(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 July 2020 (23.07.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/150396 A1

(51) International Patent Classification:
A61K 38/16 (2006.01)

(21) International Application Number:
PCT/US2020/013752

(22) International Filing Date:
15 January 2020 (15.01.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/792,660 15 January 2019 (15.01.2019) US

(71) Applicant: **CORNELL UNIVERSITY** [US/US]; Center For Technology Licensing At Cornell, University, 395 Pine Tree Road, Suite 310, Ithaca, NY 14850 (US).

(72) Inventors: **PASZEK, Matthew**; 194 Buck Road, Lansing, NY 14882 (US). **REESINK, Heidi**; 194 Buck Road, Lansing, NY 14882 (US).

(74) Agent: **DIMAIO, John T.M.** et al.; Hodgson Russ LLP, The Guaranty Building, 140 Pearl Street, Suite 100, Buffalo, NY 14202-4040 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: RECOMBINANT LUBRICINS, AND COMPOSITIONS AND METHODS FOR USING THE SAME

Part I Figures

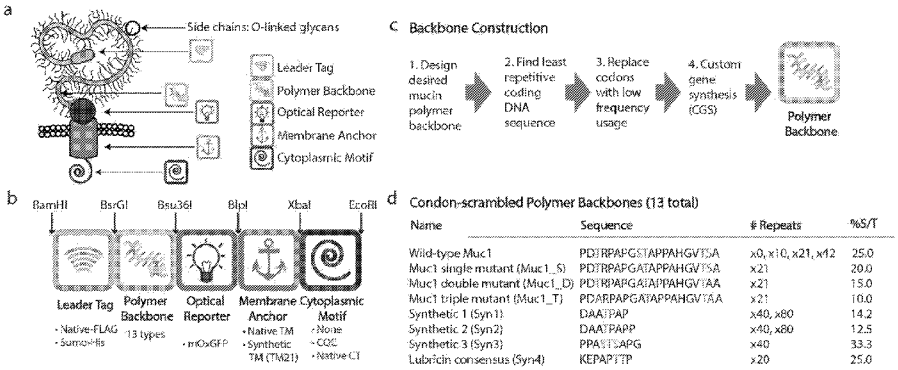


Figure 1

(57) **Abstract:** Provided are compositions and methods related to modified lubricins, methods of making the modified lubricins, and methods of using the modified lubricins for coating a variety of inanimate objects, and for prophylaxis and/or therapy of disorders where enhanced lubrication of one or more parts of a human or non-human mammal is desirable.

WO 2020/150396 A1

[Continued on next page]

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

RECOMBINANT LUBRICINS, AND COMPOSITIONS AND METHODS FOR USING THE SAME

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. provisional patent application no. 62/792,660, filed January 15, 2019, the entire disclosure of which is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under grant nos. 1DP2GM119133-01 and 1U54CA210184-01 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE DISCLOSURE

The disclosure provided improved glycoproteins, and compositions and methods related to the same.

BACKGROUND OF THE DISCLOSURE

Lubricin is a glycosylated protein found in several places in mammalian anatomy. For example, lubricin is present in synovial fluid and on the surface of cartilage. Lubricin has an important role in lubrication of joints and maintaining the correct joint environment.

Previous attempts have been made to provide recombinant forms of lubricin, but there remains an ongoing and unmet need for new lubricin and lubricin-like glycoproteins that can be employed in a wide variety of environments. The present disclosure is pertinent to this need.

SUMMARY OF THE DISCLOSURE

The present disclosure provides compositions and methods that relate to modified glycoproteins. Aspects of the disclosure pertains to modified lubricins, pharmaceutical compositions that contain the modified lubricins, cDNAs and expression vectors that encode the modified lubricins, eukaryotic cells that express the modified lubricins, and methods of using the modified lubricins and compositions comprising them for a variety of purposes. The methods include use of such agents for prophylaxis and/or therapy of a variety of conditions where improved lubrication of a surface or fluid within a human or a non-human mammal is desirable. The disclosure also includes using the compositions to provide lubrication to the surface of a variety of inanimate objects.

In certain embodiments, the modified lubricins comprise a change in a number of tandem repeats of specific amino acid sequences, and/or one or more changes in the amino acid sequences of the modified lubricins, relative to their naturally produced counterparts. In embodiments, the modified lubricins comprise amino acid sequences that are derived from human, equine, or canine lubricins, but have different functional attributes relative to previously provided recombinant versions of such sequences. In an embodiment, the modified lubricins have an increased half-life, such as an intra-articular half-life when injected into a mammal, of more than 4 days. In embodiments, the modified lubricins exhibit an intra-articular half-life of more than 15 days, or at least 30 days. In embodiments, the modified lubricins have a modified glycosylation pattern, relative to an unmodified lubricin.

In embodiments, the modified lubricins include contiguous repeated sequences that are one or a combination of KEPAPTTP (SEQ ID NO:1), KEPAPTP (SEQ ID NO:9) and KEPAPTTTP (SEQ ID NO:10). In embodiments, the repeated sequence is repeated contiguously 10-120 times. In a non-limiting embodiment, the repeated sequence is repeated 59 times.

In embodiments, the modified lubricins comprise amino acid sequences that are derivatives of lubricins produced by human or non-human mammals. In embodiments, the contiguous repeated sequences are flanked on their N- and C-terminal segments by lubricin amino acid sequences that are at least 90% identical to human, equine, or canine lubricin sequences.

In embodiments, the modified lubricins include additional components, such as an added secretory signal from a human, or a non-human mammal, or other suitable source.

BRIEF DESCRIPTION OF THE FIGURES

The figures and tables of this disclosure are divided into four Parts (Part I, Part II, Part III and Part IV), as described below.

Part I Figures

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. The Wild-type Muc1 sequence is SEQ ID NO:8. The Muc1 single mutant (Muc1_S) is SEQ ID NO:5. The Muc1 double mutant (Muc1_D) is SEQ ID NO:6. The Muc1 triple mutant (Muc1_T) is SEQ ID NO:7. The

Synthetic 1 (Syn1) is DAATPAP is SEQ ID NO:2. The Synthetic 2 (Syn2) is SEQ ID NO:3. The Synthetic 3 (Syn3) is SEQ ID NO:4. The Lubricin consensus sequence (Syn4) is SEQ ID NO:1.

Figure 2: Construction and Validation of Sequence-Specific Mucin Expression. (a) Components and features of codon-optimized Muc1 variants with GFP reporters. The amino acid sequence in (a) is SEQ ID NO:8 (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. The sequences under sMuc1 mutants (21 repeats) are from top down: SEQ ID NO:8, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 (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. The sequences shown in Figure 4 are KEPAPTTP (SEQ ID NO:1) DAATPAP (SEQ ID NO:2) DAATPAPP (SEQ ID NO:3) and PASTSAPG (SEQ ID NO:4). (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 (SEQ ID NO:1) 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. * $P<0.05$

Figure 6: Western blot analysis of MCF10A cells edited with lentivirus with native repetitive (Native_Muc1) versus codon-scrambled Muc1 cDNAs (Muc1_42).

Figure 7: Mucins with Tunable Sizes. The sequences shown in Figure 7 are PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8) (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 8: 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 9: Cell-Surface Mucin Mutants Derived from Muc1 Tandem Repeat Sequences. The sequences shown in Figure 9 under mMUC1 mutants (21 repeats) from top down are PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8), PDTRPAPGATAPPAHGVTSA (SEQ ID NO:5) PDTRPAPGATAPPAHGVTAA (SEQ ID NO:6) and PDARPAPGATAPPAHGVTAA (SEQ ID NO:7). (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 10: 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 11: Mucins Constructed with Designer Tandem Repeats. The sequences shown in Figure 11 are DAATPAP (SEQ ID NO:2) DAATPAPP (SEQ ID NO:3) and PPASTSAPG (SEQ ID NO:4). (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).

Part II Figures

Figure 12: Engineering Biopolymer-Coated Cell Lines. A transposon-based method was used to stably integrate the DNA encoding the engineered biopolymers under a doxycycline inducible promoter. **A,** Schematic representation of the all-in-one vector used for producing biopolymer-coated cell lines showing key elements. For incorporation into the cellular genome, the vector includes a tetracycline responsive element (tetO), a minimal CMV promoter, the Muc1 signal sequence (Muc1 N-terminus), the tandem repeats of the biopolymer (0, 21, or 42 repeats of PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8)), the transmembrane domain of Muc1 (Muc1 TM), the bicistronic green fluorescent protein reporter (IRES GFP), a EF-1α promoter, the reverse tetracycline transactivator (rtTA), and a second bicistronic neomycin resistance cassette (IRES NeoR). These elements were all flanked by 5' and 3' inverted terminal repeat sequences (ITRs) required for transposon-mediated incorporation into the genome. For vector replication and production in bacteria, there was also an ampicillin resistance cassette (AmpR) and an origin of replication (ori). **B,** Schematic representation of membrane bound biopolymers expressed by the cells and localized to the cells surface. **C,** Schematic of the relative size of the extracellular domain of the engineered biopolymers designated Mucin-0, Mucin-135, and Mucin-270 for their

respective length in nm. The predicted molecular weight of these proteins was 42 kDa, 81 kDa, and 120 kDa, respectively.

Figure 13: Validation of Biopolymer Coatings. Expression and cell-surface localization of biopolymer coatings was validated for the new, engineered 293-F cell lines. **A,** Representative confocal microscopy images of stable suspension adapted human embryonic kidney 293 (293-F) cell lines – wild type (w.t.), or stably expressing the Mucin-0, Mucin-135, or Mucin-270 biopolymer. Images show the cell membrane (shown in blue, CF633 Wheat Germ Agglutinin, WGA), *O*-glycans covalently attached to the Mucin-135 and Mucin-270 biopolymers (shown in red, CF568 Peanut Agglutinin, PNA), and green-fluorescent protein (shown in green, GFP) which is co-expressed on the plasmid with the Mucin-0, Mucin-135 and Mucin-270 biopolymer. **B,** Representative flow cytometry histograms showing the polydisperse population of biopolymer expressing cell lines compared to w.t. cells, y-axis is scaled to show the population distribution of GFP positive cells. >50,000 cells per histogram. **C,** Quantification of the percent of cells which are GFP positive for each cell line. Cells with GFP signal above the gray line in Fig. 2B were considered GFP positive. Mean and S.D. are shown, >50,000 cells per sample, n = 4. **D,** Representative immunoblot (left) and lectin blot (right) of whole cell lysates for each generated stable cell line compared to w.t. cells, n = 3. **E,** Viable cell concentration determined by hemocytometer counting with trypan blue exclusion, n = 3. **F,** GFP signal of Mucin-270 cells after induction of expression at t = 0 hr, measured by flow cytometry, n = 3, >15,000 cells per sample. **G,** Agarose gel showing polymerase chain reaction (PCR) product of Mucin-270 gene from DNA extracted from non-transfected cells (Mock), w.t. cells transiently transfected (Transient), or cells with the Mucin-270 gene incorporated in the genome and cultured for 2 months (2 mo.) or 12 days (12 d) after gentamycin selection. Star indicates the predicted molecular weight of Mucin-270 PCR product. #1 and #2 are biological replicates. Mean and S.D. shown, ns – not significant.

Figure 14: Biopolymer Coatings Reduced Cell Aggregation. Genetically-encoded biopolymer coatings of Mucin-135 and Mucin-270 size reduce cell aggregation in suspension cell culture. **A,** Representative phase contrast images for w.t. and biopolymer cell lines. Images were for cells grown at a concentration of $3.8 \pm 0.7 \times 10^6$ cells/mL at 72 hr post-induction. **B,** Quantification of the fraction of cells in various cluster sizes from phase contrast images such as those shown in Fig. 3A, 3 biological replicate samples, 2 technical replicate samples, 3 images analyzed per sample, samples (further discussion of replicates in Materials and Methods section). Center lines show the medians; box limits indicate the 25th

and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; crosses represent sample means. **C,** Quantification of the fraction of cells which are in clusters of various sizes from phase contrast images such as those shown in Fig. 3A. Mean and S.D. are shown. **D,** Ripleys K function versus distance calculated for the cell distribution acquired from phase contrast images such as those shown in Fig. 3A. Mean and S.E.M. are shown, replicates described in Fig. 3B. ns – not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$.

Figure 15: Mucin-270 Reduced Aggregation in High Calcium Culture Media. The Mucin-270 cell line out-performs commercial anti-clumping solution in highly aggregating conditions. **A,** Image of Mucin-270 and w.t. cultures grown in media with 2 mM $CaCl_2$ (+ Ca^{2+}). Mucin-270 expression significantly decreases cell aggregation, even compared to commercially available anti-clumping reagent (+ anti-clump). **B,** Quantification of the concentration of w.t. or Mucin-270-expressing cells in suspension for control cultures with no treatment (null), with the addition of commercial anti-clumping reagent (+ anti-clump), with the addition of 2 mM $CaCl_2$ (+ Ca^{2+}), or with both anti-clumping reagent and 2 mM $CaCl_2$ (+ anti-clump + Ca^{2+}). Statistical comparison is to null condition for each cell line. Mean and S.D. are shown, n = 3. ns – not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$.

Figure 16: Biopolymer Coating Enhanced Resistance to Shear Stresses. Expression of the stably incorporated biopolymers protects cells from shear stresses. **A,** Schematic representation of the experimental setup for shearing cells. Briefly, cells were sheared by flowing through a 500 μm Teflon tube under a constant applied force of 1 kg in gravity before being analyzed by flow cytometry with a live/dead cell stain. **B,** Quantification of the fraction of dead cells after shearing the cells for the w.t. and biopolymer cell lines, Mean and S.E.M. are shown, > 50,000 cells measured for each population, n = 6. ns – not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$.

Figure 17: Biopolymer Coated Cells can be Transfected. Transfection was determined for the biopolymer coated cell lines by transfection with a cytoplasmic red-fluorescent protein (RFP). **A,** Quantification of the number of cells for w.t. and biopolymer coated cells transiently transfected with cytoplasmic RFP. The count of transfected cells was normalized to the count of w.t. cells transfected per experiment to account for variable transfection efficiency between replicate transfections. > 50,000 cells measured for each population, n = 3. **B,** Representative flow cytometry histogram showing the distribution of expression among transfected cell populations. The peak to the left of the gray line, centered around zero, represented the non-transfected population for each cell line which is further

validated by the overlapping histogram of non-transfected w.t. cells (w.t.-null). **C,** Quantification of the geometric mean of RFP for positively transfected cells from B. Mean and S.D. shown, ns – not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$.

Figure 18: Mucin-270 cells Produced Comparable Levels of Recombinant Protein Expression. Quantification of secreted, recombinant RFP from media supernatant of w.t. or Mucin-270-expressing cultures transiently transfected with secreted RFP, n = 3. Mean and S.D. shown, ns – not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$.

Figure 19: Additional data to accompany Figure 14 acquired 24 hr prior. **A,** Quantification of the fraction of cells in various cluster sizes from phase contrast images such as those shown in Fig. 3A. Cells are grown at $3.2 \pm 0.7 \times 10^6$ cells/mL for 48 hr for all panels. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; crosses represent sample means. **B,** Quantification of the fraction of cells which are in clusters of various sizes from phase contrast images such as those shown in Fig. 3A. Mean and S.D. are shown. **C,** Ripley's K function versus distance calculated for the cell distribution acquired from phase contrast images such as those shown in Fig. 3A. Mean and S.E.M. are shown, replicates described in Fig. 3B, n =3,. ns – not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$.

Part III Figures

Figure 20: Design and synthesis of synonymous lubricin (SynLubricin). A) Overview of the design and production strategy for synthetic, codon-scrambled mucins. DNA sequences for the desired protein product were optimized through a global optimization to minimize repetitive DNA sequences by codon scrambling, followed by a second optimization that reassigned codons with infrequent usage in the host cell system. **B)** SynLubricin was constructed of 59 perfect repeats of KEPAPTTP (SEQ ID NO:1)flanked by the native human N- and C-termini of PRG4. An IgK signal sequence and SumoStar tag was fused to SynLubricin for secretion and purification. SynLubricin also retains the two somatomedin B domains (SMB 1 and 2) and the two Hemopexin domains of the native protein. **C)** Calculated repetition score for the nucleotides encoding the tandem repeats of human PRG4 isoform A (PRG4A) and SynLubricin. **D)** Alignment of amino acid sequence of human PRG4 and SynLubricin. The PRG4A sequence in the alignment is amino acids 347 - 853 of SEQ ID NO:66. The SynLub sequence in the alignment is amino acids 347-818 of SEQ ID NO:68. **E)** Vector map illustrating the tetracycline-inducible promoter, multiple cloning site (MCS) for

cDNA of interest, bicistronic GFP reporter (IRES2 CopGFP), and second expression cassette for the rtTA-M2 tetracycline transactivator and neomycin-resistance gene.

Figure 21: Sorting strategy to isolate stable polyclonal cell populations that produce high levels of SynLubricin. A) Strategy for isolation of stable cell populations expressing high levels of SynLubricin. **B)** Western blots of 293-F media supernatant showing relative SynLubricin production in unsorted and twice-sorted (2x) cell populations; 1 and 2 indicate samples from two independent experiments; probed with anti-PRG4 (MABT401) and SUMO antibodies. **C)** Quantification of the relative intensity of signal from anti-PRG4 Western blots in B. **D)** Phase-contrast and fluorescence micrographs of unsorted and twice-sorted 293-F cells expressing SynLubricin.

Figure 22: Integrated SynLubricin cDNA is stable in the cellular genome. PCR amplification of SynLubricin coding region in genomic DNA extracts of wild-type and stably integrated 293-F cells cultured continuously for 2 months. As positive controls, PCR amplifications of SynLubricin plasmid and DNA extract from SynLubricin transiently transfected 293-F cells (Transient) are shown. The expected size of full-length SynLubricin is indicated by the star.

Figure 23: Optimization of SynLubricin production. A) Western blots showing relative production of SynLubricin over time in media of control cells and sorted 293-F cells induced with 1 μg/mL doxycycline for the indicated number of days in the absence or presence of the histone deacetylase inhibitor valproic acid (VPA; 3.5 mM). **B)** Quantification of the relative intensity of signal for the blots shown in A. **C)** Time course for glucose consumption in sorted 293-F cells induced at day 0 with 1 μg /mL doxycycline with or without 3.5 mM VPA. Mean and S.D. shown, n = 3. **D)** Western blot showing lubricin in the media harvested from non-producing control cells (Mock), cells transiently transfected with SynLubricin cDNA (Transient), and two successive 1-L batch cultures of sorted 293-F cells induced for three days with 1 μg /mL doxycycline and 3.5 mM VPA (Batch #1 and Batch #2); equine synovial fluid (ESF) was loaded as a control. **E)** Representative Western blot of SynLubricin produced from stably expressing 293-F cells collected at indicated time points after 1 μg/mL doxycycline induction on day 0. **F)** Quantification of Western blot replicates represented in B, n = 3, ns – not significant.

Figure 24: Purification of SynLubricin by anionic exchange chromatography. A) Sliver stain and **B)** Western blot showed SynLubrcin eluted continuously from Q Sepharose® resin over a broad range of NaCl concentrations (concentrations indicated above lanes in mM). **C)** Sliver stain and **D)** Western blot showing harvested SynLubricin media supernatant

(M), 10-fold diluted SynLubricin media supernatant (S), wild-type 293-F conditioned media (C), flow through (FT-1x), 10-fold concentrated flow through (FT-10x), and eluted fractions at indicated salt concentration (shown above lanes in mM).

Figure 25: Lubrication of cartilage explants shows functionality of SynLubricin. Friction coefficients of NaCl-extracted cartilage explants bathed in saline (PBS), bovine synovial fluid, or SynLubricin. Prior to lubrication analysis, the SynLubricin was purified with DEAE Sepharose, eluting either without washing or after a stringent 500 mM NaCl wash. Mean and S.D. are shown with independent measurements indicated. ***$p<0.001$, ****$p<0.0001$; NA: statistical testing is not applicable due to sample size.

Figure 26: Transient expression of SynLubricin altered adherent cell morphology. A) Morphology of 293-T cells mock transfected or transfected with cDNA for bicistronic SynLubricin IRES copGFP. Images shown are a merged overlay of phase contrast and fluorescence micrographs. Note the inhibition of cell-cell adhesion near cells expressing high levels of the copGFP reporter. **B)** Western blot of equine synovial fluid (ESF) and media supernatant from mock-transfected and SynLubricin-transfected cells probed with MABT401 antibody against PRG4 tandem repeats.

Figure 27: Validation of new transposon-based gene delivery vector. Flow cytometry results showing correlation of levels of mCherry2 and the copGFP reporter.

Figure 28: The sequence shown on Figure 28 is PDTRPAPGSTAPPAHGVTSA (unmodified Muc1 repeat) (SEQ ID NO:8). **Application of the codon-scrambling strategy for Muc1. A)** Schematic of SynMuc1 with codon-scrambled tandem repeats. **B)** Calculated repetition score for the nucleotides encoding the tandem repeats of human Muc1 and SynMuc1. **C)** Western blot of media supernatant from 293-F cells transfected with SynMuc1 cDNA (+cDNA) or non-transfected cells (M), Ni-NTA resin flow through from His-affinity purification (FT), and eluted protein (Elution) probed with a Muc1 antibody. **D)** PNA-lectin blot of C. **E)** Western blot of C, probed with a SUMO antibody.

Figure 29: SynLubricin has low affinity for immobilized-metal-affinity-chromatography (IMAC) resin. A) Western blot of media supernatant and the IMAC purification flow throughs, washes, and eluted fractions from Fe^{3+} and Ni^{2+} loaded nitrotriacetic acid (NTA) resins. Elutions were performed at the indicated NaCl concentration. No non-specific binding of sialic acids to multivalent Fe^{3+} was observed. **B)** Western blot of flow through, wash, and eluted fractions from uncharged NTA resin.

Figure 30: Images showing retention of SynLubricin *in vivo*. Images depict localized SynLubricin at days 0 (injection), 2 weeks, and 7 weeks. The graph shows half life for two

rats over a period of about 40 days. For the graph, clearance kinetics of injected human SynLubricin into the left knee of adult male Sprague-Dawley rats (n=2). Purified human SynLubricin was fluorescently labeled with sulfo-Cy7.5 near-infrared fluorescent dye, and 20 μL of SynLubricin-Cy7.5 was injected into the healthy, left knee via a patellar tendon approach. Total lubricin fluorescence from the left knee was imaged on an IVIS Spectrum whole animal imaging system, quantified and reported as total radiant efficiency. Data was fitted to a bi-exponential decay model to calculate the alpha and beta decay constants. Half-life is reported as ln(2) divided by the beta decay constant.

Figure 31: Graph and chart showing MALDI-MS spectrum of O-glycans released from human SynLubricin and computed relative percentages.

Figure 32: Friction coefficients of NaCl-extracted cartilage explants bathed in saline (PBS) or SynLubricin at the indicated concentration. For these experiments, SynLubricin was purified from 293-F media supernatant by cation-exchange chromatography.

Part IV Figures

Figure 33: Glycocalyx polymers induce membrane projections. (A) Schematic and table illustrating the genetically encoded biopolymers that were constructed and used throughout this work. The gene library encoded native and synthetic mucins comprised of a central polypeptide core, sugar side chains linked to serine (S) and threonine (T) residues, and a transmembrane anchor. **(B)** Quantification of membrane tube density in epithelial cells, showing mucin polymers induce dramatic tubularization compared to wild-type (Control) cells. Number of cells analyzed is shown on the x-axis for each condition. Box notches here and elsewhere indicate 95% confidence intervals. **(C)** Scanning electron microscopy (SEM) images showing membrane morphologies of cells expressing the indicated biopolymer. **(D)** (left) Cartoons of Muc1 GFP-ΔCT polymers of varying length, as indicated by the number of tandem repeats (TR). (right) Flow cytometry data showing similar cell-surface expression levels of indicated mucins using a GFP-binding nanobody, n = 3, > 40,000 cells per population. **(E)** Representative SEM images of cells described in (D). **(F)** (left) Quantification of relative protein surface density on giant unilamellar vesicles (GUVs) with membrane-anchored Podocalyxin (Podxl) at low density, human serum albumin (HSA) at low density (Low HSA), or HSA at high density (High HSA), n = 10-20. All GUVs were formulated with 10 mole % Ni-NTA-lipid for protein anchorage. (center) Quantification of the fraction of GUVs with or without tubes; n is the number of GUVs analyzed for each

protein. (right) Representative confocal images of GUVs. *** $p < 0.001$ (*post-hoc* student's two tailed *t* test).

Figure 34: Membrane morphology of tissue synoviocytes is regulated by the glycocalyx. (A) Experimental workflow for resected equine synovial tissues. (B) Representative SEM images of hyaluronic acid synthase 3 (HAS3) expressing primary synoviocytes showing retraction of membrane tubules following 30 minutes of hyaluronidase (HyA) treatment to digest hyaluronic acid (HA). **(C)** Quantification showing tubule density was dependent on the presence of HA. **(D)** Images of freshly resected synovial tissue showing the nucleus (DAPI), surface-anchored HA (hyaluronic acid binding protein, HABP) of a representative synoviocyte, and the tissue collagen (second harmonic generation, SHG). Depth along the z-axis is coded according to the color bar. Note the HA-enriched membrane extensions protruding from the synovial tissue surface. Lower right panel shows a cartoon representation of the observed tissue synoviocyte. **(E)** Membrane tubules are visible, by SEM, on synoviocytes in freshly excised equine synovial tissue. The synoviocyte head is pseudo-colored in orange protruding from the synovial tissue. HyA treatment to digest HA resulted in the rapid retraction of synoviocyte tubules (right). *** $p < 0.001$ (*post-hoc* student's two-tailed *t* test).

Figure 35: Polymer brush model of the glycocalyx and generation of preferred membrane shapes. (A) Polymer model of membrane bending illustrating proposed spontaneous membrane curvature induced by the cellular glycocalyx. Low density polymers are non-interacting and adopt a compact structure in the "mushroom" regime. In the "brush" regime, polymers overlap (the average distance between polymers, D, is less than the twice the radius of gyration, R_G) and extend to avoid each other, increasing the height of the polymer brush (H). Entropic pressures are the basis for membrane curvature generation by polymer mushrooms and brushes. **(B)** Muc1 construct with SUMO and GFP tags flanking the polymer domain for visualization of polymer extension with expansion microscopy (ExM). Polymer extension versus polymer fluorescence intensity, a proportional measure of surface density, showing the indicated scaling relation. Dots, squares, and triangles indicate measurements from three samples. The red line shows a linear regression through all data points. **(C)** Theoretical prediction of spontaneous curvature generation by Muc1 polymer mushrooms and polymer brushes. Blue: estimated mushroom regime (mush.); pink: estimated brush regime (brush). The computational model here considers mucins of length 270 nm having monomeric segments of length 15 nm (Kuhn length). These parameters were based on experimental characterization of native Muc1-42TR and selected for comparison to

experiments below. **(D)** (left) Theoretical prediction of required pressure (Pa) as a function of mucin concentration for blebs of radii = 250 nm. The insert shows a pressure minimum near the mushroom-brush transition. (right) Theoretical prediction of the required point force (pN) as a function of mucin concentration for maintaining membrane tubules.

Figure 36: Preferred membrane shape depends on cell-surface biopolymer concentrations. (A) Strategy for sorting cells into populations with varying levels of cell surface mucin (Muc1-42TR-GFP ΔCT) using fluorescence-activated cell sorting (FACS). **(B)** Representative SEM images showing the transition of membrane morphological features of sorted cell populations with the indicated mucin surface density. Mucin densities were chosen to match the indicated points on the theoretical graphs (Fig. 3D). **(C)** Average radius of bleb structures measured in the mushroom regime and tube structures measured in the brush regime. **(D)** Observed density of membrane blebs on sorted cell populations having the indicated average mucin surface density. Significance was determined between mushroom regime and brush regime (*) or between the lowest brush regime density and all other brush mucin densities (+). **(E)** Observed density of membrane tubes on sorted cell populations having the indicated average mucin surface density. Symbols defined in (D). **(F)** Inverse predicted force from (Fig. 3D, right) versus the observed tube density from (E) exhibits a linear relationship and Pearson correlation coefficient of 0.97. Number of measurements shown on the x-axis of boxplots. Error bars indicate 95% confidence intervals. ns - not significant; */+ $p < 0.05$; **/++ $p < 0.01$; ***/+++ $p < 0.001$ (*post-hoc* student's two-tailed *t* test).

Figure 37: Glycocalyx-mediated membrane instabilities and extracellular vesicle biogenesis. (A) Representative confocal microscopy images of epithelial cells expressing Muc1-42TR ΔCT and stained with PNA (peanut agglutinin) for mucins and phalloidin for actin, n = 3. **(B)** Fluorescent intensity line trace from (A) (PNA image, red line). Values are normalized for their respective maximum intensities for phalloidin and PNA stains. **(C)** Average diameter of tubules in Muc1-42TR ΔCT expressing cells following treatment with DMSO (Vehicle) or with 10 μM Latrunculin-A (+ LatA) to disrupt actin assembly. **(D)** Representative SEM images of tubules in vehicle treated or LatA treated cells expressing Muc1-42TR ΔCT. **(E)** (left) Cartoon schematic of a proposed model in which the actin core resists the spontaneous membrane curvature driven by the glycocalyx brush. Upon actin depolymerization, membrane tubules are destabilized and predicted to relax into (right) various pearled structures and/or thin tubes that represent minimal energy surfaces.

Schematic drawings of these predictions are shown alongside representative pseudo-colored SEM images of cells expressing Muc1-42TR ΔCT. **(F)** Cartoon schematic of proposed mechanism where pearling and vesiculated membrane instabilities (left) are disrupted and lead to microvesicle shedding (right). **(G)** Representative histogram showing the average concentration and size distribution of extracellular vesicles for wild-type (Control) and Muc1-42TR ΔCT expressing cells and **(H)** showing Muc1-42TR ΔCT cells treated with DMSO (Vehicle) or Latrunculin A (+ LatA). Particle concentration is normalized to the max peak for each graph. Shaded area shows 95% confidence interval, n = 5, 5, 4, 7, respectively. **(I)** Representative cryogenic transmission electron microscopy (cryo-TEM) image of a vesicle collected from cells expressing Muc1-42TR ΔCT. Red boxes indicate pseudo-colored regions of interest shown on the right. *** $p < 0.001$ (*post hoc* two-tailed student's *t* test).

Figure 38. **Validation of genetically encoded mucins.** The sequences in (A) are from top down are PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8) and PPASTSAPG (SEQ ID NO:4). (A) Cartoon representations of the genetically-encoded glycoproteins. Mucin-1 (Muc1) contains 42 repeats of PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8) and Podocalyxin (S/T-Rich) has a serine- and threonine-rich region for *O*-glycosylation. The engineered glycoproteins lack the native cytoplasmic tail signaling domain (ΔCT) while retaining the native transmembrane domain (TM) or exchanged with a synthetic 21-amino-acid transmembrane anchor (TM21). The rationally designed mucin (Rational GFP-ΔCT) contains 80 repeats of PPASTSAPG SEQ ID NO:4) fused to a fluorescent marker (GFP) and the native stalk and TM without the native cytoplasmic tail signaling domain (ΔCT). **(B)** Representative confocal microscopy images showing membrane tubularization induced by various engineered glycoproteins compared to wild-type (Control) cells. The cell surface is visualized with lectin WGA (wheat germ agglutinin). Mucin staining with lectin PNA (peanut agglutinin) confirms glycoprotein *O*-glycosylation and surface localization on MCF10A cells, n = 3. **(C)** Quantification of endocytosis of Alexa Fluor 488 labeled transferrin (488 TNF) after 0.5 or 1 h of treatment. Quantification performed with flow cytometry, median signal reported with background subtraction, > 10,000 cells per population, n = 6, error bars are S.D. **(D)** Representative confocal microscopy images of endocytosed 488 TNF after 0.5 h of treatment. **(E)** Western blot showing polymer sizes expressed in epithelial cells, analyzed with an antibody against the green fluorescent protein (GFP) tag, n = 2. **(F)** Quantification of tube density for the indicated mucin size. Number of cells analyzed is shown on the x-axis for each condition. Box notches indicate 95%

confidence intervals. Statistical comparison is to 42TR. ns – not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (*post-hoc* student's two-tailed *t* test).

Figure 39: Hyaluronic acid localizes on the cell surface and induces cell-surface projections. (A) (left) Cartoon of hyaluronic acid (HA) extruded by the transmembrane protein hyaluronic acid synthase 3 (HAS3). (right) Blot of HA in lysates of wild-type (Cont.) and hyaluronic acid synthase 3 (HAS3) expressing human mammary epithelial cells (MECs, MCF10A). Note that the expressed HA is a giant linear polymer in the MDa range. **(B)** ELISA quantification of HA secreted by MECs into their media, normalized to the number of cells in the sample and the HA secretion of Control cells, n = 3. **(C)** Representative confocal microscopy images of human MECs, either wild-type (Control) or stably expressing HAS3. Cells are stained with Hoescht (nucleus) and Alexa Fluor 568 hyaluronic acid binding protein (HABP). **(D)** Representative SEM images showing highly elongated membrane tubules in HAS3-expressing human MECs (left) and a zoomed in region on the same cell (right). ** $p < 0.01$ (*post-hoc* student's two-tailed *t* test).

Figure 40: Mucins cause tubularization of model lipid membranes. (A) Representative confocal images of DOPC giant unilamellar vesicles (GUVs) labeled with Bodipy-PC with an increasing fraction of Ni-NTA lipids. Recombinant Alexa Fluor 568-labeled Podocalyxin (Podxl) associates with the GUV via a polyhistidine tag. Scale bar is 5 μm in each BODIPY-PC image. **(B)** (left) Quantification of fluorescent intensity (relative surface density) of Alexa Fluor 568-labeled human serum albumin (HSA) or Podxl on GUVs at different Ni-NTA lipid levels, n = 10-20. A similar HSA surface density to the mucin surface density (Low HSA) and a several-fold higher HSA surface density (High HSA) were used to control for protein crowding effects. (right) Quantification of the fraction of GUVs with tubes at different Ni-NTA lipid levels for each recombinant protein – Low HSA, High HSA, and Podxl, error bars are standard deviation, n = 20-90 GUVs over 1-3 experiments. **(C)** Representative confocal image of Alexa Fluor 568-HSA for a GUV with High HSA forming tubules.

Figure 41: Supporting information for physical characterization of individual mucins and mucin ensembles. (A) Cartoon representation of the recombinant Muc1 42 tandem repeat (Muc1-42TR) polymer fused to a 10x-histidine tag. **(B)** Western blot validation of recombinant Muc1-42TR production (Media + Muc1-42TR 10xHis), Ni-NTA resin binding of the protein (Flowthrough), wash of non-specific proteins (Wash), and purified recombinant Muc1-42TR polymer (Elution). Samples are probed with anti-Muc1 and anti-His antibodies as well as PNA (peanut agglutinin) to bind *O*-linked glycans. **(C)** SYPRO

Ruby protein gel stain for samples described in B. **(D)** Quantification of epithelial microvilli diameter for the indicated relative mucin surface densities. Box notches indicate 95% confidence intervals. **(E)** (left) Mucin construct (Muc1-42TR) with SUMO and GFP tags flanking the polymer domain for visualization of polymer extension with expansion microscopy (ExM). (right) ExM sample workflow. First, samples are stained and fixed. Then the proteins are chemically linked (anchored) to monomers which polymerize to form a gel. Proteins are then digested, and the gel is expanded to four times the original size. ns – not significant.

Figure 42: Additional polymer brush theory predictions for curvature generation by intermolecular interactions in the glycocalyx. (A) Graph for the predicted brush thickness as a function of biopolymer surface density in the brush regime. Brush thickness scales approximately as a power law with biopolymer concentration. **(B)** Plot showing energetic contributions as functions of the biopolymer density. In the mushroom regime, polymers have only elastic energy, while in an extended brush, excluded volume and electrostatic interactions contribute to biopolymer free energy. **(C)** Plot depicting variation of spontaneous curvature generated with biopolymer density and molecular length. **(D)** Graph displaying trend of spontaneous curvature as a function of biopolymer density and Kuhn length. Kuhn length, equal to twice the persistence length, is directly proportional to polymer bending stiffness, and is referred to as the length of a monomeric segment in the manuscript. Plots in (A-D) are in log-log format. Plots in (A) and (B) use biopolymer length, $l = 270$ nm, and monomeric segment length, $l_a = 15$ nm. Plot (C) employs polymer monomer segment size of 15 nm, and (D) uses biopolymer length of 270 nm. **(E)** Predicted dependence of spontaneous curvature on biopolymer length at high density. This graph uses polymers of $l_a = 15\ nm$ packed at a density of $50000\ \#/\mu m^2$.

Figure 43: Fluorescence-activated sorting and quantification of Muc1 surface densities. (A) Extended workflow for quantitative experiments at different Muc1 surface densities. **(B)** SDS-Page calibration of Alexa Fluor 647 labeled nanobody. **(C)** Calibration curve between the log value for integrated density of fluorescence signal from nanobody dilution series (shown in (B)) versus the log value of the number of molecule loaded. A linear regression fit and R^2 value are shown. **(D)** Residuals for the linear regression fit shown in (C). **(E)** Fluorescence-activated cell sorting (FACS) histogram showing the nanobody fluorescence signal and the populations 'a' through 'e' collected for these experiments. **(F)** Representative scanning electron microscopy (SEM) images of wild type cells which were

non-enzymatically detached from the substrate then re-adhered (detached control) for SEM imaging and cells which were non-enzymatically detached from the substrate, collected through the FACS, then re-adhered (FACS control). These images demonstrate that the method of FACS collection did not influence the membrane shapes observed with Muc1-42TR ΔCT expression (shown in Fig. 2F). **(G)** SDS-Page analysis of fluorescent nanobody signal in each cell population, a-e, after collection and lysis of the cells. **(H)** Table describing the integrated density signal from the fluorescence image shown in (G), the calculated number of molecules based on the calibration curve in (C), and the number of cells loaded in the protein gel, (G), based on the number of cells collected with FACS for each population, (E). **(I)** Calibration curve between the log of the nanobody mean signal from the FACS versus the number of molecules calculated for each population. The number of molecules per sample was normalized by the number of cells loaded and the approximate area per cell. Linear regression fit and R^2 values shown. **(J)** Residuals for linear regression fit shown in (I).

Figure 44: Tubular membrane shapes contain filamentous actin cores and resemble microvilli. (A) Representative confocal microscopy images of epithelial cells expressing Muc1-42TR ΔCT showing indirect microtubule staining with anti-microtubule and Alexa Fluor 568-labeled secondary antibodies. Mucins are labeled with Alexa Fluor 647 PNA (peanut agglutinin). The bottom row shows the region of interest from the composite image (yellow box), n = 3. **(B)** Fluorescent intensity line trace from (A) (bottom row, yellow line). Values are normalized for their respective maximum intensities. **(C)** Representative confocal microscopy images of epithelial cells expressing Muc1-42TR ΔCT showing actin staining with Alexa Fluor 568 phalloidin. Mucins are labeled with Alexa Fluor 647 PNA. The bottom row shows the region of interest from the composite image (yellow box), n = 3. This data repeats and elaborates on (Fig. 5A, B). **(D)** Fluorescent intensity line trace from (C) (bottom row, yellow line). Values are normalized for their respective maximum intensities. **(E)** Representative confocal microscopy images of the midplane of wild type (Control) or Muc1-42TR ΔCT cells which have been treated with 10 μM Latrunculin-A (LatA) for 1 h, n = 3. **(F)** Representative SEM image of LatA treated Muc1-42TR ΔCT cells.

DETAILED DESCRIPTION

Unless specified to the contrary, it is intended that every maximum numerical limitation given throughout this description includes every lower numerical limitation, as if such lower numerical limitations were expressly written herein. Every minimum numerical limitation given throughout this specification will include every higher numerical limitation,

as if such higher numerical limitations were expressly written herein. Every numerical range given throughout this specification will include every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein.

The disclosure includes every amino acid sequence described herein, and every polynucleotide sequence that encodes the amino acid sequences, including but not limited to cDNA sequences, and mRNA sequences. Complementary sequences, and reverse complementary sequences are also included. Expression vectors comprising such nucleotide sequences are encompassed by the disclosure.

The disclosure relates generally to improved glycoproteins, compositions comprising the proteins for use in diverse applications, and methods of making and using the glycoproteins. In embodiments, the glycoproteins are mucins and/or lubricins.

The disclosure includes cells and cell cultures that express the proteins described herein. In certain embodiments, the disclosure includes cell cultures that are improved for producing any of a variety of proteins due to reduced clumping, aggregation, etc. of the cells.

In embodiments, the cells that are used to express proteins of this disclosure are eukaryotic cells. In certain embodiments, the cells are eukaryotic cells, including but not limited to insect and mammalian cells. In embodiments, the mammalian cells are not Chinese hamster ovary (CHO) cells, although in certain instances CHO cells may be used. In embodiments, the cells are mammalian epithelial cells. In embodiments, the cells are human cells, and thus are better suited for producing, for example, human biologics, than non-human mammalian cells. In embodiments, the cells are human 293 cells. In embodiments, 293 cells are derived from 293 cells and stably express the SV40 large T antigen. In embodiments, the cells are human 293 cells adapted for growth in suspension cultures. In embodiments, the cells are human 293-F cells, which are commercially available from a variety of vendors.

In certain approaches, such as therapeutic approaches, the present disclosure includes modifying heterologous, or cells obtained from an individual, to express one or more of the glycoproteins described herein. Thus, in embodiments, human or non-human cells can be modified to, for example, correct a defect in a mucin or mucin-like protein, or the production thereof. In embodiments, cells modified according to this disclosure are totipotent, pluripotent, oligopotent stem, or multipotent stem cells. In embodiments, the cells are hematopoetic cells. In embodiments, the cells are chondrocytes. In embodiments, the cells are mesenchymal stem cells or marrow stromal cells. In embodiments, the cells are synovial cells. In embodiments, the cells are chondrogenic precursor cells. In embodiments, the cells

endogenously produce cartilage-specific gene products, such as type II collagen and/or cartilage-specific chondroitin sulfate proteoglycan (CSPG). In embodiments, the cells are epithelial cells, or precursors thereof, or are goblet cells. In embodiments, the cells are immune cells, and include but are not necessarily limited to T cells, such as CD4+ and CD8+ T cells, and dendritic cells. Cells can be modified according to any established technique, including but not limited to use of viral expression vectors, or by chromosome editing, such as by any suitable CRISPR-based gene editing approach. Modified cells can be administered to an individual in need thereof. In embodiments, transgenic non-human animals that have been created to express one or more of the modified proteins of this disclosure can be produced and used to study a wide range of biological functions, disorders and conditions.

In embodiments, any glycoprotein described herein can be present in a fusion protein. Fusion proteins are produced recombinantly and contain in a single, contiguous polypeptide, segments of distinct proteins. In embodiments, a fusion protein described herein comprises a glycoprotein or segment thereof, and a second protein or segment that is not particularly limited. In embodiments, the second protein produced a detectable signal, and thus includes, for example, fluorescent proteins.

In certain embodiments, the compositions and methods of this disclosure involve recombinantly produced proteins that have repeated amino acid sequences, such as tandem repeat sequences. In embodiments, the tandem repeat sequences are modified relative to their naturally occurring sequences, and the number of repeats may have been altered, relative to the number of repeats in a naturally occurring protein. Combinations of distinct repeats may be included in the polypeptides described herein.

In embodiments, the disclosure comprises introducing an expression vector described herein that encodes one or more proteins described herein, which may be a codon-optimized expression vector, into a suitable cell/cell culture, allowing expression of the protein(s), and recovering the protein(s) from the cells. In embodiments, cells in a cell culture are modified to express at least protein described herein using any suitable expression vector.

The expression vector may be integrated into a chromosome of the cells, or may be maintained permanently or transiently as an epigenetic element. The expression vector may be configured to express the protein(s) in a constituent or inducible manner. In one non-limiting embodiment, a transposon based expression vector can be used, or a lentiviral expression system can be used. In a non-limiting embodiment, a lentiviral system can be excluded as a tool to express the proteins described herein. In embodiments, any protein described herein may, or may not include, a signal sequence. In embodiments, a

polynucleotide, such as a cDNA encoding one or more of the proteins described herein, is randomly integrated into one or more chromosomes to produce the modified cells. In embodiments, a randomized transposition of a cDNA into the genome is used.

In embodiments, codon-optimized expression vectors comprise a threshold number of altered codons, wherein the altered codons do not change the amino acid encoded by the particular codons. Thus, optimized codons may contain, for example, changes in wobble bases. In embodiments, at least one codon is altered, and from one codon to all of the codons that encode each amino acid in the particular protein may be altered. In embodiments, the codon optimized cDNAs reduce cDNA sequence repetitiveness to improve stability of the nucleotide sequence during DNA processing, including but not necessarily limited to slippage during replication, transcription, reverse transcription and other nucleotide processing operations on repetitive nucleotide sequences which often result in deletions or amplifications of cDNAs and mRNAs. In embodiments, codons with less than a predetermined threshold of frequency of usage in the pertinent cell type are replaced with codons that have a higher frequency of usage. For example, in one embodiment codons that have less than or equal to 10% usage frequency in human cells can be replaced.

In embodiments, the mucin/lubricin protein, or a protein for which improved production may be desired, can be modified for recovery using any suitable approach, including but not limited to including one or more purification tags, including but not limited to a His-tag. In an embodiment, a His-tag is a linear sequence of *n* histidine residues where *n* is typically 6-10. His-tags achieve purification by binding specifically to nickel or cobalt ions, which may be for example, attached to a substrate, such as any suitable beads. The His-tag, or any other suitable purification tag, may be placed at the N-terminus of the protein, at the C-terminus of the protein, or interior to the protein. In embodiments, a FLAG-tag, or FLAG octapeptide, or FLAG epitope, is may be included in proteins of this disclosure. Suitable FLAG sequences are known in the art. In embodiments, a Small ubiquitin-related modifier (SUMO) tag, such as a His-SUMO tag can be included. In embodiments, protease cleavage sites can be included, such as for protein identification, separation, purification, etc. The proteins can be purified to any desired degree of purity.

In non-limiting embodiments, the tandem repeats that are included in proteins of this disclosure comprise any one or any combination of the following amino acid segments: KEPAPTTP (SEQ ID NO:1), KEPAPTP (SEQ ID NO:9) and KEPAPTTTP (SEQ ID NO:10), or a combination thereof.

In embodiments, from 2-120 repeats are included in a protein of this disclosure. In non-limiting embodiments 10, 21, 40, 42, 59 or 80 repeats are included. The repeated sequences may be fully contiguous within the polypeptide. In embodiments, any amino acid sequence described herein can be a segment of a longer tandem repeat, and thus may have additional amino acid sequences on its N- or C-terminus. In embodiments, the amino acid sequence of a tandem repeat described herein comprises or consists of from 7-80 amino acids. In embodiments, a tandem repeat described herein exhibits an estimated length of approximately 135 nm, or 270 nm. In embodiments, the repeats are perfect repeats, meaning the identical sequence is repeated in the protein, which differs from certain tandem repeats that occur naturally.

In embodiments, the disclosure includes all cDNA and amino acid sequences disclosed in Parts I-IV of the Examples, and variants thereof as described herein. From time to time, such representative sequences are referred to for convenience as "biobricks." In non-limiting embodiments, the disclosure provides polypeptides, such as glycoproteins, and codon-optimized expression vectors encoding the glycoproteins, that are described herein as SynMuc1 and SynLubricin, Syn1_40, Syn1_80, Syn2_40, Syn2_80, Syn3_40, and other constructs for use with non-human mammals as described further below.

Polypeptides comprising amino acid sequences that are at least 90% identical to the amino acid sequence of these sequences are included. In embodiments, the proteins comprise mutations, relative to an endogenous protein. An "endogenous" protein is a protein that is normally encoded by an unmodified gene. Likewise, an endogenous gene or other polynucleotide comprises a DNA sequence that is unmodified, such as by recombinant, gene editing, or other approaches. Mutations, as further described below, can include amino acid insertions, deletions, and changes, and may also include additional repeated sequences, or fewer repeated sequences, relative to an endogenous sequence.

In embodiments, tandem repeat amino acid sequences are introduced into a glycoprotein at its N-terminus, its C-terminus, or both the N-terminus and C-terminus. In one illustrative embodiment, a lubricin-like molecule is produced via fusion of the native N- and C-terminus of human lubricin or lubricin from a non-human mammal, with repeats of KEPAPTTP (SEQ ID NO:1), KEPAPTP (SEQ ID NO:9) and KEPAPTTTP (SEQ ID NO:10). In embodiments, the non-human mammal is a canine or equine or feline animal. Representative amino acid sequences of lubricins from equines and canines that are incorporated into modified lubricins are described further below. In embodiments, from 10-120 repeats are included. In embodiments, 59 repeats are included. In embodiments, repeat

sequences or other sequences can be separated by one another by sequences, or by linker sequences, such as from one-three amino acids.

In embodiments, recombinant lubricin polypeptide is provided, wherein a contiguously repeated sequence described below is located between an N-terminal amino acid sequence and a C-terminal amino segment that has at least 90% sequence identity to a human, canine, or equine sequence. Thus, the disclosure includes the described tandem repeats that include flanking sequences. The flanking sequences can comprise a human lubricin N-terminal and C-terminal derived amino acid sequence; a canine lubricin N-terminal and C-terminal derived amino acid sequence; or an equine lubricin N-terminal and C-terminal derived amino acid sequence.

In one non-limiting embodiments, wherein the contiguously repeated sequence is located between an N-terminal human lubricin derived sequence that has at least 90% sequence identity to the human lubricin sequence:

QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDFKRVCTAELSCKGRCFESFER
GRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAPPPSGASQTIKSTTKRSPK
PPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSTIRKIKSSKNSAANRELQKKLK
VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAK
PINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEK
TSAKDLAPTSKVLAKPTPKAETTTKGPALTTP (SEQ ID NO:75)

and a C-terminal human lubricin amino derived amino acid sequence that has at least 90% sequence identity to the human lubricin sequence:

SEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTT
AKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKI
TTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKP
TSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVE
VNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNIC
NGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFTRCNC
EGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYF
FKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGETTQVRRRRFERAIGPSQTHTIRIQ
YSPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYY
NIDVPSRTARAITTRSGQTLSKVWYNCP SEQ ID NO:76)

In embodiments, the lubricin sequences that flank the contiguous repeats in non-human animals can include amino acid changes that, in non-limiting embodiments, are changes of from 3-7 amino acids, relative to the native sequences at the N or C termini of the following sequences.

In an embodiments, the contiguously repeated sequence is located between an N-terminal canine derived lubricin sequence that has at least 90% sequence identity to the following sequences.

QDLPSCAGRCGEGYSRDAICNCDYNCQHYMECCPDFKKACTVELSCKGRCFESFAR
GRECDCDSDCKKYGKCCPDYEDFCGRVHNPTSPPSSKTAPPSPGASQTIKSTAKRSPK
APNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSTIRKIKSSKNSAANKELKKKPKV
KDNKKERTPKKKPPPEPPVVDEAGSGLDNGDIKLTPTPDIPTTQRNKVTTSPKFTTGK
PINPKPSLPPNTDTSKETSSTPNKETTVKSKETLANKETSSKAKEKITSAKETRSAEKTP
AKDFVPTTKAPVKSTPKAESTTKGPALTTP (SEQ ID NO:77) (wherein for example, the seven C-terminal amino acids may be changed from the native canine sequence, which is SPAPTTP (SEQ ID NO:83);

and a C-terminal canine lubricin derived amino acid sequence that has at least 90% sequence identity to the canine lubricin sequence:

SEVTTTAKDKTTEKDIIPEITTAVPKITTQETATPTEETTTESKTSTTTQVTSTTSSKNTP
KATTLAPKVMTATQKTTTTEETMNKPEETTAVPKDTATSTKVSTPRPRKPTKAPKKP
ASTKKPNTIPKRKKPKTTPTPPKMTTSTMPKLHPTSSVEAMLQTTTSPNQRPNSEIVE
VNPNEDTDAAGKKPHMFPRPPVLTPIFIPGTDILVRGSNQDIAINPMLSDETNLCNGKP
VDGLTTLRNGTMVAFRGHYFWMLSPSKPPSPPRKITEVWGIPSPIDTVFTRCNCEGKT
FFFKGSQYWRFTNDIKDAGYPKQIVKGFGGLNGRIVAALSIAKYKDRPESVYFFKRG
GSVQQYTYKQEPIKKCTGRRPAINYPVYGETTQVRRRRFERAIGPSQTHTIRIHYSPIR
VSYQDKGFLHNEVKMSSQWRGFPNVVTSAIALPNIRKPDGYDYYAFSRNQYYNIDV
PSRTARVVTTRFGRTLSNIWYNC (SEQ ID NO:78) (wherein, for example, the three N-terminal amino acids are changed, relative to the corresponding canine sequence, which is PEM).

In embodiments, the contiguously repeated sequence is located between an N-terminal equine derived lubricin sequence that has at least 90% sequence identity to the equine lubricin sequence:

QDLSSCAGRCGEGYSRDATCNCDFNCQYYMECCPDFKKVCTSELSCKGRCFESFER
GRECDCDADCKKYGKCCSDYESFCEEVHNPTSPPSSKTAPPPPGASQTIKSTAKRSPK
SNKKKTKKVIESEEIIEEHSVSENQESSSSSSSSSSTIRKVKSSKNSAANRELKKKPKVK
DSKKKRTPKKKPTPEPPVIDEAGSGLDNGDFMLIPTPKIPTTQRNKVTTSPKITTVKPI
NPKPSLPPNSDTSKETTSTPNKETTVETKETEITNKETSTSANEKTTSARKSTEKTSDK
DFAPASEVPAKSTPKAETTTKGPALTTP (SEQ ID NO:79), (wherein, for example, the seven C-terminal amino acids may differ from the native equine sequence, which is SPSLTT (SEQ ID NO:84));

and a C-terminal equine lubricin derived amino acid sequence that has at least 90% sequence identity to the equine lubricin sequence:

SEVSTTTTTMKPPTTPKNLAESTPEFPAEPTPKALENSPKEPAVPTTKAPEVTKPEVTT
TAKDKVTGKDIHTIPEITTAAPKITTETATTTEEKTTESKVTSTIMQVTSTTEDTTTSSK
ITPKATTLAPKVMTATKTTTTQETINKLEETTAIPKDTATHSKVTTPKPKKPTKAPRKP
TSTKKPKTPRKRKPKTTPIPPKITTPTTPKSNPTTLAEAMLQTTTSPNQTPNSAMIEVNP
KNEDADAAEGEKPLVILRPHVLTPIVIPGPDFLVRGPNLGIGINPMLSDETNLCNGKPV
DGLTTLRNGTLVAFRGHYFWMLRPFSPPSPPRRITEVWGIPSPIDTVFTRCNCEGKTFF
FKDSQYWRFTNDIKDAGYPKLISKGFGGLSGKIVAALSIATYKNRPESVYFFKRGGRI
QQYIYKQEPIRKCPGRRPAIHYSVYGEAPQIRRRRFERAIGPSQTHTIRIHYSPVRVSYQ
DKVPSTDFLHNEVKVSTLWRGLPDTVTSAISLPNLRKPDGYDYYAFSKDQYYNIDVP
SRTARAITTRSGQTLSKVWYNCP (SEQ ID NO:80) (wherein, for example, the three N-terminal amino acids may differ from the native equine sequence, which is SEA).

In embodiments, a recombinantly produced protein described herein comprises variants that have tandem repeats of any one or combination of the tandem repeat sequences described herein, wherein the variants comprise modifications of such sequences. Expression vectors encoding the variants are included. In embodiments, the modifications comprise amino acid segments that have between 90.0-99.9% amino acid identity, inclusive, and including all ranges of numbers there between to the first decimal point, with contiguous amino acid and polynucleotide sequences expressly described herein. In embodiments, tandem repeats comprised by recombinantly produced proteins of this disclosure have 90, 95, 97, 98, 99 or 99.5% amino acid sequence identity to the amino acid sequences described

herein, across their full length(s). A recombinant protein is a protein expressed from a polynucleotide that has been introduced to a cell that did not comprise a coding sequence for that protein prior to introducing the polynucleotide. The same applies to recombinant cDNA sequences.

As is known in the art, to determine the percent identity of two nucleotide or amino acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps may be introduced). The nucleotides or amino acids at corresponding nucleotide or amino acid positions are then compared. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity= # of identical positions/total # of positions x 100).

In certain embodiments, the tandem repeat variants described herein comprise a change of 1, 2, 3, 4, or 5 amino acids. In embodiments, an amino acid can be deleted, added, or changed. In embodiments, an amino acid that is changed is a serine, a threonine, or a combination of serine and threonine residues are changed. In embodiments, about 1-50% of serine and/or threonine residues are changed. In embodiments, a serine or threonine residue present in a native protein sequence is changed to an alanine, or to another amino acid. In embodiments, a protein of this disclosure comprises fewer, or no amino acids that are present in a native (non-modified and/or endogenous protein). In embodiments, a native protein comprises one or any combination of asparagine, aspartic acid, glycine, isoleucine, leucine, and/or serine, which can be engineered recombinantly out of representative proteins of this disclosure.

In embodiments, amino acid changes introduced into proteins of this disclosure result in changed glycosylation patterns. Thus, in embodiments, the disclosure provides for production of recombinant proteins with controllable glycosylation patterns. In embodiments, the number of O-linked oligosaccharides present on a protein of this disclosure is modified. In embodiments, the glycosylation pattern is changed relative to a control, such as a protein in which a corresponding glycosylation site is not changed. In embodiments, one or more properties of the proteins, and or cells that express the proteins, is changed. In embodiments, the stoichiometry of oligosaccharides to protein/amino acids is changed in, for example, a glycoprotein of this disclosure. In embodiments, a protein of this disclosure comprises a percentage by weight of glycosidic residues that is different from a suitable control. In embodiments, a protein of this disclosure exhibits a lubrication parameter, such as a dynamic

coefficient of friction. In embodiments, a coefficient of friction can be determined using any suitable approach, such as cartilage on cartilage friction test. In embodiments, a protein of this disclosure exhibits a lubrication parameter that is different from a suitable control.

In embodiments, a recombinantly produced protein as described herein comprises a change relative to a control in the Core 1 O-glycan structure, Galβ1-3GalNac, and/or the amount of Core 1 derivatives of Galβ1-3GalNAc, and/or the amount of terminally substituted sialic acids therein, or a change in GalNAc (N-acetylgalactosamine) monosaccharide glycosylation. In embodiments, a protein described herein can comprise the Core 2 O-glycan, GlcNAcβ1-6(Galβ1-3)GalNAc and/or the Core 2 derivatives of GlcNAcβ1-6(Galβ1-3)GalNAc, which comprise at least 5 percent of all Core 1, Core 2, Core 3, Core 4, Core 5, Core 6, Core 7, and Core 8 O-glycan structures. In embodiments, such a protein is produced by human cells that are cultured as further described herein.

In embodiments, proteins of this disclosure may be in the form of monomers, dimers, multimers, and combinations thereof. In embodiments, monomer/dimer ratios, proportions, and/or concentrations are changed, relative to suitable controls.

In embodiments, segments of proteins described herein can be separated by any suitable linking amino acids. In embodiments, linker can comprise from 1-20 amino acids, inclusive, and including all integers and ranges of integers there between. In general, linkers are comprised of a glycine, serine, or serine and glycine. In embodiments, linking amino acids do not intervene tandem repeats. In embodiments, secreted forms of glycosylation mutants are provided.

In embodiments, a modified lubricin lacks one or both of a cytoplasmic domain and a transmembrane domain. In embodiments, lubricins of this disclosure comprise a secretion signal, such as for use in producing the modified protein(s). The amino acid sequences of many suitable secretion signals are known in the art and can be used in embodiments of this. In one embodiment, a human secretory sequence comprises or consists of MAWKTLPIYLLLLLSVFVIQQVSS (SEQ ID NO:72). In one embodiment, a canine secretion signal comprises or consists of MQWKILPIYLLLLSVFLIQQVS (SEQ ID NO:73). In one embodiment, an equine secretion signal comprises or consists of MEWKILPIYLLLLLSIFSIQEVSS (SEQ ID NO:74), or another sequence as further described herein, including changes to the N- and C-terminal amino acids. In embodiments, a native secretory signal is replaced with a segment of an immunoglobulin, such as an IgG kappa light chain sequence from a human or a mouse or another mammal. In embodiments,

the secretory sequence comprises a secretory sequence from any of: IL-2, CD33, Human IgG2 H, Chymotrypsinogen, trypsinogen, Gaussia luc, Influenza Haemagglutinin, Human insulin, or Silkworm Fibroin.

In embodiments, a polypeptide of this disclosure may have one or more modified amino acids that are, for example, conjugated to another moiety. In embodiments, a polypeptide of this disclosure is conjugated to at least one azido group such that they can be readily conjugated to other moieties, such as using click chemistry, such as by modifying an O-glycan with an azide. In embodiments, a polypeptide of this disclosure is cyclized, or stapled.

In embodiments, a tandem repeat sequence described herein is incorporated into any glycoprotein. In embodiments, the glycoprotein is any mucin or lubricin protein. In embodiments, the glycoprotein is Proteoglycan 4, also referred to in the art as lubricin, which comprises a protein that in humans is encoded by the *PRG4* gene. In non-limiting the disclosure provides a modified mucin termed SynMuc1, as described further below. In another non-limiting embodiment, a modified lubricin is provided as SynLubricin, as further described below.

In an embodiment, production of protein is increased using cells modified herein, wherein the cells are present in a cell culture container, including but not limited to any cell culture dish, and bioreactors. In embodiments, modified cells according to this disclosure are used in bioreactors to produce any desired protein, or combination thereof. In non-limiting embodiments, the bioreactor comprises a suspended cell bioreactor. In embodiments, bioreactors have a volume of from 1-25,000 liters, inclusive, and including all numbers and ranges of numbers there between.

In embodiments, cDNA libraries are provided. In embodiments, the disclosure comprises providing a cDNA library as described herein, and selecting one or a combination of the cDNAs described or modifying cells by introducing the cDNA and/or an expression vector encoding the cDNA into a cell. Selection can be based upon an intended or actual use for the cells, such as for use in protein production, based on any particular protein and cell expression system. Kits encoding the proteins are also included.

In embodiments, one or more proteins described herein can be combined with other agent(s), such as biodegradable polymer(s), nanoparticles, pectin, alginate, cross-linked derivatives of poly(acrylic acid), polyvinyl alcohol, polyvinyl pyrollidone, polysaccharides, hydroxypropyl methylcellulose, carboxymethylcellulose, lectins, rheology modifiers, plasticizers, chondroitin, glucosamine, and/or any hyaluronic acid.

For use in prophylaxis and/or therapy of diseases wherein, for example, anti-adhesive agents may be of benefit, compositions described herein can be administered in a conventional dosage form prepared by mixing with a standard pharmaceutically acceptable carrier according to known techniques. Some examples of pharmaceutically acceptable carriers can be found in: *Remington: The Science and Practice of Pharmacy* (2005) 21st Edition, Philadelphia, PA. Lippincott Williams & Wilkins, the disclosure of which is incorporated herein by reference. In embodiments, pharmaceutical and other compositions comprising the proteins described herein can be provided as liquids, tablets, powders, sprays, ointments, hydrogels, and aerosols.

In embodiments, pharmaceutical compositions comprising one or more proteins of this disclosure can be administered to an individual using any suitable route, including but not necessarily limited to topically, orally and parenterally, and as further described below. For example, the proteins can be administered intravenously, by direct injection into synovial joints or other synovial structures (tendon sheaths, bursae), intraperitoneally, by direct injection into the pericardial sac, by direct injection into the pleural cavity, subdermally, subcutaneously, or by direct application to skin, mucous membranes, or the eye.

In embodiments, the disclosure includes administering an effective amount of one or more of the polypeptides described herein, and/or a composition comprising such polypeptides. An effective amount can vary depending on pharmaceutical formulation methods, administration methods, the patient's age, body weight, sex, administration time, administration route, and other factors that will be apparent to those skilled in the art. Compositions can be administered once, or over a series of administrations. In embodiments, the disclosure includes a single dose, or several doses. In non-limiting examples, for use in prophylaxis and/or therapy of a join or similar structure, a suitable concentration of a polypeptide ranges from 250ug/mL to 2mg/mL, inclusive, and including all ranges of numbers there between, and all ranges of milligrams, micrograms. In embodiments, approximately 1 to 3 mL is used for mammalian joint applications. Dosing frequency can be adjusted on an individual basis. Given the unexpected half-life polypeptides of this disclosure, a suitable dosing frequency is once every 3-6 weeks.

In embodiments, the disclosure comprises methods, compositions, and devices for treating an ocular disease, disorder or condition in a mammal. In embodiments, proteins produced by cells as described herein are used for treatment of eye disease or condition using any method or device known to those of ordinary skill in the art. In embodiments, compositions comprising the proteins are used for intracameral, intravitreous,

subconjunctival, sub-Tenon's, subretinal, or topical application to the corneal surface. The proteins may be delivered directly to the eye (for example: topical ocular drops or ointments; slow release devices in the cul-de-sac or implanted adjacent to the sclera or within the eye) using techniques well known by those skilled in the art. It is further contemplated that the proteins described herein may be formulated in intraocular insert or implant devices.

In embodiments, a pharmaceutical comprising one or more proteins described herein is used to treat an eye disorder that comprises one or more diseases or injury to the retina, including age-related macular degeneration (AMD), retinitis pigmentosa (RP), and diabetic retinopathy (DR). In an embodiment, the individual has dry, atrophic (nonexudative) age-related macular degeneration, defined as progressive age-related degeneration of the macular associated with retinal pigment epithelial changes including atrophy and drusen, which is a common cause of vision loss in adults for which therapy is limited. In embodiments, the disorder comprises one or more diseases or injury to the cornea. In embodiments, the individual has glaucoma, which may include primary, secondary and/or congenital glaucoma.

In embodiments, proteins of this disclosure can be provided in the form of eye drops. In embodiments, the eye drops comprise any one or more of steroids, antihistamines, sympathomimetics, beta receptor blockers, parasympathomimetics, parasympatholytics, prostaglandins, nonsteroidal anti-inflammatory drugs (NSAIDs), antibiotics, antifungal, or topical anesthetics. In certain embodiments, the eye drops are for use with any dry eye condition. In embodiments, the eye drops are for use in lubrication of eyes, including but not necessarily for a contact lens wearer. In embodiments, the compositions are provided as lubricating eye drops. In embodiments, the lubricating eye drops comprise artificial tears. In embodiments, the eye drops may be free of medications, and thus function only as lubricating/tear-replacement compositions. In other embodiments, the eye drops may be for treatment of ocular allergic reactions, and thus my also comprise antihistamines, and/or vasoconstriction agents. In embodiments, an eye drop formulation comprises 250ug/mL to 2mg/mL, inclusive, and including all ranges of numbers there between, and all ranges of milligrams, micrograms. Such concentrations can be used in typical eye drop volumes, such as 1-2 drops/eye at approximately 0.05 to 0.01 mL per eye.

In embodiments, compositions comprising proteins described herein can be used in conjunction with contact lenses. In embodiments, the proteins are used in a contact lens solution. Thus, proteins described herein can be mixed with any suitable contact lens solution components, which include but are not necessarily limited to saline, mild abrasives, surfactants, anti-fungal and anti-bacterial agents, which include but are not limited to

conventional amicrobial agents, or hydrogen peroxide or boric acid, and preservatives, such as ascorbic acid or edetate disodium. Contact lenses provided in a solution comprising one or more proteins described herein are included within the scope of this disclosure.

In embodiments, compositions comprising proteins described here can be directed to a mucosal lining. The mucosal lining, includes, for example, the upper and lower respiratory tract, eye, buccal cavity, nose, rectum, vagina, urogenital tract, periodontal pocket, intestines and colon. In certain embodiments, the compositions can be used for oral inhalations. In embodiments, the oral inhalation comprises nasal applications, and thus may include nasal sprays, nasal drops, and nasal ointments. In embodiments, oral inhalation may comprise bronchial sprays and inhalers. In embodiments, the proteins may be used to access mucosa through use of throat lozenges, chewing gum, mouthwashes or gargles, suppositories, or tampons.

In embodiments, compositions comprising proteins described herein are used as surgical anti-adhesives (intraperitoneal lubricants to lubricate viscera and prevent post-op intestinal and visceral adhesions during intra-abdominal surgical procedures/manipulations; intrapleural lubricants to lubricate lungs and prevent post-operative pleural adhesions during intra-thoracic surgical procedures/manipulations; intrapericardial lubricants to lubricate the cardiac surface and prevent post-op pericardial adhesions during cardiac surgical procedures/manipulations). As a post-operative synovial fluid replacement following any arthroscopic, tenoscopic, or bursoscopic procedure to maintin lubrication and prevent adhesions or pannus formation. In embodiments, the compositions are used for treating joint sepsis/infection in any mammal described herein. In certain embodiments, the compositions can be used in conjunction with wound healing, treatment of wound infection, and treatment of generalized sepsis.

In embodiments, a non-human mammal to which a composition comprising a modified lubricin is described herein is in need of any one or combination of disorders described herein. Further, equines may be in need of treatment for one or a combination of disorders to which equines are particularly susceptible. In one non-limiting embodiment, the equine animal is in need of treatment for osteochondritis dissecans (OCD). Other common equine embodiments, which may be extended to canines and felines, include treating intra-articular fracture, osteochondral fragmentation, meniscal injury, cartilage injury, synovitis, joint sepsis and post-traumatic osteoarthritis (PTOA). Other embodiments include treating tendon and ligament injuries, including but not limited to: superficial digital flexor and deep digital flexor tendonitis/tendinopathy, suspensory ligament desmitis/desmopathy,

tenosynovitis and navicular bursitis. Equine ophthalmic embodiments include: corneal ulcer, descemetocele, and fungal keratitis/keratopathy. In non-limiting embodiments, a canine animal is in need of treatment for rupture of the cranial cruciate ligament (RCCL)—analogous to anterior cruciate ligament injury in humans, elbow dysplasia, hip dysplasia, tendonitis/desmitis, and ophthalmic applications, including keratoconjunctivitis sicca (KCS), immune-mediated keratopathy, and indolent ulcer.

In embodiments, compositions comprising modified proteins such as modified lubricins may be administered to humans and non-human animals for therapeutic or prophylactic purposes. In embodiments, modified lubricins are administered to a canine, feline or an equine animal to prevent or limit the severity of injuries that are prone to arise in athletic competitions or during animal working. For example, the compositions can be administered to equines to prevent or limit joint/cartilage damage during equestrian events, or during work, such as in police work or ranching. Typical equestrian events include rodeo, dressage, show jumping, vaulting, polo, horse racing, and many others that will be apparent to those skilled in the art where the risk of joint and related injuries is high. Further, it is considered that the compositions of this disclosure will be useful for treating a variety of other non-human mammals, such as in veterinary hospitals and clinics, animal rescue facilities, and zoos. In embodiments, a composition of this disclosure is used for prophylaxis and/or therapy of an avian animal.

In embodiments, an article of manufacture may be coated and/or impregnated with a composition comprising any of the proteins described herein. In embodiments, the article of manufacture is coated on any porous or non-porous surface. In embodiments, the article comprises a medical device, including but not necessarily limited to a surgical device, a dental or orthopedic device, sutures, catheters, an intubation device, an anesthesia delivery device, a dressing, bandage, etc. In embodiments, proteins described herein are used to coat cell culture devices, including, but not necessarily limited to, cell culture plates, multiwell plates, bioreactors, and any other surface, wherein an anti-adhesive property is desirable.

In another aspect the disclosure includes a supplement product, such as a nutraceutical product, a dietary supplement, a food ingredient, etc., The supplement product can be provided in the form of, for example, a liquid, capsules, tablets, softgels, powders, and the like.

In embodiments, a pharmaceutical and/or nutraceutical product comprising one or more proteins described herein is provided in a container, such as any suitable closed or sealable container which may be sterile. In embodiments, the product comprises printed

material. The printed material can be provided as a product insert, label, or as a component of packaging. The printed material provides an indication that composition comprising the polypeptides is to be used for treating any disease, disorder, or condition as described herein, or for producing an anti-adhesive effect for any purpose. In one embodiment, polypeptides described herein are used as a supplement for treating a condition of joints, including, but not necessarily limited to joint pain, arthritis, including, but not necessarily limited to, osteoarthritis, rheumatoid arthritis, injuries to joints, menisci or cartilage, such as sports injuries, or in conjunction with joint/ligament repair surgeries. Thus, administering compositions described herein for the purposes of improving the health or well-being of an individual, are included within the disclosure. In embodiments, compositions of this disclosure can be injected directly into a joint and/or synovial fluid. In embodiments, the composition is administered directly or indirectly to any synovial structure, including but not limited to a synovial joint, and tendon sheath, or bursa. In embodiments, compositions of this disclosure can be also be used for injection directly into the tendon, tendon sheath, ligament or bursa following a tendon, ligament or bursal injury, trauma or infection. In embodiments, the compositions can be in contact with a mesothelial surface: e.g., the composition is administered to that it contacts a surface abdominally, or pericardially, for prophylaxis and or therapy of a disorder associated with one or more of such surfaces.

The disclosure may be better understood by reference to the following non-limiting Examples, which are provided as exemplary of the disclosure, divided into four Parts. The following examples are presented in order to more fully illustrate the embodiments of the disclosure and should in no way be construed, however, as limiting the broad scope of the disclosure. The reference listings of this disclosure is not an indication that any particular cited reference is material to patentability.

EXAMPLES

Part I

This Part I of the disclosure provides non-limiting and representative examples of sequence-specific mucins with controllable glycosylation patterns, and data and discussion of the same.

In particular, this Part I relates to the understanding that, prior to the present disclosure, few design guidelines existed for encoding customized mucin glycoproteins with tunable glycosylation patterns. Part I accordingly provides 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 new, rationally designed domains for synthetic mucins. Of the more than 400 possible cDNA combinations, this Part I provides a library of over 50 mucins, each with unique chemical, structural, and optical properties. The library is applied to develop general guidelines for the design and engineering of mucins, which form a part of this disclosure. Surprisingly, it was discovered that the extension of the immature α-GalNAc Tn-antigen to Core 1 and Core 2 glycan structures strongly depends on the frequency of O-glycosylation sites along the mucin backbone. As will be apparent to those skilled in the art from this disclosure, sialyation of glycan structures is readily tuned through recycling motifs on the mucin cytoplasmic tail. It is also demonstrated that the overall length of the mucin polypeptide backbone can have unexpected effects on glycosylation. Without intending to be bound by any particular theory, it is expected that that the mucin parts inventory presented here, along with the described design guidelines for making new mucins, can be broadly applied for glycocalyx research and mucin-based biotechnologies.

Introduction to Part I

Cell-surface mucins are a family of membrane-anchored biopolymers that are defined by their unstructured polypeptide backbone with a high density of sugar side chains(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). 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 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.

One of the major barriers to progress in developing such understanding has been the lack of tools for precise editing of the molecular structure of mucins. Genetic approaches that target glycosyltransferases can be highly effective in altering mucin glycosylation(10), but these approaches typically affect broad classes of glycoproteins, making any observed effects on cell behavior difficult or impossible to pinpoint to a particular mucin. To overcome the limitations of genetic approaches, libraries of bio-mimetic mucin polymers with plasma membrane anchors have been developed for glycocalyx editing(6, 11). While highly successful in unraveling some mechanistic details of mucin function, synthetic polymers are typically cleared from the cell-surface in hours to days and must be continuously replenished through media supplementation(12, 13). Thus, investigation of behaviors over longer time durations, particularly *in vivo*, are largely inaccessible with synthetic mucin mimetics.

Prior to the present disclosure, strategies for mucin engineering and glycocalyx editing that combines the important features of the synthetic chemical approach – defined backbone chemistry, tailored glycan structures, and precision glycan placement – with the power and long-term stability of genomic encoding had yet to be developed. Advances in custom gene synthesis support development of cDNA sequences to be constructed at unprecedented speed and low cost. However, custom gene synthesis is not 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(14, 15).

As described in this Part I, a 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(16, 17).

In this Part I, 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. The presently described combinatorial library of mucin parts enables facile

construction of mucin biopolymers with tunable sizes, side-chain spacing, and glycan types for glycocalyx editing.

Part I - Results

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

Part I results demonstrate 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 (SEQ ID NO:8) and KEPAPTTP (SEQ ID NO:1) have similarity to native Muc1 and Proteoglycan 4 (Lubricin), respectively. Three repeats were designed based on statistical analysis of mucin O-glycosylation sites (PPASTSAPG) (SEQ ID NO:4) or analysis of O-GalNAc transfer efficiency (DAATPAP (SEQ ID NO:2) 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 6). 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. As expected based on previous reports[27], the larger mucins formed a glycocalyx that was substantial enough to dislodge epithelial cells from their substrate.

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 8). 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 9c). 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 10). 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 (SEQ ID NO:2), DAATPAPP (SEQ ID NO:3), and PPASTSAPG and KEPAPTTP (SEQ ID NO:1) which have similarity to 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 DAATPAP (SEQ ID NO:2) and DAATPAPP (SEQ ID NO:3) 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 10b).

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 (SEQ ID NO:2) and DAATPAPP (SEQ ID NO:3) 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 occurs 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 (SEQ ID NO:11) 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).

Materials and Methods

Antibodies and Reagents

The following antibodies were used: anti-Human MUC1 (CD227) (clone HMPV; 555925, BD Biosciences), mouse anti-□-Actin (clone C4; 47778, Santa Cruz), chicken anti-SUMO/SUMOstar (AB7002, LifeSensors), mouse 6xHis (552565, BD Biosciences), mouse anti-□-tubulin (clone B-7; 5286, Santa Cruz), mouse anti-GFP (clone 4B10; 2955, Cell Signaling Technology), m-IgG□ 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' (SEQ ID NO:12) (forward) and 5'-GGCAGCTGAGCCCTTATACAGCTCGTCCATGCCGTGAGT-3' (reverse) (SEQ ID NO:13).. 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 (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) (SEQ ID NO:14) and 5'-CCACCGCCGACCGAGGTGACATCCTG-3' (reverse) (SEQ ID NO:15) primers.

The cDNA with recycling motif CQCRRK (SEQ ID NO:11) pcDNA3.1 Muc1_10 TM21 CQC was generated from pcDNA3.1 Muc1_10 TM21 through Q5 site-directed mutagenesis with 5'- CCGAAAGTAGGAATTCGGGCCCGTTTAAACCCGC-3' (forward) (SEQ ID NO:16) and 5'- CGGCACTGACATCTAGAGTACCACAACAAAGCCAGGC-3' (reverse) (SEQ ID NO:17) 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 (SEQ ID NO:2) and DAATPAPP (SEQ ID NO:3) 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 °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 proteins (600ug, each) was denatured by heating at 100 °C for 5 min. The denatured proteins were 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[54]. The samples were cooled, neutralized with 10 % acetic acid, passed through a Dowex H+ resin column, and lyophilized with borates removed under the stream of nitrogen. The glycans were permethylated for structural characterization by mass spectrometry using previously reported methods[55]. Briefly, the dried eluate was dissolved with dimethyl sulfoxide (DMSO) and methylated by using methyl iodide and NaOH-DMSO base (prepared by mixing DMSO and 50 % w/w NaOH solution). The reaction was quenched with water and the reaction mixture was extracted with methylene chloride and dried. The permethylated glycans were dissolved in

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

Cellular O-Glycome Reporter/Amplification (CORA)

All chemicals were purchased from Millipore Sigma except where noted. Solvents were of HPLC grade or higher, and 0.1% (v/v) trifluoroacetic acid was included in all chromatography steps. Benzyl 2-acetamido-2-deoxy-α-D-galactopyranoside (BnGalNAc) was peracetylated by heating in a molar excess of 33% (v/v) acetic anhydride in anhydrous pyridine for 1 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).

Discussion of Part I

The O-glycosylation of mucins determines their physical and biochemical characteristics, and, thus, their biological functions. This Part I provides a genetically encoded system to edit the mucin biopolymers, and can be used as a tool for glycocalyx engineering, among other significant utilities that are discussed above. 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]. In this Part I we modify signals and motifs in the mucin backbone sequences and cytoplasmic tails to encode mucins with varying physical features, backbone chemistries, and glycosylation patterns.

Using codon degeneracy to design mucin cDNAs with minimal repetition, we were able to apply custom gene synthesis for construction of 13 representative unique mucin repeats, each of which could be readily combined with other functional domains for cell-surface anchorage and control of trafficking. All repeat sequences tested were successfully fabricated with no failures. The disclosure therefore includes using the described design strategy to produce other constructs as described herein. By combining these cDNAs in a modular fashion with other functional cDNA “bricks,” mucins of modified structure and functionality, given the benefit of this disclosure, can readily be constructed with known molecular techniques, including Gibson Assembly, Golden Gate Assembly, and other modern DNA assembly approaches.

An observation in this Part I 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 10).

Analyses of O-glycosylation in this Part 1 were partly based on lectin blots. Controls were used 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).

We modified 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. It is considered that swapping mucin cytoplasmic tails may be a viable strategy to at least partially engineer emergent glycoforms.

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	SEQ ID NO:
pcDNA3.1 Syn1 FWD	AGGTAGCGTCTCGTCCCGCCTCAGGCATACTTTATTG	18
pcDNA3.1 Syn1 REV	AGGTAGCGTCTCGTCGGGAGCAGGGGTAGCG	19
Syn1 FWD	AGGTAGCGTCTCGCCGATGCAGCTACTCCAGCTCCGGACGCC	20
Syn1 REV	AGGTAGCGTCTCGGGGAGCAGGGGTAGCG	21
pcDNA3.1 Syn2 FWD	CTTCTGCGTCTCGTCCCGCCTCAGGCATACTTTATTGGCGA	22
pcDNA3.1 Syn2 REV	CTTCTGCGTCTCGTCGGGAGGAGCTGGTGTAGCCGCG	23
Syn2 FWD	CTTCTGCGTCTCCCCGATGCAGCTACCCCGGCTCCACCC	24
Syn2 REV	CTTCTGCGTCTCCGGGAGGAGCTGGTGTAGCCGCG	25

Summary of cDNA "Biobricks" as described in Part I

Summary of cDNA "Biobricks" as described in Part I.

Leader Tag

1. **Native-FLAG**

   Amino acid sequence:

   MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAD YKDDDDLY (SEQ ID NO:26)

   cDNA sequence:

   GGATCCATGACACCGGGCACCCAGTCTCCTTTCTTCCTGCTGCTGCTCCTC ACAGTGCTTACAGTTGTTACAGGTTCTGGTCATGCAAGCTCTACCCCAGGT GGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCCAGCTCTAC TGAGAAGAATGCTGATTACAAGGATGACGACGACCTGTACA (SEQ ID NO:27)

2. **His-SUMO**

   Amino acid sequence:

   METDTLLLWVLLLWVPGSTGDGHHHHHHGSLQDSEVNQEAKPEVKPEVKPE THINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLTFLYDGIEIQAD QAPEDLDMEDNDIIEAHREQIGGGSGSGHASSTPGGEKETSATQRSSVPSSTEK NADYKDDDDLY (SEQ ID NO:28)

   cDNA sequence:

   GGATCCGCCACCATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCT CTGGGTTCCAGGTTCCACTGGTGACGGTCATCACCATCATCATCACGGGTC CCTGCAGGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCA GAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTC AGAGATCTTCTTCAAGATCAAAAAGACCACTCCTTTAAGAAGGCTGATGG AAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAACGTTCTTG TACGACGGTATTGAAATTCAAGCTGATCAGGCCCCTGAAGATTTGGACAT GGAGGATAACGATATTATTGAGGCTCACAGAGAACAGATTGGAGGTGGCT CCGGCTCCGGTCATGCAAGCTCTACCCCAGGTGGAGAAAAGGAGACTTCG

GCTACCCAGAGAAGTTCAGTGCCCAGCTCTACTGAGAAGAATGCTGATTA
CAAGGATGACGACGACCTGTACA (SEQ ID NO:29)

In the representative polymer backbone segment sequences presented immediately below, repeat sequences are proceeded by the following sequence: LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSV (SEQ ID NO:30) with the pertinent repeat sequence designated with the pertinent SEQ ID and the number of its repeats designated in brackets with a subscript, the subscript indicating the number of repeats. The alphnuermic names given above each sequence are names of the sequences, rather than sequences themselves.

Polymer Backbone

1. **Codon-Scrambled Muc1 x42 (Muc1_42)**
2. **Amino acid sequence:**
3. LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD VTSV[PDTRPAPGSTAPPAHGVTSA]$_{42}$ASG **(SEQ ID NO:30 [SEQ ID NO:8]$_{42}$ASG**

cDNA sequence:
TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCCCGGATACGCGACCCGCCCCAGGGTCAACAGCGCCC
CCAGCCCACGGCGTTACATCTGCACCTGACACTAGACCTGCGCCAGGATC
AACAGCTCCACCGGCTCACGGGGTCACCAGTGCCCCCGACACTCGACCAG
CTCCGGGGTCTACCGCTCCCCCGGCTCATGGTGTCACTAGCGCGCCTGACA
CACGCCCGGCACCAGGGAGTACGGCCCCTCCTGCGCACGGCGTAACTTCA
GCCCCAGATACTCGACCTGCTCCGGGCTCAACAGCCCCGCCTGCACATGG
AGTTACATCAGCCCCTGATACTAGACCGGCTCCAGGTTCAACTGCTCCGCC
AGCACATGGTGTAACGTCTGCGCCCGATACTCGCCCAGCACCTGGGTCCA
CAGCTCCCCCTGCGCATGGAGTAACATCAGCACCTGATACCAGACCTGCC
CCGGGCAGCACTGCACCCCCAGCACATGGCGTAACATCAGCACCAGATAC
TCGCCCCGCTCCTGGTTCCACGGCTCCCCCCGCGCATGGCGTTACTTCAGC
TCCAGATACACGGCCGGCACCCGGCAGTACGGCTCCACCCGCACATGGAG
TAACGAGTGCTCCGGACACTCGGCCTGCTCCAGGAAGTACCGCACCTCCG
GCCCATGGCGTGACAAGTGCTCCCGACACCAGACCAGCGCCTGGTTCAAC
AGCACCGCCAGCTCATGGTGTAACCTCAGCTCCCGATACTAGACCCGCGC
CAGGTTCCACCGCTCCACCTGCACACGGGGTGACGAGCGCACCTGATACG
CGCCCGGCACCGGGAAGCACAGCGCCTCCCGCTCACGGAGTCACTAGCGC
CCCGGATACAAGACCCGCACCTGGATCTACAGCTCCTCCAGCTCACGGCG
TCACGAGTGCACCCGATACACGACCGGCCCCAGGCTCTACAGCCCCACCA
GCACATGGAGTCACGAGTGCACCTGATACTAGGCCCGCTCCGGGTTCCAC
AGCACCTCCTGCACATGGTGTTACATCCGCTCCTGATACGAGACCCGCTCC
AGGCTCTACTGCCCCACCGGCACACGGCGTGACCAGTGCTCCAGATACCC

GGCCAGCTCCTGGGAGTACTGCGCCTCCAGCTCATGGCGTCACTAGTGCA
CCTGATACAAGACCAGCCCCCGGTTCCACTGCTCCACCAGCCCATGGTGT
AACAAGTGCACCGGACACAAGGCCAGCCCCTGGTAGTACTGCTCCTCCTG
CTCACGGTGTTACTAGTGCTCCTGACACCAGACCTGCCCCTGGAAGTACTG
CACCGCCTGCTCATGGAGTCACATCAGCTCCGGATACTCGGCCGGCTCCG
GGATCAACCGCTCCTCCGGCTCATGGAGTAACCTCCGCACCGGATACTAG
GCCTGCACCGGGGAGTACAGCACCACCTGCTCATGGTGTGACTAGCGCTC
CTGACACTCGCCCCGCTCCCGGTAGCACTGCCCCCCCTGCACATGGGGTG
ACTTCAGCTCCTGATACTCGGCCTGCACCCGGAAGCACAGCCCCCCCAGC
TCATGGGGTCACAAGCGCTCCAGATACTAGGCCAGCGCCGGGAAGTACAG
CCCCTCCAGCGCACGGTGTAACTTCCGCGCCAGACACACGCCCTGCTCCC
GGATCAACGGCACCTCCAGCACACGGTGTGACGTCCGCACCCGACACAAG
ACCGGCACCTGGTTCTACTGCACCTCCCGCGCACGGAGTTACTTCAGCACC
AGATACAAGACCTGCTCCTGGCTCAACTGCCCCTCCGGCGCATGGTGTAA
CTAGTGCGCCTGATACACGCCCAGCACCGGGTAGTACGGCACCACCAGCT
CATGGAGTTACGTCAGCTCCAGATACGCGCCCTGCACCAGGCAGTACAGC
TCCGCCGGCCCACGGAGTAACTAGCGCACCAGATACCAGGCCAGCACCCG
GTAGTACCGCGCCTCCTGCCCATGGAGTAACTTCCGCCCCCGATACCCGAC
CTGCACCTGGCAGTACCGCCCCTCCCGCCCACGGGGTAACCAGTGCACCA
GACACGCGGCCCGCACCAGGATCTACTGCTCCCCCAGCGCATGGGGTAAC
TTCTGCACCAGATACGAGGCCTGCCCCAGGTAGTACAGCGCCACCTGCCC
ACGGTGTCACCTCCGCTCCTGATACAAGGCCTGCGCCTGGATCAACTGCA
CCACCGGCGCACGGGGTTACAAGTGCCCCTGACACGAGACCAGCACCAGG
TTCTACGGCGCCTCCGGCACATGGAGTGACTAGTGCCCCAGACACTAGGC
CGGCTCCTGGATCAACCGCACCACCCGCTCATGGAGTGACATCAGCGCCA
GATACTAGACCAGCTCCCGGGTCAACTGCGCCGCCCGCCCATGGGGTTAC
TTCTGCTCCAGACACTCGCCCAGCCCCAGGATCAACGGCTCCTCCCGCACA
CGGAGTGACCTCTGCTCCTGATACCAGGCCAGCTCCAGGGTCTACAGCAC
CCCCTGCTCATGGGGTAACATCTGCCGCCTCAGG (SEQ ID NO:50)

4. Codon-Scrambled Muc1 x21 (Muc1_21)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD
VTSV[PDTRPAPGSTAPPAHGVTSA]$_{21}$ASG **(SEQ ID NO:30)[SEQ ID NO:8]$_{21}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCCCAGACACTCGGCCTGCACCGGGATCAACCGCCCCA
CCGGCTCATGGTGTAACTAGTGCGCCTGATACCAGACCAGCACCAGGGAG
TACTGCACCTCCTGCTCATGGGGTTACTAGTGCCCCCGATACGCGACCTGC
TCCTGGAAGCACAGCACCGCCGGCTCACGGCGTAACGAGTGCTCCTGACA
CAAGGCCCGCTCCAGGGTCAACTGCACCACCTGCACACGGAGTGACATCA
GCGCCAGATACGAGACCTGCACCAGGAAGTACAGCGCCGCCAGCCCACG
GAGTAACTTCAGCCCCGGACACTAGGCCAGCACCTGGTTCAACGGCGCCT
CCAGCCCATGGAGTAACATCCGCTCCCGATACTCGTCCTGCTCCGGGTTCC
ACAGCTCCTCCCGCACATGGGGTGACTAGTGCTCCAGATACTCGCCCAGC
ACCCGGTAGTACCGCTCCTCCTGCACATGGCGTCACTAGTGCACCAGACA

CGCGTCCGGCTCCTGGGTCTACAGCTCCACCAGCTCACGGAGTTACCAGT
GCACCTGACACTAGACCTGCGCCCGGTTCGACGGCTCCGCCCGCCCATGG
GGTAACGTCTGCGCCGGATACACGCCCTGCACCTGGATCTACCGCACCTC
CGGCCCATGGTGTCACGAGCGCACCTGATACGAGGCCTGCTCCAGGTAGT
ACTGCTCCCCCCGCTCATGGAGTTACTAGCGCTCCTGATACTCGACCGGCA
CCTGGCAGCACTGCTCCTCCAGCACATGGTGTTACATCGGCTCCAGACAC
ACGTCCCGCGCCAGGATCGACTGCTCCACCCGCTCACGGGGTCACATCTG
CACCCGATACACGGCCAGCTCCCGGTTCCACTGCCCCGCCTGCCCATGGC
GTTACTTCGGCACCAGATACCCGACCCGCACCAGGCAGTACAGCACCTCC
AGCGCATGGTGTGACAAGCGCCCCTGATACACGACCAGCTCCAGGCTCAA
CAGCACCACCAGCACACGGTGTAACCTCAGCTCCGGATACCCGTCCAGCT
CCTGGTAGTACAGCCCCTCCTGCGCACGGAGTCACAAGTGCTCCCGACAC
AAGACCAGCCCCAGGTTCTACTGCGCCACCTGCTCACGGTGTTACCTCTGC
CCCAGATACAAGACCTGCCCCTGGCTCTACGGCACCCCCGGCACATGGAG
TCACTTCCGCACCGGATACTAGACCAGCGCCTGGGAGTACGGCCCCCCCA
GCTCATGGCGTGACTTCTGCTGCCTCAGG (SEQ ID NO:51)

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

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD
VTSV[PDTRPAPGSTAPPAHGVTSA]$_{10}$ASG **(SEQ ID NO:30)[SEQ ID NO:8]$_{10}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCCCAGATACAAGACCGGCCCCAGGATCTACGGCTCCT
CCGGCTCATGGAGTCACTTCTGCTCCAGACACAAGGCCCGCGCCGGGTTC
TACAGCACCGCCTGCTCATGGTGTTACTAGCGCACCCGATACGAGACCTG
CTCCGGGATCAACGGCACCTCCTGCCCACGGGGTAACATCTGCACCGGAC
ACTCGCCCTGCGCCCGGTTCAACCGCTCCACCCGCACACGGAGTGACAAG
CGCTCCTGACACTAGACCAGCACCAGGTTCTACAGCCCCACCAGCCCATG
GAGTTACCAGTGCACCAGATACTAGGCCAGCTCCAGGTAGTACTGCACCC
CCAGCTCATGGGGTTACATCAGCTCCCGACACGCGACCAGCTCCTGGAAG
CACTGCCCCTCCAGCTCACGGTGTGACCTCAGCACCTGATACACGCCCTGC
ACCTGGCTCTACTGCTCCCCCCGCTCATGGCGTAACTAGTGCCCCGGATAC
TCGACCCGCCCCTGGTTCCACAGCTCCGCCAGCACATGGTGTAACAAGTG
CTCCTGATACCCGACCAGCGCCTGGAAGTACCGCACCACCTGCACATGGA
GTAACTTCAGCCGCCTCAGG (SEQ ID NO:52)

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

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD
VTSVGGGGGASG (SEQ ID NO:31)

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT

GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCGGCGGTGGTGGAGGAGCCTCAGG (SEQ ID NO:99)

7. **Codon-Scrambled Muc1 Single Glycosylation Mutant x21 (Muc1_21S) Amino acid sequence:**
8. LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD VTSV[PDTRPAPGATAPPAHGVTSA]$_{21}$ASG **(SEQ ID NO:30)[SEQ ID NO:5]$_{21}$ASG**

cDNA sequence:
TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCCCAGATACCAGACCTGCGCCTGGAGCCACAGCTCCT
CCTGCCCATGGCGTCACAAGTGCCCCTGACACACGCCCAGCTCCCGGGGC
TACAGCCCCACCTGCACATGGTGTTACTAGTGCACCAGACACCAGACCGG
CTCCGGGAGCCACGGCACCCCCCGCTCATGGTGTCACTTCCGCACCGGAT
ACGAGGCCAGCACCTGGGGCCACTGCGCCGCCGGCACATGGGGTGACTAG
TGCGCCAGATACTCGCCCTGCTCCAGGGGCTACTGCCCCTCCAGCTCATGG
CGTAACCTCAGCGCCTGATACCCGACCAGCGCCAGGTGCCACTGCACCGC
CAGCCCATGGGGTCACTAGTGCTCCTGACACTAGACCTGCACCTGGAGCT
ACAGCACCTCCAGCGCATGGTGTGACAAGCGCCCCAGACACGAGACCAGC
CCCCGGTGCCACCGCTCCTCCCGCACATGGAGTTACTAGCGCTCCGGACA
CAAGACCGGCACCAGGTGCGACTGCACCACCGGCTCATGGAGTAACTTCA
GCACCAGATACACGGCCTGCTCCCGGCGCTACAGCTCCACCAGCACATGG
CGTTACCTCCGCACCTGACACGAGGCCCGCTCCAGGAGCCACTGCTCCCC
CTGCACACGGTGTTACGTCAGCTCCAGATACGCGGCCAGCTCCGGGCGCA
ACAGCTCCCCCGGCTCACGGTGTAACCAGTGCTCCCGACACAAGGCCTGC
ACCCGGAGCAACCGCACCTCCGGCCCATGGTGTAACAAGTGCACCTGATA
CTAGGCCCGCGCCTGGTGCTACTGCTCCACCTGCTCACGGCGTGACATCAG
CCCCTGATACGAGACCTGCCCCAGGGGCAACTGCACCTCCTGCTCATGGG
GTAACTAGTGCCCCCGATACAAGACCAGCACCGGGAGCGACCGCCCCCCC
AGCACACGGAGTAACGAGCGCACCCGATACTCGACCTGCACCAGGAGCG
ACGGCTCCACCCGCTCACGGAGTCACGAGTGCTCCAGACACTCGACCTGC
TCCTGGCGCGACAGCACCACCAGCTCACGGGGTTACTAGTGCTCCTGATA
CACGACCCGCACCAGGGGCGACTGCTCCTCCAGCCCACGGAGTTACATCT
GCCCCGGATACAAGGCCAGCACCCGGTGCAACTGCTCCGCCCGCCCATGG
AGTCACAAGTGCTCCGGATACTAGACCAGCTCCTGGGGCTACGGCGCCTC
CTGCGCACGGAGTGACTTCTGCTGCCTCAGG

9. **Codon-Scrambled Muc1 Double Glycosylation Mutant x21 (Muc1_21D) Amino acid sequence:**
LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD VTSV[PDTRPAPGATAPPAHGVTAA]$_{21}$ASG **(SEQ ID NO:30)[SEQ ID NO:6]$_{21}$ASG**

cDNA sequence:
TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG

ATGTCACCTCGGTCCCAGACACGCGACCCGCACCAGGCGCGACTGCTCCT
CCTGCGCATGGTGTAACAGCGGCCCCTGATACGAGGCCAGCCCCTGGAGC
CACCGCACCTCCAGCACACGGAGTGACTGCAGCTCCCGATACTAGACCCG
CGCCAGGAGCAACAGCTCCTCCAGCTCATGGTGTGACGGCCGCCCCAGAT
ACCAGACCTGCCCCAGGGGCGACAGCACCCCCCGCTCACGGCGTAACTGC
AGCCCCGGATACGAGACCAGCTCCTGGGGCCACTGCACCTCCGGCTCATG
GGGTAACAGCTGCCCCCGATACCCGACCTGCACCCGGAGCTACAGCGCCG
CCTGCACACGGTGTAACCGCAGCTCCGGATACTAGACCTGCGCCTGGAGC
AACGGCGCCTCCTGCACATGGGGTTACTGCTGCGCCAGATACAAGGCCTG
CCCCTGGTGCAACAGCACCTCCTGCTCATGGCGTGACAGCTGCACCAGAC
ACAAGACCAGCGCCAGGTGCTACTGCACCACCTGCTCACGGGGTAACTGC
TGCTCCAGATACTCGCCCTGCACCGGGAGCGACGGCTCCACCAGCTCACG
GAGTAACGGCAGCACCTGACACTAGGCCGGCTCCGGGAGCTACGGCACCG
CCCGCACATGGCGTCACTGCGGCTCCTGACACACGACCAGCACCCGGTGC
CACAGCTCCGCCAGCACATGGTGTTACGGCTGCTCCCGACACGAGACCCG
CTCCTGGAGCTACTGCTCCCCCGGCTCACGGTGTTACTGCAGCGCCTGATA
CACGCCCAGCACCGGGGGCTACAGCACCACCAGCCCATGGGGTCACAGCA
GCTCCAGACACTCGGCCAGCCCCAGGTGCAACTGCTCCACCCGCCCATGG
TGTCACTGCTGCACCTGATACCAGGCCGGCACCAGGAGCCACGGCCCCGC
CGGCACATGGAGTGACCGCGGCACCCGATACAAGACCTGCTCCGGGCGCT
ACAGCCCCCCCAGCCCACGGAGTCACCGCTGCTCCTGATACTCGACCGGC
ACCTGGTGCTACAGCTCCACCGGCCCATGGCGTTACAGCAGCACCAGATA
CGAGGCCCGCTCCAGGTGCGACCGCTCCTCCCGCTCATGGAGTAACAGCC
GCTCCGGACACTAGACCGGCTCCCGGCGCAACTGCGCCCCCTGCCCATGG
AGTTACTGCCGCACCGGATACACGCCCTGCCCCGGGAGCAACTGCCCCTC
CAGCGCACGGAGTTACAGCTGCTGCCTCAGG (SEQ ID NO:53)

10. Codon-Scrambled Muc1 Triple Glycosylation Mutant x21 (Muc1_21T) Amino acid sequence:

11. LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD VTSV[PDARPAPGATAPPAHGVTAA]$_{21}$ASG **(SEQ ID NO:30[SEQ ID NO:7]$_{21}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCCCAGATGCAAGGCCTGCCCCGGGAGCGACAGCACCA
CCAGCACATGGAGTGACGGCCGCCCCAGACGCTCGACCGGCACCAGGAG
CAACTGCTCCTCCCGCACATGGGGTCACTGCGGCCCCTGATGCGAGGCCG
GCACCTGGAGCTACTGCTCCACCGGCCCATGGTGTCACTGCAGCCCCGGA
TGCTAGACCGGCTCCGGGCGCAACTGCGCCGCCAGCCCATGGAGTTACTG
CTGCGCCAGATGCGCGGCCTGCCCCAGGTGCTACAGCCCCCCCTGCCCAT
GGCGTAACAGCTGCCCCCGATGCTCGCCCTGCACCGGGAGCAACGGCGCC
TCCAGCGCACGGAGTAACGGCAGCACCAGATGCTCGGCCAGCACCGGGG
GCTACAGCTCCACCTGCTCACGGTGTAACTGCAGCGCCTGATGCACGACC
AGCCCCTGGAGCAACAGCTCCGCCTGCACACGGAGTGACTGCTGCACCTG
ATGCTAGGCCAGCCCCAGGGGCGACTGCACCTCCAGCACACGGTGTTACA
GCTGCTCCAGACGCACGCCCAGCACCCGGTGCCACAGCTCCTCCTGCGCA
TGGTGTGACAGCTGCACCAGACGCCCGACCCGCGCCAGGAGCCACGGCTC

CACCAGCTCACGGCGTGACCGCGGCTCCTGACGCTAGGCCAGCTCCTGGA
GCCACCGCTCCTCCAGCTCATGGCGTTACAGCAGCTCCCGACGCAAGACC
CGCTCCTGGGGCCACTGCTCCCCCCGCTCACGGGGTAACAGCCGCTCCGG
ATGCAAGACCTGCCCCTGGTGCTACTGCACCACCCGCCCATGGGGTTACT
GCAGCTCCGGACGCTAGACCTGCTCCGGGAGCTACAGCGCCCCCAGCCCA
CGGAGTCACAGCAGCACCTGACGCGAGACCAGCGCCAGGTGCAACTGCCC
CTCCTGCACATGGTGTTACTGCCGCACCGGATGCCAGACCTGCACCCGGA
GCTACGGCCCCGCCGGCTCATGGGGTAACTGCTGCTCCTGATGCCCGACC
CGCTCCAGGCGCGACCGCACCTCCTGCTCATGGAGTAACAGCGGCACCCG
ATGCACGGCCGGCTCCCGGCGCTACAGCACCTCCGGCACATGGCGTCACC
GCAGCTCCAGATGCCAGGCCCGCACCAGGTGCGACGGCACCGCCCGCTCA
TGGTGTAACCGCTGCTCCCGATGCGAGACCTGCGCCTGGTGCAACAGCAC
CCCCGGCTCACGGAGTTACGGCTGCTGCCTCAGG **(SEQ ID NO:54)**

12. Lubricin consensus, KEPAPTTP x20 (Syn4_20)

Amino acid sequence:

13. LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD VTSV[KEPAPTTP]$_{20}$ASG **(SEQ ID NO:30)[SEQ ID NO:1]$_{20}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCAAGGAACCTGCACCTACAACCCCGAAGGAGCCCGCA
CCGACCACCCCAAAAGAACCTGCGCCGACAACTCCAAAGGAGCCAGCTCC
AACGACGCCAAAGGAACCAGCACCTACGACCCCCAAGGAACCCGCCCCG
ACGACTCCGAAGGAGCCTGCACCAACAACTCCTAAAGAACCAGCGCCTAC
TACGCCTAAAGAACCTGCTCCTACTACACCAAAAGAGCCAGCACCCACGA
CACCGAAAGAACCTGCCCCTACTACCCCTAAAGAACCCGCTCCTACCACA
CCAAAGGAACCGGCTCCCACTACTCCCAAAGAACCAGCCCCAACTACACC
TAAAGAACCGGCCCCCACCACTCCTAAAGAGCCGGCGCCAACTACTCCAA
AAGAACCAGCTCCTACAACTCCCAAGGAGCCGGCACCTACTACTCCGAAA
GAGCCCGCGCCCACAACACCCAAAGAGCCTGCTCCGACTACTCCTGCCTC
AGG (SEQ ID NO:55)

14. Synthetic 1, DAATPAP x40 (Syn1_40)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD VTSV[DAATPAP]$_{40}$ASG **(SEQ ID NO:30)[SEQ ID NO:2]$_{40}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCGATGCAGCTACTCCAGCTCCGGACGCCGCAACACCC
GCTCCAGACGCCGCCACCCCAGCTCCAGATGCTGCTACACCTGCACCTGA
TGCCGCAACTCCCGCGCCGGATGCCGCGACTCCAGCACCGGACGCTGCGA
CGCCAGCCCCTGATGCTGCAACACCGGCTCCTGATGCTGCGACTCCTGCGC
CAGATGCAGCTACACCAGCCCCGGATGCTGCAACGCCTGCTCCTGACGCA
GCTACTCCGGCCCCCGACGCTGCTACCCCGGCGCCTGATGCTGCTACTCCC

GCTCCTGATGCGGCCACTCCAGCCCCAGACGCAGCAACCCCAGCCCCCGA
TGCTGCTACGCCTGCACCCGACGCGGCCACACCTGCGCCGGACGCAGCGA
CACCTGCCCCTGACGCTGCCACGCCCGCACCTGATGCAGCTACGCCAGCT
CCCGATGCGGCAACACCTGCTCCAGATGCCGCCACTCCTGCTCCGGATGC
GGCGACACCAGCGCCTGACGCCGCTACGCCGGCACCTGATGCTGCCACTC
CGGCTCCAGATGCAGCGACCCCAGCGCCAGACGCGGCAACTCCAGCGCCC
GATGCAGCTACCCCAGCACCAGATGCTGCAACCCCTGCACCGGATGCAGC
AACGCCAGCACCTGACGCGGCTACTCCTGCACCAGATGCAGCAACTCCTG
CCCCGGACGCGGCGACTCCCGCACCAGACGCTGCAACTCCGGCACCAGAT
GCGGCTACCCCCGCTCCCGACGCAGCCACTCCCGCCCCAGATGCAGCCAC
ACCAGCTCCTGATGCAGCAACACCAGCACCCGATGCCGCTACCCCTGCTC
CCGCCTCAGG (SEQ ID NO:56)

15. Synthetic 1, DAATPAP x80 (Syn1_80)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD
VTSV[DAATPAP]$_{80}$ASG **(SEQ ID NO:30)[SEQ ID NO:2]$_{80}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCGATGCAGCTACTCCAGCTCCGGACGCCGCAACACCC
GCTCCAGACGCCGCCACCCCAGCTCCAGATGCTGCTACACCTGCACCTGA
TGCCGCAACTCCCGCGCCGGATGCCGCGACTCCAGCACCGGACGCTGCGA
CGCCAGCCCCTGATGCTGCAACACCGGCTCCTGATGCTGCGACTCCTGCGC
CAGATGCAGCTACACCAGCCCCGGATGCTGCAACGCCTGCTCCTGACGCA
GCTACTCCGGCCCCCGACGCTGCTACCCCGGCGCCTGATGCTGCTACTCCC
GCTCCTGATGCGGCCACTCCAGCCCCAGACGCAGCAACCCCAGCCCCCGA
TGCTGCTACGCCTGCACCCGACGCGGCCACACCTGCGCCGGACGCAGCGA
CACCTGCCCCTGACGCTGCCACGCCCGCACCTGATGCAGCTACGCCAGCT
CCCGATGCGGCAACACCTGCTCCAGATGCCGCCACTCCTGCTCCGGATGC
GGCGACACCAGCGCCTGACGCCGCTACGCCGGCACCTGATGCTGCCACTC
CGGCTCCAGATGCAGCGACCCCAGCGCCAGACGCGGCAACTCCAGCGCCC
GATGCAGCTACCCCAGCACCAGATGCTGCAACCCCTGCACCGGATGCAGC
AACGCCAGCACCTGACGCGGCTACTCCTGCACCAGATGCAGCAACTCCTG
CCCCGGACGCGGCGACTCCCGCACCAGACGCTGCAACTCCGGCACCAGAT
GCGGCTACCCCCGCTCCCGACGCAGCCACTCCCGCCCCAGATGCAGCCAC
ACCAGCTCCTGATGCAGCAACACCAGCACCCGATGCCGCTACCCCTGCTC
CCGATGCAGCTACTCCAGCTCCGGACGCCGCAACACCCGCTCCAGACGCC
GCCACCCCAGCTCCAGATGCTGCTACACCTGCACCTGATGCCGCAACTCCC
GCGCCGGATGCCGCGACTCCAGCACCGGACGCTGCGACGCCAGCCCCTGA
TGCTGCAACACCGGCTCCTGATGCTGCGACTCCTGCGCCAGATGCAGCTA
CACCAGCCCCGGATGCTGCAACGCCTGCTCCTGACGCAGCTACTCCGGCC
CCCGACGCTGCTACCCCGGCGCCTGATGCTGCTACTCCCGCTCCTGATGCG
GCCACTCCAGCCCCAGACGCAGCAACCCCAGCCCCCGATGCTGCTACGCC
TGCACCCGACGCGGCCACACCTGCGCCGGACGCAGCGACACCTGCCCCTG
ACGCTGCCACGCCCGCACCTGATGCAGCTACGCCAGCTCCCGATGCGGCA
ACACCTGCTCCAGATGCCGCCACTCCTGCTCCGGATGCGGCGACACCAGC
GCCTGACGCCGCTACGCCGGCACCTGATGCTGCCACTCCGGCTCCAGATG
CAGCGACCCCAGCGCCAGACGCGGCAACTCCAGCGCCCGATGCAGCTACC

CCAGCACCAGATGCTGCAACCCCTGCACCGGATGCAGCAACGCCAGCACC
TGACGCGGCTACTCCTGCACCAGATGCAGCAACTCCTGCCCCGGACGCGG
CGACTCCCGCACCAGACGCTGCAACTCCGGCACCAGATGCGGCTACCCCC
GCTCCCGACGCAGCCACTCCCGCCCCAGATGCAGCCACACCAGCTCCTGA
TGCAGCAACACCAGCACCCGATGCCGCTACCCCTGCTCCCGCCTCAGG
(SEQ ID NO:57)

16. Synthetic 2, DAATPAPP x40 (Syn2_40)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD
VTSV[DAATPAPP]$_{40}$ASG **(SEQ ID NO:30)[SEQ ID NO:3]$_{40}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCGATGCAGCTACCCCGGCTCCACCCGATGCGGCAACA
CCAGCCCCTCCCGATGCAGCAACACCTGCTCCCCCCGATGCTGCTACCCCT
GCTCCGCCTGATGCTGCAACTCCAGCTCCGCCCGATGCCGCTACACCTGCC
CCCCCTGACGCCGCCACGCCCGCTCCTCCGGATGCTGCAACCCCAGCACC
CCCAGACGCCGCTACCCCAGCTCCACCAGATGCTGCTACACCCGCACCAC
CTGATGCCGCAACACCGGCGCCTCCTGATGCTGCTACTCCAGCCCCACCTG
ATGCAGCAACTCCTGCGCCACCAGACGCTGCCACACCTGCACCACCAGAT
GCAGCCACACCAGCACCGCCAGACGCAGCAACGCCGGCTCCGCCAGATGC
AGCGACACCAGCGCCACCTGACGCAGCGACTCCAGCACCACCGGATGCGG
CTACCCCCGCTCCGCCGGACGCGGCGACTCCTGCCCCTCCTGACGCGGCA
ACTCCGGCCCCTCCAGATGCGGCGACCCCAGCCCCGCCGGATGCCGCGAC
TCCGGCTCCCCCGGACGCTGCAACACCCGCTCCACCTGATGCTGCCACTCC
CGCGCCTCCAGATGCTGCAACGCCAGCTCCCCCTGATGCTGCGACGCCTG
CTCCTCCAGATGCAGCTACACCGGCTCCTCCTGATGCAGCTACGCCTGCAC
CGCCTGACGCTGCTACGCCAGCACCTCCCGACGCAGCCACTCCTGCACCTC
CTGATGCGGCCACTCCAGCGCCCCCGGATGCAGCTACTCCTGCTCCACCG
GACGCCGCAACTCCCGCCCCTCCGGACGCAGCTACTCCCGCTCCCCCAGA
TGCAGCAACCCCTGCACCCCCCGACGCGGCCACCCCTGCCCCACCAGATG
CCGCCACTCCGGCACCACCCGACGCTGCGACTCCCGCACCTCCAGACGCG
GCTACACCAGCTCCTCCCGCCTCAGG (SEQ ID NO:58)

17. Synthetic 2, DAATPAPP x80 (Syn2_80)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD
VTSV[DAATPAPP]$_{80}$ASG **(SEQ ID NO:30)[SEQ ID NO:3]$_{80}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCGATGCAGCTACCCCGGCTCCACCCGATGCGGCAACA
CCAGCCCCTCCCGATGCAGCAACACCTGCTCCCCCCGATGCTGCTACCCCT
GCTCCGCCTGATGCTGCAACTCCAGCTCCGCCCGATGCCGCTACACCTGCC
CCCCCTGACGCCGCCACGCCCGCTCCTCCGGATGCTGCAACCCCAGCACC
CCCAGACGCCGCTACCCCAGCTCCACCAGATGCTGCTACACCCGCACCAC
CTGATGCCGCAACACCGGCGCCTCCTGATGCTGCTACTCCAGCCCCACCTG

ATGCAGCAACTCCTGCGCCACCAGACGCTGCCACACCTGCACCACCAGAT
GCAGCCACACCAGCACCGCCAGACGCAGCAACGCCGGCTCCGCCAGATGC
AGCGACACCAGCGCCACCTGACGCAGCGACTCCAGCACCACCGGATGCGG
CTACCCCCGCTCCGCCGGACGCGGCGACTCCTGCCCCTCCTGACGCGGCA
ACTCCGGCCCCTCCAGATGCGGCGACCCCAGCCCCGCCGGATGCCGCGAC
TCCGGCTCCCCCGGACGCTGCAACACCCGCTCCACCTGATGCTGCCACTCC
CGCGCCTCCAGATGCTGCAACGCCAGCTCCCCCTGATGCTGCGACGCCTG
CTCCTCCAGATGCAGCTACACCGGCTCCTCCTGATGCAGCTACGCCTGCAC
CGCCTGACGCTGCTACGCCAGCACCTCCCGACGCAGCCACTCCTGCACCTC
CTGATGCGGCCACTCCAGCGCCCCCGGATGCAGCTACTCCTGCTCCACCG
GACGCCGCAACTCCCGCCCCTCCGGACGCAGCTACTCCCGCTCCCCCAGA
TGCAGCAACCCCTGCACCCCCCGACGCGGCCACCCCTGCCCCACCAGATG
CCGCCACTCCGGCACCACCCGACGCTGCGACTCCCGCACCTCCAGACGCG
GCTACACCAGCTCCTCCCGATGCAGCTACCCCGGCTCCACCCGATGCGGC
AACACCAGCCCCTCCCGATGCAGCAACACCTGCTCCCCCCGATGCTGCTA
CCCCTGCTCCGCCTGATGCTGCAACTCCAGCTCCGCCCGATGCCGCTACAC
CTGCCCCCCCCTGACGCCGCCACGCCCGCTCCTCCGGATGCTGCAACCCCAG
CACCCCCAGACGCCGCTACCCCAGCTCCACCAGATGCTGCTACACCCGCA
CCACCTGATGCCGCAACACCGGCGCCTCCTGATGCTGCTACTCCAGCCCCA
CCTGATGCAGCAACTCCTGCGCCACCAGACGCTGCCACACCTGCACCACC
AGATGCAGCCACACCAGCACCGCCAGACGCAGCAACGCCGGCTCCGCCA
GATGCAGCGACACCAGCGCCACCTGACGCAGCGACTCCAGCACCACCGGA
TGCGGCTACCCCCGCTCCGCCGGACGCGGCGACTCCTGCCCCTCCTGACGC
GGCAACTCCGGCCCCTCCAGATGCGGCGACCCCAGCCCCGCCGGATGCCG
CGACTCCGGCTCCCCCGGACGCTGCAACACCCGCTCCACCTGATGCTGCC
ACTCCCGCGCCTCCAGATGCTGCAACGCCAGCTCCCCCTGATGCTGCGAC
GCCTGCTCCTCCAGATGCAGCTACACCGGCTCCTCCTGATGCAGCTACGCC
TGCACCGCCTGACGCTGCTACGCCAGCACCTCCCGACGCAGCCACTCCTG
CACCTCCTGATGCGGCCACTCCAGCGCCCCCGGATGCAGCTACTCCTGCTC
CACCGGACGCCGCAACTCCCGCCCCTCCGGACGCAGCTACTCCCGCTCCC
CCAGATGCAGCAACCCCTGCACCCCCCGACGCGGCCACCCCTGCCCCACC
AGATGCCGCCACTCCGGCACCACCCGACGCTGCGACTCCCGCACCTCCAG
ACGCGGCTACACCAGCTCCTCCCGCCTCAGG (SEQ ID NO:59)

18. Synthetic 3, PPASTSAPG x40 (Syn3_40)

Amino acid sequence:

LYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQD
VTSV[PPASTSAPG]$_{40}$ASG **(SEQ ID NO:30)[SEQ ID NO:4]$_{40}$ASG**

cDNA sequence:

TGTACATGGACATGGTCGCTGTGAGTATGACCAGCAGCGTACTCTCCAGC
CACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCT
GGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGG
ATGTCACCTCGGTCCCACCTGCATCTACCAGTGCCCCGGGTCCACCTGCCT
CTACTAGCGCCCCAGGACCTCCGGCAAGTACATCAGCGCCAGGACCCCCT
GCTTCCACTAGTGCACCCGGTCCCCCGGCATCTACGTCTGCCCCTGGCCCA
CCTGCTTCAACTTCAGCACCAGGACCACCCGCAAGCACATCAGCCCCAGG
CCCTCCCGCCTCTACAAGCGCTCCGGGGCCTCCGGCCTCTACCTCAGCTCC
AGGCCCACCAGCCAGCACTTCAGCCCCTGGTCCACCCGCTTCAACCTCAG
CACCCGGACCTCCTGCCTCAACTTCCGCTCCCGGTCCACCAGCTAGTACCT
CTGCTCCGGGCCCTCCGGCGAGCACGTCAGCACCGGGACCACCTGCGAGT

ACAAGTGCACCTGGCCCGCCCGCTAGCACAAGTGCCCCCGGTCCTCCAGC
ATCCACTAGTGCACCAGGGCCTCCAGCCAGCACTAGTGCGCCGGGTCCCC
CCGCGAGTACGTCAGCTCCGGGACCTCCAGCTTCTACATCTGCTCCTGGGC
CCCCTGCATCAACTAGTGCCCCTGGACCACCGGCTAGTACGTCAGCTCCTG
GTCCCCCTGCCAGTACTAGCGCTCCAGGGCCACCAGCAAGTACGAGCGCA
CCAGGCCCCCCAGCCTCTACGAGTGCACCGGGTCCTCCTGCAAGTACCTCC
GCTCCAGGTCCTCCGGCTTCAACGTCCGCACCTGGACCTCCCGCGTCCACA
TCAGCTCCCGGCCCTCCAGCGAGTACTTCTGCTCCCGGACCACCAGCGTCC
ACATCTGCGCCTGGTCCTCCCGCTAGTACCTCTGCACCTGGTCCGCCGGCC
AGTACAAGTGCTCCCGGGCCTCCCGCATCAACATCTGCACCAGGTCCACC
GGCGTCTACTAGTGCCCCAGGTCCCCCAGCTTCAACATCAGCACCTGGGC
CGCCTGCTAGTACATCCGCTCCTGGACCCCCAGCAAGTACTTCCGCCCCTG
GGCCTCCTGCTTCTACTTCAGCTCCTGGCCCTCCTGCGTCAACTAGTGCTC
CAGGACCGCCAGCTAGTACTTCCGCGCCCGGTGCCTCAGG (SEQ ID NO:60)

Optical Reporter

1. **mOxGFP**

Amino acid sequence:

SGSASGSAMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLT
LKFISTTGKLPVPWPTLVTTLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQE
RTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSH
NVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHY
LSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSA (SEQ ID NO:31)

cDNA sequence:

CCTCAGGCTCTGCATCAGGCTCAGCTATGGTGTCCAAGGGCGAGGAGCTG
TTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGG
CCACAAGTTCTCCGTGCGGGGCGAGGGCGAGGGCGATGCCACCAACGGC
AAGCTGACCCTGAAGTTCATCAGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGAGCTTCTCCCGCTA
CCCCGACCACATGAAGCGCCACGACTTCTTCAAGAGCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCACCTACAAG
ACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA
GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTTCAACTCCCACAACGTCTATATCACCGCCGACAAGCA
GAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACGTGGAGGAC
GGCTCCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGA
CGGCCCCGTGCTGCTGCCCGACAACCACTACCTGTCCACCCAGTCCAAGCT
GTCCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTTCTGGAATTCG
TGACCGCCGCCGGGATCACTCACGGCATGGACGAGCTGTATAAGGGCTC
AGC (SEQ ID NO:61)

Membrane Anchor

1. **Native TM**

Amino acid sequence:

SASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHHSSV
PPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFL
QIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASR
YNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLVCVLVALAIVYLIALAVC
QCRRK* (SEQ ID NO:32)

cDNA sequence:
GCTCAGCTTCTACTCTGGTGCACAACGGCACCTCTGCCAGGGCTACCACA
ACCCCAGCCAGCAAGAGCACTCCATTCTCAATTCCCAGCCACCACTCTGAT
ACTCCTACCACCCTTGCCAGCCATAGCACCAAGACTGATGCCAGTAGCAC
TCACCATAGCTCGGTACCTCCTCTCACCTCCTCCAATCACAGCACTTCTCC
CCAGTTGTCTACTGGGGTCTCTTTCTTTTTCCTGTCTTTTCACATTTCAAAC
CTCCAGTTTAATTCCTCTCTGGAAGATCCCAGCACCGACTACTACCAAGAG
CTGCAGAGAGACATTTCTGAAATGTTTTTGCAGATTTATAAACAAGGGGG
TTTTCTGGGCCTCTCCAATATTAAGTTCAGGCCAGGATCTGTGGTGGTACA
ATTGACTCTGGCCTTCCGAGAAGGTACCATCAATGTCCACGACGTGGAGA
CACAGTTCAATCAGTATAAAACGGAAGCAGCCTCTCGATATAACCTGACG
ATCTCAGACGTCAGCGTGAGTGATGTGCCATTTCCTTTCTCTGCCCAGTCT
GGGGCTGGGGTGCCAGGCTGGGGCATCGCGCTGCTGGTGCTGGTCTGTGT
TCTGGTTGCGCTGGCCATTGTCTATCTCATTGCCTTGGCTGTCTGTCAGTGC
CGCCGAAAGTAGGGAATTC (SEQ ID NO:62)

2. **Synthetic TM TM21**

Amino acid sequence:
ASGILYWRNPTESDSIVLAIIVPSLLLLLCLALLWYMRRRSM* (SEQ ID NO:49)

cDNA sequence:
CCTCAGGCATACTTTATTGGCGAAACCCAACGGAAAGTGATAGCATCGTT
TTGGCAATTATCGTCCCCAGTCTGCTCCTCTTGCTCTGCCTGGCTTTGTTGT
GGTACATGCGCCGACGAAGTATGTAGGAATTC (SEQ ID NO:63)

Cytoplasmic Motif

1. **Native CT**

Amino acid sequence:
SRCQCRRKNYGQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSTDRSPYEKVS
AGNGGSSLSYTNPAVAAASANL* (SEQ ID NO:33)

cDNA sequence:
TCTAGATGTCAGTGCCGCCGAAAGAACTACGGGCAGCTGGACATCTTTCC
AGCCCGGGATACCTACCATCCTATGAGCGAGTACCCCACCTACCACACCC
ATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTAGCCCCTATGAGAAG
GTTTCTGCAGGTAAtGGTGGCAGCAGCCTCTCTTACACAAACCCAGCAGTG
GCAGCCGCTTCTGCCAACTTGTAGGAATTC (SEQ ID NO:64)

2. **CQC**

Amino acid sequence:
SRCQCRRK* (SEQ ID NO:34)

cDNA sequence:
TCTAGATGTCAGTGCCGCCGAAAGTAGGAATTC (SEQ ID NO:65)

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

The following sequence are representative amino acid sequences for mucin and lubricin constructs, as further described herein, and for which the entire sequences, including the N-terminal signal sequence, tandem repeat domain, fluorescent optical reporter (green flourescent (GFP) in certain of these sequences), the transmembrane domain to the cytoplasmic tail domain. In embodiments, modified lubricins omit the transmembrain domain, the cytoplasmic tail, domain, and the optical reporter. It will be recognized that the GFP sequence may be, omitted or substituted by any other amino acid sequence, including but not limited to the sequence of other detectable proteins, or second polypeptides, as described above. The alphnuermic names given above each sequence are names of the sequences, rather than sequences themselves.

1. PDTRPAPGSTAPPAHGVTSA_42
Muc1_42_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGST
APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRP
APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP
DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG
VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP
PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAP
GSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD
TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT
SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA

HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGST
APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRP
APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP
DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG
VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP
PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAP
GSTAPPAHGVTSAASGSASGSAMVSKGEELFTGVVPILVELDGDVNGHKFSVRG
EGEGDATNGKLTLKFISTTGKLPVPWPTLVTTLTYGVQSFSRYPDHMKRHDFFKS
AMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKL
EYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLL
PDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSASTLVHNGT
SARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHHSSVPPLTSSNHSTSPQ
LSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNI
KFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPF
SAQSGAGVPGWGIALLVLVCVLVALAIVYLIALAVCQCRRK* (SEQ ID NO:35)

2. **PDTRPAPGSTAPPAHGVTSA_21**
Muc1_21_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGST
APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRP
APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP
DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG
VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP
PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAP
GSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD
TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAASGSASGSAMVSKGEEL
FTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTGKLPVPWPTLVT
TLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEG
DTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNV
EDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVT
AAGITHGMDELYKGSASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLASHS
TKTDASSTHHSSVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQ
ELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTINVHDVETQFNQ
YKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLVCVLVALAIVYLI
ALAVCQCRRK* (SEQ ID NO:36)

3. **PDTRPAPGSTAPPAHGVTSA_10**
Muc1_10_TM21_CT

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGST
APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRP
APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP
DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAASGILYWRNPTESDSIV
LAIIVPSLLLLLCLALLWYSRCQCRRKNYGQLDIFPARDTYHPMSEYPTYHTHGR
YVPPSSTDRSPYEKVSAGNGGSSLSYTNPAVAAASANL*(SEQ ID NO:37)

4. **PDTRPAPGSTAPPAHGVTSA_0**
Muc1_0_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVGGGGGASGSASGSAMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEG
EGDATNGKLTLKFISTTGKLPVPWPTLVTTLTYGVQSFSRYPDHMKRHDFFKSA
MPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE
YNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLP
DNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSASTLVHNGTS
ARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHHSSVPPLTSSNHSTSPQL
STGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNIKF
RPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSA
QSGAGVPGWGIALLVLVCVLVALAIVYLIALAVCQCRRK* (SEQ ID NO:38)

5. PDTRPAPGATAPPAHGVTSA_21
Muc1_21S_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVPDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTRPAPGA
TAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTR
PAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTS
APDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPA
HGVTSAPDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTRPAPGA
TAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTR
PAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTS
APDTRPAPGATAPPAHGVTSAPDTRPAPGATAPPAHGVTSAASGSASGSAMVSK
GEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTGKLPVPWP
TLVTTLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEV
KFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIR
HNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLE
FVTAAGITHGMDELYKGSASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLA
SHSTKTDASSTHHSSVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTD
YYQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTINVHDVETQ
FNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLVCVLVALAI
VYLIALAVCQCRRK* (SEQ ID NO:39)

6. PDTRPAPGATAPPAHGVTAA_21
Muc1_21D_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVPDTRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAPDTRPAPG
ATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAPD
TRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHG
VTAAPDTRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAPDTRPAPGAT
APPAHGVTAAPDTRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAPDTR
PAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVT

AAPDTRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAPDTRPAPGATAP
PAHGVTAAPDTRPAPGATAPPAHGVTAAPDTRPAPGATAPPAHGVTAAASGSAS
GSAMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTG
KLPVPWPTLVTTLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGT
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNG
IKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKR
DHMVLLEFVTAAGITHGMDELYKGSASTLVHNGTSARATTTPASKSTPFSIPSHH
SDTPTTLASHSTKTDASSTHHSSVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNS
SLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTI
NVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLV
CVLVALAIVYLIALAVCQCRRK* (SEQ ID NO:40)

7. **PDARPAPGATAPPAHGVTAA_21**
Muc1_21T_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVPDARPAPGATAPPAHGVTAAPDARPAPGATAPPAHGVTAAPDARPAPG
ATAPPAHGVTAAPDARPAPGATAPPAHGVTAAPDARPAPGATAPPAHGVTAAPD
ARPAPGATAPPAHGVTAAPDARPAPGATAPPAHGVTAAPDARPAPGATAPPAHG
VTAAPDARPAPGATAPPAHGVTAAPDARPAPGATAPPAHGVTAAPDARPAPGAT
APPAHGVTAAPDARPAPGATAPPAHGVTAAPDARPAPGATAPPAHGVTAAPDAR
PAPGATAPPAHGVTAAPDARPAPGATAPPAHGVTAAPDARPAPGATAPPAHGVT
AAPDARPAPGATAPPAHGVTAAPDARPAPGATAPPAHGVTAAPDARPAPGATAP
PAHGVTAAPDARPAPGATAPPAHGVTAAPDARPAPGATAPPAHGVTAAASGSAS
GSAMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTG
KLPVPWPTLVTTLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGT
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNG
IKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKR
DHMVLLEFVTAAGITHGMDELYKGSASTLVHNGTSARATTTPASKSTPFSIPSHH
SDTPTTLASHSTKTDASSTHHSSVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNS
SLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTI
NVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLV
CVLVALAIVYLIALAVCQCRRK*(SEQ ID NO:41)

8. **KEPAPTTP_20 (Syn4_20)**
Muc1_Syn4_20_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP
APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAP
TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPASGSASG
SAMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTGK
LPVPWPTLVTTLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGI
KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKR
DHMVLLEFVTAAGITHGMDELYKGSASTLVHNGTSARATTTPASKSTPFSIPSHH
SDTPTTLASHSTKTDASSTHHSSVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNS
SLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTI

NVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLV
CVLVALAIVYLIALAVCQCRRK*(SEQ ID NO:42)

9. **DAATPAP_40 (Syn1_40)**
Muc1_Syn1_40_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPA
PDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAAT
PAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDA
ATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAP
DAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATP
APDAATPAPDAATPAPASGSASGSAMVSKGEELFTGVVPILVELDGDVNGHKFS
VRGEGEGDATNGKLTLKFISTTGKLPVPWPTLVTTLTYGVQSFSRYPDHMKRHD
FFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG
HKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSASTLVH
NGTSARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHHSSVPPLTSSNHST
SPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLGL
SNIKFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSDVP
FPFSAQSGAGVPGWGIALLVLVCVLVALAIVYLIALAVCQCRRK* (SEQ ID
NO:43)

10. **DAATPAP_80 (Syn1_80)**
Muc1_Syn1_80_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPA
PDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAAT
PAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDA
ATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAP
DAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATP
APDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAA
TPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPD
AATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPA
PDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAAT
PAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDAATPAPDA
ATPAPDAATPAPDAATPAPDAATPAPASGSASGSAMVSKGEELFTGVVPILVELD
GDVNGHKFSVRGEGEGDATNGKLTLKFISTTGKLPVPWPTLVTTLTYGVQSFSRY
PDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGID
FKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQ
QNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELY
KGSASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHHSSV
PPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQI
YKQGGFLGLSNIKFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRYNLT
ISDVSVSDVPFPFSAQSGAGVPGWGIALLVLVCVLVALAIVYLIALAVCQCRRK*
(SEQ ID NO:44)

11. **DAATPAPP_40 (Syn2_40)**
Muc1_Syn1_40_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPD
AATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPP
DAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAP
PDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPA
PPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATP
APPDAATPAPPDAATPAPPDAATPAPPDAATPAPPASGS
ASGSAMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFIST
TGKLPVPWPTLVTTLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQK
NGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNE
KRDHMVLLEFVTAAGITHGMDELYKGSASTLVHNGTSARATTTPASKSTPFSIPS
HHSDTPTTLASHSTKTDASSTHHSSVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQF
NSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREG
TINVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLV
LVCVLVALAIVYLIALAVCQCRRK* (SEQ ID NO:45)

12. **DAATPAPP_80 (Syn2_80)**
Muc1_Syn1_40_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPD
AATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPP
DAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAP
PDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPA
PPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATP
APPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAAT
PAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAA
TPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDA
ATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPD
AATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPP
DAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAPPDAATPAP
PDAATPAPPDAATPAPPDAATPAPPDAATPAPPASGSASGSAMVSKG
EELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTGKLPVPWPT
LVTTLTYGVQSFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVK
FEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRH
NVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEF
VTAAGITHGMDELYKGSASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLAS
HSTKTDASSTHHSSVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDY
YQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTINVHDVETQF
NQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLVCVLVALAIV
YLIALAVCQCRRK* (SEQ ID NO:46)

13. **PPASTSAPG_40 (Syn3_40)**

Muc1_Syn1_40_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTS
APGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGP
PASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPAST
SAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPG
PPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPAS
TSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAP
GPPASTSAPGPPASTSAPGASGSASGSAMVSKGEELFTGVVPILVELDGDVNGHK
FSVRGEGEGDATNGKLTLKFISTTGKLPVPWPTLVTTLTYGVQSFSRYPDHMKRH
DFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSASTLV
HNGTSARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHHSSVPPLTSSNH
STSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFL
GLSNIKFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSD
VPFPFSAQSGAGVPGWGIALLVLVCVLVALAIVYLIALAVCQCRRK* (SEQ ID
NO:47)

14. **PPASTSAPG_80 (Syn3_80)**

Muc1_Syn1_40_mOxGFP_dCT_BlpI

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNADYK
DDDDLYMDMVAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWG
QDVTSVPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTS
APGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGP
PASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPAST
SAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPG
PPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPAS
TSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAP
GPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPA
STSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSA
PGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPP
ASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTS
APGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGP
PASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPASTSAPGPPAST
SAPGPPASTSAPGPPASTSAPGPPASTSAPGASGSASGSAMVSKGEELFTGVVPILV
ELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTGKLPVPWPTLVTTLTYGVQSF
SRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIEL
KGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLAD
HYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYKGSASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHH
SSVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMF
LQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRY
NLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLVCVLVALAIVYLIALAVCQCR
RK (SEQ ID NO:48)

List of Constructs used in Part I

Membrane Associated Mucin

52. pcDNA3.1+_Muc1_0_TM21
53. pcDNA3.1+_Muc1_10_TM21
54. pcDNA3.1+_Muc1_21_TM21
55. pcDNA3.1+_Muc1_42_TM21
56. pcDNA3.1+_Muc1_21S_TM21
57. pcDNA3.1+_Muc1_21D_TM21
58. pcDNA3.1+_Muc1_21T_TM21
59. pcDNA3.1+_Muc1_10_TM21_CT
60. pcDNA3.1+_Muc1_10_TM21_CQC
61. pcDNA3.1+_Muc1_10_dCT
62. pcDNA3.1+_Muc1_10_FL
63. pcDNA3.1+_Muc1_Syn4_20_TM21
64. pcDNA3.1+_Muc1_Syn1_40_TM21
65. pcDNA3.1+_Muc1_Syn2_40_TM21
66. pcDNA3.1+_Muc1_Syn3_40_TM21
67. pcDNA3.1+_Muc1_Syn1_80_TM21
68. pcDNA3.1+_Muc1_Syn2_80_TM21
69. pcDNA3.1+_Muc1_Syn3_80_TM21
70. pPB_Tet_Muc1_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
71. pPB_Tet_Muc1_42_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
72. pPB_Tet_Muc1_21_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
73. pPB_Tet_Muc1_10_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
74. pPB_Tet_Muc1_0_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
75. pPB_Tet_Muc1_21D_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
76. pPB_Tet_Muc1_21T_TM21_IRES2_copGFP_rtTAsM2_IRES_NeoR
77. pLV_puro_teton_Muc1_42_dCT
78. pLV_puro_teton_Muc1_dCT
79. pPB_Muc1_mOxGFP_dCT_BlpI
80. pPB_Muc1_42_mOxGFP_dCT_BlpI
81. pPB_Muc1_21_mOxGFP_dCT_BlpI
82. pPB_Muc1_10_mOxGFP_dCT_BlpI
83. pPB_Muc1_0_mOxGFP_dCT_BlpI
84. pPB_Muc1_21S_mOxGFP_dCT_BlpI
85. pPB_Muc1_21D_mOxGFP_dCT_BlpI
86. pPB_Muc1_21T_mOxGFP_dCT_BlpI
87. pPB_Muc1_Syn4_20_mOxGFP_dCT_BlpI
88. pPB_Muc1_Syn1_40_mOxGFP_dCT_BlpI
89. pPB_Muc1_Syn2_40_mOxGFP_dCT_BlpI
90. pPB_Muc1_Syn3_40_mOxGFP_dCT_BlpI
91. pPB_Muc1_Syn1_80_mOxGFP_dCT_BlpI
92. pPB_Muc1_Syn2_80_mOxGFP_dCT_BlpI

Secreted Mucin

93. pPB_Tet_SumoStar_Muc1_42_rtTAsM2_IRES_NeoR
94. pPB_Tet_SumoStar_Muc1_21T_rtTAsM2_IRES_NeoR
95. pPB_Tet_SumoStar_Muc1_21D_rtTAsM2_IRES_NeoR

96. pPB_Tet_SumoStar_Muc1_21S_rtTAsM2_IRES_NeoR
97. pPB_Tet_SumoStar_Muc1_21_rtTAsM2_IRES_NeoR
98. pPB_Tet_SumoStar_Muc1_0_rtTAsM2_IRES_NeoR
99. pPB_Tet_SumoStar_Muc1_Syn1_40_rtTAsM2_IRES_NeoR
100. pPB_Tet_SumoStar_Muc1_Syn2_40_rtTAsM2_IRES_NeoR
101. pPB_Tet_SumoStar_Muc1_Syn3_40_rtTAsM2_IRES_NeoR
102. pPB_Tet_SumoStar_Muc1_Syn1_80_rtTAsM2_IRES_NeoR
103. pPB_Tet_SumoStar_Muc1_Syn2_80_rtTAsM2_IRES_NeoR

References cited in Part I - references listed in any part of this disclosure is not an indication that any of the references are material to patentability.

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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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**.//doi.org/10.1021/acsbiomaterials.7b00037.

(28) Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells | Nature Methods//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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.//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//www.nature.com/articles/nmeth.2089 (accessed Oct 31, 2018).//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.

(54) Fukuda, M. Beta-Elimination for Release of O-GalNAc-Linked Oligosaccharides from Glycoproteins and Glycopeptides. *Curr Protoc Mol Biol* **2001**, *Chapter 17*, Unit17.15B.//doi.org/10.1002/0471142727.mb1715bs31.

(55) Shajahan, A.; Heiss, C.; Ishihara, M.; Azadi, P. Glycomic and Glycoproteomic Analysis of Glycoproteins-a Tutorial. *Anal Bioanal Chem* **2017**, *409* (19), 4483–4505.//doi.org/10.1007/s00216-017-0406-7.

Part II

This Part II of the disclosure illustrates mucin-coating technologies for protection and reduced aggregation of cellular production systems.

In connection with this Part II, optimization of host-cell production systems with improved yield and production reliability is desired in order to meet the increasing demand for biologics with complex post-translational modifications. Prior to the present disclosure, aggregation of suspension-adapted mammalian cells remained a significant problem that can limit the cellular density and per volume yield of bio-reactors. This Part II provides a

genetically encoded technology that directs the synthesis of anti-adhesive and protective coatings on the cellular surface. We genetically encode new cell-surface coatings through the fusion of engineered mucin domains to synthetic transmembrane anchors. Combined with appropriate expression systems, the mucin coating technology directs the assembly of thick, highly hydrated barriers to strongly mitigate cell aggregation and protect cells in suspension against fluid shear stresses. The coating technology is demonstrated on suspension adapted human 293-F cells, which resist clumping even in media formulations that otherwise would induce extreme cell aggregation and show improved performance over commercially available anti-clumping agent. The stable biopolymer coatings do not show deleterious effects on cell proliferation rate, efficiency of transient transfection with cDNAs, or recombinant protein expression. Overall, the mucin coating technology and engineered cell lines described herein exhibit the ability to improve the single-cell growth and viability of suspended cells in bioreactors.

This Part II, as well as other parts of this disclosure, pertain to biopolymers referred to in the art as mucins, which are utilized to reduce adhesion and fouling at biological interfaces. Mucins are characterized by amino acid sequences rich in serine and threonine residues, which are post-translationally modified with *O*-linked pendant glycan structures (Thornton, Rousseau, & McGuckin, 2008). The bottlebrush molecular structure of mucins confers an anti-adhesive characteristic that is used by biological systems for diverse purposes, including antifouling coatings, lubrication, and modulation of cellular interactions (Jay & Waller, 2014; Kuo, Gandhi, Zia, & Paszek, 2018; Paszek et al., 2014). Of the mucin family members, Mucin-1 (Muc1) is recognized as an anti-adhesive protein that can interfere with integrin- and cadherin-mediated cell interactions (Klinken, Dekker, Buller, & Einerhand, 1995; Wesseling, Valk, & Hilkens, 1996; Wesseling, van der Valk, Vos, Sonnenberg, & Hilkens, 1995). The anti-adhesive properties of Mucl are conferred by its large ectodomain, which is heavily *O*-glycosylated during trafficking to the cell surface. Neutral and anionic sugar residues of the glycans can coordinate with water to form a highly hydrated barrier on the cell surface (Gendler & Spicer, 1995).

In this Part II, novel mucin cDNAs and mucins encoded by them are described and used to create a genetically-encoded technology for reduction of aggregation of human-cell host production systems. In particular, the presently described mucin technology is improved, tested, and refined for use, for example, as an anti-adhesive coating on host-cell production systems. As a non-limiting demonstration, we develop new 293-F cell lines with stable anti-

adhesive coatings and evaluate their performance in regards to proliferation rate, cell aggregation, resistance to shear stress, and efficiency of transfection with plasmid DNA.

Materials and Methods

Antibodies and Reagents

The following antibodies were used: Human CD227 (555925, BD Biosciences) (Muc1), β-Actin (sc-4778, Santa Cruz), Goat anti-Mouse IgG-HRP (sc-2005, Santa Cruz). Lectins used were: Biotinylated Peanut Agglutinin (PNA; B-1075, Vector Laboratories), CF568 PNA (29061, Biotium), CF640R PNA (29063, Biotium), CF633 Wheat Germ Agglutinin (WGA; 29024, Biotium). Biotinylated lectins were detected using ExtrAvidin-Peroxidase (E2886, Sigma). To induce transactivator cell lines, doxycycline was used (sc-204734, Santa Cruz). For gentamycin selection, G418 was used (10131035, Thermo Fisher).

Constructs

A tetracycline-inducible, transposon based Piggybac expression vector with an integrated, co-expressed reverse tetracycline transactivator gene (pPB tet rtTA NeoR) was used for stable line generation. The pPB tet rtTA NeoR plasmid was modified by the insertion of the internal ribosome entry site (IRES) of the encephalomyocarditis virus followed by the fluorescent protein copGFP into the NotI and XbaI sites (pPB tet IRES GFP rtTA NeoR). Synthetic cDNAs containing either 21 or 42 tandem repeats (TR) of the amino acid sequence PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8) were codon optimized with codon scrambler (Tang & Chilkoti, 2016), generated through custom gene synthesis (General Biosystems), and cloned in place of the native tandem repeats in pcDNA3.1 Muc1 TM21 – previously described in (Paszek et al., 2014; Shurer et al., 2017) – using the BamHI and Bsu36I restriction sites. The Muc1 gene containing the engineered 21 or 42 tandem repeats was then cloned into the pPB tet IRES GFP rtTA NeoR plasmid using the BamHI and EcoRI sites to generate Muc1 42TR TM21 pPB tet IRES GFP rtTA NeoR and Muc1 21TR TM21 pPB tet IRES GFP rtTA NeoR plasmids used to make the Mucin-270 and Mucin-135 biopolymer cell lines, respectively. To produce the Mucin-0 cell line, the native Muc1 tandem repeats were deleted from the pcDNA3.1 Muc1 TM21 through Q5 site directed mutagenesis with 5'-TGGAGGAGCCTCAGGCATACTTTATTG-3' (SEQ ID NO:14) forward) and 5'-CCACCGCCGACCGAGGTGACATCCTG-3' ((SEQ ID NO:15) reverse) primers. The Muc1 gene with 0TR was then cut from the pcDNA3.1 Muc1 0TR TM21 and cloned into the pPB tet IRES GFP rtTA NeoR plasmid via the BamHI a nd EcoRI sites. The plasmid pLV puro mRuby2 was used for transient transfection experiments with cytoplasmic

red fluorescent protein (RFP). For secreted RFP experiments, SS-mScarlet-I pPB tet IRES GFP rtTA NeoR plasmid was used. To construct this plasmid, the backbone was linearized using BamHI-HF and EcoRI-HF. A dsDNA oligo encoding the Muc1 signal sequence (MTPGTQSPFFLLLLLTVLTVVTGS (SEQ ID NO:26)) fused by a linker (four Glycines followed by a Serine) to mScarlet-I was ordered from Integrated DNA Technologies. This fragment was inserted into the linearized backbone via NEB HiFi Assembly.

Cell Lines and Culture

FreeStyle 293-F Cells were obtained from Thermo Fisher Scientific. Cells were cultured and maintained according to the manufacturer's guidelines in an Eppendorf New Brunswick s41i incubator in Erlenmeyer flasks. Cells were maintained between 0.5 x 10^6 and 3 x 10^6 cells/mL at 120 rpm, 37°C, and 8% CO_2 in FreeStyle 293 Expression Medium (Thermo). Transfections were performed using polyethyleneimine (PEI) as previously reported (Durocher et al., 2002). Genetically-encoded stable cell lines were created by co-transfection of the pPB tet IRES GFP rtTA NeoR plasmids described above with a hyperactive transposase plasmid (Shurer et al., 2017) and subsequently selected with 750 μg/mL of gentamycin for two weeks. Cell proliferation was quantified by cell counting on a hemocytometer with trypan blue exclusion.

Confocal Microscopy

Samples were collected, pelleted at 200 rcf for 5 min, and fixed in 4% paraformaldehyde for 10 minutes at room temperature. Samples were washed three times with PBS. Cells were labeled with 1:1000 CF568 PNA for *O*-glycans and 1:1000 CF633 WGA for the cell membrane in PBS for 30 minutes at room temperature. Samples were washed three times with PBS and imaged on a Zeiss LSM800 with a 63x water immersion objective.

Flow Cytometry Analysis

All samples were measured using live cells, unless otherwise indicated. Cells were harvested from suspension culture, pelleted at 200 rcf for 5 min, and resuspended in 0.5% BSA PBS. Samples were filtered through a 0.22 μm filter cap and analyzed on a BD FACS Aria Fusion. For the doxycycline time-course, cells were induced with 1 μg/mL of doxycycline. Cellular samples from the cultures were taken at the indicated time points, pelleted at 200 rcf for 5 min, and fixed with 4% paraformaldehyde for 10 min at room temperature. Samples were rinsed three times with PBS and stored at 4°C until flow

cytometry analysis. Analysis of all flow cytometry data was performed using FlowJo software.

Immuno- and Lectin Blot Analysis

Cells are inoculated at 0.5 x 10^6 cells/mL and grown overnight, 16-18 hr. Biopolymer expression was then induced with 1 μg/mL doxycycline, and cells were grown with doxycycline for an additional 48 hr. After 48 hr, a sample was taken for each cell line, pelleted at 200 rcf for 5 min before the supernatant was separated, and the cell pellet was lysed by resuspending in RIPA lysis buffer (Abcam), vortexing the sample for 30 seconds, and heating to 98°C for 10 min. Lysates were frozen on liquid nitrogen and stored at -80°C. Lysates were separated on Nupage 3-8% Tris-Acetate gels (Invitrogen) and transferred to PVDF membranes. Membranes were blocked with 3% BSA TBST for 2 hr. Primary antibodies were diluted 1:1000 and lectins were diluted to 1 μg/mL in 3% BSA TBST and incubated on membranes overnight at 4°C. Secondary antibodies or ExtrAvidin were diluted 1:2000 in 3% BSA TBST and incubated for 2 hr at room temperature. Blots were developed in Clarity ECL (BioRad) substrate and imaged on a ChemiDoc (BioRad) documentation system.

PCR amplification of Mucin-270 transgene in the transfected 293F cells

To test for amplification or deletion of stably integrated Mucin-270 cDNAs in 293F genomes, PCR amplification was performed with Q5 Hot start high-fidelity DNA polymerase (New England Biolabs Inc., Ipswich, MA) using extracted genomic DNA as the template. Genomic DNA was extracted with GeneJET genomic DNA purification kit (Thermo Scientific., Waltham, MA). A total of 60 ng of genomic DNA was used for PCR amplification. Primers: Mucin-270 FWD 5'-ATGACACCGGGCACCCAGTC-3' (SEQ NO:85) and Mucin-270 REV 5'-CTACATACTTCGTCGGCGCATGTAC-3' (SEQ NO:86). Size of amplicon is 2994 bp.

Cell Clumping Analysis

Cells were inoculated at 0.75 x 10^6 cells/mL and induced with 1 μg/mL doxycycline after overnight growth (16-18 hr). Cells were then grown to a high cell density for an additional 48 or 72 hr in the presence of 1 μ/mL doxycycline. Cell density was quantified by collecting sample of the culture, mixing thoroughly to dissociate large clumps, and counting viable cells with a hemocytometer and trypan blue exclusion. For imaging, samples were drawn with wide-bore pipette tips to reduce dissociation of large clumps and diluted in PBS to approximately 6.75 x 10^4 cells/cm^2 for imaging in 2D. Phase contrast images were

acquired on an Olympus IX81 microscope with a 10x objective. Fiji was used for image processing (Schindelin et al., 2012). Two independent samples were collected and prepared as technical replicates for imaging with three regions of interest imaged per technical replicate. Three biological replicates were performed. Automated image analysis was performed using custom analysis software adapted from a previous publication (Shurer et al., 2017). Briefly, the analysis software located the center of each circular object. The coordinates of each cell's center were then used to calculate the Ripley's K function in MATLAB. The percent of single cells was calculated by counting the total number of cells which do not have any neighboring cells within 19 μm and dividing by the total number of cells in the image. Similarly, the percent of cells in various cluster sizes was calculated by binning the cells into clusters based on the number of neighboring cells within 19 μm.

To evaluate resistance to calcium induced cell aggregation, cultures were inoculated at 0.5 x 10^6 cells/mL and induced with 1 μg/mL doxycycline after overnight growth (16-18 hr). After 48 hr, cells were resuspended at 4 x 10^6 cells/mL. The culture media was then supplemented with 2 mM $CaCl_2$, 1:300 anti-clumping agent (Thermo Fisher, 0010057AE), or both. Still images and videos of the cell suspension were acquired after 24 hr of treatment by transferring the culture to a glass test tube. The concentration of cells in suspension was determined by collecting duplicate samples from each culture after allowing the largest aggregates to settle out of suspension for 20 seconds. Cell concentration was measured using a hemocytometer and Trypan blue.

Shear Stress Experiments

Cells were inoculated at 0.5 x 10^6 cells/mL, grown overnight (16-18 hr), and induced with 1 μg/mL doxycycline for 48 hr. Using a 5 mL syringe with a 16-gauge needle connected to 6.5 in of 1.02 mm silicon tubing, cell suspensions were sheared by flowing through a 500 μm constriction (Teflon tubing) at a constant force generated by a 1 kg mass applied to a syringe with gravity. Samples were passed through the constriction five times. Cells were then stained with 1 μg/mL CF640R PNA for 15 min at 4°C. Cells were washed with 0.5% BSA PBS three times and then stained with Ethidium homodimer-1 (dead cell stain, Thermo Fisher, L3224). Three biological replicates were performed, with two technical replicates for each biological replicate. Percent dead cells was determined by measuring the fraction of cells that had taken up the dead cell stain on a BD FACS Aria Fusion. A control sample without shear was used to subtract background cell death for each cell line. For Mucin-135

and Mucin-270 cell lines, only PNA positive cells were considered for analysis. Data analysis was performed using FlowJo software.

Transfection Experiments

Cells were inoculated at 0.5 x 10^6 cells/mL, grown overnight (16-18 hr), and induced with 1 μg/mL doxycycline for 48 hr. Cells were then diluted to 2 x 10^6 cells/mL in fresh medium containing 1 μg/mL doxycycline and transfected with 1 μg DNA per 10^6 cells. The next day (16-18 hr post-transfection), cells were diluted 1:1 with fresh medium containing 1 μg/mL doxycycline. To measure transfection efficiency, cells were transfected with the pLV puro mRuby2 plasmid and transfection efficiency was calculated by flow cytometry as the fraction of cells expressing RFP 72 hr post transfection. For production and secretion of recombinant RFP, cells were transfected with SS-mScarlet-I pPB tet IRES GFP rtTA NeoR. After 24 hr, secreted RFP fluorescence in the media supernatant was quantified using a Tecan M1000 Pro plate reader.

Statistical Analysis

Statistical significance was determined by ordinary one-way ANOVA or Student's *t* test (two-tailed) as appropriate using Prism (GraphPad). All graphs were generated in Prism (Graphpad) except for boxplot which were generated in R.

Results

Genetically-Encoded Biopolymers Expressed on the Surface of 293-F Cell Lines

This Part II demonstrates creation of cDNAs that encode Muc1-like biopolymers with transmembrane domains for anchorage to the cell surface. The biopolymer domains consisted of an unstructured protein backbone with 0 – 42 perfect repeats of PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8), which is recognized by the *O*-glycosylation machinery of the endoplasmic reticulum and Golgi apparatus and heavily glycosylated while trafficked to the cell surface. Each biopolymer was targeted to the extracellular space by the native Muc1 signal sequence. The biopolymers were anchored to the cell membrane with a 21-amino acid transmembrane domain (Mercanti et al., 2010; C. R. Shurer et al., 2017). By replacing the native autocatalytic domain of Muc1 (Levitin et al., 2005) with the engineered 21-amino acid transmembrane domain, we mitigated the risk of ectodomain shedding from the cell surface. The described engineered constructs also lacked a cytoplasmic tail to avoid inadvertent transduction of biochemical or physical stimuli by the mucins.

The genetic modification of the 293-F cell line was performed non-virally with an "all-in-one plasmid" that contained all necessary elements for selection and tetracycline-

inducible expression (Fig. 12A). The vector included a tetracycline-responsive promoter for expression of the biopolymer coating and an additional cassette for constitutive expression of the reverse tetracycline transactivator (rtTA-M2) and neomycin-resistance gene (Gossen, Bender, Muller, al, & Freundlieb, 1995). A bicistronic green fluorescent protein (GFP) reporter was also included for visual confirmation of transcription of the mucin cDNA. The cDNA for the biopolymers was stably incorporated into the genome at random locations by transposon mediated integration (X. Li et al., 2013; Wilson, Coates, & George, 2007; Woodard & Wilson, 2015). This approach avoided the use of any viral technology, which poses a serious safety concern in bio-manufacturing (Dumont et al., 2016). We predicted that the modified cells would be coated with a dense, inducible layer of mucin biopolymers on their surface (Fig. 12B).

We tested three different representative biopolymers size for their effects on 293-F cell aggregation. Mucin-like genes with 0, 21, and 42 tandem repeats were constructed. The contour lengths of the polymers with 21 and 42 repeats were predicted to be 135 nm and 270 nm, respectively. We therefore designated the biopolymers Mucin-0, Mucin-135, and Mucin-270 based on the relative length of the biopolymer (Fig. 12C). Because it lacks the large, glycosylated biopolymer domain, the Mucin-0 construct served as a control for any effects related to expression of the transmembrane anchor of the biopolymer.

We confirmed the expression and localization of the biopolymers to the cell surface. Fluorescent microscopy showed expression of the cDNA, reported by the bicistronic GFP signal, and the presence of *O*-glycans on the membrane of cells expressing the Mucin-135 and Mucin-270 semi-synthetic genes (Fig. 13A). We observed a large distribution of biopolymer expression levels, which without intended to be constrained by any particular theory is attributed to the randomized transposition of the cDNAs into the genome (Fig. 2B). Despite the broad distribution, a large portion of the cell populations had stably integrated the cDNA, as shown by the GFP reporter (Fig. 13A-C). The expression and size of the biopolymers was further validated by Western blot (Fig. 13D). Both the Mucin-135 and Mucin-270 could be probed with antibodies against the native Muc1 tandem repeats (Fig. 13D, left). Wild-type (w.t.) cells had no detectable level of endogenous Muc1 expression and no significant *O*-linked mucin-like glycosylation (Fig. 13D). The Mucin-135 and Mucin-270 were heavily glycosylated when expressed. This is shown by the protein bands which are detected above the protein sequence molecular weight when probing with anti-Muc1 antibodies (Fig. 13D, left; predicted molecular weights 81 kDa and 120 kDa for Mucin-135 and Mucin-270, respectively). *O*-glycosylation is further demonstrated by the detection of the

biopolymer with PNA which binds specifically to *O*-linked glycans such as those found on Muc1 (Fig. 13D, right).

No significant difference in cell proliferation rate was observed for any of our biopolymer-coated cell lines (Fig. 13E). We concluded that the additional protein load of our biopolymers did not adversely affect the rapid growth rate of parental 293-F cells. For a stable cell line, we used the well characterized reverse-tetracycline inducible promoter (Gossen et al., 1995) which initiates gene transcription upon addition of doxycycline and halts transcription on withdrawal of doxycycline. This cell line responded as predicted to induction by doxycycline, demonstrating temporal control over expression of the mucin coating (Fig. 13F).

Highly repetitive cDNAs, such as mucins, are reported to have higher frequencies of amplification and deletion in the cellular genome (Gemayel, Vinces, Legendre, & Verstrepen, 2010; Oren et al., 2016). The cDNAs for our Mucin-135 and Mucin-270 constructs were codon optimized to minimize their repetitiveness. We found that the optimized cDNAs were stable when integrated in the host cell genome. Notably, no noticeable amplification or deletion of stably integrated Mucin-270, the largest and most repetitive of our biopolymer cDNAs, was observed after 2 months of cell culture (Fig. 13G).

Biopolymer Coatings Reduced Cell Aggregation

After establishing stable populations, we analyzed whether the biopolymer coatings could reduce cell aggregation in suspension cell cultures. Phase contrast images of the cell lines qualitatively showed more cell aggregates in the w.t. and Mucin-0 cell lines than in the Mucin-135 and Mucin-270 lines (Fig. 14A). Quantification of the fraction of single cells in the sample showed an increase in the percent of single cells for the Mucin-135 and Mucin-270 coatings compared to the w.t. cells, while the Mucin-0 line showed no difference compared to w.t. cells (Fig. 12B, Fig. 19A). Correspondingly, w.t. and Mucin-0 coated cell lines were much more likely to form clusters of two or more cells than Mucin-135 or Mucin-270 cell lines (Fig. 14C, Fig. 19B).

Inspection of phase contrast images of our 293-F lines engineered with Mucin-135 or Mucin-270 revealed that the majority of cells were singlets or doublets with few detectable higher order aggregates (Fig. 14B). Because of the absence of higher order aggregates, we reasoned that the doublets in the Mcuin-135 and Mucin-270 samples may be actively dividing cells or cells that have yet to full disassociate following cytokinesis. The appearance of doublets can also result from single cells randomly settling out of suspension too near each

other to resolve in the 2D plane of the image formed on our microscope. To approximate the frequency of single cells which could randomly settle out of suspension in such a way, we created a simulated dataset of randomly placed centroids and conducted our clustering analysis. On average, the simulated centroids would be counted as singlets 66% of the time. By comparison, 57% of the Mucin-270 cells were singlets (Fig. 14B).

To quantify the extent of cell clustering, we analyzed the spatial distribution of cells in the image using the Ripley's K function, a spatial distribution statistic that counts the frequency at which neighboring particles are found within a given distance of any given particle. Using this statistical tool, we observed that the Mucin-135 and Mucin-270 biopolymers show decreased clustering compared to the w.t. and Mucin-0 cell lines (Fig. 14D, Fig. 19C).

Mucin-270 Coatings Outperformed Commercially Available Anti-Clumping Agent

We found that the Mucin-270 biopolymer coating could reduce cell aggregation even in extreme pro-clumping conditions. Suspension adapted cell lines have previously been shown to significantly aggregate under specific media conditions, such as high calcium concentrations that are known to promote engagement of cadherins (Dee et al., 1997; Han et al., 2006b; Kim, Tai, Mok, Mosser, & Schuman, 2011; Meissner et al., 2001; Peshwa et al., 1993; Sjaastad & Nelson, 1997; Tolbert et al., 1980; Yamamoto et al., 2000; Zanghi et al., 2000). When cultured in high calcium conditions (2 mM $CaCl_2$), the Mucin-270 biopolymer coated cells showed qualitatively less aggregation than w.t. cells (Fig. 15A). Notably, cultures with Mucin-270 biopolymer coatings retained their turbidity in the pro-clumping conditions, whereas unmodified cells assembled into large clusters easily visible to the naked eye (Fig. 15A). Mucin-270-coated cells show a slight decrease in concentration of cells in suspension upon calcium treatment while w.t. cells have essentially no cells remaining in suspension (Fig. 15B).

Further, the Mucin-270 coating outperforms a commercially available anti-clumping agent in highly aggregating conditions. Under high calcium conditions, anti-clumping agent had no discernable efficacy in mitigating cell clumping (Fig. 15A). Addition of commercial anti-clumping agent to Mucin-270 coated cells did not further enhance their resistance to clumping in our assays (Fig. 15B). Together, these results demonstrated the ability of the presently provided genetically-encoded biopolymer coatings to reduce cell aggregation in suspension.

Biopolymer Coatings Provided Resistance to Shear Stress

The sensitivity of suspension-adapted mammalian cells to shear stresses imposes a limit on the rate of mixing and mass transfer in typical bioreactors (Hu, Berdugo, & Chalmers, 2011). Large volume bioreactors operated at high-cell densities require increased mixing to overcome mass transfer limitations (Hu et al., 2011). Thus, cellular sensitivity to shear places another limit on bioreactor productivity. Because protection of ductal epithelial cells to shear stress is a physiological function of mucins, we considered whether, as an added benefit, our biopolymer coatings protect cells from shear stresses. To test this, suspended cells were sheared by passage through a narrow constriction and then analyzed for viability after reintroduction into culture (Fig. 16A). A 1 kg mass was applied to a vertically-oriented syringe to generate a constant and controlled pressure that drove the flow of suspended cells through a 7.6 cm length of 500 μm diameter Teflon tubing. Cell death was analyzed by flow cytometry using a live/dead cell stain. We found that the Mucin-135 and Mucin-270 biopolymer-coated cell lines had significantly greater viability after shearing compared to both w.t. and Mucin-0 cell lines (Fig. 16B), suggesting that the mucin coatings could allow for higher mixing rates in the bioreactor.

Biopolymer Coated Cell Lines can be Transiently Transfected and Produced Comparable Levels of Recombinant Protein

The use of transient transfection of cells for recombinant protein production has recently become of interest to avoid the long development times associated with selection and isolation of stable cell lines for production of new pharmaceuticals (Derouazi et al., 2004; Durocher et al., 2002; Swiech et al., 2011). Given the potential barrier effect of a mucopolysaccharide coating on the cell surface, we tested whether expression of the presently provided biopolymers would affect transfection efficiency of the cell lines. To test, we transiently transfected cell lines with a plasmid for expression of cytoplasmic red-fluorescent protein. We observed no statistically significant difference in the transfection efficiency of the Mucin-0, Mucin-135, or Mucin-270 cell lines compared to the w.t. cells (Fig. 17A). Single-cell analysis revealed similar distributions of recombinant protein production across the engineered and parental cell populations (Fig. 17B). Further, there is no significant difference in the RFP signal of transfected cells, indicating comparable expression of transiently transfected proteins in the different cell lines (Fig. 17C). We also tested the performance of the engineered cells for production of secreted recombinant proteins. As non-limiting example, we fused a signal peptide to the fluorescent protein, mScarlet-I, and

measured production of the secreted protein in medium supernatant from transiently transfected cultures. Mucin-270 coated cells produced the same quantities of secreted recombinant protein as w.t. cells (Fig. 18). Thus, the described biopolymer coatings did not adversely affect transfection efficiency and high protein production rate of the 293-F cell system.

Discussion of Part II

This Part II demonstrates, among other features, that established cell lines can be genetically modified to express engineered mucin biopolymers for anti-adhesion. Expression of these biopolymers does not negatively impact the desirable characteristics of 293-F cells, including their fast proliferation rates (Fig. 12E) and high transfection efficiencies (Fig. 15A, B). Moreover, the expression of the biopolymers significantly reduces undesirable cell clumping (Fig. 14, Fig. 15, Fig. 19) and enhances resistance of the cells to shear forces (Fig. 6). Mucin-135 coating and thicker Mucin-270 coatings performed similarly in head-to-head tests and are expected to be equally well-suited for the applications described herein.

The described biopolymer coatings provide a significant reduction of cell aggregation in serum-free media formulations that are typically used for production in bioreactor formulations. Notably, the coatings could reduce aggregation further even in media formulations that were designed to minimize cell clumping (eg. Invitrogen Freestyle 293-F media). The disclosure includes biopolymer expression on cell aggregation in media formulations that have historically been avoided due to issues of cell aggregation. For example, highly efficient transient transfections have long been performed with DNA-calcium phosphate precipitates (Jordan & Wurm, 2004). However, at the high calcium concentrations required, 293-F cells are known to form large cell aggregates (Meissner et al., 2001; Peshwa et al., 1993). Based on results of this Part II results (Fig. 15), use of the Mucin-135 or Mucin-270 coatings significantly reduce cell aggregation in such conditions for improved protein production from transiently transfected cultures.

The disclosure includes further improvements of the described mucin coating can be achieved through additional optimization of the engineered mucins and their regulated expression. Notably, excessive over-production of highly glycosylated mucin-like proteins could possibly compete with recombinant glycoproteins for the cellular glycosylation machinery and the nucleotide sugar building blocks of glycans. Shedding of the engineered mucins from the cell surface is mitigated by the described selection of a membrane anchor, which lacks a proteolytic cleavage site.

The mucin approached described herein can be employed as a solution for suspension-adapted suspension systems that tend to aggregate in the bio-reactor. But it will be recognized that the ability of these compositions to protect cells and strongly resist clumping could also benefit current bio-manufacturing platforms, like CHO cells, which can still aggregate under non-ideal reactor conditions or in non-optimal media formulations. As bio-manufacturing looks beyond CHO systems for next-generation production platforms that mitigate the risk of non-human glyco-conjugates and other antigenic epitopes, adaptation to growth in suspension remains a significant and time-consuming challenge for human, primate, and many other mammalian cell lines (Amaral et al., 2016; Rodrigues et al., 2013). By promoting cell viability and minimizing aggregation, the presently provided compositions can be expected to help overcome some of the significant barriers to suspension adaptation.

Taken together, this Part II presents a mucin coating technology for improved single-cell growth of cells in suspension. The system was largely successful in mitigating cell aggregation.

References

Amaral, R. L. F. do, Bomfim, A. de S., Abreu-Neto, M. S. de, Picanço-Castro, V., Russo, E. M. de S., Covas, D. T., & Swiech, K. (2016). Approaches for recombinant human factor IX production in serum-free suspension cultures. *Biotechnology Letters*, *38*(3), 385–394.//doi.org/10.1007/s10529-015-1991-1

Carter, P. J. (2011). Introduction to current and future protein therapeutics: A protein engineering perspective. *Experimental Cell Research*, *317*(9), 1261–1269.//doi.org/10.1016/j.yexcr.2011.02.013

Casademunt, E., Martinelle, K., Jernberg, M., Winge, S., Tiemeyer, M., Biesert, L., … Schröder, C. (2012). The first recombinant human coagulation factor VIII of human origin: human cell line and manufacturing characteristics. *European Journal of Haematology*, *89*(2), 165–176.//doi.org/10.1111/j.1600-0609.2012.01804.x

Dee, K. U., Shuler, M. L., & Wood, H. A. (1997). Inducing single-cell suspension of BTI-TN5B1-4 insect cells: I. The use of sulfated polyanions to prevent cell aggregation and enhance recombinant protein production. *Biotechnology and Bioengineering*, *54*(3), 191–205.//doi.org/10.1002/(SICI)1097-0290(19970505)54:3<191::AID-BIT1>3.0.CO;2-A

Derouazi, M., Girard, P., Van Tilborgh, F., Iglesias, K., Muller, N., Bertschinger, M., & Wurm, F. M. (2004). Serum-free large-scale transient transfection of CHO cells. *Biotechnology and Bioengineering*, *87*(4), 537–545.//doi.org/10.1002/bit.20161

Dumont, J., Euwart, D., Mei, B., Estes, S., & Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical Reviews in Biotechnology*, *36*(6), 1110–1122.//doi.org/10.3109/07388551.2015.1084266

Durocher, Y., Perret, S., & Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Research*, *30*(2), e9.

Gemayel, R., Vinces, M. D., Legendre, M., & Verstrepen, K. J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual Review of Genetics*, *44*, 445–477.//doi.org/10.1146/annurev-genet-072610-155046

Gendler, S. J., & Spicer, A. P. (1995). Epithelial Mucin Genes. *Annual Review of Physiology*, *57*(1), 607–634.//doi.org/10.1146/annurev.ph.57.030195.003135

Ghaderi, D., Zhang, M., Hurtado-Ziola, N., & Varki, A. (2012). Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnology & Genetic Engineering Reviews*, *28*, 147–175.

Gossen, M., Bender, G., Muller, G., al, et, & Freundlieb, S. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science*, *268*(5218), 1766.

Han, Y., Liu, X.-M., Liu, H., Li, S.-C., Wu, B.-C., Ye, L.-L., … Chen, Z.-L. (2006a). Cultivation of Recombinant Chinese hamster ovary cells grown as suspended aggregates in stirred vessels. *Journal of Bioscience and Bioengineering*, *102*(5), 430–435.//doi.org/10.1263/jbb.102.430

Han, Y., Liu, X.-M., Liu, H., Li, S.-C., Wu, B.-C., Ye, L.-L., … Chen, Z.-L. (2006b). *Journal of Bioscience and Bioengineering*, *102*(5), 430–435.//doi.org/10.1263/jbb.102.430

Hu, W., Berdugo, C., & Chalmers, J. J. (2011). The potential of hydrodynamic damage to animal cells of industrial relevance: current understanding. *Cytotechnology*, *63*(5), 445–460.//doi.org/10.1007/s10616-011-9368-3

Jay, G. D., & Waller, K. A. (2014). The biology of Lubricin: Near frictionless joint motion. *Matrix Biology*, *39*, 17–24.//doi.org/10.1016/j.matbio.2014.08.008

Jordan, M., & Wurm, F. (2004). Transfection of adherent and suspended cells by calcium phosphate. *Methods*, *33*(2), 136–143.//doi.org/10.1016/j.ymeth.2003.11.011

Kim, S. A., Tai, C.-Y., Mok, L.-P., Mosser, E. A., & Schuman, E. M. (2011). Calcium-dependent dynamics of cadherin interactions at cell–cell junctions. *Proceedings of the National Academy of Sciences*, *108*(24), 9857–9862.//doi.org/10.1073/pnas.1019003108

Klinken, B. J. V., Dekker, J., Buller, H. A., & Einerhand, A. W. (1995). Mucin gene structure and expression: protection vs. adhesion. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *269*(5), G613–G627.

Kuo, J. C.-H., Gandhi, J. G., Zia, R. N., & Paszek, M. J. (2018). Physical biology of the cancer cell glycocalyx. *Nature Physics*, *14*(7), 658–669.//doi.org/10.1038/s41567-018-0186-9

Leader, B., Baca, Q. J., & Golan, D. E. (2008). Protein therapeutics: a summary and pharmacological classification. *Nature Reviews Drug Discovery*, *7*(1), 21–39.//doi.org/10.1038/nrd2399

Levitin, F., Stern, O., Weiss, M., Gil-Henn, C., Ziv, R., Prokocimer, Z., … Wreschner, D. H. (2005). The MUC1 SEA module is a self-cleaving domain. *The Journal of Biological Chemistry*, *280*(39), 33374–33386.//doi.org/10.1074/jbc.M506047200

Li, L., Qin, J., Feng, Q., Tang, H., Liu, R., Xu, L., & Chen, Z. (2011). Heparin Promotes Suspension Adaptation Process of CHO–TS28 Cells by Eliminating Cell Aggregation. *Molecular Biotechnology*, *47*(1), 9–17.//doi.org/10.1007/s12033-010-9306-1

Li, X., Burnight, E. R., Cooney, A. L., Malani, N., Brady, T., Sander, J. D., … Craig, N. L. (2013). piggyBac transposase tools for genome engineering. *Proceedings of the National Academy of Sciences*, *110*(25), E2279–E2287.//doi.org/10.1073/pnas.1305987110

Liu, M., & Goudar, C. T. (2013). Gene expression profiling for mechanistic understanding of cellular aggregation in mammalian cell perfusion cultures. *Biotechnology and Bioengineering*, *110*(2), 483–490.//doi.org/10.1002/bit.24730

Meissner, P., Pick, H., Kulangara, A., Chatellard, P., Friedrich, K., & Wurm, F. M. (2001). Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells. *Biotechnology and Bioengineering*, *75*(2), 197–203.

Mercanti, V., Marchetti, A., Lelong, E., Perez, F., Orci, L., & Cosson, P. (2010). Transmembrane domains control exclusion of membrane proteins from clathrin-coated pits. *J Cell Sci*, *123*(19), 3329–3335.//doi.org/10.1242/jcs.073031

Oren, M., Barela Hudgell, M. A., D'Allura, B., Agronin, J., Gross, A., Podini, D., & Smith, L. C. (2016). Short tandem repeats, segmental duplications, gene deletion, and genomic

instability in a rapidly diversified immune gene family. *BMC Genomics*, *17*.//doi.org/10.1186/s12864-016-3241-x

Park, J. H., Lim, M. S., Woo, J. R., Kim, J. W., & Lee, G. M. (2016). The molecular weight and concentration of dextran sulfate affect cell growth and antibody production in CHO cell cultures. *Biotechnology Progress*, *32*(5), 1113–1122.//doi.org/10.1002/btpr.2287

Paszek, M. J., DuFort, C. C., Rossier, O., Bainer, R., Mouw, J. K., Godula, K., … Weaver, V. M. (2014). The cancer glycocalyx mechanically primes integrin-mediated growth and survival. *Nature*, *511*(7509), 319–325.//doi.org/10.1038/nature13535

Peshwa, M. V., Kyung, Y.-S., McClure, D. B., & Hu, W.-S. (1993). Cultivation of mammalian cells as aggregates in bioreactors: Effect of calcium concentration of spatial distribution of viability. *Biotechnology and Bioengineering*, *41*(2), 179–187.//doi.org/10.1002/bit.260410203

Rodrigues, M. E., Costa, A. R., Henriques, M., Cunnah, P., Melton, D. W., Azeredo, J., & Oliveira, R. (2013). Advances and Drawbacks of the Adaptation to Serum-Free Culture of CHO-K1 Cells for Monoclonal Antibody Production. *Applied Biochemistry and Biotechnology*, *169*(4), 1279–1291.//doi.org/10.1007/s12010-012-0068-z

Sandberg, H., Kannicht, C., Stenlund, P., Dadaian, M., Oswaldsson, U., Cordula, C., & Walter, O. (2012). Functional characteristics of the novel, human-derived recombinant FVIII protein product, human-cl rhFVIII. *Thrombosis Research*, *130*(5), 808–817.//doi.org/10.1016/j.thromres.2012.08.311

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., … Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.//doi.org/10.1038/nmeth.2019

Shukla, A. A., & Thömmes, J. (2010). Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends in Biotechnology*, *28*(5), 253–261.//doi.org/10.1016/j.tibtech.2010.02.001

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

Sjaastad, M. D., & Nelson, W. J. (1997). Integrin-mediated calcium signaling and regulation of cell adhesion by intracellular calcium. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *19*(1), 47–55.//doi.org/10.1002/bies.950190109

Swiech, K., Kamen, A., Ansorge, S., Durocher, Y., Picanço-Castro, V., Russo-Carbolante, E. M., … Covas, D. T. (2011). Transient transfection of serum-free suspension HEK 293 cell culture for efficient production of human rFVIII. *BMC Biotechnology*, *11*, 114.//doi.org/10.1186/1472-6750-11-114

Tang, N. C., & Chilkoti, A. (2016). Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nature Materials*, *15*(4), 419–424.//doi.org/10.1038/nmat4521

Thornton, D. J., Rousseau, K., & McGuckin, M. A. (2008). Structure and Function of the Polymeric Mucins in Airways Mucus. *Annual Review of Physiology*, *70*(1), 459–486.//doi.org/10.1146/annurev.physiol.70.113006.100702

Tolbert, W. R., Hitt, M. M., & Feder, J. (1980). Cell aggregate suspension culture for large-scale production of biomolecules. *In Vitro*, *16*(6), 486–490.//doi.org/10.1007/BF02626461

Tsao, Y. S., Condon, R., Schaefer, E., Lio, P., & Liu, Z. (2001). Development and improvement of a serum-free suspension process for the production of recombinant adenoviral vectors using HEK293 cells. *Cytotechnology*, *37*(3), 189–198.//doi.org/10.1023/A:1020555310558

Vink, T., Oudshoorn-Dickmann, M., Roza, M., Reitsma, J.-J., & de Jong, R. N. (2014). A simple, robust and highly efficient transient expression system for producing antibodies. *Methods*, *65*(1), 5–10.//doi.org/10.1016/j.ymeth.2013.07.018

Wesseling, J., Valk, S. W. van der, & Hilkens, J. (1996). A mechanism for inhibition of E-cadherin-mediated cell-cell adhesion by the membrane-associated mucin episialin/MUC1. *Molecular Biology of the Cell*, *7*(4), 565–577.//doi.org/10.1091/mbc.7.4.565

Wesseling, J., van der Valk, S. W., Vos, H. L., Sonnenberg, A., & Hilkens, J. (1995). Episialin (MUC1) overexpression inhibits integrin-mediated cell adhesion to extracellular matrix components. *The Journal of Cell Biology*, *129*(1), 255–265.

Wilson, M. H., Coates, C. J., & George, A. L. (2007). PiggyBac transposon-mediated gene transfer in human cells. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *15*(1), 139–145.//doi.org/10.1038/sj.mt.6300028

Woodard, L. E., & Wilson, M. H. (2015). piggyBac-ing models and new therapeutic strategies. *Trends in Biotechnology*, *33*(9), 525–533.//doi.org/10.1016/j.tibtech.2015.06.009

Wurm, F., & Bernard, A. (1999). Large-scale transient expression in mammalian cells for recombinant protein production. *Current Opinion in Biotechnology*, *10*(2), 156–159.//doi.org/10.1016/S0958-1669(99)80027-5

Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology*, *22*(11), 1393–1398.//doi.org/10.1038/nbt1026

Yamamoto, S., Matsuda, H., Takahashi, T., Xing, X.-H., Tanji, Y., & Unno, H. (2000). Aggregate formation of rCHO cells and its maintenance in repeated batch culture in the absence of cell adhesion materials. *Journal of Bioscience and Bioengineering*, *89*(6), 534–538.//doi.org/10.1016/S1389-1723(00)80052-3

Zanghi, J. A., Renner, W. A., Bailey, J. E., & Fussenegger, M. (2000). The Growth Factor Inhibitor Suramin Reduces Apoptosis and Cell Aggregation in Protein-Free CHO Cell Batch Cultures. *Biotechnology Progress*, *16*(3), 319–325.//doi.org/10.1021/bp0000353

Zhu, J. (2012). Mammalian cell protein expression for biopharmaceutical production. *Biotechnology Advances*, *30*(5), 1158–1170.//doi.org/10.1016/j.biotechadv.2011.08.022

Part III

This Part III provides representative and non-limiting approaches to stable recombinant production of codon-scrambled lubricin and mucin in human cells, and characterization of modified lubricins derived from human, equine and canine sequences. This Part III demonstrates exploitation of codon redundancy to encode desired polypeptides with minimal nucleotide repetition. The codon-scrambling strategy was applied to generate synonymous genes, or "synDNAs," for two representative mucins of commercial interest: lubricin and Muc1. Stable, long-term recombinant production in suspension-adapted human 293-F cells was demonstrated for the synonymous lubricin cDNA, which is referred to herein from time to time as "SynLubricin." Under optimal conditions, a 293-F sub-population produced recombinant SynLubricin at more than 200 mg/L of media and was stable throughout two months of continuous culture. Functionality tests confirmed that the recombinant lubricin could effectively inhibit cell adhesion and lubricate cartilage explants. Together, this Part III provides, among other apsects, a viable workflow for cDNA design and stable mucin production in mammalian host production systems.

Part III Introduction

As will be recognized from the foregoing description, mucins are membrane-bound or secreted glycoproteins containing a variable number of tandem repeats that are defined by

their densely clustered sites for *O*-glycosylation (Hang & Bertozzi, 2005). This extensive glycosylation gives rise to a bottlebrush molecular structure that confers mucins with remarkable physical properties (Kuo, Gandhi, Zia, & Paszek, 2018). Mucins at biological interfaces can coordinate with water molecules to form hydrated layers that protect delicate cellular or tissue structures, deter biofouling, and resist pathological cellular deposition (Hattrup & Gendler, 2008). For instance, transmembrane mucins such as Muc1 and Muc16 are densely grafted on the ocular surface, where they maintain hydration, resist abrasion, and provide a selective barrier to macromolecules (Gipson, Spurr-Michaud, Tisdale, & Menon, 2014; Mauris & Argüeso, 2012) Similarly, the secreted mucin-like glycoprotein called proteoglycan 4 (PRG4), or lubricin, can bind to cells and tissue interfaces, including the articular cartilage and ocular surfaces, enabling low friction lubrication and protection from pathological cellular deposition and biofouling (Rhee et al., 2005; Schmidt, Sullivan, Knop, & et al., 2013).

Alterations in mucin expression and glycosylation are observed in various pathological conditions, ranging from cancer and inflammatory bowel disease to ocular disease (Dhanisha, Guruvayoorappan, Drishya, & Abeesh, 2018). Patients with genetic mutations that preclude functional lubricin synthesis demonstrate symptoms of Camptodactyly-Arthropathy-Coxa Vara-Pericarditis (CACP) syndrome, including early-onset polyarthropathy as a result of pannus formation and impaired joint lubrication (Bahabri et al., 1998; Marcelino et al., 1999). Decreased synovial fluid lubricin concentrations have also been observed in patients with anterior cruciate ligament injury, osteoarthritis, and rheumatoid arthritis (Elsaid et al., 2008; Kosinska et al., 2015). As such, there has been significant interest in the development of recombinant lubricin and other mucins as injectable therapeutics for osteoarthritis and rheumatic diseases (Le Graverand-Gastineau, 2010)and as topical treatments for chronic dry eye and other conditions that require application of exogenous lubricants (Schmidt et al., 2013).

Despite this commercial interest, recombinant production has proven challenging for Muc1, lubricin, and other mucins that contain a high number of tandem repeats. Although highly productive clones of Chinese Hamster Ovary (CHO) cells have been isolated for a truncated Muc1 with approximately 1/3 of its native tandem repeats, similar attempts to isolate clones for full-length recombinant Muc1 have failed (Backstrom et al., 2003). Likewise, stable clones for recombinant lubricin with the complete 76-78 native tandem repeats produced the glycoprotein at low levels (Jones et al., 2007), but a modified recombinant lubricin protein construct (LUB:1), which contained only 1/3 of the tandem

repeats, was more amenable to large scale production (Flannery et al., 2009). More recently, the production of full-length recombinant human lubricin expressed in suspension-adapted CHO cells has been reported and has demonstrated potential as an ocular lubricant for treating dry eye disease or hydrating contact lenses (Samsom et al., 2014). The precise details of how recombinant production was achieved for the full-length lubricin remain proprietary, and at the time of filing of this application or patent, it is believed no published strategy for large-scale lubricin production is available.

The exact biology that underlies the difficulty of producing mucins at high levels remains unclear. However, long, repetitive DNA sequences, such as those common in the cDNAs of mucin tandem repeats, are relatively unstable in the cellular genome (Pearson, Edamura, & Cleary, 2005). The fidelity of nearly all DNA processing steps can be compromised by slippage and other errors linked to repetitive sequences (López Castel, Cleary, & Pearson, 2010). Consequently, repeats can mutate by addition or loss of their unit nucleotide sequence up to 100,000 times more frequently than point mutations in non-repetitive regions (Oren et al., 2016). The variation in tandem repeat numbers for Muc1 and other mucins in humans and mammals provides an evolutionary argument that these genomic cDNAs are mutational hotspots (Gemayel, Vinces, Legendre, & Verstrepen, 2010). Recombination and truncation of exogenous Muc1 cDNAs in bacteria have also been reported, suggesting a high level of instability for these repetitive sequences in host microbial cells, as well (Backstrom et al., 2003).

Now that advances in custom gene synthesis (CGS) enable fast and cost-effective synthesis of long cDNAs (Kosuri & Church, 2014), a new approach to providing improved genomic stability of mucins is provided herein, and in certain embodiments exploits codon redundancy to identify and use synonymous gene sequences that are less repetitive but encode the same desired polypeptide. Such codon optimization algorithms have been developed and successfully applied for elastin-like proteins and some other repetitive protein domains (Tang & Chilkoti, 2016). However, it is believed that, prior to the present disclosure, optimized synthetic cDNAs had not been designed, synthesized and tested for bio-manufacturing of large mucins of commercial interest.

Also, prior to the present disclosure, most biologics, including mucins, have been produced in CHO cells due to their fast growth, adaptability to suspension culture, and capacity for glycosylation and other important post-translational modifications. However, CHO cells can generate glycan epitopes that are now suspected to elicit adverse immunological responses in humans (Butler & Spearman, 2014). Namely, the α1,3-

galactosyltransferases of CHO and other non-primate cells produce glycans with Galα1,3-Gal residues that can be immunogenic to humans, apes, and other old-world monkeys that have lost α1,3-galactosyltransferase activity (Bosques et al., 2010; Brooks, 2004). CHO cells also can generate Neu5Gc, a terminal sialic acid that is common in most mammalian cells but has been lost in humans and primates (Ghaderi, Zhang, Hurtado-Ziola, & Varki, 2012). These glycans are of particular concern for recombinant mucins, which can consist of 75% or more carbohydrate by mass and are often highly sialylated (Estrella, Whitelock, Packer, & Karlsson, 2010). Recombinant production of the glycoproteins in human cells would avoid the risk of Galα1,3-Gal and Neu5Gc residues; but, it is beleieved that prior to the present disclosure, no successful attempts at large-scale mucin production in a human cell host production system has been reported.

Thus, the present disclosure demonstrates, in addition to other aspects, that cDNA optimization through codon scrambling is an effective strategy to achieve stable recombinant production of mucins and mucin-like glycoproteins, and that this strategy is viable in suspension-adapted human 293-F cells. Notably, the United States Food and Drug Administration (FDA) has recently approved several biologics produced in 293-F cells, establishing the cell platform as a viable alternative to CHO and other non-human systems for manufacturing specialized therapeutics (Dumont, Euwart, Mei, Estes, & Kshirsagar, 2016). In this disclosure, the codon-scrambling approach is demonstrated for Muc1 and lubricin, and the production strategy is further developed to achieve stable production of a functional, full-length recombinant lubricin. It will be recognized by those skilled in the art, when given the benefit of the present disclosure, the presently described approaches can be used for stable and robust expression of other mucins and mucin-like proteins.

Part II Results

Design and synthesis of cDNA for synonymous lubricin

As an approach for recombinant mucin production, we applied a codon-scrambling and optimization strategy to design synthetic mucin cDNAs within minimal codon repetition (Fig. 20A). A global codon optimization algorithm was applied to find the least repetitive gene sequence that encoded the desired mucin tandem repeats (Tang & Chilkoti, 2016). To tailor the sequences for production in a human host system, such as 293-F, a subsequent optimization was conducted to replace any codons with less than 10% usage frequency in humans (Fig. 20A). We envisioned that the optimized mucin cDNAs could be synthesized

through rapid and low-cost services for CGS (Kosuri & Church, 2014; Tang & Chilkoti, 2016).

We first tested this approach for human lubricin, which has approximately 59 tandem repeats with a consensus sequence of KXPXPTTX (SEQ ID NO:87), with KEPAPTTP (SEQ ID NO:1) being the most frequent repeat. For our synthetic lubricin, we optimized the codons for 59 perfect repeats of the KEPAPTTP (SEQ ID NO:1) consensus sequence (Fig. 20B). The protein sequence for the perfect repeats had approximately 88% similarity to the native human PRG4 repeats (Fig. 20C). The synthetic tandem repeats were flanked by additional sequences encoding the native N- and C- termini of human PRG4. These sequences included the native somatomedin and hemopexin domains of lubricin. We also included an IgK leader sequence, 6x histidine tag, and N-terminal SumoStar tag to aid in protein secretion and purification (Fig. 20B). We named the new semi-synthetic gene encoded by the codon-optimized cDNA "synonymous lubricin" or "SynLubricin."

The nucleotides encoding SynLubricin were significantly less repetitive than native PRG4. We analyzed the nucleotide sequences with an alignment algorithm that detects tandem repeats and scores their degree of repetitiveness based on how frequently they repeat and how closely the identified consensus matches the nucleotides of the queried sequence (Benson, 1999). The detected repeats were aligned with the queried sequence through a Smith-Waterman style local alignment, and the overall repetitiveness was scored by assigning +2 for each nucleotide match and -7 for each mismatch or indel (Benson, 1999). Thus, a higher score was indicative of more nucleotide repetition. The tandem repeats of SynLubricin had a modest score of 168, whereas the native PRG4 repeats had a much higher repetition score of 1001. The present disclosure encompasses such sequences, wherein the overall repetitiveness score of a polynucleotide is compared to a suitable control.

We also aligned the amino acids of the SynLubricin tandem repeats to the 59 tandem repeats of human PRG4 isoform A (Fig. 20D). We noted that the perfect repeats of SynLubricin and the native repeats of human PRG4-A have similar compositions of alanine, glutamic acid, lysine, and threonine, while proline content is slightly higher in the SynLubricin repeats (37% vs 30.5%; Part III Table S1). We also noted that the native repeats contain small amounts of asparagine (0.2%), aspartic acid (0.4%), glycine (0.8%), isoleucine (0.2%), leucine (1.4%) and serine (2.6%), which are not contained in SynLubricin (Part III Supplemental Table 1). Thus, in addition to a distinct coding sequence, the amino acid sequence of SynLubricin is distinct from that of human PRG4.

The low-repetition of nucleotides in the SynLubricin gene enabled synthesis of the desired cDNA using available techniques. We also had a cDNA for the native human lubricin/PRG4 sequence through a commercial venodr. However, our attempts to subsequently clone the native PRG4 cDNA sequence into a mammalian expression vector and recombinantly express the product in mammalian cells failed. Consequently, we discontinued further efforts at recombinant production of lubricin with the full-length, native cDNA.

Efforts to produce SynLubricin in transiently transfected mammalian cells were successful. The SynLubricin cDNA was fused to a bicistronic copGFP reporter and transiently transfected into adherent human embryonic kidney 293-T cells. The protein product of the SynLubricin gene was highly glycosylated, as desired, and exhibited the anti-adhesive properties that we predicted. Transfected cells maintained large gaps between cells in the monolayer, particularly at locations where visible copGFP fluorescence reported high expression levels of the bicistronic mRNA (Fig. 26A). We noted that these observations were consistent with the known anti-adhesive functionality of native lubricin (Rhee et al., 2005). In contrast, mock transfected cells grew to a highly confluent monolayer in culture (Fig. 26A). A western blot of the media supernatant from the SynLubricin-transfected cultures revealed a high molecular weight protein of approximately 460 kDa, which was similar in size to the native lubricin that we detected in equine synovial fluid (Fig. 26B). The expected molecular weight of the peptide backbone of SynLubricin was 145 kDa, indicating that SynLubricin was extensively glycosylated.

We next developed strategies for stable production of the synthetic mucins in 293-F suspension cultures. In one embodiment, we created a non-viral transposon vector for “all-in-one” inducible expression of mucins. The vector contained a tetracycline-responsive promoter for inducible expression of the desired gene and a bicistronic copGFP reporter. The vector also contained a second cassette under control of an EF1alpha promoter for expression of the rtTA-M2 tetracycline transactivator and a bicistronic neomycin resistance gene for selection (Fig. 20E). To test the performance of the expression system, we cloned mCherry2 into the vector and transfected 293-F cells with cationic polyethylenimine (PEI) condensates following standard protocols (Boussif et al., 1995; de los Milagros Bassani Molinas, Beer, Hesse, Wirth, & Wagner, 2014; Sonawane, Szoka Jr., & Verkman, 2003). Stable cell populations were isolated after two weeks of selection, and mCherry2 production was validated by flow cytometry. Based on the flow cytometric analysis, we found that stable

cells produced high levels of mCherry2, and that the fluorescence readout of the copGFP reporter was generally a good indicator of recombinant protein production (Fig. 27).

Design and synthesis of cDNA for synonymous Muc1

We tested whether the described strategy for mucin-type cDNAs was generalizable and could be applied to other mucins. We chose the mucin Muc1, which is important in the hydration and protection of the cornea and other epithelial surfaces (Mantelli & Argüeso, 2008). We noted that the native tandem repeats of Muc1 are polymorphic, with 42 perfect repeats being most frequent in humans (Nath & Mukherjee, 2014). We applied the codon optimization strategy to design a cDNA for 42-perfect Muc1 repeats, PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8). The optimized sequence was fused to the codons for the native N-terminus of human Muc1. We also added the IgK leader sequence, 6x histidine tag, and SumoStar tag, similarly to SynLubricin (Fig. 28A). We calculated a very high repetition score of 4997 for the nucleotide coding sequence of the native human Muc1 tandem repeats. The repetition score was reduced to 220 in our synthetic cDNA, which we referred to as SynMuc1 (Fig. 28B).

The optimized coding sequence for SynMuc1 was synthesized through standard CGS services, whereas efforts to synthesize the extremely repetitious sequence of the native Muc1 cDNA were not able to be carried out by commercial vendors. The custom synthesized SynMuc1 cDNA was transfected into 293-F cells. The recombinant protein was purified from the media supernatant via immobilized metal affinity chromatography (IMAC) and detected by Western blot with an antibody against the native human tandem repeats (Fig. 28C). The recombinant mucin was extensively *O*-glycosylated, as indicated by the strong signal when probed with peanut agglutinin (PNA), a lectin that is specific for a core-1, mucin-type disaccharide (Fig. 28D).

During purification, we noticed that a significant percentage of the mucin failed to bind to the IMAC resin and was detected in the flow through (Fig. 28C, D). Western blotting confirmed the presence of the 6x-histidine SumoStar purification tag on the recombinant protein in the flow through and eluted fractions, suggesting that the N-terminus and purification tag were present but inaccessible to the immobilized IMAC cations as would be the case, for example, if the tag was buried in the random coil of the mucin biopolymer (Fig. 28E). Since an objective was to demonstrate the production of the recombinant SynMuc1 and not optimize its purification, alternative chromatography approaches were not explored.

Stable host production of recombinant SynLubricin

Using a transposon system, we tested its application for SynLubricin production (Fig. 21A). Unexpectedly, we found that after selection with G418, comparatively few cells exhibited high copGFP reporter levels following doxycycline induction (Fig. 21B). To overcome the issue, we applied a two-round sorting strategy using the copGFP reporter to isolate a sub-population of cells that expressed SynLubricin at high levels. Stable cells were expanded and sorted for the top 5% copGFP expressers, which were then expanded and sorted a second time for the top 10% expressers. We found that the sorting strategy improved SynLubricin production 15-fold and did not impact the molecular weight of the glycosylated protein product (Fig. 21B, C). The sorted cell populations displayed noticeably higher levels of the copGFP reporter after induction with doxycycline, indicating successful isolation of a polyclonal population with higher gene expression levels.

To confirm the cDNA stability of the integrated SynLubricin gene in our stable 293-F cells, genomic DNA was extracted from modified 293-F cells after two months of continuous culture. The SynLubricin cDNA was then amplified by polymerase chain reaction (PCR) using primers that were specific to SynLubricin (Fig. 22). The amplified gene was approximately 4 kb in length, as expected for full-length lubricin, and indistinguishable in size from similarly amplified genes obtained using the original SynLubricin plasmid as the template or DNA extracted from transiently transfected cells (Fig. 22). Even after culture for 2 months, the polyclonal cell population exhibited no indications of SynLubricin gene application or deletion, indicating a high level of genomic stability (Fig. 22).

Optimization of SynLubricin production

We analyzed whether SynLubricin productivity could be improved through addition of the histone deacetylase inhibitor, valproic acid (VPA), which has previously been shown to drastically increase production of some recombinant proteins in 293-F cells (Backliwal et al., 2008). Our sorted cell population was induced with doxycycline in the presence or absence of 3.5 mM VPA, and media supernatants were sampled each subsequent day from batch cultures. The molecular weights of the protein products were similar, suggesting that VPA did not appreciably affect the total extent of glycosylation of the protein product (Fig. 23A). Interestingly, the recombinant protein levels peaked at approximately 2-3 days post-induction in cultures without VPA and declined rapidly thereafter (Fig. 23B). In VPA treated cultures, SynLubricin levels in the media did not decline as significantly over time. We ruled out protein degradation as a likely explanation for the decline of recombinant protein in cultures without VPA, since we saw no prominent degradation products for lubricin on

Western blots (Fig. 23A). We instead considered the possibility that the 293-F culture might consume the recombinant protein in conditions of reduced nutrient availability. Consistent with this possibility, we observed that the decline in recombinant protein levels coincided with the depletion of glucose in the cultures without VPA (Fig. 23C). Metabolic activity largely ceased in VPA treated cultures after 3 days, as indicated by a sharp decline in glucose consumption (Fig. 23C). Thus, VPA may prevent the loss of recombinant protein in batch cultures through slowing 293-F cellular metabolism.

We next scaled up production to 1-liter bioreactors operated in batch mode and conducted two independent production runs with VPA added. Each production run yielded plentiful recombinant protein that was comparable in molecular weight to both recombinant protein isolated from transiently transfected cultures and native lubricin detected in equine synovial fluid (Fig. 23D). An ELISA using purified bovine lubricin as a standard reported approximately 200 mg/L of SynLubricin in the batch runs with our stable 293-F lines. Less than 50% of the stable cell population showed strong expression of the copGFP reporter in the batch bioreactors, suggesting that increases in productivity could likely be achieved with clonal expansion of the production cell line (Fig. 21D). We noted that a limitation of our ELISA-based quantification may be the use of a bovine standard, which may over- or under-estimate SynLubricin levels.

We tested whether stable protein production could be achieved with periodic media changes to avoid nutrient depletion. Conditioned media was harvested from doxycycline-induced cultures that were maintained for 10 consecutive days in the absence of VPA. Media in the batch cultures was exchanged every 48 hrs to replenish nutrients and remove metabolic waste products. Viable cell concentration was also reduced to 1 x 10^6 cells/mL every 48 hrs. SynLubricin production levels were stable over the 10 days of culture, and the SynLubricin molecular weight was constant, indicating that glycosylation was also stable (Fig. 23E). While there appears to potentially be a slight decrease in SynLubricin production with time, there is no significant difference in protein yield (Fig. 23F).

SynLubricin is a functional biolubricant

Recombinant SynLubricin was effectively purified from conditioned 293-F media supernatant using= either anion exchange or cation exchange chromatography. Anion-exchange chromatography followed our previously reported strategy for isolation of native lubricin from equine synovial fluid, with slight modification from using DEAE-Sepharose® to using Q Sepharose® (Reesink et al., 2016). Success purification with cation exchange

purification was achieved on a column of POROS™ XS (ThermoFisher) resin with a mobile phase of 50 mM phosphate buffer, 100 mM NaCl, pH 6.8, and a linear elution gradient from 0.1 to 1 M NaCl in 50 mM phosphate buffer, pH 6.8. We also attempted IMAC to purify the native lubricin, but the recombinant SynLubricin had poor affinity to IMAC resins (Fig. 29). As for SynMuc1, we reasoned that the N-terminal histidine-tag could be buried in the large, random coil of the SynLubricin tandem repeats and abandoned the IMAC approach. In contrast, SynLubricin bound to the anion-exchange resin strongly and eluted continuously over high salt concentrations ranging from approximately 350 mM to 1.5 M (Fig. 24A, B). The continuous elution of SynLubricin was likely explained by a varying frequency of anionic sialic acids in the *O*-glycans of the recombinant SynLubricin (Estrella et al., 2010). We found that a stringent wash step of approximately 500 mM NaCl could remove most protein contaminants detectable by silver stain, although some SynLubricin was inevitably lost to this high-salt wash (Fig. 24C, D).

To ensure functionality of our recombinant SynLubricin, we tested its ability to lubricate cartilage and reduce friction. Recombinant SynLubricin was purified via anion exchange chromatography using the stringent 500 mM NaCl wash step to eliminate most protein contaminants (Fig. 24D). Following purification, SynLubricin was dialyzed in saline and diluted to physiological concentrations. Lubrication was tested on bovine articular cartilage explants where the native lubricin boundary layer had been extracted using a custom linear reciprocating tribometer (Jones et al., 2007). Compared to a saline control, we found that SynLubricin-containing solutions, as well as control synovial fluid, significantly reduced the boundary friction of cartilage explants (Fig. 25; $p < 0.001$ and 0.0001, respectively).

We also tested a small quantity of a second SynLubricin sample that was purified without the stringent wash of the anion exchange column with 500 mM NaCl. Notably, cartilage friction coefficients were markedly lower for this SynLubricin preparation than any of the measured friction coefficients for the more stringently washed SynLubricin preparations (Fig. 25). Low sample volume for the unwashed SynLubricin preparation hindered obtaining enough independent measurements for meaningful statistical comparisons (Fig. 25). However, further optimization of purification conditions using techniques that will be apparent to those skilled in the art, given the benefit of this disclsoure, are expected to produce recombinant lubricin fractions with improved performance in biolubrication. For instance, less negatively charged lubricin fractions that elute at lower salt concentrations (350-500 mM NaCl) are important for cartilage biolubriction either by acting independently or in synergy with more negatively charged lubricin fractions. Alternatively, contaminants

that are eliminated with the 500 mM NaCl wash might act synergistically with lubricin in cartilage lubrication.

This Part III exmaple provides an approach to larger-scale, mucin bio-manufacturing. Success in the design and synthesis of new semi-synthetic genes for both Muc1 and lubricin, combined with our success in isolating highly stable, lubricin-expressing cell populations, indicates that this approach may be broadly applicable for recombinant mucins with long, repetitive domains. The successful demonstration of recombinant production in a human cell system that avoids the risk of immunogenic Galα1,3-Gal and Neu5Gc epitopes. We find that the recombinant product of our SynLubricin gene is functional in its ability to resist cellular adhesion (Fig. 26A) and lubricate biological surfaces, such as cartilage (Fig. 25). Thus, SynLubricin can be expected to be suitable for diverse applications ranging from injectables for osteoarthritis to topical treatments for chronic dry eye. Moreover, given the speed and low cost of CGS, the approach described herein can be expected to be applied to rapidly prototype designer mucins with new or modified functional domains.

In addition to the foregoing, we tested SynLubricin for various properties.

As shown in Fig. 30, SynLubricin exhibits a remarkable and unexpected half-life *in vivo*. Results shown in Fig. 30 were obtained as follows. In particular, SynLubricin exhibits an intra-articular half-life when injected into a mammal of well over the 4 days that has been a previously determined value for native lubricin. (See, for example, Hurtig, et al. *Two compartment pharmacokinetic model describes the intra-articular delivery and retention of rhprg4 following ACL transection in the Yucatan mini pig*. J Orthop Res. Feb;37(2):386-396. doi: 10.1002/jor.24191. Epub 2018 Dec 17). Thus, in embodiments, a recombinant protein of this disclosure exhibits a half-life that is more than four days, and can last at least up to about 30 days, or longer. In embodiment, the half-life is up to 50 days, at least 50 days, or longer.

The results shown in Fig. 30 were obtained as follows. All animal protocols were approved by the Cornell University Institutional Animal Care and Use Committee (Protocol #2017-0084). Male Sprague-Dawley rats were purchased from Harlan Sprague-Dawley, Inc. (ENVIGO) at 10-12 weeks of age and were housed in pairs under a standard 12-hour light/dark cycle starting at 6 am. The animals were allowed to move freely in their cages, fed a commercial diet formulated without alfalfa to minimize background fluorescence (ENVIGO Teklad #2918), and allowed access to tap water. After a minimum of 3 days of acclimation upon arrival, animals were identified via ear notching and were weighed. Under isoflurane anesthesia (1-1.5L/min O_2 with 2.5% isoflurane), hair was shaved from the ventral and lateral aspects of each rat using a beard trimmer. After a sterile povidone-iodine and 70%

ethyl alcohol skin preparation, the left knee was injected with 20 μL of SynLubricin-Cy7.5 or 20 μL dextran500kD-Cy7.5 in Dulbecco's phosphate buffered saline (dPBS) via a patellar tendon approach using a 27g needle and 0.5mL syringe (Becton Dickinson) and with the knee in flexion at a 90° angle. Any residual povidone-iodine was removed following injection with ethyl alcohol. No injections were performed in the right knee so that it could be used as an internal control for background fluorescence calculations.

Each rat was imaged using an IVIS Spectrum whole animal imaging system (PerkinElmer™) at 0, 6 and 12 hours and 1, 2, 3, 5, 7, 14, 21, 28 days, up to 56 days post-injection. Both auto and 2 sec exposure times were obtained. Animals were anesthetized under isoflurane anesthesia (1-1.5L/min O_2 with 2.5% isoflurane), and hair was shaved at weekly intervals immediately prior to imaging, beginning at 7 days post-injection. Two animals were injected with 20 μL of SynLubricin-Cy7.5, and four animals were injected with 20 μL of dextran 500kD-Cy7.5 as an additional control. Data was fitted to a bi-exponential decay model to calculate the alpha and beta decay constants. The half-life of lubricin of about 45 days is reported as ln(2) divided by the beta decay constant. In contrast, dextran was cleared rapidly from the rat knee.

As shown in Fig. 31, SynLubricin produced in 293-F cells is contains a mix of Core I and Core II glycans. Analysis of glycans was performed in the positive ion mode by MALDI-TOF/TOF-MS with assignment of glycan structures done manually by using Glycoworkbench software. Notably, the Core II glycans comprised 20.3% of detected Core O-glycans. More than 29% of the O-glycan structures were sialylated.

We conducted an additional tribiological analysis to determine the dependence of cartilage-on-cartilage friction on SynLubricin concentration. As shown in Fig. 32, as little as 100 μg/mL SynLubricin in PBS effectively lowers the friction coefficient (sliding speed = 0.1 mm/s) for cartilage as compared to PBS only control. All data were obtained using a custom linear reciprocating tribometer with cylindrical cartilage explants (6mm diameter Å~ 2mm thickness) that were harvested from the femoral condyles of neonatal bovine stifles. Endogenous cartilage-bound lubricin was extracted from the explants using a 30 min incubation in 1.5M NaCl, followed by a 1 hr equilibration step in PBS, prior to testing. SynLubricin for these studies was purified from conditioned 293-F media using cation exchange chromatography with POROS™ XS (ThermoFisher) resin.

MATERIALS AND METHODS

Antibodies and Reagents

The following antibodies were used: mouse anti-human CD227 (555925, BD Biosciences) (Muc1), mouse anti-human lubricin (MABT401, EMD Millipore), goat anti-mouse IgG-HRP (sc-2005, Santa Cruz), mouse anti-SUMO (4G11E9, GenScript). Lectins used were biotinylated Peanut Agglutinin (PNA; B-1075, Vector Laboratories). Biotinylated lectins were detected using ExtrAvidin-Peroxidase (E2886, Sigma). To induce transactivator cell lines, doxycycline was used (sc-204734, Santa Cruz). For neomycin selection, G418 was used (10131035, Thermo Fisher). Valproic acid (VPA) was used as a histone deacetylase inhibitor (Sigma P4543-100G).

Constructs

A tetracycline-inducible, transposon based Piggybac expression vector with an integrated, co-expressed reverse tetracycline transactivator gene (pPB tet rtTA NeoR) was used for stable line generation. The pPB tet rtTA NeoR plasmid was modified by the insertion of the internal ribosome entry site (IRES) of the encephalomyocarditis virus followed by the fluorescent protein copGFP into the NotI and XbaI sites of the plasmid (pPB tet IRES copGFP rtTA NeoR). Synthetic cDNA for a lubricin analog with 59 perfect repeats of KEPAPTTP (SEQ ID NO:1), native N-and C-terminal domains, and an N-terminal SumoStar tag (lifesensors) were generated through custom gene synthesis (General Biosystems) and cloned into the multiple cloning site of pPB tet IRES copGFP rtTA NeoR using BamHI and EcoRI restriction sites. Similarly, cDNA for a soluble, codon-scrambled Muc1 having 42 perfect repeats of PDTRPAPGSTAPPAHGVTSA (SEQ ID NO:8) and a native human Muc1 N-terminus with SumoStar tag was generated by custom gene synthesis in the pcDNA3 plasmid. For construction of an mCherry2 IRES2 copGFP expression plasmid, an mCherry2 cDNA was isolated by EcoRI and NotI digestion of pmCherry2 N1 and cloned into the EcoRI and NotI digested pPB tet IRES copGFP rtTA NeoR vector to create pPB tet mCherry2 IRES copGFP rtTA NeoR.

Cell Lines and Culture

FreeStyle 293-F (293-F) cells were obtained from Thermo Fisher Scientific. Cells were cultured and maintained according to the manufacturer's guidelines in 100-ml Wheaton Celstir glass spinner flasks. Cells were maintained between 0.5 x 10^6 and 3 x 10^6 cells/mL at 120 rpm, 37°C, and 8% CO_2 in FreeStyle 293 Expression Medium (Thermo). 293-F transfections were performed using polyethyleneimine (PEI) as previously reported

(Durocher, Perret, & Kamen, 2002). Stable cell lines were created by co-transfection of the pPB tet IRES copGFP rtTA NeoR plasmids described above with a hyperactive transposase plasmid (Shurer et al., 2018) and subsequently selected with 750 μg/mL of G418 for two weeks. Human embryonic kidney cells transformed with the SV40 large T antigen (293-T; ATCC) were maintained in high-glucose DMEM supplemented with 10% fetal bovine serum and penicillin/streptomycin. 293-T cells were transfected through a standard calcium phosphate transfection protocol. Cell proliferation was quantified by cell counting on a hemocytometer with trypan blue exclusion.

Cell Sorting and SynLubricin Production

293-F cells with stable incorporation of SynLubricin IRES copGFP or SynLubricin IRES mNeonGreen were expanded and induced at 1 x 10^6 cells/mL with 1 μg/mL doxycycline for 24 hours. The top 5% of GFP-expressing cells were collected through Fluorescence Activated Cell Sorting (FACS) on a FACSAria Fusion (BD Biosciences). When needed, cells were sorted a second time, collecting the top 10% of GFP-expressing cells. For SynLubricin production, cells were transferred to a 1 L ProCulture glass spinner flask (Corning) and induced at 2 x 10^6 cells/mL with 1 μg/mL doxycycline and 3.5 mM VPA. Smaller scale production of lubricin was also conducted in 100-ml Wheaton Celstir glass spinner flasks for measurement of lubricin production rates and glucose consumption rates in the presence or absence of VPA. Glucose levels were recorded with a GlucCell glucose monitoring system (CESCO BioProducts).

Immuno- and Lectin Blot Analysis

Protein in culture supernatants or purified samples were separated on NuPAGE 3-8% Tris-Acetate gels (Invitrogen) and transferred to PVDF membranes. Membranes were blocked with 3% BSA TBST for 2 hours. Primary antibodies were diluted 1:1000 and lectins were diluted to 1 μg/mL in 3% BSA TBST and incubated on membranes overnight at 4°C. Secondary antibodies or ExtrAvidin were diluted 1:2000 in 3% BSA TBST and incubated for 2 hours at room temperature. Blots were developed in Clarity ECL (BioRad) substrate and imaged on a ChemiDoc (BioRad) documentation system. Fiji was used for image processing (Schindelin et al., 2012).

Enzyme-linked immunosorbent Assay (ELISA)

A custom sandwich ELISA was used to assess the concentration of SynLubricin, similarly to previous descriptions. A 96-well plate (Costar) was incubated overnight at 4°C with 10 μg/mL peanut agglutinin (Sigma) in 50mM sodium bicarbonate buffer, pH 9.5. Plates

were blocked with 3% BSA PBS for 1 hour at room temperature. Serial dilutions of FPLC-purified bovine synovial fluid lubricin were used as standards. Samples were loaded at 1:200 dilution in DPBS for 1 hour at room temperature, followed by three washes in PBS + 0.1% Tween20. The primary antibody used (Millipore MABT401) binds to the native PRG4 tandem repeats of human and bovine lubricin, which have approximately 90% sequence similarity to the repeats of SynLubricin. Primary antibody and secondary antibody (Millipore AP126P) were diluted 1:5000 and 1:2000, respectively, and each incubated for 1 hour at room temperature, with three washes with PBS-T in between antibody incubations and following the secondary antibody incubation. The ELISA was developed at room temperature with 1-Step Ultra TMB (ThermoFisher) for 9-12 minutes or until a royal blue color appeared, at which point the reaction was stopped with 2N H_2SO_4. Absorbance was measured at 450 nm with 540 nm background subtraction on a Tecan Spark® 3M microplate reader, and concentrations were calculated using Magellan software with a four parameter Marquardt fit.

Purification of recombinant SynMuc1

293-F cells were transiently transfected using the PEI protocol previously described. After 24 hours, the media supernatant was collected. The media supernatant was diluted 1:4 in 20 mM sodium phosphate, 0.5 M NaCl, pH 7.4 and incubated with 100 µL Ni Sepharose excel resin (17371201, GE) overnight at 4°C. Sample flow through was collected using a gravity column (29922, Thermo). The resin was washed with 5 mL 20 mM sodium phosphate, 0.5 M NaCl, 5 mM imidazole, pH 7.4. SynMuc1 was eluted with 5 mL of 20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH 7.4. SynMuc1 was desalted into PBS using a Zeba Spin Desalting Column (87766, Thermo).

Purification of recombinant SynLubricin

SynLubricin was purified from SynLubricin IRES copGFP or SynLubricin IRES mNeonGreen positive 293-F cell culture supernatant by fast protein liquid chromatography (FPLC) with Q Sepharose® resin (GE) or POROS™ XS (ThermoFisher) resin. For anion exchange, the supernatant was diluted 1:10 with 50mM Tri-HCl buffer, pH 7.5, and loaded onto the column. The column was washed with 50mM Tris-HCl, 525mM NaCl, pH7.5. Purified SynLubricin was collected by eluting with 50mM Tris-HCl, 1M NaCl, pH 7.5. The purified SynLubricin was dialyzed into PBS using a Tube-O-Dialyzer (G-Biosciences) overnight at 4°C. The final purified product was obtained by concentrating with a SpeedVac on the low setting. For cation exchange, supernatant from SynLubricin IRES mNeonGreen positive 293-F suspension cell cultures was first passed through a 0.8 µm pore size cellulose-

acetate filter (Sartorius) followed by desalting and capture by fast protein liquid chromatography (FPLC). The desalting operation was performed on Sephadex G-25 (GE) fine resin with a mobile phase of 50 mM phosphate buffer, 100 mM NaCl, pH 6.8. For the capture operation the desalted samples were injected onto a column of POROS™ XS (ThermoFisher) resin and eluted with a linear gradient from 0.1 to 1 M NaCl in 50 mM phosphate buffer, pH 6.8. SynLubricin was identified in the fractions eluted between 0.46 and 0.64 M NaCl, and these fractions were pooled and used without further purification.

Glycan Profiling of SynLubricin

All reagents were purchased from Sigma unless otherwise mentioned. Recombinant SynLubricin was denatured by heating at 100 °C for 5 min. The denatured proteins were 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 borates removed under the stream of nitrogen. The glycans were permethylated for structural characterization by mass spectrometry using previously reported methods. Briefly, the dried eluate was dissolved with dimethyl sulfoxide (DMSO) and methylated by using methyl iodide and NaOH-DMSO base (prepared by mixing DMSO and 50 % w/w NaOH solution). The reaction was quenched with water and the reaction mixture was extracted with methylene chloride and dried. The permethylated glycans were dissolved in methanol and crystallized with α-dihydroxybenzoic acid (DHBA, 20 mg/mL in 50% v/v methanol: water) matrix. Analysis of glycans present in the samples was performed in the positive ion mode by MALDI-TOF/TOF-MS using an AB SCIEX TOF/TOF 5800 (Applied Biosystem MDS Analytical Technologies) mass spectrometer. Permethylated glycans from the samples were infused on an Orbitrap Fusion Tribrid mass spectrometer through an ESI probe with HCD and CID fragmentation option for further structural confirmation. The MS1 and MS2 spectra of the glycans were acquired at high resolution by a simple precursor scan and respective ions were selected manually for further MS/MS scanning. Assignment of glycan structures were done manually and by using Glycoworkbench software, based on the fragmentation patterns and common biosynthetic pathways.

Tribology

The performance of SynLubricin as a boundary lubricant was assessed using a custom linear reciprocating tribometer as previously described (Gleghorn & Bonassar, 2008). Briefly, cylindrical cartilage explants (6mm diameter x 2mm thickness) were harvested from the

femoral condyles of neonatal bovine stifles. Endogenous cartilage-bound lubricin was extracted using a 30 min incubation in 1.5M NaCl, followed by a 1-hour equilibration step in PBS. Explants were incubated in either PBS, SynLubricin, or bovine synovial fluid for 15-20 min prior to loading onto a tribometer in a 1 mL bath of the respective fluid. Explants were compressed to approximately 30% strain against a glass counter-face and permitted to depressurize over the course of one hour. After reaching an equilibrium normal load, the counter-face was linearly reciprocated at a speed of 0.3 mm/s for three cycles. Simultaneously, a biaxial load recorded the normal and shear loads. For both the forward and reverse directions and at each speed, the friction coefficient was calculated as the mean shear force while sliding divided by the equilibrium normal load.

Statistical Analysis

Statistical significance was determined by one-way ANOVA or Student's *t* test (two-tailed) as appropriate using Prism (GraphPad). For the lubrication data, a one-way ANOVA with Tukey's post-hoc tests were performed to compare mean friction coefficients across all lubricants. All graphs were generated in Prism (GraphPad, La Jolla, CA).

Part III Supplemental Table 1: Amino acid compositions in the tandem repeats of human PRG4 isoform A and SynLubricin.

Human PRG4A Repeats			SynLubricin Repeats		
Amino acid composition			Amino acid composition		
Ala (A)	58	11.4%	Ala (A)	59	12.5%
Arg (R)	0	0.0%	Arg (R)	0	0.0%
Asn (N)	1	0.2%	Asn (N)	0	0.0%
Asp (D)	2	0.4%	Asp (D)	0	0.0%
Cys (C)	0	0.0%	Cys (C)	0	0.0%
Gln (Q)	0	0.0%	Gln (Q)	0	0.0%
Glu (E)	48	9.4%	Glu (E)	59	12.5%
Gly (G)	4	0.8%	Gly (G)	0	0.0%
His (H)	0	0.0%	His (H)	0	0.0%
Ile (I)	1	0.2%	Ile (I)	0	0.0%
Leu (L)	7	1.4%	Leu (L)	0	0.0%
Lys (K)	69	13.6%	Lys (K)	59	12.5%
Met (M)	0	0.0%	Met (M)	0	0.0%
Phe (F)	0	0.0%	Phe (F)	0	0.0%
Pro (P)	155	30.5%	Pro (P)	177	37.5%
Ser (S)	13	2.6%	Ser (S)	0	0.0%
Thr (T)	150	29.5%	Thr (T)	118	25.0%
Trp (W)	0	0.0%	Trp (W)	0	0.0%
Tyr (Y)	0	0.0%	Tyr (Y)	0	0.0%
Val (V)	0	0.0%	Val (V)	0	0.0%

The following sequences are Human SynLubricin, Canine SynLubricin, and Equine SynLubricin. Italics represent secretory signals. Bold nucleotides a GS between the leader and the SynLubricin sequence.

DNA:

ATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTTCCAGGTTCCACTG
*GTGA***CGGCTCC**CAGGACCTGTCTAGCTGTGCCGGAAGATGTGGCGAGGGCTACA
GCAGAGATGCCACCTGTAACTGCGACTACAACTGCCAGCACTACATGGAATGCT
GCCCCGACTTCAAGAGAGTGTGCACAGCCGAGCTGAGCTGCAAGGGCAGATGCT
TCGAGTCCTTCGAGAGGGGCAGAGAGTGCGATTGCGACGCCCAGTGCAAGAAAT
ACGACAAGTGCTGCCCTGACTACGAGAGCTTCTGTGCCGAGGTGCACAACCCCA
CATCTCCACCTAGCAGCAAGAAGGCCCCTCCACCTTCTGGCGCCTCTCAGACAAT
CAAGAGCACCACCAAGCGGAGCCCCAAGCCTCCTAACAAGAAAAAGACCAAGA
AAGTGATCGAGAGCGAGGAAATCACCGAGGAACACAGCGTGTCCGAGAATCAA
GAGAGCAGCTCCAGCAGCAGCTCCTCCAGCTCTAGCTCCACCATCCGGAAGATC
AAGTCCAGCAAGAACAGCGCCGCCAACAGAGAGCTGCAGAAAAAGCTGAAAGT
GAAGGACAACAAGAAGAACCGGACCAAGAAGAAGCCCACACCTAAGCCTCCAG
TGGTGGATGAGGCTGGCAGCGGACTGGACAACGGCGACTTCAAAGTGACCACAC
CTGACACCAGCACCACACAGCACAACAAGGTGTCCACCTCTCCTAAGATCACCA
CCGCCAAGCCTATCAACCCCAGACCTAGCCTGCCTCCAAACAGCGACACCTCCAA
AGAAACCAGCCTGACCGTGAACAAAGAGACAACCGTCGAGACAAAAGAGACTA
CCACCACCAACAAGCAGACTAGTACCGACGGCAAAGAGAAAACCACCAGCGCC
AAAGAGACTCAGAGCATCGAAAAGACCTCCGCCAAGGATCTGGCCCCTACCTCT
AAGGTGCTGGCCAAGCCAACACCAAAGGCCGAGACAACCACAAAGGGCCCTGCT
CTGACAACCCCTAAGGAGCCAGCACCCACAACGCCGAAGGAACCAGCGCCCACG
ACCCCTAAAGAACCAGCTCCTACAACGCCCAAGGAACCGGCGCCAACAACGCCT
AAGGAACCGGCACCAACAACACCCAAAGAGCCCGCCCCCACTACTCCTAAAGAA
CCGGCTCCAACTACACCGAAGGAACCTGCCCCGACAACCCCAAAGGAACCAGCC
CCTACAACCCCTAAAGAGCCAGCGCCAACCACGCCCAAAGAACCTGCGCCGACT
ACCCCGAAAGAGCCGGCACCCACTACGCCCAAAGAGCCGGCCCCCACAACCCCG
AAGGAACCGGCTCCGACGACACCAAAGGAGCCTGCGCCCACTACACCCAAGGAG
CCTGCACCAACCACTCCCAAGGAGCCAGCTCCCACAACACCAAAGGAACCCGCG
CCCACCACGCCAAAAGAGCCAGCACCTACAACACCTAAGGAACCTGCTCCAACC
ACCCCAAAGGAGCCCGCACCTACGACTCCCAAGGAACCCGCTCCAACGACGCCT
AAGGAGCCGGCACCTACCACTCCAAAGGAGCCAGCCCCGACTACTCCGAAGGAG
CCTGCCCCAACTACTCCCAAAGAGCCAGCCCCCACGACTCCTAAGGAACCAGCA
CCAACGACACCGAAAGAACCCGCTCCCACGACGCCGAAAGAACCTGCCCCTACG
ACACCCAAAGAACCAGCCCCAACAACTCCTAAAGAGCCGGCTCCCACTACCCCT
AAGGAGCCAGCGCCTACGACCCCAAAAGAGCCTGCACCGACAACGCCAAAGGA
ACCTGCACCCACCACCCCTAAGGAACCCGCACCAACTACCCCAAAAGAACCTGC
ACCTACTACTCCAAAGGAACCGGCCCCTACCACCCCCAAGGAACCTGCGCCAAC
TACGCCGAAAGAGCCCGCGCCAACGACTCCGAAAGAACCAGCGCCGACAACTCC
AAAAGAGCCCGCTCCGACCACACCGAAAGAGCCTGCTCCCACCACACCAAAAGA
ACCAGCACCGACCACTCCTAAGGAGCCTGCTCCTACTACGCCTAAAGAACCTGCT
CCGACTACACCTAAAGAACCCGCGCCTACCACGCCTAAAGAGCCTGCGCCTACA
ACTCCCAAAGAACCCGCACCGACTACGCCAAAAGAACCGGCCCCAACGACCCCG
AAAGAACCGGCACCGACGACTCCAAAAGAACCCGCCCCAACCACACCTAAAGAG
CCCGCACCCACGACACCTAAGGAGCCCGCTCCTACCACACCCAAGGAACCAGCT
CCAACAACCCCCAAAGAGCCTGCCCCCACCACTCCGAAGGAACCCGCCCCTACT

ACACCAAAAGAGCCGGCGCCTACTACCCCCAAAGAACCGGCGCCCACAACTCCG
AAAGAGCCAGCTCCGACAACACCGAGCGAAGTGTCTACCCCTACAACCACCAAA
GAGCCAACCACCATCCACAAGAGCCCCGACGAGTCTACACCTGAGCTGTCTGCC
GAGCCTACTCCTAAGGCTCTGGAAAACAGCCCCAAAGAACCCGGGGTGCCCACC
ACAAAAACACCAGCCGCCACAAAGCCCGAGATGACCACCACAGCCAAGGACAA
GACCACCGAGCGGGACCTGAGAACAACCCCTGAAACCACAACCGCCGCTCCAAA
GATGACAAAAGAAACCGCCACAACCACCGAGAAAACAACCGAGAGCAAGATCA
CCGCCACCACAACACAAGTGACCTCCACCACCACTCAGGACACCACACCTTTCAA
GATCACAACCCTCAAGACCACTACACTGGCCCCAAAAGTGACGACCACAAAGAA
AACCATCACCACGACCGAGATCATGAACAAGCCCGAGGAAACCGCTAAGCCCAA
GGACAGGGCCACCAACAGCAAGGCCACCACACCAAAGCCACAGAAGCCTACAA
AGGCCCCTAAGAAGCCAACCAGCACAAAAAAGCCCAAGACCATGCCTAGAGTGC
GGAAGCCTAAGACAACCCCAACACCTCGGAAGATGACCAGCACTATGCCCGAGC
TGAACCCCACCTCTAGAATCGCCGAAGCCATGCTGCAGACCACCACTAGACCCA
ATCAGACCCCTAACAGCAAGCTGGTGGAAGTGAACCCCAAGTCCGAAGATGCCG
GCGGAGCTGAAGGCGAGACACCTCATATGCTGCTGAGGCCCCACGTGTTCATGC
CCGAAGTGACCCCTGACATGGACTACCTGCCAAGAGTGCCCAACCAGGGCATCA
TCATCAACCCTATGCTGAGCGACGAGACAAACATCTGCAACGGCAAGCCCGTGG
ACGGCCTGACCACACTGAGAAATGGAACCCTGGTGGCTTTCCGGGGCCACTACTT
TTGGATGCTGAGCCCTTTCAGCCCTCCATCTCCTGCCAGACGGATCACAGAAGTG
TGGGGCATCCCTTCTCCAATCGACACCGTGTTCACCCGGTGCAACTGCGAGGGCA
AGACATTCTTCTTCAAGGACAGCCAGTATTGGCGGTTCACCAACGACATCAAGGA
CGCCGGCTATCCCAAGCCAATCTTCAAAGGCTTCGGAGGCCTGACCGGCCAGATT
GTGGCTGCTCTGTCTACCGCCAAGTACAAGAACTGGCCCGAGAGCGTGTACTTCT
TTAAGAGAGGCGGCTCCATCCAGCAGTACATCTACAAGCAAGAGCCCGTGCAGA
AGTGCCCCGGAAGAAGGCCAGCTCTGAATTACCCCGTGTACGGCGAGACTACCC
AAGTGCGGAGAAGAAGATTCGAGAGAGCCATCGGACCCAGCCAGACACACACC
ATCAGAATCCAGTACAGCCCCGCCAGACTGGCCTACCAGGATAAGGGCGTGCTG
CACAACGAAGTGAAAGTGTCCATCCTGTGGCGGGGACTGCCCAATGTGGTCACA
AGCGCCATCAGCCTGCCTAACATCAGAAAGCCCGACGGCTACGACTACTACGCC
TTTAGCAAGGACCAGTACTACAACATCGACGTGCCCAGCAGAACCGCCAGAGCC
ATCACAACAAGATCCGGCCAGACACTGAGCAAAGTGTGGTACAACTGTCCTTGA
(SEQ ID NO:67)

Amino acids:
*METDTLLLWVLLLWVPGSTGD***GS**QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECC
PDFKRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPP
SSKKAPPPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSS
SSTIRKIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFK
VTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTT
NKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTP**K**EP
APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT
PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP
APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT
PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP
APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT
PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP
APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT
PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP**S**EVSTPTTTKEP
TTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLR

TTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTTTLAPK
VTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKTMPR
VRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG
AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTL
RNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFTRCNCEGKTFFFKDSQ
YWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYFFKRGGSIQQYI
YKQEPVQKCPGRRPALNYPVYGETTQVRRRRFERAIGPSQTHTIRIQYSPARLAYQD
KGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDVPSRTAR
AITTRSGQTLSKVWYNCP (SEQ ID NO:68)

Canine SynLubricin

italics = Canine secretary peptide sequence

DNA:

*atgcaatggaagattctccccatatacttgttgctgctcagtgtattcctcatccaacaagtaagtagt*caagatctcccttcttgtgcag
gcaggtgtggagaaggctatagtcgggatgcgatttgtaattgtgattataactgccaacattacatggagtgctgtccggactttaagaa
agcatgtacggtcgagctcagttgtaaagggcgctgtttcgaatctttcgctagaggccgagaatgtgactgcgacagtgactgcaaaa
agtacggaaagtgttgcccagattacgaggacttttgcgggagagtacacaaccctacttcaccaccttcttccaaaactgcaccacctt
ccccgggggcctctcagacaattaagtcaacggccaaacgctcacccaaggctccgaacaaaaaaaagactaagaaggtaatagag
agtgaggaaatcaccgaggagcactctgtgtcagaaaaccaagaaagttcttcatcatcaagctcttcttcatccactattcgcaaaataa
agtcatctaagaactctgcggcgaataaagagcttaaaaagaagccaaaagtaaaggataataaaaaggagcgaacaccgaagaaa
aagccaccacctgaaccccccgtagttgatgaggcggggtcaggcttggacaatggagacattaaattgacacccacgcctgacatt
cctacgactcaacgaaataaggttactacaagtcccaaattcaccacaggtaagcccatcaacccaaaacctagtctcccaccgaaca
ccgatacgtcaaaggagacgtcatccactcccaacaaggaaacaactgtcaaaagtaaagagacacttgctaacaaggaaaccagc
agtaaagcgaaggagaaaattacgtctgctaaagagactcggtctgcggagaagaccccagcgaaggactttgtgcctacgacgaa
agcccctgtcaaatctactccgaaggcggaaagcactactAAGGGCCCTGCTCTGACAACCCCTaaggagcca
gcacccacaacgccgaaggaaccagcgcccacgacccctaaagaaccagctcctacaacgcccaaggaaccggcgccaacaac
gcctaaggaaccggcaccaacaacacccaaagagcccgccccccactactcctaaagaaccggctccaactacaccgaaggaacct
gccccgacaaccccaaaggaaccagcccctacaaccctaaagagccagcgccaaccacgcccaaagaacctgcgccgactacc
ccgaaagagccggcacccactacgcccaaagagccggccccacaaccccgaaggaaccggctccgacgacaccaaaggagc
ctgcgcccactacacccaaggagcctgcaccaaccactcccaaggagccagctcccacaacaccaaaggaacccgcgcccacca
cgccaaaagagccagcacctacaacacctaaggaacctgctccaaccaccccaaaggagcccgcacctacgactcccaaggaacc
cgctccaacgacgcctaaggagccggcacctaccactccaaaggagccagccccgactactccgaaggagcctgccccaactact
cccaaagagccagcccccacgactcctaaggaaccagcaccaacgacaccgaaagaacccgctcccacgacgccgaaagaacct
gcccctacgacacccaaagaaccagccccaacaactcctaaagagccggctcccactacccctaaggagccagcgcctacgaccc
caaaagagcctgcaccgacaacgccaaaggaacctgcacccaccacccctaaggaacccgcaccaactaccccaaaagaacctg
cacctactactccaaaggaaccggcccctaccacccccaaggaacctgcgccaactacgccgaaagagcccgcgccaacgactcc
gaaagaaccagcgccgacaactccaaaagagcccgctccgaccacaccgaaagagcctgctcccaccacaccaaaagaaccagc
accgaccactcctaaggagcctgctcctactacgcctaaagaacctgctccgactacacctaaagaacccgcgcctaccacgcctaa
agagcctgcgcctacaactcccaaagaacccgcaccgactacgccaaaagaaccggccccaacgaccccgaaagaaccggcacc
gacgactccaaaagaacccgccccaaccacacctaaagagcccgcacccacgacacctaaggagcccgctcctaccacacccaa
ggaaccagctccaacaacccccaaagagcctgcccccaccactccgaaggaacccgcccctactacaccaaaagagccggcgcct
actacccccaaagaaccggcgcccacaactccgaaagagccagctccgacaacaccgAGCGAAGTGACAacgacggc
taaagataaaacgaccgagaaagacataattccagagattaccactgctgttcccaagatcacaactcaagaaactgctacgccaacc
gaggagacgactacggaatctaagacctcaactacgacccaagtcacttctactactagtagcaaaaacactccaaaagccacgacc
ctcgcgcccaaggtgatgacagcaacacaaaaaaccacgactactgaagagaccatgaacaagcccgaagagacgacggcagtg
cctaaagatactgcaacatcaacgaaggtaagcaccccgcgaccccgaaagccaaccaaagcaccaaagaaacccgcaagtacaa
agaaacccaacacgatccctaaacgaaaaaaaccaaaaactacacctaccccgccaaagatgactacgagcactatgcctaaactcc
atcctacctcctccgttgaggcaatgctgcaaactacaacgtcccccaatcaacgacctaattctgagatagtagaggtcaaccccaac
gaggatacggacgcggctggaaagaaaccccatatgttcccgcgacctcctgttttgacccccatatttatccctggaaccgacattctt

gtgcgggggtccaatcaagatattgccataaatcccatgctttccgacgagacaaatctctgtaatggaaaacctgtcgacggattgac
aaccctccgaaatggtactatggtggcgttccgcggccattatttctggatgttgagtccttccaaaccccgagtcctccccggaagatt
acagaggtttggggcatcccctctcccatagataccgtttttacgcgatgcaattgtgagggtaaaacattcttcttcaagggcagtcagt
actggcgattcactaacgacatcaaggacgcaggctaccccaaacagatcgtcaagggtttcggaggcttgaatggtcgaattgtcgc
tgccctgtctatagctaagtacaaggaccggccagagtctgtctatttttttcaagcgcggcggctcagtgcaacaatatacttacaagca
agagccgataaaaaaatgtacagggcgccggccggcgattaactaccctgtatatggtgagactacacaagtgaggcggagacgctt
tgagagggcgataggcccttctcagacgcataccatccggatacactactcccctattcgggttagctaccaggacaagggtttcttgc
acaatgaagtaaaaatgtccagtcaatggagaggtttcccgaacgttgttacctcagcaattgcgctgcctaacatcaggaagcctgat
ggttacgactattacgcgttttctcgcaatcaatattataacattgatgttccctcccgcactgccagagttgtgactacaagatttggacga
accctctccaatatatggtacaattgcccctag (SEQ ID NO:69)

Amino acids:
*MQWKILPIYLLLLSVFLIQQVSS*QDLPSCAGRCGEGYSRDAICNCDYNCQHYMECCPDF
KKACTVELSCKGRCFESFARGRECDCDSDCKKYGKCCPDYEDFCGRVHNPTSPPSSK
TAPPSPGASQTIKSTAKRSPKAPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSTIR
KIKSSKNSAANKELKKKPKVKDNKKERTPKKKPPPEPPVVDEAGSGLDNGDIKLTPT
PDIPTTQRNKVTTSPKFTTGKPINPKPSLPPNTDTSKETSSTPNKETTVKSKETLANKET
SSKAKEKITSAKETRSAEKTPAKDFVPTTKAPVKSTPKAESTTKGPALTTPKEPAPTTP
KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA
PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP
KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA
PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP
KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA
PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP
KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA
PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPSEVTTTAKDKTTEKDII
PEITTAVPKITTQETATPTEETTTESKTSTTTQVTSTTSSKNTPKATTLAPKVMTATQK
TTTTEETMNKPEETTAVPKDTATSTKVSTPRPRKPTKAPKKPASTKKPNTIPKRKKPK
TTPTPPKMTTSTMPKLHPTSSVEAMLQTTTSPNQRPNSEIVEVNPNEDTDAAGKKPH
MFPRPPVLTPIFIPGTDILVRGSNQDIAINPMLSDETNLCNGKPVDGLTTLRNGTMVAF
RGHYFWMLSPSKPPSPPRKITEVWGIPSPIDTVFTRCNCEGKTFFFKGSQYWRFTNDI
KDAGYPKQIVKGFGGLNGRIVAALSIAKYKDRPESVYFFKRGGSVQQYTYKQEPIKK
CTGRRPAINYPVYGETTQVRRRRFERAIGPSQTHTIRIHYSPIRVSYQDKGFLHNEVK
MSSQWRGFPNVVTSAIALPNIRKPDGYDYYAFSRNQYYNIDVPSRTARVVTTRFGRT
LSNIWYNCP (SEQ ID NO:70)

(1) Equine SynLubricin

Red = Equine secretary peptide sequence

DNA:

*atggagtggaaaatcctgcctatttaccttctcctgttgctgagtatattctccatccaggaggtttcaagc*caagacctttctagttgcg
ctggtcggtgtggggagggatactctcgggatgcgacttgcaactgcgattttaattgtcaatactacatggaatgttgtccggactttaa
gaaagtctgtacatctgaattgtcttgtaaaggccgctgtttcgagagtttcgaaaggggcgagaatgcgattgcgatgctgactgtaa
gaaatacggtaagtgttgttcagattatgaaagcttctgcgaggaagtccataatcctacgtctccgccgagttccaagacagctccccc
gcctccaggggccagccagactatcaagagtacagctaaacggtcaccaaagtcaaataagaaaaaaactaaaaaagttatcgaga
gtgaagagatcatagaagaacacagtgtgtccgagaatcaggagtcatcttccagctctagctcaagttcatctaccatccgcaaggtta
agtctagcaaaaactcagcagcgaacagagaactcaaaaagaagcctaaggtcaaggattctaaaaaaaaacgaaccccgaaaaaa
aaaccgacgcctgagccaccagtcatagacgaggccgggagtggtttggataacggagacttcatgttgattcccaccccgaaaattc
caaccacgcaaagaaataaggtgacgacatcaccaaagattacaacggtaaaaccaattaaccccaagccttcccttcctcccaattcc
gacacgtcaaaagagaccactagcacacctaataaagaaactacggtcgagaccaaggagaccgagatcacaaacaaggagactt
ctacaagcgccaatgaaaagactacgagcgccaggaagagtacagagaaaacatccgacaaagattttgctccggccagcgaagta

cctgcaaaaagtacccctaaggctgaaaccaccacaaagggccctgctctgacaacccctaaggagccagcacccacaacgccga
aggaaccagcgcccacgacccctaaagaaccagctcctacaacgcccaaggaaccggcgccaacaacgcctaaggaaccggca
ccaacaacacccaaagagcccgcccccactactcctaaagaaccggctccaactacaccgaaggaacctgccccgacaaccccaa
aggaaccagcccctacaaccctaaagagccagcgccaaccacgcccaaagaacctgcgccgactaccccgaaagagccggcac
ccactacgcccaaagagccggcccccacaaccccgaaggaaccggctccgacgacaccaaaggagcctgcgcccactacaccca
aggagcctgcaccaaccactcccaaggagccagctcccacaacaccaaaggaacccgcgcccaccacgccaaaagagccagca
cctacaacacctaaggaacctgctccaaccaccccaaaggagcccgcacctacgactcccaaggaacccgctccaacgacgcctaa
ggagccggcacctaccactccaaaggagccagccccgactactccgaaggagcctgccccaactactcccaaagagccagccccc
acgactcctaaggaaccagcaccaacgacaccgaaagaacccgctcccacgacgccgaaagaacctgcccctacgacacccaaa
gaaccagccccaacaactcctaaagagccggctcccactaccctaaggagccagcgcctacgaccccaaaagagcctgcaccga
caacgccaaaggaacctgcacccaccacccctaaggaacccgcaccaactaccccaaaagaacctgcacctactactccaaagga
accggcccctaccacccccaaggaacctgcgccaactacgccgaaagagcccgcgccaacgactccgaaagaaccagcgccga
caactccaaaagagcccgctccgaccacaccgaaagagcctgctcccaccacaccaaaagaaccagcaccgaccactcctaagga
gcctgctcctactacgcctaaagaacctgctccgactacacctaaagaacccgcgcctaccacgcctaaagagcctgcgcctacaact
cccaaagaacccgcaccgactacgccaaaagaaccggccccaacgaccccgaaagaaccggcaccgacgactccaaaagaacc
cgccccaaccacacctaaagagcccgcacccacgacacctaaggagcccgctcctaccacacccaaggaaccagctccaacaacc
cccaaagagcctgcccccaccactccgaaggaacccgcccctactacaccaaaagagccggcgcctactacccccaaagaaccgg
cgcccacaactccgaaagagccagctccgacaacaccgagcgaagtgtctaccacgacgactaccatgaaacctccgacgacacc
caaaaatcttgctgaaagcaccccagagttcccagcggagccaacacccaaagcactggagaactcacccaaagaaccggctgtac
cgactacgaaggcccctgaagtaaccaaaccagaagtcacaacaaccgctaaagacaaggttacgggaaaggatattcacacgatt
cccgagataactacagcggcacctaagataacgaccgaaacggccacgacaactgaagagaaaacaacggaaagtaaggtgacct
ctactataatgcaagtgacctccacgaccgaggatacgacgacaagctccaagataacgcctaaagcaacgacattggcaccgaaa
gtgatgaccgcaacaaaaactaccacaacacaggaaacgataaacaagctggaggagacgacggctattcctaaggatacggcga
cgcacagcaaagtgactacgccaaagccgaagaagccgaccaaagcgcctcgaaagccgacatccacaaagaaaccgaaaacg
ccgcgcaagcgcaaaccaaagacaacaccgattcccccgaaaatcaccaccccgaccactcctaaaagtaaccctacgactttggc
ggaagccatgcttcagactacaacttcacctaaccagactccaaattccgctatgatagaggtcaacccgaaaaacgaggacgcgga
cgctgcggaaggggaaaagccgctcgtgatacttcgaccacacgtccttactccaatcgtcataccgggtccggactttcttgtccgcg
gtccaaacttgggaatcggaattaaccccatgcttagcgacgagacgaacttgtgtaacggtaaaccagtggacggactcaccaccct
gagaaatggaactctcgtggctttcaggggccactatttctggatgctccgaccatttagtcccccgagtccgccgaggagaatcaccg
aggtatgggggattccctctcctattgataccgtcttcactcgctgcaactgcgagggaaagacatttttcttcaaggactcacagtattgg
cgattcaccaacgacataaaggatgctggataccctaaattgattagcaagggctttgggggggcttagtggcaaaatcgtggccgctct
ttcaatagcaacgtacaagaacaggccagagagcgtttatttttttaagcgaggggggcgaatacagcaatacatctacaagcaagaa
cccataagaaagtgtccaggacgccgaccagctatacattattcagtttacggagaggctcctcagattcggaggagaaggttcgaac
gggccataggcccgtctcagacgcacaccatccgcattcactactcccccgtacgcgtatcataccaagacaaagtgccgtccactga
ctttctccacaacgaggtcaaagtaagcaccctgtggcgcggacttccagacaccgttacatccgccatttcccttcctaacttgcggaa
accagacggatacgactattatgctttttcaaaagaccaatattataatattgacgtcccgagccgaactgctcgcgcaataactacccga
agtggccagacattgagtaaggtctggtataactgtccctag (SEQ ID NO:71)

Amino acids:

*MEWKILPIYLLLLLSIFSIQEVSS*QDLSSCAGRCGEGYSRDATCNCDFNCQYYMECCPDF
KKVCTSELSCKGRCFESFERGRECDCDADCKKYGKCCSDYESFCEEVHNPTSPPSSK
TAPPPPGASQTIKSTAKRSPKSNKKKTKKVIESEEIIEEHSVSENQESSSSSSSSSSTIRK
VKSSKNSAANRELKKKPKVKDSKKKRTPKKKPTPEPPVIDEAGSGLDNGDFMLIPTP
KIPTTQRNKVTTSPKITTVKPINPKPSLPPNSDTSKETTSTPNKETTVETKETEITNKETS
TSANEKTTSARKSTEKTSDKDFAPASEVPAKSTPKAETTTKGPALTTPKEPAPTTPKE
PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT
TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT
TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE

PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT
TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT
TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPSEVSTTTTTMKPPTTPKN
LAESTPEFPAEPTPKALENSPKEPAVPTTKAPEVTKPEVTTTAKDKVTGKDIHTIPEITT
AAPKITTETATTTEEKTTESKVTSTIMQVTSTTEDTTTSSKITPKATTLAPKVMTATKT
TTTQETINKLEETTAIPKDTATHSKVTTPKPKKPTKAPRKPTSTKKPKTPRKRKPKTTP
IPPKITTPTTPKSNPTTLAEAMLQTTTSPNQTPNSAMIEVNPKNEDADAAEGEKPLVIL
RPHVLTPIVIPGPDFLVRGPNLGIGINPMLSDETNLCNGKPVDGLTTLRNGTLVAFRG
HYFWMLRPFSPPSPPRRITEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDA
GYPKLISKGFGGLSGKIVAALSIATYKNRPESVYFFKRGGRIQQYIYKQEPIRKCPGRR
PAIHYSVYGEAPQIRRRRFERAIGPSQTHTIRIHYSPVRVSYQDKVPSTDFLHNEVKVS
TLWRGLPDTVTSAISLPNLRKPDGYDYYAFSKDQYYNIDVPSRTARAITTRSGQTLSK
VWYNCP (SEQ ID NO:98)

REFERENCES

Backliwal, G., Hildinger, M., Kuettel, I., Delegrange, F., Hacker, D. L., & Wurm, F. M. (2008). Valproic acid: a viable alternative to sodium butyrate for enhancing protein expression in mammalian cell cultures. *Biotechnol Bioeng*, *101*(1), 182–189.//doi.org/10.1002/bit.21882

Backstrom, M., Link, T., Olson, F. J., Karlsson, H., Graham, R., Picco, G., … Hansson, G. C. (2003). Recombinant MUC1 mucin with a breast cancer-like O-glycosylation produced in large amounts in Chinese-hamster ovary cells. *Biochem J*, *376*(Pt 3), 677–686.//doi.org/10.1042/bj20031130

Bahabri, S. A., Suwairi, W. M., Laxer, R. M., Polinkovsky, A., Dalaan, A. A., & Warman, M. L. (1998). The camptodactyly-arthropathy-coxa vara-pericarditis syndrome: Clinical features and genetic mapping to human chromosome 1. *Arthritis & Rheumatism*, *41*(4), 730–735.//doi.org/10.1002/1529-0131(199804)41:4<730::AID-ART22>3.0.CO;2-Y

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580.//doi.org/10.1093/nar/27.2.573

Bosques, C. J., Collins, B. E., Meador, J. W., Sarvaiya, H., Murphy, J. L., DelloRusso, G., … Venkataraman, G. (2010). Chinese hamster ovary cells can produce galactose-α-1,3-galactose antigens on proteins. *Nature Biotechnology*, *28*(11), 1153–1156.//doi.org/10.1038/nbt1110-1153

Boussif, O., Lezoualc'h, F., Zanta, M. A., Mergny, M. D., Scherman, D., Demeneix, B., & Behr, J. P. (1995). A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc Natl Acad Sci U S A*, *92*(16), 7297–7301.//doi.org/10.1073/pnas.92.16.7297

Brooks, S. A. (2004). Appropriate glycosylation of recombinant proteins for human use. *Molecular Biotechnology*, *28*(3), 241–255.//doi.org/10.1385/MB:28:3:241

Butler, M., & Spearman, M. (2014). The choice of mammalian cell host and possibilities for glycosylation engineering. *Current Opinion in Biotechnology*, *30*, 107–112.//doi.org/10.1016/j.copbio.2014.06.010

de los Milagros Bassani Molinas, M., Beer, C., Hesse, F., Wirth, M., & Wagner, R. (2014). Optimizing the transient transfection process of HEK-293 suspension cells for protein production by nucleotide ratio monitoring. *Cytotechnology*, *66*(3), 493–514.//doi.org/10.1007/s10616-013-9601-3

Dhanisha, S. S., Guruvayoorappan, C., Drishya, S., & Abeesh, P. (2018). Mucins: Structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets. *Critical Reviews in Oncology/Hematology*, *122*, 98–122.//doi.org/10.1016/j.critrevonc.2017.12.006

Dumont, J., Euwart, D., Mei, B., Estes, S., & Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit Rev Biotechnol*, *36*(6), 1110–1122.//doi.org/10.3109/07388551.2015.1084266

Durocher, Y., Perret, S., & Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res*, *30*(2), E9.

Elsaid, K. A., Fleming, B. C., Oksendahl, H. L., Machan, J. T., Fadale, P. D., Hulstyn, M. J., … Jay, G. D. (2008). Decreased lubricin concentrations and markers of joint inflammation in the synovial fluid of patients with anterior cruciate ligament injury. *Arthritis & Rheumatism*, *58*(6), 1707–1715.//doi.org/10.1002/art.23495

Estrella, R. P., Whitelock, J. M., Packer, N. H., & Karlsson, N. G. (2010). The glycosylation of human synovial lubricin: implications for its role in inflammation. *Biochemical Journal*, *429*(2), 359–367.//doi.org/10.1042/bj20100360

Flannery, C. R., Zollner, R., Corcoran, C., Jones, A. R., Root, A., Rivera-Bermudez, M. A., … Glasson, S. S. (2009). Prevention of cartilage degeneration in a rat model of osteoarthritis by intraarticular treatment with recombinant lubricin. *Arthritis Rheum*, *60*(3), 840–847.//doi.org/10.1002/art.24304

Gemayel, R., Vinces, M. D., Legendre, M., & Verstrepen, K. J. (2010). Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annual Review of Genetics*, *44*(1), 445–477.//doi.org/10.1146/annurev-genet-072610-155046

Ghaderi, D., Zhang, M., Hurtado-Ziola, N., & Varki, A. (2012). Production platforms for

biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnol Genet Eng Rev*, *28*, 147–175.//doi.org/10.5661/bger-28-147

Gipson, I. K., Spurr-Michaud, S., Tisdale, A., & Menon, B. B. (2014). Comparison of the Transmembrane Mucins MUC1 and MUC16 in Epithelial Barrier Function. *PLOS ONE*, *9*(6), e100393.//doi.org/10.1371/journal.pone.0100393

Gleghorn, J. P., & Bonassar, L. J. (2008). Lubrication mode analysis of articular cartilage using Stribeck surfaces. *J Biomech*, *41*(9), 1910–1918.//doi.org/10.1016/j.jbiomech.2008.03.043

Hang, H. C., & Bertozzi, C. R. (2005). The chemistry and biology of mucin-type O-linked glycosylation. *Bioorganic & Medicinal Chemistry*, *13*(17), 5021–5034.//doi.org/10.1016/j.bmc.2005.04.085

Hattrup, C. L., & Gendler, S. J. (2008). Structure and Function of the Cell Surface (Tethered) Mucins. *Annual Review of Physiology*, *70*(1), 431–457.//doi.org/10.1146/annurev.physiol.70.113006.100659

Jones, A. R. C., Gleghorn, J. P., Hughes, C. E., Fitz, L. J., Zollner, R., Wainwright, S. D., … Flannery, C. R. (2007). Binding and localization of recombinant lubricin to articular cartilage surfaces. *Journal of Orthopaedic Research*, *25*(3), 283–292.//doi.org/doi:10.1002/jor.20325

Kosinska, M. K., Ludwig, T. E., Liebisch, G., Zhang, R., Siebert, H.-C., Wilhelm, J., … Steinmeyer, J. (2015). Articular Joint Lubricants during Osteoarthritis and Rheumatoid Arthritis Display Altered Levels and Molecular Species. *PLOS ONE*, *10*(5), e0125192.//doi.org/10.1371/journal.pone.0125192

Kosuri, S., & Church, G. M. (2014). Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods*, *11*, 499.//doi.org/10.1038/nmeth.2918

Kuo, J. C.-H., Gandhi, J. G., Zia, R. N., & Paszek, M. J. (2018). Physical biology of the cancer cell glycocalyx. *Nature Physics*, *14*(7), 658–669.//doi.org/10.1038/s41567-018-0186-9

Le Graverand-Gastineau, M. P. (2010). Disease modifying osteoarthritis drugs: facing development challenges and choosing molecular targets. *Curr Drug Targets*, *11*(5), 528–535.//doi.org/10.2174/138945010791011893

López Castel, A., Cleary, J. D., & Pearson, C. E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature Reviews Molecular Cell Biology*, *11*, 165.//doi.org/10.1038/nrm2854

Mantelli, F., & Argüeso, P. (2008). Functions of ocular surface mucins in health and disease.

Current Opinion in Allergy and Clinical Immunology, *8*(5), 477–483.//doi.org/10.1097/ACI.0b013e32830e6b04

Marcelino, J., Carpten, J. D., Suwairi, W. M., Gutierrez, O. M., Schwartz, S., Robbins, C., … Warman, M. L. (1999). CACP, encoding a secreted proteoglycan, is mutated in camptodactyly-arthropathy-coxa vara-pericarditis syndrome. *Nature Genetics*, *23*, 319.//doi.org/10.1038/15496

Mauris, J., & Argüeso, P. (2012). Mucins and Galectin-3 in Ocular Surface Health and Disease. In *Galectins and Disease Implications for Targeted Therapeutics* (Vol. 1115, pp. 409–414). American Chemical Society.//doi.org/10.1021/bk-2012-1115.ch025

Nath, S., & Mukherjee, P. (2014). Muc1: a multifaceted oncoprotein with a key role in cancer progression. *Trends in Molecular Medicine*, *20*(6), 332–342.//doi.org/10.1016/j.molmed.2014.02.007

Oren, M., Barela Hudgell, M. A., D'Allura, B., Agronin, J., Gross, A., Podini, D., & Smith, L. C. (2016). Short tandem repeats, segmental duplications, gene deletion, and genomic instability in a rapidly diversified immune gene family. *BMC Genomics*, *17*, 900.//doi.org/10.1186/s12864-016-3241-x

Pearson, C. E., Edamura, K. N., & Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nature Reviews Genetics*, *6*, 729.//doi.org/10.1038/nrg1689

Reesink, H. L., Bonnevie, E. D., Liu, S., Shurer, C. R., Hollander, M. J., Bonassar, L. J., & Nixon, A. J. (2016). Galectin-3 Binds to Lubricin and Reinforces the Lubricating Boundary Layer of Articular Cartilage. *Scientific Reports*, *6*, 25463.//doi.org/10.1038/srep25463

Rhee, D. K., Marcelino, J., Baker, M., Gong, Y., Smits, P., Lefebvre, V., … Carpten, J. D. (2005). The secreted glycoprotein lubricin protects cartilage surfaces and inhibits synovial cell overgrowth. *J Clin Invest*, *115*(3), 622–631.//doi.org/10.1172/jci22263

Samsom, M. L., Morrison, S., Masala, N., Sullivan, B. D., Sullivan, D. A., Sheardown, H., & Schmidt, T. A. (2014). Characterization of full-length recombinant human Proteoglycan 4 as an ocular surface boundary lubricant. *Experimental Eye Research*, *127*, 14–19.//doi.org/10.1016/j.exer.2014.06.015

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., … Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*.//doi.org/10.1038/nmeth.2019

Schmidt, T. A., Sullivan, D. A., Knop, E., & et al. (2013). Transcription, translation, and function of lubricin, a boundary lubricant, at the ocular surface. *JAMA Ophthalmology*,

131(6), 766–776.//doi.org/10.1001/jamaophthalmol.2013.2385

Shurer, C. R., Colville, M. J., Gupta, V. K., Head, S. E., Kai, F., Lakins, J. N., & Paszek, M. J. (2018). Genetically Encoded Toolbox for Glycocalyx Engineering: Tunable Control of Cell Adhesion, Survival, and Cancer Cell Behaviors. *ACS Biomaterials Science & Engineering*, *4*(2), 388–399.//doi.org/10.1021/acsbiomaterials.7b00037

Sonawane, N. D., Szoka Jr., F. C., & Verkman, A. S. (2003). Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes. *J Biol Chem*, *278*(45), 44826–44831.//doi.org/10.1074/jbc.M308643200

Tang, N. C., & Chilkoti, A. (2016). Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nature Materials*, *15*(4), 419–424.//doi.org/10.1038/nmat4521

Part IV

This Part IV provides, among other aspects, a description of the physical principles of membrane shape regulation by the glycocalyx.

In connection with this Part IV, it is known that cells bend their plasma membranes into highly curved forms to interact with the local environment, but how shape generation is regulated is not fully resolved. This Part IV describes a broad synergy between shape-generating processes in the cell interior and the external organization and composition of the cell-surface glycocalyx. Mucin biopolymers and long-chain polysaccharides within the glycocalyx can generate entropic forces that favor or disfavor the projection of spherical and finger-like extensions from the cell surface. A polymer brush model of the glycocalyx successfully predicts the effects of polymer size and cell-surface density on membrane morphologies. Specific glycocalyx compositions can also induce plasma membrane instabilities to generate more exotic undulating and pearled membrane structures and drive secretion of extracellular vesicles. Together, results presented in this Part IV suggest a fundamental role for the glycocalyx in regulating curved membrane features that serve in diverse modes of communication between cells and with the extracellular matrix.

Introduction to Part IV

Tubular and spherical extensions of the plasma membrane play vital roles in human development and everyday cellular functions. While curved membrane protrusions have long been recognized to increase cell-surface area for secretion, absorption, and receptor-mediated communication, modern research has provided compelling examples of much more diverse

and sophisticated functionalities (Marshall, 2012). For instance, T-cells of the adaptive immune system generate a high density of tubular microvilli to engage antigen presenting cells, and such structures may be similarly important for the recognition of tumor cells by engineered immune cell therapies (D'Aloia et al., 2018; Jung et al., 2016). Membrane projections also enable cell-to-cell communication over long ranges and at precise three-dimensional locations in tissues. During development, long and thin membrane projections called cytonemes pinpoint delivery of morphogens from 'sender' cells to specific 'receiver' cells up to 40-microns away (Bischoff et al., 2013; Kornberg and Roy, 2014). Stem cells, immune cells, and many other cell types are also known to bend their plasma membranes into spherical microvesicles that are directly shed and can deliver macromolecular cargoes over long distances (Tricarico et al., 2017). Moreover, curved membrane features are ubiquitous in physical cell behaviors, including migration and mechanotransduction. For example, spherical membrane expansions called blebs are generated by primordial germ cells, tumor cells, and other cell types for protrusion and frictional coupling with the tissue matrix during migration (Paluch and Raz, 2013).

Deregulation of membrane-shape generating processes can contribute directly to disease progression. As a notable example, aggressive tumor cells frequently extend numerous microvilli for adhesion and rolling in the vasculature (Kramer and Nicolson, 1979; Liu et al., 2018). Aggressive tumor cells can also project blebs for amoeboid migration (Bergert et al., 2015; Friedl and Wolf, 2010). Microvesicles often bud from the plasma membrane of tumor cells at abnormally high rates (Antonyak et al., 2011; Becker et al., 2016). Cargoes carried by these particles are now recognized to have diverse modulatory roles, including reprogramming of other cell types in the stroma and the preparation of distant metastatic niches for colonization (Becker et al., 2016).

Forces originating from cytoskeletal dynamics are posited to generate membrane curvature for the diverse spherical and tubular structures on the cell surface. Polymerizing cytoskeletal filaments are envisioned to push out at discrete points along the plasma membrane for extension of microvilli, cilia, filapodia and other finger-like projections (Footer et al., 2007; Gupton and Gertler, 2007; Peskin et al., 1993). Contraction of the cytoskeleton generates the hydrostatic pressure for spherical expansion of the membrane during bleb formation (Charras et al., 2005). The physical dynamics that bend sub-regions of the plasma membrane into microvesicles remain poorly understood; however, reports have implicated the actin cytoskeleton in their biogenesis (Tricarico et al., 2017).

While the cell-surface glycocalyx is not featured in canonical models of membrane shape regulation, correlations abound between glycocalyx composition and cell-surface morphology in both normal and disease states. In normal cell physiology, polypeptide and sugar co-polymers called mucins are frequently anchored at high densities on the surfaces of epithelial microvilli (Hattrup and Gendler, 2008; Kesavan et al., 2009; Kesimer et al., 2013), cilia (Button et al., 2012), and filapodia (Bennett et al., 2001); while hyaluronan polymers densely coat the microvilli of oocytes and mesothelium (Evanko et al., 2007; Makabe Sayoko et al., 2006); and long chains of sialic acid and hyaluronan decorate the highly curved surfaces of neuronal axons (Fowke et al., 2017; van den Pol and Kim, 1993; Zhang et al., 1992). T-cells and dendritic cells express cell-surface mucins upon activation or maturation, which coincides often with the dramatic changes in membrane tubularization and microvilli generation (Agrawal et al., 1998; Cloosen et al., 2004; Jung et al., 2016; Pilon et al., 2009). Aggressive tumor cells frequently produce an abundance of mucins and hyaluronan on their cell surface (Kufe, 2009; Turley et al., 2016), and the expression of these polymers has been anecdotally linked to their unique membrane features, such as extensive microvilli (Polefka et al., 1984). Mucins and hyaluronan polymers are also densely arrayed on the surfaces of enterocytes, reactive astrocytes, dendritic cells, and tumor cells that are known to secrete high levels of microvesicles (Cloosen et al., 2004, 2004; Gangoda et al.; McConnell et al., 2009; Paszek et al., 2014; Pelaseyed et al.; Tricarico et al., 2017). While the ubiquity of these correlations suggests a possible causal relationship between glycocalyx polymer composition and plasma membrane morphologies, a specific mechanism of action has not been delineated. The present disclosure contributes to an understanding of this mechanism of action.

Mucins and long-chain polysaccharides are anchored to the membrane in such a way that long polymer chains or loops are expected to extend from the cell surface (Hattrup and Gendler, 2008; Lee et al., 1993). The ensemble resembles a well-studied structure in polymer physics called a brush, where polymers are grafted on one end to a surface (Chen et al., 2017). Polymer brush theory has long recognized that steric interactions in a densely crowded brush restrict the number of molecular configurations each polymer can explore, thereby increasing the free energy of the system through reduced entropy (de Gennes, 1980). Similar to the thermodynamic basis of gas pressure, the entropic penalty associated with molecular crowding can theoretically generate sufficient pressure to deform a flexible surface, like a membrane (Hiergeist and Lipowsky, 1996; Lipowsky, 1995).

RESULTS

Glycocalyx polymers and membrane morphology:

In this Part IV, we analyzed whether glycocalyx polymers may generate an entropic bending force to favor the formation of specific membrane forms. As a corollary to this, we tested whether emergent membrane structures could be tuned through rational manipulation of the glycocalyx.

To test this, we constructed a genetically encoded library of native, semi-synthetic, and rationally designed mucin polymers of varying size, backbone sequence, and membrane anchorage (Fig. 33A and Fig. 36A). Each construct encoded a mucin polymer domain comprised of an unstructured polypeptide backbone with a high density of serine and threonine sites for *O*-glycosylation. When expressed in cells, the mucin domains were post-translationally modified with *O*-linked sugar side chains to form a bottlebrush molecular structure that defines mucins (Fig. 38A, B).

Polymer domains in the library included the 42 native tandem repeats (TR) of Mucin-1 (Muc1-42TR), the serine and threonine-rich polymer domain of Podocalyxin (Podxl; S/T-Rich), and a new synthetic mucin that we rationally designed and constructed through the tandem fusion of 80 perfect repeats based on a consensus of mucin *O*-glycosylation sequence, PPASTSAPGA (Rational) (Fig. 33A and Fig. 38A). Each polymer domain was fused to the native Muc1 transmembrane anchor with the cytoplasmic tail deleted (ΔCT), or a 21-amino acid synthetic transmembrane anchor (TM21), or a native mucin anchor with a membrane proximal green fluorescent protein for imaging (GFP-ΔCT) (Fig. 33A and Fig. 38A).

When expressed and assembled at high levels on the epithelial cell surface, each mucin polymer in our library triggered a dramatic tubularization of the plasma membrane, as observed by scanning electron microscopy (SEM) (Fig. 33B, C and Fig. 38B). Without intending to be bound by any particular theory, we concluded that this tubularization was likely a general consequence of polymer anchorage to the plasma membrane and did not require a specific biopolymer sequence or transmembrane anchor. Notably, the Muc1-42TR ΔCT was identical to native Mucin-1 except for the cytoplasmic tail, indicating that native glycocalyx constituents can influence plasma membrane morphology in addition to our rationally designed polymers. Mucin expression did not have a significant effect on endocytosis, arguing against lipid recycling and the regulation of membrane tension as a primary mechanism for the morphological changes (Fig. 38C, D).

The tubularization phenomenon was relatively insensitive to the length of the mucin polymer domain, provided that the polymers were expressed on the cell surface at moderate

to high densities. cDNAs for 0, 10, or 42 Muc1 repeats were fused with a GFP-tagged transmembrane anchor to encode cell-surface mucins with expected contour lengths of 0, 65, and 270 nm, respectively (Fig. 33D and Fig. 38E). Cell lines expressing the constructs were sorted into populations with similar mucin surface densities using a nanobody that probed cell-surface GFP (Fig. 33D). The flexible polymer domain was required for efficient membrane tubularization, and the 10- and 42-TR mucins induced comparable levels of membrane tubularization despite their size difference (Fig. 33E and Fig. 38F). We compared cells of similar spread area to rule out the possibility that changes in membrane surface tension and other effects associated with cell spreading could explain the morphological differences (Fig. 33E).

Similar to mucins, we found that a glycocalyx rich in large, linear polysaccharides could also trigger dramatic changes in plasma membrane morphology. Notably, hyaluronic acid synthase 3 (HAS3) expression increased the density of high molecular weight hyaluronic acid (HA) polymers on the cell surface and led to the protrusion of many finger-like membrane extensions (Fig. 36A-D), consistent with prior observations by others (Koistinen et al., 2015). Together, these results suggested that diverse glycocalyx polymer types and sizes might influence cell morphological states.

We next tested whether glycocalyx biopolymers could induce spontaneous curvature in model membranes independent of intracellular machinery. When anchored to the surface of giant unilamellar vesicles (GUVs), we found that the S/T-rich polymer domain of Podxl triggered spontaneous generation of spherical and tubular membrane structures (Fig. 33F and Fig. 37A, B). Tubules were also observed at very high densities of a folded protein, human serum albumin (HSA), consistent with previous findings that the extensive crowding of folded or intrinsically disordered proteins could induce spontaneous membrane curvatures in GUVs (Stachowiak et al., 2010) (Fig. 33F and Fig. 37B, C). However, the surface density required to induce spontaneous tubularization was significantly lower for Podxl mucin compared to HSA (Fig. 33F and Fig. 37B).

Specialized cells in vivo:

Motivated by these observations *in vitro*, we considered whether glycocalyx polymers might play a role in shaping the morphology of specialized cell types *in vivo*. We elected to evaluate synoviocytes, since these secretory cells are known to produce large quantities of HA for joint lubrication and, thus, are expected to display a high density of HA polymers on their surface. We isolated synovial tissues from equine carpus (Fig. 34A) and found that

primary synoviocytes expressing HAS3 were highly tubulated, but treatment with hyaluronidase (HyA) to degrade HA resulted in the rapid destabilization and disappearance of membrane tubules (Fig. 34B, C). We also evaluated synoviocyte morphology in tissues that were freshly extracted and briefly cultured *ex vivo* (< 1 h). The synoviocytes in native synovial tissue displayed an HA-rich head that appeared highly tubulated and protruded from the tissue matrix (Fig. 34D, E). Brief treatment of the tissue with HyA *ex vivo* resulted in a dramatic retraction of synoviocyte tubules, suggesting a role for the glycocalyx in the maintenance of membrane projections *in vivo* (Fig. 34E).

Polymer brush framework:

We considered whether the observed membrane shapes and their frequencies could be rationalized through the framework of polymer brush theory. We noted that two limiting regimes are classically described in polymer physics for end-grafted polymers: the "mushroom" regime, where polymers at low grafting densities have limited interactions with each other, and the "brush" regime, where crowded polymers can interact sterically and electrostatically with each other to exert larger pressures on the anchoring surface (Milner, 1991) (Fig. 38A). For mucins, we expected the transition from the mushroom to brush regime to occur at a surface density where the average distance between the polymers was approximately two times their radius of gyration in solution (Fig. 38A).

To measure the radius of gyration and flexibility of individual mucins, we produced recombinant Muc1-42TR with a terminal purification tag in place of its transmembrane anchor (Fig. 41A-C). Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) reported 32 nm ± 0.4% for the mucin radius of gyration in physiological buffer. Based on the estimated Muc1-42TR contour length of approximately 270 nm, and again without intending to be bound by any particular theory, we concluded that the mucin had a persistence length of approximately 7.5 nm and adopted the extended random coil configuration expected for a semi-flexible polymer in solution.

We next tested whether polymer brush theory could capture the physical behavior of mucin ensembles on the cell surface. We tested whether mucins stretch and extend in a predictable manner as they become progressively more crowded, a characteristic physical behavior originally described by Alexander and de Gennes in their seminal theories on polymer brushes (Alexander, 1977; Milner, 1991). We chose to evaluate mucin extension on actin-containing tubules that resembled microvilli, since the curvature of these structures was highly uniform and essentially independent of the mucin surface density (Fig. 41D). As such,

we were able to approximate the tubule surface as a rigid cylinder of fixed radius for direct comparison to classical theory. A cDNA for Muc1-42TR with complimentary epitope tags flanking the mucin polymer domain was constructed. Following cellular expression, the encoded tags were labeled with fluorophore-conjugated probes and resolved on microvilli cross-sections using a super-resolution optical technique called expansion microscopy (ExM) (Fig. 35B and Fig. 41E). We found that the mucin extension had an exponential dependence, or 'scaled,' with fluorescence intensity, and hence surface density, with an exponent of 0.48 ± 0.10 (Fig. 35B). This value compared well to the theoretically derived power law exponent of between 0.33 and 0.5 for polyelectrolytes grafted on a rigid cylindrical surface at physiological salt concentrations (Zhulina and Borisov, 1996).

We created a polymer brush model to describe the physical behavior of a mucin-rich glycocalyx assembled on the plasma membrane. The entropic pressure contributed by the mucin brush generated a spontaneous membrane curvature that strongly scaled with polymer density and weakly with polymer chain length (Hiergeist and Lipowsky, 1996) (Fig. 35C and Fig. 42). The weak dependence on polymer length was consistent with findings that mucins with 10 and 42 repeats had comparable effects on cell-surface morphology despite their 4-fold difference in size (Fig. 33E and Fig. 38F). For these two mucins, our brush model predicted only a ~20% difference in induction of spontaneous membrane curvature (Fig. 42).

Preferred membrane shapes:

We tested whether the polymer model could explain the frequency of finger-like and spherical protrusions from the cell surface. We reasoned that protrusion of a specific membrane feature would be disfavored when high intracellular forces were required to extend or maintain the protrusion and favored when these force requirements were minimal. Minimizing the standard Helfrich free energy function for membranes with induced spontaneous curvature, we calculated the equilibrium cytosolic pressure required to maintain a spherical membrane bleb and the point force required to maintain a membrane tubule (Fig. 36D). For experimental comparison, we evaluated the types, sizes, and frequencies of plasma membrane features as a function of mucin cell-surface density. Cells expressing Muc1-42TR GFP were labeled with an anti-GFP nanobody and sorted into populations of varying mucin surface levels (Fig. 36A). The average mucin surface density in each population was estimated by SDS-PAGE through interpolation using a nanobody standard curve (Fig. 43). Molecular surface densities in the sorted populations ranged from 180 to ~50,000 mucins per μm^2. For reference, we expected the mushroom to brush transition to occur around 250

mucins per μm^2 based on the measured radius of gyration of recombinant Muc1-42TR in solution.

Initially, we evaluated membrane blebs. Using physical parameters measured for Muc1-42TR, we predicted that the pressure required for maintaining a bleb with a typical radius of 250 nm would be minimal at moderate mucin densities near the mushroom-brush transition (Fig. 35D). A surprising model prediction was that the required maintenance pressure would rise sharply at higher mucin densities, quickly reaching pressures that exceed the known limits of the cell's contractile machinery (Charras et al., 2008). Thus, theory suggested that blebbing would be suppressed by a highly dense glycocalyx (Fig. 35D). Our experimental observations showed good qualitative agreement with these predictions. Cells with a mucin density near the estimated mushroom-brush transition displayed a significant number of large, bleb-like forms with an average radius of 260 ± 100 nm (Fig. 36B-D; 180 mucins per μm^2). Upon crossover into the brush regime, the bleb frequency plummeted precipitously, consistent with the model's prediction of a quadratic rise in the necessary bleb maintenance pressure (Fig. 36B, D).

The glycocalyx polymer model predicted a much different dependence of tubule projection on mucin density. The predicted point force required for maintaining an extended tubule decreased progressively with high mucin densities and exhibited no sharp transitions (Fig. 35D). Accordingly, the frequency of cell-surface tubules observed in our sorted cell populations increased steadily with mucin density throughout the mushroom and brush regimes until the cell was fully saturated with tubes at very high mucin densities (Fig. 36B-E). Notably, theory predicted that at these high densities, the required force for tubule extension is comparable to the polymerization force of a single cytoskeletal filament, ~1 pN (Footer et al., 2007). Based on the experimentally measured mucin densities, we estimated the theoretical point force, f, required to maintain tubules. Remarkably, the experimentally observed tube frequency had a nearly perfect inverse correlation with the theoretical point force (Fig. 36F). The Pearson's correlation coefficient describing the relationship between tube density and $1/f$ was 0.97.

The polymer model also predicted that the spontaneous curvatures generated by high mucin surface densities exceeded the curvature of finger-like projections that we observed on the cell surface. We noted that the tubular membrane projections on our cells typically contained a filamentous actin (F-actin) core and did not contain microtubules (Fig. 37A, B, Fig. 44A-D). Disruption of F-actin assembly with the drug Latrunculin A (LatA) led to a reduction in tubule diameter by approximately 30 nm (Fig. 37C, D and Fig. 44E, F),

indicating that the mucin-induced spontaneous curvature exceeded the curvature of the stable, actin-filled projections. It should be noted that our measurement of LatA-treated cells likely excluded very thin and delicate membrane tubules that were difficult to preserve throughout the SEM sample preparation. Nevertheless, these results clearly indicated that spontaneous curvatures generated by the glycocalyx can meet or exceed the curvature requirements for thin, finger-like projections, such as microtubules, cilia, filapodia, axons, and cytonemes, which have characteristic diameters of approximately 100-200 nm.

Membrane instabilities and extracellular vesicle generation:

We next considered whether other functional membrane shapes could be generated through actions of the glycocalyx. We noted that a progressive increase in spontaneous curvature has been known to trigger membrane instabilities and morphological changes in membrane vesicles (Campelo and Hernández-Machado, 2007; Tsafrir et al., 2001). Therefore, we reasoned that membrane instabilities could arise if the F-actin cores that physiologically resist the spontaneous curvatures of mucins were disrupted. Indeed, our model suggested that ~400 mucins per μm^2 or more would be sufficient to drive membrane instabilities in tubules. Accordingly, we observed that LatA treatment triggered formation of pearled and undulating structures that are characteristic of membrane instabilities (Fig. 37D).

Deuling, Helfrich, and others theoretically considered instabilities in membrane tubules with volume to area ratio, λ, and found that for certain spontaneous curvatures, c_0, the membrane bending energy vanished through the adoption of one of three "Delaunay" shapes: a cylinder for $c_0 = 1/2\lambda$ (Shape 1), a smoothly varying set of unduloids for $1/2\lambda < c_0 < 2/3\lambda$ (Shape 2), and a set of equal-sized "pearls" for $c_0 = 2/3\lambda$ (Shape 3) (Campelo and Hernández-Machado, 2007; Tsafrir et al., 2001). For spontaneous curvatures that exceeded $2/3\lambda$, the lowest energy shapes that satisfied the constraints of volume and surface area were found to include a set of small pearls of the preferred curvature with one or more big pearls necessary to hold excess volume (Shape 4) and a set of pearls with a gradient in size (Shape 5) (Campelo and Hernández-Machado, 2007; Tsafrir et al., 2001). We evaluated whether the minimal energy surfaces, Shapes 1-5, would be formed on cells expressing moderate to high levels of mucin without exogenous treatments, and found commonplace examples of each expected shape (Fig. 37E). The observation of these shapes provided a compelling argument that membrane instabilities can be driven by specific compositions of the glycocalyx.

Remarkably, we discovered that membrane pearling was an intermediate step towards the secretion of extracellular vesicles directly from the plasma membrane (Fig. 37F).

Compared to controls, the conditioned media from Muc1-42TR-expressing cells contained massive concentrations of particles ranging in size from approximately 100-nm to 400-nm (Fig. 5G), which is characteristic of microvesicles (Pol et al., 2016). Particle generation was further enhanced by LatA treatment to disrupt the supporting F-actin cores of surface projections and locally destabilize the plasma membrane (Fig. 37H). Cryo-transmission electron microscopy (cryo-TEM) confirmed that the secreted particles were indeed membrane vesicles and grafted with a distinct glycocalyx ultrastructure on their surfaces (Fig. 37I). These observations are consistent with previous reports of vesicle generation from microvilli in enterocytes and other mucin expressing cells (McConnell et al., 2009). However, and without intending to be bound by any particular theofy, our results now suggest a possible three-step mechanism for microvesicle generation: (1) cytoskeletal filaments help extend and stabilize long and thin protrusions from the plasma membrane in a glycocalyx-dependent manner; (2) following disassembly of the cytoskeletal core, spontaneous curvature imposed by the glycocalyx induces membrane instabilities of the tubules; and (3) membrane pearls pinch off to release vesicles (Fig. 5E, F).

DISCUSSION

The description presented in this Part IV implicates an entropic mechanism through which the glycocalyx can strongly influence the favorability of diverse plasma membrane shapes and protrusions. The morphological changes regulated by the glycocalyx could, in principle, have broad consequences on membrane processes, ranging from absorption and secretion to cellular communication, signaling, and motility (Lange, 2011; Paluch and Raz, 2013; Sauvanet et al., 2015; Schmick and Bastiaens, 2014). Given that glycosylation changes dramatically and in tandem with cell fate transitions (Buck et al., 1971; Freeze, 2013; Satomaa et al., 2009), and that the pool of monomers for construction of glycocalyx polymers is tightly coupled to specific metabolic programs (Dennis et al., 2009; Koistinen et al., 2015; Ying et al., 2012), this Part IV raises the intriguing possibility that the glycocalyx may serve as a conduit linking physical morphology to specific cell states.

Contemporary frameworks for understanding membrane shape regulation largely lack a physical description of the glycocalyx. However, long-chain biopolymers in the glycocalyx are almost universally found anchored to the surfaces of curved membrane features and cell-surface organelles (Bennett et al., 2001; Button et al., 2012; Evanko et al., 2007; Fowke et al., 2017; Hattrup and Gendler, 2008; Kesavan et al., 2009; Kesimer et al., 2013; Makabe Sayoko et al., 2006; van den Pol and Kim, 1993; Zhang et al., 1992). The results in this Part IV

suggests that the principles and theories of polymer physics can be adopted to understand, at least to a first approximation, the physical regulation of membrane shape generation by the glycocalyx. A model of end-anchored polymer mushrooms and polymer brushes is a simple physical representation of the glycocalyx. The actual glycocalyx architecture can include additional hierarchies of crosslinking, entanglement, and molecular inhomogeneity (Tammi et al., 2002). However, the nearly perfect inverse relationships between the force requirements for membrane extension, as estimated using a relatively simple model of the glycocalyx, and the experimentally observed frequencies of these extensions argue that at least some of the physical behaviors of the glycocalyx can be captured using polymer network models. Indeed, we found that glycocalyx polymer extension correlates with cell surface density according to the classic scaling laws developed by de Gennes and others for polymer brushes (Gennes, 1979; Zhulina and Borisov, 1996).

How the glycocalyx and intracellular shape-generating processes coordinate in space and time to control membrane protrusions is not fully resolved. In particular, the Rho family of GTPases are master regulators of cytoskeletal dynamics and cell-surface morphology (Hall, 1998). The description in this Part IV suggests that by modulating the barrier to membrane bending, the glycocalyx primes the membrane for expansion into specific types of spherical or tubular forms that are subject to regulation by Rho GTPases. This integrated view suggests that perturbation of normal cell-surface morphology could be achieved through deregulation of intracellular shape generating processes, glycocalyx polymer assembly, or both. For instance, deregulation of Rho GTPase signaling, cytoskeletal dynamics, and glycocalyx assembly are all common hallmarks of cancer cells (Paszek et al., 2014; Pinho and Reis, 2015; Porter et al., 2016; Yamaguchi and Condeelis, 2007) and may each contribute to the unique cell-surface dynamics that contribute to the lethality of metastatic cancer cells.

Bending of surfaces by anchored polymers is a general physical phenomenon. As such, membrane shape regulation by the glycocalyx could be a universal feature relevant in all cell types. Future efforts may unravel physical function of the glycocalyx in the biogenesis of specific membrane organelles and signaling structures, including cilia, axons, cytonemes, and microvilli. Nevertheless, the description in this Part IV supports a more holistic model of membrane shape regulation that includes consideration of forces on both the intracellular and extracellular faces of the plasma membrane.

METHODS

Antibodies and reagents. The following antibodies were used: FITC-Human CD227 (Muc1) (559774, BD Biosciences), Human CD227 (555925, BD Biosciences) (Muc1), Alexa Flour 488 Human Podocalyxin (222328, R&D Systems), Actin (sc1615, Santa Cruz), GFP (4B10, 2955S, Cell Signaling), 6xHis (9000012, BD Biosciences), Goat anti-Mouse IgG-HRP (sc-2005, Santa Cruz), Mouse anti-Goat IgG-HRP (sc-2354, Santa Cruz). Lectins used were: Biotinylated Peanut Agglutinin (PNA; B-1075, Vector Laboratories), CF568 PNA (29061, Biotium), CF640R PNA (29063, Biotium), CF633 Wheat Germ Agglutinin (WGA; 29024, Biotium). Biotinylated lectins were detected using ExtrAvidin-Peroxidase (E2886, Sigma). Hyaluronic acid (HA) was probed in blots with fluorescently labeled or biotinylated bovine nasal hyaluronic acid binding protein (HABP; Millipore). Biotin-HABP was detected with horseradish peroxidase conjugated streptavidin (HRP-streptavidin; R&D Systems). For HA ELISAs, the DuoSet Hyaluronan kit was from R&D Systems. Actin depolymerization was induced through treatment with Latrunculin A (LatA; *76343-93-6; Cayman Chemicals)*.

For formation of giant unilamellar vesicles (GUVs), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) and 1,2-dioleoyl-*sn*-glycero-3-((*N*-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl), with nickel salt (DOGS-NTA-Ni) were purchased from Avanti Polar Lipids; 2-(4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-pentanoyl)-1-hexadecanoyl-sn-glycero-2-phosphocholine (Bodipy-PC) was purchased from Invitrogen; His-tagged recombinant human Podocalyxin (Ser23 – Arg427; accession number AAB61574.1) was from R&D Systems; and His-tagged human serum albumin (accession number NP_000468) was from ACROBiosystems.

GFP binding protein (nanobody) came from Chromotek. NHS-esters of Alexa Fluor 488, Alexa Fluor 568, and Alexa Fluor 647 were from Invitrogen. Electron microscopy-grade 16% paraformaldehyde, 10% glutaraldehyde, and 2% OsO_4 for scanning electron microscopy (SEM) were obtained from Electron Microscopy Sciences.

Cloning and constructs. cDNAs for cytoplasmic-tail-deleted human Muc1 with 42 tandem repeats (Muc1-42TR ΔCT), Muc1-42TR polymer domain fusion with the TM21 synthetic membrane domain (Muc1-42TR TM21), cytoplasmic-tail-deleted human Podocalyxin (S/T-Rich ΔCT) were generated and cloned into the tetracycline-inducible PiggyBac expression vector (pPB TetOn Puro) or mammalian expression vector pcDNA3.1 as previously described (Paszek et al., 2014; Shurer et al.). To make lentiviral vector pLV Hygro TetOn HAS3, the cDNA for human HAS3 (accession NP_005320) was obtained from R&D Systems and amplified via PCR with the forward primer, 5'-GGCACCTCGAGGATGCCGGTGCAGCTGACGACA-3' (SEQ ID NO:88), and reverse

primer, 5'-GGCAGAATTCTTACACCTCAGCAAAAGCCAAGCT - 3' (SEQ ID NO:89). The PCR product was cloned into pJET1.2 (ThermoFisher) according to manufacturer's protocol, and subcloned into the AbsI and EcoRI sites of pLV Hygro TetOn (Paszek et al., 2012). For generation of pPB Muc1 GFP ΔCT TetOn Puro with varying number of tandem repeats, the cDNA for mOxGFP (Addgene #68070; heretofore mOxGFP is referred to as GFP) was amplified with primers: 5'-GGCAGCTCAGCTATGGTGTCCAAGGGCGAGGAGCTGT-3' ((SEQ ID NO:90) forward) and 5'- GGCAGCTGAGCCCTTATACAGCTCGTCCATGCCGTGAGT-3' ((SEQ ID NO:91) reverse). The PCR product was cloned into pJET1.2 and subcloned non-directionally into the BlpI site of pPB Muc1-42TR ΔCT TetOn Puro. For constructs with 10 and 42 native tandem repeats (PDTRPAPGSTAPPAHGVTSA ((SEQ ID NO:8)), synthetic cDNAs for the desired repeat units were generated through custom gene synthesis (General Biosystems) and cloned in place of the tandem repeats in pPB Muc1 GFP ΔCT TetOn Puro using the BamHI and Bsu36I restriction sites. Muc1 tandem repeats were deleted through Q5 site directed mutagenesis with 5'-TGGAGGAGCCTCAGGCATACTTTATTG-3' (SEQ ID NO:92) forward) and 5'-CCACCGCCGACCGAGGTGACATCCTG-3' ((SEQ ID NO:93) reverse) primers to generate pPB Muc1 0TR GFP ΔCT TetOn Puro. To add a SumoStar tag to the Muc1-42TR GFP ΔCT N-terminus, a cDNA encoding the IgG kappa leader sequence, SumoStar tag, and Muc1 N-terminus was generated through custom gene synthesis (General Biosystems) and inserted in place of the Muc1 N-terminus in pPB Muc1 GFP TetOn Puro using the BamHI and BsrGI restriction sites. For recombinant production of the mucin polymer domain, 42 tandem repeats from Muc1 were fused to an N-terminal S6 tag (GDSLSWLLRLLN) and C-terminal 10x-histidine purification tag to make Muc1-42TR 10X His. To insert the S6 tag, Q5 site directed mutagenesis was performed using 5'-GTTGCGACTGCTTAACGGACAGATCTCGATGGTGAGC-3' (SEQ ID NO:94) forward) AND 5'-AGCCAGCTCAGGGAATCCCCAGCATTCTTCTCAGTAGAG-3' ((SEQ ID NO:95) reverse) on a pcDNA3.1 plasmid containing the Muc1 N-terminus from pPB Muc1-42TR ΔCT TetOn Puro between BamHI and BglII sites. The S6 tag was subsequently cut at these sites and replaced in the Muc1-42TR ΔCT N-terminus in pPB Muc1-42TR ΔCT TetOn Puro. The 10x-histidine tag was added by annealing the oligos, 5'-TCAGGCCACCACCACCATCACCATCATCACCACCATTAGGG-3' (SEQ ID NO:96) and 3'-CCGGTGGTGGTGGTAGTGGTAGTAGTGGTGGTAATCCCTTAA-5' (SEQ ID NO:97), and inserting in place of the Muc1-42TR ΔCT C-terminus in pPB Muc1-42TR ΔCT TetOn Puro using the Bsu36I and EcoRI restriction sites.

Cell lines and culture. MCF10A and HEK293T cells were obtained from ATCC. MCF10A cells were cultured in DMEM/F12 media supplemented with 5% horse serum, 20 ng/mL EGF, 10 μg/ml insulin, 500 ng/mL hydrocortisone, 100 ng/mL cholera toxin and penicillin/streptomycin. HEK293T cells were cultured in DMEM high glucose supplemented with 10% fetal bovine serum and penicillin/streptomycin. Equine synoviocytes were cultured in low glucose (1.0 g/L) DMEM media supplemented with 40 mM HEPES, 4 mM L-Glutamine, 110 mg/L sodium pyruvate, 10% fetal bovine serum and penicillin/streptomycin. Subculture of the synoviocytes was performed every 3–4 days. All adherent cells were maintained at 37°C, 5% CO_2, and 90% RH. Suspension-adapted 293F cells obtained from Thermo Fisher (R79007) and were maintained in Freestyle 293F Expression Medium (Thermo Fisher, 12338018) in spinner flasks at 37°C, 8% CO_2, 120 RPM, and 80% RH according to manufacturer's protocol. Stable MCF10A, primary equine synoviocyte, and 293F cells expressing the rtTA-M2 tetracycline transactivator were prepared by lentiviral transduction using the pLV rtTA-NeoR plasmid as previously described (Paszek et al., 2012). For preparation of mucin expressing cell lines, plasmids with ITR-flanked expression cassettes (i.e. PiggyBac vectors) were co-transfected with the PiggyBac hyperactive transposase using Nucleofection Kit V (Lonza) or FreeStyle Max Reagent (Thermo Fisher) according to manufacturer's protocols and selected with 1 μg/ml puromycin or 200 μg/mL hygromycin.

Equine synovial tissue resection and primary synoviocyte isolation. Primary equine synoviocytes were obtained from the shoulder, stifle, carpal, tarsal and fetlock joints of a yearling horse (*Equus caballus*). To isolate the fibroblast-like type B synovial cells (synoviocytes), synovial membrane tissues were digested with 0.15% collagenase (Worthington Biochemical, Lakewood, NJ) supplemented with 0.015% DNase I (Roche, Indianapolis, IN) for 3 h at 37 °C in Ham's F12 media, followed by filtration and centrifugation at 250x *g* for 10 minutes as previously described (Saxer et al., 2001).

Freshly resected synovial tissues were either incubated for 30 min in Ham's F12 media with or without 1 U/mL Hyaluronidase (Sigma) and fixed or immediately fixed for 24 h with 4% paraformaldehyde and 1% glutaraldehyde in PBS. Tissues were then either processed for SEM or reduced with 0.1 mg/mL $NaBH_4$ for 20 min on ice and further processed for confocal imaging.

Scanning electron microscopy (SEM) and analysis. All samples were fixed for 24 h with 4% paraformaldehyde and 1% glutaraldehyde in PBS, post-fixed for 45 min with 1% osmium tetraoxide in dH_2O, washed and subsequently dehydrated stepwise in ethanol of

25%, 50%, 70%, 95%, 100%, 100% before drying in a critical point dryer (CPD 030, Bal-Tec). Samples were coated with gold-palladium in a Desk V sputter system (Denton Vacuum) and imaged on a field emission scanning electron microscope (Mira3 FE-SEM, Tescan or FE-SEM LEO 1550, Carl Zeiss Inc.). For actin depolymerization studies, cells were treated for 60 min with 10 μM LatA before fixation, where indicated.

Cellular tube density, diameter, and length were analyzed in ImageJ Fiji (Schindelin et al., 2012). For quantification of tube density per area, a ~2 μm x 2 μm region of interest was drawn and the encompassed tubes counted manually. Tube diameter was measured by drawing a strain line through the tube cross section at its mid-point. Tube length was measured for tubes extending approximately parallel to the image plane, as identified by visual inspection, using the ImageJ line segment tool.

Confocal microscopy for cells and tissues. Cells were plated at 5,000 cells/cm^2 and subsequently induced with 0.2 μg/mL of doxycycline for 24 h before being fixed with 4% paraformaldehyde. Antibodies were diluted 1:200 in 5% normal goat serum PBS and incubated overnight at 4°C. Lectins were diluted to 1 μg/mL in 5% normal goat serum PBS and incubated for 2 h at room temperature. For hyaluronic acid staining of cells and tissues, HABP was diluted to 0.125 μg/ml in 0.5% normal goat serum in PBS and incubated on samples for 24 h. Cell samples were imaged on a Zeiss LSM inverted 880 confocal microscope using a 40x water immersion objective (NA 1.1). In addition to HABP, $NaBH_4$-treated tissues were stained with 1 μg/mL Hoechst for 10 min and imaged on a Zeiss 880 upright confocal microscope with a 40x water dipping lens. Unstained tissue collagen was visualized with second harmonic generation using non-descan detectors.

Immuno- and lectin blot analysis. Cells were plated at 20,000 cells/cm^2 and induced with 0.2 μg/mL doxycycline for 24 h before lysis with Tris-Triton lysis buffer (Abcam). Lysates were separated on Nupage 4-12% Bis-Tris or 3-8% Tris-Acetate gels (Thermo Fisher) and transferred to PVDF membranes. Primary antibodies were diluted 1:1000 and lectins were diluted to 1 μg/mL in 3% BSA TBST and incubated 4 h at room temperature or overnight at 4°C. Secondary antibodies or ExtrAvidin were diluted 1:2000 in 3% BSA TBST and incubated for 2 h at room temperature. Blots were developed in Clarity ECL (BioRad) substrate, imaged on a ChemiDoc (BioRad) documentation system, and quantified in ImageJ Fiji (Schindelin et al., 2012).

Flow cytometry. Cells were plated at 20,000 cells/cm^2 and grown for 24 h. Cells were then induced with 0.2 μg/mL doxycycline for 24 h. Adherent cells were non-enzymatically detached by incubating with 1 mM EGTA in PBS at 37°C for 20 min and

added to the population of floating cells, if present. Antibodies were diluted 1:200 and lectins were diluted to 1 μg/mL in 0.5% BSA PBS and incubated with cells at 4°C for 30 min. The BD Accuri C6 flow cytometer was used for analysis.

Analysis of HA synthesis and molecular size. Control and lentiviral transduced MCF10A and primary equine synoviocytes were plated and induced with 0.2 μg/mL doxycycline for 24 h. Total levels of HA secreted into the cell culture media were measured via the DuoSet Hyaluronan ELISA kit following manufacturer's protocol. Briefly, a 96-well microplate was coated with recombinant human Aggrecan. HA in cell culture media was captured by the coated Aggrecan and detected with Biotin-HABP/HRP-Streptavidin. HA concentration was measured using *S. pyogenes* HA standard (R&D Systems). HA molecular mass was assayed by electrophoresis and blot analysis essentially as described (Yuan et al., 2013), using agarose instead of polyacrylamide for gel electrophoresis. Briefly, cell culture media containing HA was loaded in a 0.6% agarose gel in TBE buffer. Following electrophoresis, samples were transferred to HyBond N+ membrane (GE Healthcare). HA was probed with biotin-HABP (0.125 μg/ml in 0.1% BSA-PBS, 1 h) and subsequently detected with HRP-Streptavidin (0.025 μg/ml in 0.1% BSA-PBS, 1 h). Blots were developed in ECL substrate (Amresco), imaged on a ChemiDoc (BioRad) documentation system, and quantified in ImageJ Fiji (Schindelin et al., 2012).

Analysis of mucin radius of gyration. The Muc1 polymer domain with 42 tandem repeats (S6 Muc1-42TR 10xHis) was produced recombinantly in suspension adapted Freestyle 293F cells. Stable 293F cell lines were prepared with the pPB Muc1-42TR 10xHis Puro TetOn Puro vector as described above. Production of Muc1 biopolymer was induced with 1 μg/mL doxycycline in 30 mL of suspension culture in Freestyle 293F media. Induced media was collected after 24 h and purified on HisPur Ni-NTA resin (Thermo Fisher) according to standard protocols. Briefly, 1 mL bed volume of Ni-NTA resin was rinsed with equilibration buffer (20 mM sodium phosphate, 0.5 M NaCl, pH = 7.4). Equilibrated resin was incubated overnight at 4°C with 10 mL harvested 293F media diluted in 30 mL of equilibration buffer. Beads were washed in equilibration buffer with 5 mM imidazole and eluted in equilibration buffer with 500 mM imidazole. Eluted protein was dialyzed against PBS and analyzed by SDS-PAGE. Gels were stained with Sypro Ruby (Thermo Fisher) according to manufacturer's instructions to confirm protein size and purity. Gels were blotted and probed with Muc1 and His antibodies to confirm mucin identity and PNA lectin to confirm mucin *O*-glycosylation. Purified recombinant Muc1 was dialyzed against PBS to remove imidazole.

The radius of gyration of the recombinant Muc1 polymer domain was measured with size-exclusion chromatography-coupled to multiangle light scattering (SEC-MALS). Purified protein (40 µL of Muc1 with a concentration of 5 µg/µL) was subjected to SEC using a Superdex 200 Increase 10/300 column (GE Healthcare) equilibrated in MALS buffer (20 mM sodium phosphate, 0.5 M NaCl, pH 7.4). The SEC was coupled to a static 18-angle light scattering detector (DAWN HELEOS-II) and a refractive index detector (Optilab T-rEX, Wyatt Technology). Data were collected every second at a flow rate of 0.7 mL/min. Data analysis was carried out using ASTRA VI, yielding the molar mass, mass distribution (polydispersity), and radius of gyration of the sample (32.0 nm ± 0.4%). For normalization of the light scattering detectors and data quality control, monomeric BSA (Sigma) was used.

Variation of mucin lengths and cell-surface densities. *Mucin lengths*: MCF10As expressing Muc1 mOxGFP with 0, 10, or 42 tandem repeats were sorted for similar levels of GFP on a BD FACs Aria II. Stable populations were created from these sorted lines. Cells were plated onto 8 mm coverslips at 10,000 cells/cm^2 for 16-18 h, then induced with 0.2 µg/mL of doxycycline for 24 h and fixed for SEM analysis.

Mucin cell surface density: A nanobody with an approximate size of 2 nm (15 kDa) and picomolar affinity for GFP was obtained from ChromoTech and labeled with NHS-Alexa Fluor 647 according to manufacturer's protocol. MCF10A cells expressing Muc1 mOxGFP with 42 tandem repeats were labeled in 5 µg/ml 647-nanobody for 20 min on ice to label only cell surface mucins. Cells were sorted onto poly-l-lysine treated 8 mm coverslips at 5,000 to 10,000 cells/cm^2 for SEM, allowed to adhere for 4 h at 37°C, and fixed for SEM imaging. Alternatively, cells were sorted into 1.7 mL Eppendorf tubes, resuspended in 100 µL 0.5% BSA PBS, and lysed with 100 µL 2x RIPA lysis buffer for estimation of mucin surface densities via SDS-PAGE. Lysed samples were run simultaneously with Alexa Fluor 647-nanobody standards of known molecular concentration. Nanobody fluorescence in lysed samples and standards were imaged on a Typhoon 9400 imaging system (GE Healthcare). Total fluorescence in each sample or standard was quantified in ImageJ Fiji (Schindelin et al., 2012). A standard curve was constructed by relating fluorescence from nanobody standards to their known concentration. The number of labeled mucins in each lysate were estimated based on the standard curve. The mucin surface density was estimated by dividing the total number of mucins by the known number of cells in each sample and their average surface area of 5,000 µm^2 based on an average radius of 20 µm and spherically shaped wild-type cells in suspension. A standard curve was constructed based on the number of mucins per area and the known mean fluorescence signal from the FACS collected population. This standard

curve was then applied to calculate the number of mucins per area of populations collected subsequently.

Giant unilamellar vesicles. *Preparation*. Giant Unilamellar Vesicles (GUVs) were prepared by electroformation as described previously (Angelova and Dimitrov, 1986). Briefly, lipids and dye dissolved in chloroform were spread on glass slides coated with ITO (Indium-Tin-Oxide). The slides were placed under vacuum for 2 h to remove all traces of organic solvents. The lipid films were hydrated and swelled in 120 mM sucrose at 55°C. GUVs were electroformed by the application of an oscillating potential of 1.4 V (peak-to-peak) and 12 Hz for 3 h (Busch et al., 2015). GUVs compositions were prepared with DOPC and increasing molar fractions of DOGS-Ni-NTA lipid (5, 10, 15, and 20 mol%). Bodipy-PC was used to label the lipids at a dye/lipid ratio of 1/2500. Recombinant His-tagged Podocalyxin and human serum albumin (HSA) were conjugated with NHS-Alexa Fluor 568, and the degree of labelling quantified according to the manufacturer's protocol. GUVs were diluted in 20 mM HEPES, 50 mM NaCl, pH = 7.4 (120 mOsm) and then mixed with labeled Podocalyxin (~2 μM) or HSA (0.125 or 0.375 μM) for at least 20 minutes before imaging (GUVs/proteins = 1/1 by volume).

Imaging and analysis. GUVs were imaged on a Nikon C2plus confocal microscope using a 60x water immersion objective (NA 1.2). Lipids and (Bodipy-PC) and protein (Alexa Fluor 568) were imaged through excitation at wavelength λ= 488 and 561 nm, respectively. Dye fluorescent intensity was measured by taking 5 different line scans across the GUV in ImageJ Fiji (Schindelin et al., 2012). The intensity profile of each line was analyzed using Mathematica 10.3, where the integral of the intensity peak was calculated and averaged for 5 different lines per GUV.

Expansion microscopy. Expansion microscopy (ExM) was performed as described previously (Tillberg et al., 2016) and involved steps of anchoring fluorescent dyes and proteins, gelation, digestion and expansion to achieve dye retention and separation. Briefly, fixed and stained cells were anchored with 0.1 mg/ml Acryloyl-X, SE (6-((acryloyl)amino)hexanoic acid, succinimidyl ester (ThermoFisher) in PBS for 16 h at RT, washed twice and further incubated 1 h at 37°C in a monomer solution (1 × PBS, 2 M NaCl, 8.625% (w/w) sodium acrylate, 2.5% (w/w) acrylamide, 0.15% (w/w) N,N′-methylenebisacrylamide) mixed with ammonium persulfate 0.2% (w/w) initiator and tetramethylethylenediamine 0.2% (w/w) accelerator for gelation. For digestion, gelled samples were gently transferred into 6 well glass bottom plates (Cellvis) and treated with Proteinase K (New England Biolabs) at 8 units/mL in digestion buffer (50 mM Tris (pH 8), 1

mM EDTA, 0.5% Triton X-100, 1 M NaCl) for 16 h at room temperature. For expansion, digested gels were washed in large excess volume of ddH_2O for 1 h. This was repeated 4 – 6 times until the expansion plateaued. Samples were imaged on a Zeiss LSM inverted 880 confocal microscope using a 40x water immersion objective (NA 1.1) in Airyscan mode to optimize resolution.

Isolation of extracellular vesicles. Cell were plated at 10,000 cells/cm^2 in appropriate dishes. Following induction with 1 μg/ml doxycycline for 18 h, cells were rinsed with PBS twice then serum-starved for an additional 6 h with 1 μg/mL doxycycline treatment. Conditioned media from serum-starved cells was clarified by pelleting cellular debris through two consecutive centrifugations at 600x *g* for 5 min.

Nanoparticle tracking analysis. Extracellular vesicles in the clarified media were analyzed using a Malvern NS300 NanoSight. Imaging was performed for 60 s with five captures per sample. Particle analysis was performed using Malvern Nanoparticle Tracking Analysis software.

Plunge-freezing vitrification. From clarified media, 3-5 μl of sample was pipetted onto holey carbon-coated 200 mesh copper grids (Quantifoil Micro Tools, Jena, Germany) with hole sizes of ~2 μm. The grids were blotted from the reverse side and immediately plunged into a liquid ethane/propane mixture cooled to liquid nitrogen temperature using a custom-built vitrification device (MPI, Martinsried, Germany). The plunge-frozen grids were stored in sealed cryo-boxes in liquid nitrogen until used.

Cryogenic transmission electron microscopy. Cryogenic transmission electron microscopy (cryo-TEM) was performed on a Titan Themis (Thermo Fisher Scientific, Waltham, MA) operated at 300 kV in energy-filtered mode, equipped with a field-emission gun, and 3838x3710 pixel Gatan K2 Summit direct detector camera (Gatan, Pleasanton, CA) operating in Counted, dose-fractionated modes. Images were collected at a defoci of between -1 μm and -3 μm. Images were binned by 2, resulting in pixel sizes of 0.72 -1.1 nm.

Statistics. Statistics were calculated in Graphpad Prism. One-way ANOVA and post-hoc two-tailed student's t-test were used where appropriate as indicated by figure legends. For boxplots - center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, and notches, where shown, indicate the 95% confidence interval.

REFERENCES

Agrawal, B., Krantz, M.J., Parker, J., and Longenecker, B.M. (1998). Expression of MUC1 Mucin on Activated Human T Cells: Implications for a Role of MUC1 in Normal Immune Regulation. Cancer Res. *58*, 4079–4081.

Alexander, S. (1977). Adsorption of chain molecules with a polar head a scaling description. J. Phys. *38*, 983–987.

Angelova, M.I., and Dimitrov, D.S. (1986). Liposome electroformation. Faraday Discuss. Chem. Soc. *81*, 303–311.

Antonyak, M.A., Li, B., Boroughs, L.K., Johnson, J.L., Druso, J.E., Bryant, K.L., Holowka, D.A., and Cerione, R.A. (2011). Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fibronectin to recipient cells. Proc. Natl. Acad. Sci. *108*, 4852–4857.

Becker, A., Thakur, B.K., Weiss, J.M., Kim, H.S., Peinado, H., and Lyden, D. (2016). Extracellular Vesicles in Cancer: Cell-to-Cell Mediators of Metastasis. Cancer Cell *30*, 836–848.

Bennett, R., Järvelä, T., Engelhardt, P., Kostamovaara, L., Sparks, P., Carpén, O., Turunen, O., and Vaheri, A. (2001). Mucin MUC1 is seen in cell surface protrusions together with ezrin in immunoelectron tomography and is concentrated at tips of filopodial protrusions in MCF-7 breast carcinoma cells. J. Histochem. Cytochem. Off. J. Histochem. Soc. *49*, 67–77.

Bergert, M., Erzberger, A., Desai, R.A., Aspalter, I.M., Oates, A.C., Charras, G., Salbreux, G., and Paluch, E.K. (2015). Force transmission during adhesion-independent migration. Nat. Cell Biol. *17*, 524–529.

Bischoff, M., Gradilla, A.-C., Seijo, I., Andrés, G., Rodríguez-Navas, C., González-Méndez, L., and Guerrero, I. (2013). Cytonemes are required for the establishment of a normal Hedgehog morphogen gradient in *Drosophila* epithelia. Nat. Cell Biol. *15*, 1269–1281.

Buck, C.A., Glick, M.C., and Warren, L. (1971). Glycopeptides from the surface of control and virus-transformed cells. Science *172*, 169–171.

Busch, D.J., Houser, J.R., Hayden, C.C., Sherman, M.B., Lafer, E.M., and Stachowiak, J.C. (2015). Intrinsically disordered proteins drive membrane curvature. Nat. Commun. *6*, 7875.

Button, B., Cai, L.-H., Ehre, C., Kesimer, M., Hill, D.B., Sheehan, J.K., Boucher, R.C., and Rubinstein, M. (2012). A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia. Science *337*, 937–941.

Campelo, F., and Hernández-Machado, A. (2007). Model for Curvature-Driven Pearling Instability in Membranes. Phys. Rev. Lett. *99*, 088101.

Charras, G.T., Yarrow, J.C., Horton, M.A., Mahadevan, L., and Mitchison, T.J. (2005). Non-equilibration of hydrostatic pressure in blebbing cells. Nature *435*, 365–369.

Charras, G.T., Coughlin, M., Mitchison, T.J., and Mahadevan, L. (2008). Life and Times of a Cellular Bleb. Biophys. J. *94*, 1836–1853.

Chen, W.-L., Cordero, R., Tran, H., and Ober, C.K. (2017). 50th Anniversary Perspective: Polymer Brushes: Novel Surfaces for Future Materials. Macromolecules *50*, 4089–4113.

Cloosen, S., Thio, M., Vanclée, A., Leeuwen, V., M, E.B., Senden-Gijsbers, B.L.M.G., Oving, E.B.H., Germeraad, W.T.V., and Bos, G.M.J. (2004). Mucin-1 is expressed on dendritic cells, both in vitro and in vivo. Int. Immunol. *16*, 1561–1571.

D'Aloia, M.M., Zizzari, I.G., Sacchetti, B., Pierelli, L., and Alimandi, M. (2018). CAR-T cells: the long and winding road to solid tumors. Cell Death Dis. *9*, 282.

Dennis, J.W., Nabi, I.R., and Demetriou, M. (2009). Metabolism, Cell Surface Organization, and Disease. Cell *139*, 1229–1241.

Evanko, S.P., Tammi, M.I., Tammi, R.H., and Wight, T.N. (2007). Hyaluronan-Dependent Pericellular Matrix. Adv. Drug Deliv. Rev. *59*, 1351–1365.

Footer, M.J., Kerssemakers, J.W.J., Theriot, J.A., and Dogterom, M. (2007). Direct measurement of force generation by actin filament polymerization using an optical trap. Proc. Natl. Acad. Sci. *104*, 2181–2186.

Fowke, T.M., Karunasinghe, R.N., Bai, J.-Z., Jordan, S., Gunn, A.J., and Dean, J.M. (2017). Hyaluronan synthesis by developing cortical neurons *in vitro*. Sci. Rep. *7*, 44135.

Freeze, H.H. (2013). Understanding Human Glycosylation Disorders: Biochemistry Leads the Charge. J. Biol. Chem. *288*, 6936–6945.

Friedl, P., and Wolf, K. (2010). Plasticity of cell migration: a multiscale tuning model. J. Cell Biol. *188*, 11–19.

Gangoda, L., Boukouris, S., Liem, M., Kalra, H., and Mathivanan, S. Extracellular vesicles including exosomes are mediators of signal transduction: Are they protective or pathogenic? PROTEOMICS *15*, 260–271.

Gennes, P.-G. (1979). Scaling Concepts in Polymer Physics (Ithaca, NY: Cornell University Press).

de Gennes, P. (1980). Conformations of polymers attached to an interace. Macromolecules 1069–1075.

Gupton, S.L., and Gertler, F.B. (2007). Filopodia: the fingers that do the walking. Sci. STKE Signal Transduct. Knowl. Environ. *2007*, re5.

Hall, A. (1998). Rho GTPases and the actin cytoskeleton. Science *279*, 509–514.

Hattrup, C.L., and Gendler, S.J. (2008). Structure and Function of the Cell Surface (Tethered) Mucins. Annu. Rev. Physiol. *70*, 431–457.

Hiergeist, C., and Lipowsky, R. (1996). Elastic Properties of Polymer-Decorated Membranes. J. Phys. II *6*, 1465–1481.

Jung, Y., Riven, I., Feigelson, S.W., Kartvelishvily, E., Tohya, K., Miyasaka, M., Alon, R., and Haran, G. (2016). Three-dimensional localization of T-cell receptors in relation to microvilli using a combination of superresolution microscopies. Proc. Natl. Acad. Sci. *113*, E5916–E5924.

Kesavan, G., Sand, F.W., Greiner, T.U., Johansson, J.K., Kobberup, S., Wu, X., Brakebusch, C., and Semb, H. (2009). Cdc42-mediated tubulogenesis controls cell specification. Cell *139*, 791–801.

Kesimer, M., Ehre, C., Burns, K.A., Davis, C.W., Sheehan, J.K., and Pickles, R.J. (2013). Molecular organization of the mucins and glycocalyx underlying mucus transport over mucosal surfaces of the airways. Mucosal Immunol. *6*, 379–392.

Koistinen, V., Kärnä, R., Koistinen, A., Arjonen, A., Tammi, M., and Rilla, K. (2015). Cell protrusions induced by hyaluronan synthase 3 (HAS3) resemble mesothelial microvilli and share cytoskeletal features of filopodia. Exp. Cell Res. *337*, 179–191.

Kornberg, T.B., and Roy, S. (2014). Cytonemes as specialized signaling filopodia. Development *141*, 729–736.

Kramer, R.H., and Nicolson, G.L. (1979). Interactions of tumor cells with vascular endothelial cell monolayers: a model for metastatic invasion. Proc. Natl. Acad. Sci. U. S. A. *76*, 5704–5708.

Kufe, D.W. (2009). Mucins in cancer: function, prognosis and therapy. Nat. Rev. Cancer *9*, nrc2761.

Lange, K. (2011). Fundamental role of microvilli in the main functions of differentiated cells: Outline of an universal regulating and signaling system at the cell periphery. J. Cell. Physiol. *226*, 896–927.

Lee, G.M., Johnstone, B., Jacobson, K., and Caterson, B. (1993). The dynamic structure of the pericellular matrix on living cells. J. Cell Biol. *123*, 1899–1907.

Lipowsky, R. (1995). Bending of Membranes by Anchored Polymers. EPL Europhys. Lett. *30*, 197.

Liu, T.-L., Upadhyayula, S., Milkie, D.E., Singh, V., Wang, K., Swinburne, I.A., Mosaliganti, K.R., Collins, Z.M., Hiscock, T.W., Shea, J., et al. (2018). Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms. Science *360*, eaaq1392.

Makabe Sayoko, Naguro Tomonori, and Stallone Tiziana (2006). Oocyte–follicle cell interactions during ovarian follicle development, as seen by high resolution scanning and transmission electron microscopy in humans. Microsc. Res. Tech. *69*, 436–449.

Marshall, W.F. (2012). Organelle Size Control Systems: From Cell Geometry to Organelle-Directed Medicine. BioEssays News Rev. Mol. Cell. Dev. Biol. *34*, 721–724.

McConnell, R.E., Higginbotham, J.N., Shifrin, D.A., Tabb, D.L., Coffey, R.J., and Tyska, M.J. (2009). The enterocyte microvillus is a vesicle-generating organelle. J. Cell Biol. *185*, 1285–1298.

Milner, S.T. (1991). Polymer brushes. Science *251*, 905–914.

Paluch, E.K., and Raz, E. (2013). The role and regulation of blebs in cell migration. Curr. Opin. Cell Biol. *25*, 582–590.

Paszek, M.J., DuFort, C.C., Rubashkin, M.G., Davidson, M.W., Thorn, K.S., Liphardt, J.T., and Weaver, V.M. (2012). Scanning angle interference microscopy reveals cell dynamics at the nanoscale. Nat. Methods *9*, 825–827.

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. (2014). The cancer glycocalyx mechanically primes integrin-mediated growth and survival. Nature *511*, 319–325.

Pelaseyed, T., Bergström, J.H., Gustafsson, J.K., Ermund, A., Birchenough, G.M.H., Schütte, A., Post, S. van der, Svensson, F., Rodríguez-Piñeiro, A.M., Nyström, E.E.L., et al. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. Immunol. Rev. *260*, 8–20.

Peskin, C.S., Odell, G.M., and Oster, G.F. (1993). Cellular motions and thermal fluctuations: the Brownian ratchet. Biophys. J. *65*, 316–324.

Pilon, C., Levast, B., Meurens, F., Le Vern, Y., Kerboeuf, D., Salmon, H., Velge-Roussel, F., Lebranchu, Y., and Baron, C. (2009). CD40 engagement strongly induces CD25 expression on porcine dendritic cells and polarizes the T cell immune response toward Th1. Mol. Immunol. *46*, 437–447.

Pinho, S.S., and Reis, C.A. (2015). Glycosylation in cancer: mechanisms and clinical implications. Nat. Rev. Cancer *15*, 540–555.

Pol, E. van der, Böing, A.N., Gool, E.L., and Nieuwland, R. (2016). Recent developments in the nomenclature, presence, isolation, detection and clinical impact of extracellular vesicles. J. Thromb. Haemost. *14*, 48–56.

van den Pol, A.N., and Kim, W.T. (1993). NILE/L1 and NCAM-polysialic acid expression on growing axons of isolated neurons. J. Comp. Neurol. *332*, 237–257.

Polefka, T.G., Garrick, R.A., Redwood, W.R., Swislocki, N.I., and Chinard, F.P. (1984). Solute-excluded volumes near the Novikoff cell surface. Am. J. Physiol.-Cell Physiol. *247*, C350–C356.

Porter, A.P., Papaioannou, A., and Malliri, A. (2016). Deregulation of Rho GTPases in cancer. Small GTPases *7*, 123–138.

Satomaa, T., Heiskanen, A., Mikkola, M., Olsson, C., Blomqvist, M., Tiittanen, M., Jaatinen, T., Aitio, O., Olonen, A., Helin, J., et al. (2009). The N-glycome of human embryonic stem cells. BMC Cell Biol. *10*, 42.

Sauvanet, C., Wayt, J., Pelaseyed, T., and Bretscher, A. (2015). Structure, Regulation, and Functional Diversity of Microvilli on the Apical Domain of Epithelial Cells. Annu. Rev. Cell Dev. Biol. *31*, 593–621.

Saxer, R.A., Bent, S.J., Brower-Toland, B.D., Mi, Z., Robbins, P.D., Evans, C.H., and Nixon, A.J. (2001). Gene mediated insulin-like growth factor-I delivery to the synovium. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *19*, 759–767.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schmick, M., and Bastiaens, P.I.H. (2014). The Interdependence of Membrane Shape and Cellular Signal Processing. Cell *156*, 1132–1138.

Shurer, C., Colville, M., Gupta, V., Head, S., Kai, F., Lakins, J., and Paszek, M. A Genetically Encoded Toolbox for Glycocalyx Engineering: Tunable Control of Cell Adhesion, Survival, and Cancer Cell Behaviors. ACS Biomater. Sci. Eng.

Stachowiak, J.C., Hayden, C.C., and Sasaki, D.Y. (2010). Steric confinement of proteins on lipid membranes can drive curvature and tubulation. Proc. Natl. Acad. Sci. *107*, 7781–7786.

Tammi, M.I., Day, A.J., and Turley, E.A. (2002). Hyaluronan and Homeostasis: A Balancing Act. J. Biol. Chem. *277*, 4581–4584.

Tillberg, P.W., Chen, F., Piatkevich, K.D., Zhao, Y., Yu, C.-C. (Jay), English, B.P., Gao, L., Martorell, A., Suk, H.-J., Yoshida, F., et al. (2016). Protein-retention expansion

microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. Nat. Biotechnol. *34*, 987–992.

Tricarico, C., Clancy, J., and D'Souza-Schorey, C. (2017). Biology and biogenesis of shed microvesicles. Small GTPases *8*, 220–232.

Tsafrir, I., Sagi, D., Arzi, T., Guedeau-Boudeville, M.-A., Frette, V., Kandel, D., and Stavans, J. (2001). Pearling Instabilities of Membrane Tubes with Anchored Polymers. Phys. Rev. Lett. *86*, 1138–1141.

Turley, E.A., Wood, D.K., and McCarthy, J.B. (2016). Carcinoma Cell Hyaluronan as a "Portable" Cancerized Prometastatic Microenvironment. Cancer Res. *76*, 2507–2512.

Yamaguchi, H., and Condeelis, J. (2007). Regulation of the actin cytoskeleton in cancer cell migration and invasion. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1773*, 642–652.

Ying, H., Kimmelman, A.C., Lyssiotis, C.A., Hua, S., Chu, G.C., Fletcher-Sananikone, E., Locasale, J.W., Son, J., Zhang, H., Coloff, J.L., et al. (2012). Oncogenic Kras Maintains Pancreatic Tumors through Regulation of Anabolic Glucose Metabolism. Cell *149*, 656–670.

Yuan, H., Tank, M., Alsofyani, A., Shah, N., Talati, N., LoBello, J.C., Kim, J.R., Oonuki, Y., de la Motte, C.A., and Cowman, M.K. (2013). Molecular mass dependence of hyaluronan detection by sandwich ELISA-like assay and membrane blotting using biotinylated hyaluronan binding protein. Glycobiology *23*, 1270–1280.

Zhang, H., Miller, R.H., and Rutishauser, U. (1992). Polysialic acid is required for optimal growth of axons on a neuronal substrate. J. Neurosci. Off. J. Soc. Neurosci. *12*, 3107–3114.

Zhulina, E.B., and Borisov, O.V. (1996). Polyelectrolytes Grafted to Curved Surfaces. Macromolecules *29*, 2618–2626.

Part IV Suppemental Information

Theoretical Considerations

Glycocalyx Polymer Brush Model

Without intending to be constrained by any particular theory, the disclosure provides a model to explain how biopolymers in the glycocalyx can generate entropic driving forces for membrane curvature. The model considers long chain polymers anchored on one end to the plasma membrane. Common examples of long-chain polymers in the glycocalyx include mucins and hyaluronic acid (HA), which we model specifically here. The modeling framework could be similarly applied to other types of glycocalyx polymers, including

polysialic acid and other glycosaminoglycans. Hyaluronic acid is a semi-flexible linear polysaccharide comprised of repeating units of glucuronic acid and N-acetylglucosamine. Mucins have a more complex bottlebrush structure comprised of a central polypeptide backbone and densely clustered glycan side chains along the backbone. Although their structure is complex, bottlebrush polymers can be modelled as effective linear polymers with a monomer size on the order of the side chains (Paturej et al., 2016). Therefore, we consider all glycocalyx polymers in our model to be linear or effectively linear.

Biopolymers in the glycocalyx are anchored to the cell surface in several ways, including through transmembrane anchors, covalent conjugation to integral membrane proteins, and non-covalently to specific transmembrane receptors. Cell surface mucins are anchored directly near their carboxy terminus by a single transmembrane domain. Hyaluronic acid is anchored to the cell surface through specific transmembrane receptors on the cell surface. While it is possible for hyaluronic acid to be anchored at multiple points along the polymer backbone, for simplicity, we consider all glycocalyx polymers to have a single membrane anchor at one end.

The cell surface is also decorated with many types of integral and peripheral membrane proteins. These molecules could also contribute to an entropic pressure on the cell membrane, similar to a 2D gas pressure. To isolate the effects of glycocalyx polymers on the membrane, we did not include possible contributions from other cell surface proteins, as well as intracellular forces. However, the model could be extended to include these additional contributions to the system energy.

Biopolymers have excluded volumes accounting for steric interactions between monomers on the same polymer as well as between monomers on adjacent molecules (de Gennes, 1980). Large negative charges on acidic sugars, such as glucuronic acid and sialic acid, give rise to intramolecular and intermolecular electrostatic interactions (Israels et al., 1994). Finally, the polymers and the brush have entropic contributions due to the elastic energy, which captures the stretch of the molecules (de Gennes, 1980). Embedded in a deformable lipid membrane, the energy of this polymer glycocalyx and that of the membrane can minimize to yield the equilibrium configuration (Lipowsky, 1995; Stachowiak et al., 2012). Hence, in the present model below, we performed an energy minimization of the glycocalyx and the underlying membrane to describe the surface curvature.

Depending on surface density, polymers tethered to a surface exhibit two particular regimes of physical behavior – mushroom and brush. The Flory radius measures the approximate size of an entire polymer, and is given by $R_F \approx l_a N_a^{\nu} = l^{\nu} l_a^{1-\nu}$, where N_a is the

number of monomers in the polymer, l_a is the size of each monomer or effective monomer, l is the fully extended length of the polymer chain, and ν is called the Flory exponent. $\nu \cong 0.6$ for hydrophilic biopolymers in good solvents like water. At low densities, such that intermolecular spacing is larger than the polymer Flory radius, i.e. $C_G < 1/(R_F)^2$, where C_G is biopolymer concentration, biopolymers take up preferable conformations independent of neighbor interactions. In this regime, the flexible molecules can coil up to exhibit mushroom-like structures. On the other hand, at high surface concentrations, when the intermolecular spacing is smaller than the Flory radius, intermolecular interactions can dominate and stretch the biopolymers out into a brush-like structure. The polymer layer extension or thickness, the stored energy, and the generated membrane curvatures exhibit different scaling laws in these regimes, as described below.

In the mushroom regime, the attachment of a biopolymer to a flat, impenetrable surface reduces the number of accessible molecular conformations, cutting down the polymer shapes that penetrate the surface. Curving the impenetrable grafting surface can marginally increase the permissible configurations, and increase the entropy of the polymer. Thus, flexible biopolymers tethered to a deformable membrane can generate curvatures, as described by Lipowsky (Lipowsky, 1995). However, the additional entropy due to membrane curvature is small and consequently, curvatures generated by polymer mushrooms are also small, relative to deformations elicited by intermolecular interactions in polymer brushes. In this mushroom regime, the free energy due to the entropic contribution of each mushroom polymer tethered to a curved membrane is:

$$F_{mushroom} = -TS_{mushroom} \sim -k_B T \frac{2\pi R_{mushroom}}{R}, \quad (1)$$

where the reference configuration is the polymer tethered to a flat surface, $S_{mushroom}$ is the corresponding entropic contribution, $R_{mushroom}$ is the Flory radius of the mushroom-shaped biopolymer, and R is the radius of curvature of the underlying membrane. In the mushroom regime, we consider the formation of spherical membrane structures. The bending energy of the curved membrane is:

$$F_{membrane} = \frac{\kappa}{2C_G R^2}, \quad (2)$$

where κ is the bending stiffness of the membrane bilayer, C_G is the surface density of the biopolymers, and $1/C_G$ is the area available for each polymer. Minimizing the total energy,

$F_{total} = F_{mushroom} + F_{membrane}$ with respect to the radius of curvature, R, as $\partial F_{total}/\partial R = 0$, we obtain the following scaling law for R:

$$R \sim \frac{\kappa}{k_B T} \frac{1}{2\pi C_G l_a N_a^{\nu}}, \quad (3)$$

where l_a is the size of monomeric segments and N_a is the number of such monomers in a polymer molecule.

At high surface densities, such that neighboring polymer molecules interact with each other, grafted polymers exhibit a brush-like structure (de Gennes, 1980). In this regime, we consider the formation of tubular structures from the membrane and predict the tubule curvatures generated by intermolecular crowding effects on the cell surface. An energy minimization approach elucidates the equilibrium curvature and brush extension as follows. For a tubule with radius R, the energy of the glycocalyx per length of the tubule contains elastic, excluded volume, and electrostatic components (Borisov and Zhulina, 2002; Bracha et al., 2013; Zhulina et al., 2006):

$$F_{brush} = F_{elastic} + F_{excluded\ volume} + F_{electrostatic}, \quad (4)$$

$$F_{brush} = k_B T \int_R^{R+H} \left[\frac{3}{2 l_a^2 c_p s} + \left(w + \frac{\alpha_b^2}{2\Phi_{ion}} \right) c_p^2 s \right] dr, \quad (5)$$

where R is the radius of the tubule, H is the thickness of the glycocalyx brush, l_a is the size of monomeric segments that form the biopolymers, c_p is the monomer concentration, and s is the area per polymer. At the tubule surface, the area per polymer, $s(r = R)$ is related to the biopolymer surface density, C_G, as $s(r = R) = 1/C_G$. w is the excluded volume of monomer segments, α_b is the degree of ionization of a monomer, Φ_{ion} is the ion concentration in bulk solution, and r is a radial coordinate.

Zhulina et al. (Zhulina et al., 2006) provide expressions for c_p. Given the monomer length and diameter are similar (Paturej et al., 2016), we consider the monomeric segments to be cylinders with an aspect ratio close to 1. The energy per length of the underlying membrane bent into the tubular structure is (Helfrich, 2014):

$$F_{membrane} = \frac{\pi \kappa}{R}, \quad (6)$$

where κ is the membrane bending modulus. Thus, the total energy per tubule length is:

$$F_{total} = F_{brush} + F_{membrane}$$
$$= k_B T \int_R^{R+H} \left[\frac{3}{2 l_a^2 c_p s} + \left(w + \frac{\alpha_b^2}{2\Phi_{ion}} \right) c_p^2 s \right] dr + \frac{\pi\kappa}{R}. \quad (7)$$

Minimizing the total energy with respect to the tubule radius ($dF_{total}/dR = 0$) reveals the dependence of the spontaneous curvature on the properties of the glycocalyx and the cell membrane, including the surface density of biopolymers.

We consider the implications of this theory for native Muc1, as an example mucin. We course-grain the bottlebrush biopolymer into N_a effective monomers of size $l_{a,eff}$ (Paturej et al., 2016). In this work, we measure the radius of gyration, R_G, of Muc1 to be 32 nm. We estimate the overall stretched length, l, to be 270 nm based on electron micrographs of Muc1 purified from human HEp-2 epithelial cells (Bramwell et al., 1986). The radius of gyration is related to the Flory radius by $R_G \approx \frac{1}{\sqrt{6}} R_F = \frac{1}{\sqrt{6}} l^{\nu} l_{a,eff}^{1-\nu}$. Using estimates of R_G = 32 nm, l = 270 nm, and ν = 0.6, we estimate the mucin to be described by $N_a = 18$ effective monomeric segments each having a size of $l_{a,eff} = 15$ nm. We note that this effective monomer size is in good agreement with expectations based on estimates of the mucin side chain size to be 5-10 nm (Kesimer et al., 2013; McMaster et al., 1999). We assume that sialic acids on mucins contribute to a charge density of approximately 5 e^- per 20 amino acid tandem repeat. Our assumption is based on most mucin *O*-glycosylation sites being occupied with sialylated glycans (Bäckström et al., 2003; Müller et al., 1999).

The scaling law for the mucin mushroom regime predicts small spontaneous curvatures for low biopolymer densities (Fig. 35C). The predicted spontaneous curvatures are comparable to the curvatures of the bleb-like protrusions observed in cells expressing low surface densities of mucins, as shown in Fig. 36B, 180 mucins/μm^2. For higher densities, where the biopolymers form a brush, the corresponding model above predicts the generation of curvatures similar or greater to those observed in the tubules on the cells of Fig. 36B, 52000 mucins/μm^2. The curvature of such tubules is predicted to increase exponentially with biopolymer density. Notably, the continuous transition between mushroom and brush regimes predicted about a biopolymer density of 250 #/μm^2 accompanies a change in cell surface morphology from bleb-like to tubulated (Fig. 36B, D, E).

Similarly, HA molecules closely resemble linear polymer chains. For instance, a 1 MDa HA molecule has a length of 2.5 μm when stretched out, and can be modeled as a chain

of 250 monomeric units approximately 10 nm long (Cleland Robert L., 2004; Hayashi et al., 1995). Polymer theory predicts such a polymer to have a large Flory radius of about 1 μm, which is more than an order of magnitude larger than that of Muc1. Thus, HA is expected to have a much larger effective volume and physical presence on the cell surface than Muc1. The consequently stronger intramolecular and intermolecular interactions in HA should render it significantly more effective at bending the membrane than Muc1. Furthermore, considerably lower surface density of HA is expected to generate the same membrane curvature as a surface densely crowded with Muc1.

We also conducted numerical calculations for the specific example of HA. Adopting the approach of Bracha et al. on DNA, also a linear polyelectrolyte, we coarse grain hyaluronic acid into N_a cylindrical segments of length l_a and diameter d to allow application of polymer brush theory scaling laws (Bracha et al., 2013). The Kuhn length, l_a, of the biopolymers is twice the persistence length and the length scale at which the molecule is straight. Hyaluronic acid is semi-rigid owing to the local stiffness that arises from intrinsically large size of the sugar ring monomers and the hindered rotations about the glycosidic linkages (Day and Sheehan, 2001). Measurements of the persistence length range from 5 to 9 nm. The diameter of the hyaluronic acid chain is about 0.6 nm (Cowman et al., 2005). In this work, we measure the molecular weight of hyaluronic acid produced by the hyaluronic acid synthase 3 (HAS3) to be approximately 3 MDa. This large size corresponds to a fully stretch length of approximately 10 μm, assuming a disaccharide size of 1 nm.

Force Requirements for Cell Surface Blebs and Tubes

To predict the relative frequencies of blebs and tubes on the cell surface, we perform energetic calculations for the cell membrane. The crowding pressure of the glycopolymers effectively increases the natural curvature of the cell membrane. Hence, we lump together the crowding effects of the glycocalyx into a spontaneous membrane curvature, c_0.

Intracellular forces pushing the cell membrane out, e.g. actin polymerization, can generate cylindrical tubes (Weichsel and Geissler, 2016). Here we consider a tube of length L and radius R_{tube} generated due to a force f. On the other hand, a hydrostatic pressure difference p between inside and outside the cell can form spherical blebs of radius R_{bleb} (Charras and Paluch, 2008). The energy of the membrane in these configurations includes the bending energy, surface tension, and contributions from the pressure p or the force f (Derényi et al., 2002; Helfrich, 2014; Seifert et al., 1991):

$$F = \int_A \frac{\kappa}{2}(c_1 + c_2 - c_0)^2 dA + \sigma A - pV - fL, \quad (8)$$

where κ is the bending stiffness of the membrane, c_1 and c_2 are the principal curvatures, c_0 is the spontaneous curvature of the membrane – generated due to the crowding pressure of the biopolymers, A is the area of the membrane, and σ is the surface tension of the membrane. For tubes, $p = 0,\ f \neq 0$, and L is the length of the tube, whereas for blebs, $f = 0,\ p \neq 0$, and V is the bleb volume.

A cylindrical tube of radius R_{tube} has $c_1 = 0$ and $c_2 = 1/R_{tube}$, which simplify the energy:

$$F_{tube} = \left[\frac{\kappa}{2}\left(\frac{1}{R_{tube}} - c_0\right)^2 + \sigma\right] 2\pi R_{tube} L - fL. \quad (9)$$

The case of a spherical bleb with a very thin neck provides an upper limit on the energy of a bleb. For a bleb with radius R_{bleb}, $c_1 = c_2 = 1/R_{bleb}$, and

$$F_{bleb} = \left[\frac{\kappa}{2}\left(\frac{2}{R_{bleb}} - c_0\right)^2 + \sigma\right] 4\pi R_{bleb}^2 - \frac{4\pi R_{bleb}^3}{3} p. \quad (10)$$

At equilibrium, these energies are minimized with respect to the radii of the blebs and tubes (Derényi et al., 2002). The tube energy is also minimized with respect to the tube length L at steady state (Derényi et al., 2002). That is,

$$\frac{\partial F_{tube}}{\partial R_{tube}} = 0, \quad \frac{\partial F_{tube}}{\partial L} = 0, \quad (11)$$

and

$$\frac{\partial F_{bleb}}{\partial R_{bleb}} = 0 \quad (12)$$

at equilibrium. The equilibrium equations (Eq. 11) for the tube imply:

$$R_{tube} = \frac{1}{\sqrt{c_0^2 + 2\sigma/\kappa}}, \quad (13)$$

and

$$f = 2\pi\kappa\left(\sqrt{c_0^2 + 2\sigma/\kappa} - c_0\right). \quad (14)$$

These equilibrium calculations predict the tube radius is completely governed by the mechanical properties of the lipid bilayer and the spontaneous curvature. These calculations do not account for the structural support of actin filaments widening the tubes.

Bleb energy minimization (Eq. 12) yields the pressure requirement for a bleb of a given size:

$$p = \frac{2\sigma}{R_{bleb}} - \frac{c_0\kappa}{R_{bleb}}\left(\frac{2}{R_{bleb}} - c_0\right). \quad (15)$$

Eq.13-15 relate the force or pressure required to maintain a tube or bleb with the spontaneous curvature generated by the biopolymers. Fig. 35C details the dependence of the spontaneous curvature on biopolymer concentration. We thus graph the force and pressure requirements against the biopolymer concentration (Fig. 35D). Comparisons with typically observed forces from actin polymerization and hydrostatic pressures explain the relative densities of tubes and blebs as a function of biopolymer density.

REFERENCES

Bäckström, M., Link, T., Olson, F.J., Karlsson, H., Graham, R., Picco, G., Burchell, J., Taylor-Papadimitriou, J., Noll, T., and Hansson, G.C. (2003). Recombinant MUC1 mucin with a breast cancer-like O-glycosylation produced in large amounts in Chinese-hamster ovary cells. Biochem. J. *376*, 677–686.

Borisov, O.V., and Zhulina, E.B. (2002). Effect of Salt on Self-Assembly in Charged Block Copolymer Micelles. Macromolecules *35*, 4472–4480.

Bracha, D., Karzbrun, E., Shemer, G., Pincus, P.A., and Bar-Ziv, R.H. (2013). Entropy-driven collective interactions in DNA brushes on a biochip. Proc. Natl. Acad. Sci. *110*, 4534–4538.

Bramwell, M.E., Wiseman, G., and Shotton, D.M. (1986). Electron-microscopic studies of the CA antigen, epitectin. J. Cell Sci. *86*, 249–261.

Charras, G., and Paluch, E. (2008). Blebs lead the way: how to migrate without lamellipodia. Nat. Rev. Mol. Cell Biol. *9*, 730–736.

Cleland Robert L. (2004). Viscometry and sedimentation equilibrium of partially hydrolyzed hyaluronate: Comparison with theoretical models of wormlike chains. Biopolymers *23*, 647–666.

Cowman, M.K., Spagnoli, C., Kudasheva, D., Li, M., Dyal, A., Kanai, S., and Balazs, E.A. (2005). Extended, relaxed, and condensed conformations of hyaluronan observed by atomic force microscopy. Biophys. J. *88*, 590–602.

Day, A.J., and Sheehan, J.K. (2001). Hyaluronan: polysaccharide chaos to protein organisation. Curr. Opin. Struct. Biol. *11*, 617–622.

Derényi, I., Jülicher, F., and Prost, J. (2002). Formation and interaction of membrane tubes. Phys. Rev. Lett. *88*, 238101.

de Gennes, P. (1980). Conformations of polymers attached to an interace. Macromolecules 1069–1075.

Hayashi, K., Tsutsumi, K., Nakajima, F., Norisuye, T., and Teramoto, A. (1995). Chain-stiffness and excluded-volume effects in solutions of sodium hyaluronate at high ionic strength. Macromolecules *28*, 3824–3830.

Helfrich, W. (2014). Elastic Properties of Lipid Bilayers: Theory and Possible Experiments. Z. Für Naturforschung C *28*, 693–703.

Israels, R., Leermakers, F.A.M., Fleer, G.J., and Zhulina, E.B. (1994). Charged Polymeric Brushes: Structure and Scaling Relations. Macromolecules *27*, 3249–3261.

Kesimer, M., Ehre, C., Burns, K.A., Davis, C.W., Sheehan, J.K., and Pickles, R.J. (2013). Molecular organization of the mucins and glycocalyx underlying mucus transport over mucosal surfaces of the airways. Mucosal Immunol. *6*, 379–392.

Lipowsky, R. (1995). Bending of Membranes by Anchored Polymers. EPL Europhys. Lett. *30*, 197.

McMaster, T.J., Berry, M., Corfield, A.P., and Miles, M.J. (1999). Atomic force microscopy of the submolecular architecture of hydrated ocular mucins. Biophys. J. *77*, 533–541.

Müller, S., Alving, K., Peter-Katalinic, J., Zachara, N., Gooley, A.A., and Hanisch, F.G. (1999). High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells. J. Biol. Chem. *274*, 18165–18172.

Paturej, J., Sheiko, S.S., Panyukov, S., and Rubinstein, M. (2016). Molecular structure of bottlebrush polymers in melts. Sci. Adv. *2*, e1601478.

Seifert, U., Berndl, K., and Lipowsky, R. (1991). Shape transformations of vesicles: Phase diagram for spontaneous- curvature and bilayer-coupling models. Phys. Rev. A *44*, 1182–1202.

Stachowiak, J.C., Schmid, E.M., Ryan, C.J., Ann, H.S., Sasaki, D.Y., Sherman, M.B., Geissler, P.L., Fletcher, D.A., and Hayden, C.C. (2012). Membrane bending by protein-protein crowding. Nat. Cell Biol. *14*, 944–949.

Weichsel, J., and Geissler, P.L. (2016). The More the Tubular: Dynamic Bundling of Actin Filaments for Membrane Tube Formation. PLOS Comput. Biol. *12*, e1004982.

Zhulina, E.B., Birshtein, T.M., and Borisov, O.V. (2006). Curved polymer and polyelectrolyte brushes beyond the Daoud-Cotton model. Eur. Phys. J. E *20*, 243–256.

What is claimed is:

1. A recombinant lubricin polypeptide comprising a repeated sequence, the repeated sequence comprising or consisting of:
KEPAPTTP (SEQ ID NO:1)
KEPAPTP (SEQ ID NO:9)
KEPAPTTTP (SEQ ID NO:10), or a combination thereof.

2. The recombinant lubricin polypeptide of claim 1, wherein the repeated sequence is repeated contiguously 10-120 times.

3. The recombinant lubricin polypeptide of claim 2, wherein the repeated sequence is repeated contiguously 59 times.

4. The recombinant lubricin polypeptide of claim 1, wherein the contiguously repeated sequence is located between an N-terminal amino acid sequence and a C-terminal amino segment that has at least 90% sequence identity, respectively, to one of:
a human lubricin N-terminal and C-terminal derived amino acid sequence;
a canine lubricin N-terminal and C-terminal derived amino acid sequence; or
an equine lubricin N-terminal and C-terminal derived amino acid sequence.

5. The recombinant lubricin polypeptide of claim 4, wherein the contiguously repeated sequence is located between an N-terminal human lubricin derived sequence that has at least 90% sequence identity to the human lubricin sequence:

QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDFKRVCTAELSCKGRCFESFER
GRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAPPPSGASQTIKSTTKRSPK
PPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSTIRKIKSSKNSAANRELQKKLK
VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAK
PINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEK
TSAKDLAPTSKVLAKPTPKAETTTKGPALTTP (SEQ ID NO:75)

and a C-terminal human lubricin amino derived amino acid sequence that has at least 90% sequence identity to the human lubricin sequence:

SEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTT
AKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKI
TTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKP
TSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVE
VNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNIC
NGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFTRCNC
EGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYF
FKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGETTQVRRRRFERAIGPSQTHTIRIQ
YSPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYY
NIDVPSRTARAITTRSGQTLSKVWYNCP (SEQ ID NO:76).

6. The recombinant lubricin of claim 5, wherein the repeated sequence is KEPAPTTP (SEQ ID NO:1).

7. The recombinant lubricin polypeptide of claim 4, wherein the contiguously repeated sequence is located between an N-terminal canine derived lubricin sequence that has at least 90% sequence identity to the canine lubricin sequence:

QDLPSCAGRCGEGYSRDAICNCDYNCQHYMECCPDFKKACTVELSCKGRCFESFAR
GRECDCDSDCKKYGKCCPDYEDFCGRVHNPTSPPSSKTAPPSPGASQTIKSTAKRSPK
APNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSTIRKIKSSKNSAANKELKKKPKV
KDNKKERTPKKKPPPEPPVVDEAGSGLDNGDIKLTPTPDIPTTQRNKVTTSPKFTTGK
PINPKPSLPPNTDTSKETSSTPNKETTVKSKETLANKETSSKAKEKITSAKETRSAEKTP
AKDFVPTTKAPVKSTPKAESTTKGPALTTP (SEQ ID NO:77)

and a C-terminal canine lubricin derived amino acid sequence that has at least 90% sequence identity to the canine lubricin sequence:

SEVTTTAKDKTTEKDIIPEITTAVPKITTQETATPTEETTTESKTSTTTQVTSTTSSKNTP
KATTLAPKVMTATQKTTTTEETMNKPEETTAVPKDTATSTKVSTPRPRKPTKAPKKP
ASTKKPNTIPKRKKPKTTPTPPKMTTSTMPKLHPTSSVEAMLQTTTSPNQRPNSEIVE
VNPNEDTDAAGKKPHMFPRPPVLTPIFIPGTDILVRGSNQDIAINPMLSDETNLCNGKP
VDGLTTLRNGTMVAFRGHYFWMLSPSKPPSPPRKITEVWGIPSPIDTVFTRCNCEGKT

FFFKGSQYWRFTNDIKDAGYPKQIVKGFGGLNGRIVAALSIAKYKDRPESVYFFKRG
GSVQQYTYKQEPIKKCTGRRPAINYPVYGETTQVRRRRFERAIGPSQTHTIRIHYSPIR
VSYQDKGFLHNEVKMSSQWRGFPNVVTSAIALPNIRKPDGYDYYAFSRNQYYNIDV
PSRTARVVTTRFGRTLSNIWYNC (SEQ ID NO:78).

8. The recombinant lubricin of claim 7, wherein the repeated sequence is KEPAPTTP (SEQ ID NO:1).

9. The recombinant lubricin polypeptide of claim 4, wherein the contiguously repeated sequence is located between an N-terminal equine derived lubricin sequence that has at least 90% sequence identity to the equine lubricin sequence:

QDLSSCAGRCGEGYSRDATCNCDFNCQYYMECCPDFKKVCTSELSCKGRCFESFER
GRECDCDADCKKYGKCCSDYESFCEEVHNPTSPPSSKTAPPPPGASQTIKSTAKRSPK
SNKKKTKKVIESEEIIEEHSVSENQESSSSSSSSSTIRKVKSSKNSAANRELKKKPKVK
DSKKKRTPKKKPTPEPPVIDEAGSGLDNGDFMLIPTPKIPTTQRNKVTTSPKITTVKPI
NPKPSLPPNSDTSKETTSTPNKETTVETKETEITNKETSTSANEKTTSARKSTEKTSDK
DFAPASEVPAKSTPKAETTTKGPALTTP (SEQ ID NO:79)

and a C-terminal equine lubricin derived amino acid sequence that has at least 90% sequence identity to the equine lubricin sequence:

SEVSTTTTTMKPPTTPKNLAESTPEFPAEPTPKALENSPKEPAVPTTKAPEVTKPEVTT
TAKDKVTGKDIHTIPEITTAAPKITTETATTTEEKTTESKVTSTIMQVTSTTEDTTTSSK
ITPKATTLAPKVMTATKTTTTQETINKLEETTAIPKDTATHSKVTTPKPKKPTKAPRKP
TSTKKPKTPRKRKPKTTPIPPKITTPTTPKSNPTTLAEAMLQTTTSPNQTPNSAMIEVNP
KNEDADAAEGEKPLVILRPHVLTPIVIPGPDFLVRGPNLGIGINPMLSDETNLCNGKPV
DGLTTLRNGTLVAFRGHYFWMLRPFSPPSPPRRITEVWGIPSPIDTVFTRCNCEGKTFF
FKDSQYWRFTNDIKDAGYPKLISKGFGGLSGKIVAALSIATYKNRPESVYFFKRGGRI
QQYIYKQEPIRKCPGRRPAIHYSVYGEAPQIRRRRFERAIGPSQTHTIRIHYSPVRVSYQ
DKVPSTDFLHNEVKVSTLWRGLPDTVTSAISLPNLRKPDGYDYYAFSKDQYYNIDVP
SRTARAITTRSGQTLSKVWYNCP (SEQ ID NO:80).

10. The recombinant lubricin of claim 9, wherein the repeated sequence is KEPAPTTP (SEQ ID NO:1).

11. The recombinant lubricin of claim 4, further comprising a secretory signal that is optionally MAWKTLPIYLLLLLSVFVIQQVSS (SEQ ID NO:72).

12. The recombinant lubricin of claim 5, further comprising a secretory signal that is METDTLLLWVLLLWVPGSTGD (SEQ ID NO:81).

13. The recombinant lubricin of claim 6, further comprising a secretory signal that is METDTLLLWVLLLWVPGSTGD (SEQ ID NO:81).

14. The recombinant lubricin of claim 7, further comprising a secretory signal that is MQWKILPIYLLLLSVFLIQQVSS (SEQ ID NO:82).

15. The recombinant lubricin of claim 8, further comprising a secretory signal that is MQWKILPIYLLLLSVFLIQQVSS (SEQ ID NO:82).

16. The recombinant lubricin of claim 9, further comprising a secretory signal that is MEWKILPIYLLLLLSIFSIQEVSS (SEQ ID NO:74).

17. The recombinant lubricin of claim 10, further comprising a secretory sequence that is MEWKILPIYLLLLLSIFSIQEVSS (SEQ ID NO:74).

18. The recombinant lubricin of any one of claims 1-17, wherein the recombinant lubricin has an intra-articular half-life when injected into a mammal of more than 4 days.

19. The recombinant lubricin of claim 17, wherein the recombinant lubricin has an intra-articular half-life when injected into a mammal of at least 15 days.

20. The recombinant lubricin of claim 19, wherein the recombinant lubricin has an intra-articular half-life when injected into a mammal of at least 30 days.

21. A composition comprising a recombinant lubricin of any one of claims 1-17.

22. The composition of claim 22, wherein the composition is selected from a pharmaceutical formulation, an eye drop, and a contact lens solution.

23. The composition of claim 22, for use in treating a human in need thereof.

24. The composition of claim 22, for use in treating a non-human mammal in need thereof.

25. The composition of claim 24, wherein the composition is for use in treating a canine or equine animal in need thereof.

26. One or more mammalian cells comprising a recombinant lubricin of any one of claims 1-17.

27. The one or more mammalian cells of claim 26, wherein the cells are in a suspension culture.

28. An isolated polynucleotide and/or an expression vector encoding a recombinant lubricin of any one of claims 1-17.

29. One or more mammalian cells comprising the isolated polynucleotide of claim 28, wherein the one or more mammalian cells are optionally adapted to growth in a suspension culture.

30. A suspension culture comprising mammalian cells that express a recombinant lubricin of any one of claims 1-17.

31. A method of making a recombinant lubricin of any one of claims 1-17, the method comprising introducing into mammalian cells a polynucleotide encoding said recombinant lubricin such that the cells express said recombinant lubricin.

32. The method of claim 31, further comprising isolated the recombinant lubricin from the mammalian cells.

33. An inanimate article fully or partially coated with a composition comprising a recombinant lubricin of any one of claims 1-17.

34. The inanimate article of claim 33, wherein the inanimate article comprises a contact lens.

35. A method for treating an individual in need thereof, the method comprising introducing into the individual a composition comprising a recombinant lubricin of any one of claims 1-17.

36. The method of claim 35, wherein the recombinant lubricin persists in the individual for a period of greater than four days.

37. The method of claim 36, wherein the recombinant lubricin persists in the individual for a period of greater than 30 days.

38. The method of claim 37, wherein the recombinant lubricin persists in the individual for a period of at least 30 days.

39. The method of claim 35, wherein the individual is in need of treatment for a disorder of a synovial joint, a tendon sheath, or bursa.

40. The method of claim 39, comprising introducing the composition into the synovial joint.

41. The method of claim 35, wherein the individual is in need of treatment for a disorder of the eye.

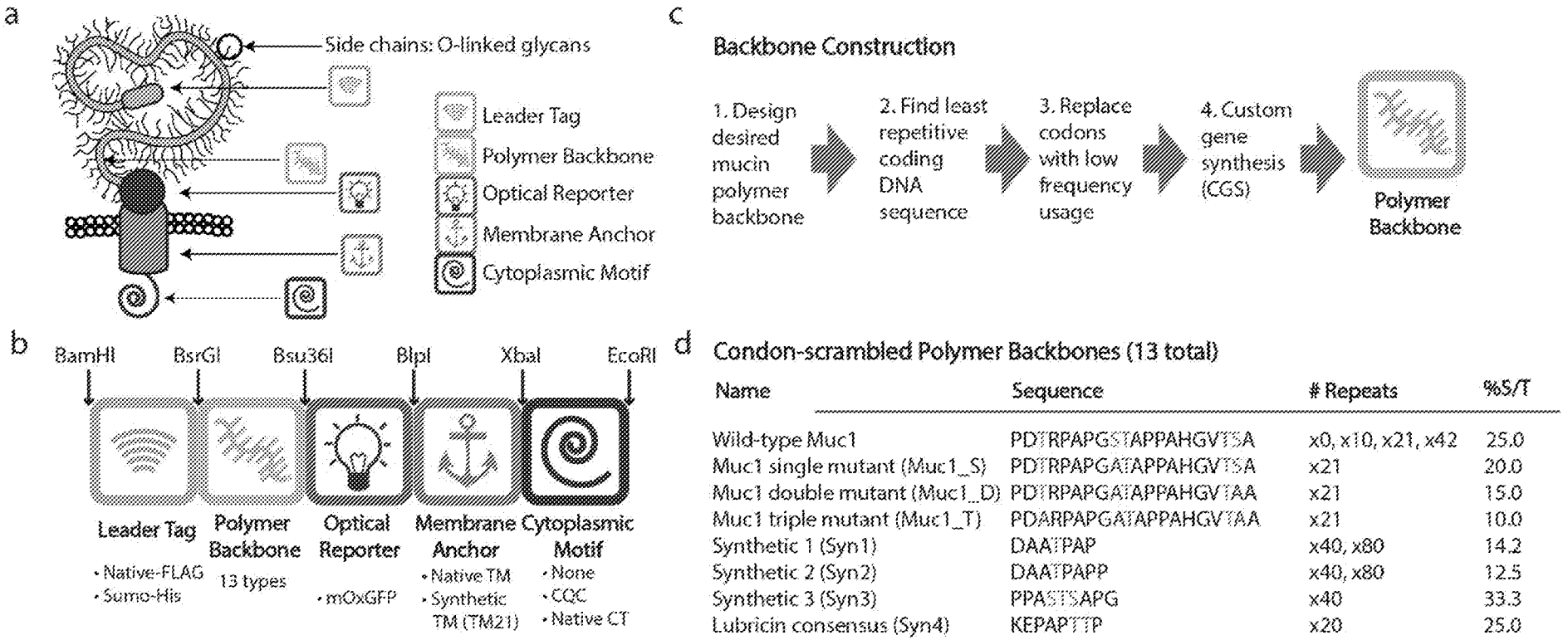
42. The method of claim 41, comprising introducing the composition into the eye of the individual.

43. The method of claim 35, wherein the individual is in need of treatment for a disorder of a mucosal surface.

44. The method of claim 43, wherein the composition is contacted with the mucosal surface.

5

Part I Figures



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

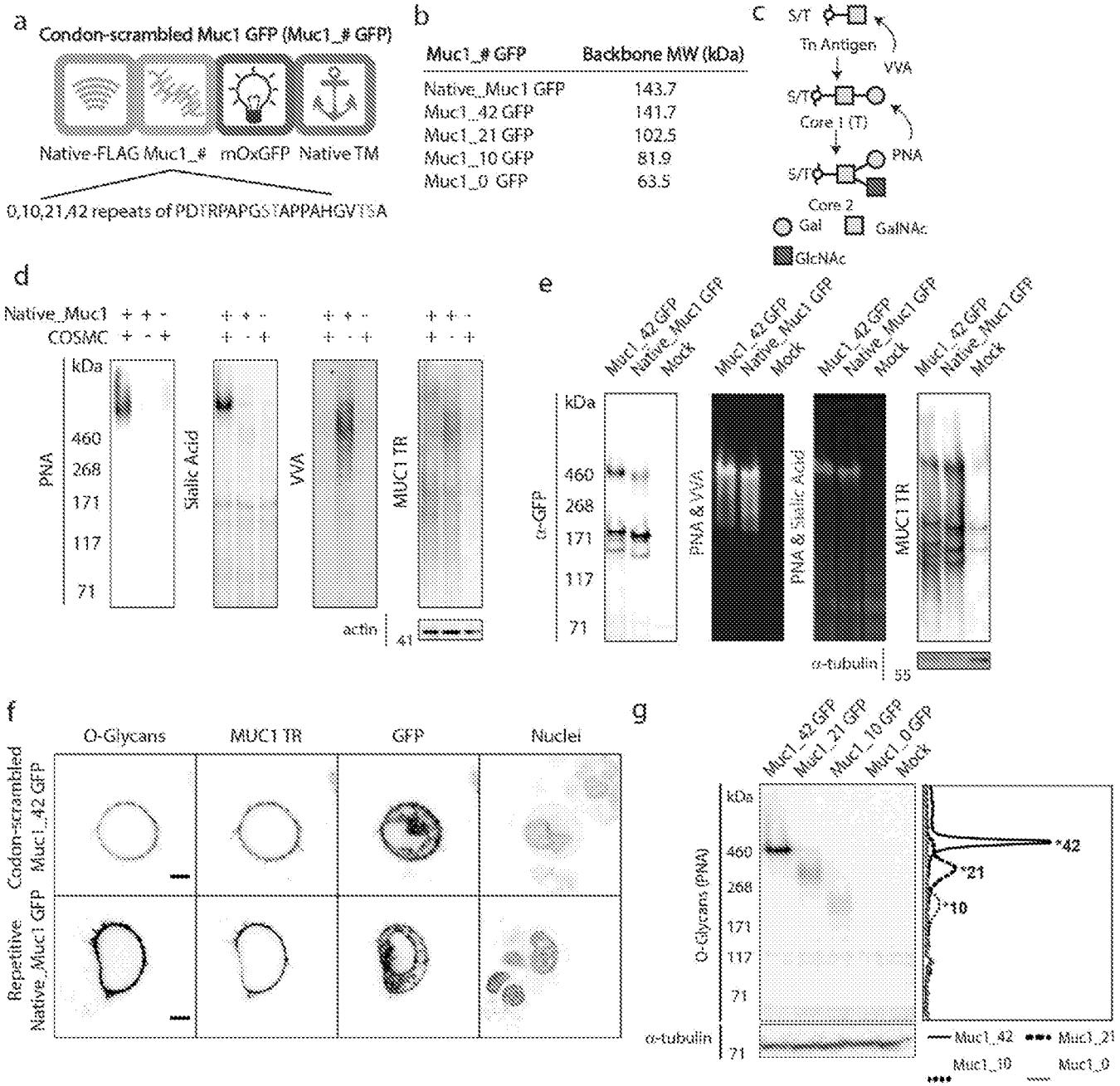

a
Condon-scrambled Muc1 GFP (Muc1_# GFP)
Native-FLAG Muc1_# mOxGFP Native TM
0,10,21,42 repeats of PDTRPAPGSTAPPAHGVTSA
b
Muc1_# GFP Backbone MW (kDa)
Native_Muc1 GFP 143.7
Muc1_42 GFP 141.7
Muc1_21 GFP 102.5
Muc1_10 GFP 81.9
Muc1_0 GFP 63.5
c
S/T
Tn Antigen
VVA
Core 1 (T)
PNA
Core 2
Gal
GalNAc
GlcNAc
d
Native_Muc1
COSMC
kDa
460
268
171
117
71
PNA
Sialic Acid
VVA
MUC1 TR
actin
41
e
Muc1_42 GFP
Native_Muc1 GFP
Mock
kDa
460
268
171
117
71
α-GFP
PNA & VVA
PNA & Sialic Acid
MUC1 TR
α-tubulin
55
f
O-Glycans
MUC1 TR
GFP
Nuclei
Codon-scrambled Muc1_42 GFP
Repetitive Native_Muc1 GFP
g
Muc1_42 GFP
Muc1_21 GFP
Muc1_10 GFP
Muc1_0 GFP
Mock
kDa
460
268
171
117
71
O-Glycans (PNA)
α-tubulin
71
*42
*21
*10
Muc1_42
Muc1_21
Muc1_10
Muc1_0

Figure 2

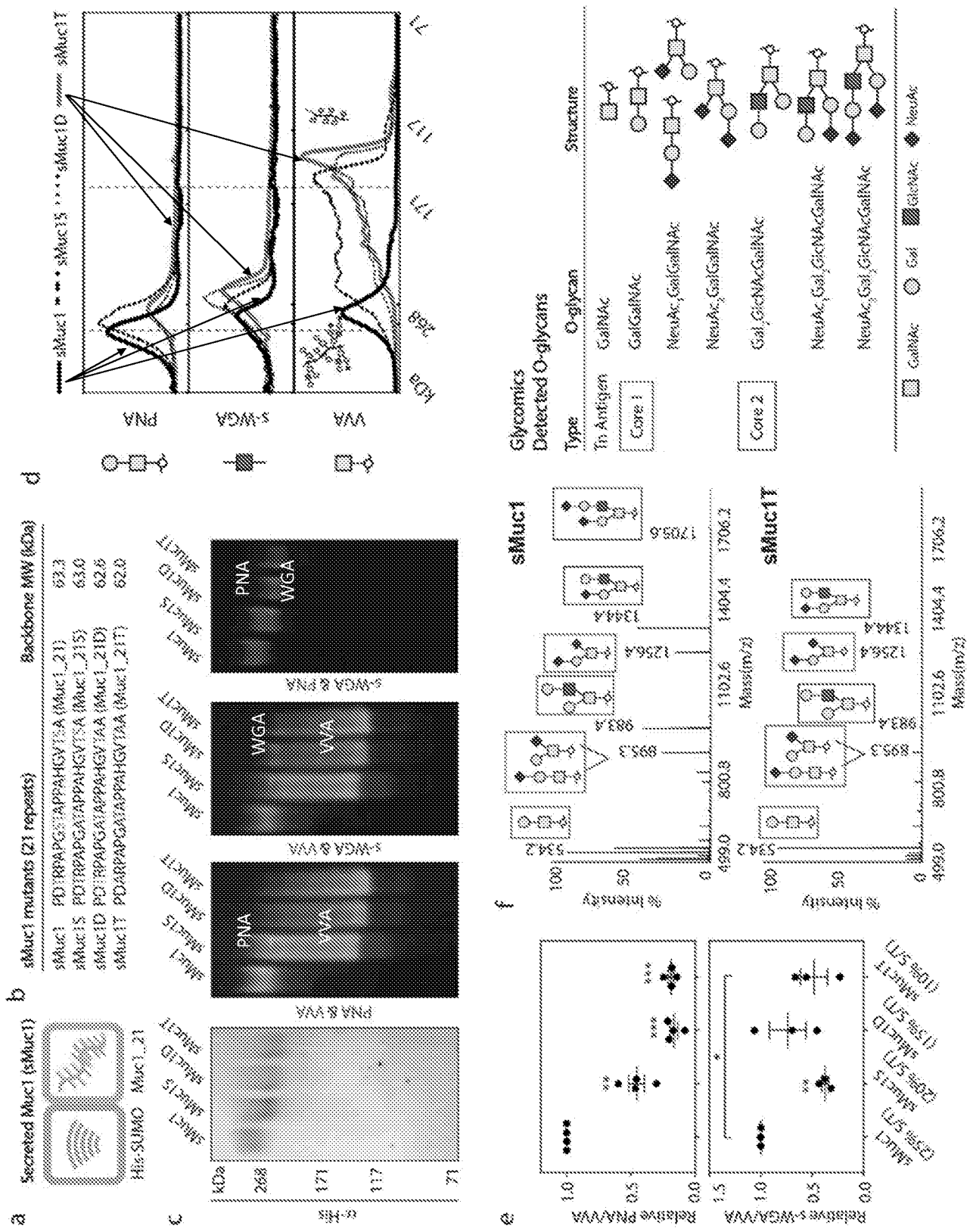

a
Secreted Muc1 (sMuc1)
His-SUMO
Muc1_21
b
sMuc1 mutants (21 repeats)
Backbone MW (kDa)
sMuc1 PDTRPAPGSTAPPAHGVTSA (Muc1_21) 62.3
sMuc1S PDTRPAPGSTAPPAHGVTSA (Muc1_21S) 63.0
sMuc1D PDTRPAPGATAPPAHGVTAA (Muc1_21D) 62.6
sMuc1T PDARPAPGATAPPAHGVTAA (Muc1_21T) 62.0
c
α-His
kDa
268
171
117
71
sMuc1
sMuc1S
sMuc1D
sMuc1T
PNA & VVA
s-WGA & VVA
s-WGA & PNA
PNA
VVA
WGA
d
sMuc1
sMuc1S
sMuc1D
sMuc1T
PNA
s-WGA
VVA
e
Relative PNA/VVA
Relative s-WGA/VVA
f
Glycomics
Detected O-glycans
Type
O-glycan
Structure
Tn Antigen
GalNAc
Core 1
GalGalNAc
NeuAc1Gal1GalNAc
NeuAc2Gal1GalNAc
Core 2
Gal1GlcNAc1GalNAc
NeuAc1Gal1GlcNAc1GalNAc
NeuAc2Gal1GlcNAc1GalNAc
GalNAc
Gal
GlcNAc
NeuAc
sMuc1
sMuc1T
% Intensity
Mass(m/z)
534.2
895.3
983.4
1256.4
1344.4
1705.6
1404.4
1706.2
1102.6
800.8
499.0

Figure 3

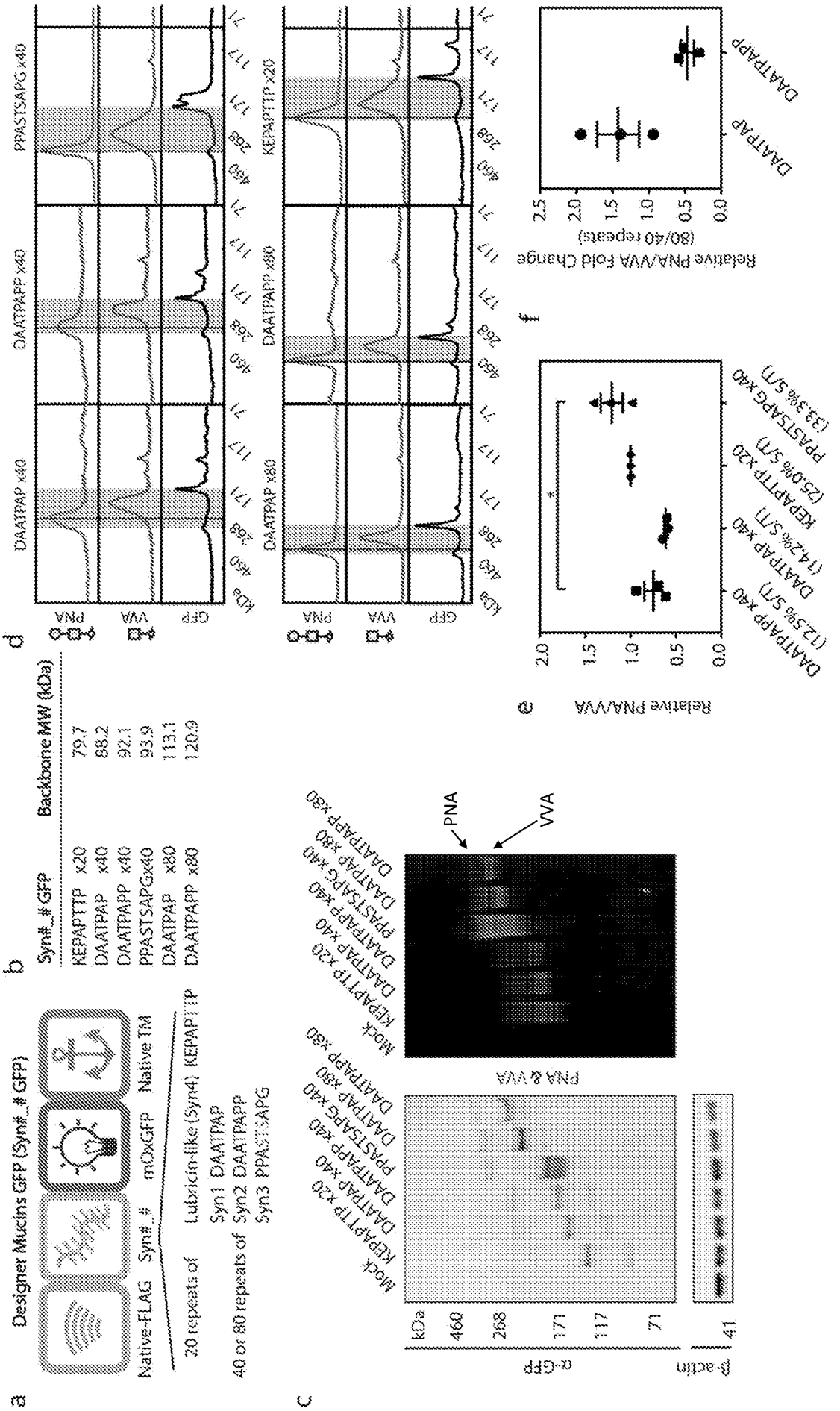

a
Designer Mucins GFP (Syn#_# GFP)
Native-FLAG
Syn#_#
mOxGFP
Native TM
20 repeats of
Lubricin-like (Syn4)
KEPAPTTP
40 or 80 repeats of
Syn1 DAATPAP
Syn2 DAATPAPP
Syn3 PPASTSAPG
b
Syn#_# GFP
Backbone MW (kDa)
KEPAPTTP x20 79.7
DAATPAP x40 88.2
DAATPAPP x40 92.1
PPASTSAPGx40 93.9
DAATPAP x80 113.1
DAATPAPP x80 120.9
c
Mock
KEPAPTTP x20
DAATPAP x40
DAATPAPP x40
PPASTSAPG x40
DAATPAP x80
DAATPAPP x80
PNA
VVA
PNA & VVA
kDa
460
268
171
117
71
41
α-GFP
β-actin
d
DAATPAP x40
DAATPAPP x40
PPASTSAPG x40
DAATPAP x80
DAATPAPP x80
KEPAPTTP x20
PNA
VVA
GFP
e
Relative PNA/VVA
DAATPAP x40 (12.5% S/T)
DAATPAPP x40 (14.2% S/T)
KEPAPTTP x20 (25.0% S/T)
PPASTSAPG x40 (33.3% S/T)
f
Relative PNA/VVA Fold Change (80/40 repeats)
DAATPAP
DAATPAPP

Figure 4

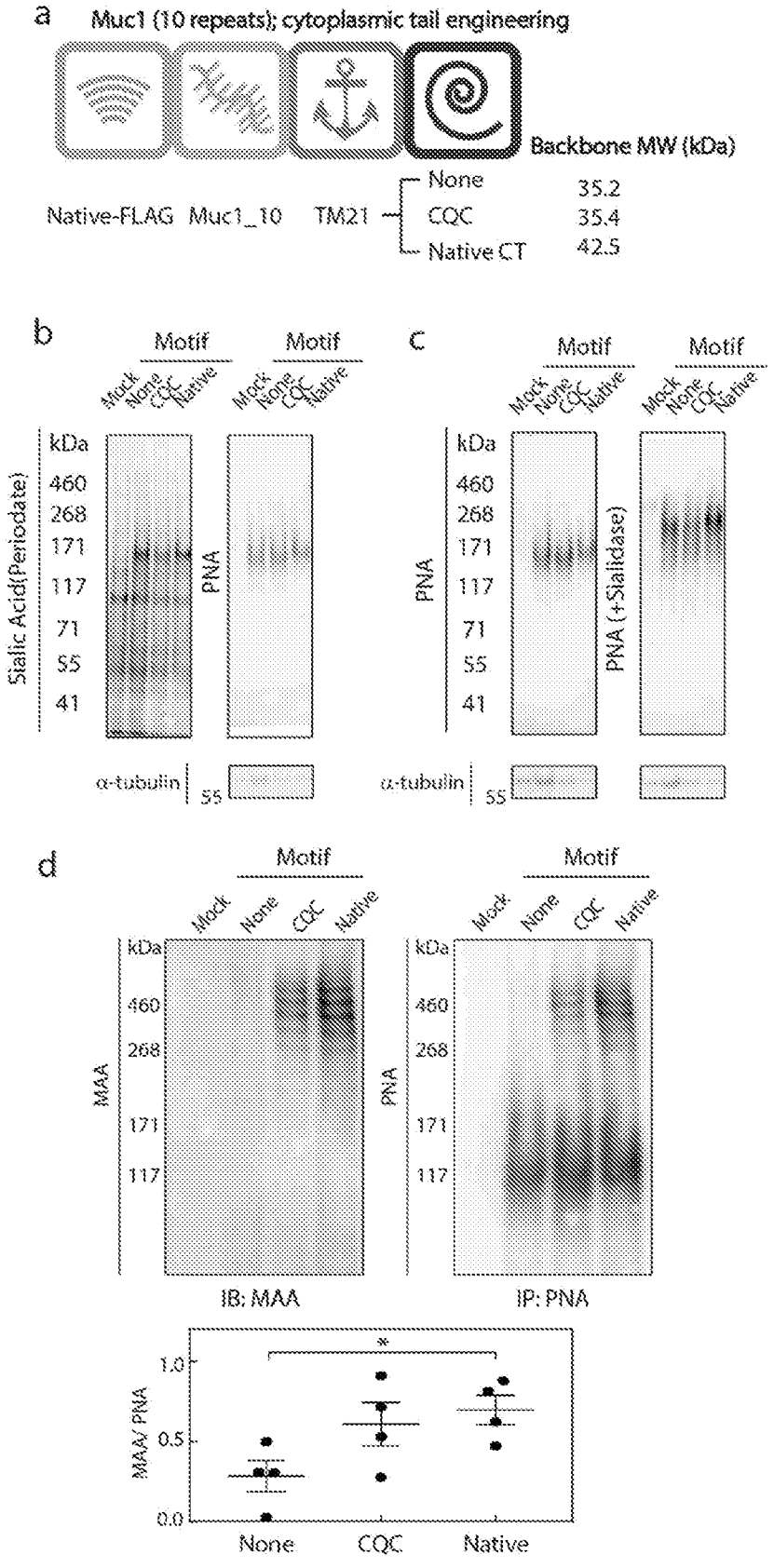

a
Muc1 (10 repeats); cytoplasmic tail engineering
Backbone MW (kDa)
Native-FLAG Muc1_10 TM21
None 35.2
CQC 35.4
Native CT 42.5
b
Motif
Motif
Mock None CQC Native
kDa
460
268
171
117
71
55
41
Sialic Acid(Periodate)
PNA
α-tubulin
55
c
PNA
PNA (+Sialidase)
d
MAA
IB: MAA
IP: PNA
MAA/ PNA
1.0
0.5
0.0
*
None
CQC
Native

Figure 5

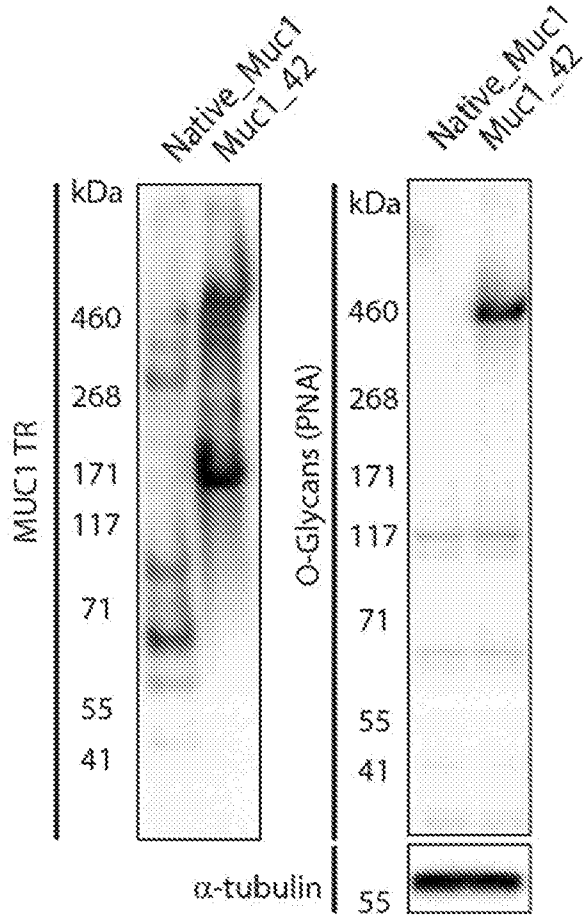

Native_Muc1
Muc1_42
kDa
460
268
171
117
71
55
41
MUC1 TR
Native_Muc1
Muc1_42
kDa
460
268
171
117
71
55
41
O-Glycans (PNA)
α-tubulin
55

Figure 6

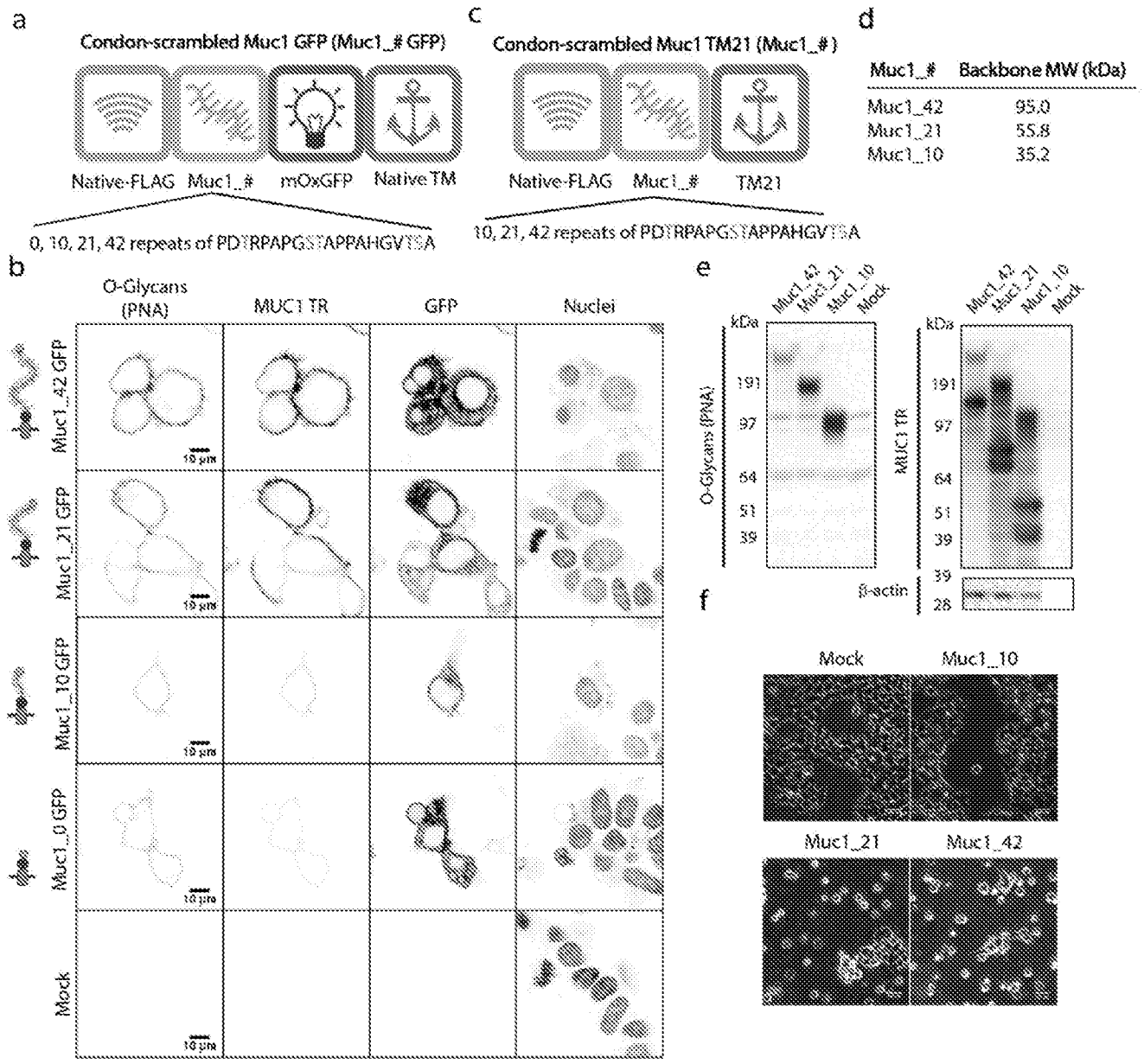


Figure 7

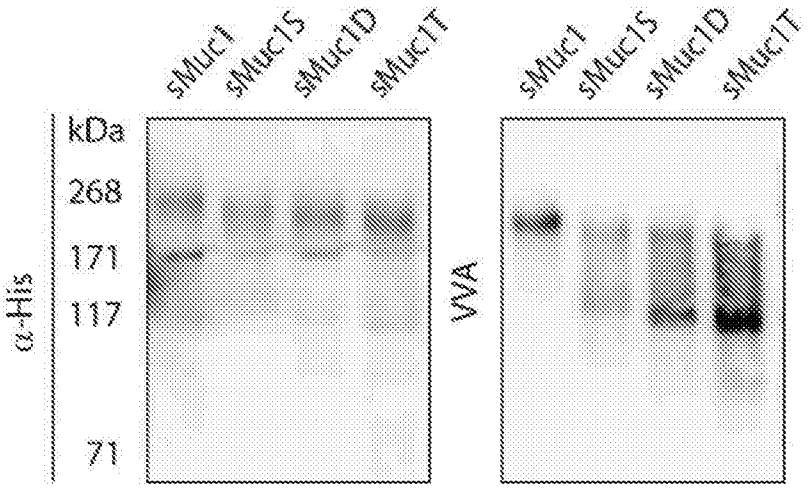


Figure 8

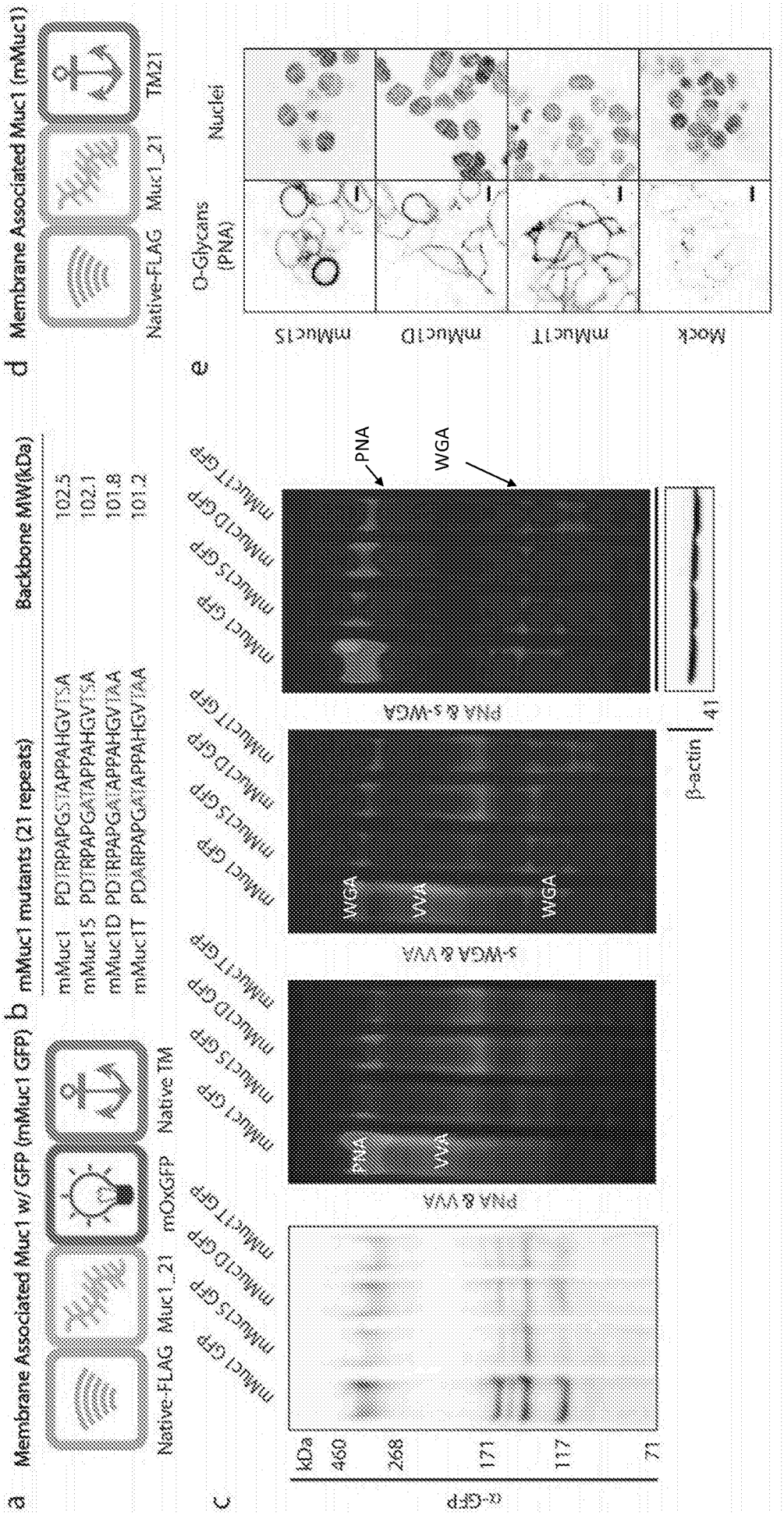

a
Membrane Associated Muc1 w/ GFP (mMuc1 GFP)
Native-FLAG
Muc1_21
mOxGFP
Native TM
b
mMuc1 mutants (21 repeats)
Backbone MW(kDa)
mMuc1 PDTRPAPGSTAPPAHGVTSA 102.5
mMuc15 PDTRPAPGATAPPAHGVTSA 102.1
mMuc1D PDTRPAPGATAPPAHGVTAA 101.8
mMuc1T PDARPAPGATAPPAHGVTAA 101.2
c
mMuc1 GFP
mMuc15 GFP
mMuc1D GFP
mMuc1T GFP
PNA & VVA
s-WGA & VVA
PNA & s-WGA
β-actin
41
kDa
460
268
171
117
71
α-GFP
d
Membrane Associated Muc1 (mMuc1)
Native-FLAG
Muc1_21
TM21
e
O-Glycans (PNA)
Nuclei
mMuc15
mMuc1D
mMuc1T
Mock

Figure 9

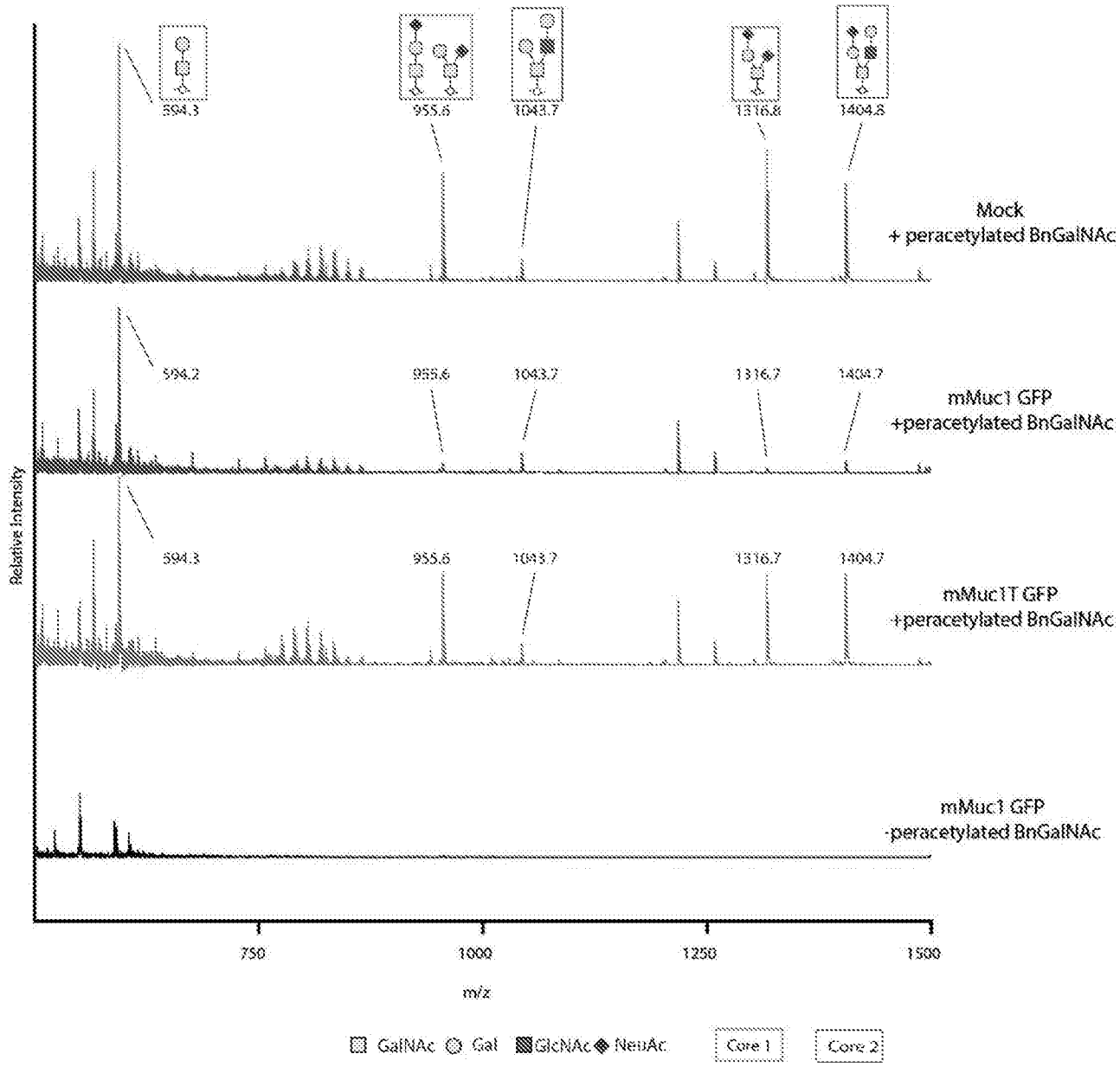

594.3
955.6
1043.7
1316.8
1404.8
Mock
+ peracetylated BnGalNAc
594.2
955.6
1043.7
1316.7
1404.7
mMuc1 GFP
+peracetylated BnGalNAc
594.3
955.6
1043.7
1316.7
1404.7
mMuc1T GFP
+peracetylated BnGalNAc
mMuc1 GFP
-peracetylated BnGalNAc
Relative Intensity
750
1000
1250
1500
m/z
GalNAc
Gal
GlcNAc
NeuAc
Core 1
Core 2

Figure 10

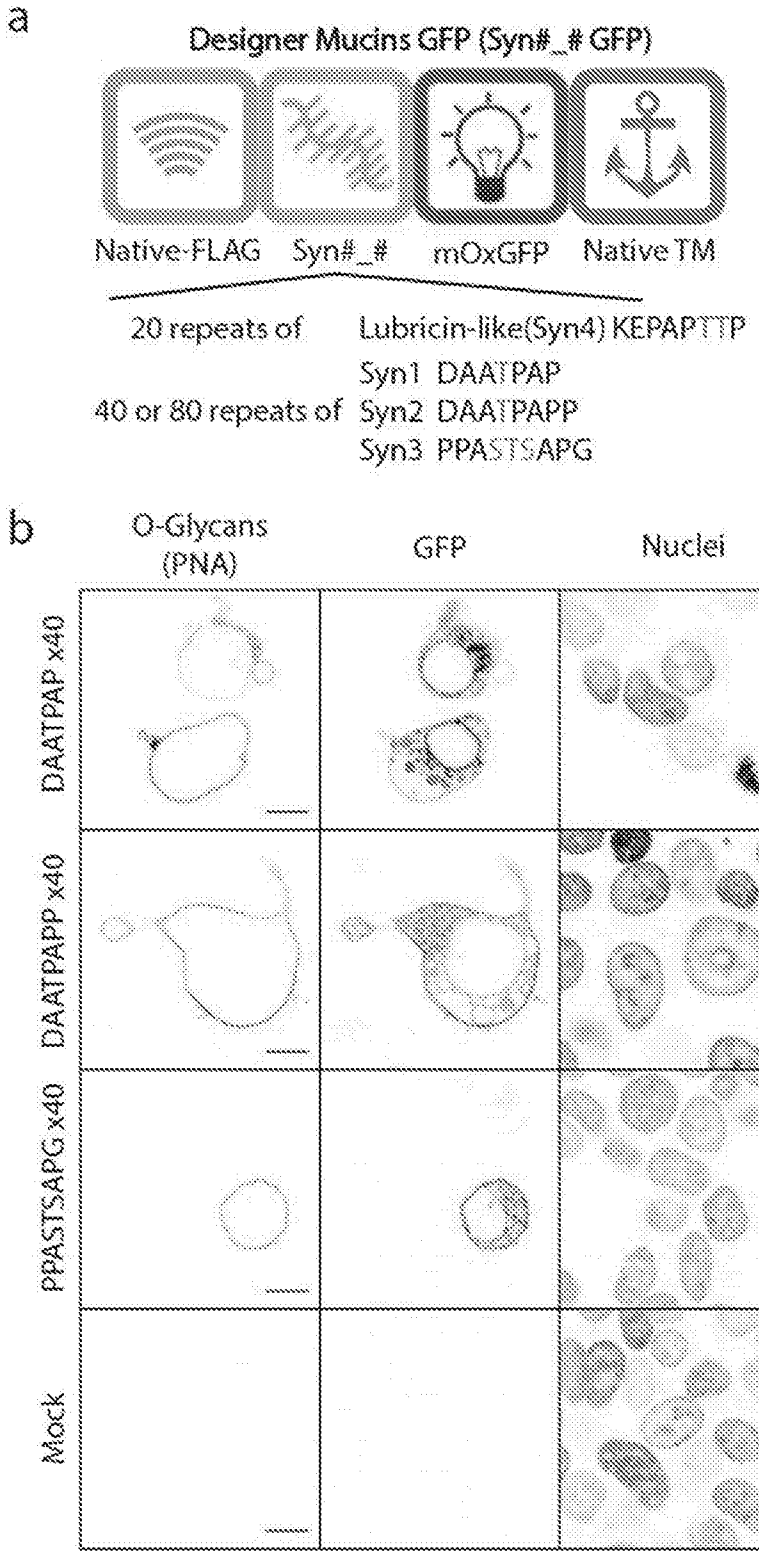

a
Designer Mucins GFP (Syn#_# GFP)
Native-FLAG
Syn#_#
mOxGFP
Native TM
20 repeats of
Lubricin-like(Syn4) KEPAPTTP
40 or 80 repeats of
Syn1 DAATPAP
Syn2 DAATPAPP
Syn3 PPASTSAPG
b
O-Glycans (PNA)
GFP
Nuclei
DAATPAP x40
DAATPAPP x40
PPASTSAPG x40
Mock

Figure 11

Part II Figures

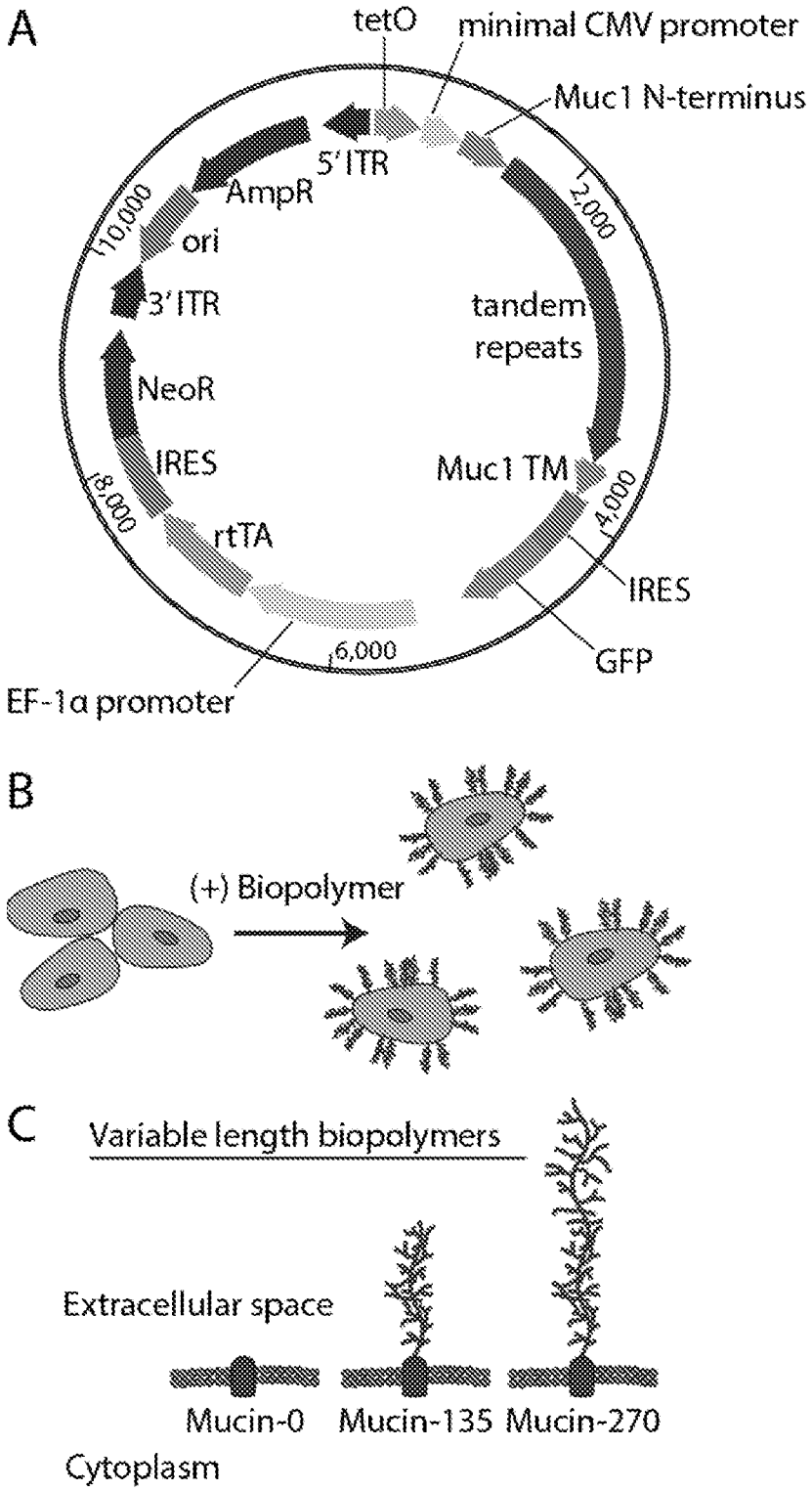


Figure 12

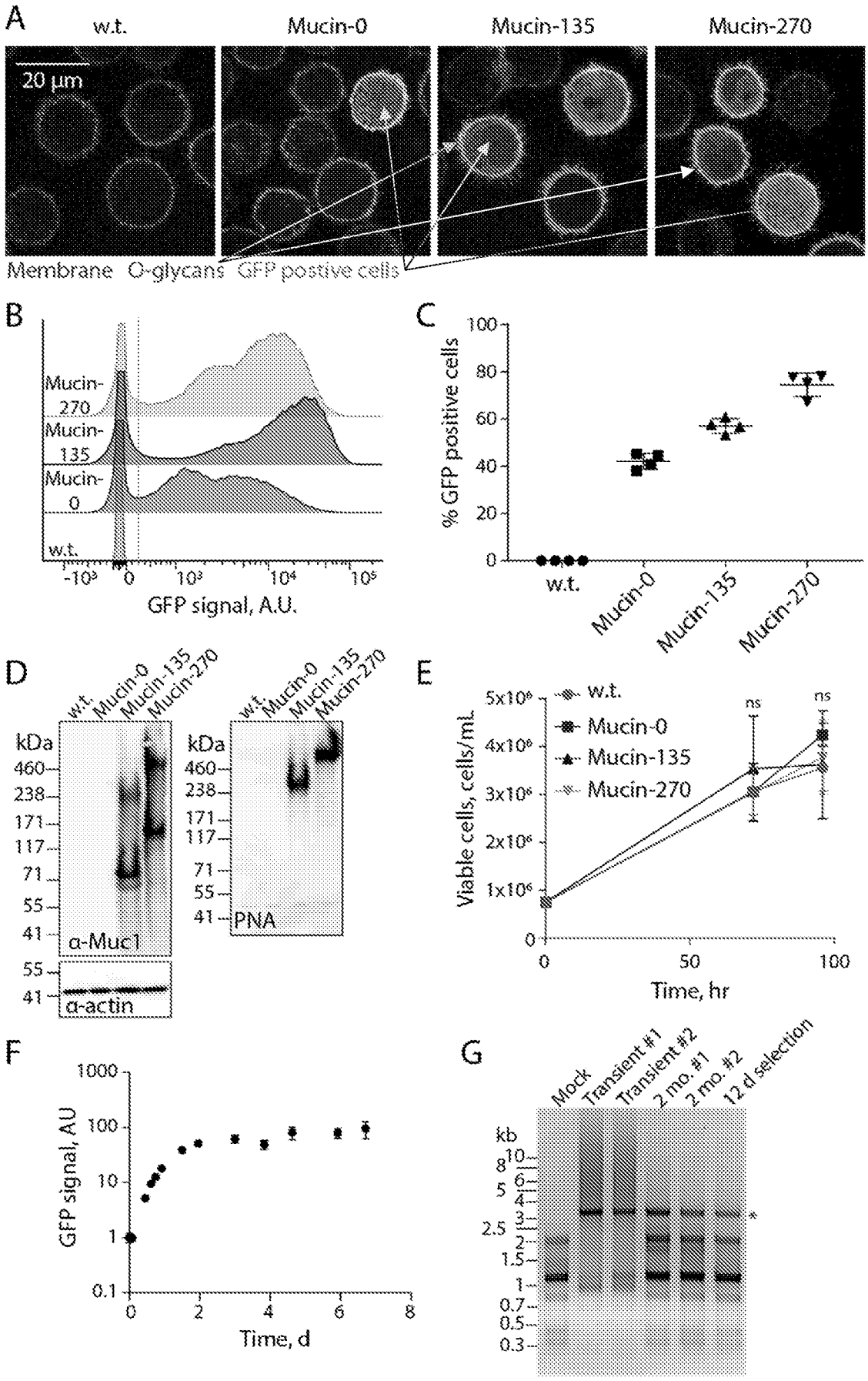

A
w.t.
Mucin-0
Mucin-135
Mucin-270
20 µm
Membrane O-glycans GFP postive cells
B
Mucin-270
Mucin-135
Mucin-0
w.t.
GFP signal, A.U.
C
% GFP positive cells
w.t.
Mucin-0
Mucin-135
Mucin-270
D
kDa
460
238
171
117
71
55
41
α-Muc1
α-actin
PNA
E
Viable cells, cells/mL
Time, hr
ns
F
GFP signal, AU
Time, d
G
Mock
Transient #1
Transient #2
2 mo. #1
2 mo. #2
12 d selection
kb

Figure 13

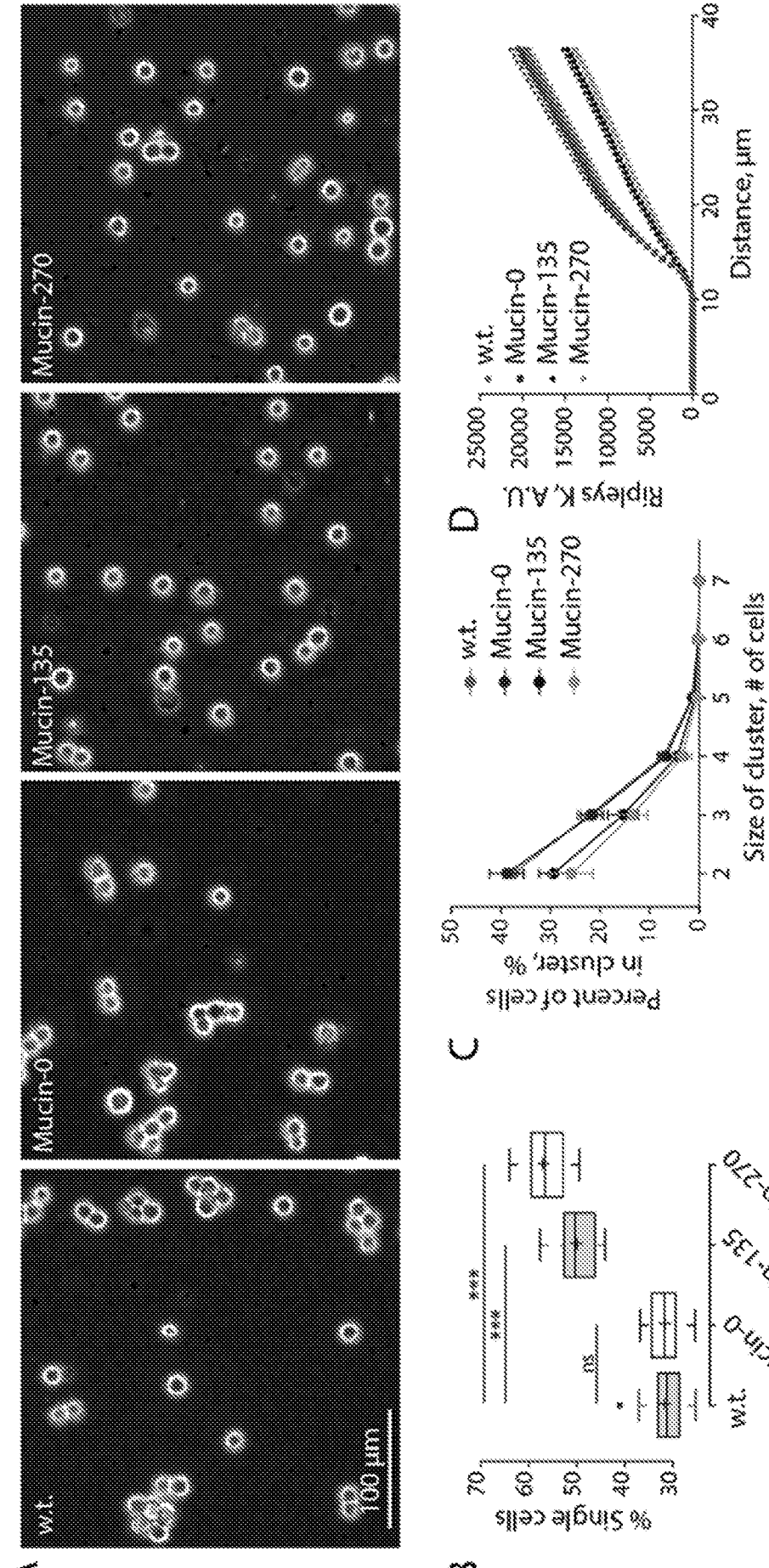

A
w.t.
Mucin-0
Mucin-135
Mucin-270
100 µm
B
% Single cells
70
60
50
40
30
***
**
ns
*
w.t.
Mucin-0
Mucin-135
Mucin-270
C
Percent of cells in cluster, %
50
40
30
20
10
0
2
3
4
5
6
7
Size of cluster, # of cells
w.t.
Mucin-0
Mucin-135
Mucin-270
D
Ripleys K, A.U.
25000
20000
15000
10000
5000
0
0
10
20
30
40
Distance, µm
w.t.
Mucin-0
Mucin-135
Mucin-270

Figure 14

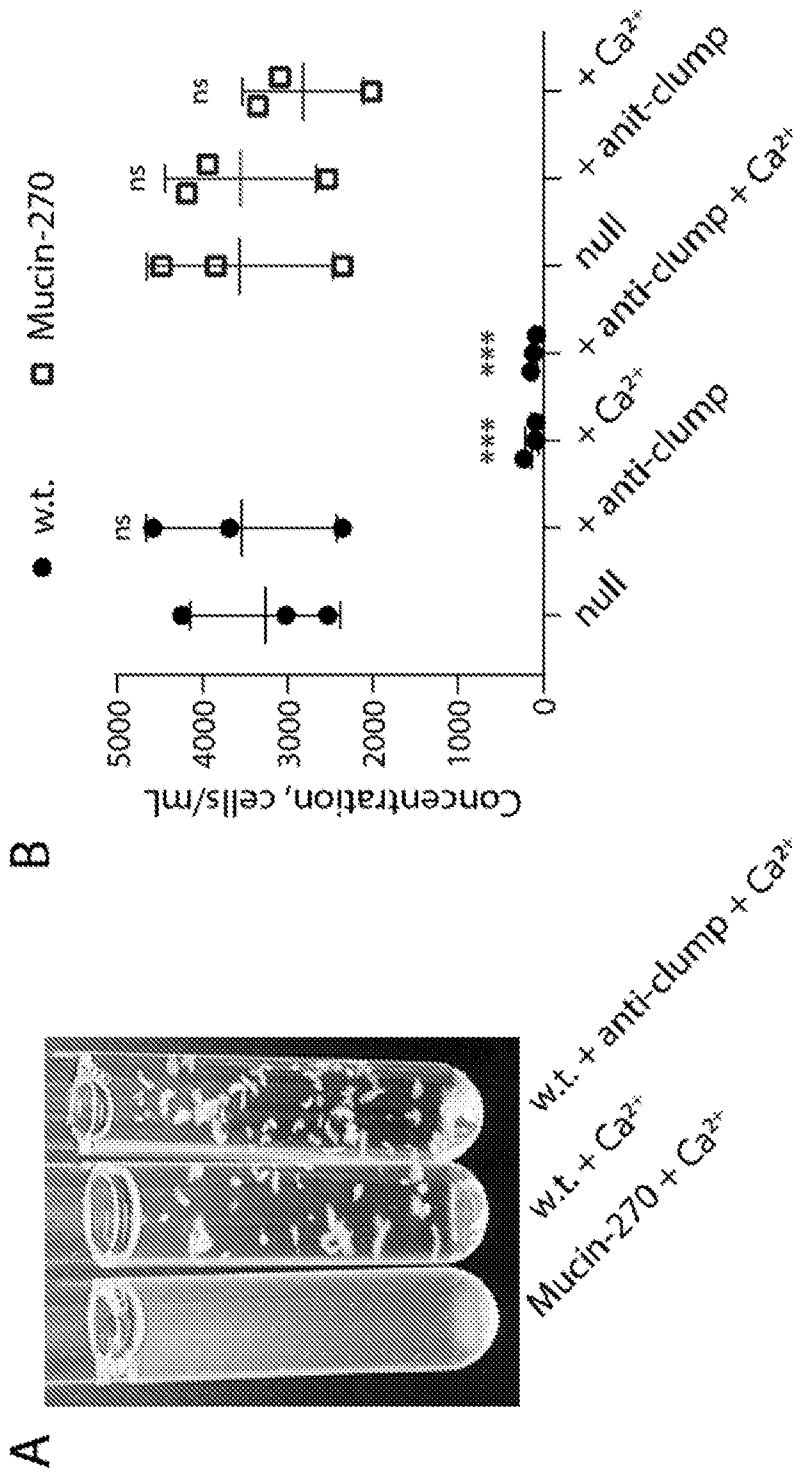

A
w.t. + anti-clump + Ca2+
w.t. + Ca2+
Mucin-270 + Ca2+
B
w.t.
Mucin-270
Concentration, cells/mL
5000
4000
3000
2000
1000
0
ns
ns
ns
***
***
null
+ anti-clump
+ Ca2+
null
+ anti-clump + Ca2+
+ anit-clump
+ Ca2+

Figure 15

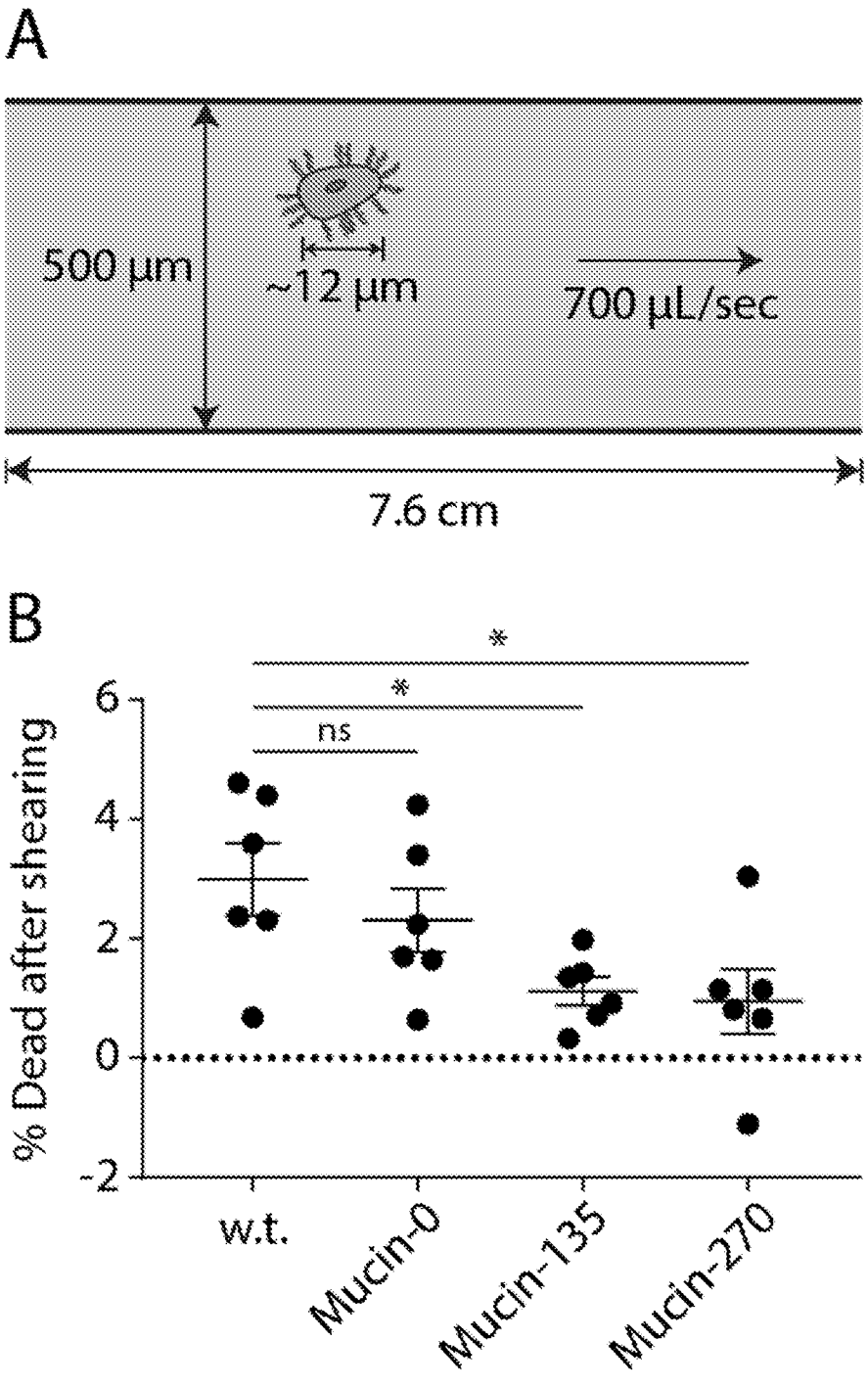

A
500 µm
~12 µm
700 µL/sec
7.6 cm
B
*
*
ns
% Dead after shearing
6
4
2
0
-2
w.t.
Mucin-0
Mucin-135
Mucin-270

Figure 16

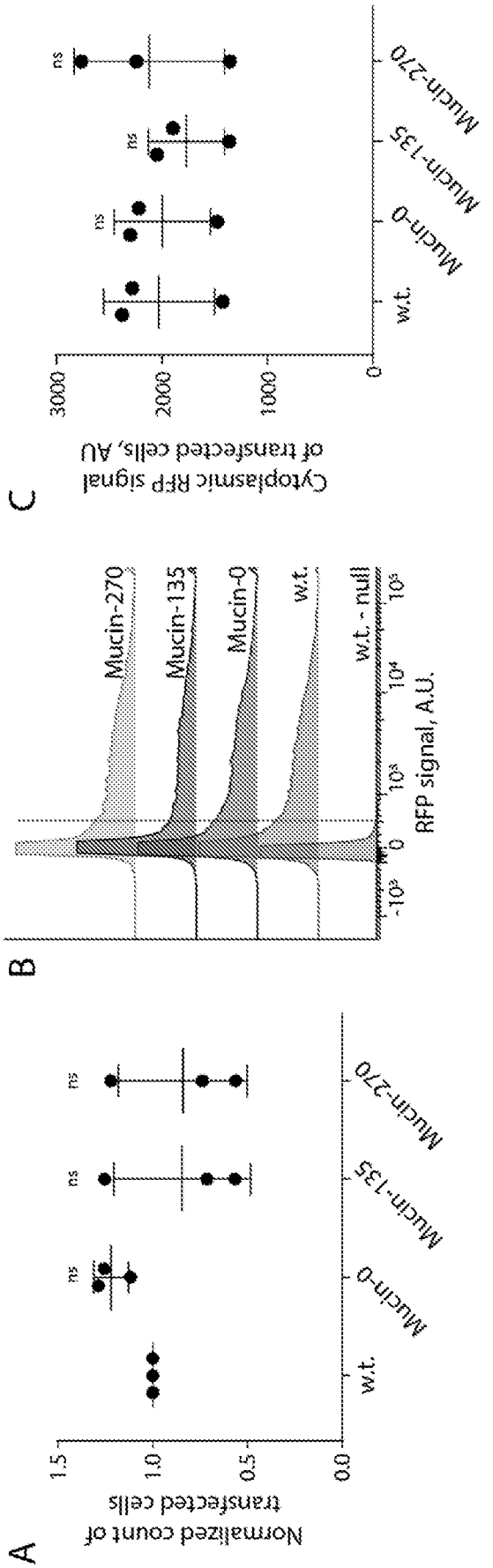


Figure 17

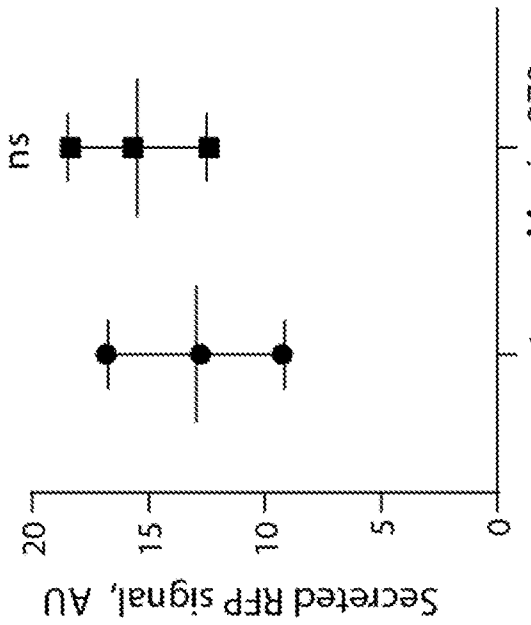


Figure 18

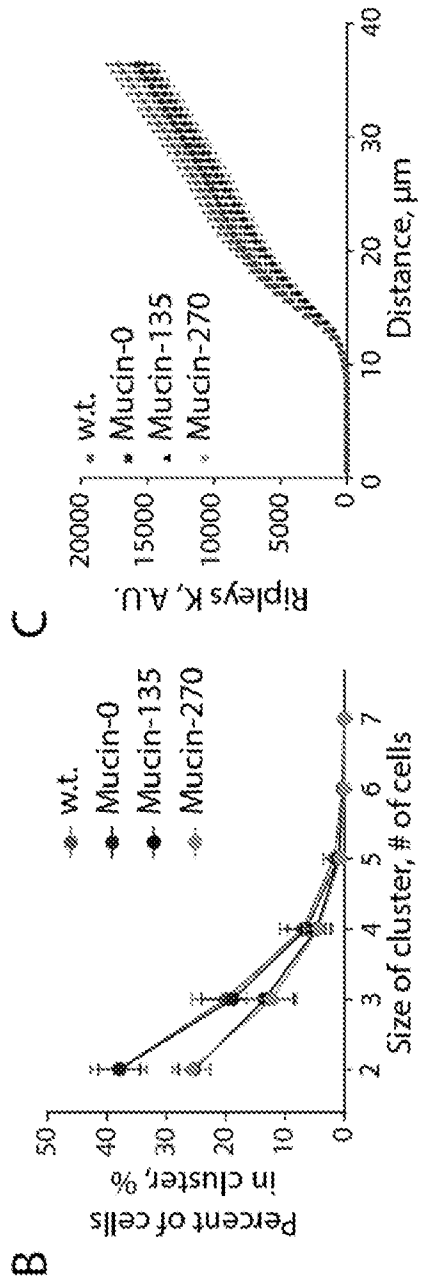

A
% Single cells
70
60
50
40
30
20
***
***
ns
w.t.
Mucin-0
Mucin-135
Mucin-270
B
Percent of cells in cluster, %
50
40
30
20
10
0
2
3
4
5
6
7
Size of cluster, # of cells
w.t.
Mucin-0
Mucin-135
Mucin-270
C
Ripleys K, A.U.
20000
15000
10000
5000
0
0
10
20
30
40
Distance, µm
w.t.
Mucin-0
Mucin-135
Mucin-270

Figure 19

Part III Figures

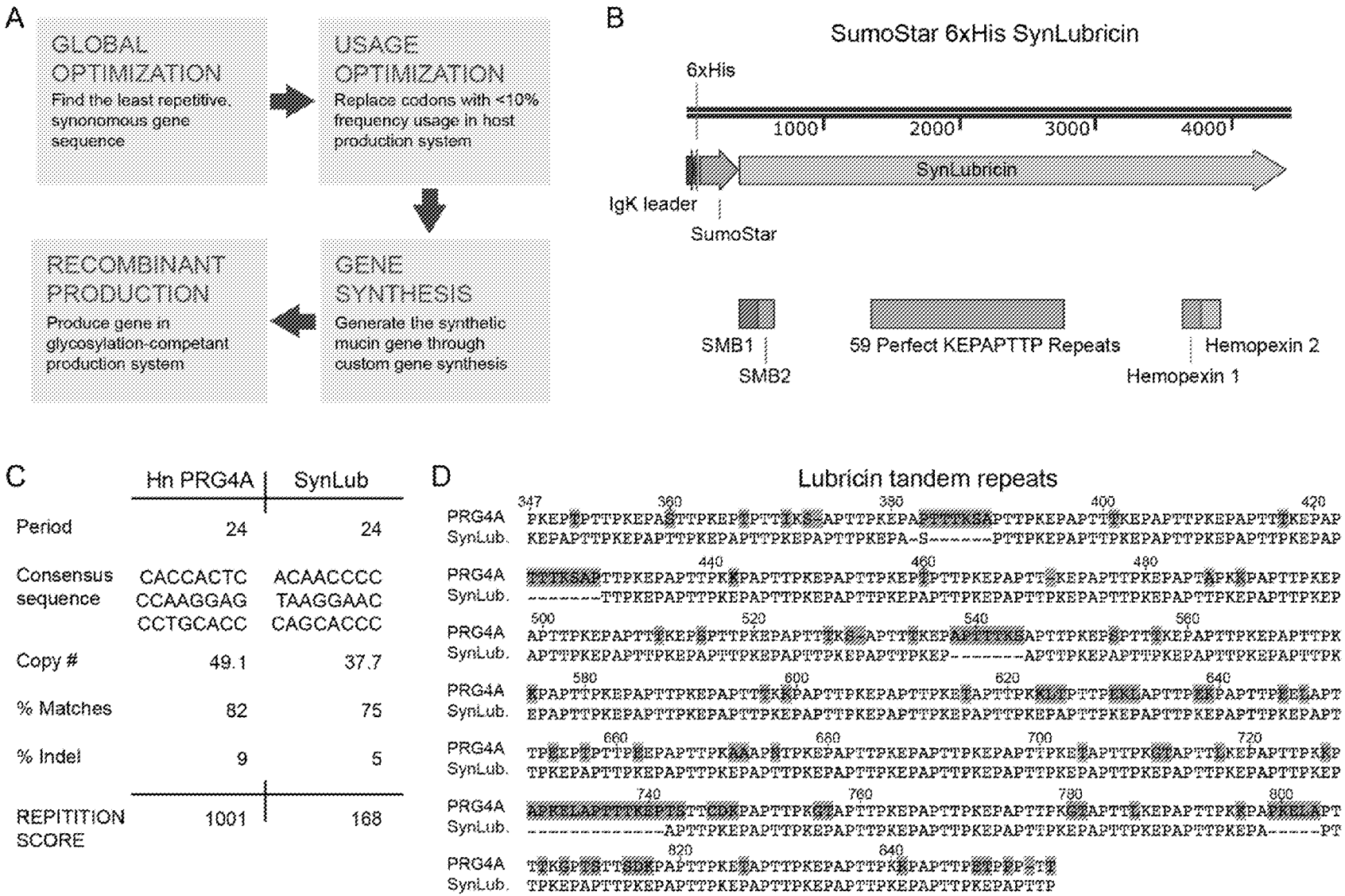


Figure 20

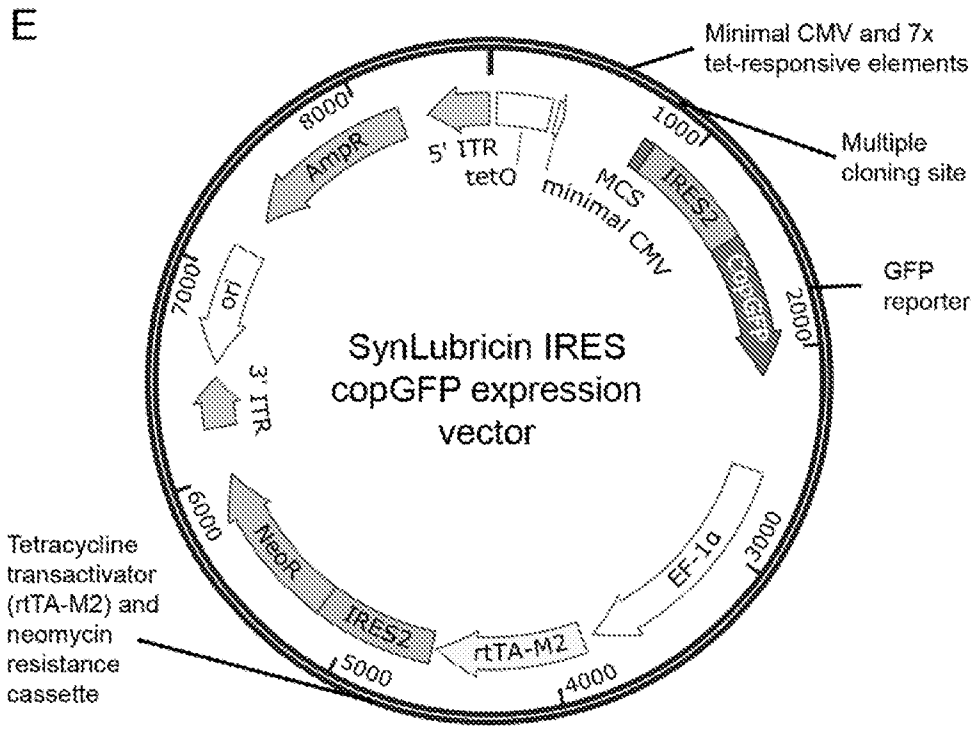


Figure 20 (continued)

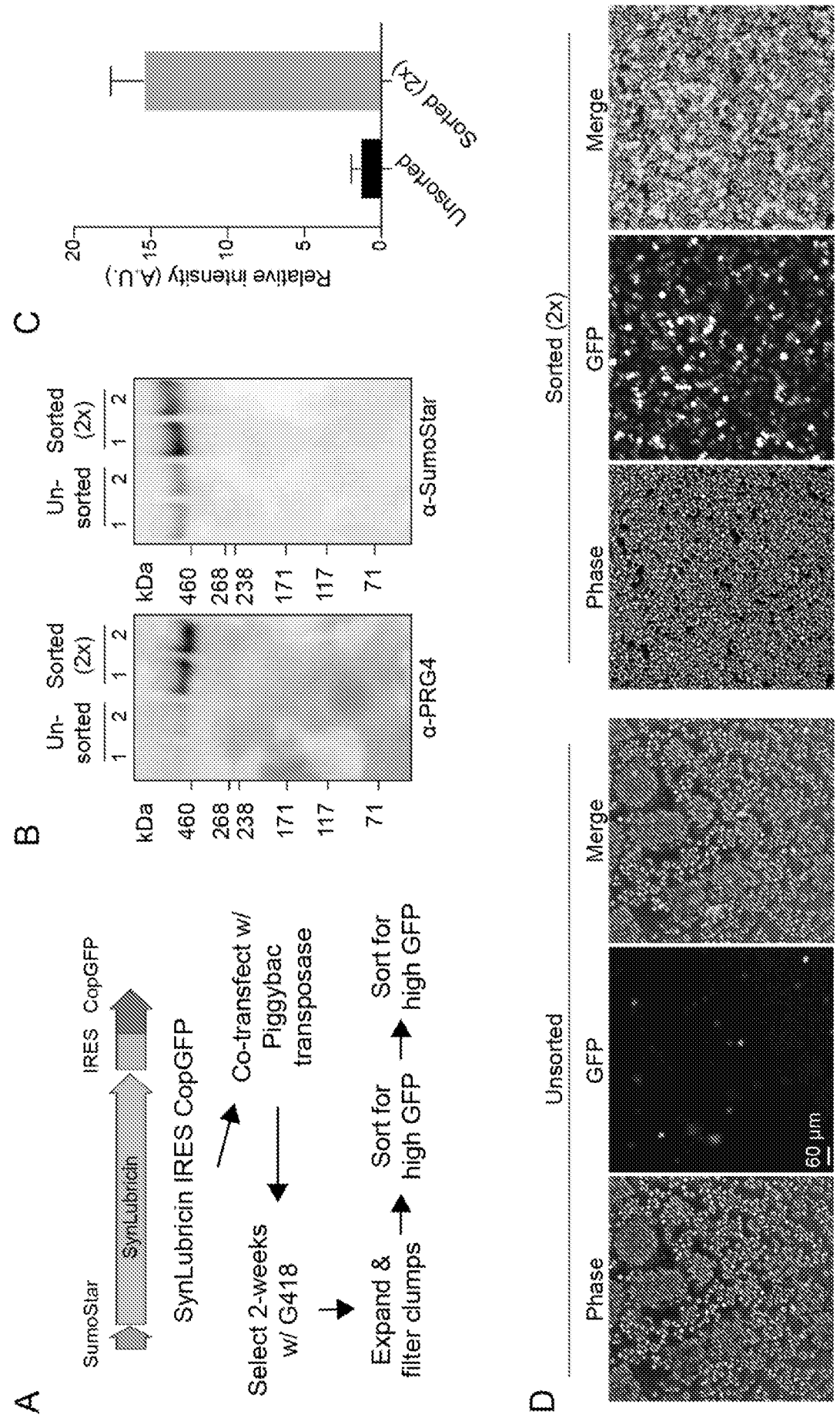

A
SumoStar
SynLubricin
IRES
CopGFP
SynLubricin IRES CopGFP
Co-transfect w/ Piggybac transposase
Select 2-weeks w/ G418
Expand & filter clumps
Sort for high GFP
Sort for high GFP
B
Un-sorted
Sorted (2x)
1 2 1 2
kDa
460
268
238
171
117
71
α-PRG4
Un-sorted
Sorted (2x)
1 2 1 2
kDa
460
268
238
171
117
71
α-SumoStar
C
Relative intensity (A.U.)
20
15
10
5
0
Unsorted
Sorted (2x)
D
Unsorted
Sorted (2x)
Phase
GFP
Merge
Phase
GFP
Merge
60 µm

Figure 21

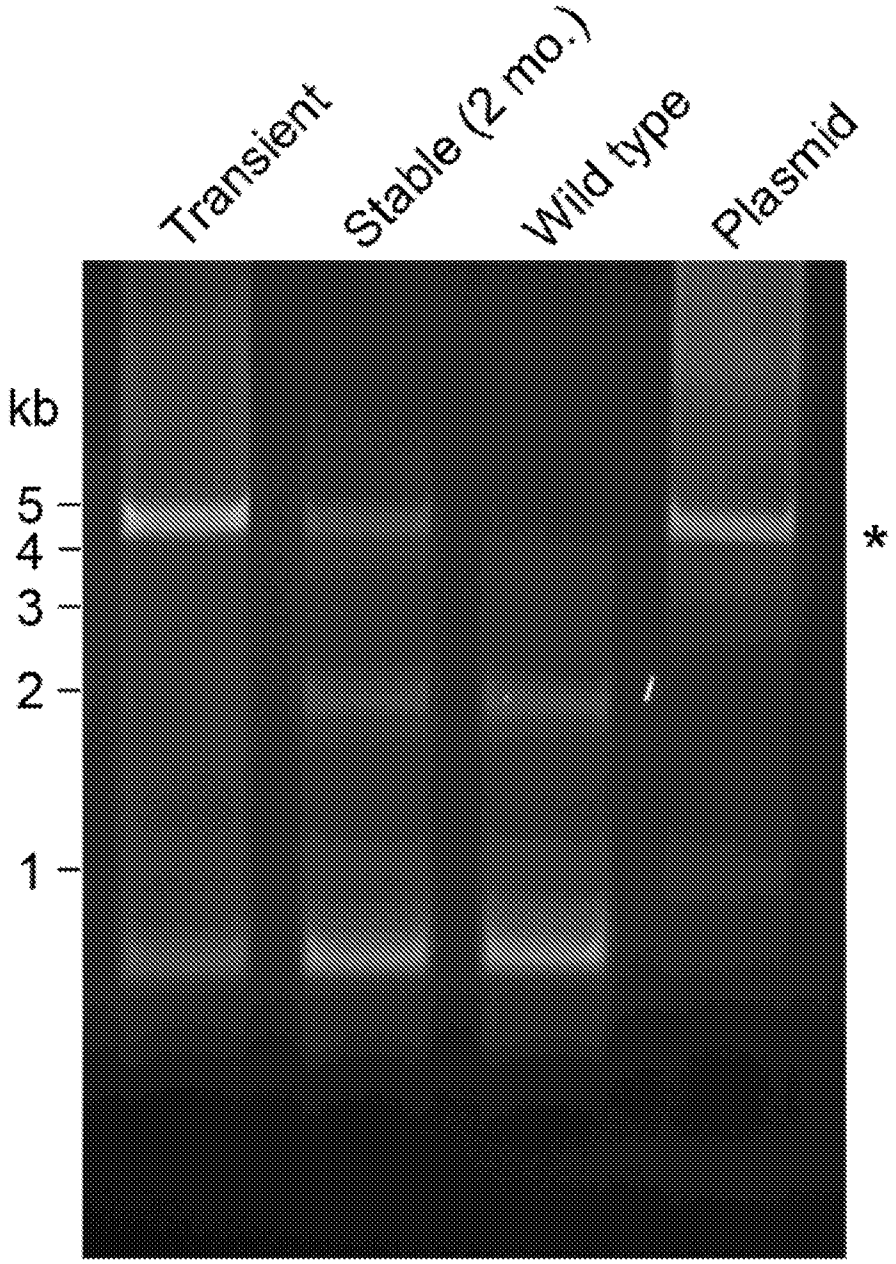


Figure 22

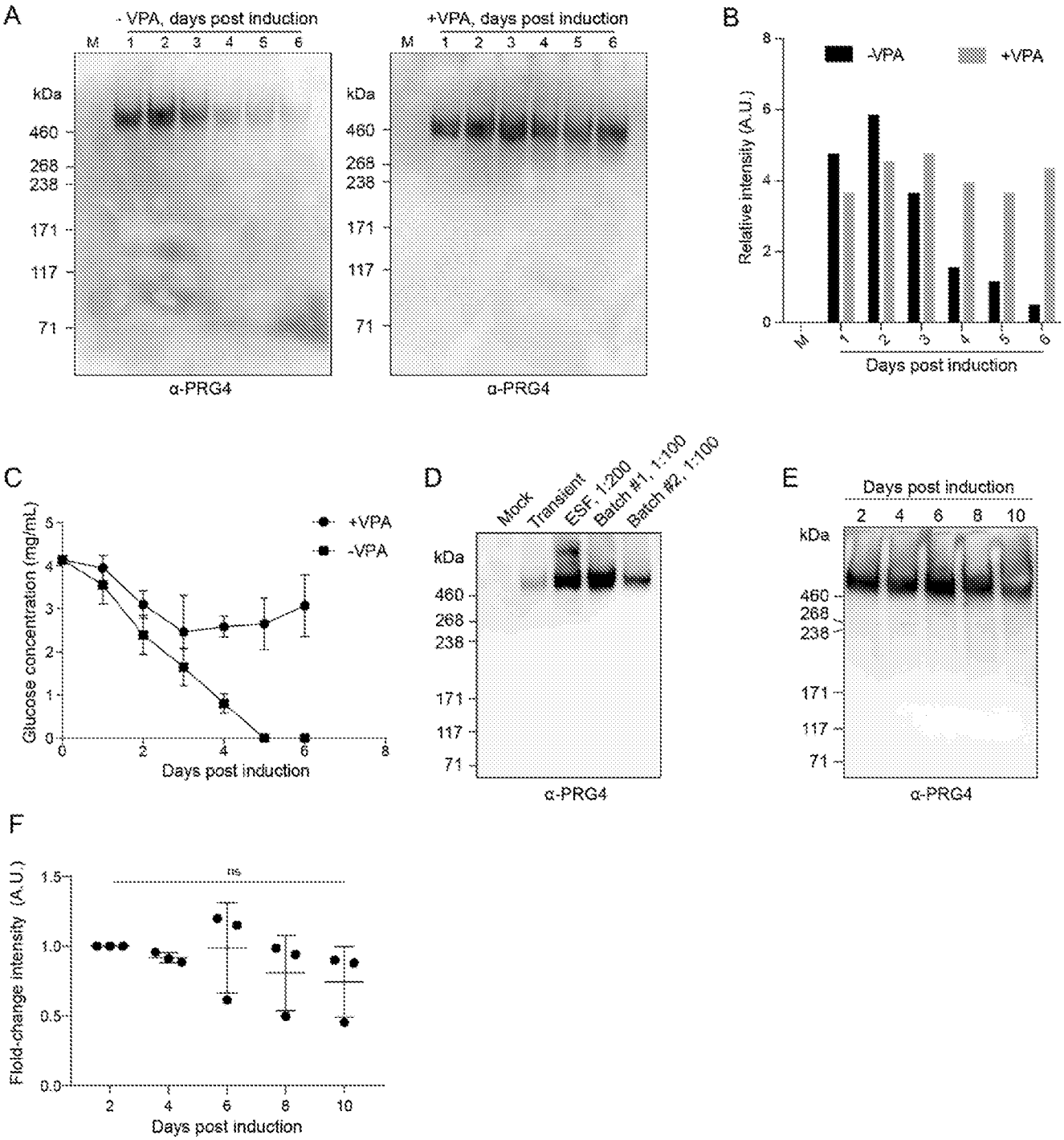

A
- VPA, days post induction
+VPA, days post induction
M 1 2 3 4 5 6
kDa
460
268
238
171
117
71
α-PRG4
B
-VPA
+VPA
Relative intensity (A.U.)
Days post induction
C
+VPA
-VPA
Glucose concentration (mg/mL)
Days post induction
D
Mock
Transient
ESF, 1:200
Batch #1, 1:100
Batch #2, 1:100
α-PRG4
E
Days post induction
2 4 6 8 10
α-PRG4
F
ns
Fold-change intensity (A.U.)
Days post induction

Figure 23

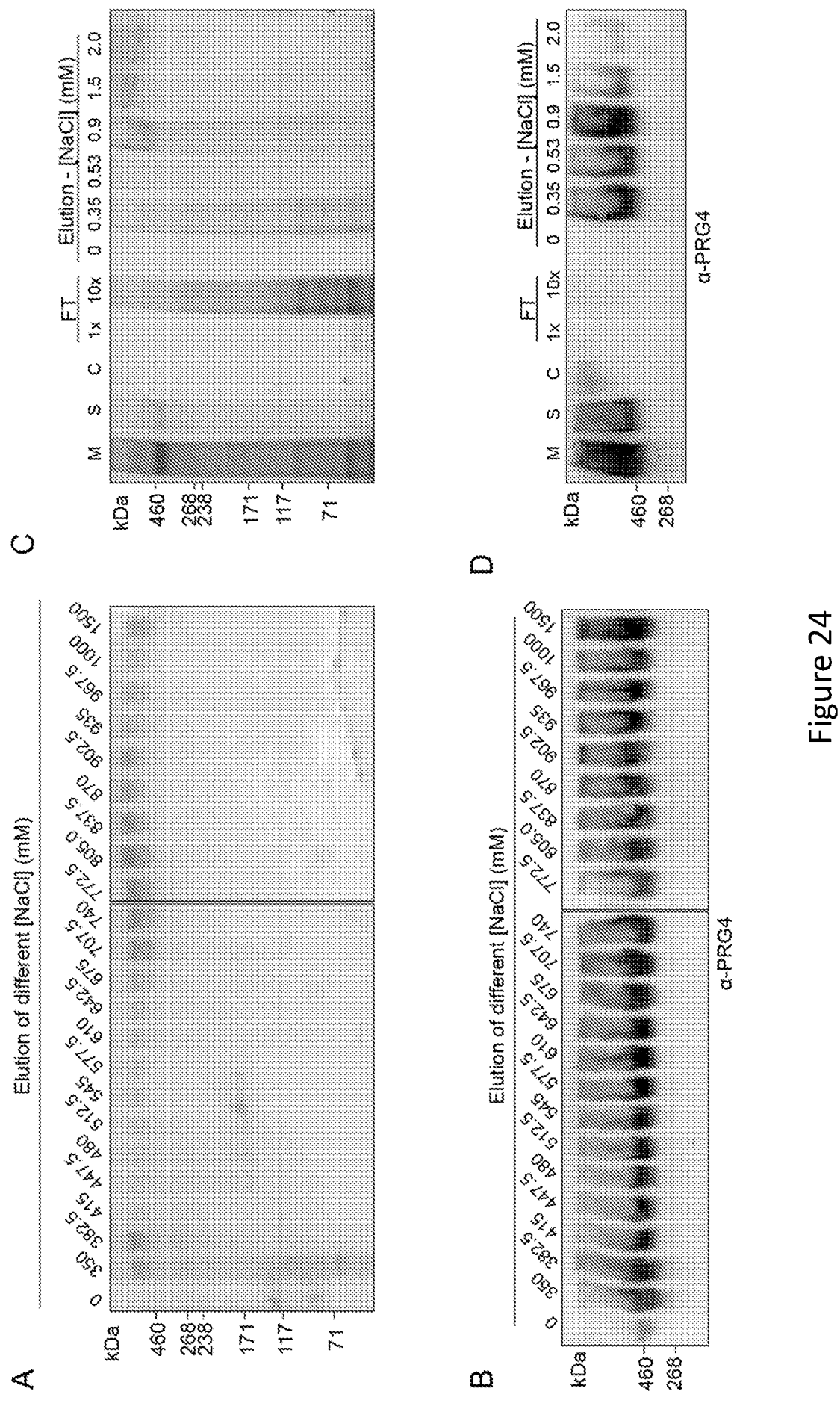

A
Elution of different [NaCl] (mM)
0
350
382.5
415
447.5
480
512.5
545
577.5
610
642.5
675
707.5
740
772.5
805.0
837.5
870
902.5
935
967.5
1000
1500
kDa
460
268
238
171
117
71
B
Elution of different [NaCl] (mM)
α-PRG4
C
M
S
C
FT
1x
10x
Elution - [NaCl] (mM)
0
0.35
0.53
0.9
1.5
2.0
D
α-PRG4

Figure 24

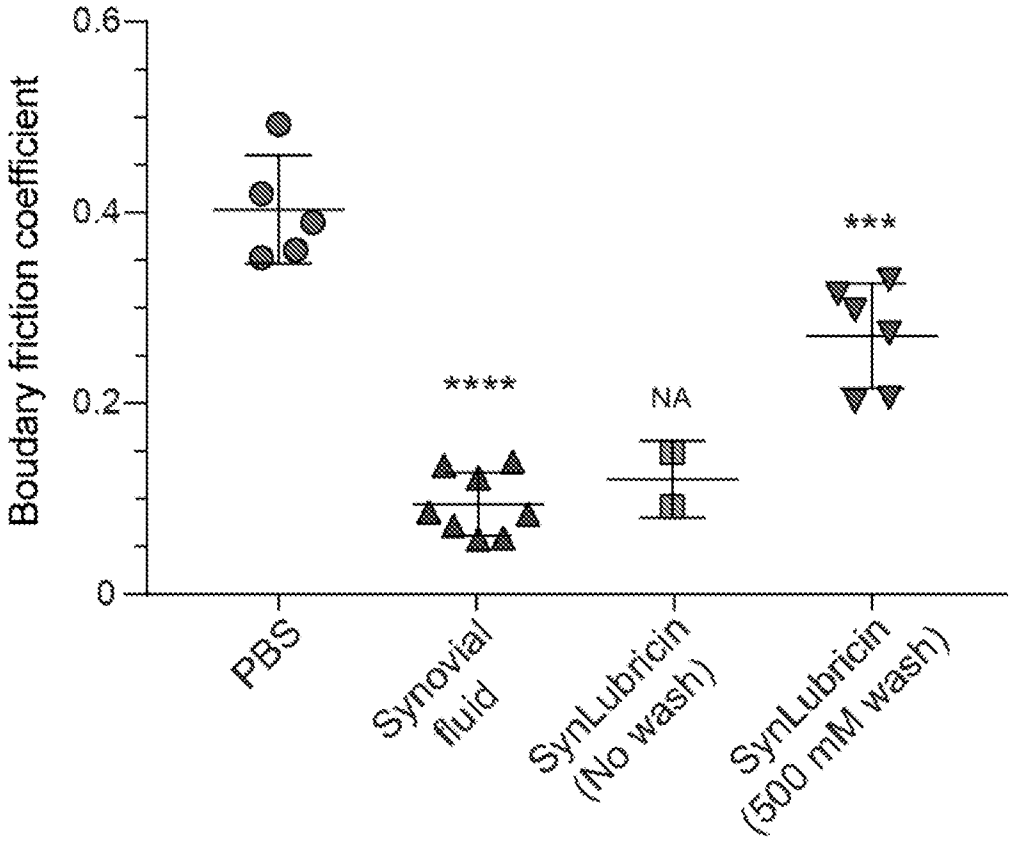

Boudary friction coefficient
0.6
0.4
0.2
0
****
NA
***
PBS
Synovial fluid
SynLubricin (No wash)
SynLubricin (500 mM wash)

Figure 25

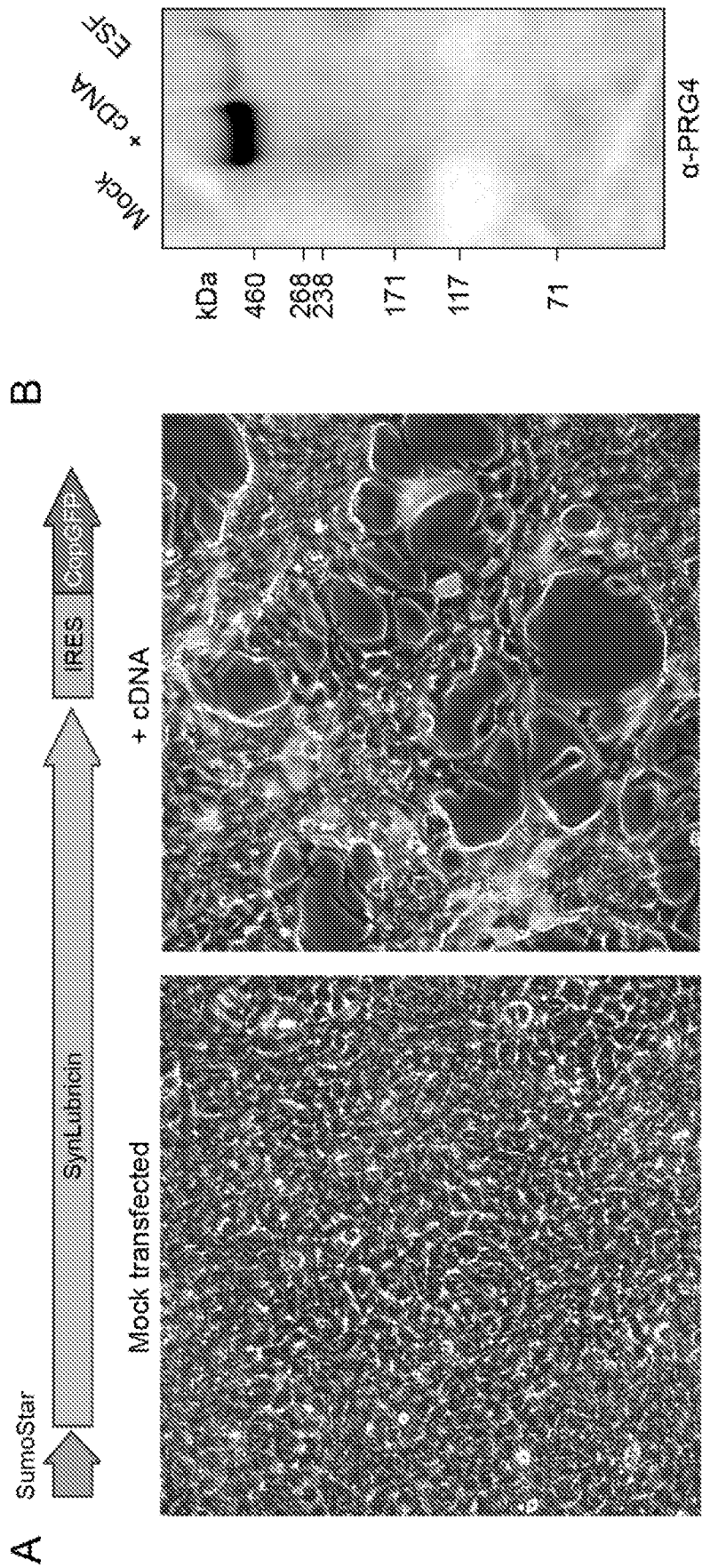

A
SumoStar
SynLubricin
IRES
eGFP
Mock transfected
+ cDNA
B
kDa
460
268
238
171
117
71
Mock
+ cDNA
ESF
α-PRG4

Figure 26

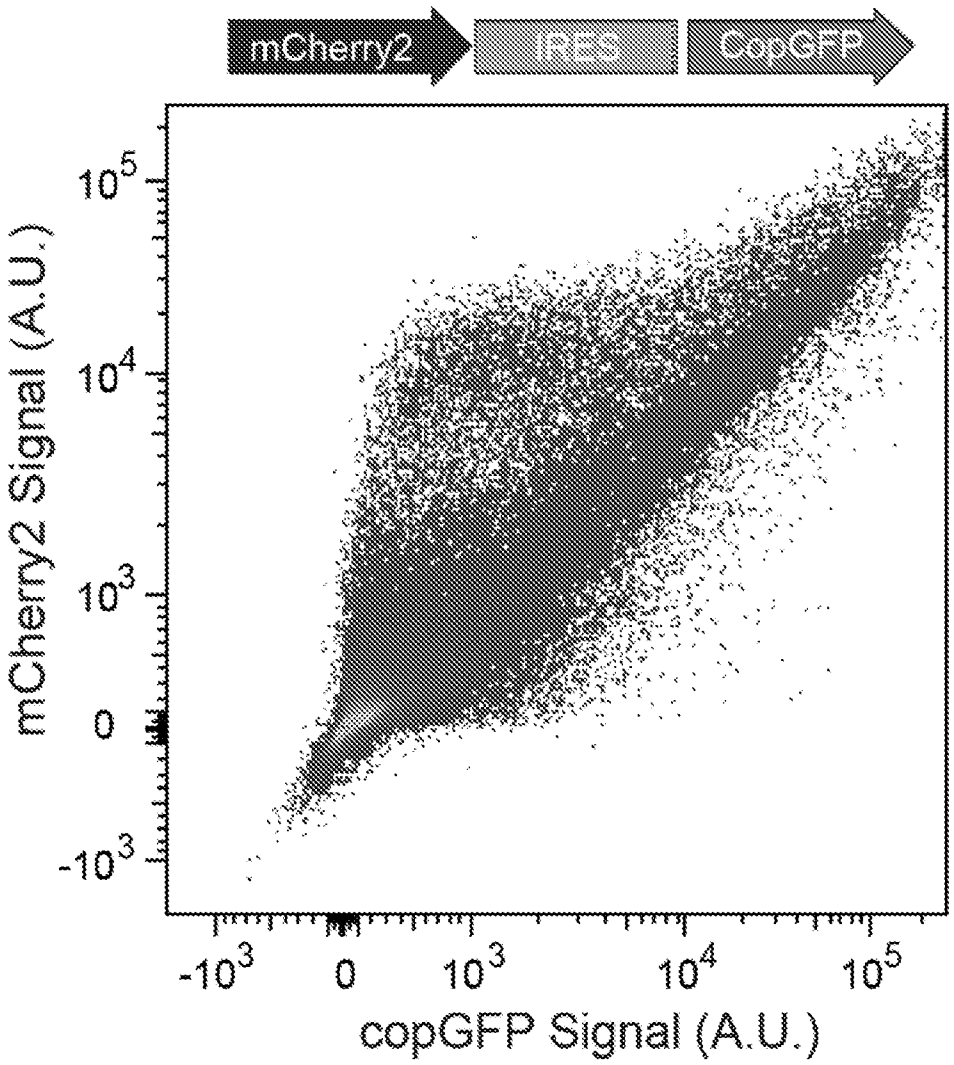

mCherry2
IRES
CopGFP
mCherry2 Signal (A.U.)
10^5
10^4
10^3
0
-10^3
-10^3
0
10^3
10^4
10^5
copGFP Signal (A.U.)

Figure 27

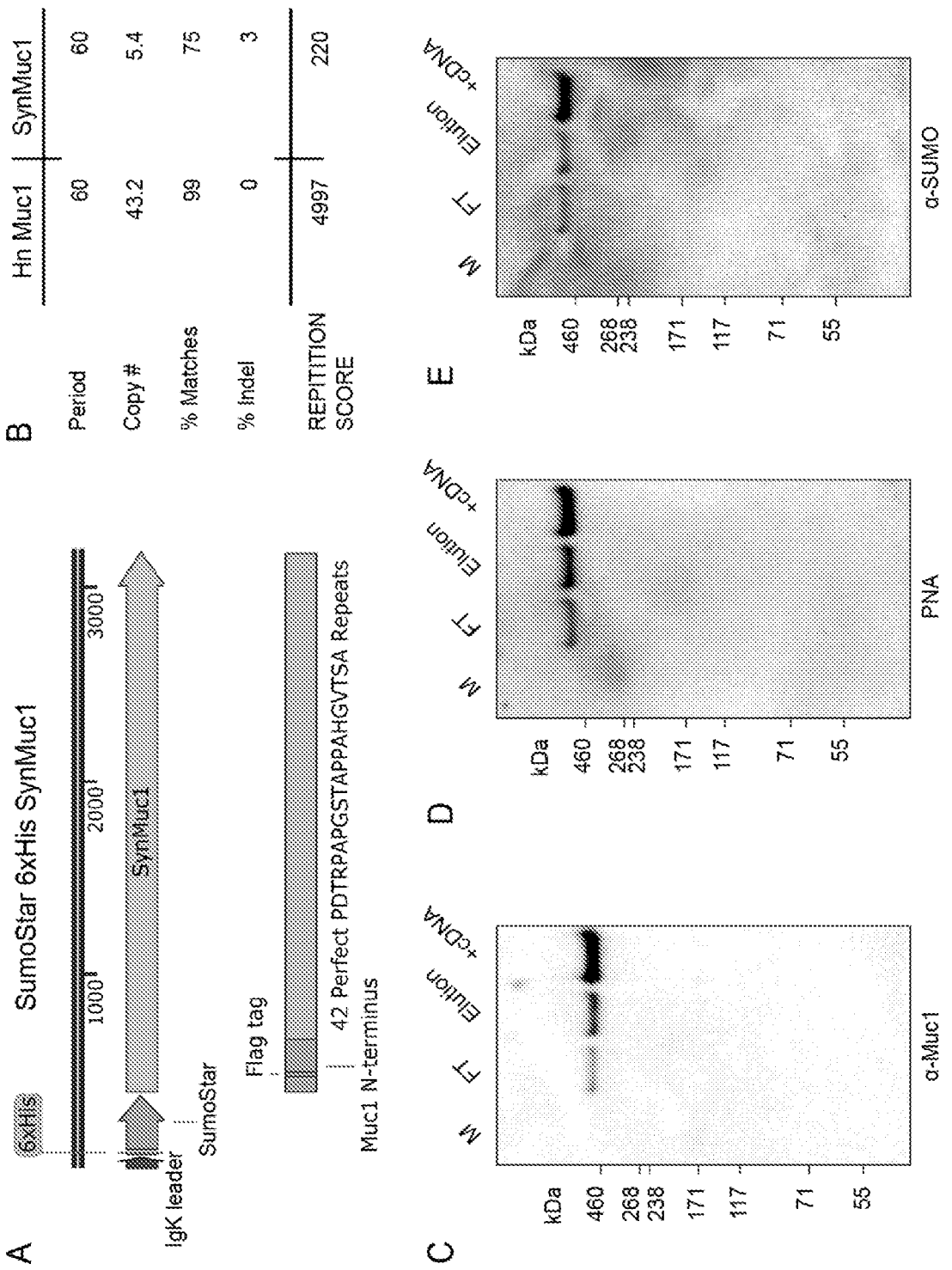

A
SumoStar 6xHis SynMuc1
6xHis
1000
2000
3000
SynMuc1
IgK leader
SumoStar
Flag tag
42 Perfect PDTRPAPGSTAPPAHGVTSA Repeats
Muc1 N-terminus
B
Hn Muc1
SynMuc1
Period 60 60
Copy # 43.2 5.4
% Matches 99 75
% Indel 0 3
REPITITION SCORE 4997 220
C
M
FT
Elution
+cDNA
kDa
460
268
238
171
117
71
55
α-Muc1
D
PNA
E
α-SUMO

Figure 28

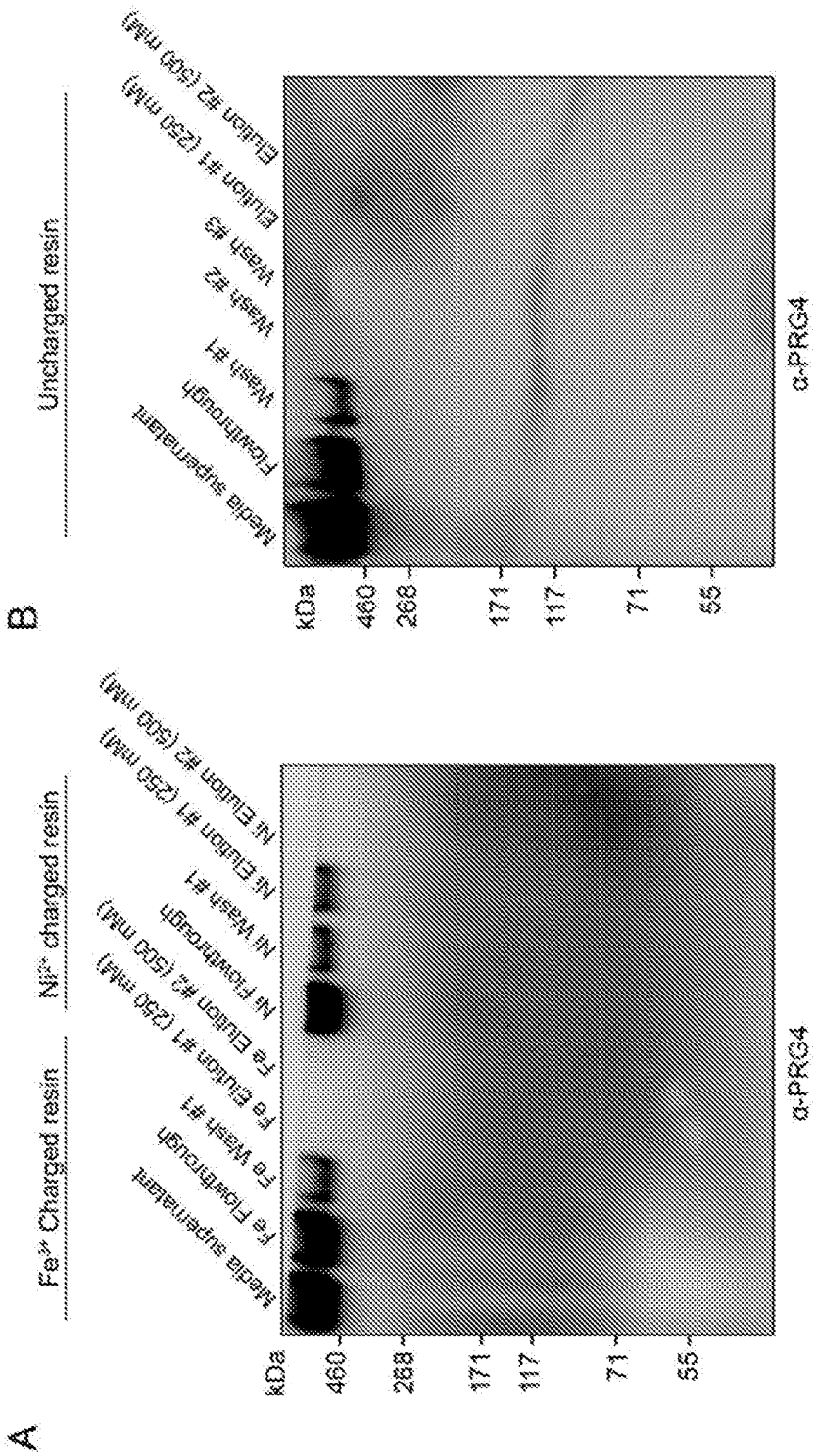

A
Fe3+ Charged resin
Ni2+ charged resin
Media supernatant
Fe Flowthrough
Fe Wash #1
Fe Elution #1 (250 mM)
Fe Elution #2 (500 mM)
Ni Flowthrough
Ni Wash #1
Ni Elution #1 (250 mM)
Ni Elution #2 (500 mM)
kDa
460
268
171
117
71
55
α-PRG4
B
Uncharged resin
Media supernatant
Flowthrough
Wash #1
Wash #2
Wash #3
Elution #1 (250 mM)
Elution #2 (500 mM)
kDa
460
268
171
117
71
55
α-PRG4

Figure 29

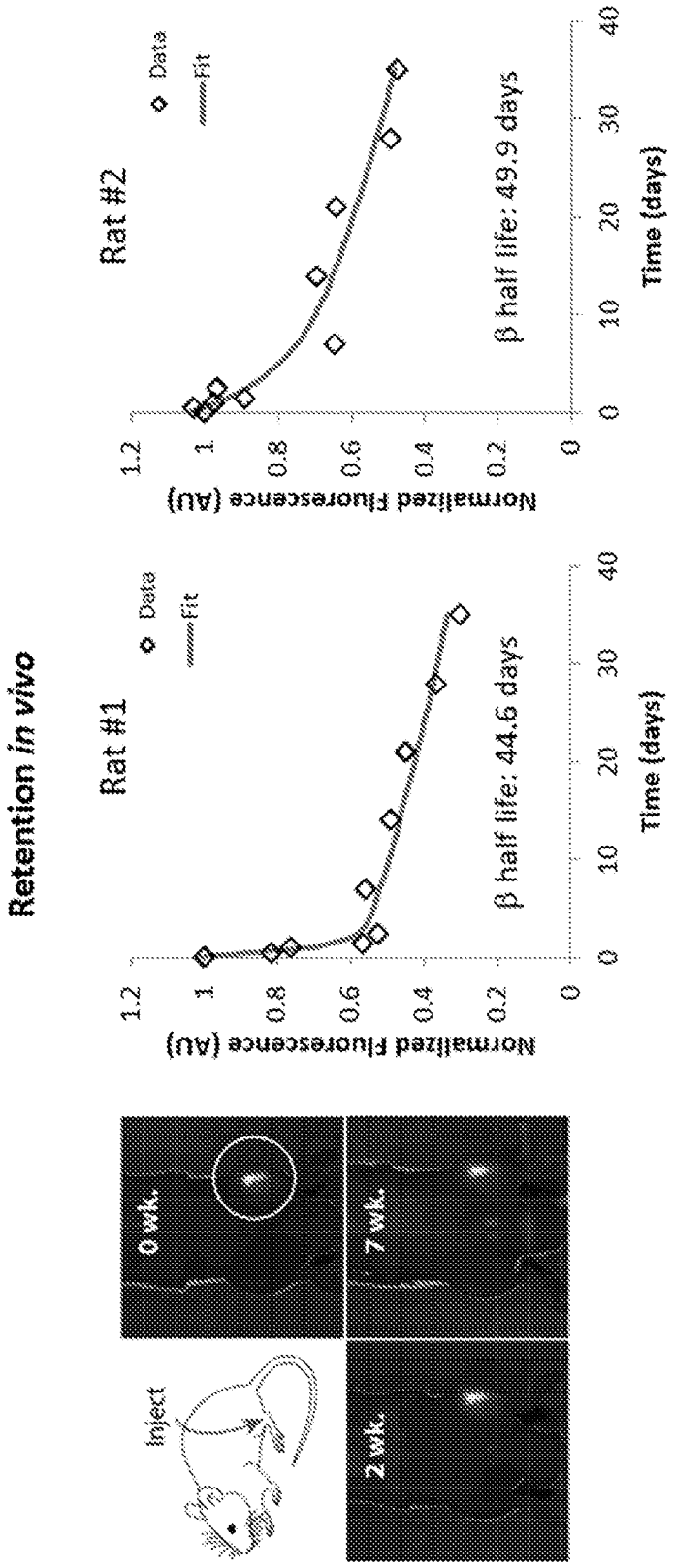
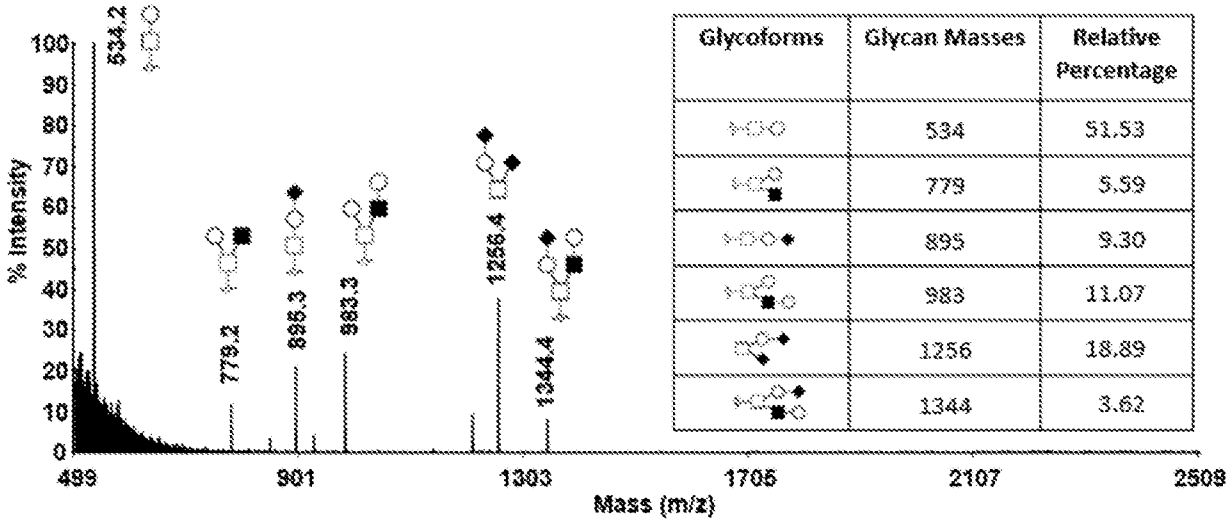

Retention in vivo
Rat #1
Data
Fit
β half life: 44.6 days
Normalized Fluorescence (AU)
Time (days)
Rat #2
Data
Fit
β half life: 49.9 days
Normalized Fluorescence (AU)
Time (days)
Inject
0 wk.
2 wk.
7 wk.

Figure 30



Glycoforms	Glycan Masses	Relative Percentage
	534	51.53
	779	5.59
	895	9.30
	983	11.07
	1256	18.89
	1344	3.62

Figure 31

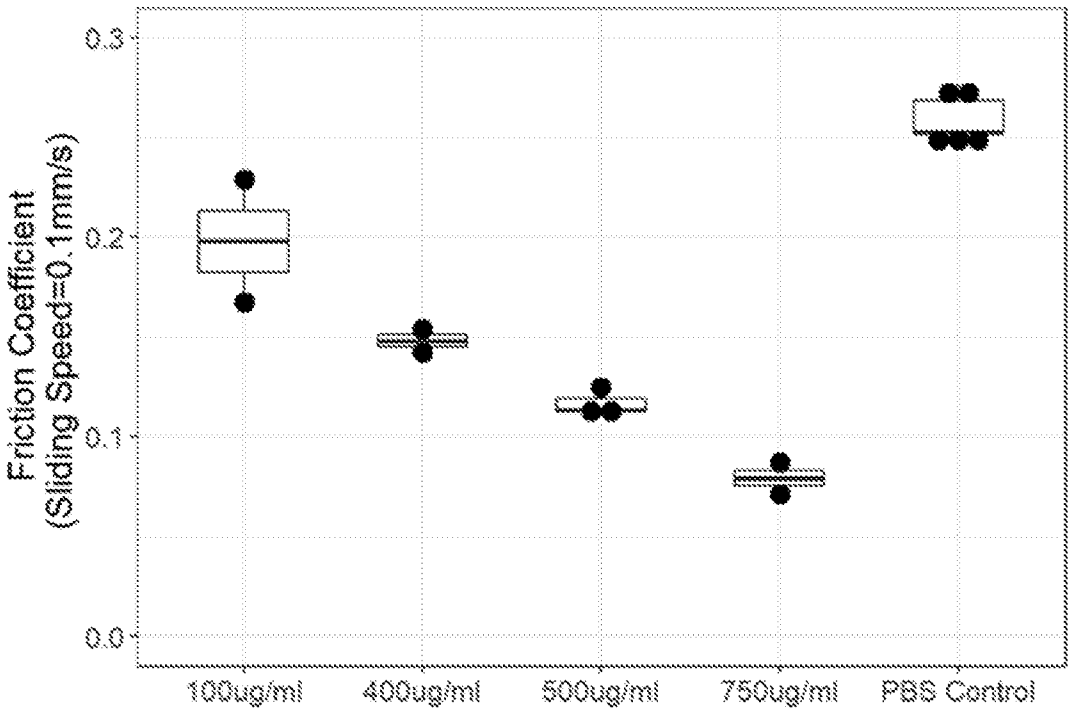

Friction Coefficient
(Sliding Speed=0.1mm/s)
0.3
0.2
0.1
0.0
100ug/ml
400ug/ml
500ug/ml
750ug/ml
PBS Control

Figure 32

Part IV Figures

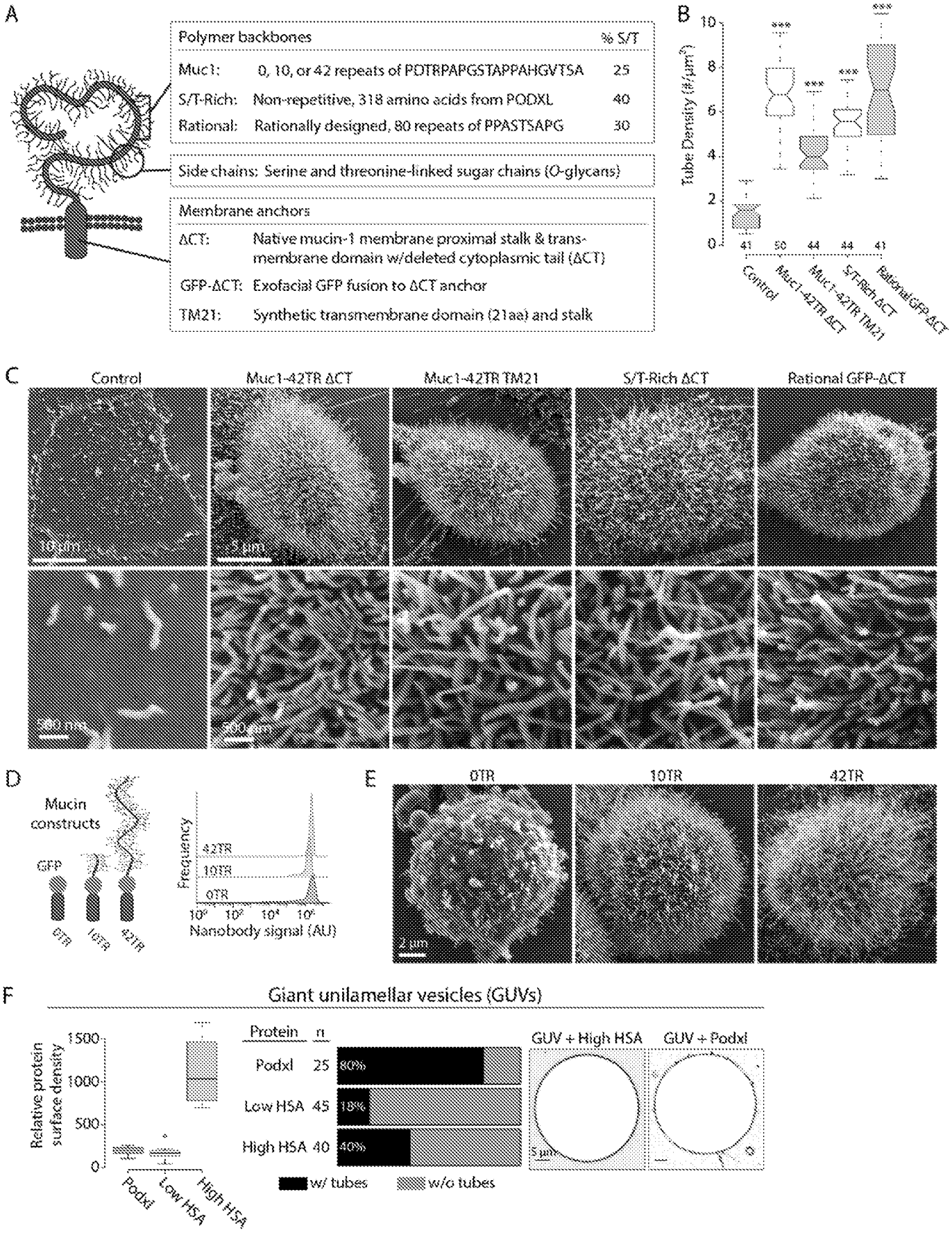


Figure 33

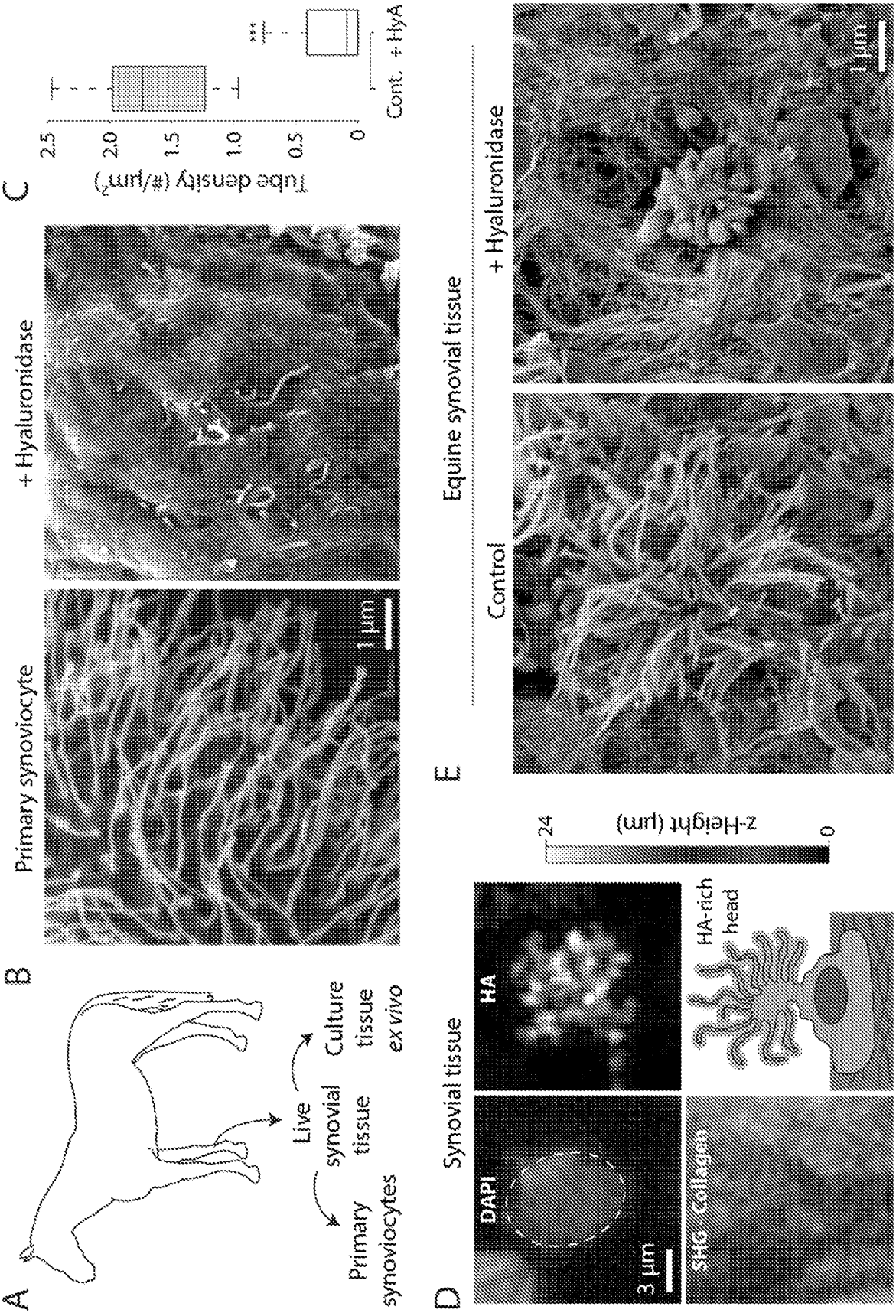

A
Live synovial tissue
Culture tissue ex vivo
Primary synoviocytes
B
Primary synoviocyte
+ Hyaluronidase
1 μm
C
Tube density (#/μm²)
2.5
2.0
1.5
1.0
0.5
0
***
Cont.
+ HyA
D
Synovial tissue
DAPI
3 μm
HA
SHG Collagen
HA-rich head
z-Height (μm)
24
0
E
Equine synovial tissue
Control
+ Hyaluronidase
1 μm

Figure 34

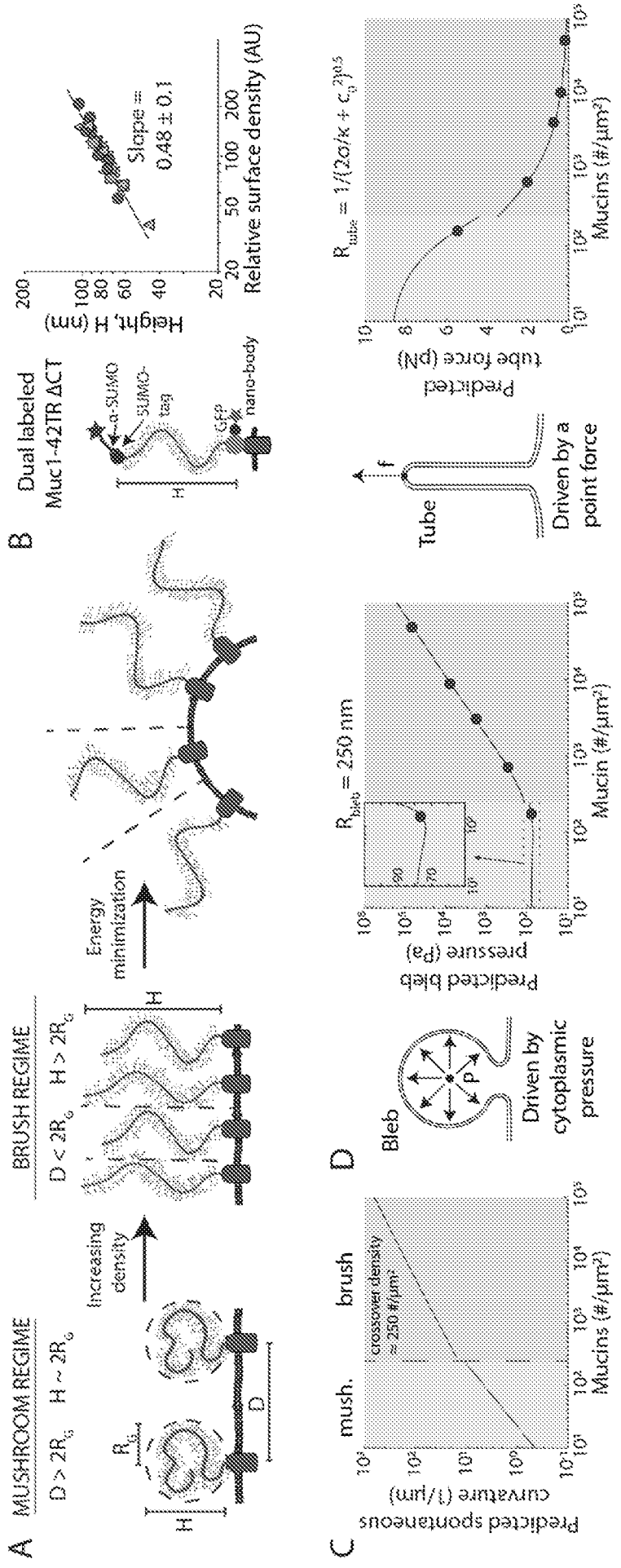

A
MUSHROOM REGIME
D > 2R_G H ~ 2R_G
BRUSH REGIME
D < 2R_G H > 2R_G
Increasing density
Energy minimization
B
Dual labeled Muc1-42TR ΔCT
a-SUMO
SUMO-tag
GFP
nano-body
Height, H (nm)
Relative surface density (AU)
Slope = 0.48 ± 0.1
C
mush.
brush
crossover density ≈ 250 #/μm²
Predicted spontaneous curvature (1/μm)
Mucins (#/μm²)
D
Bleb
Driven by cytoplasmic pressure
Predicted bleb pressure (Pa)
R_bleb = 250 nm
Mucin (#/μm²)
Tube
Driven by a point force
Predicted tube force (pN)
R_tube = 1/(2σ/κ + c_0^2)^0.5
Mucins (#/μm²)

Figure 35

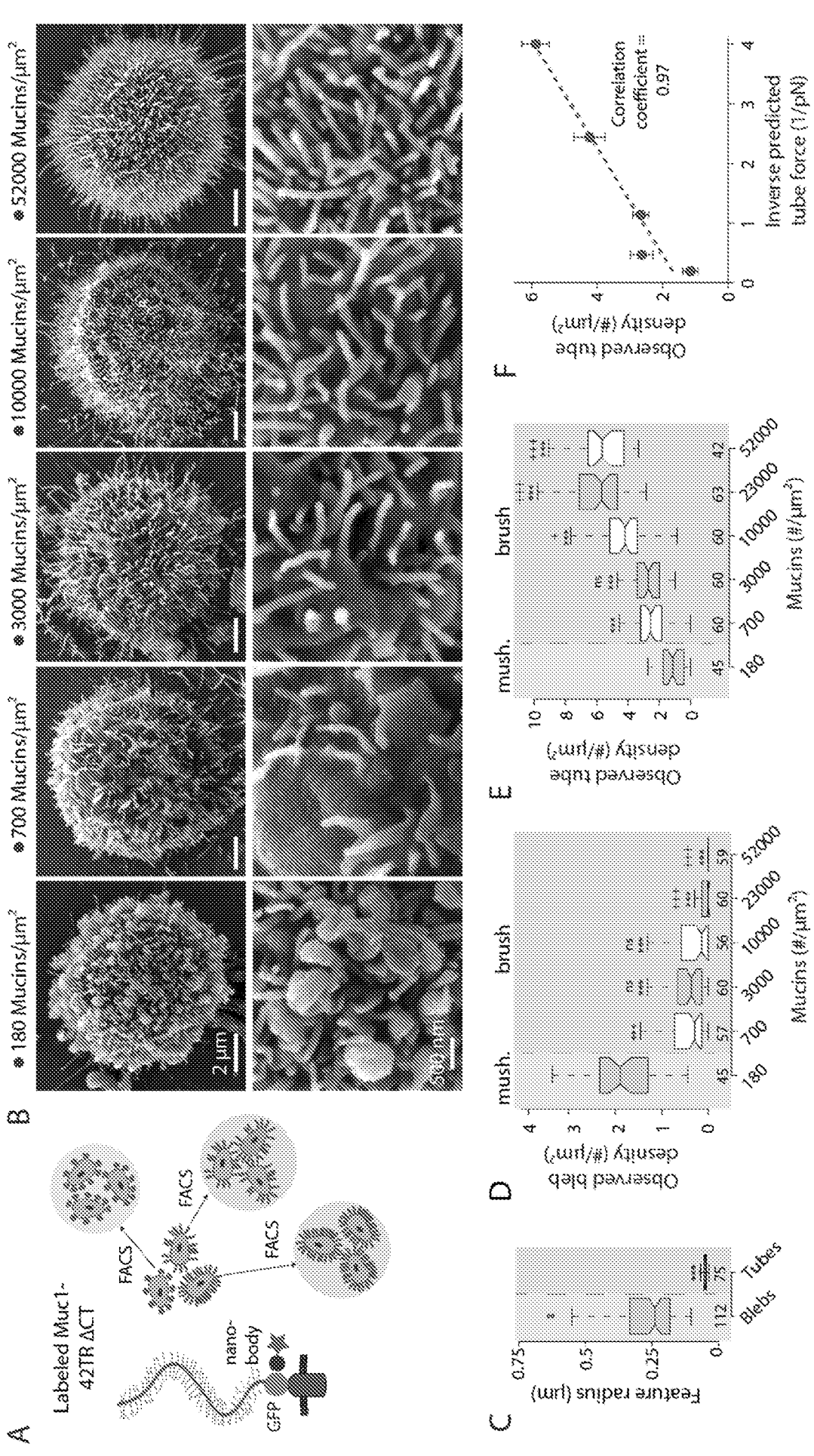

A
Labeled Muc1-
42TR ΔCT
nano-
body
GFP
FACS
FACS
FACS
B
180 Mucins/μm²
700 Mucins/μm²
3000 Mucins/μm²
10000 Mucins/μm²
52000 Mucins/μm²
2 μm
C
Feature radius (μm)
0.75
0.5
0.25
0
112
75
***
Tubes
Blebs
D
Observed bleb
density (#/μm²)
mush.
brush
Mucins (#/μm²)
45
57
60
56
60
59
180
700
3000
10000
23000
52000
E
Observed tube
density (#/μm²)
mush.
brush
Mucins (#/μm²)
45
60
60
60
63
42
180
700
3000
10000
23000
52000
F
Observed tube
density (#/μm²)
Correlation
coefficient =
0.97
Inverse predicted
tube force (1/pN)

Figure 36

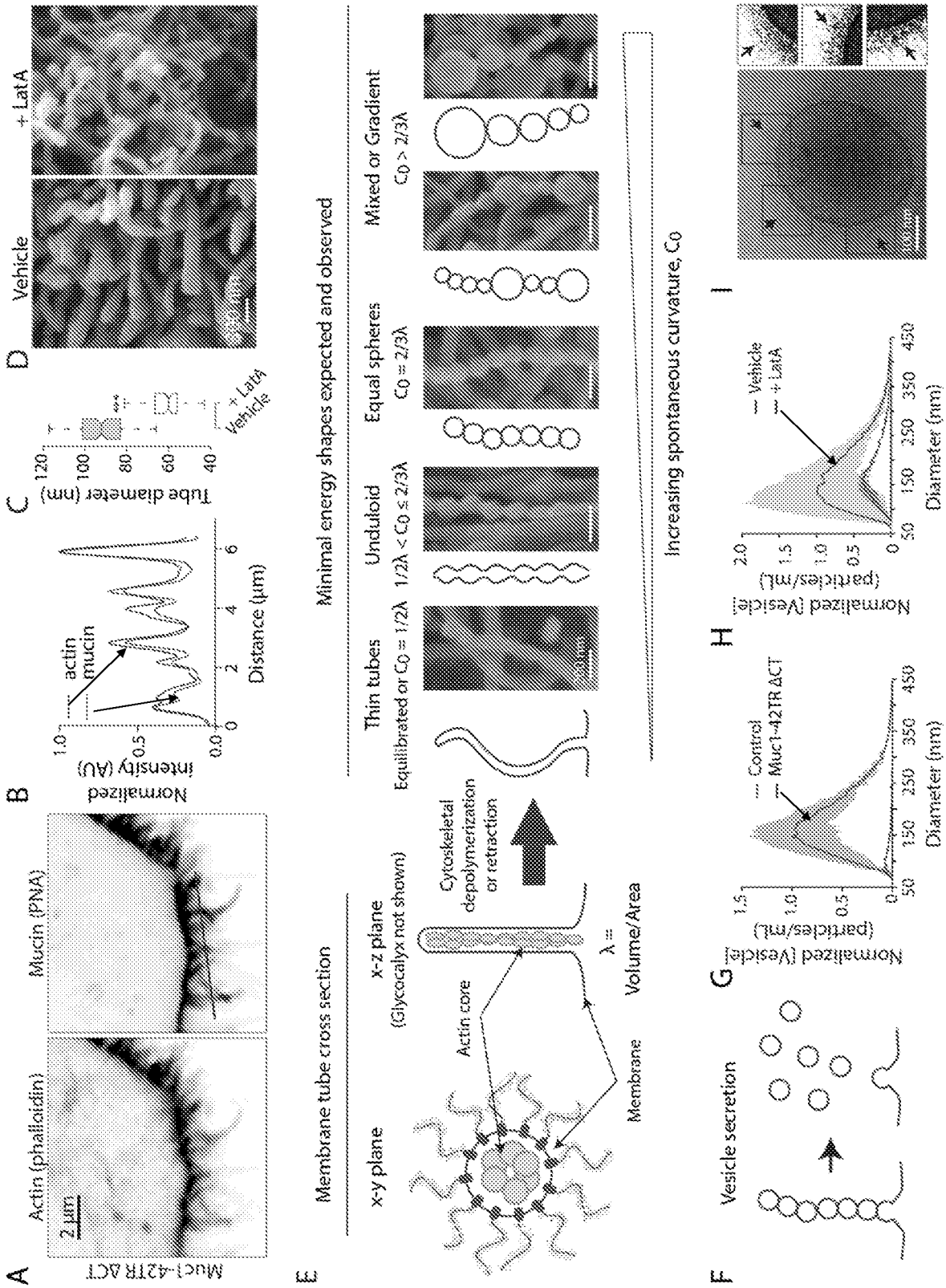

A
Actin (phalloidin)
Mucin (PNA)
2 μm
Muc1-42TR ΔCT
B
Normalized intensity (AU)
actin
mucin
Distance (μm)
C
Tube diameter (nm)
Vehicle
+ LatA
D
Vehicle
+ LatA
E
Membrane tube cross section
x-y plane
x-z plane
(Glycocalyx not shown)
Actin core
Membrane
λ = Volume/Area
Cytoskeletal depolymerization or retraction
Minimal energy shapes expected and observed
Thin tubes
Equilibrated or C0 = 1/2λ
Unduloid
1/2λ < C0 ≤ 2/3λ
Equal spheres
C0 = 2/3λ
Mixed or Gradient
C0 > 2/3λ
Increasing spontaneous curvature, C0
F
Vesicle secretion
G
Normalized [Vesicle] (particles/mL)
Control
Muc1-42TR ΔCT
Diameter (nm)
H
Normalized [Vesicle] (particles/mL)
Vehicle
+ LatA
Diameter (nm)
I

Figure 37

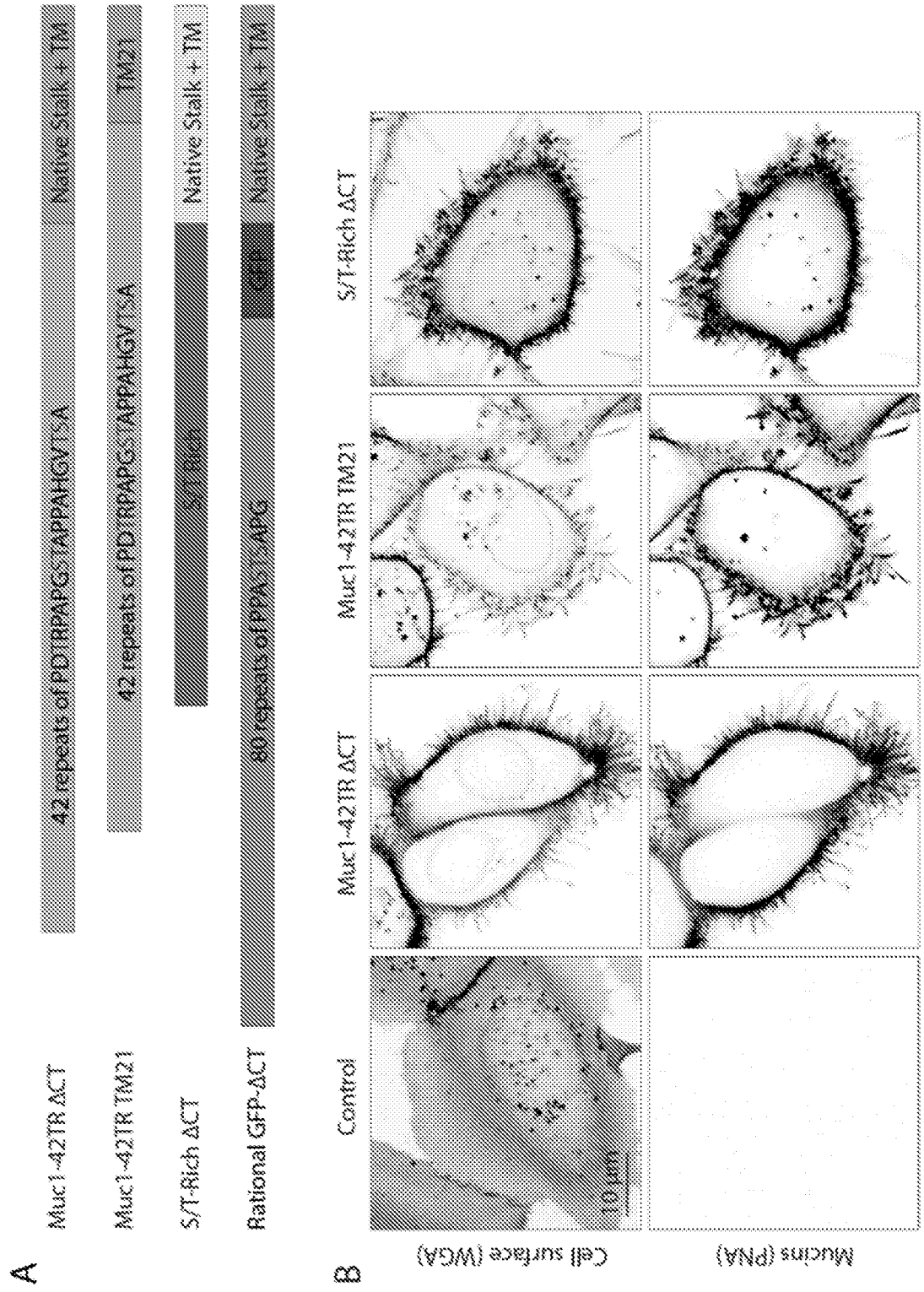

A
Muc1-42TR ΔCT
42 repeats of PDTRPAPGSTAPPAHGVTSA
Native Stalk + TM
Muc1-42TR TM21
42 repeats of PDTRPAPGSTAPPAHGVTSA
TM21
S/T-Rich ΔCT
Native Stalk + TM
Rational GFP-ΔCT
80 repeats of PPAT_APG
Native Stalk + TM
B
Control
Muc1-42TR ΔCT
Muc1-42TR TM21
S/T-Rich ΔCT
Cell surface (WGA)
Mucins (PNA)
10 μm

Figure 38

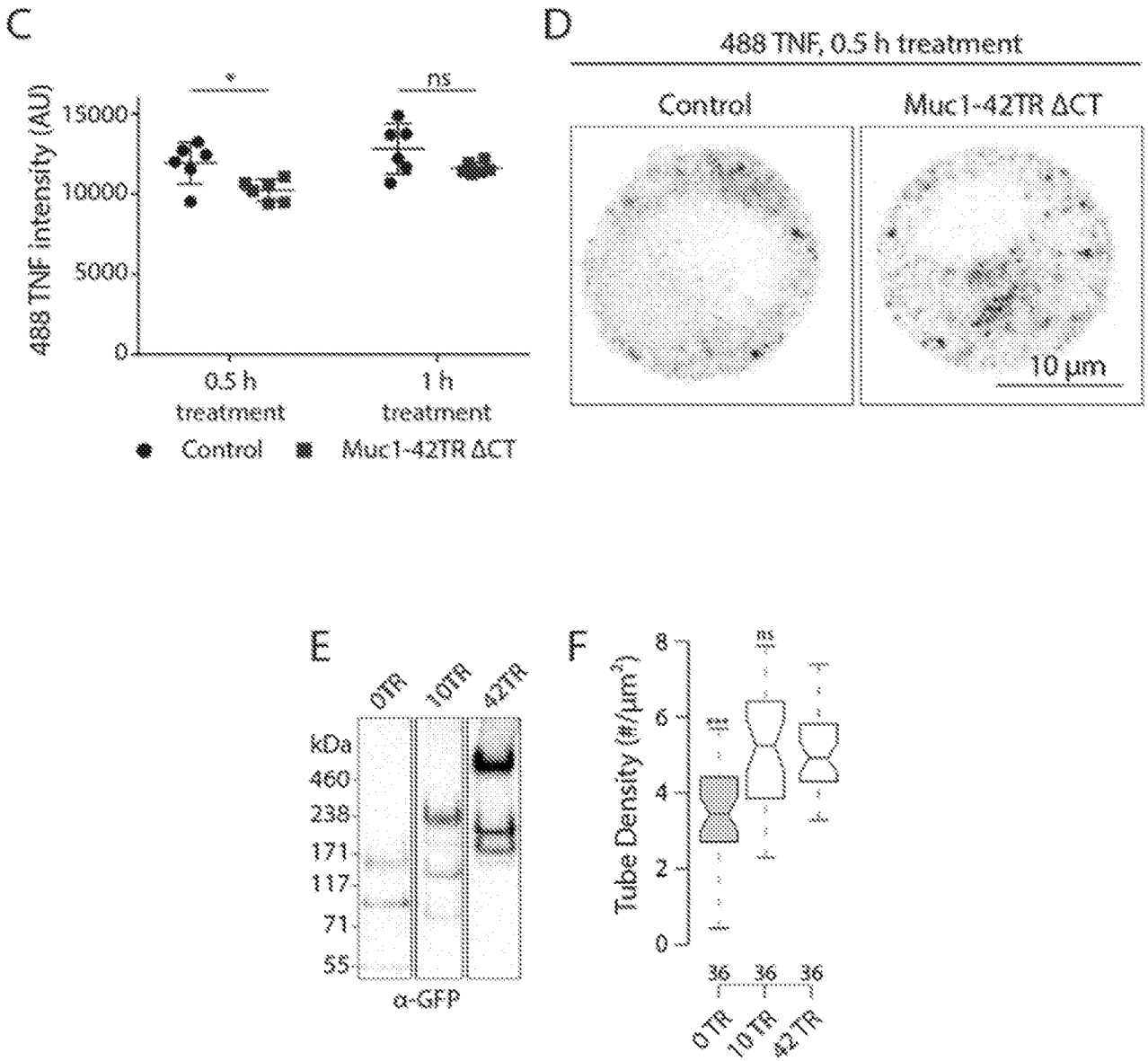


Figure 38 (continued)

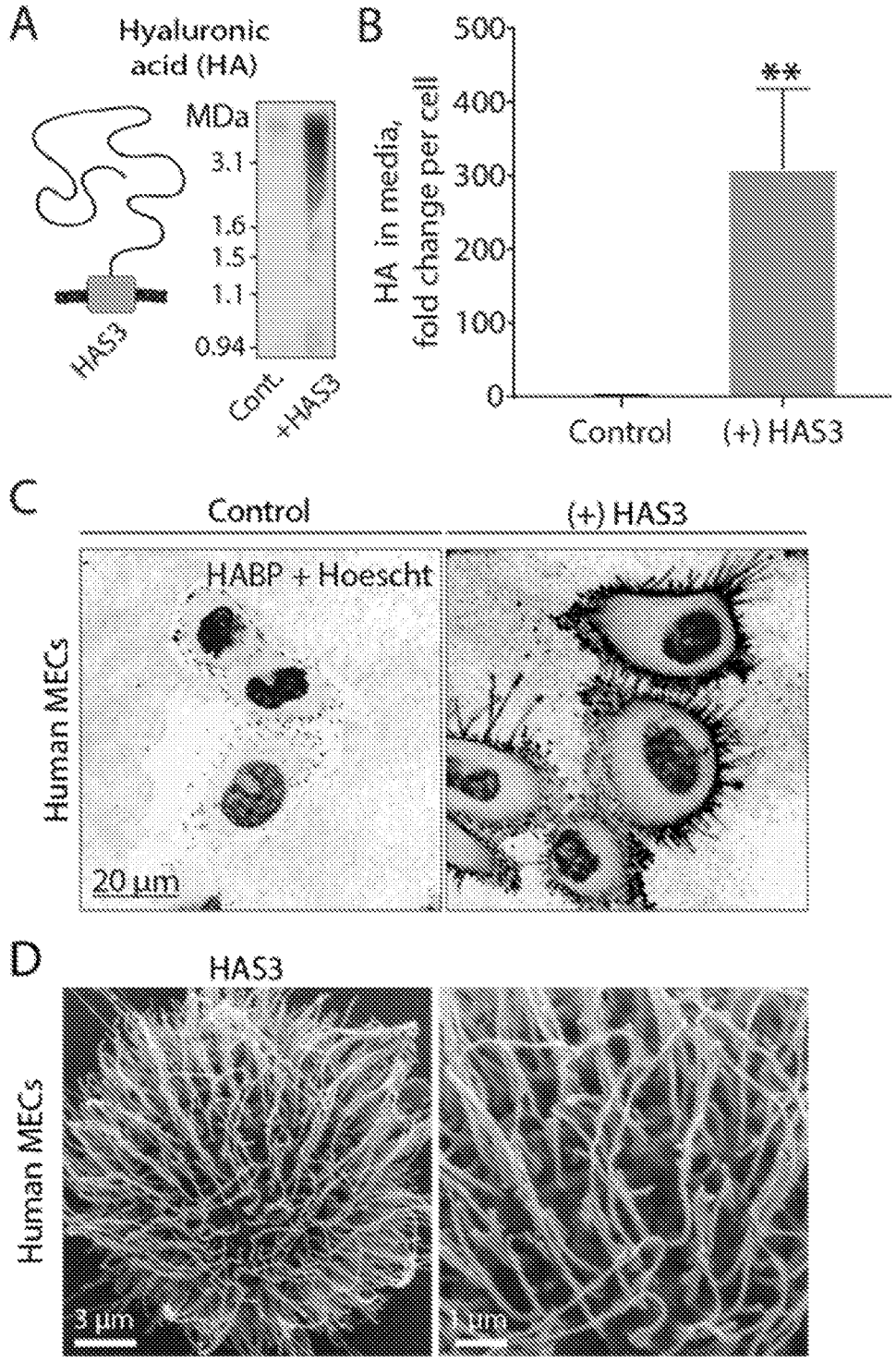

A
Hyaluronic acid (HA)
MDa
3.1
1.6
1.5
1.1
0.94
HAS3
Cont.
+HAS3
B
HA in media, fold change per cell
500
400
300
200
100
0
**
Control
(+) HAS3
C
Control
(+) HAS3
HABP + Hoescht
Human MECs
20 µm
D
HAS3
Human MECs
5 µm
5 µm

Figure 39

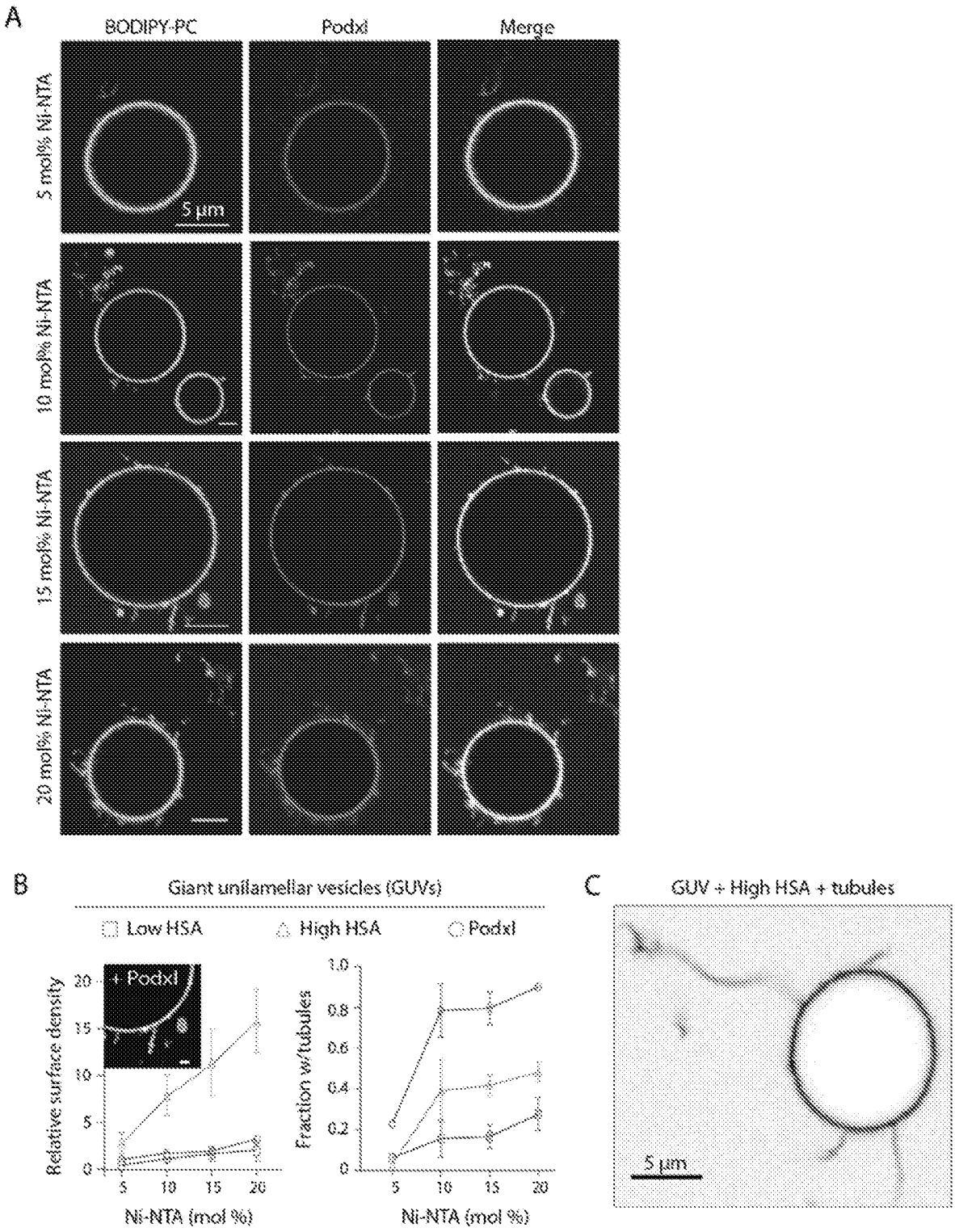

A
BODIPY-PC
Podxl
Merge
5 mol% Ni-NTA
10 mol% Ni-NTA
15 mol% Ni-NTA
20 mol% Ni-NTA
5 μm
B
Giant unilamellar vesicles (GUVs)
Low HSA
High HSA
Podxl
+ Podxl
Relative surface density
Fraction w/tubules
Ni-NTA (mol %)
Ni-NTA (mol %)
C
GUV + High HSA + tubules
5 μm

Figure 40

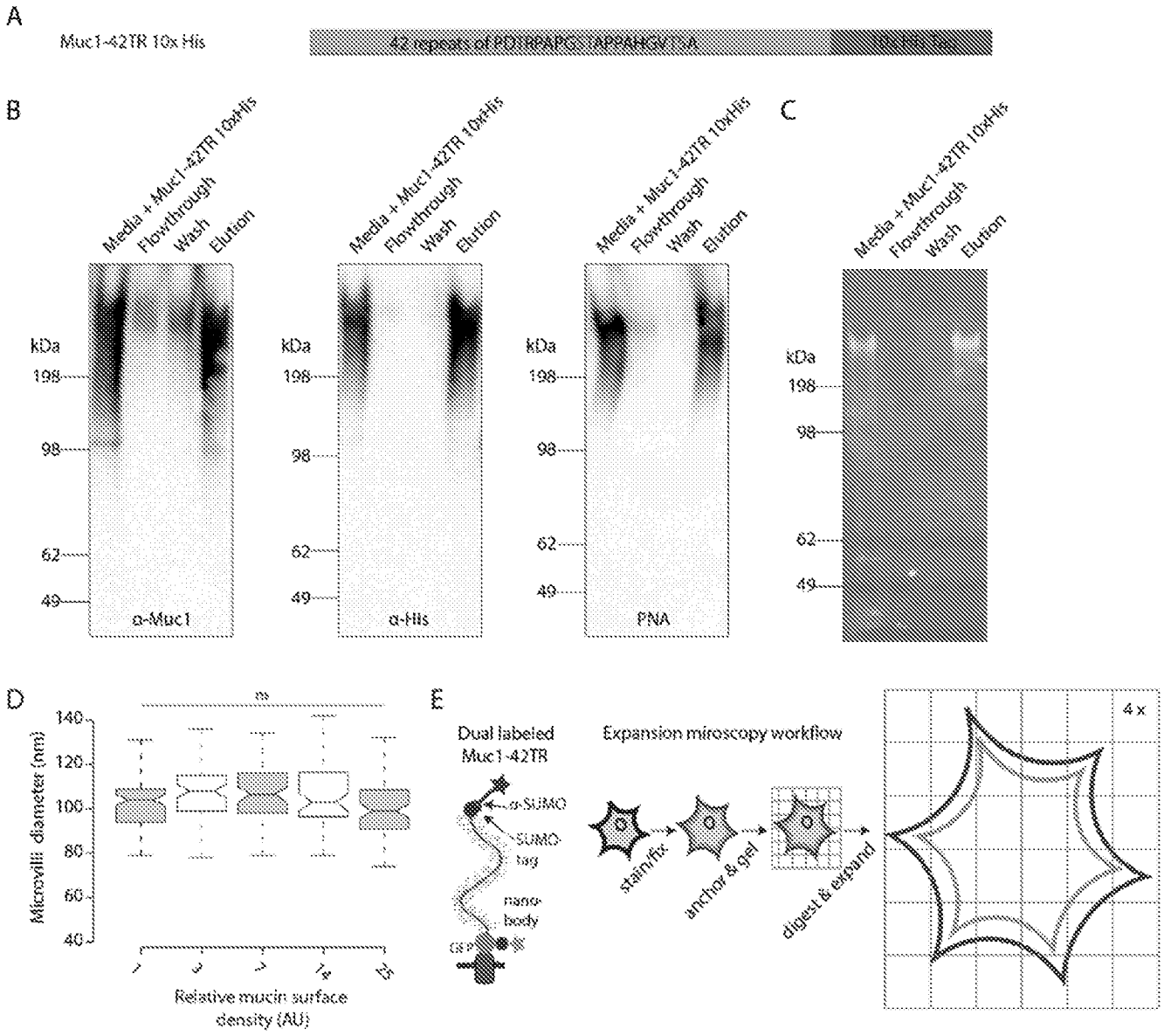

A
Muc1-42TR 10x His
B
Media + Muc1-42TR 10xHis
Flowthrough
Wash
Elution
kDa
198
98
62
49
α-Muc1
α-His
PNA
C
D
Microvilli diameter (nm)
140
120
100
80
60
40
Relative mucin surface density (AU)
E
Dual labeled Muc1-42TR
α-SUMO
SUMO-tag
nano-body
GFP
Expansion miroscopy workflow
stain/fix
anchor & gel
digest & expand
4 x

Figure 41

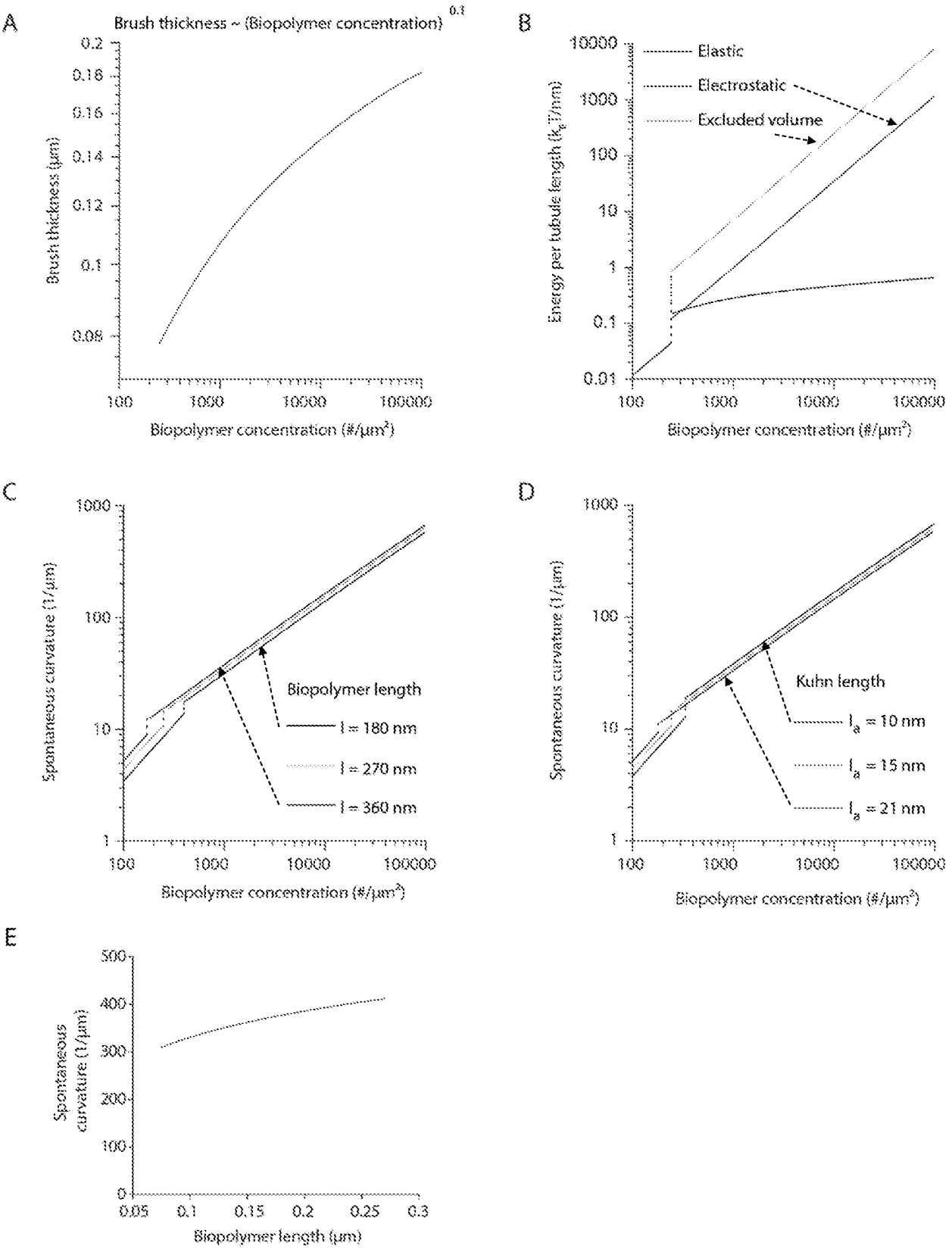
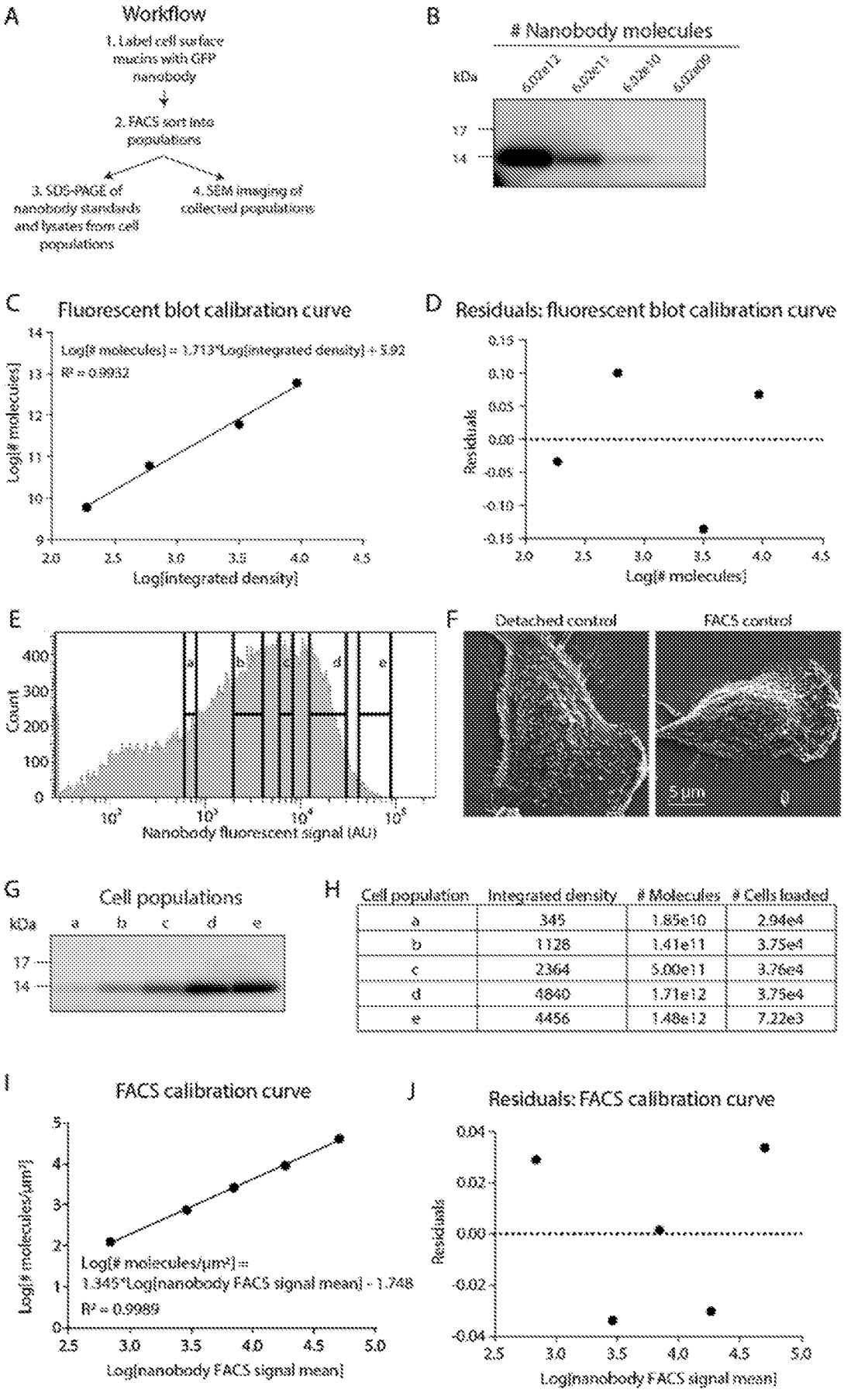

A
Brush thickness ~ (Biopolymer concentration)^0.1
Brush thickness (μm)
0.2
0.18
0.16
0.14
0.12
0.1
0.08
100
1000
10000
100000
Biopolymer concentration (#/μm²)
B
Energy per tubule length (k_BT/nm)
10000
1000
100
10
1
0.1
0.01
Elastic
Electrostatic
Excluded volume
100
1000
10000
100000
Biopolymer concentration (#/μm²)
C
Spontaneous curvature (1/μm)
1000
100
10
1
Biopolymer length
l = 180 nm
l = 270 nm
l = 360 nm
100
1000
10000
100000
Biopolymer concentration (#/μm²)
D
Spontaneous curvature (1/μm)
1000
100
10
1
Kuhn length
l_a = 10 nm
l_a = 15 nm
l_a = 21 nm
100
1000
10000
100000
Biopolymer concentration (#/μm²)
E
Spontaneous curvature (1/μm)
500
400
300
200
100
0
0.05
0.1
0.15
0.2
0.25
0.3
Biopolymer length (μm)

Figure 42



Cell population	Integrated density	# Molecules	# Cells loaded
a	345	1.85e10	2.94e4
b	1128	1.41e11	3.75e4
c	2364	5.00e11	3.76e4
d	4840	1.71e12	3.75e4
e	4456	1.48e12	7.22e3

Figure 43

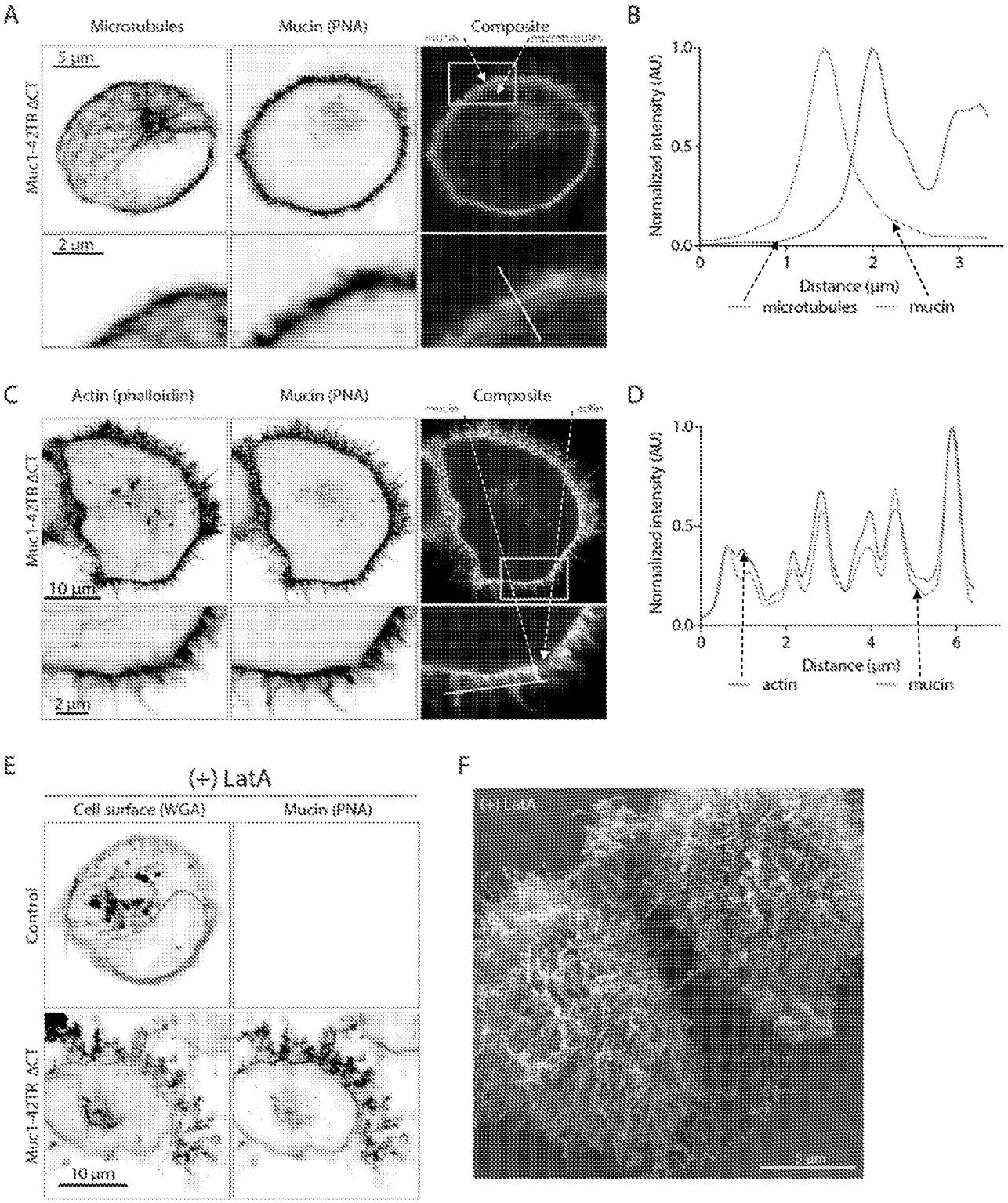

A
Microtubules
Mucin (PNA)
Composite
Muc1-42TR ΔCT
5 μm
2 μm
B
Normalized intensity (AU)
1.0
0.5
0.0
0
1
2
3
Distance (μm)
microtubules
mucin
C
Actin (phalloidin)
Mucin (PNA)
Composite
Muc1-42TR ΔCT
10 μm
2 μm
D
Normalized intensity (AU)
1.0
0.5
0.0
0
2
4
6
Distance (μm)
actin
mucin
E
(+) LatA
Cell surface (WGA)
Mucin (PNA)
Control
Muc1-42TR ΔCT
10 μm
F
(+) LatA
5 μm

Figure 44

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/013752

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A61K 31/715; A61K 38/00; A61K 38/17; C07K 14/47; C07K 14/78 (2020.01)
CPC - A61K 31/715; A61K 38/17; A61K 38/1709; A61K 47/36; C07K 14/4725 (2020.05)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 514/8; 514/12; 514/1.1; 514/16.7; 514/17.1; 514/20.9; 530/350 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015/0010606 A1 (ARTHREX, INC.) 08 January 2015 (08.01.2015) entire document	1-6, 18-44
A	US 2010/0074935 A1 (FLANNERY et al) 25 March 2010 (25.03.2010) entire document	1-6, 18-44
A	US 2018/0353571 A1 (LUBRIS LLC) 13 December 2018 (13.12.2018) entire document	1-6, 18-44
A	WO 2016/187508 A2 (THE BROAD INSTITUTE INC. et al) 24 November 2016 (24.11.2016) entire document	1-6, 18-44
A	US 2015/0343019 A1 (WYETH LLC) 03 December 2015 (03.12.2015) entire document	1-6, 18-44
A	US 2013/0196930 A1 (WYETH LLC) 01 August 2013 (01.08.2013) entire document	1-6, 18-44

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 May 2020	15 JUN 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/013752

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NOs: 1, 9, 10, 68, 72, and 74-82 were searched.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/013752

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
See extra sheet(s).

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-6 and 18-44 to the extent that they read on a recombinant lubricin polypeptide of SEQ ID NO:68.

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I+: claims 1-44 are drawn to recombinant lubricin polypeptides.

The first invention of Group I+ is restricted to a recombinant lubricin polypeptide, wherein the recombinant lubricin polypeptide is selected to be SEQ ID NO:68. It is believed that claims 1-6 and 18-44 read on this first named invention and thus these claims will be searched without fee to the extent that they read on a recombinant lubricin polypeptide of SEQ ID NO:68.

Applicant is invited to elect additional recombinant lubricin polypeptides, each with specified SEQ ID NO, to be searched in a specific combination by paying an additional fee for each set of election. An exemplary election would be a recombinant lubricin polypeptide, wherein the recombinant lubricin polypeptide is selected to be SEQ ID NO:70. Additional recombinant lubricin polypeptides will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The Groups I+ formulas do not share a significant structural element responsible for prophylaxis and/or therapy of disorders, requiring the selection of alternatives for the amino acid sequence of the recombinant lubricin polypeptides, where "...[the] recombinant lubricin polypeptide comprising a repeated sequence, the repeated sequence comprising or consisting of: KEPAPTTP (SEQ ID NO:1), KEPAPTP (SEQ ID NO: 9), KEPAPTTTP (SEQ ID NO:10), or a combination thereof".

Additionally, even if Groups I+ were considered to share the technical features of a recombinant lubricin polypeptide comprising a repeated sequence, the repeated sequence comprising or consisting of: KEPAPTTP (SEQ ID NO:1), KEPAPTP (SEQ ID NO: 9), KEPAPTTTP (SEQ ID NO:10), or a combination thereof; these shared technical features do not represent a contribution over the prior art.

Specifically, US 2013/0196930 A1 to Wyeth LLC discloses a recombinant lubricin polypeptide (recombinant lubricin molecules, Abstract) comprising a repeated sequence comprising KEPAPTTP (the amino acid sequence "EPAPTTTKSAPTTPKEPAPTTP" (SEQ ID NO:28; 22 amino acids) joins SEQ ID NO:26 to (N minus 2) repeats of SEQID NO:27 in preferred embodiments, Para. [0050]).

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.