Title

The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins.

Journal

Proceedings of the National Academy of Sciences of USA, 116(15)

Authors

Malaker, Stacy

Pedram, Kayvon

Ferracane, Michael

et al.

Publication Date

2019-04-09

DOI

10.1073/pnas.1813020116

Peer reviewed

The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins

Stacy A. Malaker[a,1], Kayvon Pedram[a,1], Michael J. Ferracane[b], Barbara A. Bensing[c], Venkatesh Krishnan[d], Christian Pett[e,f], Jin Yu[e], Elliot C. Woods[a], Jessica R. Kramer[g], Ulrika Westerlind[e,f], Oliver Dorigo[d], and Carolyn R. Bertozzi[a,h,2]

[a]Department of Chemistry, Stanford University, Stanford, CA 94305; [b]Department of Chemistry, University of Redlands, Redlands, CA 92373; [c]Department of Medicine, San Francisco Veterans Affairs Medical Center and University of California, San Francisco, CA 94143; [d]Stanford Women's Cancer Center, Division of Gynecologic Oncology, Stanford University, Stanford, CA 94305; [e]Leibniz-Institut für Analytische Wissenschaften (ISAS), 44227 Dortmund, Germany; [f]Department of Chemistry, Umeå University, 901 87 Umeå, Sweden; [g]Department of Bioengineering, University of Utah, Salt Lake City, UT 84112; and [h]Howard Hughes Medical Institute, Stanford, CA 94305

Edited by Laura L. Kiessling, Massachusetts Institute of Technology, Cambridge, MA, and approved February 25, 2019 (received for review July 30, 2018)

Mucin domains are densely *O*-glycosylated modular protein domains that are found in a wide variety of cell surface and secreted proteins. Mucin-domain glycoproteins are known to be key players in a host of human diseases, especially cancer, wherein mucin expression and glycosylation patterns are altered. Mucin biology has been difficult to study at the molecular level, in part, because methods to manipulate and structurally characterize mucin domains are lacking. Here, we demonstrate that secreted protease of C1 esterase inhibitor (StcE), a bacterial protease from *Escherichia coli*, cleaves mucin domains by recognizing a discrete peptide- and glycan-based motif. We exploited StcE's unique properties to improve sequence coverage, glycosite mapping, and glycoform analysis of recombinant human mucins by mass spectrometry. We also found that StcE digests cancer-associated mucins from cultured cells and from ascites fluid derived from patients with ovarian cancer. Finally, using StcE, we discovered that sialic acid-binding Ig-type lectin-7 (Siglec-7), a glycoimmune checkpoint receptor, selectively binds sialomucins as biological ligands, whereas the related receptor Siglec-9 does not. Mucin-selective proteolysis, as exemplified by StcE, is therefore a powerful tool for the study of mucin domain structure and function.

O-glycosylation | mucin | protease | glycoproteomics | Siglec

Mucin domains are modular protein domains characterized by a high frequency of Ser and Thr residues that are *O*-glycosylated with α-*N*-acetylgalactosamine (α-GalNAc). The initial glycosylation with α-GalNAc and elaboration to larger glycan structures occur through complex, interdependent metabolic pathways to which hundreds of genes can contribute. The resulting glycoprotein population is highly heterogeneous and dynamic, and cannot be predicted based on genomic information alone (1). Dense spacing of these heterogeneous *O*-glycans forces the underlying polypeptide to adopt a unique "mucin fold," which is rigid and extended (2) (Fig. 1*A*, *Left*). The mucin fold is found in a wide variety of cell surface and secreted proteins of various families, analogous to the common Ig fold (3).

Mucin-domain glycoproteins contribute to many biological processes. For example, they are involved in embryogenesis (4), barrier formation (5), host–pathogen interactions (6), and immune signaling (7). Due to their stiff, elongated, and highly hydrated structures, mucin domains are also important modulators of cell- and protein-level biophysics (8). For example, CD45R0 is a conserved splice variant of the T cell marker CD45, in which three exons encoding mucin domains are missing. The loss of the mucin domains promotes dimerization, abrogating intracellular tyrosine phosphatase activity needed for T cell activation, and thereby aiding activated T cells in regulated cessation of their response (9). Mucin domains also serve as receptor ligands. For example, the binding of P-selectin to its ligand, P-selectin glycoprotein ligand-1 (PSGL-1), a mucin-domain glycoprotein, is an essential component of leukocyte rolling and adhesion (10).

When strung together in "tandem repeats" mucin domains can form the large structures characteristic of mucin family proteins. Mucins can be hundreds to thousands of amino acids long and >50% glycosylation by mass (11); MUC16, one of the largest mucins, can exceed 22,000 residues and 85% glycosylation by mass (12), with a persistence length of 1–5 μm (13). Due, in part, to their presence at cellular peripheries, mucin family proteins are important mediators of cell–cell and cell–environment interactions, and have been under scrutiny for years due to their association with human cancer (14) (which must combat inhibitory cues from its surroundings at every stage of its progression). Indeed, the membrane-associated mucin MUC1 is aberrantly expressed in ∼60% of all cancers diagnosed each year in the United States (15), rendering MUC1 one of the most prominently dysregulated genes in cancer. Functionally, recent work has shown that the MUC1 ectodomain alone can drive tumor growth by enhancing cancer cell survival and promoting proliferation through biophysical pathways (16). The large mucin MUC16 (also called CA-125) is highly expressed in ovarian cancer and is used as a clinical biomarker for treatment efficacy and surveillance (17). Observations such as these have motivated numerous efforts toward mucin-based

Significance

Mucin-domain glycoproteins are found in nearly every tissue of the human body, and are important in biological processes ranging from embryogenesis to cancer. Because there are few tools to study mucin domains, their biological functions at the molecular scale remain unclear. Here, we help address a hurdle to the study of mucin-domain glycoproteins by characterizing a bacterial protease with selectivity for mucins. This mucinase selectively removes native mucins from cell surfaces and cuts them into fragments amenable to analysis.

Author contributions: S.A.M., K.P., and C.R.B. designed research; S.A.M., K.P., M.J.F., and B.A.B. performed research; V.K., C.P., J.Y., J.R.K., U.W., and O.D. contributed new reagents/analytic tools; S.A.M., K.P., M.J.F., B.A.B., E.C.W., and C.R.B. analyzed data; and S.A.M., K.P., and C.R.B. wrote the paper.

Conflict of interest statement: A patent application relating to the use of enzymes to digest mucin-domain glycoproteins has been filed by Stanford University (docket no. STAN-1510PRV). C.R.B. is a cofounder and Scientific Advisory Board member of Palleon Pharmaceuticals, Enable Bioscience, Redwood Biosciences (a subsidiary of Catalent), and InterVenn Biosciences, and a member of the Board of Directors of Eli Lilly & Company. O.D. has participated in advisory boards for Tesaro, Merck, and Geneos. O.D. is a speaker for Tesaro and AstraZeneca.

This article is a PNAS Direct Submission.

[1]S.A.M. and K.P. contributed equally to this work.

[2]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.1813020116/-/DCSupplemental.

Published online March 25, 2019.

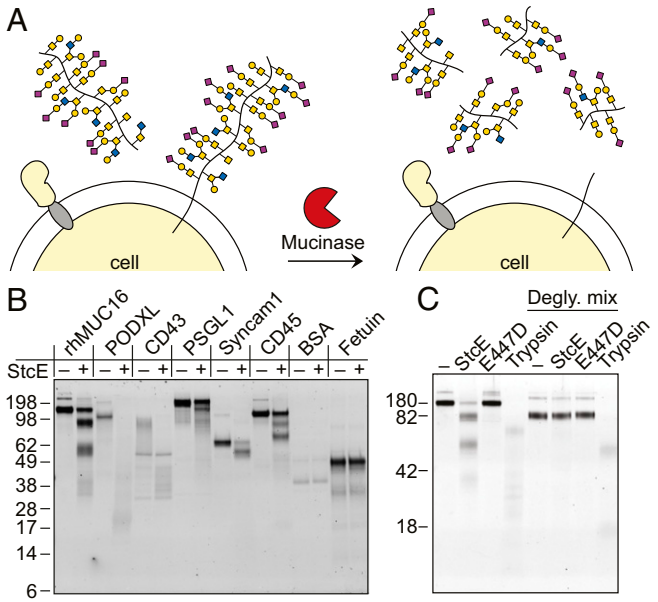

Fig. 1. StcE is a protease that selectively cleaves mucins. (*A*) A mucinase would enable mucin-domain glycoproteins to be selectively removed from biological samples and cut into fragments, facilitating their analysis. Cellular and tissue experiments are depicted in Figs. 5 and 6. (*B*) Recombinant glycoproteins were treated with StcE at a 1:10 E/S ratio for 3 h at 37 °C, and the digests were separated by SDS/PAGE. Glycoproteins and glycosylated peptide fragments were visualized with periodate-based Emerald 300 Glycoprotein Stain (ThermoFisher Scientific); silver stained gel images are shown in *SI Appendix*, Fig. S4*B*. BSA is not reported to be glycosylated. (*C*) Recombinant human MUC16 was digested with StcE, E447D, or trypsin with and without prior enzymatic deglycosylation [Deglycosylation (Degly.) Mix; Promega]. The digestion products were visualized as in *B*.

vaccines (18), small-molecule (19) and antibody (20) therapies, and, more recently, CAR-T cell therapies (21).

Despite all this attention, we know little about the molecular structures and biological activities of mucin-domain glycoproteins. This gap is due to unique challenges posed by their highly *O*-glycosylated structures. First, mucin domains are not amenable to standard sequence, glycosite, and glycoform mapping via mass spectrometry (MS). Unlike *N*-glycans, *O*-glycans lack a peptide consensus motif, are not predictable in structure, and cannot be quantitatively released in their native form via enzymatic treatment. Further, the high density of *O*-glycosylation on mucin domains makes them resistant to digestion by workhorse proteases such as trypsin, meaning the majority of their sequence space is often left unanalyzed (22). The issue of protease accessibility is compounded by poor ionizability of any densely glycosylated peptides that are generated (23). Challenges such as size, heavy glycosylation, and protease resistance have also rendered mucin-domain glycoproteins difficult to study by standard biochemical techniques. As selective depletion and isolation of mucin proteins have not been possible, a comprehensive list of mucin-domain glycoproteins does not exist (24), the biological roles of many cell surface associated mucins remain elusive (25), and signaling pathways relating to transmembrane and shed mucin structures are underexplored (7).

Systems with truncated forms of glycosylation, such as engineered "SimpleCells" lacking *O*-glycan elaboration machinery can simplify mucin analysis and glycosite identification, but functionally important glycan structures beyond the initiating *O*-GalNAc are lost (26). Recently, an "*O*-protease" was reported that cleaves N-terminally to glycosylated serine and threonine residues (27, 28). However, this enzyme requires substrate desialylation for optimal activity, and is not specific for mucins. *O*-sialoglycoprotein endopeptidase (OSGEP) is a commercial enzyme historically marketed as having mucin-degrading activity. Unfortunately, OSGEP's apparent activity has not been tied to a gene sequence that is amenable to recombinant expression, prompting the current view that OSGEP may act as a mixture of several unknown enzymes (29).

The ubiquitous and highly conserved nature of mucin structures implies an ecological need for enzymes with mucin-specific proteolytic activity. Indeed, recent evidence has pointed to the existence of families of mucin-targeting proteases comprising hundreds of enzymes largely found in organisms living in mucin-rich host environments (30–32). One such enzyme, secreted protease of C1 esterase inhibitor (StcE), is a zinc metalloprotease of human pathogenic enterohemorrhagic *Escherichia coli* (EHEC). Discovered by Welch and coworkers (33) and crystallized by Strynadka and coworkers (34), StcE promotes EHEC pathogenesis in humans by cleaving the protective mucus layers of the gut (35, 36). It is reported to cleave densely *O*-glycosylated proteins, but not *N*-glycosylated or sparsely *O*-glycosylated substrates (31). Given these reported pathogenic properties, we speculated that StcE could be transformed into a research tool to efficiently and selectively cleave human mucins. Mucin-specific proteolysis would be a valuable addition to the biochemist's toolbox, enabling mucin-domain glycoproteins to be selectively liberated from biological samples and cut into fragments amenable to analysis (Fig. 1*A*, *Right*).

Here, we report that StcE has a distinct peptide- and glycan-based cleavage motif that enables high selectivity for mucins. This "mucinase" improves sequence coverage, glycosite mapping, and glycoform analysis of recombinant mucins by MS, and enables specific release of mucins from cell lines and human tissue samples. In addition, we used StcE to provide evidence for the existence of professional mucin ligands of sialic acid-binding Ig-type lectin-7 (Siglec-7), an immune checkpoint receptor.

Results and Discussion

StcE Has Peptide- and Glycan-Based Selectivity for Mucins. We expressed StcE and its catalytically inactive point mutant (E447D) as 98-kDa soluble, N-terminal His-tagged proteins in *E. coli*, as previously described (34) (*SI Appendix*, Fig. S1). Previous reports suggested that StcE activity against a known substrate, C1 esterase inhibitor (C1INH), is detectable in a pH range of 6.1–9.0, in a temperature range of 4–55 °C, in high salt and detergent, and after days of incubation at 37 °C, consistent with its pathological activity in the mammalian gut (37). In our hands, StcE was amenable to high-yield expression (80 mg/L), active against C1INH (*SI Appendix*, Fig. S2), stable to lyophilization (*SI Appendix*, Fig. S3), and operative at nanomolar concentrations in all media types tested. We next assessed StcE's activity on clinically relevant mucin-domain glycoproteins. StcE did not cleave glycosylated but nonmucin proteins [fetuin, fibronectin, erythropoietin (EPO), transferrin, and α1-acid glycoprotein] (*SI Appendix*, Fig. S4*A*), but cleaved all tested mucin-domain glycoproteins (recombinant MUC16, podocalyxin, CD43, PSGL-1, Syncam-1, and CD45), as evidenced by gel shifts to lower molecular weights (Fig. 1*B* and *SI Appendix*, Fig. S4*B*). StcE's activity was abrogated when its substrates were enzymatically deglycosylated, indicating a glycan requirement for cleavage (Fig. 1*C*). At the same time, StcE cleavage of C1INH was not significantly influenced by desialylation of the substrate using various sialidases (*SI Appendix*, Fig. S5).

We next asked whether StcE has a preferred peptide sequence or glycan recognition motif. The recombinant mucin-domain glycoproteins listed above were digested with StcE, de–*N*-glycosylated with PNGaseF, trypsinized, and subjected to MS analysis using an optimized protocol (*SI Appendix*, Fig. S6). Given that the substrates were recombinantly expressed, the Byonic (ProteinMetrics) database search was limited to "*O*-glycan 6 most common" as variable modifications (Fig. 2*B*, *Inset*; *vide infra* for further discussion). Through manual validation

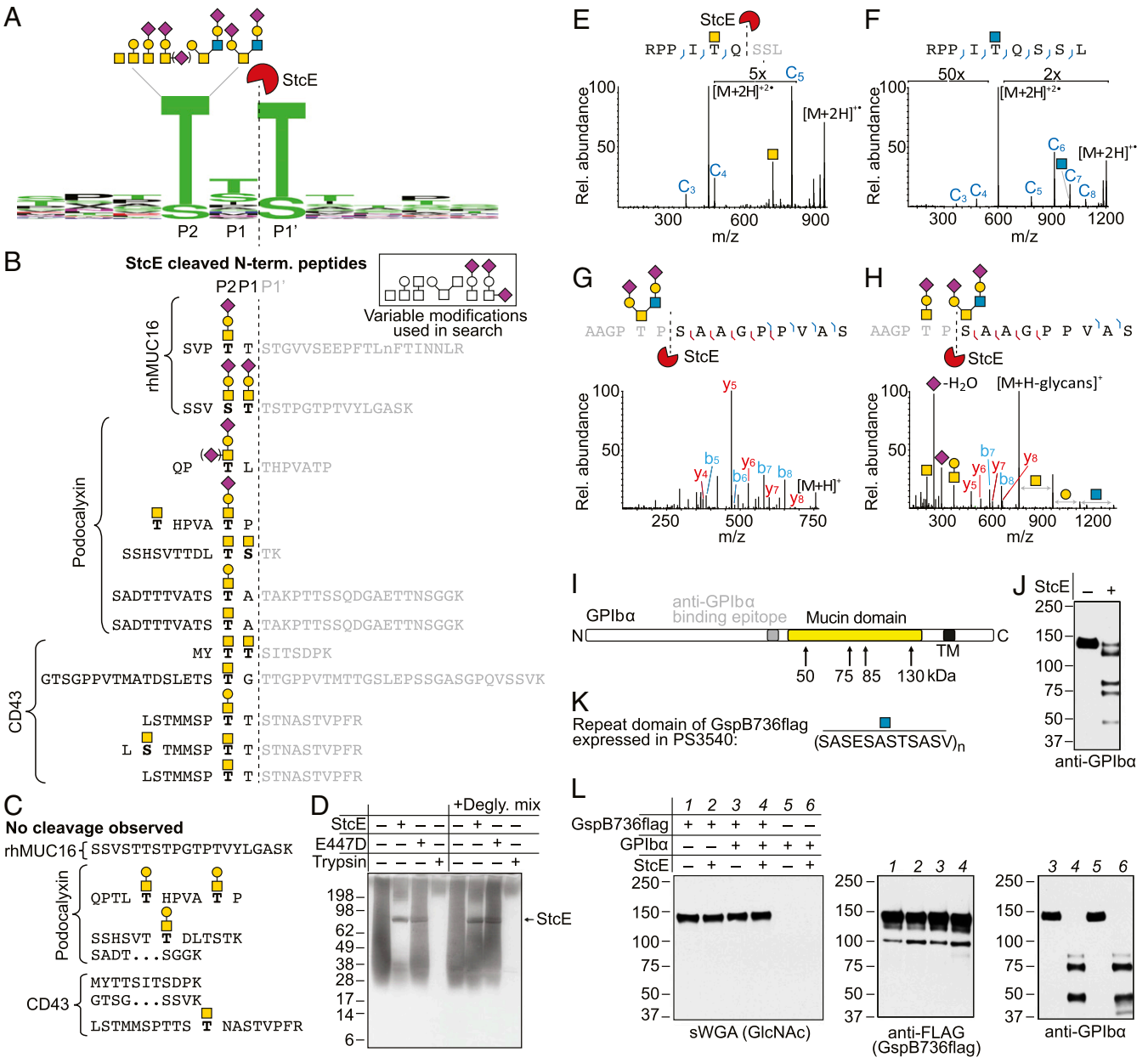

Fig. 2. StcE exhibits peptide- and glycan-based selectivity for mucins. (*A*) Glycoproteins shown in Fig. 1*B* were digested with StcE. The digest was then deglycosylated by treatment with PNGase F, trypsinized, and analyzed by MS. Sequences of the StcE-dependent cleavage products were used as WebLogo inputs (weblogo.berkeley.edu). StcE recognizes the consensus sequence S/T*–X–S/T, and cleaves the peptide backbone before the P1′ S/T only when the P2 S/T is glycosylated (indicated with an asterisk). Detected glycoforms from recombinant proteins and synthetic peptides at P2 are shown. Parentheses indicate that the linkage for the second sialic acid of the disialylated structure could not be assigned. (*B*) Examples of StcE-cleaved N-terminal peptides from several recombinant mucins, with assigned glycan structures shown. (*Inset*) Peptides were searched with the *O*-glycan 6 most common variable modification database. Gray text denotes the C-terminal sequence of detected intact peptides. (*C*) Examples of detected peptides with sequences corresponding to those shown in *B*, but with glycosylation inconsistent with the consensus motif. (*D*) StcE, E447D, and trypsin were reacted with a native peptide backbone *N*-carboxyanhydride–derived copolymer consisting of 50% GalNAc-α-*O*-serine and 50% lysine (38), with and without prior enzymatic deglycosylation [Delgycosylation (Degly.) Mix; Promega]. The arrow indicates the StcE band. StcE was incubated with RPPIT(α-GalNAc)QSSL (*E*), RPPI(β-GlcNAc)TQSSL (*F*), AGG (disialyl-core 2)TPSAAGPPVAS (*G*), and AGG(NeuAc-Gal-GalNAc)TP(disialyl-core 2)SAAGPPVAS (*H*) for 3 h at 37 °C and subjected to MS analysis. ETD spectra are shown for *E* and *F*; HCD spectra are shown for *G* and *H*. Rel., relative. (*I*) Diagram of GPIbα sequence, showing the mucin domain in yellow, along with arrows corresponding to sites of potential StcE cleavage, based on molecular weights of cleavage bands and amino acid sequence. (*J*) Lanes contain 1.5 μL of a platelet surface protein extract (1.2 μg of total protein and ∼15 ng of GPIbα), either untreated or treated with StcE protease (20 μg/mL) for 2 h at 37 °C. Proteins were separated by SDS/PAGE, transferred to nitrocellulose, and then probed with an antibody that recognizes the N-terminal domain of GPIbα. (*K*) Repeat domain of GspB736flag (a truncated, C-terminally 3xFLAG-tagged variant of *S. gordonii* GspB) is highly posttranslationally modified with unelaborated *O*-GlcNAc when expressed in PS3540. (*L*) GspB736flag and platelet GPIbα, either alone or in combination, were treated with StcE (100 μg/mL) for 17 h at 37 °C. Blots were probed with succinylated wheat germ agglutinin (sWGA), anti-FLAG, or anti-GPIbα as indicated.

of glycopeptides present in the StcE samples but not in the control samples (PNGaseF and trypsin only), we discovered that StcE has a distinct peptide consensus sequence, S/T*–X–S/T, where cleavage occurred before the second serine or threonine and X was any amino acid (Fig. 2*A*). StcE is able to cleave sequences in which the spacer X is missing, although it does so less efficiently (*SI Appendix*, Fig. S7). As seen from a table containing StcE-cleaved N-terminal peptides, the P2 (*) position was invariably glycosylated (Fig. 2*B*; all sequenced peptides are shown in Dataset S1). This glycosylation ranged from a single *O*-GalNAc residue to higher order structures such as a disialylated T antigen, indicating that StcE accepts a variety of glycans at the P2 position. Note that StcE cleavage was also permissive to glycosylation at the P1 and P1′ positions. In all cases, neither the peptide sequence nor the glycan alone was sufficient to predict cleavage, as demonstrated in Fig. 2*C*, which lists several detected peptides with sequences corresponding to those shown in Fig. 2*B*, but with glycosylation that does not match the consensus motif.

Based on MS analysis of cleaved peptides (Fig. 2*B* and Dataset S1), the minimum necessary glycoform for StcE cleavage was a single GalNAc residue. To confirm this, we incubated a synthetic glycosylated polypeptide comprising GalNAc-α-*O*-Ser residues mixed with Lys residues in a random sequence (38) together with StcE. As seen in Fig. 2*D*, StcE cleaved the glycosylated polymer, and this cleavage was mitigated when the polymer was deglycosylated. We also found that StcE cleaved a synthetic peptide containing a single GalNAc, converting RPPIT*QSSL to RPPIT*Q (Fig. 2*E*). StcE did not cleave the same peptide sequence containing β-GlcNAc instead of α-GalNAc (Fig. 2*F*). In fact, StcE did not cleave any *O*-GlcNAcylated or *O*-mannosylated peptides tested (39–42) (*SI Appendix*, Fig. S8).

We next asked whether StcE could cleave peptides with larger, more elaborated glycan structures. StcE was reacted with synthetic glycopeptides containing core 2 structures on the P2 position or the P1′ position (43). In each case, StcE was able to cleave the peptide at the predicted site (Fig. 2 *G* and *H* and *SI Appendix*, Figs. S9 and S10). We also tested cleavage of the human platelet surface protein GPIbα, which is reported to contain ∼96% core 2 glycosylation in its mucin domain (44, 45) (Fig. 2*I*). StcE cleaved GPIbα from healthy donor-derived platelets into four fragments, as detected by an antibody specific to a region N-terminal to the mucin domain (Fig. 2*J*).

Finally, we tested StcE cleavage of the repeatedly *O*-GlcNAcylated *Streptococcus gordonii* protein, GspB736flag, which bears multiple *O*-GlcNAc modifications in its SASE-SASTSASV repeat domain (46) (Fig. 2*K*). As seen from anti-Flag and succinylated wheat germ agglutinin blots in Fig. 2*L*, GspB736flag is resistant to StcE cleavage even under forcing conditions (100 μg/mL for 17 h at 37 °C). GPIb is shown as a positive control for StcE activity, and is cleaved into smaller fragments than those in Fig. 2*J* as a result of the high enzyme/substrate (E/S) ratio.

Taken together, we can conclude that (*i*) *O*-GalNAc is the minimum required glycoform on the P2 position serine or threonine residue, (*ii*) elaborated core 1 and core 2 structures do not impede cleavage, and (*iii*) *O*-GlcNAc and *O*-mannose peptides are not cleaved by StcE. As *O*-GalNAc is the first glycan found on every site of mucin-type *O*-glycosylation and S/T–X–S/T is commonly found in their characteristic proline-, threonine-, and serine-rich repeat domains, these data support the view that StcE is a protease that selectively cleaves mucins but is promiscuous within that family.

Glycopeptide Docking Supports StcE's Selectivity for α-GalNAcylated Peptides. The observed insensitivity of *O*-glycosylated but non-mucinous proteins to StcE activity suggested that secondary structure may be an additional recognition determinant. For example, fetuin was not cleaved by StcE (Fig. 1*B*), although it exhibits a correctly glycosylated StcE consensus sequence GPT*PSAA (* = sialyl T antigen, among others) (47). We sought to better visualize this possibility via docking experiments using the StcE crystal structure (34) and several podocalyxin-based model peptides (Ac-PTLTH-NMe). When docked in a manner consistent with ligands in homologous metzincin crystal structures (48), each ligand's peptide backbone made specific contacts with the catalytic zinc ion as well as three sites (on two residues, Gly427 and Gly429) of a flanking β-strand, forming a segment similar to an antiparallel β-sheet (Fig. 3*A*). The glycan moieties projected away from the enzyme active site, allowing the enzyme to accommodate different glycan structures.

Notably, much work has been done to investigate the 3D structure of analogous mucin peptides and their glycoforms (49). In general, the naked peptides were found to be conformationally mobile but biased toward turn-like conformations; peptides modified with the α-GalNAc moiety were rigid and adopted extended conformations closer to a β-strand, consistent with the conformation of our docked peptides. This change in conformational preference, which was not observed following modification with β-GalNAc (50), is proposed to be due to steric factors and specific intra- and interresidue contacts unique to the α-GalNAc moiety (49).

In our modeling experiments, we often observed a conformation containing an intramolecular hydrogen bond between the amide hydrogen of the α-GalNAc moiety and the carbonyl oxygen of the terminal acetyl group in our docked structures. In this conformation, the carbonyl oxygen of the α-GalNAc is positioned to interact with an additional residue (Gly431) of the β-strand lining the active site (Fig. 3*A*). An analogous residue is involved in ligand recognition by related metzincin enzymes (48), and it is likely important for substrate recognition in StcE. If so, this provides a molecular mechanism connecting the α-GalNAc moiety's dual role in determining the substrate's (*i*) secondary structure and (*ii*) recognition by the enzyme.

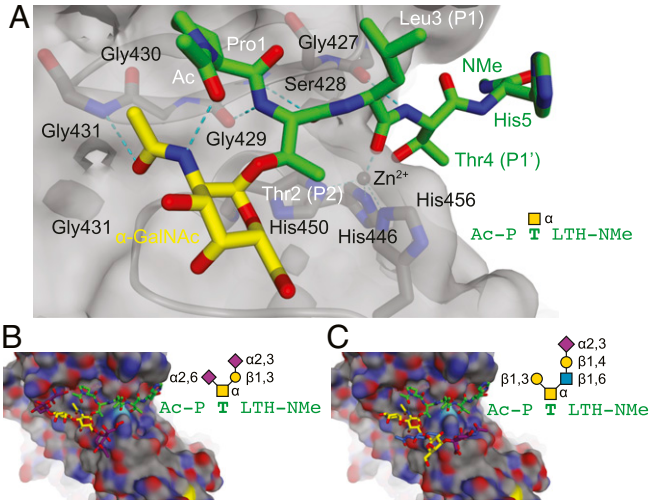

Fig. 3. Structure of StcE (34) and model peptides following docking. (*A*) Ac-P(α-GalNAc)TLTH-NMe. The peptide (green sticks) and GalNAc (yellow sticks) moieties of the ligand make contacts (cyan dashes) with the zinc ion (gray sphere) and important residues of the active site (gray sticks). The GalNAc moiety also makes an intramolecular contact (cyan dashes) with the acetyl group of the ligand. Combined, this docked structure suggests a role for α-GalNAc in affecting substrate conformation and recognition. P2, P1, and P1′ positions are indicated. StcE was able to accommodate larger glycans Ac-P(disialyl core 1)TLTH-NMe (78) (*B*) and Ac-P(sialyl core 2)TLTH-NMe (64) (*C*), and the docked species formed important interactions (cyan dashes) between the enzyme (shaded surface), peptide (green sticks), and GalNAc (yellow sticks) moieties, similar to *A*.

Importantly, larger core 1 and core 2 structures were also well accommodated by the enzyme in our docking studies, consistent with experimental findings (Fig. 3 *B* and *C*). Sugar residues attached at C6 of the core α-GalNAc moiety flanked the protein surface adjacent to the active site, whereas those branching from C3 were projected along the enzyme surface adjacent to the ligand's N terminus. Ultimately, the docked structures help visualize how StcE might contact the peptide backbone of heavily glycosylated mucin domains. We note that these models are consistent with Strynadka and coworkers' (34) proposed cleavage mechanism, involving closure of the StcE active site around substrates with an extended α-GalNAc–induced secondary structure.

StcE Improves MS Analysis of Mucin-Domain Glycoproteins. Given its selectivity for mucin domains, we predicted that StcE could be incorporated in common proteomic work flows to facilitate analysis of mucin glycoproteins. The six highly glycosylated, commercially available, and recombinant mucin-domain proteins we employed for characterization of StcE's cleavage specificity are known to be important in human disease. Podocalyxin is a sialoprotein that reportedly promotes the growth, proliferation, and metastasis of solid tumors (51, 52). MUC16, as discussed in the Introduction, is highly dysregulated in carcinomas, and its aberrant glycosylation and/or overexpression contributes to immune evasion (53). CD43 and CD45 are both regulators of several T cell functions, including their activation, proliferation, and differentiation (54, 55). PSGL-1, a sialomucin, engages with its receptor, P-selectin, to mediate leukocyte rolling and adhesion (10). Finally, Syncam-1 is important for cell adhesion and acts as a tumor suppressor in non-small cell lung cancer (56).

To explore whether StcE cleavage could improve MS analysis of these mucin-domain glycoproteins, we analyzed all peptides present in StcE-treated and control samples. As before, the data were searched using a small *O*-glycan database (*O*-glycan 6 most common; structures are listed in Fig. 2*B*, *Inset*). We chose this glycan database for several reasons. First, poor ionization of structures bearing large sialylated glycans limits their detection by MS (23). Second, charge density considerations generally limit electron transfer dissociation (ETD) fragmentation of glycopeptides with glycans larger than disialyl T (57). Third, current search algorithms are restricted in their ability to search large *O*-glycan libraries while maintaining reasonable search times. Finally, the recombinant proteins listed above were expressed in either Chinese hamster ovary (CHO; podocalyxin, MUC16, and PSGL-1), HEK293T (CD43), or NS0 (CD45, Syncam-1) cells. CHO cells are known to primarily express unelaborated core 1 structures (58), whereas HEK293T cells display slightly more elaborated core 1 and 2 glycans (59). Limited information is available on *O*-glycosylation profiles of proteins expressed in NS0 cells (60). Therefore, the *O*-glycan 6 most common database should represent a reasonable portion of the glycans that are present on the recombinant proteins and amenable to analysis.

Higher energy collision-induced dissociation (HCD) spectra corresponding to each glycopeptide identified by the database search were inspected to confirm that the HexNAc and/or NeuAc fingerprint ions were present and that the assigned monosaccharide composition was correct. Then, the ETD spectrum for the corresponding glycopeptide was manually sequenced to definitively site-localize the glycosylation. As seen in Fig. 4*A*, StcE treatment increased protein sequence coverage by up to 50%, number of glycosites by up to sixfold, and number of localized glycans by up to 11-fold, with averages of 20%, 3.5-fold, and fourfold improvement, respectively. Notably, many of these sites of glycosylation were previously unknown. Of the 77 sites we defined, 59 were reported previously, increasing the total number of sites by 31%. The observed gains were likely due to StcE's ability to break up areas of dense *O*-glycosylation, which generated smaller glycopeptides with higher charge density. This allowed for ETD spectra, which are necessary for glycosite mapping. To illustrate this concept, ETD spectra of three representative CD43 peptides are shown in Fig. 4*B*. In the untreated sample (Fig. 4*B*, *Bottom*) site localization of the three *O*-GalNAc modifications was not possible, but StcE treatment (Fig. 4*B*, *Top*) resulted in two peptides covering the same sequence, each with sufficient charge and fragmentation for site localization of the modification. We note also that in silico searches for peptides with serine or threonine at their N terminus may aid in database searches of StcE-cleaved samples (*SI Appendix*, Fig. S11).

StcE Cleaves Mucins from Biological Samples. Next, we tested StcE's ability to cleave native, human-derived mucins. We found that a commercially available semicrude preparation of MUC16 from ascites fluid of patients with cancer was sensitive to StcE cleavage (Fig. 5*A* and *SI Appendix*, Fig. S12). The density around 200 kDa in the semicrude preparation does not originate from full-length glycosylated MUC16, however, which migrates with an apparent molecular weight in the megadalton range (61, 62). To demonstrate cleavage of full-length human MUC16, we treated crude ascites, taken from a patient who had ovarian

A

	Sequence coverage (%)		No. of glycosites		No. of localized glycans	
StcE	–	+	–	+	–	+
Podocalyxin	69	82	9	27	27	91
MUC16	47	66	2	11	5	26
CD43	92	92	4	24	6	53
PSGL-1	46	72	4	5	7	8
Syncam-1	86	97	1	6	1	11
CD45	66	67	2	4	2	5
Averages	67.7	79.3	3.7	12.8	8.0	32.3
Fold Increase		**1.2**		**3.5**		**4**

B

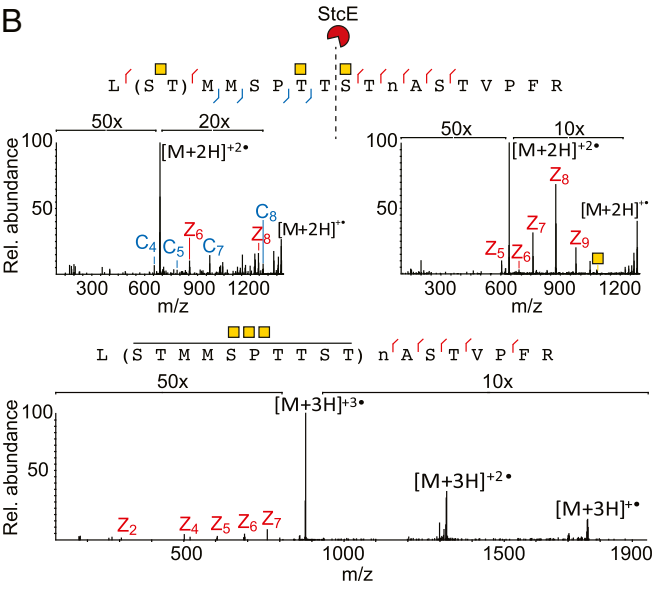

Fig. 4. StcE increased the number of assigned glycosites, number of localized glycans, and sequence coverage of every protein studied. (*A*) Recombinant substrates were digested with StcE, de–*N*-glycosylated with PNGaseF, trypsinized, and then subjected to MS using an HCD-triggered ETD instrument method. ETD spectra were used to assign glycosites. (*B*, *Top*) ETD spectra of N- and C-terminal StcE-cleaved peptides LSTMMSPTT and STNASTVPFR from CD43 from the experiment described in *A*. (*B*, *Bottom*) ETD spectrum from the control sample (PNGaseF and trypsin only) is shown. The lowercase "n" denotes deamidation. Parentheses indicate that the sites modified with GalNAc residues could not be assigned. Rel., relative.

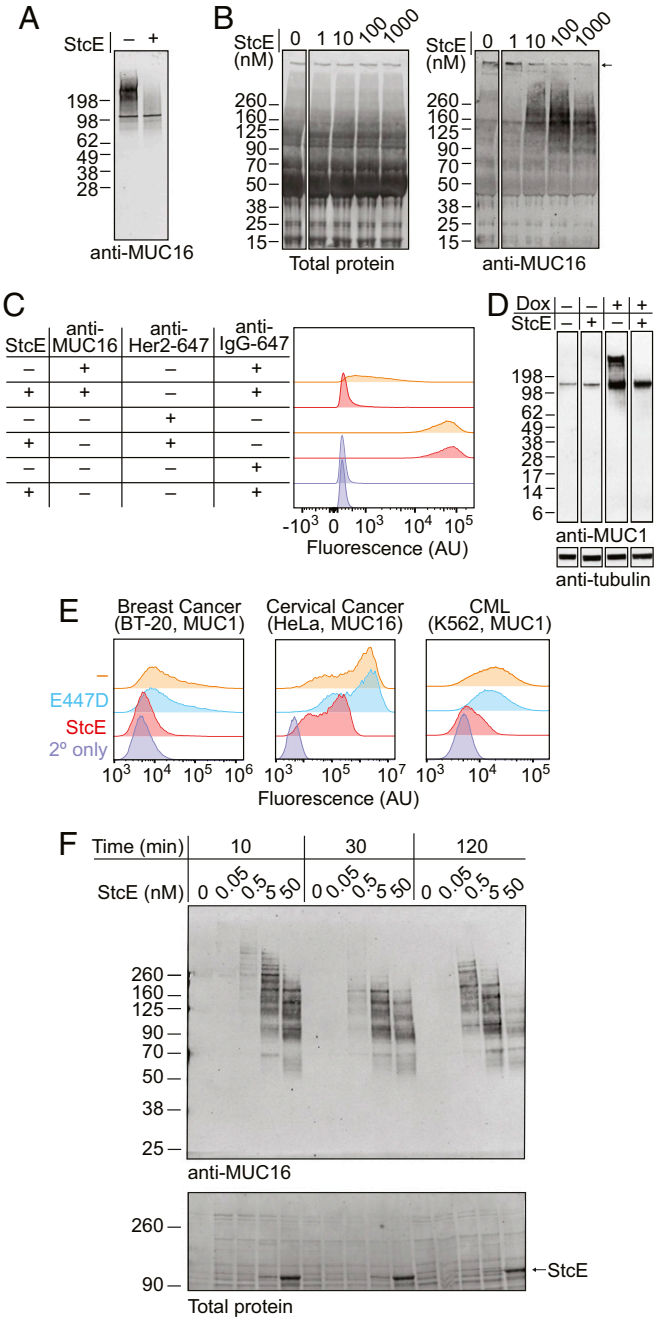

Fig. 5. StcE can cleave native mucins from ascites fluid derived from patients with cancer and cultured cell surfaces. (*A*) StcE was incubated at a 1:10 E/S ratio for 3 h at 37 °C with a semicrude patient-derived commercial preparation of MUC16 (Lee BioSolutions). An anti-MUC16 Western blot is shown, as MUC16 was a minority of the material by total protein stain (silver stain is shown in *SI Appendix*, Fig. S9). The ∼100-kDa band in both lanes is likely nonspecific antibody binding to a contaminating protein in the commercial sample of CA-125. The MUC16 antibody (X75; Abcam) binds to extracellular repeat domains. (*B*) Crude ascites fluid derived from patients with ovarian cancer was incubated with StcE for 1 h at 37 °C at the concentrations shown. (*C*) SKBR3 cells were treated with 50 nM StcE for 2 h at 37 °C, and then subjected to live cell flow cytometry with staining for MUC16 and HER2. (*D*) Western blot of MCF10A cells expressing MUC1ΔCT on a doxycycline (Dox) promoter. StcE treatment was performed on live cells as in *C*. The MUC1 antibody (VU4H5; Cell Signaling Technology) binds to extracellular repeat domains. (*E*) BT-20, HeLa, and K562 cells were treated with StcE as in *C* and subjected to live cell flow cytometry with staining for MUC1 or MUC16. CML, chronic myelogenous leukemia. (*F*) Plated HeLa cells incubated in HBSS were treated with StcE at the times and concentrations shown. Supernatants were lyophilized, resuspended in sample buffer, separated by SDS/PAGE, and immunoblotted for MUC16.

cancer, with StcE. In untreated ascites, we detected density in the stacking gel (Fig. 5*B*, arrow and *SI Appendix*, Fig. S13), consistent with a species of very high molecular weight. StcE treatment for 1 h at 37 °C resulted in a dose-dependent decrease in apparent molecular weight, demonstrating that StcE is active on human MUC16.

As noted above, cell surface mucins have been implicated as pathogenic drivers of cancer progression. If StcE could cleave such mucins, it would enable their functional analysis. We first tested cell viability after StcE treatment, and found that it was nontoxic to both adherent and suspension cell lines at all concentrations tested, and did not affect proliferation over days (*SI Appendix*, Fig. S14). Next, we treated the human breast cancer cell line SKBR3 with StcE and probed for changes in abundance of MUC16. As determined by flow cytometry analysis, StcE depleted MUC16, but had no effect on the highly abundant *N*-glycosylated but nonmucin HER2 receptor (Fig. 5*C*). We also tested StcE's effects on the breast cancer-associated mucin MUC1 using an MCF10A cell line ectopically expressing a signaling-deficient form of this cell surface mucin (MUC1ΔCT) (63). StcE readily cleaved glycosylated MUC1ΔCT but was inactive on an underglycosylated form of MUC1ΔCT (∼125 kDa) also visible on the Western blot (Fig. 5*D* and *SI Appendix*, Figs. S15 and S16). Further, StcE's activity does not appear to be cell line-dependent, as it digested cell surface MUC1 and MUC16 from cell lines derived from a variety of cancer types (Fig. 5*E*). These data confirm that StcE retains its activity *in cellulo*. Furthermore, the supernatants of StcE-treated HeLa cells, but not their vehicle-treated counterparts, stained strongly for MUC16, and the apparent molecular weight of the mucin fragments decreased with increasing enzyme concentration and treatment time (Fig. 5*F* and *SI Appendix*, Fig. S17). StcE can therefore be used as a tool to release and solubilize mucins from biological samples.

To test whether StcE could be employed to purify mucins from protein mixtures, we conjugated StcE to beads using reductive amidation. StcE effectively enriched C1INH from a mixture with BSA, using EDTA in the buffer to inhibit StcE's cleavage activity (*SI Appendix*, Fig. S18*A*). We expanded on this with a larger scale purification of MUC16 from the supernatant of a human ovarian cancer cell line, OVCAR-3, using E447D-conjugated beads. Once again, EDTA was added to abrogate any residual proteolytic activity of the E447D mutant. An anti-MUC16 Western blot revealed depletion of MUC16 from the flow-through and enrichment of MUC16 in the elution (*SI Appendix*, Fig. S18*B*). In addition, a high-molecular-weight band matching the observed molecular weight of MUC16 was visible in the elution by total protein stain, underscoring the efficiency of the enrichment. We note that further optimization is required to reduce the amount of shed E447D under the boiling elution conditions.

StcE Treatment of Cultured Cells Reveals That Siglec-7 Binds Mucin-Domain Glycoproteins. Given StcE's selectivity for mucins, we reasoned that it could be employed as a tool to discover mucin-based ligands of glycan-binding receptors whose physiological binding partners remain unknown. There is growing evidence that so-called glycan-binding proteins can recognize discrete glycoprotein or glycolipid ligands via motifs that encompass both glycan structures as well as elements of their underlying scaffolds. As a landmark example, PSGL-1 was identified as a cell surface mucin that functions as the chief ligand for P-selectin at sites of inflammatory leukocyte recruitment (10) [notably, PSGL-1 is digested with StcE (Fig. 1*B*)]. The molecular determinants of PSGL-1 that confer P-selectin binding include a specific *O*-glycan structure combined with a nearby peptide motif (64). Likewise, the immune modulatory receptor PILRα recognizes a composite mucin-derived sialoglycopeptide epitope on

cognate ligands (65). These examples hint at a rich biology for mucin glycoproteins as ligands for a variety of receptors involved in cell trafficking and immune regulation. We speculated that StcE's ability to destroy these structures on cells might help reveal mucins as binding partners of orphan receptors.

Siglecs are a glycan-binding receptor family whose physiological ligands are largely unknown (66, 67). Individual family members exhibit preferences for sialosides of various linkages to underlying glycan motifs, but the specific glycoproteins or glycolipids they interact with in biological settings are mysterious. Recent work has implicated Siglec-7 and Siglec-9 as inhibitory receptors that function similar to the immune checkpoints PD-1 and CTLA-4, the targets of several successful cancer immune therapies (68). Extracellularly, Siglec-7 and Siglec-9 each have a sialic acid-binding V-set domain (Fig. 6*A*). Intracellularly, they resemble PD-1, with C-terminal cytosolic tyrosine-based inhibitory motif (ITIM) and tyrosine-based switch motif (ITSM) domains that mediate inhibitory signaling. Enzymatic removal of sialic acids *en masse* from cancer cell surfaces enhances immune cell-mediated clearance of those cells through loss of Siglec-7 and Siglec-9 binding (69). Despite years of effort, however, professional ligands of Siglec-7 and Siglec-9 have not been identified, leading to a prevailing view that these Siglecs have broad and overlapping affinities for a multitude of sialylated cell surface molecules (70).

Using soluble Siglec-Fc fusions, we assessed the effects of StcE treatment on Siglec-7 and Siglec-9 binding to SKBR3 cells. Analysis by flow cytometry showed that StcE treatment depletes Siglec-7–Fc binding but has no effect on Siglec-9–Fc binding (Fig. 6*B*, *Top*, flow cytometry histogram and *Bottom*, biological replicates). The inactive StcE point mutant E447D had no effect on binding of either Siglec-Fc. We confirmed that Siglec-7–Fc binding and Siglec-9–Fc binding are both dependent on sialic acid via treatment with *Vibrio cholerae* sialidase (*SI Appendix*, Fig. S19). These results suggest that Siglec-7 recognizes mucin glycoproteins on SKBR3 cells but Siglec-9 binds structures that are resistant to StcE digestion.

To ensure that StcE did not simply bind to cell surface mucins and block accessibility to Siglec-7–Fc, we stained StcE- and E447D-treated SKBR3 cells with anti-His antibodies, which should bind the His-tagged enzymes (Fig. 6*C*). E447D bound cell surfaces more tightly than StcE did, but did not deplete Siglec-7–Fc binding, indicating that StcE's enzymatic activity was required for its effects. Interestingly, periodate-mediated labeling of cell surface sialic acids (71) revealed that StcE treatment had only a minor effect on total cell surface sialic acid levels (Fig. 6*D*). Thus, StcE removes a small fraction of total sialosides, while depleting a larger fraction of Siglec-7–Fc-binding structures. Finally, we tested a panel of cell lines for StcE-mediated depletion of Siglec-7 and Siglec-9 ligands. In all other cell lines tested, Siglec-7–Fc binding decreased upon StcE digestion, while Siglec-9–Fc binding remained unchanged (Fig. 6*E* and *SI Appendix*, Fig. S19). Note that previous reports that glycolipids such as GD3 can serve as ligands for Siglec-7 (72) are consistent with StcE's ability to substantially, but not completely, abrogate Siglec-7–Fc binding.

To further support the identification of Siglec-7 as a sialomucin-binding receptor, we employed *ldlD* CHO cells, which are deficient in UDP-glucose/galactose-4-epimerase (GALE) (73). GALE interconverts UDP-glucose and UDP-GlcNAc to UDP-galactose and UDP-GalNAc, respectively. Without active GALE, *ldlD* CHO cells can still take up glucose from tissue culture media and use it to biosynthesize nucleotide sugars of glucose, mannose, fucose, and sialic acid. However, they cannot initiate or elaborate their glycans with GalNAc or galactose, resulting in truncated cellular glycans. Supplementing the media with 10 μM galactose and 100 μM GalNAc rescues the phenotype, as these undergo conversion to the respective nucleotide sugars within cells (73). Unrescued *ldlD* CHO cells exhibited weak binding by both

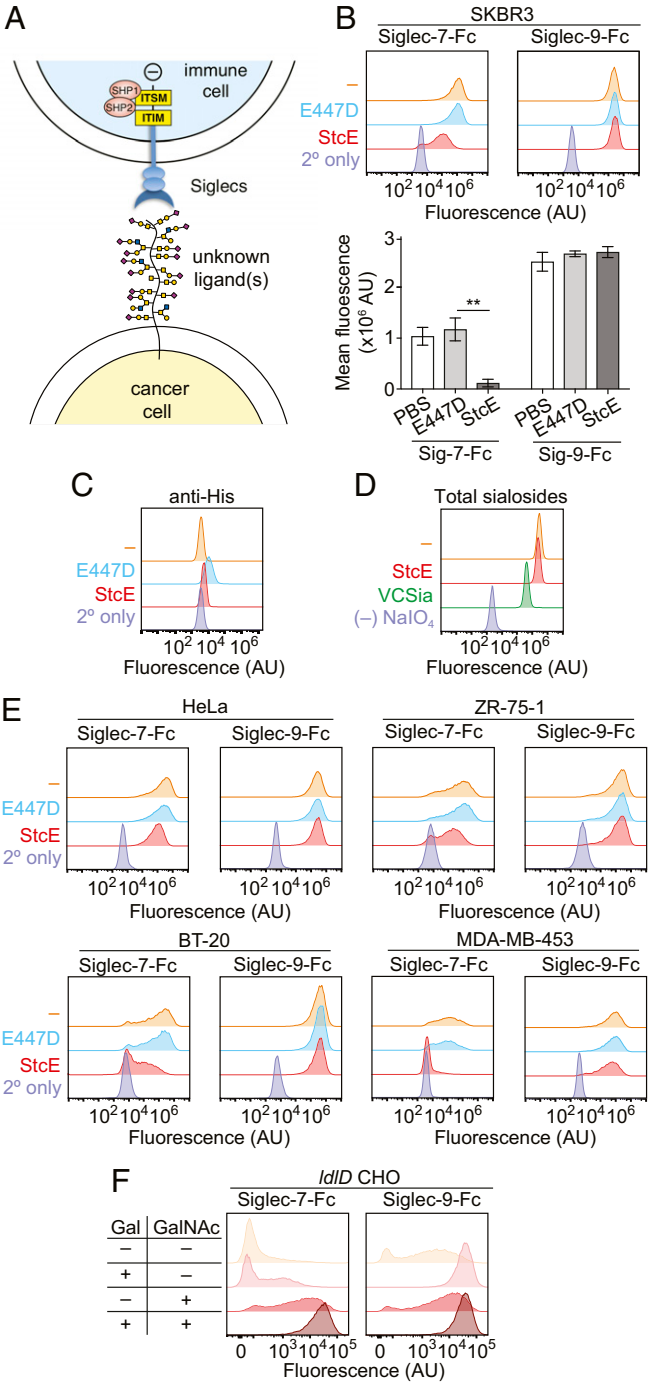

Fig. 6. StcE reveals that Siglec-7 binds mucin-domain glycoproteins. (*A*) Siglecs are a family of leukocyte receptors that bind sialylated ligands of unknown identity. Similar to PD-1, upon ligand binding, they transmit inhibitory signals through intracellular tyrosine-based switch motif (ITSM) and tyrosine-based inhibitory motif (ITIM) domains. (*B*, *Top*) SKBR3 cells were treated with 50 nM StcE or E447D for 2 h at 37 °C, stained with Siglec-7–Fc and Siglec-9–Fc, and subjected to live cell flow cytometry. (*B*, *Bottom*) Mean fluorescence intensity of three biological replicates is shown. Error bars are SDs. $^{**}P < 0.005$ by Student's two-tailed *t* test. (*C*) SKBR3 cells treated as in *B* were washed, stained with anti–His-FITC or isotype-FITC, and subjected to live cell flow cytometry. (*D*) SKBR3 cells treated with 50 nM StcE, 50 nM E447D, or 30 nM *V. cholerae* sialidase (VCSia) for 2 h at 37 °C were subjected to periodate-based sialic acid labeling (71), followed by fixed cell flow cytometry. (*E*) Flow cytometry as in *B* on HeLa, ZR-75-1, BT-20, and MDA-MB-453 cells. (*F*) Flow cytometry analysis of *ldlD* CHO cells, with galactose (Gal) and GalNAc rescue conditions shown. Staining was performed with Siglec-7–Fc or Siglec-9–Fc.

Siglec-7–Fc and Siglec-9–Fc (Fig. 6*F*). Siglec-9–Fc binding increased by approximately the same amount after rescue with galactose alone and with both galactose and GalNAc supplementation, but increased only slightly with GalNAc rescue alone (Fig. 6*F*, *Right*). These results are consistent with a view that Siglec-9 ligands are predominantly nonmucinous, as GalNAc deficiency should abrogate mucin-type *O*-glycosylation (74). Siglec-7–Fc binding was largely unaffected by galactose supplementation alone, but increased with GalNAc supplementation. Rescue with both sugars increased Siglec-7–Fc binding further (Fig. 6*F*, *Left*). In all conditions, both Siglec-7–Fc binding and Siglec-9–Fc binding were sialidase-sensitive, confirming their dependence on sialic acid (*SI Appendix*, Fig. S19). In addition, StcE treatment had no effect on Siglec-9–Fc binding across any rescue condition and only decreased Siglec-7–Fc binding in cases with GalNAc supplementation (*SI Appendix*, Fig. S16).

These data distinguish the specificities of Siglec-7 and Siglec-9 on cell surfaces. In the case of Siglec-7, it appears that professional glycoprotein ligands may exist, and that at least a subset are mucin-domain glycoproteins. If specificity for a professional glycoprotein ligand could be defined, it would pave the way for a novel class of immune checkpoint interventions. As noted above, the role that mucin-domain glycoproteins play in immunological signaling is not limited to Siglec-7. For example, receptors such as CD45 (75) and TIM-3 (76), which are emerging as critical players in healthy immune function and the immune response to cancer, contain prominent mucin domains that are thought to be necessary for their activities. Further, several members of the galectin family, which are prooncogenic glycan-binding proteins, are known mucin binders, but their specificities for discrete glycoproteins have not been fully characterized (77). Enzymatic demucination with StcE could provide a powerful tool for deorphanizing the receptors and ligands that interact with mucin-domain glycoproteins.

In conclusion, we introduce the bacterial protease StcE as a robust and efficient tool for the analysis of mucin-domain glycoproteins. After solving its peptide- and glycan-based selectivity for mucins, we demonstrated that this enzyme can be used in glycoproteomic mapping of mucin glycosites and their associated glycoforms, as a method for selective cleavage, release, and enrichment of mucins from cell and tissue material, as well as in the study of native mucin biology. The ease of purification, stability, and potency of the enzyme makes it a reagent easily distributed and utilized by a variety of laboratories. Further exploration and engineering of bacterial mucinase families may yield enzymes with complementary glycan and peptide specificities that would expand the biochemist's toolbox further. In addition, we hope that the ability to efficiently generate glycosylated mucin peptides will spur efforts to develop better methods for glycopeptide ionization.

Materials and Methods

In Vitro StcE Activity Assays. C1INH purified from human plasma was purchased from Molecular Innovations (catalog no. HC1INH). Recombinantly expressed MUC16, podocalyxin, PSGL-1, Syncam-1, and CD45 were purchased from R&D Systems (5609-MU, 1658-PD, 3345-PS, AF1459-SP, and 1430-CD, respectively). Recombinantly expressed CD43 was purchased from Origene (TP304195). Human fibronectin was purchased from Sigma–Aldrich (F0895). IL-2 expressed in insect cells was purchased from Biolegend (589104). EPO expressed in CHO cells was purchased from Peptrotech (100-64). κ-Casein from bovine milk was purchased from Sigma–Aldrich (C0406). Human transferrin was purchased from Sigma (3309). α1-Acid glycoprotein from human plasma was purchased from Sigma–Aldrich (G3642). BSA was purchased from Sigma–Aldrich (A7906-1KG). Fetuin from bovine fetal calf serum was obtained from Promega (V496A).

To test StcE's activity against mucin-like glycoproteins and nonmucins, reaction conditions were as follows: 1:10 E/S ratio, total volume of 15 μL, and 0.1% Protease Max in 50 mM ammonium bicarbonate as buffer (3 h at 37 °C). A portion of each condition (0.5 μg) was loaded onto 10% Criterion XT Bis-Tris precast gels (Bio-Rad) and run with XT-Mes (Bio-Rad) at 180 V for 1 h. Each gel was stained with silver stain or a Pro-Q Emerald 300 Glycoprotein Gel and Blot Stain Kit (ThermoFisher Scientific), according to the manufacturer's instructions. Deglycosylation of rhMUC16 was performed according to the manufacturer's instructions (Deglycosylation Mix; Promega). Mucin mimetic copolymer consisting of 50% GalNAc-Ser and 50% Lys was synthesized as previously described (38). Both deglycosylated polymer and untreated polymer were subjected to StcE cleavage and gel staining as described above for recombinant protein substrates.

The *O*-mannosylated and core 2 peptides were synthesized as previously described (39–43). Commerically synthesized *O*-GalNAcylated and *O*-GlcNAcylated peptides were kindly supplied by Don Hunt, University of Virginia, Charlottesville, VA. The *O*-mannosylated, *O*-GalNAcylated, and *O*-GlcNAcylated peptides were supplied dry and were reconstituted in 0.1% formic acid to give a final concentration of 1 μg/μL (*O*-GlcNAc/GalNAc peptides) or 2 μg/μL (*O*-mannose peptides). β-*O*-GlcNAcylated and α-*O*-GalNAcylated peptides were added to 0.5 μg of StcE in a final volume of 13 μL. The *O*-mannosylated peptides (1 μg each) were added to StcE in a separate reaction under the same conditions. For core 2 synthetic peptides, 40 μL of solutions containing 1 μM peptide was separately dried down via a speedvac. Reaction with StcE was performed with 1 μg of enzyme in a total volume of 12 μL of 50 mM ammonium bicarbonate for 3 h at 37 °C, with shaking. Negative controls substituted StcE with an equal volume of 50 mM ammonium bicarbonate. Afterward, the reaction was quenched by the addition of 0.5 μL of formic acid. All samples were desalted via C18 to remove StcE and analyzed by high-resolution MS.

MS. A description of the sample preparation is provided in *SI Appendix*. Samples (1 μg) were reconstituted in 10 μL of 0.1% formic acid (ThermoFisher Scientific) containing 25 fmol/μL angiotensin (Millipore Sigma) and vasoactive peptide (Anaspec). Samples were analyzed by online nanoflow liquid chromatography-tandem MS using an Orbitrap Fusion Tribrid mass spectrometer (ThermoFisher Scientific) coupled to a Dionex Ultimate 3000 HPLC system (ThermoFisher Scientific). A portion of the sample (4 μL) was loaded via autosampler onto a 20-μL sample loop and injected at 0.3 $\mu L \cdot min^{-1}$ onto a 75-μm × 150-mm EASY-Spray column (ThermoFisher Scientific) containing 2 μm C18 beads. The column was held at 40 °C using a column heater in the EASY-Spray ionization source (ThermoFisher Scientific). The samples were eluted at 0.3 $\mu L \cdot min^{-1}$ using a 90-min gradient and a 185-min instrument method. Solvent A was composed of 0.1% formic acid in water, whereas solvent B was 0.1% formic acid in acetonitrile. The gradient profile was as follows (min/%B): 0:3, 3:3, 93:35, 103:42, 104:98, 109:98, 110:3, and 185:3. The instrument method used an MS1 resolution of 60,000 at FWHM of 400 *m/z*, an automatic gain control (AGC) target of 3e5, and a mass range from 300 to 1,500 *m/z*. Dynamic exclusion was enabled with a repeat count of 3, repeat duration of 10 s, and exclusion duration of 10 s. Only charge states 2–6 were selected for fragmentation. MS2 resolutions were generated at top speed for 3 s. HCD was performed on all selected precursor masses with the following parameters: isolation window of 2 *m/z*, 28–30% collision energy, ion trap or orbitrap (resolution of 30,000) detection, and AGC target of 1e4 ions. ETD was performed if (*i*) the precursor mass was between 300 and 1,000 *m/z* and (*ii*) three of eight glycofingerprint ions (126.055, 138.055, 144.07, 168.065, 186.076, 204.086, 274.092, and 292.103) were present at ±0.5 *m/z* and greater than 5% relative intensity. ETD parameters were as follows: calibrated charge-dependent ETD times, 2e5 reagent target, and precursor AGC target of 1e4. MS data analysis is discussed in *SI Appendix*.

Flow Cytometry and Western Blotting of StcE-Treated Cells. Cell culture materials and methods are discussed in *SI Appendix*. Cells were treated with StcE or E447D when plated or after lifting with enzyme-free cell dissociation buffer (ThermoFisher Scientific). Typical treatment conditions were 5 μg of StcE per 1 million cells in 1 mL of complete media or Hank's buffered salt solution (HBSS) for 2 h at 37 °C. After treatment, cells were washed with PBS or HBSS. For flow cytometry, cells were resuspended in cold PBS with 0.5% BSA and transferred to a 96-well, V-bottom plate. Cells were then resuspended in the probe of interest. Flow cytometry antibody vendors and treatment conditions are discussed in *SI Appendix*. Flow cytometry data were analyzed using FlowJo v. 10.0 (TreeStar). For Western blots, supernatants posttreatment (1-mL volumes) were collected into tubes containing 75 μL of 0.5 M EDTA to quench the reaction, and then snap-frozen in liquid nitrogen and lyophilized to dryness. Posttreatment cells were washed with enzyme-free cell dissociation buffer, which contains EDTA, to quench the reaction. Cells were then washed two times with PBS, pelleted, and lysed with sample buffer [1× NuPAGE LDS Sample Buffer (ThermoFisher Scientific) supplemented with 25 mM DTT]. Genomic DNA was sheared via probe tip sonication. Lyophilized supernatants were brought up in sample buffer. Both cell lysates and supernatants were boiled for 5 min at 95 °C and spun at 14,000 × *g*

for 2 min, and 30 μL of each was loaded into an 18-well, 4–12% Criterion XT Bis-Tris precast gel (Bio-Rad). The gel was run with XT MOPS (Bio-Rad) at 180 V for 1 h. Proteins were transferred to 0.2 μm of nitrocellulose using the Trans-Blot Turbo Transfer System (Bio-Rad), at 2.5 A constant for 15 min. Total protein was quantified using REVERT stain (LI-COR Biosciences) or Ponceau-S stain (Millipore Sigma). Western blot antibody vendors and treatment conditions, as well as StcE treatment of ascites fluid derived from patients with ovarian cancer, are discussed in *SI Appendix*. Clinical samples from consenting patients were obtained from the Department of Obstetrics and Gynecology, Stanford Hospital under approved Institutional Review Board protocol 13939.

ACKNOWLEDGMENTS. We thank Jeffrey Shabanowitz and Dina Bai (University of Virginia) for technical support. We thank Donald Hunt (University of Virginia) for his gift of *O*-GlcNAc and *O*-GalNAc synthetic peptides and Natalie Strynadka (University of British Columbia) for her gift of the StcE expression plasmid. We thank Ryan Flynn and Benjamin Schumann (Stanford University) for critical reading of the manuscript. This work was supported, in part, by National Cancer Institute Grant R01CA227942 (to C.R.B.). S.A.M. and J.R.K. were supported by National Institute of General Medical Sciences F32 Postdoctoral Fellowships. K.P. was supported by a National Science Foundation Graduate Research Fellowship, a Stanford Graduate Fellowship, and the Stanford ChEM-H Chemistry/Biology Interface Predoctoral Training Program. B.A.B. was supported by the National Institutes of Health Grants IA106987 and CA221244, and the American Heart Association Grant 17SDG33660424. O.D. and V.K. were supported by the Mary Lake Polan Gynecologic Oncology Endowment for Gynecologic Cancer Research, Stanford School of Medicine, Stanford University.

1. Woo CM, Iavarone AT, Spiciarich DR, Palaniappan KK, Bertozzi CR (2015) Isotope-targeted glycoproteomics (IsoTaG): A mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat Methods* 12:561–567.
2. Hollingsworth MA, Swanson BJ (2004) Mucins in cancer: Protection and control of the cell surface. *Nat Rev Cancer* 4:45–60.
3. Hang HC, Bertozzi CR (2005) The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem* 13:5021–5034.
4. Haltiwanger RS, Lowe JB (2004) Role of glycosylation in development. *Annu Rev Biochem* 73:491–537.
5. Gendler SJ, Spicer AP (1995) Epithelial mucin genes. *Annu Rev Physiol* 57:607–634.
6. Naughton J, Duggan G, Bourke B, Clyne M (2014) Interaction of microbes with mucus and mucins: Recent developments. *Gut Microbes* 5:48–52.
7. van Putten JPM, Strijbis K (2017) Transmembrane mucins: Signaling receptors at the intersection of inflammation and cancer. *J Innate Immun* 9:281–299.
8. Kuo JC-H, Gandhi JG, Zia RN, Paszek MJ (2018) Physical biology of the cancer cell glycocalyx. *Nat Phys* 14:658–669.
9. Xu Z, Weiss A (2002) Negative regulation of CD45 by differential homodimerization of the alternatively spliced isoforms. *Nat Immunol* 3:764–771.
10. Pouyani T, Seed B (1995) PSGL-1 recognition of P-selectin is controlled by a tyrosine sulfation consensus at the PSGL-1 amino terminus. *Cell* 83:333–343.
11. Patton S, Gendler SJ, Spicer AP (1995) The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim Biophys Acta* 1241:407–423.
12. Das S, et al. (2015) Membrane proximal ectodomain cleavage of MUC16 occurs in the acidifying Golgi/post-Golgi compartments. *Sci Rep* 5:9759.
13. Felder M, et al. (2014) MUC16 (CA125): Tumor biomarker to cancer therapy, a work in progress. *Mol Cancer* 13:129.
14. Jonckheere N, Van Seuningen I (2010) The membrane-bound mucins: From cell signalling to transcriptional regulation and expression in epithelial cancers. *Biochimie* 92:1–11.
15. Kufe DW (2009) Mucins in cancer: Function, prognosis and therapy. *Nat Rev Cancer* 9:874–885.
16. Woods EC, et al. (2017) A bulky glycocalyx fosters metastasis formation by promoting G1 cell cycle progression. *eLife* 6:e25752.
17. Suh KS, et al. (2010) Ovarian cancer biomarkers for molecular biosensors and translational medicine. *Expert Rev Mol Diagn* 10:1069–1083.
18. Kimura T, et al. (2013) MUC1 vaccine for individuals with advanced adenoma of the colon: A cancer immunoprevention feasibility study. *Cancer Prev Res (Phila)* 6:18–26.
19. Zhou Y, Rajabi H, Kufe D (2011) Mucin 1 C-terminal subunit oncoprotein is a target for small-molecule inhibitors. *Mol Pharmacol* 79:886–893.
20. Chen Y, et al. (2007) Armed antibodies targeting the mucin repeats of the ovarian cancer antigen, MUC16, are highly efficacious in animal tumor models. *Cancer Res* 67:4924–4932, and erratum (2007) 67:5998.
21. Posey AD, Jr, et al. (2016) Engineered CAR T cells targeting the cancer-associated tn-glycoform of the membrane mucin MUC1 control adenocarcinoma. *Immunity* 44:1444–1454.
22. Kesimer M, Sheehan JK (2012) Mass spectrometric analysis of mucin core proteins. *Methods Mol Biol* 842:67–79.
23. Wohlgemuth J, Karas M, Eichhorn T, Hendriks R, Andrecht S (2009) Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Anal Biochem* 395:178–188.
24. Julenius K, Mølgaard A, Gupta R, Brunak S (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15:153–164.
25. Kane LP (2010) T cell Ig and mucin domain proteins and immunity. *J Immunol* 184:2743–2749.
26. Vakhrushev SY, et al. (2013) Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. *Mol Cell Proteomics* 12:932–944.
27. Yang S, et al. (2018) Deciphering protein O-glycosylation: Solid-phase chemoenzymatic cleavage and enrichment. *Anal Chem* 90:8261–8269.
28. Yang W, Ao M, Hu Y, Li QK, Zhang H (2018) Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides (EXoO). *Mol Syst Biol* 14:e8486.
29. Jiang P, Mellors A (2013) O-sialoglycoprotein endopeptidase. *Handbook of Proteolytic Enzymes*, eds Salvesen G, Rawlings ND (Academic, London), 3rd Ed, pp 1664–1666.
30. Nakjang S, Ndeh DA, Wipat A, Bolam DN, Hirt RP (2012) A novel extracellular metallopeptidase domain shared by animal host-associated mutualistic and pathogenic microbes. *PLoS One* 7:e30287.
31. Szabady RL, Welch RA (2013) StcE peptidase and the StcE-like metalloendopeptidases. *Handbook of Proteolytic Enzymes*, eds Salvesen G, Rawlings ND (Academic, London), 3rd Ed, pp 1272–1280.
32. Noach I, et al. (2017) Recognition of protein-linked glycans as a determinant of peptidase activity. *Proc Natl Acad Sci USA* 114:E679–E688.
33. Lathem WW, et al. (2002) StcE, a metalloprotease secreted by Escherichia coli O157:H7, specifically cleaves C1 esterase inhibitor. *Mol Microbiol* 45:277–288.
34. Yu ACY, Worrall LJ, Strynadka NCJ (2012) Structural insight into the bacterial mucinase StcE essential to adhesion and immune evasion during enterohemorrhagic E. coli infection. *Structure* 20:707–717.
35. Grys TE, Siegel MB, Lathem WW, Welch RA (2005) The StcE protease contributes to intimate adherence of enterohemorrhagic Escherichia coli O157:H7 to host cells. *Infect Immun* 73:1295–1303.
36. Hews CL, et al. (2017) The StcE metalloprotease of enterohaemorrhagic Escherichia coli reduces the inner mucus layer and promotes adherence to human colonic epithelium ex vivo. *Cell Microbiol* 19:e12717.
37. Grys TE, Walters LL, Welch RA (2006) Characterization of the StcE protease activity of Escherichia coli O157:H7. *J Bacteriol* 188:4646–4653.
38. Kramer JR, Onoa B, Bustamante C, Bertozzi CR (2015) Chemically tunable mucin chimeras assembled on living cells. *Proc Natl Acad Sci USA* 112:12574–12579.
39. Yu J, et al. (2017) Induction of antibodies directed against branched core O-mannosyl glycopeptides-selectivity complimentary to the ConA lectin. *Chemistry* 23:3466–3473.
40. Bartels MF, et al. (2016) Protein O-mannosylation in the murine brain: Occurrence of mono-O-mannosyl glycans and identification of new substrates. *PLoS One* 11:e0166119.
41. Yu J, et al. (2016) Distinctive MS/MS fragmentation pathways of glycopeptide-generated oxonium ions provide evidence of the glycan structure. *Chemistry* 22:1114–1124.
42. Yu J, Westerlind U (2014) Synthesis of a glycopeptide vaccine conjugate for induction of antibodies recognizing O-mannosyl glycopeptides. *ChemBioChem* 15:939–945.
43. Pett C, et al. (2018) Effective assignment of α2,3/α2,6-sialic acid isomers by LC-MS/MS-based glycoproteomics. *Angew Chem Int Ed Engl* 57:9320–9324.
44. Tsuji T, et al. (1983) The carbohydrate moiety of human platelet glycocalicin. *J Biol Chem* 258:6335–6339.
45. Bensing BA, Li Q, Park D, Lebrilla CB, Sullam PM (2018) Streptococcal Siglec-like adhesins recognize different subsets of human plasma glycoproteins: Implications for infective endocarditis. *Glycobiology* 28:601–611.
46. Seepersaud R, Sychantha D, Bensing BA, Clarke AJ, Sullam PM (2017) O-acetylation of the serine-rich repeat glycoprotein GspB is coordinated with accessory Sec transport. *PLoS Pathog* 13:e1006558.
47. Windwarder M, Altmann F (2014) Site-specific analysis of the O-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry. *J Proteomics* 108:258–268.
48. Cerdà-Costa N, Gomis-Rüth FX (2014) Architecture and function of metallopeptidase catalytic domains. *Protein Sci* 23:123–144.
49. Martínez-Sáez N, Peregrina JM, Corzana F (2017) Principles of mucin structure: Implications for the rational design of cancer vaccines derived from MUC1-glycopeptides. *Chem Soc Rev* 46:7154–7175.
50. Coltart DM, et al. (2002) Principles of mucin architecture: Structural studies on synthetic glycopeptides bearing clustered mono-, di-, tri-, and hexasaccharide glycodomains. *J Am Chem Soc* 124:9833–9844.
51. Sizemore S, Cicek M, Sizemore N, Ng KP, Casey G (2007) Podocalyxin increases the aggressive phenotype of breast and prostate cancer cells in vitro through its interaction with ezrin. *Cancer Res* 67:6183–6191.
52. Larrucea S, et al. (2008) Expression of podocalyxin enhances the adherence, migration, and intercellular communication of cells. *Exp Cell Res* 314:2004–2015.
53. Belisle JA, et al. (2010) Identification of Siglec-9 as the receptor for MUC16 on human NK cells, B cells, and monocytes. *Mol Cancer* 9:118.
54. Cannon JL, et al. (2008) CD43 regulates Th2 differentiation and inflammation. *J Immunol* 180:7385–7393.
55. Modak M, et al. (2016) Engagement of distinct epitopes on CD43 induces different co-stimulatory pathways in human T cells. *Immunology* 149:280–296.
56. Boles KS, Barchet W, Diacovo T, Cella M, Colonna M (2005) The tumor suppressor TSLC1/NECL-2 triggers NK-cell and CD8+ T-cell responses through the cell-surface receptor CRTAM. *Blood* 106:779–786.
57. Reiding KR, Bondt A, Franc V, Heck AJR (2018) The benefits of hybrid fragmentation methods for glycoproteomics. *TrAC Trends Anal Chem* 108:260–268.

58. North SJ, et al. (2010) Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *J Biol Chem* 285:5759–5775.
59. Fujitani N, et al. (2013) Total cellular glycomics allows characterizing cells and streamlining the discovery process for cellular biomarkers. *Proc Natl Acad Sci USA* 110:2105–2110.
60. Goh JB, Ng SK (2018) Impact of host cell line choice on glycan profile. *Crit Rev Biotechnol* 38:851–867.
61. Yu H, et al. (2017) Siglec-8 and Siglec-9 binding specificities and endogenous airway ligand distributions and properties. *Glycobiology* 27:657–668.
62. Karlsson NG, McGuckin MA (2012) O-Linked glycome and proteome of high-molecular-mass proteins in human ovarian cancer ascites: Identification of sulfation, disialic acid and O-linked fucose. *Glycobiology* 22:918–929.
63. Shurer CR, et al. (2017) Genetically encoded toolbox for glycocalyx engineering: Tunable control of cell adhesion, survival, and cancer cell behaviors. *ACS Biomater Sci Eng* 4:388–399.
64. Somers WS, Tang J, Shaw GD, Camphausen RT (2000) Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell* 103:467–479.
65. Kuroki K, et al. (2014) Structural basis for simultaneous recognition of an O-glycan and its attached peptide of mucin family by immune receptor PILRα. *Proc Natl Acad Sci USA* 111:8877–8882.
66. Varki A, Gagneux P (2012) Multifarious roles of sialic acids in immunity. *Ann N Y Acad Sci* 1253:16–36.
67. Macauley MS, Crocker PR, Paulson JC (2014) Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol* 14:653–666.
68. Sharma P, Allison JP (2015) Immune checkpoint targeting in cancer therapy: Toward combination strategies with curative potential. *Cell* 161:205–214.
69. Xiao H, Woods EC, Vukojicic P, Bertozzi CR (2016) Precision glycocalyx editing as a strategy for cancer immunotherapy. *Proc Natl Acad Sci USA* 113:10304–10309.
70. Fraschilla I, Pillai S (2017) Viewing Siglecs through the lens of tumor immunology. *Immunol Rev* 276:178–191.
71. Zeng Y, Ramya TNC, Dirksen A, Dawson PE, Paulson JC (2009) High-efficiency labeling of sialylated glycoproteins on living cells. *Nat Methods* 6:207–209.
72. Nicoll G, et al. (2003) Ganglioside GD3 expression on target cells can modulate NK cell cytotoxicity via siglec-7-dependent and -independent mechanisms. *Eur J Immunol* 33: 1642–1648.
73. Kingsley DM, Kozarsky KF, Hobbie L, Krieger M (1986) Reversible defects in O-linked glycosylation and LDL receptor expression in a UDP-Gal/UDP-GalNAc 4-epimerase deficient mutant. *Cell* 44:749–759.
74. Altschuler Y, et al. (2000) Clathrin-mediated endocytosis of MUC1 is modulated by its glycosylation state. *Mol Biol Cell* 11:819–831.
75. Koyama S, et al. (2016) Adaptive resistance to therapeutic PD-1 blockade is associated with upregulation of alternative immune checkpoints. *Nat Commun* 7:10501.
76. Freeman SA, et al. (2016) Integrins form an expanding diffusional barrier that coordinates phagocytosis. *Cell* 164:128–140.
77. Elola MT, et al. (2018) Galectins: Multitask signaling molecules linking fibroblast, endothelial and immune cell programs in the tumor microenvironment. *Cell Immunol* 333:34–45.
78. Wright CS, Jaeger J (1993) Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J Mol Biol* 232:620–638.