





Structural basis of mammalian high-mannose *N*-glycan processing by human gut *Bacteroides*

Beatriz Trastoy ^{1,9}✉, Jonathan J. Du^{2,8,9}, Erik H. Klontz^{2,3,4}, Chao Li ⁵, Javier O. Cifuentes ¹, Lai-Xi Wang⁵, Eric J. Sundberg ^{2,3,6,8}✉ & Marcelo E. Guerin^{1,7}✉

The human gut microbiota plays a central role not only in regulating the metabolism of nutrients but also promoting immune homeostasis, immune responses and protection against pathogen colonization. The genome of the Gram-negative symbiont *Bacteroides thetaiotaomicron*, a dominant member of the human intestinal microbiota, encodes polysaccharide utilization loci PULs, the apparatus required to orchestrate the degradation of a specific glycan. EndoBT-3987 is a key endo- β -*N*-acetylglucosaminidase (ENGase) that initiates the degradation/processing of mammalian high-mannose-type (HM-type) *N*-glycans in the intestine. Here, we provide structural snapshots of EndoBT-3987, including the unliganded form, the EndoBT-3987-Man₉GlcNAc₂Asn substrate complex, and two EndoBT-3987-Man₉GlcNAc and EndoBT-3987-Man₅GlcNAc product complexes. In combination with alanine scanning mutagenesis and activity measurements we unveil the molecular mechanism of HM-type recognition and specificity for EndoBT-3987 and an important group of the GH18 ENGases, including EndoH, an enzyme extensively used in biotechnology, and for which the mechanism of substrate recognition was largely unknown.

¹Structural Biology Unit, Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia Technology Park, Building 801A, 48160 Derio, Spain. ²Institute of Human Virology, University of Maryland School of Medicine, Baltimore, MD 21201, USA. ³Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201, USA. ⁴Program in Molecular Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201, USA. ⁵Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA. ⁶Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA. ⁷IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Spain. ⁸Present address: Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA. ⁹These authors contributed equally: Beatriz Trastoy, Jonathan J. Du. ✉email: btrastoy@cicbiogune.es; eric.sundberg@emory.edu; mrcguerin@cicbiogune.es

The composition and physiology of the gut microbiota plays a pivotal role in human health and disease^{1–4}. One major factor that influence the balance of bacterial species in the gut is the influx of glycans into the intestine, mostly from diet as well as host mucosal secretions and secreted epithelial cells^{5–8}. The processing of a broad diversity of glycans present in the human gut requires of several glycosidic linkage-specific degradative enzymes⁹. Human intestinal enzymes are capable of fully degrading a small set of glycans containing only one or two different sugar linkages⁹. Gut symbiotic microorganisms provide the complementary enzymatic machinery necessary to depolymerize glycans into their sugar components that otherwise cannot be processed by the host^{10,11}. The best studied strategy for glycan acquisition by human gut bacteria is one that is used by members of the phylum *Bacteroidetes*, which represents the main bacterial phylum of the human large bowel. The genomes of *Bacteroidetes* contain polysaccharide utilization loci (PULs)¹² that encode the apparatus required to utilize multi-subunit carbohydrates, with each PUL orchestrating the degradation of a specific glycan. The gene products of PULs have been termed Sus-like systems because they function by a similar mechanism as does the starch utilization system (Sus) but harbor enzymes that are predicted to target glycans other than starch¹³. The Sus-like systems are widespread among the *Bacteroidetes* members accounting, in many species, for up to one-fifth of their genomes to encode Sus-like pathways, as in *Bacteroides thetaiotaomicron*^{14–16}. Sus-like systems are mainly composed of (i) carbohydrate active enzymes, (ii) surface glycan binding proteins, and (iii) a porin that operate the degradation of a certain type of glycan (14). The Sus-like systems identified to date cover the recognition and specificity for all plant and animal tissue glycans that are expected to enter the human gut, as well as the breakdown of O-linked and N-linked glycans^{7,17–25}.

B. thetaiotaomicron encodes a high-mannose mammalian N-glycan (HMNG) depolymerizing system, comprised by four enzymes and two surface glycan binding proteins⁷. EndoBT-3987 is a secreted enzyme that initiates the HM degradation pathway and hydrolyzes the oligosaccharide from its polypeptide chain on the cell surface (Fig. 1)⁷. According to the currently accepted model, the released HM is held on the surface of *B. thetaiotaomicron* through the mannose-binding protein BT3986, while BT3984 recognizes the GlcNAc at the reducing end of the glycan and orientates the glycan into the outer membrane porin BT3983 for transport into the periplasm (Fig. 1)⁷. Three periplasmic α -mannosidases hydrolyze the oligosaccharide into the trisaccharide Man α 1–6Man β 1–4GlcNAc⁷. BT3990 and BT3991 hydrolyze α 1–2Man and α 1–3Man linkages, respectively, whereas the terminal undecorated α 1–6Man is hydrolyzed by BT3994, which requires GlcNAc at the reducing end for activity⁷. The *B. thetaiotaomicron* mutant lacking the extra-cellular σ factor regulator of the HM N-glycan PUL reduced the growth of the bacteria showing the importance of N-glycan metabolism in these species⁷.

EndoBT-3987 is an endo- β -N-acetylglucosaminidase (ENGase) that specifically catalyzes the hydrolysis of the β 1–4 linkage between the first two GlcNAc residues of the HM glycans^{7,26,27} but it is not able to hydrolyze complex-type (CT-type) N-glycans²⁷. ENGases are endoglycosidases that hydrolyze the chitobiose core of N-linked glycans (EC 3.2.1.96)^{28–30}. This enzyme class comprises glycoside hydrolase family 18 (GH18) and 85 (GH85) of the Carbohydrate-Active Enzymes Database (CAZy; www.cazy.org)^{31,32}. GH18 and GH85 family members have been extensively used in past decades to modify the N-linked glycan structures in proteins³³. GH18 ENGases display a wide range of protein specificities, except for EndoS and EndoS2 that are both specific for IgG antibodies^{34,35}, but more restricted glycan

specificity, and can be divided into three main subfamilies: (i) those that hydrolyze CT-type glycans (e.g., EndoF2, EndoF3, EndoS, EndoS2); (ii) those that hydrolyze HM and hybrid (Hy) N-glycans (e.g., EndoH, EndoF1, EndoS2); and (iii) those that hydrolyze HM and CT-type glycans (e.g., EndoS2, EndoBI-1). To our knowledge, EndoS2 is the only example that hydrolyzes the three major groups of N-glycans, CT-, HM-, and Hy N-glycans on IgG1 antibodies³⁶. Perhaps the ENGase most commonly used as enzymatic reagent in glycoprotein research is EndoH from *Streptomyces plicatus*³⁷. Since its characterization in 1974³⁸, this enzyme has been extensively used as a tool for glycan analysis³⁹, to monitor protein trafficking^{40–42} and deglycosylate heterogeneous glycoforms on glycoproteins for crystallographic purposes^{43–45}. EndoH has a similar glycan specificity to EndoBT-3987.

The molecular mechanism by which EndoBT-3987 or EndoH specifically recognize HM glycans has yet to be defined, even though several X-ray crystals structures of both enzymes in their unliganded forms have been determined in the past^{46,47}. Here, we provide the high-resolution X-ray crystal structures of EndoBT-3987 in its unliganded form, in complex with its substrate, Man₉GlcNAc₂Asn, and two of its products, Man₅GlcNAc and Man₉GlcNAc. In combination with alanine scanning mutagenesis and hydrolytic activity measurements of EndoBT-3987 point mutants, we present here the comprehensive structural basis of its catalytic and substrate recognition mechanisms, as well as those of other GH18 ENGases that specifically recognize HM-type glycans.

Results

Structure of EndoBT-3987 in complex with glycan substrate.

EndoBT-3987 has a predicted signal peptide (residues 1–25; SignalP-5.0) that was removed from the construct. In order to obtain the crystal structure of the enzyme-substrate complex we used a catalytically inactive version of EndoBT-3987, in which the residues Asp312 and Glu314 are mutated to alanine and leucine, respectively (EndoBT-3987_{D312A/E314L}; see below for further details). The crystal structure of EndoBT-3987_{D312A/E314L} in complex with the Man₉GlcNAc₂Asn substrate was solved by molecular replacement methods (EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn complex hereafter; pdb code 6TCV; Fig. 2a; Supplementary Figs. 1 and 2; Supplementary Table 1 and Methods section). This structure represents an example of an enzyme-substrate complex in the GH18 ENGase family. EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn crystallized in the R 3 space group with one molecule in the asymmetric unit and diffracted to a maximum resolution of 1.3 Å (Supplementary Table 1). Residues 26–41 were not visible in the structure. The full-length EndoBT-3987 comprises two domains from the N- to the C-terminus: (i) a β -sandwich domain (42–179) followed by a short linker (residues 180–193) that continues to (ii) a GH domain (residues 194–476).

The EndoBT-3987 GH domain adopts the conserved (α/β)₈-barrel topology typical of enzymes from the GH18 family, with a shallow pocket in which one molecule of the Man₉GlcNAc₂Asn substrate is unambiguously identified in the crystal structure (Fig. 2b, c). Specifically, the Man₉GlcNAc₂Asn substrate is located in the center of the (α/β)₈-barrel, flanked by the connecting loops β 10– α 2 (loop 1; residues 201–205), β 11– α 3 (loop 2; residues 228–245), β 14– α 4 (loop 3; residues 275–286), β 15– α 5 (loop 4; residues 313–330), β 16– β 17 (loop 5; residues 355–374), β 17– β 18 (loop 6; residues 379–395), β 18– α 6 (loop 7; residues 402–411), and β 19– α 7 (loop 8; residues 430–432) (Fig. 2a–c). The Asn residue is completely exposed to the solvent, accounting for the ability of EndoBT-3987 to process HM-type glycans attached to a

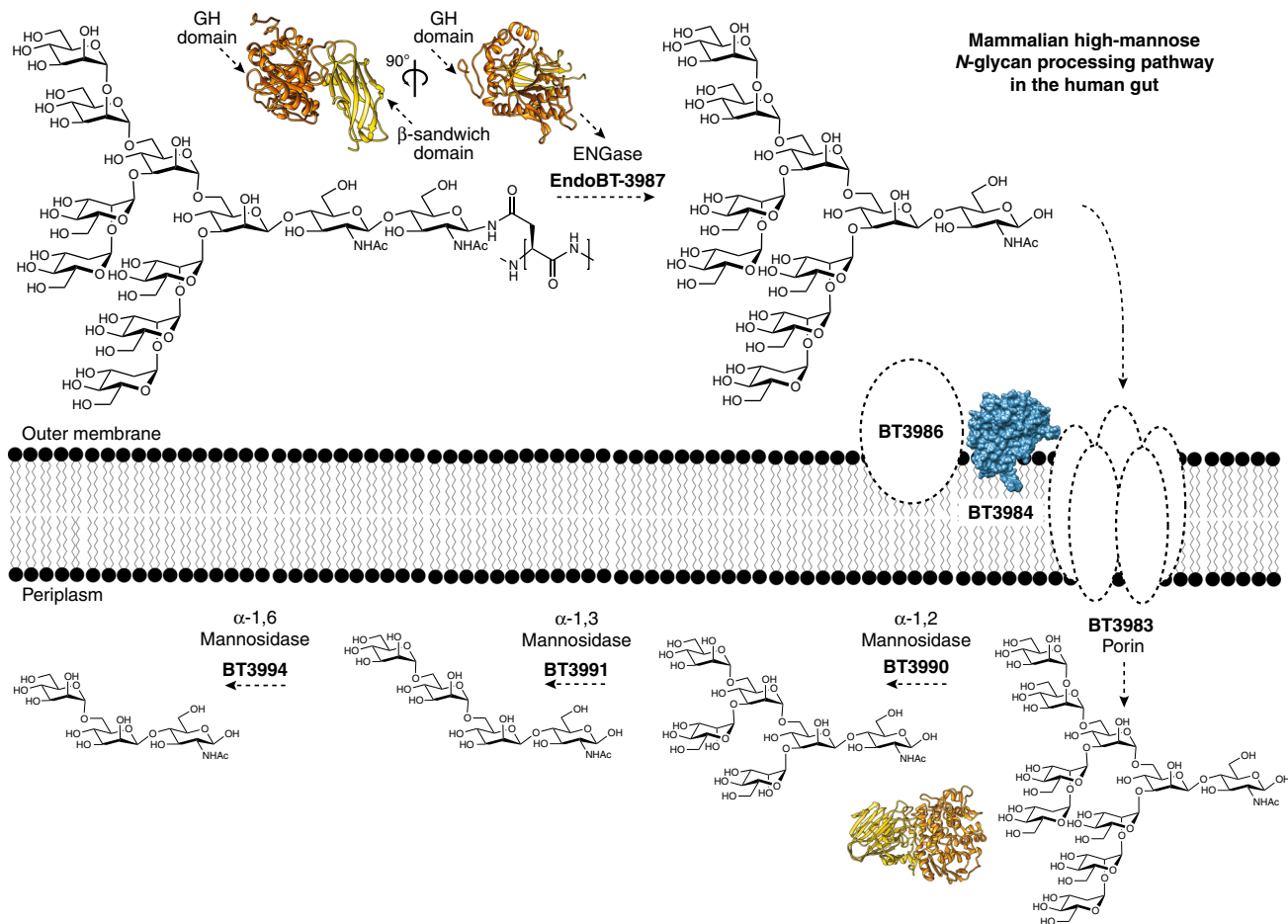


Fig. 1 Depolymerization of HM-type N-glycans by *B. thetaiotaomicron*. BT3987 deglycosylates HM-glycoproteins in the cell surface. The HM product interacts with the surface glycan binding protein (SGBP), BT3986, and the SusD homolog BT3984 interacts with the reducing end GlcNAc, orientating the glycan into the SusC homolog BT3983 porin. The HM glycan is transported to the periplasm where it is processed by three mannosidases to the trisaccharide Man- α 1,6-Man- β 1,4-GlcNAc. Cartoon representation of BT3987 (PDB CODE 6T8I) and BT3990 (PDB CODE 2WVZ), and surface representation of BT3984 (PDB CODE 3CGH; 7).

broad spectrum of proteins^{7,26,27}. The O3 and O6 atoms of the first GlcNAc (+1) residue make hydrogen bonds with the side chains of Y315 and N379, respectively. The GlcNAc (+1) residue is stabilized by an additional van der Waals interaction with the side chain of W355. The GlcNAc (-1) residue is in a skew boat conformation (1S_5), with the C2-acetamido group pointed toward the mutated catalytic residue D312A. The skew boat conformation of GlcNAc (-1) is stabilized by hydrogen bonds between the O3 atom and the side chain of Y315, the N2 atom of the acetamide group and the side chain of Y380, and the O6 atom and the side chain of E401. The hydrophobic residues F198, F227, F353, and F429 located at the top of the corresponding β 10, β 11, β 16, and β 19 β -strands of the barrel, respectively, form the floor of the binding pocket (Fig. 2d). The O₂ atom of the central Man (-2) residue makes a hydrogen bond with the side chain of E200. The Mana1-2Mana1-6 (Mana1-2Mana1-3)Mana1-6 antenna (antenna 1), mainly interacts with loops 1, 2, 3, and 4, whereas the Mana1-2Mana1-2Mana1-3 antenna (antenna 2) interacts with loops 1 and 7 (Fig. 2). The O₂ atom of Man (-3), which bisects antenna 1 into two additional sub-antennae, antenna α 1,6' (antenna 1a; Man (-4) and (-5) residues), and antenna α 1,3' (antenna 1b; Man (-6) and (-7) residues; Fig. 2), makes a hydrogen bond with the Ne₂ atom of H277. The O₄ atom makes a hydrogen bond with O₆ of Man (-7). The Man (-3) residue is stabilized by

an additional van der Waals interaction with the side chain of Y315. The Man residues of antenna 1a are solvent exposed, Man (-4) and Man (-5) make van der Waals interactions with H277. In contrast, the antenna 1b is buried in the substrate binding pocket (Fig. 2). The O₂ atom of the Man (-6) residue of antenna 1b makes electrostatic interactions with the side chains of N202 and N245; whereas the O₃ makes a hydrogen bond with the side chain of N230, located in β 12 of the β 12- β 13 hairpin, and also interacts with the side chain N245. The O₄ atom makes a hydrogen bond with the main chain of A228 and also interacts with the side chain of N230. Finally, the O₆ atom makes a hydrogen bond with the side chain of E200. The Man (-7) residue of the antenna 1b is mainly solvent exposed. The O₄ and O₆ atoms of the Man (-8) residue located in antenna 2, make hydrogen bond with the side chains of D203 and E200, respectively. In addition, the O₄ of Man (-9) makes electrostatic interaction with the side chain of N403, while Man (-10) is completely solvent exposed. For more information about the catalytic cycle of EndoBT-3987, please see Supplementary Note 1 and Supplementary Fig. 3.

Structure of EndoBT-3987 in complex with two glycan products. The strategy for capturing a native binary enzyme-product complex was to perform co-crystallization experiments with the

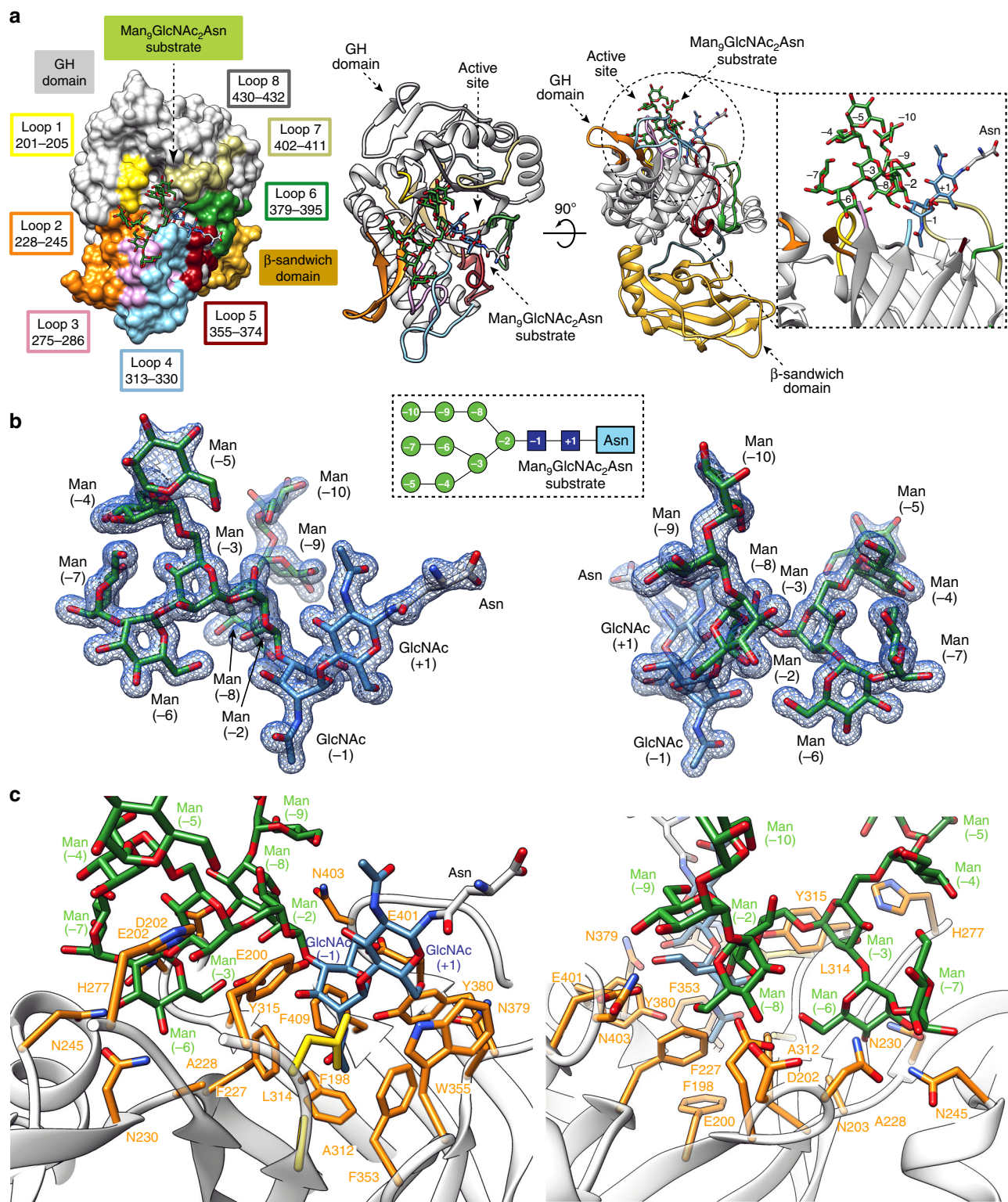


Fig. 2 The overall structure of EndoBT-3987 and the substrate $\text{Man}_9\text{GlcNAc}_2\text{Asn}$ glycan binding site. **a** Surface representation (left) with annotated domains and GH loops and cartoon representation (center) of two views of the $\text{EndoBT-3987}_{\text{D312A/E314L}}\text{-Man}_9\text{GlcNAc}_2\text{Asn}$ crystal structure. On the right panel zoom in of the carbohydrate-binding site of $\text{Man}_9\text{GlcNAc}_2\text{Asn}$. **b** Two views of the electron density of $\text{Man}_9\text{GlcNAc}_2\text{Asn}$ substrate shown at 1.0σ r. m. s. deviation. **c** Two views of the key residues of EndoBT-3987 interacting with $\text{Man}_9\text{GlcNAc}_2\text{Asn}$ substrate are colored in orange. The mutated catalytic residues D312A and E314A are colored in yellow.

full-length EndoBT-3987 ($\text{EndoBT-3987}_{\text{WT}}$) in the presence of $\text{Man}_5\text{GlcNAc}_2\text{Asn}$ or $\text{Man}_9\text{GlcNAc}_2\text{Asn}$ substrates. It worth noting that the enzyme was active against both HM *N*-glycan substrates. We thus obtained one snapshot of $\text{EndoBT-3987}_{\text{WT}}$

in complex with the $\text{Man}_5\text{GlcNAc}$ product ($\text{EndoBT-3987}_{\text{WT}}\text{-Man}_5\text{GlcNAc}$; pdb code 6TCW), and two snapshots of the $\text{EndoBT-3987}_{\text{WT}}$ in complex with the $\text{Man}_9\text{GlcNAc}$ product ($\text{EndoBT-3987}_{\text{WT}}\text{-Man}_9\text{GlcNAc-1}$, pdb code 6T8K; EndoBT-

EndoBT-3987_{WT}-Man₉GlcNAc-2 crystal form for our description since it displays the highest resolution. In contrast to what is observed in the EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn complex, the C2-acetamide group of the GlcNAc (−1) residue is oriented towards its anomeric carbon in the EndoBT-3987_{WT}-Man₉GlcNAc-2 crystal structure. The N2 and O7 atoms of the acetamide group make hydrogen bonds with the side chain of D312 and Y380 (Fig. 3d). The boat ^{1,4}B conformation of GlcNAc (−1) is further stabilized by hydrogen bond interactions between the O1 atom with the side chain of E314, the O3 atom with the side chain of Y315, and the O6 atom with the side chain of E401. The GlcNAc (−1) shows a ¹S₅ conformation in the EndoBT-3987_{WT}-Man₉GlcNAc crystal structure, but in this case a calcium cation is coordinated to O6 and O1 of GlcNAc (−1) (For more information about the catalytic cycle of EndoBT-3987, please see the Supplementary Note 1 and Supplementary Fig. 3).

Structural basis of EndoBT-3987 specificity for HM glycans. To further investigate how EndoBT-3987 specifically recognizes HM substrates at the molecular level, we performed single alanine mutations of residues in the loops that decorate the β-barrel core of the enzyme and contact the Man₉GlcNAc₂Asn substrate N-linked glycan. We studied their ability to process the N-linked HM glycans on two substrates, ribonuclease B (RNaseB) and HM-IgG. Specifically, we mutated key residues in loop 1 (E200, N202, D203, and N208), loop 2 (N230 and N245), loop 3 (H277A), loop 7 (Y403), and loop 8 (S432). As depicted in Fig. 4, H277A from loop 3 drastically reduced the hydrolytic activity against RNaseB and IgG of the enzyme. Furthermore, the N230A and N245A mutants located into the β-hairpin of loop 2, also reduced the activity of EndoBT-3987 on both substrates and E200A and N202A mutants located in loop 1 that interact with antenna 1b produced a smaller reduction of the hydrolytic activity of the enzyme. In contrast, mutations in loops 7 or 8 did not affect the activity of the enzyme against RNaseB nor IgG (Fig. 4). Collectively, the mutational analysis of the EndoBT-3987 loops that contact the HM glycan indicated that the interactions with the antennae 1b (loops 1, 2, and 3) were critical for glycan recognition, while those with antenna 2 (loops 1, 7, and 8) were effectively dispensable. Finally, we studied the influence of the β-sandwich domain into the hydrolytic activity of EndoBT-3987. Two deep cavities were identified at the interface between the β-sandwich and GH domains of the enzyme. We performed alanine mutations of residues in (i) cavity 1, located in β1 (Y49), α1 (Y95A, Y99A, and H103) and the α1-β5 loop (F107), and (ii) cavity 2, located in β3 (Y69). We also mutated Y439 to alanine in α7 of the GH domain, this residue points to the cavity one of the β-sandwich domain and could also be involved in carbohydrate binding. All mutant variants were enzymatically active as the wild-type enzyme, strongly suggesting that the two cavities are dispensable for the HM cleavage by EndoBT-3987.

Discussion

To further advance the understanding of EndoBT-3987 N-glycan specificity, we performed a structural analysis in the context of the GH18 family of ENGases. A search for structural homologs using the DALI server revealed that EndoBT-3987 shows significant structural similarity with EndoF1 from *E. meningoseptica* (pdb code 2EBN; Z-score of 37.5; r.m.s.d. value of 1.7 Å for 214 aligned residues; 42% identity)⁴⁸, and EndoH from *Streptomyces plicatus* (pdb code 1C8Y; Z-score of 30.9; r.m.s.d. value of 2.4 Å for 214 aligned residues; 29% identity)⁴⁹. EndoBT-3987, EndoF1, and EndoH are highly specific for HM and Hy-type N-glycans, but are able to process a broad range of proteins. In contrast to EndoBT-3987, both EndoF1 and EndoH do not display

an additional β-sandwich domain. A detailed comparison of the EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn and EndoBT-3987_{WT}-Man₉GlcNAc-2 binary complexes with those of the unliganded forms crystal structures of EndoF1 and EndoH reveals that the key residues that interact with the HM glycan substrate and product in EndoBT-3987 are well conserved in EndoF1 and EndoH, supporting a common recognition mechanism (Fig. 5a). EndoBT-3987 residues N230 and N245 (loop 2) and H277A (loop 3) lie in equivalent positions as 48, 63, and 95 in EndoF1, and 48, 61, and 95 in EndoH. Supporting this notion, the minimum N-glycan structure that EndoH and EndoF1 are able to hydrolyze is the core Manα1-3Manα1-6Manβ1-4GlcNAcβ1-4GlcNAc, including Man (−6), Man (−3), Man (−2), GlcNAc (−1), and GlcNAc (+1)⁵⁰. The O2 atom of Man (−3) makes a critical hydrogen bond with H277, whereas the O3 and O4 of Man (−6) makes important interactions with the side chains of N245 and N230, respectively. As a consequence, the Man (−6) residue is deeply buried and in close contact with the β-hairpin of loop 2.

The GH18 family of ENGases can be classified according to their N-glycan specificity into three main groups: (i) enzymes that hydrolyze CT-type N-glycans, (ii) enzymes that hydrolyze HM-type N-glycans, and (iii) enzymes that hydrolyze both CT- and HM-type N-glycans. Most members able to hydrolyze HM-type N-glycans also hydrolyze Hy-type N-glycans³⁰. The N-glycan specificity of the GH18 family members is due to unique interactions between the loops that decorate the β-barrel core of the enzymes and the N-glycan chemical structures^{27,51}. EndoBT-3987, EndoF1, and EndoH, which are classified in the first group, show the capacity to hydrolyze HM-type but not CT-type N-glycans²⁷. In contrast, EndoS and EndoF3 are able to hydrolyze CT-type but not HM-type N-glycans. Binding of EndoS to CT-type N-glycans is predominantly driven by GH domain loops that make contacts with the α(1,3) antenna, a mechanism that is conserved in EndoS2^{27,52}. In contrast, our experimental data clearly showed that the HM-hydrolyzing enzymes preferentially recognize the α(1,6) antenna 1b of HM-type N-glycans. In that sense, molecular docking calculations of a CT-type N-glycan into EndoBT-3987 revealed that the α(1,6) antenna makes clashes with loop 2 and 3 and cannot be accommodated into the binding grooves due to steric hindrance (Supplementary Fig. 3a). EndoS2 is able to hydrolyze CT- and HM-type N-glycans. The crystal structure of EndoS2 in complex with Man₉GlcNAc and alanine scan mutagenesis experiments revealed that EndoS2 interacts mostly with the antenna α(1,3) of HM glycans and not with the antenna α(1,6) as EndoBT-3987 does. Supporting this notion, the conformations of the antennae α(1,3) of the HM- and CT-type N-glycan in the binding site of EndoS2 are very similar between them, but different from the conformation of HM-type glycans found in the binding pocket of EndoBT-3987 (Fig. 5b).

Amino-acid sequence alignment with other GH18 ENGases that are able to hydrolyze HM-type glycans revealed that the important residues for substrate binding N230, N245 and H277 are conserved in EndoH^{37,53}, EndoF1⁵⁰, A6286²⁶, BT1285²⁶, EF2863⁵⁴, and Endo-Fsp⁵⁵, while Eng18B⁵⁶, Endo18A⁵⁶, EndoT⁵⁷, and EndoFv⁵⁸ have other residues at these positions, a histidine, phenylalanine and alanine instead of N230, N245, and H277, respectively (Fig. 5c). The superposition of the EndoT and EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn crystal structures (Supplementary Fig. 5b) shows some clashes of Man (−7) in antenna α(1,6) with loop 2. This suggests that the mechanism of HM-type glycan recognition and the glycan specificity is different than EndoBT-3987 in this subgroup of enzymes that are able to hydrolyze HM-type N-glycans vs CT-type N-glycans.

Amino-acid sequence alignment revealed that EndoBT-3987 shows high degree of sequence identity with other putative GH18 ENGases from four families of the order *Bacteroidales*, from the phylum *Bacteroidetes*: *Bacteroidaceae* (*Bacteroides*),

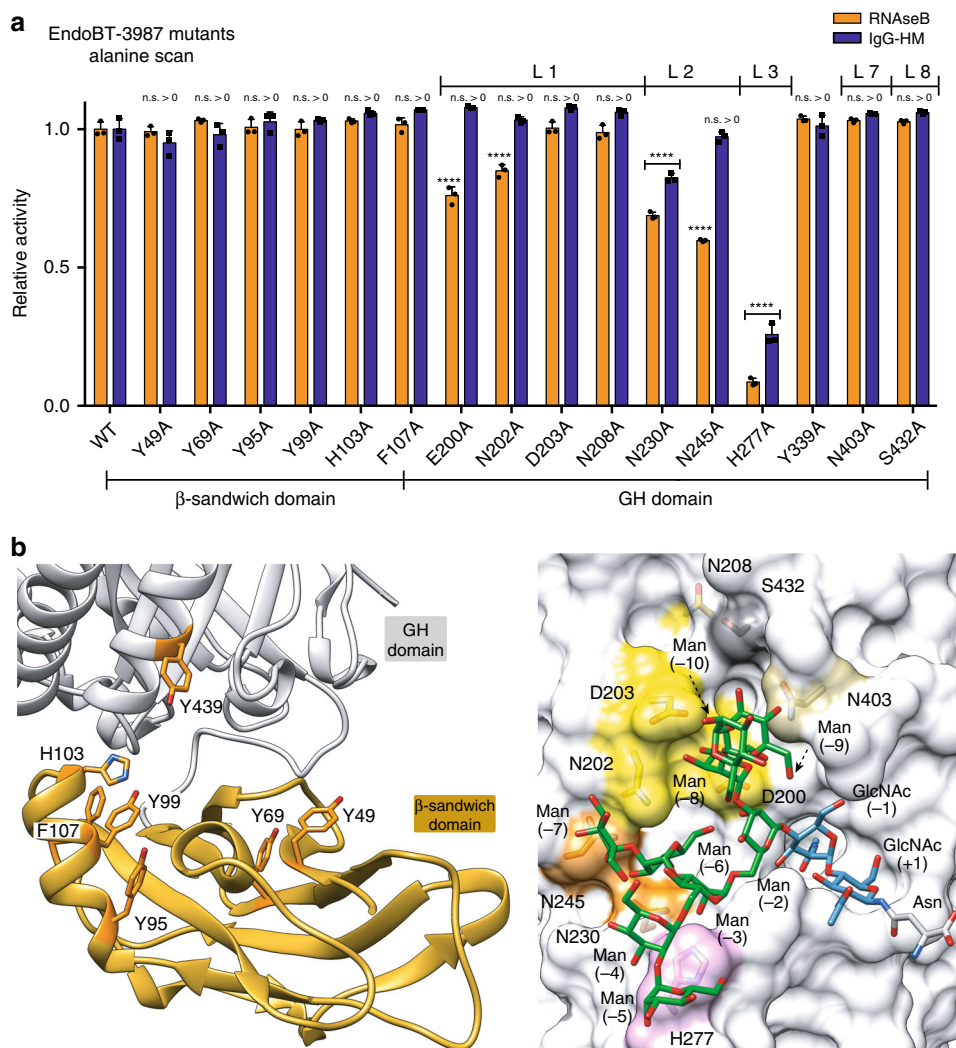
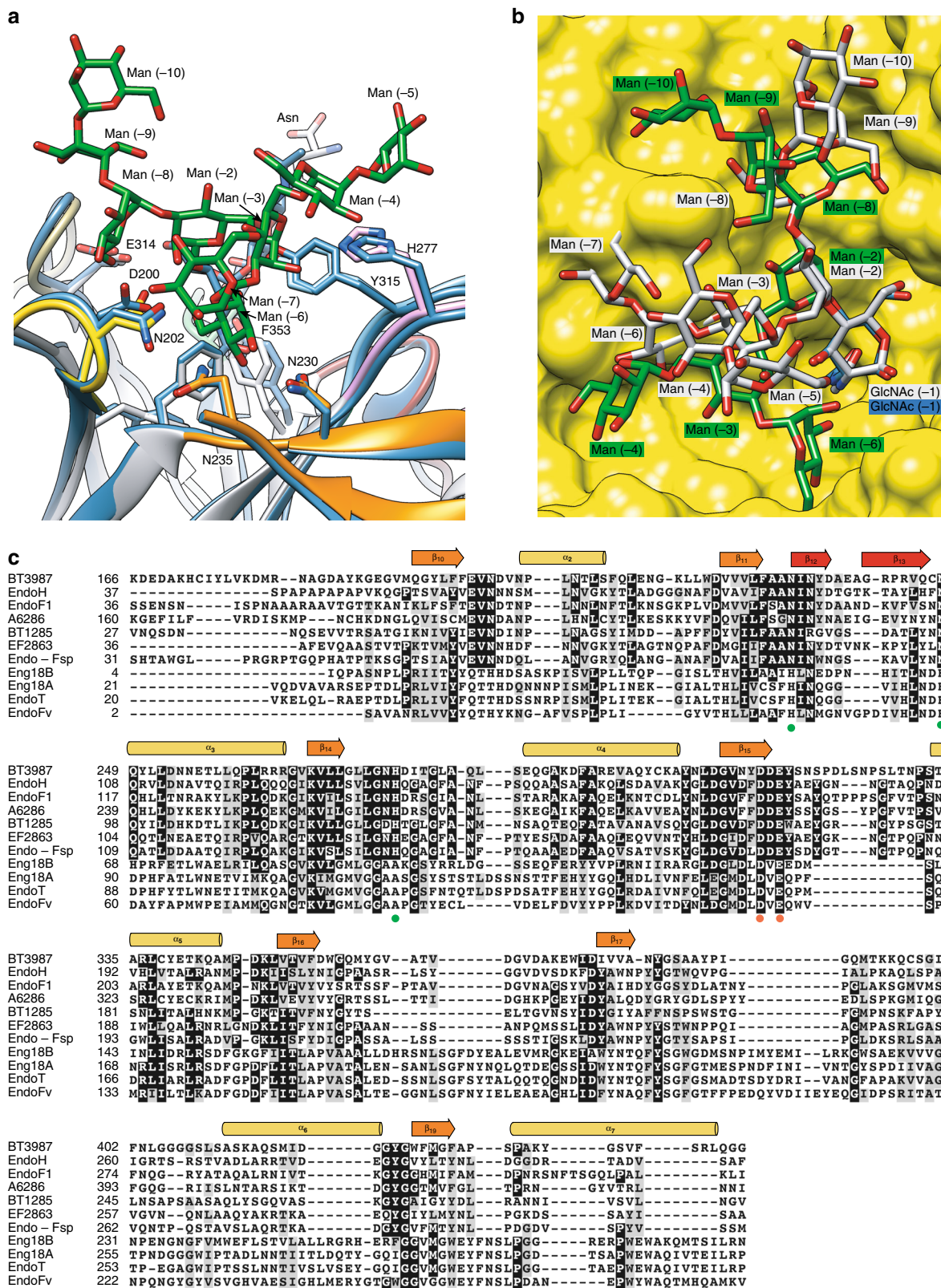


Fig. 4 Alanine scan mutagenesis of EndoBT-3987 active site for HM IgG1 and RNaseB. **a** Hydrolytic activity of EndoBT-3987 and mutants against HM-IgG1 and RNaseB is shown, as determined by LC-MS analysis, normalized to EndoBT-3987_{WT}. Statistical significance compared with wild-type EndoBT is annotated (multiple comparisons test, Tukey method; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, **** $p < 0.0001$, n.s. > 0, not significantly greater than no-enzyme control). Source data are provided as a 'Source_Data_File_Fig_4a'. **b** In the left panel cartoon representation of the β -sandwich domain. The mutated residues are highlighted in oranges. In the right panel surface representation of the EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn crystal structure showing the alanine mutations performed in loop 1 (yellow), loop 2 (orange), loop 3 (pink), loop 7 (brown), and loop 8 (gray) in the glycosidase domain (gray).

Dysgonamonadaceae (*Dysgonamodas*), *Prevotellaceae* (*Prevotella*), and *Rikenellaceae* (*Allistipes*) (Supplementary Fig. 6, Supplementary Data 1). The catalytic residues D312 and E314 and the residues important for the binding of HM-type N230, N245, and H277 in EndoBT-3987 are also conserved along the species, suggesting a similar mechanism of mammalian HM-type N-glycans degradation in this species. Interestingly, all of them have at least one additional N-terminal β -sandwich domain, attached to the GH18 ENGases domain, with conserved hydrophobic residues in all the β -strands. In some of the putative enzymes, the β -sandwich domain is classified as a BACON (*Bacteroidetes*-Associated Carbohydrate-binding often N-terminal) domain. It has been suggested that this domain binds carbohydrate-containing molecules most likely mucins⁵⁹. However, it has been shown that the β -sandwich domain of the GH BoGH5A from *Bacteroides ovatus*, also classified as a BACON domain, did not bind glycan or protein substrates²¹. Altogether, and taking into account our alanine scanning mutagenesis experiments where the activity of the enzyme was not affected by any of the mutants in the β -sandwich domain, the experimental data suggest that the

β -sandwich domain is not directly involved in substrate binding or enzymatic activity. Instead, the function of the β -sandwich domain might be to distance the GH domain from the cell surface and confer additional mobility to the catalytic domain to hydrolyze the N-glycan²¹.

In summary, we have determined the molecular mechanism by which EndoBT-3987 from *B. thetaiotaomicron*, the enzyme that initiates the degradation/processing of mammalian HM-type N-glycans in the intestine, specifically recognizes these glycans at the atomic level. *B. thetaiotaomicron* hydrolyzes HM-type N-glycans from carbohydrates in the diet but also processes HM-type N-glycans present in mucins, playing a role in mediating immune responses⁶⁰. We have presented a series of structural snapshots of the reaction center of the EndoBT-3987 from *B. thetaiotaomicron*. During this sequence of events, we visualize how the enzyme guides the substrate Man₉GlcNAc₂Asn into the reaction center where the hydrolytic reaction takes place, and unveil the mechanism of Man₉GlcNAc₂ product release. Based on these data, we have identified homologous enzymes in four families of the order *Bacteroidales*, strongly supporting the



notion of a common mechanism in the degradation of mammalian HM-type N-glycan among these species. Finally, this substrate recognition mechanism is shared by other important HM-modifying enzymes of the GH18 family, like EndoH or EndoF1, that have been extensively used in glycan analysis for more than 50 years, and for which such a mechanism of substrate

specificity was largely unknown. These findings will inform future efforts to engineer enzymes with customized glycan specificities.

Methods

Cloning, expression, and purification of EndoBT-3987. The pSpeedET vector encoding the EndoBT-3987 (*B. thetaiotaomicron* VPI-5482) gene was purchased from DNASU plasmid repository (<https://dnasu.org/DNASU/Home.do>). The

Fig. 5 Structural basis of EndoBT-3987 specificity for HM-type N-glycans. **a** Structural comparison of EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn in gray and EndoH (PDB CODE 2EBN) in blue. The residues involved in substrate recognition are colored by loops: loop 1 (yellow), loop 2 (orange), and loop 3 (pinked). **b** Structural comparison of the HM-type glycan conformation in the active site of EndoBT-3987_{WT}-Man₉GlcNAc-2 (gray) and EndoS2 (PDB CODE 6MDV) (green) on surface representation of binding site of EndoS2 (yellow). **c** Structure weighted sequence alignment of BT3987 with GH18 ENGases family with characterized endo-N-acetyl-β-D-glucosaminidase activity against HM-type N-glycans and inactive against CT-type N-glycans. Comparison of BT3987 from *B. thetaiotaomicron* VPI-5482 (Q8AON4, Uniprot code), EndoH from *Streptomyces plicatus* (P04067, Uniprot code), EndoF1 from *Elizabethkingia meningoseptica* (P36911, Uniprot code), A6286 from *Prevotella melaninogenica* (D9RSV7, Uniprot code), BT1285 from *B. thetaiotaomicron* VPI-5482 (Q8A889, Uniprot code), EF2863 from *Enterococcus faecalis* (Q830C5, Uniprot code), Endo-Fsp from *Flavobacterium* sp. (P80036, Uniprot code), Eng18B from *Hypocrea atroviride* IMI 206040 (G9P8KO, Uniprot code), Eng18A from *Hypocrea atroviride* IMI 206040 (G9NR36, Uniprot code), EndoT from *Hypocrea jecorina* (C4RA89, Uniprot code), and EndoFv from *Flammulina velutipes* (D1GA49, Uniprot code). The catalytic residues are marked with red dots and the key residues that interact with Man₉ are marked with green dots.

recombinant EndoBT-3987_{WT} protein is depicted in the Supplementary Fig. 2. Single and double-point mutations (Supplementary Table 2) were made by using the FastCloning method⁶¹. Full sequences were confirmed by Genewiz (<https://www.genewiz.com>). *Escherichia coli* BL21(DE3) pLysS cells (Novagen) transformed with the corresponding plasmid were grown in 2000 ml of LB medium supplemented with 50 μg ml⁻¹ kanamycin at 37 °C. When the culture reached an OD₆₀₀ value of 0.6–0.8, the culture was incubated at 22 °C for 1 h. The expression of EndoBT-3987_{WT} was induced by adding 0.5 mM IPTG. After ca. 16 h at 22 °C, the cells were harvested by centrifugation at 5000 × g for 20 min at 4 °C and resuspended in 50 ml of 50 mM Tris-HCl, pH 7.5, 500 mM NaCl and 10% glycerol, containing protease inhibitors (Thermo Scientific[™], A32955) and 2.5 μl of Benzamide (Merck, 71205; solution A). Cells were disrupted by sonication (12 cycles of 10 s pulses with 60 s cooling intervals between the pulses, and 60% of amplitude) at 4 °C, and the suspension was centrifuged at 10,000 × g for 10 min at 4 °C. The supernatant after being filtrated by 0.2 μm and then applied into a HisPur NiNTA column (1 ml, Thermo Scientific) equilibrated with 50 mM Tris-HCl, pH 7.5, 500 mM NaCl and 10% glycerol. The elution was performed with a linear gradient of 0 to 500 mM imidazole in 20 ml of solution A at 1 ml min⁻¹. These were concentrated in an Amicon Ultra-15 centrifugal filter unit (Millipore) with a molecular cutoff of 10 kDa at 4000 × g. EndoBT was further purified by size-exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in PBS, pH 7.4. The eluted protein was stored at -80 °C. Further purification steps were needed for crystallography experiments with EndoBT-3987_{WT} and EndoBT-3987_{D312A/E314L}. The protein was dialyzed against 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT, and 500 mM EDTA (solution C), with TEV protease (1:20 ratio), overnight at 18 °C. The completeness of the enzymatic digestion reaction was confirmed by SDS-PAGE and the solution protein was then loaded into a Superdex 200 26/60 column (350 ml; GE Healthcare), equilibrated in 20 mM Tris-HCl pH 7.5, 50 mM NaCl. The eluted protein was concentrated at 10 mg mL⁻¹ using an Amicon Ultra-15 centrifugal filter unit (Millipore) with a molecular cutoff of 30 kDa at 4000 × g.

EndoBT-3987 crystallization and data collection. EndoBT-3987_{WT} was crystallized by mixing 0.25 μL of a protein solution at 10 mg ml⁻¹ in 20 mM Tris-HCl pH 7.5, 50 mM NaCl with 0.25 μL of 200 mM sodium bromide, and 20% (w/v) PEG 3350. Crystals grew in 1–2 days. They were transferred to a cryo-protectant solution containing 20% glycerol and frozen under liquid nitrogen. Complete X-ray diffraction datasets were collected at the beamline BL13-XALOC (ALBA, Cerdanyola del Valles, Spain). EndoBT-3987_{WT} crystallized in the trigonal space group *P*3₁2₁ with one molecule in the asymmetric unit and diffracted to a maximum resolution of 1.4 Å (Supplementary Table 1). The EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn complex was crystallized by mixing 0.25 μL of a protein solution at 10 mg ml⁻¹ in 20 mM Tris-HCl pH 7.5, 50 mM NaCl and 2.5 mM Man₉GlcNAc₂Asn, with 0.25 μL of 100 mM MES pH 6, 50 mM CaCl₂ and 10% (w/v) PEG 3350. They were transferred to a cryo-protectant solution containing 20% glycerol and frozen under liquid nitrogen. Complete X-ray diffraction of both datasets were collected at beamline I24 (Diamond Light source, Oxfordshire, UK). The EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn complex crystallized in the triclinic space group *R*3 with one molecule in the asymmetric unit and diffracted to a maximum resolution of 1.3 Å (Supplementary Table 1). The EndoBT-3987_{WT}-Man₅GlcNAc complex was crystallized by mixing 0.25 μL of a protein solution at 10 mg ml⁻¹ in 20 mM Tris-HCl pH 7.5, 50 mM NaCl and 2.5 mM Man₅GlcNAc₂Asn, with 0.25 μL of 100 mM MES pH 6.0, 200 mM CaCl₂ and 20% (w/v) PEG 6000. Crystals grew in 1–2 days. They were transferred to a cryo-protectant solution containing 10% glycerol and frozen under liquid nitrogen. Complete X-ray diffraction datasets were collected at the beamline BL13-XALOC (ALBA, Cerdanyola del Valles, Spain). The EndoBT-3987_{WT}-Man₅GlcNAc complex crystallized in the orthorhombic space group *P*2₁2₁2₁ with one molecule in the asymmetric unit and diffracted to a maximum resolution of 1.6 Å (Supplementary Table 1). The EndoBT-3987_{WT}-Man₉GlcNAc complex was crystallized in two crystal forms, referred thereafter as EndoBT-3987_{WT}-Man₉GlcNAc-1 and EndoBT-3987_{WT}-Man₉GlcNAc-2. The first crystal form, EndoBT-3987_{WT}-Man₉GlcNAc-1, was obtained by mixing 0.25 μL of a protein solution at 10 mg ml⁻¹ in 20 mM Tris-HCl

pH 7.5, 50 mM NaCl and 2.5 mM Asn-GlcNAc₂Man₉ with 0.25 μL of 0.02 M sodium/potassium phosphate, 100 mM SPG (succinic acid, phosphate, glycine) system pH 9.0, and 25% (w/v) PEG 1500. The second crystal form EndoBT-3987_{WT}-Man₉GlcNAc-2 was obtained by mixing 0.25 μL of a protein solution at 10 mg ml⁻¹ in 20 mM Tris-HCl pH 7.5, 50 mM NaCl and 2.5 mM Man₉GlcNAc₂Asn with 0.25 μL of 0.02 M sodium/potassium phosphate, 100 mM Bis-Tris propane pH 8.5, and 20% (w/v) PEG 3350. Both crystal forms grew in 1–2 days. They were transferred to a cryo-protectant solution containing 10% glycerol and frozen under liquid nitrogen. Complete X-ray diffraction datasets for both crystal forms were collected at beamline I03 (Diamond Light source, Oxfordshire, UK). EndoBT-3987_{WT}-Man₉GlcNAc-1 crystal and crystallized in the triclinic space group *P*1 with two molecules in the asymmetric unit and diffracted to a maximum resolution of 2.0 Å (Supplementary Table 1). The second crystal form EndoBT-3987_{WT}-Man₉GlcNAc-2, crystallized in the orthorhombic space group *P*2₁2₁2₁ with one molecule in the asymmetric unit and diffracted to a maximum resolution of 1.7 Å (Supplementary Table 1). All datasets were integrated and scaled with XDS following standard procedures⁶².

EndoBT-3987 structures determination and refinement. Structure determination of EndoBT-3987_{WT} was carried out by molecular replacement methods implemented in Phaser⁶³ and the PHENIX suite⁶⁴, and using the pdb code 3POH as a search template. Structure determination of EndoBT-3987_{D312A/E314L}-Man₉GlcNAc₂Asn, EndoBT-3987_{WT}-Man₅GlcNAc, EndoBT-3987_{WT}-Man₉GlcNAc-1, and EndoBT-3987_{WT}-Man₉GlcNAc-2 complexes were carried out by molecular replacement using the crystal structure of EndoBT-3987_{WT} as a template model. Model rebuilding was carried out with Buccaneer⁶⁵ and the CCP4 suite⁶⁶. The final manual building was performed with Coot⁶⁷ and refinement with phenix.refine⁶⁸. The structure was validated by MolProbity⁶⁹. Data collection and refinement statistics are presented in Supplementary Table 1. Molecular graphics and structural analyses were performed with the UCSF Chimera package⁷⁰.

Chemoenzymatic preparation of HM N-glycans. The HM N-linked glycan Man₉GlcNAc₂Asn was prepared by digestion of soybean agglutinin isolated from soybean flour, and subsequent chromatographic purification. Crude soybean agglutinin (3.2 g) was obtained from 500 g of soybean flour (Sigma) through fractional precipitation with ammonium sulfate and digested thoroughly with pronase (Sigma). The digestion was filtered, and the filtrate was lyophilized. The residue was loaded onto a column (1.5 × 70 cm) of Sephadex G50 (Sigma), which was pre-equilibrated and eluted with 0.1 M AcOH. The fractions containing Man₉GlcNAc₂Asn were pooled and lyophilized. The material was finally purified by reverse-phase HPLC to afford homogeneous Man₉GlcNAc₂Asn (20 mg) as a white powder after lyophilization, which was characterized by compositional analysis, HPLC, and electron spray ionization mass spectrometry (ESI-MS)⁷¹. ESI-MS: calculated for Man₉GlcNAc₂Asn, *M* = 1996.69 Da; found (*m/z*), 999.74 [*M* + 2H]²⁺, 685.98 [*M* + H + Na + K]³⁺. The Man₅GlcNAc₂Asn N-glycan was prepared by digestion of the high-mannose N-glycan (Man₉GlcNAc₂Asn) using the α1,2-mannosidase from *B. thetaiotaomicron* (Supplementary Fig. 7)⁷². Specially, the mannosidase (10 μg) was added to a solution of Man₉GlcNAc₂Asn (10 mg, 0.005 mmol) in a buffer (PBS, 100 mM, pH 7.4, 1 mL). The reaction was monitored with LC-MS (Waters), and the desired product was purified with size-exclusion chromatography (GE Healthcare) to give the pure Man₅GlcNAc₂Asn N-glycan as a white powder after lyophilizing (6.34 mg, 94%). ESI-MS: calcd for Man₅GlcNAc₂Asn, *M* = 1349.21 Da; found (*m/z*), 675.59 [*M* + 2H]²⁺, 1349.90 [*M* + H]⁺.

Purification of high-mannose Fc and IgG1. CD4-induced (CD4i) IgG1 plasmid⁷³ was expressed in HEK293T cells (ATCC) using polyethyleneimine as a transfection agent. Kifunensine, a potent inhibitor of the mannosidase I enzyme, was used to ensure HM glycans were present on the Fc and IgG1⁴⁴. After transfection, cells were cultured for 96 h in Free-style F17 medium supplemented with GlutaMAX and Geneticin (Thermo Fisher Scientific). HM IgG1 was purified using protein A chromatography, with 20 mM sodium phosphate buffer pH 7.0 used as a binding buffer and 100 mM sodium citrate buffer pH 3.0 as elution buffer. The fractions

were neutralized with 1 M Tris pH 9.0. SDS-PAGE was used to identify fractions which contained HM IgG1. These were subsequently pooled and concentrated.

Enzymatic activity assays of alanine EndoBT-3987 mutants. Reactions for the alanine scan mutants were set up using 100 nM EndoBT-3987_{WT} for reactions with RNaseB and 500 nM for reactions with high-mannose IgG1. The enzymes were mixed with 5 μ M substrate in PBS pH 7.4 at room temperature. For the alanine scan, 10 μ l aliquots of the reaction were taken in triplicate and allowed to progress for 30 and 45 min for RNaseB and high-mannose IgG1 substrates, respectively. All reactions were quenched using 1.1 μ l of 1% trifluoroacetic acid. The quenched reactions were then mixed with 50 mM TCEP and analyzed by LC-MS using an Accela LC System attached to a LXQ linear ion trap mass spectrometer (Thermo Scientific, Waltham, MA). Relative amounts of the substrate and hydrolysis products were quantified after deconvolution of the raw data and identification of the corresponding peaks using BioWorks (Thermo Scientific, Waltham, MA). The data were plotted and statistical significance was determined using a multiple comparisons test (Tukey method) in GraphPad (GraphPad Software, LaJolla, CA).

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

Data availability

The atomic coordinates and structure factors have been deposited with the Protein Data Bank, accession codes **6T81** (EndoBT-3987_{WT}), **6TCV** (EndoBT-3987_{D312A/E314L}⁻Man₉GlcNAc₂Asn), **6TCW** (EndoBT-3987_{WT}-Man₉GlcNAc), **6T8K** (EndoBT-3987_{WT}-Man₉GlcNAc-1), and **6T8L** (EndoBT-3987_{WT}-Man₉GlcNAc-2). The source data underlying Fig. 4a and Supplementary Fig. 1b are provided as a Source Data file. Other data are available from the corresponding authors upon reasonable request.

Received: 13 January 2020; Accepted: 30 January 2020;

Published online: 14 February 2020

References

- Arpaia, N. et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451–455 (2013).
- Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: Potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* **6**, 121–131 (2008).
- Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327–336 (2011).
- Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9**, 313–323 (2009).
- Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105–1115 (2008).
- Luis, A. S. et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. *Nat. Microbiol.* **3**, 210–219 (2018).
- Cuskin, F. et al. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165–169 (2015).
- Desai, M. S. et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353 (2016).
- Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335 (2012).
- Salyers, A. A., Vercellotti, J. R., West, S. E. H. & Wilkins, T. D. Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. *Appl. Environ. Microbiol.* **33**, 319–322 (1977).
- Salyers, A. A., West, S. E. H., Vercellotti, J. R. & Wilkins, T. D. Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl. Environ. Microbiol.* **34**, 529–533 (1977).
- Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the bacteroidetes sus-like paradigm. *J. Biol. Chem.* **284**, 24673–24677 (2009).
- Tancula, E., Feldhaus, M. J., Bedzyk, L. A. & Salyers, A. A. Location and characterization of genes involved in binding of starch to the surface of *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **174**, 5609–5616 (1992).
- Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008).
- Xu, J. et al. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**, 2074–2076 (2003).
- Xu, J. et al. Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol.* **5**, e156 (2007).
- Lapébie, P., Lombard, V., Drula, E., Terrapon, N. & Henrissat, B. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat. Commun.* **10**, 2043 (2019).
- Ndeh, D. et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65–70 (2017).
- Ficko-Blean, E. et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat. Commun.* **8**, 1685 (2017).
- Rogowski, A. et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 7481 (2015).
- Larsbrink, J. et al. A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. *Nature* **506**, 498–502 (2014).
- Briliūtė, J. et al. Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat. Microbiol.* **4**, 1571–1581 (2019).
- Hehemann, J. H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
- Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc. Natl Acad. Sci. USA* **109**, 19786–19781 (2012).
- Pluvinage, B. et al. Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. *Nat. Commun.* **9**, 1043 (2018).
- Liu, C. P. et al. Glycoengineering of antibody (Herceptin) through yeast expression and in vitro enzymatic glycosylation. *Proc. Natl Acad. Sci. USA* **115**, 720–725 (2018).
- Trastoy, B. et al. Structural basis for the recognition of complex-type N-glycans by Endoglycosidase S. *Nat. Commun.* **9**, 1874 (2018).
- Maley, F., Trimble, R. B., Tarentino, A. L. & Plummer, T. H. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem.* **180**, 195–204 (1989).
- Karamanos, Y., Bourgerie, S., Barreaud, J. P. & Julien, R. Are there biological functions for bacterial endo-N-acetyl- β -D-glucosaminidases? *Res. Microbiol.* **146**, 437–443 (1995).
- Fairbanks, A. J. The ENGases: versatile biocatalysts for the production of homogeneous N-linked glycopeptides and glycoproteins. *Chem. Soc. Rev.* **46**, 5128–5146 (2017).
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
- Henrissat, B. & Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**, 637–644 (1997).
- Li, C. & Wang, L. X. Chemoenzymatic methods for the synthesis of glycoproteins. *Chem. Rev.* **118**, 8359–8413 (2018).
- Collin, M. & Olsén, A. EndoS, a novel secreted protein from *Streptococcus pyogenes* with endoglycosidase activity on human IgG. *EMBO J.* **20**, 3046–3055 (2001).
- Sjögren, J. et al. EndoS2 is a unique and conserved enzyme of serotype M49 group A *Streptococcus* that hydrolyzes N-linked glycans on IgG and α 1-acid glycoprotein. *Biochem. J.* **455**, 107–118 (2013).
- Sjögren, J. et al. EndoS and EndoS2 hydrolyze Fc-glycans on therapeutic antibodies with different glycoform selectivity and can be used for rapid quantification of high-mannose glycans. *Glycobiology* **25**, 1053–1063 (2015).
- Tarentino, A. L., Plummer, T. H. & Maley, F. The release of intact oligosaccharides from specific glycoproteins by endo-beta-N acetylglucosaminidase H. *J. Biol. Chem.* **249**, 818–824 (1974).
- Tarentino, L. & Maley, F. Purification and properties of an Endo- β -N-acetylglucosaminidase from *Streptomyces griseus*. *J. Biol. Chem.* **249**, 811–817 (1974).
- Gaunitz, S., Nagy, G., Pohl, N. L. B. & Novotny, M. V. Recent advances in the analysis of complex glycoproteins. *Anal. Chem.* **89**, 389–413 (2017).
- Freeze, H. H. & Kranz, C. Endoglycosidase and glycoamidase release of N-linked glycans. *Curr. Protoc. Mol. Biol.* **62**, 12.4.1–12.4.25 (2010).
- Taner, S. B. et al. Interactions of NK cell receptor KIR3DL1*004 with chaperones and conformation-specific antibody reveal a functional folded state as well as predominant intracellular retention. *J. Immunol.* **186**, 62–72 (2011).
- Guo, Y. et al. Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity. *J. Exp. Med.* **208**, 2083–2098 (2011).
- Lee, J. E., Fusco, M. L. & Saphire, E. O. An efficient platform for screening expression and crystallization of glycoproteins produced in human cells. *Nat. Protoc.* **4**, 592–604 (2009).
- Chang, V. T. et al. Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15**, 267–273 (2007).
- Davis, S. J. & Crispin, M. in *Functional and Structural Proteomics of Glycoproteins* (eds. Owens, R. & Nettleship, J. E.) 127–158 (Springer, Dordrecht, 2011).

46. Rao, V., Guan, C. & Van Roey, P. Crystal structure of endo-beta-N-acetylglucosaminidase H at 1.9 Å resolution: active-site geometry and substrate recognition. *Structure* **3**, 449–457 (1995).
47. Rao, V., Cui, T., Guan, C. & Van Roey, P. Mutations of endo-beta-N-acetylglucosaminidase H active site residue Asp 130 and Glu 132: activities and conformations. *Protein Sci.* **8**, 2338–2346 (1999).
48. Van Roey, P., Rao, V., Plummer, T. H. J. & Tarentino, A. L. Crystal structure of endo-beta-N-acetylglucosaminidase F1, an alpha/beta-barrel enzyme adapted for a complex substrate. *Biochemistry* **33**, 13989–13996 (1994).
49. Rao, V., Roey, P., Van, Cui, T. & Guan, C. Mutations of endo-beta-N-acetylglucosaminidase H active site residues Asp 130 and Glu 132: activities and conformations. *Protein Sci.* **8**, 2338–2346 (2008).
50. Trimble, R. B. & Tarentino, A. L. Identification of distinct endoglycosidase (endo) activities in *Flavobacterium meningosepticum*: endo F1, endo F2, and endo F3. Endo F1 and endo H hydrolyze only high mannose and hybrid glycans. *J. Biol. Chem.* **266**, 1646–1651 (1991).
51. Du, J. J., Klontz, E. H., Guerin, M. E., Trastoy, B. & Sundberg, E. J. Structural insights into the mechanisms and specificities of IgG-active endoglycosidases. *Glycobiology* **cwz042** (2019).
52. Klontz, E. H. et al. Molecular basis of broad spectrum N-glycan specificity and processing of therapeutic IgG monoclonal antibodies by Endoglycosidase S2. *ACS Cent. Sci.* **5**, 524–538 (2019).
53. Trimble, R. B., Tarentino, A. L., Plummer, T. H. & Maley, F. Asparaginyl glycopeptides with a low mannose content are hydrolyzed by endo-beta-N-acetylglucosaminidase H. *J. Biol. Chem.* **253**, 4508–4511 (1978).
54. Böhle, L. A., Mathiesen, G., Vaaje-Kolstad, G. & Eijsink, V. G. H. An endo-beta-N-acetylglucosaminidase from *Enterococcus faecalis* V583 responsible for the hydrolysis of high-mannose and hybrid-type N-linked glycans. *FEMS Microbiol. Lett.* **325**, 123–129 (2011).
55. Fujita, K. et al. Identification of amino acid residues essential for the substrate specificity of *Flavobacterium* sp. Endo-beta-N-acetylglucosaminidase. *Biosci. Biotechnol. Biochem.* **65**, 1542 (2001).
56. Tzelepis, G. D., Melin, P., Jensen, D. F., Stenlid, J. & Karlsson, M. Functional analysis of glycoside hydrolase family 18 and 20 genes in *Neurospora crassa*. *Fungal Genet. Biol.* **49**, 717–730 (2012).
57. Stals, I. et al. Identification of a gene coding for a deglycosylating enzyme in *Hypocrea jecorina*. *FEMS Microbiol. Lett.* **303**, 9–17 (2010).
58. Hamaguchi, T. et al. Purification, characterization and molecular cloning of a novel endo-beta-N-acetylglucosaminidase from the basidiomycete, *Flammulina velutipes*. *Glycobiology* **20**, 420–432 (2010).
59. Mello, L. V., Chen, X. & Rigden, D. J. Mining metagenomic data for novel domains: BACON, a new carbohydrate-binding module. *FEBS Lett.* **584**, 2421–2426 (2010).
60. Leclaire, C. et al. Molecular basis for intestinal mucin recognition by galectin-3 and C-type lectins. *FASEB J.* **32**, 3301–3320 (2018).
61. Li, C. et al. FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* **11**, 92 (2011).
62. Kabsch, W. XDS. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 125–132 (2010).
63. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
64. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 213–221 (2010).
65. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **62**, 1002–1011 (2006).
66. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 235–242 (2011).
67. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 486–501 (2010).
68. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **68**, 352–367 (2012).
69. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 12–21 (2010).
70. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
71. Wang, L. X., Ni, J., Singh, S. & Li, H. Binding of high-mannose-type oligosaccharides and synthetic oligomannose clusters to human antibody 2G12: implications for HIV-1 vaccine design. *Chem. Biol.* **11**, 127–134 (2004).
72. Toonstra, C., Wu, L., Li, C., Wang, D. & Wang, L. X. Top-down chemoenzymatic approach to synthesizing diverse high-mannose N-glycans and related neoglycoproteins for carbohydrate microarray analysis. *Bioconjug. Chem.* **29**, 1911–1921 (2018).
73. Guan, Y. et al. Diverse specificity and effector function among human antibodies to HIV-1 envelope glycoprotein epitopes exposed by CD4 binding. *Proc. Natl Acad. Sci. USA* **110**, E69–E78 (2013).

Acknowledgements

This work was supported by the MINECO/FEDER EU contracts BFU2016-77427-C2-2-R, BFU2017-92223-EXP and Severo Ochoa Excellence Accreditation SEV-2016-0644; and the Basque Government contract KK-2019/00076 (to M.E.G.); NIH T32 AI095190 (to E.H.K.); and NIH R01 GM096973 (to L.-X.W.). This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 844905 (B.T.). We acknowledge Diamond Light Source (proposals mx20113), ALBA synchrotron beamline BL13-XALOC (mx2018093013), and iNEXT (proposals 1618/2538) for providing access to synchrotron radiation facilities. We gratefully acknowledge all members of the Structural Glycobiology Group (CIC bioGUNE, Spain) for valuable scientific discussions.

Author contributions

B.T., L.-X.W., E.J.S., and M.E.G. conceived the project. B.T., J.J.D., E.H.K., C.L., and J.O.C. performed the experiments. B.T., J.J.D., E.H.K., C.L., J.O.C., E.J.S., L.-X.W., and M.E.G. analyzed the results. B.T., J.J.D., E.J.S., and M.E.G. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-14754-7>.

Correspondence and requests for materials should be addressed to B.T., E.J.S. or M.E.G.

Peer review information *Nature Communications* thanks the 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) 2020