JOC The Journal of Organic Chemistry

Article

Positional Scanning MUC1 Glycopeptide Library Reveals Importance of PDTR Epitope Glycosylation for Lectin Binding

YashoNandini Singh, Maria Carolina Rodriguez Benavente, Mohammed H. Al-Huniti, Donella Beckwith, Ramya Ayyalasomayajula, Eric Patino, William S. Miranda, Alex Wade, and Mare Cudic

J. Org. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.joc.9b02396 • Publication Date (Web): 04 Dec 2019

Just Accepted

Positional Scanning MUC1 Glycopeptide Library Reveals Importance of PDTR Epitope Glycosylation for Lectin Binding

YashoNandini Singh, Maria C. Rodriguez Benavente[#], Mohammed H. Al-Huniti, Donella Beckwith, Ramya Ayyalasomayajula, Eric Patino, William S. Miranda, Alex Wade, and Maré Cudic*

From the Department of Chemistry and Biochemistry, Charles E, Schmidt College of Science, Florida Atlantic University, 777 Glades Road, Boca Raton, Florida 33431, United States;

[#] Present Address: University of Georgia, Department of Biochemistry and Molecular Biology, Complex Carbohydrate Research Center, 120 E Green St, Athens, GA 30602, United States

* To whom correspondence should be addressed: Maré Cudic: Department of Chemistry & Biochemistry, Florida Atlantic University, Boca Raton, FL 33431; mcudic@fau.edu; Tel. (561) 297-4645.

ABSTRACT

One of the main barriers to explaining the functional significance of glycan-based changes in cancer is the natural epitope heterogeneity found on the surface of cancer cells. To help address this knowledge gap, we focused on designing synthetic tools to explore the role of tumor-associated glycans of MUC1 in the formation of metastasis *via* association with lectins. In this study, we have synthesized for the first time a MUC1-derived positional scanning synthetic glycopeptide combinatorial library (PS-SGCL) that vary in number and location of cancer-associated Tn antigen using the "tea bag" approach. The determination of the isokinetic ratios necessary for the equimolar incorporation of (glyco)amino acids mixtures to resin-bound amino acid was determined, along with developing an efficient protocol for on resin deprotection of *O*-acetyl groups. Enzyme-linked lectin assay (ELLA) was used to screen PS-SGCL against two plant lectins, *Glycine max soybean agglutinin* (SBA) and *Vicia villosa* (VVA). Results revealed a carbohydrate density-dependent affinity trend and site-specific glycosylation requirements for high affinity binding to these lectins. Hence, PS-SGCLs provide a platform to systematically elucidate MUC1-lectin binding specificities, which in long term may provide a rational design for novel inhibitors of MUC1-lectin interactions involved in tumor spread and glycopeptide-based cancer vaccines.

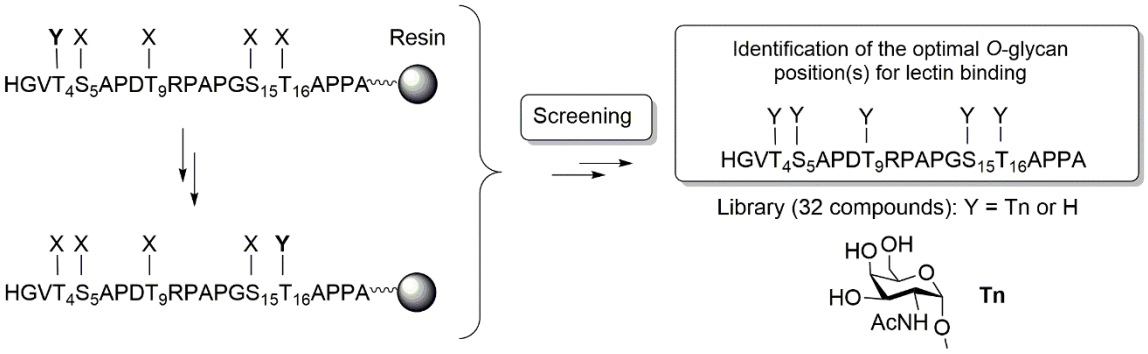
INTRODUCTION

Glycosylation, a diverse form of post-translation modification of proteins, ranges from simple monosaccharide glycans to highly complex and branched moieties.[1-4] Many glycoproteins possess several attachment sites for glycans, and such sites can be occupied by different glycans in different copies of a protein, a phenomenon termed microheterogeneity.[5] Thus, monosaccharide diversity combined with the multiple ways in which they can be arranged on proteins creates tremendous and highly complex diversity of glycoconjugate structures.[6] These glycoconjugates comprise an "information-rich" system, capable of participating in a wide range of biological functions through interaction with endogenous lectins.[7,8]

The most abundant form of *O*-linked glycosylation on cell surfaces and extracellular proteins, termed "mucin-type," is characterized by α-*N*-acetylgalactosamine (GalNAc, Tn) attached to the hydroxyl group of threonine (Thr) and serine (Ser) side chains.[9] Extension with galactose (Gal), *N*-acetylglucosamine (GlcNAc), or GalNAc produces eight different core structures, which can be further elongated or modified by sialylation, sulfation, acetylation, fucosylation, and polylactosamine-extension. Clustering of *O*-glycans in Thr/Ser-rich tandem repeat domains is characteristic of mucins, which are a family of large, heavily glycosylated proteins produced by epithelial tissue.[10] More importantly, mucins are the major carriers of altered glycosylation in carcinomas.[11-13] Aberrant glycosylation of MUC1, a highly glycosylated transmembrane mucin, includes expression of tumor-associated carbohydrate antigens (TACAs) which are often comprised of shorter and less complex *O*-glycan chains (Tn: αGalNAc-, and TF: βGal-1,3-αGalNAc-) and increased sialylation of terminal structures (sialyl-Tn: αNeuNAc-2,6-αGalNAc-).[14-17] These antigens are implicated in cell adhesion, migration, tumor proliferation, cancer progression and aggressiveness. This is seen in the interaction of MUC1 with β-galactoside binding lectin, galectin-3, which has been linked to cancer metastasis,[18] as well as the interaction with macrophage galactose-specific lectin (MGL), which plays a crucial role in tumor immune

evasion.[19,20] Even though MUC1 is the most characterized amongst mucins, the biological significance of the diversity in the type and positions of *O*-glycans on MUC1 is still unclear. A key obstacle is that due to the microheterogenicity, isolation of well-defined MUC1 from natural sources is difficult or even impossible, and as such, there is a lack of availability of rigorously controlled and structurally defined model compounds for characterizing the molecular origin of the high selectivity of endogenous lectins for MUC1.

Recent synthetic efforts have been focused towards synthesis of *O*-glycopeptide libraries and methods for their display on microarrays for further dissecting ligand-lectin interactions.[21-25] In this study, synthesis of MUC1-derived positional scanning synthetic glycopeptide libraries (PS-SGCLs) with *O*-glycans (Tn) attached to Thr or Ser at positions corresponding to the potential glycosylation sites is described (Scheme 1). This focused combinatorial library with defined structural complexity will allow us to evaluate the effect of neighboring residue glycosylation, glycan density, and/or the presence of unique patterns of *O*-glycan clusters on binding to lectins, thus helping us understand the multivalent carbohydrate-lectin recognition processes at the molecular level.



Scheme 1. A schematic diagram of the MUC1-Tn positional scanning glycopeptide combinatorial library (PS-SGCL) for use in screening assays with lectins.

Although a fully randomized one-bead-one-compound (OBOC) glycopeptide[26] and MUC1-based glycopeptide library has been synthesized,[27,28] this is the very first time that the positional scanning approach has been applied to the synthesis of glycopeptide libraries. In comparison to

the OBOC library approach, the positional scanning synthetic combinatorial library (PS-SCL) offers several advantages including: a) synthesis of a more diverse library (in the OBOC approach, each resin bead contains a single compound whereas in the PS-SCL each resin bead contains multiple compounds); b) simplified deconvolution process through positional scanning, allowing for identification of the most potent inhibitors from the mixture-based library; and c) lower number of samples required for screening, decreasing time and screening expenses.[29] In addition, the PS-SCLs are adaptable to almost any screening technique, including cell-based assays and *in vivo* models.[30]

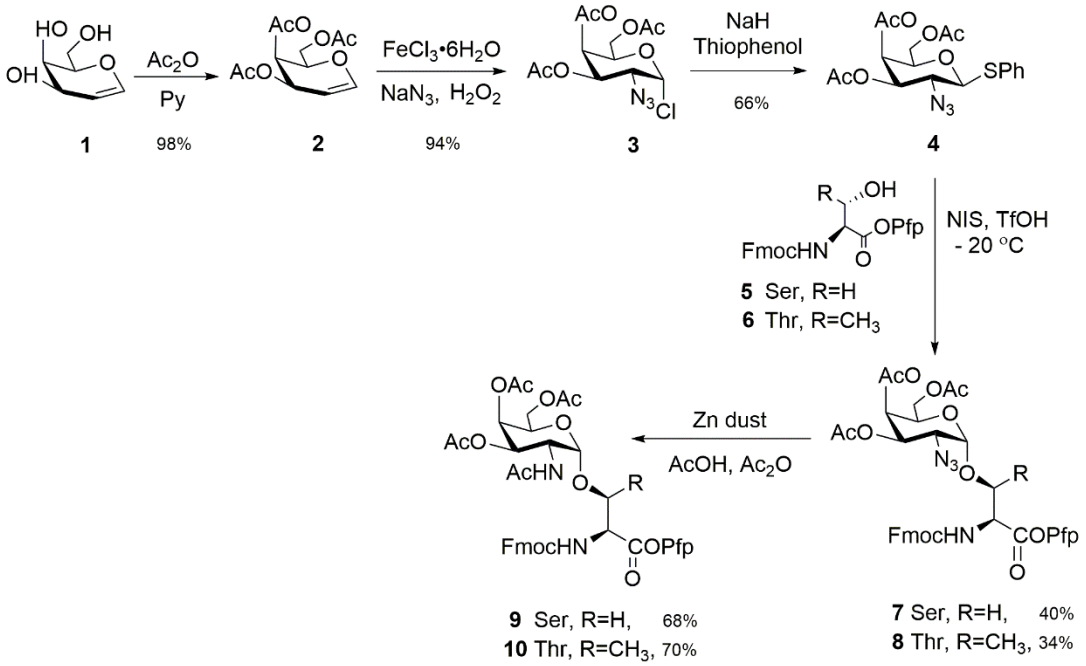
As proof of concept, we have demonstrated that a MUC1 PS-SGCL could be assembled and used to examine the interaction with two plant lectins, *Glycine max soybean agglutinin* (SBA) and *Vicia villosa* (VVA), with known specificity for Tn glycan.

RESULT AND DISCUSSION

Synthesis of Glycosylated Building Blocks.

The synthesis of the glycopeptide libraries requires access to gram quantities of glycosylated amino acids. Thus, the proposed synthetic routes must offer fast and efficient approach towards desired building blocks containing Tn antigens. The current approaches towards the glycosylated building block carrying Tn antigen mostly rely on the Koenigs-Knorr activation step for the *O*-glycoside formation.[31-33] The greatest challenge when using this approach is the low stability of glycosyl halides. Even the fast purification of the glycosyl halide on a silica column by “flash chromatography” results in significant decomposition. In addition, the reaction requires optimized low temperatures and inert conditions, and the complete stereoselectivity of the reaction is difficult to achieve despite the presence of the non-participating group such as azide at the C2 position. In order to improve on the shortcomings of this synthetic route, we are proposing an alternate synthetic path for the *O*-glycoside preparation by using the more stable phenyl thioglycoside

glycosyl donor.[34] In addition, this strategy also relied on the dual purpose, pentafluorophenyl (Pfp) group, that not only served as the C-terminus acid protecting group during synthesis but also acted as an activating group during solid-phase peptide synthesis (SPPS). The synthesis of Fmoc-protected pentafluorophenyl ester of *O*-glycosylated Ser/Thr residue **9/10** was performed according to Scheme 2. Per-acetylated D-galactal **2** was prepared from commercially available D-galactal **1** using acetic anhydride in pyridine.[35,36] The one-pot azidochlorination procedure was used to obtain 2-azido galactosyl chloride **3.**[37] Without further purification, compound **3** was converted to phenyl thioglycoside **4** through displacement of the anomeric leaving group. After purification, compound **4** was coupled with the pentafluorophenyl ester of Fmoc-protected Ser/Thr **5**/**6** under the influence of *N*-iodosuccinimide-triflic acid activation conditions to yield a mixture of anomeric *O*-glycosylated Ser/Thr building **7/8** in 53% / 46% yield, respectively. The employed coupling conditions led to an α- to β-anomer ratio of 3:1, and the change of promoter (trimethylsilyl trifluoromethaneulfonate) or temperature did not seem to affect the overall yield or selectivity of the glycosylation reaction. It is worth noting that the advantage of dealing with the more stable phenyl thioglycoside **4** may not totally compensate for the overall yield decrease and anomer selectivity in comparison to Koenigs−Knorr glycosyl donors.[37] The desired α-anomer was successfully separated from unwanted β-anomer by column chromatography. The subsequent reduction and *N*-acetylation of azido group of the α-anomer of glycosides **7** and **8** by zinc dust/acetic acid/acetic anhydride resulted in final 2-*N*-acetamido-analogs, **9** and **10** in 68% / 70% yield, respectively. The purity of the Ser/Thr building blocks **9** and **10** was confirmed by RP-HPLC chromatography (see Supporting Information, Page S11) and HRMS (ESI-TOF) (**9**: m/z $[M + Na]^+$ calcd for $C_{38}H_{35}F_5N_2O_{13}$ 845.19 Da; found 845.68 Da; **10**: m/z $[M + Na]^+$ calcd for $C_{39}H_{37}F_5N_2O_{13}$ 859.21 Da; found 859.71 Da). The NMR spectra for compounds **8**-**10** are in agreement with the previously published data and confirmed the anomeric purity of **9** and **10** (See Supporting Information, Pages S2-8). Both compounds were prepared in gram quantities for the synthesis of the positional scanning libraries.



Scheme 2. Synthesis of glycosylated amino acid building blocks **9** and **10**.

Design and Synthesis of Glycopeptide Library.

The extracellular part of MUC1 contains a domain of numerous tandem repeats of 20 amino acids (HGVTSAPDTRPAPGSTAPPA), each harboring five possible glycosylation sites (Thr^4, Ser^5, Thr^9, Ser^{15}, and Thr^{16}). Our proof of concept library was built on the 20-mer tandem repeat sequence and involved only one glycan incorporation (Tn antigen). Through the use of combinatorial chemistry, this library with varying number and location of Tn antigen can be quickly synthesized and analyzed.

The general strategy for glycopeptide combinatorial library preparation and library deconvolution is outlined in Scheme 1 and 3. The positional scanning approach was chosen for the synthesis of MUC1-derived glycopeptide library with Tn glycan attached to Thr and/or Ser at positions corresponding to the potential glycosylation sites [Thr^4 (P4), Ser^5 (P5), Thr^9 (P9), Ser^{15} (P15) and Thr^{16} (P16)] displaying native-like heterogeneous and aberrant Tn epitope presentation as seen on the surface of cancer cells. Glycopeptide library was prepared by “tea-bag” approach

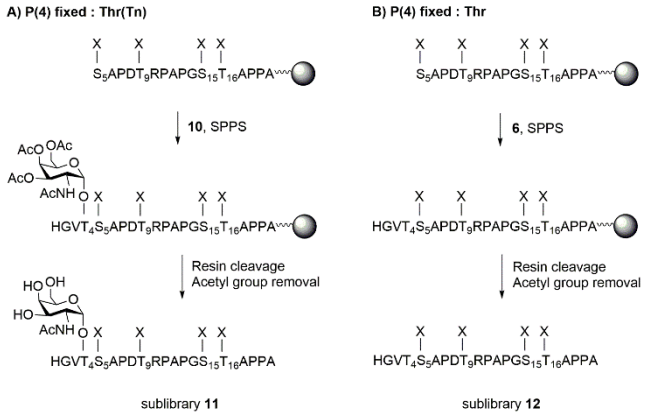
using standard Fmoc-SPPS.[38] This approach creates two sublibraries per each of the five possible glycosylation sites, resulting in a library that consists of 10 sublibraries (**11**-**20**) (Table 1). Each of the ten positional sublibraries contains the same diversity of peptides, a total of 16 glycopeptides. These glycopeptides differ only in the location of the position defined with a single (glyco)amino acid.

Fixed Position P(Y)	Number of Sublibraries	Sublibrary Sequence	Number of Glycopeptides/ Sublibrary
P4 (T* or T)	2	HGV**T*** XAPDXRPAPGXXAPPA (**11**)	16
		HGV**T**XAPDXRPAPGXXAPPA (**12**)	16
P5 (S* or S)	2	HGVX**S***APDXRPAPGXXAPPA (**13**)	16
		HGVX**S**APDXRPAPGXXAPPA (**14**)	16
P9 (T* or T)	2	HGVXXAPD**T***RPAPGXXAPPA (**15**)	16
		HGVXXAPD**T**RPAPGXXAPPA (**16**)	16
P15 (S* or S)	2	HGVXXAPDXRPAPG**S***XAPPA (**17**)	16
		HGVXXAPDXRPAPG**S**XAPPA (**18**)	16
P16 (T* or T)	2	HGVXXAPDXRPAPGX**T***APPA (**19**)	16
		HGVXXAPDXRPAPGX**T**APPA (**20**)	16

Table 1. Composition of the complete MUC1-Tn PS-SGCL library (where Y = fixed position of glycosylated (T*/S*) or non-glycosylated (T/S) building block, X = randomized position, and T*/S* corresponds to Thr/Ser bearing Tn antigen). Each sublibrary holds one position constant for either T*/S* or T/S, while the remaining positions are randomized (X).

The first set of sublibraries had a defined glycan Tn (**11**) or no glycan (**12**) attached to Thr in position 4, [Thr^4(Y)], whereas the remaining four glycosylation sites [Ser^5(X), Thr^9(X), Ser^{15}(X), andThr^{16}(X)] consisted of mixtures of the building blocks bearing Tn and no glycan attached to the Ser/Thr side chain (Scheme 3). As a result of the two fixed amino acids (glycosylated Thr and non-glycosylated Thr) incorporated in positions (Y), and the randomization at the remaining four positions (X), each sublibrary yields 16 peptides (Y^X different peptides in each mixture, 2^4). The next set of glycopeptide sublibraries had the defined glycan (**13**) or no glycan (**14**) in the second glycosylation site [Ser5(Y)], and the remaining positions [Thr4(X), Thr9(X), Ser15(X), and Thr16(X)] that will contain building block mixtures as described above. This process was repeated

for the remaining three glycosylation sites to generate a combinatorial library containing total of five glycosylation sites randomized with either non- or glycosylated Ser/Thr residues bearing Tn glycan. This resulted in a diversity of total of 32 individual glycopeptides (2^5).



Scheme 3. Principle of the positional scanning format in design and synthesis of glycopeptide library with fixed position at Thr^4 (Y) and randomized other four glycosylation sites [Ser^5(X), Thr^9(X), Ser^{15}(X), andThr^{16}(X)]. In this schematic, fixing position 4 by coupling glycosylated Thr (**10**) and nonglycosylated Thr (**6**) residues will yield the two sublibraries (**11**) and (**12**), respectively.

The non-glycosylated 20-mer MUC1 repeat (**26**) was used as a negative control and five glycosylated peptides (**21**-**25**) with difficult sequences as predicted by the Peptide Coupling Difficulty Predictor software[39] were designed to monitor efficiency of synthetic steps (Table 2). In addition to the library-derived controls by "tea bag" method, we have synthesized six controls using standard automated Fmoc-SPPS strategy.[33] These controls included mono- and multiple-glycosylated MUC1 glycopeptides with Tn glycan attached to Thr or Ser residues (**27**-**31, 33-35**) and non-glycosylated MUC1 (**32**) (Table 2). HPLC profiles and MALDI-TOF MS spectrums of synthesized MUC1-Tn control glycopeptides by "tea bag" and SPPS are provided in the Supporting Info (Pages S26-34).

Control glycopeptides by tea bag approach	Sequence
MUC1-Ser^{15},Thr^{16} (**21**)	HGVTSAPDTRPAPG**S*T***APPA
MUC1-Ser^{5},Ser^{15} (**22**)	HGVT**S***APDTRPAPG**S***TAPPA
MUC1-Thr^{9,16},Ser^{15} (**23**)	HGVTSAPD**T***RPAPG**S*T***APPA
MUC1-Thr^{4,9,16} (**24**)	HGV**T***SAPD**T***RPAPGS**T***APPA
MUC1-Thr^{4,9,16},Ser^{5,15} (**25**)	HGV**T*S***APD**T***RPAPG**S*T***APPA
MUC-1 non-glycosylated (**26**)	HGVTSAPDTRPAPGSTAPPA
Individual control glycopeptides by SPPS	**Sequence**
MUC1-Thr^{4} (**27**)	HGV**T***SAPDTRPAPGSTAPPA
MUC1-Ser^{5} (**28**)	HGVT**S***APDTRPAPGSTAPPA
MUC1-Ser^{5},Thr^{16} (**29**)	HGVT**S***APDTRPAPGS**T***APPA
MUC1-Thr^{4,9,16} (**30**)	HGV**T***SAPD**T***RPAPGS**T***APPA
MUC1-Ser^{5,15},Thr^{9,16} (**31**)	HGVT**S***APD**T***RPAPG**S*T***APPA
MUC-1 non-glycosylated (**32**)	HGVTSAPDTRPAPGSTAPPA
MUC1-Thr^{9} (**33**)	HGVTSAPD**T***RPAPGSTAPPA
MUC1-Ser^{15} (**34**)	HGVTSAPDTRPAPG**S***TAPPA
MUC1-Thr^{16} (**35**)	HGVTSAPDTRPAPGS**T***APPA

Table 2. Sequence of MUC1-Tn control glycopeptides synthesized by “tea bag” (**21**-**26**) and automated SPPS (**27**-**35**) approach.

A key component in the preparation of positional-scanning libraries is the determination of isokinetic ratios for near equimolar incorporation of each component of the mixture at randomized positions (X) with respect to a fixed position (Y).[40,41] The relative reaction ratios of Fmoc-protected amino acids used for the solid-phase synthesis of mixture based combinatorial libraries have been previously determined by competitive coupling methods.[42,43] However, the relative ratios of glycosylated amino acids necessary for equimolar coupling to the resin-bound amino acids are not known. The preassembled pentafluorophenyl esters of glycosylated amino acids **9** and **10** and their commercially available non-glycosylated counter-pairs, Fmoc-Ser/Thr(OtBu)-OPfp, were used in the evaluation of isokinetic mixtures in this study. Active esters, in particular pentafluorophenyl esters of Fmoc-protected amino acids, were already proven as the most

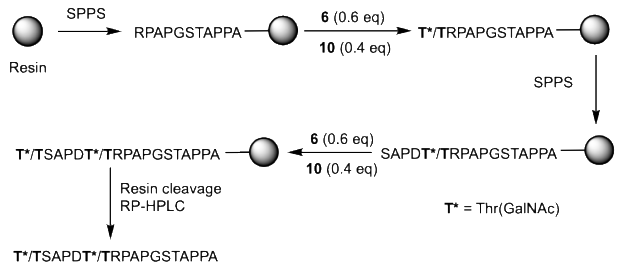
suitable option in the synthesis of PS-SCL for studying the specificity of protein kinases.[44] The simple mixture of two components, Fmoc-Ser/Thr(Tn)-OPfp (**9**/**10**) and Fmoc-Thr/Ser(OtBu)-OPfp (**5**/**6**), was incorporated in different ratios at randomized positions (X). To increase the reaction rate, couplings were performed in the presence of 1-hydroxybenzotriazole (HOBt).[45] The ratio of products formed was analyzed by RP-HPLC and compound identity was confirmed by MALDI-TOF (see Supporting Information, Pages S14-24). The isokinetic ratios for each of the five glycan positions denoted by "X" were determined within the limits of experimental error (coupling and weighing error ±5%) and results are summarized in Table 3. The optimal ratio (**6**/**10**) for positions 4, 9, and 16 was found to be 0.6 equiv. of non-glycosylated Fmoc-protected Thr (**6**) to 0.4 equiv. of glycosylated building block (**10**). The isokinetic ratios for position 4, for equimolar coupling to the resin-bound glycosylated amino acids, in this case to glycosylated Ser at position 5, was examined to determine whether the steric hindrance will play a role in the coupling kinetics. No significant difference was observed to the above-mentioned ratio for Thr-mixture (**6**/**10)** incorporation. Positions 5 and 15 occupied by Ser, required 0.6 equiv. of glycosylated building block (**9**) and 0.4 equiv. of non-glycosylated Fmoc-protected Ser (**5**) for near equimolar incorporation.

Position (P)	Peptide Name	Peptide Sequence	Isokinetic Ratio (X : X*)
P4	MUC1-Thr4(X)	**X**SAPDTRPAPGSTAPPA	0.60 : 0.40
P4	MUC1Ser5*-Thr4(X)	**X**S*APDTRPAPGSTAPPA	0.65 : 0.35
P5	MUC1-Ser5(X)	**X**APDTRPAPGSTAPPA	0.40 : 0.60
P9	MUC1-Thr9(X)	**X**RPAPGSTAPPA	0.55 : 0.45
P15	MUC1-Ser15(X)	**X**TAPPA	0.40 : 0.60
P16	MUC1-Thr16(X)	**X**APPA	0.60 : 0.40

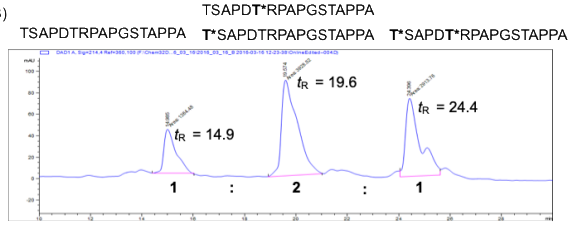
Table 3. Isokinetic ratios of Fmoc-Thr/Ser(Tn)-OPfp (X*) and Fmoc-Thr/Ser(OtBu)-OPfp (X) for equimolar coupling for all five randomized positions (X).

As a proof of concept, these ratios were used in the synthesis of a short MUC1 sequence, $MUC1_{4\text{-}20}$, (**T**^{4}SAPD**T**^{9}RPAPGSTAPPA) with positions 4 and 9 consisting of mixtures of the building blocks bearing Tn or no glycan attached to the Thr side chain (Figure 1A). Upon completion of the peptide chain assembly, peptides were cleaved from the resin, and analyzed in acetylated form of glycan by RP-HPLC analysis. The integration of total peak area of three well-separated peaks revealed 1:2:1 ratio, which would be expected for an equimolar amount of four peptides formed (Figure 1B). The peak corresponding to the non-glycosylated peptide eluted first (t_R=14.9 min), followed by the two peptides with one site glycosylated either at position 4 or 9 (t_R=19.6 min), and the double glycosylated peptide eluted the last (t_R=24.4 min). MALDI-TOF analysis confirmed the expected molecular weights of non-, mono- and double glycosylated peptides (Figure 1C). The present results were comparable with prior studies that evaluated the efficiency of isokinetic mixtures.[46]

(A)



(B)



(C)

Sequence	Expected [M + H]$^+$ (Da)	Observed [M + H]$^+$ (Da)
TSAPDTRPAPGSTAPPA	1592.77	1591.48
TSAPDT*RPAPGSTAPPA	1921.77	1921.56
T*SAPDTRPAPGSTAPPA	1921.77	1921.90
T*SAPDT*RPAPGSTAPPA	2250.77	2250.24

Figure 1. (A) General outline for the positional combinatorial synthesis of $MUC1_{4-20}$ sequence with positions 4 and 9 randomized. (B) RP-HPLC gradient used was 0-30%B in 30 min at 1 mL/min with detection at λ = 214 nm. Peptide ratio was determined by integration of peak area. (C) Products are characterized by MALDI-TOF MS.

Removal of *O*-acetyl protecting groups from the glycan portion of glycopeptides is usually done upon peptidyl cleavage from the resin with basic reagents such as sodium methoxide (NaOMe) in methanol (MeOH)[23] or aq solution of sodium hydroxide (NaOH)[33]. When comparing methanolic sodium methoxide *vs* aqueous sodium hydroxide, it has been shown that methanolic solutions were more favorable over aqueous solution in regard to β-elimination side product formation.[47] However, the removal of *O*-acetyl groups before the cleavage from the resin has been seldom explored.[48-50] In order to minimize post-cleavage steps with the library, we have explored conditions for efficient on-resin *O*-acetyl group removal. Control mono-substituted MUC1 peptide **27** was first treated with NaOMe (10 mM) in DMF-MeOH mixture (17:3, v/v) (Figure 2A). For comparison, deacetylation of the same peptide was carried out after the cleavage from the resin by treatment with NaOH solution (0.01 M) (See Supporting Information, Pages S28). The reaction progress was followed by RP-HPLC, and products analyzed by MALDI-TOF. Although the on-resin conditions were found effective in the synthesis of similar mucin-like peptides,[48] we have observed significant degree of β-elimination of glycan (over 50%) even with careful control of pH (8.5-9.5). The presence of side products was minimized by replacement of DMF with DCM (Figure 2B), and completely eliminated using MeOH as solvent (Figure 2C). The reactions in MeOH were completed in 3 h and the yield of the final product was comparable to the product obtained by treatment with aq NaOH solution (0.01 M) after the cleavage from the resin. This may not be surprising since resins with polar PEG spacers, such as the ones used in our glycopeptide synthesis, allow the resin to swell in a wide range of solvents including MeOH.[51] In addition, we have considered the possibility that in a polar aprotic solvent, such as DMF, nucleophilicity correlates well with basicity, making the α-proton more susceptible to β-elimination. On the

contrary, MeOH is a polar protic solvent, and allows for hydrogen bonding, thus the nucleophile is considerably less basic (reactive), and at the same time, the α-proton is less acidic. Furthermore, DMF may contain free amine impurities which may facilitate the base-catalyzed β-elimination of glycan. Additional basic conditions in MeOH were explored, such as NaOH (10 mM) (Figure 2D), ammonia (NH_3) (7 M) (Figure 2E), and hydrazine:MeOH (7:3, v/v) (Figure 2F). Hydrazine treatment resulted in β-elimination side product in an amount similar to values obtained with NaOMe (10 mM) in DMF-MeOH mixture (17:3, v/v) in contrast to some successful attempts reported in the literature.[49,50] Complete deprotection without presence of side products was obtained for conditions involving ammonia (7 M) and NaOH (10 mM) in MeOH (Figure 2D, E). Based on these findings, subsequent optimization studies with multi-substituted MUC1 peptides were performed exclusively by using NaOMe, NaOH, and NH_3 in MeOH as solvent. Overall on-resin treatment of di- and multi-glycosylated MUC1 peptides with mild base, NH_3 (7 M) in MeOH, was more efficient in comparison to NaOMe or NaOH in MeOH (see Supporting Information, Page S25). Thus, all sub-libraries and controls prepared by "tea bag" approach were treated with 7 M NH_3 in MeOH solution individually for 3 h prior to TFA-promoted cleavage from resin. An example of the RP-HPLC analyses and MALDI-TOF MS spectra of two glycopeptide sublibraries, **17** and **18**, with fixed position as Ser-*O*-GalNAc or Ser, respectively, is provided in the Supporting Information (S39-41). All sublibraries exhibited similar profile, since they all have one of the glycan positions fixed with a glycosylated or a non-glycosylated Thr or Ser residue, and the remaining positions are equimolar ratio of the glycosylated and non-glycosylated Thr or Ser residue. A crude sublibrary with fixed position as Thr or Ser is expected to result in 16 glycopeptides (non-glycosylated, mono-, di-, tri- and tetra-glycosylated peptides), while a crude sublibrary with fixed position as Thr-*O*-GalNAc or Ser-*O*-GalNAc should result in 16 glycopeptides (mono-, di-, tri-, tetra-, and penta-glycosylated). The presence or absence of the non-glycosylated or penta-glycosylated peptide was detectable by HPLC and MS analysis.

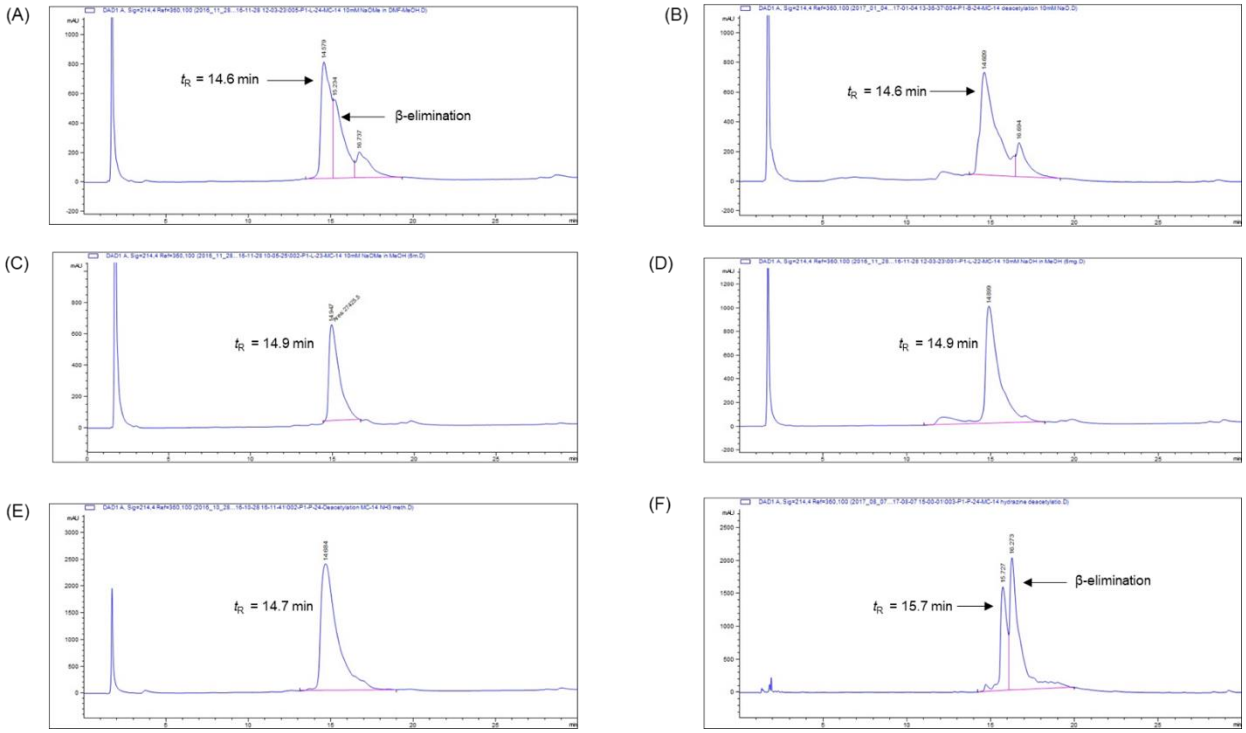


Figure 2. On-resin deacetylation methods explored for mono-glycosylated MUC1-Tn control, **27**, HGV**T(Tn)**SAPDTRPAPGSTAPPA. Conditions tested were (A) 10 mM NaOMe in DMF:MeOH (17:3) (B) 10 mM NaOH in DCM:MeOH (17:3) (C) 10 mM NaOMe in MeOH (D) 10 mM NaOH in MeOH and (E) 7 M NH_3 in MeOH (F) Hydrazine in MeOH (7:3). RP-HPLC gradient used was 0-30%B in 30 min at 1 mL/min and detection was at λ = 214 nm.

Screening and Deconvolution of Glycopeptide Library.

Enzyme-linked lectin assay (ELLA) was used to demonstrate the feasibility of our MUC1-derived glycopeptide positional scanning combinatorial libraries to study specificity of lectins.[52,53] Legume lectins ability to detect carbohydrate moieties have been useful for studying the molecular recognition properties of cellular oligosaccharides.[54] Among lectins that recognize GalNAc (Tn), SBA and VVA were chosen for this study.[55] The screening assays for BSA-monosaccharides binding to lectins, published by our group,[53] were adapted for the MUC1-derived glycopeptide

library screening assay. Briefly, microtiter plate wells were immobilized with MUC1 glycopeptides and/or glycopeptide libraries, blocked with 3% of bovine serum albumin (BSA) to minimize non-specific binding, and incubated with either biotin-conjugated SBA or VVA lectin. After removal of excess of lectin, horseradish peroxidase (HRP) conjugated streptavidin was added, followed by a HRP substrate solution. The intensity of the colored product was recorded at 450 nm using an enzyme-linked immunosorbent assay (ELISA) plate reader. The assay was performed in three replicates for each concentration point in 96-well plates. The average absorbance reading after background subtraction was plotted against glycopeptide and/or glycopeptide sublibrary concentrations. Controls **27**-**32**, prepared by automated SPPS approach, were first screened against lectins at eight concentrations (0.1, 1, 5, 10, 50, 100, 250, 500 µg/mL) (see Supporting Information, Figure S35a,b). Non-glycosylated MUC1 peptide **30** was used as a negative control and showed no binding to either SBA or VVA lectin. Mono-substituted glycopeptide controls **27** and **28** bound to both lectins in a concentration-dependent manner. However, at concentrations higher than 10 µg/mL, saturation of lectin binding sites was observed with multi-substituted MUC1 peptides **29**-**31**, and concentration dependence was lost (see Supporting Information, Figure S35a,b). It is worth mentioning that the synergy effect of mixtures cannot be excluded without more careful analysis of data. However, the similarities in the measured OD values for sublibraries and individual multivalent peptides suggest its absence. In addition, it has been shown that synergistic effect is rarely observed and doesn't usually interfere with the deconvolution of the positional scanning libraries.[56]

In order to maximize the outcome of the whole library screen, we have performed two additional preliminary screens: a) at higher concentration range (10, 100, 250, 500, and 1000 µg/mL) with library controls **21**-**23** and **27**-**32**, and b) lower concentration range (0.1, 0.5, 1, 2, 5, 10 µg/mL) with controls **31**-**32** and sublibrary **20** (see Supporting Information, Figure S36c,d). The library controls **21**-**23** showed similar binding pattern as the individually synthesized MUC1-Tn glycopeptides, where binding increased with increase in glycosylation sites. Based on the results

of preliminary absorbance read-out responses with control glycopeptides, the final glycopeptide sublibraries **11**-**20** were screened at four different concentrations (0.1, 0.5, 10, and 500 µg/mL). Control individual glycopeptides, derived by "tea bag" approach (**21**-**24**) and automated SPPS approach **25**-**30** were used as controls. MUC1-Tn PS-SGCL and control glycopeptides in general showed higher binding to VVA lectin by approximately 1 OD unit compared to SBA and results point out to the importance of Tn clusters in lectin binding (Figure 3 and 4, and Supporting Information Figure S37e,f). Concentration-dependent activity for both lectins was observed at concentrations 0.1-10 µg/mL, however saturation was achieved at 500 µg/mL. Notable exception is sublibrary **16**, which has non-glycosylated Thr fixed at position 9. This sublibrary showed large affinity loss in screening with SBA lectin compared to sublibrary **15** that has the Thr^{9} glycosylated with Tn antigen (Figure 3 and Supporting Information Figure S38g). Evidently, SBA binding specificity for MUC1 tandem repeat is significantly influenced by the presence or absence of glycosylation in the PDTR region. Strong affinity of SBA for MUC1 repeat bearing Tn glycan within PDTR motif has been reported.[57] In such study, the extended binding epitope, that in addition to glycan includes peptide backbone close to the glycosylation site was supported by the experimental evidences and molecular dynamic simulations analysis. The enhancement in affinity was attributed to the additional hydrogen bonding and electrostatic interactions with the peptide backbone and the hydrophobic pocket of SBA. In our study, a smaller decrease in the binding affinity was observed for VVA suggesting differences in recognition of glycopeptide epitopes amongst the two lectins (Figure 4 and Supporting Information Figure S37f). We also observed that at 0.5 µg/mL, SBA lectin showed the highest affinity for the Tn antigen presented at Thr^{9} (sublibrary **15**), followed by slight preference for Ser^{15} and Thr^{16} (sublibrary **17** and **19**, respectively) compared to Thr^{4} and Ser^{5} positions (sublibrary **11** and **13**, respectively) (Supporting Information Figure S37e). In accordance with these results, the affinity for SBA of control diglycosylated peptide **29** that does not contain glycan attached to Thr^{9}, was lower in affinity when compared to peptide **30** (triglycosylated) and **31** (tetraglycosylated), that contain Thr^{9}

glycosylation site. In the case of VVA lectin, these affinities were in the similar range. Nevertheless, the binding of **29** was higher than binding of monoglycosylated peptides **27** (glycan attached to Thr^{4}) and **28** (glycan attached to Ser^{5}). These results further support the relative importance of the effect Thr^{16} has to binding and multivalency. To rationalize our results from the positional scanning deconvolution of the library, we screened individually synthesized mono-glycosylated MUC1-Tn peptides that have the Tn *O*-glycan attached at Thr^{4} (**27**), Ser^{5} (**28**), Thr^{9} (**33**), Ser^{15} (**34**), and Thr^{16} (**35**) positions (Table 2) at two concentrations, 10 and 500 µg/mL, with SBA lectin (Figure 5). Our results clearly show the highest preference of SBA lectin for position Thr^{9}, followed by positions Thr^{16} and Ser^{15}, and lastly, for Ser^{5} and Thr^{4}. These findings further suggest that the binding profile of SBA-MUC1 interactions depends not only on the carbohydrate moiety but also on the peptide region surrounding the glycan site of attachment.

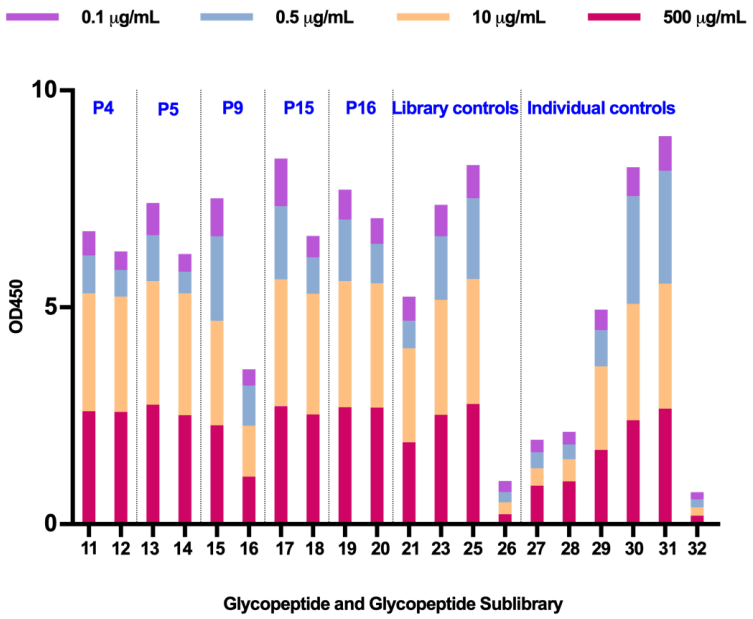


Figure 3. Screening of MUC1-Tn PS-SCL (**11**-**20**) and control glycopeptides synthesized by tea bag library approach (**21**-**26**) and automated SPPS (**27**-**32**) with SBA lectin using ELLA assay. Absorbance values at 450 nm are the average of triplicate wells after background subtraction, which are represented as stacked scale score for each compound.

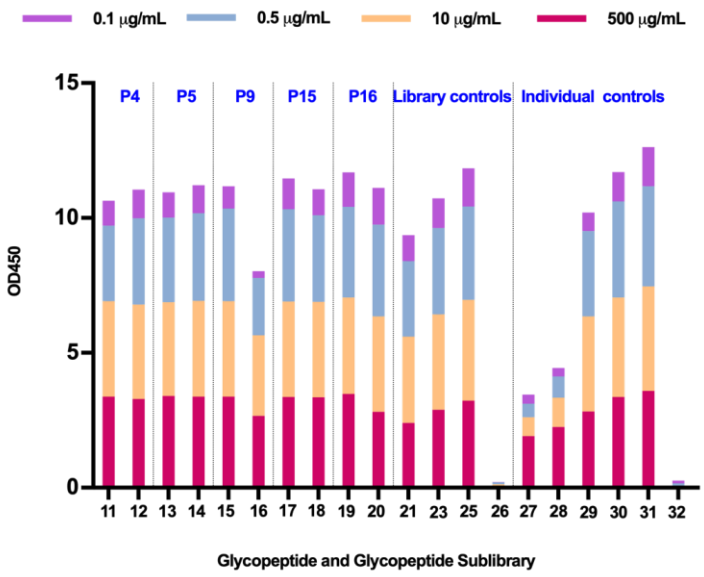


Figure 4. Screening of MUC1-Tn PS-SCL (**11**-**20**) and control glycopeptides synthesized by "tea bag" library approach (**21**-**26**) and automated SPPS (**27**-**32**) with VVA lectin using ELLA assay. Absorbance values at 450 nm are the average of triplicate wells after background subtraction, which are represented as stacked scale score for each compound.

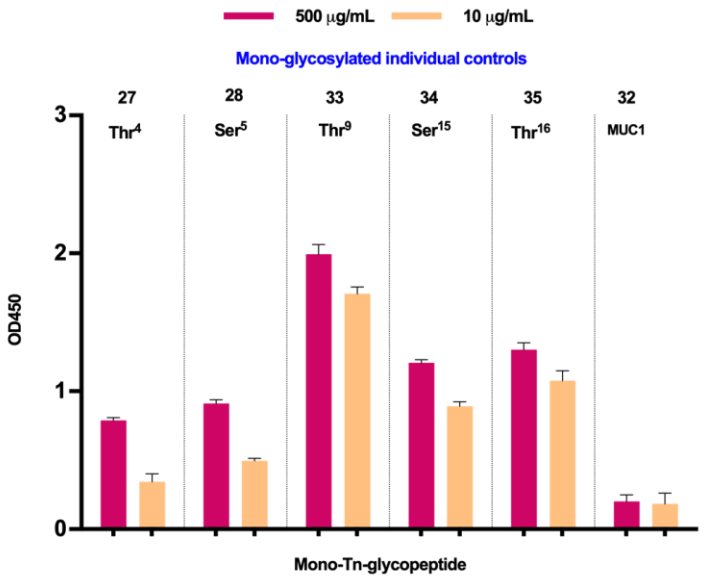


Figure 5. Binding profile of mono-Tn glycopeptides (**33**-**35**) with SBA lectin using ELLA assay. Absorbance values at 450 nm are the average of triplicate wells after background subtraction.

CONCLUSIONS

The first synthesis of PS-SGCL based on the MUC1 repeat has been described. This library with varying number and location of Tn antigen was screened against two plant lectins SBA and VVA. Important recognition epitope for SBA lectin specificity was determined. The presence of Tn glycan at position 9 within the PDTR epitope has shown to be crucial for recognition, suggesting that interaction depends not only on the carbohydrate moiety but also on the peptide region surrounding the glycan site of attachment. Thus, synthesis of glycopeptide libraries in positional scanning format is a promising tool to probe biological roles of tumor-associated MUC1.

EXPERIMENTAL SECTION

Reagents. All starting materials and reagents for organic synthesis were purchased from commercial sources and were of higher or analytical grade. Solvents were purchased as either anhydrous grade products in sealed containers or reagent grade products and used as received. HPLC grade acetonitrile was used for peptide purification and analysis by HPLC. Fmoc-protected amino acids and coupling reagents for peptide synthesis, 2-(6-chloro-1H-benzotriazol-1-yl)-1,1,3,3 tetramethylaminium hexafluorophosphate (HCTU) and 1-hydroxybenzotriazole (HOBt), were obtained from EMD Biosciences. Biochemistry reagents for ELLA, horseradish-peroxidase conjugated streptavidin (HRP-streptavidin) and 3,3',5,5'-tetramethylbenzidine (TMB), were purchased from Fisher Scientific. Biotin-labeled *Soybean Agglutinin* (biotin-SBA) and *Vicia Villosa* (biotin-VVA) were purchased from Vector Laboratories.

General Experimental Methods. Glycosylation reactions were carried under inert atmosphere using standard disposable syringes, stainless steel needles, cannula, and septa. The progress of reaction was monitored by thin-layer chromatography (TLC) on 200 μm thick silica gel F-254 coated aluminum plates, visualized by charring with 10% solution of sulfuric acid in ethanol and/or by UV light when applicable. Flash-column chromatography was performed on silica gel columns

(230-400 Mesh). ^{1}H and ^{13}C NMR spectra were obtained on Bruker Advanced III 400 MHz spectrometers. Signals are reported in terms of their chemical shifts (δ in ppm) relative to $CDCl_3$ (1H: δ 7.26 and 13C{1H}: δ 77.16). ^{1}H NMR spectra data are reported in terms of chemical shift (δ, ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), coupling constant (J values, Hz), and integration. The signal assignments given for ^{1}H- and ^{13}C NMR data are based on 2D NMR spectra such as HSQC.

Coupling constants (J values) are reported in hertz (Hz). Accurate mass (HRMS) for synthetic glycosylated building blocks was obtained from the University of Florida using an Agilent 1100 series system consisting of an ESI-TOF instrument (positive ion mode).

Synthesis of Thr/Ser Glycosylated Building Blocks Bearing Tn Antigen. Compounds **1**-**3** and **5** - **6** were prepared according to the protocol described previously by our group[33] and in case of compound **4** by others[58].

*Synthesis of 1-phenyl-3,4,6-tri-O-acetyl-2-azido-2-deoxy-α-D-galactopyranoside (**4**)*. Based on Scheme 1. To a suspension of sodium hydride (60% in mineral oil; 0.18 g, 1.3 equiv, 4.45 mmol) in anhydrous THF (13 mL) at -15°C, thiophenol (491 μL, 1.3 equiv, 4.45 mmol) was added in small portions. Reaction mixture was stirred under argon for approximately 40 min until formation of sodium thiophenolate was observed. Chloroazide sugar **3** (1.2 g, 1 equiv, 3.43 mmol) was dissolved in anhyd acetonitrile (13 mL) and was cooled down to -15°C, followed by sodium thiophenolate addition, and stirred at room temperature. The progress of the reaction was monitored by TLC [14:2:1 (v/v/v) DCM/toluene/ethyl acetate]. Once the complete disappearance of chloroazide sugar **3** had been observed (usually in 3 h), the reaction was diluted with DCM (15 mL) and washed with water in a separatory funnel. The organic layer was separated, dried over sodium sulfate (Na_2SO_4), and concentrated under vacuum. The residue was purified by flash chromatography [6:1 (v/v) toluene/ethyl acetate] to obtain 957 mg of thiophenol sugar **4** in 66%

yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.61 – 7.59 (m, 2H, Ar), 7.35 – 7.33 (m, 3H, Ar), 5.34 (d, *J* = 2.7 Hz, 1H, H-4, β), 4.86 (dd, J = 10.3, 3.1 Hz, 1H, H-3, β), 4.51 (d, J = 10.1 Hz, 1H, H-1, β), 4.13 (ddd, J = 26.7, 11.4, 6.6 Hz, 2H, H-6, β), 3.89 (d, J = 6.5 Hz, 1H, H-5, β), 3.64 (t, J = 10.2 Hz, 1H, H-2, β), 2.08 (s, 3H, CH_3–(C=O)-, β), 2.02 (d, J = 4.2 Hz, 6H, CH_3–(C=O)-, β). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$) δ 170.4, 169.9, 169.7 (CH_3–(C=O)-, α and β), 133.6, 131.0, 129.0, 128.6 (Ar, α and β), 86.5 (C-1, β), 74.4 (C-5, β), 73.0 (C-3, β), 66.5 (C-4, β), 61.6 (C-6, β), 59.4 (C-2, β), 20.9 – 20.49 (CH_3–(C=O)-, α and β). The NMR data are in agreement with published literature values.[58]

*Synthesis of N^{α}-(9H-fluoren-9-yl)-methoxycarbonyl-O-(3,4,6-tri-O-acetyl-2-azido-2-deoxy-α-D-galactopyranosyl)-L-serine pentafluorophenyl ester (**7**).* Based on Scheme 1. Thiophenol sugar **4** (0.4 g, 1 equiv, 0.95 mmol) and Fmoc-Ser-OPfp **5** (0.51 g, 1.1 equiv, 1.04 mmol) were dissolved in anhyd DCM (18 mL) in the presence of 4 Å molecular sieves. The reaction mixture was stirred for 2 h at room temperature, after which *N*-iodosuccinimide (0.32 g, 1.5 equiv, 1.42 mmol) was added. Then the reaction flask was cooled down to -20°C and trifluoromethanesulfonic acid (28 μL, 0.2 equiv, 0.32 mmol) was added. The reaction mixture was stirred at -20°C and the progress of the glycosylation reaction was monitored by TLC [2.5:1 (v/v) hexane/ethyl acetate]. Once the complete disappearance of thiophenol sugar **4** was observed, the reaction was neutralized with triethyl amine and diluted with DCM (50 mL). The mixture was filtered through Celite and the organic layer was washed with water in a separatory funnel, separated, dried over Na_2SO_4, and concentrated under vacuum. The crude residue was purified by flash chromatography [5:1 (v/v) toluene/ethyl acetate] to obtain 304 mg of the α-anomer of C-2 azide analog of Fmoc-Ser(Tn) building block **7** in 40% yield and the corresponding β-anomer was isolated in 13% yield (99 mg). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (d, J = 7.4 Hz, 2H), 7.62 (dd, J = 7.2, 3.1 Hz, 2H), 7.40 (t, J = 7.4 Hz, 2H), 7.32 (tdd, J = 7.4, 3.4, 1.1 Hz, 2H), 6.13 (d, J = 8.6 Hz, 1H), 5.46 (d, J = 2.5 Hz, 1H), 5.35 – 5.27 (m, 1H), 5.01 (d, J = 3.5 Hz, 1H), 4.96 (d, J = 8.6 Hz, 1H), 4.48 (d, J = 6.5 Hz, 2H), 4.33 (dd, J = 11.2, 3.3 Hz, 1H), 4.26 (t, J = 6.8 Hz, 1H), 4.17 (d, J = 6.3 Hz, 1H), 4.13 – 4.03 (m,

3H), 3.71 (dd, J = 11.2, 3.5 Hz, 1H), 2.17 (s, 3H), 2.07 (s, 3H), 1.98 (s, 3H). $^{13}C\{^{1}H\}$ NMR (100 MHz, $CDCl_3$) δ 170.6, 170.1, 169.9 (CH_3-(C=O)-), 166.3 (NH–C=O), 143.7, 141.5, 128.0, 127.3, 125.2, 120.2 (Ar), 99.8 (C-1), 69.9 (C-β), 68.1 (C-3), 67.7 (C-4), 67.6 (C-5), 67.6 (FmocCH_2), 62.0 (C-6), 57.6 (C–α), 54.5 (C-2), 47.2 (FmocCH), 20.8, 20.7, 20.6 (CH_3-(C=O)). HRMS (ES-TOF) m/z: calcd for $C_{36}H_{31}F_5N_4O_{12}$ $[M + H]^+$ = 807.1931, found $[M + Na]^+$ = 829.1751 and $[M + NH_4]^+$ = 824.2197.

Synthesis of N^{α}-(9H-fluoren-9-yl)-methoxycarbonyl-*O*-(3,4,6-tri-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranosyl)-L-threonine pentafluorophenyl ester ***(8)***. Based on Scheme 1. Thiophenol sugar **4** (0.4 g, 1 equiv, 0.95 mmol) and Fmoc-Thr-OPfp **6** (0.53 g, 1.1 equiv, 1.04 mmol) were dissolved in anhyd DCM (18 mL) in the presence of 4 Å molecular sieves. The reaction mixture was stirred for 2 h at room temperature, after which *N*-iodosuccinimide (0.32 g, 1.5 equiv, 1.42 mmol) was added. Then the reaction flask was cooled down to -20°C and trifluoromethanesulfonic acid (28 μL, 0.2 equiv, 0.32 mmol) was added. The reaction mixture was stirred at -20°C and the progress of the glycosylation reaction was monitored by TLC [2.5:1 (v/v) hexane/ethyl acetate]. Once the complete disappearance of thiophenol sugar **4** was observed, the reaction was neutralized with triethyl amine and diluted with DCM (50 mL). The mixture was filtered through Celite and the organic layer was washed with water in a separatory funnel, separated, dried over Na_2SO_4, and concentrated under vacuum. The crude residue was purified by flash chromatography [5:1 (v/v) toluene/ethyl acetate = 5:1] to obtain 263 mg of the α-anomer of C-2 azide analog of Fmoc-Thr(Tn) building block **8** in 44% yield and the corresponding β-anomer was isolated in 12% yield (91 mg). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, J = 7.6 Hz, 2H), 7.66 – 7.58 (m, 2H), 7.41 – 7.28 (m, 4H), 5.89 (d, J = 9.3 Hz, 1H), 5.48 (d, J = 2.4 Hz, 1H), 5.32 – 5.28 (m, 1H), 5.17 (d, J = 3.6 Hz, 1H), 4.77 (dd, J = 9.3, 2.1 Hz, 1H), 4.58 (dd, J = 6.4, 2.0 Hz, 1H), 4.45 (ddd, J = 24.4, 10.6, 7.4 Hz, 2H), 4.29 (t, J = 6.9 Hz, 2H), 4.11 (d, J = 6.4 Hz, 2H), 3.77 (dd, J = 11.1, 3.7 Hz, 1H), 2.17 (s, 3H), 2.08 (s, 3H), 2.06 (s, 3H), 1.45 (d, J = 6.4 Hz, 3H). $^{13}C\{^{1}H\}$ NMR

(100 MHz, $CDCl_3$) δ 170.5, 170.1, 170.0 (CH_3-(C=O)-), 166.9 (NH–C=O), 156.7, 143.8, 141.5, 132.6, 129.5, 127.9, 127.3, 125.3, 120.2 (Ar), 99.3 (C-1), 76.3 (C-β), 68.6 (C-3), 67.9 (C-4), 67.5 (C-5), 67.4 (FmocCH_2), 61.9 (C-6), 58.6 (C–α), 58.1 (C-2), 47.2 (FmocCH), 20.8 (CH3-(C=O)-), 19.0 (C–γ). HRMS (ES-TOF) m/z: calcd for $C_{37}H_{33}F_5N_4O_{12}$ $[M + H]^+$ = 821.2088, found $[M + Na]^+$ = 843.1907 and $[M + NH_4]^+$ = 838.2353.

Synthesis of *N*$^{\alpha}$-(9H-fluoren-9-yl)-methoxycarbonyl-*O*-(3,4,6-tri-*O*-acetyl-2-acetamido-2-deoxy-α-D-galactopyranosyl)-L-serine pentafluorophenyl ester ***(9)***. Based on Scheme 1. Compound **7** (0.3 g, 1 equiv, 0.37 mmol) was dissolved in anhyd THF (4.8 mL). The reaction flask was cooled down in an ice bath and acetic anhydride (8.2 mL, 200 equiv) and acetic acid (0.8 mL, 34 equiv) were added, followed by activated zinc (0.5 g, 20 equiv, 7.44 mmol), and the reaction mixture was stirred overnight under argon. The progress of the reaction was monitored by TLC [1:1 (v/v) toluene/ethyl acetate]. Once the complete disappearance of **7** was observed, the reaction was diluted with ethyl acetate (25 mL). The mixture was filtered through Celite and the organic layer was washed with water in a separatory funnel, separated, dried over Na_2SO_4, and concentrated under vacuum with addition of toluene at the end to azeotrope trace amounts of acid. The residue was purified by flash chromatography [1:1 (v/) toluene/ethyl acetate] to obtain 208 mg of the α-anomer of Fmoc-protected *O*-glycosylated Ser building block **9** in 68% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (d, *J* = 7.4 Hz, 2H, Ar-H Fmoc), 7.60 (m, 2H, Ar-H Fmoc), 7.40 – 7.30 (m, 4H, Ar-H Fmoc), 6.22 (d, *J* = 8.6 Hz, 1H, NH$COCH_3$), 5.72 (d, *J* = 9.4 Hz, 1H, NH-Ser), 5.36 (dd, *J* = 14.1, 2.5 Hz, 1H, Gal-H_4), 5.12 (dd, *J* = 11.4, 2.8 Hz, 1H, Gal-H_3), 4.95 (dd, *J* = 16.2, 5.9 Hz, 2H, Gal-H_1 & CHα), 4.66 – 4.60 (m, 1H, Gal-H_2), 4.51 – 4.49 (m, 2H, FmocCH_2), 4.26– 4.01 (m, 6H, FmocCH, Gal-H_5, Gal-H_{6a-6b}, CH_2β-Ser), 2.17 (s, 3H), 2.00 (s, 3H), 1.96 (s, 3H), 1.89 (s, 3H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$) δ 171.2, 170.5, 170.4 (CH3 - (C=O)-), 166.8 (NH – C=O), 155.9, 143.7, 141.5, 128.0, 127.3, 125.0, 120.2 (Ar), 99.1 (C-1), 69.1 (C- β), 68.2 (C-4), 67.6 (C-3), 67.3 (FmocCH_2 & C-5),62.2 (C – 6) 54.30 (C– α), 47.7 (C-2), 47.2 (FmocCH), 22.9, 20.9, 20.8, 20.7

(CH3 - (C=O)-). HRMS (ES-TOF) m/z: calcd for $C_{38}H_{35}F_5N_2O_{13}$ $[M + H]^+$ = 823.2132, found $[M + Na]^+$ = 845.1952.

Synthesis of N^{α}-(9H-fluoren-9-yl)-methoxycarbonyl-*O*-(3,4,6-tri-*O*-acetyl-2-acetamido-2-deoxy-α-D-galactopyranosyl)-L-threonine pentafluorophenyl ester ***(10)***. Based on Scheme 1. Compound **8** (0.3 g, 1 equiv, 0.37 mmol) was dissolved in anhyd THF (4.7 mL). The reaction flask was cooled down in an ice bath and acetic anhydride (8.1 mL, 200 equiv) and acetic acid (0.8 mL, 34 equiv) were added, followed by activated zinc (0.5 g, 20 equiv, 7.31 mmol), and the reaction mixture was stirred overnight under argon. The progress of the reaction was monitored by TLC [1:1 (v/v) toluene/ethyl acetate]. Once the complete disappearance of **8** was observed, the reaction was diluted with ethyl acetate (25 mL). The mixture was filtered through Celite and the organic layer was washed with water in a separatory funnel, separated, dried over Na_2SO_4, and concentrated under vacuum with addition of toluene at the end to azeotrope trace amounts of acid. The residue was purified by flash chromatography [1:1 (v/v) toluene/ethyl acetate] to obtain 214 mg of the α-anomer of Fmoc-protected *O*-glycosylated Thr building block **10** in 70% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, *J* = 7.5 Hz, 2H, Ar-H Fmoc), 7.63 (d, *J* = 7.4 Hz, 2H, Ar-H Fmoc), 7.43– 7.28 (m, 4H, Ar-H Fmoc), 6.05 (d, *J* = 9.5 Hz, 1H, NHCOCH$_3$), 5.77 (d, *J* = 9.9 Hz, 1H, NH-Thr), 5.36 (d, *J* = 17.6 Hz, 1H, Gal-H$_4$), 5.08 (dd, *J* = 11.5, 3.2 Hz, 1H, Gal-H$_3$), 5.01 (d, *J* = 3.5 Hz, 1H, Gal-H$_1$), 4.73 (dd, *J* = 9.5, 1.3 Hz, 1H, CHα-Thr), 4.63 – 4.52 (m, 3H, Fmoc CH2 & Gal-H$_2$), 4.43 (d, *J* = 5.0 Hz, 1H, CHβ-Thr), 4.24 (dt, *J* = 17.8, 6.1 Hz, 2H, Fmoc-CH, Gal-H$_5$), 4.15 – 4.03 (m, 2H, Gal-H$_{6a-6b}$), 2.16 (s, 3H, CH$_3$ – (C=O)), 2.04 (s, 3H, CH$_3$ – (C=O)), 1.99 (s, 3H, CH$_3$ – (C=O)), 1.68 (s, 3H, NHCOCH$_3$), 1.42 (d, *J* = 6.3 Hz, 3H, CH$_3$-Thr).^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$) δ 171.2, 170.5, 170.4, 170.1 (CH3 - (C=O)-), 167.6 (NH – C=O), 156.5, 143.7, 141.5, 127.9, 127.3, 125.0, 120.2 (Ar), 100.2 (C – 1),76.7 (C- β from 2D, signal merged with $CDCl_3$) 68.3 (C- 4), 67.6 (C-3), 67.5 (FmocCH$_2$), 67.4 (C-5), 62.3 (C – 6), 58.6 (C–

α), 47.6 (C- 2), 47.3 (FmocCH), 22.7, 20.9, 20.8, 20.7 (<u>C</u>H_3 - (C=O)-), 18.7 (C – γ). HRMS (ES-TOF) m/z: calcd for $C_{39}H_{37}F_5N_2O_{13}$ $[M + H]^+$ = 837.2289, found $[M + Na]^+$ = 859.2108.

Synthesis and Purification of (Glyco)peptides. Standard Fmoc solid-phase peptide chemistry was employed to synthesize all individual (glyco)peptide analogs on a PS3 automated peptide synthesizer (GyrosProtein Technologies Inc.) as previously reported.[59] TentaGel S RAM resin (Advanced ChemTech) was used, and amino acid couplings were done using a 4-fold excess of amino acids, HOBt, and HCTU, in the presence of 0.4 M *N*-methylmorpholine (NMM) in DMF. The pentafluorophenyl ester of glycoamino acid was coupled manually using a 1.5-fold excess, in the presence of *N*, *N*-diisopropylethylamine (DIPEA) (pH 8). The coupling was carried out for 16 h. The ninhydrin test was used to monitor completion of coupling. The fully assembled glycopeptide chains were cleaved from the resin using a TFA/thioanisole/water mixture (95:2.5:2.5) ratio for 3 hours. The cleavage solution was then precipitated in cold MTBE to yield the crude acetylated glycopeptides. Acetyl groups were removed using 0.01 M NaOH solution for 15 min, solution was neutralized with hydrochloric acid (HCl), and lyophilized to yield the final crude MUC1-Tn glycopeptides.

(Glyco)peptide purification was performed using a 1260 Infinity reversed-phase high-performance liquid chromatography (RP-HPLC) system by Agilent Technologies. The analytical RP-HPLC uses a Grace Vydac monomeric C18 column (250 x 4.6 mm, 5μm, 120Å) at 1 mL/min flow rate and Aeris Peptide C18 column (150 x 4.6 mm, 3.6μm, 100Å) at 0.8 mL/min flow rate, with 0.1% TFA in water (A) and 0.1%TFA in acetonitrile (B) as the eluents. The elution gradient for analytical RP-HPLC purification was 0 to 30%B over 30 min. Preparative RP-HPLC was utilized for final peptide purification, which uses the Grace Vydac monomeric C18 column (250 x 22 mm, 10μm, 120Å) at 10 mL/min flow rate, with 0.1% TFA in water (A) and 0.1%TFA in acetonitrile (B) as the eluents. The elution gradient for preparative RP-HPLC purification was 0 to 25% over 110 min. The peptide analogs were detected at 214 nm by the UV-Vis detector (variable

detector, Agilent Technologies). Purified peptides were characterized by MALDI-TOF MS by Voyager-DE™ STR (Applied Biosystems, Foster City, CA) using α-cyano-4-hydroxycinnamic acid as matrix.

Isokinetic ratio determination. Peptides were synthesized on TentaGel XV RAM resin (Rapp Polymere, GmbH) on 0.05 mmol scale using standard Fmoc-solid-phase peptide synthesis (SPPS). The isokinetic ratios for each of the five glycan positions denoted by "X" is summarized in Table 3. At the desired site of glycosylation, a mixture of Fmoc- Ser/Thr(Tn)-OPfp and Fmoc-Ser/Thr(tBu)-OH (2 equiv, in different ratios) was coupled manually in the presence of HOBT (2 equiv) and DIPEA (pH 8), and coupling was allowed to proceed for 16 h. The (glyco)peptide mixture was cleaved from the resin using TFA/thioanisole/water mixture (95:2.5:2.5 v/v/v) for 2 h and precipitated with cold MTBE. The precipitate was dissolved in water and the ratio of glycosylated *vs* non-glycosylated product was determined by RP-HPLC. The identity of the isolated products was confirmed by MALDI-TOF MS. Different ratio of Fmoc-Ser/Thr(Tn)-OPfp and Fmoc-Ser/Thr(tBu)-OH were explored until equimolar incorporation of the two components at "X" position was achieved.

On-resin deacetylation. A total of 5 basic solutions were tested for on-resin removal of acetyl group: (1) NaOMe (10 mM and 50 mM) in DMF/MeOH (17:3,v/v); (2) hydrazine (70 mM) in MeOH; (3) sodium methoxide (10 mM) in methanol; (4) NaOH (10 mM) in MeOH; and (5) NH_3 (7 M) in MeOH. The acetylated glycopeptide resin (100 mg) was treated with basic solution (10 mL) for 3 h. The resin was washed with methanol and DCM (3x). The deacetylated glycopeptides were cleaved from the resin using TFA/thioanisole/water mixture (95:2.5:2.5) for 2 h and precipitated by adding cold MTBE. The crude product was dissolved in water and lyophilized. The progress of deacetylation reaction was monitored by RP-HPLC and product fractions were analyzed by MALDI-TOF MS.

Synthesis of MUC1-Tn Glycopeptide Library. For the synthesis of the positional scanning combinatorial library, the peptides were synthesized on TentaGel XV RAM resin (Rapp Polymere, GmbH) on 0.43 mmol scale (resin substitution 0.27 mmol/g) using the "tea bag" approach.[60] The coupling reaction conditions for the synthesis of MUC1-Tn sublibraries and controls (total of 16 bags) were the same as described above for the individual glycopeptide synthesis except that all couplings were done manually. The randomized position "X" contained mixture of the non-glycosylated and glycosylated amino acid in the corresponding isokinetic ratios that were determined and optimized for competitive amino acid coupling (Table 3). After completion of the peptide chain, on-resin deacetylation was performed by using NH_3 (7 M) in methanol (10 mL) for 3 h. Each bag was then subjected to a TFA/thioanisole/water mixture (95:2.5:2.5) for 5 h, followed by precipitation in cold MTBE. The crude product was dissolved in water and lyophilized as powder for use in screening assays.

Screening of MUC1-Tn library. The library was screened against biotin-SBA and biotin-VVA using an enzyme-linked lectin assay (ELLA), as described previously.[52,53] A high-binding 96-well enzyme-linked immunosorbent assay (ELISA) plates (Greiner Bio-One) were coated with the glycopeptide sublibraries and controls (50 μL per well) at four different concentrations (0.1, 0.5, 10, and 500 μg/mL) in phosphate-buffered saline (PBS) (0.01 M, pH 7.4). Wells were incubated overnight to dry at 37°C, and then blocked with 3% bovine serum albumin (BSA) in PBS (300 μL) overnight on the shaker. The coated wells were incubated with either biotin-conjugated SBA or VVA lectin (100 μL, 50 μg/mL in PBS) for 2 h at room temperature on the shaker. Upon washing with PBS (2x, 100 μL), wells were treated with horseradish peroxidase (HRP) conjugated streptavidin (100 μL, 1:4000) for 1 h at room temperature on the shaker. The wells were then washed again with PBS (2x, 100 μL) and water (300 μL), upon which 3,3,5,5-tetramethylbenzidine (TMB) was added (100 μL) and incubated at room temperature for 15 min on the shaker. The reaction was terminated using 2 M sulfuric acid (100 μL). The absorbance readings were recorded

at 450 nm using an ELISA plate reader (BioTek, EPOCH plate reader). The average absorbance reading after background subtraction was plotted against glycopeptide library concentration.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Copies of NMR and HRMS spectra of **7-10**, RP-HPLC and MALDI-TOF spectra of library controls, **21**–**26** (synthesized by the "tea bag" approach), **27**–**35** (synthesized by standard automated SPPS), sublibrary examples, isokinetic ratio and on-resin deacetylation studies, and additional ELLA graphs with biotin-SBA and biotin-VVA lectins.

AUTHOR INFORMATION

Corresponding Author

* Maré Cudic: Department of Chemistry & Biochemistry, Florida Atlantic University, Boca Raton, FL 33431; mcudic@fau.edu; Tel. (561) 297-4645.

Present Addresses

Present Address: University of Georgia, Department of Biochemistry and Molecular Biology, Complex Carbohydrate Research Center, 120 E Green St, Athens, GA 30602, United States

Author Contributions

Y.S. and M.C.R. are the lead contributors to this work. M.A., D.B., R.A., E.P., W.S.M. and A.W. contributed in some parts of the project. M.C. design and supervised the research and wrote the paper.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health Grant CA242351 to M. C.

HRMS work supported by NIH S10 OD021758-01A1 (University of Florida).

REFERENCES

1.Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol 13*, 448-462.

2.Schwarz, F.; Aebi, M. (2011) Mechanisms and principles of *N*-linked protein glycosylation. *Curr Opin Struct Biol 21*, 576-582.

3.Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. (1998) Concepts and Principles of *O*-linked Glycosylation. *Crit Rev Biochem Mol Biol 33*, 151-208.

4.Varki, A. (2017) Biological roles of glycans. *Glycobiology 27*, 3-49.

5.Nwosu, C. C.; Seipert, R. R.; Strum, J. S.; Hua, S. S.; An, H. J.; Zivkovic, A. M.; German, B. J.; Lebrilla, C. B. (2011) Simultaneous and extensive site-specific N- and O-glycosylation analysis in protein mixtures. *J Proteome Res 10*, 2612-2624.

6.Fuchs, S. (2012) Chemically Modified Tandem Repeats in Proteins: Natural Combinatorial Peptide Libraries. *ACS Chem Biol 8*, 275-282.

7.Gabius, H. J., *The sugar code: Fundamentals of glycosciences*. Wiley-VCH: Weinheim, 2009.

8.RodrÍguez, E.; Schetters, S. T. T.; van Kooyk, Y. (2018) The tumour glyco-code as a novel immune checkpoint for immunotherapy. *Nature Rev Immunol 18*, 204.

9.Hang, H. C.; Bertozzi, C. R. (2005) The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem 13*, 5021-5034.

10.Corfield, A. P. (2015) Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochim Biophys Acta (BBA) - General Subjects 1850*, 236-252.

11.Tarp, M. A.; Clausen, H. (2008) Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochim Biophys Acta (BBA) - General Subjects 1780*, 546-563.

12.Hakomori, S. (2002) Glycosylation defining cancer malignancy: New wine in an old bottle. *Proc Nat Acad Sci U. S. A. 99*, 10231-10233.

13.Hollingsworth, M. A.; Swanson, B. J. (2004) Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer 4*, 45-60.

14.Ju, T.; Cummings, R. D. (2002) A unique molecular chaperone Cosmc required for activity of the mammalian core 1 β3-galactosyltransferase. *Proc Nat Acad Sci U.S.A 99*, 16613-16618.

15.Dalziel, M.; Whitehouse, C.; McFarlane, I.; Brockhausen, I.; Gschmeissner, S.; Schwientek, T.; Clausen, H.; Burchell, J. M.; Taylor-Papadimitriou, J. (2001) The Relative Activities of the C2GnT1 and ST3Gal-I Glycosyltransferases Determine *O*-Glycan Structure and Expression of a Tumor-associated Epitope on MUC1. *J Biol Chem 276*, 11007-11015.

16.Campbell, B. J.; Finnie, I. A.; Hounsell, E. F.; Rhodes, J. M. (1995) Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin. *J Clin Invest 95*, 571-6.

17.Cazet, A.; Julien, S.; Bobowski, M.; Burchell, J.; Delannoy, P. (2010) Tumour-associated carbohydrate antigens in breast cancer. *Breast Cancer Res. 12*, 204.

18.Yu, L.-G.; Andrews, N.; Zhao, Q.; McKean, D.; Williams, J. F.; Connor, L. J.; Gerasimenko, O. V.; Hilkens, J.; Hirabayashi, J.; Kasai, K.; Rhodes, J. M. (2007) Galectin-3 interaction with Thomsen-Friedenreich disaccharide on cancer-associated MUC1 causes increased cancer cell endothelial adhesion. *J Biol Chem 282*, 773-781.

19.van Vliet, S. J.; van Liempt, E.; Geijtenbeek, T. B. H.; van Kooyk, Y. (2006) Differential regulation of C-type lectin expression on tolerogenic dendritic cell subsets. *Immunobiology 211*, 577-585.

20.Van Vliet, S. J.; Paessens, L. C.; Broks-van den Berg, V. C. M.; Geijtenbeek, T. B. H.; van Kooyk, Y. (2008) The C-Type Lectin Macrophage Galactose-Type Lectin Impedes Migration of Immature APCs. *J Immunol 181*, 3148-3155.

21.Westerlind, U.; Schröder, H.; Hobel, A.; Gaidzik, N.; Kaiser, A.; Niemeyer, C. M.; Schmitt, E.; Waldmann, H.; Kunz, H. (2009) Tumor-Associated MUC1 Tandem-Repeat Glycopeptide Microarrays to Evaluate Serum– and Monoclonal–Antibody Specificity. *Angew Chem Int Ed 48*, 8263-8267.

22.Oyelaran, O.; Li, Q.; Farnsworth, D.; Gildersleeve, J. C. (2009) Microarrays with Varying Carbohydrate Density Reveal Distinct Subpopulations of Serum Antibodies. *J Proteome Res 8*, 3529-3538.

23.Blixt, O.; Cló, E.; Nudelman, A. S.; Sørensen, K. K.; Clausen, T.; Wandall, H. H.; Livingston, P. O.; Clausen, H.; Jensen, K. J. (2010) A high-throughput O-glycopeptide discovery platform for seromic profiling. *J Proteome Res 9*, 5250-5261.

24.Ohyabu, N.; Hinou, H.; Matsushita, T.; Izumi, R.; Shimizu, H.; Kawamoto, K.; Numata, Y.; Togame, H.; Takemoto, H.; Kondo, H.; Nishimura, S.-I. (2009) An Essential Epitope of Anti-MUC1 Monoclonal Antibody KL-6 Revealed by Focused Glycopeptide Library. *J Am Chem Soc 131*, 17102-17109.

25.Rangappa, S.; Artigas, G.; Miyoshi, R.; Yokoi, Y.; Hayakawa, S.; Garcia-Martin, F.; Hinou, H.; Nishimura, S.-I. (2016) Effects of the multiple *O*-glycosylation states on antibody recognition of the immunodominant motif in MUC1 extracellular tandem repeats. *MedChemComm 7*, 1102-1122.

26.Ying, L.; Liu, R.; Zhang, J.; Lam, K.; Lebrilla, C. B.; Gervay-Hague, J. (2005) A topologically segregated one-bead-one-compound combinatorial glycopeptide library for identification of lectin ligands. *J Comb Chem 7*, 372-384.

27.Kracun, S. K.; Cló, E.; Clausen, H.; Levery, S. B.; Jensen, K. J.; Blixt, O. (2010) Random glycopeptide bead libraries for seromic biomarker discovery. *J Proteome Res 9*, 6705-6714.

28.André, S.; Maljaars, C. E. P.; Halkes, K. M.; Gabius, H.-J.; Kamerling, J. P. (2007) Discovery of galectin ligands in fully randomized combinatorial one-bead-one-compound (glyco)peptide libraries. *Bioorg Med Chem Lett 17*, 793-798.

29.Pinilla, C.; Appel, J. R.; Borras, E.; Houghten, R. A. (2003) Advances in the use of synthetic combinatorial chemistry: Mixture-based libraries. *Nat Med 9*, 118-122.

30.Houghten, R. A.; Ganno, M. L.; McLaughlin, J. P.; Dooley, C. T.; Eans, S. O.; Santos, R. G.; LaVoi, T.; Nefzi, A.; Welmaker, G.; Giulianotti, M. A.; Toll, L. (2016) Direct Phenotypic Screening in Mice: Identification of Individual, Novel Antinociceptive Compounds from a Library of 734 821 Pyrrolidine Bis-piperazines. *ACS Comb Sci 18*, 51-64.

31.Schmidt, R. (1986) New methods for the synthesis of glycosides and oligosaccharides - Are there alternatives to the Koenigs-Knorr method? *Angew Chem Int Ed 25*, 212-235.

32.Cudic, M.; Burstein, G. D., Preparation of Glycosylated Amino Acids Suitable for Fmoc Solid-Phase Assembly. 2008; Vol. 494, pp 187-208.

33.Rodriguez, M. C.; Yegorova, S.; Pitteloud, J.-P.; Chavaroche, A. E.; André, S.; Ardá, A.; Minond, D.; Jiménez-Barbero, J.; Gabius, H.-J.; Cudic, M. (2015) Thermodynamic Switch in Binding of Adhesion/Growth Regulatory Human Galectin-3 to Tumor-Associated TF Antigen (CD176) and MUC1 Glycopeptides. *Biochemistry 54*, 4462-4474.

34.Garegg, P. J., Thioglycosides as Glycosyl Donors in Oligosaccharide Synthesis. In *Advances in Carbohydrate Chemistry and Biochemistry*, Horton, D., Ed. Academic Press: 1997; Vol. 52, pp 179-205.

35.Cudic, M.; Ertl, H. C. J.; Otvos Jr, L. (2002) Synthesis, conformation and T-helper cell stimulation of an *O*-linked glycopeptide epitope containing extended carbohydrate side-chains. *Bioorg Med Chem Lett 10*, 3859-3870.

36.Paulsen, H.; Peters, S.; Bielfeldt, T.; Meldal, M.; Bock, K. (1995) Synthesis of the glycosyl amino acids *N*-Fmoc-Ser[Ac_4-b-D-Galp-(1-3)-Ac_2-a-D-GalN_3p]-OPfp and *N*-Fmoc-Thr[Ac_4-b-D-Galp-(1-3)-Ac_2-a-D-GalN_3p]-OPfp and the application in the solid-phase peptide synthesis of

multiply glyosylated mucin peptides with Tn and T antigentc structures. *Carbohydr Res 268*, 17-34.

37.Plattner, C.; Hofener, M.; Sewald, N. (2011) One-pot azidochlorination of glycals. *Org Lett 13*, 545-547.

38.Dooley, C. T., Houghten, R. A., Synthesis and screening of positional scanning combinatorial libraries. In *Methods in Mollecular Biology*, Totowa, New Jersey, 1998; Vol. 87, pp 13-24.

39.Milton, R. C. d. L.; Milton, S. C. F.; Adams, P. A. (1990) Prediction of difficult sequences in solid-phase peptide synthesis. *J Am Chem Soc 112*, 6039-6046.

40.Acharya, A. N.; Ostresh, J. M.; Houghten, R. A. (2002) Determination of isokinetic ratios neccessary for equimolar incorporation of carboxylic acids in the solid-phase synthesis of mixture-based combinatorial libraries. *Biopolymers 65*, 32-39.

41.Giulianotti, M. A.; Debevec, G.; Santos, R. G.; Maida, L. E.; Chen, W.; Ou, L.; Yu, Y.; Dooley, C. T.; Houghten, R. A. (2012) A novel method for the determination of isokinetic ratios and its application in the synthesis of two new positional scanning libraries. *ACS Comb Sci 14*, 503-512.

42.Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. (1991) Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature 354*, 84-86.

43.Dooley, C. T.; Houghten, R. A. (1997) Synthesis and Screening of Positional Scanning Combinatorial Libraries. *Methods Mol Biol 87*, 13-24.

44.Ljungdahl, T.; Veide-Vilg, J.; Wallner, F.; Tamás, M. J.; Grøtli, M. (2010) Positional Scanning Peptide Libraries for Kinase Substrate Specificity Determinations: Straightforward and Reproducible Synthesis Using Pentafluorophenyl Esters. *J Comb Chem 12*, 733-742.

45.Atherton, E.; Cameron, L. R.; Sheppard, R. C. (1988) Peptide synthesis: Part 10. Use of pentafluorophenyl esters of fluorenylmethoxycarbonylamino acids in solid phase peptide synthesis. *Tetrahedron 44*, 843-857.

46.Ostresh, J. M.; Winkle, J. H.; Hamashin, V. T.; Houghten, R. A. (1994) Peptide libraries: determination of relative reaction rates of protected amino acids in competitive couplings. *Biopolymers 34*, 1681-1689.

47.Sjölin, P.; Elofsson, M.; Kihlberg, J. (1996) Removal of Acyl Protective Groups from Glycopeptides: Base Does Not Epimerize Peptide Stereocenters, and β-Elimination Is Slow. *J Org Chem 61*, 560-565.

48.Liu, M.; Barany, G.; Live, D. (2005) Parallel solid-phase synthesis of mucin-like glycopeptides. *Carbohydr Res 340*, 2111-2122.

49.Compañón, I.; Guerreiro, A.; Mangini, V.; Castro-López, J.; Escudero-Casao, M.; Avenoza, A.; Busto, J. H.; Castillón, S.; Jiménez-Barbero, J.; Asensio, J. L.; Jiménez-Osés, G.; Boutureira, O.; Peregrina, J. M.; Hurtado-Guerrero, R.; Fiammengo, R.; Bernardes, G. J. L.; Corzana, F. (2019) Structure-Based Design of Potent Tumor-Associated Antigens: Modulation of Peptide Presentation by Single-Atom O/S or O/Se Substitutions at the Glycosidic Linkage. *J Am Chem Soc 141*, 4063-4072.

50.Thompson, P.; Lakshminarayanan, V.; Supekar, N. T.; Bradley, J. M.; Cohen, P. A.; Wolfert, M. A.; Gendler, S. J.; Boons, G.-J. (2015) Linear synthesis and immunological properties of a fully synthetic vaccine candidate containing a sialylated MUC1 glycopeptide. *Chem Comm 51*, 10214-10217.

51.Grøtli, M.; Gotfredsen, C. H.; Rademann, J.; Buchardt, J.; Clark, A. J.; Duus, J. Ø.; Meldal, M. (2000) Physical Properties of Poly(ethylene glycol) (PEG)-Based Resins for Combinatorial Solid Phase Organic Chemistry: A Comparison of PEG-Cross-Linked and PEG-Grafted Resins. *J Comb Chem 2*, 108-119.

52.Jakas, A.; Perc, M.; Sućˊ, J.; Rodriguez, M. C.; Cudic, M.; Cudic, P. (2016) Synthesis of anthrose lipidic derivative as mimic of B. anthracis BclA glycoprotein for use in ELISA-like binding assays. *J Carbohydr Chem 35*, 69-85.

53.Rodriguez, M. C.; Yongye, A. B.; Cudic, M.; Martinez Mayorga, K.; Liu, E.; Mueller, B. M.; Ainsley, J.; Karabencheva-Christova, T.; Christov, C. Z.; Cudic, M.; Cudic, P. (2017) Targeting cancer-specific glycans by cyclic peptide lectinomimics. *Amino Acids 49*, 1867-1883.

54.Lagarda-Diaz, I.; Guzman-Partida, A. M.; Vazquez-Moreno, L. (2017) Legume Lectins: Proteins with Diverse Applications. *Int J Mol Sci 18*, 1242.

55.Piller, V.; Piller, F.; Cartron, J.-P. (1990) Comparison of the carbohydrate-binding specificities of seven N-acetyl-D-galactosamine-recognizing lectins. *Eur J Biochem 191*, 461-466.

56.Houghten, R. A.; Pinilla, C.; Appel, J. R.; Blondelle, S. E.; Dooley, C. T.; Eichler, J.; Nefzi, A.; Ostresh, J. M. (1999) Mixture-based synthetic combinatorial libraries. *J Med Chem 42*, 3743-3778.

57.Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; Coelho, H.; Valero-González, J.; Castro-López, J.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Corzana, F.; Peregrina, J. M. (2015) Detection of tumor-associated glycopeptides by lectins: the peptide context modulates carbohydrate recognition. *ACS Chem Biol 10*, 747-756.

58.Komba, S.; Terauchi, T.; Machida, S. (2009) Automated Synthesis of a Tri-branched Pentasaccharide: the Application of the Uni-chemo Hydroxyl Protection Method to the Automated Synthesis of Oligosaccharides. *J Appl Glycoscience 56*, 193-206.

59.Rodriguez, M. C.; Yegorova, S.; Pitteloud, J.-P.; Chavaroche, A. E.; Andre, S.; Arda, A.; Minond, D.; Jimenez-Barbero, J.; Gabius, H.-J.; Cudic, M. (2015) Thermodynamic switch in binding of adhesion/growth regulatory human galectin-3 to tumor-associated TF antigen (CD176) and MUC1 glycopeptides. *Biochemistry 54*, 4462−4474.

60.Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc Natl Acad Sci U. S. A. 82*, 5131-5135.

For Table of Contents Only