Synthesis of Complex Glycopeptides Toward Novel Cancer Marker Discovery

(複雑な糖ペプチドの合成による新しい癌マーカーの探索)

A Thesis

Submitted for the degree of

Doctor of Life Science

By

YOGESH K V

Transdisciplinary Life Science Course

Graduate School of Life Science

Hokkaido University

September 2018

DECLARATION

I hereby declare that the matter embodied in this thesis entitled is of investigations carried out by me under the supervision of Prof. Shin-Ichiro Nishimura at the Laboratory of Advanced Chemical Biology, Transdisciplinary Life Science Course, Graduate School of Life Science, Hokkaido University, Japan and it has not been submitted elsewhere for the award of any other degree or diploma.

In keeping with general practice of reporting scientific observations, due acknowledgement has been made whenever the work described has been based on the findings of the other investigators. Any omission that might have occurred by oversight of error of judgement is regretted.

YOGESH K V

CERTIFICATE

I hereby declare that the matter embodied in this thesis entitled has been carried out by YOGESH K V, under my supervision at the Laboratory of Advanced Chemical Biology, Transdisciplinary Life Science Course, Graduate School of Life Science, Hokkaido University, Japan.

Prof. Shin-Ichiro Nishimura

(Research Supervisor)

Contents

Abbreviation

AcOH	Acetic acid
Ac_2O	Acetic anhydride
ACN	Acetonitrile
MeCN	Acetonitrile
Ac	Acetyl
AGP	Alpha-1-acid glycoprotein
AGP1	Alpha-1-acid glycoprotein 1
AGP2	Alpha-1-acid glycoprotein 2
Asn	Asparagine
BDA	Benzaldehyde dimethyl acetal
Bn	Benzyl
BnOH	Benzyl alcohol
BnBr	Benzyl Bromide
$BF_3{\cdot}Et_2O$	Borantriflouride diethyl ether
CSA	(±)-10-Camphor sulfonic acid
COOH	Carboxylic
CAN	Ceric ammonium nitrate
$CHCl_3$	Chloroform
DMC	2-chloro-1,3-dimethylimidazolinium chloride
CXP	Collision Cell Exit Potential
CE	Collision Energy
CID	Collision Induced Dissociation
CDP	Cytidine diphosphate
DP	Declustering Potential
DNA	Deoxyribose Nucleic Acid
$CDCl_3$	Deuterated chloroform
CH_2Cl_2	Dichloromethane
DCM	Dichloromethane
DIBAL	Diisobutylaluminium hydride
ESI-MS	Electrospray Ionization-Mass spectrometry

ER lumen	Endoplasmic Reticulum lumen
EMS	Enhanced Mass
EPI	Enhanced Product Ion
ER	Enhanced Resolution
EP	Entrance Potential
Et	Ethyl
EtOAc	Ethyl acetate
US FDA	United States Food and Drug Administration
Gtfs	Glycosyltransferases
GDP	Guanosine diphosphate
HSQC	Heteronuclear Single Quantum Coherence
HPLC	High Performance Liquid Chromatography
hrs	Hours
HSA	Human Serum Albumin
HOBt	1-Hydroxybenzotriazole
Tris	Hydroxymethyl aminomethane
IgG	Immunoglobulin G
LC	Liquid Chromatography
LC-SRM	Liquid Chromatography-Selected Reaction Monitoring
LiBr	Lithium bromide
$MgSO_4$	Magnesium sulfate
$MnCl_2$	Manganese chloride
Man	Mannose
MS	Mass Spectroscopy
MALDI	Matrix Assisted Laser Desorption Ionization
MALDI-TOF MS	Matrix Assisted Laser Desorption Ionization - Time of Flight Mass Spectroscopy
MeOH	Methanol
MUC1	Mucin 1
GlcNAc	*N*-Acetylglucosamine
Neu5Ac	*N*-Acetylneuraminic acid
DIEA	*N*, *N*-Diisopropylethylamine
DMF	*N*, *N*-Dimethylformamide

NMR	Nuclear Magnetic Resonance
OSTase	Oligosaccharyltransferase
$Pd(OH)_2/C$	Palladium hydroxide on carbon (Pearlman's catalyst)
KOH	Potassium hydroxide
RT	Retention Time
RP-HPLC	Reverse Phase-High Performance Liquid Chromatography
RNA	Ribonucleic acid
SRM/MRM	Selected Reaction Monitoring/ Multiple Reaction Monitoring
Ser/Thr	Serine/Threonine
Ag_2CO_3	Silver carbonate
$NaHCO_3$	Sodium bicarbonate
Na_2CO_3	Sodium carbonate
NaH	Sodium hydride
NaOMe	Sodium methoxide
Na_2SO_4	Sodium sulfate
$Na_2S_2O_3$	Sodium thiosulphate
TMS	Tetramethylsilane
TLC	Thin Layer Chromatography
SPh	Thiophenyl
TOF	Time of Flight
Troc	2,2,2-Trichloroethoxycarbonyl
TEA	Triethylamine
Et_3N	Triethylamine
TFA	Trifluoroacetic acid
TMSOTf	Trimethylsilyltrifluoromethanesulfonate
Trp	Tryptophan
UDP	Uridine diphosphate
H_2O	Water

CHAPTER 1

General Introduction:

1.1. Proteins

Proteins are large macromolecules like polysaccharides and nucleic acids in bio system perform a huge array of functions within organisms such as catalyzation, DNA replication, response to stimuli, and transportation of molecules from one location to another. Proteins consisting of one or more long chains of amino acid residues, differ from one another mainly in their sequence of amino acids usually results in protein folding into a specific three-dimensional structure that determines its activity.

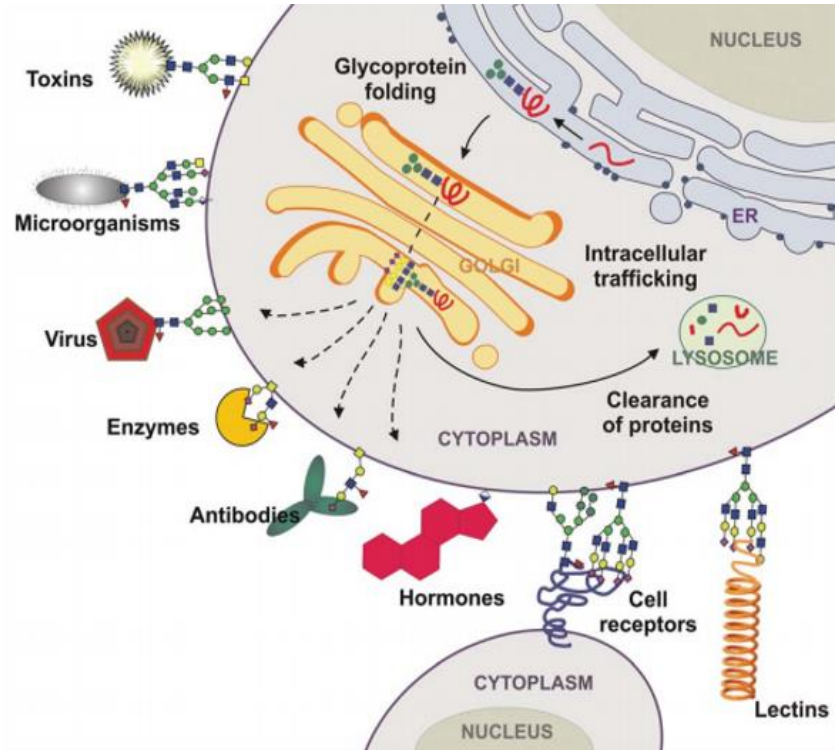


Fig 1.1.1. Role of carbohydrate chains in biological functions.

A protein contains at least one long polypeptide, a linear chain of amino acid residues. Peptides or oligopeptides are the most commonly used terms for the sequences containing less than 20-30 amino acid residues and are rarely considered to be proteins. Each amino acid in a polypeptide chain is referred as a residue linked to another via peptide bond and the linked series of carbon, nitrogen, and oxygen atoms are known as the main chain or protein backbone. All amino acid residues that are incorporated biosynthetically into proteins during translation process possess common structural features, including an α-carbon to which an amino group, a carboxyl group, and a variable side chain are bonded except proline, which only differs from this basic structure as it contains an unusual

ring to the N-end amine group, which forces the CO–NH amide moiety into a fixed conformation. The synthesis of the sequence of the protein naturally occurs from N-terminus to C-terminus and therefore written and represent usually as such. These N-terminus and C-terminus are free amino group of one end and free carboxyl group of another end of peptide backbone respectively.

Glycosylation is a form of co-translational and post-translational modification catalyzed by various glycosyltransferase enzymes which are mostly located in the Golgi apparatus in cells refers to the reaction in which a glycosyl donor is attached to a hydroxyl or other functional group of another molecule usually a glycosyl acceptor. For instance, attachment of glycans to proteins is one which results in the alteration of physical and chemical properties, folding, stability, activity, and ultimately, the function of the proteins. In biology, these glycosyl donors are usually glycan molecules, refers to different forms of sugars such as monosaccharides, oligosaccharides and polysaccharides, which are the common carbohydrate part of glycoconjugates such as glycoprotein, glycolipid or a proteoglycan. Glycoproteins are the proteins that contains these glycan chains covalently attached to polypeptide side-chains. In glycan classification, two major glycans are *N*-glycans and *O*-glycans, which are well known structural components of carbohydrates of these glycoproteins. For instance, AGP (alpha-1-acid glycoprotein) is reported for the presence of *N*-glycans in its structure whereas, MUC1 for *O*-glycans respectively.

There are two major types of glycosylation; *N*-linked glycosylation and *O*-linked glycosylation. However, *N*-glycosylation is the major one covers almost 90% glycoproteins presents three major types of *N*-glycan molecules such as high mannose, hybrid and complex type which are covalently attached to a nitrogen of asparagine side-chains of peptide backbone of proteins, that is present as a part of Asn–X–Ser/Thr consensus sequence (where X is any amino acid except proline).[1]

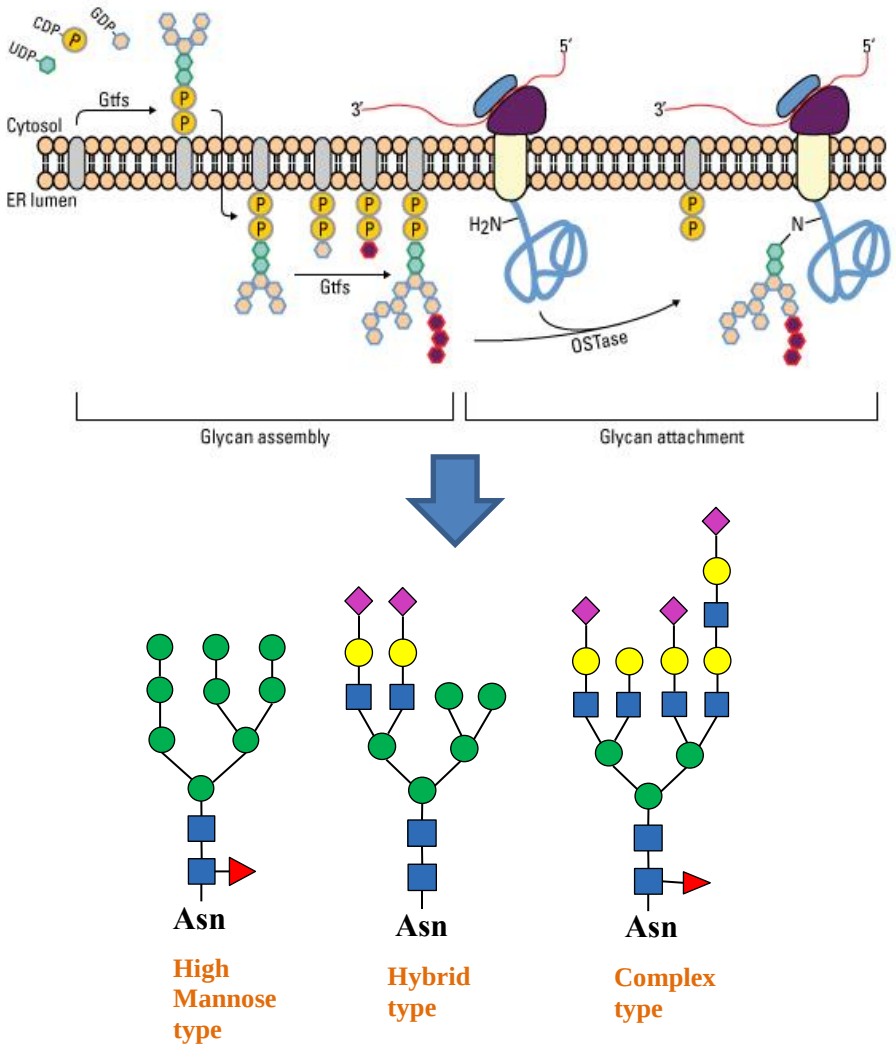


Fig 1.1.2. Types of *N*-glycans.

Glycans are essential biologically active sugars found enormously in nature and the term glycomics was coined in a field of research to investigate its crucial role of these glycans in biology. These molecules are found on the surface of many cells are important in cell-cell interaction in disease-related processes such as inflammation, infection and cancer. Hence these molecules are being investigated since three decades as novel targets for drug and vaccine development.

Generation of Cancer cells takes place when the genes responsible for regulating cell division are damaged and the process is called carcinogenesis.[2] These cancer cells are transported through the circulating system to distant sites induces dynamic structural changes in the *N*-glycans of human serum glycoproteins. Among these *N*-glycans, hyper-branched glycans such as tri- and tetra-antennary *N*-glycans are increased distinctly in the sera of various cancer patients such as hepatocellular carcinoma,[3a] renal pancreatic cancer,[3b] cellular carcinoma,[3c] prostate cancer,[3d] etc. and even in other diseased patients shows their potential for serum biomarker in many types of cancer and other diseases.

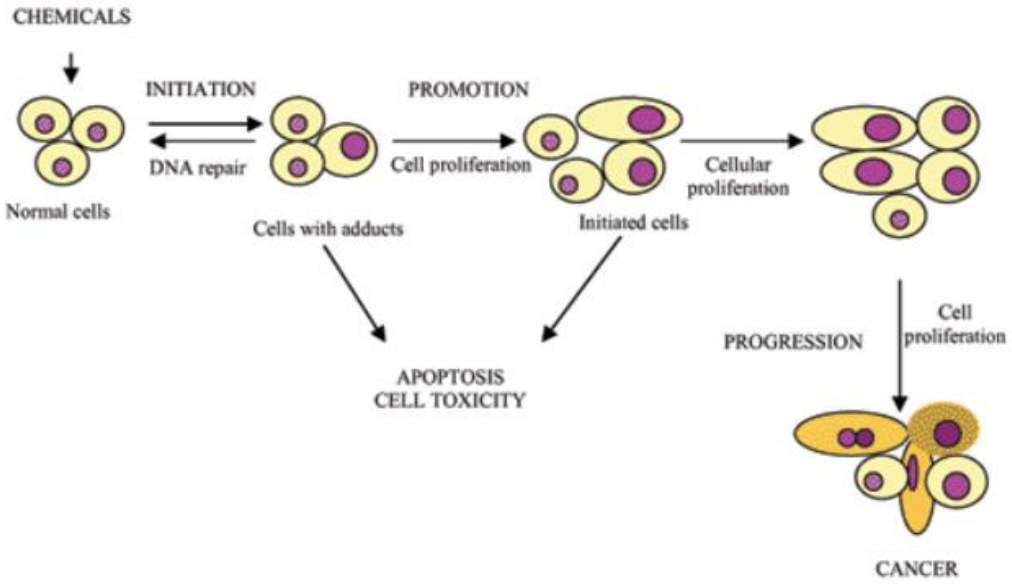


Fig. 1.1.3. Mechanism of progression in cancer.

It was also demonstrated that alteration of serum levels of these multi-antennary *N*-glycoforms often correlates with efficacy and adverse side effects to current treatment options,[4] suggesting that serum *N*-glycans can be a new class of companion biomarkers in the personalized medicine. These observations imply that serum levels of the multi-antennary *N*-glycoforms themselves might become sensitive biomarkers indicating changes in the metabolic and homeostatic balance of entire human bodies caused by cancer cells proliferation and malignant alterations. However, it should also be noted that such *N*-glycan biomarkers are not suited for the discrimination between different cancer patients and even patients suffering from other diseases exhibiting a quite similar serum *N*-glycan profile such as multi-antennary *N*-glycoforms.

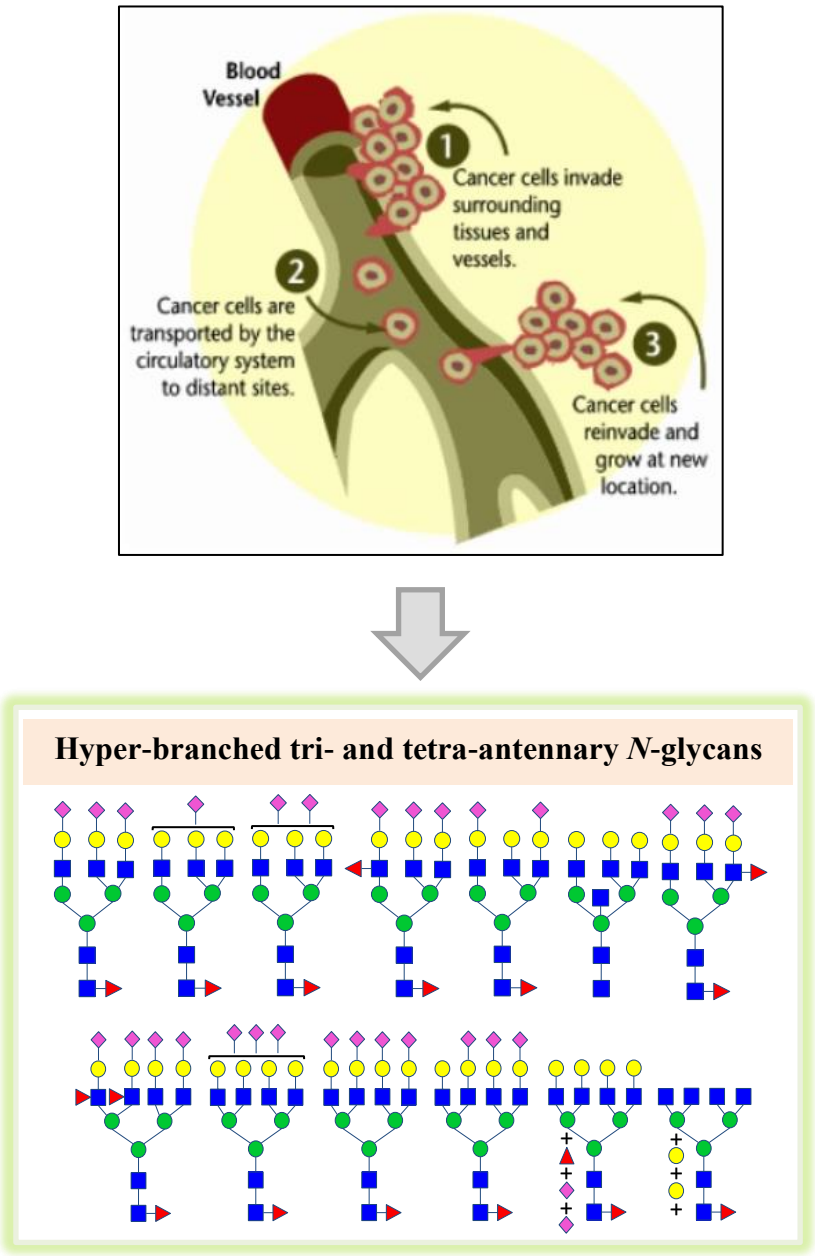


Fig 1.1.4. Complex *N*-glycan profile in various types of diseases including cancer.

1.2. SRM/MRM based targeted glycoproteomics

Recently, MS based qualitative and quantitative methods are familiar and more appropriate trend in a field of proteomics. The first mass spectrometer was constructed in 1912 and since then new designs

in the format of analytical tools of this technique has been established spontaneously to develop the technology with robustness, good reproducibility, high sensitivity, high throughput to achieve an analysis with dynamic range of proteins of biological fluids, for example, plasma, serum etc.

Liquid chromatography interfaced to mass spectrometry (LC-MS and LC-MS/MS) is the most routinely applied method for the identification and quantitation of protein glycosylation. In this approach, glycoproteins are first enzymatically digested (i.e., trypsin), separated using LC, and then subjected to MS and tandem MS (MS/MS) analysis. Spontaneous efforts made in the literature on glycosylation studies using LC-MS and LC-MS/MS have demonstrated the applicability of these methods to decipher aberrations in glycosylation associated with diseased samples.[5] The main challenge of MS-based proteomic analysis in plasma is the analysis of extremely low concentration of proteins of interest with a very high level of biochemical background. Direct liquid chromatography–MS (LC-MS) analysis includes tryptic digestion of peptides derived from serum usually results in the detection of predominant peptides from most abundant proteins such as human serum albumin (HSA) and immunoglobulin G (IgG). Mass spectrometry using electrospray ionization (ESI-MS) and MALDI-TOF MS is applied for LC−MS based glycomic analysis but unfortunately, these techniques are not fully exploited for large-scale biomarker discovery studies because of lack of appropriate computational tools. As an alternative, techniques like lectin affinity chromatography and peptide labeling are used in conjunction with LC-MS to enrich samples of interest. In addition, among several formats of MS technology for proteomics established so far, “shotgun proteomics” is one,[6] where proteins are usually digested to peptides preferably by a sequence specific enzyme such as trypsin, which are then separated, analyzed and thus thousands of proteins can be quantified without any prior knowledge offers only hypothesis free information since it does not focus on specific protein or peptide of interest during analysis.

Despite LC-MS and LC-MS/MS are widely used to enable effective characterization of posttranslational modification, weakened intensity due to microheterogeneity of peptides and lesser ionization efficiency of glycopeptides than those of peptides makes these techniques to suffer from several limitations. Targeted proteomics technique has emerged as a powerful protein quantitation tool recently as a complement of shotgun proteomics.[7] It is capable of test specific proteins and

peptides of which their information is known prior to analysis. These techniques are currently implemented with a variety of mass analyzers, including combinations of quadrupoles, orbital ion traps (orbitrap), and time-of-flight mass analyzers.

The most widely used targeted proteomics technique based on SRM/MRM,[12] selected reaction monitoring/ multiple reaction monitoring is an alternative approach for untargeted shotgun proteomics because of its high selectivity and high sensitivity with sample preparation, such as depletion of human blood sera and subproteome enrichment to reduce the biochemical background, which results usually from most abundant proteins such as albumin, Immunoglobulin (IgG), etc.

LC-SRM analysis integrated with stable isotope labeling methods are routinely used not only in small-molecule analyses such as pharmacokinetics studies[8] but also in protein quantification, through which it is possible to establish biomarker candidates in clinical samples[9]. SRM method for small molecules focuses on direct measurement of compounds. In contrast, the method for proteins involves indirect measurement using proteotypic peptides derived from target proteins as its substitutes and therefore synthetic isotopically labeled homologs as reference peptides are added into the sample (e.g., clinical samples) in well-defined concentrations prior to perform quantification. These synthetic isotopically labeled homologs are the peptides labeled with a stable isotope, e.g. ^{13}C, or ^{15}N which are chemically identical to their native counterparts provides identical chromatographic behavior with only a small specific mass difference.[10] The location for heavy isotope-labeling is more often on leucine, proline, valine, phenylalanine or tyrosine and C-terminal arginine or lysine rather than any in tryptic peptides.[11] The amount of corresponding peptides, which are originating from within an organism can then be determined by comparing their signal intensities with the signal intensities of their isotopically labeled counterparts. Advanced technical facility including increasing number of peptides in one LC-MS experiment with scheduling each transition (a pair of precursor ions and one of its fragment ions) with projected retention time for each peptide and analytes punctuality associated with acquire additional transitions has increased the confidence and robustness of this technique recently.

SRM/MRM (Selected Reaction Monitoring/Multiple Reaction Monitoring) is a special type of powerful mass spectrometer complement to untargeted shotgun methods as targeted mass

spectrometry enabling an absolute quantitation of the predetermined target proteins across multiple and heterogeneous samples in a highly precise and reproducible manner.[12] It is a triple quadrupole instrument consists of three quadrupoles; the first quadrupole (Q1) which can select target precursor ion predefined by the user allows to undergo fragmentation with collision induced dissociation (CID) at second quadrupole (Q2), usually referred as a collision cell gives different fragments of precursor ion and the fragments of interest is then selected at third quadrupole (Q3). Coeluting background ions are rejected by two quadrupoles (Q1 and Q3), acts as mass filters to specifically select predefined m/z values of precursor ion as well as fragments of interest needed for quantitation which provides higher selectivity during MRM analysis. Resulting fragments flows towards third quadrupole (Q3) from the collision cell (Q2) are commonly referred as MRM transitions and are specific, represents and characterizes its corresponding precursor ion. The selection of fragment ion and precursor ion is essential for MRM based quantitation. Although an optimization of several gas and voltage parameters is essential, number of transitions, MRM time segmentation, and normalized collision energy are the important MRM parameters needed to optimize for effective quantitation. MRM analysis relies on the ionization efficiency as well as fragmentation efficiency, however the ionization efficiencies of different glycoforms with the same peptide moiety are very similar in ESI and therefore the major concern for efficient quantitation is the difference in fragmentation efficiencies and this is strongly depending on optimization of collision energy. MRM based quantitation can be performed with and without using time segments, which are termed as segmented (SMRM) and non-segmented MRM (NSMRM) respectively. In NSMRM mode, the triple quadrupole mass spectrometer is programmed to scan all the selected ions with specific transitions during any given time and this mode is advantageous when the specific elution time of a particular peak is not known. In the SMRM mode, the mass spectrometer is programmed to scan for only those ions which elute during the predetermined time segment specified for that particular ion results in increasing measurement sensitivity because of decrease of cycle time of the mass spectrometer for those ions.

1.3. Objective of the thesis

Our previous studies on the reverse glycoblotting method[13] focusing on the terminal sialic acid residues demonstrated that 25 multiple reaction monitoring (MRM) channels made by the tryptic sialo glycopeptides enriched from mouse serum enable quantitative glycoproteomics targeting glycoproteins having the sialylated *N*-glycoforms of interest in diabetic model mice in comparison with normal mice.[14] Importantly, it is impossible to determine the precise "glycan structures" of the tryptic peptides having a sialylated *N*-glycoform used for each MRM channel setting due to the structural diversity elaborated by GlcNAc-branching, subsequent galactosylation and positional sialylation during their biosynthesis. It is the case not only in our report but also in all reports of SRM/MRM based quantitation of proteins or peptides unless used structure defined synthetic compounds as an internal standard during overall analysis. Therefore, it is obvious that availability of the structure-defined synthetic glycopeptides is essential for successful SRM/MRM-based quantitative glycoproteomics.

Interestingly, protein glycosylation can occur on multiple glycosylation sites of peptide backbone of proteins involving the attachment of different glycans to each and same site results in the structural variability of glycans from site to site and structural microheterogeneity at the same site during biosynthesis, which makes the process of purification of glycopeptide of target protein candidate in a complex biological sample such as plasma, serum etc. is impossible. This issue approach for alternative method to know the concentration of each glycopeptide of target protein to evaluate further study including post-translational modification behaviour, biomarker discovery. Combining robust databases of human serum glycomics and proteomics focusing on major glycoproteins is one of the most efficient approaches to discover potent candidates of the target tryptic peptides displaying a focused glycoform. The present study communicates for the first-time potentials of the tryptic glycopeptides derived from serum α1-acid glycoprotein (AGP) as an example of new class biomarkers when structurally-defined synthetic glycopeptides are employed as designated standards in SRM assay for the targeted quantitative glycoproteomics.

1.4. References

1.4. References

1. Drickamer, K.; Taylor, M. E. Introduction to Glycobiology, **2006**.
2. (a) Laura, D. W.; Parsons, D. W.; Jones, S.; Lin J.; Sjoblom, T.; Leary, R. J.; Shen, D.; Boca, S. M.; Barber, T.; Ptak, J.; Silliman, N.; Szabo, S.; Dezso, Z.; Ustyanksky, V.; Nikolskaya, T.; Nikolsky, Y.; Karchin, R.; Wilson, P. A.; Kaminker, J. S.; Zhang, Z.; Croshaw, R.; Willis, J.; Dawson, D.; Shipitsin, M.; Willson, J. K. V.; Sukumar, S.; Polyak, K.; Park, B. H.; Pethiyagoda, C. L.; Pant. P. V. K.; Ballinger, D. G.; Sparks, A. B.; Hartigan, J.; Smith, D. R.; Suh, E.; Papadopoulos, N.; Buckhaults, P.; Markowitz, S. D.; Parmigiani, G.; Kinzler, K. W.; Velculescu, V. E.; Vogelstein, B. The genomic landscapes of human breast and colorectal cancers. *Science* **2007,** 318**,** 1108–1113. (b) Knudson, A. G. Two genetic hits to cancer. *Nature Reviews Cancer* **2001,** 1, 157-162. (c) Fearon, E. R.; Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **1990,** 61, 759–767.
3. (a) Kamiyama, T.; Yokoo, H.; Furukawa, J. I.; Kurogochi, M.; Togashi, T.; Miura, N.; Nakanishi, K.; Kamachi, H.; Kakisaka, T.; Tsuruga, Y.; Taketomi, A.; Nishimura, S -I.; Todo, S. Identification of novel serum biomarkers of hepatocellular carcinoma using glycomic analysis. *Hepatology* **2013**, *57*, 2314–2325. (b) Nouso, K.; Amano, M.; Miyahara, K.; Ito, M. I.; Morimoto, Y.; Kato, H.; Tsutsumi, K.; Tomoda, T.; Yamamoto, N.; Nakamura, S.; Kobayashi, S.; Kuwaki, K.; Hagihara, H.; Onishi, H.; Miyake, Y.; Ikeda, F.; Shiraha, H.; Takaki, A.; Nakahara, T.; Nishimura, S.-I.; Yamamoto, K. Clinical utility of high-throughput glycome analysis in patients with pancreatic cancer. *J. Gastroenterol.* **2013**, *48*, 1171-1179. (c) Hatakeyama, S.; Amano, M.; Tobisawa, Y.; Yoneyama, T.; Tsuchiya, N.; Habuchi, T.; Nishimura, S-I.; Ohyama, C. Serum *N*-glycan alteration associated with renal cell carcinoma detected by high throughput glycan analysis. *J. Urology* **2014**, *191*, 805–813. (d) Ishibashi, Y.; Tobisawa, Y.; Hatakeyama, S.; Ohashi, T.; Tanaka, M.; Narita, S.; Koike, T.; Habuchi, T.; Nishimura, S. -I.; Ohyama, 30 C.; Yoneyama, T. Serum tri- and tetra-antennary *N*-glycan is a potential predictive biomarker for castration-resistant prostate cancer. *Prostate* **2014**, *74*, 1521-1529.

4. Miyahara, K.; Nouso, K.; Miyake, Y.; Nakamura, S.; Obi, S.; Amano, M.; Hirose, K.; Nishimura, S. -I.; Yamamoto, K. Serum glycan as a prognostic marker in patients with advanced hepatocellular carcinoma treated with sorafenib. *Hepatology* **2014**, *59*, 355-356.
5. (a) Balog, C. I. A.; Mayboroda, O. A.; Wuhrer, M.; Hokke, C. H.; Deelder, A. M.; Hensbergen, P. J. Mass Spectrometric Identification of Aberrantly Glycosylated Human Apolipoprotein C-III Peptides in Urine from Schistosoma mansoni-infected Individuals. *Mol. Cell. Proteomics* **2010,** 9, 667. (b) Whelan, S. A.; Lu, M.; He, J.; Yan, W.; Saxton, R. E.; Faull, K. F.; Whitelegge, J. P.; Chang, H. R. Mass Spectrometry (LC-MS/MS) site mapping of N-glycosylated membrane proteins for breast cancer biomarkers. *J. Proteome Res.* **2009,** 8, 4151.
6. John, R. Y. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. *J. Am. Chem. Soc.* **2013,** 135, 1629−1640.
7. (a) Eva, B.; Eduard, S. What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics* **2017,** 17, 17–18. (b) Tujin, S.; Ehwang, S.; Song, N.; Karin, D. R.; Tao, L.; Qian, W.-J.; Richard, D. S. Advances in targeted proteomics and applications to biomedical research. *Proteomics* **2016,** 16, 2160-2182.
8. Hoke, S. H.; Morand, K.; Greis, K. D.; Baker, T. R.; Harbol, K.; Dobson R. L. M. Transformations in pharmaceutical research and development, driven by innovations in multidimensional mass spectrometry-based technologies. *Int. J. Mass Spectrom.* **2001,** 212, 135–196.
9. (a) Gallien, S.; Duriez, E.; Domon, B. Selected reaction monitoring applied to proteomics. *J. Mass Spectrom.* **2011,** 46, 298–312. (b) Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **2007,** 6, 2212–2229. (c) Pan, S.; Aebersold, R.; Chen, R.; Rush J.; Goodlett D. R.; McIntosh M. W.; Zhang J.; Brentnall, T. A. Mass spectrometry based targeted protein quantification: methods and applications. *J. Proteome Res.* **2009,** 8, 787–797. (d) Desiderio, D. M.; Kai, M. Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed. Mass Spectrom.* **1983,** 10, 471–479. (e) Anderson, L.; Hunter,

C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **2006,** 5, 573–588.

10. Tambor, V.; Fucikova, A.; Lenco, J.; Kacerovsky, M.; Rehacek, V. Application of proteomics in biomarker discovery: a primer for the clinician. *Physiol. Res.* **2010,** 59, 471–497.
11. Han, B.; Higgs, R. E. Proteomics: from hypothesis to quantitative assay on a single platform. Guidelines for developing MRM assays using ion trap mass spectrometers. *Brief Funct. Genomic Proteomic* **2008**, 7, 340–54.
12. (a) Picotti, P.; Aebersold, R. Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **2012**, *9*, 555-566. (b) Harlan, R.; Zhang, H. Targeted proteomics: a bridge between discovery and validation. *Expert Rev. Proteomics* **2014**, *11*, 657-661.
13. Kurogochi, M.; Amano, M.; Fumoto, M.; Takimoto, A.; Kondo, H.; Nishimura, S. -I. Reverse glycoblotting allows rapid enrichment glycoproteomics of biopharmaceuticals and disease-related biomarkers. *Angew. Chem. Int. Ed.* **2007**, *46*, 8808-8813.
14. Kurogochi, M.; Matsushita, T.; Amano, M.; Furukawa, J.; Shinohara, Y.; Aoshima, M.; Nishimura, S. -I. Sialic acid-focused quantitative mouse serum glycoproteomics by multiple reaction monitoring assay. *Mol. Cell. Proteomics* **2010**, *9*, 2354-2368.

CHAPTER 2

Synthesis of AGP (alpha-1-acid glycoprotein) glycopeptides

2.1. Introduction

AGP Structure:

AGP (alpha-1-acid glycoprotein), also known as Orosomucoid is named because of its migration with the α-1 protein group during plasma protein electrophoresis.[1] The peptide backbone structure of AGP is comprised of 183 amino acid residues as a single polypeptide precursor, covers roughly 59% of AGP's structure.[2] It is a highly acidic glycoprotein with negative charge and low pI (2.8-3.8) as its carbohydrate content has up to 16 sialic acid residues which covers 11-12% of its total molecular mass. The low pI of AGP is due to the presence of COOH group at C1 of sialic acid residues and many acidic amino acids in its actual peptide sequence. An approximate molecular mass of each sialic acid residue is 314 Da, however AGP becomes inactive with higher pI of 4.97 if this mass is lost through desialylation. It is reported that the 183 single polypeptide precursors of AGP contain two disulfide bridges which are formed between cysteine residues 5-147, 72-164[3] and found to be buried completely or partially with 5-7 tyrosine, 2 tryptophan and most of phenylalanine residues in its native state. Many amino acids such as tryptophan (Trp), certain tyrosine (Tyr) and most of phenylalanine (Phe) residues have preferences of location within AGP's folded structure due to molecular interactions within the polypeptide backbone.

Human AGP was firstly thought to be comprised of 181 amino acids with a single polypeptide chain.[4] Later, it is proved[5] that the addition of Lys and Arg residues at positions 173 and 174 gives 183 amino acids in length as an actual peptide sequence of AGP. It consists of two genetic variants namely AGP1 and AGP2, which are differ by almost 22 amino acid residues substitutions within the molecules with the gene products ratio of 3:1 in plasma under normal physiological condition. However, it is expressed from a cluster of three adjacent genes on chromosome 9, namely AGP-A (ORM-1) which encodes variants F1, F2, S and AGP-B, AGP-B' (ORM-2) code for A.[6] Two alleles of AGP-A encode F1, F2 and S variants and these variants are known as F1*S with a difference of less than five amino acids each. However, variant A has approximately 20 amino acids substituted and this is the only minor difference with all variants containing 183 amino acids in total.

Like its family lipocalin, which shows many β-sheets and capable of showing significant immunomodulatory effects, the molecular structure of AGP is also found to be observed with α-helix (15%), β-sheet (41%), reversed β-turns (12%), bands (8%) and unordered structures (24%) and it folds as a highly symmetrical β-sheet structure containing a single antiparallel β-sheet with eight strands.[7]

There are five glycosylation sites in a single polypeptide chain of AGP gives a platform to share over remaining 40% of its molecular mass by carbohydrate content and the corresponding glycosylation sites are denoted as Asn-15, -38, -54, -75 and -85,[8] which carry highly sialylated heteropolysaccharide complex-type *N*-linked glycans attached to Asn residues usually in the first half of the polypeptide. However, it is difficult to conclude a precise glycan structure of each glycosylation sites mainly because of glycan micro-heterogeneity, which is associated with glycan chains rather than peptide backbone due to great diversity of the well-known terminating sugars namely sialic acid and fucose residues.

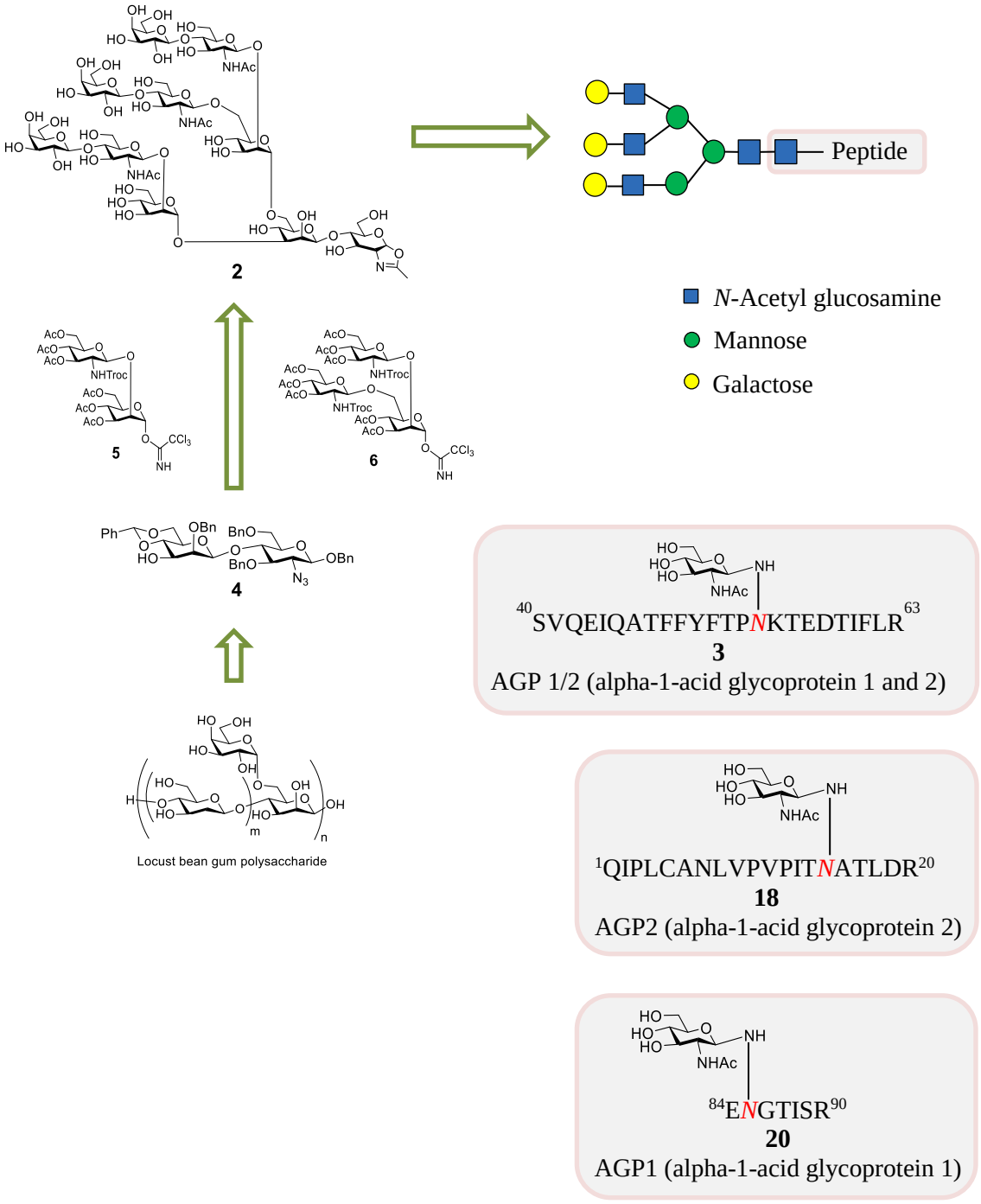
Role of AGP in cancer:

AGP (alpha-1-acid glycoprotein) is a well-known human plasma protein, synthesized in the liver and secreted into the blood presents its concentration of approximately 2.5 mg/mL in healthy normal individuals.[9] It is one of the acute phase proteins, reported for considerable increase in its plasma concentration during acute infections or inflammations. Its plasma concentration is usually between 0.4 and 1.1 g/L in healthy people. Initial investigation of AGP is focused on alteration of its serum/plasma concentration and as a result, significant alteration is reported in various types of cancers and other diseases. For instance,[6] the plasma concentrations of AGP in the breast and lung cancer groups were found to be increased 2.5 times, while in the ovary cancer group it is 1.6 times with a range between 0.45 and 2.85 g/L in these cancers. It is one of the major serum *N*-linked glycoproteins, comprised of bi- and multi-antennary *N*-glycans with exclusive expression which occurs naturally in plasma through secretion by hepatocytes, breast, cardiac cells, testes, gastrointestinal tract and immune cells such as monocytes and lymphocytes.

AGP is a highly heterogeneous glycoprotein presents greatest number of *N*-glycopeptides. For instance, recently its site-specific glycopeptides analysis by LC-MS/MS platforms as a model *N*-glycoprotein represents almost 165-site specific *N*-glycopeptides as its isoforms.[10] Tri- and tetra-antennary complex type *N*-glycans are the major glycoforms of AGP[11] and a huge alteration in its glycosylation, fucosylation and sialylation pattern of these *N*-glycans are of great interest as a serum biomarker with various types of cancers,[12] many pathophysiological, and physiological states such as inflammation, pregnancy.[13] It is one of the natural glycoproteins, exist as a mixture of glycoforms differing in oligosaccharide structures, and it is almost impossible to isolate each corresponding glycoforms for understanding its functions in detail, which is mainly hindered due to the structural microheterogeneity[10] of these *N*-glycans caused by the diverse patterns of glycosylation.

2.2. Synthesis of AGP glycopeptides

Scheme 1. Synthetic strategy of AGP glycopeptides using naturally occurring polysaccharide, locust bean gum galactomannan.



Among 109 human serum protein analytes approved through 2008 by US Food and Drug Administration (FDA),[14] we selected AGP as such a candidate providing tryptic glycopeptides because of accumulated evidence that (i) AGP displays often tri- and tetra-antennary *N*-glycans at 5 potential glycosylation sites, and (ii) not only the patterns of AGP glycoforms differ widely in the various cancer patient groups compared with the healthy control groups, but alteration of the glycan structures depends on each patient's clinical status.[15]

Scheme 1 outlined a synthetic route of the AGP glycopeptides **1, 17,** and **19** based on the chemical and enzymatic approach using the triantennary *N*-glycan derivative **2** as a donor substrate and an acceptor glycopeptides **3, 18,** and **20** in the *trans*-glycosylation catalyzed by a recombinant endo- M (N175Q).[16] Oxazoline **2** was synthesized by using a key intermediate, benzyl (2-*O*-benzyl- 4,6-*O*-benzylidine-β-D-mannopyranosyl)-(1→4)-3,6-di-*O*-benzyl-2-azido-2-deoxy-β-Dglucopyranoside (**4**), derived through β1,4-mannobiose octaacetate obtained from locust bean gum polysaccharide as an abundant material according to the method reported previously[17] with a slight modification to enhance efficiency and versatility of the synthesis of related derivatives. It is important to note that compound **4** bearing totally *O*-benzyl/benzylidene protective groups facilitates greatly a seamless approach based on chemical coupling with glycosyl donors **5**, **6**, and subsequent enzymatic galactosylation toward the synthesis of the decasaccharide oxazoline **2**. As shown in Scheme 2, a known 2- azido derivative **7**[17] was converted readily into benzyl glycoside **8** and further chemical manipulations through the di-*O*-benzylidene derivatives **9** and **10** afforded efficiently the key disaccharide azidoalcohol **4.** (see also, Experimental Section).

Scheme 3 indicates a synthetic process of asialo-triantennary decasaccharide **16**, notably a stable precursor of oxazoline **2**, from the key intermediate **4**. Use of the novel disaccharide azidoalcohol **4** facilitated greatly further chemical manipulation of the triantennary *N*-glycan derivatives **11**~**16**. Coupling of **4** with disaccharide imidate **5**[17] proceeded smoothly in the presence of trimethylsilyl trifluoroacetic acid to give the tetrasaccharide **11** in 88% yield. Next, diol **12** was subjected to the glycosylation with trisaccharide imidate **6**[17] in the presence of $BF_3{\cdot}Et_2O$ and gave a desired heptasaccharide **13** in 90% yield. After de-*O*-acetylation and conversion of an azido into acetamido

group of **13**, agalacto-triantennary heptasaccharide **14** was treated with recombinant human β1,4-galactosyltransferase (GalT) in the presence of uridine 5'-diphosphate-D-galactose (UDP-Gal) to yield asialo-triantennary decasaccharide **15** in 93% yield. As expected, removal of four *O*-benzyl protective groups in **15** was achieved by a simple hydrogenation in the presence of $Pd(OH)_2/C$ to provide the free decasaccharide **16** in 95% yield (see also, Experimental Section).

According to the method discovered by Shoda *et al*.,[18] compound **16** was treated with aqueous solution of 2-chloro-1,3-dimethylimidazolinium chloride (DMC) and triethylamine at 0°C for 40 min (Scheme 4), then unstable decasaccharide oxazoline **2** was employed directly for the *trans*-glycosylation reaction with acceptor peptides **3, 18,** and **20** having a proximal GlcNAc at Asn54 (AGP1 and AGP2), Asn15 (AGP2), and Asn85 (AGP1) residues respectively. It was demonstrated that *trans*-glycosylation reaction of the oxazoline **2** by a recombinant *endo*-M (N175Q)[16,17] in the presence of an acceptor **3, 18,** and **20** proceeded at 30°C and afforded the AGP glycopeptides **1, 17,** and **19** after 10hrs, 4hrs, and 4hrs incubation in approximately 20%, 46%, and 44% yield respectively (see also, Scheme 5, 6 and 7 in the Experimental Section). The acceptor peptides **3**, **18** and **20** were synthesized by microwave assisted solid phase peptide synthesis using 2-chlorotrityl chloride resin (see also, Experimental Section).

Finally, α-2,6 sialylation reaction of glycopeptide **1** was performed using CMP-NANA, α-2,6 sialyltransferase in presence of TRIS-HCl buffer (pH = 7.5) at r.t. and the reaction was monitored by MALDI-TOF MS analysis (see also, Scheme 8 in Experimental Section).

List of chemoenzymatically synthesized AGP glycopeptides carrying β-1,6-GlcNAc linkage containing complex type tri-antennary *N*-glycans

(i) Amino acid sequence of alpha-1-acid glycoprotein isoforms (AGP 1 and AGP 2)

Alpha-1-acid glycoprotein 1 (AGP 1): P02763

1 QIPLCANLVP VPIT*N*ATLDQ ITGKWFYIAS AFRNEEY*N*KS VQEIQATFFY FTP*N*KTEDTI
61 FLREYQTRQD QCIY*N*TTYLN VQR*EN*GTISR YVGGQEHFAH LLILRDTKTY MLAFDVNDEK
121 NWGLSVYADK PETTKEQLGE FYEALDCLRI PKSDVVYTDW KKDKCEPLEK QHEKERKQEE GES

Alpha-1-acid glycoprotein 2 (AGP 1): P19652

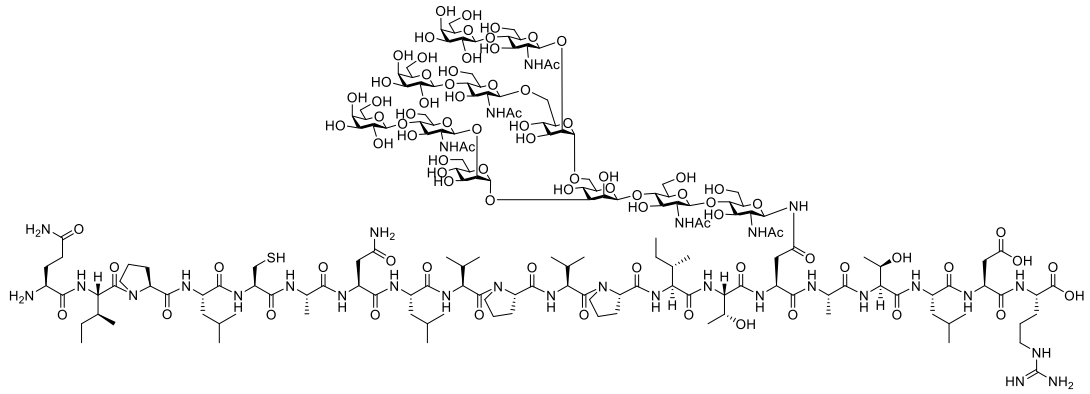
1 QIPLCANLVP VPIT*N*ATLDR ITGKWFYIAS AFRNEEY*N*KS VQEIQATFFY FTP*N*KTEDTI
61 FLREYQTRQN QCFY*N*SSYLN VQRE*N*GTVSR YEGGREHVAH LLFLRDTKTL MFGSYLDDEK
121 NWGLSFYADK PETTKEQLGE FYEALDCLCI PRSDVMYTDW KKDKCEPLEK QHEKERKQEE GES

(ii) List of selected peptide sequence of AGP to synthesize corresponding complex type tri-antennary *N*-glycopeptides

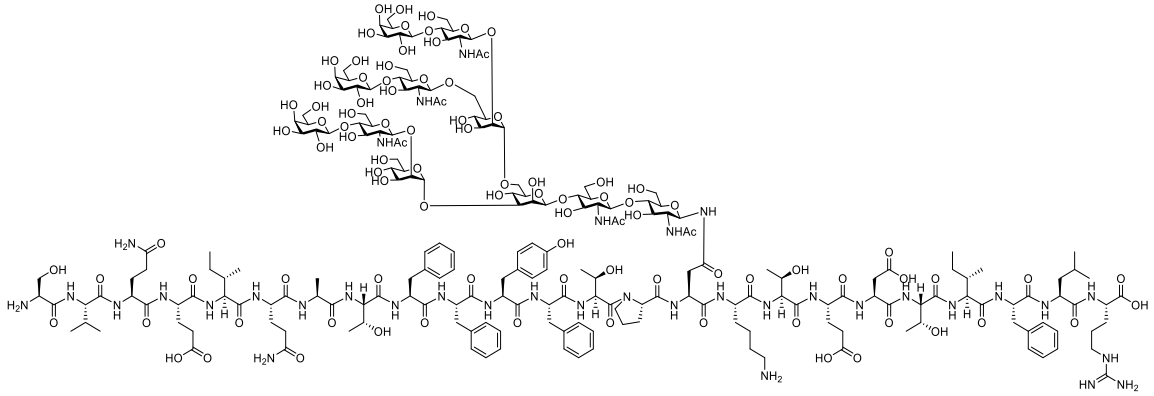
N-Glycosylation site	Alpha-1-acid glycoprotein 1 (AGP 1)	Alpha-1-acid glycoprotein 2 (AGP 2)
Site 1	---	QIPLCANLVPVPIT*N*ATLDR (1-20, Asn^{15})
Site 2	---	
Site 3	SVQEIQATFFYFTP*N*KTEDTIFLR (40-63, Asn^{54})	
Site 4	---	---
Site 5	E*N*GTISR (84-90, Asn^{85})	---

Figure 1. (i) Amino acid sequences of alpha-1-acid glycoprotein isoforms (AGP 1 and 2). Italic and bold red letters: *N*-glycosylation sites; blue letters: sites of genetic; underlined letters: tryptic digests selected to synthesis glycopeptides having tri-antennary *N*-glycan. (ii) List of synthetic *N*-glycopeptides from AGP 1 and 2. Italicized and bold red letters: *N*-glycosylation sites

a)



b)



c)

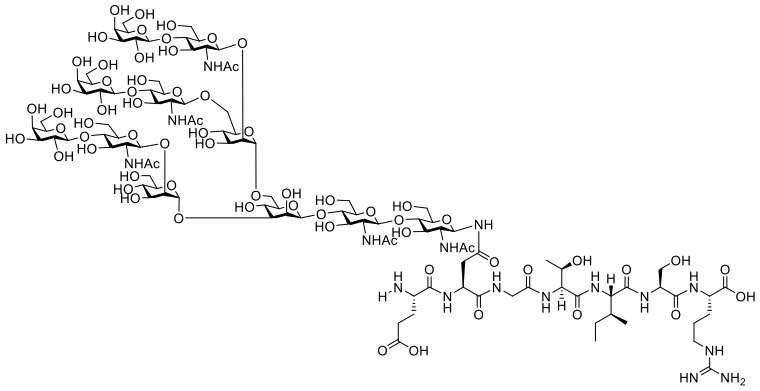
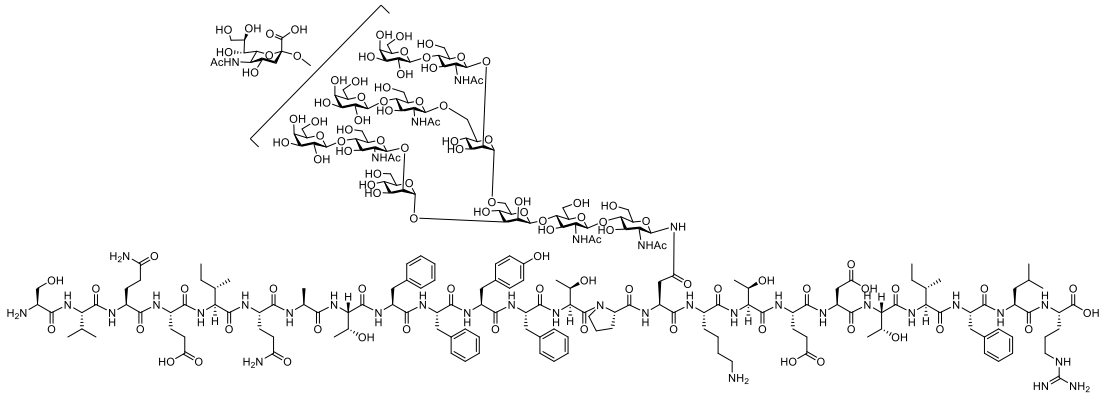


Figure 2. List of asialo form of glycopeptides of AGP synthesized. Italic, underlined and red letters: glycosylation sites (a) 20 mer glycopeptide of AGP2, glycosylation site 1: Asn[15], Peptide sequence: [1]QIPLCANLVPVPIT*N*ATLDQITGK[20] (b) 24 mer glycopeptide of AGP1/2, glycosylation site 2: Asn[54], Peptide sequence: [40]SVQEIQATFFYFTP*N*KTEDTIFLR[63] (c) 7 mer glycopeptide of AGP1, glycosylation site 5: Asn[85], Peptide sequence: [84]E*N*GTISR[90]

(a)



(b)

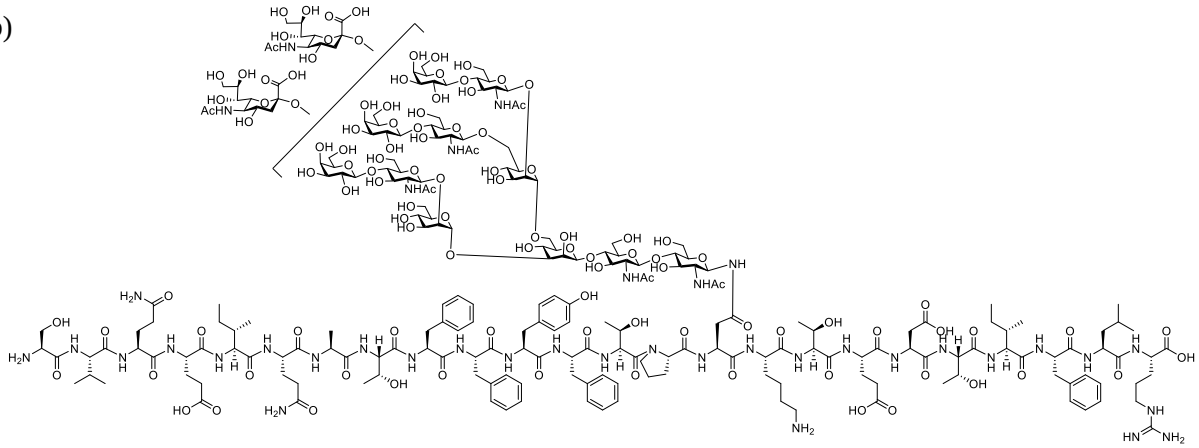
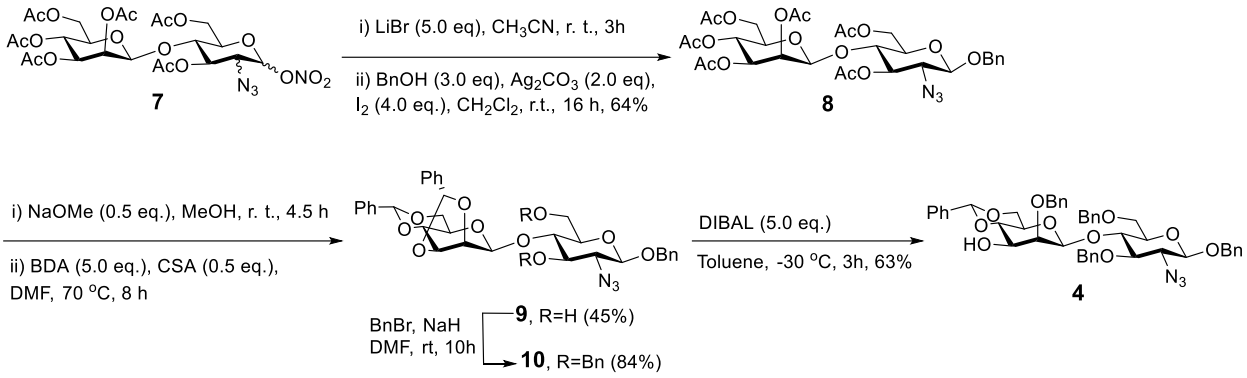


Figure 3. List of AGP α-2,6-sialyl glycopeptides synthesized; 24 mer glycopeptide of AGP1/2, glycosylation site 2: Asn[54], Peptide sequence: [40]SVQEIQATFFYFTP*N*KTEDTIFLR[63]. (a) Mono-sialylated (b) Di-sialylated

2.3. Experimental section

Scheme 2. Synthetic route to Key intermediate 4



Benzyl 2,3,4,6-tetra-*O*-acetyl-β-D-mannopyranosyl-(1→4)-3,6-di-*O*-acetyl-2-azido-2-deoxy-β-D-glucopyranoside (8).

2-Azido derivative **7** obtained readily by azido nitration of glycal[17] (4.64 g, 6.9 mmol) dissolved in dry acetonitrile was added lithium bromide (3.03g, 34.9 mmol) and stirred at r.t. for 3 h. The reaction mixture was diluted with ethyl acetate and extracted with water, washed with brine solution, dried with Na_2SO_4 and concentrated. The crude product was purified by flash column chromatography (hexane/EtOAc= 10:90) and pure 2-azidobromide (3.47g, 5.09 mmol) was co-evaporated with toluene two times, added 3.47 g of molecular sieves and once more co-evaporated with toluene, dried under vacuum for 12 h. To the residue, added dry dichloromethane (51 mL) and stirred at r.t for 15 min. To the resulting solution was added iodine (2.6 g, 20.3 mmol) and the solution was stirred at r.t. for 1 h and then silver carbonate (2.8 g, 10.1 mmol) and benzyl alcohol (1.58 mL, 15.2 mmol) were added and stirred at r.t. for 16 h. The mixture was diluted with chloroform followed by celite filtration, organic layer was washed with $Na_2S_2O_3$ solution, $NaHCO_3$ solution, brine solution, and dried over Na_2SO_4. The residual product was purified by flash column chromatography (hexane/EtOAc=65:35) to give benzyl glycoside **8** (1.8 g, 61%) as a white powder.

^{1}H NMR (500 MHz, $CDCl_3$, 25°C, TMS)**:** δ 7.37-7.31 (m, 5 H), 5.37 (d, J = 3.16 Hz, 1 H;), 5.19 (t, J = 9.9 Hz, 1 H), 5.00 (dd, J = 3.44 Hz, 2.87 Hz, 1 H), 4.96- 4.89 (m, 2 H), 4.68 (d, J = 12.05 Hz, 1 H), 4.63 (s, 1 H), 4.41-4.36 (m, 2 H), 4.33 (dd, J = 5.45 Hz, 5.17 Hz, 1 H), 4.23 (dd, J = 4.02 Hz, 4.30 Hz,

1 H), 4.11 (dd, *J* = 2.58 Hz, 1.15 Hz, 1 H), 3.77 (t, *J* = 9.76 Hz, 1 H), 3.61 (m, 1 H), 3.55 (m, 1 H), 3.45 (t, *J* = 9.04 Hz, 1 H), 2.155, 2.145, 2.14, 2.08, 2.04, 1.98 (s each, 3H each, 6xCH_3CO).

^{13}C NMR (125 MHz, $CDCl_3$, 25°C, TMS): δ 128.4, 100.2, 97.4, 74.7, 72.6, 72.5, 71.5, 71.2, 71.2, 70.7, 68.2, 65.8, 63.9, 62.3, 62.3, 62.3, 62.3, 20.8, 20.8, 20.8, 20.8, 20.8, 20.8.

MALDI-TOF MS: *m/z* calculated for $C_{31}H_{39}N_3NaO_{16}$, $[M+Na]^+$ = 732.222, found 732.248.

Benzyl (2,3:4,6-di-*O*-benzylidine-β-D-mannopyranosyl)-(1→4)-2-azido-2-deoxy-β-D-glucopyranoside (9).

To a solution of **8** (8.1 g, 11.46 mmol) in dry methanol (114 mL, 0.1 M), added sodium methoxide (309 mg, 5.73 mmol) and stirred at r.t., for 12 h. Then the solution was neutralized by Dowex 50W-X8 [H^+] resin. After the filtration, the reaction mixture was concentrated and dried in a vaccum for 3 h. Dry *N*,*N*-dimethylformamide (114 mL, 0.1 M), benzaldehyde dimethyl acetal (5.1 mL, 34.39 mmol) and (±)-10-camphorsulfonic acid (799 mg, 3.43 mmol) were added under nitrogen gas and stirred at 70°C under reduced pressure for 3 h. Added aq. $NaHCO_3$ slowly at r.t. and extracted with EtOAc. The organic phase was dried (Na_2SO_4) and concentrated. The crude residue was purified by flash column chromatography (hexane/EtOAc=72:28) to give diastereomeric mixture **9** (3.32 g, 45%) as a white powder. Characterization data of the pure *endo*-isomer isolated were listed as follows.

(Endo isomer): ^{1}H NMR (500 MHz, $CDCl_3$, 25°C, TMS): δ 7.52-7.48 (m, 2 H), 7.47-7.43 (m, 2 H), 7.41-7.31 (m, 11 H), 6.30 (s, 1 H), 5.61(s, 1 H), 5.05 (d, *J* = 2.63 Hz, 1 H), 4.90 (d, *J* = 11.96 Hz, 1 H), 4.72 (d, *J* = 11.96 Hz, 1 H), 4.58 (dd, *J* = 6.13 Hz, 5.83 Hz, 1 H), 4.46-4.40 (m, 3 H), 4.12 (dd, *J* = 7.88 Hz, 1 H), 3.89-3.93 (m, 1 H), 3.86 (t, *J* = 10.50 Hz, 1 H), 3.76-3.80 (m, 1 H), 3.73 (t, *J* = 9.33 Hz, 1 H), 3.64-3.55 (m, 2H), 3.47-3.35 (m, 2H), 2.35 (s, 1H), 1.90 (dd, *J* = 4.38 Hz, 4.67 Hz, 1H).

^{13}C NMR (125 MHz, $CDCl_3$, 25°C, TMS): δ 128.3, 126.2, 126.2, 104.7, 101.9, 100.7, 100.5, 81.2, 76.8, 76.7, 74.3, 73.8, 73.3, 71.7, 71.6, 68.7, 68.7, 65.7, 65.6, 61.2, 61.2.

MALDI-TOF MS: *m/z* calculated for $C_{33}H_{35}N_3NaO_{10}$, $[M+Na]^+$ = 656.221, found 655.878.

Benzyl (2,3:4,6-di-*O*-benzylidine-β-D-mannopyranosyl)-(1→4)-2-azido-3,6-di-*O*-benzyl-2-deoxy-β-D-glucopyranoside (10).

To a solution of mixture **9** (3.32g, 5.23 mmol) in dry *N*,*N*-dimethyl formamide (52 mL, 0.1 M), added 50% sodium hydride (377 mg, 15.7 mmol) at -15°C and stirred for 15 min. Then added benzyl bromide (1.86 mL, 15.7 mmol) and stirred for o.n., added aq. $NaHCO_3$ was slowly and extracted with EtOAc. The organic phase was washed with water, brine, dried (Na_2SO_4) and concentrated. The crude residue was purified by flash column chromatography (hexane/EtOAc=85:15) to give a white powder as a diastereo-mixture **10** (3.59 g, 84%). Characterization data of the pure *endo*-isomer isolated were listed as follows.

(Endo isomer): ^{1}H NMR (500 MHz, $CDCl_3$, 25°C, TMS): δ 7.53-7.49 (m, 2 H), 7.46-7.43 (m, 2 H), 7.42-7.26 (m, 21 H), 5.98 (s, 1 H), 5.28 (s, 1 H), 5.0 (d, J = 10.81 Hz, 1 H), 4.96-4.91 (m, 2 H), 4.80-4.72 (m, 2 H), 4.68 (d, J = 11.97 Hz, 1 H), 4.46 (d, J = 11.97 Hz, 1 H), 4.29 (d, J = 8.18 Hz, 1 H), 4.23 (t, J = 6.72 Hz, 1 H), 4.14-4.03 (m, 2 H), 4.04 (dd, J = 2.04 Hz, 1 H), 3.81 (dd, J = 3.21 Hz, 3.92 Hz, 1H), 3.73-3.67 (m, 2 H), 3.52 (t, J = 9.35 Hz, 1H), 3.45-3.36 (m, 2 H), 3.32 (t, J = 10.51 Hz, 1 H), 3.16 (m, 1 H).

^{13}C-NMR (125 MHz, $CDCl_3$, 25°C, TMS): δ 129.4, 128.3, 127.9, 127.7, 126.6, 126.1, 105.0, 101.6, 100.8, 98.5, 81.1, 79.8, 76.8, 76.4, 76.1, 75.0, 75.0, 74.6, 73.8, 73.8, 71.1, 71.1, 68.6, 68.6, 67.9, 67.9, 65.9, 65.4.

MALDI-TOF MS: *m/z* calculated for $C_{47}H_{47}N_3NaO_{10}$, $[M+Na]^+$ = 836.315, found 835.698.

Benzyl (2-*O*-benzyl-4,6-*O*-benzylidine-β-D-mannopyranosyl)-(1→4)-2-azido-3,6-di-*O*-benzyl- -2-deoxy-β-D-glucopyranoside (4).

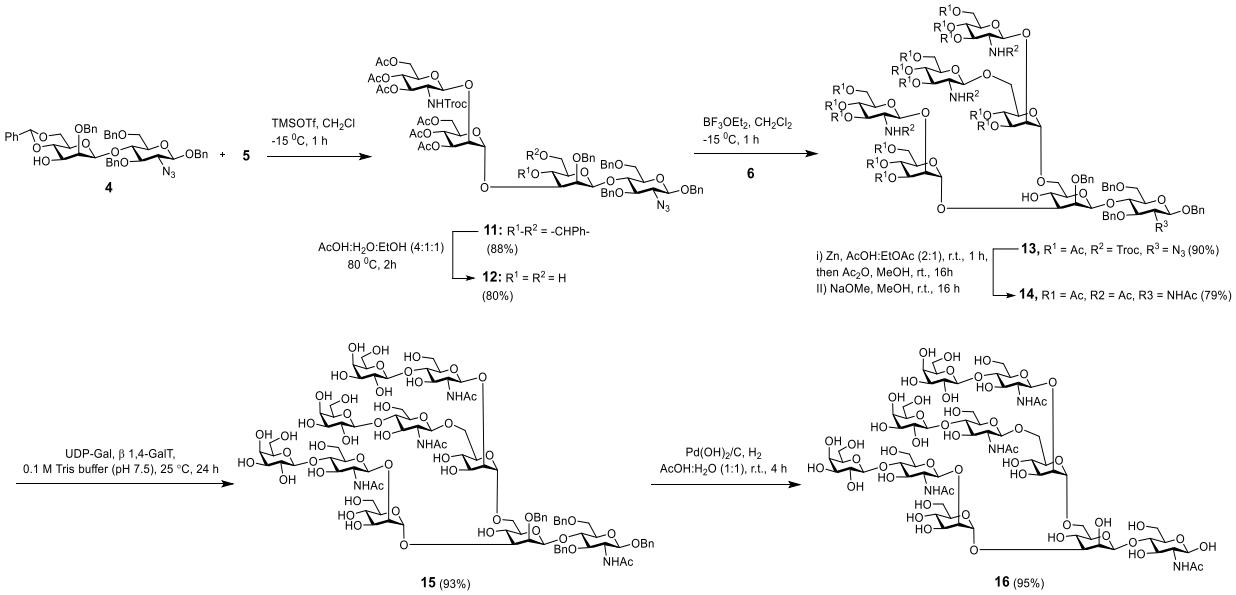
To a solution of **10** (1 g, 1.22 mmol) in toluene (6.14 mL, 0.2 M), added 1 M DIBAL (540 µL, 0.54 mmol) in toluene and stirred at -30°C for 2 h. Then, added 1 M DIBAL (360 µL, 0.36 mmol) in toluene and stirred at -30°C for another 3 h. The reaction mixture was quenched with 10% aq. KOH at 0^0C and extracted with diethyl ether, dried ($MgSO_4$) and concentrated. The crude residue was purified by flash column chromatography (hexane/EtOAc = 80:20) to give **4** (640mg, 63%) as a white powder.

^{1}H-NMR (500 MHz, $CDCl_3$, 25°C, TMS): δ 7.45 (d, J = 7.26 Hz, 2 H; H-Ar), 7.43-7.25 (m, 23 H; H-Ar), 5.45 (s, 1 H; H-CPh), 5.05 (d, J = 10.45 Hz, 1 H; H_2CPh), 4.96-4.89 (t, J = 11.61 Hz, 10.16 Hz, 2 H; H_2CPh), 4.77-4.67 (t, J = 13.93 Hz, 13.06 Hz, 2 H; H_2CPh), 4.67-4.61 (t, J = 11.03 Hz, 10.16 Hz, 2 H; H_2CPh), 4.47 (d, J = 11.9 Hz, 1 H; H_2CPh), 4.59 (s, 1 H; H'-1), 4.3 (d, J = 8.13 Hz, 1 H; H-1), 4.10-4.05 (dd, J = 4.64 Hz, 4.64 Hz, 1 H; H'-6a), 4.01 (t, J = 9.43 Hz, 1 H; H-4), 3.74-3.64 (m, 4 H; H'-2, H'-4, H-6ab), 3.57-3.43 (m, 3 H; H-2, H'-3, H'-6b), 3.35-3.30 (m, 2H; H-3, H-5), 3.08 (m, 1H; H'-5), 2.32 (d, J = 8.42 Hz, 1 H; OH).

^{13}C-NMR (125 MHz, $CDCl_3$, 25°C, TMS): δ 128.2, 127.7, 126.2, 101.9, 101.7, 100.5, 81.6, 79.1, 79.0, 77.6, 75.9, 75.9, 75.3, 75.3, 74.9, 73.9, 73.9, 71.0, 70.9, 70.9, 68.4, 68.3, 67.0, 65.8.

MALDI-TOF MS: *m/z* calculated for $C_{47}H_{49}N_3NaO_{10}$, $[M+Na]^+$ = 838.331, found 838.330.

Scheme 3. Synthesis of compound **16** using **4, 5** and **6.**



Benzyl [(3,4,6-tri-*O*-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl)-(1→3)-2-*O*-benzyl-4,6-*O*-benzylidine-β-D-mannopyranosyl]-(1→4)-2-azido-2-deoxy-3,6-di-*O*-benzyl-β-D-glucopyranoside (20).

A glycosyl donor **5**[17] (830 mg, 912 µmol) and glycosyl acceptor **4** (496 mg, 608 µmol) was dissolved and co-evaporated with toluene two times and once more in presence of magnetic stirrer and molecular sieves (830 mg), dried under vacuum for 12 h. Dry DCM (6 mL, 0.1 M) was added and cooled to -15°C. After 10 min, TMSOTf (11.8 µL, 60 µmol) was added at -15°C and the reaction mixture was stirred for 1 hour at -15°C and then quenched with triethylamine. The mixture was diluted with EtOAc, washed with sat. $NaHCO_3$, brine, dried (Na_2SO_4), filtered and concentrated. The crude residue was purified by flash column chromatography on silica gel (Toluene/EtOAc=70:20) to give compound **11** (845 mg, 88%).

^{1}H NMR (600 MHz, $CDCl_3$, 25°C)**:** δ ppm 7.52 (s, 5 H), 7.42-7.25 (m, 20 H), 5.42 (s, 1 H), 5.21 (t, 1 H, *J* = 9.7 Hz), 5.06 (d, 1 H, *J* = 10.0 Hz), 5.01-4.99 (m, 2 H), 4.94 (d, 1 H, *J* = 12.0 Hz), 4.85 (t, 1 H, *J* = 9.4 Hz), 4.82- 4.62 (m, 8 H), 4.58 (s, 1 H), 4.50 (d, 1 H, *J* = 11.7 Hz), 4.32 (d, 1 H, *J* = 7.7 Hz),

4.21 (brd, 1 H, *J* = 8.8 Hz), 4.16-4.01 (m, 8 H), 3.84 (d, 1 H, *J* = 12.0 Hz), 3.77-3.63 (m, 6 H), 3,57 (m, 1 H), 3,51 (t, 1 H, *J* = 8.8 Hz), 3.42- 3.34 (m, 3 H), 3.03 (m, 1 H), 2.74 (brd, 1 H, *J* = 8.8 Hz), 2.07-2.06 (m, 9 H), 2.04 (s, 3 H), 2.01 (s, 3 H), 2.00 (s, 3 H) ; see, Figure S9.

^{13}C NMR (150 MHz, $CDCl_3$, 25°C)**:** δ ppm 128.8, 128.4, 128.2, 126.9, 102.3, 101.0, 100.6, 98.7, 81.7, 78.8, 77.7, 76.6, 76.6, 75.7, 75.4, 75.2, 75.1, 74.7, 74.1, 73.9, 71.9, 70.9, 70.9, 71.2, 69.9, 68.8, 68.5, 68.4, 68.3, 68.3, 68.2, 66.9, 65.9, 65.8, 62.8, 61.6, 61.6, 55.4, 20.9, 20.9; see, Figure S10.

MALDI-TOF MS: *m/z* calculated for $C_{74}H_{83}Cl_3N_4NaO_{27}$, $[M+Na]^+$ = 1587.420, found 1587.654.

Benzyl [(3,4,6-tri-*O*-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl)-(1→3)-2-*O*-benzyl-β-D-mannopyranosyl]-(1→4)-2-azido-2-deoxy-3,6-di-*O*-benzyl-β-D-glucopyranoside (21).

A solution of Compound **11** (400 mg) in acetic acid: water: ethanol = 4:1:1 (12 mL) was stirred at 80°C. After reaction completion by TLC, the reaction mixture was evaporated. The crude residue was purified by flash column chromatography on silica gel (hexane/EtOAc=40:60) to give diol **12** (300 mg, 80%).

^{1}H NMR (600 MHz, $CDCl_3$, 25°C)**:** δ ppm 7.40-7.25 (m, 20 H), 5.34 (brs, 1 H), 5.13 (t, 1 H, *J* = 9.0 Hz), 5.08 (d, 1 H, *J* = 9.5 Hz), 5.03 (d, 1 H, *J* = 10.8 Hz), 4.96 (d, 1 H, *J* = 11.3 Hz), 4.95-4.89 (m, 3 H), 4.86 (brd, 1 H, *J* = 9.8 Hz), 4.73-4.64 (m, 4 H), 4.60 (d, 1 H, *J* = 12.1 Hz), 4.55-4.53 (m, 3 H), 4.29 (d, 1 H, *J* = 8.1 Hz), 4.19 (brd, 1 H, *J* = 12.1 Hz), 4.13-3.93 (m, 8 H), 3.76-3.66 (m, 4 H), 3.56 (m, 1 H), 3.49-3.46 (m, 3 H), 3.36 (d, 1 H, *J* = 9.5 Hz), 3.29-3.26 (m, 2 H), 3.21 (brs, 1 H), 3.06 (m, 1 H), 2.09 (s, 3 H), 2.04 (s, 6 H), 2.02 (s, 9 H) ; see, Figure S11.

^{13}C NMR (150 MHz, $CDCl_3$, 25°C)**:** δ ppm 128.3, 128.2, 127.7, 126.7, 100.9, 100.4, 100.4, 96.8, 81.3, 81.0, 79.1, 77.1, 75.7, 75.1, 75.0, 75.0, 74.9, 74.5, 74.1, 73.9, 73.9, 73.9, 71.4, 71.4, 70.9, 70.8, 70.2, 70.1, 70.1, 69.6, 69.1, 68.6, 68.3, 66.9, 66.3, 66.3, 65.8, 64.1, 63.3, 63.1, 62.7, 62.6, 62.0, 62.0, 55.5, 20.8, 20.8; see, Figure S12.

MALDI-TOF MS: *m/z* calculated for $C_{67}H_{79}Cl_3N_4O_{27}$, $[M+H]^+$ = 1477.407, $C_{67}H_{79}Cl_3N_4NaO_{27}$, $[M+Na]^+$ =1499.389, found 1476.030 and 1502.174.

Benzyl {(3,4,6-tri-*O*-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl)-(1→3)-[(3,4,6-tri-*O*-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)- (3,4,6-tri-*O*-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl)-(1→6)]-3,4-di-*O*-acetyl-α-D-mannopyranosyl)-(1→6)-2-*O*-benzyl-β-D-mannopyranosyl}-(1→4)-2-azido-2-deoxy-3,6-di-*O*-benzyl-β-D-glucopyranoside (22).

A solution of glycosyl acceptor **12** (236 mg, 160 μmol) and glycosyl donor **6**[17] (319 mg, 239 μmol) was co-evaporated with toluene twice and once more in presence of molecular sieves 4 A° (320mg), magnetic stirrer and dried under vacuum for 12 h. Dry DCM (32 mL, 0.005 M) was added and cooled to -35°C. After 30 min, $BF_3.Et_2O$ (8 μL, 63 μmol) was added at -35°C and the reaction mixture was stirred for 2 h at -35°C, and then the mixture was filtered, diluted with EtOAc and washed with sat. $NaHCO_3$, brine, dried (Na_2SO_4), filtered and concentrated. The crude residue was purified by column chromatography on silica gel (toluene/EtOAc=65:35) to give compound **13** (387 mg, 90%).

^{1}H NMR (600 MHz, $CDCl_3$, 25°C)**:** δ ppm 7.40-7.15 (m, 20 H), 6.27 (brs, 1 H), 5.83 (d, 1 H, *J* = 9.5 Hz), 5.54 (t, 1 H, *J* = 9.2 Hz), 5.35 (brs, 1 H), 5.54 (t, 1 H, *J* = 10.1 Hz), 5.20-4.52 (m, 30H), 4.31-4.26 (m, 3 H), 4.21-4.06 (m, 11 H), 4.02 (brd, 1 H, *J* = 7.8 Hz), 3.98 (d, 1 H, *J* = 9.5 Hz), 3.95-3.89 (m, 5 H), 3.80 (m, 1 H), 3.74-3.58 (m, 9 H), 3.52-3.46 (m, 2 H), 3.37-3.28 (m, 4 H), 3.22 (m, 1 H), 3.15-3.07 (m, 2 H), 2.95 (brd, 1 H, *J* = 10.4 Hz), 2.11 (s, 3 H), 2.09 (s, 6 H), 2.04-1.99 (m, 36 H) ; see, Figure S13.

^{13}C NMR (150 MHz, $CDCl_3$, 25°C)**:** δ ppm 128.4, 128.2, 128.2, 128.0, 126.5, 101.9, 100.5, 100.3, 100.3, 97.3, 96.8, 96.5, 81.0, 81.0, 78.8, 76.7, 74.9, 74.9, 74.8, 74.7, 74.6, 74.2, 74.0, 73.9, 73.9, 73.8, 73.2, 72.9, 71.9, 71.8, 71.6, 71.3, 71.1, 71.0, 71.0, 70.9, 70.6, 70.1, 70.1, 70.0, 69.3, 69.2, 69.2, 68.6,

68.4, 68.3, 67.6, 67.5, 67.4, 67.4, 66.9, 66.0, 65.9, 65.8, 62.2, 62.1, 62.1, 61.8, 61.7, 56.3, 56.1, 56.0, 55.7, 21.0, 20.8, 20.8; see, Figure S14.

MALDI-TOF MS: *m/z* calculated for $C_{107}H_{129}Cl_9N_6NaO_{52}$, $[M+Na]^+$ = 2667.472, found 2665.197.

Benzyl {(2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→2)-2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→3)-α-D-mannopyranosyl)-(1→3)-[(2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→2)-2-acetimido-2-deoxy-β-D-glucopyranosyl)-(1→6)]-α-D-mannopyranosyl)-(1→6)-2-*O*-benzyl-β-D-mannopyranosyl}-(1→4)-2-azido-2-deoxy-3,6-di-*O*-benzyl-β-D-glucopyranoside(23).

To a solution of compound **13** (381 mg, 143 μmol) in EtOAc (13.0 mL), added zinc (3.5 g, 53.19 mmol), acetic acid (750 μL, 35.94 mmol) and the solution was stirred at r.t., for 1 h. The reaction mixture was filtered through celite and washed with EtOAc. To the filtrate, added Ac_2O (800 μL) and the resulting mixture were stirred at r.t., for 15 h. The reaction mixture was concentrated and co-evaporated with toluene. The residue (300 mg) was dissolved in MeOH (4 mL). To the resulting solution, added NaOMe (6.9 mg, 131 mmol) and the mixture was stirred at room temperature for 16 h. The reaction mixture was neutralized using 20% acetic acid and concentrated. The residue was dissolved in water, lyophilized, and purified by RP-HPLC to give compound **14** (184 mg, 79%).

^{1}H NMR (600 MHz, D_2O, 25°C)**:** δ ppm 7.41-7.20 (m, 13 H), 6.98 (m, 2 H), 5.08 (brs, 1 H), 4.77 (brs, 1 H), 4.70-4.64 (m, 3 H), 4.56 (d, 1 H, *J* = 11.7 Hz), 4.49 (d, 1 H, *J* = 12.3 Hz), 4.48 (brs, 1 H), 4.47 (d, 1 H, *J* = 11.2 Hz), 4.44 (d, 1 H, *J* = 8.2 Hz), 4.41 (d, 1 H, *J* = 8.8 Hz), 4.33 (d, 1 H, *J* = 12.3 Hz), 4.30 (d, 1 H, *J* = 8.2 Hz), 4.07 (m, 1 H), 4.03 (d, 1 H, *J* = 10.0 Hz), 3.97 (d, 1 H, *J* = 8.3 Hz), 3.93 (t, 1 H, *J* = 10.0 Hz), 3.91-3.88 (m, 2 H), 3.82-3.79 (m, 3 H), 3.72 (dd, 1 H, *J* = 3.5, 9.4 Hz), 2.47 (m, 1 H), 1.98 (s, 3 H), 1.94 (s, 3 H), 1.93 (s, 3 H), 1.47 (s, 3 H) ; see, Figure S15.

^{13}C NMR (150 MHz, D_2O, 25°C)**:** δ ppm 128.6, 128.5, 128.5, 128.4, 128.4, 128.3, 128.7, 101.3, 101.0, 99.6, 99.4, 99.3, 99.0, 96.3, 79.4, 79.4, 78.3, 77.8, 76.5, 76.4, 75.6, 75.0, 74.6, 74.5, 74.0, 73.9,

73.5, 73.5, 73.4, 73.1, 73.0, 72.7, 71.5, 71.4, 71.4, 69.8, 69.7, 69.7, 69.4, 69.4, 69.0, 68.0, 67.9, 67.4, 67.1, 65.9, 65.1, 65.0, 61.4, 60.5, 60.5, 59.8, 59.7, 55.5, 55.4, 54.5, 22.5, 22.4, 22.1; see, Figure S16.

MALDI-TOF MS: *m/z* calculated for $C_{78}H_{108}N_4NaO_{36}$, $[M+Na]^+$ = 1699.664, found 1699.299.

Benzyl {(β-D-galactopyranosyl-(1→4)-2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→2)- β-D-galactopyranosyl-(1→4)-2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→4)-α-D-mannopyranosyl)-(1→3)-[(β-D-galactopyranosyl-(1→4)-2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→2)-β-D-galactopyranosyl-(1→4)-2-acetimido-2-deoxy--β-D-glucopyranosyl)-(1→6)]-α-D-mannopyranosyl)-(1→6)-2-*O*-benzyl-β-D-manno pyranosyl}-(1→4)-2-azido-2-deoxy-3,6-di-*O*-benzyl-β-D-glucopyranoside (24).

To a solution of compound **14** (3.7 mg, 2.2 μmol) in 110.2 μL of water (20 mM; theoretical concentration), added a mixture of 1 M Tris buffer (pH 7.5) (24 μL), 1 M $MnCl_2$ (2.4 μL), 200 mM UDP-galactose (90 μL), 4 units/mL human recombinant β1, 4-galactosyltransferase (GalT) (24 μL) and water (99.6 μL). After incubation at 25°C for 24 h, the crude product was purified by RP-HPLC (column: Inertile ODS-3, 20x250 mm) eluted with H_2O and CH_3CN to afford decasaccharide **15** (2.4 mg, 93%).

^{1}H NMR (600 MHz, D_2O, 25°C)**:** δ ppm 7.41-7.19 (m, 13 H), 6.97 (m, 2 H), 5.08 (brs, 1 H), 4.70-4.64 (m, 3 H), 4.56 (d, 1 H, *J* = 11.7 Hz), 4.49 (d, 1 H, *J* = 12.6 Hz), 4.47-4.45 (m, 3 H), 4.43 (d, 1 H, *J* = 8.3 Hz), 4.37 (d, 1 H, *J* = 7.8 Hz), 4.36 (d, 1 H, *J* = 7.8 Hz),4.32 (d, 1 H, *J* = 7.9 Hz), 4.30 (d, 1 H, *J* = 7.7 Hz), 4.07 (m, 1 H), 4.04 (d, 1 H, *J* = 9.3 Hz), 3.99 (m, 1 H), 3.97 (d, 1 H, *J* = 8.3 Hz), 3.93 (t, 1 H, *J* = 9.9 Hz), 3.92-3.86 (m, 4 H), 3.82-3.79 (m, 5 H), 3.75 (m, 1 H), 2.56 (m, 1 H), 1.97 (s, 3 H), 1.93 (s, 6 H), 1.47 (s, 3 H) ; see, Figure S17.

^{13}C NMR (150 MHz, D_2O, 25°C)**:** δ ppm 128.6, 128.5, 128.5, 128.4, 127.6, 102.8, 102.8, 102.6, 101.2, 101.0, 99.5, 99.4, 99.2, 98.9, 96.2, 79.4, 79.3, 78.2, 77.8, 77.6, 76.4, 76.3, 75.2, 74.6, 74.5, 74.4, 74.0, 73.9, 73.9, 73.5, 73.1, 73.0, 72.5, 71.4, 71.3, 70.9, 69.9, 69.9, 69.3, 68.4, 67.9, 67.8, 67.4,

67.0, 66.0, 65.1, 65.0, 61.4, 61.4, 60.9, 60.0, 59.9, 59.2, 59.2, 54.9, 54.7, 22.5, 22.5, 22.2; see, Figure S18.

MALDI-TOF MS: *m/z* calculated for $C_{96}H_{138}N_4NaO_{51}$, $[M+Na]^+$ = 2185.822, found 2186.200.

β-D-galactopyranosyl-(1→4)-2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→2)-β-D-galactopyranosyl-(1→4)-2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→4)-α-D-mannopyranosyl)-(1→3)-[(β-Dgalactopyranosyl-(1→4)-2-acetimido-2-deoxy-β-D-glucopyranosyl-(1→2)-β-D-galactopyranosyl-(1→4)-2-acetimido-2-deoxy--β-D-glucopyranosyl)-(1→6)]-α-D-mannopyranosyl)-(1→6)-β-D-mannopyranosyl-(1→4)-2-acetamido-deoxy-D-glucopyranose (16).

To a solution of compound **15** (9 mg, 3.6 μmol) in 50% AcOH (aq.) (4.0 mL), added 20% $Pd(OH)_2/C$ (5 mg) and stirred for 4 h under H_2 atmosphere. The reaction mixture was filtered through a celite pad, the filtrate was concentrated and then the residue was dissolved in water and lyophilized. The crude product was purified on a Sephadex G15 column by elution with H_2O. Fractions containing the product were pooled and lyophilized to give the free dodecasaccharide **16** (7.3 mg, 95%) as white solid.

^{1}H NMR (600 MHz, D_2O, 25°C)**:** δ ppm 5.10 (d, 1 H, *J* = 3.2 Hz), 5.02 (brs, 1 H), 4.77 (brs, 1 H), 4.67-4.61 (m, 1 H), 4.50-4.47 (m, 2 H), 4.45 (d, 1 H, *J* = 8.2 Hz), 4.38-4.35 (m, 3 H), 4.15 (m, 1 H), 4.11-4.09 (m, 2 H), 3.99 (m, 1 H), 3.30 (m, 1 H), 1.95 (s, 6 H), 1.94 (brs, 6 H) ; see, Figure S19.

^{13}C NMR (150 MHz, D_2O, 25°C)**:** δ ppm 102.8, 102.8, 102.8, 101.4, 100.1, 99.3, 99.3, 97.0, 94.8, 90.4, 80.2, 79.6, 78.3, 76.4, 76.2, 75.2, 74.5, 72.4, 71.9, 70.9, 70.2, 70.1, 68.4, 67.4, 67.2, 65.5, 65.3, 65.3, 61.6, 61.6, 61.0, 60.0, 59.8, 54.9, 53.5, 23.3, 23.3, 22.5, 22.3; see, Figure S20.

MALDI-TOF MS: *m/z* calculated for $C_{68}H_{114}N_4NaO_{51}$, $[M+Na]^+$ = 1825.634, found 1826.191.

Synthesis of peptide fragments (3, 18 and 20) of AGP carrying GlcNAc on asparagine residue.

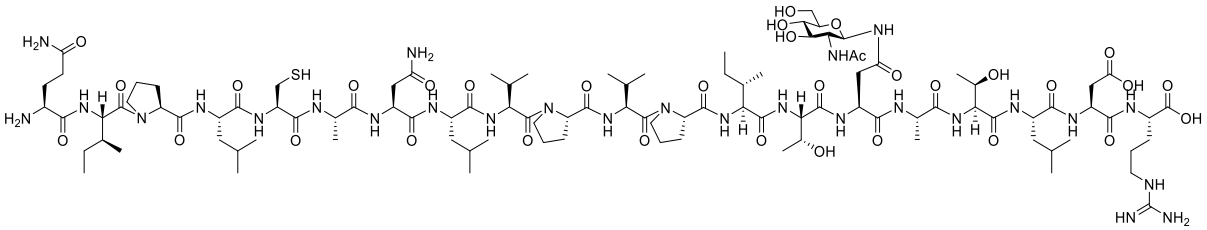
2-Chloro trityl chloride resin (1.58 mmol/g, 100mg, 158 μmol), Fmoc-amino acids (96 μmol, 4.0 eq.), and Fmoc-N(Ac_3AcNH-β-Glc)-OH (189.6 μmol, 1.2 eq.) were used. Fmoc amino acids used are Fmoc-Arg(Pbf)-OH, Fmoc-Leu-OH, Fmoc-Phe-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Lys(Boc)-OH, Fmoc-Tyr(t-Bu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Val-OH, Fmoc-Ser(t-Bu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Ala-OH, Fmoc-Ile-OH and Fmoc-Thr(t-Bu)-OH. 50 mg of 2-Chloro trityl chloride resin was placed in a 10 mL Libra tube and allowed to swell in DCM for a period of 2h at r.t., removed the DCM after that. Very first Fmoc amino acid was coupled to the resin by adding a base, DIEA (144 μmol, 6 eq.) under microwave irradiation for 9 min and methanol capping was performed for 10 secs at room temperature. From 2nd Fmoc amino acids coupling onwards, each corresponding Fmoc amino acid (632 μmol, 4.0 eq.) dissolved in a mixture of HBTU, HOBt in DMF (96 μmol, 4.0 eq.), and DIEA (144 μmol, 6.0 eq.) (final concentration of amino acid, 0.4 M) was added to the resin and the mixture was shaken under microwave irradiation for 10 min. For an introduction of the Fmoc-glycosylated amino acid, Fmoc-Asn(Ac_3AcNH-β-Glc)-OH (36 μmol, 1.5 eq.) dissolved in a mixture of PyBOP, HOBT in DMF (36 μmol, 1.5 eq.), and DIEA (72 μmol, 3 eq.) (final concentration of amino acid, 0.2 M) was treated with a resin under microwave irradiation for 15 min. After coupling of each amino acid, acetyl capping of unreacted amino acids was performed with a solution of Ac_2O:DIEA:DMF (1.0:0.5:8.5) followed by the Fmoc removal reaction conducted with 20% piperidine in DMF (2 mL) and the mixture was shaken under microwave irradiation for 3 min. Following filtration and washing with DMF and DCM (3 mL, three times each), Fmoc-removal, coupling, and capping procedures as described above were carried out repeatedly. After completion of the synthesis, the glycopeptidyl-resin was treated with TFA:H_2O:TIS (95.0:2.5:2.5) (2.0 mL) at r.t. for 2 h and the resin was filtered. The resin was washed twice with the same cocktail and the filtrates were combined and concentrated by streaming of nitrogen gas. The GlcNAc peptide was precipitated by adding cold *tert*-butyl methyl ether, discarded the liquid which is separated after centrifugation and the residue obtained was dissolved in 50% aq. acetonitrile and lyophilized. Then, the residue was dissolved in methanol (5.0 mL) and the pH was adjusted to 12.8

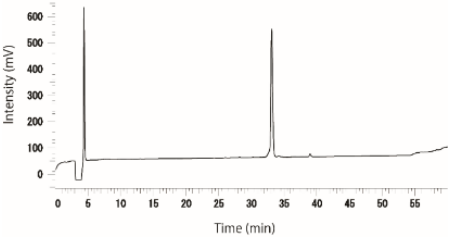
using 1 N sodium hydroxide, stirred for 2 h at r.t. The mixture was neutralized by an addition of 1 N acetic acid and evaporated. The crude material was purified by RP-HPLC to get pure compounds.

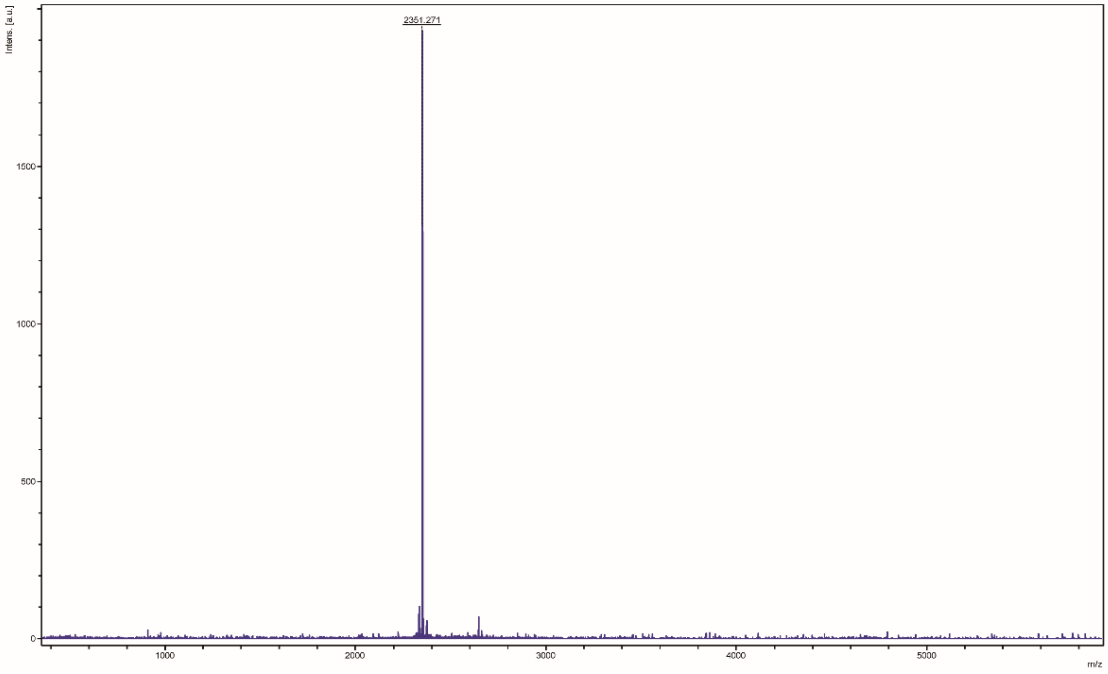
20 mer peptide fragment (18) of alpha-1-acid glycoprotein 2 (AGP2) carrying *N*-acetyl glucosamine residue on asparagine of glycosylation site Asn15. (5.2 mg, 23% overall yield calculated from the resin employed, 9.6 µmol)

Analytical HPLC: A:B (0.1% TFA in H_2O:0.1% TFA in CH_3CN), A/B gradient 0-50 min, 15-40%

MALDI-TOF MS: *m/z* calculated for $C_{103}H_{176}N_{27}O_{33}S$, $[M+H]^+$ = 2351.264, found = 2351.271



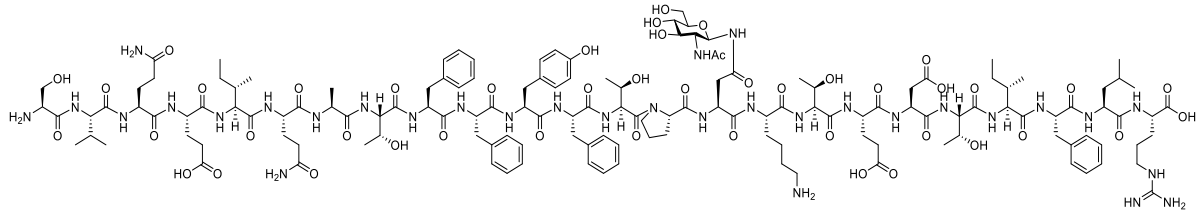


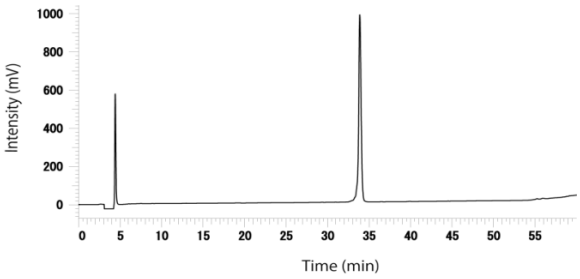


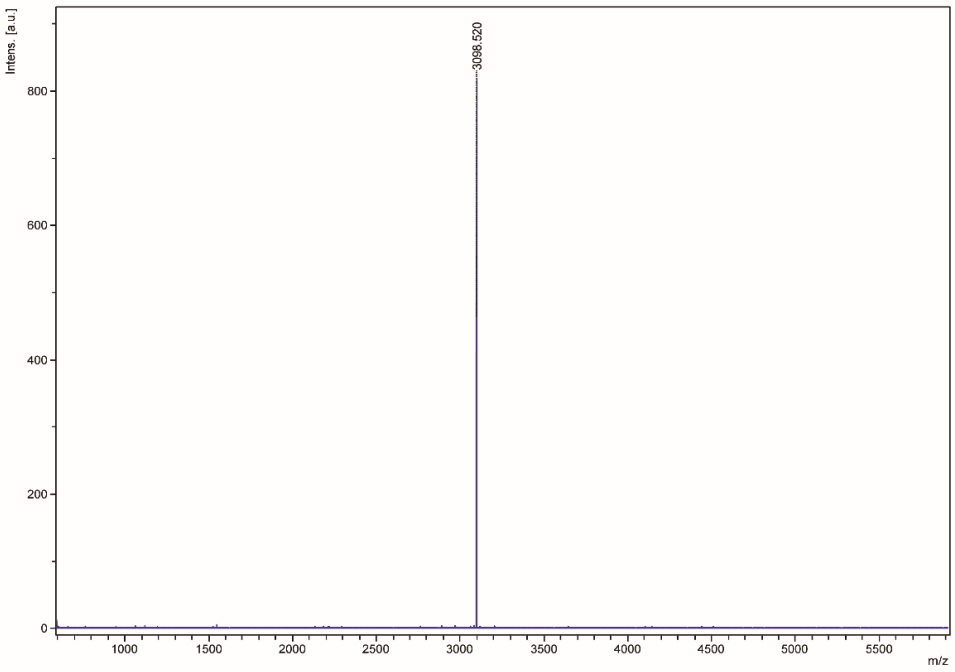
24 mer peptide fragment (3) of alpha-1-acid glycoprotein 1 (AGP1) and alpha-1-acid glycoprotein 2 (AGP2) carrying *N*-acetyl glucosamine residue on asparagine of glycosylation site Asn54. (1.6 mg, 5.37% overall yield calculated from the resin employed, 9.6 µmol)

Analytical HPLC: A:B (0.1% TFA in H_2O:0.1% TFA in CH_3CN), A/B gradient 0-50 min, 15-45%

MALDI-TOF MS: *m/z* calculated for $C_{143}H_{213}N_{32}O_{45}$, $[M+H]^+$ = 3098.536, found = 3098.520



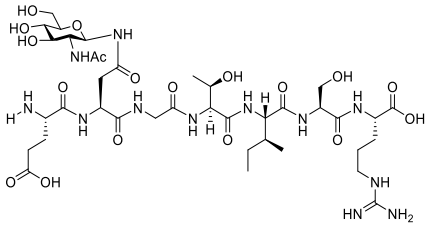


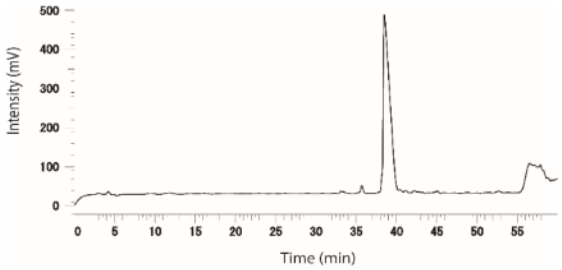


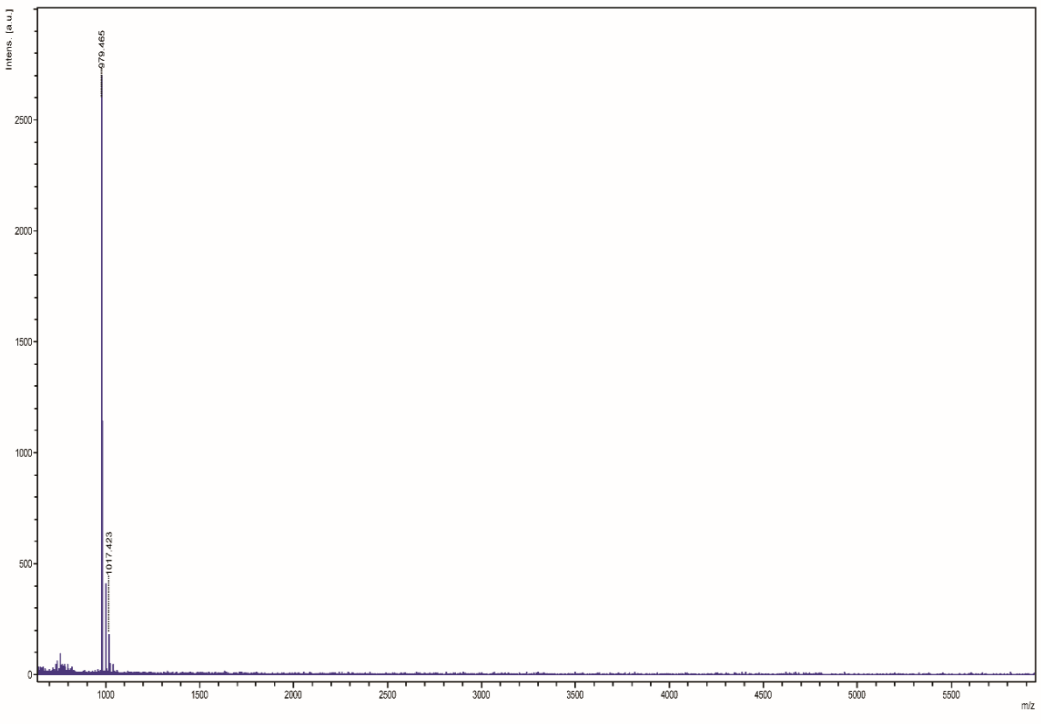
7 mer peptide fragment (20) of alpha-1-acid glycoprotein 1 (AGP1) carrying *N*-acetyl glucosamine residue on asparagine of glycosylation site Asn85. (3.8 mg, 81% overall yield calculated from the resin employed, 4.8 µmol)

Analytical HPLC: A:B (0.1% TFA in H_2O:0.1% TFA in CH_3CN), A/B gradient 0-50 min, 0-10%

MALDI-TOF MS: *m/z* calculated for $C_{38}H_{67}N_{12}O_{18}$, $[M+H]^+$ = 979.469, found = 979.465

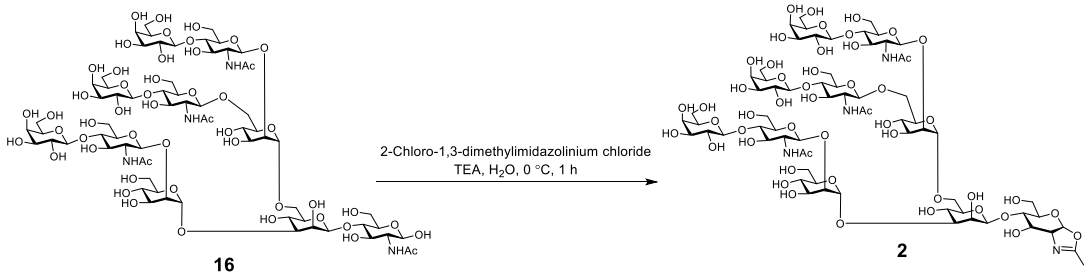






Recombinant *endo*-M (N175Q) catalyzed trans-glycosylation reaction using GlcNAc peptide fragments 3, 18 and 20 of AGP as acceptors and oxazoline derivative (2) as donor.

Scheme 4. Synthesis of decasaccharide axazoline derivative (2) from decasaccharide (16)



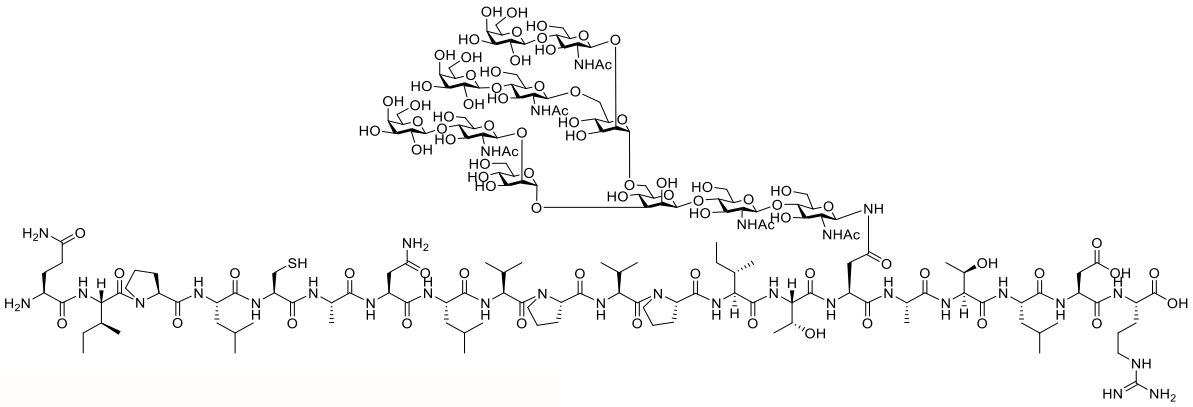
To a solution of decasaccharide **16** (7.3 mg, 3.4 μmol) in H_2O (1.0 mL) was added Et_3N (43 μL, 305 μmol) and 2-chloro-1,3-dimethylimidazolinium chloride (DMC) (22 mg, 119 μmol). The solution was stirred at 0 °C for 40 min, then the reaction mixture was purified directly by gel filtration on a Sephadex G10 column eluted with 0.5% aq. NH_3. The fractions containing the product was evaporated *in vacuo* and the residue was lyophilized to afford crude decasaccharide oxazoline **2** (approximately 7.0 mg) as a white solid. According to the general method reported previously,[16] unstable oxazoline derivative **2** was used directly for the next step without further purification.

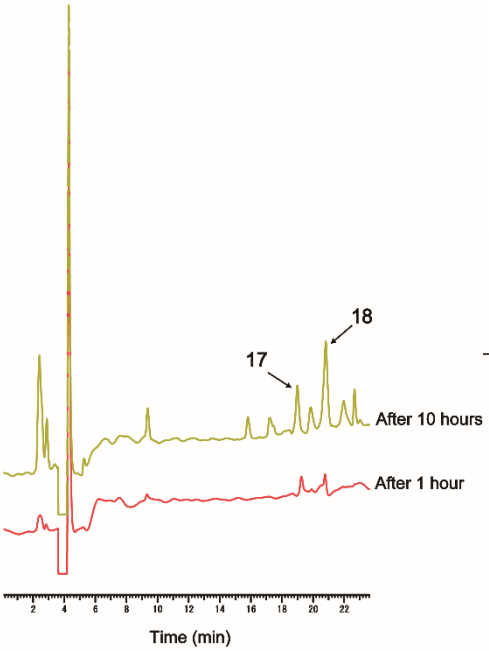
Scheme 6. Synthesis of 20 mer Glycopeptide fragment (**17**) of alpha-1-acid glycoprotein 2 (AGP2) carrying complex type tri-antennary oligosaccharide.

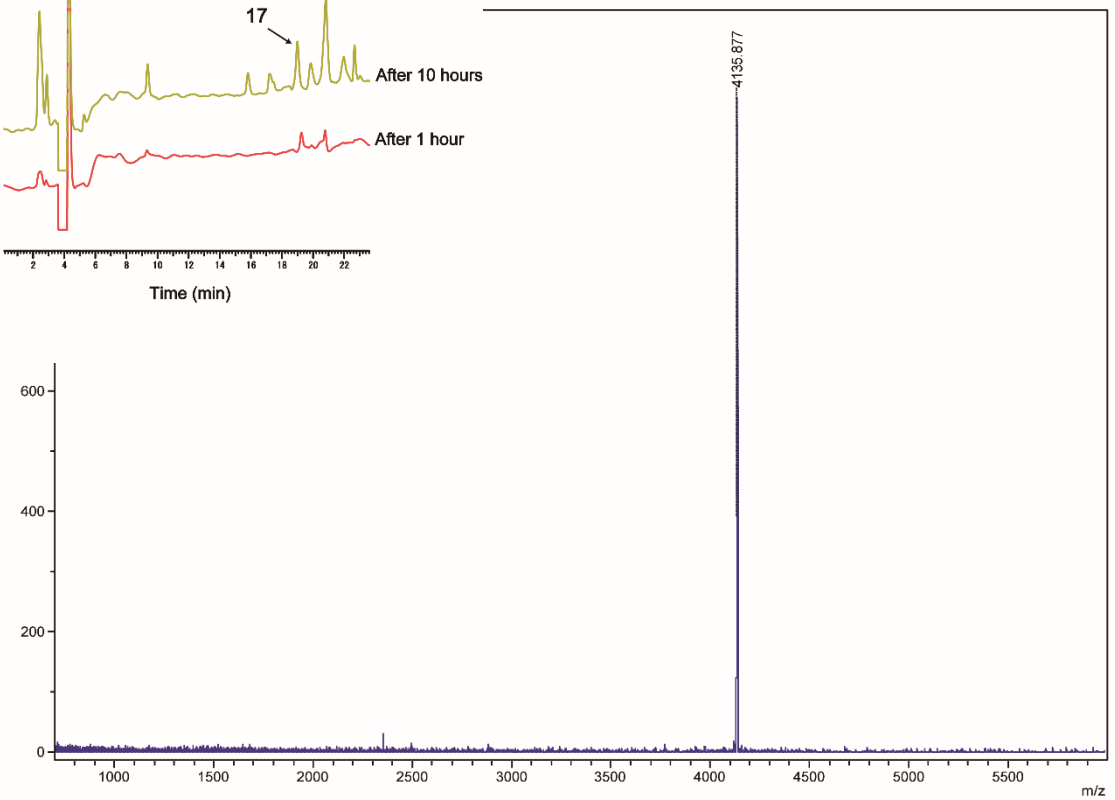
To a solution of 2 μL of 6.5 mM glycopeptide **18** in Milli Q H_2O, 5.2 μL (6 eq.) of 15mM decasaccharide oxazoline **2** was added in presence of phosphate buffer, pH = 5.0 of final concentration of 25 mM and mutant *endo*-M (N175Q) of final concentration of 50mU/mL, incubated at 30 °C for 4 h. The reaction was monitored by reverse-phase analytical HPLC.

Analytical HPLC: A:B (0.1% TFA in H_2O:0.1% TFA in CH_3CN), A/B gradient 0-50 min, 2-60%

MALDI-TOF MS: *m/z* calculated for $C_{171}H_{288}N_{31}O_{83}S$, $[M+H]^+$ = 4135.898, found = 4135.877, Yield = 46% from reverse phase analytical HPLC data.




18
17
After 10 hours
After 1 hour
2
4
6
8
10
12
14
16
18
20
22
Time (min)

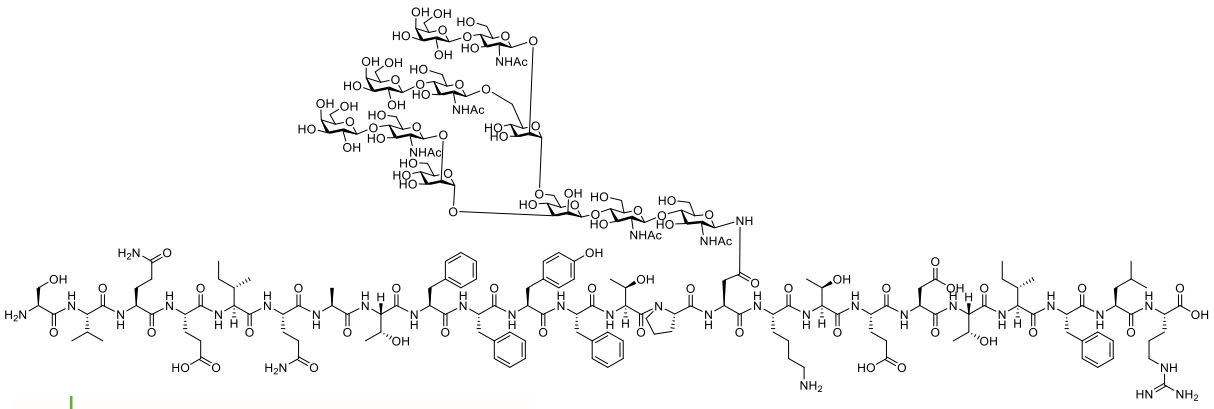

4135.877
600
400
200
0
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
m/z

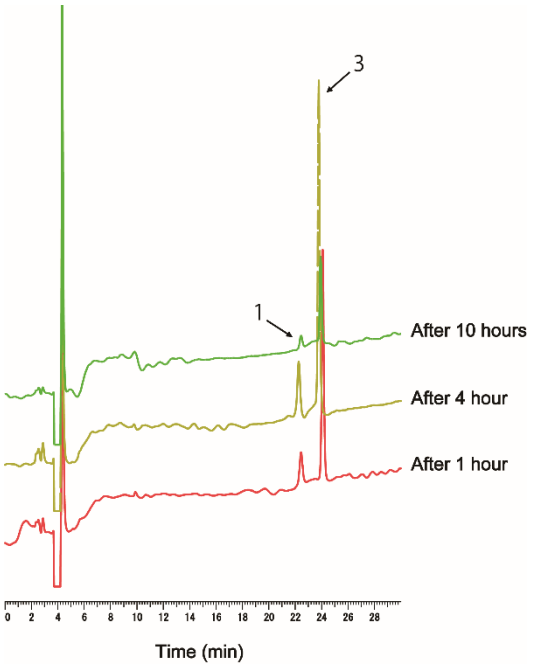
Scheme 5. Synthesis of 24 mer Glycopeptide fragment (**1**) of alpha-1-acid glycoprotein 1 (AGP1) and alpha-1-acid glycoprotein 2 (AGP2) carrying complex type tri-antennary oligosaccharide.

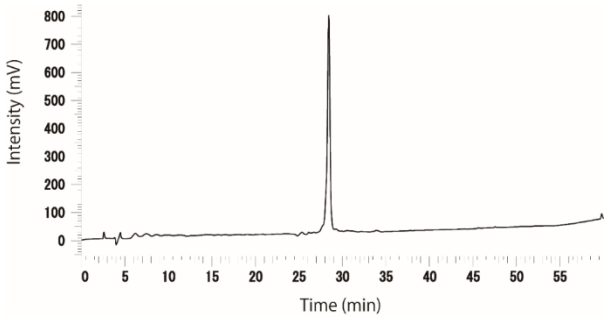
To a solution of 2 µL of 6.5 mM glycopeptide **3** in Milli Q H_2O, 2.6 µL (3 eq.) of 15mM decasaccharide oxazoline **2** was added in presence of phosphate buffer, pH = 6.0 of final concentration of 25 mM and mutant *endo*-M (N175Q) of final concentration of 50mU/mL, incubated at 30 °C for 10 h. The reaction was monitored by reverse-phase analytical HPLC.

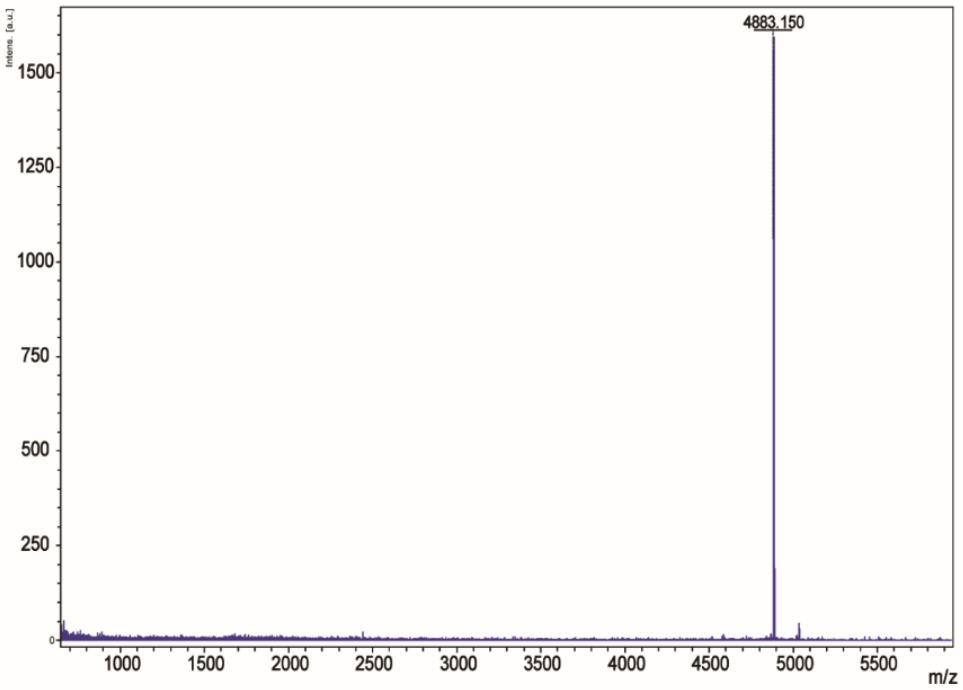
Analytical HPLC: A:B (0.1% TFA in H_2O:0.1% TFA in CH_3CN), A/B gradient 0-50 min, 2-60%

MALDI-TOF MS: *m/z* calculated for $C_{211}H_{325}N_{36}O_{95}$, $[M+H]^+$ = 4883.170, found = 4883.150, Yield = 20% from reverse phase analytical HPLC data.







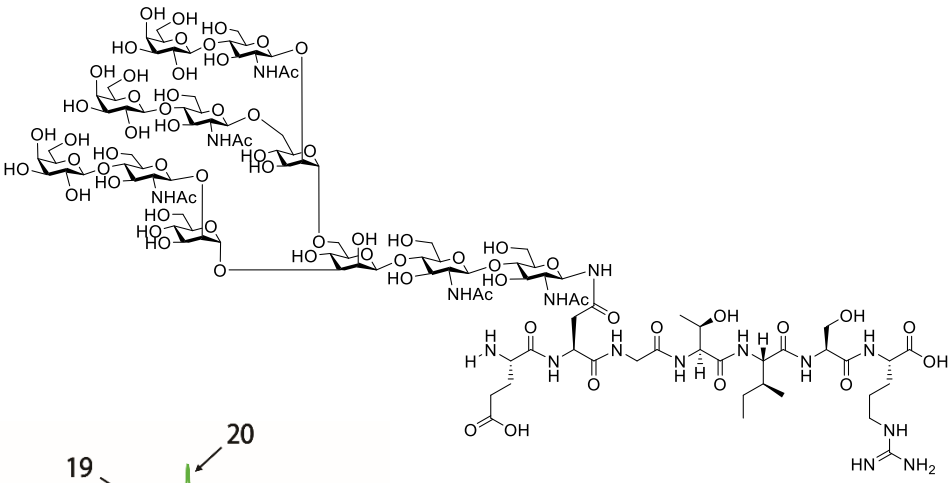


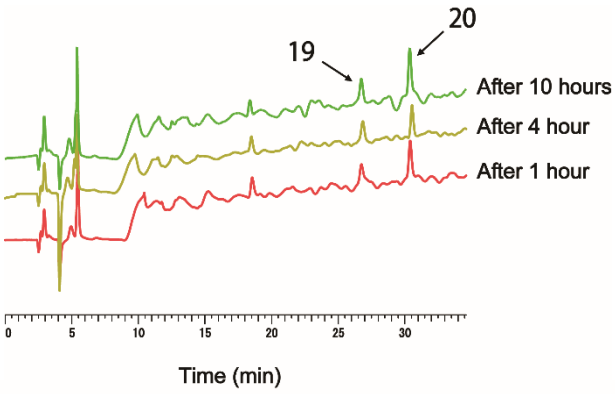
Scheme 7. Synthesis of 7 mer Glycopeptide fragment (**19**) of alpha-1-acid glycoprotein 1 (AGP1) carrying complex type tri-antennary oligosaccharide.

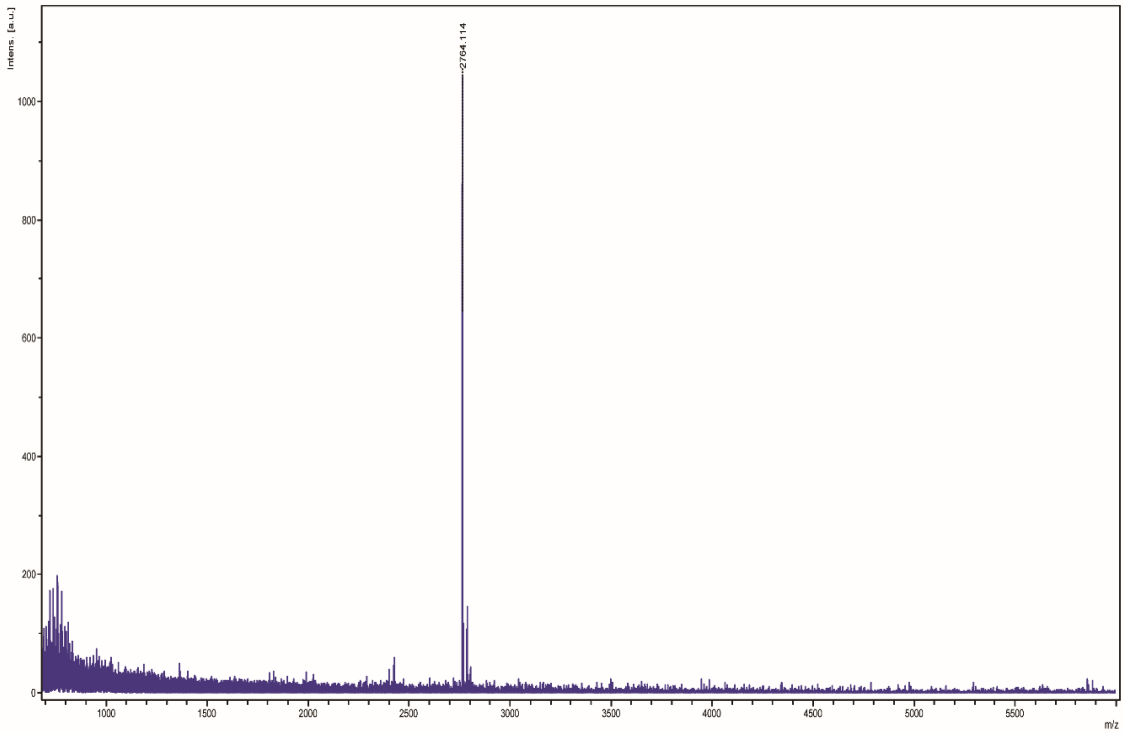
To a solution of 2 µL of 6.5 mM glycopeptide **20** in Milli Q H_2O, 5.2 µL (6 eq.) of 15mM decasaccharide oxazoline **2** was added in presence of phosphate buffer, pH = 5.0 of final concentration of 25 mM and mutant *endo*-M (N175Q) of final concentration of 50mU/mL, incubated at 30 °C for 4 h. The reaction was monitored by reverse-phase analytical HPLC.

Analytical HPLC: A:B (0.1% TFA in H_2O:0.1% TFA in CH_3CN), A/B gradient 0-50 min, 0-16%

MALDI-TOF MS: *m/z* calculated for $C_{106}H_{179}N_{16}O_{68}$, $[M+H]^+$ = 2764.104, found = 2764.114, Yield = 44% from reverse phase analytical HPLC data.




20
19
After 10 hours
After 4 hour
After 1 hour
0
5
10
15
20
25
30
Time (min)

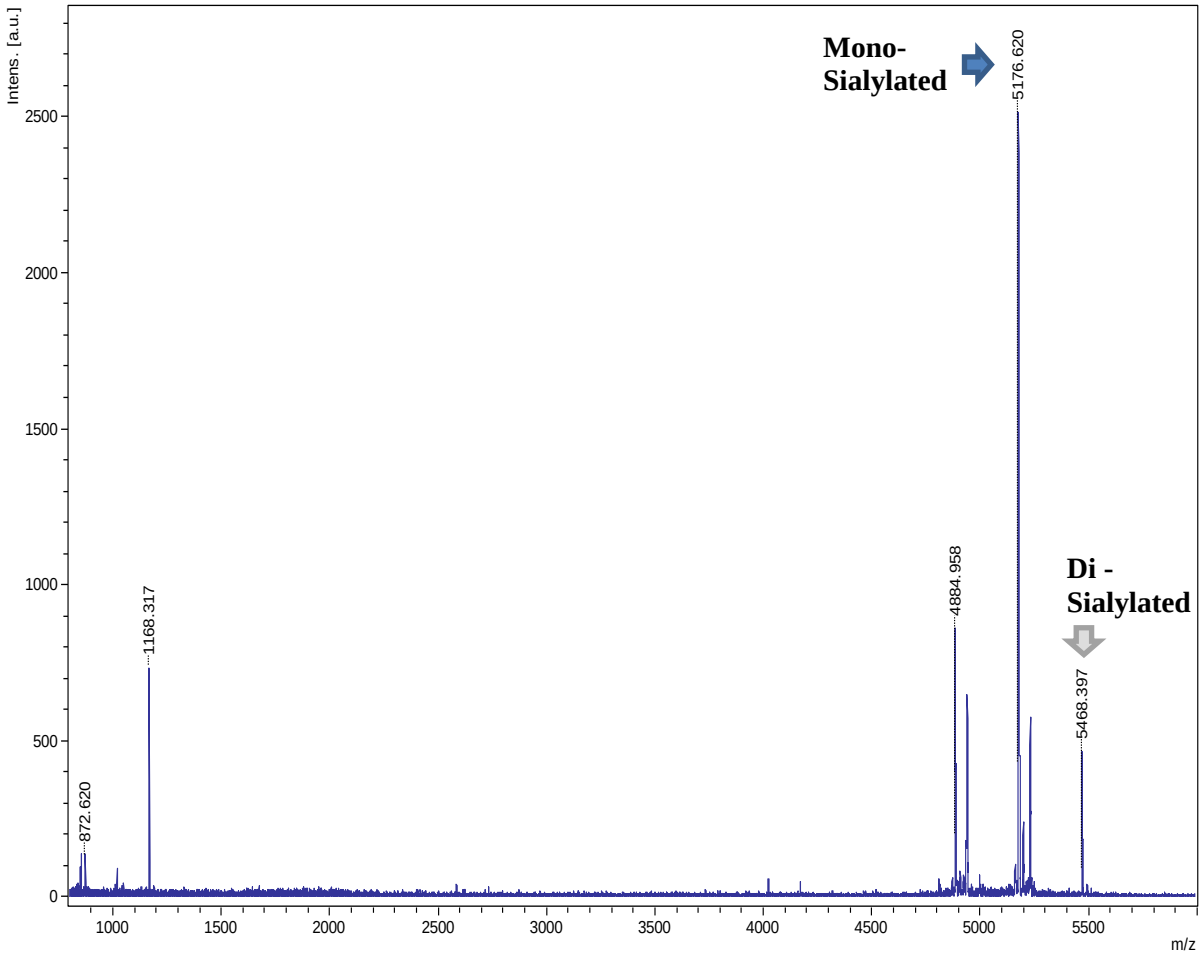

Intens. [a.u.]
2764.114
1000
800
600
400
200
0
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
m/z

Synthesis of sialyl glycopeptides of AGP

Scheme 8. Synthesis of 24 mer sialyl glycopeptide fragments of alpha-1-acid glycoprotein 1/2 carrying complex type tri-antennary oligosaccharide.

To a solution of 1 μL of 4 mM glycopeptide **1** in milli Q H_2O, 7 μL (106 eq.) of 100 mM of CMP-NANA was added in presence of final concentration, 91 mM of TRIS-HCl buffer (pH = 7.5), 0.0083% of triton X-100, and 129 mM of α-2,6-sialyltransferase from photobacterium damsela, incubated at r.t. for 45 h. The reaction was monitored by MALDI-TOF MS.

MALDI-TOF MS: *m/z* calculated for mono-sialylated glycopeptide, $C_{222}H_{342}N_{37}O_{103}$, $[M+H]^+$ = 5174.266, found = 5176.620; di-sialylated glycopeptide, $C_{233}H_{359}N_{38}O_{111}$, $[M+H]^+$ = 5465.361, found = 5468.397.



2.4. References

2.4. References

1. Nakano, M.; Kakehi, K.; Tsai, M.-H.; Yuan, C. L. Detailed structural features of glycan chains derived from α1-acid glycoproteins of several different animals: the presence of hypersialylated, O-acetylated sialic acids but not disialyl residues. *Glycobiology* **2004,** 14, 431-441.
2. Thierry, F.; Najet, M.-N.; Dominique, P. Alpha-1-acid glycoprotein. *Biochimica et Biophysica Acta* **2000,** 1482, 157-171.
3. Schmid, K.; Burgi, W.; Collins, J. H.; Nanno, S. The Disulfide Bonds of α_1-Acid Glycoprotein. Biochemistry 1974, 13, 2694-2696.
4. Schmid, K.; Kaufmann, H.; Isemura, S.; Bauer, F.; Emura, J.; Motoyama, T.; Ishiguro, M.; Nanno, S. Structure of α_1-acid glycoprotein. Complete amino acid sequence, multiple amino acid substitutions, and homology with the immunoglobulins. *Biochemistry,* **1973,** 12, 2711-2724.
5. Dente, L.; Ciliberto, G.; Cortese, R. Structure of the human $\alpha_1$acid glycoprotein gene: sequence homology with other human acute phase protein genes. *Nucleic Acids Research* **1985,** 13, 3941–3952.
6. Duche, J.-C.; Saik, U.; Nicolas, S.; Emmanuelle, M.; Isabelle, M.; Jerome, B. Expression of the genetic variants of human alpha-1-acid glycoprotein in cancer. *Clinical Biochemistry,* **2000,** 33, 197-202.
7. Kopecky, V. Jr.; Ettrich, R.; Hofbauerova, K.; Baumruk, V. Structure of human α1-acid glycoprotein and its high-affinity binding site. *Biochemical and Biophysical Research Communications* **2003,** 300, 41-46.
8. Huang, R.Y.-C.; Hudgens, J. W. Effects of Desialylation on Human α1-Acid Glycoprotein–Ligand Interactions. *Biochemistry* **2013,** 52, 7127-7136.
9. Luciana, D.; Maria, G. P.; Andres, M.; Riccardo, C. *The EMBO Journal* **1987,** 6, 2289-2296.
10. Ju, Y. L.; Hyun, K. L.; Gun, W. P.; Heeyoun, H.; Hoi, K. J.; Ki, N. Y.; Eun, S. J.; Kwang, H. K.; Jun, S. K.; Jong, W. K.; Sung, H. Y.; Choi, C.-W.; Seung, I. K.; Lim, J.-S.; Jeong, S.-K.; Paik, Y.-K.; Lee, S.-Y.; Park, J.; Kim, S. Y.; Choi, Y.-J.; Kim, Y.-I.; Jawon, S.; Cho, J.-Y.; Myoung, J. O.; Nari, S.; Hyun, J. A.; Jin, Y. K.; Jong, S. Y. Characterization of Site-Specific *N*-Glycopeptide

isoforms of α-1-acid glycoprotein from an Interlaboratory Study Using LC-MS/MS. *J. Proteome Res.* **2016,** 15, 4146-4164.

11. Michael, J. T.; Catherine, E. C.; Halsall, H. B. Analysis of the five glycosylation sites of human alpha 1-acid glycoprotein. *Biochem. J.* **1992,** 283, 105-112.
12. (a) Astrom, E.; Stal, P.; Zenlander, R.; Edenvik, P.; Alexandersson, C.; Haglund, M.; Ryden, I.; Pahlsson, P. Reverse lectin ELISA for detecting fucosylated forms of α-1-acid glycoprotein associated with hepatocellular carcinoma. *PLOS ONE* **2017,** 12, 1-15. (b) Tanabe, K.; Kitagawa, K.; Kojima, N.; Iijima, S. Multifucosylated Alpha-1-acid glycoprotein as a Novel Marker for Hepatocellular Carcinoma. *J. Proteome Res.* **2016,** 15, 2935-2944. (c) Balmana, M.; Gimenez, E.; Puerta, A.; Llop, E.; Figueras, J.; Fort, E.; Victoria, S.-N.; Bolos, C. D.; Rizzi, A.; Barrabes, S.; Frutos, M. D.; Peracaula, R. Increased α1-3 fucosylation of α-1-acid glycoprotein (AGP) in pancreatic cancer. *Journal of Proteomics* **2016,** 132, 144-154. (d) Croce, M. V.; Salice, V. C.; Lacunza, E.; Segal-Eiras, A. α_1-acid glycoprotein (AGP): a possible carrier of Sialyl lewis X (slewis X) antigen in colorectal carcinoma. *Histol. Histopathol.* **2005,** 20, 91-97.
13. (a) Ongay, S.; Martin-Alvarez, P. J.; Neusub, C.; Frutos, M. D. Statistical evaluation of CZE-UV and CZE-ESI-MS data of intact α-1-acid glycoprotein isoforms for their use as potential biomarkers in bladder cancer. *Electrophoresis* **2010,** 31, 3314-3325. (b) Paltrinieri, S.; Gelain, M. E.; Ceciliani, F.; Ribera, A. M.; Battilani, M. SHORT COMMUNICATION Association between faecal shedding of feline coronavirus and serum α1-acid glycoprotein sialylation. *Journal of Feline Medicine and Surgery* **2008,** 10, 514-518. (c) Magdalena, O.-P.; Hirnle, L.; Iwona, K.-P. Alternation of *N*-glycan branching and expression of sialic acid on amniotic fluid alpha-1-acid glycoprotein derived from second and third trimesters of normal and prolonged pregnancies. *Clinica Chimica Acta* **2006,** 367, 86-92. (d) Ceciliani, F.; Grossi, C.; Giordano, A.; Pocacqua, V.; Paltrinieri, S. Decreased sialylation of the acute phase protein α1-acid glycoprotein in feline infectious peritonitis (FIP). *Veterinary Immunology and Immunopathology* **2004,** 99, 229-236. (e) Kremmer, T.; Szollosi, E.; Boldizsar, M.; Vincze, B.; Ludanyi, K.; Imre, T.; Schlosser, G.; Vekey, K. Liquid chromatographic and mass spectrometric analysis of human serum acid alpha-1-glycoprotein. *Biomedical Chromatography* **2004,** 18, 323-329.

14. Anderson, N. L. The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clin. Chem*. **2010**, *56*, 177-185.

15. (a) Luo, Z.; Lei, H.; Sun, Y.; Liu, X.; Su, D.-F. Orosomucoid, an acute response protein with multiple modulating activities. *J. Physiol. Biochem*. **2015**, *71*, 329-340. (b) Hashimoto, S.; Asao, T.; Takahashi, J.; Yagihashi, Y.; Nishimura, T.; Saniabadi, A. R.; Poland, D. C. W.; van Dijk, W.; Kuwano, H.; Kochibe, N.; Yazawa, S. α1-Acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis. *Cancer* **2004**, *101*, 2825-2836. (c) Tanabe, K.; Kitagawa, K.; Kojima, N.; Iijima, S. Multifucosylated alpha-1-acid glycoprotein as a novel marker for hepatocellular carcinoma. *J. Proteomic Res*. **2016,** *15*, 2935-2944.

16. Umekawa, M.; Li, C.; Higashiyama, T.; Huang, W.; Ashida, H.; Yamamoto, K.; Wang, L-X. Efficient glycosynthase mutant derived from *Mucor hiemalis* endo-β *N*-acetylglucosaminidase capable of transferring oligosaccharide from both sugar oxazoline and natural *N*-glycan. *J. Biol. Chem*. **2010**, *285*, 511-521.

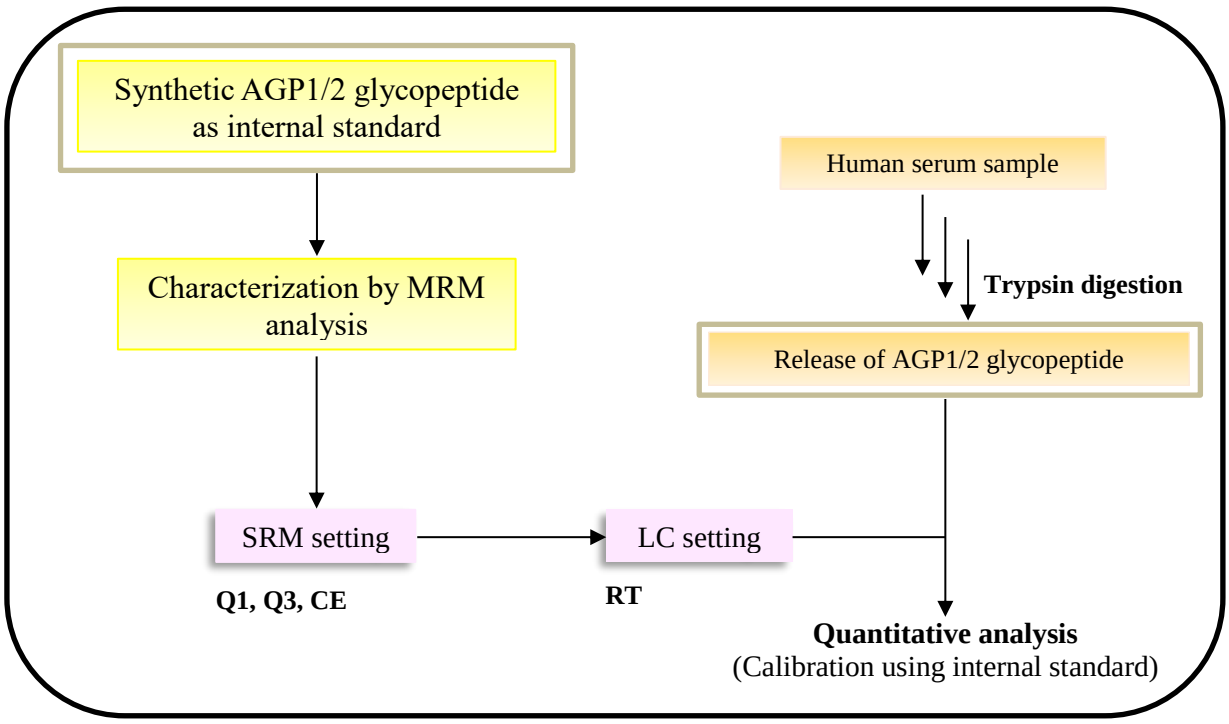
17. Ravi, K. H. V.; Naruchi, K.; Miyoshi, R.; Hinou, H.; Nishimura, S. -I. A new approach for the synthesis of hyperbranched N-glycan core structures from locust bean gum. *Org. Lett*. **2013**, *15*, 6278-6281.

18. (a) Noguchi, M.; Tanaka, T.; Gyakushi, H.; Kobayashi, A.; Shoda, S. -I. Efficient synthesis of sugar oxazolines from unprotected *N*-acetyl-2-amino sugars by using chloroformamidinium reagent in water. *J. Org. Chem*. **2009**, *74*, 2210-2212. (b) Umekawa, M.; Higashiyama, T.; Koga, Y.; Tanaka, T.; Noguchi, M.; Kobayashi, A.; Shoda, S.-I.; Huang, W.; Wang, L-X.; Yamamoto, K. Efficient transfer of sialooligosaccharide onto proteins by combined use of a glycosynthase-like mutant of *Mucor hiemalis* endoglycosidase and synthetic sialo-complex-type sugar oxazoline. *Biochim. Biophys. Acta* **2010**, *1800*, 1203-1209.

Chapter 3.

SRM-based targeted glycoproteomics

Strategy for SRM-based absolute quantification of AGP1/2 glycopeptide (1)



SRM/MRM LC-MS/MS analysis:

AB Sciex 4000Q TRAP® TurboIonSpray systems and Dionex HPLC were used.

Separation:

LPG-3x00 pump, WPS-3000 auto sampler, FLM-3100 column component, and WVD-3400 detector under control of Chromeleon 6.80.

Column:

Inertsil Diol 4.6x250 mm, and Inertsil ODS-3 2.1x150 mm (GL Sciences Inc.).

Data was analyzed by a series of software: Analyst 1.5 and MultiQuant 1.1.0.26.

3.1. SRM/MRM channel setting for the synthetic AGP glycopeptide

Optimization of SRM/MRM parameters for synthetic glycopeptide 1 of AGP.

Table S1. Amino acid analysis of synthetic glycopeptide **1** of AGP.

VIS1 (570nm) Result No.	**Time of elution in min**	**Name of amino acid residue**	**Peak height**	**Peak area**	**Concentration in nmol/40µL**
1	16.587	Asp	33245	1420379	0.36622
2	20.133	Thr	53760	2558219	0.64245
3	21.807	Ser	18137	953817	0.22957
4	27.100	Glu	49141	3019402	0.72700
5	39.360	Gly	5089	374586	0.09312
6	42.547	Ala	11122	853469	0.20598
7	46.027	Val	24471	776801	0.18551
8	51.693	Ile	32585	1379442	0.33574
9	53.320	Leu	16945	785332	0.19209
10	55.153	Nle	93790	4253808	1.05682
11	56.627	Tyr	18197	705397	0.17840
12	58.420	Phe	70987	2617778	0.67015
15	68.347	Lys	21701	877271	0.19496
18	71.127	His	1669	71335	0.01761
19	73.873	NH_3	200436	15110489	4.64983
20	85.120	Arginine	10151	656013	0.16964
VIS2 (440nm) Result No.	**Time of elution in min**	**Name of amino acid residue**	**Peak height**	**Peak area**	**Concentration in nmol/40µL**
5	29.473	Proline	2561	177471	0.20194
Total					4.41039

Final concentration of the stock solution of synthetic glycopeptide **1** of AGP was estimated as 35.2 nmol/mL (172µg/mL) by calculated from the above total concentration of the tested solution (4.41 nmol/40 µL).

The accurate concentration of compound **1** was defined on the basis of the results of amino acid analysis as summarized in Table S1. A solution of synthetic AGP glycopeptide **1** (800 fmol/µL) prepared in 1:1 ratio of 0.1% formic acid aq. and acetonitrile was used to analyze its mass and fragments by infusion method using 4000Q-TRAP mass spectrometry and syringe pump with a flow of 5µL/min using 1mL syringe of a diameter of 4.61 mm. The EMS (Enhanced mass) mode measurement showed *m/z* 1630.0 which is believed to be of the synthetic AGP glycopeptide **1**; ER (Enhanced Resolution) mode of measurement which is usually gives higher resolution compared to EMS mode of measurement and hence found the *m/z* 1628.500 its isotopomer found in the TIC (Total Ion Chromatogram) are 1628.5, 1628.8, 1629.2, 1629.5, 1629.8, 1630.2, 1630.5, 1630.9 shown in Figure S1. These isotopomer with a difference of almost 0.3 Da indicates the charge of the glycopeptide **1** is 3+.

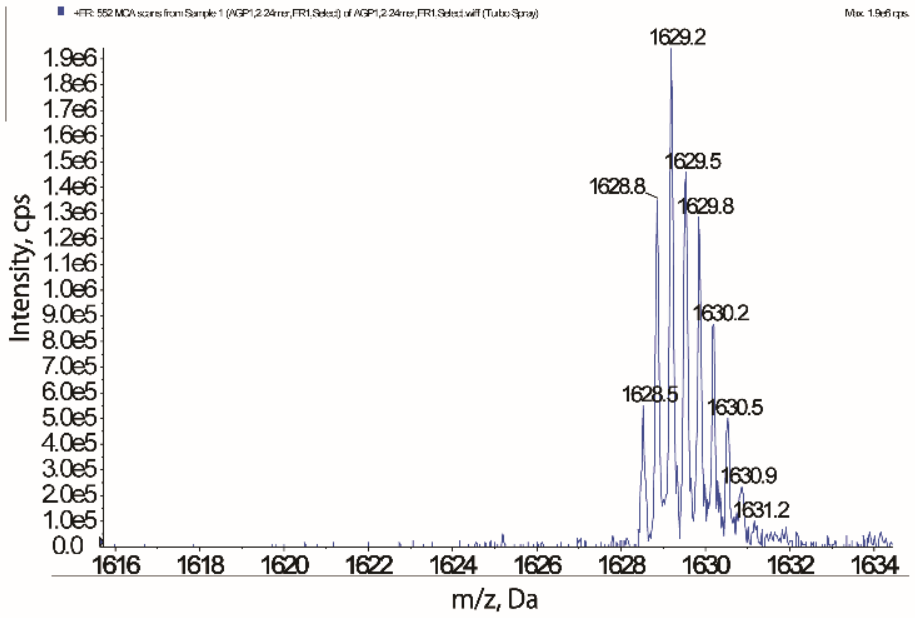


Figure S1. Mass and charge analysis of synthetic glycopeptide **1** of AGP by ER (Enhanced Resolution) mode of measurement.

Herein the highest intensity of isotopomer observed at *m/z* 1629.2 was selected as a precursor ion (Q1) of glycopeptide fragment **1** of AGP. Thus, the isotopomer of *m/z* 1629.2 was selected to analyze fragmentation by collision with inert N_2 gas at Q2 quadrupole in EPI (Enhanced Product Ion) mode of

measurement at collision energy of 80 V and found that the fragments detected are of *m/z* 366.2, 649.6, 1448.5, 1550.4, 1652.0, 1733.0, 1814.2, 1894.9, 1996.6, 2077.6 as shown in Figure S2.

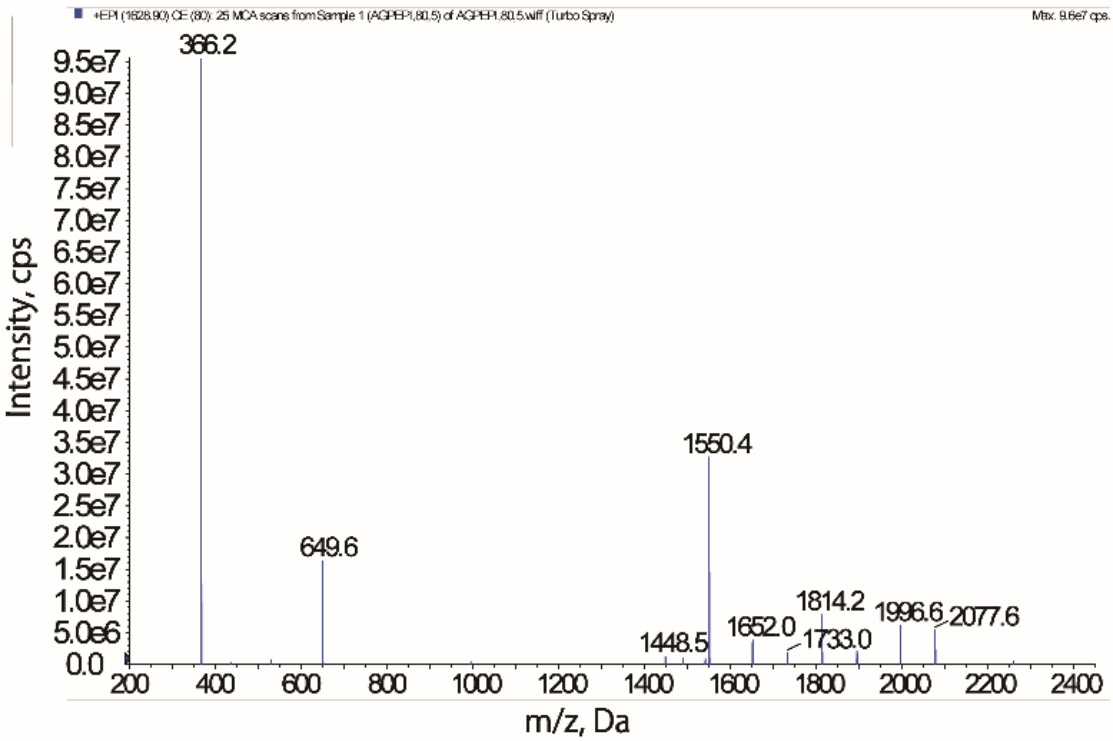


Figure S2. Analysis of fragments or m/z values of Q3 of synthetic glycopeptide **1** of AGP by EPI (Enhanced Product Ion) mode of measurement.

The selected fragments of glycopeptide **1** with *m/z* 1629.200 as Q1 for an optimization of MRM parameters were of *m/z* 1550.400, 366.200, 2077.600, 1996.600, 1894.900, 1814.200, 1733.000, 1652.000, 1448.500 and 649.600 (Figure S2). Out of which, fragments of *m/z* 366.200, 2077.600, 1996.000, 1550.400, 649.000 showed better intensity in cps with an increasing order of almost 366.200 > 2077.600 > 1996.000 > 1550.400 > 649.000 with optimized collision energy listed in Table S3.

Table S2. List of fragments or Q3 values of synthetic glycopeptide **1** of AGP from the precursor ion Q1 of m/z 1629.2 by EPI (Enhanced Product Ion) mode of measurement at collision energy 80 volts.

Precursor ion (Q1) in Da	**Fragment ion (Q3) in Da**
1629.200	366.200
1629.200	2077.600
1629.200	1996.600
1629.200	1550.400
1629.200	649.600
1629.200	1894.900
1629.200	1814.200
1629.200	1733.000
1629.200	1652.000
1629.200	1448.500

Table S3. MRM channel candidates with optimized parameters for the precursor ion Q1 of m/z 1629.200.

Precursor ion (Q1) in Da	**Fragment ion (Q3) in Da**	**Declustering Potential (DP) in Volts**	**Collision Energy (CE) in Volts**	**Collision Cell Exit Potential (CXP) in Volts**	**Entrance Potential (EP) in Volts**
1629.200	366.200	186.0	57.0	8.0	10.0
1629.200	2077.600	186.0	57.0	54.0	10.0
1629.200	1996.600	186.0	63.0	50.0	10.0
1629.200	1550.400	186.0	65.0	40.0	10.0
1629.200	649.600	186.0	111.0	16.0	10.0

Manual CE (collision energy) optimization was performed for a few parameters by step of 1 CE difference ramping. Finally, SRM/MRM transition of 1629.2/366.2 was obtained as an optimized channel for the quantitation of AGP glycopeptide **1**. DP (Declustering potential), CXP (collision cell exit potential), and EP (entrance potential).

3.2. Targeted glycoproteomics by SRM assay focusing human AGP glycopeptide (1)

Tryptic digestion of whole serum glycoproteins to release glycopeptide fragments.

Ten micro liters (10 μL) of human serum was added to 0.33 M ammonium bicarbonate (15 μL) and Milli Q H2O (30 μL) and the solution was incubated at 60 °C for 10 min. To the solution was added 120 mM dithiothreitol (DTT) (5 μL) and incubated at 60 °C for 30 min, subsequently added 123 mM iodoacetamide (IAA) (10 μL) and incubated for 1 h in a dark condition at room temp. Then, the mixture was added 5 μL of trypsin (40 U/μL) in 1 mM HCl and incubated at 37 °C for 16 h. The reaction mixture was heated at 90 °C for 10 min to inactivate an enzyme. Finally, the mixture was dried in a SpeedVac concentrator.

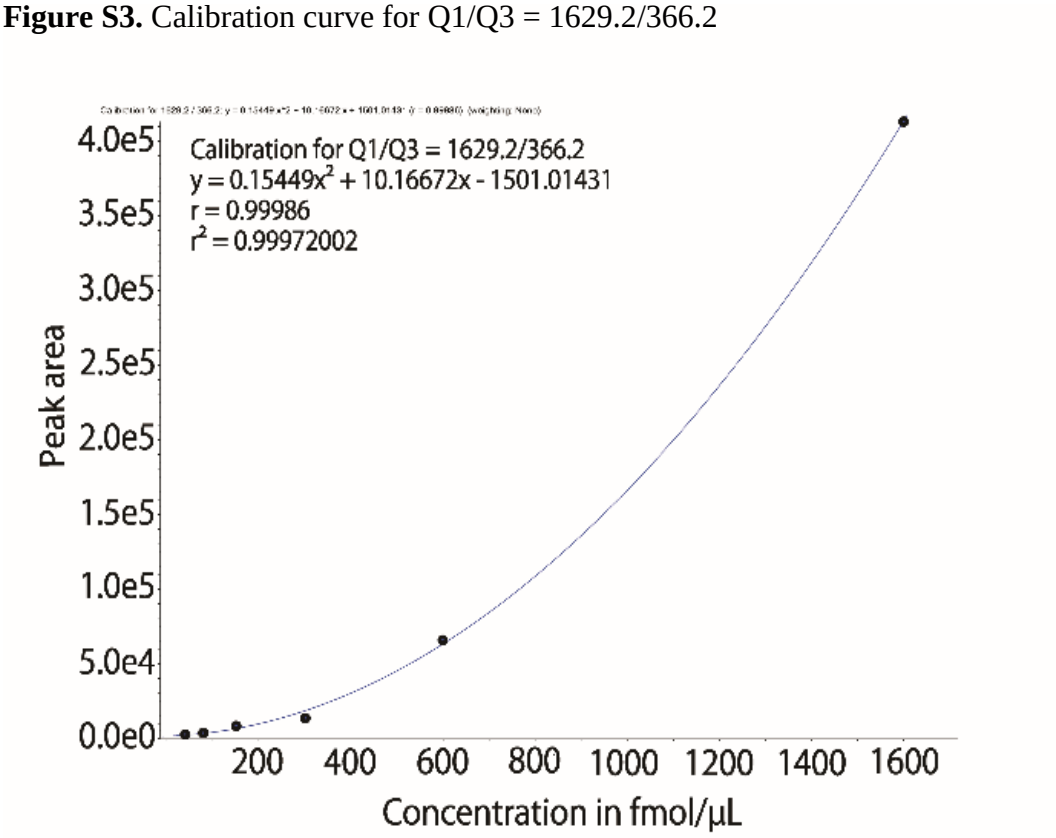
SRM/MRM-based LC-MS/MS quantification of serum AGP glycopeptide fragment (1).

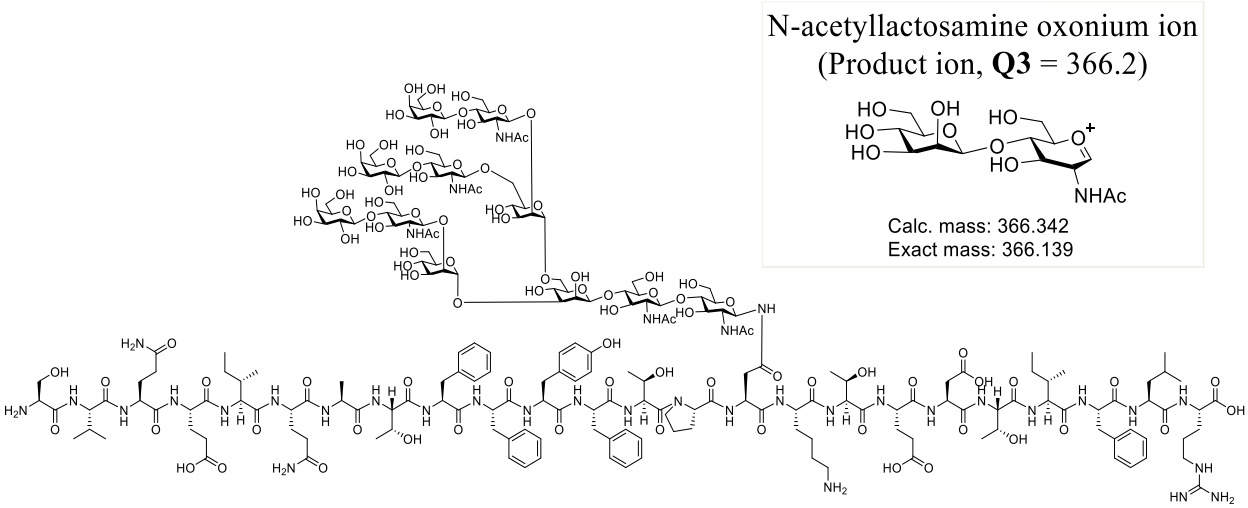
LC-MS/MS analysis was conducted with Dionex UltimateTM 3000 HPLC and AB Sciex 4000Q Trap®

TurboIonSpray system. Separation was performed with LPG-3×00 pump, WPS-3000 auto sampler, FLM-3100 column component and WVD-3400 detector under the control of software: Chromeleon 6.80. Column: Inertsil Diol 4.6×250 mm and Inertsil ODS-3 2.1×150 mm (GL Sciences Inc.). Acquired data was analyzed by a series of software: Analyst 1.5 and MultiQuant 1.1.0.26. LC condition was optimized for the SRM/MRM channel of Q1/Q3 (1629.2/366.2) for the synthetic AGP glycopeptide fragment **1** in advance with a concentration of 1000 fmole/μL: flow of 200 μL/min; multi-step gradient; (A):(B) = 0.1% formic acid in aqueous solution: 0.1% formic acid in acetonitrile; (A)/(B), 0 min: 97/3→18 min: 72/28→19 min: 10/90→23 min: 10/90→23.1 min: 97/3→30 min: 97/3; column temperature: 60 °C; inject volume: 1.5 μL, respectively. As a result, Q1/Q3 = 1629.2/366.2 gave satisfactory S/N spectra at elution time of 17.35 ± 0.05 min. A series of samples of the concentrations of 40, 80, 150, 300, 600 and 1600 fmole/μL was prepared from above mentioned stock solution of 35.2 nmole/mL (172μg/mL) of the synthetic AGP glycopeptide **1** as internal standards with 0.1% formic acid in Milli Q H2O. Pretreated human serum samples were dissolved in 30 μL of 0.1% formic acid in Milli Q H_2O, separately. Finally, 1.5 μL of all standard samples of the

synthetic AGP glycopeptide **1** and 10 µL of all pretreated human serum samples of healthy, HCC and RCC patients were submitted to the SRM-based LC-MS/MS quantification and the total ion chromatogram (TIC) of acquired quantification data were shown in Figure S4-S7.

Figure S3. Calibration curve for Q1/Q3 = 1629.2/366.2





^{40}Ser-Val-Gln-Glu-Ile-Gln-Ala-Thr-Phe-Phe-Tyr-Phe-Thr-Pro-Asn-Lys-Thr-Glu-Asp-Thr-Ile-Phe-Leu-Arg^{63}

^{40}SVQEIQATFFYFTP*N*KTEDTIFLR^{63}

1, (Precursor ion, **Q1** = 1629.2)
Synthetic glycopeptide of AGP (alpha-1-acid glycoprotein)

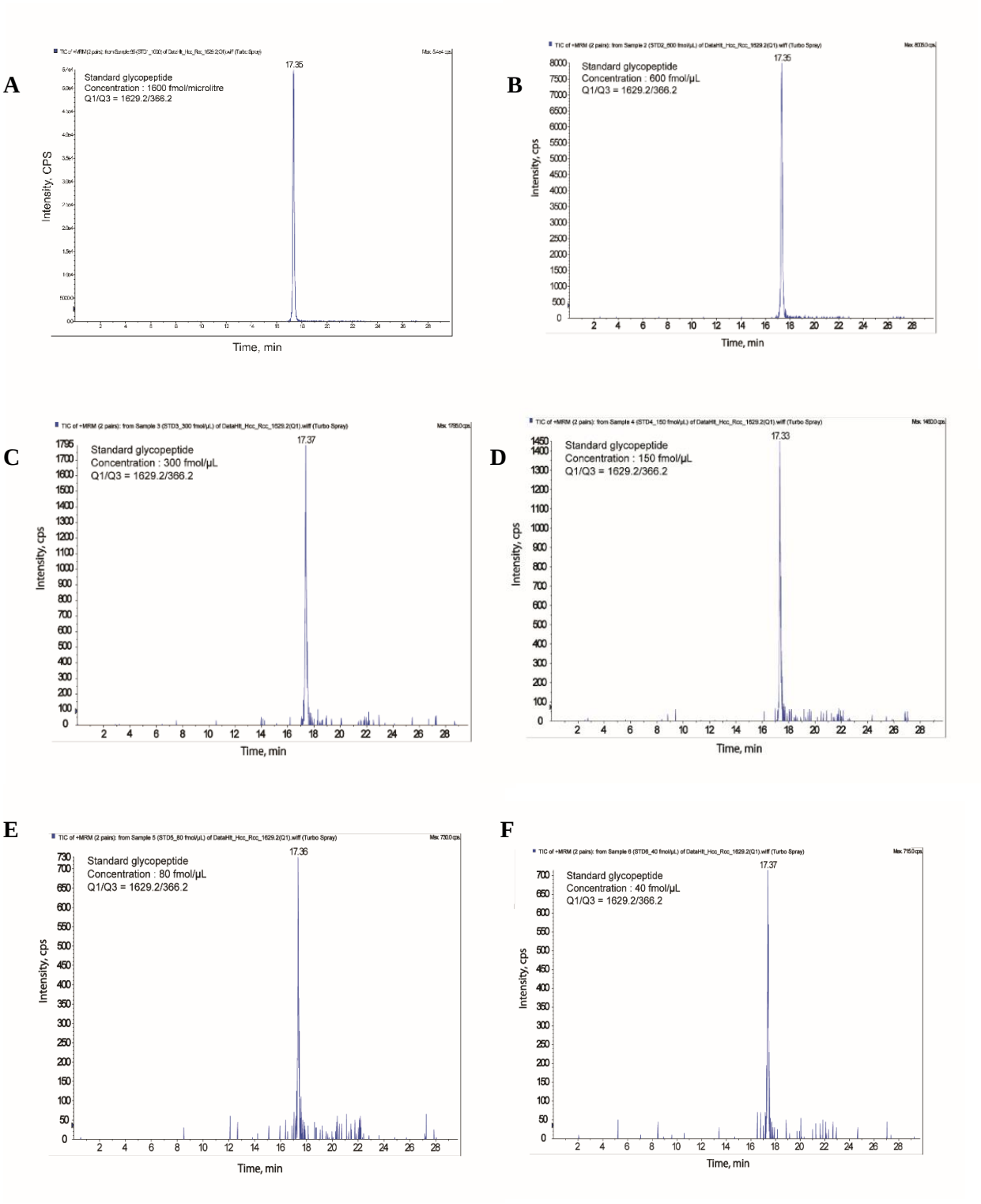


Figure S4. TIC of synthetic glycopeptide **1** of AGP at 6 different concentrations for the calibration curve: (A) 1600 fmole/µL, (B) 600 fmole/µL, (C) 300 fmole/µL, (D) 150 fmole/µL, (E) 80 fmole/µL, and (F) 40 fmole/µL, respectively.

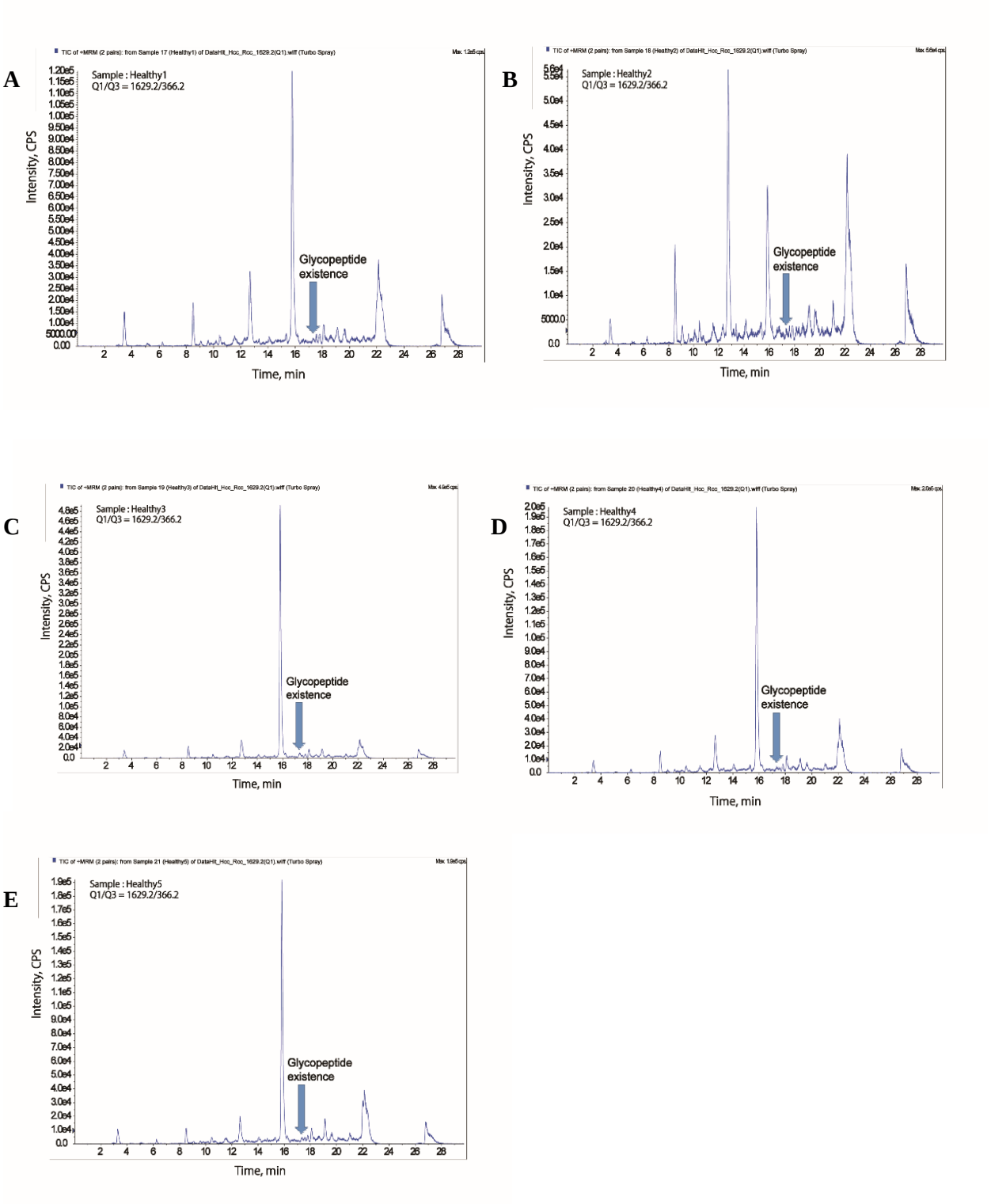


Figure S5. TIC of synthetic glycopeptide **1** derived from sera of 5 different healthy controls A~E, respectively.

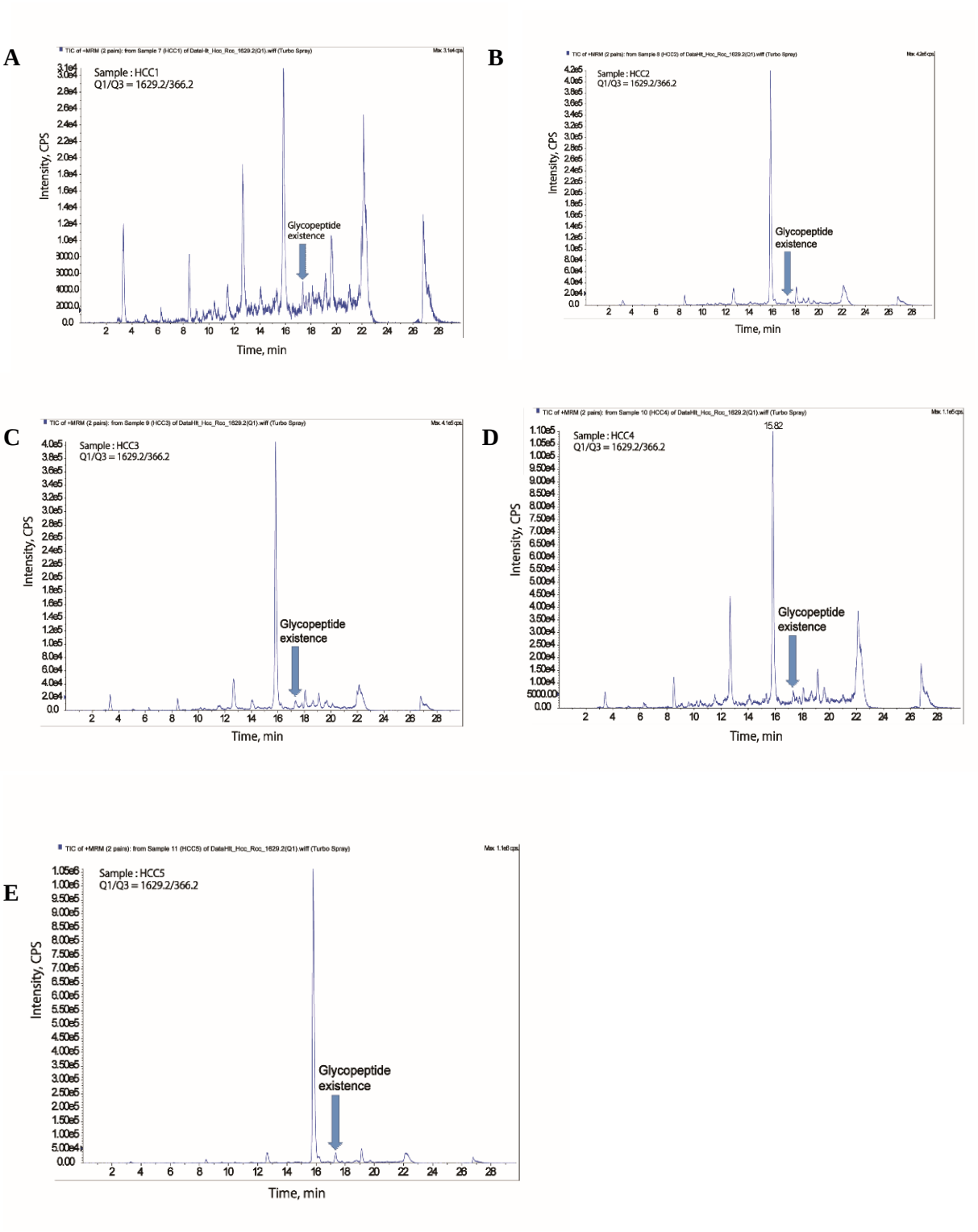


Figure S6. TIC of synthetic glycopeptide **1** derived from sera of 5 different HCC patients A~E, respectively.

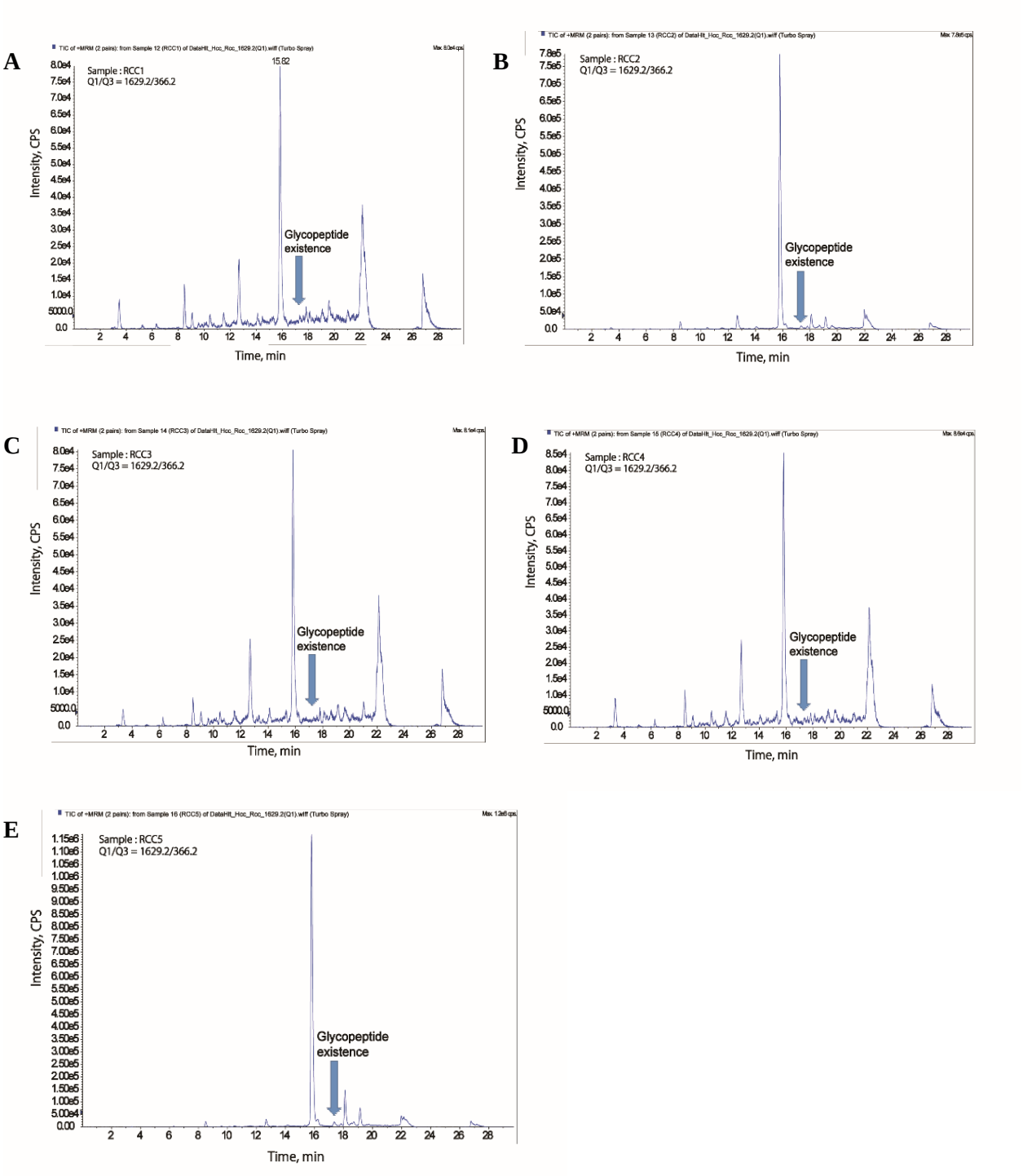


Figure S7. TIC of synthetic glycopeptide **1** derived from sera of 5 different RCC patients A~E, respectively.

Table S4. Results table.

Entry	Sample Type	Component Name (Q1/Q3)	Actual Concentration	Peak Area	Peak Height	Retention Time	Calculated Concentration	Accuracy
1	Standard1	1629.2 / 366.2	1600	4.13E+05	4.99E+04	17.34	1.60E+03	99.98
2	Standard2	1629.2 / 366.2	600	6.58E+04	7.14E+03	17.33	6.13E+02	102.19
3	Standard3	1629.2 / 366.2	300	1.33E+04	1.66E+03	17.37	2.45E+02	81.65
4	Standard4	1629.2 / 366.2	150	8.21E+03	1.14E+03	17.35	1.78E+02	118.73
5	Standard5	1629.2 / 366.2	80	3.84E+03	6.17E+02	17.36	9.44E+01	118
6	Standard6	1629.2 / 366.2	40	2.71E+03	5.81E+02	17.37	6.16E+01	153.97
7	HCC1	1629.2 / 366.2	N/A	2.70E+04	3.49E+03	17.37	3.75E+02	N/A
8	HCC2	1629.2 / 366.2	N/A	1.01E+05	9.11E+03	17.34	7.71E+02	N/A
9	HCC3	1629.2 / 366.2	N/A	1.35E+05	1.09E+04	17.34	8.96E+02	N/A
10	HCC4	1629.2 / 366.2	N/A	4.00E+04	4.78E+03	17.34	4.67E+02	N/A
11	HCC5	1629.2 / 366.2	N/A	3.54E+05	3.53E+04	17.35	1.48E+03	N/A
12	RCC1	1629.2 / 366.2	N/A	5.57E+03	1.82E+03	17.32	1.33E+02	N/A
13	RCC2	1629.2 / 366.2	N/A	7.87E+04	6.96E+03	17.36	6.75E+02	N/A
14	RCC3	1629.2 / 366.2	N/A	2.79E+03	1.30E+03	17.36	6.42E+01	N/A
15	RCC4	1629.2 / 366.2	N/A	4.53E+03	1.97E+03	17.34	1.11E+02	N/A
16	RCC5	1629.2 / 366.2	N/A	1.71E+05	1.61E+04	17.36	1.02E+03	N/A
17	Healthy1	1629.2 / 366.2	N/A	1.42E+04	1.93E+03	17.34	2.56E+02	N/A
18	Healthy2	1629.2 / 366.2	N/A	1.13E+04	2.07E+03	17.31	2.21E+02	N/A
19	Healthy3	1629.2 / 366.2	N/A	5.48E+04	8.22E+03	17.34	5.55E+02	N/A
20	Healthy4	1629.2 / 366.2	N/A	1.72E+04	3.06E+03	17.34	2.88E+02	N/A
21	Healthy5	1629.2 / 366.2	N/A	2.27E+04	2.75E+03	17.35	3.39E+02	N/A

Table S5. Summary of results analysis.

Entry	Status of Sera	Sex	Age	TNM Classification	Stage	TIC peak area at 17.35 min	Serum level of AGP glycopeptide 1 (fmol/µL)
1	Healthy1	M	50's	-	-	1.42×10^4	**255.8**
2	Healthy2	M	50's	-	-	1.13×10^4	**221.0**
3	Healthy3	M	50's	-	-	5.48×10^4	**555.3**
4	Healthy4	M	50's	-	-	1.72×10^4	**288.0**
5	Healthy5	M	50's	-	-	2.27×10^4	**338.9**
6	HCC1	M	50's	T4 N0 M0	IIIC	2.70×10^4	**374.9**
7	HCC2	M	40's	T4 N1 M1	IVB	1.01×10^5	**771.0**
8	HCC3	M	50's	T4 N0 M0	IIIC	1.35×10^5	**895.7**
9	HCC4	M	60's	T2 N0 M0	II	4.00×10^4	**467.3**
10	HCC5	M	50's	T1 N0 M0	I	3.54×10^5	**1479.0**
11	RCC1	M	50's	T1 N0 M0	I	5.57×10^3	**132.6**
12	RCC2	M	50's	T3b N0 M1	IV	7.87×10^4	**674.8**
13	RCC3	M	50's	T1b N0 M0	I	2.79×10^3	**64.2**
14	RCC4	M	60's	T1a N0 M0	I	4.53×10^3	**110.9**
15	RCC5	M	60's	T3b N0 M1	IV	1.71×10^5	**1016.0**

10 µL of sera were used for the tryptic digestion. As for a general protocol for the pretreatment of serum samples**.**
HCC: Hepatocellular carcinoma; **RCC:** Renalcell carcinoma; **M:** Male (Gender).

T1: Solitary tumor without vascular invasion; **T2**: Solitary tumor with vascular invasion or multiple tumors, none > 5 cm; **T4**: Tumor (s) with direct invasion of adjacent organs other than gallbladder or with visceral peritoneum; **N0**: No regional lymph node metastasis; **N1**: Regional lymph node metastasis; **M0**: No distant metastasis; **M1**: Distant metastasis

T1: Tumor ≤ 7 cm in greatest dimension, limited to the kidney; **T1a**: Tumor ≤ 4 cm in greatest dimension, limited to the kidney; **T1b**: Tumor > 4 cm but ≤ 7 cm in greatest dimension, limited to the kidney; **T3b**: Tumor grossly extends into the vena cava below the diaphragm; **N0**: No regional lymph node metastasis; **M0**: No distant metastasis; **M1**: Distant metastasis

Chapter 4

Concluding remarks

Conclusion:

We demonstrated for the first-time high potentials of quantitative glycoproteomics targeting serum tryptic glycopeptides as new class biomarkers that can be monitored directly without any enrichment process of the parent glycoproteins. Remarkably, use of structure-defined synthetic glycopeptide **1** of AGP as a calibration standard in SRM assay allowed for the absolute quantitation of the focused glycopeptides elaborated during tryptic digestion of the whole serum glycoproteins.

N-glycopeptides of AGP might be significant clues to elucidate its biological functions since it is a well-characterized glycoprotein related to many diseases including cancer. However, analysis of glycoproteins and their respective glycopeptides is difficult and challenging because of the microheterogeneity of glycans, the diversity of glycosylation site, and the low proportions of these glycopeptides in a biological sample. Therefore, we have reported the synthesis of AGP glycopeptides as its tryptic digests carrying complex type tri-antennary *N*-glycans (compounds **1, 17,** and **19**) and the selected peptide sequences are ^{40}SVQEIQATFFYFTPNKTEDTIFLR^{63} (AGP1/2, Asn^{54}), ^{1}QIPLCANLVPVPITNATLDR^{20} (AGP2, Asn^{15}), and ^{84}ENGTISR^{90} (AGP1, Asn^{85}). This effort on the synthesis of complex glycopeptides of AGP noticed an importance of key disaccharide azidoalcohol **4** derived from abundant locust bean gum galactomannan would enable the synthesis of a variety of branched *N*-glycoforms both triantennary and tetra-antennary *N*-glycans when combined with a series of enzyme-assisted modifications. Particularly, *trans*-glycosylation activities of engineered *endo*-glycosidases to the GlcNAc-peptides as acceptors using preformed oligosaccharide oxazolines as donor substrates might be a key to expand the feasibility of this approach, which provides nice tools not only to discover novel disease biomarkers but to understand the significance and molecular mechanism in the dynamic and site-specific protein glycosylation.

ACKOWLEDGEMENTS:

The present dissertation is described authors' study from 2014-2018 at Graduate School of Live Science, Hokkaido University.

Immeasurable appreciation deepest gratitude for the help and support extended to the following persons who is in one way or another have contributed in making this study possible.

Professor Shin-Ichiro Nishimura, principle investigator for his support, advices, guidance, valuable comments, suggestion and provision that benefited me much in the completion and success of this study.

Professor Kenji Monde, Associate Professor Hiroshi Hinou, Dr. Ravikumar, and Dr. Fayna Maria Garcia Martin for their encourage, knowledge sharing and help for my entire research, analysis of data and its statistical computations.

I am grateful to our group secretary Ms. Morita and Mrs. Kana Masui, they kept me organized and their indispensable help in dealing with scholarship and administration during my research so that I could optimally carry out my research and travels despite my poor ability to speak Japanese.

To all my Nishimura's Laboratory colleagues for their cooperation, daily support and encouragement.

I would like to pay tribute to the Hokkaido University, Japan for provided me a opportunity to study in Japan. I gratefully acknowledge "IGP-MEXT", Japanese government for provided me MEXT scholarship during Ph.D. course.

My family deserves special mention for their incredible support and prayers, so that I completed my doctoral course successfully. I would like to thank my family members, my mother Padma, my father Venkatesh, my sisters Pavithra K V and Pallavi K V and my beloved cousins including Somashekhar, Srinivas and more for their warm wishes to my study.

I would like to give my special thanks to my colleagues Dr. Sridhara S. N., Dr. Divakar, Ms. Swarna sangam, Ms. Jyothipriyadarshini, Ms. Niroopa Y. Mr. Madhusudana K. N. for their sincere support and wishes.

Last but not least, I would like to thank everybody who has been an important part in one or the other way for the successful realization of my thesis as well as my career.

Prof. Shin-Ichiro Nishimura

Prof. Kenji Monde

Prof. Hiroshi Hinou

Prof. K. S. Rangappa

Prof. Kempegowda

Dr. Ravikumar H. V.

Dr. Fayna Maria Garcia

Mr. Kazuki Hammura

Mr. Hayashi Naomi

Mr. Yasuhiro Yokoi

Mr. Toshifumi Suga

Mr. Hajime Wakui

Mr. Takashi Sakai

Ms. Yukina Umemoto

Ms. Maki Morita

Ms. Kana Masui

Mr. Nozomi Hirane

Dr. Shun Hayakawa

Mr. Ryosuke Koide

Mr. Risho Miyoshi

Dr. Shobith Rangappa

Mr. Islam Md. Jahirul

Mr. Arpan Chowdary

Mr. Abrha Gebreselema

Mr. Jurgen Sanes

Ms. Risa Takayama

Mr. Venkatesh

Mrs. Padma Venkatesh

Ms. Pavithra K. V.

Ms. Pallavi K. V.

Prof. D. Channegowda

Prof. Sridhara S.N.

Prof. Divakara T. R.

Ms. Swarna S.

Ms. Jyothipriyadarshini B. P.

Ms. Niroopa Y.

Prof. Mohan Kumar

Dr. Praveen Kumar

Mr. Deepak H. V.

Dr. Sunil Kumar K. R.

Mrs. Chaitra Sunil Kumar

Dr. Siddubasavegowda

Dr. Divyavani

Mr. Vishwanatha H. M.

Mrs. Maya Ravikumar

Dr. Halley Manzene

Mr. Somashekhar

Mr. Srinivas

Mrs. Sunanda siddaraju

Dr. Mahadevswamy M. M.

Mr. Timma Shetty

Mr. Ramesh murthy K. T.

Mr. Umesh murthy K. T.

Mr. Mahadeva K. M.

Mr. Rajashekhar murthy K. V.

Finally, I am expressing my apology that I could not mention personally one by one.

I sincerely thank you for all your help and encouragement.

YOGESH K V