Mucins are a family of large heavily O-glycosylated proteins with tandem repeated sequences (TRs) containing a high proportion of threonines and serine residues serving as O-glycan attachment sites. Mucin TRs undergo GalNAc-type O-glycosylation (hereafter simply O-glycosylation) by a large family of polypeptide GalNAc-transferases and O-glycosylation is one of the most abundant and diverse types of posttranslational modifications (PTMs). The mucin TR domains, which are densely covered with O-glycans, form a rigid bottlebrush-
like structure that is largely resistant to general degradation by traditional proteases. Mucins line all mucosal surfaces and are the major macromolecules in body fluids serving essential functions in clearance, containment, feeding, orienting, and continuously replenishing our microbiomes and selecting for commensals to repress pathogenic microorganisms. Changes in expression and glycosylation of mucins are associated with human diseases including cancer, and the mucins MUC1, MUC4, and MUC16 serve as circulating biomarkers of different types of cancers. Mucins of cancer cells are found with truncated or aberrant O-glycans, well-known as specific human tumor-associated carbohydrate antigens (TACAs). Among TACAs, the Tn (GalNAcα1-O-Ser/Thr), the T (Galβ1-3GalNAcα1-O-Ser/Thr), and STn (NeuSAcα2-6GalNAcα1-O-Ser/Thr) antigens stand out as the most prevalent and expression of these epitopes are thought to promote tumorigenesis and metastasis.

The large gel-forming mucins form oligomeric networks or extended bundles through disulfide bridging in their C/N-termini as part of mucus layers to provide protection for epithelial cells by limiting activation of inflammatory cascades and bacterial contact. In the intestine, cross-linked networks of mucin MUC2 form a dense mucus layer that serves as a barrier for microorganisms and a loose mucus layer that traps and contains the microbiome. In the airways, mucins MUC5B with MUC5AC form long thick bundles that sweep surfaces by cilia movements. Most gut bacteria use dietary fibers and starches as a nutrient source, while a subset of these species can also degrade and ferment mucins. By removing sugars such as sialic acid, these enzymes can create new epitopes that elicit inflammatory responses. Mucins are important for the continuous renewal process and homeostasis of the mucus layers, but they are also used by pathogens to degrade the mucin layers and invade the mucus and reach the underlying epithelium. Understanding the structure and molecular mechanisms of microbial mucinases in the human gut and their substrate specificities is therefore of utmost importance.

Although the definition of mucinases is vague in the literature, they may be considered as a subclass of O-glycoproteases that serve a more limited number of substrates such as mucins and mucin domain-containing glycoproteins in comparison to general O-glycoproteases, such as A. muciniphila OgpA and Acinetobacter CpaA that target all types of O-glycoproteins. Moreover, mucinases may contain separate mucin-binding modules, such as the X409 module identified on the Escherichia coli O157:H7 StcE mucinase, which is believed to drive StcE mucinase to its mucin substrates. Bacterial O-glycoproteases and mucinases mostly cleave glycopeptides N-terminally to contiguous glycosylated Ser or Thr residues (e.g. B. thetaiotaomicron BT4244 and A. muciniphila AM0908), with exceptions for StcE and Streptococcus pneumoniae ZmpC. These last two enzymes cleave glycopeptides C-terminally to glycosylated Ser/Thr residues that are near but not contiguous to the cleavage point (Fig. 1b). Most mucinases appear to recognize part of the O-glycan(s) adjacent to the O-glycopeptide cleavage site. Interestingly the mucinases characterized to date have different tolerance for the structure of the O-glycans guiding the peptide bond cleavage, with several mucinases accepting only truncated O-glycans such as the simple Tn and T structures without sialic acid capping, while for example StcE functions better with elaborated core 1 and core 2 O-glycan structures and blocked by α2–6 linked sialic acid in the core 1 di-sialyl-T (dST) O-glycans as well as by core 3 O-glycans (see Fig. 1a, b for the structures of some of these O-glycans). These differences in O-glycan substrate specificities are likely related to their biological roles. Mucinases from commensals may participate in foraging mucins after their O-glycans have been trimmed by exoglycosidases including sialidases to regulate normal mucus homeostasis. However, mucinases like StcE from invading pathogens may instead participate in the pathogenesis by destroying the most nascent mucins with intact O-glycans and penetrating the dense mucus layer that protects the microbiome from reaching the epithelium. Mucinases are expressed by commensal bacteria such as...
B. thetaiotaomicron (BT4244), A. muciniphila (AM0908, AM1514 and AM0627), and pathogenic bacteria such as Clostridium perfringens (ZmpB), Pseudomonas aeruginosa (IMPa), S. pneumoniae (ZmpC), E. coli O157:H7 (StcE), and enteroaggregative E. coli (Pic)4. Some StcE and ZmpC have been shown to cleave glycopeptides containing complex O-glycans (Fig. 1a, b). Mucinases and O-glycoproteases appear to be promiscuous in terms of substrate peptide sequence specificity4, although refinement of substrate peptide sequences is still in progress24.

While most of the enzymes listed above depend on a single O-glycan for cleavage of O-glycopeptides, A. muciniphila AM0627 was recently described to require two adjacent truncated O-glycans and cleave in between the two O-glycosites (Fig. 1a, b). Mucinases and O-glycoproteases appear to be promiscuous in terms of substrate peptide sequence specificity4, although refinement of substrate peptide sequences is still in progress24.

Here, we applied a multidisciplinary approach encompassing biophysical and computational techniques to reveal the molecular basis for catalysis and O-glycopeptide selectivity by AM0627. We report the structure of AM0627 in complex with a bis-T glycopeptide and demonstrate that the enzyme interacts with both glycans in the bis-O-glycopeptide, preferentially recognizing the Galβ1-3GalNAc disaccharide (T O-glycan) at the G1 and G2 subunits and the GalNAc sugar (Tn O-glycan) at the G1' subunit (Fig. 1b; see also the accepted subsite nomenclature for the sugar moieties). Moreover, we uncover the enzyme catalytic mechanism by QM/MM simulations, showing that a nucleophilic water and a catalytic base Glu residue are required for efficient catalysis. We demonstrate, using well-defined mucin reporter substrates, that AM0627 prefers T over Tn O-glycan substrates and is essentially inactive with sialylated complex O-glycosylated substrates.

Finally, we identify a Tyr residue in AM0627 that is conserved in BT4244, which participates in a key interaction with the T O-glycan at S1 and provides a basis for the bis-O-glycan substrate specificity of AM0627 and likely also BT4244 mucinases.

**Results**

**Architecture of the AM0627-bis-T glycopeptide-Zn\(^{2+}\) complex**

To gain insights into the structure of AM0627 and its recognition and cleavage of bis-O-glycans-containing glycopeptides, we pursued the determination of its structure by X-ray crystallography. Initially, we chemoenzymatically synthetized a P-selectin glycoprotein ligand 1 (PSGL-1)-like bis-T glycopeptide (hereafter P1 and defined by the

---

**Fig. 2 | Crystal structure of the AM0627E326A–P1–Zn\(^{2+}\) complex.**

(a) Ribbon structure of AM0627E326A complexed to P1 and Zn\(^{2+}\) (upper panel). In the lower panel, the monomeric form is displayed in a different view. The IgG fold and the catalytic domains are colored in orange and cyan, respectively. The secondary structures are only indicated in the upper panel and α6 is highlighted in red. The P1 amino acids are depicted with green carbon atoms whereas the GalNAc and Gal are shown as yellow and magenta carbon atoms, respectively. While the GalNAc moieties are shown in yellow according to the SNFG6, the Gal moieties are shown in magenta for illustration purposes. The Zn\(^{2+}\) and a water molecule are shown as a pink and red sphere, respectively.

(b) Close-up view of the active site showing the bound Zn\(^{2+}\), a water molecule, and P1. His325, Ala326, His329, and Glu343 are shown as cyan sticks. Electron density maps are Fo-Fc (blue) contoured at 2.2σ for P1 and 2Fo-Fc (blue) contoured at 1σ for Zn\(^{2+}\) and a water molecule. Only the residues of P1 with well-defined density (Ala3 to Thr6) and partial density are displayed (Thr1, Glu2, Pro7, and the amine group of Pro8). Close-up view of the active site showing the pentagonal geometry formed by the Zn\(^{2+}\) and the coordinating residues (see dotted yellow lines for the interactions between the metal and the amino acids). The hydrogen bonds between the water molecule and P1 are indicated as dotted brown lines.

---

B. thetaiotaomicron (BT4244), A. muciniphila (AM0908, AM1514 and AM0627), and pathogenic bacteria such as Clostridium perfringens (ZmpB), Pseudomonas aeruginosa (IMPa), S. pneumoniae (ZmpC), E. coli O157:H7 (StcE), and enteraggregative E. coli (Pic)4. So far only StcE and ZmpC have been shown to cleave glycopeptides containing complex O-glycans (Fig. 1a, b). Mucinases and O-glycoproteases appear to be promiscuous in terms of substrate peptide sequence specificity4, although refinement of substrate peptide sequences is still in progress24.

While most of the enzymes listed above depend on a single O-glycan for cleavage of O-glycopeptides, A. muciniphila AM0627 was recently described to require two adjacent truncated O-glycans and cleave in between the two O-glycosites (Fig. 1b; see also the accepted subsite nomenclature for the sugar moieties). Moreover, we uncover the enzyme catalytic mechanism by QM/MM simulations, showing that a nucleophilic water and a catalytic base Glu residue are required for efficient catalysis. We demonstrate, using well-defined mucin reporter substrates, that AM0627 prefers T over Tn O-glycan substrates and is essentially inactive with sialylated complex O-glycosylated substrates. Finally, we identify a Tyr residue in AM0627 that is conserved in BT4244, which participates in a key interaction with the T O-glycan at S1 and provides a basis for the bis-O-glycan substrate specificity of AM0627 and likely also BT4244 mucinases.
domain, which was active and comparable in activity to the wild-type (wt) AM0627A21-Pro71 and the Tn moiety was designed a shorter construct of the E326A-inactive mutant that started in Glu506; see Supplementary Fig. 1a and see the methods section. Despite crystals appeared, these diffracted poorly and prompted us to determine the structure of the AM0627E326A (P71-E506) enzyme construct produced better diffracting crystals. We, therefore, tested if the truncated AM0627E326A (AM0627P71-E506) enzyme construct produced better diffracting crystals. We, therefore, tested if the truncated AM0627E326A (AM0627P71-E506) enzyme was active and comparable in activity to the wild-type (wt) AM0627E326A (P71-E506) enzyme using an artificial Tn bis-O-glycan reporter based on our previously reported mucin reporter design. This reporter contains multiple residues with bis-O-glycosyls (AEAAATTTPAPAKm)1n, and the Tn O-glycoform of this reporter was produced homogeneously in HEK293 cells with KO of CIGALTI. The WT AM0627E326A (P71-E506) and the AM0627P71-E506 exhibited similar efficient cleavage of this O-glycan substrate when decorated by Tn O-glycans (Supplementary Fig. 1b), confirming that the truncated AM0627P71-E506 represented a fully active enzyme suitable for structural studies. The crystal structure of AM0627E326A was solved at 1.9 Å using zinc single-wavelength anomalous dispersion (Zn-SAD; see the Methods section). The model obtained from SHELXE allowed us to solve the structure of the AM0627E326A–Zn2+ complex at a higher resolution (1.5 Å) by molecular replacement using PHASER (see Supplementary Table 1 and see the “Methods” section). Although the asymmetric unit (AU) of P212121 crystals contained two molecules of AM0627E326A that partially contacted each other, gel filtration chromatography (Supplementary Fig. 2) showed that AM0627E326A was monomeric, which was further confirmed by the PISA server. The AM0627E326A crystal structure revealed two distinct domains formed by an Ig-like fold domain and the M60-like catalytic domain (Fig. 2a). Ig-like fold domains were also found in previous mucinases crystal structures such as the ones for BT4244, ZmpB and IMPa29. With respect to the M60-like catalytic domain (Fig. 2a), the M60-like catalytic domain30 which, according to MEROPS database, is also present in AM0908. The root-mean-square deviation (RMSD) between both molecules belonging to chains A and B in the AU is 0.44 Å on 437 equivalent Cα atoms. Hereafter we will discuss only molecule B because it contains a better-defined density for the AM0627E326A structure, Glu326 was mutated to Ala326) and His329 was replaced each other, gel filtration chromatography (Supplementary Fig. 2) showed that AM0627E326A was monomeric, which was further confirmed by the PISA server. The AM0627E326A crystal structure revealed two distinct domains formed by an Ig-like fold domain and the M60-like catalytic domain (Fig. 2a). Ig-like fold domains were also found in previous mucinases crystal structures such as the ones for BT4244, ZmpB and IMPa29. With respect to the M60-like catalytic domain, BT4244 and ZmpB structures share a similar M60-like catalytic domain (Fig. 2a). Ig-like fold domains were also found in previous mucinases crystal structures such as the ones for BT4244, ZmpB and IMPa29. With respect to the M60-like catalytic domain, BT4244 and ZmpB structures share a similar M60-like catalytic domain30 which, according to MEROPS database, is also present in AM0908. The root-mean-square deviation (RMSD) between both molecules belonging to chains A and B in the AU is 0.44 Å on 437 equivalent Cα atoms. Hereafter we will discuss only molecule B because it contains a better-defined density for the AM0627E326A structure, Glu326 was mutated to Ala326) and His329 was replaced each other, gel filtration chromatography (Supplementary Fig. 2) showed that AM0627E326A was monomeric, which was further confirmed by the PISA server. The AM0627E326A crystal structure revealed two distinct domains formed by an Ig-like fold domain and the M60-like catalytic domain (Fig. 2a). Ig-like fold domains were also found in previous mucinases crystal structures such as the ones for BT4244, ZmpB and IMPa29. With respect to the M60-like catalytic domain, BT4244 and ZmpB structures share a similar M60-like catalytic domain30 which, according to MEROPS database, is also present in AM0908. The root-mean-square deviation (RMSD) between both molecules belonging to chains A and B in the AU is 0.44 Å on 437 equivalent Cα atoms. Hereafter we will discuss only molecule B because it contains a better-defined density for the AM0627E326A structure, Glu326 was mutated to Ala326) and His329 was replaced each other, gel filtration chromatography (Supplementary Fig. 2) showed that AM0627E326A was monomeric, which was further confirmed by the PISA server. The AM0627E326A crystal structure revealed two distinct domains formed by an Ig-like fold domain and the M60-like catalytic domain (Fig. 2a). Ig-like fold domains were also found in previous mucinases crystal structures such as the ones for BT4244, ZmpB and IMPa29. With respect to the M60-like catalytic domain, BT4244 and ZmpB structures share a similar M60-like catalytic domain30 which, according to MEROPS database, is also present in AM0908. The root-mean-square deviation (RMSD) between both molecules belonging to chains A and B in the AU is 0.44 Å on 437 equivalent Cα atoms. Hereafter we will discuss only molecule B because it contains a better-defined density for the AM0627E326A structure, Glu326 was mutated to Ala326) and His329 was replaced each other, gel filtration chromatography (Supplementary Fig. 2) showed that AM0627E326A was monomeric, which was further confirmed by the PISA server. The AM0627E326A crystal structure revealed two distinct domains formed by an Ig-like fold domain and the M60-like catalytic domain (Fig. 2a). Ig-like fold domains were also found in previous mucinases crystal structures such as the ones for BT4244, ZmpB and IMPa29. With respect to the M60-like catalytic domain, BT4244 and ZmpB structures share a similar M60-like catalytic domain30 which, according to MEROPS database, is also present in AM0908. The root-mean-square deviation (RMSD) between both molecules belonging to chains A and B in the AU is 0.44 Å on 437 equivalent Cα atoms. Hereafter we will discuss only molecule B because it contains a better-defined density for the AM0627E326A structure, Glu326 was mutated to Ala326) and His329 was replaced each other, gel filtration chromatography (Supplementary Fig. 2) showed that AM0627E326A was monomeric, which was further confirmed by the PISA server. The AM0627E326A crystal structure revealed two distinct domains formed by an Ig-like fold domain and the M60-like catalytic domain (Fig. 2a). Ig-like fold domains were also found in previous mucinases crystal structures such as the ones for BT4244, ZmpB and IMPa29. With respect to the M60-like catalytic domain, BT4244 and ZmpB structures share a similar M60-like catalytic domain30 which, according to MEROPS database, is also present in AM0908. The root-mean-square deviation (RMSD) between both molecules belonging to chains A and B in the AU is 0.44 Å on 437 equivalent Cα atoms.
shown to be variable in length and suggested to underlie the difference in substrate specificity between these enzymes. An additional conserved Glu residue (Glu343 in AM0627) together with the two His residues of the HEXXH motif and the Thr5 carbonyl group coordinate the zinc ion to form a pentagonal geometry (Fig. 2c), a feature that fits with AM0627 belonging to the gluzincin-like family of zinc metallopeptidases. The zinc ion in the Zn$^{2+}$-bound AM0627 crystal structure was only coordinated in a trigonal geometry by His325, His329, and Glu343. In addition, a water molecule did not coordinate the zinc ion and therefore did not replace the role of the Thr5 carbonyl group in coordinating the metal. However interestingly, in our structure, we observed the presence of a potential catalytic water molecule, further discussed below, which was visualized in the active site establishing hydrogen bonds with the Thr5 carbonyl group, the
AM0627 recognizes bis-T O-glycans within a specific peptide sequence

The AM0627-active site is formed by the zinc-binding site, discussed above, and the glycopeptide binding site (Fig. 3a). At the level of the peptide sequence of P1, the main enzyme–peptide interactions are as follows. The Glu2 side chain makes a hydrogen bond with the Asp292 side chain, the Ala3 methyl group establishes CH–π interactions with the side chains of Trp149/Phe290, and Gln4 forms hydrogen bonds with Arg291 (both side chain and backbone). Regarding the two glycosylated Thr residues (Thr5 and Thr6 in Fig. 3a), the Thr5 backbone forms a hydrogen bond with the Tyr470 side chain, the Thr5 methyl group forms a CH–π interaction with the Tyr287 side chain and the Thr6 backbone makes a hydrogen bond with the Tyr287 side chain (Fig. 3a).

AM0627 preferably cleaves bis-T over bis-Tn substrates and depends on the GalNAc at G1 as the minimal O-glycan structure for activity

AM0627 was recently reported to cleave bis-O-glycopeptides containing T O-glycans as well as Tn and sialylated core I (mSta) O-glycans (Fig. 1b) and further evidence for this was presented in the recent report. To get further insights into the type of O-glycans being recognized and the minimal O-glycan structure required for cleavage, we synthesized a battery of glycopeptides containing GalNAc, galactosamine (GalN), and Gal-GalNAc (see Fig. 4a). The synthesis of the glycopeptides with different O-glycan positions and structures (P2–P9) were based on the P1 peptide sequence (TEATQTPPPPA). The P2 glycopeptide contained a bis-Tn while P5 and P6 glycopeptides had one single GalNAc moiety at G1 and G1, respectively. In addition, we synthesized one diglycopeptide containing a GalNAc moiety at G1 and a GalN at G1 (P8), and another one with the inverse order of sugar units (P7). The two mono-T glycopeptides P3 and P4, which contain the mono-T O-glycan in the SI and SI′ positions, respectively, were chemoenzymatically synthesized using P3 and P6 as templates. A naked peptide, P9, was also made to confirm whether cleavage might take place in the absence of the O-glycans. The use of GalN in place of GalNAc was considered in order to explore the role of the acetyl group in substrate recognition (Fig. 4a).

The MALDI-TOF time-course assays with the wt AM0627 enzyme and different glycopeptide substrates showed that the P1 and P2 glycopeptides with bis-T and bis-Tn O-glycans, respectively, served as the best substrates with cleavage of P1 being slightly faster than that of P2 at the early time points expected with maximum velocities of reactions (e.g. at 5 min, ~40% P1 and ~25% P2 were cleaved while at 15 min, ~70% P1 and ~35% P2 were cleaved). Note that the highest specific activity for AM0627 using the P1 was estimated because under a higher concentration of substrate (500 μM) and short time points (5–30 min), the reaction was almost linear, reaching ~50% cleavage of P1 at 10 min and 100% cleavage at 120 min. This was likely due to substrate inhibition.
taking place at a high concentration for P1 (compare Figs. 3c and 4b). Based on the same rationale described above, the specific activity for the second-best substrate, P2, was 1.6-fold lower than that of P1 (1.2 U/mg for P2 versus 1.9 U/mg for P1). The P5 mono-Tn (T*-T-) glycopeptide was a considerably poorer substrate (<10% cleavage at 5 min and <15% at 15 min), followed by P8 (T^+T^-) with a modified bis Tn (<5% cleavage at 5 min and <10% at 15 min) and the P3 mono-T (T^+T^-) with barely detectable cleavage (<5% at 15 min). The P4 mono-T (T^-T^-), the P6 mono-Tn (T^+T^-), and the P7 glycopeptide (T^-T^-) were completely resistant to cleavage. Finally, the naked P9 peptide without O-glycans served as the non-substrate control demonstrating that the presence of O-glycans was absolutely required for catalysis (Fig. 4b and Supplementary Fig. 4). Comparison of the enzyme activity with P7 (T^-T^-) versus P8 (T^+T^-) indicated that the acetyl group of the GalNAc at the G1 subunit is indispensable for activity. These results indicate that AM0627 has a preference for glycopeptide substrates with bis-T O-glycans compared to bis-Tn O-glycans, and only cleaved glycopeptides with one O-glycan when the O-glycan was positioned in the G1 subunit and not G1'. Moreover, the GalNAc residue at the G1 subunit is essential for cleavage since substituting the GalNAc at G1 with GalN abrogated cleavage (P7), while substitution at G1' only reduced cleavage (P8).

To rationalize the above findings at the molecular level, we performed molecular dynamics simulations of enzyme complexes with P1, P2, and P9 (three independent simulations were performed, for a total of 1.5 μs for each complex). The systems were built from the structure of the AM0627-P2 complex, with P1-Zn^2+ complex as a template, upon reversing Ala326 to Glu326 (see the “Methods” section). The relative movement between the peptide and the enzyme was analyzed by monitoring selected intermolecular distances that are only formed when the peptide adopts reactive conformations. We selected two of the Zn^2+ coordination distances (His324-Nε2-Zn^2+ and His329-Nε2-Zn^2+) and the interaction distances between Arg291 and the side chain of the peptide Gln (Nε2NZ2-OQ3a and Ne2NZ2-Nε3a). The results showed that the naked peptide P9 is highly dynamic and unstable, not being able to keep all these interactions at any time during the simulations, and frequently adopting conformations in which the peptide is almost detached from the enzyme (Fig. 4d). In contrast, P2 and P1 were stable for significant periods of time, both keeping these relevant interactions during the simulations (Fig. 4c–f and Supplementary Figs. 5–8). These results agree with the experimental results reported above, showing that the enzyme preferentially recognizes P1 and P2 and easily binds P1 in a suitable configuration for the catalytic reaction to take place.

QM/MM metadynamics simulations of the reaction catalyzed by AM0627 suggest a water molecule acting as a nucleophile and a Glu residue as the catalytic base

To address the catalytic mechanism of the enzyme, we performed QM/MM metadynamics simulations of the AM0627–P1–Zn^2+ complex. Two distinct mechanisms have been proposed in the literature for Zn-metalloproteases, depending on whether the nucleophile residue (supposedly Glu326 in AM0627) directly attacks the carbonyl carbon atom of the scissile peptide bond (hereafter named as C*) or it does it indirectly, via a water molecule (Supplementary Fig. 9)31,32. The active site configuration of AM0627–P1–Zn^2+ obtained from QM/MM MD simulations show that the Zn^2+ cation keeps the usual tetrahedral coordination and one oxygen atom of the Glu326 carboxylate group remains close to the C* atom (∼4 Å) (Supplementary Fig. 10a). This is a possible reactive configuration for direct nucleophilic attack of Glu326 on the peptide carbonyl. However, QM/MM metadynamics simulations of the chemical reactions considering direct attack from Glu326, using a collective variable corresponding to the Glu326–O–C* distance, resulted in a high energy barrier and the formation of an unstable complex (Supplementary Fig. 10b–d). This suggests that the nucleophilic attack of Glu326 is not direct but probably mediated by a water molecule. In fact, a water molecule was found to fit perfectly in the active site (Fig. 5b and Supplementary Fig. 11), being stable all along the QM/MM MD simulations. Such putative catalytic water is coordinated with the Zn^2+ ion and forms a hydrogen bond with Glu326. The water molecule oxygen atom is at ∼3 Å from C*, being also well oriented for nucleophilic attack on the C* atom. To drive the reaction, we again used QM/MM metadynamics simulations of the nucleophilic attack, using the distance between the water oxygen and C* as a collective variable (Supplementary Fig. 11). The simulations led to the formation of a stable intermediate (INT, Fig. 5b) in which the C* atom is four-coordinated. The C*–N bond at INT is stretched with respect to its value in the Michaelis complex (from 1.39 Å at MC to 1.52 Å at INT) but not (yet) cleaved. During the reaction, a proton transfers from the attacking water molecule to Glu326, which thus acts as a general base. The free energy barrier of the reaction (12.1 kcal/mol, Fig. 5a) is
indicative of a feasible reaction, in line with values previously reported for other Zn-dependent proteases. A subsequent QM/MM metadynamics simulation was performed for the second reaction step of the enzymatic reaction, starting from the tetrahedral intermediate (INT), using the peptide bond distance (C-N) as a collective variable (Supplementary Fig. 11). The simulations show that peptide bond cleavage is consistent with the transfer of the C=O proton to the N atom. The whole process involves a much lower energy barrier (5.6 kcal/mol, Fig. 5a) than the first reaction step and leads to the complete cleavage of the C=N bond (Fig. 3b). Therefore, AM0627 can effectively hydrolyze the TS-T6 peptide bond of P1 in a two-step reaction, with the formation of a tetrahedral intermediate, via a nucleophlic attack by a water molecule and the assistance of Glu326 as a general base.

Structural analysis of AM0627 and other mucinases can predict those displaying bis-O-glycan preferences

To infer why AM0627 recognizes clustered O-glycans and in turn cleaves glycopeptides containing bis-O-glycans, we analyzed the active sites of previously reported mucinases and the O-glycoprotease OgpA compared to that of AM0627. The structure of BT-4244 (PDB entry: 5KDB, ZmpB (PDB entries: 5KDU/6XT1), IMPa2 (PDB entry: 5KDX), and OgpA (PDB entry: 6Z2P) were previously solved in complex with Tn, mSTb, and T O-glycans, and T glycopeptide (i.e., a peptide containing a T O-glycan), respectively (Fig. 6a). Inspection of their active site reveals that the residues that are around the GalNAc moiety at the G1 subunit (Trp321/Asn347/Arg362/Tyr288AM0627, Trp570/Asn593/Arg611/Trp538BT4244, Trp572/Asn775/Arg790/Ph272ZmpB, Trp692/Asn720/Arg742/Ph272IMPa, and Trp747/Asn770/Arg785/Ph272ZmpC) and even their interactions with the GalNAc moiety are mostly conserved (Fig. 6a). Exceptions occur with Ph272ZmpB/Ph272ZmpC that establish CH–π interactions with the Sia moiety of mSTb, and Trp685IMPa that does not interact with any sugar moiety. Apart from these interactions at the G subunits, other interactions have been described thoroughly before. Regarding the degree of conservation of the residues at the G subunits, only BT4244 shares the key tyrosine residue that was found to be crucial in AM0627 (Tyr470AM0627 and Tyr723BT4244). We showed above that Tyr470AM0627 interacts with both GalNAc and Gal located at the G1 and G2 subunits, respectively, and its mutation to Ala abolishes completely the AM0627 activity (Fig. 3b). Therefore, there are important structural similarities among these mucinases, indicating that some of them could behave similarly regarding O-glycan recognition.

To get more insights into the potential function of Tyr723 in BT4244, we superimposed the AM0627 crystal structure with those of BT4244, ZmpB, IMPa, and ZmpC (Fig. 6b; in this figure, only the P1 from the AM0627(329–424)–P1–Zn2+ complex is shown for illustration purposes). As expected, the lowest root-mean-square deviation (RMSD) and the greater number of aligned residues were found between AM0627 and BT4244 (2.02 Å and 394 aligned residues, respectively) followed by ZmpB/ZmpC and AM0627 (2.25/2.26 Å and 331/342 aligned residues), and IMPa with AM0627 (2.99 Å and 342 aligned residues). A closer inspection of the active site of the complex between BT4244 and the Tn O-glycan (Trp632AM0627, Trp752AM0627, Tyr116OgpA, Tyr288AM0627, Tyr538BT4244, Phe727ZmpB, Trp685IMPa, Phe722ZmpC), this residue in OgpA (Tyr116) prefers to interact with the Gal at the G2 subunit, while residues such as Phe727ZmpB/Phe722ZmpC–glycans. AM0627 was previously suggested to cleave a peptide bond in a bis-O-glycan patch. At the level of the recognition of the substrate peptide sequence by OgpA, it is important to highlight that OgpA mostly recognizes the amino acid backbones of the T glycopeptide, except for a hydrogen bond interaction between Arg6 side chain and Leu213 backbone (Fig. 6a), suggesting that OgpA, as a typical O-glycoprotease, might be less specific for the peptide sequences of glycoproteins substrates than AM0627. Yet, AM0627 is promiscuous towards different glycoprotein substrates, as shown very recently.

BT4244 preferably acts on glycopeptides with bis-Tn and bis-T O-glycans present in MUC1

To get more insights into the activity of AM0627 and BT4244 on bis-O-glycans, we compared the activity of BT4244 and wt AM0627 towards different O-glycan forms using a recombinant mucin reporter O-glycoprotein substrate, which contains 6.5 TRs of the 20 amino acid human MUC1 TR sequence (GVTSAVPDRAPCGSTPAHH) with five O-glycosites. Note that the MUC1 TR contains two bis-O-glycosites and one isolated glycosite. We previously showed that all five O-glycosites are fully O-glycosylated when expressed in glycoengineered HEK293 cells and that BT4244 efficiently cleaves the Tn glycoform of this MUC1 reporter with predominant cleavage in between the bis-O-glycan at the VTSa and GSTA motifs. We tested the wt glycoform of the MUC1 TR reporter (containing a mixture of mono and disialylated core 1 and 2 structures), and three engineered more homogeneous glycoforms with mSTa, T, and Tn O-glycans. AM0627 was previously suggested to cleave bis-mSTa O-glycans. However, here we found that neither AM0627 nor BT4244 efficiently cleaved the sialylated MUC1 reporters (wt MUC1 and mSTa-MUC1), although slight degradation of the mSTa glycoform by AM0627 at the highest concentration (1:5 enzyme/substrate ratio) was apparent. In contrast, both mucinases efficiently cleaved the T and Tn-MUC1 reporters (cleavage observed from 1:100 enzyme/substrate ratio) (Fig. 7), which is in agreement with the cleavage studies using the P glycopeptides (Fig. 4b). AM0627 exhibited a preference for the T glycoform compared to Tn, while BT4244 revealed the inverse pattern with a preference for the Tn glycoform. To rationalize this data, we performed MD simulations on the complex of BT4244 with a MUC1 glycopeptide (AHGVTSA) containing a bis-T and a bis-Tn O-glycan. The results show that the hydrogen bonds established between BT4244 and the bis-Tn GalNAc at the

Article

https://doi.org/10.1038/s41467-022-32021-9
Fig. 6 | Analysis of BT4244, ZmpB, IMPa, ZmpC, OgpA, and AM0627 active sites.

a Close-up view of the active sites of BT4244-Tn O-glycan, ZmpB, mSTb, IMPa-T O-glycans, ZmpC-mSTb, and OgpA-T glycopeptide complexes. Note that the zinc ion is not present in the structure of OgpA in complex with the T-glycopeptide because an inactive version of OgpA was used in which the His205, a residue that coordinates the Zn²⁺, and the catalytic Glu206 was mutated to Ala20. The ligands and the Zn²⁺ in these structures are shown as olive carbon atoms and as a pink sphere, respectively. The amino acids of the enzymes are shown as cyan carbon atoms. The dotted lines are displayed as brown and yellow for the hydrogen bonds and the interactions between the metal and the amino acids, respectively.

b Active sites of BT4244-Tn O-glycan, ZmpB, mSTb, IMPa-T O-glycan, ZmpC-mSTb, and OgpA-T glycopeptide complexes showing the superimposed P1 from AM0627-E326A–P1–Zn²⁺ complex. P1 is colored as in Figs. 2 and 3. The red arrow indicates the presence of the conserved Tyr723BT4244 with Tyr470AM0627, and the red circle indicates that the Tyr is not present in ZmpB/ZmpC and that a steric clash might occur between Thr775/Tyr776IMPa and the sugar units at G1 and G2.
the GI subunit are more stable during the simulation than the hydrogen bonds between BT4244 and the bis-T GalNAc at the GI subunit, which might explain why BT4244 slightly prefers to act on glycopeptides containing bis-Tn over bis-T patches (Supplementary Fig. 12).

To further explain why both mucinases did not act on glycopeptides containing sialic acids (bis-mSTa), we inspected the active site of the structure of AM0627 in complex with P1. Our analysis reveals that the Sia bound to the OH3 of Gal at the G2 subunit would likely clash with residues such as Lys387, Asp388, and Val389. Note that Val389 is at 3.38 Å from the Gal OH3 (Supplementary Fig. 13a). Regarding the other Gal at the G2 subunit, it is likely that the Sia moiety would be repulsed by the negatively AM0627 surface close to the Gal moiety (Supplementary Fig. 13b). Finally, we also addressed why the STn glycoform blocks peptide cleavage, as reported for the StcE mucinase5 and other mucinases19. A closer look at the AM0627-active site reveals that α2-6-linked Sia residues on the GalNAc moieties will likely clash with Trp149, Tyr288, and Asp318. Note that the OH6 of GalNAc at the G1 subunit is at 5.25 Å of Trp149, and the OH6 of GalNAc at the GI subunit is at 3.38 Å of Tyr288, and 2.47 Å of Asp318 (Supplementary Fig. 13a). Overall, our experimental results clearly demonstrate that BT4244 has the same preferences for bis-O-glycans as those of AM0627, although with somewhat different cleavage rates. In addition, our MD simulations suggest that steric factors are the reason for the observation that these two mucinases do not efficiently act on glycopeptides with O-glycans capped by Sia as demonstrated in our experiments with the MUC1 mucin reporters (Fig. 7). Recent studies have suggested that AM0627 can cleave native MUC2 isolated from
Caco-2 cells although with increased efficiency after desialylation and analysis of select cleaved fragments have shown mST O-glycans at the cleavage site. However, these studies were performed with extremely high enzyme-to-substrate ratio and extended digestion times (24 h). In addition, native isolated mucins are highly heterogeneous substrates for which it is not possible to efficiently monitor the degradation (e.g. cleavage at one or few sites versus cleavage at all potential bis-O-glycan sites). The present work with well-defined O-glycosylated substrates and normalized enzyme to substrate ratios clearly support the structural data that sialic acids and more elaborate O-glycan structures are much poorer substrates.

Discussion

Degradation of mucins is important for the normal process of renewal and clearance of mucins in the mucus layers lining mucosal surfaces and essential for pathogenic bacteria like enterohemorrhagic Escherichia coli (EHEC) relying on the penetration of the protective mucus layers to reach the underlying epithelium. While several mucinases have been reported and characterized so far, only two of these, StcC and ZmpC, derived from pathogenic bacteria, appear to be able to cleave mucins covered by complex elaborated O-glycans mainly produced by the microbiota as part of the process of degrading mucins for nutrient sourcing. In contrast, mucinases such as AM0627 and BT4244 that only cleave mucins with short truncated O-glycans are likely dependent on prior trimming of O-glycans down to Tn or T by bacterial glycosidases produced by commensal and symbiotic bacteria. Although further insights are clearly needed, we envision that mucinases like AM0627 and BT4244 may serve in the last steps of degradation of the mucus layer during its continuous renewal to ensure homeostasis of the process. Akkermansia muciniphila is a gut commensal that relies on mucins as the main source of carbon and nitrogen, and this bacteria highly expresses AM0627 as well as multiple glycosidases and putative sulfatases in response to mucins.

Mucinases and O-glycoproteases characterized thus far appear to rely on a single O-glycan adjacent to the peptide cleavage site, and here we unveiled molecular mechanisms by which the mucinases AM0627 and BT4244 instead exhibit preference for cleavage in between two or more O-glycans. Clustered O-glycan patches are particularly found in mucins and mucin-like domains, with MUC2 and MUC5AC being among the mucins with the highest density of O-glycans including stretches of 3–6 adjacent O-glycans. While the sequence and spacing of O-glycans in mucins may not be conserved, the density of O-glycans does appear to be conserved. On the contrary, the distribution of bis-O-glycans is found among most mucin TRs and many O-glycoproteins. Our structure of an inactive form of AM0627 with P1 shed light into how this mucinase recognizes bis-O-glycans and achieves catalysis. AM0627 clearly prefers substrates containing bis-T O-glycans and to lesser degree bis-Tn with a critical need for a GalNAc moiety at the G1 subunit for catalysis. Based on the AM0627 structure and its comparison with other mucinases and O-glycoproteases, we also inferred that a Tyr residue (Tyr470) is key to explain the recognition of bis-O-glycans, a feature shared in the BT4244 mucinase with similar substrate preference. The recently reported crystal structure of the Zn²⁺-bound AM0627 did not reveal the molecular basis for the bis-O-glycan substrate preference, likely due to misplacement of the mono-T- O-glycan at S1 with sugar moieties located at G1 and G2 subunits. This study also concluded that the O-glycans were in close contact with Trp149 and Phe290 residues, which clearly is not the case in our crystal structure. Misplacement of the docked glycopeptide substrate may also have been the reason for missing the key function of Tyr470 in recognition of the mono-T O-glycan at S1 in driving the preference for bis-O-glycan substrate sites. Interestingly, this recent study included activity analyses of enzyme mutants (mutations at residues Trp149, Tyr287, and Phe290) using different glycoprotein substrates that showed the importance of these residues in the recognition of the peptide backbone of glycoprotein substrates, which further support our interpretation of the AM0627 crystal structure presented here. Overall, this demonstrates the importance of obtaining experimental structures of protein–ligand complexes for reliable information on protein recognition and mechanisms that may be further complemented with computational studies.

The finding that the AM0627 and BT4244 mucinases does not use MUC1 TR reporter substrates with O-glycans capped by Sia is supported by our structural analysis. Shon et al. previously inferred that AM0627 cleaved glycoproteins with Sia moieties. However, this interpretation was based on enzyme assays with highly heterogeneous glycoprotein substrates and excessive enzyme ratios and incubation times. The study presented here, using mucin reporter substrates with defined O-glycan structures, clearly confirms that AM0627 and BT4244 have strong specificities for bis-O-glycans with either T or Tn glycoform preferences, and are expected to cleave more widely when substrates are presented with unsialylated T and Tn O-glycans.

Zinc metalloproteases follow two distinct mechanisms for the cleavage of peptide sequences depending on whether the nucleophile is a water molecule or a dioxirane species assisted by our QM/MM dynamics simulations, we found that AM0627, and likely other mucinases, follow a two-step mechanism in which a nucleophilic water molecule is firstly activated by a Glu residue acting as the catalytic base (Glu326 in AM0627). The first and most important step of catalysis leads to the formation of an intermediate with tetrahedral coordination of the carbon atom of the T5–T6 peptide bond, which is effectively cleaved in the second catalytic step, assisted by the transfer of a proton from T5 to T6. A water molecule that establishes a hydrogen bond with the general base residue (Glu326) is well poised for nucleophilic attack on the carbonyl group of the scissile peptide bond during the first reaction step. Interestingly, a water molecule was also observed near the GalNAc acetamide NH at G1 subunit in the X-ray structure of AM0627 in complex with P1, thus we speculate that this water molecule is held by Glu326 in the complex of the WT enzyme and acts as the nucleophile. The kinetics data with the inactive P7, which contains a galactosamine at G1, thus lacking the acetamido group, suggested that this group is crucial for catalysis. It is possible that the acetamido group has a role in sequestering the catalytic water molecule through the NH substituent. In addition, the absence of the acetyl group in the galactosamine moiety might lead to the formation of a protonated amine group that might not be efficient in trapping the catalytic water molecule, and/or influence the position of the sugar moiety bound to AM0627. Both scenarios would certainly affect negatively the catalytic properties of AM0627.

In conclusion, we provide structural and mechanistic insights into a mucinase that recognizes bis-O-glycans, which have led to the identification of another mucinase, BT4244, with the same requirement for bis-O-glycans, albeit with a different preference for O-glycan structures. We identified a key conserved Tyr residue in AM0627 and BT4244 positioned close to the substrate G subunits responsible for the preference for bis-O-glycans. The expanding repertoire of mucinases provides new tools to break a barrier in studying mucins and O-glycoproteins with dense O-glycodomains that cannot be digested by traditional proteases. Deeper knowledge of the substrate specificities of these mucinases, both with respect to peptide backbone and O-glycan positions and structures, will aid in the design of digestion strategies for select glycoprotein substrates.

Methods

Protein expression and purification

The DNA sequence encoding amino acid residues 21–506 of the AM0627 (Amuc_0627) was codon optimized and synthesized by
GenScript (USA) for expression in E. coli. At the 5′-end, the construct also contained a sequence encoding a 12xHis tag and a Tobocato Ethic Virus (TEV) cleavage site. The DNA, containing at the 5′-end a recognition sequence for EcoRI, and at the 3′-end a stop codon and a recognition sequence for SalI, was cloned into pMALC2x, rendering the vector pMALC2x-12Hist-TEV-AM0627\(^{71,70}\). In the vector, the TEV cleavage site is located between the maltose binding protein (MBP)-12Hist and the protein of interest. All mutants in AM0627 were generated following a standard site-directed mutagenesis protocol by GenScript (quick change) using the vector pMALC2x-12Hist-TEV-AM0627\(^{71,70}\). The vector pMALC2x-12Hist-TEV-AM0627\(^{71,70}\)-E326A and pMALC2x-12Hist-TEV-AM0627\(^{71,70}\) were also generated by GenScript and by using as a template the AM0627\(^{71,70}\)-E326A construct from pMALC2x-12Hist-TEV-AM0627\(^{71,70}\)-E326A and the AM0627\(^{71,70}\)-E326A from pMALC2x-12Hist-TEV-AM0627\(^{71,70}\)-E326A, respectively.

Each plasmid was transformed into E. coli BL21(DE3) and grown in 2XTY medium (1.6% (v/v) tryptone, 1% (w/v) yeast extract powder and 0.5% (w/v) NaCl), containing 100 μg/ml of ampicillin at 37 °C. When the O.D at 600 nm reached 0.6 to 0.8, the culture was induced with 1 mM isopropyl 1-thio-ß-D-galactopyranoside (IPTG) at 18 °C. After 16 h incubation, the cells were harvested by centrifugation at 17,000 × g at 4 °C for 10 min. Cells were lysed using buffer A (25 mM Tris pH 8, 500 mM NaCl, 10 mM Imidazole) and loaded into a HisTrap Column (GE Healthcare). Proteins were eluted with an imidazole gradient from 10 mM up to 500 mM and then the buffer was exchanged to buffer B (25 mM Tris pH 8, 150 mM NaCl) using a HiPrep 26/10 Desalting Column (GE Healthcare). Thereafter, the TEV recognition site was cleaved using TEV protease.

TEV protease and MBP-12Hist were later removed from the solution using a His-Trap Column (GE Healthcare), and isolated proteins were then loaded into a HiLoad 26/60 Superdex 75 Column (GE Healthcare), previously equilibrated with buffer B. The proteins were concentrated using Amicon Ultra-15 mL and quantification was carried out by absorbance at 280 nm using their theoretical extinction coefficient (\(ε_{280\text{nm}}\)) for the wt AM0627\(^{71,70}\) and mutants \(-78370\) to \(-78980\) M\(^{-1}\) cm\(^{-1}\), and \(ε_{280\text{nm}}\) wt AM0627\(^{71,70}\) and AM0627\(^{71,70}\)-E326A, mutant \(-77240\) M\(^{-1}\) cm\(^{-1}\).

Recombinant BT4244 enzyme was produced in E. coli and purified as reported previously\(^{35}\). Briefly, the recombinant BT4244 (residues 35–557) was gene synthesized with a codon-optimized sequence (Twist bioscience, USA) and cloned in a PET28-based vector. The plasmid was transformed in T7 Express (NEB) bacterial strains, grown at 37 °C for 2 h, induced with 1 mM IPTG, and cultured at 16 °C overnight. Proteins were purified by nickel-nitrilotriacetic acid (Ni-NTA) chromatography and followed by gel-filtration chromatography with a Superdex 200 16/60 column. The fraction containing the enzyme were pooled and dialyzed in PBS.

Crystalization and data collection

AM0627\(^{71,70}\)-E326A was concentrated to 15 mg/ml and co-crystallized with 0.5 mM ZnCl\(_2\) and 5 mM PI. Appropriate size of crystals appeared at 37 °C for 10 min. Cells were harvested by centrifugation at 17,700 × g at 4 °C. The crystals were harvested by centrifugation at 17,700 × g at 4 °C. The crystals were then washed with two zinc cation and to reveal any association with the carbohydrate molecules. The LEaP module of Amber20 was used to generate the topology and coordinate files for the classical MD simulations, which were carried out using the CUDA version of the PMemd module\(^{21}\) of the AMBER20 simulation package.

Molecular dynamics (MD) simulations in explicit water

We used our crystal structure of AM0627\(^{71,70}\) and that of BT4244 (PDB entry: 5KD8\(^{29}\)) as the starting structures for all simulations reported in this work. The mutation of the catalytic residue (E326A) was mutared back to Chimeran \(^{34}\). The mutation states of His were chosen based on the hydrogen bond network and metal coordination manually check with Chimeran. The Zn\(^{2+}\)-coordinating His325 and His329 residues were inserted as His, protons while the rest histidine residues were inserted as His, protonated. The simulations were performed at pH 7, thus all Asp and Glu residues were negatively charged while all Arg and Lys residues were positively charged. The complex system was placed in the center of a cubic box 98 × 98 × 98 Å\(^3\) with a distance of at least 10 Å between the surface of the solute and the edge of the box. The box was then solvated with TIP3P water molecules, and counterions were added to neutralize the system. The protein was described using the Amber ff14SB force field\(^{30}\), while the GLYCAM06 force field\(^{31}\) was used to describe the carbohydrate molecules. The LEAP module of AMBER20 was used to generate the topology and coordinate files for the classical MD simulations, which were carried out using the CUDA version of the PMemd module\(^{21}\) of the AMBER20 simulation package. The solvated system was first subjected to 5000 steps steepest descent minimization, followed by 5000 steps conjugate gradient minimization with positional restraints on all heavy atoms of the solute, using a 50 kcal mol\(^{-1}\) Å\(^{-2}\) harmonic potential. The minimized system was then heated up to 300 K using the Berendsen thermostat, with a time constant of 1 ps for the coupling, and 50 kcal mol\(^{-1}\) Å\(^{-2}\) positional restraints applied over three 500 ps steps of the heating process. The positional restraints were then gradually decreased to 5 kcal mol\(^{-1}\) Å\(^{-2}\) over four 500 ps steps of NPT equilibration, using the Berendsen thermostat and barostat to keep the system at 300 K and 1 atm. For the production runs, each system was subjected to either 200 or 400 ns of sampling in an NPT ensemble at constant temperature (300 K) and constant pressure (1 atm), controlled by the Langevin thermostat, with a collision frequency of 2.0 ps\(^{-1}\), and the Berendsen barostat with a coupling constant of 1.0 ps. The SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms. A cut-off of 10 Å was applied to all nonbonded interactions, with the long-range electrostatic interactions being treated with the particle mesh Ewald (PME) approach. A time step of 2 fs was used for all the classical simulations, and coordinates were saved from the simulation every 10 ps. Three independent runs were performed.

QM/MM metadynamics

One representative snapshot extracted from the classical MD trajectory was used for the subsequent QM/MM MD simulations, which combines Born–Oppenheimer MD simulation, based on density

Nature Communications | (2022) 13:4324

https://doi.org/10.1038/s41467-022-32021-9
functional theory (DFT), with force-field MD methodology. The QM region consists of the Zn$^{2+}$ ion and its coordinated residues (His325, His326, and Glu343) as well as parts of the substrate peptide (Thr5, Thr6, and Pro7) and the catalytic water molecule and Glu326, resulting in a total number of 76 QM atoms (including 7 capping hydrogens), as shown in Supplementary Fig. 1I. The dangling bonds between the QM and MM region were capped with hydrogen atoms. The QM region was enclosed in an isolated supercell of size 20.0 × 20.0 × 20.0 Å$^3$. All QM/MM MD and metadynamics simulations were performed using CP2K v7.1 interfaced with PLUMED v2.5\textsuperscript{23-25} combining the QM program QUICKSTEP and the MM driver FIST. In this code, a real space multigrid technique is used to compute the electrostatic coupling between the QM and MM\textsuperscript{44} region. The QM region was treated at the DFT (BLYP) level, employing the dual basis set of Gaussian and plane-waves (GPW) formalism, whereas the remaining part of the system was modeled at the classical level using the same parameters as in the classical MD simulations. The Gaussian triple-$\zeta$ valence polarized (TZV2P) basis set was used to expand the wave function, while the auxiliary plane-wave basis set with a density cut-off of 350 Ry and GTH pseudopotentials\textsuperscript{55} was utilized to converged the electron density. All QM/MM MD simulations were performed under the NVT ensemble using a coupling constant of 10 fs and an integration time step of 1.0 fs. First, the system was equilibrated without any constraint for 10.0 ps. Then, the metadynamics\textsuperscript{53} method was used to explore the free energy profile for each reaction step. The distance between water oxygen and carbonyl carbon of Thr5 (C$^\alpha$) was used as a collective variable (CV1) for the first reaction step, while the distance between carbonyl carbon of Thr5 and amide nitrogen of Thr6 was taken as a collective variable for the second reaction step (Supplementary Fig. 1I). The proton transfers happened spontaneously during the metadynamics simulations, so no CV was needed for activating them. The Gaussian height was set to 1.0 kcal/mol, which was reset to 0.1 kcal/mol when it was about to cross the transition state, and the time deposition interval between two consecutive Gaussians was set to 25 fs. Gaussian widths were tuned according to the oscillations of each collective variable (0.2 Å for both CV1 and CV2). Recrossing over the TS$^\ast$ was observed in the first reaction step, but not the second step. Here, the proton transferred from Thr5 to the amide group of Thr6 spontaneously but did not return to the intermediate state. For this reason, the free energy of the P state is likely to be overestimated, while the reaction mechanism is not affected. The MD trajectories obtained in the simulation were analyzed by VMD and PyMOL (PyMOL 2.4.2), and distance calculation and clustering analysis were done by CPTRAJ\textsuperscript{4} from Amber 20. Plots were made with Matplotlib, while figures of structures were rendered with Chimera\textsuperscript{4} and PyMOL, plots and figures were combined with Inkscape v0.92.3.  

**Solid-phase (glyco)peptide synthesis (SPPS)**

(Glyco)peptides were synthesized by stepwise microwave-assisted solid-phase synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). Fmoc-Thr[GalNAc(\alpha(\text{Ac})\text{-Gal})]-OH (2.0 equiv) or Fmoc-Thr[GalNAc(\alpha(\text{Ac})\text{-Gal})]-OH (2.0 equiv) was synthesized as described in the literature\textsuperscript{15} and manually coupled using HBTU [(2H-benzotriazol-1-yl)-1,1,3,3-tetramethyloxylum hexafluorophosphate], while all other Fmoc amino acids (5.0 equiv.) were automatically coupled using oxyma pure/DIC (N,N-diisopropylcarbodiimide). The O-acetyl groups of GalNAc moiety were removed in a mixture of NH$_2$NH$_2$/MeOH (7:3). In the case of peptides P7 and P8, the azido group was transformed into the corresponding amino group by standard Pd/C hydrogenation. (Glyco)peptides were then released from the resin, and all acid-sensitive side-chain protecting groups were simultaneously removed using TFA 95%, TIS (trisopropylsilane) 2.5% and H$_2$O 2.5%, followed by precipitation with cold diethyl ether. The crude products were purified by HPLC on a Phenomenex Luna C18(2) column (10 μm, 250 mm × 21.2 mm) and a dual absorbance detector, with a flow rate of 10 mL/min.

**Glycopeptide preparation**

All the glycopeptides used in this work were dissolved at 100 mM in buffer 25 mM Tris pH 7.5. The pH of each solution was measured with pH strips and when needed adjusted to pH 7–8 through the addition of 0.1–5 μL of 2 M NaOH.

**Synthesis of glycopeptides containing the T O-glycan**

The glycopeptide P1 was incubated overnight at 37 °C in 23 mM with 51 μM D. melanogaster CIGaTI\textsuperscript{16,17}, 200 μM MnCl$_2$ and 97 mM UDP-Gal in buffer B in a final volume of 180 μL. To generate P3 and P4, we used similar conditions to the described above but used the substrates P5 and P6, respectively. Purification of the new glycopeptides was performed as described above.

**In vitro enzyme cleavage of glycopeptides**

In vitro glycopeptide cleavage activity was measured by MALDI-TOF MS semi-quantitatively. The reaction of the wild-type (wt) AM0627\textsuperscript{42-45} and all mutants were performed by adding 0.4 μM of the wt AM0627\textsuperscript{44-46} and mutants with 57 μM of P1 and incubated at 37 °C in 50 mM ammonium bicarbonate buffer (pH 8.0). The experiments with P1–P9 were performed by adding 0.4 μM of the wt AM0627\textsuperscript{42-45} with 500 μM of the (glyco)peptides in 50 mM ammonium bicarbonate buffer and incubated at 37 °C. The reaction mixtures were taken at the indicated time points and product development was detected by MALDI-TOF MS.

**Proteolytic Cleavage Assay with MUC1-6.5xTRs reporter**

The design and construction of the MUC1-6.5xTRs reporters were previously reported. Glycoengineered HEK293 cell lines (HEK293 cells were originally purchased from GIBCO) with O-glycan designs for D. melanogaster (KO CIGALTI), core 1 (KO GCNT1, ST3GAL1/2, ST6GALNAC2/3/4), mono-sialylT (mSTa) (KO GCNT1, ST6GALNACT2/3/4) and wildtype HEK293\textsuperscript{41} were used for the stable expression of MUC1 TR reporter and are available as part of the cell-based glycan array resource\textsuperscript{50}. All isogenic HEK293 cells stably expressing MUC1-TR reporters were seeded at a density of 0.25 × 10$^5$ cells/ml and cultured for 5 days on an orbital shaker in F17 medium (Gibco) supplemented with 0.1 Kolliphor Pl88 (Sigma-Aldrich) and 2% Glutamax. Culture media were purified by Ni-NTA affinity (Qiagen) chromatography (pre-equilibration with 25 mM sodium phosphate, 0.5 M NaCl, 10 mM imidazole pH 7.4, and eluted with the addition of 200 mM imidazole). Purified reporters were further desalted followed by buffer exchange to MilliQ using Zea spin columns (Thermo Fisher Scientific) and quantified using a Pierce\textsuperscript{48} BCA Protein Assay Kit (Thermo Fisher Scientific). Proteolytic cleavage assays with purified glycoengineered MUC1 reporters (500 ng) were performed by incubating serial dilutions of the wt AM0627\textsuperscript{45} and mutants AM0627\textsuperscript{E326Q}, AM0627E326D, or BT4244 for 2 h at 37 °C in 50 mM ammonium bicarbonate buffer pH 8.0. Reactions were terminated by heat inactivation at 95 °C for 5 min. Samples were run on NuPAGE Novex gels (Bis–Tris 4–12%) at 200 V for 45 min followed by staining with Krypton Fluorescent Protein Stain (Thermo Fisher Scientific) and imaged with an ImageQuant LAS4000 system (GE Healthcare).

**Reporting summary**

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

**Data availability**

The crystal structure of the AM0627\textsuperscript{E326A}–PI–Zn$^{2+}$ complex was deposited at the RCSB PDB with accession code 7YX8. Previously published PDB structures used in this study are available under the accession codes 5KDX, 5KCU, 6XT1, 5DX2 and 7SCI. The
trajectory files of the classical MD simulation and QM/MM metadynamics simulations have been deposited to Zenodo at https://doi.org/10.5281/zenodo.6521230. Other data 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. de Las Rivas, M., Lira-Navarrete, E., Gerken, T. A. & Hurtado-Guerrero, R. Poly peptide GalNAc-Ts: from redundancy to specificity. Curr. Opin. Struct. Biol. 56, 87–96 (2019).
3. 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).
4. Shon, D. J., Kuo, A., Ferracane, M. J. & Malaker, S. A. Classification, structural biology, and applications of mucin domain-targeting proteases. Biochem. J. 478, 1585–1603 (2021).
5. Nason, R. et al. Display of the human mucinome with defined O-glycans by gene engineered cells. Nat. Commun. 12, 4070 (2021).
6. Nath, S. & Mukherjee, P. MUC1: a multifaceted oncoprotein with a key role in cancer progression. Trends Mol. Med. 20, 332–342 (2014).
7. Bafna, S., Kaur, S. & Batra, S. K. Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells. Oncogene 29, 2893–2904 (2010).
8. Ju, T., Otto, V. I. & Cummings, R. D. The 1n antigen-structural simplicity and biological complexity. Angew. Chem. Int. Ed. Engl. 50, 1770–1791 (2011).
9. Kudelka, M. R., Ju, T., Heimbarg-Molinaro, J. & Cummings, R. D. Simple sugars to complex disease—mucin-type O-glycans in cancer. Adv. Cancer Res. 126, 53–135 (2015).
10. Wandall, H. H., Nielsen, M. A. I., King-Smith, S., de Haan, N. & Bagdonait, I. Global functions of O-glycosylation; promises and challenges in O-glycobiology. FEBS J. 288, 7183–7212 (2021).
11. Radhakrishnan, P. et al. Immature truncated O-glycophenotype of cancer directly induces oncogenic features. Proc. Natl Acad. Sci. USA 111, E4066–E4075 (2014).
12. Hauselmann, I. & Borsig, L. Altered tumor-cell glycosylation promotes metastasis. Front. Oncol. 4, 28 (2014).
13. Kufe, D. W. Mucins in cancer: function, prognosis and therapy. Nat. Rev. Cancer 9, 874–885 (2009).
14. Belzer, C. Nutritional strategies for mucosal health: the interplay between microbes and mucin glycans. Trends Microbiol. 30, 3–21 (2021).
15. Hansson, G. C. Mucus and mucins in the intestinal and respiratory tracts. J. Intern. Med. 285, 479–490 (2019).
16. Paone, P. & Cani, P. D. Mucus barrier, mucins and gut microbiota: the expected slimy partners? Gut 69, 2232–2243 (2020).
17. Laville, E. et al. Investigating host microbiota relationships through functional metagenomics. Front. Microbiol. 10, 1286 (2019).
18. Bagdonait, I., Pallesen, E. M. H., Nielsen, M. I., Bennett, E. P. & Wandall, H. H. Mucin-type O-GalNAc glycosylation in health and disease. Adv. Exp. Med. Biol. 1325, 25–60 (2021).
19. Shon, D. J. et al. An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. Proc. Natl Acad. Sci. USA 117, 21295–21307 (2020).
20. Trastoy, B., Naegel, A., Anso, I., Sijogren, J. & Guerin, M. E. Structural basis of mammalian mucin processing by the human gut O-glycopeptidase OgpaA from Akkermansia muciniphila. Nat. Commun. 11, 4844 (2020).
21. Haurat, M. F. et al. The glycopeptidase CpaA secreted by medically relevant Acinetobacter species targets multiple O-linked host glycoproteins. mBio 11, e02033-20 (2020).
22. Pluvainage, B. et al. Architecturally complex O-glycopeptidases are customized for mucin recognition and hydrolysis. Proc. Natl Acad. Sci. USA 118, e2019220118 (2021).
23. Malaker, S. A. et al. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. Proc. Natl Acad. Sci. USA 116, 7278–7287 (2019).
24. Konstantinidi, A. et al. Exploring the glycosylation of mucins by use of O-glycoidomain reporters recombiantly expressed in glycoengineered HEK293 cells. J. Biol. Chem. 298, 101784 (2022).
25. Shon, D. J., Fernandez, D., Riley, N. M., Ferracane, M. J. & Bertozzi, C. R. Structure-guided mutagenesis of a mucin-selective metalloprotease from Akkermansia muciniphila alters substrate preferences. J. Biol. Chem. 298, 101917 (2022).
26. Uson, I. & Sheldrick, G. M. An introduction to experimental phasing of macromolecules illustrated by SHELX: new autotracing features. Acta Crystallogr. D Struct. Biol. 74, 106–116 (2018).
27. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).
28. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 714–797 (2007).
29. Noach, I. et al. Recognition of protein-linked glycans as a determinant of peptidase activity. Proc. Natl Acad. Sci. USA 114, E679–E688 (2017).
30. Rawlings, N. D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptides in the PANTHER database. Nucleic Acids Res. 46, D624–D632 (2018).
31. Lipscamb, W. N. Carboxypeptidase A mechanisms. Proc. Natl Acad. Sci. USA 77, 3875–3878 (1980).
32. Christianson, D. W. & Lipscamb, W. N. Carboxypeptidase A. Acc. Chem. Res. 22, 62–69 (1989).
33. Laitaoja, M., Valjakka, J. & Janis, J. Zinc coordination spheres in protein structures. Inorg. Chem. 52, 10983–10991 (2013).
34. Xu, D. & Guo, H. Quantum mechanical/molecular mechanical and density functional theory studies of a prototypical zinc peptidase (carboxypeptidase A) suggest a general acid-general base mechanism. J. Am. Chem. Soc. 131, 9780–9788 (2009).
35. Wu, S., Zhang, C., Xu, D. & Guo, H. Catalysis of carboxypeptidase A: promoted-water versus nucleophilic pathways. J. Phys. Chem. B 114, 9259–9267 (2010).
36. Blumberger, J., Lamoureux, G. & Klein, M. L. Peptide hydrolysis in thermolysin: ab initio QM/MM investigation of the Glu143-assisted water addition mechanism. J. Chem. Theory Comput. 3, 1837–1850 (2007).
37. Coelho, H. et al. Atomic and specificity details of mucin 1 O-glycosylation process by multiple poly peptide GalNAc-transferase isoforms unveiled by NMR and molecular modeling. JACS Au. 2, 631–645 (2022).
38. Chiavolini, D. et al. The three extra-cellular zinc metalloproteinases of Streptococcus pneumoniae have a different impact on virulence in mice. BMC Microbiol. 3, 14 (2003).
39. Grys, T. E., Siegel, M. B., Latham, 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).
40. Wang, B. X., Wu, C. M. & Ribbeck, K. Home, sweet home: how mucus accommodates our microbiota. FEBs J. 288, 1789–1799 (2021).
41. Wagner, C. E., Wheeler, K. M. & Ribbeck, K. Mucins and their role in shaping the functions of mucous barriers. Annu. Rev. Cell Dev. Biol. 34, 189–215 (2018).
42. Davey, L. et al. Mucin Foraging Enables Akkermansia muciniphila to Compete Against Other Microbes in the Gut and to Modulate Host Sterol Biosynthesis (Research Square, 2022).
43. Cerda-Costa, N. & Gomis-Ruth, F. X. Architecture and function of metallopeptidase catalytic domains. Protein Sci. **23**, 123–144 (2014).

44. Yang, J. Y. et al. Characterization of a new M13 metallopeptidase from deep-sea Shewanella sp. e525-6 and mechanistic insight into its catalysis. Front. Microbiol. **6**, 1498 (2015).

45. Kabach, W. Xds. Acta Crystallogr. D Biol. Crystallogr. **66**, 125–132 (2010).

46. Winn, M. D. et al. Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. **67**, 235–242 (2011).

47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60**, 2126–2132 (2004).

48. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. **25**, 1605–1612 (2004).

49. 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).

50. Kirschner, K. N. et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J. Comput. Chem. **29**, 622–655 (2008).

51. Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle mesh Ewald. J. Chem. Theory Comput. **9**, 3878–3888 (2013).

52. Ensing, B., Laio, A., Parrinello, M. & Klein, M. L. A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. J. Phys. Chem. B **109**, 6676–6687 (2005).

53. Laio, A. & Parrinello, M. Escaping free-energy minima. Proc. Natl Acad. Sci. USA **99**, 12562–12566 (2002).

54. Laino, T., Mohamed, F., Laio, A. & Parrinello, M. An efficient real space multigrid QM/MM electrostatic coupling. J. Chem. Theory Comput. **1**, 1176–1184 (2005).

55. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. Phys. Rev. B Condens. Matter **54**, 1703–1710 (1996).

56. Roe, D. R. & Cheatham, T. E. 3rd PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J. Chem. Theory Comput. **9**, 3084–3095 (2013).

57. Plattner, C., Hofener, M. & Sewald, N. One-pot azidochlorination of glycals. Org. Lett. **13**, 545–547 (2011).

58. Gonzalez-Ramirez, A. M. et al. Structural basis for the synthesis of the core 1 structure by C1GalT1. Nat. Commun. **13**, 2398 (2022).

59. Bull, C., Joshi, H. J., Clausen, H. & Narimatsu, Y. Cell-based glycan to new biological targets, the European Research Council (ERC-2020-SyG-951231 "CARBOCENTRE" to C.R.) and the Marie Skłodowska-Curie Innovative Training Networks (H2020-MSCA-ITN-2018-814102 “Sweet Crosstalk” to C.R.). V.T. thanks the Spanish Ministry of Science, Innovation and Universities for the FPI fellowship. Q.L. thanks AGAUR for the Beatriu de Pinos postdoctoral fellowship (No. 2019 BFP 00129). RJB thanks to FAPESP (2016/24191-8). Y.N. thanks the Mizutani Foundation for Glycoscience (grant 210066). The research leading to these results has also received funding from the FP7 (2007-2013) under BioStruct-X (grant agreement No. 283570 and BIOSTRUCTX_51856). F.C. thanks the Mizutani Foundation for Glycoscience (grant 220115) and the EU (Marie-Sklodowska Curie ITN, DIRNANO, grant agreement No. 958544).

**Author contributions**

R.H.-G. designed the crystallization construct and refined the crystal structure. R.H.-G., V.T., and R.J.B. solved the crystal structure. V.T. and A.G.-G. performed the expression and purification of AMO627 forms and the mutants. V.T. and A.G.-G. crystallized the complex. A.M.G.-R. performed the expression and purification of DmCIGaT1. Y.N. performed the activity assays with the glycopolymers and the MUC1-5xTRs reporters. F.C. and Q.L. performed the MD calculations. Q.L. also performed the QM/MM metadynamics experiments. I.C. synthesized the glycopolymers. R.H.-G. wrote the article with contributions mainly by C.R., F.C., Y.N., H.C., R.J.B., and Q.L. All authors read and approved the final manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-32021-9.

**Correspondence** and requests for materials should be addressed to Carme Rovira or Ramon Hurtado-Guerrero.

**Peer review information** Nature Communications thanks Elisa Fadda, and the other anonymous reviewer(s) for their contribution to the peer review of this work.

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

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022