric procedures for the same procedure and same incubation time in order to ensure the efficiency of the pressure-enhanced proteolysis procedure.

The optimal atmospheric digestion protocol has been determined to be at 37°C with tris-buffer and denaturation by boiling, with an enzyme to substrate ratio of 1:40 for the trypsin and 1:20 for the Chymotrypsin. The same conditions have therefore been applied to 2 samples of 1 mg/ml SBA, one being but under shaking in the incubator, and one being put at 37°C and 40kpsi cycles (50 seconds on, 10 seconds off) in the barocycler.

After 2 hours of hydrolysis, the samples have been ziptip and analyzed by MALDI-MS. We can observe 90% sequence coverage for the sample that has been put in the

barocyler (Figure 20a.) for 2 hours, against 32% sequence coverage for the one that underwent standard atmospheric pressure digestion (figure 20b.).



**Figure 20.** Comparison of MALDI-TOF mass spectra of SBA proteolytic digestion after 120 cycles of 50 s at 40 kpsi and 10 s at atmospheric pressure and 37 °C (a) and standard proteolytic digestion at atmospheric pressure and 37 °C (b). All peptide fragments produced singly charged ions and are labeled T for tryptic and C for chymotryptic cleavage products. In addition, peaks corresponding to sodium and potassium adducts ( $[M+Na]^+$ ,  $[M+K]^+$ ) were observed in all spectra. The region with  $m/z > 1300$  Da has been magnified.

Pressure-assisted enzymatic hydrolysis proved therefore to be an efficient and reliable tool to increase the hydrolysis speed of highly folded and hydrolysis-resistant proteins.



### 2.1.3. Structural characterization of lectins by proteolytic peptide mapping - mass spectrometry

The previously described method have been applied to hydrolyze SBA and to characterize the resulting tryptic and chymotryptic peptides, in order to apply later on the peptide mixture into an affinity column for separation and isolation of the target peptides.

Figure 21 shows sequence of SBA with the MALDI-TOF identified peptides from figure 20 highlighted in red. All predicted tryptic or chymotryptic fragments could be detected, except for the small fragments (2-5 amino acids) and the Asp-75 glycosylated peptide [71–75]. Detailed list of the obtained peptide is shown in appendix 3.

```
1    AETVSESWNK FVPKQPNMIL QGDAIVTSSG KLQLNKVDEN GTPKPSSSLGR
51   ALYSTPIHIW DKETGSVSASF AASFNFTFYA PDKRLADGL AFFLAPIDTK
101  PQTHAGYLGL FNENESGDQV VAVEFDTFRN SWDPPNPHIG INVNSIRSIK
151  TTSWDLANNK VAKVLITYDA STSLLVASLV YPSQRTSNIL SDVVDLKTSL
201  PEWVRIGFSA ATGLDIPGES HDVLSWSFAS NLPHASSNID PLDLTSFVLH
251  EAI
```

**Figure 21.** Sequence of SBA, in which the identified peptides are highlighted in red.

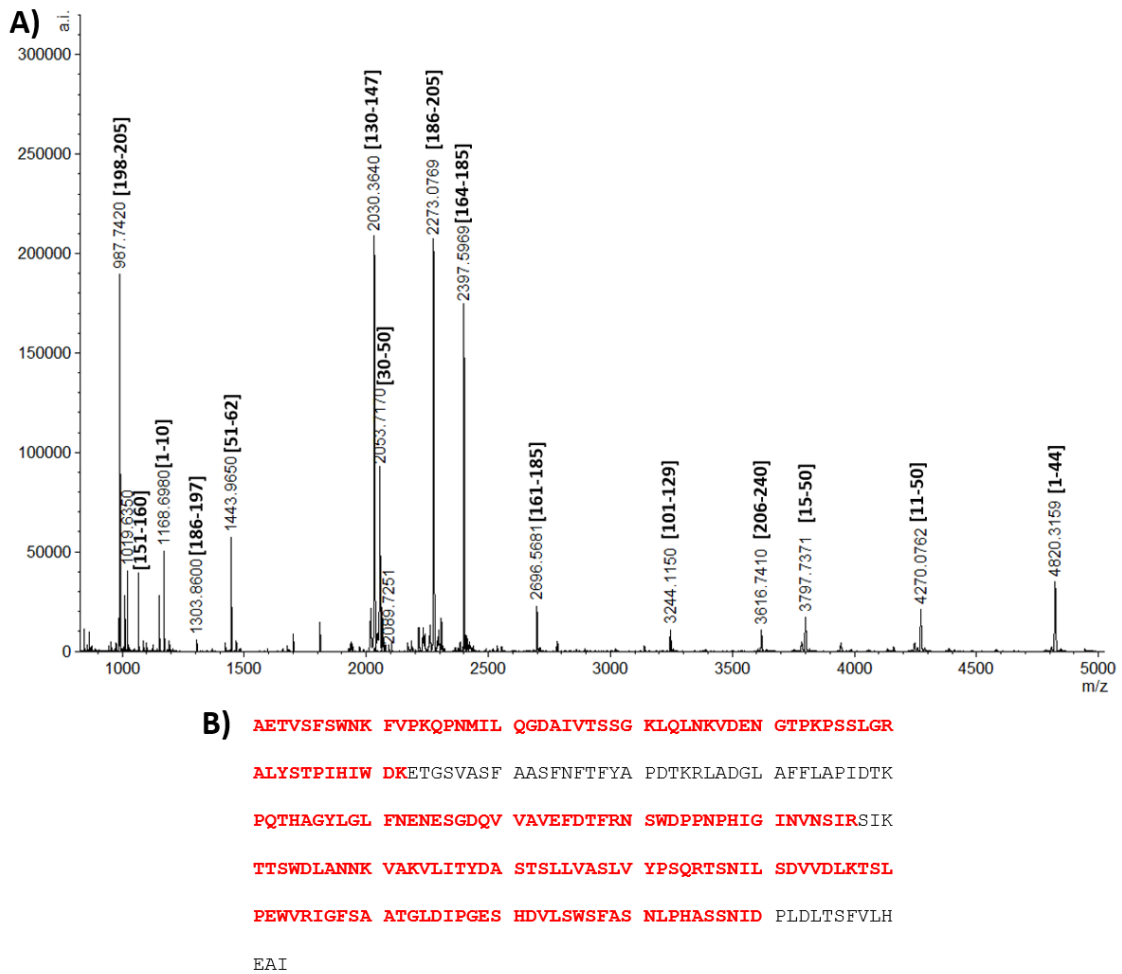
One of the main drawbacks of the use of multiple enzymes such as a mixture of trypsin and chymotrypsin is the relatively small size of the peptide obtained. Indeed, together these two enzymes cover a lot of possible cleavage sites, thus rendering the analysis of epitopes by MALDI-MS difficult as they could consist of small fragments of 2 to 5 amino-acids.

In order to overcome such a situation, the use of Clostripain was investigated. Clostripain is a cysteine-activated protease. It selectively hydrolyzes arginyl bonds, and lysyl bonds at a lower rate, and is dependent on thiol and calcium ions

(115). Unlike Trypsin, Clostripain accepts proline in the S1' subsite (116), making its use a valuable addition to the number and size of generated peptides.

Since the activity of Clostripain depends upon a cysteine thiol group, a reducing agent (1 mM DTT) was added to the digestion buffer (100 mM Tris, 10 mM CaCl<sub>2</sub>). 100 µg SBA was digested with an enzyme to substrate ratio of 1:20 and incubated in the barocycler for 2 hours (50°C, 30 kpsi, 50 seconds on, 10 seconds off, 120 cycles). The digestion was then stopped with 2 µL TFA and analyzed by MALDI-MS (figure 22).

80% of the SBA sequence has been recovered. Some cleavages after lysine have been observed as well as peptides cleaved due to the C-terminal truncation of the lectin. The absence of the fragment [63-83] can be explained by the Asp75 glycosylation and the low ionization efficiency of glycopeptides, as well as the ionization suppression caused by the nonglycosylated peptides.

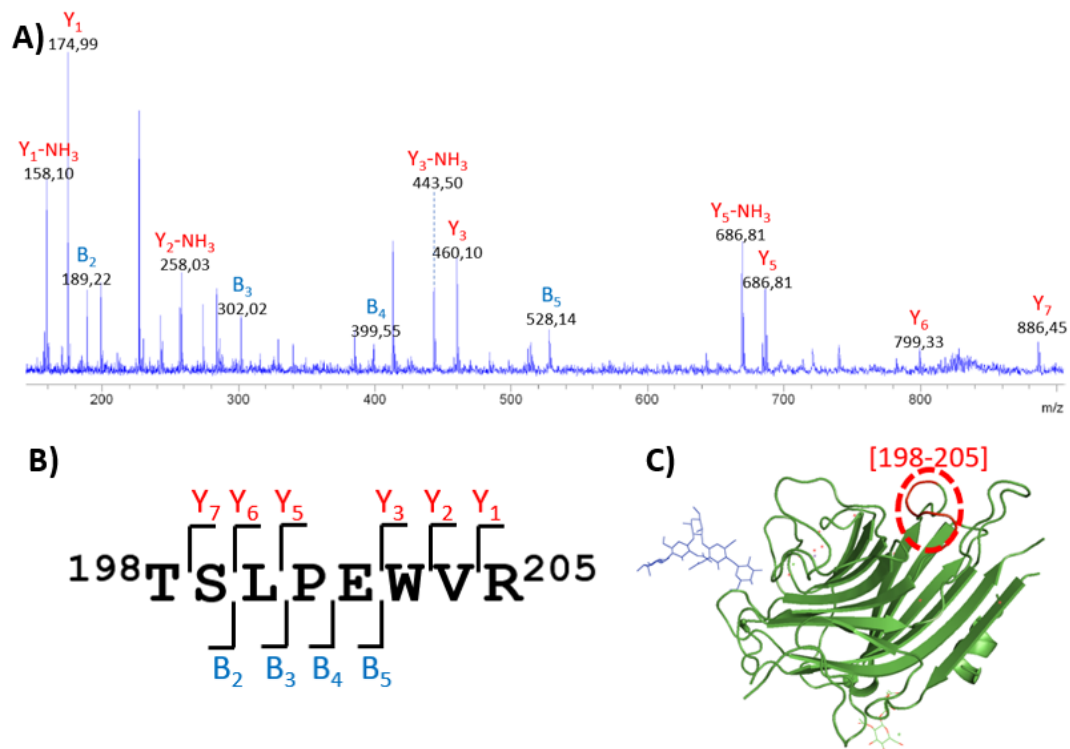


**Figure 22.** (a) MALDI-TOF MS of intact SBA Clostripain digest. All assigned peptides were singly charged. (b) Sequence of SBA, in which the identified peptides are highlighted in red.

The use of this method would be useful to extract longer peptides fragments from affinity-MS and confirm the results that can be obtained by other proteolytic methods.

In order to ensure the proper identification of the peptides, MS/MS has also been performed. Due to the relative low sensitivity of the instrument, MS/MS was performed on some of the most intensive and critical signals, such as the signal  $m/z=987.7$  Da [198-205], which is a fragment showing tryptic cleavages and found in both trypsin/chymotrypsin method and Clostripain method.

As it can be seen in figure 23, MALDI-TOF/TOF using collision induced dissociation (CID) provided full sequence coverage of the target peptide.



**Figure 23.** (a) CID MS/MS of peptide [198-205] during MALDI-MS of the proteolytic peptide mapping. (b) Sequence of SBA[198-205] showing the identified fragmentation sites. (c) Position of the fragment [198-205] on the 3D structure of SBA.

#### **2.1.4. Synthesis and purification of a peptide spacer for glycan immobilization**

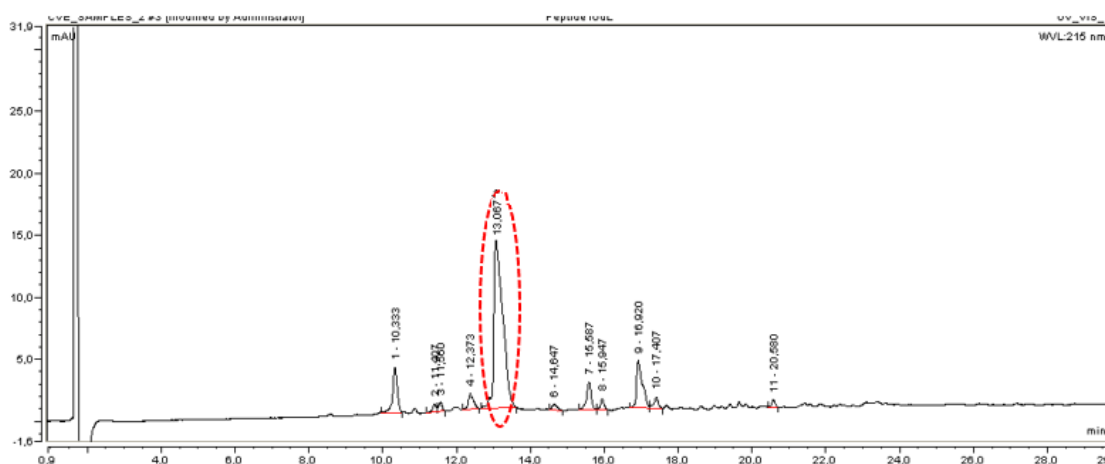
In recent years, techniques for screening interaction partners has been increasingly studied, mostly on protein–protein binding, especially by using surface plasmon resonance (SPR) to capture new binding partners prior to characterization by mass spectrometry (117) or HPLC profiling (118).

Versatile methods for surface modification and bioconjugation are essential to the biosensor and microarray communities. Approaches with either immobilized lectins or glycans for highly parallel analysis, high sensitivity and specificity of detection have become increasingly popular with applications in diagnostics and for the detection of a low level of disease biomarkers (119).

A variety of strategies have been used to achieve non-covalent and covalent attachment of biomolecules on a solid support. Direct covalent immobilization of sugars, through their reducing end, on sensor surfaces is challenging because the methods designed for these procedures involve complex and time-consuming chemically derivatization steps. Moreover, the small size of the glycans compared to the lectins lead to difficult stereo-interaction of the protein-glycan complex and difficult reliable biosensor monitoring.

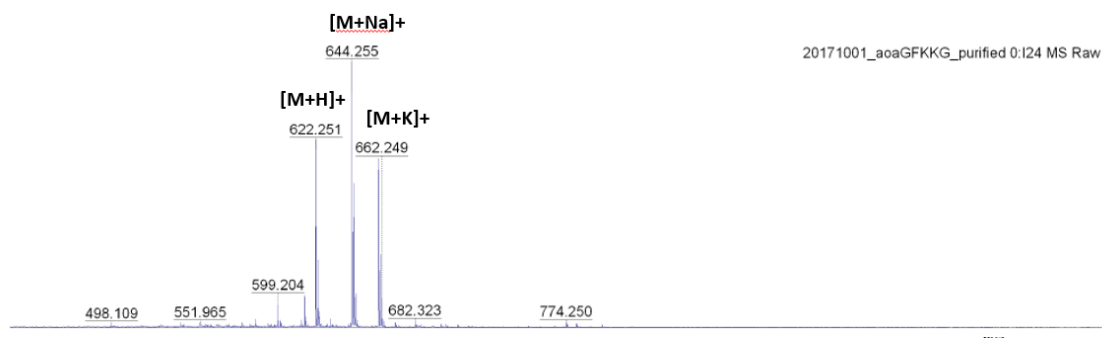
Here, we investigated the use of a small aminoxyacetyl (aoa) containing peptide carrier, N[Me]-O-Aoa-GFKKG-amide (figure 24), as linker between the biosensor surface and the glycan. Indeed, oxime chemoselective reaction between the reducing end of an oligosaccharide and an AOA-containing peptide has been shown to be an efficient tool for glycoconjugation (120). The presence of the methyl group ensures correct exposure of the carbohydrate on the chip surface as well as the conformational integrity of the monosaccharide unit proximal to the surface, a particularly relevant point for short (mono- and disaccharide) epitopes. Sequences longer than the 5 amino acid peptide were investigated by Vila-Perello et al (70) with poorer results.





**Figure 26.** HPLC-UV spectrum of an aliquot of the N[Me]-O-AOA-GFKKG peptide during purification step. The peak at 13 minutes has been confirmed by LC-MS to be the target peptide.

The main peak was then collected and the ACN removed by rotary evaporation, in order to the formation of a Schiff base between the purified peptide and acetone impurities that can result from HPLC-grade ACN. The purified peptide was then analyzed by MALDI-MS. The analysis confirmed the successful purification of the peptide (figure 27), with the presence of sodium and potassium ions, which are common to observe by MALDI-MS.



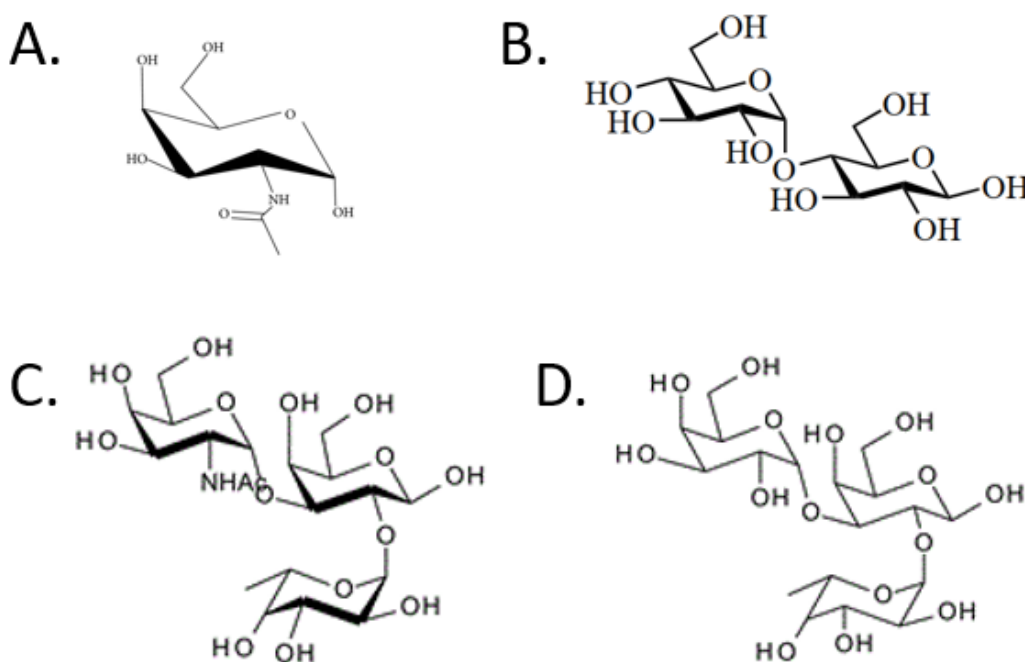
**Figure 27.** MALDI-TOF MS spectra of the purified N[Me]-O-AOA-GFKKG ( $m/z$  622 Da) peptide.  $[M+Na]^+$  and  $[M+K]^+$  are also observed. Impurities observed in the MS spectra prior purification (figure 25) are absent.

The peptide was the lyophilized and stored at  $-20^{\circ}\text{C}$  for further uses.

## 2.2. Identification of carbohydrate recognition epitopes in lectins

### 2.2.1. Pressure enhanced proteolytic excision and extraction of lectins

In a first variant of the analytical approach termed proteolytic excision (121) (122), several carbohydrates (GalNAc, maltose, blood group A and blood group B trisaccharides, figure 28) were immobilized on an sepharose-based affinity column via a hydroxyl group, and the carbohydrate-binding SBA added on the column and incubated to form a protein-carbohydrate complex. The column was then washed using the binding buffer to remove any excess of non-bound SBA and other impurities, leaving only the affinity-bound protein onto the column.



**Figure 28.** Chemical structures of the carbohydrate ligands investigated : (A) N-Acetylgalactosamine (2-acétamido-2-désoxy-D-galactopyranose), (B) Maltose ( $\alpha$ -D-glucopyranosyl(1-4)-D-glucopyranose), (C) blood group A-trisaccharides ( $\alpha$ -L-Fuc-(1-2)-[ $\alpha$ -D-GalNAc-(1-3)]- $\beta$ -D-Gal), (D) blood group B-trisaccharides ( $\alpha$ -L-Fuc-(1-2)-[ $\alpha$ -D-Gal-(1-3)]- $\beta$ -D-Gal)



The previously described proteolytic excision experiment used standard atmospheric pressure digestion (121). The interest of this approach is that the epitope fragments recognized by the protein are thereby shielded by it against digestion and will remain bound to the immobilized protein. However, one of the drawbacks of this approach is the long duration needed for the immobilized protein (up to 12 hours) to be completely digested.

Here, we investigated a new approach in which the immobilized complex is directly subjected to pressure-enhanced proteolytic digestion, by subjecting the sepharose-ligand-lectin complex to the Barocycler and to the procedure described above. A mixture of trypsin and chymotrypsin as described in chapter 2.1.2. Upon using carbohydrate-bound sepharose, the length of the incubation can be varied as carbohydrates are inherently stable to the action of the proteases.

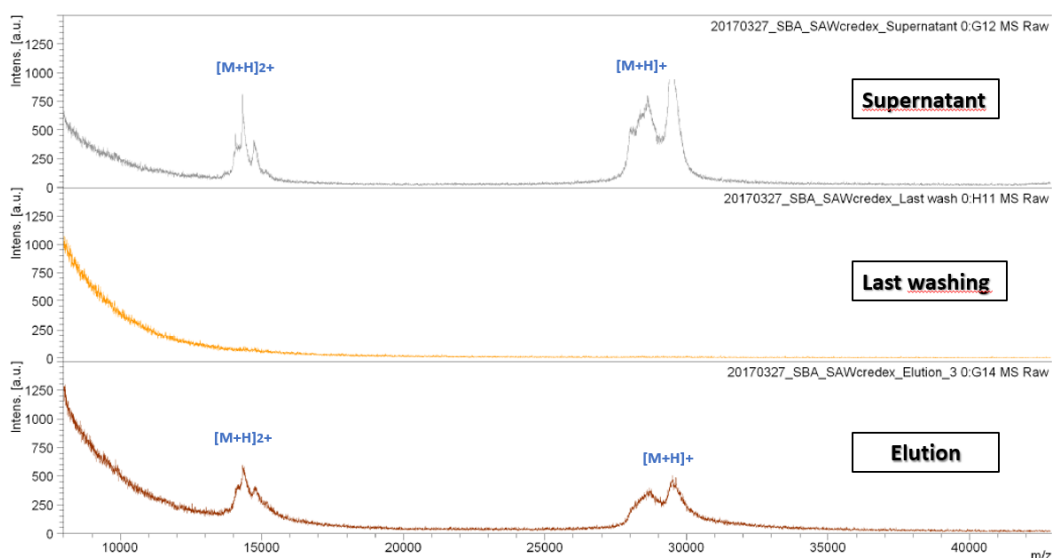
On the following step, the non-bound peptides were washed away by using the binding buffer and the remaining affinity-bound peptide ligand fragments were eluted, washed by Ziptip and analyzed by MALDI-MS.

In a second variant, proteolytic extraction, SBA was first subjected to pressure-enhanced proteolytic digestion as described in chapter 2.1.2. and the mixture of peptide fragments was then presented to the immobilized carbohydrates and incubated in order for the peptides presenting affinity towards the carbohydrate to bind to it. The washing and elution step were then performed similarly as the proteolytic excision method.

In order to verify the proper binding of the sugar employed to the sepharose column, control experiments were done using the entire SBA prior to digestion. In short, a 1 mg/ml SBA solution in HEPES buffer was incubated on the affinity matrix for 2 hours at 37°C. The supernatant was then removed and analyzed by MALDI-MS. Any remaining non-bound SBA was then washed away by using 400 µL of the binding buffer and shaking for 15 minutes. After 5 washing steps, the last fraction was cleaned by C4 ziptip and analyzed by MALDI-MS and showed no trace of SBA. The affinity column was then incubated two times in 400 uL ACN/0.1% TFA 2:1 (v/v) solution

under agitation for 15 minutes and eluted. The two fractions were mixed, concentrated and washed by 4C ziptip and analyzed by MALDI-MS (figure 29).

The absence of SBA in the last washing fraction and its identification in the elution fraction strongly indicate that the column preparation was successful and that the lectin bind efficiently to the affinity column and do not elute even with extensive washing.



**Figure 29.** MALDI-MS measurement of the control experiment. SBA was bound to a GalNAc-sepharose column. After incubation, the supernatant was washed away and measured by MALDI-MS (top). The affinity-matrix was then washed with binding buffer until no SBA can be measured (middle). The remaining bound SBA was then eluted from the column and analyzed (bottom).

### 2.2.2. Development of buffer and elution system for lectin affinity interactions

Lectins bind carbohydrates noncovalently, reversibly and relatively weakly, at least in the case of mono- and disaccharides which are the most widely used ligands. Due to the particularly low affinity of such interactions compared to protein-protein or protein-antibody, particular care has to be taken when eluting peptides bound to

carbohydrate as well as to the pH of the solutions to maximize the recovery of the peptides and the stability of the column during repeated uses. Elution of the affinity-bound peptides can be performed with a concentrated solution of the identical carbohydrate immobilized on the affinity column; the carbohydrate in solution thereby displacing the affinity-bound peptides from the immobilized carbohydrate in a competitive manner. A different carbohydrate may be used for elution if the affinity of the protein or peptide is comparable to the immobilized carbohydrate. To prevent formation of cysteine-containing peptide dimers through disulfide bonds, DTT (1 g/L) can be added to the eluent without adverse effects. For the elution of affinity-bound SBA peptides from 200  $\mu$ L matrix with immobilized GalNAc, the column was shaken with 400  $\mu$ L 0.3 M GalNAc in HEPES buffer and eluted, and this procedure repeated twice. However, due to the high concentration of carbohydrate required for quantitative elution of affinity-bound peptides, sample cleaning prior to MS analysis that could lead to interference in the desalting step, and time-consuming HPLC-MS manipulation, this elution method was not prioritized. Another drawback of the carbohydrate elution system is the relative complexity and high prices of the sugar employed.

As SBA is dependent to calcium and magnesium ions for its activity, removing them from the binding buffer did lead to a limited recovery of bound peptides.

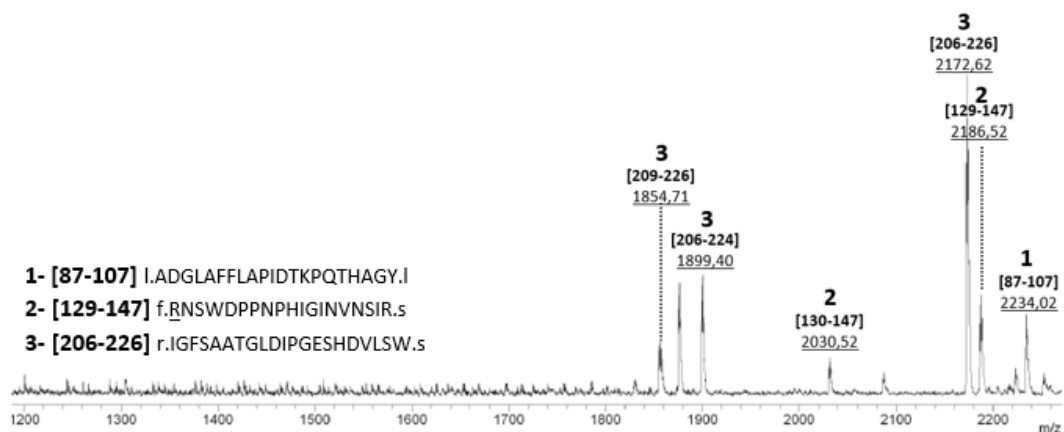
Elution with aqueous-organic solvents is advantageous because of the lower price of chemicals used and the easier sample cleaning procedures. Several combinations of TFA, HCl, acetic acid, THF, ACN, and methanol containing elution solutions were initially tested with varying results. The most successful solvent systems were based on acetonitrile, particularly ACN/0.1% TFA, 2:1 (v/v), pH 3, or 60–80% ACN in water, pH 6. Elution was performed in the same manner as for competitive elution, by shaking with 400  $\mu$ L eluent for 15 min and repeating twice this procedure. Since acetonitrile-containing solutions have a dehydrating effect on agarose-based affinity matrices depending on the percentage of ACN, the matrix had to be directly equilibrated in HEPES buffer immediately after the elution step. Indeed, the use of a 100% ACN elution solution did not allow to recover the bound peptides and completely dehydrated and irreversibly damaged the affinity matrix. Other elution

systems such as 500mM acetic acid or 5–10% ACN in water did not perform well, and also solutions based on non-volatile substances such as glycine or chaotropic agents were found unsuitable.

### 2.2.3. Identification of saccharide and disaccharide binding sites in lectins

As SBA shows affinity towards GalNAc and glycosides and oligosaccharides containing terminal GalNAc and Galactose and shows no affinities towards N-acetyl-D-glucosamine and glucopyranosyl. GalNAc was therefore selected as a ligand in order to identify the binding sites of SBA. Parallely, a negative control was prepared using a second affinity column and maltose as ligand, in order to differentiate between specific SBA-GalNAc binding sites and possible unspecific protein-matrix or protein-sugar binding. The conditions (buffers, binding and elution times, number of washing steps) were identical in both experiments.

MALDI-MS analyses of the elution fractions upon proteolytic excision and extraction from the affinity columns performed with immobilized GalNAc showed ions corresponding to 3 specific peptide sequences [87–107] (1), [129–147] (2), and [206–224] (3) (Figure 30).



**Figure 30.** MALDI-TOF mass spectra of elution fractions from proteolytic excision of glycine max lectin from GalNAc affinity column. Peptides (1) correspond to SBA[87–107]; peptides (2) correspond to SBA[129–144], SBA[130–144], and SBA[130–147]; peptides (3) correspond to SBA[206–224] and SBA[206–226]. In addition, low intensity peaks corresponding to sodium and potassium adducts ( $[M + Na^+]^+$ ,  $[M + K^+]^+$ ) were observed.

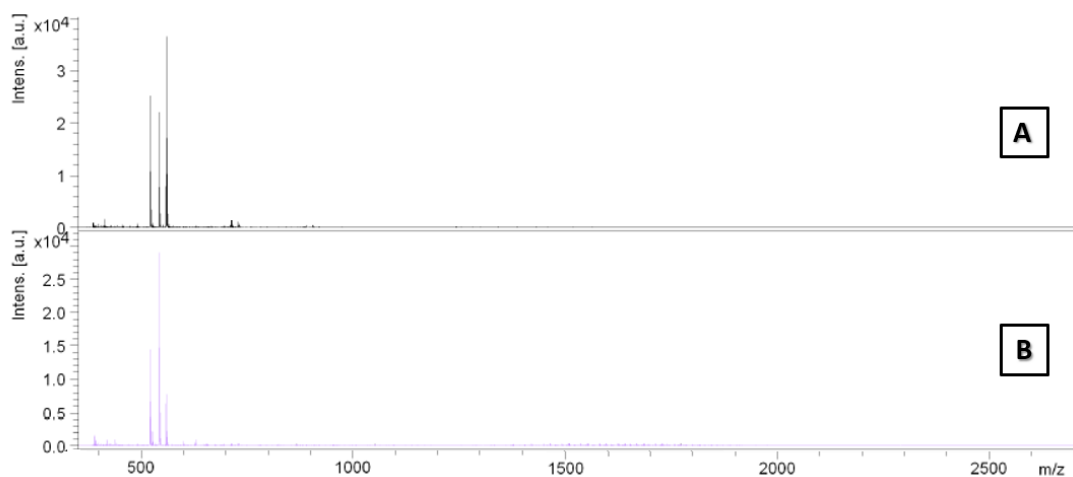
These epitope peptides identified in the elution fractions contain the amino acid residues Asp-88, Gly-106, Asn-130, and Asp-215 in direct contact with the carbohydrate ligand, in complete agreement with the X-ray crystal structure data (Figure 31). In the crystal structure, Rao et al. (123) reported that Asp-88 forms a bidentate hydrogen bond with C3-0H and C4-0H hydroxyl groups, while the nitrogen group ND2 of Asn-130 and of Gly-106 form hydrogen bonds with C2-0H and C3-0H hydroxyl groups, respectively, and the 4-0H hydroxyl group forms another hydrogen bond with the backbone-NH of Leu-214.



**Figure 31.** X-ray crystal structure of SBA in complex with a biantennary blood group antigen analog (PDB entry 1G9F) showing SBA[87–107], SBA[129–147], and SBA[206–226] in red. Amino acids known to be in direct contact with the carbohydrate are highlighted in blue.

MALDI-MS analyses of the elution fractions upon proteolytic excision and extraction from the affinity columns performed with immobilized maltose (figure 32) did not show any SBA peptide, both with competitive elution and with strong acid/organic elution. A blank MALDI spot, containing only matrix and solvent, was also analyzed and showed the same signals observed at  $m/z < 600$  Da, thus confirming that the signals observed in the elution fraction of the control experiments issued from impurities and matrix interferences.

The negative control thus confirmed the specificity of the peptides obtained by proteolytic excision and extraction and dismissed the presence of any unspecific binding that would be issued from protein – matrix interaction.



**Figure 32.** MALDI-TOF mass spectra of last washing and elution fractions from proteolytic excision of glycine max lectin from maltose affinity column. (a) Competitive elution (b) ACN/0.1% TFA 2:1 (v/v) elution.

#### 2.2.4. Identification of binding sites of blood group oligosaccharides in SBA

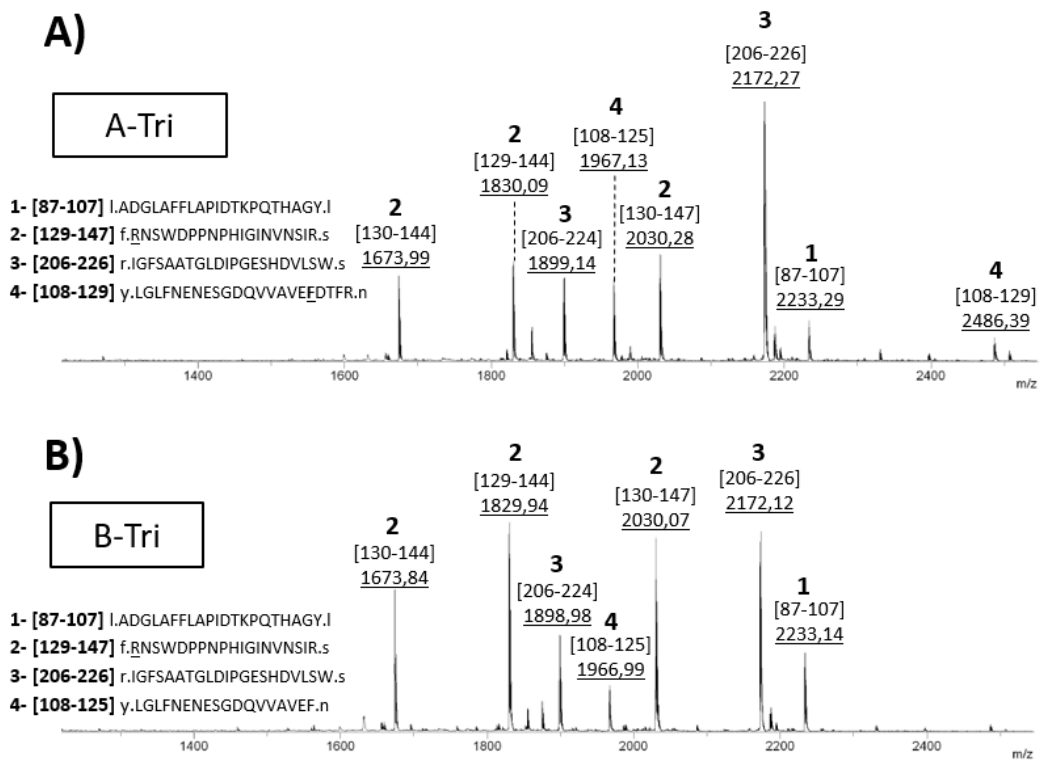
After the mass spectrometric identification of the SBA epitope specific to GalNAc, a further goal of this dissertation was to observe specific epitope peptides residing in the structures of lectins and showing a molecular differentiation between saccharides and oligosaccharide epitopes that should be applicable to a blood group determination assay.

Therefore, blood group A and B trisaccharides, which contains galactose and GalNAc moieties, were employed in order to observe possible difference in epitopes obtained, both between the two trisaccharides and the between the blood group sugar and a simple GalNAc ligand.

During these experiments, competitive elution using GalNAc were unsuccessful, even at high concentration. This may be explained by the difference of affinities between the AB blood group epitope peptides and the GalNAc peptides, leading to the inability for GalNAc to compete with AB oligosaccharides. Utilization of high concentration of GalNAc furthermore led to interference with the desalting procedure. Due to the high cost of AB oligosaccharides, competitive elution using these sugars was not possible. The standard elution system with ACN:0.1% TFA 2:1 (v/v) was therefore used for all AB oligosaccharides affinity experiments, followed by direct re-equilibration with binding buffer in order to avoid the dehydration effect of ACN on the sepharose matrix.

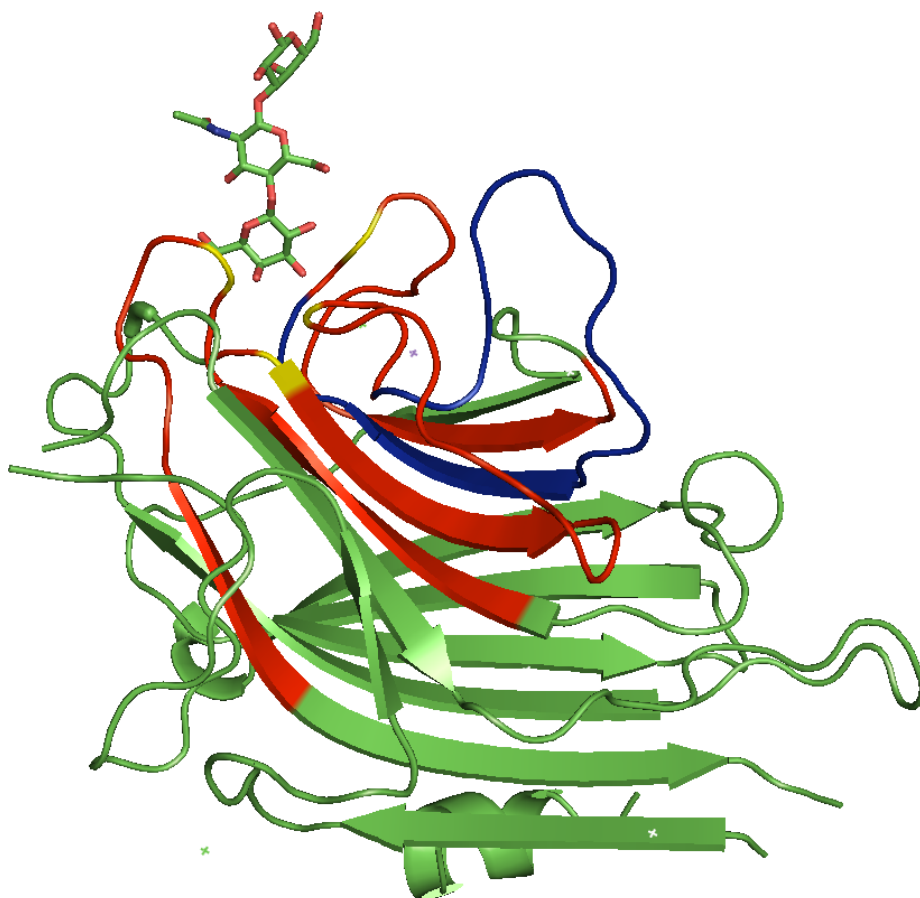
Immobilized A- and B- trisaccharide were incubated with SBA and subjected to a mixture of trypsin and chymotrypsin under cycles of high pressure and atmospheric pressure as described in chapters 2.1.2 and 2.3.1. The columns were then washed until no peptide signal was observed by MALDI-MS and the bound peptides were then eluted.

Identical epitope ligands were found between the elution fractions from immobilized GalNAc and from A- and B-trisaccharides, with an additional specific peptide sequence [108–125] (Figure 33).



**Figure 33.** MALDI-TOF mass spectra of elution fractions from proteolytic excision of glycine max lectin from A-Tri (A) and B-tri (B) affinity column. Peptides (1) correspond to SBA[87–107]; peptides (2) correspond to SBA[129–144], SBA[130–144], and SBA[130–147]; peptides (3) correspond to SBA[206–224] and SBA[206–226], and peptides (4) correspond to SBA[108–125] and SBA[108–129]. In addition, low intensity peaks corresponding to sodium and potassium adducts ( $[M + Na^+]^+$ ,  $[M + K^+]^+$ ) were observed.





**Figure 34.** X-ray crystal structure of SBA in complex with a biantennary blood group antigen analog (PDB entry 1G9F) showing SBA[87–107], SBA[129–147], and SBA[206–226] in red and MS-identified peptides specific for A-tri and B-tri, SBA[108–128], in blue. Amino acids known to be in direct contact with the carbohydrate are highlighted in yellow.

The fourth peptide domain (4) identified only in the elution fractions of A- and B-trisaccharide include the loop [112–119], where electron density has been found to be weak, and is located close to the sugar binding site (123) (figure 34). The presence of (4) in the elution fraction can therefore be explained by an interaction between the branched oligosaccharides and the peptide loop. These results indicate the potential of pressure-assisted proteolytic excision mass spectrometry for identifying affinity-derived lectin epitope peptides, and to differentiate between different carbohydrate structures.

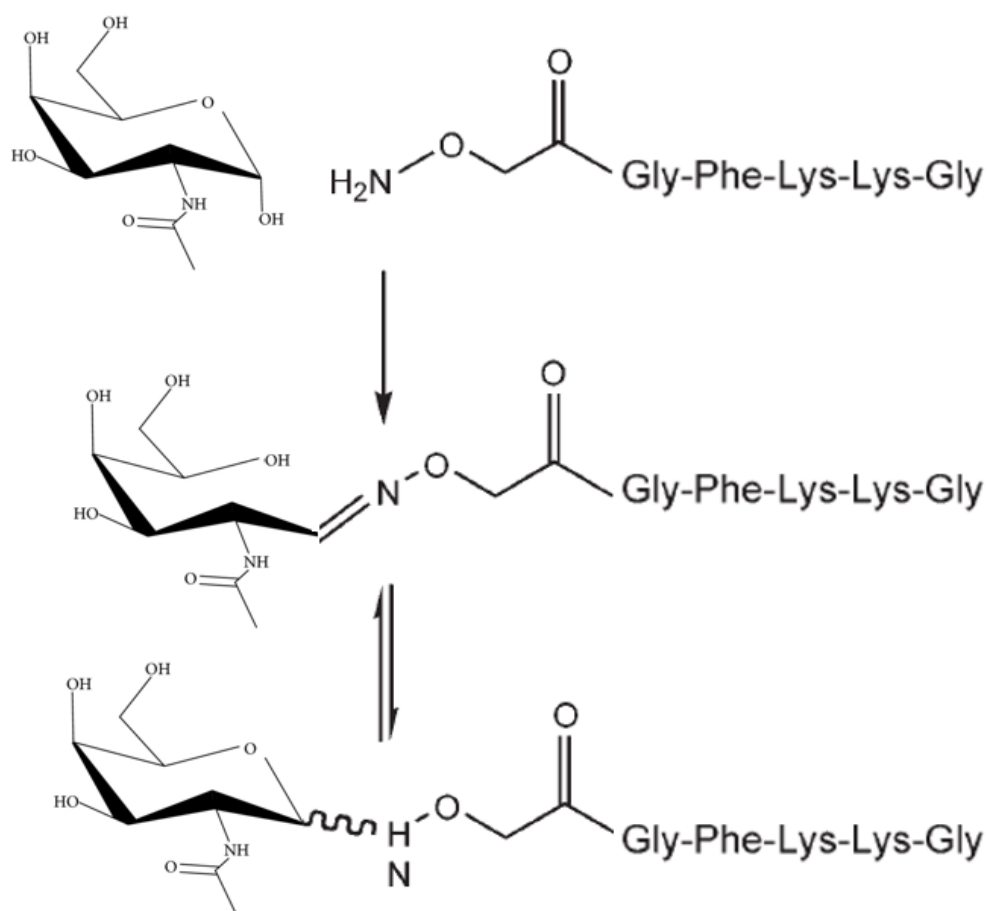
### **2.3. Determination of dissociation constants of lectin - carbohydrates interactions**

Although pressure-enhanced proteolytic excision mass spectrometry allowed us to characterize qualitatively the epitopes binding to the ligands, little is known about the precise dissociation constant ( $K_D$ ) of these lectin – carbohydrate complexes.

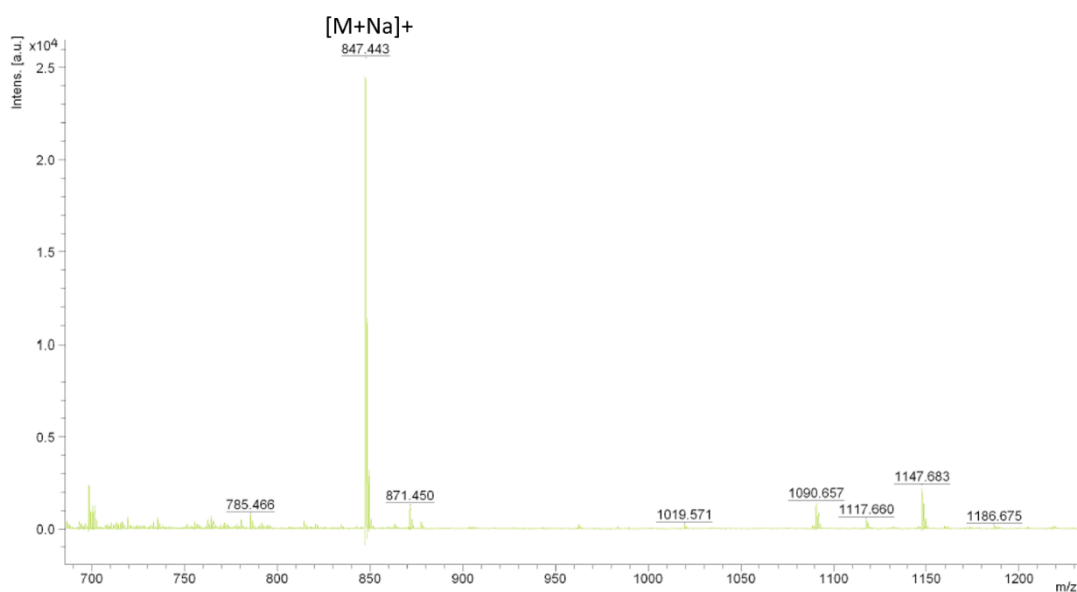
In order to measure the affinity of SBA to the saccharides, different approaches were investigated to attach the ligand to the surface, using a DVS activation approach, and using a tailor-made peptide carrier.

Using the 5 amino-acid peptide carrier described in chapter 2.2, GalNAc was attached to the amidated peptide carrier via oxyme ligation through an aminoxyacetic acid (AOA) linker (Figure 35) and the resulting glycoprobe was immobilized on an N-hydroxysuccinimide (NHS)- activated gold chip, creating a highly reactive succinimide ester which reacts with primary amines on peptides and proteins. After immobilization, the remaining activated carboxymethyl groups were blocked with ethanolamine, as described in the Experimental Section (Chapter 3.4.4.).

This tailor-made peptide fulfills several necessary criteria to display properly glycans over a gold surface. The two Lys residues ensure proper guidance to the negatively charged sensor surface and an efficient covalent C-terminal binding to the activated carboxyl group on the gold chip. The addition of a Phe increase the hydrophobic character of the peptide and facilitate subsequent purification/isolation steps, while the Gly residues provide flexibility and distance between the surface of the gold chip and the glycan epitope. All-in-all, this peptide was large enough to provide a substantial mass enhancement effect on the chip surface, thus enhancing the biosensor's signal response.



**Figure 35.** Oxime chemical ligation reaction process between the peptide linker Aoa-GFKKG and a GalNAc saccharide.

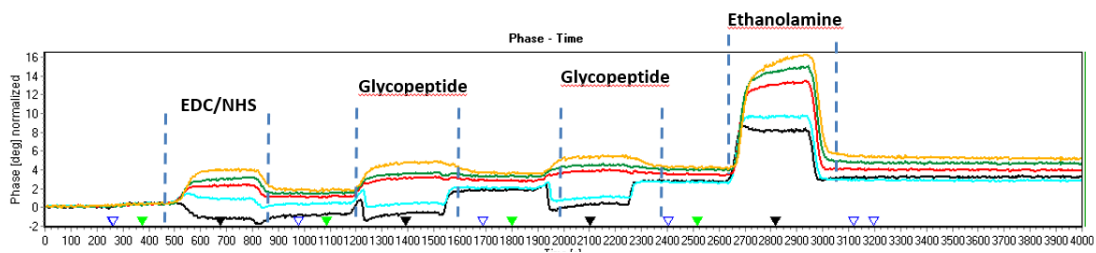


**Figure 36.** MALDI-TOF MS spectra of the GalNAc-N[Me]-O-Aoa-peptide. The peak at  $m/z=847.443$  correspond to the mass of the glycopeptide  $[m+Na]^+$ .

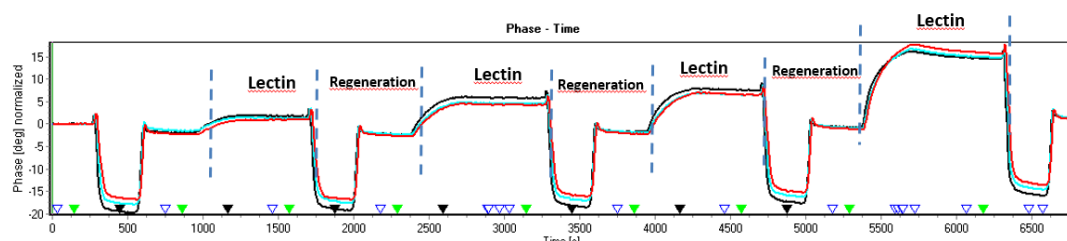
In order to confirm the versatility of the method, 2 different lectin-carbohydrates couples were selected for this experiment: SBA – GalNAc and Concanavalin A (ConA) – Mannose.

Glycoconjugation have been performed with equimolar concentration of glycans and peptide in 0.1M AcOH, pH4.6, at 37°C, and the resulting glycopeptide have been purified by preparative chromatography, analyzed by MALDI-MS (figure 36), and lyophilized.

The glycopeptide have then been successfully immobilized on the gold chip surface (Figure 37), and series of different concentrations of lectins were injected over the immobilized glycoprobe (Figure 38).

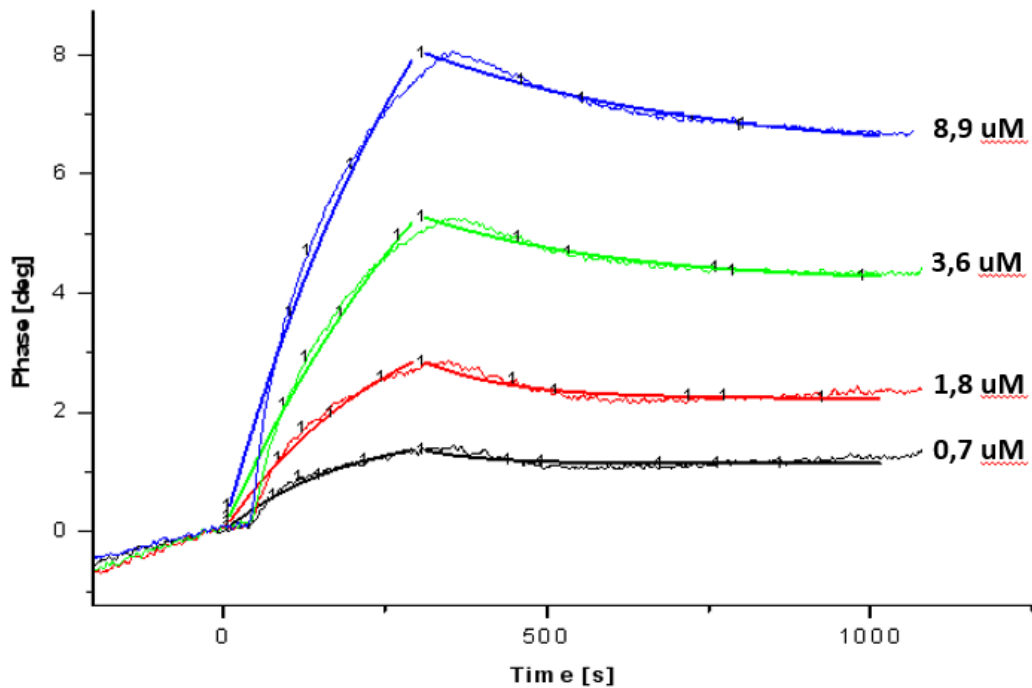


**Figure 37.** EDC/NHS immobilization of the GalNAc glycopeptide over the SAW gold chip and ethanolamine capping.

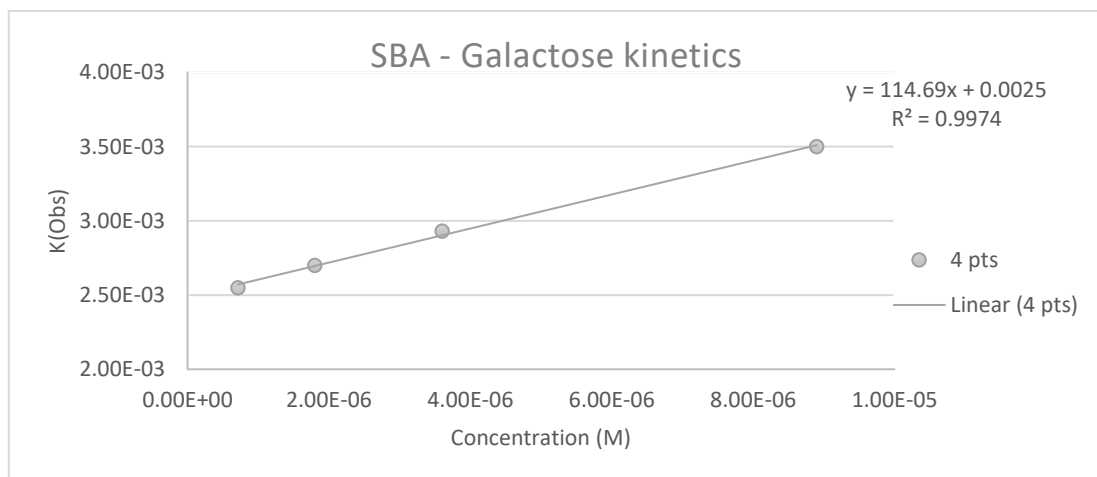


**Figure 38.** Series of injection of increasing concentration of SBA followed each time by an injection of a regeneration solution.

After fitting the recorded data according to the theoretical 1:1 Langmuir binding model (figure 39), the extracted observed rate constants ( $k_{\text{obs}} = k_{\text{off}} + C_{\text{analyte}} * k_{\text{on}}$ ) were plotted versus the analyte concentrations (Figure 40). Linear regression of the data points yielded the association ( $k_{\text{on}}$ ) and dissociation ( $k_{\text{off}}$ ) rate constants that were used to obtain the equilibrium dissociation constant  $K_D = k_{\text{off}} * k_{\text{on}}^{-1}$ .



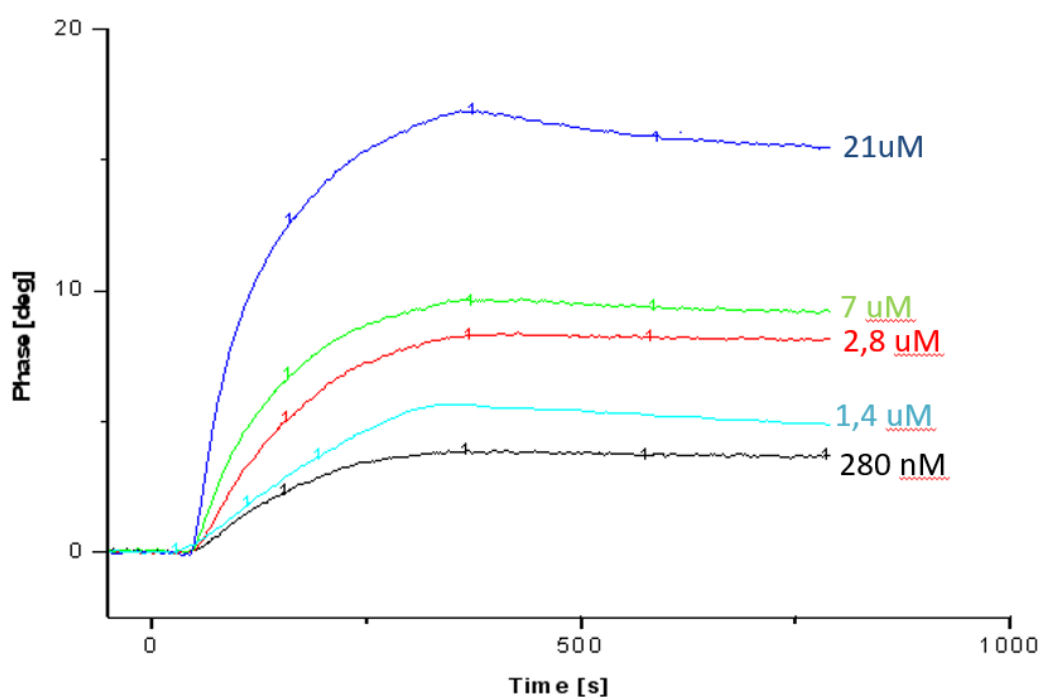
**Figure 39:** Fitted association and dissociation curves of SVA in a series of concentrations (0.7 – 8.9  $\mu\text{M}$ ) with the immobilized glycoprobe



**Figure 40:** The plot of  $k_{\text{obs}}$  values against SBA concentration concentrations provided a  $K_D$  of 22  $\mu\text{M}$ .

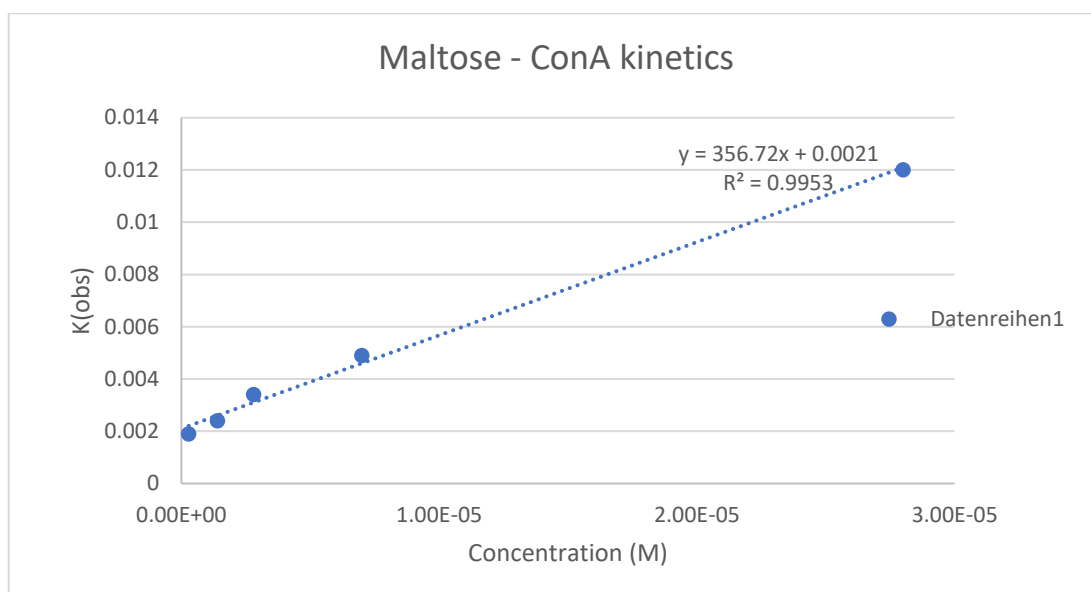
The  $K_D$  obtained for the interaction between galactose and SBA was 22  $\mu\text{M}$ , in accordance with the range observed for protein – carbohydrate interactions.

Similar results have been observed for the Concanavalin A – Mannose affinity experiment. However, here, the 1:1 Langmuir binding model + buffer offset fitting was applied (figure 41). In this model, the constituents that are accompanying the buffer and the injected molecule are taken into account and a permanently bound residue is assumed. This model provided a better fitting that the standard 1:1 model, and could be explained by the use of unpurified samples, by aberrant binding behavior or by the higher affinity observed.



**Figure 41:** Fitted association and dissociation curves of ConA in a series of concentrations (0.28 – 21  $\mu\text{M}$ ) with the immobilized glycoprobe.

Indeed, after plotting of the  $k_{\text{obs}}$  values against the concentration (figure 42), the  $K_D$  has been calculated to be 5.9  $\mu\text{M}$ .



**Figure 42:** The plot of  $k_{\text{obs}}$  values against ConA concentration concentrations provided a  $K_D$  of 5.9  $\mu\text{M}$ .

#### 2.4. Surface immobilization procedure of carbohydrate for the identification of protein binding sites

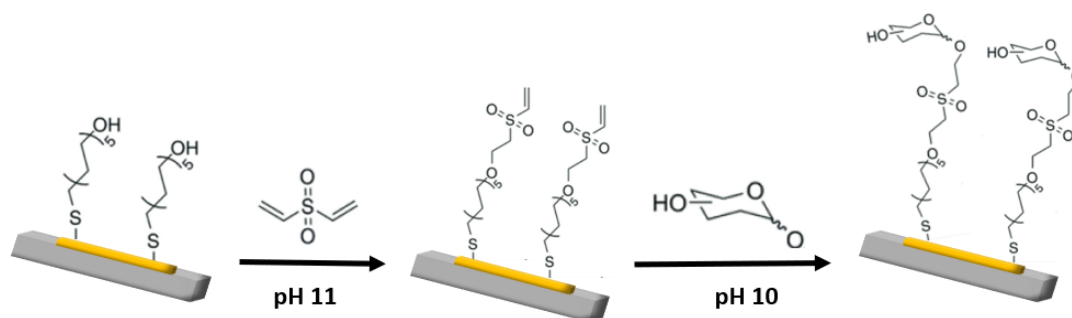
In order to overcome the time-consuming procedure for the immobilization of carbohydrate on sepharose microcolumns and to develop a quicker and less sample-expensive method for the determination of both carbohydrate binding sites of lectin and the measurement of carbohydrate-lectin thermodynamics, the proteolytic extraction mass-spectrometry method was adapted to be used inside the microfluidics of a biosensor instrument through the use of a gold chip as support for affinity matrix and epitope extraction.

As the peptide linker used in the above chapters includes lysine residues, its use was not possible for the immobilization of saccharides on the surface of the chip when proceeding with epitope extraction, because it would lead to cleavages and loss of fragments from the remaining trypsin present in the SBA digestion solution injected in the microfluidic systems.



To overcome this, the divinyl sulphone approach for carbohydrate immobilization on sepharose column was adapted and used to covalently immobilize the glycans over the gold chip surface. The process involves two sequential nucleophilic 1,4-additions (i.e., Michael reactions) in basic solution at ambient temperature, wherein DVS serves as a linchpin between the surface and captured biomolecules (124). The first step modifies hydroxyl groups present on the surface to generate an activated vinyl sulfone-modified substrate. In the second step, the DVS is reacted with biomolecules bearing a hydroxyl group. The functional biomolecules are thus covalently bound to the surface in a site selective manner based on the nucleophilicities of the reactive groups.

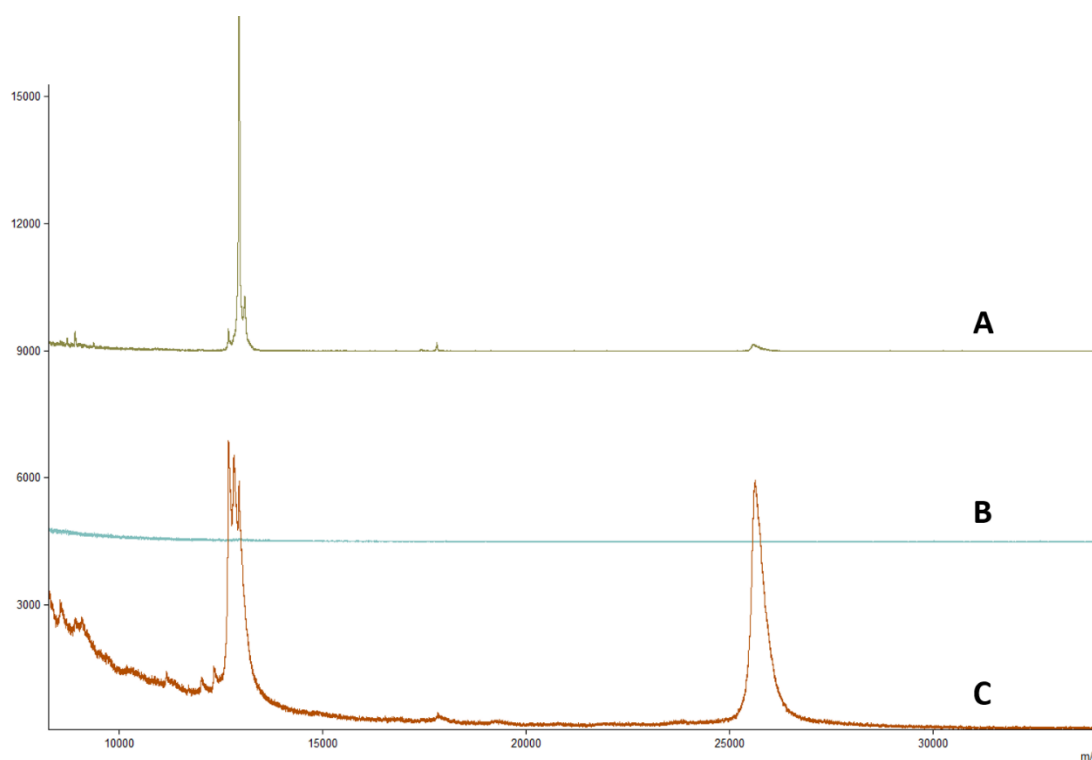
The freshly cleaned gold chip was first immersed in 16-mercaptohexadecanoic acid, cleaned and dried as previously described. The chip was then immersed in 10% DVS (v/v, 0.5 M carbonate buffer, pH 11) solution for 1 h at ambient temperature, followed by thoroughly rinsing with water and dried under a stream of argon. The DVS-modified surface was then immersed in a solution of 10% GalNAc in carbonate buffer, pH 10, in the dark and at ambient temperature, and under slow shaking for 16 hours. The chip was then washed with water, dried and stored in the dark at 4°C. The procedure is illustrated in figure 43.



**Figure 43.** DVS procedure for glycan-functionalized gold chip

The chip was then placed in the SAW biosensor sensor cell and a solution of 0.1% BSA in HEPES buffer was flown over it at 30  $\mu$ L.min for 2 hours in order to passivate the surface. The buffer was then changed to HEPES for all further steps.

A first trial has been performed using the maltose and concanavalin A (ConA) couple. Maltose has been immobilized on the surface and a solution of  $1 \text{ mg.mL}^{-1}$  concanavalin in HEPES has been injected in the SAW instrument. Upon detection of the arrival of the Concanavalin A over the complete surface of the gold chip, the flow rate has been stopped. The chip has then been put to  $37^\circ\text{C}$  for two hours. The flow has then been resumed and the sample washed away from the surface. The surface has then been washed and the remaining bound ECA eluted according to the method described in chapter 3.7. Supernatant, last washing fraction and elution fractions has been collected accordingly, and analyzed by MALDI-MS (figure 44).

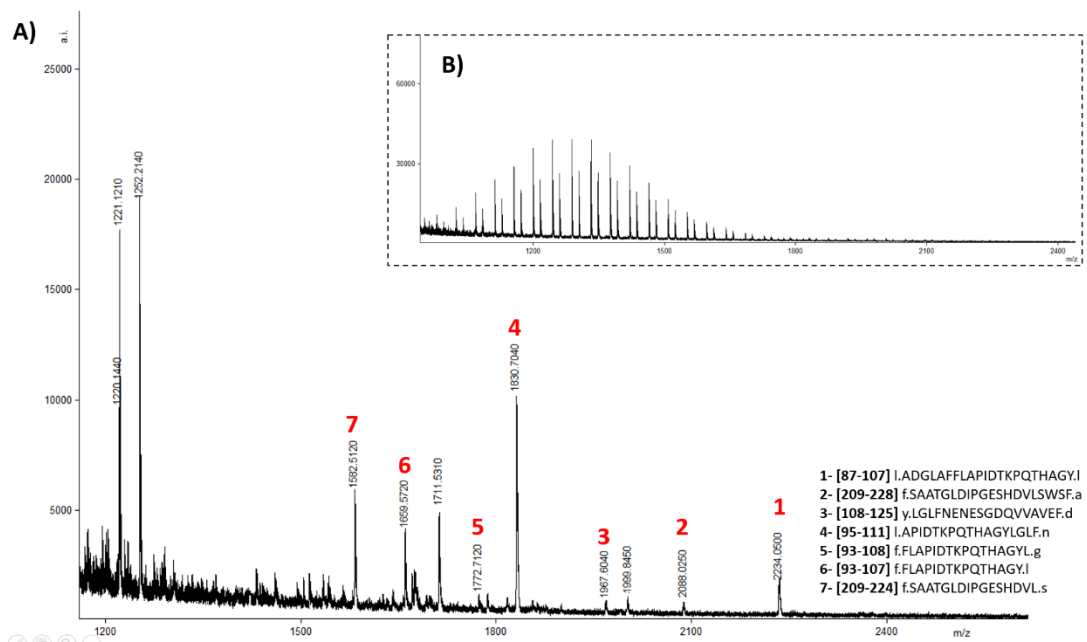


**Figure 44.** MALDI-MS spectra of supernatant (A), last wash fraction (B) and elution (C) of ConA over the functionalized SAW gold chip. An offset of 9000 and 4500 a.i. has been applied to A and B respectively.

As the method has been proven to be reliable by showing the presence of ConA in the elution fraction, the same principle has therefore been applied with a solution of digested SBA.

In short, GalNAc was immobilized on the surface following the above described procedure. In order to verify the affinity of the functionalized surface to SBA, several solutions of 0.1 to 10  $\mu$ M SBA in HEPES buffer were injected over the chip, and strong response signals were observed (appendix A2). This confirmed the successful binding of the GalNAc sugar to the surface.

A solution of pressure-enhanced digested SBA as previously described, was then diluted 1:1 with HEPES and injected over the SAW system following the method described above. Upon detection of the digest solution over the complete surface of the chip, the flow was stopped, the temperature increased to 37°C and the chip left to incubate for 12 hours. The chip was then washed and the bound peptides eluted, concentrated, washed and analyzed by MALDI-MS (figure 45) following the protocol described in chapter 3.7.



**Figure 45.** A) MALDI-MS of the elution fraction of a digest mixture of SBA peptides extracted from a SAW gold chip. Peptides 1 to 6 are signals corresponding to identified peptides. Some known contaminants (human Keratin) and unknown signals are also observed. B) MALDI-MS measurement of the last washing fraction. The signal observed correspond to polyethylene glycol (PEG) contamination. All signals were singly charged.

In both washing and elution, a strong background contamination could be observed for  $m/z < 1000$  Da, rendering identification of the smaller epitope difficult. The last washing fraction showed however no signals for  $m/z > 1000$  Da, except for Polyethyleneglycol (PEG) contamination signals. PEG that is a common contaminant in mass spectrometry easily recognized with its repeated signal of 44 Da intervals.

The elution fraction, however, showed several signals of interest. 2 sections of the SBA sequence could be observed, with different cleavages sites: SBA [87-111], which includes Asp-88 and Gly-106, and SBA [209-228], which includes Asp-215. These epitopes peptides are in complete agreement with the X-ray crystal structure data and with the results previously obtained.

Improvements remains yet to be made to obtain clearer signal and to get rid of the impurities that were observed. The signal observed at 1999 and 1711 Da do not correspond to any tryptic or chymotryptic SBA peptide, suggesting an external contamination of sample that can come from numerous places such as the instruments microfluidics, injector, or during the sampling process. Impurities at 1221 Da and 1252 Da does however match chymotryptic peptides from human Keratin (Ker [306-314] and Ker [266-276]).

All in all, these results are showing the potential of the surface immobilization of carbohydrate for the simultaneous identification of carbohydrate binding sites and measurements of thermodynamics between lectin-glycan complexes.

### 3. EXPERIMENTAL PART

#### 3.1. Materials and Reagents

##### Lectins, carbohydrates and Proteolytic Enzymes

Recombinant human galectin-3 was obtained from Carbosynth, Compton (UK). Hgal-3 was purified from extracts of bacteria by affinity chromatography on lactosylated Sepharose 4B and analyzed for purity by one- and two-dimensional gel electrophoresis and gel filtration. It was checked for activity by solid-phase and cell-based assays. Glycine max lectin isolated from soy bean was purchased from Medicago AB (Uppsala, Sweden) and was purified by affinity chromatography. Trypsin and chymotrypsin, sequencing grade, were obtained from Promega (Mannheim, Germany). Clostripain was ordered from Worthington biochemicals (Lakewood, NJ, USA). GalNAc, Mannose and  $\beta$ -lactose were obtained from Sigma Aldrich (Steinheim, Germany). Blood group A- and B-trisaccharides and -tetrasaccharide were obtained from Carbosynth Ltd. (Compton, UK).

##### Solid phase peptide synthesis

Amino acids (N- $\alpha$ -Fmoc protected), TGR resin, Benzotriazol-1-yl-N-oxy-trispyrrolidino-phosphonium-hexafluoro-phosphate (PyBOP), N-methylmorpholine (NMM), piperidine and Dimethylformamide (DMF), and Methylpyrrolidine were ordered from Applied Biosystem (Foster City, CA, USA). (Boc-methylaminoxy)acetic acid was ordered from polypeptide (Strasbourg, France).

## Reagents

The following commercially available reagents were used in this work: Sepharose 4B, Urea,  $\alpha$ -cyano-4-hydroxycinnamic acid (HCCA), di-hydroxy benzoic acid (DHB), Electrospray tuning mix solution, Urea, Dithiothreitol (DTT), TEMED, Coomassie Brilliant Blue G250, polyoxyethylen-sorbitanmonolaureat (Tween 20), iodoacetamide, 16-Mercaptohexadecanoic acid, and N-hydroxysuccinimide (NHS) were purchased from Sigma Aldrich (Hamburg, Germany). Acrylamide/bis solution (30 % acrylamide), sodium dodecyl sulfate (SDS), acetonitrile (MeCN), tris-(hydroxymethyl)-aminomethane (Tris), Rotiphorese Gel 30 and ethanol were purchased from Roth (Karlshruhe, Germany). N-(3-dimethylaminopropyl)-N-ethylcarbodiimide hydrochloride (EDC), sodium dihydrogen phosphate monohydrate (NaH<sub>2</sub>PO<sub>4</sub>), sodium acetate trihydrate were purchased from Merck (Darmstadt, Germany). The Gold chips were purchased from SAW Instruments (Bonn, Germany). Trifluoroacetic acid (TFA), dihydroxy benzoic acid (DHB) were purchased from Fluka Chemie (Hamburg, Germany). All reagents were ordered from the highest purity available

## 3.2. Chemical modification and hydrolysis of proteins

### 3.2.1. Reduction and alkylation

Cysteine-containing proteins were reduced and alkylated by first denaturing 1 mg of protein in 100  $\mu$ L of a 6M urea, 100 mM Tris buffer solution. The sample was then reduced for 1 hour by the use of 5  $\mu$ L of 200 mM DTT, 100 mM Tris solution under gentle vortex. 20  $\mu$ L of alkylating agent (200 mM iodoacetamide and 100 mM Tris) was then added to the mixture and mixed by gentle vortex for 1 hour in the dark. All unreacted iodoacetamide was then consumed by adding 20  $\mu$ L of the reducing agent and gentle mixing for 1 hour. Urea concentration was then lowered by diluting the sample with 775  $\mu$ L MQ water, in order for the enzymes to retain activity.

### 3.2.2. Proteolytic digestion

#### 3.2.2.1. Trypsin

Trypsin is a serine protease that cleaves after lysine and arginine residues, except when either of them are followed by proline. The ordered trypsin is from so called “gold” grade, with lysine residues modified by methylation, allowing a higher activity and higher resistance to enzyme autolysis. The specificity of the purified trypsin is further improved by TPCK treatment, which inactivates chymotrypsin. The treated trypsin is then purified by affinity chromatography and lyophilized.

Proteins digested in-solution were diluted in 100 mM Tris buffer, pH 7.8 at concentrations of about 0.5-1.5  $\mu\text{g. } \mu\text{L}^{-1}$ . The digestions were carried out by adding trypsin to the solution with an enzyme-to-substrate ratio from 1:20 to 1:50 depending on the protein, and incubated for 6-18h at 37°C under gentle shaking. Aliquots were taken at frequent intervals during the incubation and analyzed by 1D-SDS-PAGE and MALDI-MS in order to control the optimal digestion yield. Once the protein has been completely digested, the hydrolysis is stopped by lowering the pH of the mixture to pH = 2-3 with TFA, and 100  $\mu\text{L}$  aliquots were kept at -20°C.

#### 3.2.2.2. Chymotrypsin

Sequencing grade chymotrypsin is highly-purified serine endopeptidase derived from bovine pancreas that preferentially cleaves at the carboxyl side of aromatic amino acids Tyr, Phe and Trp, except when either is followed by proline. Cleavages at low rate can also be observed after Met, Leu, and His amino acids residues (125) (126).

In-solution digestion of proteins using chymotrypsin alone, or in a mixture with trypsin, has been carried out using the same procedure as trypsin alone but with 100 mM Tris, 10 mM  $\text{CaCl}_2$  pH 8.0 as digestion buffer, as calcium has been shown to enhance chymotrypsin activity (127).

### 3.2.2.3. Clostripain

Clostripain (endoproteinase Arg-C) is a two-chain cysteine proteinase highly specific for the carboxyl peptide bonds of arginine. Cleavages at low rate can also be observed after lysine amino acids (128). Clostripain has a sulfhydryl requirement; it is activated by dithiothreitol, cysteine, or other sulfhydryl containing reagents, and requires calcium ions to release its activity (129). Optimal pH for the enzyme is pH 7.4-7.8.

Proteolytic digestion of glycine max lectin has been performed according to the following protocol: 100  $\mu$ g SBA was diluted to 100  $\mu$ L with 100 mM Tris, 10 mM  $\text{CaCl}_2$ , pH 7.6. 1 mM DTT was then added to the solution and shaken gently. Clostripain was then added to a ratio of 1:50.

### 3.2.3. Pressure-enhanced digestion

In-solution digestion procedure and proteolytic excision has been performed under pressure cycles with the use of a Barocycler. Cycles of high pressure during proteolytic digestion uses alternating high hydrostatic pressure to facilitate thermodynamic perturbation of molecular interactions and has been shown to increase hydrolysis speed and recovery and improve sequence coverage.

Typically, a mixture of protein and enzyme as described above was put in the barocycler at 30-50°C, with pressure cycles of 30-40 kpsi for 50 seconds and atmospheric pressure during 10 seconds, repeated for 30 to 120 cycles, depending on the size and folding of the protein.



### 3.3. Separation methods

#### 3.3.1. Electrophoresis

1D-SDS-Page was used to analyze fractions from affinity chromatography and to control the rate of protein hydrolysis. In this method, samples are weighed and dissolved in sodium dodecyl sulfate (SDS), a negatively charged detergent that has both hydrophilic and hydrophobic regions. SDS bind to the hydrophobic part of the proteins, and upon application of an electric field, the negative charge of the SDS causes the proteins to move through a clear acrylamide matrix toward the positive electrode. Large proteins move more slowly through the gel matrix than smaller proteins, thus separating proteins by molecular weight. Molecular weight standards are then used in order to determine the mass of the separated proteins.

A Mini-PROTEAN 3 system and a Powerpac 1000 power supply (Bio-Rad, München, Germany) were employed for casting and running the gels. The running gels and the stacking gels were made fresh before use according to the description in table 2. The stacking gel has typically a lower concentration of Acrylamide and therefore large sized pores that allow the proteins to migrate freely and get stacked at the interface between stacking gel and running gel, in order to make sure that the proteins start migrate from the same level.

**Table 2.** Composition of running and stacking gels

Solvent	Stacking gel (ml)	Running gel (ml)
H <sub>2</sub> O	3,075	7,200
1,5M Tris-HCl, pH 8,8	-	7,50
0,5M Tris-HCl, pH6,8	1,25	-
20% (w/v) SDS	0,025	0,150
Acrylamide/Bis-acrylamide (30%/0,8% w/v)	0,670	15,000
10% (w/v) ammonium persulfate (APS)	0,025	0,150
TEMED	0,005	0,020

The samples were solubilized in 50 mM Tris-HCl pH 6.8, 2 mM Dithiothreitol, 2 % SDS, 5 % glycerol, 0.01 % bromophenol blue, pH 6.8. and denatured for 5-10 minutes at 95°C. The running buffer was composed of 25 mM Tris, 200 mM Glycine and 0.1 % SDS. Gel electrophoresis was carried out using at a constant voltage of 60 V for about 15 minutes, until the tracking dye entered the separating gel, and at 120 V for ~2 h, until the tracking dye reached the end of the separating gel.

The bands were then stained by incubating the gels in a Coomassie brilliant blue solution (500mg Coomassie Brilliant Blue in a mixture of 20 ml acetic acid, 90 ml methanol and 90 ml water) for 30 minutes with shaking. The gels were then destained by incubating for 60 minutes in a solution of acetic acid, methanol and water (10:45:45) with shaking.

### 3.3.2. Liquid chromatography

Reversed - phase high performance liquid chromatography (RP-HPLC) is the most common chromatographic mode for identification, quantification and separation of any type of proteins, peptides or polymers. RP-HPLC separates molecules based on differences in hydrophobicity by using a hydrophobic stationary phase.

Semi-preparative RP-HPLC was performed on a Waters 2795 Separation module at a flow rate of 1 mL/min on a Waters Symmetry C18 column (300Å, 5 µm, 4.6 mm X 250 mm). Chromatograms were recorded with an UV-detector at 215 nm. The solvent used were 0.1 % TFA (solvent A) and 90% ACN, 0.1 % TFA (solvent B). The gradient used is resumed in table 3.

**Table 3.** Gradient used for the separation and purification of the synthetic peptide

Time (min)	%B
0	5
5	5
30	30

The injection volume was set to 50  $\mu\text{L}$  and the column temperature to 40°C. Upon detection of the peptide by the UV detector, the PEEK cable exiting the detector was manually disconnected and the peptide which was exiting the detector cell was collected in an Eppendorf. The experiment has been reproduced several times in order to obtain a large amount of purified peptide.

The identity of the purified peptide was confirmed by MALDI-MS and LC-MS.

### 3.3.3. Sample cleaning and desalting

Pierce C18 Spin Tips (Thermo Scientific, Karlsruhe) were used prior to MALDI-MS analysis in order to concentrate the samples and to remove interference from the buffer or the samples (e.g., urea, guanidine, NaCl, Tris, phosphate).

Each tip contains a C18 reversed-phase sorbent that can bind up to 10 $\mu\text{g}$  of total peptide. The desalting procedure consist of inserting the spin tip into a spin adapter seated in an Eppendorf tube. The tip is then wetted by adding 20 $\mu\text{L}$  of 0.1% TFA in 80% ACN and centrifuging (Biofuge 13, Hereaus) at 1000 turns per minute. The tip was then equilibrated by adding 20 $\mu\text{L}$  of 0.1% TFA and centrifuging again at 1000  $\times$  g for 1 minute.

The spin tip and adapter were then transferred to a new microcentrifuge tube and 20-50 $\mu\text{L}$  of sample were added, and the spin tip centrifuged at 1000  $\times$  g for 1 minute.

The tip was then washed by adding 20 $\mu\text{L}$  of 0.1% TFA and centrifuging at 1000  $\times$  g for 1 minute. This step was repeated 2-3 times.

The tip and adapter were then transferred to new Eppendorf tube and the sample was eluted by adding 20 $\mu\text{L}$  of 0.1% TFA in 80% ACN and centrifuging at 1000  $\times$  g for 1 minute. This step was repeated 2-3 times

Additionally, the sample were then concentrated by centrifuging and evaporating (Concentrator 5301, Eppendorf) at 30°C under vacuum.

### 3.4. Affinity methods

#### 3.4.1. Preparation of immobilized carbohydrate Sepharose columns

Thoroughly washed Sepharose 4B (5ml) is resuspended in 5 ml 0.5M Na<sub>2</sub>CO<sub>3</sub> pH=11 and 0.6 ml divinyl sulfone is added dropwise at RT under constant stirring. To avoid beads impairments, sample rotator is preferable to a magnetic stirrer in the following steps. The activation using divinyl sulphone takes 70 minutes at room temperature, in the dark with constant mixing followed by washing with 0.5M Na<sub>2</sub>CO<sub>3</sub> pH=11, and the activated material is stable for up to 12 months in aqueous suspension at 4 °C (130). The gel beds must never be sucked dry at this stage.

Resuspension was done with a 5-10% carbohydrate solution 0.5M NaHCO<sub>3</sub> pH=10. The vinyl groups are highly reactive with hydroxylic compounds, therefore coupling takes place at lower pH and temperatures than with other methods. The coupling with carbohydrates takes 15 hours, at room temperature in the dark.

The affinity matrix was then washed with 5 mL 0.5 M Na<sub>2</sub>CO<sub>3</sub> pH=11. Blocking of unreacted vinyl groups was carried out by incubation with 1 M ethanolamine, pH=8.5 for 1 h at 25 °C. Next, the matrix was washed sequentially with 0.1M ammonium acetate containing 0.5 M NaCl, pH 4 and 0.1 M Tris-HCl / 0.5 M NaCl, pH 8. Finally, the matrix was equilibrated in 25 mM Na<sub>2</sub>HPO<sub>4</sub> / 20 mM NaCl, pH 7.5 and stored at 4 °C.

In addition to glycoprotein and glycopeptides, p-aminophenyl glycosides and mono and di saccharides can be attached to the matrix. In this case, conjugation of the sugar oligomers to the free end of a divinyl sulfone molecule takes place via the hemiacetal hydroxyl group, providing an adequate spacer between the receptor-reactive ligand and the matrix.

The Sepharose blank used for the control experiments were prepared as described above, except without incubation with a carbohydrate solution.

### 3.4.2. Proteolytic extraction and excision

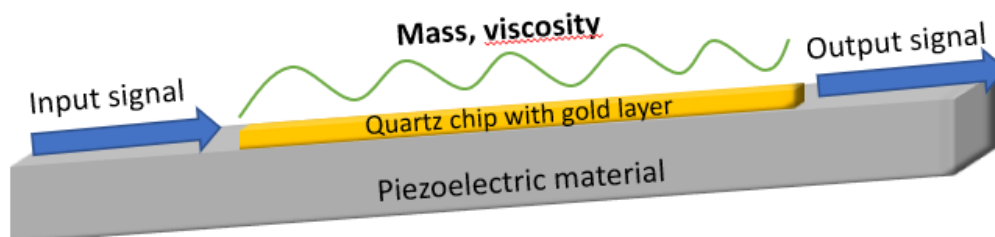
The SBA digest mixtures (see chapters 3.2.2. and 3.2.3.) were diluted 1:2, respectively, using HEPES buffer prior to the addition onto the affinity column containing immobilized carbohydrates. After incubation for 12 h at 37 °C the supernatant was removed and analyzed by MALDI-MS. Any traces of free peptides were washed out from the column with four times 400 µL HEPES, and the washing fraction was analyzed by MALDI-MS until no peptide was observed. The peptides bound to the immobilized carbohydrate were eluted under shaking with 400 µL ACN/0.1% TFA (2:1) at 37 °C for 15 min, and the procedure repeated twice. The elution fractions were then combined, concentrated, and analyzed by MALDI-MS.

### 3.4.3. Pressure enhanced proteolytic excision

A solution of 20 µg SBA in respectively 200 µL HEPES buffer, pH 7.5, was added to 200 µL affinity matrix and allowed to bind for 12 h at 37 °C. Any free (unbound) lectin was washed out with 2 × 400 µL binding buffer and the washing fractions were analyzed by MALDI-MS. The binding buffer was then switched to Tris-buffer (with 5 mM MnCl<sub>2</sub> added to allow carbohydrates to interact with SBA) and proteolytic digestion of bound lectin was performed by application to high-pressure cycles (30 kpsi for 50 s, atmospheric pressure for 10 s, 50 °C, 90 cycles) with a mixture of trypsin and chymotrypsin (1:20 and 1:30). Following digestion, the supernatant was analyzed by MALDI-MS. Free (unbound) peptides were washed out with 4 × 400 µL binding buffer and the washing fractions analyzed by MALDI-MS. The affinity-bound peptides were then eluted with two times 200 µL ACN/0.1% TFA 2:1 at 37 °C for 15 min under shaking.

#### 3.4.4. Surface acoustic wave biosensor

Binding studies were performed with a surface acoustic wave (SAW) biosensor (K5-Ssens; SAW-Instruments, Bonn, Germany; Nanotemper GmbH, München, Germany). SAW instruments are based on surface acoustic waves to measure the binding events on the surface of the sensor (figure 46). The waves are produced through inverse piezoelectric effect on the surface of quartz chip covered with a thin layer of gold. Viscosity changes and mass loadings on the chip's surface affect the phase and amplitude of the acoustic waves, which are transformed back into electrical signal through direct piezoelectric effect (131). The biosensor employs special shear waves of Love type to achieve high sensitivity in detecting interactions that take place in solution. Love waves, proper to a very thin layer of substance, coupled with displacement of matter parallel to the interface solid-liquid permit a high conservation of wave energy making them very sensible to surface effects (i.e. mass loading and viscosity changes) and lowers the noise level in the signal.



**Figure 46.** Schematic of the SAW biosensor principle. The piezoelectric material is used to convert the electric field to a mechanical wave. Upon changes of mass or viscosity on the surface during binding events, the output signal will change in amplitude and phase and the difference will be converted into an electrical signal for the monitoring.

The central quartz chip is covered with a thin layer of gold in order to immobilize Sulphur containing compounds. Any changes in mass loading on the surface will cause phase shifts in the output signal, which are used to monitor the binding events. Changes

in viscosity, however, will produce changes in both phase and amplitude and can therefore be monitored separately (132).

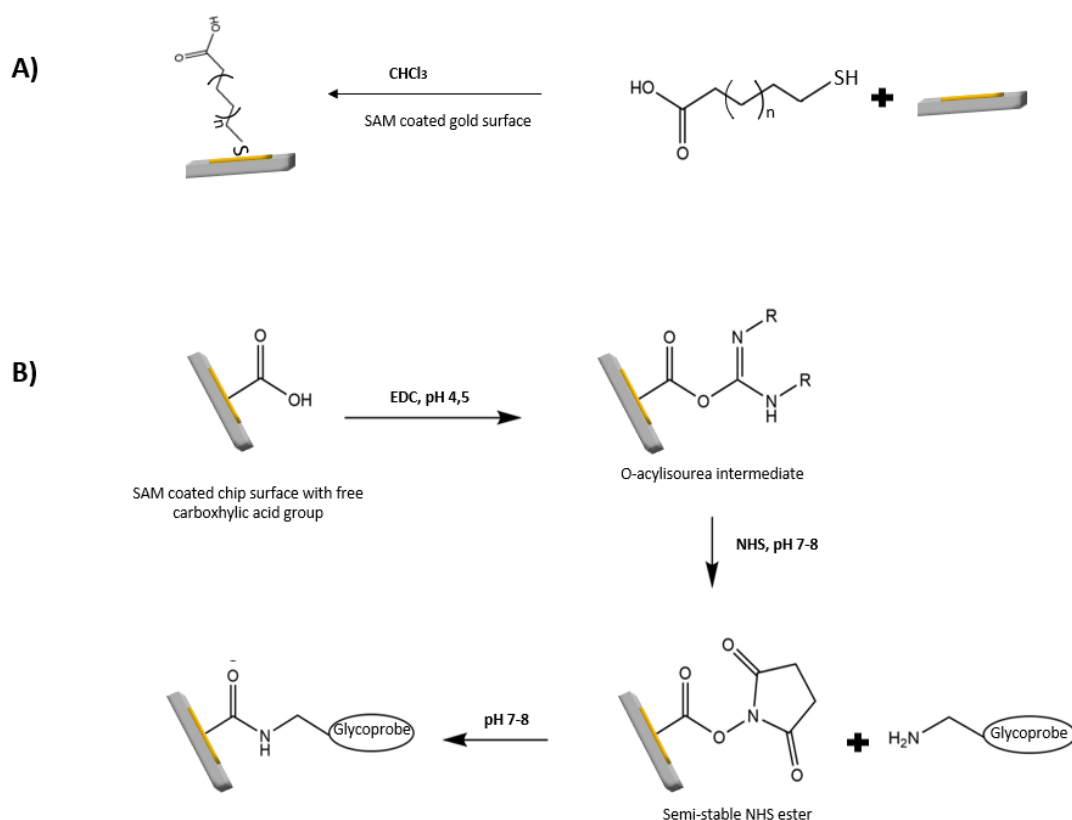
In the present work, a self-assembled monolayer (SAM) of 16-mercaptohexadecanoic acid was used as a linker on the surface of the chip (133).

Prior to use, all gold chips were cleaned by inserting them in a solution of 1:1  $\text{H}_2\text{SO}_4:\text{H}_2\text{O}_2$  (30%) and agitated for 30 min-1 hour (max) at room temperature. The chips were then washed with MQ- $\text{H}_2\text{O}$  and ethanol and then dried.

The cleaned SAW chips were inserted in a 10 mM 16-mercaptohexadecanoic acid in  $\text{CHCl}_3$  solution and softly agitated for 12-16h at room temperature (figure 47a). The chips were then washed with ethanol and dried under a stream of nitrogen air before use, to avoid the formation of any bubbles that would be formed in the flow cell.

The SAM coated chip was then inserted into the instrument and wetted with a  $200 \mu\text{L min}^{-1}$  flow of deionized and degassed water to expel the air trapped in the flow cell. The flow was then switched to  $30 \mu\text{L min}^{-1}$  for all subsequent operations.

The glycoprobe was immobilized on the SAM by carboxyl-group activation with a 1:1 mixture (v/v) of 200 mM (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and 50 mM N-hydroxysuccinimide (NHS). A  $10 \mu\text{M}$  aqueous glycoprobe solution (pH 6.5) was used for immobilization (figure 47b), followed by capping of unreacted carboxyl groups with 1 M ethanolamine, pH 8.5.



**Figure 47.** Chemistry of the glycoprobe immobilization on the SAW gold chip. (a) Immobilization of the SAM linker on the gold surface. (b) Activation of the carboxyl groups of SAM (simplified representation) and covalent attachment of the glycoprobe via amino group.

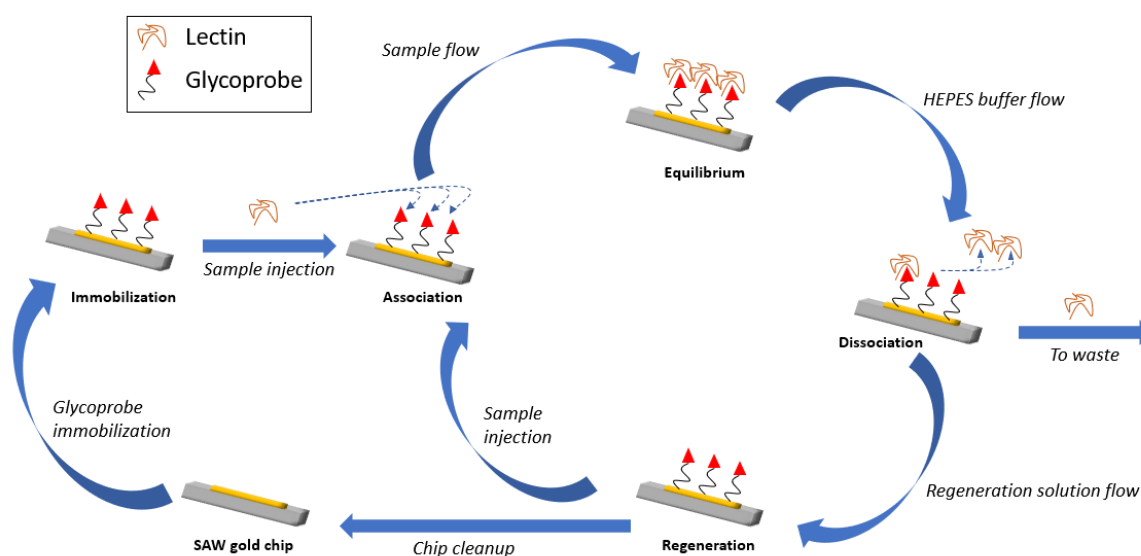
After immobilization of the glycoprobe, the running buffer is changed to HEPES buffer (10 mM HEPES, 25 mM NaCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, pH 7.4) for all subsequent operations.

150  $\mu\text{L}$  solutions of increasing concentrations of intact lectins were injected on the chip with the immobilized glycoprobe at a flow rate of 30  $\mu\text{L min}^{-1}$ . After each injection, the affinity pair was left to dissociate for 5 min and the chip surface was regenerated with 60% ACN, pH 6.5. After 5 min equilibration time the next injection was allowed. The general workflow of a SAW experiment is illustrated in figure 48.



After fitting the recorded data according to the theoretical 1:1 Langmuir binding model using the program OriginPro 7.5 (OriginLab Corporation, Northampton, USA) and FitMaster addon for Origin (SAW instruments, Bonn, Germany).

The extracted observed rate constants ( $k_{\text{obs}} = k_{\text{off}} + C_{\text{analyte}} \times k_{\text{on}}$ ) were plotted versus the analyte concentrations. Linear regression of the data points yielded the association ( $k_{\text{on}}$ ) and dissociation ( $k_{\text{off}}$ ) rate constants that were used to obtain the equilibrium dissociation constant  $K_D = k_{\text{off}} \times k_{\text{on}}^{-1}$ . A negative control for determination of unspecific binding was performed by injecting a solution of 0.5–20  $\mu\text{M}$  myoglobin in HEPES at identical conditions as described above.



**Figure 48.** Scheme of the general SAW workflow. The plain SAW gold chip is coated with SAM and the glycprobe is immobilized on it. The analyte is then injected and association takes place between the analyte and the glycprobe, until it reaches an equilibrium level, where the amount of associating analyte is equal to the amount of analyte dissociating. Once the sample injection is over, HEPES buffer is flown over the sensor chip and the analyte is slowly dissociated from the glycprobe. The association and dissociation events are monitored by the SAW software. Once the analytes are dissociated, the surface of the chip is being regenerated to get rid of the remaining bound analyte, and to prepare the surface for a subsequent injection.

Because of the relative stability of the glycoprobes compared to antibodies, a SAW gold chip with immobilized glycoprobe has been shown to be stable and reusable for up to 10 days if kept at room temperature under a slow flow of HEPES buffer ( $20\mu\text{L}\cdot\text{min}^{-1}$ ).

### 3.4.5. Carbohydrate-glycoprobe conjugation

Conjugation of N[Me]-O-Aoa-GFKKG-amide to disaccharides was done at 20 and 25 mM, respectively in 0.1 M sodium acetate, at pH 3.5 for NAc-disaccharides and pH 4.6 for lactose. After 72 h at 37 °C, the glycopeptides were purified by semipreparative HPLC on SphereClone C18 (Phenomenex,  $250 \times 10$  mm; 5  $\mu\text{m}$ ) using a 10–20% linear gradient of acetonitrile into water (both eluents with 0.1% TFA). Glycopeptide-containing fractions were neutralized with 10 mM ammonium bicarbonate to prevent acid degradation and lyophilized. All disaccharide-N[Me]-O-GFKKG-amide glycopeptides had the expected mass by MALDI-TOF MS.

## 3.5. Mass spectrometric methods

### 3.5.1. ESI-Ion trap MS

ESI-MS measurements were carried out with an Esquire 3000+ ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). The samples were dissolved in 50 % MeOH/ 1% AcCOOH, at 10-50  $\mu\text{M}$  concentrations, and measured by direct infusion to MS from a Harvard Apparatus Model 44 programmable syringe pump (Harvard Apparatus, Kent, UK) set at a flow rate of 5-10  $\mu\text{L}/\text{min}$ . Mass spectra were recorded in positive-ion mode, scanning from  $m/z$  50 to 2500. Ion source parameters were: nebulizing gas (nitrogen) pressure 15-20 psi, drying gas (nitrogen) flow rate 6 L/min, drying gas temperature 200-250 °C, capillary voltage -3.5 kV, end plate offset 500 V, capillary exit 80-120 V.

### 3.5.2. MALDI-MS

A Bruker Autoflex Smartbeam mass spectrometer equipped with external fully automated X-Y target stage MALDI source with pulsed collision gas was employed for MALDI-MS analyses. The pulsed nitrogen laser was operated at 337 nm and ions generated by 10 laser shots were accumulated for 0.5–1 s at 30 V and extracted at 15 V. Mass spectra were obtained using 10 laser shots for each scan and accumulating 30–60 scans. The matrix solution was prepared by dissolving 10 mg of 2,5-dihydroxybenzoic acid in 1 mL acetonitrile/0.1% TFA in water (2:1 v/v). A 0.5  $\mu$ L aliquot of matrix solution was deposited on the MALDI target, mixed with 0.5  $\mu$ L of sample solution and allowed to dry in ambient air. External calibration was performed using the monoisotopic masses of singly protonated ions of human angiotensins I and II, bradykinin, neurotensin, bovine insulin  $\beta$ chain (oxidized), and bovine insulin.

Aliquots of 1  $\mu$ L of the sample solution and the saturated matrix solution were mixed on the stainless steel MALDI target and allowed to dry. Acquisition of spectra was carried out at an acceleration voltage of 20 kV and a detector voltage of 1.5 kV.

#### 3.5.3.1. ESI-Quadrupole mass spectrometry

ESI-quadrupole mass spectrometry (ESI-QMS) was performed using a Waters Quattro Ultima with an electrospray source. Samples were typically dissolved in 50% ACN with 1% formic acid and introduced at a flow rate of 5–10  $\mu$ L.min<sup>-1</sup> into the ESI with a syringe pump (Harvard Apparatus Model 44, Kent, UK), where ionization takes place in the source at atmospheric pressure. These ions are sampled through a series of orifices and ion optics into the first quadrupole where they are filtered according to their mass to charge ratio ( $m/z$ ). The mass separated ions then pass into the ion tunnel collision cell, with axial field, where they either undergo Collision Induced Decomposition (CID), or pass unhindered to the second quadrupole. The fragment ions are then mass analyzed by the second quadrupole. Finally, the transmitted ions are detected by a conversion dynode, phosphor, and photomultiplier detection system. The

output signal is amplified, digitized, and presented to the data system, where Mass spectra were recorded in positive-ion mode, scanning from  $m/z$  50 to 3000.

Ion source parameters were: Capillary voltage 3.2 kV, sample cone 80V, Source temperature 100°C and desolvation temperature 250°C. Gas flow rate was set to 200 l/h.

### 3.5.3.2. High-performance liquid chromatography mass spectrometry

High-performance liquid chromatography quadrupole mass spectrometry (LC-QMS) measurements were performed using a Waters 2795 Separation module (Waters, Eschborn, Germany) coupled to the Waters Quattro Ultimate quadrupole mass spectrometer. A flow splitter was used at the exit of the HPLC column to divert the flow both to the mass spectrometer through a low diameter PEEK cable and to the UV-detected using a standard PEEK tube. The difference in PEEK cable diameter at the exit of the flow splitter allowed both the use of higher flow rate during the chromatographic separation, necessary for a proper separation, and the obtention of a lower flow rate inside the small PEEK diameter cable, which is more suitable to the MS system.

The solvent system employed was 0.5 % formic acid in MQ-water (solvent A) and 0.3 % formic acid in 95 % acetonitrile (solvent B). The separation gradient was from 2 % to 30-80 % Solvent B over 20-60 minutes, depending on the samples analyzed. The columns were heated at 30-40 °C. Peptides usually eluted at 10-30% solvent B when using C18 reversed phase column (see also chapter 3.3.2.).

The flow rate used was 0.8 mL/min. The samples were dissolved in solvent A and 5-50  $\mu$ L were used for one injection. The columns employed for separation were a Waters Symmetry C18 column (300Å, 5  $\mu$ m, 4.6 mm X 250 mm) for large amount of peptide separation and an Acclaim Pepmap C18 (4.6x150 mm, 300Å, 5  $\mu$ m) for precise small amount peptide identification.

Mass spectra were recorded in positive-ion mode, scanning from  $m/z$  50 to 2500. Ion source parameters were: Capillary voltage 3.2 kV, sample cone 80V, Source temperature 120°C and desolvation temperature 250°C. Gas flow rate was set to 250 l/h.

UV spectrums were recorded in dual 215 nm and 280 nm channel.

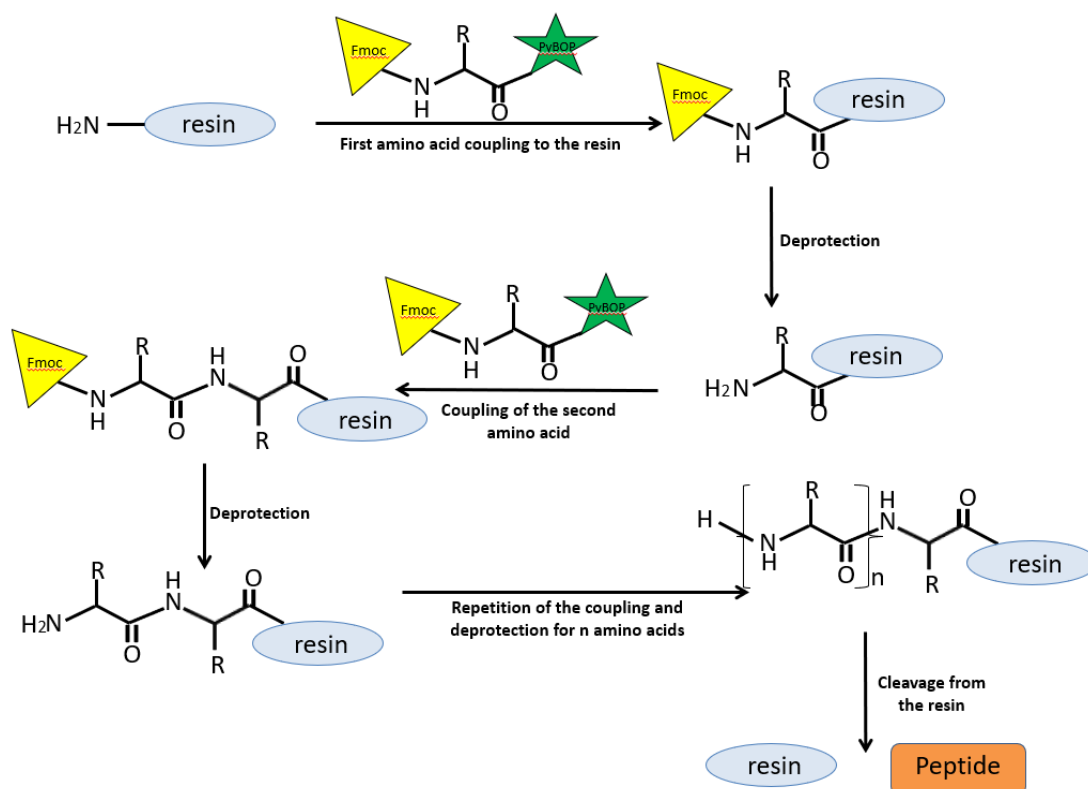
### **3.6. Solid phase peptide synthesis**

The peptides were synthesized by Fmoc-based solid-phase synthesis on a NovaSyn TGR resin using Fmoc/tBu chemistry at a 0.1 mmol scale in an ABI 433A peptide Synthesizer (ThermoFisher, Darmstadt, Germany) and using the Synthassist Software.

The resin was first swelled with DMF for 30 minutes to reach the maximum volume and to assure the entire surface accessibility, then the peptide was synthesized following several steps of coupling and deprotection as illustrated in figure 49.

Removal of the Fmoc group was performed with a series of washing using 20% piperidine in DMF and was followed by washing with DMF to remove side products.

Coupling of a new amino acid was done after activation of the Fmoc-amino acid with 0.9 M PyBOP and 1.3 M NMM in DMF for 30 minutes, followed by washing with DMF, DCM, and again DMF.



**Figure 49.** Schematic representation of Fmoc-based solid-phase synthesis

After coupling of the last amino acid, the resin was washed with ethanol and dried under vacuum. Cleavage of the peptide from the resin was done together with the deprotection of the side chains with a mixture of trifluoroacetic acid / triethylsilane / water (95:2,5:2,5, v:v:v) for 2 hours at room temperature. The peptide was then precipitated with 40 mL cold diethylether. The resin and the crude peptide were separated by filtration using 10 % acetic acid and then the peptide collected and lyophilized.

### **Glycopeptide synthesis**

The N[Me]-O-Aoa-GFKKG peptide was obtained by conversion of Boc-methylaminoxyacetic acid.DCHA (Boc-N[Me]-Aoa-OH/DHCA) salt into a free carboxyl form by acid extraction with 0.1M HCl and ethyl acetate. The organic phase containing Boc-N[Me]-Aoa-OH was recovered from three consecutive extractions, dissolved in DCM and lyophilized.

Boc-N[Me]-Aoa-OH was coupled manually in the same conditions as the other amino acids.

Conjugation between N[Me]-O-Aoa-GFKKG peptide and the saccharides was done at 20 and 25 mM respectively, in 0.1 M NaOAc pH=4.6 for 72 hours at 37°C. All conjugations were controlled by MALDI-MS for correct mass and purified by preparative RP-HPLC, as described above.

### **3.7. Offline coupling of SAW with MALDI-MS**

Extensive washing of the SAW instrument microfluidics was done prior to any measurement by flowing 0.1M Glycine, 0.1M HCl and ethanol for 30 minutes each at 200  $\mu\text{L}\cdot\text{min}$

Analysis of the elution fractions from immobilized carbohydrates on a SAW gold chip by MALDI-MS was done by injecting the aliquot of SBA digest (see also chapter 3.2.2. and 3.2.3.) into the microfluidic system of the SAW instrument in the same way as described before, with the chip being heated to 37°C. Once the signal detects the passage of the sample over the surface of the chip, the flow was stopped and the sample allowed to incubate over the chip surface for up to 12 hours. The flow was then resumed to 20  $\mu\text{L}\cdot\text{min}^{-1}$  until all the sample was passed over the chip. The surface was then washed with HEPES buffer at a flow rate of 40  $\mu\text{L}\cdot\text{min}^{-1}$  for 2 hours.

Aliquots were sampled at the waste exit of the instrument every 30 min, concentrated, desalted and check for any remaining unbound peptide.

---

After 2 hours, two times 150  $\mu\text{L}$  of elution solution were injected (ACN : 0.1% TFA, 2:1 v/v). Once the passage of the elution solution was detected on the surface of the gold chip, the elution solutions were collected in an Eppendorf tube at the waste exit of the instrument.

The solution solutions were concentrated, the ACN evaporated, cleaned by Ziptip and analyzed by MALDI-MS.



## 4. Zusammenfassung

Die meisten menschlichen Zellen und Krankheitserreger sind mit einer Vielzahl von Kohlenhydraten überzogen, und diese Kohlenhydrate liegen in vielen verschiedenen molekularen Formen vor, z. B. als Glykoproteine, Proteoglykane oder Glykolipide. Diese Kohlenhydrate bestehen aus Anordnungen von Monosaccharid-Bausteinen, die an verschiedenen Positionen ihrer Pyranose- oder Furanoseringe miteinander verbunden sein können, und diese strukturelle Komplexität ist an wichtigen physiologischen und pathophysiologischen Vorgängen beteiligt.

Obwohl sie oft unterschätzt werden, spielen Protein-Kohlenhydrat-Wechselwirkungen eine Schlüsselrolle in zahllosen und vielfältigen biologischen Prozessen, die von der zellulären Erkennung und den intrazellulären Regulationswegen bis hin zu immunologischen Reaktionen oder sogar den Wechselwirkungen zwischen Gameten und Keimzellen reichen, die die Befruchtung einleiten. So werden beispielsweise C-Typ-Lektin-Rezeptoren auf einer Vielzahl von Immunzellen wie Antigen-präsentierenden Zellen (APC) und Endothelzellen exprimiert, die an der Erkennung von Krankheitserregern beteiligt sind. Der erste Kontakt zwischen zwei Zellen oder zwischen einer Zelle und einem Krankheitserreger läuft also mit ziemlicher Sicherheit auf Protein-Kohlenhydrat-Interaktionen hinaus, was die zentrale Bedeutung dieser Interaktion für die Entwicklung von auf Glykane ausgerichteten Arzneimitteln, diagnostischen Zielen und Biomarkern unterstreicht.

Die Untersuchung von Protein-Kohlenhydrat-Wechselwirkungen ist jedoch aufgrund der strukturellen Komplexität der Glykane, der Permissivität und Multiplexität der Lektine und der relativ niedrigen Affinitätskonstanten besonders schwierig. Röntgenkristallographie, NMR und ortsgerichtete Mutagenese sind die gebräuchlichsten Techniken, um die strukturellen Details der Erkennungsvorgänge zu untersuchen. Ihre Anwendung ist jedoch durch die notorischen Schwierigkeiten bei der Kristallisierung von Protein-Glykan-Komplexen mit ausreichend hoher Auflösung sowie durch die hohen Anforderungen an relativ große Mengen und hoch gereinigte Proben begrenzt.

In dieser Arbeit wurde ein neuer Ansatz für die molekulare Identifizierung von Kohlenhydrat-Erkennungsepitopen in Lektinen mit hoher Spezifität, Empfindlichkeit und geringen Anforderungen an die Probenreinheit entwickelt, der eine Kombination aus proteolytischem Verdau und Massenspektrometrie verwendet. Bei einer ersten Form des analytischen Ansatzes, der so genannten proteolytischen Exzision, wird das Kohlenhydrat auf einer Affinitätssäule immobilisiert und das kohlenhydratbindende Protein (Lektin, Antikörper usw.) hinzugefügt, um einen Komplex zu bilden, der dann einem proteolytischen Verdau unterzogen wird. Bei einer zweiten Variante, der proteolytischen Extraktion, wird das Protein zunächst einem proteolytischen Verdau unterzogen und das Gemisch aus Peptidfragmenten dem immobilisierten Kohlenhydrat vorgelegt. Nach dem Auswaschen der nicht bindenden Peptide werden die verbleibenden affinitätsgebundenen Peptidligandenfragmente eluiert und massenspektrometrisch analysiert. Dieser Ansatz wird allgemein als *Carbohydrate Recognition Domain Excision* (CREDEX) bezeichnet.

Bei der CREDEX-Methode war der proteolytische Schritt oft ein großes Problem, da intakte Lektine sehr stabil gegenüber enzymatischer Hydrolyse sind und einen zeitaufwändigen proteolytischen Verdau erfordern. Hier demonstrieren wir die Anwendung des proteolytischen Hochdruckverdaus als effizientes Werkzeug für die massenspektrometrische Identifizierung von kohlenhydratbindenden Peptiden in tierischen und pflanzlichen Lektinen. Im letzten Elutionsschritt ist die kompetitive Elution mit einer Kohlenhydratlösung im Allgemeinen für die spezifische Isolierung von Ligandenepitopen geeignet; sie erfordert jedoch hohe Konzentrationen. Hier zeigen wir, dass eine breite Palette von organisch-wässrigen Eluentenmischungen als Ersatz für die Elution in Frage kommt. In dieser Studie haben wir die CREDEX-MS-Methode mit druckverstärkter Proteolyse zur Identifizierung von Ligandenepitopen von Blutgruppenoligosacchariden an Glycin max Lektin (Soja-Agglutinin, SBA) angewendet. Die Ergebnisse zeigen spezifische Epitoppeptide, die sich in den Strukturen der Lektine befinden und eine molekulare Differenzierung der Oligosaccharidepitope ermöglichen.

Es wurden auch verschiedene Techniken zur Immobilisierung von Zuckern untersucht, sowohl auf einer reinen Goldoberfläche als auch unter Verwendung einer

Sepharosematrix. Die Oberflächenimmobilisierung von Zuckern hat das Potenzial, als wertvolles Werkzeug für die Hochdurchsatzanalyse von Kohlenhydrat-Protein-Wechselwirkungen in der Glykomikforschung zu dienen und ist für die schnelle, quantitative und gleichzeitige Analyse einer großen Anzahl biomolekularer Wechselwirkungen unerlässlich. Acetyl-d-Galactosamin wurde daher auf einer Goldchip-Oberfläche immobilisiert. Dieser Chip wurde als einzige Schnittstelle sowohl für thermodynamische Messungen als auch für die Erfassung kohlenhydratbindender Peptide verwendet.

## 5. References

1. Andre, S., Kaltner, H., Manning, J.C., Murphy, P.V. and Gabius, H.J. (2015) *Lectins: getting familiar with translators of the sugar code. Molecules*;20(2):1788-1823.
2. Sumner, J. B.; Howell, S. F. and Zeissig, A., *Concanavalin A and hemagglutination. Science, 1935*,65-66.
3. Varki A, Cummings RD, Esko JD. *Essentials of Glycobiology. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. Chapter 28.*
4. Sharon, N., *Lectins, in eLS. 2001, John Wiley & Sons, Ltd.*
5. Van Damme EJM, Peumans WJ, Pusztai A, Bardocz S: *Handbook of plant lectins: properties and biomedical applications. 1998, John Wiley & Sons, Chichester, UK.*
6. Rudiger H, Gabius HJ: *Plant lectins: occurrence, biochemistry, functions and applications. Glycoconj J. 2001, 18: 589-613. .*
7. Boyd, W. C. (1954). In "The Proteins", Vol. 2, Part 2, 756-844. Academic Press: New York.
8. Farhud DD, Zarif Yeganeh M. *A brief history of human blood groups. Iran J Public Health. 2013;42(1):1-6. .*
9. Roggenbuck D, Mytilinaiou MG, Lapin SV, Reinhold D, Conrad K (December 2012). *Asialoglycoprotein receptor (ASGPR): a peculiar target of liver-specific autoimmunity. Auto-Immunity Highlights. 3 (3): 119-25.*
10. Edelman, G.M., Reeke, G.N. Jr., Becker, J.W, and Wang, J.L. (1972) *The covalent and three-dimensional structure of concanavalin A. Proc. Natl Acad. Sci. USA, 69, 2580-2584.*
11. Hardman, K.D. and Ainsworth, C.F. (1972) *Structure of concanavalin A at 2.4-Å resolution. Biochemistry, 11, 4910-4919.*

12. Nathan Sharon, Halina Lis, *History of lectins: from hemagglutinins to biological recognition molecules*, *Glycobiology*, Volume 14, Issue 11, November 2004, Pages 53R–62R.
13. Crennel, S., Garman, E., Laver, G., Vimr, E., and Taylor, G. (1994) *Crystal structure of Vibrio cholerae neuraminidase reveals dual lectin-like domains in addition to the catalytic domain*. *Structure*, 2, 535-544.
14. Cummings RD, McEver RP. *C-Type Lectins*. 2017. In: Varki A, Cummings RD, Esko JD, et al., editors. *Essentials of Glycobiology 3rd edition*. Cold Spring Harbor; 2015-2017.
15. Rini, J.M., *Lectin structure*. *Annu. Rev. Biophys. Biomol. Struct.*, 1995. 24: p. 551-77.
16. Nagata, S., Nakanishi, M., Nanba, R., Fujita, N. (2003) *Developmental expression of XEEL, a novel molecule of the Xenopus oocyte cortical granule lectin family*. *Dev. Genes Evol.*, 213, 368-370.
17. Chang, N-C.A., Hung, S-I., Hwa, K-Y., Kato, I., Chen, J-E., Liu, C-H., Chang, A.C. (2001) *A macrophage protein, Ym1, transiently expressed during inflammation is a novel mammalian lectin*. *J. Biol. Chem.*, 276, 17497-17506.
18. Matsushita, M., Fujita, T. (2001) *Ficolins and the lectin complement pathway*. *Immunological Reviews*, 180, 78-85.
19. Yoshida, Y., Tokunaga, F., Chiba, T., Iwai, K., Tanaka, K., Tai, T. (2003) *Fbs2 a new member of the E3 ubiquitin ligase family that recognizes sugar chains*. *J. Biol. Chem.*, 278, 43877-43884.
20. Saito, T., Hatada, M., Iwanaga, S., Kawabata, S. (1997) *A newly identified horseshoe crab lectin with binding specificity to O-antigen of bacterial lipopolysaccharides*. *J. Biol. Chem.*, 272, 30703-30708.
21. I. Ofek, D.L. Hasty and R.J. Doyle, *Bacterial Adhesion to Animal Cells and Tissues*, ASM Press, Washington, DC, 2003, 416.
22. K.A. Karlsson, *Biochem. Soc. Trans.*, 1999, 27, 471.

23. Coelho, Luana Cassandra. (2014). *Lectins: Function, structure, biological properties and potential applications*. *Research Trends*, 15, 41-62. .
24. El Dib R., Gomaa H., Carvalho R.P., Camargo S.E., Bazan R., Barretti P., Barreto F.C. *Enzyme replacement therapy for Anderson-Fabry disease*. *Cochrane Database Syst Rev*. 2016 Jul 25;7:CD006663.
25. Linthorst, Gabor E. et al. *Enzyme therapy for Fabry disease: Neutralizing antibodies toward agalsidase alpha and beta*, *Kidney International* , Volume 66 , Issue 4 , 1589 - 1595.
26. Belický Š, Katrik J, Tkáč J. *Glycan and lectin biosensors*. Estrela P, ed. *Essays in Biochemistry*. 2016;60(1):37-47.
27. Olsen L.R., Dessen A., Gupta D., Sabesan S., Sacchettini J.C., Brewer C.F. *X-ray crystallographic studies of unique cross-linked lattices between four isomeric biantennary oligosaccharides and soybean agglutinin*. *Biochemistry*, Vol. 36, No. 49, 1997.
28. Sinha, S., G. Gupta, M. Vijayan, and A. Surolia, *Subunit assembly of plant*.
29. F.A. Quioco, *Biochem. Soc. Trans.*, 1993, 21, 445.
30. Y. Bourne, P. Rouge and C. Cambillau, *J. Biol. Chem.*, 1992, 267, 97.
31. J. Bouckaert, J. Berglund, M. Schembri, E. De Genst, L. Cools, M. Wuhrer, C.S. Hung, R. Slattegard, A. Zavialov, D. Choudhury, S. Langermann, S.J. Hultgren, L. Wyns, P. Klemm, S. Oscarson, S.D. Knight, H. De Greve, *Mol. Microbiol.*, 2005, 55, 411.
32. C. Cambillau, in *Glycoproteins*, J. Montreuil, H. Schachter and J.F.G. Vliegthart (eds), Elsevier, Amsterdam, 1995, 29.
33. H. Lis and N. Sharon, *Chem. Rev.*, 1998, 98, 637.
34. Varki A, Schnaar RL, Crocker PR. *I-Type Lectins*. 2017. In: Varki A, Cummings RD, Esko JD, et al., editors. *Essentials of Glycobiology [Internet]*. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. Chapter 35.

35. Varki A, Gagneux P. *Biological Functions of Glycans*. 2017. In: Varki A, Cummings RD, Esko JD, et al., editors. *Essentials of Glycobiology* [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. Chapter 7.
36. Lutteke, T., *The use of glycoinformatics in glycochemistry*. *Beilstein J. Org.*
37. Gabius, H. , Siebert, H. , André, S. , Jiménez-Barbero, J. and Rüdiger, H. (2004), *Chemical Biology of the Sugar Code*. *ChemBioChem*, 5: 740-764.
38. Gabius, HJ. *Naturwissenschaften* (2000) 87: 108.
39. Rini J, Esko J, Varki A. *Glycosyltransferases and Glycan-processing Enzymes*. In: Varki A, Cummings RD, Esko JD, et al., editors. *Essentials of Glycobiology*. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. Chapter 5.
40. An HJ, Froehlich JW, Lebrilla CB. *Determination of Glycosylation Sites and Site-specific Heterogeneity in Glycoproteins*. *Current opinion in chemical biology*. 2009;13(4):421-426. .
41. Drickamer K, Taylor ME (2006). *Introduction to Glycobiology* (2nd ed.). Oxford University Press, USA.
42. Vliegenthart, J.F., Casset, F.: *Novel forms of protein glycosylation*. *Curr. Opin. Struct. Biol.*8, 565-571 (1998) .
43. Magnani J. L., Ernst B., *glycomimetic drugs--a new source of therapeutic opportunities*. *Discovery medicine*. 2009, 8(43):247-52.
44. Bernardi, A., Carrettoni, L., Ciponte, A.G., Monti, D.,Sonnino, S., 2000. *Second generation mimics of ganglioside GM1 as artificial receptors for cholera toxin: replacement of the sialic acid moiety*. *Bioorg. Med. Chem. Lett.* 10, 2197–2200.
45. Fazli, A., Bradley, S.J., Kiefel, M.J., Jolly, C., Holmes, I.D., von Itzstein, M., 2001. *Synthesis and biological evaluation of sialylmimetics as rotavirus inhibitors*. *J. Med. Chem.* 44, 3292–3301.

46. Fan J.-Q., Ishii S., *Active-site-specific chaperone therapy for Fabry disease Yin and Yang of enzyme inhibitors. FEBS J.* 274, 4962-4971.
47. R.M. Zinkernagel and *Annu. Rev. Immunol.*, 2003, 21, 515.
48. *State of the Art of New Vaccines Research & Development, Initiative for Vaccine Research, World Health Organization, Geneva, 2003.*
49. Sozzani, P., Arisio, R., Porpiglia, M., and Benedetto, C. (2008) *Is Sialyl Lewis x antigen expression a prognostic factor in patients with breast cancer? Int. J. Surg. Pathol.* 16, 365–374.
50. K.J. Yarema and C.R. Bertozzi, *Curr. Opin. Chem. Biol.*, 1998, 2, 49.
51. Feng D., Shaikh A. S., Wang F., *Recent Advance in Tumor-associated Carbohydrate Antigens (TACAs)-based Antitumor Vaccines. ACS Chem. Biol.* 2016, 11, 850–863.
52. Haurum, J. S., Arsequell, G., Dwek, R. A., McMichael, A. J., and Elliott, T. (1994) *Recognition of carbohydrate by major histocompatibility complex class I-restricted, glycopeptide-specific cytotoxic T lymphocytes. J. Exp. Med.* 180, 739–744.
53. Wilkinson, B. L., Day, S., Malins, L. R., and Payne, R. J. (2011) *Self-adjuvanting multicomponent cancer vaccine candidates combining per-glycosylated MUC1 glycopeptides and the Toll-like receptor 2 agonist Pam3CysSer. Angew. Chem.*, 50, 1635–1639.
54. Zhou Z., Liao G., Mandal S.S., Suryawanshi S., Guo Z., *A fully synthetic self-adjuvanting globo H-Based vaccine elicited strong T cell-mediated antitumor immunity. Chemical Science* 6.12 (2015): 7112–7121.
55. Carl R., Alving A., *Lipopolysaccharide, lipid A, and liposomes containing lipid A as immunologic adjuvants. Immunobiology*, 187, 430-446.
56. Rietschel E.T., Brade L., Lindner B., Zähringer U., in *Bacterial Endotoxic Lipopolysaccharides, Vol. 1, D.C. Morrison and J.L. Ryan (eds), CRC Press, Boca Raton, FL, 1992, p. 3.*



57. Provenzano D., Kovac P., Wade W.F., *The ABCs (Antibody, B Cells, and Carbohydrate Epitopes) of Cholera Immunity: Considerations for an Improved Vaccine. Microbiology and Immunology*, 50, 12, 899-927.
58. Islam K, Hossain M, Kelly M, et al. *PLoS Neglected*; 12(4):e0006376.
59. Sayeed MA, Bufano MK, Xu P, et al. *PLoS Neglected Tropical Diseases*. 2015;9(7):e0003881. .
60. Moise, A., Maeser, S., Rawer, S., Eggers, F., Murphy, M., Bornheim, J., Przybylski, M., *Substrate and Substrate-Mimetic Chaperone Binding Sites in Human  $\alpha$ -Galactosidase A Revealed by Affinity-Mass Spectrometry. J. Am. Soc. Mass Spectrom.* (2016) 27: 1071. .
61. Pihiková D, Kasák P, Tkac J., *Glycoprofiling of cancer biomarkers: Label-free electrochemical lectin-based biosensors. Open Chem.* 2015 Jan; 13(1):636-655.
62. Smith E. A., Thomas W. D., Kiessling L. L., Corn R. M., *Surface plasmon resonance imaging studies of protein-carbohydrate interactions. J. Am. Chem. Soc.* 2003, 125 (20), 6140–6148.
63. Park S., Lee M.R., Pyo S.J., Shin I. *Carbohydrate chips for studying high-throughput carbohydrate-protein interactions. 2004, 126 (15), 4812–4819.*
64. Frison N, Marceau P, Roche A-C, Monsigny M, Mayer R. *Oligolysine-based saccharide clusters: synthesis and specificity. Biochemical Journal.* 2002;368(Pt 1):111-119.
65. Shinohara Y. , Sota H., Kim F., Shimizu M., Gotoh M., Tosu M., Hasegawa Y., *Use of a biosensor based on surface plasmon resonance and biotinyl glycans for analysis of sugar binding specificities of lectins, J. Biochem. (Tokyo)* 117 (1995) 1076–1082.
66. McCoy, J.P., Varani, J. and Goldstein, I.J. *Enzyme-linked lectin assay (ELLA): use of alkaline phosphatase-conjugated Griffonia simplicifolia B 4 isolectin for the detection of  $\alpha$ -d-galactopyranosyl end groups. Anal. Biochem.* 1983, 130, 437–444.

67. Taniguchi, N. and Kizuka, Y. (2015) *Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics*. *Adv. Cancer Res.* 126, 11–51.
68. Reuel, N.F., Mu, B., Hinckley, A. and Strano, M.S. *Nanoengineered glycan sensors enabling native glycoprofiling for medicinal applications: Towards profiling glycoproteins without labeling or liberation steps*. *Chem. Soc. Rev.* 2012, 41, 5744–57.
69. De Boer A. R., Hokke C. H., Deelder A. M., Wuhrer M., *General Microarray Technique for Immobilization and Screening of Natural Glycans*. *Analytical Chemistry*, 2007, 79 (21), 8107-8113.
70. Vila-Perelló, M. , Gutiérrez Gallego, R. and Andreu, D. (2005), *A Simple Approach to Well-Defined Sugar-Coated Surfaces for Interaction Studies*. *ChemBioChem*, 6: 1831-1838.
71. Chai W., Feizi T., *Oligosaccharide microarrays to decipher the glyco code*. *Nature Reviews Molecular Cell Biology*, 2004, 5, 582–588.
72. Havliš, J., Thomas, H., Šebela, M., & Shevchenko, A. *Fast-Response Proteomics by Accelerated In-Gel Digestion of Proteins*. *Analytical Chemistry*, 2003, 75(6), 1300–1306. .
73. Hervey, Strader, M. B., & Hurst, G. B. *Comparison of Digestion Protocols for Microgram Quantities of Enriched Protein Samples*. *Journal of Proteome Research*. 2007, 6(8), 3054–3061.
74. ANO, M. P., HERNANDEZ, A. and ANCOS, B. *High Pressure and Temperature Effects on Enzyme Inactivation in Strawberry and Orange Products*. *Journal of Food Science*. 1997, 62: 85-88.
75. López-Ferrer, D., Petritis, K., Hixson, K.K., Heibeck, T.H., Moore, R.J., Belov, M.I., Camp, .G., Smith, R.D., *Application of pressurized solvents for ultrafast trypsin hydrolysis in proteomics: proteomics on the fly*. *J. Proteome Res.* 2008, 7(8), 3276–3281.

- 
76. McCoy J, Hubbell WL. High-pressure EPR reveals conformational equilibria and volumetric properties of spin-labeled proteins *Proc Natl Acad Sci USA*. 2011; 108(4):1331-6.
77. Dufour E., Hervé G., Haertle T. Hydrolysis of  $\beta$ -lactoglobulin by thermolysin and pepsin under high hydrostatic pressure. *Biopolymers*. 1995;35:475–483.
78. Baschung, Y., Lupu, L., Moise, A. et al. Epitope Ligand Binding Sites of Blood Group Oligosaccharides in Lectins Revealed by Pressure-Assisted Proteolytic Excision Affinity Mass Spectrometry *J. Am. Soc. Mass Spectrom.* (2018) 29: 1881. .
79. Balny C. *Biochimica et Biophysica Acta-Proteins and Proteomics* 1764 (2006) 632–639.
80. Hillenkamp, F., M. Karas, R.C. Beavis, and B.T. Chait, Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.*, 1991. 63(24): p. 1193A-1203A.
81. Karas, M. and F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, 1988. 60(20): p.2299-301.
82. Przybylski, M. and M.O. Glocker, Electrospray mass spectrometry of biomacromolecules complexes with non-covalent interactions - new analytical perspectives for supramolecular chemistry. *Angew. Chem. Int. Ed.*, 1996. 35: p. 806-826.
83. Loo, J.A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.*, 1997. 16(1): p. 1-23.
84. Ho C., Lam C., Chan M., et al. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *The Clinical Biochemist Reviews*. 2003;24(1):3-12.
85. Bruins, A. P. Mechanistic aspects of electrospray ionization. *Journal of Chromatography A*, 1998, 794(1-2), 345–357.

- 
86. Fenn, J.B., M. Mann, C.K. Meng, S.F. Wong, Whitehouse C. M., *Electrospray ionization for mass spectrometry of large biomolecules. Science*, 1989. 246(4926): p. 64-71.
87. Knochenmuss, R., *Ion formation mechanisms in UV-MALDI. Analyst*, 2006. 131(9): p. 966-986.
88. Karas, M., M. Gluckmann, and J. Schafer, *Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. J. Mass Spectrom.*, 2000. 35(1): p. 1-12.
89. Zhou, J. and T.D. Lee, *Charge state distribution shifting of protein ions observed in matrix-assisted laser desorption ionization mass spectrometry. J. Am. Soc. Mass Spectrom.*, 1995. 6(12): p. 1183-1189.
90. Wells J.M., McLuckey S.A. *Collision-induced dissociation (CID) of peptides and proteins. 2005. Meth. Enzymol. Methods in Enzymology. 402: 148–85.*
91. Suckau D., Kohl J., Karwath G., Bitter-Suermann D., Przybylski M., *Molecular epitope identification by limited proteolysis of an immobilized antigen-antibody complex and mass spectrometric peptide mapping, Proc. Natl. Acad. Sci. 1990, 87: 9848-9852.*
92. J. McLaurin, R. Cecal, M.E. Kierstead, X. Tian, A. L Phinney, M. Manea, J.E. French, M.H. Lambermon, A.A. Darabie, M.E. Browni, H.T. J. Mount, M. Przybylski, P. St George-Hyslop (2002) *Nature Med.* 8: 1263-1269.
93. R. Stefanescu, R.E. Iacob, E.N. Damoc, A. Marquardt, E. Amstalden, M. Maftai, G. Paraschiv, M. Przybylski (2007) *Eur. J. Mass. Spectrom.* 13: 69-75.
94. Parker C. E., Tomer K. B., *Epitope Mapping by a Combination of Epitope Excision and MALDI-MS, 2000, 146. 185-201.*
95. Kukacka Z. et al, *Antibody Epitope of Human  $\alpha$ -Galactosidase A Revealed by Affinity Mass Spectrometry: A Basis for Reversing Immunoreactivity in Enzyme Replacement Therapy of Fabry Disease. ChemMedChem* 2018, 13, 909.

96. Marquardt A., Muyldermans S., Przybylski M., *A synthetic camel anti-lysozyme peptide antibody (peptibody) with flexible loop structure identified by high-resolution affinity mass spectrometry*, *Chemistry*, 2006, 12: 1915-1923.
97. Moise A., André S., Eggers F., Krzeminski M., Przybylski M., Gabius HJ., *Toward bioinspired galectin mimetics: identification of ligand-contacting peptides by proteolytic-excision mass spectrometry*. *J. Am. Chem. Soc.*, 2011, 133 (38), 14844–14847.
98. Almanza, M., Vega, N. & Pérez, G.. *Isolating and characterising a lectin from Galactia lindenii seeds that recognises blood group H determinants*. *Archives of Biochemistry and Biophysics*, 2004, 429, 2, 180-190.
99. Lavanya Latha, V., Nagender Rao, R. & Nadimpalli, S. K. *Affinity purification, physicochemical and immunological characterization of a galactose-specific lectin from the seeds of Dolichos lablab*. *Prot. Exp. and Pur.*, 2006, 45, 296-306.
100. Pohleven, J., Brzin, J., Leonardi, A., Čokl, A., Štrukelj, B., Kos, J. & Sabotič, J. *Basidiomycete Clitocybe nebularis rich in lectins with insecticidal activities*. *Applied Microbiology and Biotechnology*, 2011, 91, 4, 1141-1148.
101. Kazuo Y., Yukiko K., Kaoru K., Toshiaki O., *Purification and characterization of a carbohydrate-binding peptide from Bauhinia purpurea lectin*, *FEBS Letters*, 1991, 281.
102. Lehrer, R.I., G. Jung, P. Ruchala, S. Andre, H.J. Gabius, and W. Lu, *Multivalent binding of carbohydrates by the human alpha-defensin, HD5*. *J. Immunol.*, 2009. 183(1): p. 480-90.
103. Sinz, A., *Investigation of protein-ligand interactions by mass spectrometry*. *ChemMedChem*, 2007. 2(4): p. 425-31.
104. Jin Lee, Y., *Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes*. *Mol. Biosyst.*, 2008. 4(8): p. 816-23.
105. Coales, S.J., S.J. Tuske, and Y. Hamuro, *Epitope mapping by amide hydrogen/deuterium exchange coupled with immobilization of antibody, on-line*

- proteolysis, liquid chromatography and mass spectrometry. Rapid Commun. Mass Spectrom.*, 2009. 23(639-647).
106. Elyssia S. Gallagher, Jeffrey W. Hudgens, *Mapping Protein–Ligand Interactions with Proteolytic Fragmentation, Hydrogen/Deuterium Exchange-Mass Spectrometry, Methods in Enzymology*, 566 (2016), 357-404.
107. Rappsilber, J., *The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. J. Struct. Biol.*, 2011. 173(3): p. 530-40.
108. Petrotchenko, E.V. and C.H. Borchers, *Crosslinking combined with mass spectrometry for structural proteomics. Mass Spectrom. Rev.*, 2010. 29(6): p.862-76.
109. Dorland L., Van Halbeek H., Vliegenthart J., *Primary Structure of the Carbohydrate Chain of Soybean Agglutinin. J. Biol. Chem.* 1981, 256, 7708-7711.
110. Mandal D. K., Nieves E., Bhattacharyya L., Orr G., ROBOZ J., Yu Q., Brewer C., *Purification and characterization of three isolectins of soybean agglutinin. Eur. J. Biochem.*, 1994, 221, 547-553.
111. Grant, G.; Watt, W. B.; Stewart, J. C.; Bardocz, S.; Pusztai, A. *Intestinal and pancreatic responses to dietary soybean (Glycine-max) proteins. Biochem. Soc. Trans.* 1988, 16, 610-611.
112. Hisayasu, S., H. Orimo, Y. Ikeda, K. Satoh, S. Shinjo, Y. Hirai, and Y. Yoshino. *Soybean protein isolate and soybean lectin inhibit iron-absorption in rats. J. Nutrition*, 1992, 122: 1190-1196. .
113. Jaffe, C. L., S. Ehrlichrogozinski, H. Lis, and N. Sharon. 1977. *Transition-metal requirements of soybean agglutinin. FEBS Letters* 82: 191-196.
114. Liener, I. E.; Hill, E. G. *The effect of heat treatment on the nutritive value and hemagglutinating activity of soybean oil meal. J. Nutr.* 1953, 49, 609-620.
115. Kembhavi, A., Buttle, D., Rauber, P., Barrett, A. *Clostripain: Characterization of the Active Site , F.E.B.S. Lett.* 283, 277, 1991.

116. E., Labrou N. *Handbook of Proteolytic Enzymes (Third Edition)*, s.l. : Academic press, 2013. pp. 2323-2327.
117. Slamnoiu S, Vlad C, Stumbaum M, Moise A, Lindner K, Engel N et al (2014) *Identification and affinity-quantification of  $\beta$ -amyloid and  $\alpha$ -synuclein polypeptides using on-line SAW-biosensor-mass spectrometry. J Am Soc Mass Spectrom 8:1472–1481.*
118. Gutiérrez-Gallego, R.; Haseley, S. R.; van Miegem, V. F.; Vliegthart, J. F.; Kamerling, J. P. *Identification of carbohydrates binding to lectins by using surface plasmon resonance in combination with HPLC profiling. Glycobiology 2004, 14, 373–386.*
119. Katrlík, J., Svitel, J., Gemeiner, P., Kozar, T. and Tkac, J. (2010) *Glycan and lectin microarrays for glycomics and medicinal applications. Med. Res. Rev. 30, 394–418.*
120. Guillaumie, F., Thomas O. R., Jensen K. J., *Immobilization of Pectin Fragments on Solid Supports: Novel Coupling by Thiazolidine Formation Bioconjugate Chem. 2002, 13, 2, 285-294.*
121. Stefanescu, R., Born, R., Moise, A., Ernst, B., Przybylski, M.: *Epitope structure of the CRD of Asialoglycoprotein receptor revealed by high-resolution proteolytic excision mass spectrometry. J. Am. Soc. Mass Spectrom. 2011, 22, 148–157.*
122. Stefanescu, R., Damoc, E.N., Marquardt, A., Perdivara, I., Maftai, M., Paraschiv, G., Przybylski, M., *Mass spectrometric approaches for elucidation of antigen-antibody recognition structures in molecular immunology. Eur. J. Mass. Spectrom. 2007,13, 69-75.*
123. Rao, V.S.R., King, L., Pradman, K.: *Three dimensional structure of the soybean agglutinin-Gal/GalNAc complexes by homology modeling. J. Biomol. Struct. Dyn. 15(5), 853–860 (1998).*
124. Porath J, Låås T, Janson JC. *Agar derivatives for chromatography, electrophoresis and gel-bound enzymes. III Rigid agarose gels cross-linked with divinyl sulphone (dvs). J Chromatogr. 1975; 103:49–62.*

125. Szabó, A., Sahin-Tóth, M. Determinants of chymotrypsin C cleavage specificity in the calcium-binding loop of human cationic trypsinogen. *FEBS J*, 2012, 279: 4283-4292.
126. Schilling, O., Overall, C. M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nature Biotechnology*, 2008, 26(6), 685–694. .
127. Feng C. W., Laskowski M., The effect of calcium on chymotrypsins  $\alpha$  and B. *Biochimica et biophysica acta*, 1956, 19, 110-115.
128. Gilles A. M., Imhoff J. M., Keil B.,  $\alpha$ -Clostripain. Chemical characterization, activity, and thiol content of the highly active form of clostripain. *J. Biol. Chem.* 1979 254: 1462-1468.
129. Mitchell W. M., Harrington W. F., Purification and Properties of Clostridiopeptidase B (Clostripain). *J. Biol. Chem.* 1968 243: 4683-4692.
130. Levi G., Teichberg V. I., Isolation and physicochemical characterization of electrolectin, a beta-D-galactoside binding lectin from the electric organ of *Electrophorus electricus*. *J. Biol. Chem.* 256, 5735-5740.
131. Gronewold, T.M.: Surface acoustic wave sensors in the bioanalytical field: recent trends and challenges. *Anal. Chim. Acta.* 603, 119–128 (2007).
132. Klauke TN, Gronewold TM, Perpeet M, Plattes S, Petersen B: Measurement of porcine haptoglobin in meat juice using surface acoustic wave biosensor technology. *Meat Sci* 2013, 95(3):699-703.
133. Drăgușanu, M., Petre, B.A., Przybylski, M.: Epitope motif of an antinitrotyrosine antibody specific for tyrosine-nitrated peptides revealed by a combination of affinity approaches and mass spectrometry. *J. Pept. Sci.* 17, 184–191 (2011).
134. Koshland, D. E. (1995), *The Key–Lock Theory and the Induced Fit Theory.* *Angew. Chem. Int. Ed. Engl.*, 33: 2375-2378.
135. Breitling J, Aebi M. N-Linked Protein Glycosylation in the Endoplasmic Reticulum. *Cold Spring Harbor Perspectives in Biology.* 2013;5(8):a013359.



- 
136. Belický Š, Katrlík J, Tkáč J. *Glycan and lectin biosensors*. Estrela P, ed. *Essays in Biochemistry*. 2016;60(1):37-47. .
137. Karas, M. and F. Hillenkamp, *Matrix-Assisted Laser Desorption Ionization Mass Spectrometry - Fundamentals and Applications*. *FEMS Microbiol Rev.*, 1994. 24: p. 447-458.
138. Moise, A., S. Andre, F. Eggers, M. Krzeminski, M. Przybylski, and H.J. Gabius, *Toward Bioinspired Galectin Mimetics: Identification of Ligand-Contacting Peptides by Proteolytic-Excision Mass Spectrometry*. *J. Am. Chem. Soc.*, 2011. 133(38): p. 14844-14847.
139. Boyd, W. C., *The Proteins*, Academic Press: New York, 1954 and Vol. 2, pp 756-844.
140. arki A, Cummings R, Esko J, et al., editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1999.

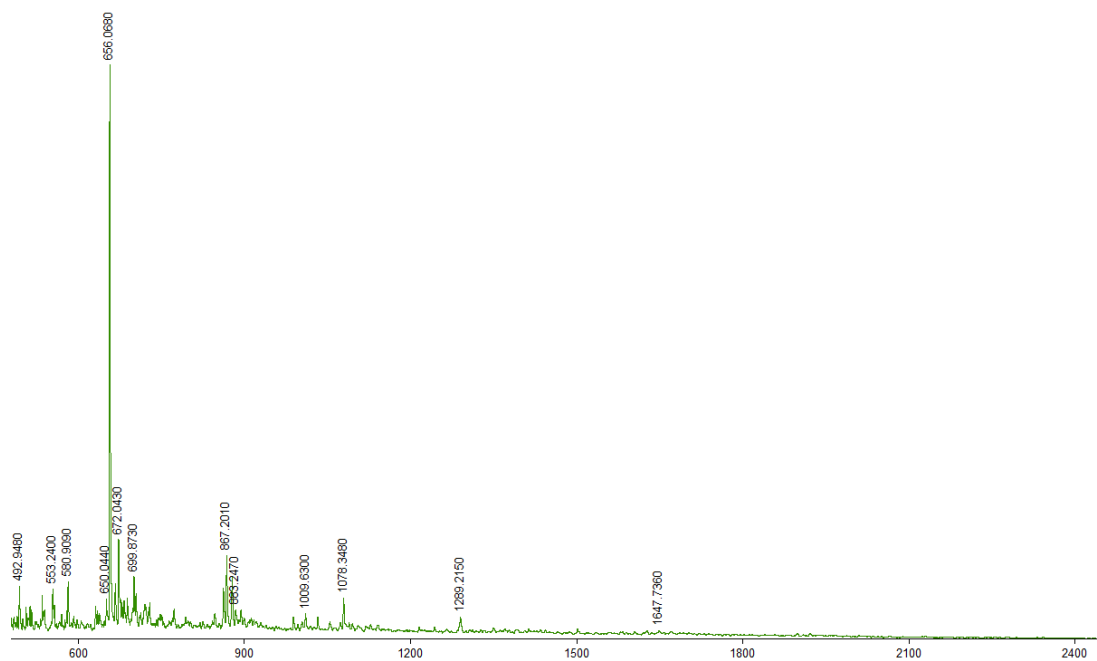
## 6. Appendix

### Appendix 1

**Table A1.** Amino acid table.

<b>Amino Acid</b>	<b>Formula</b>	<b>Monoisotopic Masses (m/z)</b>	<b>Average mass (m/z)</b>	<b>Letter</b>	<b>Code</b>
Alanine	C <sub>3</sub> H <sub>5</sub> NO	71,0371138	71,07801961	A	Ala
Arginine	C <sub>6</sub> H <sub>12</sub> N <sub>4</sub> O	156,101111	156,1859222	R	Arg
Asparagine	C <sub>4</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub>	114,0429275	114,1028044	N	Asn
Aspartic acid	C <sub>4</sub> H <sub>5</sub> NO <sub>3</sub>	115,0269431	115,0875654	D	Asp
Cysteine	C <sub>3</sub> H <sub>5</sub> NOS	103,0091845	103,1428066	C	Cys
Glutamine	C <sub>5</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub>	128,0585775	128,1294218	Q	Gln
Glutamic acid	C <sub>5</sub> H <sub>7</sub> NO <sub>3</sub>	129,0425931	129,1141828	E	Glu
Glycine	C <sub>2</sub> H <sub>3</sub> NO	57,02146373	57,05140221	G	Gly
Histidine	C <sub>6</sub> H <sub>7</sub> N <sub>3</sub> O	137,0589119	137,1395152	H	His
Isoleucine	C <sub>6</sub> H <sub>11</sub> NO	113,084064	113,1578718	I	Ile
Leucine	C <sub>6</sub> H <sub>11</sub> NO	113,084064	113,1578718	L	Leu
Lysine	C <sub>6</sub> H <sub>12</sub> N <sub>2</sub> O	128,094963	128,1725158	K	Lys
Methionine	C <sub>5</sub> H <sub>9</sub> NOS	131,0404847	131,1960414	M	Met
Phenylalanine	C <sub>9</sub> H <sub>9</sub> NO	147,0684139	147,174198	F	Phe
Proline	C <sub>5</sub> H <sub>7</sub> NO	97,05276386	97,11537291	P	Pro
Serine	C <sub>3</sub> H <sub>5</sub> NO <sub>2</sub>	87,03202843	87,07742455	S	Ser
Threonine	C <sub>4</sub> H <sub>7</sub> NO <sub>2</sub>	101,0476785	101,104042	T	Thr
Tryptophan	C <sub>11</sub> H <sub>10</sub> N <sub>2</sub> O	186,079313	186,2103138	W	Trp
Tyrosine	C <sub>9</sub> H <sub>9</sub> NO <sub>2</sub>	163,0633286	163,1736029	Y	Tyr
Valine	C <sub>5</sub> H <sub>9</sub> NO	99,06841392	99,13125442	V	Val

## Appendix 2.



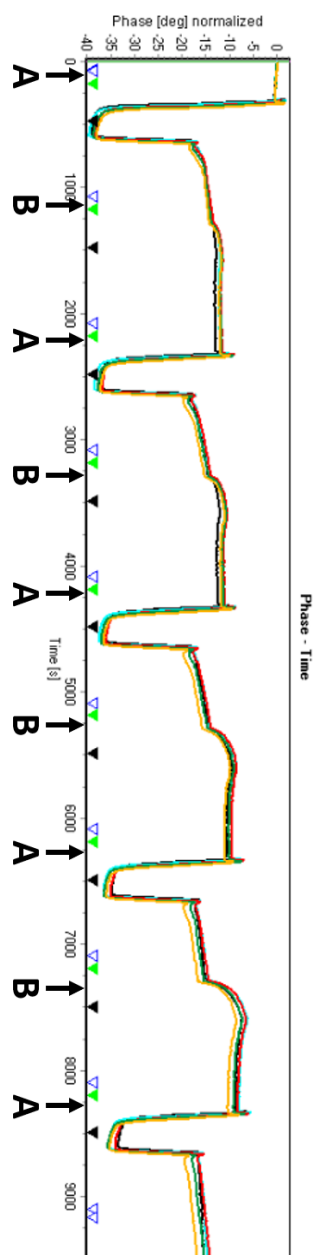
**Figure A1.** MALDI-MS spectra of SBA unsuccessfully digested for 12 hours using procedure 1.

**Appendix 3.**

**Table A2.** Peptide fragments and sequence details of identified tryptic and chymotryptic peptides obtained from the pressure-enhanced proteolytic digestion of SBA.

No.	Peptide fragment	Peptide <u>sequence</u>
1	[80-84]	<u>y.APDTK.r</u>
2	[1-6]	<u>.AETVSF.s</u>
3	[157-163]	<u>l.ANNKVAK.v</u>
4	[75-79]	<u>f.NFTFY.a</u>
5	[80-86]	<u>y.APDTKRL.a</u>
6	[63-70]	<u>k.ETGSVASF.a</u>
7	[86-93]	<u>r.LADGLAFF.l</u>
8	[85-92]	<u>k.RLADGLAF.f</u>
9	[155-163]	<u>w.DLANNKVAK.v</u>
10	[198-205]	<u>k.TSLPEWVR.i</u>
11	[85-93]	<u>k.RLADGLAFF.l</u>
12	[130-138]	<u>r.NSWDPPNPH.i</u>
13	[54-62]	<u>y.STPIHIWDK.e</u>
14	[176-185]	<u>l.VASLVYPSQR.t</u>
15	[151-160]	<u>k.TTSWDLANNK.v</u>
16	[11-20]	<u>k.FVPKQPNMIL.g</u>
17	[186-197]	<u>r.TSNILSDVVDLK.t</u>
18	[18-31]	<u>n.MILQGDAIVTSSGK.l</u>
19	[37-50]	<u>k.VDENGTPKPSSLGR.a</u>
20	[94-107]	<u>f.LAPIDTKPQTHAGY.l</u>
21	[155-168]	<u>w.DLANNKVAKVLITY.d</u>
22	[209-224]	<u>f.SAATGLDIPGESHDVL.s</u>
23	[49-62]	<u>l.GRALYSTPIHIWDK.e</u>
24	[93-107]	<u>f.FLAPIDTKPQTHAGY.l</u>
25	[130-144]	<u>r.NSWDPPNPHIGINVN.s</u>
26	[35-50]	<u>l.NKVDENGTPKPSSLGR.a</u>
27	[129-144]	<u>f.RNSWDPPNPHIGINVN.s</u>
28	[209-226]	<u>f.SAATGLDIPGESHDVLSW.s</u>
29	[206-224]	<u>r.IGFSAATGLDIPGESHDVL.s</u>
30	[108-125]	<u>y.LGLFNENESGDQVVAVEF.d</u>
31	[229-247]	<u>f.ASNLPHASSNIDPLDLTSE.v</u>
32	[130-147]	<u>r.NSWDPPNPHIGINVNSIR.s</u>
33	[206-226]	<u>r.IGFSAATGLDIPGESHDVLSW.s</u>
34	[87-107]	<u>l.ADGLAFFLAPIDTKPQTHAGY.l</u>

## Appendix 4.



**Figure A2.** Injection of regeneration solutions (A) and SBA solutions (B) over an DVS-activated GalNAc functionalized SAW gold chip.

**Appendix 5.****ABBREVIATIONS**

ACN	Acetonitrile
CID	Collision-induced dissociation
Da	Dalton
DMF	Dimethylformamide
DTT	Dithiothreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
ESI	Electrospray ionization
Fmoc	9-Fluorenylmethoxycarbonyl
HCCA	4-Hydroxy- $\alpha$ -cynamic acid
HDX	hydrogen/deuterium amide exchange
HPLC	High performance liquid chromatography
MALDI-MS	Matrix-assisted Laser desorptions/Ionizations-Mass spectrometry
Min	Minutes
MS	Mass spectrometry
MW	Molecular weight
m/z	Mass over charge ration
NHS	N-hydroxysuccinimide
NMM	N-methyl-morpholine
PBS	Phosphate buffered saline
pH	Negative logarithm of H <sub>3</sub> O <sup>+</sup> ions concentration

---

PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
RP	Reversed-phase
SAM	Self assembled monolayer
SDS-Page	Sodiumdodecylsulfate-Polyacrylamide-Gel electrophoresis
SPR	Surface plasmon resonance
TEMED	N,N,N',N'-tetraethylendiamine
TFA	Trifluoroacetic acid
TIC	Total ion chromatogram
ToF	Time of flight
Tris	Tris-(hydroxymethyl-)aminomethane
UV	Ultraviolet
°C	Grad Celsius

---

**Appendix 6.****Declaration of independence**

I declare that the thesis “Development and application of affinity-mass spectrometry to identify protein-carbohydrate interactions” was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.