Sequence-specific Mucins for Glycocalyx Engineering

Journal:	*ACS Synthetic Biology*
Manuscript ID	sb-2019-00127s.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Pan, Hao; Cornell University, Field of Biophysics Colville, Marshall; Cornell University, Biophysics Paszek, Matthew; Cornell University, Robert Frederick Smith School of Chemical and Biomolecular Engineering

Sequence-specific Mucins for Glycocalyx Engineering

Hao Pan[1], Marshall J. Colville[1] and Matthew J. Paszek[1,2,3*]

[1]Field of Biophysics, Cornell University, Ithaca, NY 14853, USA

[2]Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA
[3]Nancy E. and Peter C. Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853

*Corresponding author

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.

Key words

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 bio-active 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 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 bio-mimetic 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 KEPAPTTP 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 labelled 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 β3-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 460kDa band at the top of the smear (Figure 2e). Based on these results, we concluded that the 460kDa band was fully glycosylated Muc1, while the smear represented a heterogenous 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 non-repetitive 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 amongst individuals[25] 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 spectroscopy.

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 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, co-stained 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 towards 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 non-fully 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 KEPAPTTP 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 12-33%. 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 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 grey 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 heterogenous 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 recycling 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 tail[31] 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 labelled the sialic acids on the cell surface with PAL. On lectin blots, the PAL sialic acid signal was strongest at approximately 171kDa, 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 ultra-high molecular weight glyoforms that were reactive to MAA (Figure 5d Top). The MAA-reactive, ultra-high molecular weight glycoforms were promoted by recycling motifs. We found that the inclusion of the CQC motif led to a 2-fold increase in 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 ultra-high molecular weight glycoforms were enriched in a2,3-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-glyan 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. Knock-out 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 ultra-high 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 SO_4-3-Galβ[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 glycocalyx 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-IgGκ binding protein – horseradish peroxidase (HRP; 516102, Santa Cruz), goat anti-mouse IgG (Alexa Fluor™ 647 conjugated, A-21235; Alexa Fluor™ 488 conjugated, A-11001; Alexa Fluor™ 568 conjugated, A-11004; ThermoFisher) and goat anti-

chicken 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), fluorescein-labeled 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 piggybac 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 BamHI and EcoRI restriction sites. For generation of pPB Muc1 mOxGFP dCT TetOn Puro, the cDNA for mOxGFP (Addgene #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 non-directionally into the BlpI 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-SUMOStar tag (6x His Sumostar Muc1)was created through custom gene synthesis (General Biosystems) and cloned into the tetracycline-inducible piggybac 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/codon-scrambler)[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 BamHI and Bsu36I restriction sites, pcDNA3.1 Muc1 TM21 using the BsrGI and Bsu36I restriction sites, or pPB 6x His Sumostar Muc1 using BsrGI and Bsu36I restriction sites (See 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 BamHI and Bsu36I 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'-

TGGAGGAGCCTCAGGCATACTTTATTG-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'-CCGAAAGTAGGAATTCGGGCCCGTTTAAACCCGC-3' (forward) and 5'-CGGCACTGACATCTAGAGTACCACAACAAAGCCAGGC-3' (reverse) primers. The cDNA of native CT was subcloned into the XbaI and EcoRI 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 Golden Gate Assembly. Two pairs of custom primers for tandem repeats and complete mucin vector were designed to attach BsmbI 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). 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 (ThermoFisher) supplemented with 5% horse serum (ThermoFisher), 20 ng/mL EGF (Peprotech), 10 μg/ml insulin (Sigma), 500 ng/mL hydrocortisone (Sigma), and 100 ng/mL cholera toxin (Sigma). HEK293T cells were cultured in DMEM (ThermoFisher) supplemented with 10% fetal bovine serum (ThermoFisher). Cells were maintained at 37 ^{o}C,

5% CO_2, and 90% Relative humidity (RH). FreeStyle™ 293-F cells were cultured in suspension in FreeStyle™ 293 Expression Medium (ThermoFisher). Suspension cultures were maintained in an orbital shaker at 37 °C, 8% CO_2, 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/cm^2 and transfected with calcium phosphate for 24-36 hrs before lysis with cell lysis buffer. MCF10A cells were plated at 20,000 cells/cm^2 and induced with 0.2 μg/mL doxycycline for 24 hrs 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 μg/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 μg/mL in 5% BSA TBST and incubated for 1 hr 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 36hrs of transfection. Cells were washed with cold DPBS with $Ca^{2+}$and Mg^{2+} followed by a 10-minute 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 μM 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/cm^2 and transfected with the calcium phosphate method for 24-36 hrs 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 manufacturer's instructions and the suspension was incubated at 4 °C for 3 hrs. The beads were washed 2 times with lysis buffer and then resuspended in 4x LDS loading buffer. The resuspension was subsequently analyzed by Western blot.

Sialidase Treatment of HEK293Ts

HEK293T cells were collected 24 hrs after transfection and incubated with *Arthrobacter ureafaciens* sialidase (Roche, 10mU, 100μl final volume) in sialidase buffer[53] for 30mins at 37°C before lysis with cell lysis buffer.

Immunofluorescence

HEK293T cells were plated at 45,000 cells/cm^2 and transfected with calcium phosphate for 24 hrs 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 hrs at room temperature. Samples were imaged on a Zeiss LSM inverted 880 confocal microscope using a 40x water immersion objective (NA 1.1).

Secreted Mucin Protein Expression, Purification

16.25 µg pPB 6x His Sumostar Muc1 DNAs were transfected into HEK293T cells in 10-cm culture dishes for 48 hrs. 30 µg pPB 6x His Sumostar Muc1 DNAs were transfected into 20mL 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, 250mM imidazole). The eluted sample was diafiltrated into PBS with 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 protein was denatured by heating at 100 °C for 5 min. The denatured protein was subsequently treated with 19 mg sodium borohydride ($NaBH_4$) in 500 µL of 50 mM sodium hydroxide (NaOH) solution at 45 °C for 18 hrs. The samples were cooled, neutralized with 10 % acetic acid, passed through a Dowex H+ resin column, and lyophilized with sites removed under the stream of nitrogen. The glycans were permethylated for structural characterization by mass spectrometry. 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.

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 hour 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 μM peracetylated BnGalNAc. After 48 hours the media was aspirated and loose cells and debris 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 permethylation2. A sodium hydroxide slurry in DMSO was freshly prepared and 200 μL added to each dry sample followed by 100 μL methyl iodide (ACROS). The

samples were mixed continuously for 10 mins then the reaction halted by the addition of 600 µL deionized water. Permethylated benzyl glycans were recovered by extraction with 200 µL chloroform then washed 4 times with 800 µL deionized water. The samples were further purified by C18 gravity chromatography (Waters Sep-Pak C18 1 cc) and dried by 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).

Acknowledgement

We thank V. Weaver and J. Lakins for the lentiviral and transposon plasmids, as well as helpful discussions. This investigation was supported by National Institute of Health New Innovator DP2 GM229133 (M.J.P.) and National Cancer Institute U54 CA210184 (M.J.P.). Imaging data was acquired through the Cornell University Biotechnology Resource Center, with NYSTEM C029155 and NIH 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 (Parastoo Azadi). MALDI-TOF-MS analysis of CORA samples

was performed by the Biopolymers and Proteomics laboratory at Koch Institute for Integrative Cancer Research, MIT.

Author Information

Matthew J. Paszek (mjp31@cornell.edu)

Address: 120 Olin Hall, 113 Ho Plaza, Cornell University, Ithaca NY 14850

Conflict of Interest: This work is part of a provisional patent filed by Cornell University.

Supporting Information

Figure S1-6; Table S1-2; DNA and amino acid sequences of "biobricks"; summary of constructs available

References

(1) Brockhausen, I.; Schachter, H.; Stanley, P. O-GalNAc Glycans. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2009.

(2) Lichtenberger, L. M. The Hydrophobic Barrier Properties of Gastrointestinal Mucus. *Annu. Rev. Physiol.* **1995**, *57* (1), 565–583. https://doi.org/10.1146/annurev.ph.57.030195.003025.

(3) Hudak, J. E.; Canham, S. M.; Bertozzi, C. R. Glycocalyx Engineering Reveals a Siglec-Based Mechanism for NK Cell Immunoevasion. *Nature Chemical Biology* **2014**, *10* (1), 69–75. https://doi.org/10.1038/nchembio.1388.

(4) Paszek, M. J.; DuFort, C. C.; Rossier, O.; Bainer, R.; Mouw, J. K.; Godula, K.; Hudak, J. E.; Lakins, J. N.; Wijekoon, A. C.; Cassereau, L.; et al. The Cancer Glycocalyx Mechanically Primes Integrin-Mediated Growth and Survival. *Nature* **2014**, *511* (7509), 319–325. https://doi.org/10.1038/nature13535.

(5) Polefka, T. G.; Garrick, R. A.; Redwood, W. R.; Swislocki, N. I.; Chinard, F. P. Solute-Excluded Volumes near the Novikoff Cell Surface. *American Journal of Physiology-Cell Physiology* **1984**, *247* (5), C350–C356. https://doi.org/10.1152/ajpcell.1984.247.5.C350.

(6) Kramer, J. R.; Onoa, B.; Bustamante, C.; Bertozzi, C. R. Chemically Tunable Mucin Chimeras Assembled on Living Cells. *PNAS* **2015**, *112* (41), 12574–12579. https://doi.org/10.1073/pnas.1516127112.

(7) Coltart, D. M.; Royyuru, A. K.; Williams, L. J.; Glunz, P. W.; Sames, D.; Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Danishefsky, S. J.; Live, D. H. Principles of Mucin Architecture:

Structural Studies on Synthetic Glycopeptides Bearing Clustered Mono-, Di-, Tri-, and Hexasaccharide Glycodomains. *J. Am. Chem. Soc.* **2002**, *124* (33), 9833–9844. https://doi.org/10.1021/ja020208f.

(8) Dennis, J. W.; Granovsky, M.; Warren, C. E. Protein Glycosylation in Development and Disease. *BioEssays* **1999**, *21* (5), 412–421. https://doi.org/10.1002/(SICI)1521-1878(199905)21:5<412::AID-BIES8>3.0.CO;2-5.

(9) Reis, C. A.; Osorio, H.; Silva, L.; Gomes, C.; David, L. Alterations in Glycosylation as Biomarkers for Cancer Detection. *Journal of Clinical Pathology* **2010**, *63* (4), 322–329. https://doi.org/10.1136/jcp.2009.071035.

(10) Steentoft, C.; Vakhrushev, S. Y.; Vester-Christensen, M. B.; Schjoldager, K. T.-B. G.; Kong, Y.; Bennett, E. P.; Mandel, U.; Wandall, H.; Levery, S. B.; Clausen, H. Mining the O-Glycoproteome Using Zinc-Finger Nuclease–Glycoengineered SimpleCell Lines. *Nature Methods* **2011**, *8* (11), 977–982. https://doi.org/10.1038/nmeth.1731.

(11) Julien, S.; Adriaenssens, E.; Ottenberg, K.; Furlan, A.; Courtand, G.; Vercoutter-Edouart, A.-S.; Hanisch, F.-G.; Delannoy, P.; Le Bourhis, X. ST6GalNAc I Expression in MDA-MB-231 Breast Cancer Cells Greatly Modifies Their O-Glycosylation Pattern and Enhances Their Tumourigenicity. *Glycobiology* **2006**, *16* (1), 54–64. https://doi.org/10.1093/glycob/cwj033.

(12) Pérez-Garay, M.; Arteta, B.; Pagès, L.; Llorens, R. de; Bolòs, C. de; Vidal-Vanaclocha, F.; Peracaula, R. A2,3-Sialyltransferase ST3Gal III Modulates Pancreatic Cancer Cell Motility and Adhesion In Vitro and Enhances Its Metastatic Potential In Vivo. *PLOS ONE* **2010**, *5* (9), e12524. https://doi.org/10.1371/journal.pone.0012524.

(13) Parthasarathy, R.; Rabuka, D.; Bertozzi, C. R.; Groves, J. T. Molecular Orientation of Membrane-Anchored Mucin Glycoprotein Mimics. *J. Phys. Chem. B* **2007**, *111* (42), 12133–12135. https://doi.org/10.1021/jp072136q.

(14) Rabuka, D.; Forstner, M. B.; Groves, J. T.; Bertozzi, C. R. Noncovalent Cell Surface Engineering: Incorporation of Bioactive Synthetic Glycopolymers into Cellular Membranes. *J. Am. Chem. Soc.* **2008**, *130* (18), 5947–5953. https://doi.org/10.1021/ja710644g.

(15) Woods, E. C.; Yee, N. A.; Shen, J.; Bertozzi, C. R. Glycocalyx Engineering with a Recycling Glycopolymer That Increases Cell Survival In Vivo. *Angewandte Chemie International Edition* **2015**, *54* (52), 15782–15788. https://doi.org/10.1002/anie.201508783.

(16) Brakenhoff, R. H.; Schoenmakers, J. G.; Lubsen, N. H. Chimeric CDNA Clones: A Novel PCR Artifact. *Nucleic Acids Res* **1991**, *19* (8), 1949.

(17) Meyerhans, A.; Vartanian, J.-P.; Wain-Hobson, S. DNA Recombination during PCR. *Nucleic Acids Res* **1990**, *18* (7), 1687–1691. https://doi.org/10.1093/nar/18.7.1687.

(18) Tang, N. C.; Chilkoti, A. Combinatorial Codon Scrambling Enables Scalable Gene Synthesis and Amplification of Repetitive Proteins. *Nat Mater* **2016**, *15* (4), 419–424. https://doi.org/10.1038/nmat4521.

(19) Ferrari, F. A.; Cappello, J. Biosynthesis of Protein Polymers. In *Protein-Based Materials*; Bioengineering of Materials; Birkhäuser Boston, 1997; pp 37–60. https://doi.org/10.1007/978-1-4612-4094-5_2.

(20) Yoshida, A.; Suzuki, M.; Ikenaga, H.; Takeuchi, M. Discovery of the Shortest Sequence Motif for High Level Mucin-Type O-Glycosylation. *J. Biol. Chem.* **1997**, *272* (27), 16884–16888. https://doi.org/10.1074/jbc.272.27.16884.

(21) Pei-Xiang, X.; Prenzoska, J.; Mckenzie, I. F. C. Epitope Mapping of Anti-Breast and Anti-Ovarian Mucin Monoclonal Antibodies. *Molecular Immunology* **1992**, *29* (5), 641–650. https://doi.org/10.1016/0161-5890(92)90201-8.

(22) Zeng, Y.; Ramya, T. N. C.; Dirksen, A.; Dawson, P. E.; Paulson, J. C. High-Efficiency Labeling of Sialylated Glycoproteins on Living Cells. *Nature Methods* **2009**, *6* (3), 207–209. https://doi.org/10.1038/nmeth.1305.

(23) Wang, Y.; Ju, T.; Ding, X.; Xia, B.; Wang, W.; Xia, L.; He, M.; Cummings, R. D. Cosmc Is an Essential Chaperone for Correct Protein O-Glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (20), 9228–9233. https://doi.org/10.1073/pnas.0914004107.

(24) Bzymek, M.; Lovett, S. T. Instability of Repetitive DNA Sequences: The Role of Replication in Multiple Mechanisms. *PNAS* **2001**, *98* (15), 8319–8325. https://doi.org/10.1073/pnas.111008398.

(25) Swallow, D. M.; Gendler, S.; Griffiths, B.; Corney, G.; Taylor-Papadimitriou, J.; Bramwell, M. E. The Human Tumour-Associated Epithelial Mucins Are Coded by an Expressed Hypervariable Gene Locus PUM. *Nature* **1987**, *328* (6125), 82–84. https://doi.org/10.1038/328082a0.

(26) Carvalho, F.; Seruca, R.; David, L.; Amorim, A.; Seixas, M.; Bennett, E.; Clausen, H.; Sobrinho-Simoes, M. MUC1 Gene Polymorphism and Gastric Cancer–an Epidemiological Study. *Glycoconj J* **1997**, *14* (1), 107–111. https://doi.org/10.1023/A:1018573201347.

(27) Shurer, C. R.; Colville, M. J.; Gupta, V. K.; Head, S. E.; Kai, F.; Lakins, J. N.; Paszek, M. J. Genetically Encoded Toolbox for Glycocalyx Engineering: Tunable Control of Cell Adhesion, Survival, and Cancer Cell Behaviors. *ACS Biomater. Sci. Eng.* **2017**. https://doi.org/10.1021/acsbiomaterials.7b00037.

(28) Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells | Nature Methods https://www.nature.com/articles/nmeth.3675 (accessed Jun 1, 2019).

(29) Litvinov, S. V.; Hilkens, J. The Epithelial Sialomucin, Episialin, Is Sialylated during Recycling. *J. Biol. Chem.* **1993**, *268* (28), 21364–21371.

(30) Kinlough, C. L.; McMahan, R. J.; Poland, P. A.; Bruns, J. B.; Harkleroad, K. L.; Stremple, R. J.; Kashlan, O. B.; Weixel, K. M.; Weisz, O. A.; Hughey, R. P. Recycling of MUC1 Is Dependent on Its Palmitoylation. *J. Biol. Chem.* **2006**, *281* (17), 12112–12122. https://doi.org/10.1074/jbc.M512996200.

(31) Mercanti, V.; Marchetti, A.; Lelong, E.; Perez, F.; Orci, L.; Cosson, P. Transmembrane Domains Control Exclusion of Membrane Proteins from Clathrin-Coated Pits. *J Cell Sci* **2010**, *123* (19), 3329–3335. https://doi.org/10.1242/jcs.073031.

(32) Kinlough, C. L.; Poland, P. A.; Bruns, J. B.; Harkleroad, K. L.; Hughey, R. P. MUC1 Membrane Trafficking Is Modulated by Multiple Interactions. *J. Biol. Chem.* **2004**, *279* (51), 53071–53077. https://doi.org/10.1074/jbc.M409360200.

(33) Geisler, C.; Jarvis, D. L. Letter to the Glyco-Forum: Effective Glycoanalysis with Maackia Amurensis Lectins Requires a Clear Understanding of Their Binding Specificities. *Glycobiology* **2011**, *21* (8), 988–993. https://doi.org/10.1093/glycob/cwr080.

(34) Brockhausen, I.; Yang, J.-M.; Burchell, J.; Whitehouse, C.; Taylor-Papadimitriou, J. Mechanisms Underlying Aberrant Glycosylation of MUC1 Mucin in Breast Cancer Cells. *European Journal of Biochemistry* **1995**, *233* (2), 607–617. https://doi.org/10.1111/j.1432-1033.1995.607_2.x.

(35) Gerken, T. A.; Gilmore, M.; Zhang, J. 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.* **2002**, *277* (10), 7736–7751. https://doi.org/10.1074/jbc.M111690200.

(36) Gerken, T. A. 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* **2004**, *43* (14), 4137–4142. https://doi.org/10.1021/bi036306a.

(37) Clausen, H.; Bennett, E. P. A Family of UDP-GalNAc: Polypeptide N-Acetylgalactosaminyl-Transferases Control the Initiation of Mucin-Type O-Linked Glycosylation. *Glycobiology* **1996**, *6* (6), 635–646. https://doi.org/10.1093/glycob/6.6.635.

(38) Granovsky, M.; Bielfeldt, T.; Peters, S.; Paulsen, H.; Meldal, M.; Brockhausen, J.; Brockhausen, I. 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. *European Journal of Biochemistry* **1994**, *221* (3), 1039–1046. https://doi.org/10.1111/j.1432-1033.1994.tb18822.x.

(39) Brockhausen, I.; Dowler, T.; Paulsen, H. Site Directed Processing: Role of Amino Acid Sequences and Glycosylation of Acceptor Glycopeptides in the Assembly of Extended Mucin Type O-Glycan Core 2. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2009**, *1790* (10), 1244–1257. https://doi.org/10.1016/j.bbagen.2009.05.020.

(40) Huang, K. M.; Snider, M. D. Glycoprotein Recycling to the Galactosyltransferase Compartment of the Golgi Complex. *J. Biol. Chem.* **1993**, *268* (13), 9302–9310.

(41) Engelmann, K.; Kinlough, C. L.; Müller, S.; Razawi, H.; Baldus, S. E.; Hughey, R. P.; Hanisch, F.-G. Transmembrane and Secreted MUC1 Probes Show Trafficking-Dependent Changes in O-Glycan Core Profiles. *Glycobiology* **2005**, *15* (11), 1111–1124. https://doi.org/10.1093/glycob/cwi099.

(42) Elhammer, A. P.; Poorman, R. A.; Brown, E.; Maggiora, L. L.; Hoogerheide, J. G.; Kézdy, F. J. 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.* **1993**, *268* (14), 10029–10038.

(43) Hema Thanka Christlet, T.; Veluraja, K. Database Analysis of O-Glycosylation Sites in Proteins. *Biophysical Journal* **2001**, *80* (2), 952–960. https://doi.org/10.1016/S0006-3495(01)76074-2.

(44) Gerken, T. A.; Owens, C. L.; Pasumarthy, M. 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.* **1998**, *273* (41), 26580–26588. https://doi.org/10.1074/jbc.273.41.26580.

(45) Sihlbom, C.; van Dijk Härd, I.; Lidell, M. E.; Noll, T.; Hansson, G. C.; Bäckström, M. Localization of O-Glycans in MUC1 Glycoproteins Using Electron-Capture Dissociation

Fragmentation Mass Spectrometry. *Glycobiology* **2009**, *19* (4), 375–381. https://doi.org/10.1093/glycob/cwn144.

(46) Nicholls, J. M.; Bourne, A. J.; Chen, H.; Guan, Y.; Peiris, J. M. 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* **2007**, *8* (1), 73. https://doi.org/10.1186/1465-9921-8-73.

(47) Lee, M. E.; DeLoache, W. C.; Cervantes, B.; Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **2015**, *4* (9), 975–986. https://doi.org/10.1021/sb500366v.

(48) Paszek, M. J.; DuFort, C. C.; Rubashkin, M. G.; Davidson, M. W.; Thorn, K. S.; Liphardt, J. T.; Weaver, V. M. Scanning Angle Interference Microscopy Reveals Cell Dynamics at the Nanoscale. *Nat Meth* **2012**, *9* (8), 825–827. https://doi.org/10.1038/nmeth.2077.

(49) Subedi, G. P.; Johnson, R. W.; Moniz, H. A.; Moremen, K. W.; Barb, A. High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension. *J Vis Exp* **2015**, No. 106. https://doi.org/10.3791/53568.

(50) Shurer, C. R.; Kuo, J. C.-H.; Roberts, L. M.; Gandhi, J. G.; Colville, M. J.; Enoki, T. A.; Pan, H.; Su, J.; Noble, J. M.; Hollander, M. J.; et al. Physical Principles of Membrane Shape Regulation by the Glycocalyx. *Cell* **2019**, *177* (7), 1757-1770.e21. https://doi.org/10.1016/j.cell.2019.04.017.

(51) Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nature Methods* **2012**, *9* (7), 676–682. https://doi.org/10.1038/nmeth.2019.

(52) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis https://www.nature.com/articles/nmeth.2089 (accessed Oct 31, 2018). https://doi.org/10.1038/nmeth.2089.

(53) Reichner, J. S.; Whiteheart, S. W.; Hart, G. W. Intracellular Trafficking of Cell Surface Sialoglycoconjugates. *J. Biol. Chem.* **1988**, *263* (31), 16316–16326.

For Table of Contents Only

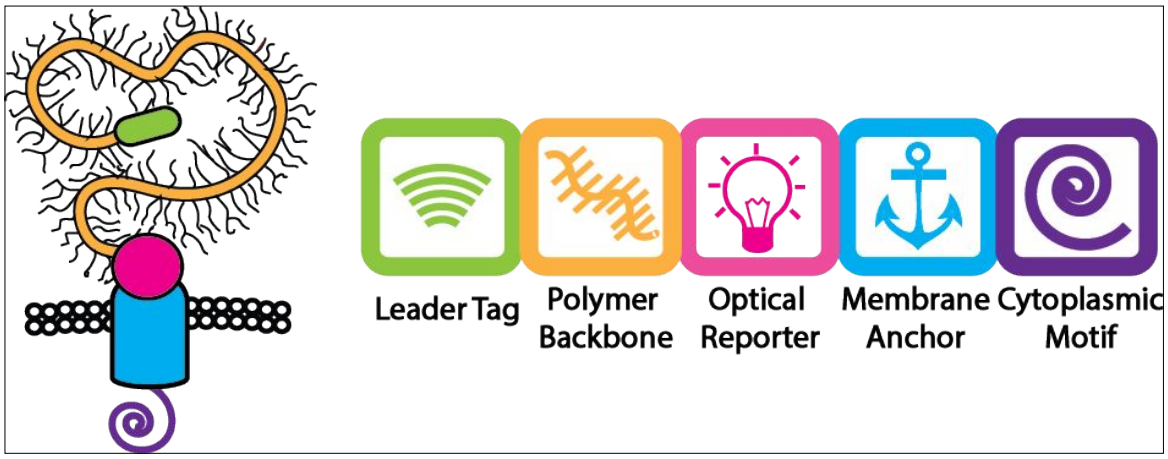
Sequence-specific Mucins for Glycocalyx Engineering

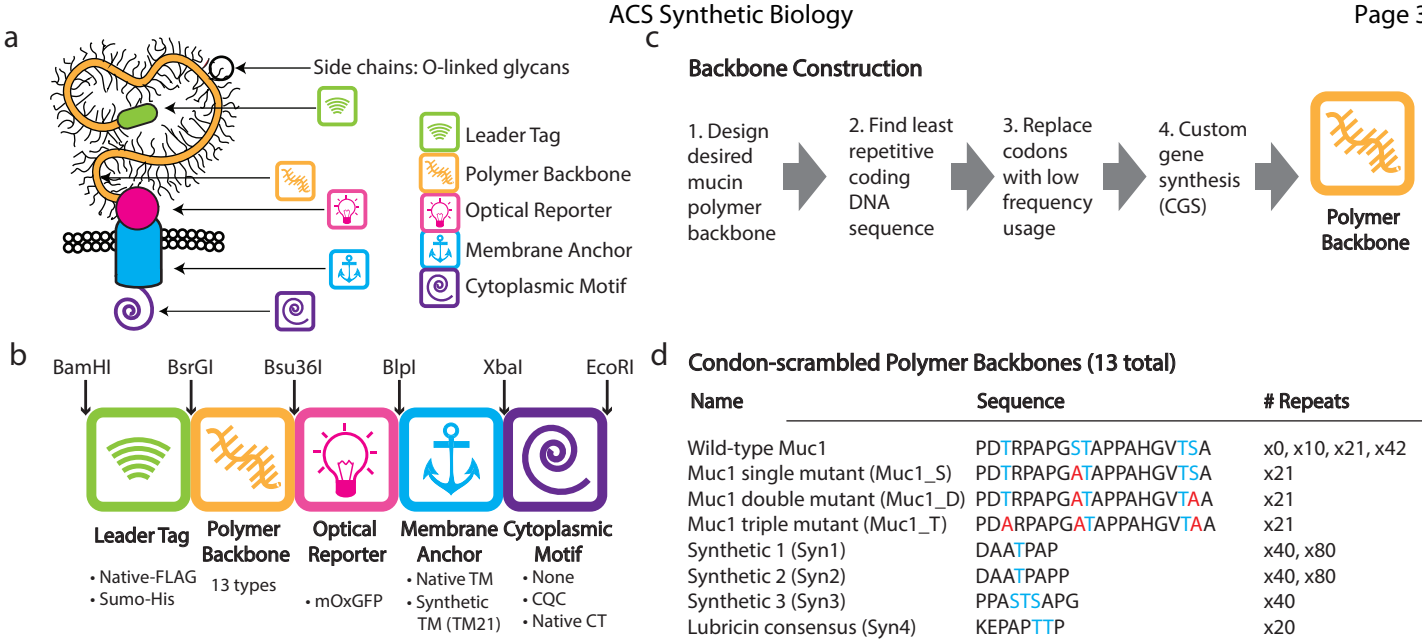
Hao Pan[1], Marshall J. Colville[1] and Matthew J. Paszek[1,2,3*]

[1]Field of Biophysics, Cornell University, Ithaca, NY 14853, USA

[2]Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA
[3]Nancy E. and Peter C. Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853

*Corresponding author





d

Condon-scrambled Polymer Backbones (13 total)

Name	Sequence	# Repeats	%S/T
Wild-type Muc1	PDTRPAPGSTAPPAHGVTSA	x0, x10, x21, x42	25.0
Muc1 single mutant (Muc1_S)	PDTRPAPGATAPPAHGVTSA	x21	20.0
Muc1 double mutant (Muc1_D)	PDTRPAPGATAPPAHGVTAA	x21	15.0
Muc1 triple mutant (Muc1_T)	PDARPAPGATAPPAHGVTAA	x21	10.0
Synthetic 1 (Syn1)	DAATPAP	x40, x80	14.2
Synthetic 2 (Syn2)	DAATPAPP	x40, x80	12.5
Synthetic 3 (Syn3)	PPASTSAPG	x40	33.3
Lubricin consensus (Syn4)	KEPAPTTP	x20	25.0

Figure 1: Combinatorial Genetic Encoded Library for Sequence-Specific Mucins. (a) Schematic diagram of the combinatorial sequence-specific mucins. (b) Schematic shows the swappable bio-bricks and flanking restriction sites for complete mucin construction. (c) Work flow for the design and fabrication of cDNAs for the mucin tandem-repeat backbones. (d) Summary of codon-scrambled mucin backbones in the library.

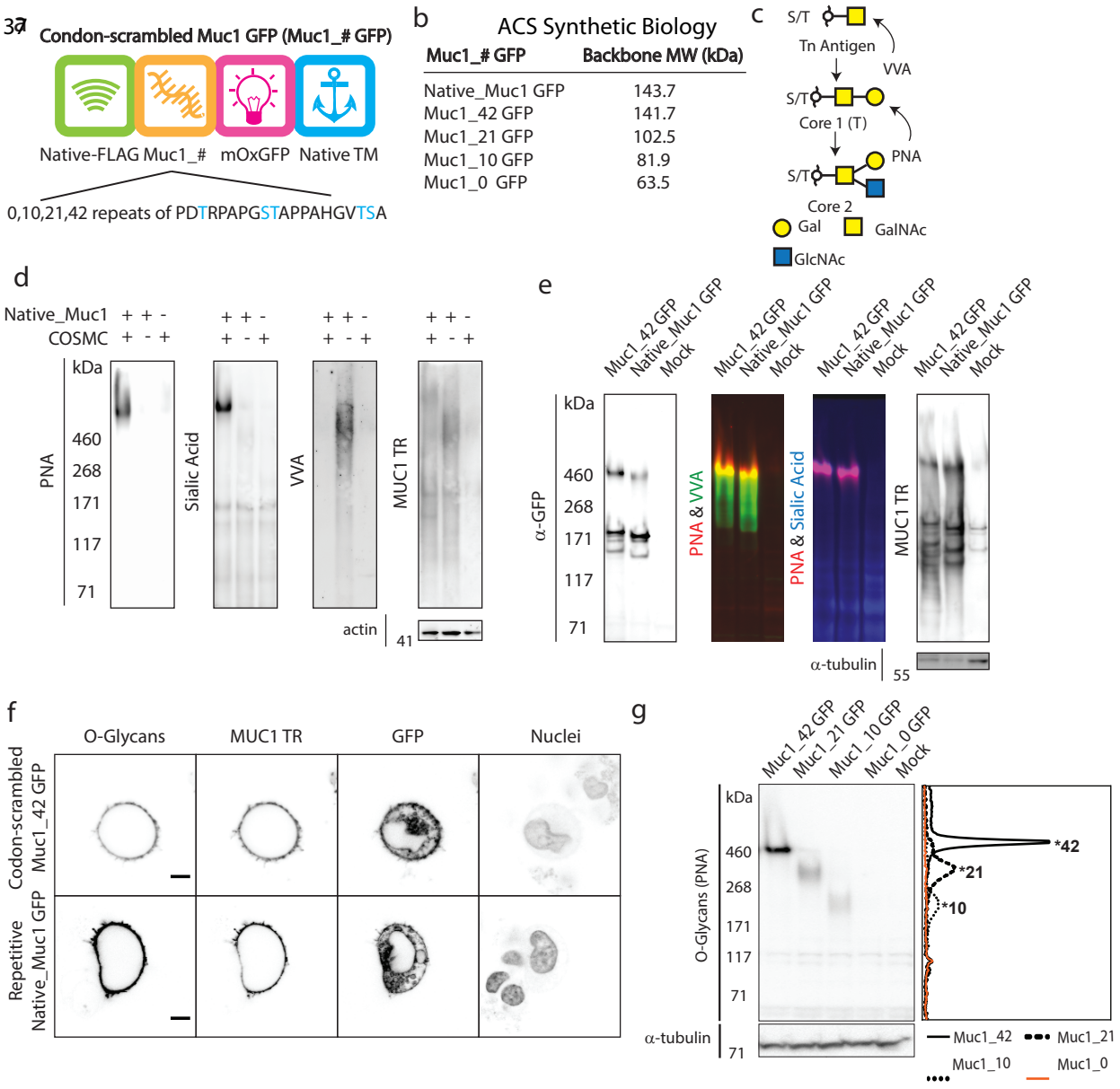


Figure 2: Construction and Validation of Sequence-Specific Mucin Expression. (a) Components and features of codon-optimized Muc1 variants with GFP reporters. (b) Predicted Molecular Weight of the polypeptide backbone. (c) Biosynthesis of Tn antigen, Core 1, and Core 2 glycans, and specificity of relevant lectins for their detection. (d)Western Blot analysis of Native Muc1 expression and glycosylation in wild-type and Core-1 β3-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 trnsfected 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.

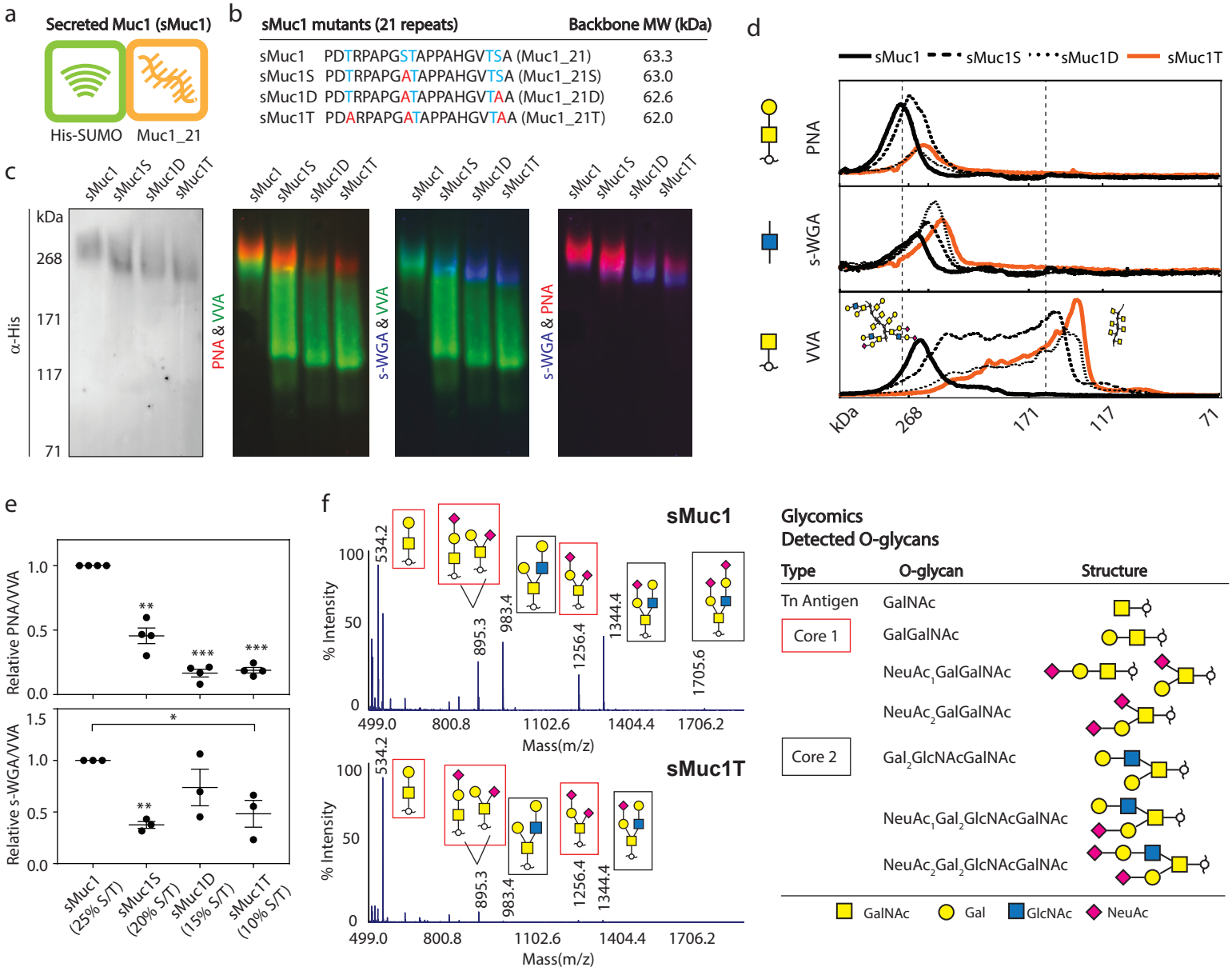


Figure 3: Engineering the Frequency of Glycosylation Sites in the Muc1 Polymer Backbone Tunes O-glycan Maturation. (a) Components and features of secreted Muc1 and engineered variants each with 21 tandem repeats. (b) Tandem repeat sequences of secreted mucin mutants and the molecular weight of the polypeptide 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 co-stained 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 (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 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.

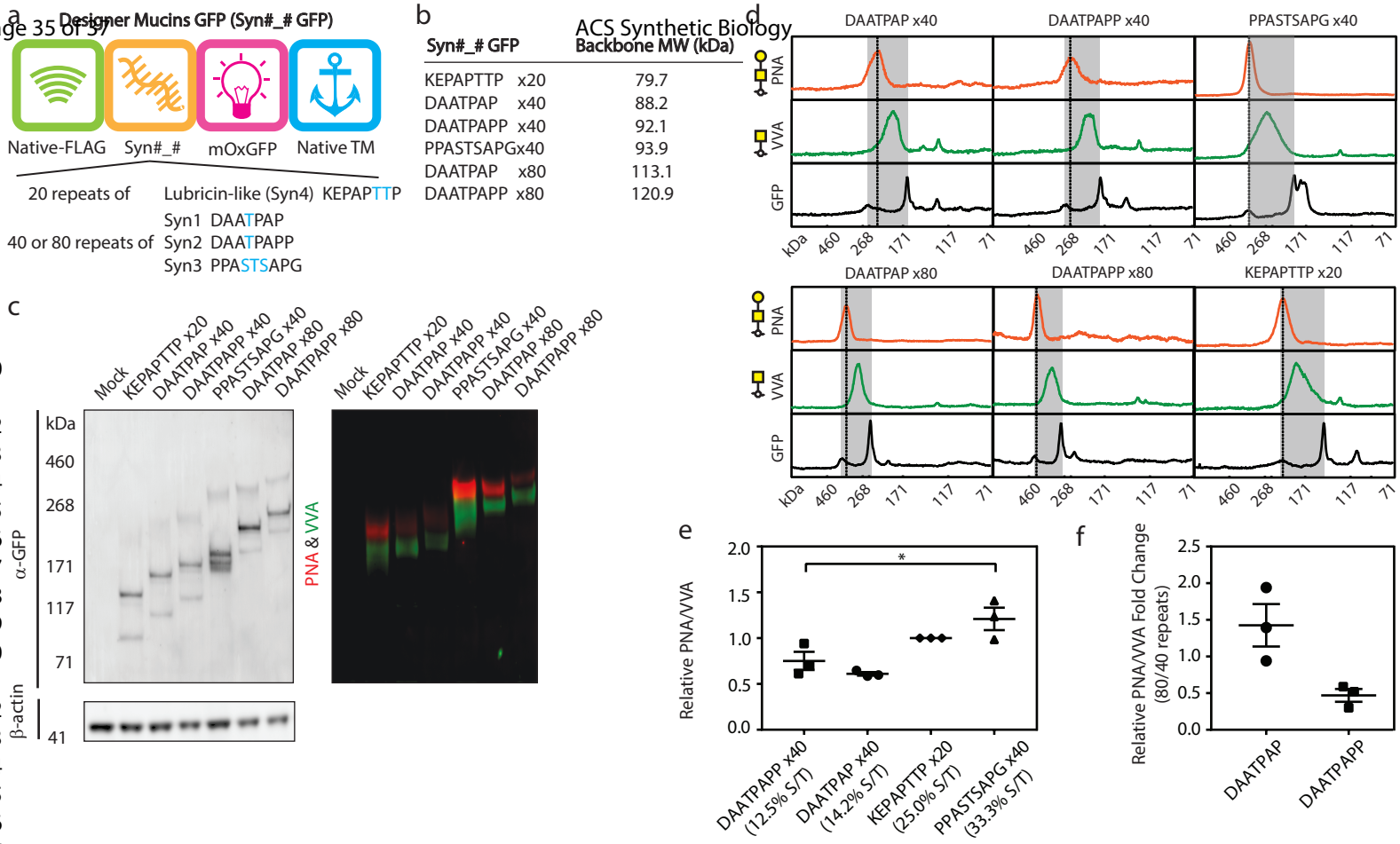


Figure 4: Designer Mucin Domains Reveal Sequence-Specific Effects on Glycosylation. (a) Components and features of designer mucins. (b) Predicted Molecular Weight of the mucin polypeptide backbones. (c) Representatie Western blot analysis (from three independent experiments) of indicated constructs in extracts of transiently transfected HEK293T cells probed with anti-GFP antibody or co-stained with PNA and VVA lectins. (d) Representative Fluorescence intensity electrophoretograms of the western blots in (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 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 KEPAPTTP x20 mucin; data presented as the mean and SEM from three independent experiments. (f) The fold change in 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. * $p<0.05$

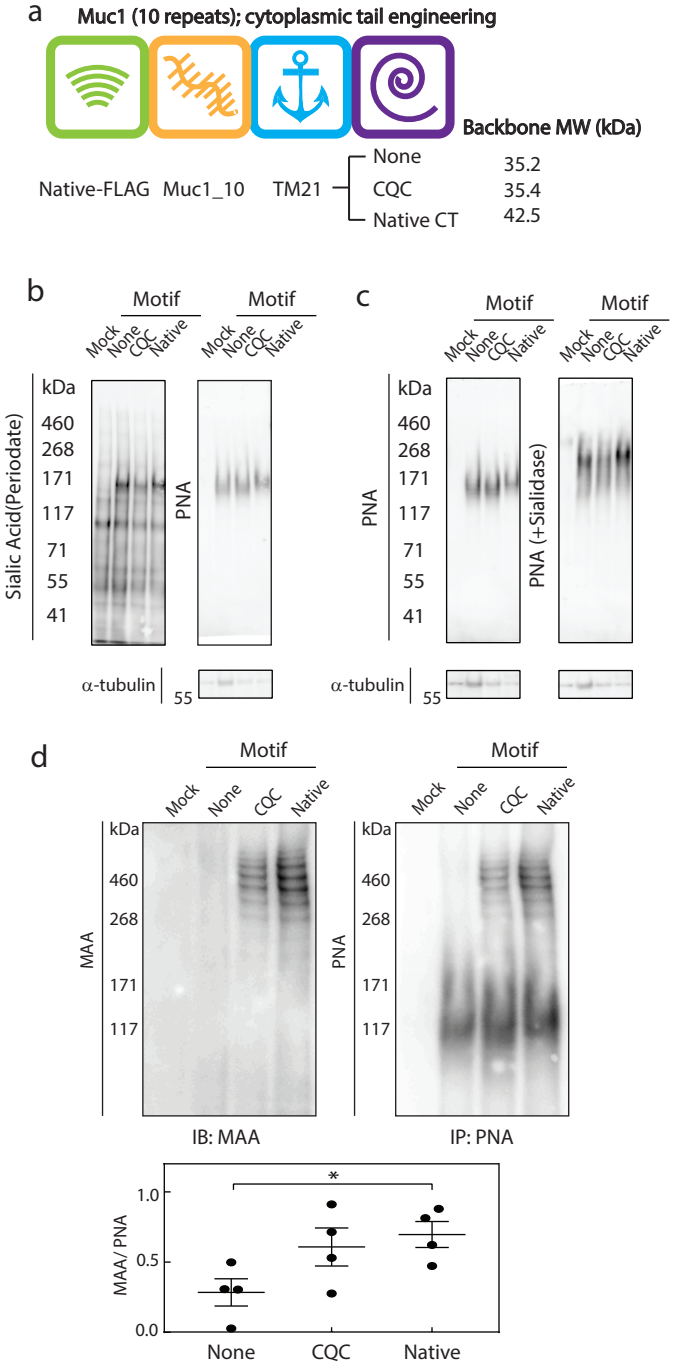


Figure 5: Tuning Mucin Glycosylation through Cytoplasmic Tail Engineering. (a) Components and features of cell-surface mucins with synthetic 21-amino-acid transmembrane anchors (TM21) and engineered cytoplasmic motifs; native CT refers to a native cytoplasmic tail adapted from Muc1. (b) Lectin blot analysis 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

For Table of Contents Only

Supporting Information for

Sequence-specific Mucins for Glycocalyx Engineering

Hao Pan[1], Marshall J. Colville[1], Nitin T. Supekar[2], Parastoo Azadi[2] and Matthew J. Paszek[1,3,4*]

[1]Field of Biophysics, Cornell University, Ithaca, NY 14853, USA

[2]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA
[3]Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA
[4]Nancy E. and Peter C. Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA

*Corresponding author

Matthew J. Paszek
Email: mjp31@cornell.edu

This PDF file includes:

Figs. S1 to S6
Table S1 to S2
Summary of cDNA "biobricks"
List of Constructs

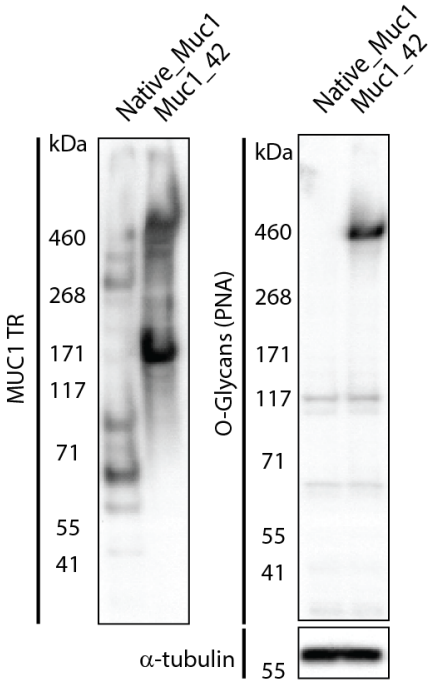


Figure S1: Western blot analysis of MCF10A cells edited with lentivirus with native repetitive (Native_Muc1) versus codon-scrambled Muc1 cDNAs (Muc1_42).

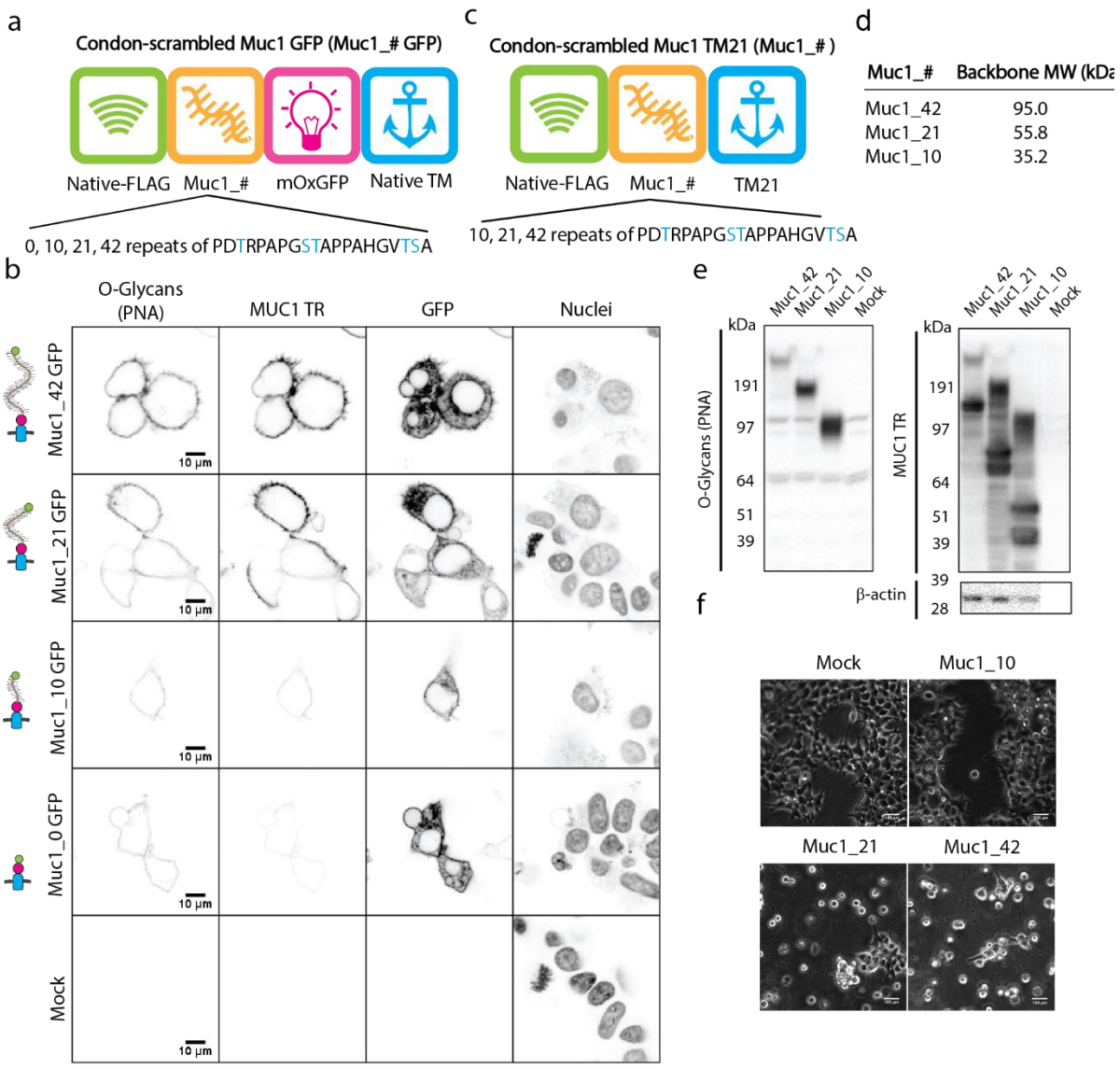


Figure S2: Mucins with Tunable Sizes. (a) Components and features of mucin constructs with GFP reporter, native Muc1 transmembrane anchor, and codon-scrambled Muc1 tandem repeats. (b) Representative immunofluorescence images of transiently transfected HEK293T cells expressing the GFP-tagged Muc1 constructs illustrated in (a) and co-stained with PNA, anti-Muc1 antibody, and Hoechst nuclear stain (scale bar 10 μm) from three independent experiments. (c) Components and features of mucin constructs with synthetic 21-amino-acid transmembrane anchor (TM21) and codon-scrambled Muc1 repeats. (d) Predicted molecular weight for mucin polypeptide backbone illustrated in (c). (e) Representative Western blot analysis (of three independent experiments) of TM21 constructs illustrated in (c) from extracts of transiently transfected HEK293T cells and probed with PNA lectin or anti-Muc1 antibody. (f) Representative phase-contrast images of HEK293Ts expressed indicated constructs in (c) from three independent experiments (scale bar 100 μm).

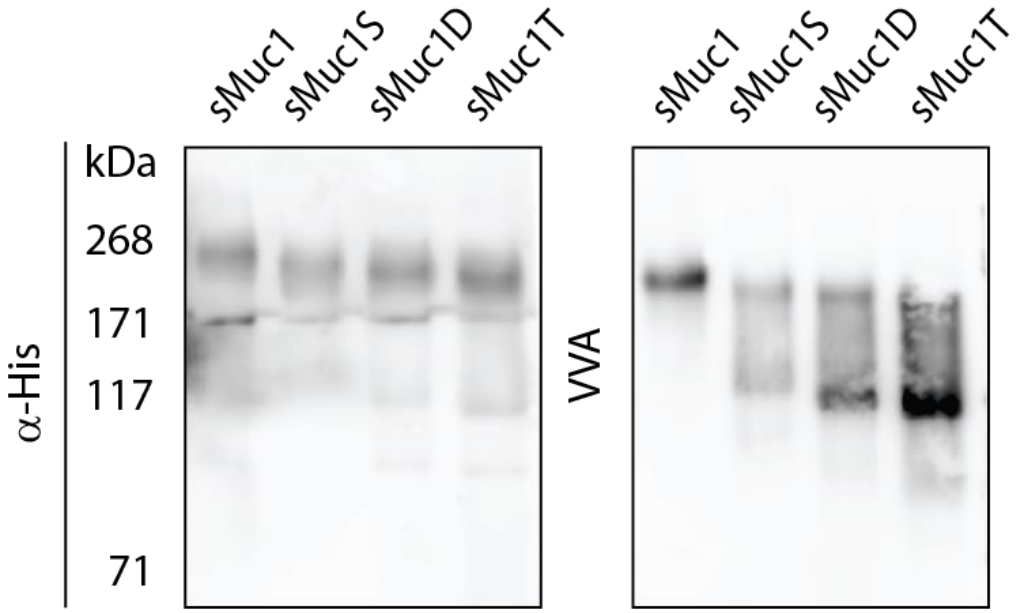


Figure S3: Western blot Image of affinity-purified recombinant secreted mucins from FreeStyle™ 293-F cell culture media probed with anti-6x His antibody and VVA lectin**.**

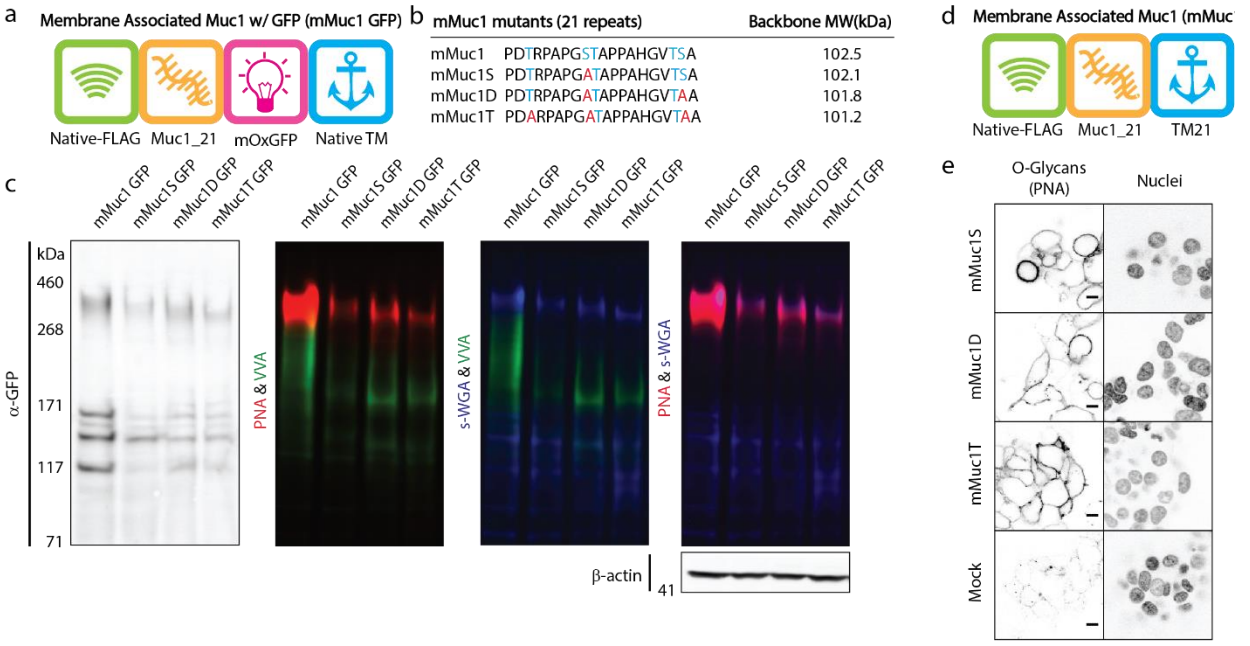


Figure S4: Cell-Surface Mucin Mutants Derived from Muc1 Tandem Repeat Sequences. (a) Components and features of mucins constructed with 21 native or engineered Muc1 repeats, GFP reporter and native Muc1 transmembrane anchor. (b) Tandem repeats and predicted backbone molecular weight of native Muc1 (mMuc1) or engineered variants with single, double, or triple serine/threonine to alanine substitutions (mMuc1S, mMuc1D, or mMuc1T). (c) Representative Western and lectin blot analysis of indicated constructs in (a) from extracts of transiently transfected HEK293T cells and probed with anti-GFP antibody or co-stained with PNA, VVA and s-WGA lectins from three independent experiments. (d) Components and features of mucins constructed with 21 native or engineered Muc1 repeats and a synthetic 21-amino-acid transmembrane anchor (TM21). (e) Representative immunofluorescence images of transiently transfected HEK293T cells expressing the indicated constructs in (d) and co-stained with PNA lectin and Hoechst nuclear stain from three independent experiments (scale bar 10 μm).

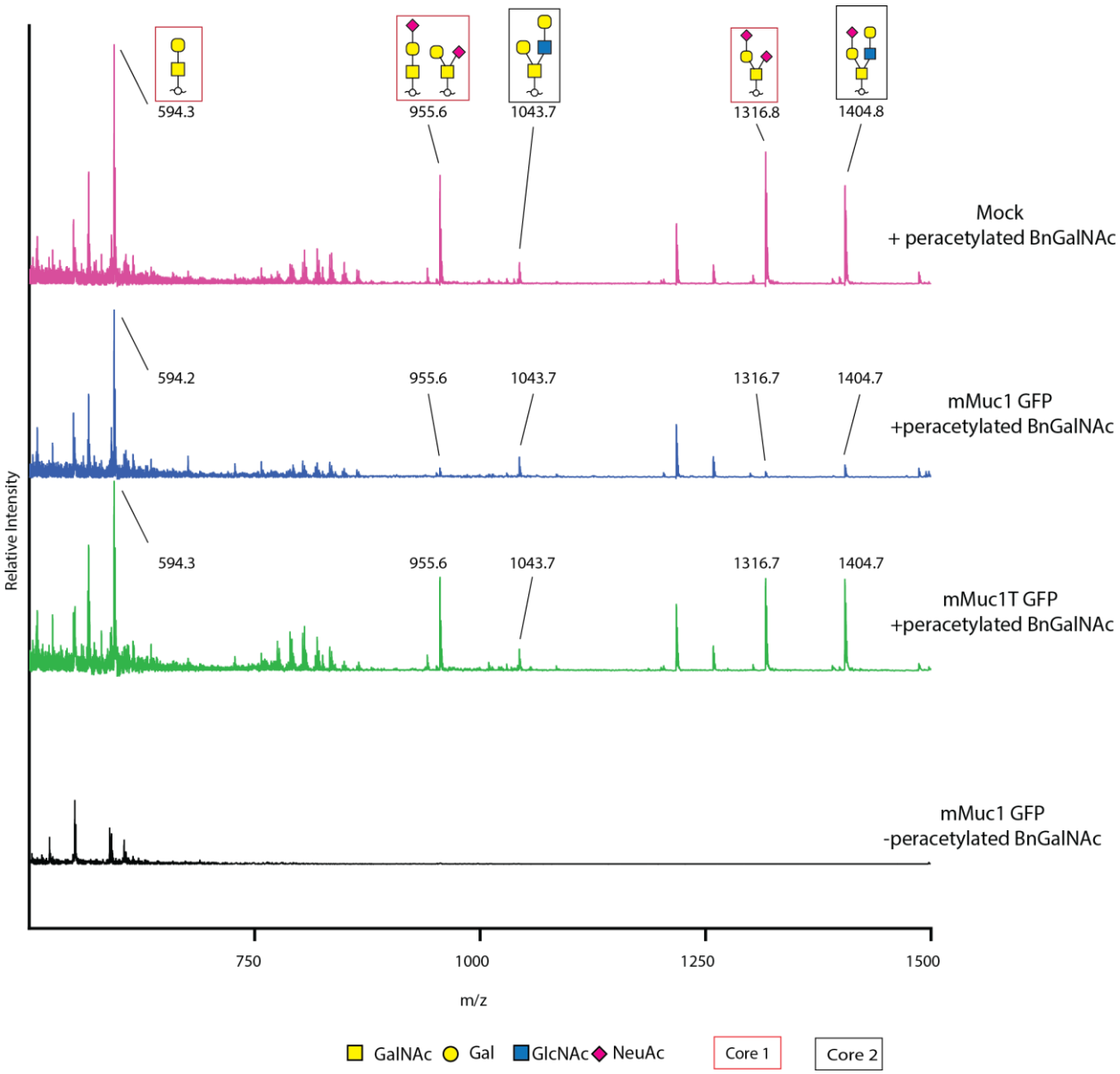


Figure S5: MALDI-TOF_MS spectra of mucin-type O-glycans as reported by Cellular O-Glycome Reporter/Amplification (CORA). HEK293T cells were transiently transfected with the indicated synthetic mucin constructs or mock vehicle. Spectra were normalized to the matrix peak at m/z = 550.

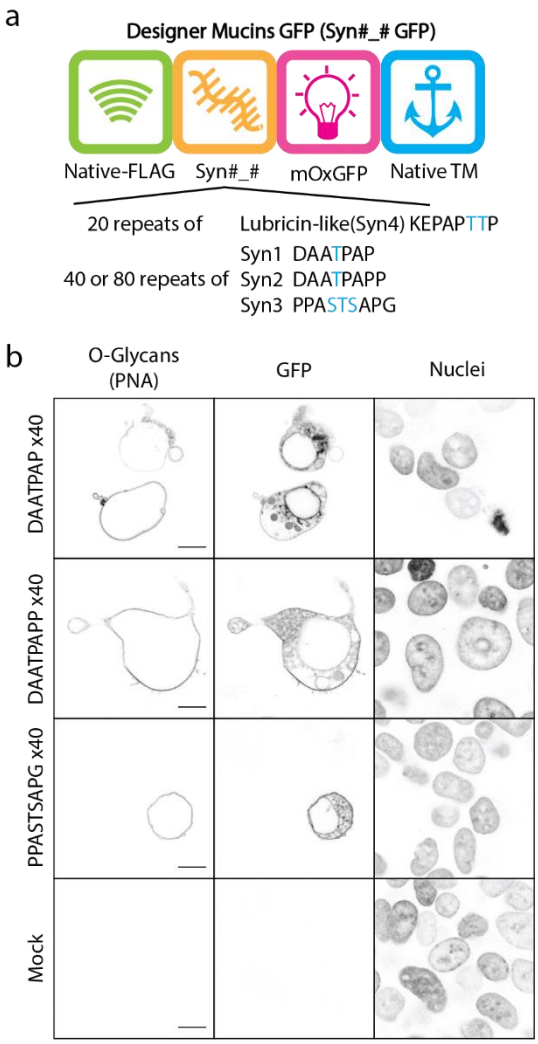


Figure S6: Mucins Constructed with Designer Tandem Repeats. (a) Components and features of mucin constructs with designer tandem repeats, GFP reporter and native Muc1 transmembrane anchor. (b) Representative immunofluorescence images of transiently transfected HEK293T cells expressing the indicated GFP-tagged constructs and co-stained with PNA lectin and Hoescht nuclear stain from three independent experiments (scale bar 10 μm).

Table S1: Repetitiveness Analysis of Mucin cDNA sequences

Repetition analysis of native and codon-scrambled cDNAs were conducted with the Tandem Repeat Finder algorithm[1]. Agreement between the queried sequence and detected tandem repeats were weighed by assigning alignment scores of +2 for nucleotide sequence matches and -7 for mismatches and indels. The high alignment score indicates high-level repetitiveness of the repeats.

Native_Muc1

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score
6-2577	60	42.9	60	99	0	4982

Muc1_42

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score
146-468	60	5.4	60	75	3	220
146-468	120	2.7	120	80	2	328
149-513	120	3.0	120	79	5	294
728-897	60	2.8	60	80	1	171
746-984	60	4.0	60	75	4	169
1013-1233	60	3.7	60	77	0	208
794-1200	120	3.4	120	75	4	273
1205-1347	60	2.4	59	74	8	135
1097-1530	180	2.4	180	77	2	379
1304-1521	60	3.6	59	76	2	175
1514-1714	60	3.3	60	78	0	204
1709-1965	120	2.1	120	80	1	273
1781-2067	60	4.8	60	71	5	177
1733-2067	120	2.8	120	77	1	269

2150-2406	60	4.3	60	73	3	140
2222-2439	120	1.8	120	79	2	258

Table Explanation:

- Indices of the repeat relative to the start of the sequence.
- Period size of the repeat.
- Number of copies aligned with the consensus pattern.
- Size of consensus pattern (may differ slightly from the period size).
- Percent of matches between adjacent copies overall.
- Percent of indels between adjacent copies overall.
- Alignment score.

Reference:

(1) Benson, G. Tandem Repeats Finder: A Program to Analyze DNA Sequences. *Nucleic Acids Res* **1999**, *27* (2), 573–580.

Table S2: Golden Gate assembly primers.

Name	Sequence
pcDNA3.1 Syn1 FWD	AGGTAGCGTCTCGTCCCGCCTCAGGCATACTTTATTG
pcDNA3.1 Syn1 REV	AGGTAGCGTCTCGTCGGGAGCAGGGGTAGCG
Syn1 FWD	AGGTAGCGTCTCGCCGATGCAGCTACTCCAGCTCCGGACGCC
Syn1 REV	AGGTAGCGTCTCGGGGAGCAGGGGTAGCG
pcDNA3.1 Syn2 FWD	CTTCTGCGTCTCGTCCCGCCTCAGGCATACTTTATTGGCGA
pcDNA3.1 Syn2 REV	CTTCTGCGTCTCGTCGGGAGGAGCTGGTGTAGCCGCG
Syn2 FWD	CTTCTGCGTCTCCCCGATGCAGCTACCCCGGCTCCACCC
Syn2 REV	CTTCTGCGTCTCCGGGAGGAGCTGGTGTAGCCGCG

Summary of cDNA "Biobricks"

Leader Tag

1. **Native-FLAG**
   Amino acid sequence:
   MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYKDDDDL
   Y
   cDNA sequence:
   GGATCCATGACACCGGGCACCCAGTCTCCTTTCTTCCTGCTGCTGCTCCTCACAGTGCTTA
   CAGTTGTTACAGGTTCTGGTCATGCAAGCTCTACCCCAGGTGGAGAAAAGGAGACTTCGG
   CTACCCAGAGAAGTTCAGTGCCCAGCTCTACTGAGAAGAATGCTGATTACAAGGATGACG
   ACGACCTGTACA
2. **His-SUMO**
   Amino acid sequence:
   METDTLLLWVLLLWVPGSTGDGHHHHHHGSLQDSEVNQEAKPEVKPEVKPETHINLKV
   SDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLTFLYDGIEIQADQAPEDLDMEDNDIIE
   AHREQIGGGSGSGHASSTPGGEKETSATQRSSVPSSTEKNADYKDDDDLY
   cDNA sequence:
   GGATCCGCCACCATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTTCCA
   GGTTCCACTGGTGACGGTCATCACCATCATCATCACGGGTCCCTGCAGGACTCAGAAGTC
   AATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTT
   AAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTTTAAG
   AAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAACGTTCT
   TGTACGACGGTATTGAAATTCAAGCTGATCAGGCCCCTGAAGATTTGGACATGGAGGAT
   AACGATATTATTGAGGCTCACAGAGAACAGATTGGAGGTGGCTCCGGCTCCGGTCATGCA
   AGCTCTACCCCAGGTGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCCAGC
   TCTACTGAGAAGAATGCTGATTACAAGGATGACGACGACCTGTACA

Polymer Backbone

1. **Codon-Scrambled Muc1 x42 (Muc1_42)**
   Amino acid sequence:
   LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[PDTR
   PAPGSTAPPAHGVTSA]$_{42}$ASG
   cDNA sequence:
   TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
   GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
   CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCCCGGATACGCGACCCG
   CCCCAGGGTCAACAGCGCCCCCAGCCCACGGCGTTACATCTGCACCTGACACTAGACCTGC
   GCCAGGATCAACAGCTCCACCGGCTCACGGGGTCACCAGTGCCCCCGACACTCGACCAGC
   TCCGGGGTCTACCGCTCCCCCGGCTCATGGTGTCACTAGCGCGCCTGACACACGCCCGGCA
   CCAGGGAGTACGGCCCCTCCTGCGCACGGCGTAACTTCAGCCCCAGATACTCGACCTGCT

CCGGGCTCAACAGCCCCGCCTGCACATGGAGTTACATCAGCCCCTGATACTAGACCGGCT
CCAGGTTCAACTGCTCCGCCAGCACATGGTGTAACGTCTGCGCCCGATACTCGCCCAGCA
CCTGGGTCCACAGCTCCCCCTGCGCATGGAGTAACATCAGCACCTGATACCAGACCTGCC
CCGGGCAGCACTGCACCCCCAGCACATGGCGTAACATCAGCACCAGATACTCGCCCCGCT
CCTGGTTCCACGGCTCCCCCCGCGCATGGCGTTACTTCAGCTCCAGATACACGGCCGGCAC
CCGGCAGTACGGCTCCACCCGCACATGGAGTAACGAGTGCTCCGGACACTCGGCCTGCTC
CAGGAAGTACCGCACCTCCGGCCCATGGCGTGACAAGTGCTCCCGACACCAGACCAGCGC
CTGGTTCAACAGCACCGCCAGCTCATGGTGTAACCTCAGCTCCCGATACTAGACCCGCGC
CAGGTTCCACCGCTCCACCTGCACACGGGGTGACGAGCGCACCTGATACGCGCCCGGCAC
CGGGAAGCACAGCGCCTCCCGCTCACGGAGTCACTAGCGCCCCGGATACAAGACCCGCAC
CTGGATCTACAGCTCCTCCAGCTCACGGCGTCACGAGTGCACCCGATACACGACCGGCCCC
AGGCTCTACAGCCCCACCAGCACATGGAGTCACGAGTGCACCTGATACTAGGCCCGCTCC
GGGTTCCACAGCACCTCCTGCACATGGTGTTACATCCGCTCCTGATACGAGACCCGCTCC
AGGCTCTACTGCCCCACCGGCACACGGCGTGACCAGTGCTCCAGATACCCGGCCAGCTCCT
GGGAGTACTGCGCCTCCAGCTCATGGCGTCACTAGTGCACCTGATACAAGACCAGCCCCC
GGTTCCACTGCTCCACCAGCCCATGGTGTAACAAGTGCACCGGACACAAGGCCAGCCCCT
GGTAGTACTGCTCCTCCTGCTCACGGTGTTACTAGTGCTCCTGACACCAGACCTGCCCCT
GGAAGTACTGCACCGCCTGCTCATGGAGTCACATCAGCTCCGGATACTCGGCCGGCTCCG
GGATCAACCGCTCCTCCGGCTCATGGAGTAACCTCCGCACCGGATACTAGGCCTGCACCG
GGGAGTACAGCACCACCTGCTCATGGTGTGACTAGCGCTCCTGACACTCGCCCCGCTCCC
GGTAGCACTGCCCCCCCTGCACATGGGGTGACTTCAGCTCCTGATACTCGGCCTGCACCCG
GAAGCACAGCCCCCCCAGCTCATGGGGTCACAAGCGCTCCAGATACTAGGCCAGCGCCGG
GAAGTACAGCCCCTCCAGCGCACGGTGTAACTTCCGCGCCAGACACACGCCCTGCTCCCG
GATCAACGGCACCTCCAGCACACGGTGTGACGTCCGCACCCGACACAAGACCGGCACCTG
GTTCTACTGCACCTCCCGCGCACGGAGTTACTTCAGCACCAGATACAAGACCTGCTCCTG
GCTCAACTGCCCCTCCGGCGCATGGTGTAACTAGTGCGCCTGATACACGCCCAGCACCGG
GTAGTACGGCACCACCAGCTCATGGAGTTACGTCAGCTCCAGATACGCGCCCTGCACCAG
GCAGTACAGCTCCGCCGGCCCACGGAGTAACTAGCGCACCAGATACCAGGCCAGCACCCG
GTAGTACCGCGCCTCCTGCCCATGGAGTAACTTCCGCCCCCGATACCCGACCTGCACCTGG
CAGTACCGCCCCTCCCGCCCACGGGGTAACCAGTGCACCAGACACGCGGCCCGCACCAGGA
TCTACTGCTCCCCCAGCGCATGGGGTAACTTCTGCACCAGATACGAGGCCTGCCCCAGGT
AGTACAGCGCCACCTGCCCACGGTGTCACCTCCGCTCCTGATACAAGGCCTGCGCCTGGA
TCAACTGCACCACCGGCGCACGGGGTTACAAGTGCCCCTGACACGAGACCAGCACCAGGT
TCTACGGCGCCTCCGGCACATGGAGTGACTAGTGCCCCAGACACTAGGCCGGCTCCTGGA
TCAACCGCACCACCCGCTCATGGAGTGACATCAGCGCCAGATACTAGACCAGCTCCCGGG
TCAACTGCGCCGCCCGCCCATGGGGTTACTTCTGCTCCAGACACTCGCCCAGCCCCAGGAT
CAACGGCTCCTCCCGCACACGGAGTGACCTCTGCTCCTGATACCAGGCCAGCTCCAGGGT
CTACAGCACCCCCTGCTCATGGGGTAACATCTGCCGCCTCAGG

2. **Codon-Scrambled Muc1 x21 (Muc1_21)**

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[PDTR
PAPGSTAPPAHGVTSA]$_{21}$ASG

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCCCAGACACTCGGCCTG
CACCGGGATCAACCGCCCCACCGGCTCATGGTGTAACTAGTGCGCCTGATACCAGACCAG
CACCAGGGAGTACTGCACCTCCTGCTCATGGGGTTACTAGTGCCCCCGATACGCGACCTG
CTCCTGGAAGCACAGCACCGCCGGCTCACGGCGTAACGAGTGCTCCTGACACAAGGCCCG
CTCCAGGGTCAACTGCACCACCTGCACACGGAGTGACATCAGCGCCAGATACGAGACCTG
CACCAGGAAGTACAGCGCCGCCAGCCCACGGAGTAACTTCAGCCCCGGACACTAGGCCAG
CACCTGGTTCAACGGCGCCTCCAGCCCATGGAGTAACATCCGCTCCCGATACTCGTCCTGC
TCCGGGTTCCACAGCTCCTCCCGCACATGGGGTGACTAGTGCTCCAGATACTCGCCCAGC
ACCCGGTAGTACCGCTCCTCCTGCACATGGCGTCACTAGTGCACCAGACACGCGTCCGGC
TCCTGGGTCTACAGCTCCACCAGCTCACGGAGTTACCAGTGCACCTGACACTAGACCTGC
GCCCGGTTCGACGGCTCCGCCCGCCCATGGGGTAACGTCTGCGCCGGATACACGCCCTGCA
CCTGGATCTACCGCACCTCCGGCCCATGGTGTCACGAGCGCACCTGATACGAGGCCTGCT
CCAGGTAGTACTGCTCCCCCCGCTCATGGAGTTACTAGCGCTCCTGATACTCGACCGGCA
CCTGGCAGCACTGCTCCTCCAGCACATGGTGTTACATCGGCTCCAGACACACGTCCCGCGC
CAGGATCGACTGCTCCACCCGCTCACGGGGTCACATCTGCACCCGATACACGGCCAGCTCC
CGGTTCCACTGCCCCGCCTGCCCATGGCGTTACTTCGGCACCAGATACCCGACCCGCACCA
GGCAGTACAGCACCTCCAGCGCATGGTGTGACAAGCGCCCCTGATACACGACCAGCTCCA
GGCTCAACAGCACCACCAGCACACGGTGTAACCTCAGCTCCGGATACCCGTCCAGCTCCT
GGTAGTACAGCCCCTCCTGCGCACGGAGTCACAAGTGCTCCCGACACAAGACCAGCCCCA
GGTTCTACTGCGCCACCTGCTCACGGTGTTACCTCTGCCCCAGATACAAGACCTGCCCCTG
GCTCTACGGCACCCCCGGCACATGGAGTCACTTCCGCACCGGATACTAGACCAGCGCCTG
GGAGTACGGCCCCCCCAGCTCATGGCGTGACTTCTGCTGCCTCAGG

3. **Codon-Scrambled Muc1 x10(Muc1_10)**

   Amino acid sequence:

   LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[PDTR
   PAPGSTAPPAHGVTSA]$_{10}$ASG

   cDNA sequence:

   TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
   GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
   CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCCCAGATACAAGACCGG
   CCCCAGGATCTACGGCTCCTCCGGCTCATGGAGTCACTTCTGCTCCAGACACAAGGCCCGC
   GCCGGGTTCTACAGCACCGCCTGCTCATGGTGTTACTAGCGCACCCGATACGAGACCTGC
   TCCGGGATCAACGGCACCTCCTGCCCACGGGGTAACATCTGCACCGGACACTCGCCCTGC
   GCCCGGTTCAACCGCTCCACCCGCACACGGAGTGACAAGCGCTCCTGACACTAGACCAGC
   ACCAGGTTCTACAGCCCCACCAGCCCATGGAGTTACCAGTGCACCAGATACTAGGCCAGC
   TCCAGGTAGTACTGCACCCCCAGCTCATGGGGTTACATCAGCTCCCGACACGCGACCAGC
   TCCTGGAAGCACTGCCCCTCCAGCTCACGGTGTGACCTCAGCACCTGATACACGCCCTGC
   ACCTGGCTCTACTGCTCCCCCCGCTCATGGCGTAACTAGTGCCCCGGATACTCGACCCGCC
   CCTGGTTCCACAGCTCCGCCAGCACATGGTGTAACAAGTGCTCCTGATACCCGACCAGCG
   CCTGGAAGTACCGCACCACCTGCACATGGAGTAACTTCAGCCGCCTCAGG

4. **Codon-Scrambled Muc1 x0 (Muc1_0)**

   Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVGGGG
GASG

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCGGCGGTGGTGGAGGAG
CCTCAGG

5. **Codon-Scrambled Muc1 Single Glycosylation Mutant x21 (Muc1_21S)**

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[PDTR
PAPGATAPPAHGVTSA]$_{21}$ASG

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCCCAGATACCAGACCTG
CGCCTGGAGCCACAGCTCCTCCTGCCCATGGCGTCACAAGTGCCCCTGACACACGCCCAGC
TCCCGGGGCTACAGCCCCACCTGCACATGGTGTTACTAGTGCACCAGACACCAGACCGGC
TCCGGGAGCCACGGCACCCCCCGCTCATGGTGTCACTTCCGCACCGGATACGAGGCCAGC
ACCTGGGGCCACTGCGCCGCCGGCACATGGGGTGACTAGTGCGCCAGATACTCGCCCTGC
TCCAGGGGCTACTGCCCCTCCAGCTCATGGCGTAACCTCAGCGCCTGATACCCGACCAGC
GCCAGGTGCCACTGCACCGCCAGCCCATGGGGTCACTAGTGCTCCTGACACTAGACCTGC
ACCTGGAGCTACAGCACCTCCAGCGCATGGTGTGACAAGCGCCCCAGACACGAGACCAGC
CCCCGGTGCCACCGCTCCTCCCGCACATGGAGTTACTAGCGCTCCGGACACAAGACCGGC
ACCAGGTGCGACTGCACCACCGGCTCATGGAGTAACTTCAGCACCAGATACACGGCCTGC
TCCCGGCGCTACAGCTCCACCAGCACATGGCGTTACCTCCGCACCTGACACGAGGCCCGCT
CCAGGAGCCACTGCTCCCCCTGCACACGGTGTTACGTCAGCTCCAGATACGCGGCCAGCT
CCGGGCGCAACAGCTCCCCCGGCTCACGGTGTAACCAGTGCTCCCGACACAAGGCCTGCA
CCCGGAGCAACCGCACCTCCGGCCCATGGTGTAACAAGTGCACCTGATACTAGGCCCGCG
CCTGGTGCTACTGCTCCACCTGCTCACGGCGTGACATCAGCCCCTGATACGAGACCTGCCC
CAGGGGCAACTGCACCTCCTGCTCATGGGGTAACTAGTGCCCCCGATACAAGACCAGCAC
CGGGAGCGACCGCCCCCCCAGCACACGGAGTAACGAGCGCACCCGATACTCGACCTGCAC
CAGGAGCGACGGCTCCACCCGCTCACGGAGTCACGAGTGCTCCAGACACTCGACCTGCTC
CTGGCGCGACAGCACCACCAGCTCACGGGGTTACTAGTGCTCCTGATACACGACCCGCAC
CAGGGGCGACTGCTCCTCCAGCCCACGGAGTTACATCTGCCCCGGATACAAGGCCAGCAC
CCGGTGCAACTGCTCCGCCCGCCCATGGAGTCACAAGTGCTCCGGATACTAGACCAGCTC
CTGGGGCTACGGCGCCTCCTGCGCACGGAGTGACTTCTGCTGCCTCAGG

6. **Codon-Scrambled Muc1 Double Glycosylation Mutant x21 (Muc1_21D)**

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[PDTR
PAPGATAPPAHGVTAA]$_{21}$ASG

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCCCAGACACGCGACCCG
CACCAGGCGCGACTGCTCCTCCTGCGCATGGTGTAACAGCGGCCCCTGATACGAGGCCAG

CCCCTGGAGCCACCGCACCTCCAGCACACGGAGTGACTGCAGCTCCCGATACTAGACCCGC
GCCAGGAGCAACAGCTCCTCCAGCTCATGGTGTGACGGCCGCCCCAGATACCAGACCTGC
CCCAGGGGCGACAGCACCCCCCGCTCACGGCGTAACTGCAGCCCCGGATACGAGACCAGC
TCCTGGGGCCACTGCACCTCCGGCTCATGGGGTAACAGCTGCCCCCGATACCCGACCTGCA
CCCGGAGCTACAGCGCCGCCTGCACACGGTGTAACCGCAGCTCCGGATACTAGACCTGCG
CCTGGAGCAACGGCGCCTCCTGCACATGGGGTTACTGCTGCGCCAGATACAAGGCCTGCC
CCTGGTGCAACAGCACCTCCTGCTCATGGCGTGACAGCTGCACCAGACACAAGACCAGCG
CCAGGTGCTACTGCACCACCTGCTCACGGGGTAACTGCTGCTCCAGATACTCGCCCTGCA
CCGGGAGCGACGGCTCCACCAGCTCACGGAGTAACGGCAGCACCTGACACTAGGCCGGCT
CCGGGAGCTACGGCACCGCCCGCACATGGCGTCACTGCGGCTCCTGACACACGACCAGCA
CCCGGTGCCACAGCTCCGCCAGCACATGGTGTTACGGCTGCTCCCGACACGAGACCCGCTC
CTGGAGCTACTGCTCCCCCGGCTCACGGTGTTACTGCAGCGCCTGATACACGCCCAGCACC
GGGGGCTACAGCACCACCAGCCCATGGGGTCACAGCAGCTCCAGACACTCGGCCAGCCCC
AGGTGCAACTGCTCCACCCGCCCATGGTGTCACTGCTGCACCTGATACCAGGCCGGCACC
AGGAGCCACGGCCCCGCCGGCACATGGAGTGACCGCGGCACCCGATACAAGACCTGCTCC
GGGCGCTACAGCCCCCCCAGCCCACGGAGTCACCGCTGCTCCTGATACTCGACCGGCACCT
GGTGCTACAGCTCCACCGGCCCATGGCGTTACAGCAGCACCAGATACGAGGCCCGCTCCA
GGTGCGACCGCTCCTCCCGCTCATGGAGTAACAGCCGCTCCGGACACTAGACCGGCTCCC
GGCGCAACTGCGCCCCCTGCCCATGGAGTTACTGCCGCACCGGATACACGCCCTGCCCCGG
GAGCAACTGCCCCTCCAGCGCACGGAGTTACAGCTGCTGCCTCAGG

7. Codon-Scrambled Muc1 Triple Glycosylation Mutant x21 (Muc1_21T)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[PDAR
PAPGATAPPAHGVTAA]$_{21}$ASG

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCCCAGATGCAAGGCCTG
CCCCGGGAGCGACAGCACCACCAGCACATGGAGTGACGGCCGCCCCAGACGCTCGACCGG
CACCAGGAGCAACTGCTCCTCCCGCACATGGGGTCACTGCGGCCCCTGATGCGAGGCCGG
CACCTGGAGCTACTGCTCCACCGGCCCATGGTGTCACTGCAGCCCCGGATGCTAGACCGG
CTCCGGGCGCAACTGCGCCGCCAGCCCATGGAGTTACTGCTGCGCCAGATGCGCGGCCTG
CCCCAGGTGCTACAGCCCCCCCTGCCCATGGCGTAACAGCTGCCCCCGATGCTCGCCCTGC
ACCGGGAGCAACGGCGCCTCCAGCGCACGGAGTAACGGCAGCACCAGATGCTCGGCCAGC
ACCGGGGGCTACAGCTCCACCTGCTCACGGTGTAACTGCAGCGCCTGATGCACGACCAGC
CCCTGGAGCAACAGCTCCGCCTGCACACGGAGTGACTGCTGCACCTGATGCTAGGCCAGC
CCCAGGGGCGACTGCACCTCCAGCACACGGTGTTACAGCTGCTCCAGACGCACGCCCAGC
ACCCGGTGCCACAGCTCCTCCTGCGCATGGTGTGACAGCTGCACCAGACGCCCGACCCGCG
CCAGGAGCCACGGCTCCACCAGCTCACGGCGTGACCGCGGCTCCTGACGCTAGGCCAGCTC
CTGGAGCCACCGCTCCTCCAGCTCATGGCGTTACAGCAGCTCCCGACGCAAGACCCGCTCC
TGGGGCCACTGCTCCCCCCGCTCACGGGGTAACAGCCGCTCCGGATGCAAGACCTGCCCCT
GGTGCTACTGCACCACCCGCCCATGGGGTTACTGCAGCTCCGGACGCTAGACCTGCTCCG
GGAGCTACAGCGCCCCCAGCCCACGGAGTCACAGCAGCACCTGACGCGAGACCAGCGCCA
GGTGCAACTGCCCCTCCTGCACATGGTGTTACTGCCGCACCGGATGCCAGACCTGCACCC
GGAGCTACGGCCCCGCCGGCTCATGGGGTAACTGCTGCTCCTGATGCCCGACCCGCTCCA

GGCGCGACCGCACCTCCTGCTCATGGAGTAACAGCGGCACCCGATGCACGGCCGGCTCCC
GGCGCTACAGCACCTCCGGCACATGGCGTCACCGCAGCTCCAGATGCCAGGCCCGCACCA
GGTGCGACGGCACCGCCCGCTCATGGTGTAACCGCTGCTCCCGATGCGAGACCTGCGCCT
GGTGCAACAGCACCCCCGGCTCACGGAGTTACGGCTGCTGCCTCAGG

8. **Lubricin consensus, KEPAPTTP x20 (Syn4_20)**

   Amino acid sequence:

   LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[KEPA
   PTTP]$_{20}$ASG

   cDNA sequence:

   TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
   GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
   CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCAAGGAACCTGCACCTA
   CAACCCCGAAGGAGCCCGCACCGACCACCCCAAAAGAACCTGCGCCGACAACTCCAAAGG
   AGCCAGCTCCAACGACGCCAAAGGAACCAGCACCTACGACCCCCAAGGAACCCGCCCCGA
   CGACTCCGAAGGAGCCTGCACCAACAACTCCTAAAGAACCAGCGCCTACTACGCCTAAAG
   AACCTGCTCCTACTACACCAAAAGAGCCAGCACCCACGACACCGAAAGAACCTGCCCCTA
   CTACCCCTAAAGAACCCGCTCCTACCACACCAAAGGAACCGGCTCCCACTACTCCCAAAG
   AACCAGCCCCAACTACACCTAAAGAACCGGCCCCCACCACTCCTAAAGAGCCGGCGCCAA
   CTACTCCAAAAGAACCAGCTCCTACAACTCCCAAGGAGCCGGCACCTACTACTCCGAAAG
   AGCCCGCGCCCACAACACCCAAAGAGCCTGCTCCGACTACTCCTGCCTCAGG

9. **Synthetic 1, DAATPAP x40 (Syn1_40)**

   Amino acid sequence:

   LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[DAAT
   PAP]$_{40}$ASG

   cDNA sequence:

   TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
   GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
   CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCGATGCAGCTACTCCAG
   CTCCGGACGCCGCAACACCCGCTCCAGACGCCGCCACCCCAGCTCCAGATGCTGCTACACC
   TGCACCTGATGCCGCAACTCCCGCGCCGGATGCCGCGACTCCAGCACCGGACGCTGCGACG
   CCAGCCCCTGATGCTGCAACACCGGCTCCTGATGCTGCGACTCCTGCGCCAGATGCAGCT
   ACACCAGCCCCGGATGCTGCAACGCCTGCTCCTGACGCAGCTACTCCGGCCCCCGACGCTG
   CTACCCCGGCGCCTGATGCTGCTACTCCCGCTCCTGATGCGGCCACTCCAGCCCCAGACGC
   AGCAACCCCAGCCCCCGATGCTGCTACGCCTGCACCCGACGCGGCCACACCTGCGCCGGAC
   GCAGCGACACCTGCCCCTGACGCTGCCACGCCCGCACCTGATGCAGCTACGCCAGCTCCCG
   ATGCGGCAACACCTGCTCCAGATGCCGCCACTCCTGCTCCGGATGCGGCGACACCAGCGCC
   TGACGCCGCTACGCCGGCACCTGATGCTGCCACTCCGGCTCCAGATGCAGCGACCCCAGCG
   CCAGACGCGGCAACTCCAGCGCCCGATGCAGCTACCCCAGCACCAGATGCTGCAACCCCT
   GCACCGGATGCAGCAACGCCAGCACCTGACGCGGCTACTCCTGCACCAGATGCAGCAACT
   CCTGCCCCGGACGCGGCGACTCCCGCACCAGACGCTGCAACTCCGGCACCAGATGCGGCTA
   CCCCCGCTCCCGACGCAGCCACTCCCGCCCCAGATGCAGCCACACCAGCTCCTGATGCAGC
   AACACCAGCACCCGATGCCGCTACCCCTGCTCCCGCCTCAGG

10. **Synthetic 1, DAATPAP x80 (Syn1_80)**

    Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[DAAT
PAP]$_{80}$ASG

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCGATGCAGCTACTCCAG
CTCCGGACGCCGCAACACCCGCTCCAGACGCCGCCACCCCAGCTCCAGATGCTGCTACACC
TGCACCTGATGCCGCAACTCCCGCGCCGGATGCCGCGACTCCAGCACCGGACGCTGCGACG
CCAGCCCCTGATGCTGCAACACCGGCTCCTGATGCTGCGACTCCTGCGCCAGATGCAGCT
ACACCAGCCCCGGATGCTGCAACGCCTGCTCCTGACGCAGCTACTCCGGCCCCCGACGCTG
CTACCCCGGCGCCTGATGCTGCTACTCCCGCTCCTGATGCGGCCACTCCAGCCCCAGACGC
AGCAACCCCAGCCCCCGATGCTGCTACGCCTGCACCCGACGCGGCCACACCTGCGCCGGAC
GCAGCGACACCTGCCCCTGACGCTGCCACGCCCGCACCTGATGCAGCTACGCCAGCTCCCG
ATGCGGCAACACCTGCTCCAGATGCCGCCACTCCTGCTCCGGATGCGGCGACACCAGCGCC
TGACGCCGCTACGCCGGCACCTGATGCTGCCACTCCGGCTCCAGATGCAGCGACCCCAGCG
CCAGACGCGGCAACTCCAGCGCCCGATGCAGCTACCCCAGCACCAGATGCTGCAACCCCT
GCACCGGATGCAGCAACGCCAGCACCTGACGCGGCTACTCCTGCACCAGATGCAGCAACT
CCTGCCCCGGACGCGGCGACTCCCGCACCAGACGCTGCAACTCCGGCACCAGATGCGGCTA
CCCCCGCTCCCGACGCAGCCACTCCCGCCCCAGATGCAGCCACACCAGCTCCTGATGCAGC
AACACCAGCACCCGATGCCGCTACCCCTGCTCCCGATGCAGCTACTCCAGCTCCGGACGCC
GCAACACCCGCTCCAGACGCCGCCACCCCAGCTCCAGATGCTGCTACACCTGCACCTGATG
CCGCAACTCCCGCGCCGGATGCCGCGACTCCAGCACCGGACGCTGCGACGCCAGCCCCTGA
TGCTGCAACACCGGCTCCTGATGCTGCGACTCCTGCGCCAGATGCAGCTACACCAGCCCC
GGATGCTGCAACGCCTGCTCCTGACGCAGCTACTCCGGCCCCCGACGCTGCTACCCCGGCG
CCTGATGCTGCTACTCCCGCTCCTGATGCGGCCACTCCAGCCCCAGACGCAGCAACCCCAG
CCCCCGATGCTGCTACGCCTGCACCCGACGCGGCCACACCTGCGCCGGACGCAGCGACACC
TGCCCCTGACGCTGCCACGCCCGCACCTGATGCAGCTACGCCAGCTCCCGATGCGGCAACA
CCTGCTCCAGATGCCGCCACTCCTGCTCCGGATGCGGCGACACCAGCGCCTGACGCCGCTA
CGCCGGCACCTGATGCTGCCACTCCGGCTCCAGATGCAGCGACCCCAGCGCCAGACGCGGC
AACTCCAGCGCCCGATGCAGCTACCCCAGCACCAGATGCTGCAACCCCTGCACCGGATGC
AGCAACGCCAGCACCTGACGCGGCTACTCCTGCACCAGATGCAGCAACTCCTGCCCCGGA
CGCGGCGACTCCCGCACCAGACGCTGCAACTCCGGCACCAGATGCGGCTACCCCCGCTCCC
GACGCAGCCACTCCCGCCCCAGATGCAGCCACACCAGCTCCTGATGCAGCAACACCAGCA
CCCGATGCCGCTACCCCTGCTCCCGCCTCAGG

11. Synthetic 2, DAATPAPP x40 (Syn2_40)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[DAAT
PAPP]$_{40}$ASG

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCGATGCAGCTACCCCGG
CTCCACCCGATGCGGCAACACCAGCCCCTCCCGATGCAGCAACACCTGCTCCCCCCGATGC
TGCTACCCCTGCTCCGCCTGATGCTGCAACTCCAGCTCCGCCCGATGCCGCTACACCTGCC
CCCCCTGACGCCGCCACGCCCGCTCCTCCGGATGCTGCAACCCCAGCACCCCCAGACGCCG

CTACCCCAGCTCCACCAGATGCTGCTACACCCGCACCACCTGATGCCGCAACACCGGCGCC
TCCTGATGCTGCTACTCCAGCCCCACCTGATGCAGCAACTCCTGCGCCACCAGACGCTGCC
ACACCTGCACCACCAGATGCAGCCACACCAGCACCGCCAGACGCAGCAACGCCGGCTCCG
CCAGATGCAGCGACACCAGCGCCACCTGACGCAGCGACTCCAGCACCACCGGATGCGGCT
ACCCCCGCTCCGCCGGACGCGGCGACTCCTGCCCCTCCTGACGCGGCAACTCCGGCCCCTC
CAGATGCGGCGACCCCAGCCCCGCCGGATGCCGCGACTCCGGCTCCCCCGGACGCTGCAAC
ACCCGCTCCACCTGATGCTGCCACTCCCGCGCCTCCAGATGCTGCAACGCCAGCTCCCCCT
GATGCTGCGACGCCTGCTCCTCCAGATGCAGCTACACCGGCTCCTCCTGATGCAGCTACG
CCTGCACCGCCTGACGCTGCTACGCCAGCACCTCCCGACGCAGCCACTCCTGCACCTCCTG
ATGCGGCCACTCCAGCGCCCCCGGATGCAGCTACTCCTGCTCCACCGGACGCCGCAACTCC
CGCCCCTCCGGACGCAGCTACTCCCGCTCCCCCAGATGCAGCAACCCCTGCACCCCCCGAC
GCGGCCACCCCTGCCCCACCAGATGCCGCCACTCCGGCACCACCCGACGCTGCGACTCCCG
CACCTCCAGACGCGGCTACACCAGCTCCTCCCG<u>CCTCAGG</u>

12. Synthetic 2, DAATPAPP x80 (Syn2_80)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[DAAT PAPP]$_{80}$ASG

cDNA sequence:

<u>TGTACA</u>TGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCGATGCAGCTACCCCGG
CTCCACCCGATGCGGCAACACCAGCCCCTCCCGATGCAGCAACACCTGCTCCCCCCGATGC
TGCTACCCCTGCTCCGCCTGATGCTGCAACTCCAGCTCCGCCCGATGCCGCTACACCTGCC
CCCCCTGACGCCGCCACGCCCGCTCCTCCGGATGCTGCAACCCCAGCACCCCCAGACGCCG
CTACCCCAGCTCCACCAGATGCTGCTACACCCGCACCACCTGATGCCGCAACACCGGCGCC
TCCTGATGCTGCTACTCCAGCCCCACCTGATGCAGCAACTCCTGCGCCACCAGACGCTGCC
ACACCTGCACCACCAGATGCAGCCACACCAGCACCGCCAGACGCAGCAACGCCGGCTCCG
CCAGATGCAGCGACACCAGCGCCACCTGACGCAGCGACTCCAGCACCACCGGATGCGGCT
ACCCCCGCTCCGCCGGACGCGGCGACTCCTGCCCCTCCTGACGCGGCAACTCCGGCCCCTC
CAGATGCGGCGACCCCAGCCCCGCCGGATGCCGCGACTCCGGCTCCCCCGGACGCTGCAAC
ACCCGCTCCACCTGATGCTGCCACTCCCGCGCCTCCAGATGCTGCAACGCCAGCTCCCCCT
GATGCTGCGACGCCTGCTCCTCCAGATGCAGCTACACCGGCTCCTCCTGATGCAGCTACG
CCTGCACCGCCTGACGCTGCTACGCCAGCACCTCCCGACGCAGCCACTCCTGCACCTCCTG
ATGCGGCCACTCCAGCGCCCCCGGATGCAGCTACTCCTGCTCCACCGGACGCCGCAACTCC
CGCCCCTCCGGACGCAGCTACTCCCGCTCCCCCAGATGCAGCAACCCCTGCACCCCCCGAC
GCGGCCACCCCTGCCCCACCAGATGCCGCCACTCCGGCACCACCCGACGCTGCGACTCCCG
CACCTCCAGACGCGGCTACACCAGCTCCTCCCGATGCAGCTACCCCGGCTCCACCCGATGC
GGCAACACCAGCCCCTCCCGATGCAGCAACACCTGCTCCCCCCGATGCTGCTACCCCTGCT
CCGCCTGATGCTGCAACTCCAGCTCCGCCCGATGCCGCTACACCTGCCCCCCCTGACGCCG
CCACGCCCGCTCCTCCGGATGCTGCAACCCCAGCACCCCCAGACGCCGCTACCCCAGCTCC
ACCAGATGCTGCTACACCCGCACCACCTGATGCCGCAACACCGGCGCCTCCTGATGCTGCT
ACTCCAGCCCCACCTGATGCAGCAACTCCTGCGCCACCAGACGCTGCCACACCTGCACCAC
CAGATGCAGCCACACCAGCACCGCCAGACGCAGCAACGCCGGCTCCGCCAGATGCAGCGA
CACCAGCGCCACCTGACGCAGCGACTCCAGCACCACCGGATGCGGCTACCCCCGCTCCGCC
GGACGCGGCGACTCCTGCCCCTCCTGACGCGGCAACTCCGGCCCCTCCAGATGCGGCGACC

CCAGCCCCGCCGGATGCCGCGACTCCGGCTCCCCCGGACGCTGCAACACCCGCTCCACCTG
ATGCTGCCACTCCCGCGCCTCCAGATGCTGCAACGCCAGCTCCCCCTGATGCTGCGACGCC
TGCTCCTCCAGATGCAGCTACACCGGCTCCTCCTGATGCAGCTACGCCTGCACCGCCTGAC
GCTGCTACGCCAGCACCTCCCGACGCAGCCACTCCTGCACCTCCTGATGCGGCCACTCCAG
CGCCCCCGGATGCAGCTACTCCTGCTCCACCGGACGCCGCAACTCCCGCCCCTCCGGACGC
AGCTACTCCCGCTCCCCCAGATGCAGCAACCCCTGCACCCCCCGACGCGGCCACCCCTGCC
CCACCAGATGCCGCCACTCCGGCACCACCCGACGCTGCGACTCCCGCACCTCCAGACGCGG
CTACACCAGCTCCTCCCGCCTCAGG

13. **Synthetic 3, PPASTSAPG x40 (Syn3_40)**

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV[PPAS
TSAPG]$_{40}$ASG

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCG
GTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAG
CTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCCCACCTGCATCTACCA
GTGCCCCGGGTCCACCTGCCTCTACTAGCGCCCCAGGACCTCCGGCAAGTACATCAGCGCC
AGGACCCCCTGCTTCCACTAGTGCACCCGGTCCCCCGGCATCTACGTCTGCCCCTGGCCCA
CCTGCTTCAACTTCAGCACCAGGACCACCCGCAAGCACATCAGCCCCAGGCCCTCCCGCCT
CTACAAGCGCTCCGGGGCCTCCGGCCTCTACCTCAGCTCCAGGCCCACCAGCCAGCACTTC
AGCCCCTGGTCCACCCGCTTCAACCTCAGCACCCGGACCTCCTGCCTCAACTTCCGCTCCC
GGTCCACCAGCTAGTACCTCTGCTCCGGGCCCTCCGGCGAGCACGTCAGCACCGGGACCAC
CTGCGAGTACAAGTGCACCTGGCCCGCCCGCTAGCACAAGTGCCCCCGGTCCTCCAGCATC
CACTAGTGCACCAGGGCCTCCAGCCAGCACTAGTGCGCCGGGTCCCCCCGCGAGTACGTC
AGCTCCGGGACCTCCAGCTTCTACATCTGCTCCTGGGCCCCCTGCATCAACTAGTGCCCCT
GGACCACCGGCTAGTACGTCAGCTCCTGGTCCCCCTGCCAGTACTAGCGCTCCAGGGCCAC
CAGCAAGTACGAGCGCACCAGGCCCCCCAGCCTCTACGAGTGCACCGGGTCCTCCTGCAA
GTACCTCCGCTCCAGGTCCTCCGGCTTCAACGTCCGCACCTGGACCTCCCGCGTCCACATC
AGCTCCCGGCCCTCCAGCGAGTACTTCTGCTCCCGGACCACCAGCGTCCACATCTGCGCCT
GGTCCTCCCGCTAGTACCTCTGCACCTGGTCCGCCGGCCAGTACAAGTGCTCCCGGGCCTC
CCGCATCAACATCTGCACCAGGTCCACCGGCGTCTACTAGTGCCCCAGGTCCCCCAGCTTC
AACATCAGCACCTGGGCCGCCTGCTAGTACATCCGCTCCTGGACCCCCAGCAAGTACTTC
CGCCCCTGGGCCTCCTGCTTCTACTTCAGCTCCTGGCCCTCCTGCGTCAACTAGTGCTCCA
GGACCGCCAGCTAGTACTTCCGCGCCCGGTGCCTCAGG

Optical Reporter

1. **mOxGFP**

Amino acid sequence:

SGSASGSAMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTGK
LPVPWPTLVTTLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAE
VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHG
MDELYKGSA

cDNA sequence:

CCTCAGGCTCTGCATCAGGCTCAGCTATGGTGTCCAAGGGCGAGGAGCTGTTCACCGGGG
TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCTCCGTGCGGG
GCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCATCAGCACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGAGCTT
CTCCCGCTACCCCGACCACATGAAGCGCCACGACTTCTTCAAGAGCGCCATGCCCGAAGG
CTACGTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACTCCCACAACGTCTA
TATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACGT
GGAGGACGGCTCCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGG
CCCCGTGCTGCTGCCCGACAACCACTACCTGTCCACCCAGTCCAAGCTGTCCAAAGACCCC
AACGAGAAGCGCGATCACATGGTCCTTCTGGAATTCGTGACCGCCGCCGGGATCACTCAC
GGCATGGACGAGCTGTATAAGGGCTCAGC

Membrane Anchor

1. **Native TM**

   Amino acid sequence:
   SASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHHSSVPPLTSSN
   HSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNI
   KFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSG
   AGVPGWGIALLVLVCVLVALAIVYLIALAVCQCRRK*

   cDNA sequence:
   GCTCAGCTTCTACTCTGGTGCACAACGGCACCTCTGCCAGGGCTACCACAACCCCAGCCA
   GCAAGAGCACTCCATTCTCAATTCCCAGCCACCACTCTGATACTCCTACCACCCTTGCCAG
   CCATAGCACCAAGACTGATGCCAGTAGCACTCACCATAGCTCGGTACCTCCTCTCACCTC
   CTCCAATCACAGCACTTCTCCCCAGTTGTCTACTGGGGTCTCTTTCTTTTTCCTGTCTTTT
   CACATTTCAAACCTCCAGTTTAATTCCTCTCTGGAAGATCCCAGCACCGACTACTACCAA
   GAGCTGCAGAGAGACATTTCTGAAATGTTTTTGCAGATTTATAAACAAGGGGGTTTTCT
   GGGCCTCTCCAATATTAAGTTCAGGCCAGGATCTGTGGTGGTACAATTGACTCTGGCCTT
   CCGAGAAGGTACCATCAATGTCCACGACGTGGAGACACAGTTCAATCAGTATAAAACGG
   AAGCAGCCTCTCGATATAACCTGACGATCTCAGACGTCAGCGTGAGTGATGTGCCATTTC
   CTTTCTCTGCCCAGTCTGGGGCTGGGGTGCCAGGCTGGGGCATCGCGCTGCTGGTGCTGG
   TCTGTGTTCTGGTTGCGCTGGCCATTGTCTATCTCATTGCCTTGGCTGTCTGTCAGTGCC
   GCCGAAAGTAGGGAATTC

2. **Synthetic TM TM21**

   Amino acid sequence:
   ASGILYWRNPTESDSIVLAIIVPSLLLLLCLALLWYMRRRSM*

   cDNA sequence:
   CCTCAGGCATACTTTATTGGCGAAACCCAACGGAAAGTGATAGCATCGTTTTGGCAATTA
   TCGTCCCCAGTCTGCTCCTCTTGCTCTGCCTGGCTTTGTTGTGGTACATGCGCCGACGAA
   GTATGTAGGAATTC

Cytoplasmic Motif

1. **Native CT**

   Amino acid sequence:

   SRCQCRRKNYGQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSTDRSPYEKVSAGNGGSSLS
   YTNPAVAAASANL*

   cDNA sequence:

   TCTAGATGTCAGTGCCGCCGAAAGAACTACGGGCAGCTGGACATCTTTCCAGCCCGGGAT
   ACCTACCATCCTATGAGCGAGTACCCCACCTACCACACCCATGGGCGCTATGTGCCCCCTA
   GCAGTACCGATCGTAGCCCCTATGAGAAGGTTTCTGCAGGTAAtGGTGGCAGCAGCCTCT
   CTTACACAAACCCAGCAGTGGCAGCCGCTTCTGCCAACTTGTAGGAATTC

2. **CQC**

   Amino acid sequence:

   SRCQCRRK*

   cDNA sequence:

   TCTAGATGTCAGTGCCGCCGAAAGTAGGAATTC

List of Constructs

Membrane Associated Mucin

1. pcDNA3.1+_Muc1_0_TM21
2. pcDNA3.1+_Muc1_10_TM21
3. pcDNA3.1+_Muc1_21_TM21
4. pcDNA3.1+_Muc1_42_TM21
5. pcDNA3.1+_Muc1_21S_TM21
6. pcDNA3.1+_Muc1_21D_TM21
7. pcDNA3.1+_Muc1_21T_TM21
8. pcDNA3.1+_Muc1_10_TM21_CT
9. pcDNA3.1+_Muc1_10_TM21_CQC
10. pcDNA3.1+_Muc1_10_dCT
11. pcDNA3.1+_Muc1_10_FL
12. pcDNA3.1+_Muc1_Syn4_20_TM21
13. pcDNA3.1+_Muc1_Syn1_40_TM21
14. pcDNA3.1+_Muc1_Syn2_40_TM21
15. pcDNA3.1+_Muc1_Syn3_40_TM21
16. pcDNA3.1+_Muc1_Syn1_80_TM21
17. pcDNA3.1+_Muc1_Syn2_80_TM21
18. pPB_Tet_Muc1_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
19. pPB_Tet_Muc1_42_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
20. pPB_Tet_Muc1_21_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
21. pPB_Tet_Muc1_10_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
22. pPB_Tet_Muc1_0_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
23. pPB_Tet_Muc1_21D_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
24. pPB_Tet_Muc1_21T_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
25. pLV_puro_teton_Muc1_42_dCT
26. pLV_puro_teton_Muc1_dCT
27. pPB_Muc1_mOxGFP_dCT_BlpI
28. pPB_Muc1_42_mOxGFP_dCT_BlpI
29. pPB_Muc1_21_mOxGFP_dCT_BlpI
30. pPB_Muc1_10_mOxGFP_dCT_BlpI
31. pPB_Muc1_0_mOxGFP_dCT_BlpI
32. pPB_Muc1_21S_mOxGFP_dCT_BlpI
33. pPB_Muc1_21D_mOxGFP_dCT_BlpI
34. pPB_Muc1_21T_mOxGFP_dCT_BlpI
35. pPB_Muc1_Syn4_20_mOxGFP_dCT_BlpI
36. pPB_Muc1_Syn1_40_mOxGFP_dCT_BlpI
37. pPB_Muc1_Syn2_40_mOxGFP_dCT_BlpI

38. pPB_Muc1_Syn3_40_mOxGFP_dCT_BlpI
39. pPB_Muc1_Syn1_80_mOxGFP_dCT_BlpI
40. pPB_Muc1_Syn2_80_mOxGFP_dCT_BlpI

Secreted Mucin

41. pPB_Tet_SumoStar_Muc1_42_rtTAsM2_IRES_NeoR
42. pPB_Tet_SumoStar_Muc1_21T_rtTAsM2_IRES_NeoR
43. pPB_Tet_SumoStar_Muc1_21D_rtTAsM2_IRES_NeoR
44. pPB_Tet_SumoStar_Muc1_21S_rtTAsM2_IRES_NeoR
45. pPB_Tet_SumoStar_Muc1_21_rtTAsM2_IRES_NeoR
46. pPB_Tet_SumoStar_Muc1_0_rtTAsM2_IRES_NeoR
47. pPB_Tet_SumoStar_Muc1_Syn1_40_rtTAsM2_IRES_NeoR
48. pPB_Tet_SumoStar_Muc1_Syn2_40_rtTAsM2_IRES_NeoR
49. pPB_Tet_SumoStar_Muc1_Syn3_40_rtTAsM2_IRES_NeoR
50. pPB_Tet_SumoStar_Muc1_Syn1_80_rtTAsM2_IRES_NeoR
51. pPB_Tet_SumoStar_Muc1_Syn2_80_rtTAsM2_IRES_NeoR