Published in final edited form as:

J Proteome Res. 2010 December 3; 9(12): 6705–6714. doi:10.1021/pr1008477.

Random Glycopeptide Bead Libraries for Seromic Biomarker Discovery

Stjepan K. Kračun¹, Emilano Cló¹, Henrik Clausen¹, Steven B. Levery¹, Knud J. Jensen², and Ola Blixt^{1,*}

¹Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3b, DK-2200, Copenhagen N, Denmark

²Department of Basic Sciences and Environment/Bioorganic Chemistry, University of Copenhagen, Thorvaldsensvej 40, DK-1871, Frederiksberg, Denmark

Abstract

Identification of disease specific biomarkers is important to address early diagnosis and management of disease. Aberrant post-translational modifications (PTM) of proteins such as O-glycosylations (O-PTMs) are emerging as triggers of autoantibodies that can serve as sensitive biomarkers. Here we have developed a random glycopeptide bead library screening platform for detection of autoantibodies and other binding proteins. Libraries were built on biocompatible PEGA beads including a safety-catch C-terminal amide linker (SCAL) that allowed mild cleavage conditions (I₂/NaBH₄ and TFA) for release of glycopeptides and sequence determination by ESI-MSⁿ. As proof-of principle, tumor specific glycopeptide reporter epitopes were built-in into the libraries and were detected by tumor specific monoclonal antibodies and autoantibodies from cancer patients. Sequenced and identified glycopeptides were re-synthesized at preparative scale by automated parallel peptide synthesis and printed on microarrays for validation and broader analysis with larger sets of sera. We further showed that chemical synthesis of the monosaccharide O-glycopeptide library (Tn-glycoform) could be diversified to other tumor glycoforms by on-bead enzymatic glycosylation reactions with recombinant glycosyltransferases. Hence, we have developed a high-throughput flexible platform for rapid biomarker discovery O-glycopeptides and the method has applicability in other types of assays like lectin/antibody/enzyme specificity studies as well as investigation of other PTMs.

Keywords

glycopeptide; post-translational modification (PTM); one-bead-one-compound (OBOC); split-mix; microarray; O-glycosylation; autoantibodies; enzymatic

Introduction

Glycosylation is one of the most abundant posttranslational modifications (PTMs) of proteins and is involved in many important physiological processes like recognition, adherence, motility, and signaling processes^{1, 2}. In cancer, aberrantly modified O-glycoproteins are able to evoke host immune responses³ and the potential for identification of new biomarkers in this area is a promising and important complement to peptide and

Corresponding author: Ola Blixt olablixt@sund.ku.dk.

Supporting Information Available: This material is available free of charge via the Internet at <http://pubs.acs.org>

S1 Data from additional experiments

S2 Mass spectrometry data for all sequences

protein biomarkers⁴⁻⁶. While enormous efforts are now devoted to proteomics, one of the next “-omics”, glycomics, is a rapidly emerging field advanced by new synthetic and analytical developments⁷.

O-glycosylation in disease, particularly in cancer, is frequently truncated and aberrantly presented as Tn (GalNAc α 1-O-Ser/Thr), T (Gal β 1-3GalNAc α 1-O-Ser/Thr), and STn (NeuAc α 2-6GalNAc α 1-O-Ser/Thr), so called tumor-associated antigens (TAAs)⁸. These structures may also be dislocated at altered densities on the carrier protein and expose novel immunogenic neoepitopes⁹. The natural human repertoire of anti-carbohydrate antibodies is substantial¹⁰ but carbohydrates are mostly recognized by natural IgM antibodies¹¹. However, the truncated O-glycans in combination with an exposed neo-peptide epitope can induce IgG antibodies and may lead to identification of promising biomarker candidates^{12, 13}. Such combined O-glycopeptide epitopes have been characterized for a number of mouse monoclonal antibodies, including the epitope for an autoantibody in a spontaneous mouse tumor model¹⁴, as well as to human tumor auto-antibodies^{15, 16}. It is therefore important to develop technologies for production and display of aberrant glycoproteins and/or glycopeptides for high throughput screening strategies of autoantibodies.

We have previously established high-throughput O-glycopeptide microarray screening strategies using chemical and enzymatic solid-phase parallel peptide synthesis¹⁵. One-bead-one compound (OBOC) combinatorial libraries of small molecules like peptides are powerful tools that can be used to synthesize vast numbers of compounds to screen for binding proteins and antibodies¹⁷. Here we further developed the OBOC method to include O-glycans generating random O-PTM bead libraries for serological screening. Our proof of concept library was designed around the tumor-associated glycopeptide epitope Tn-MUC1 for detecting tumor specific monoclonal antibodies and autoantibodies in cancer patient sera. Mass spectrometric sequence analysis of bead-selected and released glycopeptides yielded the sequence autoantibody targeted antigens, which were re-synthesized and validated on our O-glycopeptide microarray platform¹⁵.

Materials and Methods

Materials

The Fmoc-protected amino acids (apart from GalNAc α 1-threonine and GalNAc α 1-serine), MeOH, NMP, DMF, piperidine, DIEA, TFA, HBTU, HOBt, spacers ({2-[2-(Fmoc-amino)ethoxy]ethoxy}acetic acid and N1-(9-Fluorenylmethoxycarbonyl)-1,13-diamino-4,7,10-trioxatridecan-succinamic acid) were from Iris Biotech (Marktredwitz, Germany). Fmoc-GalNAc α 1-threonine and Fmoc-GalNAc α 1-serine (N^α-Fmoc-O- β -(2-acetamido-2-deoxy-3,4,6-tri-O-acetyl- α -D-galactopyranosyl)-L-threonine and N^α-Fmoc-O- β -(2-acetamido-2-deoxy-3,4,6-tri-O-acetyl- α -D-galactopyranosyl)-L-serine) were from Sussex Research (OT, Canada). The resin, PL-PEGA (300-500 μ m, 0.2 mmol/g loading) was from Varian (Palo Alto, CA, USA). DCM, Et₂O, Ac₂O, THF, formic acid, the SCAL linker (4,4'-Bis(methylsulfinyl)-2-(4-carboxybutoxy)-N-Fmoc-benzhydrylamine), 0.5M solution of NaOMe in MeOH, Goat anti-mouse-IgG (Fc-specific) alkaline phosphatase conjugate, goat anti-mouse-IgG (Fc-specific) Cy-3 conjugate, goat anti-human-IgG (Fc-specific) Cy-3 conjugate, the biotinylated HPA lectin, CHAPS, BSA, NaBH₄ and UDP-GlcNAc were from Sigma (MO, USA). The Zymax Streptavidin-Cy-3 conjugate was from Invitrogen (Carlsbad, CA, USA). The BCIP/NBT ready-to-use substrate solution was from KemEnTec Diagnostics (Taastrup, Denmark). Printing was performed on Schott Nexterion® Slide H or Schott Nexterion® Slide H MPX 16 (Schott AG, Mainz, Germany). All salts for all the buffers, including TES, Triton-X-100 Tween 20, iodine and ethanolamine were from Merck (NJ, USA). The monoclonal 5E5 antibody was produced as described from a wild-type Balb/c mouse immunized with a fully GalNAc-glycosylated Tn-MUC1 60mer

glycopeptide coupled to KLH18. The 1E10 and 5F7 monoclonal antibodies were produced as described¹⁵ whereas the VU3C2 mAb were from Chemicon, Millipore, MA19.

Buffer solutions and reagents

Antibodies were diluted in the staining buffer (0.5M NaCl, 3mM KCl, 1.5mM KH₂PO₄, 6.5mM Na₂HPO₄, 1% BSA, 1% Triton-X-100, pH=7.4). Washings after staining were done with PBST. The reduction cocktail was a solution containing 6.6mM NaBH₄ and 6mM I₂ in THF. The cleavage cocktail contained 95% TFA, 3% water and 2% TES. Blocking of the microarray slides was performed with a blocking buffer (50mM ethanolamine, 0.05M Na₂B₄O₇, pH=8.5). Compounds for printing were dissolved in the printing buffer (133mM Na₂HPO₄, 17mM NaH₂PO₄, 0.005% CHAPS, 0.03% NaN₃, pH=8.5).

Combinatorial OBOC solid phase glycopeptide library synthesis

The combinatorial part of the synthesis was performed in a custom-made teflon cylinder reactor with 20 wells²⁰. The rest of the synthesis was performed in polypropylene syringes fitted with polyethylene filters. The synthesis of the library was performed using standard Fmoc-SPPS methodology^{15, 21} on 1.5 g of PL-PEGA resin (dry resin mass). Based on the fact that the product of a split-and-mix synthesis approach is a Poisson distribution of products²², we used a 5-fold excess of beads with respect to the number of compounds to be synthesized. Before synthesis, the resin was swollen in DCM. All steps were performed at room temperature with shaking. The coupling of the SCAL-linker, the spacer, the first 3 C-terminal alanines and the last 3 N-terminal alanines was performed in polypropylene syringes fitted with polyethylene filters. The rest of the synthesis was performed in the Teflon reactor. Side-chain protecting groups were *tert*-butyl (Ser, Thr, Tyr), 2,2,4,6,7-pentamethyl-dihydrobenzofuran-5-sulfonyl (Pbf, for Arg), and trityl (Trt, for Asn, Gln, His). The Fmoc-protected amino acids, Fmoc-glyco-amino acids, linker and spacer were incorporated in NMP with 4 equivalents of Fmoc-amino acid (with respect to the resin loading), 3.6 equivalents of HBTU, 4 equivalents of HOBt and 7.2 equivalents of DIEA, with a 10-minute pre-activation step. Coupling times were 2 hours in syringes and 3 hours in the Teflon reactor. Fmoc-deprotection was performed using piperidine-DMF (1:4). The deprotection time in syringes was 30 minutes and 1 hour in the reactor. N-terminal capping was performed after the 1st, 4th and last coupling with 50% Ac₂O in DCM treatment for 15 minutes for the first library and after every second coupling for the second library. The final product was not N-acetylated. Washings in-between steps were done with DCM and NMP. The split-and-mix step was accomplished by flooding the reactor with DCM where the resin would rise from the wells and float on top of the solvent where it was stirred and mixed. The DCM was then withdrawn by vacuum-suction and the resin would sink back into the wells. When the synthesis was complete, side-chain deprotection was performed by incubating the resin in 95% TFA, 3% water and 2% TES for 2 hours. Then the de-acetylation of the carbohydrate moiety on the glycopeptide was performed by incubating the resin in a 0.5M solution of NaOMe in MeOH for 2 hours. Finally the resin was washed with DCM, NMP and finally MeOH and was kept as a MeOH slurry at 4 degrees centigrade.

Synthesis of MUC1-20mer and Tn-MUC1-20mer

The synthesis of the peptide and the glycopeptide (VTSAPDTRPAPGSTAPPAHG and VTSAPDTRPAPGSTAPPAHG respectively) was performed using standard Fmoc-SPPS methodology on 300 mg of PL-PEGA resin (dry resin mass). As with the library, SCAL was used as a linker. Before synthesis, the resin was swollen in DCM. All steps were performed at room temperature with shaking. The Fmoc-protected amino acids, Fmoc-glyco-amino acids, linker and spacer were incorporated in NMP with 4 equivalents of Fmoc-amino acid (with respect to the resin loading), 3.6 equivalents of HBTU, 4 equivalents of HOBt and 7.2 equivalents of DIEA, with a 10-minute pre-activation step. Coupling time was 1 hour.

Fmoc-deprotection was done using piperidine-DMF (1:4). Deprotection time was 15 minutes. N-terminal capping was performed after the 1st, 6th, 11th, 15th and last coupling with 50% Ac₂O in DCM treatment for 15 minutes. The final product was not N-acetylated. Washings in-between steps were done with DCM and NMP. When the synthesis was complete, side-chain deprotection was performed by incubating the resin in 95% TFA, 3% water and 2% TES for 2 hours. De-acetylation of the carbohydrate on the Tn-MUC1-20mer was performed by incubating the resin in a 0.5M solution of NaOMe in MeOH for 2 hours. Finally the resin was washed with DCM, NMP and MeOH and was kept as a MeOH slurry at 4 degrees centigrade.

Synthesis of GST_{XX}PP Tn-MUC1-20-mer variants and Tn-MUC1-20-mer variants with alanine-walk mutations and different glycosylation patterns

The synthesis of the VTSAPDTRPAGSTAPPAHG variants with spacers ({2-[2-(Fmoc-amino)ethoxy]ethoxy}acetic acid) and (N1-(9-Fmoc)-1,13-diamino-4,7,10-trioxatridecan-succinamic acid) together with the alanine-mutation walk peptides and T_{XX}P mutation peptides was performed using an automated peptide synthesizer as described [15]. The variants with different spacers were synthesized on the PEGA resin and the other peptides were synthesized on the Tentagel resin. The peptides were prepared by automated peptide synthesis on a Syro II peptide synthesizer (MultiSynTech) by standard SPPS on TentaGel S Rink Amide and PEGA resins with Fmoc for protection of N^α-amino groups. N^α-Fmoc amino acids (4.0 equiv) were coupled using HBTU (3.8 equiv), HOBt (4.0 equiv) and N,N-DIEA (8.0 equiv) as coupling agents in DMF for 45 min, except the non-essential amino acids which reacted for 120 min. N^α-Fmoc deprotection was performed using piperidine-DMF (2 : 3) for 3 minutes, followed by piperidine-DMF (1 : 4) for 12 min.

General protocol for bead staining

About 2mL of the bead slurry (approximately 20,000 beads, 10,000 beads per mL of slurry) in MeOH were taken and the MeOH was drained. The beads were then incubated in PBST (PBS, 0.5% Tween-20) for 15 minutes prior to staining. The PBST was drained prior to adding the antibody solution. The antibody dilutions for HMFG2, 5E5, VU3C2 and 5F7 were 1 mg/mL. The dilution of the biotinylated HPA and HAA lectins was 1:1000. The dilution for the 1E10 antibody was 1:100. All dilutions were made in the staining buffer. The dilution for sera was 1:25 in staining buffer. The dilutions of the secondary antibodies (goat anti-mouse-IgG-Cy3, goat anti-human-IgG-Cy3 and streptavidin-Cy-3 were 1:1000. The dilution for goat anti-mouse-IgG-AP was 1:500. The dilution of sera was 1:200. All dilutions were made in the staining buffer. Incubation times for antibodies, lectins and streptavidin were 1 hour at room temperature with shaking. All washings were done with PBST. Alkaline phosphatase conjugate stained beads were incubated with the BCIP/NBT ready-to-use substrate and the color developed within 10 minutes and that's when the positive hits were isolated. Beads stained with the secondary antibody-AP-conjugate could be reused if the purple color was removed by repeated washing of the beads with neat TFA, methanol and dichloromethane. Fluorescently labeled beads were examined under a Zeiss fluorescence microscope equipped with 75W xenon lamp and a 50W HBO lamp. Beads were selected and picked out from a plastic Petri dish using a pipette. To investigate the influence of spacer length on staining, 1E10-stained Core-3-MUC1-20mer beads consequently stained with goat anti-mouse-IgG-Cy3 were deposited on a microarray slide and their fluorescence was measured (Supplementary information, Figure S1).

Standard protocol for peptide cleavage

The isolated bead was first washed with MeOH (50 μ L) and then incubated in the reduction cocktail (NaBH₄, 6.6mM and 6mM I₂ in THF, 50 μ L) for 20 minutes. The liquid was removed and the bead was washed with MeOH (50 μ L) and then incubated in cleavage

cocktail (TFA, water and TES 95:3:2%, 5 μ L) for 30 minutes. The liquid containing the cleaved peptide was withdrawn and transferred to another vial, blow-dried and kept at 4 degrees centigrade before MS-analysis.

Mass Spectrometry

Electrospray-ionization mass spectrometry (ESI-MS) was performed on a linear ion trap-Orbitrap hybrid instrument²³ (LTQ-Orbitrap XL, Thermo-Scientific, Bremen, Germany) equipped for multistage fragmentation (MSⁿ) via conventional collision-induced dissociation (CID) higher energy CID (HCD)²⁴ in an external octopole collision cell²⁵ and electron-transfer dissociation (ETD)²⁶ using fluoranthene anion generated in an external chemical ionization (CI) source, with the capability of supplemental activation in the LTQ ion trap²⁷. The instrument was controlled using Thermo LTQ Orbitrap XL Tune Plus 2.5.5 (Thermo Fischer Scientific). Acquired spectra were processed and analyzed using Xcalibur Qual Browser 2.0.7 (Thermo Fischer Scientific). Samples were introduced by direct infusion via a TriVersa NanoMate ESI-Chip interface (Advion BioSystems, Ithaca, NY, USA) controlled by ChipSoft 8.1.0 (Advion Biosciences). All glycopeptide MS¹ and MS² spectra were acquired in positive ion Orbitrap Fourier transform (FT) mode at a nominal resolving power of 30,000. In addition to frequent m/z calibration of the Orbitrap detector according to manufacturer's instructions, a polydimethylcyclsiloxane ion (m/z 445.1200) was used as an internal calibration standard for MS¹ spectra²⁸. Each dried sample was dissolved in 25 μ L of 50% MeOH in water containing 1% of formic acid, and applied to a well of the NanoMate sample plate kept at 10°C; a sample volume of 5 μ L was delivered to the mass spectrometer via the chip interface at a flow rate of ~100 nL/min using nitrogen gas at a pressure of 0.30 psi and an electrospray potential of 1.40 kV. Following acquisition of each MS¹ spectrum, CID-, HCD- and, where necessary, ETD-MS² spectra were acquired on selected glycopeptide precursors with suitable charge states (≥ 2); sodiated as well as protonated precursors were considered for MS² analysis, depending on their relative abundance. Selected glycopeptide precursors were isolated with a width of 3 mass units (mu) for CID and 3-5 mu for HCD, and activated for 30 ms using 35% normalized collision energy for CID and 20-60% for HCD and an activation Q of 0.25 for both. ETD of the selected glycopeptide precursors was performed using an isolation width of 5 mu, an activation time of 150-250 ms, and supplemental activation of 20% normalized collision energy.

In general, a cursory analysis of the residue-specific increments between product ions in HCD spectra, and in ETD spectra where obtained, together with the known limits of the library composition, was sufficient to propose a likely peptide sequence for each glycopeptide precursor. The precise glycopeptide fragment m/z values were then analyzed by comparison with theoretical m/z values for ions (typically b, y, c, c-1, z-, z+1, and z+1; values for multiply charged and sodiated fragment ions were also considered) calculated using Protein Prospector software (<http://prospector.ucsf.edu/>) and an in-house computer program written in VBA for Microsoft Excel 2007. Errors were calculated in ppm by standard mathematical procedure from the differences between calculated and experimental m/z values.

Standard protocol for microarray validation

Printing of the microarray slides was performed using a BioRobotics MicroGrid II spotter (Genomics Solution) using Stealth 3B Micro Spotting Pins with deposit volume of approx. 6 nL of glycopeptide in print buffer (150 mM phosphate, 0.005% CHAPS pH 8.5). The compounds were distributed (20 μ L per well) in 384-well source plates (BD Falcon MicrotestTM 384-well 30 μ L assay plates from BD Biosciences, Le Pont De Claix, France) and printed in 3 replicates 3 replicates using an 8-pin (2 $\times$ 4) configuration within a 28 $\times$ 28

subgrid at a 0.21 mm pitch between each spot. The pin dwell time in the wells was of 4 seconds and the pins underwent 3 wash cycles in between source plate visits. The complete 4x2 array pattern was printed on a 16 well slide in duplicate, distributed in two columns and eight rows. Immediately after printing the slides were incubated at 80% humidity for 60 min. Remaining NHS groups on the slides were blocked by immersion in the blocking buffer (50 mM ethanolamine in 50 mM borate buffer, pH 9.2) for 1 h. Slides were rinsed in Millipore water, dried by centrifuging and probed as described below. Slides that were not to be probed immediately were stored at -18°C before the blocking step. Scanning of the slides was performed on ProScanArray HT Microarray Scanner (PerkinElmer) followed by image analysis with ProScanArray Express 4.0 software (PerkinElmer). Data were analyzed and plotted using Microsoft® Excel or GraphPad Prism software. (Telechem International ArrayIt Division)^{15, 29}. Staining of slides with the antibodies were performed at a 1 $\mu\text{g/mL}$ dilution for 30 minutes at room temperature. The dilution for lectins was 1:1000 and the dilution for sera was 1:20. All dilutions were made in the staining buffer. All washings were done with PBST. Slides were then incubated with the goat anti-mouse-IgG-Cy3, goat anti-human-IgG-Cy3 (Fc-specific) antibody or streptavidin-Cy3 (1:1000 dilution) for 30 minutes at room temperature. Finally the slides were washed with PBST and the liquid was centrifuged off (200 g) before scanning.

General protocol for on-bead enzymatic glycosylation

The beads were taken as a MeOH slurry, the MeOH was drained and the beads were incubated in PBST for 15 minutes. The PBST was then drained and the glycosylation reaction mixture was added. A batch of beads bearing Tn-glycopeptides were incubated in a 25mM cacodylic acid buffer, pH=7.4, with 10mM MnCl_2 and 0.25% Triton-X-100 overnight with shaking at room temperature with 10mM UDP-GlcNAc as a donor sugar and with a recombinant $\beta 3\text{GlcNAc-T6}$ glycosyltransferase³⁰.

General procedure for on-chip enzymatic glycosylation

Slides with immobilized peptides were blocked with ethanolamine (1hr, RT), rinsed thoroughly with milli-Q purified water, and then spun dry on a Galaxy mini-array tabletop slide centrifuge (VWR, West Chester, PA, USA). 16-well superstructures (Schott AG, Mainz, Germany) were applied and slides were treated overnight at 37°C with 50 μL glycosylation mixture for Core-3 glycosylation: 10 μM UDP-GlcNAc, 25 $\mu\text{g/mL}$ $\beta 1-3\text{GlcNAcT}$ enzyme, 10mM MnCl_2 and 0.25% Triton-X-100, 25mM cacodylic acid (pH 7.4). Immediately following glycosylation, slides were washed with PBST (5 min, shaking), PBS (5 min, shaking) and then treated with citrate buffer (pH 2.5) for 15 min, rocking. Following acid wash, slides were again washed with PBST and PBS as before and then blocked with 1% BSA, 0.5% NP40, PBS 20min, shaking. Slides were again washed with PBST, PBS, rinsed thoroughly with MQ, and dried and used in the next step.

Lectin, mAb and serum microarray analysis

Incubation volumes for each MPX16 well were performed with adhesive superstructures at 50 μL /well. Lectins were diluted to 2-10 $\mu\text{g/mL}$, mAbs to 1 $\mu\text{g/mL}$ and human sera were analyzed at 1:20 dilution. All samples were incubated on the slide for 1 μh , followed by 1 h incubation with appropriate secondary antibodies at a 1 $\mu\text{g/mL}$ dilution. All dilutions were made in staining buffer pH 7.4. Murine monoclonal antibodies were detected with Cy3-conjugated goat anti- mouse IgG (H+L) diluted 1:1000. Human IgG antibodies were detected with Cy3-conjugated goat anti-human IgG (Fc specific)(1:1000) and biotinylated lectins detected with streptavidin-Cy3 (1:1000). All incubation steps were separated by two wash steps in PBS with 0.05% Tween-20 (PBST) and one in PBS. After the final wash, slides were rinsed in H_2O , dried by centrifugation (200 g) and scanned followed by image analysis and quantification.

Results and discussion

1st generation random glycopeptide bead library including a monoclonal antibody reporter epitope

We first designed a limited prototype OBOC glycopeptide bead library with a fixed glycosylation site (Tn-Thr) at position (T_0) (Figure 1). By using the split-mix approach³¹ a dodecamer glycopeptide bead library was synthesized with randomized amino acids (A, P, E, F, I, V, S, G, M, Y, C, Q, W) at positions X_{-2} to X_{+3} flanking the GalNAc α 1-O-T residue, T_0 . The amino acids alanine (A) and proline (P) at positions X_{+1} and X_{+2} respectively were specifically selected to cover the tumor specific MUC1 monoclonal antibody 5E5 that requires a site-specific GalNAc α 1-O-threonine (-STAP-) glycosylation site in the tandem repeat for binding (Figure 1)^{15, 32}. Both N- and C-terminal ends of the randomized sequence were flanked with triple-alanine sequences. Standard Fmoc-SPPS chemistries were used²¹ and *N*-acetylation-capping steps (Ac₂O/DCM) were employed after certain steps prior to subsequent coupling to minimize formation of deletion peptides during synthesis¹⁵. The C-terminal end was attached via a spacer (sp) to a cleavable SCAL (safety-catch amide) linker³³ immobilized on PEGA beads. PEGA beads were chosen because they are more porous and hydrophilic, performing better in an aqueous environment than e.g. TentaGel, and are more suitable for screening of biomolecules such as antibodies, lectins and proteins, either in purified form or from biofluids such as sera³⁴. The spacer was employed to avoid potential steric interference of the bulky SCAL moiety with bound antibodies. Two spacers were evaluated, namely ([2-(2-Amino-ethoxy)-ethoxy]-acetic acid) and N1-(9-Fluorenylmethoxycarbonyl)-1,13-diamino-4,7,10-trioxatridecan-succinamic acid); the first one performed better in staining experiments (Supplementary information 1, Figure S1). A library containing 3125 (55) unique sequences on approximately 200,000 beads was synthesized, thus, containing high numbers of theoretical replicates, suitable for statistical accumulation of multiple staining selections for sequencing with mass spectrometry (hybrid ion trap/Orbitrap system).

After completion of the synthesis of the bead-bound glycopeptides, amino acid side-chain protecting groups were removed (TFA-TES-H₂O, 95:2:3 by percentage), and de-acetylation of the GalNAc-moiety was carried out using Zemplén conditions (NaOMe, 0.5M in MeOH). About 200 randomly selected beads were stained with biotinylated GalNAc-specific lectins, HPA (*Helix pomatia* agglutinin) and HAA (*Helix aspersa* agglutinin), to statistically confirm successful synthesis; >98% of the beads stained positively under a fluorescence microscope with Cy3-labeled Streptavidin (data not shown), demonstrating that >98% of the beads contained the fixed GalNAc α 1-O-threonine residue. Portions of the bead library (approx. 20,000 beads) were stained with the 5E5 mAb followed by colorimetric detection with the secondary anti mouse-IgG-conjugated alkaline phosphatase (AP) antibody and BCIP/NBT ready-to-use substrate solution (KemEnTec Diagnostics, Denmark). Colorimetric detection enables fast screening of the bead library without a need for costly instrumentation, as, upon addition of the AP-substrate, the positive beads turned dark purple within 10 minutes and were easily singled out using a pipette in a petri-dish³⁵. The optimal staining conditions at 1:1000 dilution of the secondary antibody positively colored beads were fast enough without significant background staining of the negative beads. In addition, beads could be de-stained by washing with TFA, MeOH and chloroform and re-stained multiple times, enabling use of the same library with different biomolecules (the staining/color removal process has been repeated 3 times without loss in fidelity instaining assays). After colorimetric bead staining, there were altogether about 10% positive hits from a given portion of the library. 20 positive glycopeptide-beads were isolated individually and released from solid support first by chemoselective deoxygenation of SCAL linker sulfoxides to thioethers using I₂/NaBH₄ in THF, followed by hydrolytic cleavage of SCAL with TFA/TES/water³⁶. These cleavage conditions avoided PEGA polymer degradation that

commonly occurs with other more harsh methods such as TFMSA treatment, and helped to maintain high sample quality for downstream mass spectrometric sequencing. The amount of glycopeptide material released from a single isolated bead was estimated to be ~1 nmol (from measured resin loading estimates), which is more than sufficient for MS-sequencing and identification of the target glycopeptides with the hybrid ion-trap/Orbitrap system (Figure 2). After MS-sequencing, there was material left from a single bead to perform a pilot print on the microarray and quickly evaluate the target glycopeptide (data not shown). In the MS¹ spectra, we frequently observed both protonated and sodiated molecular species (see, e.g., Figure 2A), and found both to be useful precursors for fragmentation analysis (discussed further below). In MS² analysis, ordinary CID mode spectra (not shown) exhibited almost exclusively changes of charge state accompanied by abundant deglycosylated peptide products, which was useful for quick confirmation that the precursor was glycosylated, but the sparse production of peptide sequence-related b and y fragments was inadequate for routine peptide sequencing. However, use of HCD mode^{24, 25} yielded nearly complete precursor glycan loss, accompanied by abundant peptide backbone sequence-related fragment ions, as illustrated in Figure 2, Panels B and C. Since the first generation library was designed so that only one GalNAc-ylated residue (**T**) occurs at the same position in each sequence, the use of ETD was unnecessary at this stage, since only the remainder of the peptide backbone sequence is indeterminate. In later versions, which incorporated possibilities for indeterminate glycosylation, the additional use of ETD-MS² became essential (see below).

In contrast to an ordinary ion trap instrument, Fourier-transform mass spectrometric (FT-MS) detection in the Orbitrap results in high resolving power and consequent mass accuracy, enabling resolution and unambiguous assignment of closely spaced fragment ions (see insets, Figure 2B), and hence dramatically improves sequencing of target glycopeptides²³.

Identified sequences from 5E5 staining were exclusively related to the MUC1 glycopeptide epitope containing the GalNAc1-O-T (position T₀) with adjacent proline residues (P) at positions X₊₂ and X₊₃. In addition, several amino acid modifications at position X₊₁ in the GalNAc1-O-TXPP region were identified, such as alanine (A), serine (S), valine (V), and glutamine (Q), but not proline (P), of those selected in the random sequence. We also identified incompletely synthesized deletion peptides terminated with an N-acetylated terminus, but this did not affect their immunogenicity as they contained the epitope required by 5E5. An extended target list is available in Supplementary information 1, Table S1.

To validate identified sequences and further refine the specificity of 5E5 we first chemically synthesized Tn-MUC1 tandem repeat glycopeptides with an alanine-mutation walk through the GST(αGalNAc)APP region in a larger scale with standard Fmoc-SPPS on TentaGel15. The glycopeptide mutants were further selectively immobilized onto N-hydroxysuccinimide-(NHS-) activated glass slides using a microarray contact printer²⁹. Arrayed glycopeptides were stained with 5E5 mAb followed by Cy3 labeled secondary anti-mouse IgG antibody for quantification using a confocal fluorescence microarray scanner. Interestingly, binding of 5E5 to the random bead library was only absent if either of the residues of (T₀) T(αGalNAc) or P (X₊₂) were mutated (Figure 3B), thus confining the minimal 5E5 binding epitope to **TXP**. Next, the native MUC1-20mer tandem repeat glycopeptides with 18 different point mutations at (X₊₁) were synthesized. As illustrated in Figure 3C, glycopeptides with amino acid mutations Q, S and V were bound by 5E5 but minimally to the P-mutation, which confirms high selectivity during bead staining and selection. Several additional amino acid mutations in the X₊₁ position not covered within the random sequences of the library were also tolerated by the 5E5 mAb (Supplementary information 1,

Figure S2.)^{15, 18}. These results clearly demonstrate that our O-PTG bead selection and re-synthesis approach are useful to specifically detect and identify selective antibody binding to O-glycopeptide epitopes. The data collected for the 5E5 epitope also shows that any possible truncations of random glycopeptides that may occur during OBOC combinatorial library synthesis (like N-Ac terminated peptides and deletion peptides) do not interfere with the quality of collected data, because as long as there is a sufficient excess of beads with respect to the number of compounds that are synthesized and as long as the epitope is synthetically built, regardless of any possible truncations in its proximity, the relevant epitopes will appear in a sufficient amount to be detected with this analytical approach.

2nd generation random glycopeptide bead library to specifically detect autoantibody epitopes in vaccinated patient sera

To further investigate the analytical competence of this platform in more complex biological samples, specifically, autoantibody reactivity of cancer sera to MUC1 peptide glycoforms, we designed a second library with random amino acid selections that covers in part autoantibody epitopes of the MUC1 TR (Figure 4A)¹⁵. Besides the selected amino acids, (S, A, P, D, R, G) at positions X₋₂ to X₊₃, a GalNAc α 1-O-serine was included as a random amino acid to allow for multiple glycosylation sites, and a library containing 16,807 unique compounds (75) were formed. By replacing the C-terminal triple alanine flanking sequence with the AHGV sequence, both a 10mer MUC1 TR peptide GSTAPPAHGV and a 7-mer APDTRPA sequence can be generated. These two epitopes on the MUC1 TR were the ones that displayed highest seroreactivity and were most commonly picked up by autoantibodies from vaccinated cancer patient sera, as well as by autoantibodies when carrying Core-3 and in some cases STn-glycoforms^{15, 16}, and, thus, should be a suitable collection of autoantibody reporter epitopes to determine utility of the approach. The second library also contained the 5E5 T(GalNAc)AP epitope in order to confirm reproducibility and alignment with the previous library. A small batch of the library (5000 beads) was screened with the 5E5 mAb, hits were sequenced by MS analysis (Figure 4B) and identified structures confirmed the previous findings described above (see Figure 3A). Mass spectral data for these sequences can be found in Supplementary information 2. With the introduction of serine residues, which could be either glycosylated or non-glycosylated, use of HCD-MS² alone became insufficient for complete sequencing of peptides, as many of the hits were not only multiply glycosylated but had additional serine residues that were not glycosylated. For full characterization of the sequences including the location of all glycosyl residues, the use of the ETD-MS² technique^{26, 27}, which yields sequence specific c and z⁺ ions (and/or related ions produced by proton transfer), with retention of GalNAc attached to S and T residues, was necessary and usually sufficient. In these cases, ETD-MS² analysis elucidated the correct positions of the glycosylated and the nonglycosylated serine residues within the sequence (see, e.g., Supplemental information, Figure S3).

To explore serum samples several sequential experiments were conducted. To deplete non-cancer related antigenic sequences, a batch of approx. 30,000 beads was screened with a pool of ten normal sera (dilution of 1:100 in staining buffer for each serum) and bound serum antibodies were detected with a goat anti-human-IgGAP secondary antibody. After de-selecting positive beads (less than 0.01% of the whole batch) from the normal serum stain, the library was sequentially re-stained with sera from a breast cancer patient pre- and post vaccination with a fully glycosylated Tn-MUC1-TR-106-mer respectively. The pre-vaccination and post-vaccination (five s.c. injections biweekly of 2 to 4 μ g 25Tn-106mer-MUC1-KLH conjugate) sera were from n = 20 breast cancer patients (stage III/IV after treatment and disease-free) enrolled in a phase I study^{15, 37}. Hits from both the pre- and post-vaccinated sera were selected and sequenced accordingly. As shown in Figure 5C, vaccinated serum generated autoantibody reactivity towards the AAA-G-S-T-X-P-X-AHGV

motif and a somewhat lesser preference for the AAA-P-D-**T**-X-X-X-AHGV motif (Figure 5C) both of which are predominant autoantibody MUC1 epitopes in cancer patients¹⁵. Hits isolated from binding of the secondary antibody (goat anti-human-IgG (Fc-specific)-AP-conjugate) (Fig 4C), normal serum (Fig 5A) or pre-vaccinated serum (Fig 5B) did not relate to any sequences obtained with the post-vaccinated sera (Figure 5C). Although a hit was obtained with screening of the secondary antibody with the sequence AAA-**S-S-T**-A-P-R-AHGV-sp (Figure 3C, sequence 16), it was not observed in screening with aforementioned sera (Figure 4C, sequences 26-45).

To validate the sequences obtained with the vaccinated patient, 31 variants of the Tn-MUC1-20mer (varying in glycosylation patterns) were re-synthesized and printed onto the microarray platform as described¹⁵. As shown in Figure 5, the sequences detected in the microarray validation experiment match the sequences obtained from the random glycopeptide bead library with the same post-vaccination patient serum. The selected patient had mixed autoantibody reactivity to PDT(aGalNAc)R- and mono-glycosylated epitopes - GS(aGalNAc)TAP- (Figure 5D) which are in good agreement with the identified sequences obtained from the bead staining. The results of this experiment show that by using the random glycopeptide bead library method we can selectively and specifically detect the host immune response to an antigen directly from a patient serum.

Rapid diversification of random glycopeptide library using glycosyltransferases

Enzymatic glycosylation is a powerful approach to synthesize complex oligosaccharides that are otherwise difficult or time consuming to obtain by chemical means³⁸. We as well as others have shown that glycosyltransferases can be used to efficiently elongate oligosaccharides on solid-phase beads and microarray surfaces^{15, 39-41}. Here we also demonstrate that such enzymes can efficiently be used to expand glycoforms directly on peptides linked to PEGA beads. We recently found that Core3 (GlcNAc β 1-3GalNAc α -) MUC1 glycopeptides among other glycoforms were frequent detected by autoantibodies in cancer sera^{15, 16}. Therefore a batch of the second glycopeptide bead library (approx. 50,000 beads) was enzymatically glycosylated with the Core-3 synthetase GlcNAc-T6) to introduce a GlcNAc β q-3 moiety to all sequences containing a GalNAc residue. To confirm successful glycosylation, a small batch of the library (approx. 5,000 beads) was screened with 1E10 mAb, a Core-3-MUC1 specific monoclonal antibody (epitope: -GST(core3)APP-)¹⁵. Several reactive beads were isolated and sequenced, which confirmed correct identity of glycopeptides as well as specificity of the 1E10 mAb (Figure 6A)¹⁵. During MS1 analysis of obtained hits, it was observed that a portion of peptide material cleaved from the bead that bore only the Tn-glycoform, meaning that not all of the peptide material on the bead was extended to Core-3, which is not surprising as slower kinetics from less degrees of freedom are expected on solid-phase reactions^{15, 40}. Other tumor glycoforms such as STn (NeuAc α 2-6GalNAc α -) could also be synthesized and used in bead staining (data not shown). However, sialosides are sensitive to acid conditions and only partially survived TFA treatment during cleavage, meaning that identification of additional glycosylation sites is not as straightforward.

Proceeding to biological samples, a batch of the Core-3-library (approx 30,000 beads) was screened with a pool of ten normal sera, and positive hits were depleted and sequenced accordingly (Figure 6B). The remaining beads were screened with a pool of three stage I breast cancer sera; some of the identified sequences resembled the Core-3 MUC1 epitope (GST(core3)APP-), as seen in Figure 6C. Identified sequences were re-synthesized, printed onto the microarray platform and evaluated with the same sera from the pool that was used on the bead platform. The results of this experiment are summarized in Figure 6D and show autoantibody binding from individual sera from the serum pool used for bead selection. and are in agreement with our previous studies¹⁵. From this limited selection, it is obvious that

within the pool, some sera contribute more than others when hits arise, proving that re-synthesis and microarray validation are very important in determining the value of a detected biomarker. Other cancer hits that were obtained with the Core-3-extended library have not been confirmed by microarray validation; a list is available in Supplementary information 1, Table S2. It should be noted that during the microarray validation procedure, naked variants of all of the sequences were synthesized and no reactivity was observed to these epitopes on the microarray, proving that the reactivity is truly dependent on the sugar moiety present on the glycopeptides (data not shown). In the case of Core-3 extended structures, it was clear that the reactivity was only due to the Core-3 glycopeptides, not the Tn-glycopeptides or naked variants also present as controls (data not shown).

Conclusion

We have demonstrated a powerful strategy to produce random O-glycopeptide bead libraries for screening of serum autoantibodies. By selecting hydrophilic PEGA beads and including a safety-catch linker for mild and clean release of glycopeptides from beads, complex biological samples such as sera could be analyzed with no non-specific binding issues and without interference of impurities during MS analysis, two important key elements in the study. Libraries containing glycopeptides reporter epitopes were correctly identified with autoantibodies from either vaccinated serum or from cancer patients. Furthermore, the glycans on the random PEGA bead library could efficiently be extended by on-bead enzymatic glycosylation reactions and were successfully detected using cancer sera. Positively stained beads, detected either by fluorescence or color, can be isolated and released glycopeptides de-convoluted by single bead mass spectrometric sequencing methodology using a hybrid ion trap/Orbitrap system which facilitates more accurate and sensitive measurements than a conventional ion trap system, along with correct assignment of glyco-amino acids within the sequence by employing the ETD technique.

Identified glycopeptide sequences were further re-synthesized and validated with our microarray chip technology with extended numbers of sera. This approach can dramatically speed up biomarker discovery efforts as well as allow for a much broader screening of potential immunogenic targets. In addition, this strategy can be applied not only to new O-PTG biomarker discovery, but also to other PTMs which may be of importance in health and disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by The Benzon Foundation, The Carlsberg Foundation, The Danish Research Councils, Danish Agency for Science, Technology and Innovation (FTP), NIH PO1 CA 052477 NIH (1U01CA128437-01), EU FP7-HEALTH-2007-A 201381, and University of Copenhagen Programme of Excellence.

References

1. Krueger KE, Srivastava S. Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics. *Mol Cell Proteomics*. 2006; 5(10):1799–810. [PubMed: 16844681]
2. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA. Glycosylation and the immune system. *Science*. 2001; 291(5512):2370–6. [PubMed: 11269318]
3. Schietinger A, Philip M, Schreiber H. Specificity in cancer immunotherapy. *Seminars in Immunology*. 2008; 20(5):276–285. [PubMed: 18684640]

4. Mintz PJ, Kim J, Do KA, Wang X, Zinner RG, Cristofanilli M, Arap MA, Hong WK, Troncso P, Logothetis CJ, Pasqualini R, Arap W. Fingerprinting the circulating repertoire of antibodies from cancer patients. *Nat Biotechnol.* 2003; 21(1):57–63. [PubMed: 12496764]
5. Chapman C, Murray A, Chakrabarti J, Thorpe A, Woolston C, Sahin U, Barnes A, Robertson J. Autoantibodies in breast cancer: their use as an aid to early diagnosis. *Ann Oncol.* 2007; 18(5):868–73. [PubMed: 17347129]
6. Ninkovic T, Hanisch FG. O-glycosylated human MUC1 repeats are processed in vitro by immunoproteasomes. *J Immunol.* 2007; 179(4):2380–8. [PubMed: 17675499]
7. Paulson JC, Blixt O, Collins BE. Sweet spots in functional glycomics. *Nat Chem Biol.* 2006; 2(5):238–48. [PubMed: 16619023]
8. Brockhausen I. Pathways of O-glycan biosynthesis in cancer cells. *Biochim Biophys Acta.* 1999; 1473(1):67–95. [PubMed: 10580130]
9. Burchell J, Taylor-Papadimitriou J. Effect of modification of carbohydrate side chains on the reactivity of antibodies with core-protein epitopes of the MUC1 gene product. *Epithelial Cell Biol.* 1993; 2(4):155–62. [PubMed: 7505698]
10. von Gunten S, Smith DF, Cummings RD, Riedel S, Miescher S, Schaub A, Hamilton RG, Bochner BS. Intravenous immunoglobulin contains a broad repertoire of anticarbohydrate antibodies that is not restricted to the IgG2 subclass. *J Allergy Clin Immunol.* 2009; 123(6):1268–76 e15. [PubMed: 19443021]
11. Springer GF. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *J Mol Med.* 1997; 75(8):594–602. [PubMed: 9297627]
12. Karsten U, von Mensdorff-Pouilly S, Goletz S. What makes MUC1 a tumor antigen? *Tumour Biol.* 2005; 26(4):217–20. [PubMed: 16006776]
13. Reuschenbach M, von Knebel Doeberitz M, Wentzensen N. A systematic review of humoral immune responses against tumor antigens. *Cancer Immunol Immunother.* 2009; 58(10):1535–44. [PubMed: 19562338]
14. Schietinger A, Philip M, Yoshida BA, Azadi P, Liu H, Meredith SC, Schreiber H. A mutant chaperone converts a wild-type protein into a tumor-specific antigen. *Science.* 2006; 314(5797):304–8. [PubMed: 17038624]
15. Blixt O, Clo E, Nudelman AS, Sørensen KK, Clausen T, Wandall HH, Livingston PO, Clausen H, Jensen KJ. A High-throughput O-Glycopeptide Discovery Platform for Seromic Profiling. *Journal of Proteome Research.* 2010 In Press.
16. Wandall HH, Blixt O, Tarp MA, Pedersen JW, Bennett EP, Mandel U, Ragupathi G, Livingston PO, Hollingsworth MA, Taylor-Papadimitriou J, Burchell J, Clausen H. Cancer biomarkers defined by autoantibody signatures to aberrant O-glycopeptide epitopes. *Cancer Res.* 2010; 70(4):1306–13. [PubMed: 20124478]
17. Lam KS, Lebl M, Krchnak V. The “One-Bead-One-Compound” Combinatorial Library Method. *Chem Rev.* 1997; 97(2):411–448. [PubMed: 11848877]
18. Sorensen AL, Reis CA, Tarp MA, Mandel U, Ramachandran K, Sankaranarayanan V, Schwientek T, Graham R, Taylor-Papadimitriou J, Hollingsworth MA, Burchell J, Clausen H. Chemoenzymatically synthesized multimeric Tn/STn MUC1 glycopeptides elicit cancer-specific anti-MUC1 antibody responses and override tolerance. *Glycobiology.* 2006; 16(2):96–107. [PubMed: 16207894]
19. Price MR, Rye PD, Petrakou E, Murray A, Brady K, Imai S, Haga S, Kiyozuka Y, Schol D, Meulenbroek MF, Snijdwint FG, von Mensdorff-Pouilly S, Verstraeten RA, Kenemans P, Blockzijl A, Nilsson K, Nilsson O, Reddish M, Suresh MR, Koganty RR, Fortier S, Baronie L, Berg A, Longenecker MB, Hilgers J, et al. Summary report on the ISOBM TD-4 Workshop: analysis of 56 monoclonal antibodies against the MUC1 mucin. San Diego, Calif., November 17–23, 1996. *Tumour Biol.* 1998; 19(Suppl 1):1–20. [PubMed: 9422084]
20. Meldal M. Multiple column synthesis of quenched solid-phase bound fluorogenic substrates for characterization of endoprotease specificity. *Methods: A Companion to Methods in Enzymology.* 1994; 6:417–424.
21. Chan, W.; White, P. Fmoc solid phase synthesis: A practical Approach. Oxford University Press; 2000.

22. Nestler HP. Combinatorial chemistry and fragment screening--two unlike siblings? *Curr Drug Discov Technol.* 2005; 2(1):1–12. [PubMed: 16472236]
23. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, Horning S. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem.* 2006; 78(7):2113–20. [PubMed: 16579588]
24. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods.* 2007; 4(9):709–12. [PubMed: 17721543]
25. Zhang Y, Ficarro SB, Li S, Marto JA. Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. *J Am Soc Mass Spectrom.* 2009; 20(8):1425–34. [PubMed: 19403316]
26. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A.* 2004; 101(26):9528–33. [PubMed: 15210983]
27. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JE, Coon JJ. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem.* 2007; 79(2):477–85. [PubMed: 17222010]
28. Olsen JV, de Godoy LM, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics.* 2005; 4(12):2010–21. [PubMed: 16249172]
29. Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, Alvarez R, Bryan MC, Fazio F, Calarese D, Stevens J, Razi N, Stevens DJ, Skehel JJ, van Die I, Burton DR, Wilson IA, Cummings R, Bovin N, Wong CH, Paulson JC. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A.* 2004; 101(49):17033–8. [PubMed: 15563589]
30. Iwai T, Inaba N, Naundorf A, Zhang Y, Gotoh M, Iwasaki H, Kudo T, Togayachi A, Ishizuka Y, Nakanishi H, Narimatsu H. Molecular cloning and characterization of a novel UDP-GlcNAc:GalNAc-peptide beta1,3-Nacetylglucosaminyltransferase (beta 3Gn-T6), an enzyme synthesizing the core 3 structure of Oglycans. *J Biol Chem.* 2002; 277(15):12802–9. [PubMed: 11821425]
31. Lam KS, Lehman AL, Song A, Doan N, Enstrom AM, Maxwell J, Liu R. Synthesis and screening of “one-bead one-compound” combinatorial peptide libraries. *Methods Enzymol.* 2003; 369:298–322. [PubMed: 14722961]
32. Tarp MA, Sorensen AL, Mandel U, Paulsen H, Burchell J, Taylor-Papadimitriou J, Clausen H. Identification of a novel cancer-specific immunodominant glycopeptide epitope in the MUC1 tandem repeat. *Glycobiology.* 2007; 17(2):197–209. [PubMed: 17050588]
33. Brust A, Tickle AE. High-throughput synthesis of conopeptides: a safety-catch linker approach enabling disulfide formation in 96-well format. *J Pept Sci.* 2007; 13(2):133–41. [PubMed: 17160973]
34. Auzanneau FI, Meldal M, Bock K. Synthesis, characterization and biocompatibility of PEGA resins. *J Pept Sci.* 1995; 1(1):31–44. [PubMed: 9222982]
35. Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature.* 1991; 354(6348):82–4. [PubMed: 1944576]
36. Karimi B, Zareyee D. Rapid, efficient and chemoselective deoxygenation of sulfoxides to thioethers using NaBH₄/I-2. *Synthesis-Stuttgart.* 2003; (3):335–336.
37. Sabbatini PJ, Ragupathi G, Hood C, Aghajanian CA, Juretzka M, Iasonos A, Hensley ML, Spassova MK, Ouerfelli O, Spriggs DR, Tew WP, Konner J, Clausen H, Abu Rustum N, Dansiehesky SJ, Livingston PO. Pilot study of a heptavalent vaccine-keyhole limpet hemocyanin conjugate plus QS21 in patients with epithelial ovarian, fallopian tube, or peritoneal cancer. *Clin Cancer Res.* 2007; 13(14):4170–7. [PubMed: 17634545]
38. Blixt O, Razi N. Chemoenzymatic synthesis of glycan libraries. *Methods Enzymol.* 2006; 415:137–53. [PubMed: 17116472]
39. Laurent N, Voglmeir J, Wright A, Blackburn J, Pham NT, Wong SC, Gaskell SJ, Flitsch SL. Enzymatic glycosylation of peptide arrays on gold surfaces. *Chembiochem.* 2008; 9(6):883–7. [PubMed: 18330850]

40. Blixt O, Norberg T. Solid-phase enzymatic synthesis of a sialyl Lewis x tetrasaccharide on a sepharose matrix. *Journal of Organic Chemistry*. 1998; 63(8):2705–2710. [PubMed: 11672140]
41. Tolborg JF, Petersen L, Jensen KJ, Mayer C, Jakeman DL, Warren RA, Withers SG. Solid-phase oligosaccharide and glycopeptide synthesis using glycosynthases. *J Org Chem*. 2002; 67(12): 4143–9. [PubMed: 12054949]

Figure 1.

The design of the 5E5-prototype glycopeptide random bead library and the screening approach. **A)** Randomized amino acid residues at positions X₋₂ to X₊₃ relative to the constant GalNAcα1-O-T (**T**) at position X₀ are linked to a spacer containing two ethylene-glycol units at the C-terminal end and a cleavable linker anchoring the glycopeptide to the PEGA bead; **B)** Bead staining (**a**) with the 5E5 antibody and the secondary anti-mouse-IgG-AP antibody for colorimetric development with BCIP/NBT ready-to-use substrate solution followed by selection and reductive acid cleavage of single bead with I₂/NaBH₄ solution (**b**). Sequence determination by MS and re-synthesis of selected glycopeptides for microarray validation (**c**).

Figure 2.

Mass spectral data for sequence 8, AAA-A-S-T-A-P-P-AAA-sp. **A**) MS¹ spectrum of the sequence, **B**, **C**) HCD-MS² spectrum of the precursor m/z 658.8398, with b and y fragment ions assigned. Closely spaced ions magnified in insets demonstrate the capability of the Orbitrap to distinguish between very closely spaced fragments; the asterisk denotes an internal fragmentation product (either APPAA or PPAAA).

Figure 3.

Sequencing of 5E5 antibody stained beads and microarray validation of re-synthesized glycopeptides with 5E5 antibody. **A)** a numbered list of sequences isolated through 5E5 antibody screening together with their observed $[M+2H]^{2+}$ ion masses (full MS experimental data for all sequences available in the Supplementary Information 2); **B)** microarray data for an alanine mutation walk through the GSTAPP region displaying the significance of the GalNAc α 1-O-T and the proline residue at the +2 position with respect to the glycosylation site; **C)** microarray data for the TXP mutations in the GSTAPP region showing the binding preferences of the 5E5 antibody with respect to the X amino acid that was randomized in the library synthesis; **D)** a microarray scan image for the graph shown in D). Spot-to-spot variations for three replicates of each compound are represented by the error bars.

Figure 4.

The design of the second generation library, 5E5 mAb data and secondary antibody hits. **A)** The design of the second random glycopeptide bead library built to resemble mucin-type structures especially those of MUC1. The inclusion of GalNAc α 1-O-serine allows for multiple glycosylation sites and the inclusion of other amino acids allows for rebuilding parts of the MUC1 tandem repeat – sequences like GSTAPPAHGV and APDTRPA. The remaining chemistry like the spacer, linker and solid support is identical to that of the first library (Figure 1.) **B)** Sequences picked up from screening the second library with the 5E5 antibody which show the same preference of the antibody as compared to the first library **C)** Sequences picked out when screening the second library with only the secondary antibody (goat anti-human-IgGAP (Fc-specific) to avoid picking up false positives.

Figure 5.

Sequence information and microarray validation data for screening of the 2nd generation library with normal sera and pre-vaccination and post-vaccination sera from a disease-free breast cancer patient vaccinated with the Tn-MUC1-106mer coupled to KLH. **A)** Sequences obtained from the second library by screening it with a pool of 10 normal Asterand sera, **B)** Sequences obtained (after screening out the normals) from the patient serum before vaccination where no sequences resembling Tn-MUC1 are observed, **C)** Sequences obtained (after screening out the normals) from the patient serum after vaccination where a high incidence of sequences resembling Tn-MUC1 is observed (sequences 26-45), **D)** microarray validation data for the post-vaccination sequences. A general preference for the GSTAPP region is observed on both platforms and to the PDTR region to a lesser extent. Spot-to-spot variations for three replicates of each compound are represented by the error bars.

Figure 6.

Sequences and microarray validation data obtained by screening the 2nd generation library that has been chemoenzymatically extended to bear the Core-3 glycoform with the 1E10 mAb, normal sera and 3 stage I breast cancer sera. **A**) sequences obtained by screening the 2nd generation Core-3 extended library with the 1E10 mAb which is specific for the GSTAPP region of Core-3-MUC1 (**T** now stands for Core-3-T), **B**) sequences obtained by screening the aforementioned library with a pool of 10 Asterand normal sera, **C**) sequences that resemble Core-3-MUC1 that have been obtained by screening the aforementioned library with 3 stage 1 breast cancer sera, **D**) microarray data for the re-synthesized sequences (52-54) which had been printed onto a microarray platform and stained with the pool as well as each of the cancer serum from the pool. CS1-3 are cancer sera 1-3 respectively. Spot-to-spot variations for three replicates of each compound are represented by the error bars.