Published in final edited form as:

ACS Synth Biol. 2019 October 18; 8(10): 2315–2326. doi:10.1021/acssynbio.9b00127.

Sequence-Specific Mucins for Glycocalyx Engineering

Hao Pan[†], Marshall J. Colville[†], Nitin T. Supekar[‡], Parastoo Azadi[‡], Matthew J. Paszek^{*,†,§,||}

[†]Field of Biophysics, Cornell University, Ithaca, New York 14853, United States

[‡]Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States

[§]Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853, United States

^{||}Nancy E. and Peter C. Meinig School of Biomedical Engineering, Cornell University, Ithaca, New York 14853, United States

Abstract

Few approaches exist for the stable and controllable synthesis of customized mucin glycoproteins for glycocalyx editing in eukaryotic cells. Taking advantage of custom gene synthesis and a biology-by-parts approach to cDNA construction, we build a library of swappable DNA bricks for mucin leader tags, membrane anchors, cytoplasmic motifs, and optical reporters, as well as codon-optimized native mucin repeats and newly designed domains for synthetic mucins. We construct a library of over 50 mucins, each with unique chemical, structural, and optical properties and describe how additional permutations could readily be constructed. We apply the library to explore sequence-specific effects on glycosylation for engineering of mucins. We find that the extension of the immature α -GalNAc Tn-antigen to Core 1 and Core 2 glycan structures depends on the underlying peptide backbone sequence. Glycosylation could also be influenced through recycling motifs on the mucin cytoplasmic tail. We expect that the mucin parts inventory presented here can be broadly applied for glycocalyx research and mucin-based biotechnologies.

GRAPHICAL ABSTRACT

*Corresponding Author mjp31@cornell.edu.

Author Contributions

All authors contributed to the design of experiments, interpretation of results, and preparation of the manuscript. H.P. constructed the mucin cDNA library and analyzed the expressed mucin glycoproteins. H.P. and M.J.C. performed the Core O-Glycan Reporter/Amplification (CORA) analysis. N.T.S. and P.A. conducted the O-glycomics analysis of purified mucins. H.P. and M.J.P. wrote the manuscript.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acssyn-bio.9b00127](https://doi.org/10.1021/acssyn-bio.9b00127). Data figures and tables, DNA and amino acid sequences of “biobricks”, and summary of constructs (PDF)

The authors declare the following competing financial interest(s): The technology described in the reported manuscript is currently proceeded with a provisional patent application by Cornell University.

Keywords

mucin; glycan; glycosylation; synthetic biology; engineering; custom gene synthesis

Cell-surface mucins are a family of membrane-anchored biopolymers that are defined by their unstructured polypeptide backbone and dense O-glycosylation.¹ While historically viewed as simple structural molecules that protect the cellular surface and resist pathological cell deposition,² cell-surface mucins are now recognized to have more sophisticated roles in regulating cellular life. In the cellular glycocalyx, mucin ensembles present bioactive glycan epitopes that mediate adhesion and communication between cells and with their external world. For instance, mucin sialic acids can modulate immune cell function through ligation of SIGLEC receptors on natural killer cells and other cell types in the microenvironment.³ Mucins can also physically regulate the spatiotemporal dynamics of receptor activation and signaling responses.⁴ For instance, dense crowding of mucins in the glycocalyx is proposed to control the diffusion and activation of receptors on the cell surface, and to have a sieving effect that controls the passage of soluble factors from the microenvironment to the cell surface.⁵

A key feature of mucins is that their molecular architecture can change dynamically through modulation of the types and frequencies of O-glycan side chains that are appended along the polypeptide backbone. For instance, the charge, size, and arrangement of glycans are proposed to control the extension and rigidity of the mucin backbone.^{6,7} Glycosylation often changes dramatically with cell-state transitions, including differentiation and transformation.^{8,9} As such, both the chemical and physical character of mucins is intimately coupled to the cellular state, contributing to the diverse modulatory roles that mucins can play in cellular adhesion, communication, and signaling. However, how precise backbone sequences and glycosylation patterns contribute to the function of individual mucins and the collective behaviors of mucins in the glycocalyx is largely unresolved.

Contemporary genetic approaches that target glycan modifying enzymes provide powerful tools for editing the glycans on individual mucins.¹⁰⁻¹² However, the mucin backbone also contributes to the overall molecular structure of mucins and their biophysical functionality but approaches for the design and genetic encoding of mucins with tailored backbone sequences remain limited. In the absence of such approaches, libraries of biomimetic mucin polymers with plasma membrane anchors have been developed for glycocalyx editing.^{6,13}

While highly successful in unraveling some mechanistic details of glycocalyx function, synthetic polymers are typically cleared from the cell-surface in hours to days and must be continuously replenished through media supplementation.^{14,15} Thus, investigation of behaviors over longer time durations, particularly *in vivo*, are largely inaccessible with synthetic mucin mimetics.

A strategy for mucin engineering and glycocalyx editing that combines the best features of the synthetic chemical approach, defined backbone chemistry, tunable glycan structures, and glycan placement, with the power and long-term stability of genomic encoding has yet to be developed. Advances in custom gene synthesis now enable new cDNA sequences to be constructed at unprecedented speed and low cost. However, custom gene synthesis has not been readily applicable for the highly repetitive DNA sequences that are characteristic of most mucins. Repetitive gene sequences impede DNA fragment assembly in custom gene synthesis and are challenging to amplify through polymerase chain reaction (PCR) due to primer mispairing.^{16,17} One possible solution is to exploit codon redundancy to construct synonymous gene sequences with minimal codon repetitiveness, an approach that has been successfully applied for elastin-like proteins.^{18,19}

Here we take advantage of codon redundancy to develop an efficient strategy to design, genetically encode, and fabricate cDNAs for synthesis of sequence-specific mucins in cells. Our combinatorial library of mucin parts enables facile construction of mucin biopolymers with tunable sizes, side-chain spacing, and glycan types for glycocalyx editing.

RESULTS

Schematic Representation of Combinatorial Genetically Encoded Library for Sequence-Specific Mucins.

We developed a modular biology-by-parts approach for combinatorial mucin cDNA construction. Each functional motif in the mucin coding sequence was flanked by restriction sites, so that unique cDNA “bricks” for mucin leader sequences, tandem repeats, optical reporters, transmembrane domains, and cytoplasmic domains could be readily swapped to construct mucins of altered functionality (Figure 1a,b). The cDNA parts catalogue included 13 unique tandem repeats for mucin biopolymers of varying size, backbone chemistry, and frequency of serine and threonine (S/T) glycosylation sites (Figure 1d). The cDNAs for the mucin polymer domains were fabricated through custom gene synthesis following codon optimization (Figure 1c). For optimization, codon redundancy was exploited to find synonymous gene sequences that coded the desired polypeptide with minimal codon repetition. The “codon-scrambled” cDNA sequences were synthesized through standard custom gene synthesis services offered by commercial vendors.

The tandem repeats that form the mucin polymer backbone were adapted from native mucins or newly designed (Figure 1d). The repeats PDTRPAPGSTAPPAHGVTSA and KEPAPTTTP were inspired by native Muc1 and Proteoglycan 4 (Lubricin), respectively. Three repeats were designed based on statistical analysis of mucin O-glycosylation sites (PPASTSAPG) or analysis of O-GalNAc transfer efficiency (DAATPAP and DAATPAPP).²⁰ The base Muc1 repeat was further modified through alanine substitutions to create Muc1-

like tandem repeats with altered frequencies of S/T potential glycosylation sites (Muc1_21S, D, T). Across the library, the percentage of S/T sites in the mucin backbones varied from 10% to 33% (Figure 1d).

Constructing and Validating the Surface Expression of Sequence-Specific Mucins.

We compared the expression of codon-scrambled, synonymous mucin cDNAs to native mucin repetitive cDNAs, and evaluated the glycosylation of the protein products. We fused the cDNAs of the native and synonymous Muc1 tandem repeats with a signal/leader sequence, membrane anchor, and GFP reporter (Figure 2a). Each construct was transiently expressed in HEK293Ts. We analyzed the glycosylation patterns of the mucins through lectin blotting. Blots were probed with peanut agglutinin (PNA) to detect Core 1 glycans, Vicia villosa lectin (VVA) to detect the unextended Tn antigen (α -GalNAc) and Muc1 mAb (clone HMPV) to probe MUC1 tandem repeat peptide core (Muc1 TR).²¹ We also labeled Muc1 sialic acids on our blots through mild periodate oxidation to generate aldehydes on sialic acids, followed by aniline-catalyzed oxime ligation (PAL) with a hydroxylamine-AF568 probe.²² The GFP reporter were also probed *via* Western blot to detect expressed mucins. In order to validate the use of lectins PNA and VVA (Figure 2c), we knocked out the Core 1 β 3-T specific molecular chaperone (COSMC) in native Muc1 overexpressing MCF10As to inhibit elongation of the primary O-linked GalNAc.²³ We compared the glycosylation pattern of overexpressed native Muc1 (Native_Muc1) in wild-type and knockout cells. Mucin in the COSMC knockout cells had lower PNA reactivity, while VVA binding dramatically increased, presumably due to abrogation of glycan extension (Figure 2d). The result confirmed that PNA can be a good indicator for extended Core 1 glycans and VVA for the unextended Tn antigen on the mucins.

Western blot analysis on native and codon-scrambled mucins confirmed that the codon-scrambled, synonymous Muc1 repeats (Muc1_42 GFP) had a molecular weight and glycosylation pattern comparable to the native repetitive Muc1 repeats (Native_Muc1 GFP) (Figure 2e). Mucins ran as a nearly continuous smear in SDS-PAGE with the Muc1 TR antibody, indicating a heterogeneous mix of glycoforms (Figure 2e; Muc1 TR). Predominant glycoforms with apparent molecular weights of approximately 470, 210, and 170 kDa were observed for each expression construct on the GFP blot (Figure 2e; GFP). VVA staining was strong in the smeared region between the upper and lower bands, whereas PNA and sialic acid signal was strongest near the 460 kDa band at the top of the smear (Figure 2e). Based on these results, we concluded that the 460 kDa band was fully glycosylated Muc1, while the smear represented a heterogeneous mix of Muc1 glycoforms containing unextended O-glycan structures. The lower bands on the GFP blot were also observed on the Muc1 TR blots, but not with lectin or sialic acid probes, indicating that these bands likely represent underglycosylated full-length Muc1. Both native and codon-scrambled Muc1 were successfully trafficked to the cell surface and incorporated into the cellular glycocalyx (Figure 2f).

One advantage of the codon-scrambled mucin cDNAs was the potential to improve the stability of the nucleotide sequence during some DNA processing operations. Slippage during replication, transcription, reverse transcription, and other nucleotide processing

operations on repetitive nucleotide sequences often results in deletions or amplifications of cDNAs and mRNAs.²⁴ We conducted a lentiviral stability assay in which we evaluated the fidelity of cDNAs incorporated into the cellular genome following viral delivery and reverse transcription. In cells virally transduced with the native, non-optimized Muc1 cDNA, the Muc1 glycoprotein product had a significantly lower molecular weight than expected, consistent with the cDNAs being truncated. Cells transiently transfected with the native Muc1 cDNA, or those virally modified with codon-scrambled Muc1 cDNA, produced glycoproteins of the expected size (Figure S1). While the lentiviral assay was not a direct test of genomic stability, the results indicated that nonrepetitive mucin sequences are more stable throughout at least some types of nucleotide processing operations.

The tandem repeats of native mucins are often polymorphic in number in humans, resulting in a variation of mucin size among individuals²⁵ and short alleles of Muc1 have been shown to be associated with gastric cancer.²⁶ Inspired by this natural variation and to further validate our approach, we designed and constructed a series of synonymous mucins with variable numbers of tandem repeats (x42, x21, x10, x0; Figure 2a). The polymorphic cDNAs expressed well on the cell surface and displayed the expected differences in size and extent of glycosylation (Figure 2g and Figure S2b). As expected based on previous reports,²⁷ the larger mucins formed a glycocalyx that was substantial enough to dislodge epithelial cells from their substrate (Figure S2f).

Substituting the Potential Glycosylation Sites with Alanine in the Mucin Polymer Backbone Tunes O-Glycan Maturation.

We next tested whether mucins with altered patterns of glycosylation, including differences in glycan extension, could be encoded by mutating away the S/T sites in the mucin backbone. Our overall strategy was to create secreted Muc1 tandem repeats in which alanine was substituted for S/T in one, two, or three of the five potential glycosylation sites in each repeat (Figure 3a,b). We envisioned that the secreted mucins could then be harvested from cell culture media for subsequent glycan analysis with lectin blotting and mass spectrometry.

cDNAs for the desired Muc1 mutants with 21 repeats each were optimized through codon scrambling and fabricated through custom gene synthesis. The single (Muc1_21S), double (Muc1_21D), and triple (Muc1_21T) glycosylation mutants had 21, 42, and 63 total S/T to alanine substitutions, respectively, and varied in potential glycosylation frequency at 20%, 15%, and 10%. An IgK signal peptide and 6x-His-SUMOStar tag was fused to the 21 copies of the wild-type Muc1 repeat or the three mutant repeats (Figure 3a). No transmembrane protein anchor was included, so that the IgK signal peptide would direct secretion of the recombinant mucin protein.

The secreted mucins were harvested from the media supernatant of HEK293 cells and analyzed by Western and lectin blot. The wild-type and glycosylation mutants had a considerably higher apparent molecular weight than the theoretical molecular mass of the undecorated peptide backbones (Figure 3b,c and Figure S3). The potential glycosylation site mutants migrated faster in SDS-PAGE, indicating that they had fewer glycan chains or that their glycans were shorter and, thus, less obstructive to their electrophoretic mobility (Figure 3c).

We found that substituting the S/T tuned the O-glycan maturation. The secreted Muc1 glycoproteins were blotted and probed with VVA for Tn antigen, PNA for Core 1 glycans, and s-WGA for GlcNAc, a building block of Core 2, 3, 4, and 6 glycans (Figure 3c). We constructed electrophoretograms by recording the fluorescence intensity of glycan probes along each lane of a single, costained blot (Figure 3d). Core 1 (PNA) and GlcNAc-containing (s-WGA) glycans were abundant in the mucin glycoforms with the highest apparent molecular weights. The lower apparent molecular weight glycoforms contained abundant VVA-reactive glycans and minimal Core 1 and GlcNAc containing glycans. Gradual alanine substitution clearly shifted the glycoform distribution toward mucins with more unextended, VVA-reactive glycans and fewer extended Core 1 and GlcNAc containing glycans (Figure 3d,e). Surprisingly, substitution of even one serine (see sMuc1S) dramatically changed the glycosylation pattern, leading to generation of more nonfully extended glycoforms (Figure 3c,d).

To validate our lectin analysis and catalogue the specific glycan structures on the mucins, we conducted mass spectrometry to profile the O-glycans on the wild-type mucin repeats (sMuc1) and the mutant with three S/T alanine mutations per repeat (sMuc1T). We identified similar Core 1 and Core 2 glycans in both samples (Figure 3f). However, the signal of extended glycans was much stronger in wild-type mucin (sMuc1) compared to the triple mutant (sMuc1T), consistent with our lectin blots. We also fused the glycosylation mutant cDNAs to a transmembrane anchor for cell-surface expression and observed a similar trend of suppression of glycan extension in the glycosylation-site mutants (Figure S4c). To ensure that the overexpression of mucin constructs did not impact functionality of the glycotransferases for glycan extension, we used Cellular O-glycome Reporter/Amplification (CORA), a method which allows protein-free profiling of the overall cellular O-glycome.²⁸ Similar Core 1 and Core 2 glycan structures were detected in both wild-type and Muc1 overexpressing HEK293T cells, indicating that the activity of T synthase and other glycosyltransferases involved in mucin extension are not inhibited by mucin overexpression (Figure S5). Overall, these data demonstrated that extension of glycans in both cell-surface and secreted mucins was sensitive to the alanine substitution along the polymer backbone.

Designer Mucin Domains Reveal Sequence-Specific Effects on Glycosylation.

We next tested whether new types of sequence-specific mucins could be created for editing the glycocalyx. A parallel goal was to further explore the impact of specific backbone features, including glycosylation site frequency and proline number, on mucin glycosylation pattern. Cell-surface mucin cDNAs with GFP reporters were constructed for our three designer mucin repeats, DAATPAP, DAATPAPP, and PPASTSAPG, and KEPAPTTTP inspired by secreted human Proteoglycan 4 (Figure 4a). The three designer mucin repeats were expected to be fully glycosylated based on *in vitro* results.²⁰ The backbones varied in frequency of glycosylation sites (S/T) from 12 to 33%. We also created extended variants of the DAATPAP and DAATPAPP mucins through PCR-amplification of the tandem repeats and reassembly with the original cDNAs to double the number of repeats to 80. All mucins were expressed well, trafficked appropriately to the cell surface, and were extensively decorated with O-glycans (Figure 4c and Figure S6b).

We analyzed the glycosylation patterns of the mucins through lectin blotting. Multiple bands were visible for each mucin on the anti-GFP blot, revealing a complex distribution of mucin glycoforms on and within the cell (Figure 4c). The heavily glycosylated mucins, as indicated by high PNA and VVA reactivity, typically ran as a smear between the highest and second highest molecular weight bands on the anti-GFP blot (Figure 4c,d). These regions were shaded in gray on the electrophoretograms to aid visualization (Figure 4d). The highest molecular weight glycoforms were heavily decorated with Core 1 glycans (Figure 4d; see PNA). The glycoforms enriched in unextended O-glycans were heterogeneous in apparent molecular weight and ran in a smear just below the Core 1 decorated mucins (Figure 4d; compare VVA and PNA).

We then evaluated whether the frequency of O-glycosylation sites might influence the maturation and extension of O-glycans. We quantified the relative Core 1 to Tn antigen ratio among our synthetic mucins through ratiometric analysis of integrated PNA and VVA signals on our lectin blots (Figure 4e). For mucins with 20 or 40 repeats, we saw a notable increase in Core 1 structures compared to Tn-antigen in mucin backbones with a higher S/T content. However, the glycoform distribution was broader for backbones with higher S/T content, as indicated by more pronounced smearing on the lectin blots and the increased width of the PNA and VVA peaks on the electrophoretograms (Figure 4c,d).

We also considered whether proline content might influence the glycosylation of the mucin backbone, since proline has previously been reported to promote glycosyltransferase interactions with mucin backbones.⁷ We compared glycosylation of the DAATPAP and DAATPAPP mucins, which only differed by a single proline per tandem repeat. For mucins with 40 copies of each repeat, the ratio of Core 1 glycans to unextended Tn-antigens was not significantly different between the two mucins (Figure 4e). However, for mucins with 80 copies of the repeats, the relative Core 1 glycan content was significantly lower in the mucin with an extra proline per repeat (Figure 4f). These results suggested that proline content may affect glycosylation in a manner that depends on the overall size of the mucin backbone.

Tuning Mucin Glycosylation through Cytoplasmic Tail Engineering.

Sialylation of O-glycans has previously been reported to occur at least partially in the endosome and trans-Golgi network following endocytosis of cell-surface mucins.²⁹ In an attempt to exploit endocytosis and trafficking as a potential tool to alter mucin glycosylation, we created cDNA “bricks” for mucin cytoplasmic tails with different endocytosis and trafficking signals. We noted that the Muc1 cytoplasmic domain can signal for clathrin-mediated endocytosis, while the Muc1 sequence CQCRRK at the boundary of transmembrane and cytoplasmic domain signals for Muc1 recycles back to the plasma membrane.³⁰ We adopted a synthetic 21-amino-acid transmembrane anchor (TM21) that could anchor mucins to the plasma membrane without a cytoplasmic tail³¹ or with the two different cytoplasmic tails in our library. The first cytoplasmic tail was a simple CQC motif to direct mucin recycling. The second was based on the native Muc1 cytoplasmic tail that contains the CQC motif as well as additional motifs, YHPM and YTNP, to direct more efficient endocytosis.³²

To test their functionality, we fused the TM21 anchor with or without the cytoplasmic tails to a codon-scrambled Muc1 with 10 tandem repeats (Muc1_10) (Figure 5a). All mucin cDNAs were transiently transfected into HEK293Ts. We labeled the sialic acids on the cell surface with PAL. On lectin blots, the PAL sialic acid signal was strongest at approximately 171 kDa, overlapping with a strong PNA signal, suggesting the PNA-reactive isoforms were also sialic-acid-abundant (Figure 5b). To confirm, we treated the cell lysates with sialidase prior to lectin blot analysis and analyzed the PNA-staining pattern to detect a shift in electrophoretic mobility due to removal of negatively charged sialic acids. Regardless of the cytoplasmic tail motif, the PNA reactive band in the mucins was higher and broader following sialidase treatment, indicating that the dominant PNA-reactive isoforms in all constructs were sialylated (Figure 5c).

To further analyze the sialylated isoforms, we pulled down the Core-1-rich mucin glycoforms with PNA and then probed with *Maackia amurensis* lectin (MAA), which prefers to bind sialic acids in an (α -2,3) linkage.³³ Surprisingly, we did not see any MAA signal near 171 kDa but noted ultrahigh molecular weight glycoforms that were reactive to MAA (Figure 5d, top). The MAA-reactive, ultrahigh molecular weight glycoforms were promoted by recycling motifs. We found that the inclusion of the CQC motif led to a 2-fold increase in the MAA/PNA ratio compared to the TM21 anchor only, and the longer cytoplasmic tail based on Muc1 increased the MAA/PNA ratio 3-fold (Figure 5d, bottom). Since MAA is also reactive to sulfated glycans, we could not definitely conclude whether the ultrahigh molecular weight glycoforms were enriched in a 2,3-linked sialic acids, sulfated glycans, or some other potential glycan structure that is reactive to MAA. Nevertheless, the results suggested that engineering of cytoplasmic motifs could be a viable strategy to at least partially tune mucin glycosylation.

DISCUSSIONS

The O-glycosylation of mucins determines their physical and biochemical characteristics and, thus, their biological functions. Our aim was to develop a genetically encoded system to edit the mucin biopolymers as a tool for glycocalyx engineering. Factors that are known to influence mucin glycosylation include the cellular repertoire of glycosyltransferases and their substrates,^{1,34} frequency of O-glycosylation sites on the polypeptide backbone,^{35,36} primary peptide sequences around the O-glycosylation sites,³⁷⁻³⁹ and trafficking of the glycoprotein.^{32,40,41} To expand the glycocalyx editing toolkit beyond targeting glycosyltransferase and their substrates, we attempted to leverage signals and motifs in the mucin backbone sequences and cytoplasmic tails to encode mucins with varying physical features, backbone chemistries, and glycosylation patterns.

Advances in custom gene synthesis now enable rapid construction of sequence-specific cDNAs at relatively low costs. By taking advantage of codon degeneracy to design mucin cDNAs with minimal repetition, we were able to apply custom gene synthesis for construction of 13 unique mucin repeats, each of which could be readily combined with other functional domains for cell-surface anchorage and control of trafficking. Given that all repeat sequences tested were successfully fabricated through standard gene synthesis services with no failures, we expect that the design strategy can be applied to expand the

catalogue of mucin repeats beyond the 13 validated in this work. By combining these cDNAs in a modular fashion with other functional cDNA “bricks,” mucins of modified structure and functionality should easily be constructed with basic molecular techniques, including Gibson Assembly, Golden Gate Assembly, and other modern DNA assembly approaches.

One of the more striking observations in this study was that extension of O-glycans from the Tn antigen to Core 1/2 glycans is discouraged by alanine substitution along the polymer backbone. Given that the effect was observed in both membrane-associated and secreted mucins, altered endocytosis and trafficking likely do not account for the differences in glycan maturation. Differences in glycosylation also are not likely explained by potential effects of mucin overexpression on the functionality of T-synthase and other glycosyltransferases involved in early O-glycan extension. As shown in the Cellular O-Glycome Reporter/Amplification analysis, similar Core 1 or Core 2 glycan structures were observed for both mucin-overexpressing and wild-type HEK293Ts (Figure S5).

We propose several viable explanations for how single or multiple alanine mutations in the Muc1 backbone might affect O-glycan maturation. First, alanine is commonly found adjacent to glycosylation sites and viewed to allow flexibility for the polypeptide backbone.^{42,43} Alanine substitution may alter the polypeptide backbone conformation, restricting glycosyltransferase docking and leading to reduced glycan extension. Second, the O-glycan structures built by cells could be site-specific.⁴⁴ Different glycosylation sites may each carry a particular complement of O-glycan structures. Mutating away that S/T may alter the range of glycoforms for the mucin, resulting in a sudden change in its glycosylation pattern, such as we observed with a single alanine substitution in sMuc1S. Third, the glycosyltransferases may operate in a yet unidentified cooperative manner, which is sensitive to the frequency of O-glycan sites along the mucin backbone.

Many of our analyses of O-glycosylation were based on lectin blots, which have some important limitations. The blot-based glycosylation analysis does not provide a quantitative evaluation of the O-glycosylation site occupancy along the mucin backbones. On our SDS-PAGE gels, mucins of varying molecular weight could be due to differences in glycosylation efficiency, glycan elongation, and sialylation, all of which would affect the electrophoretic mobility of glycoproteins in SDS-PAGE. Further evaluation with a full glycopeptide analysis would clear and strengthen the observations here.⁴⁵ Second, the promiscuous carbohydrate binding reactivity of lectins may cloud the interpretation of the results. However, controls were taken in this work to validate the main lectin-based analyses. Knockout of COSMC to abrogate glycan extension lead to decreased PNA binding and elevated VVA staining, suggesting the appropriateness of these lectins for detecting Core 1 O-glycans and Tn-antigen, respectively (Figure 2d). O-Glycomic analysis on purified mucins also validated conclusions that were based on lectin analysis regarding the types of glycan structures present on mucins (Figure 3f). Thus, in spite of the limitations mentioned, the lectin-blot-based analysis could be a viable strategy to get a visual image of the glycosylation pattern.

Based on earlier reports that O-glycans are sialylated in the endosome, we investigated the potential of modifying the mucin cytoplasmic tail for glyco-engineering. Based on a shift in

electrophoretic mobility following sialidase treatment, we concluded that recycling motifs were not required for mucin sialylation. However, inclusion of recycling motifs promoted the generation of ultrahigh molecular weight mucin glycoforms that react with MAA lectin. Since MAA can bind to glycans containing an α -2,3-linked sialic acids as well as SO₄-3-Gal β ⁴⁶ and potentially other glycan epitopes, the precise glycan structures contained in the high molecular weight mucins remain to be identified. Nevertheless, swapping mucin cytoplasmic tails could be a viable strategy to at least partially engineer emergent glycoforms.

In summary, this work presents a parts inventory for design and construction of sequence-specific mucins. We anticipate that the current mucin library can be applied and extended for glycoalyx editing in research and biotechnology.

MATERIALS AND METHODS

Antibodies and Reagents.

The following antibodies were used: anti-Human MUC1 (CD227) (clone HMPV; 555925, BD Biosciences), mouse anti- β -Actin (clone C4; 47778, Santa Cruz), chicken anti-SUMO/sUMOstar (AB7002, LifeSensors), mouse 6xHis (552565, BD Biosciences), mouse anti- α -tubulin (clone B-7; 5286, Santa Cruz), mouse anti-GFP (clone 4B10; 2955, Cell Signaling Technology), m-IgG κ binding protein–horseradish peroxidase (HRP; 516102, Santa Cruz), goat antimouse IgG (Alexa Fluor 647 conjugated, A-21235; Alexa Fluor 488 conjugated, A-11001; Alexa Fluor 568 conjugated, A-11004; ThermoFisher), and goat antichick IgY (Alexa Fluor 488 conjugated; A-11039, ThermoFisher). Lectins used were unconjugated *Arachis hypogaea* lectin/peanut agglutinin (PNA; L0881, Sigma), biotin-conjugated PNA (B-1075, Vector Laboratories), biotin-conjugated *Maackia amurensis* lectin (MAA; BA-7801, EY Lab), fluoresceinlabeled succinylated Wheat Germ Agglutinin (s-WGA; FL-1021S, Vector Lab), and biotin-conjugated *Vicia villosa* lectin (VVL, VVA; B-1235, Vector Lab). Fluorescent dyes used were Alexa Fluor 647 NHS Ester (A20006, Invitrogen), Alexa Fluor 568 NHS Ester (A20003, Invitrogen), and AFDye 568 Hydroxylamine. Biotinylated lectins were detected using ExtrAvidin-Peroxidase (E2886, Sigma) or NeutrAvidin Protein (Dylight 650 conjugated; 84607, ThermoFisher). For tetracycline-inducible systems, doxycycline was used for induction (204734, Santa Cruz). Streptavidin Sepharose beads (3419, Cell Signaling Technology) was used for immunoprecipitation assays. Cell lysis buffer (9803) and LumiGLO reagent and peroxide (7003) were from Cell Signaling Technology. Normal goat serum (S-1000) for sample blocking was from Vector Lab. Polyethylenimine (PEI) (25 kDa linear PEI, 23966, Polysciences) was used for FreeStyle 293-F cell transfection.

Gene Design and Assembly of MUC1 Tandem Repeat Domains.

cDNAs for cytoplasmic-tail-deleted human Muc1 (Muc1 dCT) and Muc1 tandem-repeat fusion with the synthetic membrane domain TM21 (Muc1 TM21) were generated and cloned into the tetracycline-inducible piggyback expression vector with Puromycin resistance cassette (pPB tetOn Puro) as previously described.²⁷ cDNA of Muc1 TM21 was also inserted into the pcDNA3.1 vector using *Bam*HI and *Eco*RI restriction sites. For

generation of pPB Muc1 mOxGFP dCT TetOn Puro, the cDNA for mOxGFP (Addgene no. 68070) was first amplified with primers: 5'-GGCAGCTCAGCTATGGTGTCCAAGGGCGAGGAGCTGT-3' (forward) and 5'-GGCAGCTGAGCCCTTATACAGCTCGTCCATGCCGTGAGT-3' (reverse). The PCR product was then cloned into pJET1.2 and subcloned nondirectionally into the BlnI site of pPB Muc1 dCT TetOn Puro. To fabricate the cDNAs of secreted mucins (sMuc1), synthetic oligos containing a IgK signal peptide and 6x-His-SUMOSStar tag (6x His Sumostar Muc1) was created through custom gene synthesis (General Biosystems) and cloned into the tetracycline-inducible piggyback expression vector with Neomycin resistance cassette (pPB tetOn Neo). The lentiviral vector pLV puro Muc1 dCT was fabricated as previously reported.⁴

cDNAs for mutant and rationally designed mucins tandem repeats were generated through custom gene synthesis following codon optimization. The least repetitive gene sequence for the desired mucin repeats was found using Codon Scrambler (<http://chilkotilab.pratt.duke.edu/codonscrambler>).¹⁸ The scrambled DNA sequence was adjusted for human codon bias by swapping any codons with less than 10% frequency usage in humans for randomly selected synonymous codons with higher usage. Synthetic oligos for the desired tandem repeats were then synthesized by custom gene synthesis (General Biosystems and Genscript) and cloned in place of the Muc1 tandem repeats in either pPB Muc1 mOxGFP dCT TetOn Puro using the *Bam*HI and *Bsu*36I restriction sites, pcDNA3.1 Muc1 TM21 using the *Bsr*GI and *Bsu*36I restriction sites, or pPB 6x His Sumostar Muc1 using *Bsr*GI and *Bsu*36I restriction sites (see the Supporting Information for cDNA sequences). To generate a lentiviral vector for Muc1 dCT with 42 codon-optimized tandem repeats pLV Muc1_42 dCT construct, the synthesized cDNA for the codon-optimized repeats was inserted into pLV puro Muc1 dCT using *Bam*HI and *Bsu*36I restriction sites. The Muc1 construct with 0 tandem repeats was generated through deletion of the tandem repeats in pcDNA3.1 Muc1_10 TM21 through Q5 site-directed mutagenesis with 5'-TGGAGGAGCCTCAGGCATACCTTTATTG-3' (forward) and 5'-CCACCGCCGACCGAGGTGACATCCTG-3' (reverse) primers.

The cDNA with recycling motif CQCRRK pcDNA3.1 Muc1_10 TM21 CQC was generated from pcDNA3.1 Muc1_10 TM21 through Q5 site-directed mutagenesis with 5'-CCGAAAGTAGGAATTCGGGCCCCGTTTAAACCCGC-3' (forward) and 5'-CGGCACTGACATCTAGAGTACCACAACAAAGCCAGGC-3' (reverse) primers. The cDNA of native CT was subcloned into the *Xba*I and *Eco*RI site of pcDNA3.1 Muc1_10 TM21 CQC.

PCR and Golden Gate Assembly of Extended Synthetic Tandem Repeats.

The 40 tandem repeats of DAATPAP and DAATPAPP mucin cDNAs in pcDNA3.1 were doubled in size to 80 repeats using the Golden Gate Assembly. Two pairs of custom primers for tandem repeats and complete mucin vector were designed to attach *Bsm*BI recognition sites with unique 4bp overhangs so that the PCR products of the 40 tandem repeats and complete mucin expression vector would ligate in a Golden Gate Assembly reaction to

amplify the tandem repeat number (Table S2). The Golden Gate Assembly reaction was conducted as previously reported.⁴⁷

Cell Lines, Culture, and Transfection.

MCF10A human mammary epithelial cells and HEK293T SV40-transformed human embryonic kidney cells were obtained from ATCC. MCF10A cells were cultured in DMEM/F12 media (Thermo-Fisher) supplemented with 5% horse serum (ThermoFisher), 20 ng/mL EGF (Peprotech), 10 μ g/mL insulin (Sigma), 500 ng/mL hydrocortisone (Sigma), and 100 ng/mL cholera toxin (Sigma). HEK293T cells were cultured in DMEM (Thermo-Fisher) supplemented with 10% fetal bovine serum (Thermo-Fisher). Cells were maintained at 37 °C, 5% CO₂, and 90% relative humidity (RH). FreeStyle 293-F cells were cultured in suspension in FreeStyle 293 Expression Medium (Thermo-Fisher). Suspension cultures were maintained in an orbital shaker at 37 °C, 8% CO₂, and 90% RH. Lentiviral transduction was conducted as previously reported in MCF10A cells with stably integrated gene cassettes for expression of the tetracycline transactivator, rtTA-M2, and neomycin resistance gene.⁴⁸ HEK293T cells were transiently transfected with the calcium phosphate method according to standard protocols. FreeStyle 293-F cells were transiently transfected with PEI as previously described.⁴⁹ CRISPR/Cas9 mediated knockout of COSMC in MCF10A Muc1 dCT cells were generated as previously reported.⁵⁰

Western Blot Analysis.

HEK293T cells were plated at 55 000 cells/cm² and transfected with calcium phosphate for 24–36 h before lysis with cell lysis buffer. MCF10A cells were plated at 20 000 cells/cm² and induced with 0.2 μ g/mL doxycycline for 24 h before lysis with cell lysis buffer. Lysates were separated on NuPAGE 3–8% or 7% Tris-Acetate gels and transferred to PVDF membranes. Primary antibodies were diluted at 1:1000 and fluorophore-conjugated or biotinylated lectins were diluted to 2 μ g/mL in 5% BSA TBST and incubated overnight at 4 °C. Secondary antibodies, ExtrAvidin-HRP or Neutravidin-Dylight 650, were diluted at 1:2000 or 1 μ g/mL in 5% BSA TBST and incubated for 1 h at room temperature. Blots were either imaged on a ChemiDoc MP Imaging System (Bio-Rad) or after being developed in LumiGLO reagent and peroxide. Integrated blot intensity was quantified with the FIJI distribution of ImageJ.^{51,52} The statistical significance of the differences among the data was calculated using a one-way ANOVA with repeated measures or two-tailed *t* test.

Periodate Labeling of Cell Surface Sialic Acids.

HEK293T cells were collected after 36 h of transfection. Cells were washed with cold DPBS with Ca²⁺ and Mg²⁺ followed by a 10 min incubation with 1 mM sodium periodate (Sigma) in DPBS. The periodate was quenched by 1 mM glycerol in cold DPBS and washed with cold DPBS. Samples were stained with 25 μ M AFDye-568-hydroxylamine (Fluoroprobes) in the presence of 10 mM aniline (Sigma) in sterile filtered DPBS + 5% FBS pH 6.7 for 30 min at 4 °C in the dark with gentle agitation.

Immunoprecipitation.

HEK293T cells were plated at 55 000 cells/cm² and transfected with the calcium phosphate method for 24–36 h before lysis with cell lysis buffer. The lysates were incubated with 125 µg/mL biotinylated lectin PNA at 4 °C with gentle rocking overnight. Streptavidin Sepharose beads were added to the cell lysates following the manufacturer's instructions, and the suspension was incubated at 4 °C for 3 h. The beads were washed 2 times with lysis buffer and then resuspended in 4× LDS loading buffer. The resuspension was subsequently analyzed by Western blot.

Sialidase Treatment of HEK293Ts.

HEK293T cells were collected 24 h after transfection and incubated with *Arthrobacter ureafaciens* sialidase (Roche, 10 mU, 100 µL final volume) in sialidase buffer⁵³ for 30 min at 37 °C before lysis with cell lysis buffer.

Immunofluorescence.

HEK293T cells were plated at 45 000 cells/cm² and transfected with calcium phosphate for 24 h before being fixed with 4% paraformaldehyde. Antibodies were diluted at 1:100 in 5% normal goat serum in PBS and incubated overnight at 4 °C. Lectins were diluted to 2 µg/mL in 5% normal goat serum in PBS and incubated for 2 h at room temperature. Samples were imaged on a Zeiss LSM inverted 880 confocal microscope using a 40× water immersion objective (NA 1.1).

Secreted Mucin Protein Expression and Purification.

In total, 16.25 µg of pPB 6x His Sumostar Muc1 DNAs were transfected into HEK293T cells in 10 cm culture dishes for 48 h. A total of 30 µg of pPB 6x His Sumostar Muc1 DNAs were transfected into 20 mL of FreeStyle 293-F cell culture for 4 days. Culture media was collected and clarified by centrifugation at 2000 rpm for 5 min. The clarified culture media was bound to Ni-NTA agarose (Qiagen) at 4 °C overnight, washed (20 mM sodium phosphate pH 8.0, 0.5 M sodium chloride (NaCl), 20 mM imidazole), and eluted with imidazole (20 mM sodium phosphate pH 8.0, 0.5 M NaCl, 250 mM imidazole). The eluted sample was dialyzed into PBS with an Amicon Ultra-4 Centrifugal Filter (10 kDa cutoff) and then desalted by using Zeba Spin desalting columns (7k MWCO). The salt-free protein solution was lyophilized and stored at –80 °C.

O-Glycan Profiling of Secreted Mucin Protein.

All reagents were purchased from Sigma unless otherwise mentioned. Purified mucin proteins (600 µg, each) was denatured by heating at 100 °C for 5 min. The denatured proteins were subsequently treated with 19 mg of sodium borohydride (NaBH₄) in 500 µL of 50 mM sodium hydroxide (NaOH) solution at 45 °C for 18 h.⁵⁴ The samples were cooled, neutralized with 10% acetic acid, passed through a Dowex H+ resin column, and lyophilized with borates removed under the stream of nitrogen. The glycans were permethylated for structural characterization by mass spectrometry using previously reported methods.⁵⁵ Briefly, the dried eluate was dissolved with dimethyl sulfoxide (DMSO) and methylated by using methyl iodide and NaOH–DMSO base (prepared by mixing DMSO and 50% w/w

NaOH solution). The reaction was quenched with water and the reaction mixture was extracted with methylene chloride and dried. The permethylated glycans were dissolved in methanol and crystallized with α -dihydroxybenzoic acid (DHBA, 20 mg/mL in 50% v/v methanol/water) matrix. Analysis of glycans present in the samples was performed in the positive ion mode by MALDI-TOF/TOF-MS using an AB SCIEX TOF/TOF 5800 (Applied Biosystem, MDS Analytical Technologies) mass spectrometer. Permethylated glycans from the samples were infused on an Orbitrap Fusion Tribrid mass spectrometer through an electrospray (ESI) probe with HCD and CID fragmentation option for further structural confirmation. The MS1 and MS2 spectra of the glycans were acquired at high resolution by a simple precursor scan, and respective ions were selected manually for further MS/MS scanning. Assignment of glycan structures were done manually and by using Glycoworkbench software, based on the fragmentation patterns and common biosynthetic pathways.

Cellular O-Glycome Reporter/Amplification (CORA).

All chemicals were purchased from Millipore Sigma except where noted. Solvents were of HPLC grade or higher, and 0.1% (v/v) trifluoroacetic acid was included in all chromatography steps. Benzyl 2-acetamido-2-deoxy- α -D-galactopyranoside (BnGalNAc) was peracetylated by heating in a molar excess of 33% (v/v) acetic anhydride in anhydrous pyridine for 1 h at 65 °C. The product was dried by speedvac (Thermo Scientific SPD1010) and used without further purification. Peracetylation was confirmed by LC-MS (Agilent 1100 series LC and G1956B MS, m/z calculated, 438.18; observed, 438.10 [M + H]⁺).

CORA was performed as previously reported.²⁸ Briefly, 500 000 HEK293T cells were plated in a 6 cm culture dish and transfected as above. Following transfection cultures were incubated in full media supplemented with 50 μ M peracetylated BnGalNAc. After 48 h, the media was aspirated and loose cells and debris were removed by centrifugation. The supernatant was then filtered (Millipore Amicon Ultra 4, 10 kDa MWCO) and benzyl glycans collected by gravity chromatography (Waters Sep-Pak C18 3 cc). The eluent was dried by speedvac before permethylation. A sodium hydroxide slurry in DMSO was freshly prepared and 200 μ L added to each dry sample followed by 100 μ L of methyl iodide (ACROS). The samples were mixed continuously for 10 min, and then the reaction was halted by the addition of 600 μ L of deionized water. Permethylated benzyl glycans were recovered by extraction with 200 μ L of chloroform, then washed 4 times with 800 μ L of deionized water. The samples were further purified by C18 gravity chromatography (Waters Sep-Pak C18 1 cc) and dried by a speedvac. Dried samples were dissolved in 50% methanol and spotted 1:1 (v/v) with a matrix of 10 mg/mL 2,5-dihydrobenzoic acid in 50% acetonitrile. Benzyl glycans were analyzed using a MicroFlex MALDI-TOF-MS (Bruker) in positive ion mode. Two external standards of permethylated maltotetraose (Cayman Chemical, m/z calculated, 885.43; observed, 885.65 [M+Na]⁺) and maltoheptaose (Cayman Chemical, m/z calculated, 1497.73; observed, 1497.90 [M + Na]⁺) were included to confirm instrument performance and calibration. Benzyl glycan compositions were assigned on the basis of predicted masses of the sodium adducts of known structures ([M + Na]⁺). Data was analyzed using Mnova (Mestrelab Research) and prepared for presentation with Prism8 (GraphPad).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank V. Weaver and J. Lakins for the lentiviral and transposon plasmids as well as helpful discussions. This investigation was supported by the National Institute of General Medical Sciences Ruth L. Kirschstein National Research Service Award 2T32GM008267 (M.J.C.), National Science Foundation Graduate Research Fellowship DGE-1650441 (M.J.C.), National Institute of Health New Innovator Grant DP2 GM229133 (M.J.P.), National Cancer Institute Grant U54 CA210184 (M.J.P.), and National Science Foundation Grant 1752226 (M.J.P.). Imaging data was acquired through the Cornell University Biotechnology Resource Center, with NYSTEM Grant C029155 and NIH Grant S10OD018516 funding for the shared Zeiss LSM880 microscope. The engineered mucin glycosylation analysis was supported in part by the National Institutes of Health Grants 1S10OD018530 and P41GM10349010 to the Complex Carbohydrate Research Center. MALDI-TOF-MS analysis of the CORA samples was performed by the Biopolymers and Proteomics laboratory at Koch Institute for Integrative Cancer Research, MIT.

REFERENCES

- (1). Brockhausen I, Schachter H, and Stanley P (2009) O-GalNAc Glycans, in *Essentials of Glycobiology* (Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, and Etzler ME, Eds.) 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- (2). Lichtenberger LM (1995) The Hydrophobic Barrier Properties of Gastrointestinal Mucus. *Annu. Rev. Physiol* 57, 565–583. [PubMed: 7778878]
- (3). Hudak JE, Canham SM, and Bertozzi CR (2014) Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. *Nat. Chem. Biol* 10, 69–75. [PubMed: 24292068]
- (4). Paszek MJ, DuFort CC, Rossier O, Bainer R, Mouw JK, Godula K, Hudak JE, Lakins JN, Wijekoon AC, Cassereau L, Rubashkin MG, Magbanua MJ, Thorn KS, Davidson MW, Rugo HS, Park JW, Hammer DA, Giannone G, Bertozzi CR, and Weaver VM (2014) The cancer glycocalyx mechanically primes integrin-mediated growth and survival. *Nature* 511, 319–325. [PubMed: 25030168]
- (5). Polefka TG, Garrick RA, Redwood WR, Swislocki NI, and Chinard FP (1984) Solute-excluded volumes near the Novikoff cell surface. *American Journal of Physiology-Cell Physiology* 247, C350–C356.
- (6). Kramer JR, Onoa B, Bustamante C, and Bertozzi CR (2015) Chemically tunable mucin chimeras assembled on living cells. *Proc. Natl. Acad. Sci U. S. A* 112, 12574–12579. [PubMed: 26420872]
- (7). Coltart DM, Royyuru AK, Williams LJ, Glunz PW, Sames D, Kuduk SD, Schwarz JB, Chen X-T, Danishefsky SJ, and Live DH (2002) Principles of Mucin Architecture: Structural Studies on Synthetic Glycopeptides Bearing Clustered Mono-, Di-, Tri-, and Hexasaccharide Glycodomains. *J. Am. Chem. Soc* 124, 9833–9844. [PubMed: 12175243]
- (8). Dennis JW, Granovsky M, and Warren CE (1999) Protein glycosylation in development and disease. *Bioessays* 21, 412–421. [PubMed: 10376012]
- (9). Reis CA, Osorio H, Silva L, Gomes C, and David L (2010) Alterations in glycosylation as biomarkers for cancer detection. *J. Clin. Pathol* 63, 322–329. [PubMed: 20354203]
- (10). Steentoft C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KT-BG, Kong Y, Bennett EP, Mandel U, Wandall H, Levery SB, and Clausen H (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered Simple-Cell lines. *Nat. Methods* 8, 977–982. [PubMed: 21983924]
- (11). Julien S, Adriaenssens E, Ottenberg K, Furlan A, Courtand G, Vercoutter-Edouart A-S, Hanisch F-G, Delannoy P, and Le Bourhis X (2006) ST6GalNAc I expression in MDA-MB-231 breast cancer cells greatly modifies their O-glycosylation pattern and enhances their tumorigenicity. *Glycobiology* 16, 54–64. [PubMed: 16135558]
- (12). Perez-Garay M, Arteta B, Pages L, de Llorens R, de Bolos C, Vidal-Vanaclocha F, and Peracaula R (2010) α 2,3-Sialyltransferase ST3Gal III Modulates Pancreatic Cancer Cell Motility and

Adhesion In Vitro and Enhances Its Metastatic Potential In Vivo. PLoS One 5, No. e12524. [PubMed: 20824144]

- (13). Parthasarathy R, Rabuka D, Bertozzi CR, and Groves JT (2007) Molecular Orientation of Membrane-Anchored Mucin Glycoprotein Mimics. J. Phys. Chem. B 111, 12133–12135. [PubMed: 17915910]
- (14). Rabuka D, Forstner MB, Groves JT, and Bertozzi CR (2008) Noncovalent Cell Surface Engineering: Incorporation of Bioactive Synthetic Glycopolymers into Cellular Membranes. J. Am. Chem. Soc 130, 5947–5953. [PubMed: 18402449]
- (15). Woods EC, Yee NA, Shen J, and Bertozzi CR (2015) Glycocalyx Engineering with a Recycling Glycopolymer that Increases Cell Survival In Vivo. Angew. Chem., Int. Ed 54, 15782–15788.
- (16). Brakenhoff RH, Schoenmakers JG, and Lubsen NH (1991) Chimeric cDNA clones: a novel PCR artifact. Nucleic Acids Res. 19, 1949. [PubMed: 2030976]
- (17). Meyerhans A, Vartanian J-P, and Wain-Hobson S (1990) DNA recombination during PCR. Nucleic Acids Res. 18, 1687–1691. [PubMed: 2186361]
- (18). Tang NC, and Chilkoti A (2016) Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. Nat. Mater 15, 419–424. [PubMed: 26726995]
- (19). Ferrari FA, and Cappello J (1997) Biosynthesis of Protein Polymers, in Protein-Based Materials, pp 37–60, Birkhäuser Boston, Cambridge, MA.
- (20). Yoshida A, Suzuki M, Ikenaga H, and Takeuchi M (1997) Discovery of the Shortest Sequence Motif for High Level Mucin-type O-Glycosylation. J. Biol. Chem 272, 16884–16888. [PubMed: 9201996]
- (21). Pei-Xiang X, Prenzoska J, and McKenzie IFC (1992) Epitope mapping of anti-breast and anti-ovarian mucin monoclonal antibodies. Mol. Immunol 29, 641–650. [PubMed: 1374842]
- (22). Zeng Y, Ramya TNC, Dirksen A, Dawson PE, and Paulson JC (2009) High-efficiency labeling of sialylated glycoproteins on living cells. Nat. Methods 6, 207–209. [PubMed: 19234450]
- (23). Wang Y, Ju T, Ding X, Xia B, Wang W, Xia L, He M, and Cummings RD (2010) Cosmc is an essential chaperone for correct protein O-glycosylation. Proc. Natl. Acad. Sci. U. S. A 107, 9228–9233. [PubMed: 20439703]
- (24). Bzymek M, and Lovett ST (2001) Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. U. S. A 98, 8319–8325. [PubMed: 11459970]
- (25). Swallow DM, Gendler S, Griffiths B, Corney G, Taylor-Papadimitriou J, and Bramwell ME (1987) The human tumour-associated epithelial mucins are coded by an expressed hypervariable gene locus PUM. Nature 328, 82–84. [PubMed: 3600778]
- (26). Carvalho F, Seruca R, David L, Amorim A, Seixas M, Bennett E, Clausen H, and Sobrinho-Simoes M (1997) MUC1 gene polymorphism and gastric cancer—an epidemiological study. Glycoconjugate J. 14, 107–111.
- (27). Shurer CR, Colville MJ, Gupta VK, Head SE, Kai F, Lakins JN, and Paszek MJ (2018) Genetically Encoded Toolbox for Glycocalyx Engineering: Tunable Control of Cell Adhesion, Survival, and Cancer Cell Behaviors. ACS Biomater. Sci. Eng 4, 388–399. [PubMed: 29805991]
- (28). Kudelka MR, Antonopoulos A, Wang Y, Duong DM, Song X, Seyfried NT, Dell A, Haslam SM, Cummings RD, and Ju T (2016) Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells. Nat. Methods 13, 81–86. [PubMed: 26619014]
- (29). Litvinov SV, and Hilken J (1993) The epithelial sialomucin, episialin, is sialylated during recycling. J. Biol. Chem 268, 21364–21371. [PubMed: 8407976]
- (30). Kinlough CL, McMahan RJ, Poland PA, Bruns JB, Harkleroad KL, Stremple RJ, Kashlan OB, Weixel KM, Weisz OA, and Hughey RP (2006) Recycling of MUC1 Is Dependent on Its Palmitoylation. J. Biol. Chem 281, 12112–12122. [PubMed: 16507569]
- (31). Mercanti V, Marchetti A, Lelong E, Perez F, Orci L, and Cosson P (2010) Transmembrane domains control exclusion of membrane proteins from clathrin-coated pits. J. Cell Sci 123, 3329–3335. [PubMed: 20826467]
- (32). Kinlough CL, Poland PA, Bruns JB, Harkleroad KL, and Hughey RP (2004) MUC1 Membrane Trafficking Is Modulated by Multiple Interactions. J. Biol. Chem 279, 53071–53077. [PubMed: 15471854]

- (33). Geisler C, and Jarvis DL (2011) Letter to the Glyco-Forum: Effective glycoanalysis with *Maackia amurensis* lectins requires a clear understanding of their binding specificities. *Glycobiology* 21, 988–993. [PubMed: 21863598]
- (34). Brockhausen I, Yang J-M, Burchell J, Whitehouse C, and Taylor-Papadimitriou J (1995) Mechanisms Underlying Aberrant Glycosylation of MUC1Mucin in Breast Cancer Cells. *Eur. J. Biochem* 233, 607–617. [PubMed: 7588808]
- (35). Gerken TA, Gilmore M, and Zhang J (2002) Determination of the Site-specific Oligosaccharide Distribution of the O-Glycans Attached to the Porcine Submaxillary Mucin Tandem Repeat Further Evidence for the Modulation of O-Glycan Side Chain Structures by Peptide Sequence. *J. Biol. Chem* 277, 7736–7751. [PubMed: 11777921]
- (36). Gerken TA (2004) Kinetic Modeling Confirms the Biosynthesis of Mucin Core 1 (β -Gal(1–3) α -GalNAc-O-Ser/Thr) O-glycan Structures Are Modulated by Neighboring Glycosylation Effects. *Biochemistry* 43, 4137–4142. [PubMed: 15065856]
- (37). Clausen H, and Bennett EP (1996) A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology* 6, 635–646. [PubMed: 8922959]
- (38). Granovsky M, Bielfeldt T, Peters S, Paulsen H, Meldal M, Brockhausen J, and Brockhausen I (1994) UDPgalactose:glycoprotein-N-acetyl-d-galactosamine 3- β -d-galactosyltransferase activity synthesizing O-glycan core 1 is controlled by the amino acid sequence and glycosylation of glycopeptide substrates. *Eur. J. Biochem* 221, 1039–1046. [PubMed: 8181460]
- (39). Brockhausen I, Dowler T, and Paulsen H (2009) Site directed processing: Role of amino acid sequences and glycosylation of acceptor glycopeptides in the assembly of extended mucin type O-glycan core 2. *Biochim. Biophys. Acta, Gen. Subj* 1790, 1244–1257.
- (40). Huang KM, and Snider MD (1993) Glycoprotein recycling to the galactosyltransferase compartment of the Golgi complex. *J. Biol. Chem* 268, 9302–9310. [PubMed: 8486626]
- (41). Engelmann K, Kinlough CL, Müller S, Razawi H, Baldus SE, Hughey RP, and Hanisch F-G (2005) Transmembrane and secreted MUC1 probes show trafficking-dependent changes in O-glycan core profiles. *Glycobiology* 15, 1111–1124. [PubMed: 15972891]
- (42). Elhammer AP, Poorman RA, Brown E, Maggiora LL, Hoogerheide JG, and Kézdy FJ (1993) The specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase as inferred from a database of *in vivo* substrates and from the *in vitro* glycosylation of proteins and peptides. *J. Biol. Chem* 268, 10029–10038. [PubMed: 8486674]
- (43). Hema Thanka Christlet T, and Veluraja K (2001) Database Analysis of O-Glycosylation Sites in Proteins. *Biophys. J* 80, 952–960. [PubMed: 11159462]
- (44). Gerken TA, Owens CL, and Pasumarthy M (1998) Site-specific Core 1 O-Glycosylation Pattern of the Porcine Submaxillary Gland Mucin Tandem Repeat Evidence for the Modulation of Glycan Length by Peptide Sequence. *J. Biol. Chem* 273, 26580–26588. [PubMed: 9756896]
- (45). Sihlbom C, van Dijk Härd I, Lidell ME, Noll T, Hansson GC, and Bäckström M (2009) Localization of O-glycans in MUC1 glycoproteins using electron-capture dissociation fragmentation mass spectrometry. *Glycobiology* 19, 375–381. [PubMed: 19095697]
- (46). Nicholls JM, Bourne AJ, Chen H, Guan Y, and Peiris JM (2007) Sialic acid receptor detection in the human respiratory tract: evidence for widespread distribution of potential binding sites for human and avian influenza viruses. *Respir. Res* 8, 73. [PubMed: 17961210]
- (47). Lee ME, DeLoache WC, Cervantes B, and Dueber JE (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol* 4, 975–986. [PubMed: 25871405]
- (48). Paszek MJ, DuFort CC, Rubashkin MG, Davidson MW, Thorn KS, Liphardt JT, and Weaver VM (2012) Scanning angle interference microscopy reveals cell dynamics at the nanoscale. *Nat. Methods* 9, 825–827. [PubMed: 22751201]
- (49). Subedi GP, Johnson RW, Moniz HA, Moremen KW, and Barb A (2015) High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension. *J. Vis. Exp*, No. e53568. [PubMed: 26779721]
- (50). Shurer CR, Kuo JC-H, Roberts LM, Gandhi JG, Colville MJ, Enoki TA, Pan H, Su J, Noble JM, Hollander MJ, O'Donnell JP, Yin R, Pedram K, Möckl L, Kourkoutis LF, Moerner WE, Bertozzi

- CR, Feigenson GW, Reesink HL, and Paszek MJ (2019) Physical Principles of Membrane Shape Regulation by the Glycocalyx. *Cell* 177, 1757–1770. [PubMed: 31056282]
- (51). Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, and Cardona A (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. [PubMed: 22743772]
- (52). Schneider CA, Rasband WS, and Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. [PubMed: 22930834]
- (53). Reichner JS, Whiteheart SW, and Hart GW (1988) Intracellular trafficking of cell surface sialoglycoconjugates. *J. Biol. Chem* 263, 16316–16326. [PubMed: 3182795]
- (54). Fukuda M (2001) Beta-elimination for release of O-GalNAc-linked oligosaccharides from glycoproteins and glycopeptides, in *Current Protocols in Molecular Biology*, Vol. 31, pp 17.15.4–17.15.9, John Wiley & Sons, Inc.: Hoboken, NJ, .
- (55). Shajahan A, Heiss C, Ishihara M, and Azadi P (2017) Glycomic and glycoproteomic analysis of glycoproteins-a tutorial. *Anal. Bioanal Chem* 409, 4483–4505. [PubMed: 28585084]

Figure 1.

Combinatorial genetic encoded library for sequence-specific mucins. (a) Schematic diagram of the combinatorial sequence-specific mucins. (b) Schematic shows the swappable biobricks and flanking restriction sites for complete mucin construction. (c) Workflow for the design and fabrication of cDNAs for the mucin tandem-repeat backbones. (d) Summary of codon-scrambled mucin backbones in the library.

Figure 2.

Construction and validation of sequence-specific mucin expression. (a) Components and features of codon-optimized Muc1 variants with GFP reporters. (b) Predicted molecular weight of the polypeptide backbone. (c) Biosynthesis of Tn antigen, Core 1, and Core 2 glycans, and specificity of relevant lectins for their detection. (d) Western blot analysis of Native Muc1 expression and glycosylation in wild-type and Core-1 β -3-T specific molecular chaperone (COSMC) knockout MCF10A cells. The MCF10A cells were stably transfected with native Muc1. The surface sialic acids were labeled with AFDye 568 through periodate labeling prior to lysate collection. The blot was stained in multiple colors with MUC1 TR (CD227 HPMV), Ab-FITC, and PNA-CF640 or biotinylated VVA (Secondary: NeutrAvidin-Dylight 650). (e) Western blot analysis of native and codon-scrambled Muc1 in extracts of transiently transfected HEK293T cells. (f) Immunofluorescence images of transiently transfected HEK293T cells expressing indicated constructs and probed with PNA lectin (left), anti-Muc1 antibody (center left), GFP (center right), and Hoescht nuclear stain (right)

(scale bar 10 μm). (g) PNA lectin blot analysis (left) and intensity profiles (right) of mucins of varying sizes in extracts of transiently transfected HEK293T cells.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3.

Engineering the frequency of glycosylation sites in the Muc1 polymer backbone tunes O-glycan maturation. (a) Components and features of secreted Muc1 and engineered variants each with 21 tandem repeats. (b) Tandem repeat sequences of secreted mucin mutants and the molecular weight of the polypeptide backbones. Single, double, and triple glycosylation mutants (sMuc1S, sMuc1D, and sMuc1T) have one, two, or three, serine/threonine (S/T) to alanine substitutions per repeat, respectively. (c) Representative Western blot analysis of affinity-purified recombinant secreted mucins from FreeStyle 293-F cell culture media probed with anti-SUMOstar antibody and PNA, s-WGA, and VVA lectins (of three independent experiments). The lectin blot was costained in multiple colors with PNA-Alexa Fluor 568, s-WGA-FITC, and biotinylated VVA (Secondary: NeutrAvidin-Dylight 650). (d) Representative fluorescence intensity electrophoretograms of the blots in part c. (e) Ratiometric intensity analysis of PNA to VVA signal (upper) and s-WGA to VVA signal (lower) for the indicated mucins and their corresponding frequency of S/T glycosylation sites in the polymer backbone. Ratiometric fluorescence intensity was quantified along each lane and normalized to the signal from the secreted mucin with wild-type Muc1 tandem repeats (sMuc1); data presented as the mean and SEM from at least three independent experiments. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. (f) Left: MALDI-TOF mass spectra registered for samples of permethylated glycan alditols from secreted mucins with wild-type

Muc1 tandem repeats (sMuc1) and triple mutant (sMuc1T) from HEK293T cell culture media. The ion signals were annotated with respect to the relative masses of molecular ions (m/z) detected as sodium adducts and by assignment of the respective core structure (red for Core 1 and black for Core 2). Right: Schematic presentation of O-linked glycans detected on the secreted mucins.

Figure 4.

Designer mucin domains reveal sequence-specific effects on glycosylation. (a) Components and features of designer mucins. (b) Predicted molecular weight of the mucin polypeptide backbones. (c) Representative Western blot analysis (from three independent experiments) of indicated constructs in extracts of transiently transfected HEK293T cells probed with anti-GFP antibody or costained with PNA and VVA lectins. (d) Representative fluorescence intensity electrophoretograms of the Western blots in part c for indicated constructs from three independent experiments. Dashed lines indicate the peak of the glycoform visible in the PNA blot. Shaded boxes indicate the regions between the bands on the anti-GFP blot with the highest and second highest apparent molecular weights. (e) Ratiometric intensity analysis of PNA to VVA staining for the indicated mucins and their corresponding frequency of serine and threonine glycosylation sites in the polymer backbone. Fluorescence intensity was quantified along each lane of the dual-probed lectin blot, and the PNA/VVA ratio was normalized to that of the KEPAITTP x20 mucin; data presented as the mean and SEM from three independent experiments. (f) The fold change in the PNA/VVA ratio with doubling the indicated mucin backbone size from 40 to 80 tandem repeats; data presented as the mean and SEM from three independent experiments. * $p < 0.05$

Figure 5.

Tuning mucin glycosylation through cytoplasmic tail engineering. (a) Components and features of cell-surface mucins with synthetic 21-amino-acid transmembrane anchors (TM21) and engineered cytoplasmic motifs; native CT refers to a native cytoplasmic tail adapted from Muc1. (b) Lectin blot analysis of the indicated mucin isoforms from transiently transfected HEK293T cells to detect sialylated O-glycans by periodate oxidation and Core-I structures by PNA; blots are representative of three independent experiments. (c) PNA-lectin blot analysis of the indicated mucin isoforms before and after sialidase treatment; blots are representative of three independent experiments. (d) Top: Representative

MAA and PNA lectin blot analysis (from four independent experiments) of the indicated mucin isoforms immunoprecipitated from transiently transfected HEK293T cells. Bottom: Ratiometric intensity of sialic acid to Core 1 glycan signal (MAA/PNA); data presented as the mean and SEM from four independent experiments. * $P < 0.05$.