

# Chemically tunable mucin chimeras assembled on living cells

Jessica R. Kramer<sup>a</sup>, Bibiana Onoa<sup>b</sup>, Carlos Bustamante<sup>b,c,d,e,f,g</sup>, and Carolyn R. Bertozzi<sup>a,h,1</sup>

<sup>a</sup>Department of Chemistry, Stanford University, Stanford, CA 94305; <sup>b</sup>Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720; <sup>c</sup>Department of Chemistry, University of California, Berkeley, CA 94720; <sup>d</sup>Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; <sup>e</sup>Department of Physics, University of California, Berkeley, CA 94720; <sup>f</sup>Kavli Energy Nanosciences Institute at Berkeley, University of California, Berkeley, CA 94720; <sup>g</sup>Howard Hughes Medical Institute, University of California, Berkeley, CA 94720; and <sup>h</sup>Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305

Edited by Barbara Imperiali, Massachusetts Institute of Technology, Cambridge, MA, and approved September 4, 2015 (received for review August 14, 2015)

**Mucins are a family of secreted and transmembrane glycoproteins characterized by a massive domain of dense O-glycosylation on serine and threonine residues. Mucins are intimately involved in immunity and cancer, yet elucidation of the biological roles of their glycodomains has been complicated by their massive size, domain polymorphisms, and variable glycosylation patterns. Here we developed a synthetic route to a library of compositionally defined, high-molecular weight, dual end-functionalized mucin glycodomain constructs via *N*-carboxyanhydride polymerization. These glycopolypeptides are the first synthetic analogs to our knowledge to feature the native  $\alpha$ -GalNAc linkage to serine with molecular weights similar to native mucins, solving a nearly 50-year synthetic challenge. Physical characterization of the mimics revealed insights into the structure and properties of mucins. The synthetic glycodomains were end-functionalized with an optical probe and a tetrazine moiety, which allowed site-specific bioorthogonal conjugation to an engineered membrane protein on live mammalian cells. This strategy in protein engineering will open avenues to explore the biological roles of cell surface mucins.**

mucin | polypeptide | glycoprotein | protein engineering | *N*-carboxyanhydride

**M**ucins are a family of glycoproteins that play a central role in multicellular biology. They are widely expressed by epithelial cells that exist in exposed environments such as the stomach, intestinal tract, lungs, eyes, and other specialized organs (1). Misregulation of mucin production and glycosylation has provided important links between immunity and cancer. It was recently discovered that cell surface mucins can regulate the assembly of signaling complexes at cell–cell and cell–extracellular (ECM) junctions (2). Changes in the elaboration and patterning of mucin O-glycans are correlated with Wiskott–Aldrich syndrome, hematological disorders, and cancer (3). For example, the biomarker mucin 1 (MUC1) is overexpressed and aberrantly glycosylated in >90% of epithelial cancers and is a highly attractive target for anticancer therapeutics (4, 5).

Both secreted and membrane-bound mucin glycoproteins exhibit genetic heterogeneity in their glycodomain length and amino acid sequence, yet all are characterized by dense O-glycosylation (6). The glycodomains are rich in serine (Ser) and threonine (Thr) residues that are initially glycosylated with  $\alpha$ -*N*-acetylgalactosamine ( $\alpha$ -GalNAc) by polypeptide  $\alpha$ -GalNAc transferases (Fig. 1A) (3). The  $\alpha$ -GalNAc residues can be further elaborated by downstream Golgi-resident glycosyltransferases to generate mucins with sugar components often comprising greater than 50% of their molecular weight.

Studies of mucin glycodomain biology have been complicated by their sheer size (150–300 kDa), varied glycosylation patterns, and polymorphic expression in domain length (7). To complicate matters, transmembrane mucins such as MUC1 contain cytoplasmic domains with diverse roles in signaling and gene regulation (8). The biology of MUC1's cytosolic tail has been probed independently by expressing a genetically truncated construct (9); however, the role of the extracellular glycodomain has been

difficult to probe. With current biological techniques, it has not been possible to deconvolute the glycosylation-dependent properties from the biochemical actions of the cytoplasmic region.

Here we describe a rapid and scalable route to fully synthetic mucin-type glycodomains with readily tunable glycan densities and chain lengths (Fig. 1B). Using this method, we prepared a library of glycopolypeptides with varied physical size, glycan density, and peptide backbone charge and hydrophobicity. Studies on the physical properties of these glycopolypeptides revealed a uniquely rigid and extended peptide structure, which may help explain the effects of changes in mucin expression and glycosylation with disease. We also investigated a method to remodel the cell surface via covalent attachment of the mucin mimics to an engineered membrane protein. These structurally authentic mucin constructs hold great potential for elucidating mucin biology.

## Results

**Design and Synthesis of Mucin Glycodomains.** In previous work, the synthesis of mucin-type glycopeptides has generally been accomplished by solid-phase peptide synthesis and native chemical or expressed protein ligation methods using glycosylated amino acid building blocks (3). These methods have proved particularly useful for recapitulating mucin structure and function in the context of eliciting an immune response for vaccine development (10). However, the low efficiency and multistep nature of these techniques has restricted peptide size to ca. 50–150 residues and limited the ability to quickly modify structure and synthesize peptides on scale (3). Nishimura and coworkers used step growth polymerization of glycosylated tripeptide monomers to produce glycopeptides with homogeneous repeating sequences (11).

## Significance

**Mucins are a diverse and heterogeneous family of glycoproteins that have been implicated in immunity and cancer. This work establishes a rapid and scalable route to synthetic mucin constructs with precisely defined glycan densities and chain lengths. Physical characterization indicates that the constructs both chemically and structurally emulate biological mucins and that dense mucin-type O-glycosylation imparts a remarkable degree of rigidity to the peptide backbone. Dual end-functionalized mucins were covalently conjugated to an engineered membrane protein on live mammalian cells. This strategy allows systematic variation in display of cell surface mucins with numerous applications in understanding the many biological roles of this unusual glycoprotein family.**

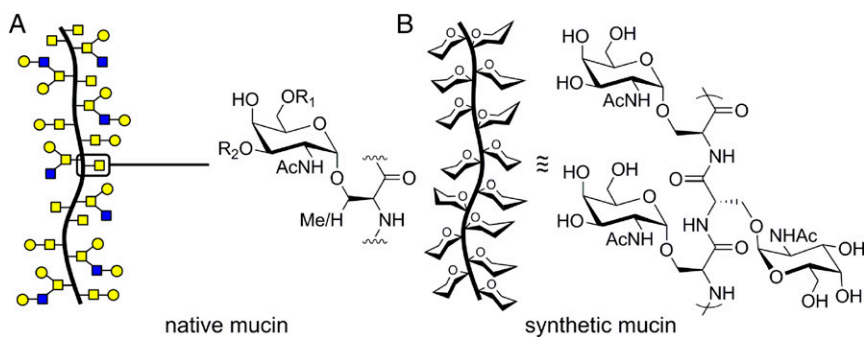
Author contributions: J.R.K. and C.R.B. designed research; J.R.K. and B.O. performed research; J.R.K. and C.B. contributed new reagents/analytic tools; J.R.K. and B.O. analyzed data; and J.R.K. and C.R.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. Email: bertozzi@stanford.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516127112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516127112/-DCSupplemental).



**Fig. 1.** (A) Native mucins are rich in clustered O-glycans that initiate with  $\alpha$ -GalNAc residues attached to Ser or Thr residues that can be further elaborated with glycans ( $R_1$ ,  $R_2$ ). (B) Cartoon representation and repeat unit of our synthetic mucins.

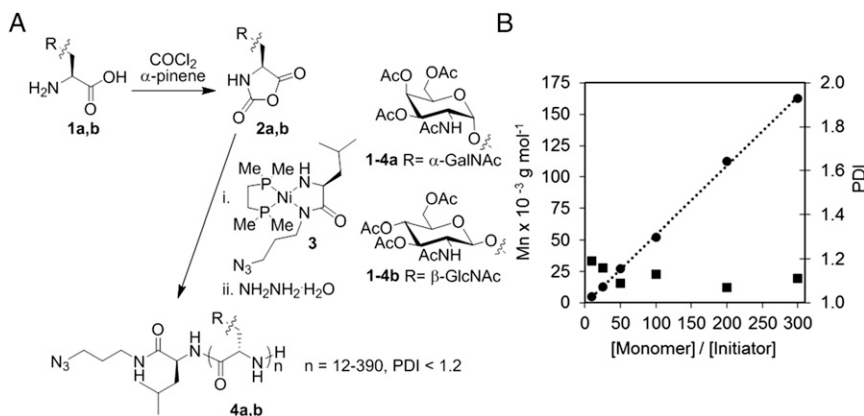
However, this technique yielded highly polydisperse [polydispersity index (PDI) 1.6–1.8, a measure of the heterogeneity of sizes of molecules in a mixture] oligomeric products of only ca. 30 residues in length, an order of magnitude smaller than native mucin glycodomains.

To more efficiently generate mucin mimics with high molecular weights and glycan densities, many investigators have turned to glycan-bearing nonpeptide synthetic polymers, which is the subject of recent reviews (12, 13). Although promising for certain applications, nonpeptide glycopolymers are unlikely to mimic the physical properties of bona fide mucin proteins. Indeed, our own recent work indicated that the bulk physical properties of mucin glycodomains is an important feature in itself and plays a role in cancer progression (2).

We envisioned developing a new route to synthetic glycopolymers in which we could access the native peptide backbone while simultaneously achieving low PDIs, high molecular weights, and exquisite control over glycosylation density. To this end, we investigated the utility of  $\alpha$ -amino acid *N*-carboxyanhydride (NCA) polymerization. The polymerization of NCAs to give polypeptides was first reported over 100 years ago and has since been widely used for rapid, scalable polypeptide synthesis and investigated for production of protein mimics (14, 15). For several decades, researchers have attempted to polymerize glycosylated NCA monomers to yield biomimetic glycopolypeptides (16–18). However, past efforts found that glycosylated Ser/Thr NCAs either could not be polymerized (19) or gave only highly polydisperse oligomeric peptides (20, 21) not useful as mimics of high-molecular weight mucins. Further, previous efforts failed to use the native  $\alpha$ -GalNAc-Ser/Thr linkage known to impart important structural features to the mucin peptide backbone. Native mucin  $\alpha$ -GalNAc-ylation of Ser/Thr is thought to extend and rigidify the peptide backbone, whereas  $\beta$ -GalNAc-ylation does not appear to have this effect (22). For studies on the physical properties of mucin glycodomains, use of the native linkage is essential.

Based on recent advances in NCA purification (23) and polymerization (24), we reexamined the glycosylated Ser NCA route.  $\alpha$ -GalNAc-Ser **1a** (25) and  $\beta$ -*N*-acetylglucosamine-Ser ( $\beta$ -GlcNAc-Ser) **1b** (26) were prepared according to literature routes and converted to the NCAs **2a,b** by treatment with phosgene via the Fuchs–Farthing method (27) (Fig. 2A).  $\beta$ -GlcNAc-Ser is a biologically relevant glycosidic  $\beta$ -linkage intended for structural comparison in this work (28). Typical NCA reaction conditions are addition of excess phosgene to the desired amino acid in THF or dioxane, followed by heating at 50 °C for several hours. We found these conditions to result in unacceptably low yields of the pure NCAs, likely due to acidic deglycosylation and incomplete cyclization. We found that a longer reaction time and lower temperature (36 h at 20 °C), plus addition of  $\alpha$ -pinene as an acid scavenger, increased yield of NCA **2a** to an average of 75%. NCA **2b** was isolated with an average yield of 50%. The lower yield is likely due to the poorer solubility of **2b** in organic solvents which likely slowed cyclization during NCA formation. It is important to note our final yields are for highly pure NCAs obtained by anhydrous chromatography, a method recently developed to purify NCAs that are low melting or otherwise difficult to crystallize (23). Previously reported glyco-Ser NCAs were purified only by precipitation (16–20) and likely contained impurities that undermined the polymerization process (23).

We next sought a method to achieve dual end-functionalization of our glycopolypeptides to simultaneously detect and functionalize biological targets. To this end, we designed azide-bearing nickel initiator **3** (Fig. 2A) to impart a clickable handle at the peptide chain initiation site, while the polypeptide terminus features a single reactive amine for functionalization via isothiocyanate chemistry. Polymerization initiator **3** is based on the work of Deming and coworkers, who described several transition metal initiators capable of controlled NCA polymerization, yielding high-molecular weight polypeptides with very low polydispersities (<1.2) (24, 29). Initiator **3** was prepared via treatment of 1,2-bis[(dimethylphosphino)ethane]Nickel(0)(cyclooctadiene) with



**Fig. 2.** (A) Preparation and polymerization of  $\alpha$ -GalNAc-Ser and  $\beta$ -GlcNAc-Ser NCAs and resulting glycopolypeptides. (B) Molecular weight ( $M_n$ ; ●) and polydispersity index (PDI; ■) of peracetylated poly( $\alpha$ -GalNAc-Ser) as functions of monomer to initiator ratio ( $[M]/[I]$ ) using nickel initiator **3** in DMF at 20 °C.

**Table 1. Molecular weight ( $M_n$ ) of peracetylated poly( $\alpha$ -GalNAc-Ser) as a function of  $\alpha$ -GalNAc-Ser NCA monomer to initiator **3** ( $[M]/[I]$ ) ratio in DMF at 20 °C**

$[M]/[I]^*$	$M_n^\dagger$	DP <sup>‡</sup>	PDI <sup>§</sup>	Yield, % <sup>¶</sup>
10	4,997	12	1.19	87
25	12,492	30	1.16	91
50	27,066	65	1.09	98
100	52,050	125	1.13	99
200	112,428	270	1.07	95
300	162,396	390	1.11	98

\*Number indicates equivalents of monomer per initiator.

<sup>†</sup>Molecular weight after polymerization as determined by <sup>1</sup>H NMR and GPC.

<sup>‡</sup>DP, number average degree of polymerization.

<sup>§</sup>Polydispersity index ( $M_w/M_n$ ) after polymerization as determined by GPC.

<sup>¶</sup>Total isolated yield of glycopolypeptide.

an azido alloc-L-leucine derivative according to a modification of Deming's procedure (*SI Appendix*) (29).

Polymerization of  $\alpha$ -GalNAc-Ser NCA **2a** was initiated by treatment with initiator **3** in DMF at room temperature. Reactions were typically complete within 2 hours even for high-molecular weight samples. By contrast, previous work with amine initiated polymerization of  $\beta$ -glyco-Ser NCAs gave only oligomeric products [degrees of polymerization (DP) 12–46] in 3–6 days (20). Variation of monomer to initiator ratios gave glycopolypeptides whose lengths increased linearly with stoichiometry and which possessed narrow chain length distributions (Fig. 2*B* and Table 1). Soluble homoglycopolypeptides could be prepared in near-quantitative conversion with DPs of nearly 400 residues, an order of magnitude longer than previously reported  $\beta$ -glucosyl polypeptides (20).  $\beta$ -GlcNAc-Ser NCA **2b** was also found to efficiently polymerize with initiator **3** and could yield high-molecular weight glycopolypeptides with similarly low PDIs (*SI Appendix*, Table S1).

Glycopolypeptides were readily deacetylated by treatment with hydrazine monohydrate in methanol to give fully water soluble poly( $\alpha$ -GalNAc-Ser) (**4a**) and poly( $\beta$ -GlcNAc-Ser) (**4b**). Glycosylation density could easily be controlled by stoichiometric addition of other amino acid NCAs. For initial trials, we chose to blend in L-Ala NCA, or protected L-Glu or L-Lys NCAs, to display functionalities found in native mucin proteins (see *SI Appendix*, Tables S2–S4, for copolymerization data) (30). Copolypeptide structures correlated well to monomer feed ratios, and the native  $\alpha$ -glycosidic linkage was found to be stable to the acidic and basic deprotection conditions required to remove the *t*Bu and TFA groups from Glu and Lys, respectively (according to <sup>1</sup>H NMR; *SI Appendix*, Tables S2–S4). Study of the polymerization kinetics of  $\alpha$ -GalNAc-Ser NCA compared with TFA-Lys NCA revealed the rate of peptide formation is essentially identical ( $t_{1/2} = 42.7$  min vs.  $t_{1/2} = 40.1$  min; *SI Appendix*, Fig. S1), indicating a statistical distribution of residues in the copolypeptides.

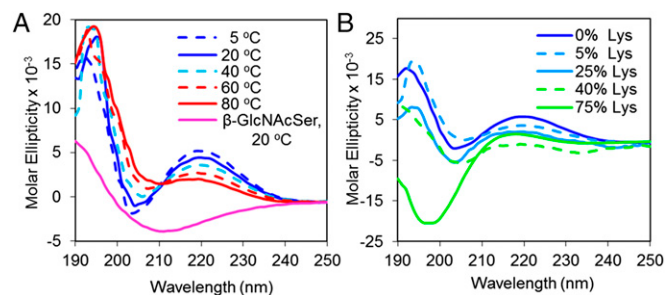
**Physical Properties of Synthetic Mucin Glycodomains.** Mucin-type glycosylation is thought to rigidify and extend the peptide backbone; however, structural studies have focused on short synthetic peptide oligomers (22, 31) or poorly defined and heterogeneous mucins extracted from biological sources (*vide infra*). To better understand the structural implications of dense O-glycosylation in mucins, we investigated the secondary structure of our mucin domain constructs by infrared spectroscopy (IR), circular dichroism spectroscopy (CD), transmission electron microscopy (TEM), and atomic force microscopy (AFM). Examination of poly( $\alpha$ -GalNAc-Ser) [DP = 300, poly( $\alpha$ -GalNAc-Ser)<sub>300</sub>, PDI = 1.11] by attenuated total reflectance FTIR did not indicate  $\beta$ -sheet structures because the amide I and II peak frequencies were observed at 1,651 and 1,547  $\text{cm}^{-1}$ , respectively, rather than ca. 1,630 and 1,530  $\text{cm}^{-1}$  (*SI Appendix*, Fig. S2) (32).  $\beta$ -sheets are typical of poly(Ser)s and have even been observed for poly(Ser)

functionalized with bulky, hydrophilic polyethylene glycol units (33). The greater steric bulk of the hydrated glycan and hydrogen bonds to the peptide backbone likely prevent sheet formation in these glycopolypeptides. The observed amide IR frequencies are typically assigned to helical or disordered protein structures and are similar to data observed with bovine submaxillary mucins (34) and mucins from human colonic adenocarcinomas (35).

To investigate the solution conformation of our synthetic mucins, CD spectra were measured in deionized water (Fig. 3*A*) from 5 to 80 °C. All CD spectra had positive maxima at ~220 nm, negative minima at 202 nm, and positive maxima at 194 nm. Similar data were observed in PBS buffer (*SI Appendix*, Fig. S3*B*). These data do not fit the patterns of canonical peptide structures in  $\alpha$ -helix or  $\beta$ -sheet conformations, and although similar to a disordered random coil, absorption intensities and positions differ (36). Ellipticities increased and decreased in magnitude with changes in temperature, a feature uncharacteristic of disordered peptides (37, 38). Similar spectra have been observed with short synthetic glycopeptides with two (31) or three (22)  $\alpha$ -GalNAc-Ser/Thr residues, native mucin samples (39), and highly glycosylated antifreeze glycoproteins featuring  $\alpha$ -GalNAc-Ser linkages (40). These data have been proposed to indicate highly organized and rigid backbone structures such as the threefold left-handed polyproline II-type helical structure, which was shown to be stabilized by glycosylation (31). Experimental studies on model compounds show that CD spectra of rigid molecules are generally more intense than those of flexible molecules (41), and theoretical studies of CD spectra as a function of peptide length confirm that the dichroism per residue is greater the longer the structure (42). In accordance with these studies, the CD spectra of a poly( $\alpha$ -GalNAc-Ser) 25-mer is moderately less intense than that of the 300-mer (*SI Appendix*, Fig. S3*A*). This peptide structure and associated CD signature appears particular to the native mucin  $\alpha$ -glycosidic linkage because we found that similar to previously reported data (22), poly( $\beta$ -GlcNAcSer)<sub>290</sub> has no positive maxima near 220 nm. These data indicate that use of the native  $\alpha$ -linkage in synthetic mucin mimics is essential for emulation of the biological structure.

Copolypeptides comprising 5–40% L-Lys showed decreasing ellipticities with increasing L-Lys content, indicating that spacing of the glycosylated residues mitigates their rigidifying effects and promotes chain relaxation (Fig. 3*B*; see *SI Appendix*, Table S3, for copolypeptide polymerization data). Only at 75% L-Lys was the conformation dominated by the L-Lys residues as indicated by the canonical CD spectrum of a random coil with a minor maximum at 218 and major minimum at 198. A similar trend was observed for the Glu and Ala copolypeptides (*SI Appendix*, Fig. S4). Based on these observations, it is interesting to speculate that mucin hyperglycosylation associated with disease may generate highly rigid, ordered, and extended peptide structures.

To directly visualize the conformation and morphology of our synthetic glycodomains, we imaged poly( $\alpha$ -GalNAc-Ser)<sub>300</sub> by

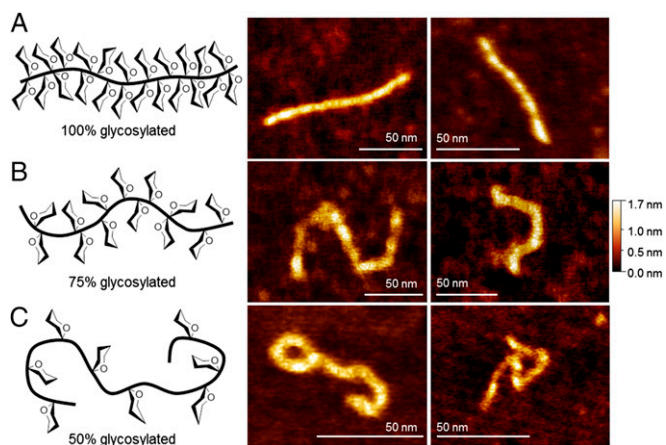


**Fig. 3.** Circular dichroism spectrum of (A) poly( $\alpha$ -GalNAc-Ser)<sub>300</sub> at various temperatures and (B) 0% Lys = poly( $\alpha$ -GalNAc-Ser)<sub>125</sub>, 5% Lys = poly( $\alpha$ -GalNAc-Ser)<sub>0.95</sub>/Lys<sub>0.05</sub>124, 25% Lys = poly( $\alpha$ -GalNAc-Ser)<sub>0.75</sub>/Lys<sub>0.25</sub>118, 40% Lys = poly( $\alpha$ -GalNAc-Ser)<sub>0.6</sub>/Lys<sub>0.4</sub>111, and 75% Lys = poly( $\alpha$ -GalNAc-Ser)<sub>0.25</sub>/Lys<sub>0.75</sub>117 at 20 °C. All spectra were recorded in Millipore water at 0.5 mg/mL.

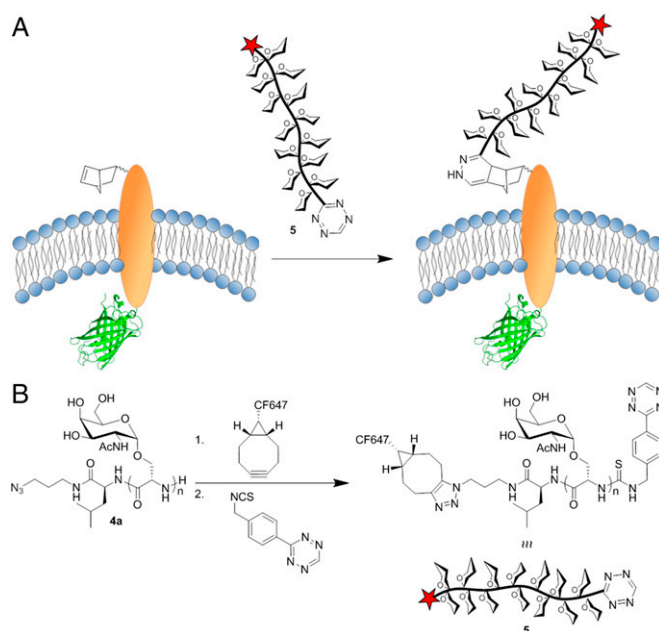
AFM and TEM. Films formed by deposition from a 10- $\mu$ M glycopolypeptide solution revealed spherical aggregates similar to those observed for mucins extracted from porcine stomach mucus (43) and human gastric mucins (44) (*SI Appendix*, Fig. S5). At 10–20 nM, we were able to directly image single glycopolypeptide molecules by AFM (Fig. 4A). Samples in aqueous solution were deposited on bare mica, dried, and imaged in air. Extended rod-like morphologies were observed throughout the sample, correlating well with AFM data reported for mucins extracted from biological sources (43–47). Poly( $\alpha$ -GalNAc-Ser)<sub>300</sub> had an average end-to-end length ( $R$ ) of 109 nm, contour length ( $L$ ) of 117 nm, and height of 1.1 nm ( $n = 12$ ) (see *SI Appendix*, Fig. S7, for example measurement). To confirm the observation by CD that spacing of glycosylated residues promotes chain relaxation, we examined 75% glycosylated poly( $\alpha$ -GalNAc-Ser<sub>0.75</sub>/Lys<sub>0.25</sub>)<sub>317</sub> and 50% glycosylated poly( $\alpha$ -GalNAc-Ser<sub>0.5</sub>/Lys<sub>0.5</sub>)<sub>325</sub> by AFM (Fig. 4B and C; see *SI Appendix*, Table S3, for polymerization data). A clear increase in molecular curvature was observed with decreasing glycosylation density, correlating well with the CD data and indicating that peptide flexibility can be readily tuned with our technique. Average molecular height and  $L$  values were similar between the three samples; however,  $R$  values differed significantly. We calculated an average  $R = 69$  and  $L = 127$  for the 75% glycosylated polypeptide ( $n = 9$ ) and  $R = 47$  and  $L = 131$  for the 50% glycosylated polypeptide ( $n = 7$ ).

AFM has been widely used to quantify the persistence length of various biomolecules including mucins. Persistence length (i.e., how far a polymeric chain persists in a given direction) is a proxy for molecular stiffness. The 100% glycosylated extended rods have  $R$  values similar to their  $L$  values, indicating that polypeptide adsorption on the mica was equilibrated and the worm-like chain model could be used to calculate the persistence length ( $P$ ) (see *SI Appendix* for discussion of calculations) (48). From the measured  $R$  and  $L$  values, we calculated  $P = 208$  nm for 100% glycosylated poly( $\alpha$ -GalNAc-Ser), a sizable value indicating that dense  $\alpha$ -GalNAc-ylation imparts a remarkable inflexibility to the peptide backbone. Spacing of the glycans resulted in  $P = 24.6$  nm for 75% glycosylated poly( $\alpha$ -GalNAc-Ser<sub>0.75</sub>/Lys<sub>0.25</sub>) and  $P = 6.1$  nm for 50% glycosylated poly( $\alpha$ -GalNAc-Ser<sub>0.5</sub>/Lys<sub>0.5</sub>). Only equilibrated molecules fully contained in the images and with clearly defined chain ends were included in the calculations.

**Cell Surface Display of Synthetic Mucins via an Engineered Membrane Protein.** Using the pyrrolysine/Amber stop codon method, many groups have reported the site-specific incorporation of unnatural amino acids (AAs) into mammalian proteins (49). We envisioned



**Fig. 4.** Cartoon representations and AFM images of synthetic mucins deposited on mica and imaged in air. (A) 100% glycosylated = poly( $\alpha$ -GalNAc-Ser)<sub>300</sub>, (B) 75% glycosylated = poly( $\alpha$ -GalNAc-Ser<sub>0.75</sub>/Lys<sub>0.25</sub>)<sub>317</sub>, and (C) 50% glycosylated = poly( $\alpha$ -GalNAc-Ser<sub>0.5</sub>/Lys<sub>0.5</sub>)<sub>325</sub>.

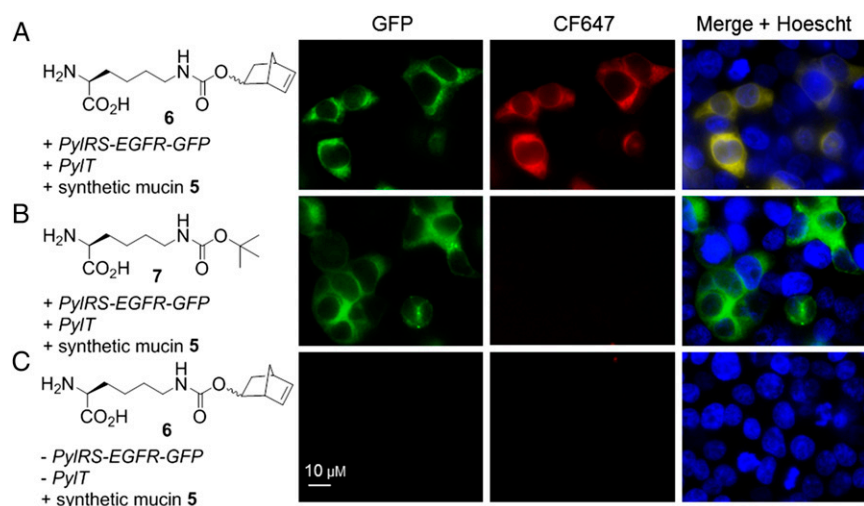


**Fig. 5.** (A) Remodeling of live mammalian cell surfaces via site-specific ligation of tetrazine-modified synthetic mucin **5** to a norbornene-bearing engineered membrane protein. (B) Synthetic procedure for modification of dual end-functionalized poly( $\alpha$ -GalNAc-Ser)<sub>n</sub> with a CF647 optical probe for visualization and a tetrazine moiety for covalent attachment to norbornene.

using this system to covalently attach our chemically precise synthetic glycodomains to a membrane protein anchor, an approach that would allow defined remodeling of cell surface mucins (Fig. 5A). In mammalian cells, Chin and coworkers expressed a membrane-resident epidermal growth factor receptor (EGFR)-green fluorescent protein (GFP) fusion construct featuring an extracellular norbornene-AA (50, 51). The cell surface norbornene moiety was labeled with a tetrazine probe based on colocalization of GFP with the probe. Their construct is an ideal model for our protein chimera concept because the synthetic mucins could be attached site-specifically via a rapid and bioorthogonal tetrazine ligation (52), ideal for reaction of two macromolecules of low solution concentration.

For preparation of protein chimeras, dual end-functional poly( $\alpha$ -GalNAc-Ser)<sub>300</sub> was modified with tetrazine and CF647 moieties as shown in Fig. 5B. Strain promoted azide-alkyne cycloaddition reaction of bicyclo[6,1,0]non-4-yn-9-yl-CF647 (photostable Cy5 equivalent, purchased from Biotium) with the azide group, followed by reaction of a tetrazine isothiocyanate with the peptide terminal amine gave the fluorophore and tetrazine modified mucin, CF647-poly( $\alpha$ -GalNAc-Ser)<sub>300</sub>-Tz **5**. According to the methods reported by Chin and coworkers, HEK 293T cells were simultaneously transfected with vectors *p4CMVE-U6-PylT* (*PylT*), which encodes four copies of the *Methanosarcina mazzellii* (*Mm*) aminoacyl-tRNA *MmPyltRNA*<sub>CUA</sub>, and *pMmPylRS-EGFR(128TAG)-GFP* (*PylRS-EGFR-GFP*), which encodes both the *Mm* pyrrolysyl-tRNA synthetase gene and the EGFR-GFP fusion gene containing the amber stop codon.

Transfections in the presence of 1 mM norbornene-AA **6**, or control *N* $\epsilon$ -Boc-L-lysine **7**, produced full-length EGFR-GFP as determined through visualization of GFP by fluorescence microscopy (Fig. 6A and B, *Left*). No GFP was observed in an experiment with **6** but with no plasmid DNA added (Fig. 6C, *Left*). Next, cells were treated with **5** (5  $\mu$ M, 4 h, 37  $^{\circ}$ C) then imaged again. Colocalization of synthetic mucin **5** (red fluorescence) and GFP (green fluorescence) was observed only for cells displaying EGFR-**6**-GFP (Fig. 6A, *Middle and Right*). No labeling with **5** was observed for cells displaying EGFR-**7**-GFP (Fig. 6B, *Middle and Right*), indicating the tetrazine ligation was indeed specific for the



**Fig. 6.** Remodeling of live mammalian cell surfaces by site-specific preparation of synthetic mucin-membrane protein chimeras. HEK 293T cells were incubated with norbornene-AA 6 (A and C) or control Boc-AA 7 (B), transfected with plasmids *PyIT* and *pMmPyIRS-EGFR(128TAG)-GFP* (A and B) or with no added plasmid DNA (C), and then treated with synthetic mucin 5 (A–C) [5  $\mu$ M CF647-poly( $\alpha$ -GalNAc-Ser)<sub>300</sub>-Tz, 4 h]. (Scale bar, 10  $\mu$ M.)

norbornene partner. Cells subjected to AA 6 but no plasmid DNA displayed no labeling with 5 under the same conditions (Fig. 6C, *Middle* and *Right*), confirming that our glycodomains are not endocytosed or nonspecifically adhered. These observations demonstrate that our synthetic mucin was site-specifically and covalently conjugated to the cell surface. Cell viability assays indicated that neither the synthetic mucin alone nor the conjugation reaction cause any appreciable cell death (*SI Appendix*, Fig. S8).

## Discussion

We have synthesized mucin glycodomain constructs that both chemically and structurally emulate biological mucins. Further, our ability to prepare and polymerize  $\alpha$ -GalNAc-Ser NCAs solves a nearly 50-year synthetic challenge. Contrary to previous reports, our results indicate that hydrogen bonding and steric hindrance are not limitations in polymerization of these monomers, particularly remarkable results considering that the native  $\alpha$ -linkage limits conformational rotation of the peptide backbone. Our route is rapid, scalable, and gives precise control over glycan density and chain length. Further, our azide-bearing initiator allows glycodomain dual end-functionalization for simultaneous detection and reaction with biological targets. Importantly, our poly( $\alpha$ -GalNAc-Ser) glycodomains may serve as substrates for enzymatic elaboration of more complex glycans (53, 54).

Physical characterization of poly( $\alpha$ -GalNAc-Ser)s by IR, CD, TEM, and AFM indicate that our materials are authentic surrogates for native mucin glycodomains and may shed light on the properties of aberrantly glycosylated cancer-associated mucins for which few physical data have been reported. The CD spectra of poly( $\alpha$ -GalNAc-Ser), and copolypeptides with L-Lys, L-Glu, or L-Ala, are particularly interesting as they indicate that hyper- $\alpha$ -GalNAc-ylation results in a highly organized and rigid backbone structure. This proposal is supported by direct observation via AFM of rod-like glycopolypeptide molecules that relax with spacing of glycosylated residues. From the AFM images, we calculated a persistence length of 208 nm for 100% glycosylated peptides and 24.6 and 6.1 nm for 75% and 50% glycosylated peptides, respectively. These data indicate that the stiffness of our synthetic mucins can easily be tuned by varying the amino acid composition and glycan density to allow precise modulation

of peptide physical properties. Persistence lengths measured by AFM for mucins extracted from healthy human ocular tissues range from 8.9 to 36 nm, apparently influenced by the degree and nature of glycosylation (46, 47). Relevant to our data, the persistence length of malignancy-associated hyperglycosylated mucins is expected to be much higher than that of mucins derived from healthy tissue (39, 44). This may have consequences related to altered focal adhesion formation and signaling (2).

As proof-of-concept in application of the synthetic mucins, we prepared glycopolypeptide–protein chimeras on live mammalian cells. Tetrazine end-modified glycopolypeptides were site-specifically attached to norbornene-bearing EGFR membrane proteins via the method of Chin and coworkers (50). This strategy could readily be adapted to replace the native glycodomain in mucin proteins with synthetic analogs whose chemical and structural properties can be systematically varied. Synthetic mucin glycodomains attached to the cell surface in this manner resemble transmembrane mucins, whereas free glycodomains are comparable to secreted mucins. We envision numerous applications toward studying how the unusual properties of mucins contribute to cell surface phenomena and other biological processes.

## Methods

Experimental procedures and characterization of all new compounds can be found in *SI Appendix*. Also included are details of NCA polymerization kinetics, copolypeptide polymerization data, additional circular dichroism spectroscopy, ATR-IR, TEM, persistence length calculations, and details of cell culture and cell viability assays.

**ACKNOWLEDGMENTS.** We thank Christina Woo, Chelsea Gordon, and Elliot Woods for a critical reading of the manuscript, and Piere Rodriguez for helpful discussions regarding persistence length calculations. Circular dichroism spectroscopy was performed at the Molecular Foundry within the Lawrence Berkeley National Laboratory and was supported by the Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract DE-AC02-05CH11231. J.R.K. was supported by a National Institutes of Health Fellowship and a University of California Chancellor's Fellowship. This work was supported by National Institute of Health Grant R01 GM059907 (to C.R.B.).

- Brayman M, Thathiah A, Carson DD (2004) MUC1: A multifunctional cell surface component of reproductive tissue epithelia. *Reprod Biol Endocrinol* 2:4.
- Paszek MJ, et al. (2014) The cancer glycocalyx mechanically primes integrin-mediated growth and survival. *Nature* 511(7509):319–325.
- Hang HC, Bertozzi CR (2005) The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem* 13(17):5021–5034.
- Baldus SE, Engelmann K, Hanisch F-G (2004) MUC1 and the MUCs: A family of human mucins with impact in cancer biology. *Crit Rev Clin Lab Sci* 41(2):189–231.

- Kufe DW (2009) Mucins in cancer: Function, prognosis and therapy. *Nat Rev Cancer* 9(12):874–885.
- Hollingsworth MA, Swanson BJ (2004) Mucins in cancer: Protection and control of the cell surface. *Nat Rev Cancer* 4(1):45–60.
- Hanisch FG, Müller S (2000) MUC1: The polymorphic appearance of a human mucin. *Glycobiology* 10(5):439–449.
- Carson DD (2008) The cytoplasmic tail of MUC1: A very busy place. *Sci Signal* 1(27):pe35.
- Mahanta S, Fessler SP, Park J, Bamdad C (2008) A minimal fragment of MUC1 mediates growth of cancer cells. *PLoS One* 3(4):e2054.

10. Gaidzik N, Westerlind U, Kunz H (2013) The development of synthetic antitumour vaccines from mucin glycopeptide antigens. *Chem Soc Rev* 42(10):4421–4442.
11. Tsuda T, Nishimura S-I (1996) Synthesis of an antifreeze glycoprotein analogue: Efficient preparation of sequential glycopeptide polymers. *Chem Commun* (24): 2779–2780.
12. Kiessling LL, Grim JC (2013) Glycopolymer probes of signal transduction. *Chem Soc Rev* 42(10):4476–4491.
13. Miura Y (2012) Design and synthesis of well-defined glycopolymers for the control of biological functionalities. *Polym J* 44(7):679–689.
14. Kricheldorf HR (2006) Polypeptides and 100 years of chemistry of  $\alpha$ -amino acid N-carboxyanhydrides. *Angew Chem Int Ed Engl* 45(35):5752–5784.
15. Woodward RB, Schramm CH (1947) Synthesis of protein analogs. *J Am Chem Soc* 69(6):1551–1552.
16. Rude E, Westphal O, Hurwitz E, Fuchs S, Sela M (1966) Synthesis and antigenic properties of sugar-polypeptide conjugates. *Immunochemistry* 3(2):137–151.
17. Rude E, Meyer-Delius M (1968) Synthesis of the N-carboxy- $\alpha$ -amino acid anhydrides of several O-acetylated serine glycosides. *Carbohydr Res* 8(2):219–232.
18. Kramer JR, Deming TJ (2014) Recent advances in glycopolypeptide synthesis. *Polym Chem* 5(3):671–682.
19. Gibson MI, Hunt GJ, Cameron NR (2007) Improved synthesis of O-linked, and first synthesis of S-linked, carbohydrate functionalised N-carboxyanhydrides (glycoNCAs). *Org Biomol Chem* 5(17):2756–2757.
20. Aoi K, Tsutsumiuchi K, Okada M (1994) Glycopeptide synthesis by an  $\alpha$ -amino acid N-carboxyanhydride (NCA) method: Ring opening polymerization of a sugar-substituted NCA. *Macromolecules* 27(3):875–877.
21. Tsutsumiuchi K, Aoi K, Okada M (1997) Synthesis of polyoxazoline-(glyco)peptide block copolymers by ring-opening polymerization of (sugar-substituted)  $\alpha$ -amino acid N-carboxyanhydrides with polyoxazoline macroinitiators. *Macromolecules* 30(14): 4013–4017.
22. Coltart DM, et al. (2002) Principles of mucin architecture: Structural studies on synthetic glycopeptides bearing clustered mono-, di-, tri-, and hexasaccharide glyco-domains. *J Am Chem Soc* 124(33):9833–9844.
23. Kramer JR, Deming TJ (2010) General method for purification of  $\alpha$ -amino acid N-carboxyanhydrides using flash chromatography. *Biomacromolecules* 11(12):3668–3672.
24. Cheng J, Deming TJ (2012) Synthesis of polypeptides by ring-opening polymerization of  $\alpha$ -amino acid N-carboxyanhydrides. *Top Curr Chem* 310:1–26.
25. Winans KA, King DS, Rao VR, Bertozzi CR (1999) A chemically synthesized version of the insect antibacterial glycopeptide, diptericin, disrupts bacterial membrane integrity. *Biochemistry* 38(36):11700–11710.
26. Campo VL, Carvalho I, Allman S, Davis BG, Field RA (2007) Chemical and chemo-enzymatic synthesis of glycosyl-amino acids and glycopeptides related to *Trypanosoma cruzi* mucins. *Org Biomol Chem* 5(16):2645–2657.
27. Kricheldorf HR (1987) *Amino Acid-N-Carboxyanhydrides and Related Heterocycles* (Springer, Berlin), pp 11–16.
28. Hart GW, Akimoto Y (2009) *The O-GlcNAc Modification. Essentials of Glycobiology*, eds Varki A, Cummings RD, Esko JD (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), 2nd Ed, Chap 18, pp 263–289.
29. Curtin SA, Deming TJ (1999) Initiators for end-group functionalized polypeptides via tandem addition reactions. *J Am Chem Soc* 121(32):7427–7428.
30. Kinlough CL, et al. (2011) Core-glycosylated mucin-like repeats from MUC1 are an apical targeting signal. *J Biol Chem* 286(45):39072–39081.
31. Spencer DIR, et al. (1997) A polyproline II motif is determined in MUC1 mucin synthetic (glyco)-peptides using CD and NMR. *Spectroscopy of Biological Molecules: Modern Trends*, eds Carmona P, Navarro R, Hernanz A (Springer, Dordrecht, The Netherlands), pp 35–36.
32. Susi H, Byler DM (1983) Protein structure by Fourier transform infrared spectroscopy: Second derivative spectra. *Biochem Biophys Res Commun* 115(1):391–397.
33. Hwang J, Deming TJ (2001) Methylated mono- and di(ethylene glycol)-functionalized  $\beta$ -sheet forming polypeptides. *Biomacromolecules* 2(1):17–21.
34. Liu F, et al. (1988) Influence of water on the infrared spectra of mucin. *Mikrochim Acta* 94(1):357–359.
35. Travo A, et al. (2010) IR spectral imaging of secreted mucus: A promising new tool for the histopathological recognition of human colonic adenocarcinomas. *Histopathology* 56(7):921–931.
36. Kelly SM, Price NC (2000) The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci* 1(4):349–384.
37. Tiffany ML, Krimm S (1972) Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. *Biopolymers* 11(11):2309–2316.
38. Siligardi G, Drake AF (1995) The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides. *Biopolymers* 37(4):281–292.
39. Shogren R, Gerken TA, Jentoft N (1989) Role of glycosylation on the conformation and chain dimensions of O-linked glycoproteins: Light-scattering studies of ovine submaxillary mucin. *Biochemistry* 28(13):5525–5536.
40. Tachibana Y, et al. (2002) Efficient and versatile synthesis of mucin-like glycoprotein mimics. *Tetrahedron* 58(51):10213–10224.
41. Goodman M, Toniolo C, Falcetta J (1969) Conformational aspects of polypeptide structure. XXX. Rotatory properties of cyclic and bicyclic amides. Restricted and rigid model compounds for peptide chromophores. *J Am Chem Soc* 91(7):1816–1822.
42. Woody RW, Tinoco I (1967) Optical rotation of oriented helices. III. Calculation of the rotatory dispersion and circular dichroism of the alpha- and 310-helix. *J Chem Phys* 46(12):4927–4945.
43. Yakubov GE, Papagiannopoulos A, Rat E, Waigh TA (2007) Charge and interfacial behavior of short side-chain heavily glycosylated porcine stomach mucin. *Biomacromolecules* 8(12):3791–3799.
44. Hong Z, et al. (2005) Atomic force microscopy reveals aggregation of gastric mucin at low pH. *Biomacromolecules* 6(6):3458–3466.
45. McMaster TJ, Berry M, Corfield AP, Miles MJ (1999) Atomic force microscopy of the submolecular architecture of hydrated ocular mucins. *Biophys J* 77(1):533–541.
46. Round AN, et al. (2002) Heterogeneity and persistence length in human ocular mucins. *Biophys J* 83(3):1661–1670.
47. Round AN, Berry M, McMaster TJ, Corfield AP, Miles MJ (2004) Glycopolymer charge density determines conformation in human ocular mucin gene products: An atomic force microscope study. *J Struct Biol* 145(3):246–253.
48. Rivetti C, Guthold M, Bustamante C (1996) Scanning force microscopy of DNA deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J Mol Biol* 264(5):919–932.
49. Fekner T, Chan MK (2011) The pyrrolysine translational machinery as a genetic-code expansion tool. *Curr Opin Chem Biol* 15(3):387–391.
50. Lang K, et al. (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* 4(4):298–304.
51. Uttamapinant C, et al. (2015) Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins. *J Am Chem Soc* 137(14):4602–4605.
52. Debets MF, van Hest JCM, Rutjes FPJT (2013) Bioorthogonal labelling of biomolecules: New functional handles and ligation methods. *Org Biomol Chem* 11(38):6439–6455.
53. Leppanen A, Penttila L, Renkonen O, McEver RP, Cummings RD (2002) Glycosulfopeptides with O-glycans containing sialylated and polyfucosylated poly lactosamine bind with low affinity to P-selectin. *J Biol Chem* 277(42):39749–39759.
54. Matsushita T, et al. (2006) Construction of highly glycosylated mucin-type glycopeptides based on microwave-assisted solid-phase syntheses and enzymatic modifications. *J Org Chem* 71(8):3051–3063.