nature communications

Article

https://doi.org/10.1038/s41467-025-63756-w

Microbial binding module employs sophisticated clustered saccharide patches to selectively adhere to mucins

Received: 17 March 2025

Accepted: 26 August 2025

Published online: 13 October 2025

Check for updates

Thapakorn Jaroentomeechai[1,11], Billy Veloz[2,11], Cátia O. Soares[3,4], Felix Goerdeler[1], Ana Sofia Grosso[3,4], Christian Büll[1,5], Rebecca L. Miller[1], Sanae Furukawa[1], Irene Ginés-Alcober[2], Víctor Taleb[2], Pedro Merino[2], Mattia Ghirardello[6], Ismael Compañón[6], Helena Coelho[3,4], Jorge S. Dias[3,4], Renaud Vincentelli[7], Bernard Henrissat[8], Hiren Joshi[1], Henrik Clausen[1], Francisco Corzana[6], Filipa Marcelo[3,4], Ramon Hurtado-Guerrero[1,2,9] & Yoshiki Narimatsu[1,10]

The mucus lining wet body surfaces forms the interphase and barrier for the microbiota and resident microbiomes. Large mucin proteins densely decorated with O-glycans make up the mucus lining to entrap, feed and shape the microbiota, and repress biofilm formation and virulence. How mucins exert these effects is poorly understood and critical is how the microbiota recognize, sense, and break down mucins. Here, we provide structural molecular evidence that a small mucin-binding module designated X409 recognizes clustered saccharide patches comprised of rows of inner monosaccharides in adjacent O-glycans. These patches are unique to mucins and binding to these provides an elegant mechanism to retain adherence to mucins despite trimming of O-glycans during microbial scavenging of monosaccharides from mucins. Realization of clustered saccharide patch-binding motifs provides a hitherto overlooked scenario of contextual glycan epitopes and impetus for discovery of new classes of glycan-binding proteins.

Mucins are Nature's solution to protect and clear body surfaces of potentially harmful microorganisms while retaining, entrapping, feeding, and governing our symbiosis with essential commensal microbes[1–5]. Mucins constitute the major components of body fluids, of the mucus lining wet surfaces, and of the epithelia lining all body surfaces[1,6]. Mucins are large O-glycoproteins essentially comprised of extended unstructured protein regions densely decorated with O-glycans. These unstructured O-glycodomain regions can encompass thousands of amino acids arranged in tandem repeated sequence (TR) motifs that are distinct in different human mucins and display

[1]Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [2]The Institute for Biocomputation and Physics of Complex Systems (BIFI); Mariano Esquillor s/n, Campus Rio Ebro, Zaragoza, Spain. [3]UCIBIO – Applied Molecular Biosciences Unit, Department of Chemistry, NOVA School of Science and Technology, NOVA University Lisbon, Caparica, Portugal. [4]Laboratory i4HB - Institute for Health and Bioeconomy, NOVA School of Science and Technology, NOVA University Lisbon, Caparica, Portugal. [5]Department of Biomolecular Chemistry, Institute for Molecules and Materials, Radboud University, Nijmegen, the Netherlands. [6]Departamento de Química and Instituto de Investigación en Química de la Universidad de La Rioja (IQUR), Universidad de La Rioja, Logroño, Spain. [7]Architecture et Fonction des Macromolecules Biologiques, Centre National de la Recherche Scientifique and Aix-Marseille University, Marseille, France. [8]Department of Biotechnology and Biomedicine (DTU Bioengineering), Technical University of Denmark; Søltofts Plads, Lyngby, Denmark. [9]Fundación ARAID, Zaragoza, Spain. [10]GlycoDisplay ApS, Copenhagen, Denmark. [11]These authors contributed equally: Thapakorn Jaroentomeechai, Billy Veloz. e-mail: rhurtado@bifi.es; yoshiki@sund.ku.dk

O-glycans in unique clustered patterns[7]. The mucin O-glycodomains govern most binding interactions with bacteria, and a remarkable lack of conservation in TR sequences among orthologous mucins suggests host-microbial adaptation[8]. However, a fundamental question is how mucin glycans are selectively recognized.

Studies of interactions with mucins have largely focused on binding to the glycans[9,10], and studies employing mucins have only been achievable with heterogeneous and poorly defined isolated mucins. To overcome these challenges, we developed cell-based mucin arrays that enable display and production of mucins with defined O-glycans[7,11]. Studies with these are beginning to highlight that the contextual presentation of glycans in clusters and patterns on proteins drives specificity of interactions, with illustrative examples of select binding to mucins found with microbial adhesins[11,12], microbial mucinases[13], and human innate immune receptors[14]. Recently, we showed that *Akkermansia muciniphila*, an intestinal symbiont and well-known mucin-binder, binds an O-glycan motif found on select human mucins[15]. However, the molecular bases for recognition of mucins and the nature of the binding epitopes are essentially unknown. The microbiota has evolved an arsenal of proteins (adhesins, lectins, carbohydrate-binding modules (CBMs)) to recognize glycans and mucins, as well as enzymes to degrade glycans (glycoside hydrolases) and mucins (O-glycoproteases/mucinases). These proteins and enzymes allow the microbiota to scavenge glycans, penetrate and break down the mucus barrier[1,16]. CBMs are frequently appended to microbial carbohydrate-active enzymes (CAZymes) as well as mucinases for efficient substrate targeting[17,18]. We previously identified the unique mucin-binding module (MBM), designated X409, appended to the enterohemorrhagic *E. coli* (EHEC) StcE mucinase and demonstrated that the X409 module is dispensable for the StcE mucinase activity but alone mediates the mucin-binding properties of the StcE mucinase[7,19,20]. This X409 module (~100 amino acids) is not a CBM that binds simple glycans, but exhibits remarkably selectivity and high affinity binding to human mucins with dense O-glycans without being dependent on particular O-glycan structures[7]. The X409 module, therefore, offers a model for studying how specific mucin-binding properties are achieved. X409 is a mobile element found on other microbial proteins. For example, the *Vibrio cholerae* biofilm matrix adhesion protein RbmC contains two X409 modules that may support colonization in the mucus and virulence[21].

Here, we show that the unique mucin-binding properties of X409 are achieved through recognition of a row of inner monosaccharides in a distinct cluster of adjacent O-glycans. This row of inner monosaccharides is comprised mainly of the inner obligate GalNAc residues attached to Ser and Thr in the STTT motif and forms what has previously been defined as a clustered saccharide patch[22], largely independent of the terminal parts of the O-glycans attached. This binding mode explains how X409 acquires binding to specific human mucins and binding largely independent of the terminal structures of O-glycans[7], and it provides an elegant way to overcome the transient nature of O-glycans on mucins due to degradation by microbial enzymes. The existence of clustered saccharide patches formed of multiple discontinuous glycans was originally envisioned as a mechanism to obtain sufficient specificity in biological interactions with common glycans[22]. Our structural molecular data now authenticate this concept and show that the presentation of such glycan clusters can drive highly selective and advantageous binding to mucins.

Results

X409 binds O-glycan clusters

As a starting point for obtaining molecular insights into how mucins are selectively recognized, we took advantage of the small X409 module that selectively binds mucins MUC2 and MUC5AC, but not, e.g., MUC1, which has a lower density of O-glycans[7]. We employed the cell-based mucin platform[7] to produce secreted fusion-glycoprotein reporters containing representative mucin TR sequences (150–200 amino acids) (Supplementary Fig. 1) decorated with custom-designed Tn or a mixture of sialylated core1/2 O-glycans (labeled WT as this is produced by HEK293 WT cells) (Supplementary Fig. 2a). We used single molecule mass photometry to analyze the binding properties of X409 (MBP-X409) to these glycoforms of different mucin reporters using 10-fold excess of X409 (mol/mol) (Fig. 1a). This confirmed that X409 does not bind MUC1 and showed that X409 bound a single motif on MUC2-2 and multiple motifs on MUC2-1 (Note, MUC2 has two distinct TR regions with different O-glycan patterns covered independently in the MUC2-1 and MUC2-2 reporters) and MUC5AC reporters. Importantly, the observed bound complexes between MBP-X409 and mucin reporters were similar in terms of number of bound X409 ($n = 1–7$) molecules per reporter as well as peak intensities at this ratio of molecules. To further evaluate binding, we used ELISA with titration of the mucin reporters (Supplementary Fig. 2b, c), which confirmed differential binding to MUC2 and MUC5AC and further demonstrated that X409 bound better to reporters with elaborate O-glycan structures compared to Tn O-glycans that only consist of the initial αGalNAc monosaccharide. Analysis of the mucin reporter sequences bound by X409 suggested that recognition involved a common motif of three to five adjacent O-glycans and requires the presence of both threonine (T) and serine (S) residues. Since X409 bound multiple times to MUC5AC with the highly conserved repeat TTSTTS O-glycan cluster, we tested ideal MUC5AC reporters (with identical TR sequences) designed with split motifs (TTS/TTS), but these did not bind (Fig. 1b and Supplementary Fig. 4). This then suggested that the X409-binding motif indeed comprised an O-glycan cluster of four or more S/T residues. We next employed ideal repeat TR reporters with TTTT or STTT O-glycan clusters, which remarkably revealed that X409 only bound the STTT motif (Fig. 1c and Supplementary Fig. 5). Expanding these ideal reporters to combinations of S/T residues in 3–6 adjacent O-glycans, indicated that specific orders of 3-4 O-glycans on S and T residues are needed for efficient binding (Supplementary Fig. 6). Thus, X409-binding requires a clustered O-glycan motif of 3-4 residues with at least one S residue and the inner αGalNAc residues.

X409 recognizes a clustered saccharide patch

We generated co-crystal structures of X409 in complex with three different glycopeptides at resolutions ranging from 1.2 to 1.75 Å in primitive orthorhombic and monoclinic space groups (Fig. 2 and Supplementary Table 1). We first analyzed the complex with the Tn-MUC5AC glycopeptide (AEAPT*T*S*T*T*SAPK, * denotes GalNAc residues, note that Ser10 was not glycosylated, see Supplementary Fig. 1b for all short glycopeptides used in the study), which was produced in glycoengineered HEK293 cells employing our recently developed Glycocarrier technology[23] (Supplementary Fig. 7). This structure revealed interactions with two GalNAc moieties (bold) in the T*T***S***T***T***S sequence as well as interactions with the four last Ser/Thr amino acids (STTS backbone and side chains) (Fig. 2c and Supplementary Table 2), suggesting that the X409-binding motif is determined by the **STTT/S** O-glycan cluster (hereafter numbered $\mathbf{S_1T_2T_3T_4}$, residue/GalNAc). We therefore synthesized an ideal 4xTn-STTT (AAS$_1$*T$_2$*T$_3$*T$_4$*PAPA) glycopeptide, and co-crystallization revealed interactions with three of the four GalNAc moieties at positions 1, 3 and 4, but not position 2 of the $\mathbf{S_1}$*T$_2$*$\mathbf{T_3}$*$\mathbf{T_4}$* sequon (Fig. 2c and Supplementary Table 2). These structures revealed that the X409-binding site employs solvent-exposed residues including aromatic residues Y828, W854, and Y859 for interactions. Most of the X409 amino acids involved in binding are located within two loop regions (between strands β2-β3 and β4-β5) with only one residue (R894) positioned in strand β6 (Figs. 2b and 3a–e). X409 achieves binding mainly through interactions with the GalNAc moieties (GalNAc$_{1,3,4}$) as well as the obligate S/T residues to which they are attached. The first GalNAc$_1$ on S ($\mathbf{S_1}$TTT) forms hydrogen bonds with R825/N856 (side chains) and R894

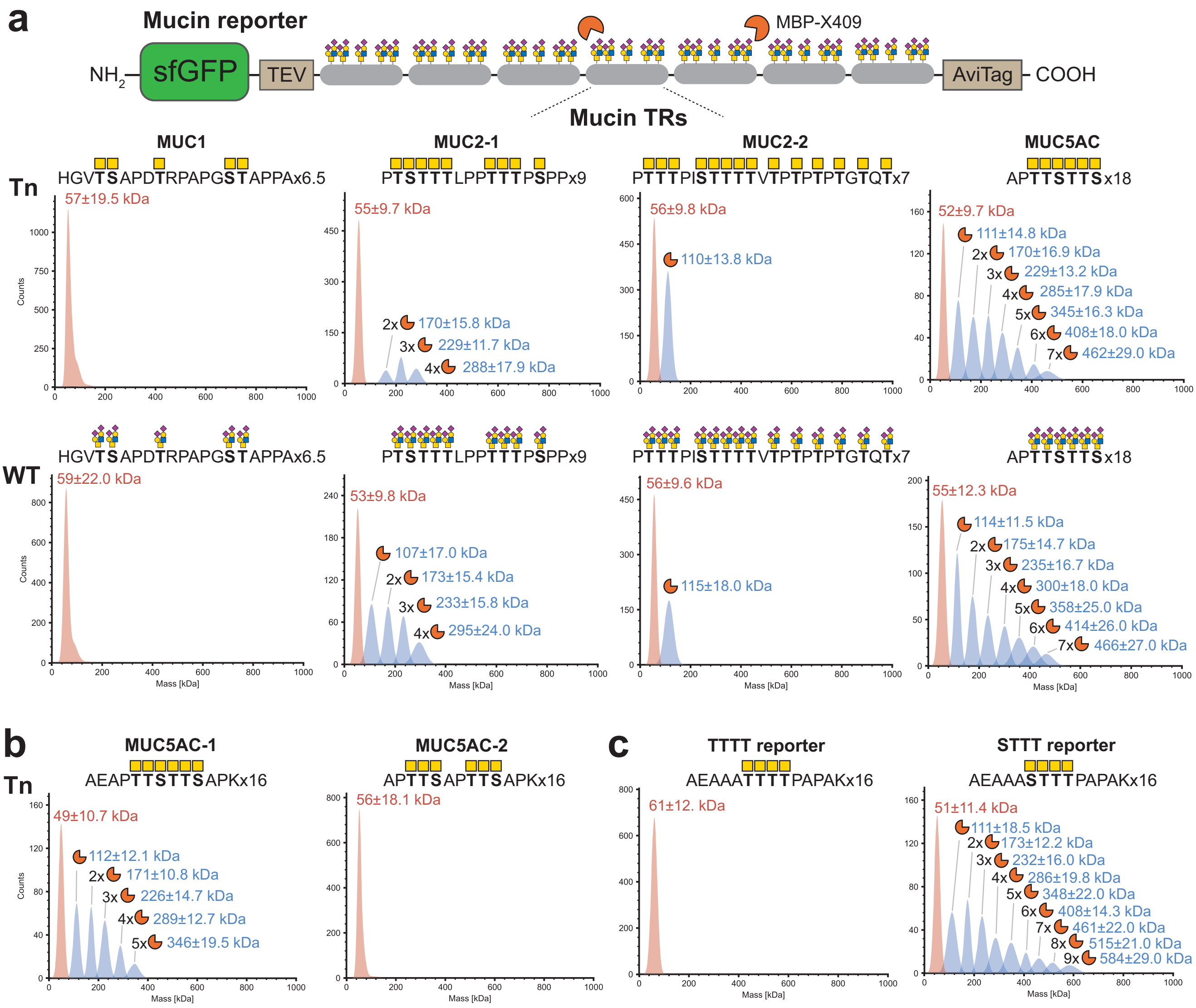


Fig. 1 | Single-molecule mass photometry analysis of the mucin-binding properties of X409. a Mass photometry histogram for MBP-X409 binding to isolated mucin reporters with Tn (upper panels) or sialylated core1/2 (WT) O-glycans (lower panels). The top is a schematic representation of the mucin reporter design with N-terminal GFP and the mucin tandem repeat (TR) region comprised of ~200 amino acids derived from human mucins as indicated. A representative amino acid sequence for TRs and the number of TRs included in the mucin reporters tested (MUC1, two different regions of MUC2, and MUC5AC) are shown in the panels above. All potential O-glycosites (Ser/Thr residues) are in bold with a representative O-glycan structure above. The histograms are color-coded to distinguish unbound MBP-X409 and mucin reporters (orange) and bound X409-reporter complexes (blue), with the estimated number of X409 molecules bound per reporter indicated. **b** Mass photometry histogram for MBP-X409 binding to an ideal MUC5AC reporter with identical TR repeat sequences without (TTSTTS) or with split motifs (TTS/TTS) and bearing Tn O-glycans. **c** Mass photometry histogram for X409 binding to ideal mucin-like reporters with identical repeated TTTT or STTT motifs in TRs with Tn O-glycans. Data are representations of at least three replicates. See Supplementary Fig. 1 for full sequences of all mucin and ideal mucin-like reporters used, and Supplementary Fig. 2 for the general design of reporters. Control data for mass photometry analysis are provided in Supplementary Fig. 3.

(side chains) through the endocyclic oxygen and OH3/acetamide carbonyl group respectively. The third GalNAc$_3$ on T (ST**T$_3$**T) forms hydrogen bonds between the acetamide NH group and Y859 (sidechain) and interacts with D855 (backbone) and Y859/W854 (sidechain) through hydrogen bonds and CH-π interactions. Finally, the fourth GalNAc$_4$ moiety (STT**T$_4$**) interacts with Y828 (sidechain) via a CH-π interaction. The underlying S/T residues contribute to interactions. S$_1$ (backbone) forms hydrogen bonds with Y859 (backbone). T$_2$ (sidechain) engages in CH-π interactions with Y859 (sidechain), while its backbone forms a hydrogen bond with N856 (sidechain). T$_3$ (backbone) forms hydrogen bonds with Y859 (sidechain), and T$_4$ forms hydrogen bonds with Y828, along with an additional CH-π interaction with Y828 (sidechain). Note that the structure with Tn-MUC5AC revealed interactions with the last S residue (STT**S$_4$**), even though this residue did not carry a GalNAc moiety. This also showed that the fourth position may interchange between S/T residues. We note that of the critical amino acid residues involved in interactions with GalNAc$_1$ and GalNAc$_3$ in the S$_1$T$_2$T$_3$T$_4$ motif, several residues only interact with the GalNAc moieties (R825/R894 with GalNAc$_1$ and W854/D855 with GalNAc$_3$), while other residues (N856 with GalNAc$_1$ and Y859 with GalNAc$_3$) also interact with the underlying S/T residues (side chains/backbone). Thus, N856 interacts with GalNAc$_1$ and T$_2$, while Y859 interacts with GalNAc$_3$ and the three S$_1$, T$_2$, T$_3$ residues in the **STT**T sequon. GalNAc$_4$ only interacts with Y828, and Y828 also interacts with T$_4$ (or S$_4$ in STT**S**), but in this case, independent of whether this position carries an O-glycan.

Since our binding data clearly indicate improved binding to mucins with elongated O-glycans (Supplementary Fig. 2b, c)[7], we

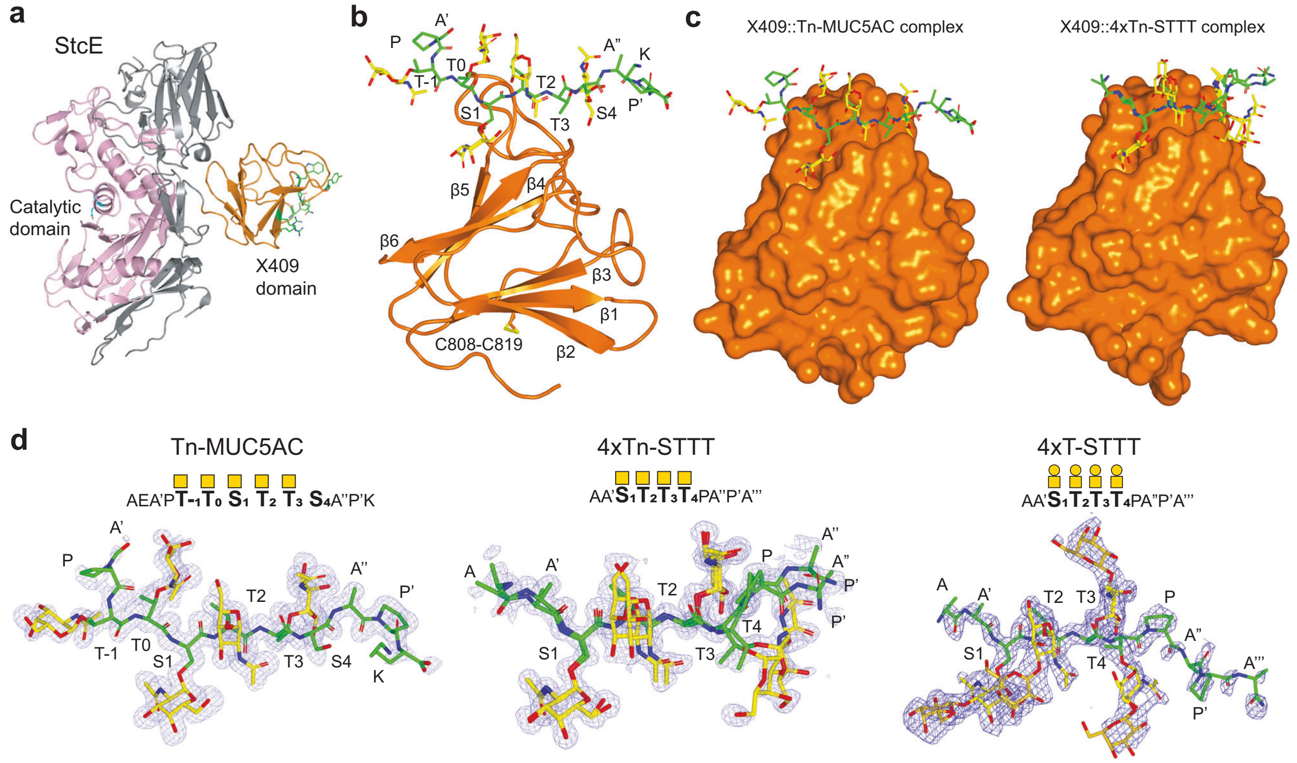
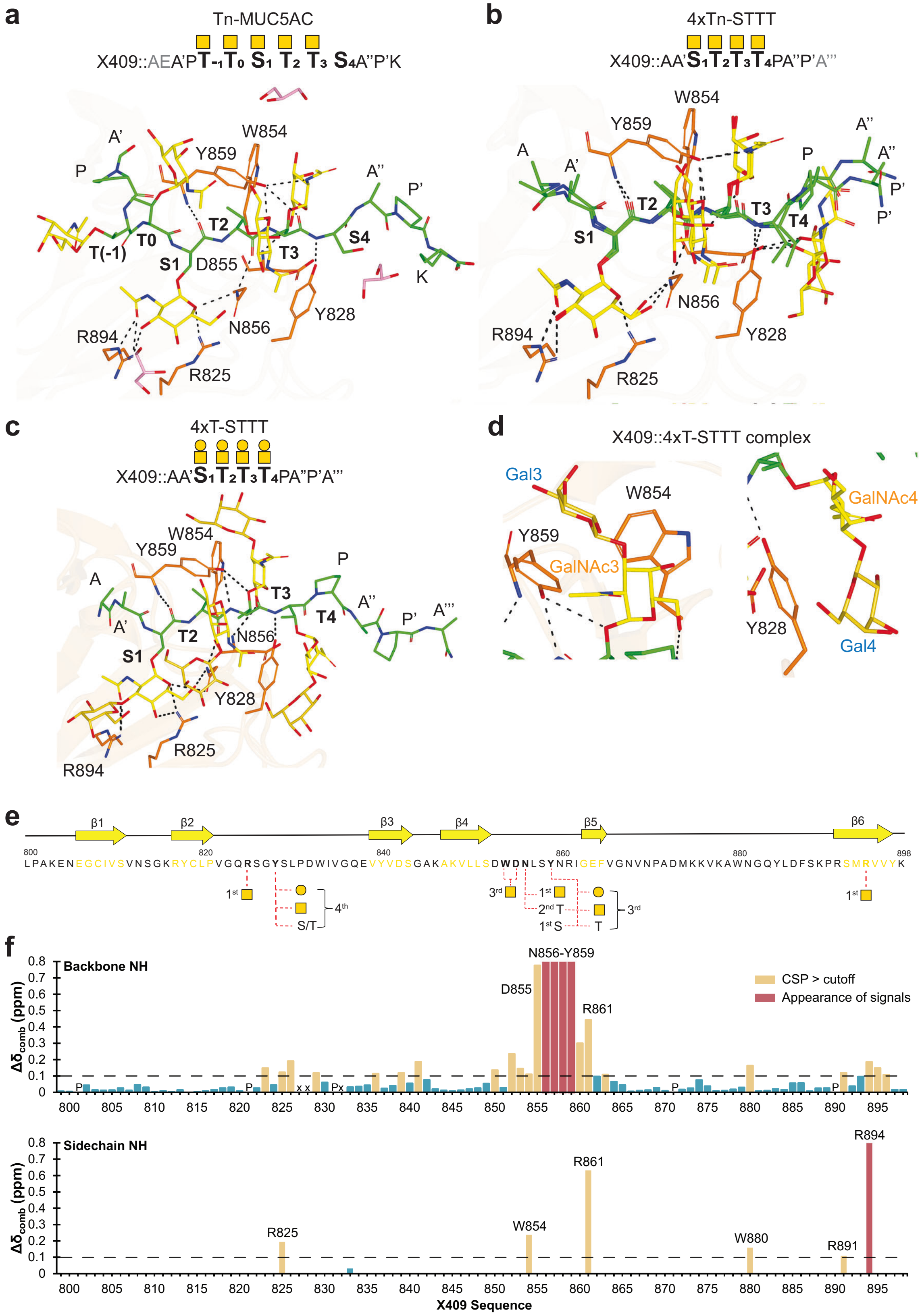


Fig. 2 | Crystal structure of X409 complexed to glycopeptides. a Ribbon structure of enterohemorrhagic *E. coli* StcE mucinase (PDB: 3UJZ) with the catalytic domain and the mucin-binding domain (X409) shown in pink and orange, respectively[19]. The catalytic residue Glu447 is highlighted in cyan, while selected residues of the X409-binding site are shown in green, illustrating that the catalytic site and X409-binding site are spatially separated. **b** Ribbon structure of the X409 complexed to the Tn-MUC5AC glycopeptide (PDB: 9GRJ) (See Supplementary Fig. 1b for glycopeptide nomenclature). Amino acids of the glycopeptide substrate are depicted with green carbon atoms, and the GalNAc residues as yellow carbon atoms. GalNAc residues are labeled according to the binding motif of the X409, with the first interacting serine labeled as S1. Upstream and downstream glycosylation sites are labeled as −1 and +1 positions, respectively, counting from S1. **c** Surface presentation of the binding site of X409 showing binding cavity that interacts with Tn-MUC5AC or 4xTn-STTT glycopeptides (PDB: 9GRJ and 9GRF). **d** Electron density maps of the Tn-MUC5AC, 4xTn-STTT, and 4xT-STTT at 2.2σ are Fo−Fc contoured (blue). The Gal residues of the core1 O-glycans in 4xT-STTT are indicated as dark yellow carbon atoms. In the 4xTn-STTT glycopeptide, the entire glycopeptide adopts dual conformations, with the GalNAc residue linked to Thr4 being the most prominent example, highlighting the local flexibility of this moiety.

analyzed the co-crystal of X409 with a 4xT-STTT glycopeptide (T O-glycan, Galβ1-3GalNAcα1-O-Ser/Thr) (Fig. 3c, d). Most of the interactions found with the Tn glycopeptides (4xTn-STTT and Tn-MUC5AC) were preserved, but additional interactions with the Gal moieties at T_3 and T_4 ($S^{**}T^{**}T_3^{**}T_4^{**}$, ** denotes Galβ1-3GalNAc) formed CH-π interactions with Y859 and Y828, respectively (Fig. 3c, d). These additional interactions likely account for the enhanced binding to mucins with elaborate O-glycans, which we confirmed by SPR isothermal measurements of X409-binding affinity with the 4xTn- and 4xT-STTT glycopeptides showing improved K_d with the T glycoform (Supplementary Fig. 8). Notably, the enhanced affinity of the T (core1) glycopeptide arises from both a faster association rate (k_{on}) and a slower dissociation rate (k_{off}). Specifically, the k_{on} increases by ~4.8-fold and the k_{off} decreases by ~7.5-fold relative to the Tn glycoform, resulting in an overall ~36-fold improvement in binding affinity. Although the Gal residue does not form hydrogen bonds, it engages in CH−π interactions that contribute to improved molecular recognition during association and greater stability of the final complex. These findings suggest that the affinity gain involves both entropic and enthalpic contributions, reflecting cooperative kinetic effects.

Next, we analyzed X409 binding in solution by NMR spectroscopy. X409 was expressed with uniform ^{13}C,^{15}N-labeling for NMR assignment of the backbone resonances following a standard triple resonance approach (BMRB ID 53125). ^{1}H,^{15}N-HSQC-based titrations of ^{15}N-X409 were performed with the 4xTn-STTT and Tn-MUC5AC glycopeptides. Significant chemical shift perturbations (CSPs) of several X409 amide resonances were observed upon addition of 4xTn-STTT indicating binding (Supplementary Fig. 9a). We found a clear decrease in intensity of signals of the free-state of X409 concomitant with appearance of new signals of the bound state, and with excess of 4xTn-STTT (1:1.5 molar ratio) the fully bound form was reached (Supplementary Fig. 9b). The observed slow exchange regime in the chemical shift NMR timescale in the ^{1}H,^{15}N-HSQC titrations is suggestive of strong affinity interactions. Several CSPs of X409 residues in presence of 4xTn-STTT (Fig. 3f) were identified. Specifically, the backbone and side chains of R894 and R825, both involved in engagement of $GalNAc_1$, exhibited perturbations (0.19 ppm CSP for backbone and 0.12 ppm for the lateral sidechain of R825 and appearance of sidechain cross-peak for R894). The D855 and the sidechain of W854, previously found to interact with $GalNAc_3$, suffered significant CSPs (0.78 and 0.24 ppm, respectively). Interestingly, the resonances from residues N856 to Y859 (in the loop between β4-β5 strands) only became observable in presence of the ligand (Fig. 3f), which suggests that the dynamics of this loop is altered, with decreased flexibility and stabilization of a conformation upon ligand binding. This alteration on the loop dynamics also induced CSPs on residues not directly interacting with the ligand, e.g. N860 and R861. We could not assign the Y828 residue, which directly interacts with $GalNAc_4$. We also performed a ^{1}H,^{15}N-HSQC based titration of ^{15}N-X409 with Tn-MUC5AC, which produced similar CSPs and dynamics' alterations on X409 structure as found with 4xTn-STTT (Supplementary Fig. 10). Thus, the binding site of X409 employs three residues (R825, N856, R894) to interact with



the first GalNAc$_1$, two residues (W854, D855) for GalNAc$_3$, and two tyrosine residues, Y859 that interacts with S$_1$T$_2$T$_3$ and GalNAc$_3$/Gal$_3$, and Y828 that interacts with S$_4$/T$_4$ and GalNAc$_4$/Gal$_4$ (Fig. 3e).

We then used the coordinates of the co-crystal structures to construct putative 3D models of X409 complexes with MUC5AC glycopeptides. These models incorporated full O-glycan occupancy of all S/T positions (AEAPT*T*S*T*T*S*APK) in all analyses, and we used 1 μs molecular dynamics (MD) simulations to obtain fully equilibrated structures in an aqueous environment (Supplementary Figs. 11 and 12). We selected truncated O-glycans (Tn, T, and the sialylated variants) for MD simulation studies because our previous study demonstrated that X409 binding to mucins was increased by core1 O-glycans while

Fig. 3 | Structural features of the X409 mucin-binding site. a–c Views of the active site for the X409-glycopeptide complexes as indicated. Residues forming the binding site are depicted in orange. Colors for sugar moieties are as in Fig. 2. The dotted lines of the hydrogen-bond interactions are depicted in black. Glycerol moieties are indicated as pink carbon atoms in (**a**). **d** Close-up view of the complex between the 4xT-STTT glycopeptide and X409 showing the essential residues (Y828 and Y859, and to an extent Y859) interacting with the glycopeptide. **e** Primary sequence of X409 overlaid with its secondary structure and depicting residues that interact with 4xT-STTT and Tn-MUC5AC glycopeptides according to X-ray crystal structure in (**a**–**d**). **f** Histogram of the combined ^{1}H,^{15}N chemical shift ($\Delta\delta_{comb}$) of the NH resonances of X409 residues' backbone and some sidechain residues (Arg and Trp) upon interacting with 4xTn-STTT glycopeptide. A cutoff line at 0.1 ppm, displayed as a black dashed line, is considered to distinguish the residues that experienced major differences in chemical shift perturbation (CSP). Strongly perturbed residues (CSP > 0.1 ppm) are in yellow bars and weakly perturbed in blue bars. In the histogram, it is also displayed as red bars the residues N856–Y859 whose resonances appeared in the ^{1}H,^{15}N-HSQC spectra upon 4xTn-STTT glycopeptide addition. Letter P denotes proline residues (lacking NH group in the backbone), and letter x indicates unassigned/missing residues. Source data are provided as a Source Data file.

further O-glycan elaborations did not appear to affect binding[7]. The MD simulations revealed that the hydrogen bonds and CH-π interactions between the glycopeptides and X409 were preserved throughout the MD simulations (Supplementary Figs. 13 and 14). Our simulations predict that the Gibbs binding energy (ΔG) of the X409 complex with Tn-MUC5AC is −48.3 ± 7.7 kcal/mol, while the binding energy significantly increases when the O-glycans are elongated with Gal residues to −58.4 ± 5.8 kcal/mol, in accordance with X409's preferential binding to mucins with elaborate O-glycans[7], and the obtained K_d SPR values (Supplementary Fig. 8). Our simulations of X409 bound to the Tn-AEAPTTSTTSAPK glycopeptide confirmed substantial contributions from Y859, Y828, and W854 that are engaged in CH-π interactions with $GalNAc_3$/$GalNAc_4$ with binding energies of −7.6 ± 1.1, −4.8 ± 1.1, and −4.4 ± 2.5 kcal/mol, respectively. In complex with the T O-glycoform of this glycopeptide, these values for Y859, Y828, and W854 changed to −10.0 ± 1.2, −6.9 ± 1.3, and −3.4 ± 0.8 kcal/mol, respectively. In this latter complex, other residues such as N856 and L857 also play a substantial role in binding with energies around −4.0 and −1.8 kcal/mol, respectively. The Gal residues at the last two O-glycans ($Gal_{3/4}$) contribute ca.−1.0 ± 0.9 kcal/mol each, and it is notable that in both complexes, the peptide backbone of the first three residues in the STTS motif contributes with binding energies ranging from −4.4 ± 0.7 to −6.2 ± 0.7 kcal/mol. Moreover, our calculations indicate that hydrogen bonds and CH-π interactions occur more frequently (with higher occupancy throughout the simulation trajectory) in complexes containing glycopeptides with further elongated *O*-glycans, e.g., sialylated T *O*-glycans (mSTa and dST) (Supplementary Fig. 11), suggesting increased structural stability in these complexes. Consistently, the root-mean square deviation (RMSD) values for these complexes, particularly in GalNAc moieties and peptide backbone, showed reduced fluctuations, further supporting this observation (Supplementary Fig. 11b). The enhanced stability for glycopeptides with elongated O-glycans can be partly attributed to transient additional stabilising interactions, such as hydrogen bonds and salt bridges involving sialic-acid residues. Based on MD simulations, we identified R891 as a critical residue that reinforces the overall stability of the complex not identified in our X-ray studies.

A key finding was that X409 specifically binds O-glycopeptides with the STTT cluster motif, but not the related TTTT (Fig. 1d). We explored this further by ^{1}H,^{15}N-HSQC adding a 4xTn-TTTT glycopeptide (sequence otherwise as 4xTn-STTT) at the molar ratio that yielded fully bound 4xTn-STTT (1:1.5, i.e. 100 µM X409/150 µM glycopeptide), and this produced only decrease in the intensities of a subset of ^{15}N-X409 resonances (Supplementary Fig. 15). Residues with decreased intensity were those involved in interactions with $GalNAc_3$ (D855, W854), while residues involved in interactions with $GalNAc_1$ (R825, N856, R894) were unaffected, which agrees with the MD simulations results (compare Tn-AEAPTTSTTSAPK vs Tn-AEAPTTTTTSAPK in Supplementary Figs. 13 and 14). Moreover, resonances from N856–Y859 residues did not appear upon ligand addition, which indicates that the loop between β4-β5 strands is not stabilized by 4xTn-TTTT. Analysis of the naked MUC5AC peptide (corresponding to Tn-MUC5AC) showed no significant CSPs, confirming critical binding to the glycans (Supplementary Fig. 16). Next, we deduced the overall conformations of the Tn-STTT and Tn-TTTT glycopeptides in solution by combining NMR and MD simulation studies. It was previously reported that the linkage of GalNAc to Ser (GalNAcα1-O-Ser) and Thr (GalNAcα1-O-Thr) adopts different conformations[24]. The linkage to Thr forced by its methyl group typically adopts the eclipsed conformation (with $\psi \approx 120°$) with the GalNAc residue almost perpendicular to the peptide backbone, while the linkage to Ser typically prefers a staggered conformation (with $\psi = 180°$) with the GalNAc residue aligned parallel to the peptide. Here, we confirmed that all GalNAc-Thr linkages in the 4xTn-STTT and 4xTn-TTTT glycopeptides adopt the characteristic eclipsed conformation in solution in the presence of the key NOE cross-peak between the NH of GalNAc and NH of Thr (Supplementary Fig. 17). The absence of NOE contact for the Ser with the 4xTn-STTT glycopeptide suggests that the glycosidic linkage for S_1 is much more flexible. MD simulations of the glycopeptides 4xTn-STTT and 4xTn-TTTT show a very good agreement between the experimental and theoretically derived distances for the peptide and GalNAc presentation (Supplementary Figs. 17 and 18). Thus, the key feature enabling X409 binding to 4xTn-STTT over 4xTn-TTTT glycopeptides is the orientation and flexibility of $GalNAc_1$ on S_1 (ψ ≈ 180° ± 10°) (Supplementary Fig. 18). MD simulations of free-state 4xTn-STTT show that GalNAc moieties adopt spatial distributions similar to those in the X-ray structure. The RMSD of GalNAc units, relative to the X-ray conformer, remained below 3 Å (heavy atoms) for 20% and under 4 Å for ~50% of the trajectory. In contrast, these conformations were absent in 4xTn-TTTT, suggesting that 4xTn-STTT preferentially adopts a crystal-compatible conformation, while 4xTn-TTTT rarely assumes the bound state. The higher entropic penalty for 4xTn-TTTT binding to X409, combined with the non-optimised interactions between 4xTn-TTTT and X409, may explain the lower affinity of X409 to the all-Thr-containing glycopeptide.

Finally, we experimentally validated most of the critical binding residues in X409 by single residue mutation and analysis of binding to isolated Tn- and T-ideal-MUC5AC reporters using ELISA (Supplementary Fig. 19). Interestingly, we found that several mutations differentially affected binding to Tn and T glycoforms. Thus, while Y859A and R825A mutations completely abrogated binding, Y828A, W854A, R894A, and R861A mutations selectively abrogated binding to Tn and not/less to T glycoforms. This is likely because binding to Tn glycoforms in general is weaker, and interactions with the Gal_3 (Y859A) and Gal_4 (Y828A) residues in elongated O-glycans can compensate for affinity loss from mutation. Indeed, our MD simulations showed that the Gal residues in elongated core1 O-glycans stabilize the structure by forming a hydrogen bond between Gal-OH6 and S829-OH next to Y828, which helps maintain the conformation of the GalNAc. We also performed MD simulations of the Tn-MUC5AC glycopeptide and a Y828A mutant of X409, which revealed a drop in binding energy to −36.2 ± 6.0 kcal/mol, and this drop was increased in simulations with the T glycoform to −46.5 ± 6.2 kcal/mol.

Since the X409 module is found to be appended to other microbial proteins, we also performed a sequence similarity search and selected four X409 homologs for recombinant expression and determination of mucin-binding properties by use of the cell-based mucin display (Supplementary Fig. 20). Only one natural X409 sequence variant with a Y828P polymorphism exhibited altered binding, and interestingly, similar to our Y828A mutant, this selectively resulted in complete loss of binding to Tn but not T glycoforms.

Discussion

Our X-Ray, NMR, and MD studies collectively demonstrate that a discontinuous clustered saccharide patch across the first, third, and fourth O-glycan in an STTT sequon is the critical X409 mucin-binding motif (Fig. 4a–d). The first GalNAc must be attached to a Ser residue, as only this provides flexibility and the conformational fit of the

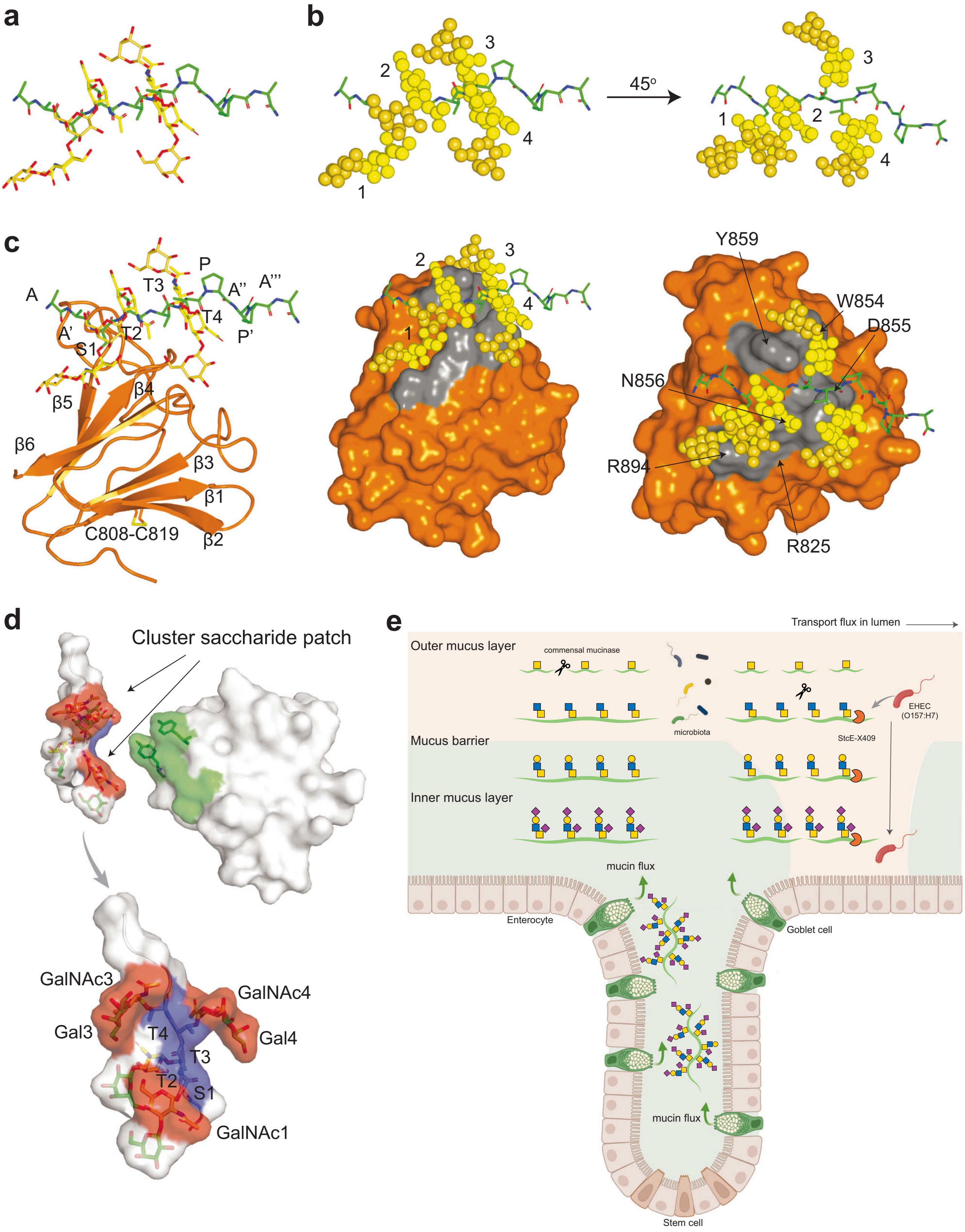


Fig. 4 | The clustered saccharide patch recognized by X409 and its potential biological roles in the intestinal mucus. a View of the 4xT-STTT glycopeptide. **b** The 4xT-STTT glycopeptide with glycan moieties depicted as light (GalNAc) and dark yellow (Gal) spheres. **c** Ribbon structure (left) or surface presentation (right) of X409 complexed to the 4xT-STTT glycopeptide (PDB: 9GRJ). Amino acids of the glycopeptide substrate are depicted with green carbon atoms, the GalNAc and Gal residues as light- and dark yellow carbon atoms, respectively. **d** Views of the clustered saccharide patch formed in the 4xT-STTT glycopeptide and the corresponding X409-binding site. **e** Schematic illustration of the proposed function of X409 in the intestinal mucus and its role for the StcE mucinase associated with EHEC. In the healthy gut (left), the commensal microbiota forage on mucin O-glycans, resulting in mucins with trimmed O-glycans (Tn/T) in the outer mucus layers. Commensal mucinases can only cleave mucins once their O-glycans are trimmed. In the EHEC infectious state (right), the potent secreted StcE mucinase with appended X409 can cleave mucins with nascent O-glycans to break the mucin barrier and facilitate penetration to the underlying epithelium for delivery of toxins. X409 provides StcE with binding to the inner clustered saccharide patch in the MUC2 mucin and hence binding throughout the mucus layers despite the trimming of O-glycans. The higher affinity of X409 towards mucins with elaborate O-glycans is predicted to "drive" StcE towards the origin of mucin synthesis (goblet cells) and the underlying epithelium, facilitating the destruction of the mucus barrier. Part of this figure is created in BioRender. Jaroentomeechai, T. (2025) https://BioRender.com/w20w589.

saccharide patch for X409 binding. The third and fourth O-glycans are required for binding, and it is the elongation of these O-glycans that mediates increased binding affinities towards nascent mucins compared to mucins with partially trimmed O-glycans. The unique mucin-binding mode of X409, to our knowledge, provides the first structural validation of the original concept of clustered saccharide patches[22] and demonstrates how a binding epitope is constructed of multiple O-glycans in a select protein sequence context. The X409-binding mode highlights that clustered saccharide patches can involve inner monosaccharides independent of the structures of the glycans, and that distinct patterns of O-glycans are recognised and govern selectivity in binding to mucins. Authentication of clustered saccharide patch epitopes provides for a potential major expansion of the glycan epitome with ramifications for analysis and discovery of the glycan interactome, as clustered saccharide epitopes escape detection by traditional experimental approaches based on individual glycans without protein context.

The STTT sequence motif is only found in select human mucins (mainly MUC2, MUC5AC, MUC2-1), and among the mucins found in the intestine (MUC1, MUC2, MUC3, MUC13, MUC17)[1] only MUC2 contains TRs with STTT motifs in ~20% while all MUC2 TRs contain TTTT motifs[25]. X409 is well-suited for the transient nature of the MUC2 mucins that are constantly scavenged for O-glycans, degraded, and replenished to maintain the integrity of the mucus-microbe interface. Binding to terminal O-glycan epitopes can only support transitory interactions, as these are progressively trimmed upon digestion by microbes that utilize monosaccharides as nutrients[3]. Instead, the X409-binding mode supports interactions with mucins, whether nascent (intact) or progressively eroded to the innermost GalNAc residues before degraded by mucinases that selectively digest mucins with truncated O-glycans (Tn/T)[13,26,27]. Interestingly, these types of mucinases contain CBMs that bind Tn/T O-glycans, and they are mainly produced by commensals[13,26,27]. An elegant aspect is that X409 achieves higher affinity to nascent mucins by interactions with the Gal and Neu5AC residues of extended O-glycans, which, despite the continuous outward flow of mucins in the mucus, provides for a sophisticated mechanism to "drive" X409 towards the underlying epithelium and source of nascent mucins (Fig. 4e). X409 is found appended to the StcE mucinase that is secreted by the highly pathogenic EHEC, causing hemorrhagic colitis by delivery of Shiga-like toxins to the mucosal epithelium[28]. StcE is a potent mucinase that can degrade nascent mucins with nascent O-glycans, and X409 thus provides this enzyme with critical mucin targeting properties in the mucus[28] (Fig. 4e). Interestingly, StcE does not cleave mucins with the core3-based O-glycans found as the major O-glycoform of MUC2 in the healthy human intestine[7]. Nonetheless, our previous binding studies[7] and the present MD simulations (Supplementary Fig. 12) clearly demonstrate that X409 binds well to the core3 O-glycosylated STTT motif, highlighting that X409 independently can target StcE to mucins with a wide range of O-glycan structures. Note, though, that due to technical limitations, binding to larger branched, e.g. core4, O-glycans was not tested. The catalytic unit of StcE itself does not have mucin-binding properties as originally suggested[20], and X409 does not serve directly in the StcE enzyme activity[7]. StcE digests mucins in the T*XT* sequence motif[20] with some sequence limitations[29], and since this is part of X409's binding motif (S*T*T*T*), X409 will target StcE to its substrates and mucin cleavage in this motif will release X409.

The clustered saccharide patch concept was originally proposed to resolve the apparent paradox that the repertoire of simple glycan structures in the human glycome appears too limited to produce a ligand repertoire with sufficient diversity to govern the many biological functions attributed to glycan-binding proteins[22]. While the structural space of human glycans is quite large[30], the main diversity in human oligosaccharides is generated through repeated use of common sequence motifs and scaffolds, and hence the repertoire of distinct glycan epitopes (often terminal) is much more limited and furthermore common to many cells[31]. Studies have provided support for the clustered saccharide patch concept. For example, the malaria parasite *P. falciparum* employs the erythrocyte binding antigen 175 (EBA−175) to mediate adherence and invasion of erythrocytes by binding to multiple sialylated O-glycans selectively on human glycophorin A[32]. Similarly, several Streptococci employ serine-rich repeat (SRR) adhesins to bind multiple O-glycans on select salivary mucins and the major platelet mucin-like O-glycoprotein GP1bα[11,12]. Evidence also indicates that sialic-acid binding Siglec immune receptors bind select O-glycoproteins that display clusters of O-glycans[14,33]. However, the structural bases for these interactions are still unknown. The cell-based glycan array has provided support for the contextual recognition of O-glycans displayed on mucins and mucin-like O-glycodomains as exemplified above, and more recently, also an example of contextual recognition of N-glycans by the Multifunctional Autoprocessing Repeats-in-Toxin (MARTX) toxins from *Vibrio* species was found[34]. MARTX binds to select N-glycoproteins, e.g. L1CAM, through recognition of a putative motif of inner monosaccharides (i.e., dependent only on GlcNAc residues and not terminal structures) in multiple biantennary N-glycans. Thus, clustered saccharide patch motifs are conceivably found on all types of glycoconjugates.

In conclusion, our studies of the small X409 module provide the first structural insight into the molecular nature of a clustered saccharide patch. Recognition and binding to clustered saccharide patches provide unique opportunities for specifically targeting select glycoproteins and a potentially enormous expansion of the glycan epitome with contextual epitopes on glycoconjugates. X409 highlights new sophisticated ways the microbiota has evolved to target and adhere to mucins in the dynamic mucus environment, and our appreciation of the binding mode may reveal novel vulnerabilities that may be exploited therapeutically. We also envision that X409 may have utility for targeted delivery of bioactive molecules to the mucus, providing improved residence time and penetration within the mucus[7,20]. Finally, our results add to the discussion of sequence diversity of mucin TR regions between species and anatomical locations, providing support for selection pressure rather than chaos.

Methods

Bacterial strains and cell lines

E. coli strain DH5α was used for all molecular cloning and plasmid storage. *E. coli* strain BL21(DE3) (Invitrogen) was used for protein expression. *E. coli* cells were cultured in LB media supplemented with antibiotics (50 µg mL^{-1} kanamycin or 100 µg mL^{-1} ampicillin) for plasmid maintenance. Adherent HEK293 cells were maintained in DMEM/F-12 (1:1 v/v) (Gibco) supplemented with 10% heat-inactivated fetal bovine serum (SigmaAldrich) and 2 mM GlutaMAX (Gibco). Suspension HEK293 cells were cultured in serum-free F17 media (Invitrogen) supplemented with 4 mM GlutaMAX (Gibco) and 0.1% Kolliphor P188 (Sigma) under agitation (120 rpm). Suspension CHO cells were maintained in EX-CELL CHO CD Fusion serum-free media (Sigma) and BalanCD CHO Growth A media (FUJIFILM) (1:1 v/v) supplemented with 4 mM GlutaMAX (Gibco). All mammalian cell culture was performed at 37 °C and 5% CO_2. Authentication of each cell line used in this study included PCR assays with species-specific primers as previously described[11,35], and routine mycoplasma analysis. All bacterial and mammalian cells used in this study are listed in Supplementary Table 3.

Mucin reporter construct designs

Membrane-bound reporters were designed by fusion of human MUC1 signal peptide (amino acids 1–62, Uniprot P15921) with human MUC1 membrane anchor domain (amino acids 1042–1138), following by FLAG tag, sfGFP, and multiple cloning sites. For secreted reporters, the membrane anchor domain was replaced by 12xHis Tag. These constructs were then cloned into either pIRES or pGS for HEK293 or CHO expression, respectively[11,35]. Exchangeable reporter gene inserts were synthesized with BamHI/NotI cloning sites for pIRES or BglII/BamHI cloning sites for pGS (Genscript, USA) and inserted into the vector using standard restriction enzyme-based molecular cloning. Full sequences of glycomodules in the reporter are shown in Supplementary Fig. 1.

Cloning and purification of *E. coli* StcE X409 and mutants

The DNA sequence encoding amino acid residues 799–898 of *E. coli* StcE (X409 domain) was codon-optimised for *E. coli* expression and synthesised by GenScript (USA). This construct was subcloned into the pMALC2x vector, generating pMALC2x-MBP-10xHis-TEV-X409. QuikChange Site-Directed Mutagenesis Kit (Agilent) was used to generate X409 mutant plasmids. All plasmid was confirmed by Sanger sequencing. For expression and purification, the plasmid was transformed into BL21 (DE3) Gold cells, grown at 37 °C in 2XTY medium (16 g L^{-1} tryptone, 10 g L^{-1} yeast extract, 5 g L^{-1} NaCl, pH 7.5) supplemented with 100 µg mL^{-1} ampicillin. Protein expression was induced at optical density ~0.6 with 1.0 mM IPTG and culture was incubated at 18 °C for 18 h. Cells were harvested by centrifugation (6000 × *g*, 15 min, 4 °C), resuspended in buffer A (50 mM Tris pH 7.5, 500 mM NaCl), and lysed using sonication (Vibracell Sonics, 10 cycles of 30 s sonication (8 s on/2 s off) and 30 s resting with 80% amplitude). ~2 mg lysozyme, protease inhibitors (0.4 mM PMSF, 4 mM benzamidine, and 20 nM leupeptin at final concentration) and 750 units of benzonase (Novagen) were then added. Cell lysate was centrifuged (48,000 × *g*, 20 min, 4 °C). Supernatant was collected, filtered (0.45 µm), and applied to a Nickel His-Trap column. The column was washed and eluted using an imidazole gradient (5–500 mM). Following fraction collection, TEV protease (1 mg TEV:50 mg fusion protein) was added to cleave MBP at 18 °C for 3 days, and the solution was dialyzed in desalting buffer (25 mM Tris, pH 7.5, 250 mM NaCl) for 1–3 days. The cleaved protein was reapplied to the His-Trap column, concentrated, and purified by size exclusion chromatography using a HiLoad 26/60 Superdex 75 Prep Grade column. Despite the column's volume of 330 mL, the protein eluted at ~380 mL, facilitating an efficient purification. The final protein was concentrated to 40–50 mg mL^{-1} in crystallization buffer (20 mM Tris, pH 7.5). Protein concentration was confirmed using absorbance at 280 nm (ε = 25,440 M^{-1} cm^{-1}). Typically, 12–15 mg of purified protein was obtained from a 2.0 L culture. All purification steps were verified by SDS-PAGE analysis.

Cell-based production of mucin and Glycocarrier reporters

Glycoengineered HEK293 and CHO cells stably expressing secreted reporters were generated as previously described[23]. Briefly, stably expressing HEK293 and CHO cells were selected in media supplemented with 1 µg mL^{-1} puromycin (InvivoGen) or in media omitting GlutaMAX, respectively. Stable cells were seeded at 0.2 × 10^6 cells mL^{-1} and cultured for 5–7 days on an orbital shaker (120 rpm). Media containing secreted Glycocarrier was then harvested (1000 × *g*, 3 min, 10,000 × *g*, 20 min), diluted (100 mM sodium phosphate, pH 7.4, 2.0 M NaCl, 40 mM imidazole), and incubated with nickel-nitrilotriacetic acid (Ni-NTA) affinity resin (ThermoFisher Scientific) overnight at 4 °C. For CHO-based culture, prior to incubation with Ni-NTA, the media was incubated with ion exchanger Amberlite resin (MB-6113, SigmaAldrich) overnight to reduce non-specific interaction with the agarose bead. Resin was collected into 5 mL gravity column (ThermoFisher Scientific), washed (25 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM imidazole), and reporters were eluted with 300 mM imidazole. Yields were quantified by NuPAGE Coomassie and Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific). The mucin reporters were characterized as reported by combinations of intact MS, bottom-up MS, and/or O-glycoprofiling when possible[7,23,29].

ELISA

MaxiSorp 96-well plates (Nunc) were coated with purified reporters overnight in 50 µl carbonate-bicarbonate buffer (pH 9.6) at 4 °C, blocked (PO_4, Na/K, 1% Triton-X-100, 1% BSA, pH 7.4) for 1 h at RT, and incubated with binders for 1 h at RT. For inhibition ELISA, X409 binder was incubated with inhibitor in a separate 96-well assay plate for 30 minutes at RT prior to being added to the ELISA plate. Plates were then washed in PBS containing 0.05% Tween-20, incubated with 50 µl of secondary antibody for 1 h at RT, followed by development with TMB (Dako), termination with 0.5 M H_2SO_4, and measurement (450 nm) with Synergy LX (BioTek). All mAbs and lectins used in the study are provided in Supplementary Table 4.

Mass photometry analysis

Microscope cover glasses (No 1.5H, 24 × 50 mm, Paul Marienfeld GmbH) were prepared by washing, five times each, with MQ water and HPLC-grade isopropanol and dried under an air stream. A silicon gasket was fixed on clean cover glasses by gently pressing with forceps. Native protein marker (InvitroGen) was used to create mass calibration curves at 66, 146, 480, and 1048 kDa. For quality control analysis prior mixing, each protein stock was diluted to 100 nM in PBS and analyzed in mass photometry at 5 nM (Supplementary Fig. 3). Binding analysis was performed by mixing MBP-X409 with mucin or O-Glycocarrier reporters at 5 nM reporters to 50 nM MBP-X409 in a 20-µL reaction volume. The mixture was incubated at room temperature for 15 mins before analysis. For each acquisition, a new well in a silicon gasket was used, and the sample was introduced into the well. Following autofocus stabilization, a movie was recorded for 90 s at RT. All data acquisition was performed using AcquireMP software (Refeyn Ltd), and data were analysed using DiscoverMP (Refeyn Ltd). Data were presented as kernel density estimates with a 5 kDa bandwidth. The error value was estimated as a Gaussian curve fitting error as determined by DiscoverMP.

Intact mass spectrometry analysis

Purified glycomodules were resuspended in 50 µL MQ water and mixed with 50 µL methanol before direct-infused into Synapt G2 mass spectrometer equipped with a T-wave ion mobility cell (Waters Ltd) using in-house fabricated Pd/Pt-coated borosilicate tip. Samples were ionized using a capillary voltage of 1.0 kV, sample cone voltage of 45 V and an extraction cone voltage of 5 V. All samples were acquired in positive ion mode in mass range setting in *m/z* rang 500 – 5000. Data were analyzed using UniDec software[36]. At least two technical replications per sample were analysed.

Glycopeptide bottom-up analysis

Purified glycopeptides were resuspended in MQ water (50 ng/μL) and formic acid added (0.1% v/v), and analyzed by EASY-nLC 1200 UHPLC (ThermoFisher Scientific) interfaced via nanoSpray Flex ion source to an Orbitrap Fusion Lumos MS (ThermoFisher Scientific). The nLC was operated in an analytical column set up using PicoFrit Emitters (New Objectives, 75 mm inner diameter) packed in-house with Reprosil-Pure-AQ C18 phase (Dr. Maisch, 1.9-mm particle size, 19–21 cm column length). Each sample was injected onto the column and eluted in gradients from 3 to 32% B for glycopeptides, and 10 to 40% for released and labeled glycans in 45 min at 200 nL/min (Solvent A, 100% water; Solvent B, 80% acetonitrile; both containing 0.1% (v/v) formic acid). A precursor MS1 scan (m/z 350–2000) of intact peptides was acquired in the Orbitrap at the nominal resolution setting of 120,000, followed by Orbitrap HCD-MS2 and ETD-MS2 at the nominal resolution setting of 60,000 of the five most abundant multiply charged precursors in the MS1 spectrum; a minimum MS1 signal threshold of 50,000 was used for triggering data-dependent fragmentation events. Targeted MS/MS analysis was performed by setting up a targeted MSn (tMSn) Scan Properties pane.

MS data analysis

Glycopeptide analysis was performed from m/z features extracted from LC-MS data using in-house written SysBioWare software[37]. For m/z feature recognition from full MS scans Minora Feature Detector Node of the Proteome discoverer 2.2 (ThermoFisher Scientific) was used. The list of precursor ions (m/z, charge, peak area) was imported as ASCII data into SysBioWare and compositional assignment within 3 ppm mass tolerance was performed. The main building blocks used for the compositional analysis were: Neu5AC, Hex, HexNAc, dHex and the theoretical mass increment of the most prominent peptide corresponding to each potential glycosites. Upon generation of the potential glycopeptide list, each glycosite was ranked for the top 10 most abundant candidates, and each candidate structure was confirmed by doing targeted MS/MS analysis, followed by manual interpretation of the corresponding MS/MS spectrum.

Flow cytometry analysis

Binding analysis of reporters displayed on the cell surface can be performed either with a stably or a transiently expressing culture. For a stable expressing culture, 50,000 cells were seeded into 96 wells immediately before cytometer analysis. For transient expression culture, cells were seeded in a 24-well culture plate at 50–60% confluency one day before transfection. The next day, transfection reagent was prepared by diluting 0.5 μg DNA into 25 μL of 150 mM NaCl solution. In a separate tube, 2.5 μL polyethylenimine hydrochloride MAX (PEI, MW 40,000, Polysciences) solution was diluted into 22.5 μL Opti-MEM media (Gibco). PEI solution was then added into DNA solution and incubated at room temperature for 5 mins to allow DNA-PEI complex formation. The DNA-PEI complex solution was then slowly added to the culture well. Cells were kept in the incubator overnight to allow reporter expression. The next day, 50,000 cells were harvested and deposited into a 96-well plate. Cells were washed once with 200 μL PBA (1% BSA in PBS) buffer. Cells were resuspended in PBA containing X409 or other control binders and incubated on ice for 1 h. Cells were then washed twice with PBA buffer and incubated with secondary antibody in PBA for 1 h. Cells were washed three times, resuspended in 200 μL PBA and analyzed by flow cytometer (Fortessa X20, BD Biosciences). Results were analyzed using FlowJo software (Flowjo LLC) and expressed as geometric mean fluorescence of the activated gated cells.

Cell-based production of glycopeptides for structural study

20 mg purified reporter was digested with in-house purified TEV protease at 1:200 weight ratio at 30°C for 18 h in 50 mM ammonium bicarbonate buffer (pH 8.0). The reaction mixture was then incubated for 1 h with Ni-NTA resin at room temperature. Mixture was then centrifuged through a filter column, and flow-through containing glycodomain was collected and further C4-HPLC purified (Jupiter 4 μ Proteo™ WIDEPORE C4, 4 μm, 90 Å, 150 × 4.6 mm, Phenomenex) using a 0–100% gradient of 90% acetonitrile in 0.1% TFA. Fractions containing glycomodules were pooled, verified by ELISA using glycoform-specific lectins or mAbs, dried, and resuspended in 1.0 mL 50 mM ammonium bicarbonate buffer. The glycomodule was further digested with MS-grade trypsin (Roche) at a 1:200 (w/w) ratio for 18 h at 37°C. After trypsin inactivation by adding TFA to 0.1% v/v, glycopeptide was C18-HPLC purified (Kinetex C18, 2.6 μm, 100 Å, 100 × 4.6 mm, Phenomenex) using a 0–100% gradient of 90% acetonitrile in 0.1% TFA.

Chemical synthesis of glycopeptides

The glycopeptide 4xTn-STTT was synthesized following our well-stablished methodology[38]. The crude glycopeptide with *O*-acetylated glycosides was redissolved in CH_3CN/H_2O mixture and purified by reverse phase HPLC on a Phenomenex Luna C18(2) column (10 μm, 250 mm × 21.2 mm) with UV detection monitoring at 212 nm using a linear gradient of 0.1% TFA/CH_3CN from 90/10 to 60/40 v:v over 21 minutes (rt: 17.8 min) and lyophilized. HRMS ESI+ Calc. for $C_{93}H_{141}N_{15}O_{46}$ $[M + 2H]^{2+}/2$, 1101.9577; found 1101.9602. The peracetylated glycopeptide was resuspended in MeOH (10 mL) and the pH was adjusted to 9 with a methanolic solution of NaOMe (0.5 M) to cleave the acetyl groups and stirred for 2 h. The reaction was quenched using freshly activated and washed Amberlyst 15 (H^+) resin until pH < 7. The solution was filtered and concentrated under reduced pressure, furnishing 4xTn-STTT, which was used in the next step without further purification. Analytical UPLC on a Phenomenex BioZen C18 column (1.7 μM, 100 mm × 2.1 mm), 30 °C, Rt = 3.2 min (linear gradient: 0.1% formic acid in acetonitrile/0.1% formic acid, (1:99) → (20:80) over 10 min, $\lambda = 214$ nm). HRMS HRMS ESI+ calcd. for $C_{69}H_{117}N_{15}O_{34}$ $[M + 2H]^{2+}/2$, 849.8938; found: 849.8952.

The 4xT-STTT glycopeptide was produced from the 4xTn-STTT by enzymatic addition of Gal residues using the *Drosophila melanogaster* C1GalT1 enzyme. The 4xTn-STTT glycopeptide (50 mM) was incubated with purified C1GalT1 enzyme (50 μM) in 50 mM Tris-Cl buffer (pH 7.5), 150 mM NaCl, 200 μM $MnCl_2$, and 100 mM of UDP-Gal, at 37 °C for 2–3 days. The product was purified by C18-HPLC (Phenomenex Luna C18(2) column 10 μm, 250 mm × 21.2 mm) using a linear gradient of H_2O (containing 0.1% TFA)/CH_3CN from 99/1 to 85/15 v:v over 30 minutes (rt: 10.2 min) and lyophilized. MALDI MS Calc. for $C_{93}H_{155}N_{15}O_{54}Na$ $[M+Na]^+$, 2368.974; found 2368.847.

Crystallization

Crystals of X409 complexes were grown using sitting drop methods at 18 °C by mixing 0.25–0.5 μL of protein solution (45 mg mL^{-1} in crystallization buffer and peptide excess) with an equal volume of a reservoir solution. Glycopeptides were dissolved in crystallization buffer and titrated with NaOH 2.0 M until pH 8 was reached, added to protein solution, and following at least 20 min incubation, the mixture was centrifuged (30,000 × *g*, 20 min at 4 °C). Crystals were collected between 4 days and two weeks after the sitting drop experiment began. Crystals of X409-Tn-MUC5AC (AEAPT*T*S*T*T*SAPK, * denotes GalNAc residues) complex were obtained at 1:4 molar ratio (45 mg/L or 4 mM X409 and 16 mM glycopeptide) in 20% w/v Polyethylene glycol monomethyl ether (PEG MME) 5000, 200 mM di-sodium hydrogen phosphate and 20% glycerol as cryoprotectant. Crystals of X409-4xTn-STTT complex were obtained at 1:3 molar ratio (4 mM X409 and 13 mM peptide) in 20% w/v PEG MME 5000, 200 mM lithium chloride, and 25% glycerol. Crystals of X409-4xT-STTT complex were obtained at 1:3 molar ratio (4 mM and 13 mM peptide) in 20% w/v PEG MME 5000, 200 mM sodium fluoride and 20% glycerol.

Structure determination and refinement

Diffraction data for the three crystals of X409 were collected on beamline XALOC at the ALBA synchrotron (Barcelona, Spain) at a wavelength of 0.97 Å and a temperature of 100 K. XDS[39] and CCP4 software packages[40] were used for data processing and scaling. Relevant statistics are presented in Supplementary Table 1. Molecular replacement with Phaser and X409 coordinates from PDB entry 3UJZ as a template was used to solve the crystal structures. Initial phases were further improved by cycles of manual model building in Coot[41] and restrained refinement with REFMAC5[40]. Further rounds of model building in Coot with restrained refinement in REFMAC5 were performed for all complexes. The crystal structures were validated with PROCHECK, and model statistics are presented in Supplementary Table 1. The Ramachandran plot for the X409:4xTn-STTT complex shows that 88.1%, 11.9%, 0.0%, and 0.0% of the amino acids are in most favored, allowed, generously allowed and disallowed regions, respectively. The Ramachandran plot for the X409:4xTn-STTT complex shows that 91.8%, 8.2%, 0.0%, and 0.0% of the amino acids are in most favored, allowed, generously allowed and disallowed regions, respectively. The Ramachandran plot for the X409:4xT-STTT complex shows that 88.6%, 11.4%, 0.0%, and 0.0% of the amino acids are in most favored, allowed, generously allowed and disallowed regions, respectively.

Surface plasmon resonance (SPR) analysis

SPR experiments were performed at 25 °C with a Biacore X-100 apparatus (Biacore, GE) in running buffer (Tris 20 mM, pH 7.5). The protein X409 was immobilized on a CM5 sensor chip (Biacore, GE) following standard amine coupling method[42]. The carboxymethyl dextran surface of the flow cell 2 was activated with a 7-min injection of a 1:1 ratio of 0.4 M EDC and 0.1 M NHS. The proteins were coupled to the surface with a 7-minute injection at 100 μg ml^{-1} in 10 mM sodium acetate, pH 5.0. The unreacted *N*-hydroxysuccinimide esters were quenched by a 7-minute injection of 0.1 M ethanolamine-HCl (pH 8.0). The level of immobilization was ~2000 RUs. Flow cell 1 was treated as flow cell 2 (amine coupling procedure) without protein, which was used as a reference. For kinetic measurements, the concentration range was 5–60 nM for the 4xTn-STTT glycopeptide and 1–10 nM for the glycopeptide 4xT-STTT The glycopeptides were injected onto the sensor chip a flow rate of 30 µl min^{-1} for a period of 100 sec followed by a dissociation period of 600–900 sec. No regeneration was needed. Sensograms data were double-referenced and solvent corrected using the Biaevaluation X-100 software (Biacore, GE). The experimental data of affinity measurements were fitted to 1:1 binding model using Prism software. The association (k_{on}) and dissociation (k_{off}) rate constants were determined by nonlinear regression fitting using Biacore Evaluation Software, and the corresponding standard errors (δk_{on}, $\delta \mathrm{k}_{\mathrm{off}}$) were used to calculate the equilibrium dissociation constant (K_{d}) and its propagated error (δK_{d}) as: $\delta K_{\mathrm{d}} = K_{\mathrm{d}} \times \mathrm{sqrt}[(\delta k_{\mathrm{off}}/k_{\mathrm{off}})^2 + (\delta k_{\mathrm{on}}/k_{\mathrm{on}})^2]$.

Protein expression and purification for NMR analysis

To produce differently isotopic labeled X409 (^{15}N-X409, ^{13}C,^{15}N-X409) we expressed a fusion His-MBP-TEV-X409 in *E. coli* BL21(DE3) grown in M9 minimal medium with $^{15}NH_4Cl$ and ^{13}C-glucose supplementation. Induction was done at $OD_{600} = 0.6$ with 1 mM IPTG, at 37°C for 4 h. Cells were harvested by centrifugation (6371 × *g*, 15 min, 8 °C), and the cell pellet resuspended in 20 mM Tris (pH 8.0), 100 mM NaCl, 0.05% NaN_3, followed by lysis by sonication. Soluble X409 (12,857 × *g*, 20 min, 8 °C) was isolated by His-Trap HP (GE) using an imidazole gradient (20 mM −1 M). The eluate was dialyzed (10 mM phosphate (pH 7.4), 100 mM NaCl, 0.05% NaN_3) and digested with a TEV protease following His-Trap separation to remove the His-MBP and analysis by SDS-PAGE.

NMR sequential backbone assignment

All NMR assignment experiments were acquired in a Bruker Avance III 600 MHz spectrometer equipped with a Cryoprobe TCI (^{1}H, ^{13}C, ^{15}N). The data was processed with Bruker TopSpin 3.5 (Bruker) and analyzed with the assistance of computer aided resonance assignment (CARA) software version 1.9.1.7[43]. The resonance of 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid (TSP) was used as a chemical shift reference in the ^{1}H-NMR experiments (δ TSP = 0 ppm), and ^{13}C and ^{15}N chemical shifts were referenced indirectly via gyromagnetic ratios. Sequence-specific backbone assignment of X409 in apo-state was performed (BMRB ID 53125) using data from the following experiments (standard Bruker pulse sequences in parenthesis): 2D ^{1}H,^{15}N-HSQC (hsqcetfpf3gpsi2), ^{1}H,^{13}C-HSQC (hsqcetgpsisp2), 3D HNCO (hncogp3d), HN(CA)CO (hncacogp3d), HNCA (hncagpwg3d), HN(CO)CA (hncocagpwg3d), HNCACB (hncacbgpwg3d), CBCA(CO)NH (cbcaconhgp3d), (H)C(CCO)NH (ccconhgp3d) and ^{15}N-edited NOESY-HSQC (noesyhsqcfpf3gpsi3d, mixing time 90 ms). These experiments were acquired at 293 K using ^{13}C,^{15}N-X409 at 290 μM in 10 mM phosphate buffer pH 7.4, 105 mM NaCl and 0.01% NaN_3 in H_2O/D_2O 90:10. A total of 89 out of 99 amide backbone resonances of the apo-state protein were assigned (90% amide resonances) (X409 has 104 amino acids and with 5 Pro residues, 99 amide backbone resonances are expected), while ten residues (G795-G798; S814; Y828; and N856–Y859 could not be assigned). Sidechain amino groups (HE-NE) of Trp and Arg residues were also assigned. The backbone assignment of X409 bound to the 5xTn-MUC5AC glycopeptide also carried out (BMRB 53185, 94% amide resonances) with standard Bruker experiments, namely with 2D ^{1}H,^{15}N-HSQC, ^{1}H,^{13}C-HSQC, 3D HNCO, HN(CA)CO, HNCACB and ^{15}N-edited NOESY-HSQC (mixing time, 90 ms) experiments, acquired at 293 K using ^{13}C,^{15}N-X409 (215 μM) and Tn-MUC5AC (645 μM) at 1:3 molar ratio in 10 mM Phosphate buffer (pH 7.4), 105 mM NaCl, and 0.01% NaN_3 in H_2O/D_2O 90:10. Additionally, four backbone amide resonances were assigned in the bound state (N856–Y859). Amide resonances assignment of X409 in presence of 4xTn-STTT was achieved transferring the assignment from X409 in presence of Tn-MUC5AC, since the chemical shift perturbations of both bound forms were similar. However, 2D ^{1}H,^{15}N-HSQC and 3D ^{15}N-edited NOESY-HSQC experiments were acquired to clarify some assignments, in a Bruker Ascend 500-MHz spectrometer with a Prodigy cryoprobe CRPN2-TR-^{1}H&^{19}F/^{13}C/^{15}N-5mm-EZ, at 293 K, using a sample of ^{15}N-X409 (100 μM) and 4xTn-STTT (150 μM) at 1:1.5 molar ratio in 10 mM Phosphate buffer (pH 7.5), 105 mM NaCl, and 0.01% NaN_3 in H_2O/D_2O 90:10. NH and sidechain assignments of X409 in free and bound states are detailed in Supplementary Figs. 21 and 22.

NMR assignment of glycopeptides

Glycopeptides, Tn-MUC5AC, 4xTn-STTT and 4xTn-TTTT, were characterized in 10 mM phosphate buffer (pH 6.7), 150 mM NaCl and 0.01% NaN_3 in H_2O/D_2O 90:10. The concentration of the glycopeptides was 828 μM for 4xTn-STTT, 165 μM for 4xTn-TTTT and 400 μM for Tn-MUC5AC. ^{1}H-NMR assignment was performed through standard 2D-TOCSY (30 and 60 ms mixing time), 2D-NOESY (150 ms mixing time) and 2D ^{1}H,^{13}C-HSQC experiments (Supplementary Table 5). The resonance of TSP was used as a chemical shift reference in the ^{1}H-NMR experiments (δ TSP = 0 ppm). All spectra were acquired in a Bruker Avance III 600 MHz spectrometer equipped with a Cryoprobe TCI (^{1}H, ^{13}C, ^{15}N) at 287 K, processed and calibrated using the software TopSpin

4.4.0. The assignments were done using the CARA software version 1.9.1.7[43].

NMR ^{1}H,^{15}N-HSQC titrations

Titrations were carried out with increasing molar ratios of three different glycopeptides (Tn-MUC5AC, 4xTn-STTT and 4xTn-TTTT) in the presence of ^{15}N-X409 at 100 μM (with Tn-MUC5AC at 10, 25, 50, 75, 100, 200, and 500 μM, 4xTn-STTT at 10, 25, 50, 75, 100, 150, and 500 μM, or 4xTn-TTTT at 10, 25, 50, 75, 100, and 150 μM final concentration). The titrations involved the preparation of two samples, one of ^{15}N-X409 at 100 μM, the other of the highest molar ratio of the mixture (100 μM ^{15}N-X409 and stock concentration of glycopeptide); the points of the titration were done by adding calculated amounts of the mixture to the sample with only protein. All samples were prepared in 10 mM phosphate buffer (pH 7.3–7.6), 105 mM NaCl and 0.01% NaN_3 in H_2O/D_2O 90:10, with 50 μM TSP for chemical shift reference. Spectra from the Tn-MUC5AC titration and MUC5AC experiments were acquired in a Bruker Avance II + 600 14.1 T spectrometer equipped with a Cryoprobe TCI (^{1}H, ^{13}C, ^{15}N), at 293 K with 2048 × 256 points and 8 scans in a spectral window of 9615.4 Hz (center at 2802 Hz) × 2067.8 Hz (center at 7236 Hz), in ^{1}H and ^{15}N sweep's width, respectively. Spectra from the 4xTn-STTT and 4xTn-TTTT titrations were acquired in a Bruker Ascend 500-MHz spectrometer with a Prodigy cryoprobe CRPN2-TR-^{1}H&^{19}F/^{13}C/^{15}N-5mm-EZ, at 293 K with 2048 × 256 points and 16 scans in a spectral window of 8196.7 Hz (center at 2351.6 Hz) × 1724 Hz (center at 6033 Hz), in ^{1}H and ^{15}N sweep's width, respectively. The data was processed with Bruker TopSpin 4.1.1 and CCPN Analysis Version 3.2.2[44]. The chemical shift perturbation analysis for the Tn-MUC5AC and 4xTn-STTT titrations was done according to previous work[45], using the expression $\Delta\delta\text{comb} = \sqrt{((\Delta\delta H)^2 + (0.102\Delta\delta N)^2)}$. The intensity perturbation analysis was done by estimating the % of volume reduction of each peak by subtracting the volume of the peak of the bound form (Vb) to the free form (Vf) and dividing it by the volume of the free form ((Vf-Vb)/Vf)*100). The volumes were extracted using CARA version 1.9.1.7[43].

MD simulations of the free glycopeptides

MD simulations were performed for the 4xTn-STTT and 4xTn-TTTT glycopeptides free in solution using the AMBER package of NMRBox[46,47]. The peptide sequence was originated by PyMOL (Version 2.0 Schrödinger, LLC), while the sugars build-up with Carbohydrate Builder (https://glycam.org/cb/). The dihedral torsion angles were adjusted using the NMR data obtained from the conformational analysis of the free glycopeptides in solution with the ViewerPro 4.2 software (Accelrys Inc., San Diego, CA, USA). The peptide sequence was parameterized with the ff14SB force field[48], while the sugars were parameterized with the GLYCAM06j-1 force field[49]. The glycopeptides were immersed in a 10 Å water box with TIP3P water molecules[50] and the charge was neutralized with explicit counter ions. A two-stage geometry optimization approach was carried out. The first step minimized only the solvent molecules and ions using a 500 kcal mol^{-1} Å^{-2} harmonic potential, followed by a second minimization step involving all atoms of the system. The system temperature was gently heated from 0 to 300 K under constant pressure of 1 atm and periodic boundary conditions for 0.1 ns. Harmonic restraints of 10 kcal mol^{-1}Å^{-2} were applied to the solute, and the Andersen temperature coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages to self-adjust potential inhomogeneities. The long-range electrostatic effects were modeled using the particle-mesh-Ewald method[51]. The SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms, without applying hydrogen mass repartitioning. An 8 Å cutoff was applied to the Lennard-Jones and electrostatic interactions. The systems were equilibrated for 1 ns with a 2 fs time step at a constant volume and temperature of 300 K. The trajectories were run for 500 ns under the same simulation conditions, and triplicate runs were performed. One additional simulation was performed where the dihedral angles of some glycosylated residues were altered in both glycopeptides, in the 4xTn-STTT glycopeptide, the S1 and T4 were modeled with the eclipsed and staggered conformation, respectively; while for the 4xTn-TTTT glycopeptide, the T1 and T4 dihedrals were altered to a staggered conformation. The simulation trajectories were then analyzed using the CPPTRAJ module of AMBER 22 and the dihedrals as well as the residue-wise RMSF were calculated. MD simulation conditions for the free glycopeptides are provided in the Supplementary Table 6.

MD simulations of the complexes

Calculations were carried out with AMBER 22 package[46] implemented with GLYCAM06 force field[49]. The X-ray structures reported in this work were used as initial structures. The glycopeptides were build-up with Carbohydrate Builder (https://glycam.org/cb/). The complexes were immersed in a water box with a 10 Å buffer of TIP3P water molecules[50] and the system was neutralized by adding explicit counter ions (Cl^-). A two-stage geometry optimization approach was performed with the PMEMD module. The first stage minimizes only the positions of solvent molecules and ions, using a 50 kcal mol^{-1} Å $^{-2}$ harmonic potential, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. In both stages, 2500 steps of steepest descent minimization were followed by 2500 steps of conjugate gradient minimization. The systems were then heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions for 2 ns. Harmonic restraints of 10 kcal mol^{-1} were applied to the solute, and the Andersen temperature coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages. The SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method[51]. A real-space cutoff of 8.0 Å was applied to electrostatic and Lennard-Jones interactions. Each system was equilibrated for 2 ns with a 2-fs time step at a constant volume and temperature of 300 K. Production trajectory was then run for an additional 1 μs under the same simulation conditions. The trajectories were analyzed using the CPPTRAJ module of AMBER 22. Protonation states for all titratable residues were assigned using the default AMBER tleap settings at pH 7.0: Asp and Glu were modeled as deprotonated, Lys and Arg as protonated, and histidine residues were automatically assigned as HIE or HID based on their local hydrogen-bonding environment. Conformational analysis of GalNAc-Thr and GalNAc-Ser residues was performed to validate conformations sampled during simulation compared with the observed conformation in the crystal structure. Note, that all MD simulations were initiated from the experimentally validated X-ray crystal structure to ensure accurate ligand positioning and avoid artifacts from random starting poses. This approach allowed observed differences in ligand dynamics and stability to be attributed to the intrinsic molecular properties rather than initial placement bias. Equilibration state was assessed by monitoring the RMSD of both the protein backbone (Cα atoms) and the ligand over the entire duration of each trajectory (Supplementary Fig. 23). MD simulation conditions for glycopeptide in the complex are provided in the Supplementary Table 6.

Theoretical binding energy

Binding free energy estimations were carried out using the Generalized Born surface area (MM-GBSA) method along 1 μs MD trajectories. The analysis employed the MMPBSA.py module from the AMBER 22 software suite. Snapshots were extracted every 100 ps from the production phase of the simulation.

MM-GBSA calculations were performed in serial mode using the Generalized Born model igb = 5, with a salt concentration of 0.1 M. Per-residue energy decomposition (idecomp = 1) was enabled to evaluate the contribution of individual residues to the overall binding free energy. These per-residue values were averaged across all selected frames to obtain representative energy profiles. To manage storage efficiently, temporary files generated during the analysis were automatically deleted after processing each frame (keep_files = 2). The verbosity setting was configured to produce concise summary outputs while omitting detailed intermediate logs.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data generated in the study are included in this article and its supplementary information files. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository under accession number PXD066932. Atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers: 9GRJ, 9GRF, and 9GSM. NMR backbone assignments were deposited in Biological Magnetic Resonance Bank under BMRB ID 53125 and 53185, and titration data is available in the Source Data File. MD simulation structures are available on Zenodo database [https://doi.org/10.5281/zenodo.16741958]. Due to large data size, mass photometry video recordings are available upon request. Materials including X409 binder and Glycocarrier reporters are available from the corresponding author upon request. Source data are provided with this paper.

References

1. Hansson, G. C. Mucins and the microbiome. *Annu. Rev. Biochem.* **89**, 769–793 (2020).
2. Sonnenburg, J. L., Angenent, L. T. & Gordon, J. I. Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? *Nat. Immunol.* **5**, 569–573 (2004).
3. Luis, A. S. & Hansson, G. C. Intestinal mucus and their glycans: a habitat for thriving microbiota. *Cell Host Microbe* **31**, 1087–1100 (2023).
4. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **6**, 81 (2015).
5. Gustafsson, J. K. & Hansson, G. C. Immune regulation of goblet cell and mucus functions in health and disease. *Annu. Rev. Immunol.* **43**, 169–189 (2025).
6. Hollingsworth, M. A. & Swanson, B. J. Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer* **4**, 45–60 (2004).
7. Nason, R. et al. Display of the human mucinome with defined O-glycans by gene engineered cells. *Nat. Commun.* **12**, 4070 (2021).
8. Alberdi, A., Andersen, S. B., Limborg, M. T., Dunn, R. R. & Gilbert, M. T. P. Disentangling host-microbiota complexity through hologenomics. *Nat. Rev. Genet.* **23**, 281–297 (2022).
9. Rillahan, C. D. & Paulson, J. C. Glycan microarrays for decoding the glycome. *Annu. Rev. Biochem.* **80**, 797–823 (2011).
10. Smith, D. F. & Cummings, R. D. Application of microarrays for deciphering the structure and function of the human glycome. *Mol. Cell Proteom.* **12**, 902–912 (2013).
11. Narimatsu, Y. et al. An atlas of human glycosylation pathways enables display of the human glycome by gene-engineered cells. *Mol. Cell* **75**, 394–407 e395 (2019).
12. Deng, L. et al. Oral streptococci utilize a Siglec-like domain of serine-rich repeat adhesins to preferentially target platelet sialoglycans in human blood. *PLoS Pathog.* **10**, e1004540 (2014).
13. Narimatsu, Y. et al. A family of di-glutamate mucin-degrading enzymes that bridges glycan hydrolases and peptidases. *Nat. Catal.* **7**, 386–400 (2024).
14. Bull, C. et al. Probing the binding specificities of human Siglecs by cell-based glycan arrays. *Proc. Natl. Acad. Sci. USA* **118**, e2026102118 (2021).
15. Elzinga, J. et al. Binding of Akkermansia muciniphila to mucin is O-glycan specific. *Nat. Commun.* **15**, 4582 (2024).
16. Pluvinage, B. et al. Architecturally complex O-glycopeptidases are customized for mucin recognition and hydrolysis. *Proc. Natl. Acad. Sci. USA* **118**, e2019220118 (2021).
17. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781 (2004).
18. Nakjang, S., Ndeh, D. A., Wipat, A., Bolam, D. N. & Hirt, R. P. A novel extracellular metallopeptidase domain shared by animal host-associated mutualistic and pathogenic microbes. *PLoS ONE* **7**, e30287 (2012).
19. Yu, A. C., Worrall, L. J. & Strynadka, N. C. Structural insight into the bacterial mucinase StcE essential to adhesion and immune evasion during enterohemorrhagic E. coli infection. *Structure* **20**, 707–717 (2012).
20. Shon, D. J. et al. An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc. Natl Acad. Sci. USA* **117**, 21299–21307 (2020).
21. Huang, X. et al. Vibrio cholerae biofilms use modular adhesins with glycan-targeting and nonspecific surface binding domains for colonization. *Nat. Commun.* **14**, 2104 (2023).
22. Varki, A. Selectin ligands. *Proc. Natl Acad. Sci. USA* **91**, 7390–7397 (1994).
23. Jaroentomeechai, T. et al. Mammalian cell-based production of glycans, glycopeptides and glycomodules. *Nat. Commun.* **15**, 9668 (2024).
24. Corzana, F. et al. Serine versus threonine glycosylation: the methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *J. Am. Chem. Soc.* **129**, 9458–9467 (2007).
25. Svensson, F., Lang, T., Johansson, M. E. V. & Hansson, G. C. The central exons of the human MUC2 and MUC6 mucins are highly repetitive and variable in sequence between individuals. *Sci. Rep.* **8**, 17503 (2018).
26. Medley, B. J. et al. A previously uncharacterized O-glycopeptidase from Akkermansia muciniphila requires the Tn-antigen for cleavage of the peptide bond. *J. Biol. Chem.* **298**, 102439 (2022).
27. Taleb, V. et al. Structural and mechanistic insights into the cleavage of clustered O-glycan patches-containing glycoproteins by mucinases of the human gut. *Nat. Commun.* **13**, 4324 (2022).
28. Grys, T. E., Siegel, M. B., Lathem, W. W. & Welch, R. A. The StcE protease contributes to intimate adherence of enterohemorrhagic Escherichia coli O157:H7 to host cells. *Infect. Immun.* **73**, 1295–1303 (2005).
29. Konstantinidi, A. et al. Exploring the glycosylation of mucins by use of O-glycodomain reporters recombinantly expressed in glycoengineered HEK293 cells. *J. Biol. Chem.* **298**, 101784 (2022).
30. Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5**, 1087–1104 (2009).
31. Schjoldager, K. T., Narimatsu, Y., Joshi, H. J. & Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **21**, 729–749 (2020).
32. Tolia, N. H., Enemark, E. J., Sim, B. K. & Joshua-Tor, L. Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. *Cell* **122**, 183–193 (2005).
33. Wisnovsky, S. et al. Genome-wide CRISPR screens reveal a specific ligand for the glycan-binding immune checkpoint receptor Siglec-7. *Proc. Natl. Acad. Sci. USA* **118**, e2015024118 (2021).
34. Chen, J. et al. *Vibrio* MARTX toxin binding of biantennary N-glycans at host cell surfaces. *Sci. Adv.* **11**, eadt0063 (2025).
35. Yang, Z. et al. Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.* **33**, 842–844 (2015).

36. Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).
37. Vakhrushev, S. Y., Dadimov, D. & Peter-Katalinic, J. Software platform for high-throughput glycomics. *Anal. Chem.* **81**, 3252–3260 (2009).
38. Bermejo, I. A. et al. Water sculpts the distinctive shapes and dynamics of the tumor-associated carbohydrate Tn antigens: implications for their molecular recognition. *J. Am. Chem. Soc.* **140**, 9952–9960 (2018).
39. Kabsch, W. Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
40. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760–763 (1994).
41. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
42. Johnsson, B., Lofas, S. & Lindquist, G. Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Anal. Biochem.* **198**, 268–277 (1991).
43. Keller, R. The Computer Aided Resonance Assignment Tutorial, Vol. 1. (CANTINA Verlag, Switzerland; 2004).
44. Skinner, S. P. et al. CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* **66**, 111–124 (2016).
45. Lima, C. D. L. et al. Structural insights into the molecular recognition mechanism of the cancer and pathogenic epitope, LacdiNAc by immune-related lectins. *Chemistry* **27**, 7951–7958 (2021).
46. Case, D. A. et al. AmberTools. *J. Chem. Inf. Model* **63**, 6183–6191 (2023).
47. Maciejewski, M. W. et al. NMRbox: a resource for biomolecular NMR computation. *Biophys. J.* **112**, 1529–1534 (2017).
48. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
49. Kirschner, K. N. et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655 (2008).
50. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
51. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

Acknowledgements

This work was supported by the NEYE Foundation and the EMBO postdoctoral fellowship (ALTF 336-2021 to T.J., ALTF 105-2023 to F.G.), the Novo Nordisk Foundation (NNF24OC0088218 to Y.N., NNF21OC0071658 to H.C., NNF22OC0073736 to R.L.M., NNF22OC0076899 to H.J.), the Carlsberg Foundation (CF20-0412 to R.L.M.), and the Lundbeck Foundation, the Danish National Research Foundation (DNRF107, 196), and a Veni grant from the Dutch Research Council (NWO VI.Veni.202.045 to C.B.). We also thank the Spanish Ministry of Science, Innovation and Universities (PID2022-136362NB-I00 to R.H-G., PID2022-137973NB-I00 to P.M., and PID2021-127622OB-I00 to F.C.), Universidad de La Rioja (REGI22/16 Project) and Gobierno de Aragón (E34_R17, E34_R23, and LMP58_18) with FEDER (2014–2020) funds for "Building Europe from Aragón" for financial support (to R.H-G.). We thank the ALBA (Barcelona, Spain) synchrotron beamline XALOC. M.G. acknowledges the Marie Skłodowska-Curie fellowship program (no. 101034288) for financial support. F.M., C.O.S., A.S.G, J.S.D. H.Co thank Fundação para a Ciência e a Tecnologia (FCT-Portugal) for MGL4Life project 10.54499/PTDC/QUI-OUT/2586/2020. F.M. and H.Co. acknowledge FCT-Portugal for the CEECINST/00042/2021/CP1773/CT0011 and 10.54499/2020.03261CEECIND/CP1586/CT0012 research contracts, respectively. C.O.S. thanks FCT-Portugal for the PhD grant 10.54499/2022.11723.BD. F.M., C.O.S., A.S.G., J.S.D., and H.Co. acknowledge FCT-Portugal for the funding UIDP/04378/2020 and UIDB/04378/2020 to UCIBIO research unit and LA/P/0140/2020 to the i4HB Associate Laboratory. The NMR spectrometers are part of the National NMR Network (PTNMR) and are partially supported by Infrastructure Project No 22161. Part of MDs were performed using NMRbox, which is supported by NIH grant P41GM111135 (NIGMS). Part of MDs were performed using supercomputer Agustina, which is supported by BIFI-ZCAM, University of Zaragoza, Spain.

Author contributions

T.J., F.G., C.B., R.L.M., R.V., and S.F. performed research and analyzed data. B.V., V.T., and R.H-G. solved and analyzed the X-ray structural studies. C.O.S., A.S.G., J.S.D., H.Co., and F.M. performed and analyzed the NMR studies. I.G-A., P.M., A.S.G., and F.C. performed the MD simulations. M.G., I. C., and F.C. synthetized the glycopeptides. T.J., B.H., H.J.J., F.M., H.C., H.J., R.H-G., and Y.N. conceived the project, designed experiments, and contributed to writing of the manuscript. All authors read and approved of the final manuscript.

Competing interests

Y.N. and H.C. have a financial interest in GlycoDisplay Aps, Y.N. and H.C.'s interests are reviewed and managed by the University of Copenhagen in accordance with their conflict of interest policies. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-63756-w.

Correspondence and requests for materials should be addressed to Ramon Hurtado-Guerrero or Yoshiki Narimatsu.

Peer review information *Nature Communications* thanks Xu Yang and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints