PNAS RESEARCH ARTICLE | CELL BIOLOGY

Steric pressure between glycosylated transmembrane proteins inhibits internalization by endocytosis

Sadhana Gollapudi[a], Sabah Jamal[a], Advika Kamatar[a], Feng Yuan[a], Liping Wang[b], Eileen M. Lafer[b], Brian Belardi[c], and Jeanne C. Stachowiak[a,d,1]

Edited by David Drubin, University of California Berkeley, Berkeley, CA; received September 15, 2022; accepted March 2, 2023

Clathrin-mediated endocytosis is essential for the removal of transmembrane proteins from the plasma membrane in all eukaryotic cells. Many transmembrane proteins are glycosylated. These proteins collectively comprise the glycocalyx, a sugar-rich layer at the cell surface, which is responsible for intercellular adhesion and recognition. Previous work has suggested that glycosylation of transmembrane proteins reduces their removal from the plasma membrane by endocytosis. However, the mechanism responsible for this effect remains unknown. To study the impact of glycosylation on endocytosis, we replaced the ectodomain of the transferrin receptor, a well-studied transmembrane protein that undergoes clathrin-mediated endocytosis, with the ectodomain of MUC1, which is highly glycosylated. When we expressed this transmembrane fusion protein in mammalian epithelial cells, we found that its recruitment to endocytic structures was substantially reduced in comparison to a version of the protein that lacked the MUC1 ectodomain. This reduction could not be explained by a loss of mobility on the cell surface or changes in endocytic dynamics. Instead, we found that the bulky MUC1 ectodomain presented a steric barrier to endocytosis. Specifically, the peptide backbone of the ectodomain and its glycosylation each made steric contributions, which drove comparable reductions in endocytosis. These results suggest that glycosylation constitutes a biophysical signal for retention of transmembrane proteins at the plasma membrane. This mechanism could be modulated in multiple disease states that exploit the glycocalyx, from cancer to atherosclerosis.

endocytosis | glycosylation | glycocalyx

Significance

Maintenance of the glycocalyx, the sugar-rich layer of the cell surface, is essential for cell–cell interactions and defense against pathogens. Glycosylated transmembrane proteins, like MUC1, which have covalently attached sugar chains, are major constituents of the glycocalyx and are often long lived on cell surfaces. However, all transmembrane proteins are subject to removal from the cell surface by endocytosis. How do glycosylated proteins escape endocytosis to maintain the glycocalyx? We use live-cell imaging in real time to examine endocytosis of glycosylated transmembrane proteins. Our data show that glycosylation increases the effective size of transmembrane proteins, making them substantially more difficult to internalize. This effect explains how glycosylated proteins accumulate at the plasma membrane, a key requirement for cellular health.

Internalization of membrane proteins by clathrin-mediated endocytosis (CME) is essential for diverse cellular functions including the modulation of receptor signaling pathways and recycling of transmembrane proteins (1–3). During the initiation of an endocytic structure, transmembrane proteins, which are the "cargo" of endocytic vesicles, are recruited when they bind to adaptor proteins such as AP2, which in turn recruit the clathrin coat (4–6). The resulting clathrin-coated structure grows and matures as more transmembrane proteins and adaptor proteins are recruited. Once the vesicle is fully formed, scission proteins such as dynamin cleave the neck of the clathrin-coated structure, allowing a clathrin-coated vesicle to bud into the cytoplasm (7).

Previous work has identified the biochemical determinants of membrane protein internalization by CME. These include specific amino acid motifs found within the cytoplasmic portions of transmembrane proteins, such as the YXXΦ and dileucine motifs, which are recognized the adaptor protein, AP2 (8–10). In addition to these biochemical factors, it is increasingly clear that the biophysical characteristics of transmembrane proteins also play important roles in modulating the extent of their internalization by CME. Specifically, a transmembrane protein's steric bulk (11), multimerization state (12, 13), and the extent to which it competes with other transmembrane proteins having similar biochemical internalization motifs (14, 15), can each have a substantial impact on its endocytosis. In particular, increasing the steric bulk of a transmembrane protein has been shown to proportionally reduce its recruitment into endocytic structures, owing to the limited capacity of these structures to accommodate transmembrane proteins (11).

One of the main factors that determines the steric bulk of a transmembrane protein is the degree to which it is glycosylated. There are two major types of glycosylation, N-linked and O-linked. In N-linked glycosylation, glycans are attached to asparagine residues (16–19), whereas in O-linked glycosylation, they are attached to serine and threonine residues (20–23). A negatively charged sialic acid glycan often terminates both N- and O-glycan structures (24). As one example of highly glycosylated cargo proteins, mucins, major constituents of the glycocalyx, are heavily O-glycosylated (25). Dysregulation of mucins is associated with multiple pathologies. For example, overexpression of mucins has been associated with lung diseases such as asthma, chronic obstructive pulmonary

Author contributions: S.G., B.B., and J.C.S. designed research; S.G., S.J., A.K., F.Y., L.W., E.M.L., B.B., and J.C.S. performed research; S.G., E.M.L., B.B., and J.C.S. analyzed data; and S.G., E.M.L., B.B., and J.C.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jcstach@austin.utexas.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2215815120/-/DCSupplemental.

Published April 6, 2023.

disease (COPD), cystic fibrosis, and some cancers (9, 26–29). In particular, lower levels and aberrant forms of mucin glycosylation are common features of tumor cells. Toward a better understanding of the role of endocytosis in such processes (1–3), here we probe the impact of glycosylation on the internalization of transmembrane proteins by CME.

To study the effect of glycosylation on the endocytosis of transmembrane proteins, we studied variants of mucin 1 (MUC1), a heavily glycosylated transmembrane mucin that is known to be taken up by CME (30–32). In particular, MUC1 is a single-pass type I transmembrane protein. It has a transmembrane domain that tethers it to the plasma membrane, a 72-amino acid cytoplasmic tail, and a heavily O-glycosylated ectodomain. The ectodomain contains a variable number of 20-amino acid TRs (33). These TRs are rich in threonine and serine residues, which are O-glycosylated (21). In vivo, MUC1 has many isoforms, each containing a different number of TRs. A MUC1 membrane protein with a greater number of TRs has more potential sites for glycosylation. Importantly, glycosylation contributes significantly to the molecular weight of MUC1. Specifically, each MUC1 tandem repeat has a molecular weight of about 2 kDa and has 5 serine and threonine residues available for glycosylation. The glycans attached to each site are typically 6 to 7 monosaccharides in length and have a molecular weight ranging from 500 to 1,300 Da. Therefore, if every site on the tandem repeat were glycosylated, its molecular weight would increase by twofold to fivefold (34–37).

Glycosylation of MUC1 influences the hydrodynamic radius of its tandem repeat domain (38). In the absence of glycosylation, the persistence length of a peptide chain, which is the approximate distance over which it can curve, is less than a nanometer (39). In contrast, the persistence length of the glycosylated MUC1 tandem repeat domain has been estimated at 7 to 8 nm (38), presumably owing to steric clashes among the O-glycans and electrostatic repulsion among sialic acid residues at their termini. Because hydrodynamic radius increases with increasing persistence length, glycosylation is expected to substantially increase the hydrodynamic radius of MUC1's tandem repeat domain.

Early work using bulk assays showed that endocytosis of MUC1 was significantly higher in glycosylation-deficient CHO cells in comparison to wild-type CHO cells (31). More recently, it has been reported that overexpression of MUC1 is capable of crowding the surface of the plasma membrane of mammalian cells, producing steric pressure that induced spontaneous assembly of finger-like membrane protrusions (38). Further, the incidence of these protrusions was found to increase with increasing MUC1 expression level, consistent with a density-based steric effect. Similarly, another recent paper showed that glycosylated MUC1 proteins were enriched in regions of high outward membrane curvature, where their steric bulk may be more easily accommodated, avoiding areas of inward membrane curvature, such as endocytic structures (40). These findings collectively suggest that the steric bulk associated with glycosylation enables transmembrane proteins to escape endocytosis. However, the mechanisms responsible for this phenomenon remain unknown, largely because the effect of glycosylation on the dynamics and content of endocytic structures has never been examined. Here, we use live-cell imaging to study a large ensemble of individual clathrin-mediated endocytic events with the goal of understanding the impact of glycosylation on endocytosis.

Results

Glycosylated Transmembrane Fusion Proteins Partition Weakly into Clathrin-Coated Structures. To evaluate the effect of glycosylation on the endocytosis of transmembrane proteins, we designed chimeric membrane proteins consisting of the N-terminal ectodomain of MUC1 fused to green fluorescent protein (GFP), fused to the transmembrane and intracellular domains of the transferrin receptor (TfR), (Fig. 1 *A*–*D*). We chose the intracellular and transmembrane domains of TfR because TfR's internalization by CME is strong and well characterized (8, 41–43). Notably, the native intracellular domain of MUC1 contains a YXXΦ motif that mediates internalization by the clathrin pathway (30, 31). Owing to diffraction-limited blurring, the extent of colocalization between native MUC1 and endocytic structures was barely above the noise level in our images (*SI Appendix*, Fig. S1). Therefore, to promote strong recruitment of proteins into endocytic structures that we could more easily study and modulate, we created chimeric "transmembrane fusion proteins" consisting of the tandem repeat ectodomain of MUC1 and the intracellular and transmembrane domains of TfR.

Using this approach, we generated transmembrane fusion proteins with ectodomains that contained different numbers of tandem repeats (TRs) in the ectodomain as follows: zero (0TR), two (2TR), five (5TR), or 10 TRs (10TR) (Fig. 1 *A*–*D*). The illustrations of the transmembrane fusion proteins in Fig. 1 *A*–*D* are drawn to scale based on previously determined estimates of the radius of gyration of MUC1 tandem repeat domains (44). Notably, the transmembrane fusion proteins only include the intracellular and transmembrane domains of the TfR, such that any glycosylation normally associated with the extracellular domain of the TfR is not present in these chimeras. Each of the transmembrane fusion proteins was separately expressed in retinal pigmented epithelial cells (RPE). The cells also stably expressed clathrin light chain tagged with mCherry, for the visualization of clathrin-mediated endocytic structures (45). RPE cells are commonly used in studies of endocytosis due to their large and well-spread lamellipodia, which enable visualization of the plasma membrane (46). The plasma membranes of these cells, proximal to the coverslip surfaces on which they were cultured, were imaged using spinning disk confocal microscopy. First, we examined the plasma membrane of live RPE cells expressing the zero tandem repeat protein (0TR). Fig. 1*E* shows that the plasma membrane has a punctate appearance in the transmembrane protein channel. In particular, the images showed strong colocalization of the transmembrane fusion protein (GFP) with clathrin (mCherry), suggesting that the 0TR protein was incorporated into clathrin-coated structures.

Similarly, we examined confocal images of cells expressing the 2TR, 5TR, and 10TR fusion proteins (Fig. 1 *F*–*H*). Interestingly, as the number of TRs increased, the intensity of the fusion proteins (GFP) within puncta that colocalized with clathrin (mCherry) appeared to decrease relative to the surrounding plasma membrane intensity in the fusion protein channel (Fig. 1 *F*–*H*). This observation suggests that the presence of an ectodomain of increasing molecular weight may have opposed endocytosis of the transmembrane fusion proteins.

Why might glycosylation oppose endocytosis of transmembrane proteins? We considered two distinct hypotheses. First, we considered whether reduced endocytosis of glycosylated membrane proteins could be due to slower recruitment of the proteins into endocytic structures within the brief 20 to 120-s lifetime of each endocytic event. Second, we investigated the impact of steric interactions between transmembrane proteins on endocytosis. Specifically, we asked whether the increased steric bulk of transmembrane proteins with larger numbers of TRs could result in reduced endocytosis owing to limited space available within endocytic structures. We present our findings with respect to each of these hypotheses in the following sections.

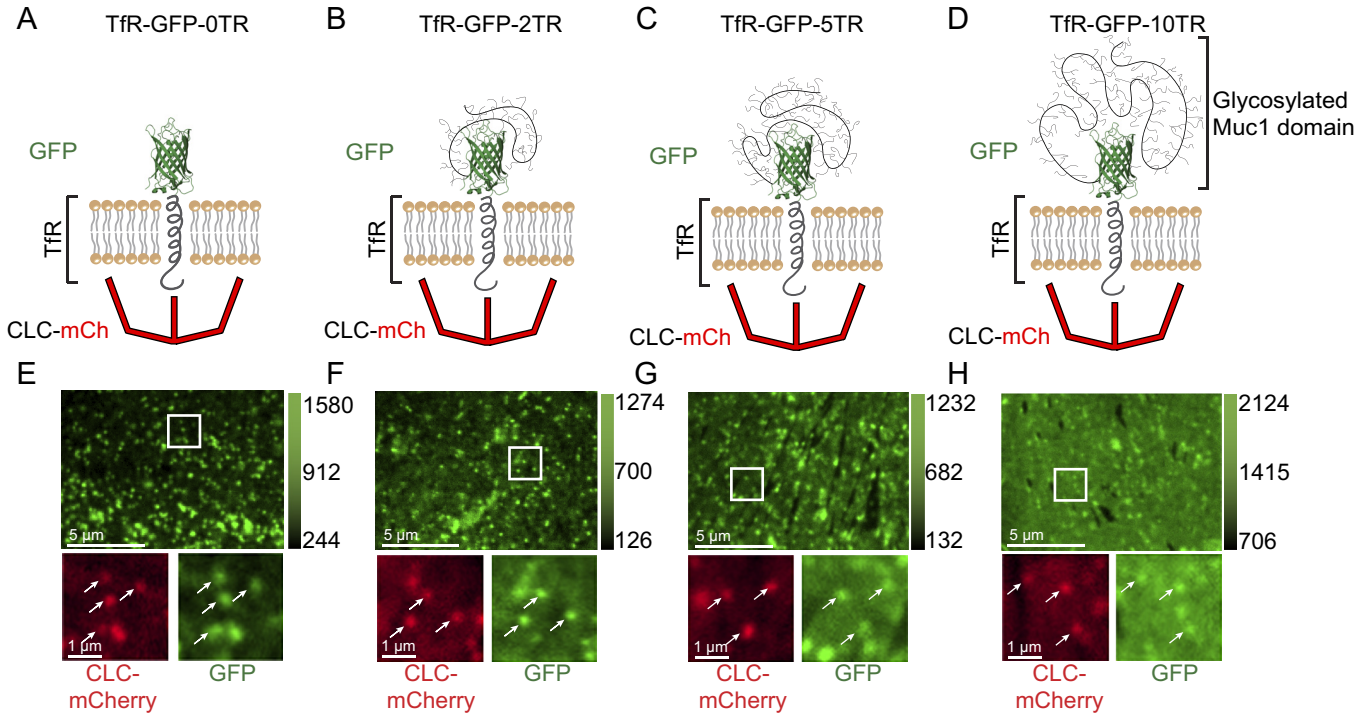

Fig. 1. Glycosylated transmembrane fusion proteins partition weakly into clathrin-coated structures. (*A*–*D*) A schematic of transmembrane fusion proteins with an incremental increase in the number of MUC1 tandem repeats from (*A*) zero tandem repeats, (*B*) two tandem repeats, (*C*) five tandem repeats, (*D*) ten tandem repeats. (*E*–*H*) Spinning disk confocal images of the plasma membrane of RPE cells transiently expressing the transmembrane fusion proteins with (*E*) zero tandem repeats, (*F*) two tandem repeats, (*G*) five tandem repeats, (*H*) 10 tandem repeats. The white box in the top image indicates the location of the smaller *Insets*.

The Timescale of Protein Loading into Clathrin-Coated Structures Is Unaffected by MUC1's Ectodomain. To evaluate the impact of MUC1's ectodomain on the dynamics of transmembrane fusion protein recruitment into endocytic structures, we imaged the recruitment of fusion proteins into growing endocytic structures in real time using TIRF (total internal reflection fluorescence) microscopy. TIRF microscopy is a preferred technique for tracking the dynamics of endocytic structures because the evanescent field of the internally reflected beam has a shallow penetration depth (~100 nm) that illuminates the plasma membrane, while largely excluding fluorescence intensity originating from the cellular cytoplasm and organelles (46, 47). Using this approach, we collected images every two seconds for a total of 10 min. These image series were collected in the same two fluorescent channels used in Fig. 1: i) clathrin light chain (mCherry) and ii) transmembrane fusion protein (GFP). The time series was analyzed to identify and track individual endocytic events, from initiation to departure (Fig. 2 *B* and *D*). Specifically, we used a publicly available algorithm, CMEAnalysis (46) to detect and track the fluorescent intensities of hundreds of endocytic structures per cell, where the clathrin light chain signal was the "master channel" used to identify endocytic structures, and the transmembrane fusion protein channel was the "subordinate channel", from which the intensity of each structure, relative to the local background signal, was estimated (46). To quantify transmembrane fusion protein partitioning within the images, we used CMEAnalysis. Furthermore, this software detects clathrin-coated structures by fitting a 2D Gaussian function to the fluorescent puncta in the clathrin-light chain channel (mCherry). Once these puncta were detected in the mCherry channel, a 2D Gaussian function was fit to the corresponding fluorescent puncta in the transmembrane fusion protein channel (GFP). The amplitudes from these fits were used to estimate the relative concentration of clathrin-light chain and transmembrane fusion proteins within each clathrin-coated structure. Next, the fluorescence intensity surrounding the detected structure was averaged to estimate the relative concentration of the proteins at the plasma membrane. Notably, *SI Appendix*, Fig. S2 shows the fraction of detected endocytic structures with transmembrane fusion protein fluorescence below the detection threshold increases as the number of TRs on the protein ectodomain increases. Finally, each detected clathrin-coated structure was tracked over its lifetime on the plasma membrane by linking the corresponding locations of each detection between consecutive frames.

Clathrin-mediated endocytic events have a broad range of lifetimes at the plasma membrane, from tens of seconds to minutes, with most structures lasting less than 120 s (12, 46). Therefore, we grouped endocytic events into cohorts based on their lifetimes at the plasma membrane, the time from appearance to disappearance. The cohorts included 10 to 19 s, 20 to 39 s, 40 to 59 s, 60 to 79 s, 80 to 99 s, and 100 to 120 s. The distribution of clathrin-coated structures across these cohorts was not substantially different between cells expressing either the 0TR or the 10TR fusion proteins, suggesting that the presence of MUC1's ectodomain did not shift the underlying dynamics of CME (Fig. 2*E*).

To evaluate the dynamics of transmembrane fusion protein entry into endocytic structures, we plotted the intensity of endocytic structures over time, for both clathrin light chain (mCherry) and the fusion protein (GFP), during individual endocytic events. For ease of comparison of dynamics across conditions, we averaged the intensity profiles of all clathrin-coated structures with lifetimes ranging 10 to 120 s. In Fig. 2 *F* and *G*, these intensity profiles were plotted over a percentage of the clathrin-coated structure's lifetime. In these plots, we observed that the intensity of the clathrin signal increased during the first 30% of the average structure's lifetime, remained relatively constant for the next 40% of its lifetime, and then decreased during the final 30% of its lifetime. A similar pattern was observed for the intensity of the transmembrane fusion protein at endocytic structures. Specifically, for endocytic structures

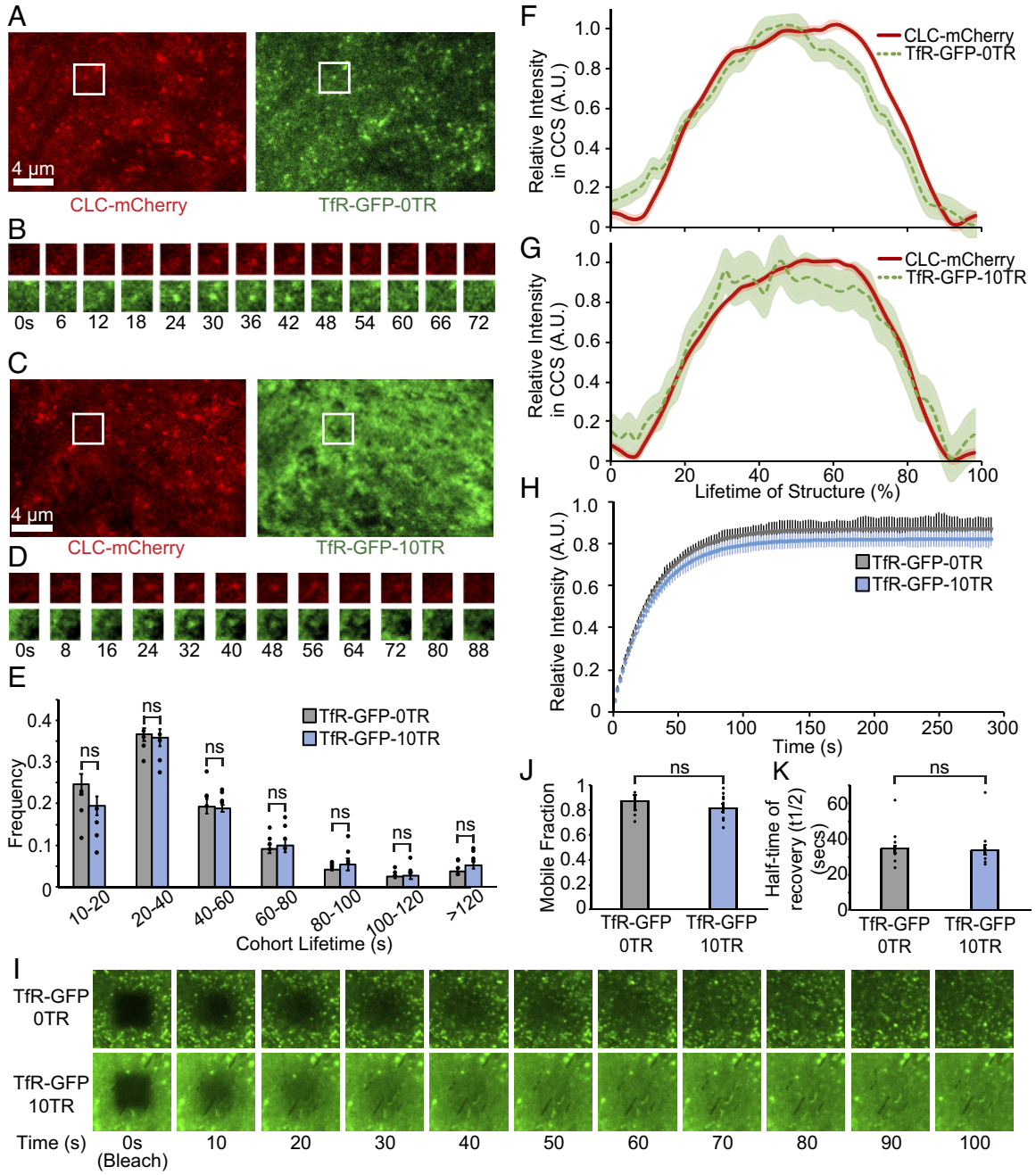

Fig. 2. The dynamics of protein loading into clathrin-coated structures are unaffected by MUC1's ectodomain. (*A* and *C*) TIRF microscopy images of the plasma membrane of RPE cells transiently expressing either (*A*) TfR-GFP-0TR or (*C*) TfR-GFP-10TR. (*B* and *D*) A sequence of images showing the maturation of one CCP tracked through its lifetime on the plasma membrane in cells expressing either (*B*) TfR-GFP-0TR or (*D*) TfR-GFP-10TR. (*E*) The lifetime distribution of clathrin-coated structures for cells expressing either TfR-GFP-0TR or TfR-GFP-10TR. The data points represent individual cells. The average lifetime of all CLC-mCherry positive tracks was 37.0 s for cells expressing TfR-GFP-0TR and 39.7 s for TfR-GFP-10TR. A two-tailed *t* test confirmed that the difference is statistically not significant ($P > 0.05$) between these data. (*F* and *G*) Fluorescence intensity of clathrin-coated structures tracked over their lifetime on the plasma membrane in RPE cells expressing either (*F*) TfR-GFP-0TR or (*G*) TfR-GFP-10TR. The red curves track the fluorescence signal of mCherry-tagged clathrin light chain and the green curves track the GFP-tagged transmembrane fusion proteins. Seventeen cells were analyzed resulting in 8143 CCSs tracked in (*F*), and 16 cells were analyzed resulting in 6,698 CCS tracked in (*G*). The shaded areas represent the mean ± SE for the intensity profiles of CLC-mCherry and transmembrane fusion protein. (*H*) FRAP recovery curves for multiple cells were averaged for both the conditions, TfR-GFP-0TR and TfR-GFP-10TR. Data from 12 cells were averaged for TfR-GFP-0TR, and data from 13 cells were averaged for TfR-GFP-10TR. A two-tailed *t* test was conducted for the best-fit values of the mobile fraction and half-time of recovery. *P*-value >0.05 for half-time of recovery (*P*-value = 0.95) and mobile fraction (*P*-value = 0.43) suggesting that difference between the datasets is statistically not significant. (*I*) Image series of fluorescence recovery at the plasma membrane of cells expressing TfR-GFP-0TR (*Top*) or TfR-GFP-10TR (*Bottom*). A square region of 5.3 μm on each side was bleached, and fluorescence recovery was tracked over 5 min. (*J*) Bar plots of the averaged mobile fraction and (*K*) half-time of recovery for TfR-GFP-0TR and TfR-GFP-10TR.

taken from cells expressing either the 0TR or 10TR fusion proteins, the intensity in the fusion protein channel also reached its steady-state value within the first 30% of its lifetime, similar to the rise in intensity in the clathrin channel. For both fusion proteins, it is clear that the steady-state intensity in the fusion protein channel was reached well before the clathrin signal began to decrease. These results suggest that entry of transmembrane fusion proteins into growing endocytic structures is a rapid process with a timescale substantially less than the time required for growth and maturation of endocytic structures. Thus, a dynamic equilibrium likely exists

between the population of transmembrane fusion proteins within clathrin-coated structures and the population on the surrounding plasma membrane, consistent with our previous work (11). This dynamic equilibrium appears to be established within the early stages of initiation of clathrin-coated structures, suggesting that these structures are filled to their equilibrium capacity with transmembrane proteins well before they are ready to depart from the plasma membrane surface.

To examine mobility of the transmembrane fusion proteins on the plasma membrane, fluorescence recovery after photobleaching (FRAP) measurements were made. The plasma membrane of RPE cells expressing either the 0TR or 10TR fusion proteins was photobleached, and fluorescence recovery was tracked (Fig. 2 *H* and *I*). The FRAP measurements showed that the recovery fraction, or mobile fraction, and the recovery time were very similar (Fig. 2 *J* and *K*) for both fusion proteins. This result suggests that the mobility of the transmembrane fusion proteins on the plasma membrane was not substantially impacted by the presence of the MUC1 ectodomain. Notably, these results are in line with the Saffman–Delbrück model, which predicts that the diffusion constant of a transmembrane protein should scale with the size of the transmembrane domain rather than that of the ectodomain (48).

Taken together, these data suggest that the reduced recruitment of the 10TR fusion protein into endocytic structures, relative to 0TR, cannot be explained by any of the following factors related to cellular dynamics: i) reduced mobility of 10TR on the membrane surface, ii) slower diffusion of 10TR into endocytic structures, or iii) altered dynamics of endocytosis in cells expressing 10TR.

Recruitment of Transmembrane Fusion Proteins into Clathrin-Coated Structures Decreases as the Molecular Weight of the Glycosylated Ectodomain Increases. If the reduced recruitment of the 10TR fusion protein relative to 0TR cannot be explained by slower dynamics on the plasma membrane surface, then perhaps the greater steric bulk of 10TR could be responsible for the reduction. To investigate this hypothesis, we measured the steady-state partitioning of each of the transmembrane fusion proteins (0TR, 3TR, 5TR, and 10TR) between endocytic structures and the surrounding plasma membrane. To make this measurement, we expressed each of the fusion proteins in RPE cells and acquired images of the plasma membrane at a single time point (Fig. 1 *E*–*H*).

To quantify transmembrane fusion protein partitioning within the images, we used CMEAnalysis (46). As previously described, this software fits a 2D Gaussian to puncta in the clathrin-light chain channel. Next, it fits a 2D Gaussian in the corresponding location in the transmembrane fusion protein channel. It also produces statistics such as the amplitude of the gaussian fit. These amplitudes represent raw, nonnormalized values and can therefore be quantitatively compared between transmembrane fusion proteins in the same plot. The amplitude of the fit is interpreted to be roughly proportional to the number of fusion proteins per endocytic structure. Using these data, we plotted the relative number of fusion proteins per endocytic structure as a function of the relative concentration of the fusion protein on the surrounding plasma membrane. In the resulting plots, the relative number of fusion proteins within each clathrin-coated structure initially increased with an increase in the relative concentration of fusion proteins on the surrounding plasma membrane (Fig. 3*A*). Eventually the relative number of transmembrane fusion proteins within clathrin-coated structures began to plateau toward a maximum value. This maximum value represents the relative number of transmembrane fusion proteins that are required to saturate a clathrin-coated structure, as described previously (11).

From these data, it is clear that the saturated capacity of clathrin-coated structures for transmembrane fusion proteins declined considerably as the number of TRs increased from 0 to 10. To estimate the saturated capacity, we applied a simple physical model, which we reported previously (11), Eq. **1**. This model describes the loading of transmembrane proteins into endocytic structures as a simple, multivalent binding problem, where the average number of fusion proteins per structure, <n>, depends on the saturated capacity per endocytic structure (N_{max}), the relative concentration of fusion proteins on the surrounding plasma membrane (C_{mem}), and the dissociation constant of binding between the fusion protein and the endocytic structure (K_{deff}).

$$< n > = \frac{NmaxCmem}{Kdeff + Cmem}. \quad [1]$$

In line with our findings in Fig. 2, this model assumes that the number of fusion proteins per endocytic structure is determined by a dynamic equilibrium between the population of fusion proteins inside and outside the structure. We applied equation Eq. **1** to the data in Fig. 3*A* for the 0TR fusion protein, leaving both N_{max} and K_{deff} as free parameters. We assume that K_{deff} should have the same value for each of the four fusion proteins (10TR, 5TR, 2TR, and 0TR), because they each display the same binding domain for the endocytic machinery. In contrast, we expected N_{max} to decrease as the number of TRs increased, owing to the increased bulk of the transmembrane fusion protein. Therefore, we held the value of K_{deff}, determined from fitting the 0TR data, constant, and fit the data for the remaining transmembrane fusion proteins (2TR, 5TR, and 10TR), with N_{max} as the only free parameter (Fig. 3*A*). Notably, the horizontal axis in Fig. 3*A* represents the local, background-subtracted fluorescence intensity at the membrane surface immediately surrounding each punctum. We interpret this intensity as being roughly proportional to the local expression level of transmembrane fusion proteins. Data for each fusion protein are plotted over the same range, ensuring that the range of expression levels in all experiments is the same. Fig. 3*B* shows the resulting values of N_{max}, which declined approximately fivefold as the number of TRs increased from 0 to 10, with more modest reductions for 2TR and 5TR, relative to 0TR. Notably, the fluorescence intensity of TfR-GFP-10TR within endocytic structures was just above the threshold for reliable detection by CMEAnalysis. Therefore, we did not attempt to quantity the recruitment of transmembrane fusion proteins containing more than 10 TRs. In vivo, MUC1 can have as many as 42 TRs. Based on our results, we would expect these larger ectodomains to further restrict the uptake of MUC1 by endocytosis. Furthermore, we cannot rule out the possibility of coupling between N_{max} and K_{deff}, perhaps through electrostatic effects, yet our data are reasonably well fit by assuming them to be independent. These results demonstrate that increasing the number of TRs in the ectodomain of the transmembrane fusion protein results in a reduced ability of endocytic structures to accommodate the proteins, likely owing to a corresponding increase in the steric bulk, as depicted in Fig. 3*C*.

Notably, owing to the moderate expression levels of the transmembrane fusion proteins used in these experiments, we do not expect the chain to substantially straighten due to steric pressure. Therefore, we have not accounted for crowding induced changes to N_{max}. However, we cannot rule out that steric pressure may influence the true capacity of endocytic structures for transmembrane proteins.

Furthermore, we made a similar observation for the native MUC1 transmembrane protein, where a decrease of approximately twofold in the uptake of MUC1-GFP-10TR was measured

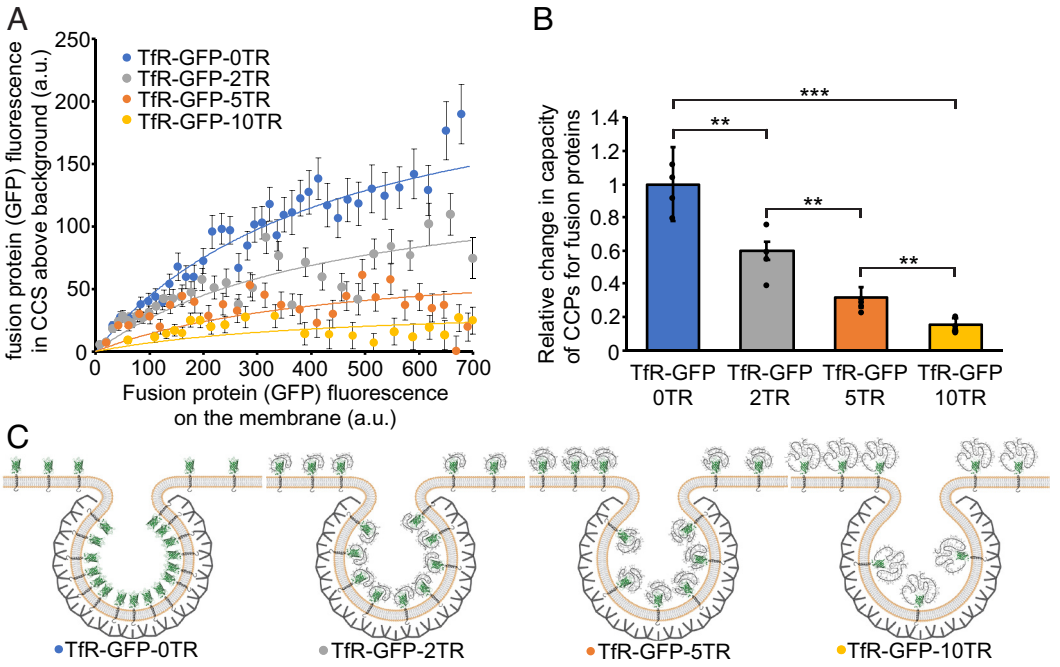

Fig. 3. Recruitment of transmembrane fusion proteins into clathrin-coated structures decreases as the molecular weight of the glycosylated ectodomain increases. (*A*) The relative number of transmembrane fusion proteins within clathrin-coated structures is plotted as a function of the relative concentration of fusion proteins on the plasma membrane surrounding each structure. Each point represents the average of data from 200 clathrin-coated structures binned by the relative concentration of the transmembrane fusion protein on the membrane. A total of 10,123 CCSs were detected from 88 cells expressing, TfR-GFP-0TR, 9,964 CCSs from 101 cells expressing TfR-GFP-2TR, 11,690 CCSs were detected from 75 cells expressing TfR-GFP-5TR, and 18,420 CCSs were detected from 80 cells expressing TfR-GFP-10TR. Error bars represent mean ± SE. Solid lines are model predictions using the best-fit values of K_{deff} and N_{max}. (*B*) Bar plot of relative CCS capacities for each of the transmembrane fusion proteins. The individual data points represent separate samples. The error bars represent 95% CI of the best-fit values of N_{max}. A two-sample *t* test was conducted on the model-predicted values of N_{max}. *P*-values were <0.05 between each pair of fusion proteins suggesting a statistically significant difference between their N_{max} values. $^{**}P < 0.01$, $^{***}P < 0.001$ (*C*) Cartoon schematic illustrating the decreased capacity of CCSs as the number of MUC1 tandem repeats increases.

compared to MUC1-GFP-0TR (*SI Appendix*, Fig. S3). These results confirm that glycosylated TRs on the MUC1 ectodomain impact its recruitment into endocytic structures. Notably, in line with these results, the apparent lower affinity of native MUC1 for endocytic structures in comparison to our transmembrane fusion proteins does not imply that native MUC1 proteins are free from the influence of steric pressure, which is present throughout the plasma membrane surface. Instead, we interpret the low copy number of native MUC1 proteins at endocytic structures to indicate that native MUC1 is largely outcompeted, through a combination of steric and biochemical contributions, by the myriad of other transmembrane proteins present at these sites (11). Specifically, the low affinity of native MUC1 for endocytic sites makes its internalization more vulnerable to steric exclusion as the size of its tandem repeat domain increases. Nonetheless, as noted above, the signal to background ratio in experiments with native MUC1 was barely above the threshold for detection, *SI Appendix*, Fig. S3. Therefore, we chose to use the TfR fusion proteins introduced above to study the mechanism by which the tandem repeat domain inhibits localization of transmembrane proteins to endocytic structures.

Taking together the results from Fig. 3, both the increasing length of the tandem repeat protein backbone and its increasing potential for glycosylation could contribute to the steric bulk of the transmembrane fusion proteins. We next sought to determine the extent to which each of these factors inhibit recruitment of fusion proteins into endocytic structures.

A Model Receptor/Ligand System Can Be Used to Compare Tandem Repeat Domains Derived from Bacteria and Mammalian Cells. To distinguish the relative impacts on endocytosis of i) the peptide backbone of the tandem repeat domain, and ii) the glycosylation of the tandem repeat domain, we devised a strategy to generate tandem repeat domains with and without glycosylation. In particular, because bacteria lack O-glycosylation machinery (49–52), tandem repeat domains produced in *Escherichia Coli.* should have little or no glycosylation in comparison to tandem repeat domains expressed in mammalian cells. To compare domains produced in bacterial vs. mammalian hosts, we expressed the tandem repeat domains as soluble "ligands", which bound to a model "receptor" expressed on the plasma membrane surfaces of RPE cells. Similar to the 0TR fusion protein described above, the model receptor consisted of the intracellular and transmembrane domains of the TfR, followed by a blue fluorescent protein (BFP) domain. Additionally, the C terminus of the model receptor was fused to a single-domain antibody against GFP (53), such that the receptor was capable of recruiting GFP-tagged "ligands" to the plasma membrane surface. As a control for the absence of the tandem repeat domain, GFP alone was used as the ligand (Fig. 4*A*). This control ligand was produced in bacteria as described in the methods section. To test the impact of the tandem repeat domain, a model ligand consisting of an N-terminal GFP domain fused to the 10 tandem repeat domain of MUC1 was used. When produced in bacteria, to avoid glycosylation, we refer to this ligand as bact-GFP-10TR (Fig. 4*B*). Here, the absence of significant glycosylation was confirmed by mass spectrometry (*SI Appendix*, Fig. S5). When glycosylation was desired, we produced the ligand by coexpressing it in RPE cells alongside the model receptor. The resulting ligand, mam-GFP-10TR, was secreted by RPE cells into the extracellular solution, where it was free to bind to the model receptor on the outer cell surface (Fig. 4*C*).

Fig. 4 *D*–*F* shows fluorescent images of the model ligands, GFP, bact-GFP-10TR, and mam-GFP-10TR, recruited to the surfaces of RPE cells by the model receptor. For each ligand, clear colocalization with the model receptor (BFP) was observed, suggesting

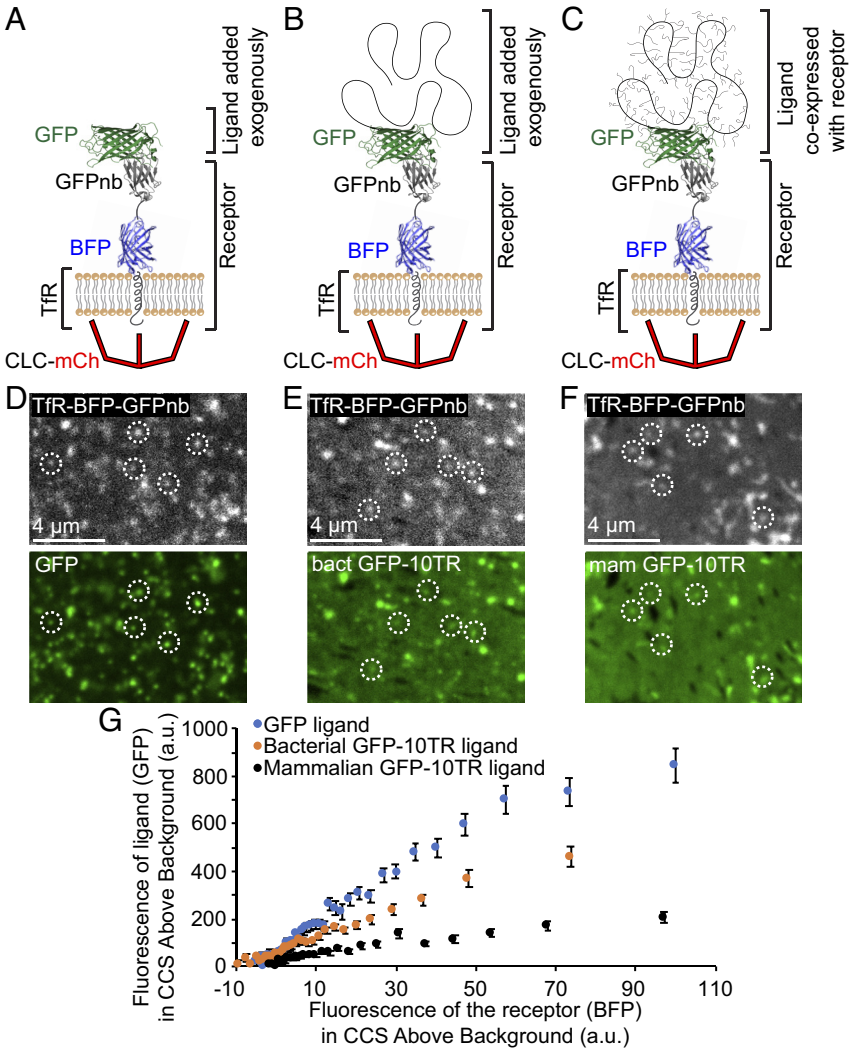

Fig. 4. A model receptor/ligand system can be used to compare tandem repeat domains derived from bacteria and mammalian cells. (*A*–*C*) A Schematic of the model receptor bound to the ligands: (*A*) GFP, (*B*) bact-GFP-10TR, or (*C*) mam-GFP-10TR. (*D*–*F*) Fluorescent images of the plasma membrane of RPE cells expressing the model receptor, TfR-BFP-GFPnb (the *Top*), and the ligands: (*D*) GFP, (*E*) bact-GFP-10TR, and (*F*) mam-GFP-10TR (*Bottom*). The images show strong colocalization between the ligands and the model receptors. (*G*) Relative intensity of puncta in the ligand channel, plotted versus the intensity of corresponding puncta in the receptor channel. Each point represents the average of 200 puncta, binned by the intensity of puncta in the model receptor channel. Error bars represent mean ± SE.

that the ligands were recruited to the cell surface by the GFP–nanobody interaction, as expected. To further confirm recruitment of the ligands to the model receptor, Fig. 4*G* plots the intensity at endocytic structures (mCherry-positive puncta) in the ligand channel, relative to the intensity of the same puncta in the model receptor channel. Each of these trends displays a clear positive slope, indicating that endocytic structures with a greater number of model receptors recruited a greater number of ligands, as expected. Notably, the slope is somewhat higher for the GFP ligand, compared to the bact-GFP-10TR, which has a greater slope in comparison to the mam-GFP-10TR ligand. This finding suggests that the presence of the tandem repeat domain, and its glycosylation, may lower the receptor–ligand affinity, likely owing to steric inhibition. Nonetheless, all three ligands were strongly recruited to endocytic structures on the cell surface by the model receptor. To prevent the apparent differences in the binding affinity of the ligands and the effective concentrations of the ligand binding the membrane from impacting our conclusions, the analysis in the following sections (Figs. 5 and 6) compares groups of cells and endocytic structures with equivalent ligand binding, rather than equivalent expression of the model receptor.

Tandem Repeat Domains Purified from Bacterial Cells Have Little or No Glycosylation in Comparison to Those Excreted by Mammalian Cells. Having confirmed recruitment of the ligands to endocytic structures containing the model receptor, we further confirmed the in-situ glycosylation of the recruited ligands by staining the cell surface with peanut agglutinin (PNA). PNA binds specifically to galactose residues, which are abundant in O-linked glycans, such as those on the MUC1 TRs (54, 55). Specifically, RPE cells expressing the model receptor were first exposed to one of three ligands, either by addition to the culture (GFP, bact-GFP-10TR) or by coexpression with the model receptor (mam-GFP-10TR). Then PNA-Alexa 647 was added to the culture, where it stained the surfaces of the cell. Fig. 5 *A*–*D* shows images of the plasma membrane surface in the ligand (GFP), and PNA (Alexa 647) channels. As a positive control, cells expressing the 10TR fusion protein (GFP) are also included (Fig. 5*D*). Because 10TR is expressed entirely within the mammalian RPE cells, we expect it to be O-glycosylated. To compare PNA recruitment among these conditions, cells with similar levels of GFP fluorescence at the plasma membrane were imaged and compared. The images indicate that cells expressing 10TR (positive control) recruited

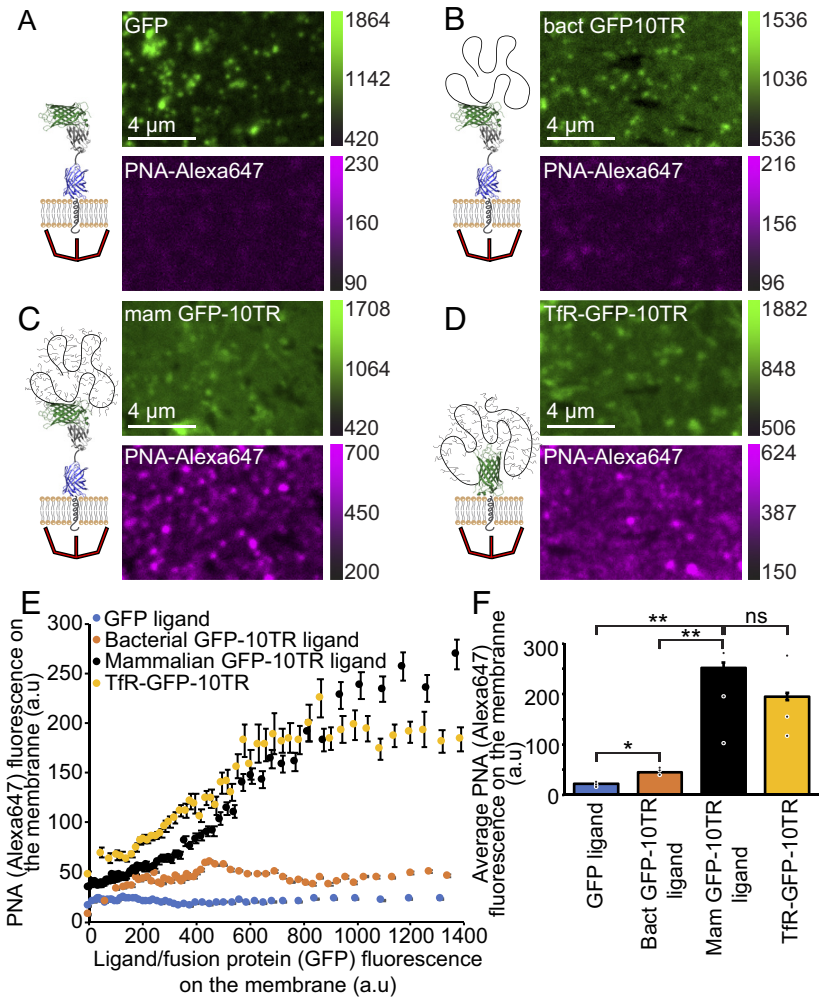

Fig. 5. Tandem repeat domains purified from bacterial cells have little or no glycosylation in comparison to those excreted by mammalian cells. (*A–D*) Spinning disk images of the plasma membrane of RPE cells stained with PNA-Alexa647 and transiently expressing the chimeric model receptor TfR-BFP-GFPnb and incubated with either (*A*) GFP ligand, (*B*) bact-GFP-10TR ligand or (*C*) coexpressed with mam-GFP-10TR. (*D*) Spinning disk image of the plasma membrane of RPE cells stained with PNA-Alexa647 and transiently expressing the transmembrane fusion protein TfR-GFP-10TR. (*E*) Plot showing the amount of PNA-Alexa647 staining the plasma membrane for each of the ligands (*A–D*), as a function of the local fluorescence of the ligands or 10TR fusion protein on the plasma membrane around the clathrin-coated structures. Each point on the plot represents the average of 200 clathrin-coated structures binned by the local membrane concentration of the proteins. A total of 12,013 CCSs were detected from 79 cells incubated with GFP ligand, 11,658 CCSs were detected from 70 cells incubated with bact-GFP-10TR, 13,861 CCSs were detected from 111 cells expressing mam-GFP-10TR, and 10,636 CCSs were detected from 87 cells expressing TfR-GFP-10TR. Error bars represent mean ± SE. (*F*) Bar plot representing the average fluorescence of PNA-Alexa647 on the membrane surrounding all clathrin-coated structures with local ligand or 10TR fusion protein fluorescence on the membrane greater than the median value (700 a.u.). The individual data points represent separate samples. The error bars represent mean ± SE. A two-sample *t* test was done on the average values of the PNA fluorescence on the membrane for each pair of ligands. *P*-value was = 0.19 for PNA average values of mam GFP-10TR ligand and TfR-GFP-10TR, suggesting that the difference between the datasets is statistically not significant. *P*-values were <0.05 when comparing the other pairs of ligands, indicating a statistically significant difference between their average PNA fluorescence. $*P < 0.05$, $**P < 0.01$.

a substantially greater amount of PNA in comparison to cells recruiting the GFP ligand (negative control), compare Fig. 5 *A* and *D*. This result suggests that overexpression of the 10TR fusion protein substantially increased the incidence of O-glycosylation at the plasma membrane surface, in agreement with previous studies in which the MUC1 tandem repeat domain was overexpressed (38). In comparison, cells that recruited the bact-GFP-10TR ligand bound low levels of PNA (Fig. 5*B*), similar to the negative control, while cells that recruited mam-GFP-10TR bound substantially higher levels of PNA (Fig. 5*C*), approaching that of the positive control.

To quantify the extent of PNA recruitment by the ligands and the 10TR fusion protein, we analyzed our images and plotted the concentration of PNA (Alexa647) versus the concentration of the ligand or 10TR fusion protein (GFP) on the plasma membrane surrounding endocytic structures (Fig. 5*E*). Here, the positive control (10TR) had a substantially higher slope relative to the negative control (GFP ligand), demonstrating that, for a given concentration at the cell surface, 10TR recruited substantially more PNA in comparison to the GFP ligand. The slope of the corresponding curve for cells displaying the bact-GFP-10TR ligand is similar to that of the negative control, while cells displaying the mam-GFP-10TR ligand produced a curve with a substantially higher slope, similar to that of the positive control. Taken together, these results suggest that the GFP-10TR ligand bares a substantial degree of O-glycosylation when produced in mammalian cells (mam-GFP-10TR) and little or no glycosylation when produced in bacteria (bact-GFP-10TR). Similarly, we confirmed the glycosylation of the transmembrane fusion proteins used in the assays in Figs. 1–3. Specifically, we found a linear increase in PNA staining with an increasing number of TRs in the transmembrane fusion proteins (*SI Appendix*, Fig. S4

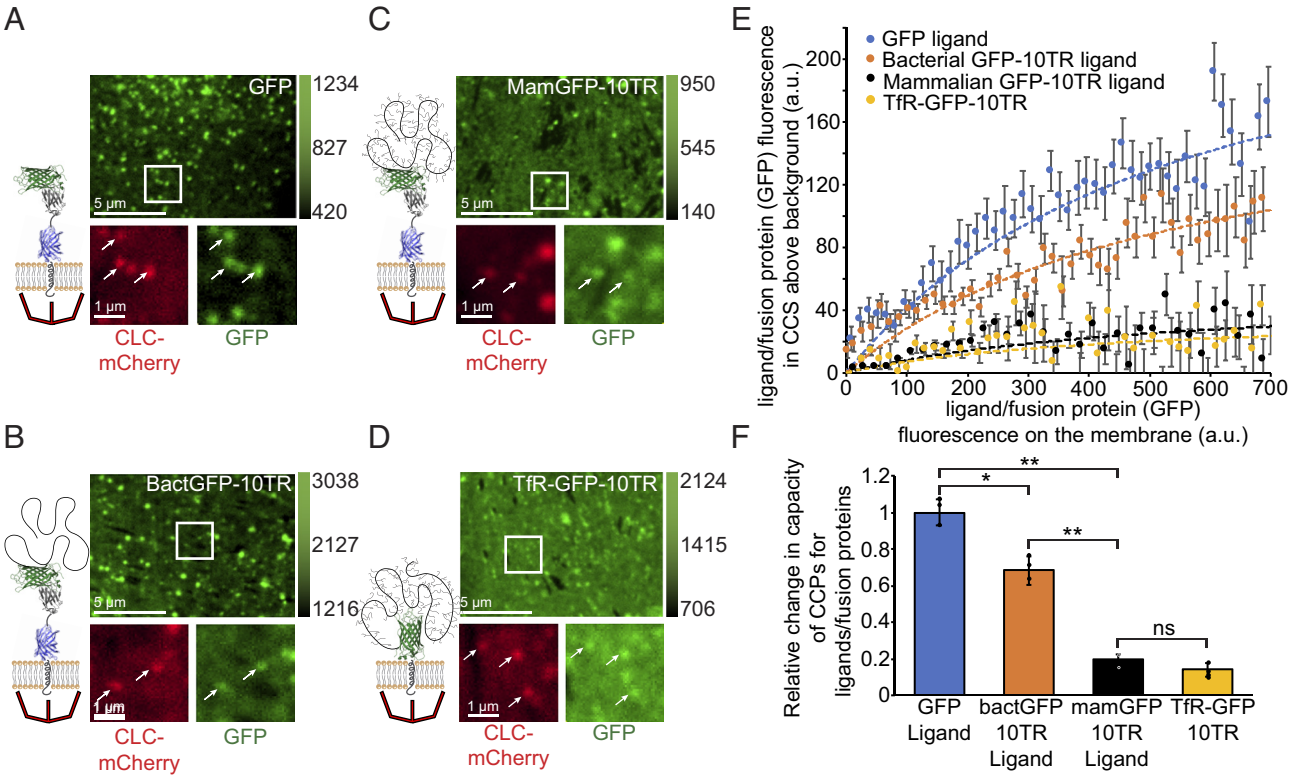

Fig. 6. Glycosylation contributes significantly to the reduced endocytosis of ligands that contain the MUC1 tandem repeat domain. Spinning disk confocal images of the plasma membrane of RPE cells transiently expressing the chimeric model receptor TfR-BFP-GFPnb and incubated with either (*A*) GFP ligand, (*B*) bact-GFP-10TR ligand, (*C*) coexpressed with mam-GFP-10TR, or (*D*) RPE cells expressing the transmembrane fusion protein, TfR-GFP-10TR, in mammalian cells. (*E*) The number of model ligands within clathrin-coated structures plotted as a function of the local concentration of ligands on the plasma membrane surrounding the clathrin-coated structures. Each point on the plot represents the average of 200 clathrin-coated structures binned by the local membrane concentration of the model ligand. A total of 20,165 CCSs were detected from 114 cells incubated with the GFP ligand, 21,893 CCSs were detected from 117 cells incubated with bact-GFP-10TR, 20,856 CCSs were detected from 122 cells expressing mam-GFP-10TR, and 18,420 CCSs were detected from 80 cells expressing TfR-GFP-10TR. Error bars represent mean ± SE. Solid lines are Boltzmann lattice model predictions using the best-fit values of K_{deff} and N_{max}. (*F*) Bar plot of relative CCS capacities for each of the ligands or 10TR fusion protein. The individual data points represent separate samples. The error bars represent 95% CI of the best-fit values of N_{max}. A two-sample *t* test was also conducted comparing the best-fit values of N_{max}. *P*-value was >0.05 for mam GFP-10TR and TfR-GFP-10TR (*P*-value was = 0.2) indicating that the difference in N_{max} was not significant. *P*-values were <0.05 between the other pairs of the ligands, indicating a statistically significant difference between their N_{max} values. $^{*}P < 0.05$, $^{**}P < 0.01$.

E and *F*). Furthermore, we performed a western blot to evaluate the extent of glycosylation that occurred on mam-GFP-10TR and found an increase in its effective molecular weight of about 70 kDa (*SI Appendix*, Fig. S6). We next evaluated the relative ability of clathrin-mediated endocytic structures to recruit model receptors bound to the two ligands.

Glycosylation Contributes Significantly to the Reduced Endocytosis of Ligands That Contain the MUC1 Tandem Repeat Domain. To evaluate endocytosis of the ligand-bound model receptors, we imaged cells recruiting each ligand: GFP, bact-GFP-10TR, and mam-GFP-10TR. Here, again we used cells expressing the 10TR fusion protein as a positive control. In each case, we compared cells with a similar overall intensity of GFP at the plasma membrane, and observed the partitioning of the GFP signal between endocytic structures and the surrounding plasma membrane (Fig. 6 *A*–*D*). As seen from comparing Fig. 6 *A* and *D*, partitioning of the GFP signal to endocytic structures was substantially weaker for cells that expressed the 10TR fusion protein compared to cells that recruited the GFP ligand. Meanwhile, cells that recruited the mam-GFP-10TR ligand had relatively low contrast, similar to 10TR, while cells that recruited the bact-GFP-10TR ligand appeared to have intermediate contrast.

To quantify these observations, we constructed recruitment curves similar to Fig. 3*A*, above. Specifically, we quantified the intensity of each ligand within clathrin-coated structures, as well as the intensity of the ligand on the surrounding plasma membrane. Fig. 6*E* shows the results of this analysis. As described above, the relative number of ligands per endocytic structure initially increased linearly with increasing relative ligand concentration on the surrounding plasma membrane before plateauing toward the saturated capacity of endocytic structures for the ligand-bound receptor. From these data, it is evident that the saturated capacity is highest for the GFP-bound receptor and lowest for the 10TR fusion protein. When the receptor bound to the mam-GFP-10TR ligand, the relative number of ligand-bound receptors was approximately the same as for 10TR. In contrast, for receptors bound to the bact-GFP-10TR ligand, the relative number of ligand-bound receptors was significantly greater, falling midway between the data for the GFP (negative control) and mam-GFP-10TR ligands (Fig. 6*E*). These results suggest that the ability of the tandem repeat domain to inhibit endocytosis of transmembrane proteins is derived in part from the steric bulk of the tandem repeat domain itself, and in part from the glycosylation of this domain, which would be expected to significantly increase its steric bulk and net charge (38). Notably, intracellular signal from the ligands, which usually appears as puncta within the endosomes, did not colocalize with clathrin-coated structures, and were therefore excluded from the analysis in Fig. 6 *E* and *F*.

Conclusion

Here, we have used live-cell imaging of large ensembles of individual endocytic events to study the impact of MUC1's tandem repeat domain on endocytosis. Surprisingly, we find that expression of MUC1 TRs at the plasma membrane has little, if any impact on endocytic dynamics or the timescale over which transmembrane proteins enter endocytic structures. In contrast, our results reveal that the steric bulk of the MUC1 ectodomain limits the number of transmembrane fusion proteins that can be accommodated within each clathrin-coated structure. Specifically, we demonstrate that increasing the length and glycosylation state of the MUC1 tandem repeat domain collectively decreases the capacity of endocytic structures for the transmembrane fusion proteins by more than fivefold.

Our results are in close agreement with a recent report, which suggests that the MUC1 ectodomain chain length influences glycocalyx properties. Notably, Park et al., 2022 found that the MUC1 tandem repeat length, in addition to other factors, impacts the crowding and extension of the ectodomain, ultimately altering the glycocalyx thickness (44). Further, we found that nearly half of the decrease in recruitment associated with MUC1's ectodomain was due to the molecular weight of the tandem repeat domain itself, in the absence of glycosylation, while the remaining portion arose directly from glycosylation. Although Shurer et al., 2019 reported a relatively small impact of the glycosylation state of MUC1 on the formation of membrane tubules, they observed a decrease in its membrane density (38). Nonetheless, their results are largely in agreement with our findings.

MUC1, a clinical biomarker for cancer, experiences differential glycosylation in tumor cells. Some tumor cells exhibit increased expression of glycosylated MUC1 and other mucins, which leads to a dense glycocalyx that is thought to promote metastasis by inhibiting integrin-mediated adhesion of cells to the extracellular matrix (56). Our results help to explain the connection between increased glycosylation and accumulation of glycosylated proteins at the plasma membrane. Specifically, we have shown that as glycosylation increases, glycocalyx proteins, such as MUC1, become more difficult to remove from the plasma membrane by endocytosis, setting up a positive feedback loop that would be expected to increase glycocalyx density.

In other contexts, tumor cells express MUC1 with truncated O-glycans (31, 57), which correlates with accumulation of MUC1 intracellularly, rather than at the plasma membrane (31). This accumulation is thought to promote multiple types of oncogenic signaling (58). Our results help to explain the connection between truncated glycans and intracellular accumulation of MUC1. In particular, we have shown that as glycosylation decreases, glycocalyx proteins, such as MUC1, can be more easily removed from the plasma membrane by endocytosis, setting up a negative feedback loop that would be expected to deplete the glycocalyx. Through similar mechanisms, loss of cell surface glycans owing to endocytosis could play a role in diseases such as COPD and atherosclerosis, where the integrity of the glycocalyx is progressively compromised (59, 60). Owing to the critical role of both glycosylation and endocytosis in many cellular processes, our findings could have broad implications for normal and aberrant cellular physiology.

Materials and Methods

Cell Culture and Transfection. Human RPE (ARPE-19) cells expressing mCherry-tagged clathrin light chain (RPE-CLC-mCherry) were received as a gift from Allen Liu (University of Michigan) and Sandra Schmid (University of Texas Southwestern). Cells were cultured in 50% Dulbecco's modified eagle medium, 50% of F12 nutrient mixture, supplemented with 10% fetal bovine serum, 20 mM HEPES, 1% Pen/Strep/L-glutamine. Cells were grown at 37 °C with 5% CO_2. For fluorescence microscopy assays, RPE-CLC-mCherry cells were seeded on acid-washed coverslips at a density of 50,000 cells per coverslip. Cells were transfected 24 h after seeding on coverslips; 3 μL Fugene HD (Promega, Madison, WI) transfection reagent was used to transfect 1 μg of each plasmid. Protein expression of the MUC1 ectodomain variants (TfR-GFP-10TR, TfR-GFP-5TR, TfR-GFP-1TR, TfR-GFP-0TR) was induced with 0.05 μg/mL Doxycycline Hyclate (Santa Cruz Biotechnology). Doxycycline Hyclate was added to cell culture media in the wells containing the seeded coverslips 16 to 18 h after transfection.

Fluorescence Microscopy. Cells were imaged 36 to 40 h after transfection using confocal microscopy or TIRF microscopy. Transfection media used for imaging lacked pH indicator (phenol red) and was supplemented with 1 μL OxyFluor (Oxyrase, Mansfield, OH) per 33 μL media to decrease photobleaching during live-cell fluorescence imaging. To prevent disulfide bond formation among the proteins, 1 mM TCEP was added to media for conditions including his-GFP and his-GFP10TR. The protein ligands were added at 500 nM to the cells, 10 min before imaging.

A spinning disk confocal microscope with a Yokogawa CSU-W1 SoRa confocal scanner unit, Olympus IX83 microscope body and an Olympus 100× plan-apochromat 1.5 NA oil-immersion objective was used to image the plasma membrane of live cells. The microscope was equipped with a Hamamatsu ORCA C13440-20CU CMOS camera for measuring fluorescence emission. Lasers with excitation wavelengths of 405 nm for BFP, 488 nm for GFP, 561 nm for mCherry, and 640 nm for Alexa647 were used.

Movies of live cells were collected on a TIRF microscope. The plasma membrane was imaged for 10 min at 2-s intervals. A Zeiss plan-apochromat 100×, 1.46 NA oil immersion TIRF objective and Photometrics Evolve delta EMCCD camera were fitted onto an Olympus IX73 microscope body. An excitation laser of wavelength 473 nm was used to excite GFP, while a 532-nm laser was used to excite mCherry. A 635-nm laser was used for autofocus correction. The cell samples were maintained at 37 °C throughout the imaging experiments.

The clathrin-coated structures, visible as fluorescent puncta in the confocal and TIRF images, were detected using CMEAnalysis (Danuser lab) (46). 2D Gaussian functions were fit to local intensity maxima in the CLC-mCherry channel (master channel), which marks the clathrin-coated structures. The SD of the Gaussian was calculated from the physical parameters of the microscope to approximate the point spread function. Additionally, the Anderson-darling test was performed on the residuals of the fit to validate the goodness of the fit. The Gaussian amplitude, representative of the fluorescence intensity of the detected punctum, and the location of the puncta were recorded. For a punctum to be considered a valid clathrin-coated structure, it had to be diffraction limited and significantly brighter than the local membrane surrounding the puncta, as described previously (46). For valid puncta in the master channel, a 2D Gaussian was then fit to the corresponding puncta in the transmembrane fusion protein, receptor, and/or ligand channels. A gaussian curve was fit in the subordinate channels within a 3σ pixel radius of the corresponding location in the master channel. Notably, the CLC-mCherry fluorescence within an endocytic structure above the background fluorescence must have a value greater than the square root of the camera noise to be measurable. In our system, CLC-mCherry fluorescence signal of approximately 10 a.u. can be differentiated above the noise, which is approximately 100 a.u. Additionally, all detections with mCherry-tagged clathrin were included in our analyses and plots.

Online Materials and Methods. *SI Appendix, Materials and Methods* includes details about plasmid constructs, gene fragments, protein purification, thermodynamic model fitting, FRAP assays, PNA staining of live cell for fluorescence microscopy, mass spectroscopy and western blot analysis.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank the M. Paszek lab at Cornell University for generously sharing plasmids for MUC1. We thank Carl Hayden for insightful discussions about this manuscript. This research was supported by the NIH

through R35GM139531 to J.C.S. and R35GM142941 to B.B., and the Welch Foundation through grant F-2047 to J.C.S and grant F-2055-20210327 to B.B. It was also supported through a grant from the NSF, 2218467, to B.B. and J.C.S.

Author affiliations: [a]Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX; [b]Department of Biochemistry and Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX; [c]McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX; and [d]Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX

1. A. Sorkin, M. von Zastrow, Endocytosis and signalling: Intertwining molecular networks. *Nat. Rev. Mol. Cell Biol.* **10**, 609–622 (2009), 10.1038/nrm2748.
2. L. K. Goh, A. Sorkin, Endocytosis of receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.* **5**, a017459 (2013), 10.1101/cshperspect.a017459.
3. A. V. Vieira, C. Lamaze, S. L. Schmid, Control of EGF receptor signaling by clathrin-mediated endocytosis. *Science* **274**, 2086–2089 (1996), 10.1126/science.274.5295.2086.
4. E. Cocucci, F. Aguet, S. Boulant, T. Kirchhausen, The first five seconds in the life of a clathrin-coated pit. *Cell* **150**, 495–507 (2012), 10.1016/j.cell.2012.05.047.
5. M. Kaksonen, A. Roux, Mechanisms of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* **19**, 313–326 (2018), 10.1038/nrm.2017.132.
6. E. M. Schmid, H. T. McMahon, Integrating molecular and network biology to decode endocytosis. *Nature* **448**, 883–888 (2007), 10.1038/nature06031.
7. B. Antonny *et al.*, Membrane fission by dynamin: What we know and what we need to know. *Embo J.* **35**, 2270–2284 (2016), 10.15252/embj.201694613.
8. H. Ohno *et al.*, Interaction of tyrosine-based sorting signals with the medium chains of clathrin-associated protein. *Mol. Biol. Cell* **6**, 398–398 (1995).
9. B. T. Kelly *et al.*, A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. *Nature* **456**, 976–979 (2008), 10.1038/nature07422.
10. L. M. Traub, Tickets to ride: Selecting cargo for clathrin-regulated internalization. *Nat. Rev. Mol. Cell Biol.* **10**, 583–596 (2009), 10.1038/nrm2751.
11. A. C. M. DeGroot *et al.*, Entropic control of receptor recycling using engineered ligands. *Biophys. J.* **114**, 1377–1388 (2018), 10.1016/j.bpj.2018.01.036.
12. A. P. Liu, F. Aguet, G. Danuser, S. L. Schmid, Local clustering of transferrin receptors promotes clathrin-coated pit initiation. *J. Cell Biol.* **191**, 1381–1393 (2010), 10.1083/jcb.201008117.
13. C. Zhao *et al.*, Receptor heterodimerization modulates endocytosis through collaborative and competitive mechanisms. *Biophys. J.* **117**, 646–658 (2019), 10.1016/j.bpj.2019.07.012.
14. A. C. M. DeGroot, C. Zhao, M. F. LaMonica, C. C. Hayden, J. C. Stachowiak, Molecular thermodynamics of receptor competition for endocytic uptake. *Soft Matter* **15**, 7448–7461 (2019), 10.1039/c9sm00876d.
15. A. C. M. DeGroot, S. Gollapudi, C. Zhao, M. F. LaMonica, J. C. Stachowiak, Weakly internalized receptors use coated vesicle heterogeneity to evade competition during endocytosis. *Biochemistry* **60**, 2195–2205 (2021), 10.1021/acs.biochem.1c00268.
16. R. Kornfeld, S. Kornfeld, Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664 (1985), 10.1146/annurev.bi.54.070185.003215.
17. T. W. Rademacher, R. B. Parekh, R. A. Dwek, Glycobiology. *Annu. Rev. Biochem.* **57**, 785–838 (1988), 10.1146/annurev.bi.57.070188.004033.
18. A. Varki, H. H. Freeze, A. E. Manzi, Preparation and analysis of glycoconjugates. *Curr. Protoc. Mol. Biol.* **88**, 17.0.1–17.0.12 (2009), 10.1002/0471142727.mb1700s88.
19. G. Opdenakker, P. M. Rudd, C. P. Ponting, R. A. Dwek, Concepts and principles of glycobiology. *FASEB J.* **7**, 1330–1337 (1993), 10.1096/fasebj.7.14.8224606.
20. H. Clausen, H. H. Wandall, M. P. DeLisa, P. Stanley, R. L. Schnaar, "Glycosylation engineering" in *Essentials of Glycobiology*, A. Varki *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2022).
21. P. Van den Steen, P. M. Rudd, R. A. Dwek, G. Opdenakker, Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **33**, 151–208 (1998), 10.1080/10409239891204198.
22. S. Mulagapati, V. Koppolu, T. S. Raju, Decoding of O-linked glycosylation by mass spectrometry. *Biochemistry* **56**, 1218–1226 (2017), 10.1021/acs.biochem.6b01244.
23. P. M. Rudd, R. A. Dwek, Glycosylation: Heterogeneity and the 3D structure of proteins. *Crit. Rev. Biochem. Mol. Biol.* **32**, 1–100 (1997), 10.3109/10409239709085144.
24. X. Chen, A. Varki, Advances in the biology and chemistry of sialic acids. *ACS Chem. Biol.* **5**, 163–176 (2010), 10.1021/cb900266r.
25. C. L. Hattrup, S. J. Gendler, Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol.* **70**, 431–457 (2008), 10.1146/annurev.physiol.70.113006.100659.
26. D. W. Kufe, Mucins in cancer: Function, prognosis and therapy. *Nat. Rev. Cancer* **9**, 874–885 (2009), 10.1038/nrc2761.
27. R. Gupta, F. Leon, S. Rauth, S. K. Batra, M. P. Ponnusamy, A systematic review on the implications of O-linked glycan branching and truncating enzymes on cancer progression and metastasis. *Cells* **9**, 446 (2020), 10.3390/cells9020446.
28. E. Denneny *et al.*, Mucins and their receptors in chronic lung disease. *Clin. Transl. Immunol.* **9**, e01120 (2020), 10.1002/cti2.1120.
29. M. C. Rose, J. A. Voynow, Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol. Rev.* **86**, 245–279 (2006).
30. C. L. Kinlough, P. A. Poland, J. B. Bruns, K. L. Harkleroad, R. P. Hughey, MUC1 membrane trafficking is modulated by multiple interactions. *J. Biol. Chem.* **279**, 53071–53077 (2004), 10.1074/jbc.M409360200.
31. Y. Altschuler *et al.*, Clathrin-mediated endocytosis of MUC1 is modulated by its glycosylation state. *Mol. Biol. Cell* **11**, 819–831 (2000).
32. X. Liu, Z. Yuan, M. Chung, MUC1 intra-cellular trafficking is clathrin, dynamin, and Rab5 dependent. *Biochem. Biophys. Res. Commun.* **376**, 688–693 (2008), 10.1016/j.bbrc.2008.09.065.
33. W. Chen *et al.*, MUC1: Structure, function, and clinic application in epithelial cancers. *Int. J. Mol. Sci.* **22**, 6567 (2021), 10.3390/ijms22126567.
34. H. Razawi *et al.*, Evidence for core 2 to core 1 O-glycan remodeling during the recycling of MUC1. *Glycobiology* **23**, 935–945 (2013), 10.1093/glycob/cwt030.
35. S. J. Storr *et al.*, The O-linked glycosylation of secretory/shed MUC1 from an advanced breast cancer patient's serum. *Glycobiology* **18**, 456–462 (2008), 10.1093/glycob/cwn022.
36. Y. Andersch-Björkman, K. A. Thomsson, J. M. Holmén Larsson, E. Ekerhovd, G. C. Hansson, Large scale identification of proteins, mucins, and their O-glycosylation in the endocervical mucus during the menstrual cycle*. *Mol. Cell. Proteomics* **6**, 708–716 (2007), 10.1074/mcp.M600439-MCP200.
37. S. Müller *et al.*, High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells*. *J. Biol. Chem.* **274**, 18165–18172 (1999), 10.1074/jbc.274.26.18165.
38. C. R. Shurer *et al.*, Physical principles of membrane shape regulation by the glycocalyx. *Cell* **177**, 1757–1770.e21 (2019), 10.1016/j.cell.2019.04.017.
39. H. Hofmann *et al.*, Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16155–16160 (2012), 10.1073/pnas.1207719109.
40. C.-H. Lu *et al.*, Membrane curvature regulates the spatial distribution of bulky glycoproteins. *Nat. Commun.* **13**, 1–16 (2022), 10.1038/s41467-022-30610-2.
41. C. Harding, J. Heuser, P. Stahl, Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes. *J. Cell Biol.* **97**, 329–339 (1983), 10.1083/jcb.97.2.329.
42. J. F. Collawn *et al.*, Transferrin receptor internalization sequence YXRF implicates a tight turn as the structural recognition motif for endocytosis. *Cell* **63**, 1061–1072 (1990).
43. J. F. Collawn *et al.*, YTRF is the conserved internalization signal of the transferrin receptor, and a second Ytrf signal at position 31–34 enhances endocytosis. *J. Biol. Chem.* **268**, 21686–21692 (1993).
44. S. Park *et al.*, Mucins form a nanoscale material barrier against immune cell attack. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.01.28.478211. Accessed 29 January 2022.
45. D. J. Busch *et al.*, Intrinsically disordered proteins drive membrane curvature. *Nat. Commun.* **6**, 7875 (2015), 10.1038/ncomms8875.
46. F. Aguet, C. N. Antonescu, M. Mettlen, S. L. Schmid, G. Danuser, Advances in analysis of low signal-to-noise images link dynamin and AP2 to the functions of an endocytic checkpoint. *Dev. Cell* **26**, 279–291 (2013).
47. D. Axelrod, Total internal reflection fluorescence microscopy in cell biology. *Traffic.* **2**, 764–774 (2001), 10.1034/j.1600-0854.2001.21104.x.
48. P. G. Saffman, M. Delbrück, Brownian motion in biological membranes. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3111–3113 (1975).
49. G. L. Rosano, E. A. Ceccarelli, Recombinant protein expression in Escherichia Coli: Advances and Challenges. *Front. Microbiol.* **5**, 172 (2014), 10.3389/fmicb.2014.00172.
50. A. A. Tokmakov *et al.*, Multiple post-translational modifications affect heterologous protein synthesis *. *J. Biol. Chem.* **287**, 27106–27116 (2012), 10.1074/jbc.M112.366351.
51. N. Jenkins, R. B. Parekh, D. C. James, Getting the glycosylation right: Implications for the biotechnology industry. *Nat. Biotechnol.* **14**, 975–981 (1996), 10.1038/nbt0896-975.
52. O. Letourneur, S. Sechi, J. Willette-Brown, M. W. Robertson, J.-P. Kinet, Glycosylation of human truncated FcϵRI α chain is necessary for efficient folding in the endoplasmic reticulum ($*$). *J. Biol. Chem.* **270**, 8249–8256 (1995), 10.1074/jbc.270.14.8249.
53. M. H. Kubala, O. Kovtun, K. Alexandrov, B. M. Collins, Structural and thermodynamic analysis of the GFP:GFP-nanobody complex. *Protein Sci.* **19**, 2389–2401 (2010), 10.1002/pro.519.
54. V. Sharma *et al.*, Molecular basis of recognition by Gal/GalNAc specific legume lectins: Influence of Glu 129 on the specificity of peanut agglutinin (PNA) towards C2-substituents of galactose. *Glycobiology* **8**, 1007–1012 (1998), 10.1093/glycob/8.10.1007.
55. M. H. Mansour, H. I. Negm, A. H. Saad, A. Y. Zahran, N. Badir, Identification of peanut agglutinin-binding glycoproteins on lizard lymphocytes. *Zoolog. Sci.* **12**, 79–85 (1995), 10.2108/zsj.12.79.
56. J. Wesseling, S. W. van der Valk, H. L. Vos, A. Sonnenberg, J. Hilkens, Episialin (MUC1) overexpression inhibits integrin-mediated cell adhesion to extracellular matrix components. *J. Cell Biol.* **129**, 255–265 (1995), 10.1083/jcb.129.1.255.
57. K. O. Lloyd, J. Burchell, V. Kudryashov, B. W. T. Yin, J. Taylor-Papadimitriou, Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines: Demonstration of simpler and fewer glycan chains in tumor cells *. *J. Biol. Chem.* **271**, 33325–33334 (1996), 10.1074/jbc.271.52.33325.
58. Y. Li, D. Kufe, The human DF3/MUC1 carcinoma-associated antigen signals nuclear localization of the catenin p120(Ctn). *Biochem. Biophys. Res. Commun.* **281**, 440–443 (2001), 10.1006/bbrc.2001.4383.
59. K. Kato *et al.*, MUC1 contributes to goblet cell metaplasia and MUC5AC expression in response to cigarette smoke in vivo. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **319**, L82–L90 (2020), 10.1152/ajplung.00049.2019.
60. J. Qu, Y. Cheng, W. Wu, L. Yuan, X. Liu, Glycocalyx impairment in vascular disease: Focus on inflammation. *Front. Cell Dev. Biol.* **9**, 730621 (2021), 10.3389/fcell.2021.730621.