

Steric pressure between glycosylated transmembrane proteins inhibits internalization by endocytosis

Authors

Sadhana Gollapudi¹, Sabah Jamal¹, Advika Kamatar¹, Feng Yuan¹, Liping Wang³, Eileen M. Lafer³, Brian Belardi⁴, Jeanne C. Stachowiak^{1,2*}

¹*Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX*

²*Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX*

³*University of Texas Health Science Center at San Antonio, Department of Biochemistry and Structural Biology, San Antonio, TX*

⁴*McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX*

**To whom correspondence should be addressed: jcstach@austin.utexas.edu*

COMPETING INTERESTS

The authors declare no competing interests.

ABSTRACT

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.

SIGNIFICANCE STATEMENT

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, such as 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? Here 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 helps to explain how glycosylated proteins accumulate at the plasma membrane, a key requirement for cellular health.

INTRODUCTION

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 proteins¹⁻³. 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 coat⁴⁻⁶. 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 cytoplasm⁷.

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 by the adaptor protein, AP2⁸⁻¹⁰. 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 bulk¹¹, 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 proteins¹¹.

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 structures²⁴. As one example of highly glycosylated cargo proteins, mucins, major constituents of the glycocalyx, are heavily O-glycosylated²⁵. 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 disease (COPD), cystic fibrosis, and some cancers^{926–29}. 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 clathrin-mediated endocytosis.

To study the effect of glycosylation on the endocytosis of transmembrane proteins, we studied variants of MUC1, a heavily glycosylated transmembrane mucin that is known to be taken up by clathrin-mediated endocytosis^{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 tandem repeats³³. These tandem repeats are rich in threonine and serine residues, which are O-glycosylated²¹. *In vivo*, MUC1 has many isoforms, each containing a different number of tandem repeats. A MUC1 membrane protein with a greater number of tandem repeats 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 6-7 monosaccharides in length and have a molecular weight ranging from 500 to 1300 Da. Therefore, if every site on the tandem repeat were glycosylated, its molecular weight would increase by 2-5 fold^{34–37}.

Glycosylation of MUC1 influences the hydrodynamic radius of its tandem repeat domain³⁸. 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 nanometer³⁹. In contrast, the persistence length of the glycosylated MUC1 tandem repeat domain has been estimated at 7-8 nm³⁸, 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 cells³¹. 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 protrusions³⁸. 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 structures⁴⁰. 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), (Figure 1A-D). We chose the intracellular and transmembrane domains of TfR because TfR's internalization by clathrin-mediated endocytosis is strong and well-characterized^{8,41–43}. Notably, the native intracellular domain of MUC1 contains a YXX Φ 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 (Figure 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 ten tandem repeats (10TR) (Figure 1A-D). The illustrations of the transmembrane fusion proteins in Figure 1A-D are drawn to scale based on previously determined estimates of the radius of gyration of MUC1 tandem repeat domains⁴⁴. Notably, the transmembrane fusion proteins only include the intracellular and transmembrane domains of the transferrin receptor, such that any glycosylation normally associated with the extracellular domain of the transferrin receptor is not present in these chimeras. Each of the transmembrane fusion proteins were 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 structures⁴⁵. RPE cells are commonly used in studies of endocytosis due to their large and well-spread lamellipodia, which enable visualization of the plasma membrane⁴⁶. 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). Figure 1E 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 (Figure 1F-H). Interestingly, as the number of tandem repeats 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 (Figure 1F-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 20-120s 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 tandem repeats 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.

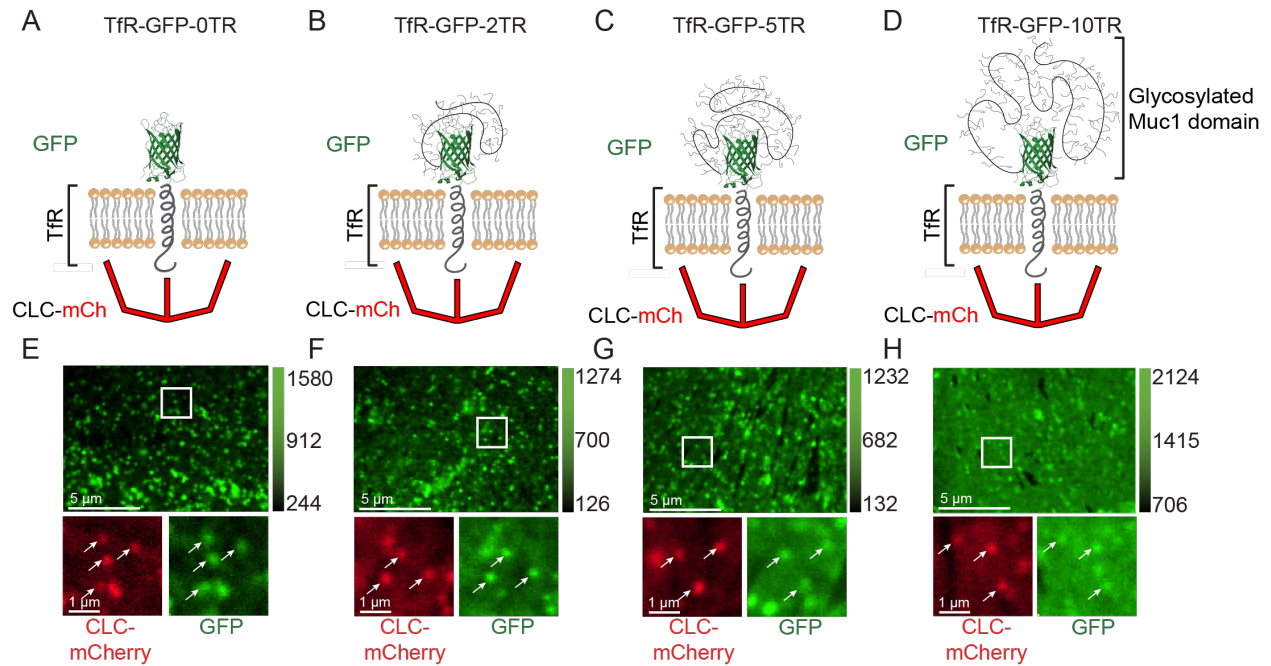


FIGURE 1. Glycosylated transmembrane fusion proteins partition weakly into clathrin-coated structures. (A-D) A schematic of transmembrane fusion proteins with an incremental increase in the number of MUC1 tandem repeats from (A) zero tandem repeats, (B) two tandem repeats, (C) five tandem repeats, (D) 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) ten 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 minutes. These image series were collected in the same two fluorescent channels used in Figure 1: (i) clathrin light chain (mCherry), and (ii) transmembrane fusion protein (GFP). The time series were analyzed to identify and track individual endocytic events, from initiation to departure (Figure 2B, D). Specifically, we used a publicly available algorithm, CMEAnalysis⁴⁶ 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 estimated⁴⁶. 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, Figure S2 shows the fraction of detected endocytic structures with transmembrane fusion protein fluorescence below the detection threshold increases as the number of tandem repeats 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 10-19 s, 20-39 s, 40-59 s, 60-79 s, 80-99 s, and 100-120 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 clathrin-mediated endocytosis (Figure 2E).

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 10-120s. In Figure 2F, 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 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 work¹¹. 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 (Figure 2H, I). The FRAP measurements showed that the recovery fraction, or mobile fraction, and the recovery time were very similar (Figure 2J, 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.

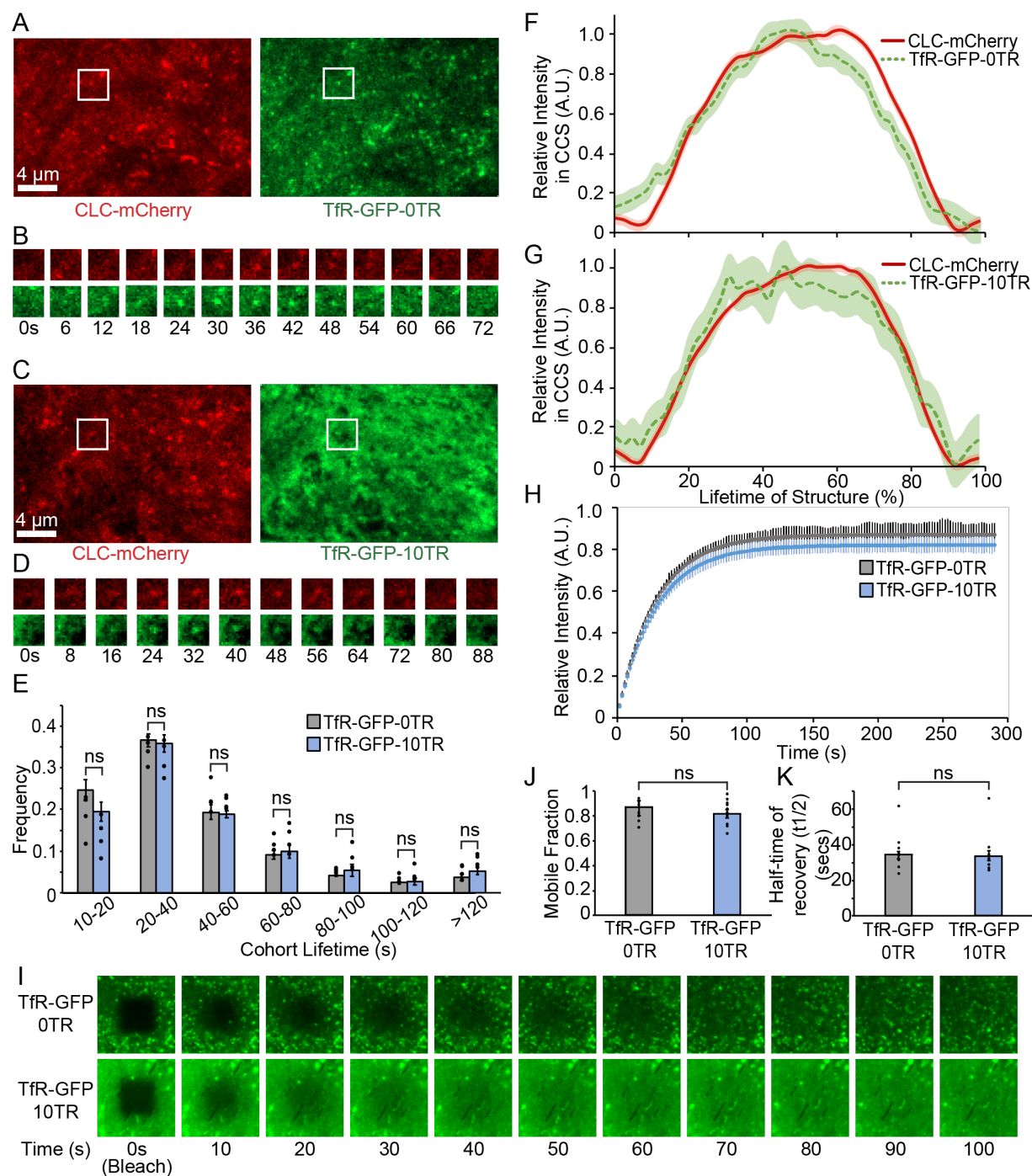


FIGURE 2. The dynamics of protein loading into clathrin-coated structures are unaffected by MUC1's ectodomain. (A,C) TIRF microscopy images of the plasma membrane of RPE cells transiently expressing either (A) TfR-GFP-0TR or (C) TfR-GFP-10TR. (B,D) A sequence of images showing the maturation of one CCP 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 seconds for cells expressing TfR-GFP-0TR and 39.7 seconds for TfR-GFP-10TR. A two tailed t-test confirmed that the difference is statistically not significant ($P > 0.05$) between these data. (F, 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. 17 cells were analyzed resulting in 8143 CCSs tracked in (F), and 16 cells were analyzed resulting in 6698 CCS tracked in (G). The shaded areas represent the mean $\pm$ 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 was averaged for TfR-GFP-0TR, and data from 13 cells was 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 data sets 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 minutes. (J) Bar plots of the averaged mobile fraction and (K) half-time of recovery for TfR-GFP-0TR and TfR-GFP-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, 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 (Figure 1E-H).

To quantify transmembrane fusion protein partitioning within the images, we used CMEAnalysis⁴⁶. 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, non-normalized 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 (Figure 3A). Eventually the relative

number of transmembrane fusion proteins within clathrin-coated structures began to plateau towards 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 previously¹¹.

From these data it is clear that the saturated capacity of clathrin-coated structures for transmembrane fusion proteins declined considerably as the number of tandem repeats increased from 0 to 10. To estimate the saturated capacity, we applied a simple physical model, which we reported previously¹¹, equation 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, $\langle n \rangle$, 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}).

$$\langle n \rangle = \frac{N_{\max} C_{\text{mem}}}{K_{\text{deff}} + C_{\text{mem}}} (1)$$

In line with our findings in Figure 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 1 to the data in Figure 3A 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, 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 tandem repeats 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, 10TR), with $N_{\max}$ as the only free parameter (Figure 3A). Notably, the horizontal axis in Figure 3A 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. Figure 3B shows the resulting values of $N_{\max}$, which declined approximately 5-fold as the number of tandem repeats 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 quantify the recruitment of transmembrane fusion proteins containing more than ten tandem repeats.

In vivo, MUC1 can have as many as 42 tandem repeats. 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_{eff} , 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 tandem repeats 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 Figure 3C. 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 two-fold in the uptake of MUC1-GFP-10TR was measured compared to MUC1-GFP-0TR (Figure S3). These results confirm that glycosylated tandem repeats 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 sites¹¹. 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, Figure 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 Figure 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.

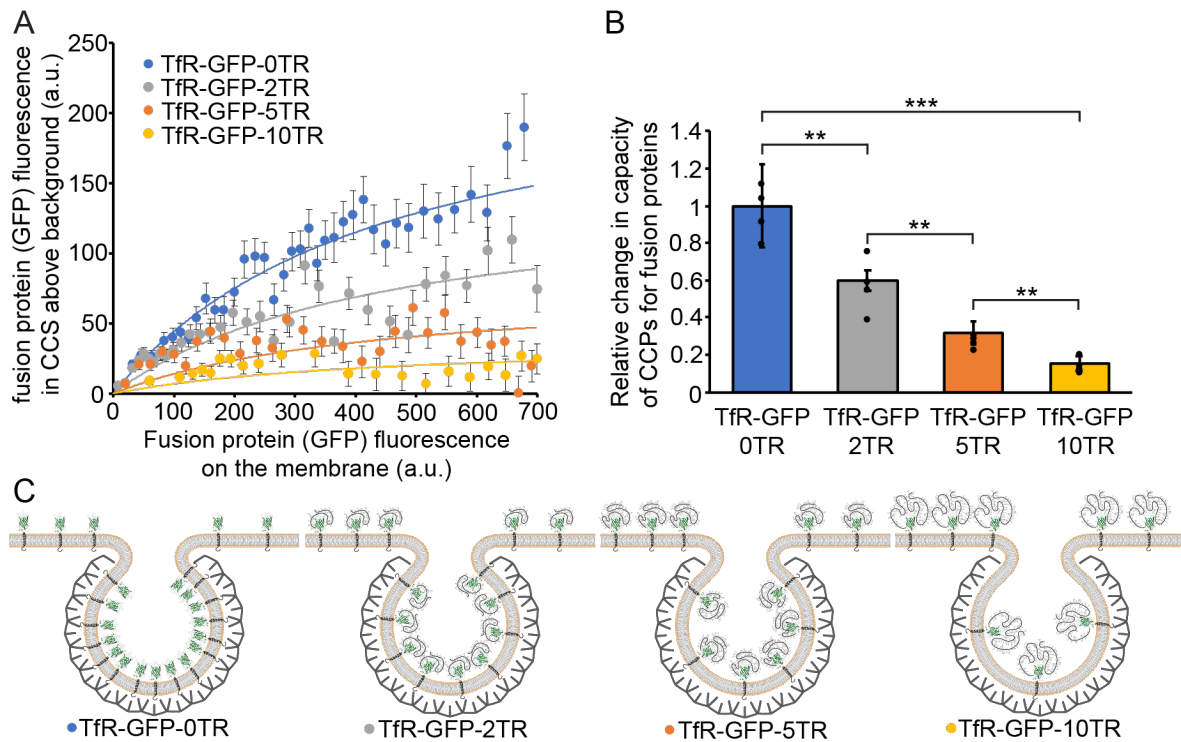


FIGURE 3. Recruitment of transmembrane fusion proteins into clathrin-coated structures decreases as the molecular weight of the glycosylated ectodomain increases. (A) The relative number of transmembrane fusion proteins within clathrin-coated structures is plotted as a function of the relative concentration 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 10123 CCSs were detected from 88 cells expressing TfR-GFP-0TR, 9964 CCSs from 101 cells expressing TfR-GFP-2TR, 11690 CCSs were detected from 75 cells expressing TfR-GFP-5TR and 18420 CCSs were detected from 80 cells expressing TfR-GFP-10TR. Error bars represent mean $\pm$ SE. Solid lines are model predictions using the best-fit values of K_{def} 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.

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 *E. 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 OTR fusion protein described above, the model receptor consisted of the intracellular and transmembrane domains of the transferrin receptor, followed by a blue fluorescent protein (BFP) domain. Additionally, the C-terminus of the model receptor was fused to a single domain antibody against GFP⁵³, 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 (Figure 4A). 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 (Figure 4B). Here the absence of significant glycosylation was confirmed by mass spectrometry (Figure S5). When glycosylation was desired, we produced the ligand by co-expressing 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 (Figure 4C).

Figure 4D-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 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, Figure 4G 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 (Figures 5, 6) compares groups of cells and endocytic structures with equivalent ligand binding, rather than equivalent expression of the model receptor.

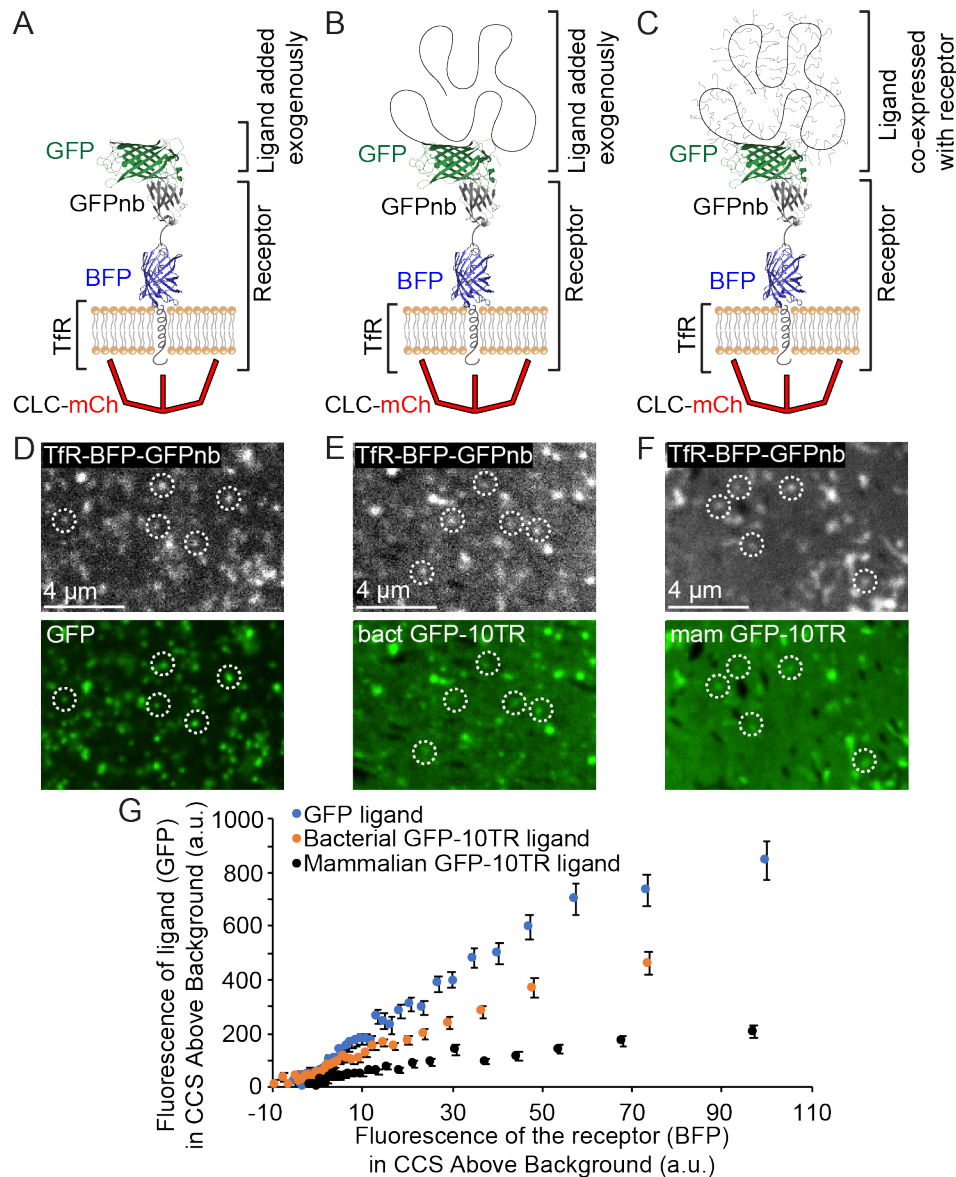


FIGURE 4. A model receptor/ligand system can be used to compare tandem repeat domains derived from bacteria and mammalian cells. (A-C) A Schematic of the model receptor bound to the ligands: (A) GFP, (B) bact-GFP-10TR, or (C) mam-GFP-10TR. (D-F) Fluorescent images of the plasma membrane of RPE cells expressing the model receptor, TfR-BFP-GFPnb (the top panels), and the ligands: (D) GFP, (E) bact-GFP-10TR, and (F) mam-GFP-10TR (bottom panels). 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 $\pm$ SE.

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 tandem repeats^{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 co-expression with the model receptor (mam-GFP-10TR). Then PNA-Alexa 647 was added to the culture, where it stained the surfaces of the cell. Figure 5A-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 (Figure 5D). 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 a substantially greater amount of PNA in comparison to cells recruiting the GFP ligand (negative control), compare Figure 5A, 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 overexpressed³⁸. In comparison, cells that recruited the bact-GFP-10TR ligand bound low levels of PNA (Figure 5B), similar to the negative control, while cells that recruited mam-GFP-10TR bound substantially higher levels of PNA (Figure 5C), 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 (Figure 5E). 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 Figures 1-3. Specifically, we found a linear increase in PNA staining with an increasing number of tandem repeats in the transmembrane fusion proteins (Figure S4 E, 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 (Figure S6). We next evaluated the relative ability of clathrin-mediated endocytic structures to recruit model receptors bound to the two ligands.

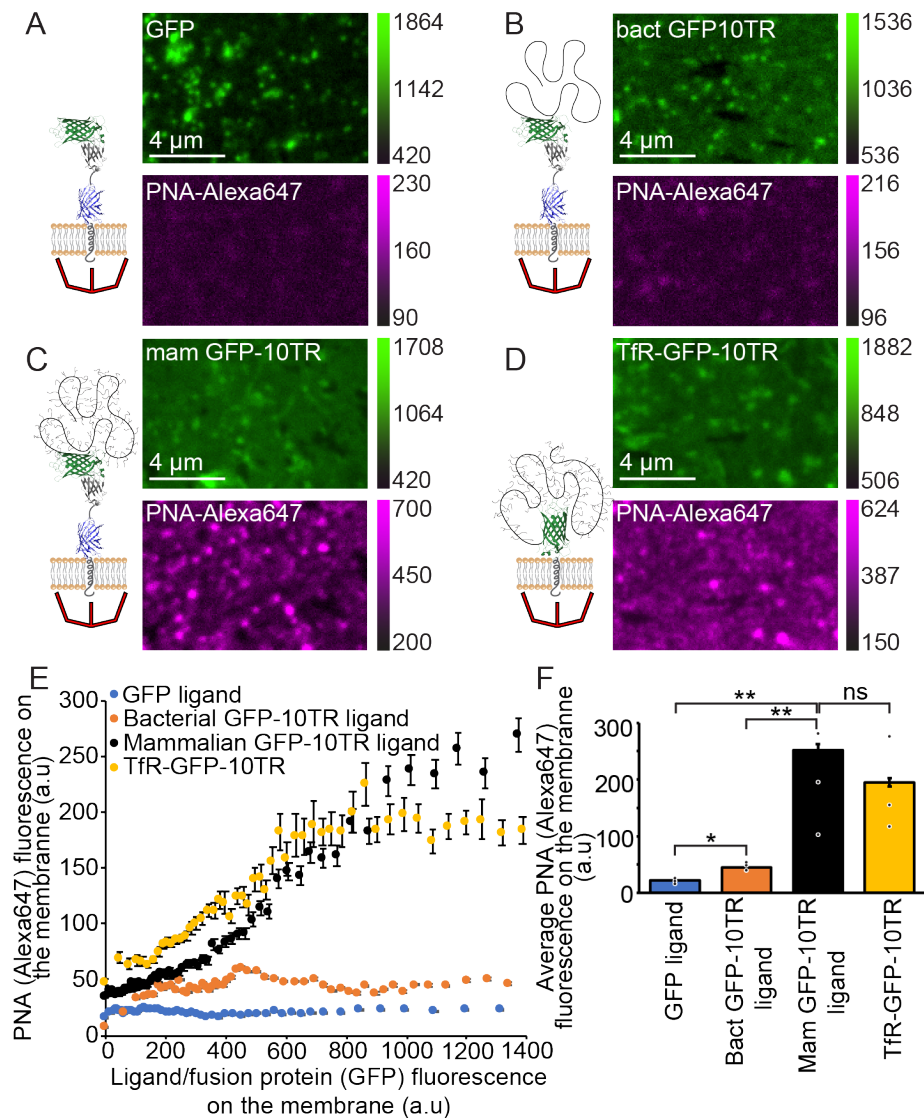


FIGURE 5. Tandem repeat domains purified from bacterial cells have little or no glycosylation in comparison to those excreted by mammalian cells. (A-D) Spinning disk images of the plasma membrane of RPE cells stained with PNA-Alexa647 and transiently expressing the chimeric model receptor TfR-BFP-GFPnb and incubated with either (A) GFP ligand, (B) bact-GFP-10TR ligand or (C) co-expressed 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 12013 CCSs were detected from 79 cells incubated with GFP ligand, 11658 CCSs were detected from 70 cells incubated with bact-GFP-

10TR, 13861 CCSs were detected from 111 cells expressing mam-GFP-10TR, and 10636 CCSs were detected from 87 cells expressing TfR-GFP-10TR. Error bars represent mean $\pm$ 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 $\pm$ 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 data sets 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. * $p < 0.05$, ** $p < 0.01$.

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 (Figure 6A-D). As seen from comparing Figure 6A, 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 Figure 3A, 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. Figure 6E 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 (Figure 6E). 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 charge³⁸. 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 Figure 6E, F.

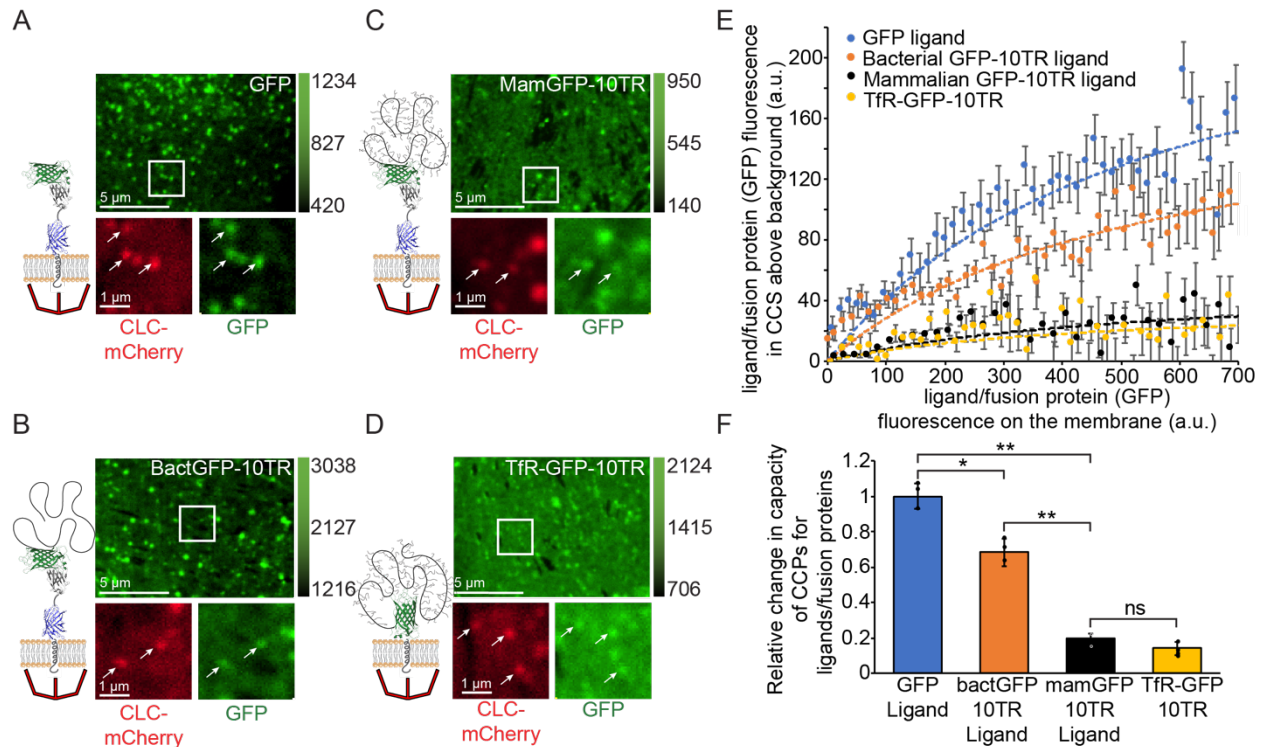


FIGURE 6. Glycosylation contributes significantly to the reduced endocytosis of ligands that contain the MUC1 tandem repeat domain. Spinning disk confocal images of the plasma membrane of RPE cells transiently expressing the chimeric model receptor TfR-BFP-GFPnb and incubated with either (A) GFP ligand, (B) bact-GFP-10TR ligand, (C) co-expressed 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 20165 CCSs were detected from 114 cells incubated with the GFP ligand, 21893 CCSs were detected from 117 cells incubated with bact-GFP-10TR, 20856 CCSs were detected from 122 cells expressing mam-GFP-10TR, and 18420 CCSs were detected from 80 cells expressing TfR-GFP-10TR. Error bars represent mean $\pm$ SE. Solid lines are Boltzmann lattice model predictions using the best-fit values of K_{def} 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. * $p < 0.05$, ** $p < 0.01$.

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 tandem repeats 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 five-fold.

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 thickness⁴⁴. 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 density³⁸. 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 matrix⁵⁶. 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 membrane³¹. This accumulation is thought to promote multiple types of oncogenic signaling⁵⁸. 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

Plasmid Constructs

Plasmids for the expression of TfR -GFP-10TR, TfR-GFP-5TR, TfR-GFP-2TR, and TfR-0TR were generated by inserting TfR-GFP and one of the above MUC1 tandem repeat fragments by Gibson Assembly cloning. The gene for the transmembrane fusion protein was inserted into a Tetracycline inducible PiggyBac expression vector. backbone vector used, pPB_MUC1_10_mOXGFP_dCT GFP BlnI RRK, was a gift from the Paszek lab (Cornell University). The MUC1_10_mOXGFP gene in the open reading frame of the vector was replaced with the genes of interest. The pEGFPN1-TfR-GFP⁴⁵ plasmid described previously was used as the template for PCR amplification of TfR-GFP DNA fragment. The plasmid encoded the intracellular and transmembrane domains of the transferrin receptor, amino acids 1-88 of GenBank accession number AAA61153. The transferrin receptor domains were fused to GFP with a 9-amino acid sequence linker (GKGDPVPVAT). The TfR-GFP gene was amplified by PCR from the plasmid using the forward primer, CTCTTAAGGCTAGAGGATCCATGGATCAAGCTAGATCAGCATTCTCT and reverse primer, GACTGGGTGCCCCGGTGTTCATCTTGTACAGCTCGTCCATGCC. The forward primer contained a 20-amino acid overlap on the 5-prime end, and the reverse primer contained a 20-amino acid overlap with the tandem repeat domain to be inserted. For the TfR-Δecto-GFP-10TR variant, a gene fragment for the 10TR domain was purchased from Integrated DNA Technologies (gBlocks). Similarly, 5TR and 2TR domain gene fragments were also purchased. Each of the gene fragments had a 20-amino acid overlap with the TfR-GFP PCR amplicon on the 5-prime end and with the vector on the 3-prime end. The vector, TfR-GFP PCR amplification product and one of the three tandem repeat domain gene fragments were ligated using the NEB Gibson Assembly master mix (NEB E2611L). The Gibson Assembly reaction product was then transformed into DH5 bacterial cells, spread on Ampicillin plates and grown for 16 hours. The colonies were screened for successful insertion of the DNA fragments.

The plasmid for expressing his-GFP-10TR in bacteria was generated by inserting the his-GFP-10TR gene into the pEGFPN1 vector by restriction cloning into the pET28a(+) vector. A gene fragment for his-GFP-10TR was purchased from Integrated DNA Technologies (gBlocks). The fragment included restriction enzyme cut sites for BamHI and EcoRI on the 5-prime and 3-prime ends respectively. The gene fragment and the pET28a(+) vector were digested with BamHI-HF (NEB R3101S) and EcoRI-HF (NEB R3136S) enzymes at 37°C for 15 minutes. The digested products were ligated with Quick Ligase (NEB M2200S) at room temperature for 10 minutes. The ligation reaction product was then transformed into DH5 bacterial cells, spread on Ampicillin plates and grown for 16 hours. The colonies were screened for successful insertion of the DNA fragments.

The plasmid for expressing GFP-10TR in mammalian cells was generated by inserting the his-GFP-10TR gene into the pEGFPN1 vector by restriction cloning. A gene fragment for his-GFP-10TR was purchased from IDT. The gene fragment included restriction enzyme cut sites for XhoI and NotI on the 5-prime and 3-prime ends respectively. For the gene to be secreted from mammalian cells, the EGFR signal sequence (MRPSGTAGAALLALLAALCPASRA) was included on the N-terminus of the gene. The vector, and the gene fragment were digested with XhoI (NEB R0146S) and NotI-HF (NEB R3189S) enzymes at 37°C for 1 hour.

The plasmid for expression of TfR-BFP-GFPnb was generated as previously described¹¹.

Gene Fragments

10TR repeat domain for TfR-GFP-10TR

```
CGGCCTGTACAAGCCAGATACAAGACCGGCCCCAGGATCTACGGCTCCTCCGGCT
CATGGAGTCACTTCTGCTCCAGACACAAGGCCCGCGCCGGGTTCTACAGCACCGC
CTGCTCATGGTGTTACTAGCGCACCCGATACGAGACCTGCTCCGGGATCAACGGC
ACCTCCTGCCCCACGGGGTAACATCTGCACCGGACACTCGCCCTGCGCCCGGTTCA
ACCGCTCCACCCGCACACGGAGTGACAAGCGCTCCTGACACTAGACCAGCACCA
GGTTCTACAGCCCCACCAGCCCATGGAGTTACCAGTGCACCAGATACTAGGCCAG
CTCCAGGTAGTACTGCACCCCCAGCTCATGGGGTTACATCAGCTCCCGACACGCG
ACCAGCTCCTGGAAGCACTGCCCTCCAGCTCACGGTGTGACCTCAGCACCTGAT
ACACGCCCTGCACCTGGCTCTACTGCTCCCCCGCTCATGGCGTAACTAGTGCCC
CGGATACTCGACCCGCCCTGGTTCCACAGCTCCGCCAGCACATGGTGTAACAAG
TGCTCCTGATACCCGACCAGCGCCTGGAAGTACCGCACCCACCTGCACATGGAGTA
ACTTCAGCCGCCTCAGGCTCTGCATCAGGCTCAGCTTAGGAATTCCGGCC
```

5TR repeat domain for TfR-GFP-5TR

CGGCCTGTACAAGCCAGATACAAGACCGGCCCCAGGATCTACGGCTCCTCCGGCT
CATGGAGTCACTTCTGCTCCAGACACAAGGCCCCGCGCCGGGTTCTACAGCACCGC
CTGCTCATGGTGTTACTAGCGCACCCGATACGAGACCTGCTCCGGGATCAACGGC
ACCTCCTGCCCCACGGGGTAACATCTGCACCGGACACTCGCCCTGCGCCCCGGTTCA
ACCGCTCCACCCGCACACGGAGTGACAAGCGCTCCTGACACTAGACCAGCACCA
GGTTCTACAGCCCCACCAGCCCATGGAGTTACCAGTGCAGCCTCAGGCTCTGCAT
CAGGCTCAGCTTAGGAATTCCGGCC

2TR repeat domain TfR-GFP-2TR

CGGCCTGTACAAGCCAGATACAAGACCGGCCCCAGGATCTACGGCTCCTCCGGCT
CATGGAGTCACTTCTGCTCCAGACACAAGGCCCCGCGCCGGGTTCTACAGCACCGC
CTGCTCATGGTGTTACTAGCGCAGCCTCAGGCTCTGCATCAGGCTCAGCTTAGGA
ATTCCGGCC

his-GFP-10TR fragment for the bact-GFP-10TR protein

CGGCCGGATCCATGGTGAGCAAGGGGCGAGGAGCTGTTACACGGGGTGGTGCCCCA
TCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCG
AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA
CCGGCAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCG
TGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTC
CGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAA
GCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAG
AACGGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGC
AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCACCCAGTCCAACTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGCCAGATACAAGACCGGCCCCAGGATCTACGGCT
CCTCCGGCTCATGGAGTCACTTCTGCTCCAGACACAAGGCCCCGCGCCGGGTTCTA
CAGCACCGCCTGCTCATGGTGTTACTAGCGCACCCGATACGAGACCTGCTCCGGG
ATCAACGGCACCTCCTGCCCCACGGGGTAACATCTGCACCGGACACTCGCCCTGCG
CCCGGTTCAACCGCTCCACCCGCACACGGAGTGACAAGCGCTCCTGACACTAGAC
CAGCACCAAGTTCTACAGCCCCACCAGCCCATGGAGTTACCAGTGCACCAGATAC
TAGGCCAGCTCCAGGTAGTACTGCACCCCCAGCTCATGGGGTTACATCAGCTCCC
GACACGCGACCAGCTCCTGGAAGCACTGCCCCTCCAGCTCACGGTGTGACCTCA
GCACCTGATACACGCCCTGCACCTGGCTCTACTGCTCCCCCGCTCATGGCGTAA
CTAGTGCCCCGGATACTCGACCCGCCCTGGTTCCACAGCTCCGCCAGCACATGG
TGTAACAAGTGCTCCTGATACCCGACCAGCGCCTGGAAGTACCGCACCACTGCA

CATGGAGTAACTTCAGCCGCCTCAGGCTCTGCATCAGGCTCAGCTTAGGAATTCC
GGCC

his-GFP-10TR fragment for the mammalian expressed GFP-10TR protein

CGGCCCTCGAGCATCATCATCATCACATGGTGAGCAAGGGCGAGGAGCTGTT
CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA
GTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT
GAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGACC
ACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT
CTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGA
CACCTTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATG
GCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCG
ACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCAACTGAG
CAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCC
GCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGCCAGATACAAGACCGGCC
CCAGGATCTACGGCTCCTCCGGCTCATGGAGTCACTTCTGCTCCAGACACAAGGC
CCGCGCCGGGTTCTACAGCACCGCCTGCTCATGGTGTACTAGCGCACCCGATAC
GAGACCTGCTCCGGGATCAACGGCACCTCCTGCCCACGGGGTAACATCTGCACC
GGACACTCGCCCTGCGCCCCGGTTCAACCGCTCCACCCGCACACGGAGTGACAAG
CGCTCCTGACACTAGACCAGCACCAGGTTCTACAGCCCCACCAGCCCATGGAGTT
ACCAGTGCACCAGATACTAGGCCAGCTCCAGGTAGTACTGCACCCCCAGCTCATG
GGGTTACATCAGCTCCCGACACGCGACCAGCTCCTGGAAGCACTGCCCTCCAGC
TCACGGTGTGACCTCAGCACCTGATACACGCCCTGCACCTGGCTCTACTGCTCCC
CCCGCTCATGGCGTAACTAGTGCCCCGGATACTCGACCCGCCCCCTGGTTCCACAG
CTCCGCCAGCACATGGTGTAAACAAGTGCTCCTGATACCCGACCAGCGCCTGGAAG
TACCGCACCACTGCACATGGAGTAACTTCAGCCGCCTCAGGCTCTGCATCAGGC
TCAGCTTAGGCGGGCCGCGGCC

Cell Culture and Transfection

Human RPE (ARPE-19) cells expressing mCherry-tagged clathrin light chain (RPE-CLC-mCherry) were received as a gift from Dr. Allen Liu (University of Michigan) and Dr. 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% CO₂. 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 hours after seeding on coverslips. 3 µL of 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 of Doxycycline Hyclate (Santa Cruz Biotechnology). Doxycycline Hyclate was added to cell culture media in the wells containing the seeded coverslips 16-18 hours after transfection.

Fluorescence Microscopy

Cells were imaged 36-40 hours after transfection using confocal microscopy or total internal reflection fluorescence (TIRF) microscopy. Transfection media used for imaging lacked pH indicator (phenol red) and was supplemented with 1 µL of OxyFluor (Oxyrase, Mansfield, OH) per 33 µL of media to decrease photobleaching during live cell fluorescence imaging. 1 mM of TCEP was added to media for conditions including his-GFP and his-GFP10TR, to prevent disulfide bond formation among the proteins. The protein ligands were added at 500 nM to the cells, 10 minutes 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 minutes at 2 second intervals. A Zeiss plan-apochromat 100x, 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)⁹. 2D gaussian functions were fit to local intensity maxima in the CLC-mCherry channel (master channel), which marks the clathrin-coated structures. The standard deviation 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 previously⁹. 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.

Protein Purification

The his-GFP protein was expressed and purified from BL21 pLysS bacterial cells as previously described^{45,61}.

his-GFP-10TR protein was expressed and purified from BL21 pLysS bacterial cells. The cells were grown in 1 L 2XTY medium for 2.5h at 37 °C to an optical density (~0.7 OD600) and then protein expression was induced with 1mM IPTG at 16 °C overnight. The cells were centrifuged at 11,000 x g for 15 minutes at 4 °C. The rest of the protocol was performed at 4 °C. The cell pellet was resuspended with a mortar and pestle in lysis buffer consisting of 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM β -ME, 1% Triton X-100 and EDTA-free protease inhibitor cocktail tablet (Thermo Scientific, Cat. # A32965) for 5 minutes on ice and then sonicated. The mixture was then clarified by ultracentrifugation at 134000 xg for 40 mins. The supernatant, which contained the protein, was incubated with 12 mL Ni-NTA resin slurry that contains 50% beads (GenScript, Cat. # L00223-25). The resin and supernatant mixture was gently stirred for 2 hours. The protein bound Ni-NTA resin was transferred into a chromatography column and allowed to settle. The Ni-NTA beads were thoroughly washed with a buffer containing 20 mM sodium phosphate and 150 mM NaCl pH 7.4, 20 mM imidazole, 20 mM β -ME, and 0.1 mM PMSF (Wash buffer). Then, the protein was eluted with the elution buffer (wash buffer plus 500 mM imidazole). The protein was eluted in fractions of 1 mL. The fractions with the highest concentration of the protein were combined and buffer-exchanged into phosphate-buffered saline using 10K Amicon Ultra Centrifugal filters (Sigma-Aldrich, Cat. # UFC9010). Protein concentration was measured using UV/VIS spectroscopy. The 20 μ L aliquots of the protein were then flash-frozen and stored at -80 °C. The purified protein was run on an SDS-PAGE gel (Bio-Rad, Cat. # 4561096) to check for purity and any protein degradation.

Thermodynamic model fitting

The data in Figures 3A and 6E, were fit with the thermodynamic lattice model as described previously¹¹. Equation 1 was used to fit the data. The data was fit on

MATLAB with using the `nlinfit` function as previously described^{13,14}. The detection software, CMEAnalysis, reports the fluorescent intensity of transmembrane fusion proteins, receptors, and/or ligands within clathrin-coated structures above the local membrane fluorescence intensity. The model in eq. (1) predicts the number of transmembrane fusion proteins within clathrin-coated structures as a function of the concentration of fusion proteins on the surrounding plasma membrane. To account for the difference in the detection software output and model prediction, a correction factor of $N_{\max} * C_{\text{mem}} * A_p$ was subtracted from Eq. 1. Here, A_p is the projected area of the fusion protein on the plasma membrane, $N_{\max}$, is the max number of fusion proteins that can be accommodated within clathrin-coated structures, and C_{mem} is the fluorescence intensity of fusion proteins on the plasma membrane surrounding the clathrin-coated structure. The best fit values for $N_{\max}$ corresponded to the maximum capacity of clathrin coated structures for each of the transmembrane fusion proteins.

Fluorescence Recovery after Photobleaching (FRAP) Assays

For the FRAP studies, a square region, 5.3 μm on each side, was photobleached on the plasma membrane of RPE cells expressing either TfR-GFP-OTR or TfR-GFP-10TR. After bleaching, the fluorescence recovery was imaged on the spinning disk confocal microscope for 5 minutes at 2 second framerate. The FRAP movies were analyzed in Fiji. The FRAP Profiler plugin was used to fit single exponential curves to the fluorescence recovery plots. The recovery plots were averaged across all cells for each condition. A two-tailed t-test conducted on the fit values of the mobile fraction and half-time of recovery.

Peanut agglutinin staining of live cell for fluorescence microscopy

PNA-Alexa647 (Invitrogen L32460) lyophilized powder was resuspended in DI water at a concentration of 1mg/mL and stored at -80°C in small aliquots. PNA-Alexa647 was diluted to 1 $\mu\text{g/mL}$ in 0.5% BSA + transfection media solution with 5 mM TCEP. Cells were treated with 1 $\mu\text{g/mL}$ PNA-Alexa647 and incubated for 10 minutes before imaging on a spinning disk confocal microscope.

Mass Spectroscopy

15 μL of the his-GFP-10TR protein purified from bacteria (210 μM) was diluted in 15 μL of phosphate buffered saline. This protein sample was buffer exchanged with 0.1% formic acid using a 10 kDa Princeton separations Centri Spin column (CS101). The buffer exchanged protein solution (71.9 μM) was then analyzed by electrospray ionization (ESI) technique using the ion trap detector. The output data was deconvoluted with Thermofisher Protein Deconvolution software. The measured mass of the most abundant fragment was 49754.08 Da (Figure S5), which is roughly equal to the calculated molecular mass of the protein (49908.37 Da). Notably, no molecules with

a mass significantly higher than the calculated mass of the protein were detected. This confirmed that the protein was deficient in glycosylation.

Protein identification was provided by the UT Austin Center for Biomedical Research Support Biological Mass Spectrometry Facility (RRID:SCR_021728). Research Resource Identifiers (RRID) can be used to easily identify and access details on the equipment utilized at core facilities by searching the RRID number on www.rrids.org.

Western blot analysis

For the western blot in Figure S6, GFP and bact-GFP-10TR were expressed and purified from E-Coli, as explained in the protein purification section. Mammalian GFP-10TR was expressed in RPE cells seeded in 6-well plates at 50,000 cells per well. The cells were lysed with RIPA buffer and were pooled across 18 wells. The cell lysate was agitated for 30 minutes at 4°C. Finally, the lysate was clarified by centrifuging the sample in a table-top centrifuge at 4°C for 20 minutes. All protein samples were resolved on a pre-cast SDS-PAGE gel (Bio-Rad Mini-PROTEAN TGX polyacrylamide gel). The protein gel was transferred onto a nitrocellulose membrane with Bio-Rad's Trans-Blot Turbo Transfer System at 1.5 Amps for 10 minutes. The membrane was blocked with 4% BSA-TPBS for 1 hour at room temperature, followed by incubation with primary antibodies at a 1:5000 dilution in 2% BSA-TPBS buffer overnight at 4°C. The secondary antibodies were diluted 1:3000 in 2% BSA-TPBS and incubated with the membrane for 1 hour at room temperature. The blot was developed with Pierce ECL western blotting substrate (Thermofisher). Chemiluminescence was imaged using the ImageQuant LAS 4000 imaging system.

ACKNOWLEDGEMENTS AND FUNDING SOURCES

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 National Institutes of Health 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 National Science Foundation, 2218467, to B.B. and J.C.S.

REFERENCES

- (1) Sorkin, A.; von Zastrow, M. Endocytosis and Signalling: Intertwining Molecular Networks. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (9), 609–622. <https://doi.org/10.1038/nrm2748>.
- (2) Goh, L. K.; Sorkin, A. Endocytosis of Receptor Tyrosine Kinases. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (5), a017459. <https://doi.org/10.1101/cshperspect.a017459>.
- (3) Vieira, A. V.; Lamaze, C.; Schmid, S. L. Control of EGF Receptor Signaling by Clathrin-Mediated Endocytosis. *Science* **1996**, *274* (5295), 2086–2089. <https://doi.org/10.1126/science.274.5295.2086>.

- (4) Cocucci, E.; Aguet, F.; Boulant, S.; Kirchhausen, T. The First Five Seconds in the Life of a Clathrin-Coated Pit. *Cell* **2012**, *150* (3), 495–507. <https://doi.org/10.1016/j.cell.2012.05.047>.
- (5) Kaksonen, M.; Roux, A. Mechanisms of Clathrin-Mediated Endocytosis. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (5), 313–326. <https://doi.org/10.1038/nrm.2017.132>.
- (6) Schmid, E. M.; McMahon, H. T. Integrating Molecular and Network Biology to Decode Endocytosis. *Nature* **2007**, *448* (7156), 883–888. <https://doi.org/10.1038/nature06031>.
- (7) Antonny, B.; Burd, C.; De Camilli, P.; Chen, E.; Daumke, O.; Faelber, K.; Ford, M.; Frolov, V.; Frost, A.; Hinshaw, J.; Kirchhausen, T.; Kozlov, M.; Lenz, M.; Low, H.; McMahon, H.; Merrifield, C.; Pollard, T.; Robinson, P.; Roux, A.; Schmid, S. Membrane Fission by Dynamin: What We Know and What We Need to Know. *Embo J.* **2016**, *35* (21), 2270–2284. <https://doi.org/10.15252/embj.201694613>.
- (8) Ohno, H.; Stewart, J.; Fournier, M. C.; Bosshart, H.; Rhee, I.; Miyatake, S.; Saito, T.; Gallusser, A.; Kirchhausen, T.; Bonifacino, J. S. Interaction of Tyrosine-Based Sorting Signals with the Medium Chains of Clathrin-Associated Protein. *Mol. Biol. Cell* **1995**, *6*, 398–398.
- (9) Kelly, B. T.; McCoy, A. J.; Späte, K.; Miller, S. E.; Evans, P. R.; Höning, S.; Owen, D. J. A Structural Explanation for the Binding of Endocytic Dileucine Motifs by the AP2 Complex. *Nature* **2008**, *456* (7224), 976–979. <https://doi.org/10.1038/nature07422>.
- (10) Traub, L. M. Tickets to Ride: Selecting Cargo for Clathrin-Regulated Internalization. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (9), 583–596. <https://doi.org/10.1038/nrm2751>.
- (11) DeGroot, A. C. M.; Busch, D. J.; Hayden, C. C.; Mihelic, S. A.; Alpar, A. T.; Behar, M.; Stachowiak, J. C. Entropic Control of Receptor Recycling Using Engineered Ligands. *Biophys. J.* **2018**, *114* (6), 1377–1388. <https://doi.org/10.1016/j.bpj.2018.01.036>.
- (12) Liu, A. P.; Aguet, F.; Danuser, G.; Schmid, S. L. Local Clustering of Transferrin Receptors Promotes Clathrin-Coated Pit Initiation. *J. Cell Biol.* **2010**, *191* (7), 1381–1393. <https://doi.org/10.1083/jcb.201008117>.
- (13) Zhao, C.; DeGroot, A. C. M.; Hayden, C. C.; Houser, J. R.; Ali, H. A.; LaMonica, M. F.; Stachowiak, J. C. Receptor Heterodimerization Modulates Endocytosis through Collaborative and Competitive Mechanisms. *Biophys J* **2019**, *117* (4), 646–658. <https://doi.org/10.1016/j.bpj.2019.07.012>.
- (14) DeGroot, A. C. M.; Zhao, C.; LaMonica, M. F.; Hayden, C. C.; Stachowiak, J. C. Molecular Thermodynamics of Receptor Competition for Endocytic Uptake. *Soft Matter* **2019**, *15* (37), 7448–7461. <https://doi.org/10.1039/c9sm00876d>.
- (15) DeGroot, A. C. M.; Gollapudi, S.; Zhao, C.; LaMonica, M. F.; Stachowiak, J. C. Weakly Internalized Receptors Use Coated Vesicle Heterogeneity to Evade Competition during Endocytosis. *Biochemistry* **2021**, *60* (27), 2195–2205. <https://doi.org/10.1021/acs.biochem.1c00268>.
- (16) Kornfeld, R.; Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annu. Rev. Biochem.* **1985**, *54* (1), 631–664. <https://doi.org/10.1146/annurev.bi.54.070185.003215>.
- (17) Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. Glycobiology. *Annu. Rev. Biochem.* **1988**, *57* (1), 785–838. <https://doi.org/10.1146/annurev.bi.57.070188.004033>.
- (18) Varki, A.; Freeze, H. H.; Manzi, A. E. Preparation and Analysis of Glycoconjugates. *Curr. Protoc. Mol. Biol.* **2009**, *88* (1), 17.0.1–17.0.12. <https://doi.org/10.1002/0471142727.mb1700s88>.
- (19) Opdenakker, G.; Rudd, P. M.; Ponting, C. P.; Dwek, R. A. Concepts and Principles of Glycobiology. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **1993**, *7* (14), 1330–1337. <https://doi.org/10.1096/fasebj.7.14.8224606>.
- (20) Clausen, H.; Wandall, H. H.; DeLisa, M. P.; Stanley, P.; Schnaar, R. L. Glycosylation Engineering. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J.

- H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2022.
- (21) Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. Concepts and Principles of O-Linked Glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33* (3), 151–208. <https://doi.org/10.1080/10409239891204198>.
 - (22) Mulagapati, S.; Koppolu, V.; Raju, T. S. Decoding of O-Linked Glycosylation by Mass Spectrometry. *Biochemistry* **2017**, *56* (9), 1218–1226. <https://doi.org/10.1021/acs.biochem.6b01244>.
 - (23) Rudd, P. M.; Dwek, R. A. Glycosylation: Heterogeneity and the 3D Structure of Proteins. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32* (1), 1–100. <https://doi.org/10.3109/10409239709085144>.
 - (24) Chen, X.; Varki, A. Advances in the Biology and Chemistry of Sialic Acids. *ACS Chem. Biol.* **2010**, *5* (2), 163–176. <https://doi.org/10.1021/cb900266r>.
 - (25) Hattstrup, C. L.; Gendler, S. J. Structure and Function of the Cell Surface (Tethered) Mucins. *Annu. Rev. Physiol.* **2008**, *70*. <https://doi.org/10.1146/annurev.physiol.70.113006.100659>.
 - (26) Kufe, D. W. Mucins in Cancer: Function, Prognosis and Therapy. *Nat. Rev. Cancer* **2009**, *9* (12), 874–885. <https://doi.org/10.1038/nrc2761>.
 - (27) Gupta, R.; Leon, F.; Rauth, S.; Batra, S. K.; Ponnusamy, M. P. A Systematic Review on the Implications of O-Linked Glycan Branching and Truncating Enzymes on Cancer Progression and Metastasis. *Cells* **2020**, *9* (2). <https://doi.org/10.3390/cells9020446>.
 - (28) Denneny, E.; Sahota, J.; Beatson, R.; Thornton, D.; Burchell, J.; Porter, J. Mucins and Their Receptors in Chronic Lung Disease. *Clin. Transl. Immunol.* **2020**, *9* (3). <https://doi.org/10.1002/cti2.1120>.
 - (29) Rose, M. C.; Voynow, J. A. Respiratory Tract Mucin Genes and Mucin Glycoproteins in Health and Disease. *Physiol. Rev.* **2006**, *86* (1), 245–279.
 - (30) Kinlough, C. L.; Poland, P. A.; Bruns, J. B.; Harkleroad, K. L.; Hughey, R. P. MUC1 Membrane Trafficking Is Modulated by Multiple Interactions. *J. Biol. Chem.* **2004**, *279* (51). <https://doi.org/10.1074/jbc.M409360200>.
 - (31) Altschuler, Y.; Kinlough, C. L.; Poland, P. A.; Bruns, J. B.; Apodaca, G.; Weisz, O. A.; Hughey, R. P. Clathrin-Mediated Endocytosis of MUC1 Is Modulated by Its Glycosylation State. *Mol. Biol. Cell* **2000**, *11* (3), 819–831.
 - (32) Liu, X.; Yuan, Z.; Chung, M. MUC1 Intra-Cellular Trafficking Is Clathrin, Dynamin, and Rab5 Dependent. *Biochem. Biophys. Res. Commun.* **2008**, *376* (4). <https://doi.org/10.1016/j.bbrc.2008.09.065>.
 - (33) Chen, W.; Zhang, Z.; Zhang, S.; Zhu, P.; K., Ko. J.; Yung, K. MUC1: Structure, Function, and Clinic Application in Epithelial Cancers. *Int. J. Mol. Sci.* **2021**, *22* (12). <https://doi.org/10.3390/ijms22126567>.
 - (34) Razawi, H.; Kinlough, C. L.; Staubach, S.; Poland, P. A.; Rbaibi, Y.; Weisz, O. A.; Hughey, R. P.; Hanisch, F.-G. Evidence for Core 2 to Core 1 O-Glycan Remodeling during the Recycling of MUC1. *Glycobiology* **2013**, *23* (8), 935–945. <https://doi.org/10.1093/glycob/cwt030>.
 - (35) Storr, S. J.; Tumour Immunology Group, U. of N. D. of B. S. N. C. H. H. R. N. N. G. P. B.; Royle, L.; Department of Biochemistry, O. G. I. O. U. S. P. R. O. O. X. Q. U.; Chapman, C. J.; Hamid, U. M. A.; Robertson, J. F.; Murray, A.; Oncimmune Ltd, C. S. B. N. C. H. H. R. N. N. G. P. B. U. K.; Dwek, R. A.; Rudd, P. M. The O-Linked Glycosylation of Secretory/Shed MUC1 from an Advanced Breast Cancer Patient's Serum. *Glycobiology* **2008**, *18* (6), 456–462. <https://doi.org/10.1093/glycob/cwn022>.
 - (36) Andersch-Björkman, Y.; Thomsson, K. A.; Holmén Larsson, J. M.; Ekerhovd, E.; Hansson, G. C. Large Scale Identification of Proteins, Mucins, and Their O-Glycosylation in the Endocervical Mucus during the Menstrual Cycle*. *Mol. Cell. Proteomics* **2007**, *6* (4), 708–716. <https://doi.org/10.1074/mcp.M600439-MCP200>.

- (37) Müller, S.; Alving, K.; Peter-Katalinic, J.; Zachara, N.; Gooley, A. A.; Hanisch, F.-G. High Density O-Glycosylation on Tandem Repeat Peptide from Secretory MUC1 of T47D Breast Cancer Cells*. *J. Biol. Chem.* **1999**, 274 (26), 18165–18172. <https://doi.org/10.1074/jbc.274.26.18165>.
- (38) Shurer, C. R.; Kuo, J. C.-H.; Roberts, L. M.; Gandhi, J. G.; Colville, M. J.; Enoki, T. A.; Pan, H.; Su, J.; Noble, J. M.; Hollander, M. J.; O'Donnell, J. P.; Yin, R.; Pedram, K.; Möckl, L.; Kourkoutis, L. F.; Moerner, W. E.; Bertozzi, C. R.; Feigenson, G. W.; Reesink, H. L.; Paszek, M. J. Physical Principles of Membrane Shape Regulation by the Glycocalyx. *Cell* **2019**, 177 (7), 1757–1770.e21. <https://doi.org/10.1016/j.cell.2019.04.017>.
- (39) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer Scaling Laws of Unfolded and Intrinsically Disordered Proteins Quantified with Single-Molecule Spectroscopy. *Proc. Natl. Acad. Sci.* **2012**, 109 (40), 16155–16160. <https://doi.org/10.1073/pnas.1207719109>.
- (40) Lu, C.-H.; Pedram, K.; Tsai, C.-T.; Jones, T.; Li, X.; Nakamoto, M. L.; Bertozzi, C. R.; Cui, B. Membrane Curvature Regulates the Spatial Distribution of Bulky Glycoproteins. *Nat. Commun.* **2022**, 13 (1), 1–16. <https://doi.org/10.1038/s41467-022-30610-2>.
- (41) Harding, C.; Heuser, J.; Stahl, P. Receptor-Mediated Endocytosis of Transferrin and Recycling of the Transferrin Receptor in Rat Reticulocytes. *J. Cell Biol.* **1983**, 97 (2), 329–339. <https://doi.org/10.1083/jcb.97.2.329>.
- (42) Collawn, J. F.; Stangel, M.; Kuhn, L. A.; Esekogwu, V.; Jing, S. Q.; Trowbridge, I. S.; Tainer, J. A. Transferrin Receptor Internalization Sequence YXRF Implicates a Tight Turn as the Structural Recognition Motif for Endocytosis. *Cell* **1990**, 63 (5), 1061–1072.
- (43) Collawn, J. F.; Lai, A.; Domingo, D.; Fitch, M.; Hatton, S.; Trowbridge, I. S. YTRF Is the Conserved Internalization Signal of the Transferrin Receptor, and a Second Ytrf Signal at Position 31-34 Enhances Endocytosis. *J. Biol. Chem.* **1993**, 268, 21686–21692.
- (44) Park, S.; Colville, M. J.; Shurer, C. R.; Huang, L.-T.; Kuo, J. C.-H.; Paek, J. H.; Goudge, M. C.; Su, J.; DeLisa, M. P.; Lammerding, J.; Zipfel, W. R.; Fischbach, C.; Reesink, H. L.; Paszek, M. J. Mucins Form a Nanoscale Material Barrier against Immune Cell Attack. *bioRxiv* January 29, 2022, p 2022.01.28.478211. <https://doi.org/10.1101/2022.01.28.478211>.
- (45) Busch, D. J.; Houser, J. R.; Hayden, C. C.; Sherman, M. B.; Lafer, E. M.; Stachowiak, J. C. Intrinsically Disordered Proteins Drive Membrane Curvature. *Nat. Commun.* **2015**, 6, 7875. <https://doi.org/10.1038/ncomms8875>.
- (46) Aguet, F.; Antonescu, C. N.; Mettlen, M.; Schmid, S. L.; Danuser, G. Advances in Analysis of Low Signal-to-Noise Images Link Dynamin and AP2 to the Functions of an Endocytic Checkpoint. *Dev. Cell* **2013**, 26 (3), 279–291.
- (47) Axelrod, D. Total Internal Reflection Fluorescence Microscopy in Cell Biology. *Traffic Cph. Den.* **2001**, 2 (11). <https://doi.org/10.1034/j.1600-0854.2001.21104.x>.
- (48) Saffman, P. G.; Delbrück, M. Brownian Motion in Biological Membranes. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, 72 (8), 3111–3113.
- (49) Rosano, G. L.; Ceccarelli, E. A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. *Front. Microbiol.* **2014**, 5, 172. <https://doi.org/10.3389/fmicb.2014.00172>.
- (50) Tokmakov, A. A.; Kurotani, A.; Takagi, T.; Toyama, M.; Shirouzu, M.; Fukami, Y.; Yokoyama, S. Multiple Post-Translational Modifications Affect Heterologous Protein Synthesis *. *J. Biol. Chem.* **2012**, 287 (32), 27106–27116. <https://doi.org/10.1074/jbc.M112.366351>.
- (51) Jenkins, N.; Parekh, R. B.; James, D. C. Getting the Glycosylation Right: Implications for the Biotechnology Industry. *Nat. Biotechnol.* **1996**, 14 (8), 975–981. <https://doi.org/10.1038/nbt0896-975>.

- (52) Letourneur, O.; Sechi, S.; Willette-Brown, J.; Robertson, M. W.; Kinet, J.-P. Glycosylation of Human Truncated FcεRI α Chain Is Necessary for Efficient Folding in the Endoplasmic Reticulum (*). *J. Biol. Chem.* **1995**, *270* (14), 8249–8256. <https://doi.org/10.1074/jbc.270.14.8249>.
- (53) Kubala, M. H.; Kovtun, O.; Alexandrov, K.; Collins, B. M. Structural and Thermodynamic Analysis of the GFP:GFP-Nanobody Complex. *Protein Sci. Publ. Protein Soc.* **2010**, *19* (12), 2389–2401. <https://doi.org/10.1002/pro.519>.
- (54) Sharma, V.; Molecular Biophysics Unit, I. I. of S. B. I.; Srinivas, V. R.; Adhikari, P.; Vijayan, M.; Surolia, A. 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* **1998**, *8* (10), 1007–1012. <https://doi.org/10.1093/glycob/8.10.1007>.
- (55) Mansour, M. H.; Negm, H. I.; Saad, A. H.; Zahran, A. Y.; Badir, N. Identification of Peanut Agglutinin-Binding Glycoproteins on Lizard Lymphocytes. *Zoolog. Sci.* **1995**, *12* (1), 79–85. <https://doi.org/10.2108/zsj.12.79>.
- (56) Wesseling, J.; van der Valk, S. W.; Vos, H. L.; Sonnenberg, A.; Hilken, J. Episialin (MUC1) Overexpression Inhibits Integrin-Mediated Cell Adhesion to Extracellular Matrix Components. *J. Cell Biol.* **1995**, *129* (1), 255–265. <https://doi.org/10.1083/jcb.129.1.255>.
- (57) Lloyd, K. O.; Burchell, J.; Kudryashov, V.; Yin, B. W. T.; Taylor-Papadimitriou, J. 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.* **1996**, *271* (52), 33325–33334. <https://doi.org/10.1074/jbc.271.52.33325>.
- (58) Li, Y.; Kufe, D. The Human DF3/MUC1 Carcinoma-Associated Antigen Signals Nuclear Localization of the Catenin P120(Ctn). *Biochem. Biophys. Res. Commun.* **2001**, *281* (2), 440–443. <https://doi.org/10.1006/bbrc.2001.4383>.
- (59) Kato, K.; Chang, E. H.; Chen, Y.; Lu, W.; Kim, M. M.; Niihori, M.; Hecker, L.; Kim, K. C. MUC1 Contributes to Goblet Cell Metaplasia and MUC5AC Expression in Response to Cigarette Smoke in Vivo. *Am. J. Physiol. - Lung Cell. Mol. Physiol.* **2020**, *319* (1), L82–L90. <https://doi.org/10.1152/ajplung.00049.2019>.
- (60) Qu, J.; Cheng, Y.; Wu, W.; Yuan, L.; Liu, X. Glycocalyx Impairment in Vascular Disease: Focus on Inflammation. *Front. Cell Dev. Biol.* **2021**, *9*, 730621. <https://doi.org/10.3389/fcell.2021.730621>.
- (61) C., S. J.; Schmid, E. M.; Ryan, C. J.; Ann, H. S.; Sasaki, D. Y.; Sherman, M. B.; Geissler, P. L.; Fletcher, D. A.; Hayden. Membrane Bending by Protein-Protein Crowding. *Nat. Cell Biol.* **2012**, *14* (9). <https://doi.org/10.1038/ncb2561>.

Supporting Information for

Steric pressure between glycosylated transmembrane proteins inhibits internalization by endocytosis

Authors

Sadhana Gollapudi¹, Sabah Jamal¹, Advika Kamatar¹, Feng Yuan¹, Liping Wang³, Eileen M. Lafer³, Brian Belardi⁴, Jeanne C. Stachowiak^{1,2*}

¹*Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX*

²*Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX*

³*University of Texas Health Science Center at San Antonio, Department of Biochemistry and Structural Biology, San Antonio, TX*

⁴*McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX*

**To whom correspondence should be addressed: jcstach@austin.utexas.edu*

This PDF file includes:

Figures S1 to S5

Supplementary Figures

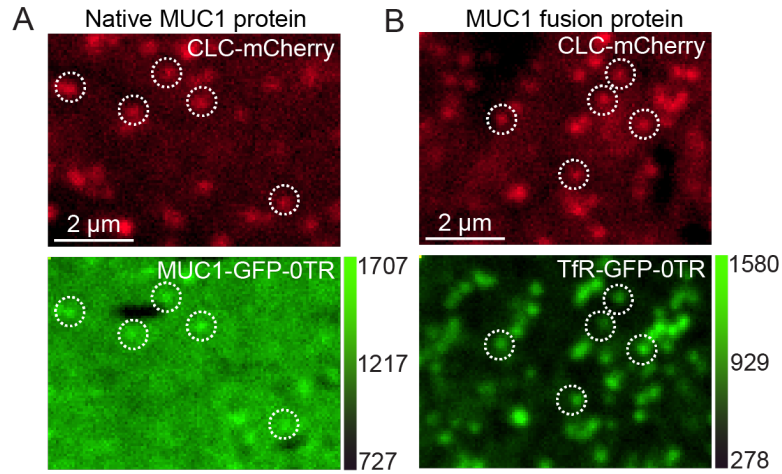


FIGURE S1 Native MUC1 displayed weak colocalization with clathrin-coated structures. Spinning disk confocal images of the plasma membrane of RPE-CLC-mCherry cells transiently expressing (A) the native MUC1 transmembrane and cytoplasmic domains fused to GFP, or (B) the transmembrane and cytoplasmic domains of transferrin receptor fused to GFP. Both these transmembrane proteins in (A) and (B) had zero MUC1 tandem repeats on the ectodomain. The images in (A) and (B) have to very similar image brightness and contrast settings.

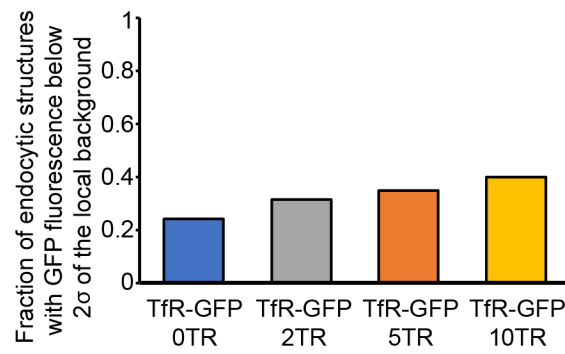


FIGURE S2. Fraction of endocytic structures that have GFP fluorescence below the detection threshold for cells expressing transferrin transmembrane proteins with different number of MUC1 tandem repeats. The bar plot shows the fraction of endocytic structures detected by CMEAnalysis with a GFP fluorescence lower than two times the standard deviation of the local background fluorescence. As the signal-to-noise ratio of the images increases, the number of detections with fluorescence below the threshold increases.

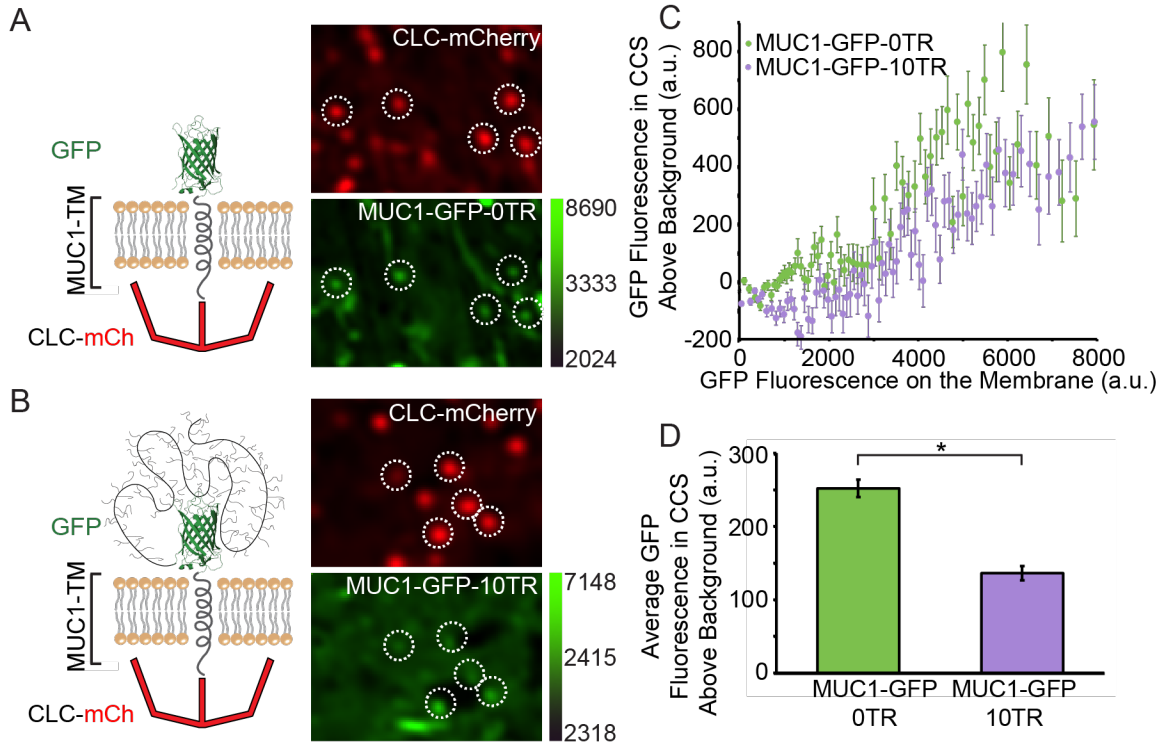


FIGURE S3. Recruitment of native MUC1 protein into clathrin-coated structures increases in the absence of the ten tandem repeat domain. (A), (B) are representative deconvoluted spinning disk confocal images of the plasma membrane of RPE cells expressing either (A) MUC1-GFP-0TR or (B) MUC1-GFP-10TR. (C) The relative number of MUC1 proteins within clathrin-coated structures is shown as a function of the relative concentration of the protein on the plasma membrane surrounding each structure. Each point represents the average data from 200 clathrin-coated structures binned by the relative concentration of the MUC1 protein on the membrane. A total of 28791 CCSs were detected from 33 RPE cells expressing MUC1-GFP-0TR, and 30883 CCSs were detected from 32 RPE cells expressing MUC1-GFP-10TR. Error bars represent mean $\pm$ SE. (D) Bar plot of the average protein fluorescence in CCSs for both MUC1-GFP-0TR and MUC1-GFP-10TR. A two-sample t-test was conducted on the average GFP fluorescence in CCSs. The p-values were < 0.05 between MUC1-GFP-0TR and MUC1-GFP-10TR suggesting a statistically significant difference between the average fluorescence values. The error bars represent the mean $\pm$ SE.

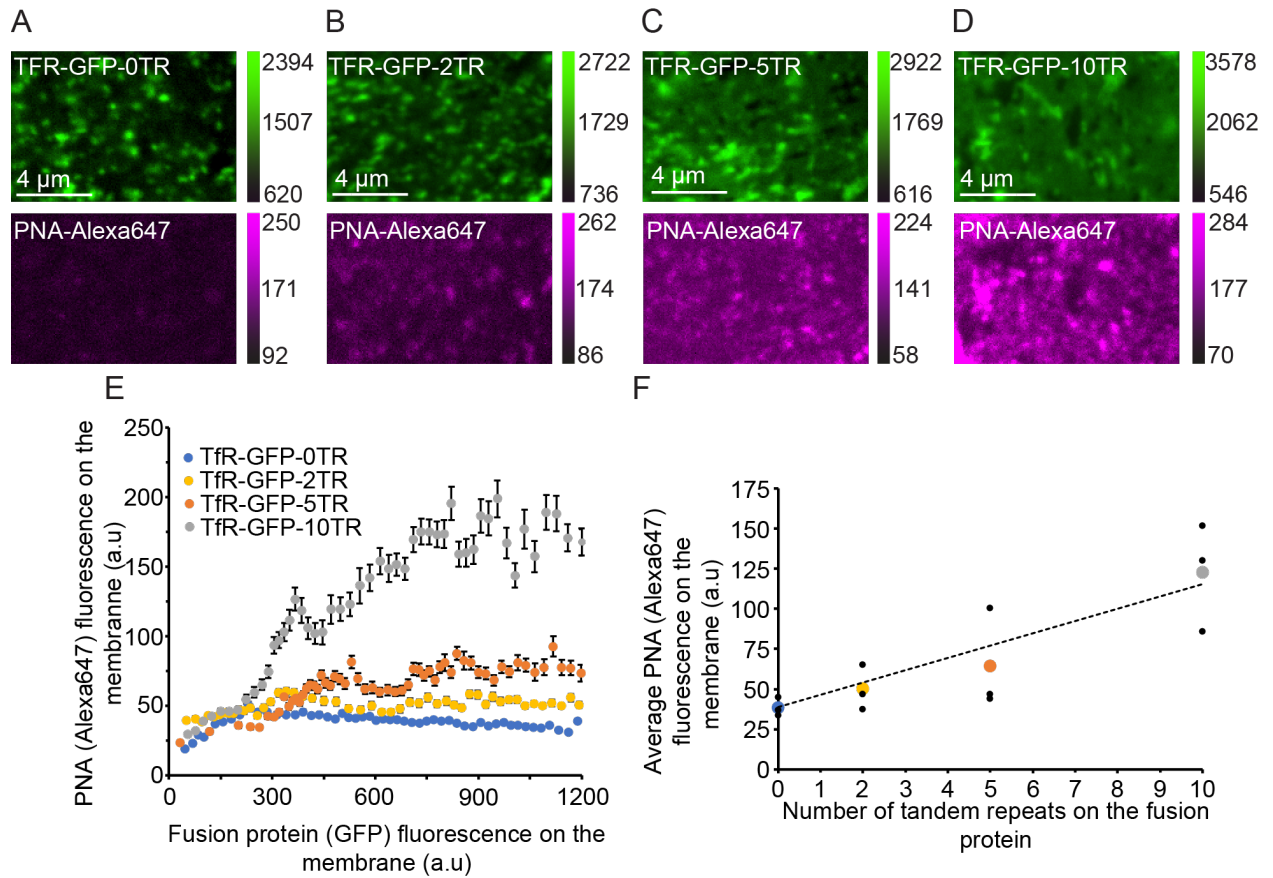


FIGURE S4. Glycosylation of transmembrane fusion proteins increases as the number of tandem repeats on the ectodomain increases. (A-D) Spinning disk images of the plasma membrane of RPE cells stained with PNA-Alexa647 and transiently expressing (A) TfR-GFP-0TR, (B) TfR-GFP-2TR, (C) TfR-GFP-5TR or (D) TfR-GFP-10TR. (E) Plot showing the amount of PNA-Alexa647 staining of the plasma membrane for each of the transmembrane fusion proteins (A-D), as a function of the local fluorescence of proteins 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 protein. A total of 24422 CCSs were detected from 47 RPE cells expressing TfR-GFP-0TR, 23754 CCSs were detected from 52 cells expressing TfR-GFP-2TR, 16334 CCSs were detected from 46 cells expressing TfR-GFP-5TR, 19051 CCSs detected from 53 cells expressing TfR-GFP-10TR. Error bars represent mean $\pm$ SE. (F) Scatterplot of the average fluorescence of PNA-Alexa647 on the membrane surrounding all clathrin-coated structures. The error bars represent mean $\pm$ SE.

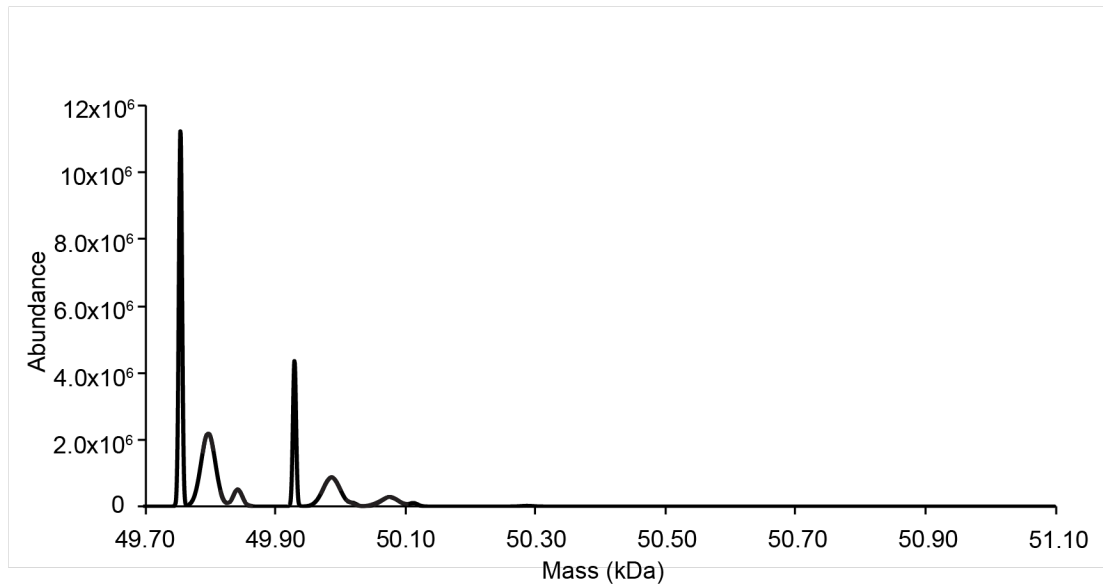


FIGURE S5 Molecular weight mass spectroscopy of the bacterially expressed ligand his-GFP-10TR. The purified protein sample was buffer exchanged into 0.1% formic acid at a concentration of 71.9 μ M. The protein sample was then run by electrospray ionization. The assay revealed that the most abundant protein molecule had a molecular weight of 49754.08 kDa. The above plot also shows the presence of very little high molecular weight molecules implying that the ligand molecule was very minimally glycosylated.

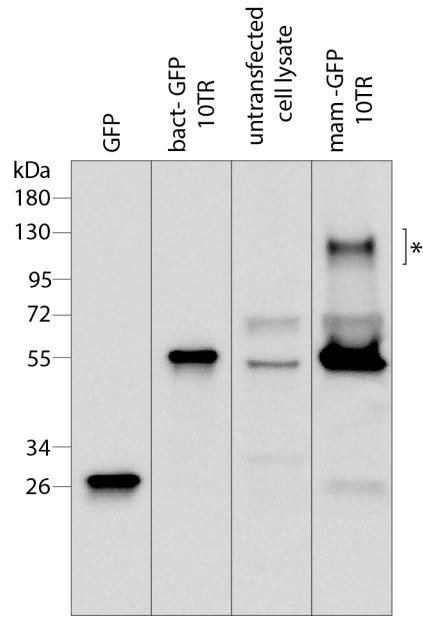


FIGURE S6. Western blot of GFP, bact-GFP-10TR, and mam-GFP-10TR illustrate an increase in molecular weight of mammalian GFP-10TR due to glycosylation. Anti-GFP primary antibodies were used to blot against the GFP in the three ligands followed by HRP conjugated secondary antibodies. Chemiluminescence was detected using ImageQuant LAS 4000 to take images at 60s exposure. The band at ~55 kDa in the lane with untransfected cell lysate is due to non-specific binding of the antibodies to the proteins within the cell lysate. *The higher molecular weight bands, which are absent in the in the lane with the untransfected cell lysate, indicate the presence of glycosylated mam-GFP-10TR ligand.

Methods and Materials

Plasmid Constructs

Plasmids for the expression of TfR -GFP-10TR, TfR-GFP-5TR, TfR-GFP-2TR, and TfR-0TR were generated by inserting TfR-GFP and one of the above MUC1 tandem repeat fragments by Gibson Assembly cloning. The gene for the transmembrane fusion protein was inserted into a Tetracycline inducible PiggyBac expression vector. backbone vector used, pPB_MUC1_10_mOXGFP_dCT GFP BlnI RRK, was a gift from the Paszek lab (Cornell University). The MUC1_10_mOXGFP gene in the open reading frame of the vector was replaced with the genes of interest. The pEGFPN1-TfR-GFP⁴⁵ plasmid described previously was used as the template for PCR amplification of TfR-GFP DNA fragment. The plasmid encoded the intracellular and transmembrane domains of the transferrin receptor, amino acids 1-88 of GenBank accession number AAA61153. The transferrin receptor domains were fused to GFP with a 9-amino acid sequence linker (GKGDPPVAT). The TfR-GFP gene was amplified by PCR from the plasmid using the forward primer, CTCTTAAGGCTAGAGGATCCATGGATCAAGCTAGATCAGCATTCTCT and reverse primer, GACTGGGTGCCCGGTGTCATCTTGTACAGCTCGTCCATGCC. The forward primer contained a 20-amino acid overlap on the 5-prime end, and the reverse primer contained a 20-amino acid overlap with the tandem repeat domain to be inserted. For the TfR- Δ ecto-GFP-10TR variant, a gene fragment for the 10TR domain was purchased from Integrated DNA Technologies (gBlocks). Similarly, 5TR and 2TR domain gene fragments were also purchased. Each of the gene fragments had a 20-amino acid overlap with the TfR-GFP PCR amplicon on the 5-prime end and with the vector on the 3-prime end. The vector, TfR-GFP PCR amplification product and one of the three tandem repeat domain gene fragments were ligated using the NEB Gibson Assembly master mix (NEB E2611L). The Gibson Assembly reaction product was then transformed into DH5 bacterial cells, spread on Ampicillin plates and grown for 16 hours. The colonies were screened for successful insertion of the DNA fragments.

The plasmid for expressing his-GFP-10TR in bacteria was generated by inserting the his-GFP-10TR gene into the pEGFPN1 vector by restriction cloning into the pET28a(+) vector. A gene fragment for his-GFP-10TR was purchased from Integrated DNA Technologies (gBlocks). The fragment included restriction enzyme cut sites for BamHI and EcoRI on the 5-prime and 3-prime ends respectively. The gene fragment and the pET28a(+) vector were digested with BamHI-HF (NEB R3101S) and EcoRI-HF (NEB R3136S) enzymes at 37°C for 15 minutes. The digested products were ligated with Quick Ligase (NEB M2200S) at room temperature for 10 minutes. The ligation reaction product was then transformed into DH5 bacterial cells, spread on Ampicillin plates and grown for 16 hours. The colonies were screened for successful insertion of the DNA fragments.

The plasmid for expressing GFP-10TR in mammalian cells was generated by inserting the his-GFP-10TR gene into the pEGFPN1 vector by restriction cloning. A gene fragment for his-GFP-10TR was purchased from IDT. The gene fragment included restriction enzyme cut sites for XhoI and NotI on the 5-prime and 3-prime ends respectively. For the gene to be secreted from mammalian cells, the EGFR signal sequence (MRPSGTAGAALLALLAALCPASRA) was included on the N-terminus of the gene. The vector, and the gene fragment were digested with XhoI (NEB R0146S) and NotI-HF (NEB R3189S) enzymes at 37°C for 1 hour.

The plasmid for expression of TfR-BFP-GFPnb was generated as previously described¹¹.

Gene Fragments

10TR repeat domain for TfR-GFP-10TR

CGGCCTGTACAAGCCAGATACAAGACCGGCCCCAGGATCTACGGCTCCTC
CGGCTCATGGAGTCACTTCTGCTCCAGACACAAGGCCCGCGCCGGGTTCTA
CAGCACCGCCTGCTCATGGTGTTACTAGCGCACCCGATACGAGACCTGCTC
CGGGATCAACGGGCACCTCCTGCCCCACGGGGTAACATCTGCACCGGACACT
CGCCCTGCGCCCGGTTCAACCGCTCCACCCGCACACGGAGTGACAAGCGC
TCCTGACACTAGACCAGCACCAAGGTTCTACAGCCCCACCAGCCCATGGAGT
TACCAGTGCACCAGATACTAGGCCAGCTCCAGGTAGTACTGCACCCCCAGC
TCATGGGGTTACATCAGCTCCCGACACGCGACCAGCTCCTGGAAGCACTGC
CCCTCCAGCTCACGGTGTGACCTCAGCACCTGATACACGCCCTGCACCTGG
CTCTACTGCTCCCCCGCTCATGGCGTAAGTAGTGCCCCGGATACTCGACC
CGCCCCTGGTTCCACAGCTCCGCCAGCACATGGTGTAACAAGTGCTCCTGA
TACCCGACCAGCGCCTGGAAGTACCGCACCACTGCACATGGAGTAAGTTC
AGCCGCCTCAGGCTCTGCATCAGGCTCAGCTTAGGAATTCCGGCC

5TR repeat domain for TfR-GFP-5TR

CGGCCTGTACAAGCCAGATACAAGACCGGCCCCAGGATCTACGGCTCCTC
CGGCTCATGGAGTCACTTCTGCTCCAGACACAAGGCCCGCGCCGGGTTCTA
CAGCACCGCCTGCTCATGGTGTTACTAGCGCACCCGATACGAGACCTGCTC
CGGGATCAACGGGCACCTCCTGCCCCACGGGGTAACATCTGCACCGGACACT
CGCCCTGCGCCCGGTTCAACCGCTCCACCCGCACACGGAGTGACAAGCGC
TCCTGACACTAGACCAGCACCAAGGTTCTACAGCCCCACCAGCCCATGGAGT
TACCAGTGCAGCCTCAGGCTCTGCATCAGGCTCAGCTTAGGAATTCCGGCC

2TR repeat domain TfR-GFP-2TR

CGGCCTGTACAAGCCAGATACAAGACCGGCCCCAGGATCTACGGCTCCTC
CGGCTCATGGAGTCACTTCTGCTCCAGACACAAGGCCCGCGCCGGGTTCTA
CAGCACCGCCTGCTCATGGTGTTACTAGCGCAGCCTCAGGCTCTGCATCAG
GCTCAGCTTAGGAATTCCGGCC

his-GFP-10TR fragment for the bact-GFP-10TR protein

CGGCCGGATCCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGT
GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
TGTCGGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
TTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGAC
CACCTGACCTACGGCGTGCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
GCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG
AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAAA
CAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGT
GAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCG
ACCACTACCAGCAGAACACCCCCATCGGGCGACGGCCCCGTGCTGCTGCC
GACAACCACTACCTGAGCACCCAGTCCAACTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCAC
TCTCGGCATGGACGAGCTGTACAAGCCAGATACAAGACCGGCCCCAGGAT
CTACGGCTCCTCCGGCTCATGGAGTCACTTCTGCTCCAGACACAAGGCCCG
CGCCGGGTTCTACAGCACCGCCTGCTCATGGTGTTACTAGCGCACCCGATA
CGAGACCTGCTCCGGGATCAACGGCACCTCCTGCCACGGGGTAACATCT
GCACCGGACACTCGCCCTGCGCCCCGGTTCAACCGCTCCACCCGCACACGG
AGTGACAAGCGCTCCTGACACTAGACCAGCACCAGGTTCTACAGCCCCACC
AGCCCATGGAGTTACAGTGACACAGATACTAGGCCAGCTCCAGGTAGTAC
TGCACCCCCAGCTCATGGGGTTACATCAGCTCCCGACACGCGACAGCTC
CTGGAAGCACTGCCCCTCCAGCTCACGGTGTGACCTCAGCACCTGATACAC
GCCCTGCACCTGGCTCTACTGCTCCCCCGCTCATGGCGTAACTAGTGCCC
CGGATACTCGACCCGCCCTGGTTCCACAGCTCCGCCAGCACATGGTGTA
CAAGTGCTCCTGATACCCGACCAGCGCCTGGAAGTACCGCACACCTGCA
CATGGAGTAACTTCAGCCGCCTCAGGCTCTGCATCAGGCTCAGCTTAGGAA
TTCCGGCC

his-GFP-10TR fragment for the mammalian expressed GFP-10TR protein

CGGCCCTCGAGCATCATCATCATCACATGGTGAGCAAGGGCGAGGAG
CTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC
GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCC
CTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAAGTCTCAGCC
GCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCG
AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACA
GCTGGAGTACAACATAACAGCCACAACGTCTATATCATGGCCGACAAGCA
GAAGAACGGCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGG
CAGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACG
GCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAACTGA
GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTG
ACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGCCAGATAC
AAGACCGGCCCCAGGATCTACGGCTCCTCCGGCTCATGGAGTCACTTCTGC

TCCAGACACAAGGCCCGCGCCGGGTTCTACAGCACCGCCTGCTCATGGTG
 TTACTAGCGCACCCGATACGAGACCTGCTCCGGGATCAACGGGCACCTCCTG
 CCCACGGGGTAACATCTGCACCGGACACTCGCCCTGCGCCCGGTTCAACC
 GCTCCACCCGCACACGGAGTGACAAGCGCTCCTGACACTAGACCAGCACC
 AGGTTCTACAGCCCCACCAGCCCATGGAGTTACCAGTGCACCAGATACTAG
 GCCAGCTCCAGGTAGTACTGCACCCCCAGCTCATGGGGTTACATCAGCTCC
 CGACACGCGACCAGCTCCTGGAAGCACTGCCCCTCCAGCTCACGGTGTGA
 CCTCAGCACCTGATACACGCCCTGCACCTGGCTCTACTGCTCCCCCGCTC
 ATGGCGTAAGTAGTGCCCCGGATACTCGACCCGCCCTGGTTCCACAGCTC
 CGCCAGCACATGGTGTAACAAGTGCTCCTGATACCCGACCAGCGCCTGGAA
 GTACCGCACCACTGCACATGGAGTAACTTCAGCCGCCTCAGGCTCTGCAT
 CAGGCTCAGCTTAGGCGGCGCCGGCC

Protein Purification

The his-GFP protein was expressed and purified from BL21 pLysS bacterial cells as previously described^{45,61}.

his-GFP-10TR protein was expressed and purified from BL21 pLysS bacterial cells. The cells were grown in 1 L 2XTY medium for 2.5h at 37 °C to an optical density (~0.7 OD₆₀₀) and then protein expression was induced with 1mM IPTG at 16 °C overnight. The cells were centrifuged at 11,000 x g for 15 minutes at 4 °C. The rest of the protocol was performed at 4 °C. The cell pellet was resuspended with a mortar and pestle in lysis buffer consisting of 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM β-ME, 1% Triton X-100 and EDTA-free protease inhibitor cocktail tablet (Thermo Scientific, Cat. # A32965) for 5 minutes on ice and then sonicated. The mixture was then clarified by ultracentrifugation at 134000 xg for 40 mins. The supernatant, which contained the protein, was incubated with 12 mL Ni-NTA resin slurry that contains 50% beads (GenScript, Cat. # L00223-25). The resin and supernatant mixture was gently stirred for 2 hours. The protein bound Ni-NTA resin was transferred into a chromatography column and allowed to settle. The Ni-NTA beads were thoroughly washed with a buffer containing 20 mM sodium phosphate and 150 mM NaCl pH 7.4, 20 mM imidazole, 20 mM β-ME, and 0.1 mM PMSF (Wash buffer). Then, the protein was eluted with the elution buffer (wash buffer plus 500 mM imidazole). The protein was eluted in fractions of 1 mL. The fractions with the highest concentration of the protein were combined and buffer-exchanged into phosphate-buffered saline using 10K Amicon Ultra Centrifugal filters (Sigma-Aldrich, Cat. # UFC9010). Protein concentration was measured using UV/VIS spectroscopy. The 20 μL aliquots of the protein were then flash-frozen and stored at -80 °C. The purified protein was run on an SDS-PAGE gel (Bio-Rad, Cat. # 4561096) to check for purity and any protein degradation.

Thermodynamic model fitting

The data in Figures 3A and 6E, were fit with the thermodynamic lattice model as described previously¹¹. Equation 1 was used to fit the data. The data was fit on MATLAB with using the nlinfit function as previously described^{13,14}. The detection

software, CMEAnalysis, reports the fluorescent intensity of transmembrane fusion proteins, receptors, and/or ligands within clathrin-coated structures above the local membrane fluorescence intensity. The model in eq. (1) predicts the number of transmembrane fusion proteins within clathrin-coated structures as a function of the concentration of fusion proteins on the surrounding plasma membrane. To account for the difference in the detection software output and model prediction, a correction factor of $N_{\max} * C_{\text{mem}} * A_p$ was subtracted from Eq. 1. Here, A_p is the projected area of the fusion protein on the plasma membrane, $N_{\max}$, is the max number of fusion proteins that can be accommodated within clathrin-coated structures, and C_{mem} is the fluorescence intensity of fusion proteins on the plasma membrane surrounding the clathrin-coated structure. The best fit values for $N_{\max}$ corresponded to the maximum capacity of clathrin coated structures for each of the transmembrane fusion proteins.

Fluorescence Recovery after Photobleaching (FRAP) Assays

For the FRAP studies, a square region, 5.3 μm on each side, was photobleached on the plasma membrane of RPE cells expressing either TfR-GFP-0TR or TfR-GFP-10TR. After bleaching, the fluorescence recovery was imaged on the spinning disk confocal microscope for 5 minutes at 2 second framerate. The FRAP movies were analyzed in Fiji. The FRAP Profiler plugin was used to fit single exponential curves to the fluorescence recovery plots. The recovery plots were averaged across all cells for each condition. A two-tailed t-test conducted on the fit values of the mobile fraction and half-time of recovery.

Peanut agglutinin staining of live cell for fluorescence microscopy

PNA-Alexa647 (Invitrogen L32460) lyophilized powder was resuspended in DI water at a concentration of 1mg/mL and stored at -80°C in small aliquots. PNA-Alexa647 was diluted to 1 $\mu\text{g}/\text{mL}$ in 0.5% BSA + transfection media solution with 5 mM TCEP. Cells were treated with 1 $\mu\text{g}/\text{mL}$ PNA-Alexa647 and incubated for 10 minutes before imaging on a spinning disk confocal microscope.

Mass Spectroscopy

15 μL of the his-GFP-10TR protein purified from bacteria (210 μM) was diluted in 15 μL of phosphate buffered saline. This protein sample was buffer exchanged with 0.1% formic acid using a 10 kDa Princeton separations Centri Spin column (CS101). The buffer exchanged protein solution (71.9 μM) was then analyzed by electrospray ionization (ESI) technique using the ion trap detector. The output data was deconvoluted with Thermofisher Protein Deconvolution software. The measured mass of the most abundant fragment was 49754.08 Da (Figure S5), which is roughly equal to the calculated molecular mass of the protein (49908.37 Da). Notably, no molecules with a mass significantly higher than the calculated mass of the protein were detected. This confirmed that the protein was deficient in glycosylation.

Protein identification was provided by the UT Austin Center for Biomedical Research Support Biological Mass Spectrometry Facility (RRID:SCR_021728).

Research Resource Identifiers (RRID) can be used to easily identify and access details on the equipment utilized at core facilities by searching the RRID number on www.rrids.org.

Western blot analysis

For the western blot in Figure S6, GFP and bact-GFP-10TR were expressed and purified from E-Coli, as explained in the protein purification section. Mammalian GFP-10TR was expressed in RPE cells seeded in 6-well plates at 50,000 cells per well. The cells were lysed with RIPA buffer and were pooled across 18 wells. The cell lysate was agitated for 30 minutes at 4°C. Finally, the lysate was clarified by centrifuging the sample in a table-top centrifuge at 4°C for 20 minutes. All protein samples were resolved on a pre-cast SDS-PAGE gel (Bio-Rad Mini-PROTEAN TGX polyacrylamide gel). The protein gel was transferred onto a nitrocellulose membrane with Bio-Rad's Trans-Blot Turbo Transfer System at 1.5 Amps for 10 minutes. The membrane was blocked with 4% BSA-TPBS for 1 hour at room temperature, followed by incubation with primary antibodies at a 1:5000 dilution in 2% BSA-TPBS buffer overnight at 4°C. The secondary antibodies were diluted 1:3000 in 2% BSA-TPBS and incubated with the membrane for 1 hour at room temperature. The blot was developed with Pierce ECL western blotting substrate (Thermofisher). Chemiluminescence was imaged using the ImageQuant LAS 4000 imaging system.