Structural and functional characterization of a multi-domain GH92 α-1,2-mannosidase from *Neobacillus novalis*

Bartłomiej M. Kołaczkowski[1,4], Olga V. Moroz[2], Elena Blagova[2], Gideon J. Davies[2], Marie Sofie Møller[3], Anne S. Meyer[3], Peter Westh[3], Kenneth Jensen[4], Keith S. Wilson[*,2], Kristian B.R.M. Krogh[*,4]

[1] Roskilde University, Department of Science and Environment, Universitetsvej 1, Building 28, 4000 Roskilde, Denmark

[2]York Structural Biology Laboratory, Department of Chemistry, University of York, York, YO10 5DD, UK

[3]Technical University of Denmark, Department of Biotechnology and Biomedicine, Building 224, 2800 Kongens Lyngby, Denmark

[4]Novozymes A/S, Biologiens Vej 2, 2800 Kongens Lyngby, Denmark

[*]correspondence: kbk@novozymes.com

[*]correspondence: keith.wilson@york.ac.uk

Keywords: glycans; α-mannosidase; *Neobacillus novalis;* glycoside hydrolase family 92; carbohydrate-binding module family 32;

Abstract

Many secreted eukaryotic proteins are *N*-glycosylated with oligosaccharides comprised of a high mannose *N*-glycan core and, in the specific case of yeast cell wall proteins, an extended backbone of α-1,6-mannan carrying a number of α-1,2 and α-1,3 mannose substituents of varying lengths. α-Mannosidases from CAZy family GH92 release terminal mannose residues from these *N*-glycans providing access for the α-endomannanases which then degrade the α-mannan backbone. Most characterized GH92 α-mannosidases consist of a single catalytic domain while a few have extra domains including putative carbohydrate binding modules (CBM). To date, neither the function nor structure of a multi-domain GH92 α-mannosidase CBM has been characterized. Here we report the biochemical investigation and crystal structure of the full-length five-domain GH92 α-1,2-mannosidase from *Neobacillus novalis* (*Nn*GH92) with mannoimidazole bound in the active site and an additional mannoimidazole bound to the N-terminal CBM32. The structure of the catalytic domain is very similar to that reported for the *Bt*3990 GH92 α-mannosidase from *Bacteroides thetaiotaomicron*, with the substrate binding site being highly conserved. The function of the CBM32s and other *Nn*GH92 domains was investigated by their sequential deletion and suggested that whilst their binding to the catalytic domain was crucial for the overall structural integrity of the enzyme, they appear to have little impact on the binding affinity to yeast α-mannan substrate. These new findings provide a better understanding of how to select and optimize other multi-domain bacterial GH92 α-mannosidases for yeast α-mannan or mannose-rich glycan degradation.

Introduction

The fungal cell wall consists of polysaccharide layers, including chitin and β-glucans, which provide a scaffold for the mannoproteins present in the outer layer. These glycoproteins are composed of a protein moiety decorated with either *O*- or *N*-linked glycans. In *Saccharomyces cerevisiae*, the *N*-glycans have a high mannose core backbone (Figure 1A) that is extended to ~200 α-1,6-linked mannose units (Figure 1B).[1] This α-1,6-linked mannose backbone is decorated with side chains composed of first α-1,2-linked mannose units, followed by terminal α-1,3-linked mannose units.[2]

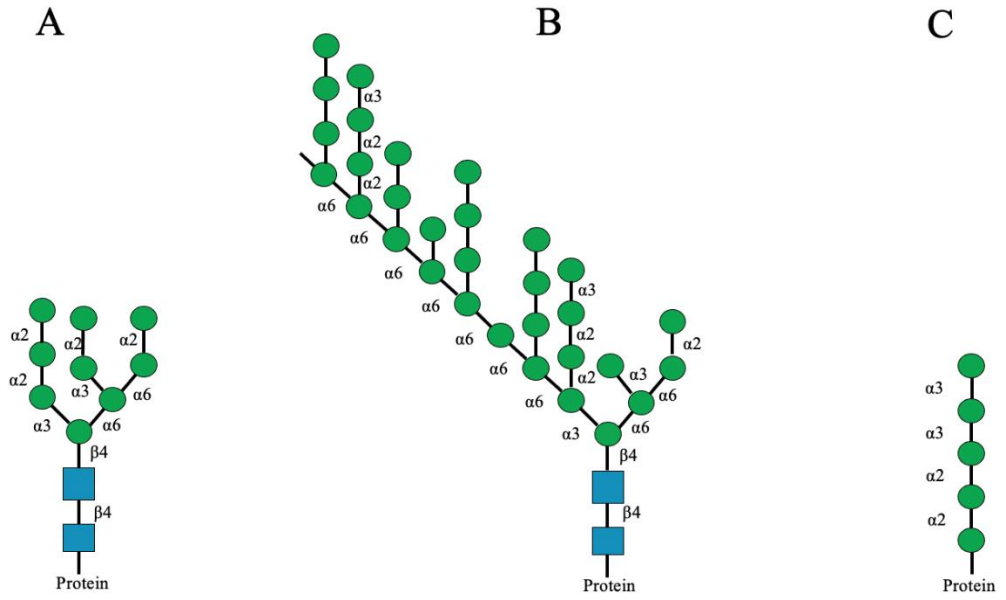

Figure 1. Schematic illustration of the structures targeted by GH92 α-mannosidases. (A) the high-mannose *N*-glycan core[3], (B) α-mannan found in the outer layer of the yeast cell wall; the α-1,6-linked backbone can reach a degree of polymerization ~200[2], (C) an *O*-linked glycan[4]. The proteins produced in *S. cerevisiae* are often decorated with the structures (A) and (C) that contribute to the protein *N*- and *O*-glycosylation, respectively. The scheme was inspired by[3]. The green circles and the blue squares are mannose and N-acetyl-glucosamine units, respectively, depicted according to the SNFG guidelines [5,6]

A key requirement for the complete enzymatic hydrolysis of α-mannan is the removal of the side chains that obstruct access to the α-1,6-linked mannan backbone.[7] α-Mannosidases capable of

hydrolysing α-1,2-/α-1,3-glycosidic bonds in the α-mannan side chains belong to the CAZy[8] glycoside hydrolase family GH92. In 2015, Cuskin *et al*.[7] published a study showing that the human gut bacterium, *Bacteroides thetaiotaomicron*, had developed a highly specialized enzymatic machinery to degrade yeast α-mannan releasing short α-oligomannosaccharides and single mannose residues which it can use as a sole carbon source of energy. This allowed the identification of multiple genes, organized in polysaccharide utilization loci, encoding enzymes with α-mannosidase or α-mannanase activity. Other α-mannosidases, belonging to GH99 and GH38, have been identified which facilitate hydrolysis of the side chains in the degradation of yeast α-mannan by *B. thetaiotaomicron*.[3,7]

All known family GH92 members are Ca^{2+}-dependent *exo*-α-mannosidases that perform hydrolysis of terminal non-reducing mannose residues with inversion of anomeric configuration.[9] Based on the characterization of the *B. thetaiotaomicron* GH92 α-mannosidases, a wide range of glycosidic bond specificities have been identified, including α-1,2-, α-1,3-, α-1,4-linked mannose linkages.[9] GH92 α-mannosidases were found to play an important role in the depolymerization of the *S. cerevisiae* cell wall α-mannan[7], and mannose-rich *N*-glycans[10,11] or O-glycans[12] found in glycoproteins produced by fungi (Figure 1).

The structures of several bacterial GH92 α-mannosidases have been reported including *Bt*3990 (PDB: 2WVX), *Bt*2199 (PDB: 2WVY), *Bt*3130 (PDB: 6F8Z), *Bt*3965 (PDB: 6F91) from *Bacteroides thetaiotaomicron*[9,13], *Cc*GH92 from *Cellulosimicrobium cellulans* (PDB: 2XSG)[14], *Ef*Man-I from *Enterococcus faecalis* (PDB: 6DWO)[11] and *Sp*GH92 from *Streptococcus pneumoniae* (PDB: 5SWI)[15]. All known structures of GH92 α-mannosidases catalytic domains have a highly conserved two-subdomain composition, a N-terminal β-sandwich and a C-terminal $(\alpha/\alpha)_6$-barrel, with both subdomains contributing to a pocket-like active site with distinctive –1 and +1 sugar binding subsites.[9,16] Among these bacterial GH92 structures, a common pattern was identified in the active site, with a highly conserved –1 subsite accommodating the mannosyl non-reducing end and a divergent +1 subsite [9,13]. The poorly conserved substrate-binding amino acid residues at the +1 subsite were found to be a structural factor for differentiating enzyme preference towards mannosyl linkages including α-1,2- (*Bt*3990), α-1,3- (*Bt*3130) α-1,4-linkages (*Bt*3965).[9,13]

In the CAZy database[8], there are currently 31 "characterized" GH92 α-mannosidases. Among them, there is only one enzyme, namely the GH92 αMan2 from *Microbacterium* sp. M-90[17], which possesses an extra domain in addition to the catalytic core domain, and this extra domain belongs to carbohydrate-binding module family 32 (CBM32). CBM32 domains display a β-sandwich fold containing a metal ion, usually Ca^{2+}, with an exposed shallow cleft carbohydrate binding site.[18–20] The CBM32s are a family with a wide range of ligand specificities, primarily targeting non-reducing ends of complex glycans such as the mucin-type[21] including galactose, lactose[18] and N-acetyl-lactosamine (LacNAc)[22]. Recently, a few CBM32s have been identified appended to GH enzymes catalysing degradation of plant cell wall polysaccharides, including pectin[23] and β-mannan[24]. No CBM32 was present in any of the 22 GH92 α-mannosidases from *B. thetaiotaomicron* and the binding specificity of the CBM32s was not assigned to α-mannan or other glycan containing α-mannooligosaccharides.

Here we report the biochemical characterization and crystal structure of the full-length multi-domain GH92 α-mannosidase from *Neobacillus novalis* (*Nn*GH92), a bacterium identified in agricultural soils[25,26]. *Nn*GH92 is additionally important as it has recently been used as an enzymatic technique to map fungal high-mannose structures.[12] The 3-D structure was solved with the known GH92 inhibitor mannoimidazole (ManI) bound in the active site. Sequence and structural alignments were performed with the known *Bt*3990 GH92 α-1,2-mannosidase from *B. thetaiotaomicron* to identify the general acid[27], the Brønsted base[27] and the ligand interactions in the Ca^{2+}-containing active site. Domain deletion variants were designed and expressed to evaluate the influence of each non-catalytic domain on the activity and binding ability of the enzyme on yeast α-mannan and the isolated yeast cell wall from *S. cerevisiae*.

Results

The full-length sequence of *Nn*GH92 was deposited in GenBank with the code LR963497.1. In contrast to *B. thetaiotaomicron*, which produces GH92 α-mannosidases as solely catalytic domains, *Nn*GH92 has five domains as discussed below. A gene encoding the full-length *Nn*GH92 was cloned and expressed in *Bacillus subtilis*. The construct used for the structural study excludes the signal peptide and therefore started from Ser34, with an N-terminal His-tag: HHHHHHPR. After purification, it showed a single band in SDS-PAGE analysis (Figure S1, lane 2).

Overall structure of *Nn*GH92

With the aim of confirming how *Nn*GH92 accommodates and interacts with mannopyranosides, and to establish the location and possible roles of the associated non-catalytic domains, the crystal structure of the wild type was solved in complex with the mannosidase inhibitor mannoimidazole (ManI) (PDB entry: 7NSN) (Figure 2). The structure was solved by molecular replacement using the published structure of *Bt*3990 (PDB: 2WZS)[9] as a template and refined at 2.3Å resolution (Table 1). *Nn*GH92 only crystallized in the presence of ManI, suggesting that the ligand-binding improved the structure stability leading to the formation of high-quality crystals. Most importantly, the structure corresponded to the full-length enzyme including all the non-catalytic domains, Figure 2, at least in Chain A.

Table 1. Data collection and refinement of *Nn*GH92

Data set[a]	**MVL-*Nn*GH92**
Beamline	I04-1
Wavelength (Å)	0.92Å
Space group	$P2_1$
Unit cell parameters (Å)	*a=94.61, b=151.94, c=114.01, β=94.63°*
Total reflections	410285 (21040)
Unique reflections	141182 (7109)
Completeness (%)	98.1 (99.5)
Multiplicity	2.9 (3.0)
R_{merge} (%)	13.2 (77.7)
R_{meas} (%)	18.6 (109.0)
R_{pim}[b]	13.1 (76.3)
< I >/< σ*(I)*>	5.1 (1.1)
Resolution range (Å)	47.92-2.29 (2.33-2.29)
$CC_{1/2}$[c]	0.981 (0.505)
Wilson *B*-factor (Å^2)	24.4
No. of reflections, working set	141151
No. of reflections, test set	6948
Final R_{cryst}	0.21
Final R_{free}	0.25
Cruickshank DPI	0.31
No. of non-H atoms	20260
R.m.s. deviations	
Bonds (Å)	0.007
Angles ()	1.484
Average *B* factors (Å^2)	
Chain A protein	31
MVL -1	27
MVL -2	36
Chain B protein	36
MVL-1	23
Molprobity score	2.09
Ramachandran plot	
Most favoured (%)	96.25
Outliers (%)	0.15
PDB code	7NSN

[a]Values for the outer shell are given in parentheses.

[b]$R_{pim} = \Sigma_h[m/(m-1)]^{1/2}\Sigma_i \mid I_{h,i} - < I_h > |/\Sigma_h\Sigma_i I_{h,i}$; $R_{meas} = \Sigma_h[1/(m-1)]^{1/2}\Sigma_i \mid I_{h,i} - < I_h >|/\Sigma_h\Sigma_i I_{h,I}$.

[c]$CC_{1/2}$ is defined in[28]

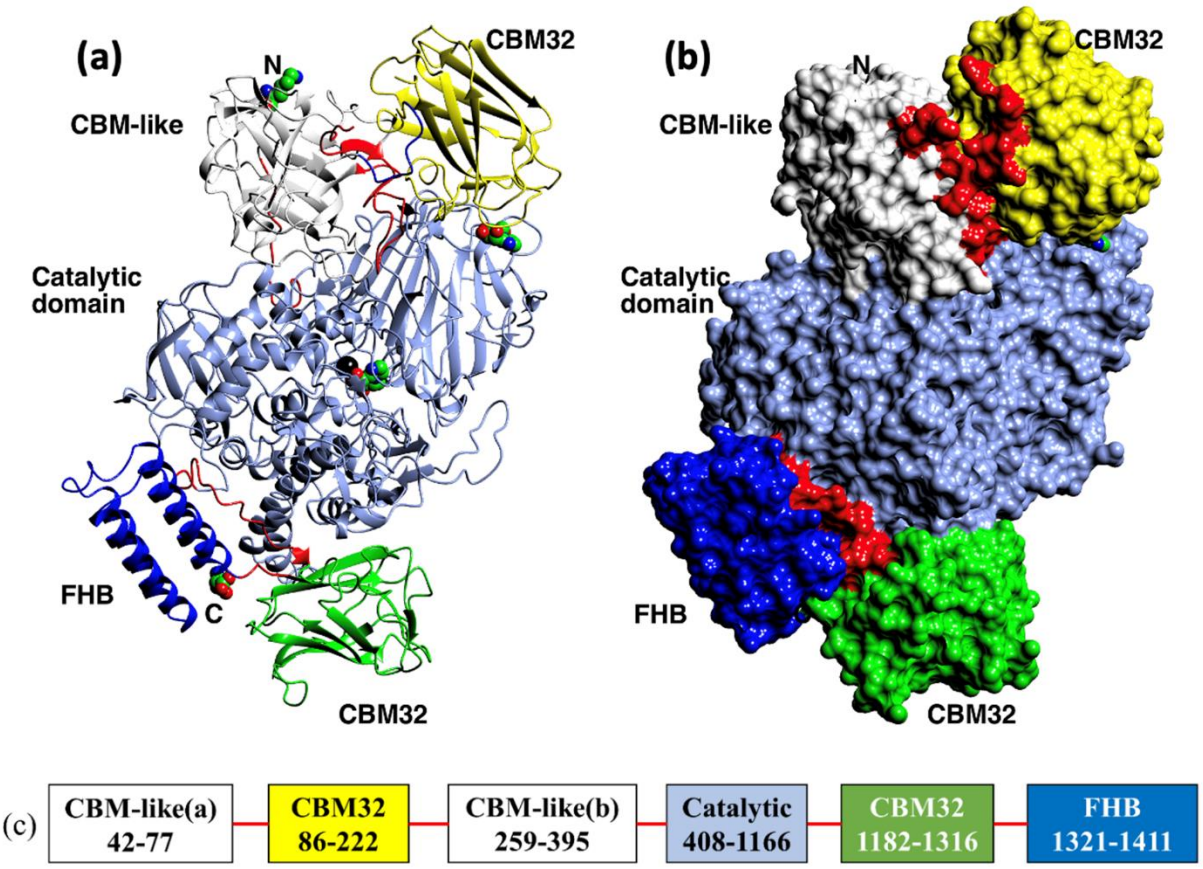

Figure 2. The structure of *Nn*GH92 (PDB: 7NSN). (a) The fold of Chain A in ribbon format. The domains are coloured, starting from the N-terminus CBM-like (white, whose fold is split by the first CBM32 insert), CBM32 (lemon), catalytic domain (ice blue), second CBM32 (green) and 4-helix bundle (FHB, blue). The linkers between the domains are in red and are ordered in the structure. Both mannoimidazoles (ManI) are shown as spheres coloured by atom type, the first one at the active site between the two subdomains of the catalytic domain, the second between the catalytic and N-terminal CBM32. The calcium ion adjacent to the active site ManI is shown as a black sphere. The N terminal residue (Lys42) and C terminal residue (Asp1411) are shown as spheres. (b) The surface of the domains coloured as in (a). The extensive packing surfaces of the domains is evident. The ordered linkers between domains are highlighted in red. The images were created with CCP4mg[29]. (c) Schematic representation of the domain structure of *Nn*GH92. CBM-like(a) and CBM-like(b) correspond to the same domain structure.

There are two independent monomers in the asymmetric unit. Chain A consists of residues 42-1411. A few residues at the N-terminus including the His-Tag are disordered with no electron density. The rest of A chain is well ordered in the crystal, with the exception of a short loop of five

residues, 224-227. At the N-terminus there is the start of an all beta sheet domain reminiscent of a CBM and termed "CBM-like", which is made up of residues 42-77 and 259-395. A CBM32 (residues 86-222) is inserted into a loop of the CBM-like domain. These are followed by the catalytic domain (residues 408-1166), the C-terminal CBM32 (residues 1182-1316) and a 4-helix bundle domain (FHB) (residues 1321-1411). In Chain B, there were only very poor fragments of density for the N-terminal CBM32 domain, and residues 79-227 are missing from the model. The rest of the fold of the A and B chains is essentially identical with an r.m.s.d. of 0.27Å over 1221 equivalent C_α atoms. This supports a stable set of interactions for the extra domains surrounding the core catalytic domain. The rest of the description will focus on the better ordered Chain A. The catalytic domain was traced without any breaks and its fold, as expected, was very similar to that of *Bt*3990 with an r.m.s.d. of 2.01Å over 712 equivalent C_α atoms reflecting the moderate sequence identity (41.6%). The chain of the catalytic domain adopted the expected two-subdomain structure: an N-terminal β-sandwich and a C-terminal $(\alpha/\alpha)_6$-barrel (Figure 2A). Superposition of the *Bt*3990 and *Nn*GH92 crystal structures, showed no difference in structural elements for the catalytic domains. The active site is a shallow pocket, with both N- and C-terminal domains contributing to its shape.

The reaction mechanism of GH92 enzymes, catalysis occurring with inversion of anomeric configuration, demands two residues; an acid to assisting leaving-group departure and a base to enhance the nucleophilic attack of water.[9] The catalytic residues, the Brønsted acid Glu944 (Glu533 in *Bt*3990) and the Brønsted base Asp1058 (Asp644 in *Bt*3990), are conserved (Figure 3A, Figure S2).[27] To investigate further the local features in the active site, the structure of *Nn*GH92 was superimposed on that of *Bt*3990 in complex with thio-linked α-1,2-mannobiose (MSM) (PDB entry: 2WW3[9]) (Figure 3). Like *Bt*3990, *Nn*GH92 aligns MSM with a clear boundary between the –1 and +1 subsites indicated by the non-hydrolysable glycosidic S atom. Moreover, the ManI mannose ring superimposes on the mannose ring of the MSM at the –1 subsite. The residues providing interactions in both subsites are highly conserved, with nearly identical orientation and position in both enzymes (Figure 3B). At the –1 subsite, the Ca^{2+} ion supports the positioning of the sugar ring by interacting with the O2 and O3 of the mannose. The specificity of *Bt*3990 α-1,2-mannosidase is primarily driven by three residues at the +1 subsite: the His584-Glu585 pair of

residues forming hydrogen bonds with O3 and O4 of the MSM mannose and Trp88 providing a hydrophobic interaction.[13] Equivalent residues (His996-Glu997, Trp477) are present in the +1 subsite of *Nn*GH92 (Figure 3B, Figure S2). These structural findings, together with the observed activity on α-1,2-mannobiose[12] (Figure S3, Figure S4), confirmed the assignment of *Nn*GH92 as α-1,2-mannosidase. Interestingly, two residues, Leu581 and Leu793, enter the +1 subsite from the right (Figure 3B). In the structures of *Bt*3990, *Bt*2199 and *Bt*3130, there is Cys399 instead of Leu793 in *Nn*GH92 (Figure 3B). Despite the Cys399 not having been identified as involved in any interaction with the sugar moiety at the +1 subsite, the hydrophobic nature of the two Leu residues may be involved in coordinating other extended glycan chains, such as yeast α-mannan side chains. The absence of the residues conferring α-1,3-mannosidase specificity (identified in *Bt*3130[13]) in the +1 subsite of *Nn*GH92, probably explains its low activity towards α-1,3-mannobiose[12].

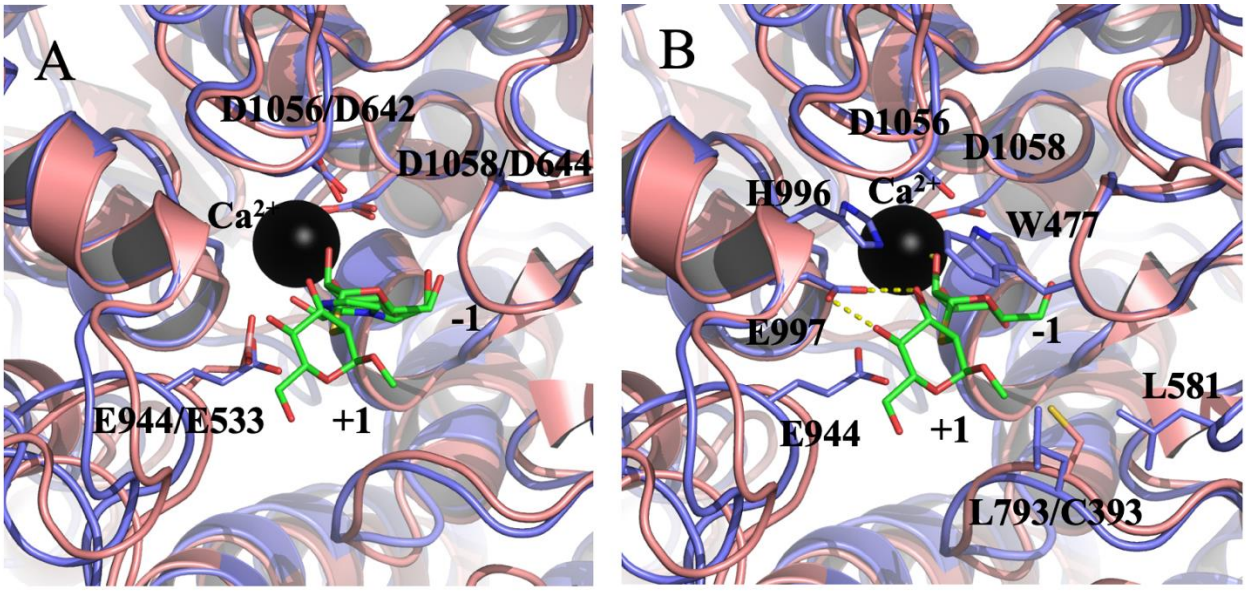

Figure 3. Comparison of the active sites of *Nn*GH92 in blue and *Bt*3990 in pink with α-mannosidase inhibitors. The *Bt*3990-MSM complex (PDB: 2WW3) was superimposed on the *Nn*GH92-ManI complex using PyMOL and its built-in function cealign. The calcium ion is coloured in black. (A) The general acid (Glu944/533) and two general bases Asp1056/642 and Asp1058/644 align at equivalent positions. At the –1 subsite, the mannose ring of ManI aligned with the non-reducing end of the MSM. (B) Amino acid residues shaping binding subsites and interacting with the ligands. At the +1 subsite, Leu793 in *Nn*GH92 was at the equivalent position to Cys393 in *Bt*3990, whereas the other residues were highly conserved in similar

orientations. The structures were visualized using PyMOL (The PYMOL Molecular Graphics System, version 2.3.2, Schrodinger, LLC).

The structure of the *Nn*GH92 non-catalytic domains

The structure of *Nn*GH92 contained several domains in addition to the catalytic domain (Figure 2, Figure 4). Both CBM32s had a β-sandwich fold architecture typical for this family with five- and three-stranded antiparallel β-sheets opposing one another (Figure 4 A, D). Ca^{2+} ions were buried within the structure of these CBM32s, a common feature of the CBM32 family[19,22,30,31].

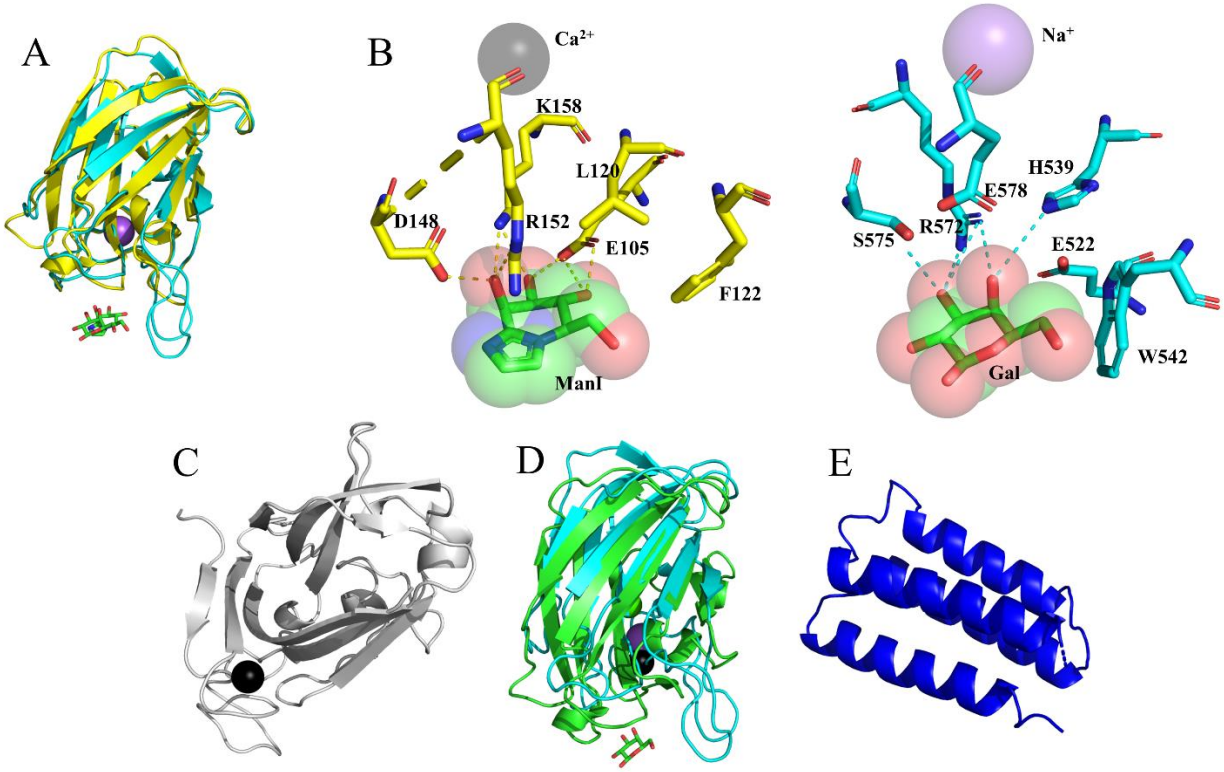

Figure 4. The structure of the *Nn*GH92 non-catalytic domains. (A) The N-terminal CBM32 (yellow) superimposed on that of the CBM32 of the GH33 sialidase from *Micromonospora viridifaciens* (*Mv*GH33), (cyan, PDB: 1euu[32]). (B) The ManI (coloured by atom type) bound to the N-terminal CBM32 (yellow) and galactose (Gal, coloured by atom type) bound to the CBM32 (cyan) of the *Mv*GH33. The highlighted

residues shape the binding sites of both CBM32. The Na^+ ion (purple) is in an equivalent position to the Ca^{2+} atom (black) in *Nn*GH92. (C) The CBM-like domain (grey) with a bound Ca^{2+} atom (black). (D) The C-terminal CBM32 (green) superimposed on the CBM32 of the *Mv*GH33 (cyan, PDB: 1euu[32]). (E) The FHB domain (blue) of unknown function.

There was clear electron density for a single ManI bound to the N-terminal CBM32. A CBM32-galactose complex from *Micromonospora viridifaciens* GH33 (*Mv*GH33) sialidase (PDB: 1euu[32]) was overlaid on this domain with an r.m.s.d. of 2.29Å over 128 equivalent C_α atoms and a sequence identity of 31.1% (Figure 4A). In the overlaid structures of *Nn*GH92-CBM32 and *Mv*GH33-CBM32, both ligands (ManI and galactose) and the metal ions (Ca^{2+} and Na^+) lie in equivalent positions, respectively, confirming the ligand binding site of the NnGH92 N-terminal CBM32. In Figure 4A, the major difference is visible in the loop covering the C6 group of the galactose and ManI, where *Mv*CBM32 is more extended with Trp542 (Figure 4B). While the loop is much shorter for the N-terminal CBM32, it has an aromatic residue, Phe122 (Figure 4B), with a similar orientation, providing stacking interactions with the C6 group of ManI. In *Mv*CBM32, residues Arg572 and Glu578 form hydrogen bonds with O3/O4 and O4 of the galactose ring, respectively. In the N-terminal CBM32, Arg152 has a similar orientation to Arg572 but is more distantly located, forming a hydrogen bond only with the O2 of ManI. In addition, Asp148-O2 and Glu105-O3/O4 form hydrogen bonds with the mannose ring of the ManI. The two ManI ligands, at the active site and binding site of N-terminal CBM32, are ~35 Å from one another and lie along the same axial plane, probably to facilitate the movement of the N-terminal CBM32 and the catalytic domain (CD) to bring the substrate closer to the active site.

Following the N-terminal CBM32, there is another domain, which despite its β–sheet, similar to "CBM-like" architecture including Ca^{2+}, cannot be assigned to any known CBM family based on sequence alignment (Figure 4C). Structural alignment with Gesamt[33] led to the identification of the closest homologue (PDB: 1PMH, r.m.s.d. 2.9Å over 157 equivalent C_α, Q-score 0.34) belonging to the CBM27-mannohexaose complex associated with a GH26 β-mannanase[34]. This CBM27 interacts with β-mannooligosaccharides through an aromatic platform formed by tryptophan residues (W23, W60, W113). Despite having the Ca^{2+} atoms at equivalent positions, the aromatic platform is absent in the *Nn*GH92 CBM-like domain. Based on the very low sequence identity (7%)

between the CBM27 and CBM-like domain, and the absence of bound ligand, it was not possible to assign a carbohydrate binding site for this domain. Therefore, the CBM-like domain could have either a purely structural function, or function to block access of unfavourable glycan chains to the enzyme active site, prioritizing only shorter glycan chains. A sequence homologue of the 'CBM-like' domain is also found in family GH38 α-mannosidases from *Clostridia* spp. (BCI61027.1 and BCI60986.1).

While the C-terminal CBM32 has a high structural similarity with the N-terminal CBM32 (r.m.s.d. 1.9Å over 128 equivalent C_α), they share only limited sequence identity (36%). The C-terminal CBM32 was also superimposed on that of *Mv*CBM32 to indicate a potential binding site (Figure 4D). Similar to the N-terminal CBM32, the C-terminal CBM32 putative binding site is found at the same axial plane, with a distance of ~40 Å to the active site of *Nn*GH92.

The FHB domain is composed of a bundle of four α-helices (Figure 4E), oriented towards the exterior of the *Nn*GH92 structure (Figure 2). Its closest structural homologue is the bacterial protein EntA from *Enterococcus faecium* (PDB:2BL8, r.m.s.d. 2.8Å over 72 equivalent C_α), which belongs to a group of immunity proteins conferring protection of bacteriocin-producing organisms against their own bacteriocins[35]. Due to the very low sequence identity (~6%), it is not possible to assign any function to the FHB. Other close structural homologues with low sequence identity (below 10%) were identified but no functions were provided for these domains (PDB: 2QZG, 2QSB, 2RLD). A three α-helix bundle is also appended at the C-terminus, in series with CBM32s, to a GH84 β-N-acetylglucosaminidase from *Clostridium perfringens*[36].

Design of *Nn*GH92 variants and biochemical characterization

As previously demonstrated, among the tested α-1,2-, α-1,3-, and α-1,6-linked mannobioses, *Nn*GH92 exhibited preference towards α-1,2-mannobiose, confirming it is α-1,2-mannosidase.[12] Upon longer incubation (19 h), α-1,3-mannobiose was also partially hydrolysed while no activity was detected on α-1,6-mannobiose.[12] *Nn*GH92 had a pH optimum between pH 6 and 7 and a temperature optimum between 42 and 52°C. The thermal stability, measured as melting temperature (T_m), was found to be 60°C in the presence of Ca^{2+} at pH 6 (Table 2). As expected, *Nn*GH92 was active on yeast α-mannan, reaching a degree of conversion of approximately 50%

after 1 h (Figure 5A). This suggests that it is likely to hydrolyse both the α-1,3- and α-1,2-glycosidic bonds present in the α-mannan side chains, as previously demonstrated for other GH92 α-mannosidases[7,9,17]. Furthermore, *Nn*GH92 did not exhibit activity on α-mannan from the *S. cerevisiae* yeast *mnn2* mutant[37], comprised of only the α-1,6-linked mannan backbone without side chains. Based on the high degree of conversion of yeast α-mannan by *Nn*GH92 (Figure 5A), the enzyme was tested in the presence of GH76 endo-α-1,6-mannanase and GH125 *exo*-α-1,6-mannosidase to investigate whether the combination of all three enzyme can boost α-mannan degradation. Indeed, all three enzymes improved the degree of conversion, indicating a complete depolymerization of the yeast α-mannan (Figure S5). Thin-layer chromatography (TLC) was used to provide a qualitative profile of the products generated by *Nn*GH92 on yeast α-mannan. After 1 h hydrolysis, only a single band was observed on the TLC plate. This confirms that *Nn*GH92 releases monosaccharide as its main enzymatic reaction product (Figure 5B), consistent with the exoglycosidase action of GH92 enzymes[7,9].

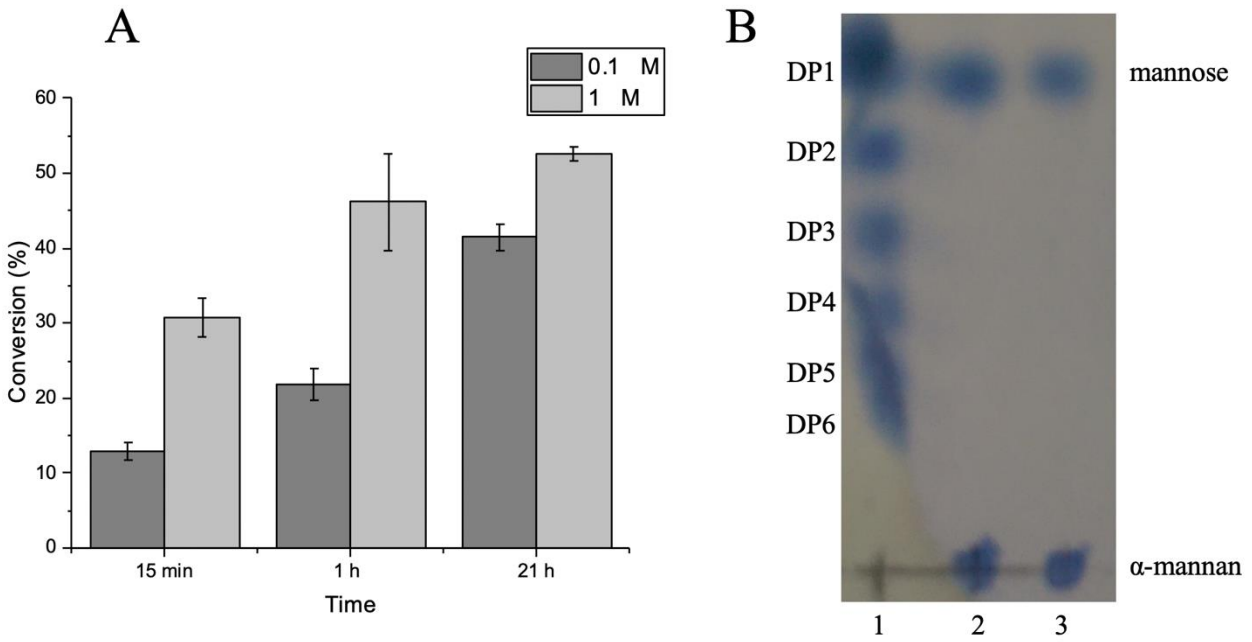

Figure 5. Yeast α-mannan hydrolysis by *Nn*GH92 (A) The hydrolysates were analysed at three different time points using the reducing sugar assay (PAHBAH). The degree of conversion was calculated based on the total mannose released after strong acid hydrolysis (see Materials and Methods). (B) Product release pattern analysed with TLC. Lane 1 represents standards composed of β-1,4-mannooligosaccharides with a

degree of polymerization (DP) between 1 and 6. Lanes 2 and 3 correspond to the α-mannan hydrolysates after 1 h at enzyme concentrations of 0.1 μM and 1 μM. Only a single band was observed, corresponding to DP1.

The Michaelis-Menten kinetic parameters with α-1,2-mannobiose were determined (Figure S3) and compared to the model *Bt*3990 GH92 α-1,2-mannosidase[9]. The maximum turnover (k_{cat}) of *Nn*GH92 was found to be approximately three times higher ($13.6 \times 10^3 \pm 3.7 \times 10^3$ min^{-1}) as compared to *Bt*3990 ($5.2 \times 10^3 \pm 3 \times 10^2$ min^{-1})[9] . Thus, the presence of the *Nn*GH92 associated domains appears to have an impact on the overall rate of hydrolysis. In addition, the Michaelis constant (K_M) was calculated, indicating a slightly lower substrate affinity of *Nn*GH92 (0.46 ± 0.22 mM) as *Bt*3990 (0.76 ± 0.11 mM)[9].

Table 2. The *Nn*GH92 wild-type and the truncated variants investigated in this study.

*Nn*GH92 variant name	CBM-like domain[b]	N-terminal CBM32[b]	C-terminal CBM32[b]	FHB[b]	MW[c] (kDa)	T_m[d] (°C)	Relative activity[e] (%)
wild-type (LR963497.1)[a]	+	+	+	+	151.7	60	100
ΔFHB	+	+	+	–	141.9	53	81 ± 0.7
ΔFHBCBM32	+	+	–	-	125.2	48	11 ± 2.0
ΔN-CBM	–	–	+	+	112.8	50	55 ± 2.0
core	–	–	–	–	85.5	49	4 ± 0.4
inactive (E944Q)	+	+	+	+	151.7	60	4 ± 0.5

[a] European Nucleotide Archive accession number. [b] (+) present (–) truncated. [c]Theoretical. [d]Melting temperature at pH 6. [e] activity measured on yeast α-mannan and calculated using wild type (WT) as the reference (Figure 6); ± corresponds to the standard deviation of triplicate sample reaction.

To elucidate the biochemical role of the associated domains in *Nn*GH92, a number of N- and C-terminally truncated variants were designed (Table 2). Deletions of the selected domains were designed based on the full-length structure. The inactive variant was created by mutating the general acid Glu944 to Asp, previously demonstrated to suppress the activity of *Bt*3990[9]. The variants were successfully expressed and purified. Based on SDS-PAGE analysis (Figure S1), each deletion affected the structural integrity of *Nn*GH92 to a varying degree. For the variants with the C-terminal deletions (ΔFHB and ΔFHBCBM32) the top band corresponded to the molecular

weight of the enzyme variants, respectively, followed by the two additional bands with a lower molecular weight (Figure S1, lanes 3-4). The N-terminal deletions (ΔN-CBM and core variants) had a much more severe impact, showing only a very low-intensity band corresponding to the full-length variant followed by multiple bands corresponding to proteins with different molecular weight (Figure S1, lanes 5-6). In-gel digest of the selected bands from the SDS-PAGE (Figure S1) followed by mass spectrometry analysis indicated that all the protein bands had the expected molecular weight calculated for that variant (data not shown).

The T_m of the variants was significantly lower, ~10 °C, than that of the wild-type and inactive variant (Table 2). As expected, the activity of the variants was also impacted by the absence of the associated domains (Figure 6A). The activity on the yeast α-mannan was least impaired for the ΔFHB variant, reaching approximately 80% of WT activity at 1 μM enzyme concentration. The deletion of the C-terminal CBM32 (ΔFHBCBM32 and core variants) completely inactivated the enzyme whereas the ΔN-CBM variant retained a striking ~50% activity of the wild-type. The activity results are in keeping with the mass spectrometry data, suggesting that the N-terminal deletion did not affect the structural stability of the catalytic domain to the same extent as the deletion of C-terminal CBM32, which probably exposed crucial structural elements of the catalytic domain, making it more susceptible to proteolytic attack. The same pattern was observed when the activity was tested on α-1,2-mannobiose (Figure S4). Hence, the deletions impacted the overall catalytic efficiency rather than local structural elements that could influence the hydrolysis of more complex polysaccharides, such as yeast α-mannan. Furthermore, the residual activity (25°C, 2 days) of all the variants did not change (Figure S6). The inactive variant did not demonstrate any activity on yeast α-mannan but residual activity was found on α-1,2-mannobiose at high enzyme concentration (1 μM). This is in agreement with results for *Bt*3990 alanine and glutamine variants of the general acid E533 (equivalent to E944 in *Nn*GH92), which substantially impacted the catalytic efficiency against α-1,2-mannobiose but did not completely inactivate the enzyme[9]. Moreover, the core variant did not demonstrate any activity on yeast α-mannan (Figure 6A) and α-1,2-mannobiose (Figure S4). This is interesting when compared to Bt3990 which consists only of the catalytic core without any accessory domain and exhibits high activity on both yeast α-mannan and α-1,2-mannobiose.[9]

To investigate a possible α-mannan binding function for the FHB and CBM32 domains, variants were mixed with insoluble yeast cell wall extracts from *S. cerevisiae* and the amounts of unbound protein were quantified in the supernatant (see Materials and Methods) (Figure 6B, Table S1). Removal of the FHB reduced the population of the bound enzyme by ~50% in comparison to the wild-type. Interestingly, the C-terminal deletion of both FHB and CBM32 increased the amount of bound enzyme. Considering the absence of CBM32 and its complete loss of activity on all the substrates tested including the yeast cell wall (Figure 6A, Figure S7), the ΔFHBCBM32 variant showed unusual behaviour. This may suggest that the C-terminal CBM32 serves the role of maintaining the integrity of the *Nn*GH92 architecture, especially the structural elements involved in its catalytic efficiency. Possibly, such an impaired *Nn*GH92 still preserved sufficient structural stability to perform binding, but probably much less specific and not limited to α-mannan. This is also observed for the core variant where a small population of enzyme was bound to the yeast cell wall extract (Figure 6B). Despite conducting the binding studies at 4 °C to limit the activity of the wild-type and the variants on insoluble yeast cell wall, there is a risk that the binding moieties were partially hydrolysed by the variants which demonstrated activity on yeast cell wall (wild-type, ΔX216) (Figure S7). Therefore, inactive and ΔFHBCBM32 variants might have appeared to show better binding (Figure 6B, Table S1) because yeast cell wall motifs remained intact due to lack of activity of these variants on the yeast cell wall (Figure S7). Interestingly, similar relative activity of *Nn*GH92 variants were detected towards α-1,2-mannobiose (Figure S4) and yeast α-mannan (Figure S6). ΔFHBCBM32 did not exhibit any activity towards either α-1,2-mannobiose nor α-mannan, and hence better binding to the yeast cell wall could be explained by the maintenance of the intact yeast cell wall motif.

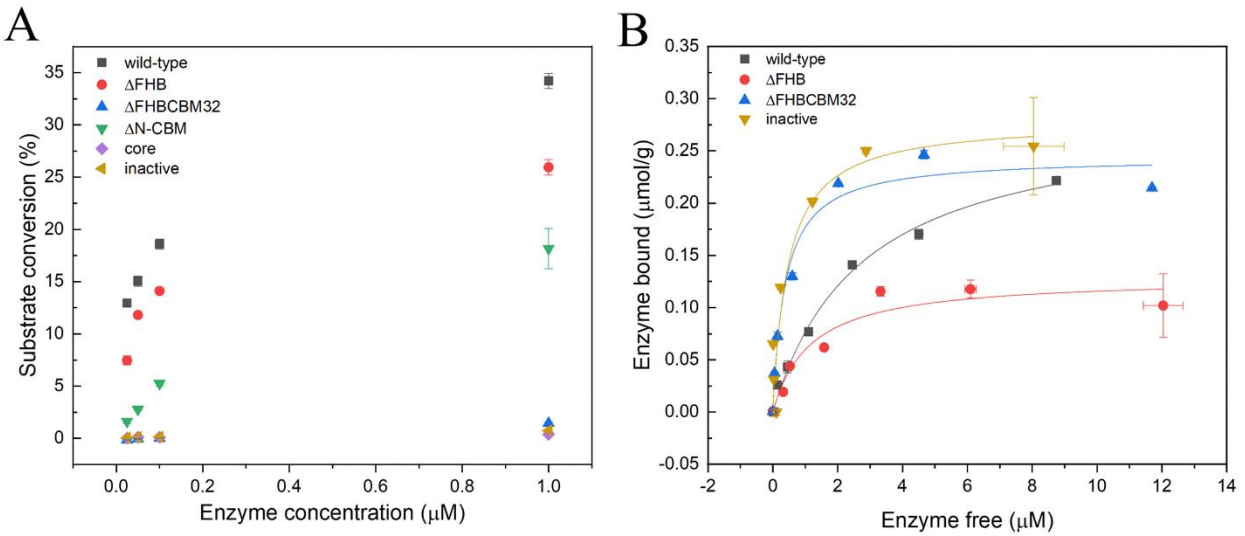

Figure 6. Activity and binding profile of the *Nn*GH92 variants. (A) Yeast α-mannan hydrolysis. The variants, at different enzyme concentration, were mixed with 2.5 g/L yeast α-mannan and incubated for 1 h at 37 °C. The activity was calculated based on the release of the reducing sugar ends (PAHBAH). (B) the variants binding to yeast cell wall extracts (see Materials and Methods). No binding was found for the ΔN-CBM and core variants and control (BSA). Solid lines represent the fitted Langmuir equation. Error bars represent standard deviations from triplicate measurements.

The affinity of the variants to α-mannan was also qualitatively evaluated using native affinity gel electrophoresis[38] (Figure S8). The electrophoretic mobility was reduced for all the variants by the presence of soluble α-mannan in the gel matrix. Compared to a control, the movement of the variants was retarded due to the binding to the polysaccharide. Due to the large molecular weight of *Nn*GH92 (Table 2) and the presence of multiple bands in the SDS-PAGE analysis (Figure S1), the influence of the investigated domains was not obvious. However, the ΔFHBCBM32 variant with deletion of the C-terminal CBM32, seemed to regain affinity since in the presence of α-mannan, the migration of the bands corresponding to this variant were reduced more than for the bands corresponding to the ΔFHB variant. This suggests a stronger affinity for ΔFHBCBM32, consistent with the binding affinity results demonstrated on the yeast cell wall (Figure 6B).

Discussion

The study presented here describes the structural and biochemical investigation of the *Nn*GH92 α-1,2-mannosidase wild-type and domain-deletion variants (Table 2). The catalytic domains of all the GH92 α-1,2-mannosidases in the current CAZy database[8] and the *Nn*GH92 wild-type, have the same characteristic fold with the active site having its binding residues highly conserved in the –1 subsite and more divergent in the +1 subsite, which was previously proposed to determine the α-mannosidase specificity[9,13]. The presence of the binding triad Trp477, Glu997 and His996 at the active site of *Nn*GH92 (Trp88, Glu585, and His584 in *Bt*3990[9], Trp70, Glu541, and His540 in *Sp*GH92[15]), interacting with the mannose residues of the leaving group (+1 subsite), and the biochemical characterization on α-1,2-mannobiose[12] (Figure S3, Figure S4) and fungal *O*-glycans containing mannose residues linked through α-1,2-glycosidic bonds[12] confirmed the classification of *Nn*GH92 as a α-1,2-mannosidase. An attempt to solve the crystal structure of the inactive variant (E944Q) to trap α-mannobiose ligands at the active site failed because only poor-quality crystals that were not suitable for diffraction data collection could be grown.

The structure of full length *Nn*GH92 was solved. The presence of a CBM32 was previously established in the GH92 α-1,2-mannosidases from *Microbacterium* sp. M-90 (aman2)[17] and *Cellulosimicrobium cellulans* (*Cc*GH92)[14], but none of the studies investigated the structural or biochemical influence of these CBM32s on the enzymes. The absence of CBMs in GH92 α-mannosidase from *B. thetaiotaomicron*, suggests the presence of a CBM may be driven by the natural habitat of the host organism.

CBMs glycan specificity is often guided by the catalytic domain specificity to which the CBM is appended[20]. This has been demonstrated for the CBM32s, connected to the catalytic domains of various GH families[18,19,24,39], which sometimes are found in multiple copies within the same enzyme architecture[31]. CBM32 has not been documented to modulate activity of GH92 α-mannosidases against α-mannans. However, it has been found to increase the activity of the GH5 β-mannanase from *Clostridium thermocellum* against insoluble β-mannans.[24] The biochemical studies of the *Nn*GH92 variants did not provide an obvious answer to how the appended domains modulate binding to α-mannooligosaccharides. Since the CBM32 domains were appended to the

403 catalytic domain, the differences in the binding affinity could be hindered by the binding properties
404 of the catalytic domain, especially since CBM32s exhibit low binding affinity (in the range of mM^{-1}
405 and low uM^{-1})[24,30]. However, the presence of ManI at the binding site of the N-terminal CBM32,
406 indicated by the overlay with the CBM32 from *Mv*GH33[32], provides the first structural suggestion
407 of CBM32 binding to mannopyranoside rings. This ligand binding by the CBM32 from *Nn*GH92
408 might not be its only major function. The impaired structural integrity of the catalytic core upon
409 removal of the CBM32s strongly suggests a role for this CBM in protecting structural elements
410 that might be easily accessible for protease attack or simply not fully functional without the
411 appended domains.

412 This study provides the biochemical and the structural investigation of the multi-domain *Nn*GH92
413 α-1,2-mannosidase targeting yeast α-mannan and fungal protein mannose-rich glycans. The
414 structural comparison to *Bt*3990 confirmed the important amino acids involved in the interaction
415 with ManI. A second ManI was bound to the N-terminal CBM32 which allowed the identification
416 of its binding site. This appeared to provide strong evidence for the *Nn*GH92 CBM32's ability to
417 bind to α-mannooligosaccharides, however, it was not possible to demonstrate this in binding
418 studies. A better binding profile might be obtained by studying these CBM32s expressed separately
419 from the catalytic domain. Understanding the role of the non-catalytic domains is important for the
420 future design of more stable and active bacterial GH92 α-1,2-mannosidases. In particular, *Nn*GH92
421 can be optimized for the efficient enzymatic *N*- and *O*-deglycosylation of fungal glycoproteins.[12]

Methods

Materials

α-1,2-mannobiose, α-1,3-mannobiose, α-1,6-mannobiose were purchased from Dextra. All other chemicals, unless otherwise stated, were purchased from Sigma-Aldrich. α-mannan from *S. cerevisiae mnn2* mutant (α-1,6-linked mannan backbone without side chains) was extracted as described previously[37]. The same extraction was done for *S. cerevisiae* WT to use it as a control α-mannan.

Cloning, expression, and purification of *Nn*GH92 wild-type and variants

The data for GH92 α-1,2-mannosidase from *Neobacillus novalis* (*Nn*GH92) was deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number LR963497.1 (GenBank sequence ID). The design of the variants was based on the structure of *Nn*GH92, targeting the linker regions between the catalytic domain and the associated non-catalytic domains or between both associated non-catalytic domains. The following deletions were introduced to the variants – ΔFHB variant: Δ1319-1411; ΔFHBCBM32: Δ1168-1411; ΔN-CBM: Δ34-397; core: Δ34-397 and Δ1168-1411; where the numbers correspond to the deleted range of amino acid residues from the wild-type. All the constructs were verified by sequencing.

The wild-type and the variants were cloned and expressed as extracellular enzymes in *Bacillus subtilis* in a similar setup as described previously[40] with the following modifications. The native signal peptide was replaced by the Alcalase signal peptide followed by a histidine tag (6xHis + PR) resulting in a N-terminal sequence of MKKPLGKIVASTALLISVAFSSSIASAHHHHHHPR. The fermentation broth was sterile filtrated and then 500 mM NaCl was added and adjusted to pH 7.5/NaOH. The sample was loaded onto a Ni-Sepharose™ 6 Fast Flow column (GE Healthcare, Piscataway, NJ, USA) equilibrated in 50 mM HEPES, pH 7.5 with 500 mM NaCl (buffer A). After loading, the column was washed with 10 CV of buffer A, and bound proteins were eluted with 500 mM imidazole in buffer A. The fractions containing the enzyme were pooled and applied to a Sephadex™ G-25 (medium) (GE Healthcare, Piscataway, NJ, USA) column equilibrated and

eluted in 50 mM HEPES pH 7.5. Fractions were analysed by SDS-PAGE, and fractions containing the enzyme were combined. Protein concentrations were determined by measuring absorption at 280 nm with a spectrophotometer (NanoDrop 8000, Thermo Scientific) using extinction coefficients based on the amino acid sequence of *Nn*GH92 enzymes. The identity of the purified enzymes was verified by excising the protein bands from SDS-PAGE gel (Figure S1) and analysing a tryptic digest with mass spectrometry.

Enzyme assay conditions

All enzyme activity assays were conducted in an assay buffer composed of 50 mM MES pH 6.0, 50 mM NaCl, 2 mM $CaCl_2$, 0.01 % Triton X-100 unless otherwise stated. Enzymatic hydrolysates were quenched with 0.15 M NaOH and analysed with a reducing sugar assay (PAHBAH) to quantify reducing sugar ends[41] or HPAEC-PAD to quantify the released mannose concentration. The detailed experimental procedure is described elsewhere[42]. The absorption of the coloured products was measured at 405 nm using a plate reader (Spectra Max 3; Molecular Devices). The absorbance readouts were recalculated to reducing ends' concentration using a mannose standard curve (0-5 mM). The substrate conversion was calculated as the actual yield / theoretical yield × 100%. The hydrolysis of yeast α-mannan by *Nn*GH92 resulted in DP1 products and only mannose was identified in HPAEC-PAD. Thus, the measured sugar reducing ends concentration was assumed to be equal to the mannose reducing ends (actual yield). The yeast α-mannan was acid hydrolysed and the released monosaccharides were quantified with HPAEC-PAD as described elsewhere[43] and the mannose concentration was calculated. This concentration was divided by the initial yeast α-mannan concentration corrected for the monomeric units (180/162) to obtain the theoretical yield. The sample analyses using HPAEC-PAD followed the procedure described earlier[43].

Thermal stability

The studied enzymes were analysed by nano differential scanning fluorimetry (nanoDSF, Prometheus NT.48, NanoTemper) to determine the melting temperature (T_m). The enzymes were diluted to a concentration of 2 mg/mL in 50 mM MES pH 6.0. Thermal stability was tested with a

heating scan range from 20 to 90 °C at a scan rate of 2 °C/min. The data analysis and calculation of T_m were done using the PR.ThermControl (NanoTemper) software.

pH and temperature optima

The pH profile was calculated by enzymatic hydrolysis of 5 mg/mL α-mannan solubilized in assay buffers with different buffer components: 50 mM sodium acetate (pH 3.6, 4.0, 5.0) 50 mM MES (pH 6.0), 50 mM HEPES (pH 7.0, 8.0) and TRIS (pH 9.0); with 13 nM *Nn*GH92 over incubation time of 2 h at 25 °C. Enzymatic hydrolysates were withdrawn at different time points (15, 30, 45, 75 and 120 min) and analysed with the reducing sugar assay (PAHBAH). Based on the absorbance measurement at 405 nm, the linear range of reaction progress curve was calculated, and the highest activity was set to 1 and the rest normalized with the same factor.

The temperature profile was calculated at five different points: 20, 33, 42, 52 and 60 °C; using the assay buffer pH 6. The other calculations were done following the same procedure as for the pH profile.

*Nn*GH92 wild type kinetics with α-mannobiose

The kinetic constants (k_{cat} and K_M) for *Nn*GH92 wild-type were determined by assaying the initial hydrolysis rate at 37 °C on different α-1,2-mannobiose concentrations. The release of mannose was quantified using Megazyme International kit for D-mannose assay kit (Megazyme, K-MANGL) and mannose standard curve. The corresponding substrate concentrations were prepared by dissolving α-1,2-mannobiose in the assay buffer (see Enzyme assay conditions). The enzyme concentration used for the assay was 7 nM. The initial rates were plotted as a function of α-1,2-mannobiose concentrations and fitted with the Michaelis-Menten equation. A similar experiment was attempted for α-1,3-mannobiose except for using 66 nM *Nn*GH92 wild-type, however, it was not possible to fit the Michaelis-Menten equation due to insufficient initial rate points.

Thin layer chromatography (TLC)

A few droplets (3-4 µL) of the completed hydrolysis reaction mixture by *Nn*GH92 were spotted on silica gel TLC plates (stationary phase). Once the sample spots dried, the plates were immersed in

a solution of butanol:acetic acid:water mixed in the ratio 2:1:1 (mobile phase). The plate was developed until the mobile phase reached 80-90% of the full height of the plate. The plate with the separated components was dried with a hot air gun and carbohydrates were detected by immersing the plate in chemical stain (5% ammonium molybdate 0.02% cerium sulfate, 5% sulfuric acid). After a few seconds, the plate was removed and again dried until the development of blue bands.[7] Standards were prepared by solubilizing 1 mg/mL β-mannooligosaccharides (Megazyme) with a degree of polymerization of 2-6 and 1 mg/mL mannose (Sigma). The sample's DP was estimated by comparing the sample's bands to the lane with sugar standards.

Crystallization

Crystallization experiments were carried out in the presence or absence of 8 mM $CaCl_2$ and 5 mM mannoimidazole. Hits were only obtained for the mannoimidazole complex with $CaCl_2$, in PACT premier™ HT-96 (Molecular Dimensions), conditions B7 (0.2 M NaCl, 0.1 M MES pH 6.0, 20% PEG6K) and E6 (0.2 M Na-formate, 20% PEG 3350). The crystals were imperfect and were used to make seeding stock. The seeding stock was prepared and microseed matrix screening (MMS, recent review in [44]) carried out using an Oryx robot (Douglas instruments) according to published protocols[45,46]. Briefly, crystals were crushed and diluted with ~50 µl of mother liquor. The solution was transferred into a seed bead containing reaction tube and vortexed for three minutes. The seeding stock was used straight away, and any remaining seeds were frozen and kept at -20°C. MMS was carried out in the PACT screen, giving in increased number of better hits. Crystals from condition F11 (0.2 M Na-citrate, 0.1 M Bis-Tris-propane pH 6.5, 20% PEG 3350), were used to make a seeding stock for the next seeding rounds, into optimization screens based on the successful conditions, with different seed dilutions. The crystallization drops contained 150 nl protein (including 8 mM $CaCl_2$ and 5 mM mannoimidazole) + 50 nl seeding stock + 100 nl mother liquor from a new random screen. The final crystal was obtained in 21% PEG3350, 0.1 M BTP pH 6.6, 0.2 M Na-citrate.

Data collection, structure solution and refinement

All computation was carried out using programs from the CCP4 suite[47], unless otherwise stated. Data to 2.3 Å resolution were collected at the Diamond Light Source, beamline I04-1, integrated

using XDS[48] within the Xia2 pipeline[49] and scaled with Aimless[50]. The space group was $P2_1$, a=94.61, b=151.940, c=114.01Å, β=94.63°. The structure was solved by molecular replacement using Molrep[51] with 2wzs (Family GH92 Inverting Mannosidase *Bt*3990 from *Bacteroides thetaiotaomicron* VPI-5482 in complex with Mannoimidazole) as a model. 60 cycles of jelly body refinement with Refmac[52] were followed by density modification with Parrot[53], and the initial model was built with Buccaneer[54]. Further refinement was carried out with Refmac with the TLS option iterated alternated with manual model correction in Coot[55]. The quality of the final model was validated using Molprobity[56] as part of the Phenix package[57]. The final data processing and refinement processing statistics are given in (Table 1).

Yeast cell wall extraction

Yeast cell walls were extracted from *S. cerevisiae* cells grown in sterile YPD medium for 3 days at 32 °C, 150 rpm. The yeast cells were harvested and washed 3 times with cold deionized water by centrifugation at 4000 rpm for 10 min. Yeast cells were diluted in 10 mM Tris-HCl pH 8.0 at a concentration of 50 mg cell wet mass/mL. Extraction of yeast cell walls was conducted with a cell disruptor (CF1 model, Constant Systems) at a pressure of 18 kpsi. Four passages were applied to ensure complete disruption of the yeast cells. Subsequently, the extracted cell walls were pelleted at 3800 g for 5 min, washed with cold water until the supernatant became clear and stored at 4 °C.[58] Then the yeast cell walls were washed three times and resuspended in 50 mM MES pH 6.0, 50 mM NaCl, 2 mM $CaCl_2$. The final concentration of the yeast cell wall stock was calculated as a dry cell weight and used for the activity and binding assays.

Native affinity gel electrophoresis

The ability of *Nn*GH92 variants to bind to a soluble yeast α-mannan was evaluated by the native affinity gel electrophoresis. The materials, assay, and the data analysis was according to the protocol demonstrated elsewhere[59] with the following changes: the gel was composed of 10% acrylamide and 0.1% yeast α-mannan in 50 mM Tris pH 8.7. Each lane was loaded with 4 µg of enzyme. Both a control and polysaccharide gels were run in 50 mM Tris pH 8.7 for 20h at 4 °C, with a constant 75V. In the control gel, the yeast α-mannan was substituted with 50 mM Tris pH 8.7. The gels were prepared without the stacking layer.

Yeast cell wall binding and activity assay

*Nn*GH92 variants at different enzyme concentration were mixed with 20 g/L (dry cell weight, DWC) insoluble yeast cell wall extract from *S. cerevisiae* solubilized in the assay buffer, except for 0.01 % Triton X-100, and equilibrated for 1 h at +4 °C, 1110 rpm. The resulting mixtures were centrifuged (16800 rpm) at +4 °C and the amounts of unbound protein were obtained using a spectrophotometric method, measuring absorbance at 280 nm. The calculation of free enzyme and bound enzyme and the fitting with the Langmuir isotherm was performed as described elsewhere[60]. The activity assays of *Nn*GH92 variants were performed on the same substrate. The enzymes at two enzyme concentration 0.1 and 1 μM were mixed with 60 g/L DWC extracted yeast cell wall solubilized in the assay buffer and incubated at 37 °C for 1 h. The hydrolysates were analysed with the reducing sugar assay (PAHBAH) and absorbance measurements were recalculated to mannose concentration using a mannose standard curve.

Data Availability Statement

The data for a GH92 α-1,2-mannosidase from *Neobacillus novalis* (*Nn*GH92) was deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number LR963497.1 (GenBank sequence ID). The structure and coordinates files for *Nn*GH92 in complex with mannoimidazole were deposited in the Protein Data Bank under accession code 7NSN. All other data are included in the main article and supplementary materials.

Acknowledgements

The authors thank the Diamond Light Source for access to beamline I04-1 (proposal number mx-mx-18598) that contributed to the results presented here. The authors thank Dr Johan Turkenburg and Sam Hart for assistance during data collection.

Funding

This work has been supported by Roskilde University, Novozymes A/S, Innovation Fund Denmark [Grant number: 5150-00020B], the Novo Nordisk Foundation [Grant number: NNF15OC0016606 and NNFSA170028392] and the Carlsberg Foundation. GJD is supported by the Royal Society "Ken Murray" Research Professorship.

Conflict of interests

Bartłomiej M. Kołaczkowski, Kristian B. R. M. Krogh and Kenneth Jensen work for Novozymes A/S, a major manufacturer of industrial enzymes. Work at York was supported by funding from Novozymes A/S.

Author contributions

BMK conducted the overall study, performed experiments, wrote the manuscript. OVM and EB performed crystallization, data collection, structure solution and refinement and interpretation. GJD and KSW obtained funding, supervised OVM and EB, provided experimental design and interpretation and aided in manuscript drafting and illustration. MSM supervised data acquisition for native affinity gel electrophoresis and provided suggestions for substrate affinity experiments. ASM, PW, KJ, KBRMK, conceived the study, supervised many aspects of the work, and participated in manuscript preparation. All authors read and approved the final manuscript.

References

1. Abbott, D. W., Martens, E. C., Gilbert, H. J., Cuskin, F. & Lowe, E. C. Coevolution of yeast mannan digestion: Convergence of the civilized human diet, distal gut microbiome, and host immunity. *Gut Microbes* **6**, 334–339 (2015).

2. Orlean, P. Architecture and biosynthesis of the Saccharomyces cerevisiae cell wall. *Genetics* **192**, 775–818 (2012).

3. Hakki, Z. *et al.* Structural and kinetic dissection of the endo-α-1,2-mannanase activity of bacterial GH99 glycoside hydrolases from Bacteroides spp. *Chem. - A Eur. J.* **21**, 1966–1977 (2015).

4. Goto, M. Protein O-glycosylation in fungi: Diverse structures and multiple functions. *Biosci. Biotechnol. Biochem.* **71**, 1415–1427 (2007).

5. Varki, A. *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323–1324 (2015).

6. Neelamegham, S. *et al.* Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **29**, 620–624 (2019).

7. Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165–169 (2015).

8. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).

9. Zhu, Y. *et al.* Mechanistic insights into a Ca2+-dependent family of α-mannosidases in a human gut symbiont. *Nat. Chem. Biol.* **6**, 125–132 (2010).

10. Liu, F. F., Kulinich, A., Du, Y. M., Liu, L. & Voglmeir, J. Sequential processing of mannose-containing glycans by two α-mannosidases from Solitalea canadensis. *Glycoconj. J.* **33**, 159–168 (2016).

11. Li, Y. *et al.* Enterococcus faecalis α1–2-mannosidase (EfMan-I): an efficient catalyst for glycoprotein N-glycan modification. *FEBS Lett.* **594**, 439–451 (2020).

12. Kołaczkowski, B. M. *et al.* Analysis of fungal high mannose structures using CAZymes. *Glycobiology* (2021) doi:10.1093/GLYCOB/CWAB127.

13. Thompson, A. J. *et al.* Bacteroides thetaiotaomicron generates diverse α-mannosidase activities through subtle evolution of a distal substrate-binding motif. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 394–404 (2018).

14. Tiels, P. *et al.* A bacterial glycosidase enables mannose-6-phosphate modification and improved cellular uptake of yeast-produced recombinant human lysosomal enzymes. *Nat.*

Biotechnol. **30**, 1225–1231 (2012).

15. Robb, M. *et al.* Molecular Characterization of N-glycan Degradation and Transport in Streptococcus pneumoniae and Its Contribution to Virulence. *PLOS Pathog.* **13**, e1006090 (2017).

16. Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**, 557–559 (1997).

17. Maruyama, Y., Nakajima, T. & Ichishima, E. A 1,2-α-d-mannosidase from a Bacillus sp.: purification, characterization, and mode of action. *Carbohydr. Res.* **251**, 89–98 (1994).

18. Newstead, S. L., Watson, J. N., Bennet, A. J. & Taylor, G. Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 1483–1491 (2005).

19. Ficko-Blean, E. *et al.* Carbohydrate recognition by an architecturally complex α-N-acetylglucosaminidase from Clostridium perfringens. *PLoS One* **7**, (2012).

20. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781 (2004).

21. Ficko-Blean, E. & Boraston, A. B. The interaction of a carbohydrate-binding module from a Clostridium perfringens N-acetyl-β-hexosaminidase with its carbohydrate receptor. *J. Biol. Chem.* **281**, 37748–37757 (2006).

22. Boraston, A. B., Ficko-Blean, E. & Healey, M. Carbohydrate recognition by a large sialidase toxin from Clostridium perfringens. *Biochemistry* **46**, 11352–11360 (2007).

23. Lyu, Q. *et al.* Structural and biochemical characterization of a multidomain alginate lyase reveals a novel role of CBM32 in CAZymes. *Biochim. Biophys. Acta - Gen. Subj.* **1862**, 1862–1869 (2018).

24. Mizutani, K. *et al.* Influence of a mannan binding family 32 carbohydrate binding module on the activity of the appended mannanase. *Appl. Environ. Microbiol.* **78**, 4781–4787 (2012).

25. Heyrman, J. *et al.* Bacillus novalis sp. nov., Bacillus vireti sp. nov., Bacillus soli sp. nov., Bacillus bataviensis sp. nov. and Bacillus drentensis sp. nov., from the Drentse A grasslands. *Int. J. Syst. Evol. Microbiol.* **54**, 47–57 (2004).

26. Patel, S. & Gupta, R. S. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus bacillus: Proposal for six new genera of bacillus species, peribacillus gen. nov., cytobacillus gen. nov., mesobacillus gen. nov., neobacillus gen. nov., metabacillus gen. nov. and alkalihalobacillus gen. nov. *Int. J. Syst. Evol. Microbiol.* **70**, 406–438 (2020).

27. Ly, H. D. & Withers, S. G. Mutagenesis of Glycosidases. *Annu. Rev. Biochem.* **68**, 487–522 (1999).

28. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science*

(80-.). **336**, 1030–1033 (2012).

29. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 386–394 (2011).

30. Ficko-Blean, E. & Boraston, A. B. N-Acetylglucosamine Recognition by a Family 32 Carbohydrate-Binding Module from Clostridium perfringens NagH. *J. Mol. Biol.* **390**, 208–220 (2009).

31. Abbott, D. W., Eirín-López, J. M. & Boraston, A. B. Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules. *Mol. Biol. Evol.* **25**, 155–167 (2008).

32. Gaskell, A., Crennell, S. & Taylor, G. The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* **3**, 1197–1205 (1995).

33. Krissinel, E. Desktop and Web-based Gesamt Software for Fast and Accurate Structural Queries in the PDB. *J. Comput. Sci. Appl. Inf. Technol.* **2**, 1–7 (2017).

34. Roske, Y., Sunna, A., Pfeil, W. & Heinemann, U. High-resolution crystal structures of Caldicellulosiruptor strain Rt8B.4 carbohydrate-binding module CBM27-1 and its complex with mannohexaose. *J. Mol. Biol.* **340**, 543–554 (2004).

35. Johnsen, L., Dalhus, B., Leiros, I. & Nissen-Meyer, J. 1.6-Å crystal structure of EntA-im: A bacterial immunity protein conferring immunity to the antimicrobial activity of the pediocin-like bacteriocin enterocin A. *J. Biol. Chem.* **280**, 19045–19050 (2005).

36. Ficko-Blean, E. *et al.* Portrait of an enzyme, a complete structural analysis of a multimodular β-N-acetylglucosaminidase from Clostridium perfringens. *J. Biol. Chem.* **284**, 9876–9884 (2009).

37. Raschke, W. C., Kern, K. A., Antalis, C. & Ballou, C. E. Genetic control of yeast mannan structure. Isolation and characterization of mannan mutants. *J. Biol. Chem.* **248**, 4660–4666 (1973).

38. Cockburn, D. *et al.* Using carbohydrate interaction assays to reveal novel binding sites in carbohydrate active enzymes. *PLoS One* **11**, (2016).

39. Rao, F. V. *et al.* Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *EMBO J.* **25**, 1569–1578 (2006).

40. Jensen, K., Østergaard, P. R., Wilting, R. & Lassen, S. F. Identification and characterization of a bacterial glutamic peptidase. *BMC Biochem.* **11**, 47 (2010).

41. Lever, M. Calorimetric Determination and Acid Fluorometric with Hydrazide. *Anal. Biochem.* **281**, 274–281 (1973).

42. Sørensen, T. H. *et al.* Temperature effects on kinetic parameters and substrate affinity of

Cel7A cellobiohydrolases. *J. Biol. Chem.* **290**, 22193–22202 (2015).

43. Schiano-di-Cola, C. *et al.* Structural and biochemical characterization of a family 7 highly thermostable endoglucanase from the fungus Rasamsonia emersonii. *FEBS J.* **287**, 2577–2596 (2020).

44. D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W. & Marsh, M. Microseed matrix screening for optimization in protein crystallization: What have we learned? *Acta Crystallogr. Sect. Struct. Biol. Commun.* **70**, 1117–1126 (2014).

45. Shah, A. K. *et al.* On increasing protein-crystallization throughput for X-ray diffraction studies. *Acta Crystallogr. Sect. D* **61**, 123–129 (2005).

46. Shaw Stewart, P. D., Kolek, S. A., Briggs, R. A., Chayen, N. E. & Baldock, P. F. M. Random microseeding: A theoretical and practical exploration of seed stability and seeding techniques for successful protein crystallization. *Cryst. Growth Des.* **11**, 3432–3441 (2011).

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

48. Kabsch, W. Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–32 (2010).

49. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1260–1273 (2013).

50. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).

51. Vagin, A. & Teplyakov, A. Molecular replacement with {\it MOLREP}. *Acta Crystallogr. Sect. D* **66**, 22–25 (2010).

52. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).

53. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 470–478 (2010).

54. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1002–1011 (2006).

55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

56. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).

57. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).

58. Dallies, N., Francois, J. & Paquet, V. A new method for quantitative determination of polysaccharides in the yeast cell wall. Application to the cell wall defective mutants of

738 Saccharomyces cerevisiae. *Yeast* **14**, 1297-1306. (1998).

739 59. Cockburn, D., Wilkens, C. & Svensson, B. Affinity Electrophoresis for Analysis of Catalytic
740 Module-Carbohydrate Interactions. in *Protein-Carbohydrate Interactions: Methods and*
741 *Protocols* (eds. Abbott, D. W. & Lammerts van Bueren, A.) 119–127 (Springer New York,
742 2017). doi:10.1007/978-1-4939-6899-2_9.

743 60. Kołaczkowski, B. M. *et al.* Removal of N-linked glycans in cellobiohydrolase Cel7A from
744 Trichoderma reesei reveals higher activity and binding affinity on crystalline cellulose.
745 *Biotechnol. Biofuels* **13**, 136 (2020).

746

Supporting information for

Structural and functional characterization of a multi-domain GH92 α-1,2-mannosidase from *Neobacillus novalis*

Bartłomiej M. Kołaczkowski[1,4], Olga V. Moroz[2], Elena Blagova[2], Gideon J. Davies[2], Marie Sofie Møller[3], Anne S. Meyer[3], Peter Westh[3], Kenneth Jensen[4], Keith S. Wilson[*, 2], Kristian B.R.M. Krogh[*,4]

[1] Roskilde University, Department of Science and Environment, Universitetsvej 1, Building 28, 4000 Roskilde, Denmark
[2]York Structural Biology Laboratory, Department of Chemistry, University of York, York, YO10 5DD, UK
[3]Technical University of Denmark, Department of Biotechnology and Biomedicine, Building 224, 2800 Kongens Lyngby, Denmark
[4]Novozymes A/S, Biologiens Vej 2, 2800 Kongens Lyngby, Denmark

[*]correspondence: kbk@novozymes.com
[*]correspondence: keith.wilson@york.ac.uk

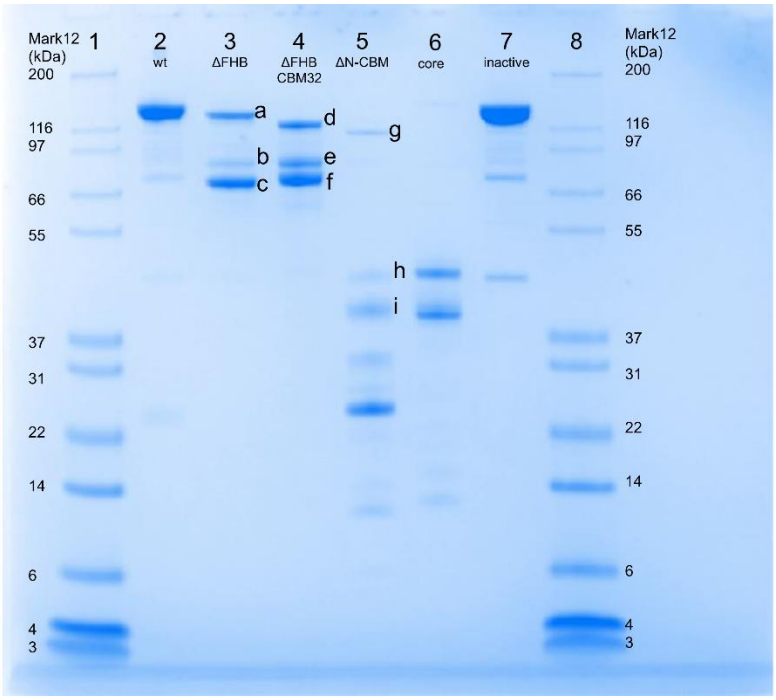

Figure S1. SDS-PAGE gel of the purified *Nn*GH92 wild-type and variants used in this study. Lane 1, 8: Molecular weight standard (Mark12, Invitrogen); lane 2: *Nn*GH92 wild-type; lane 3: ΔFHB; lane 4: ΔFHBCBM32; lane 5: ΔN-CBM32, lane 6: core; lane 7: inactive (E944Q) The gel was stained with Coomassie Blue and 1 µg of each enzyme was loaded on the gel. The bands indicated by the letters (a-i) were excised and digested with trypsin. The isolated peptides were analyzed with mass spectrometry and matched with the protein sequence of the corresponding *Nn*GH92 variants. The bottom bands (b, c, e, f) corresponding to the C-terminal deletion variants (ΔFHB and ΔFHBCBM32), were randomly truncated from the protein C-terminus with poor coverage in the catalytic domain region. The bottom bands (h, i) of the N-terminal deletion variant (ΔCBM) maintained good protein coverage within the catalytic domain. No common cleavage pattern was identified.

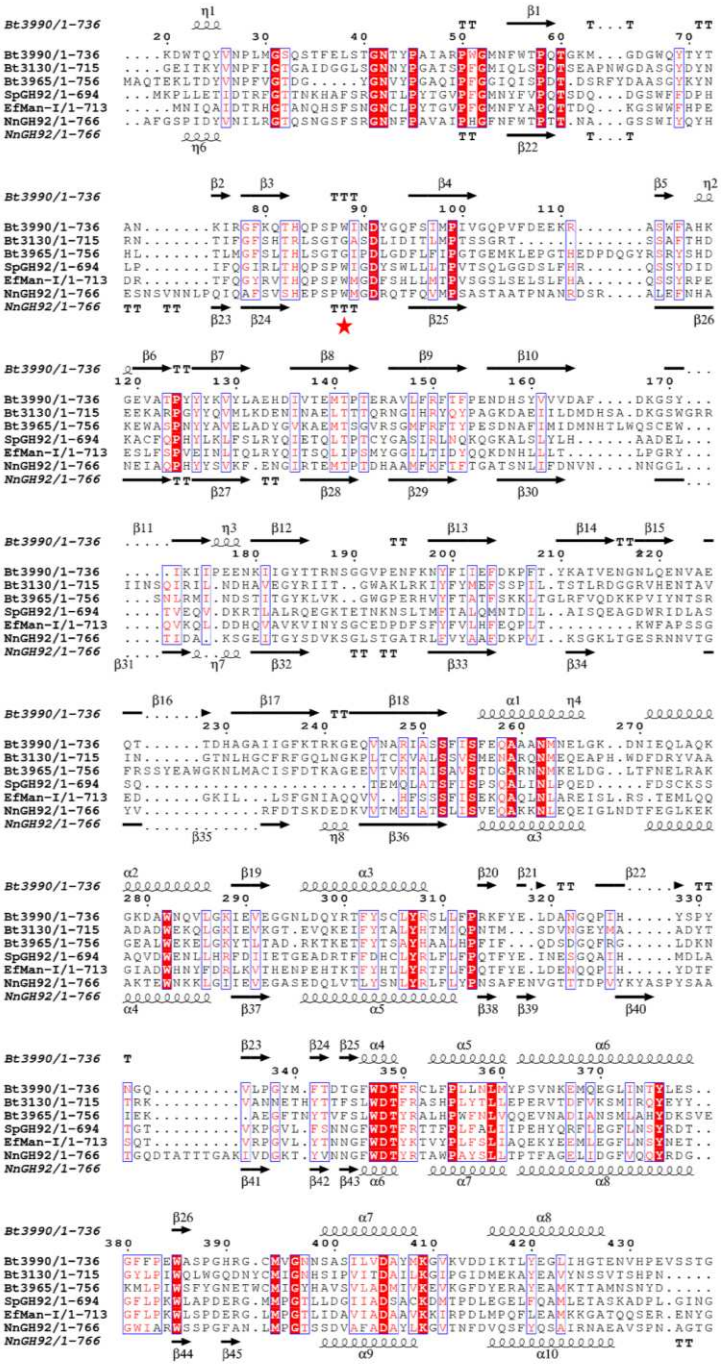

Figure S2. Multiple sequence alignment of catalytic domains from 6 GH92 α-mannosidases with experimentally determined structures. Secondary structure elements for *Bt*3990 (PDB 2WZS) and *Nn*GH92 (PDB: 7NSN) are shown on top and bottom of the alignment, respectively. The numbering is based on the sequence of *Bt*3990. The general acid (E533) and Brønstead base (D644) are marked with a blue star whereas the residues driving specificity of α-1,2-mannosidase are marked with a red star (W88, H584-E585). Identical residues are marked in white characters on a red background. Highly similar residues are framed in a blue box. The alignment was created with MUSCLE algorithm[1] and the figure prepared using ESPript 3.0 web server with default parameters[2].

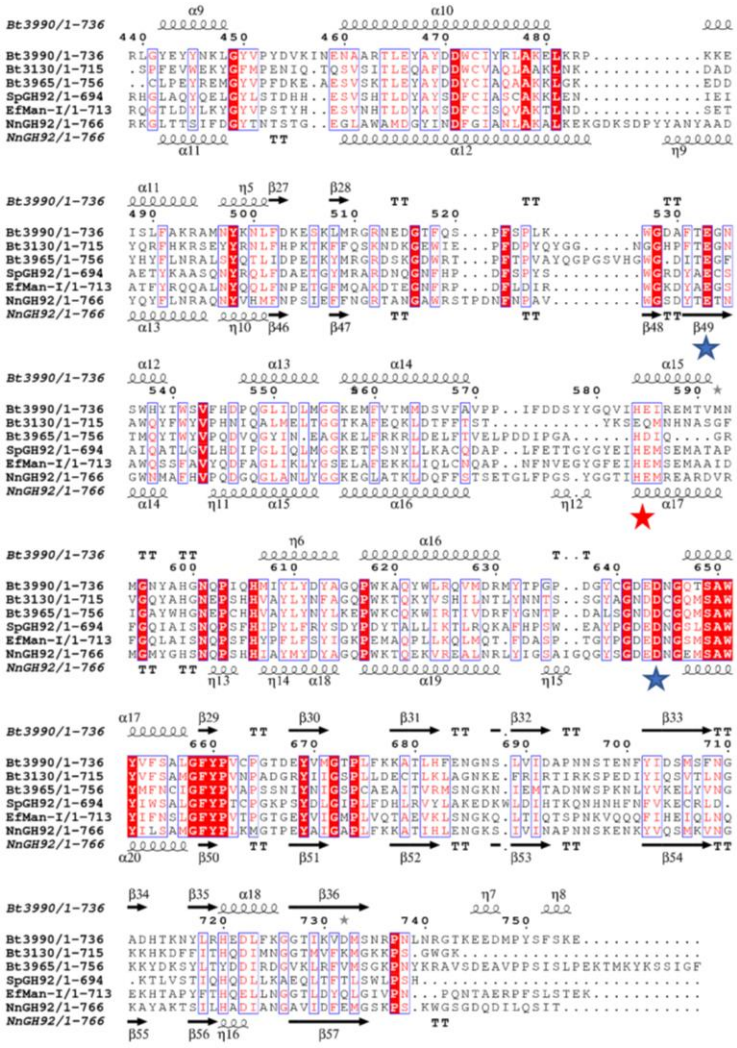

Figure S2. Continued

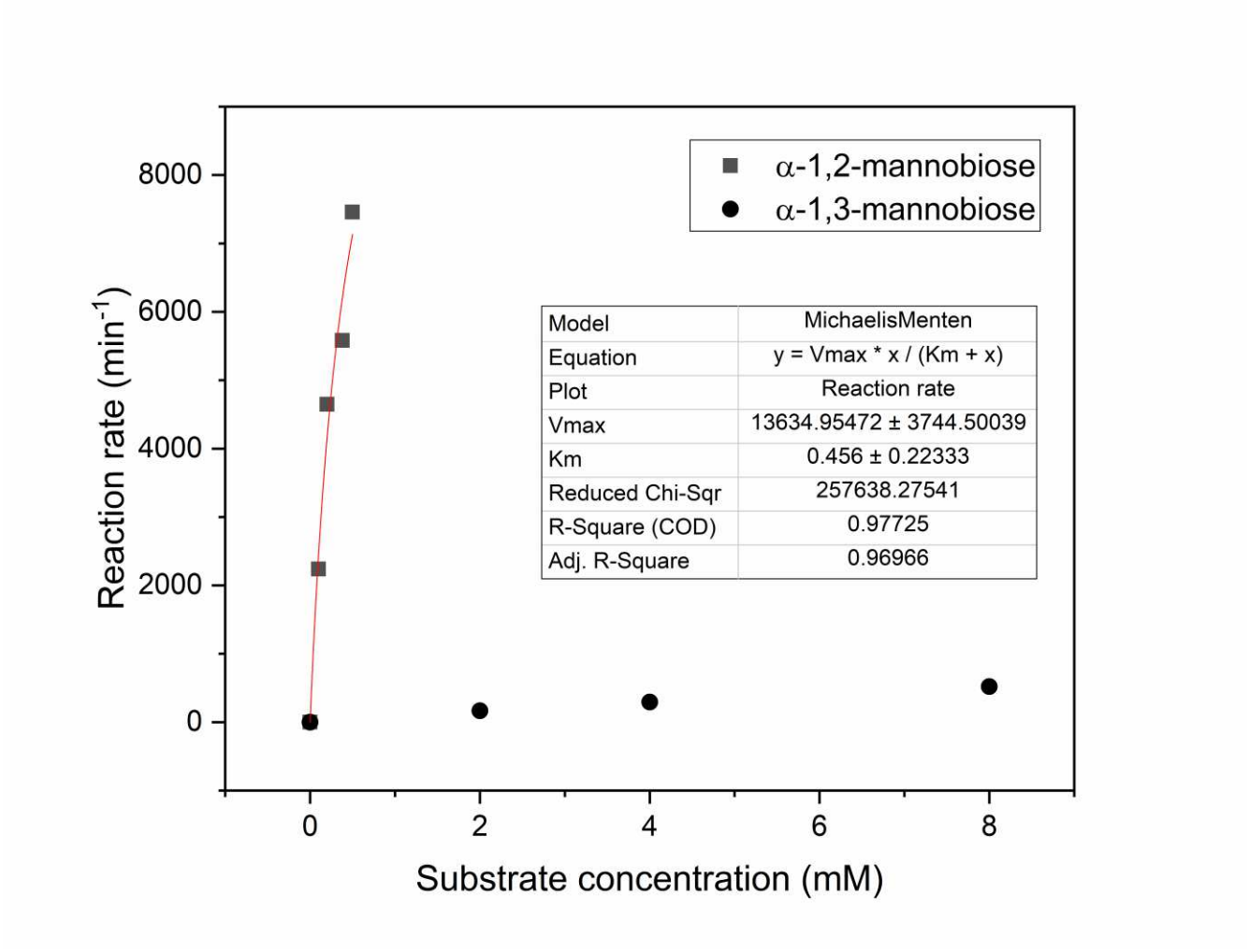

Figure S3. *Nn*GH92 Michaelis-Menten kinetics on α-1,2-mannobiose and α-1,3-mannobiose. Each dot represents initial hydrolysis rate measured at different concentrations of substrates. Mannose release was quantified by Megazyme International kit for D-mannose assay kit using mannose standard curve. Solid lines for α-1,2-mannobiose represent a non-linear fit of Michaelis-Menten equation.

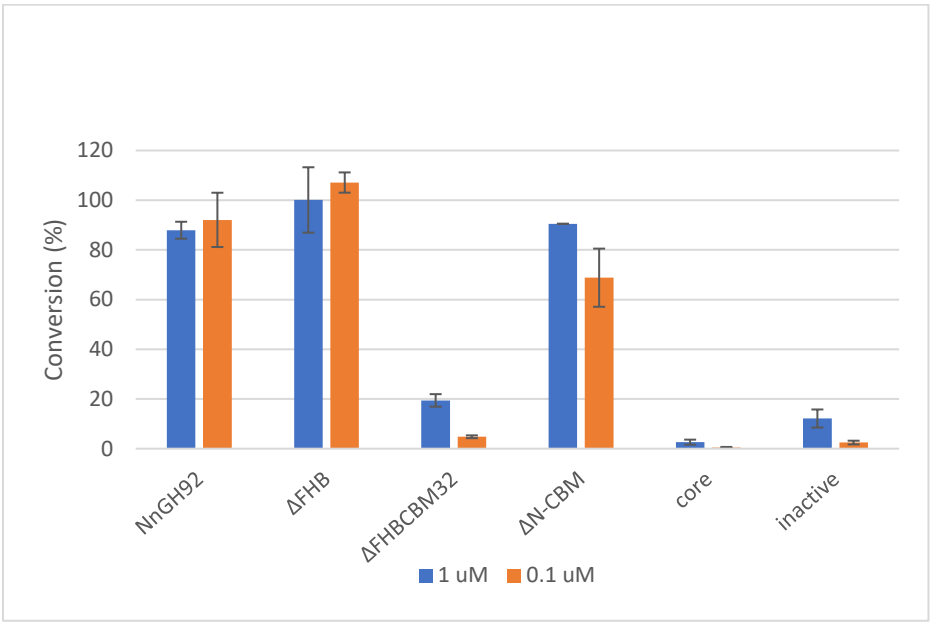

Figure S4. Activity profile of the *Nn*GH92 variants on 1.46 mM α-1,2-mannobiose dissolved in the assay buffer.

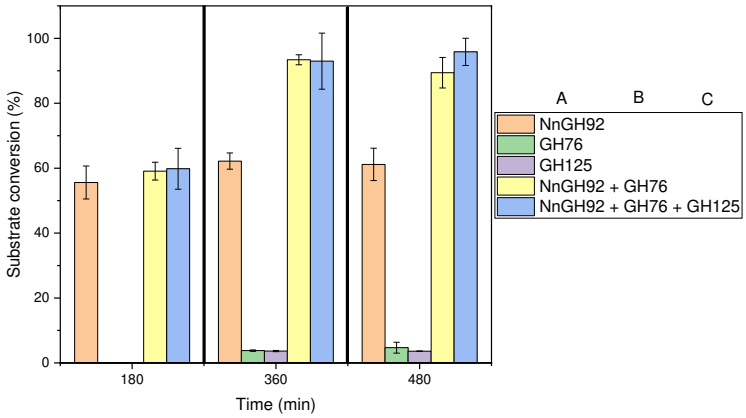

Figure S5. Yeast α-mannan hydrolysis by *Nn*GH92 (E_0 = 0.13 µM), the GH76 endo-α-1,6-mannanase (E_0 = 0.03 µM) and the GH125 exo-α-1,6-mannosidase (E_0 = 0.34 µM). The yeast α-mannan (S_0 = 2.5 mg/mL) was dissolved in the assay. The reaction was initiated by adding the enzymes in the order of t1=0 min (A), t2=180 min (B), t3=360 min (C), as shown in the figure legend. The blank spaces under the corresponding letter indicate the addition of an equivalent volume of the assay buffer. The extent of hydrolysis was followed by reducing sugar assay (PAHBAH).

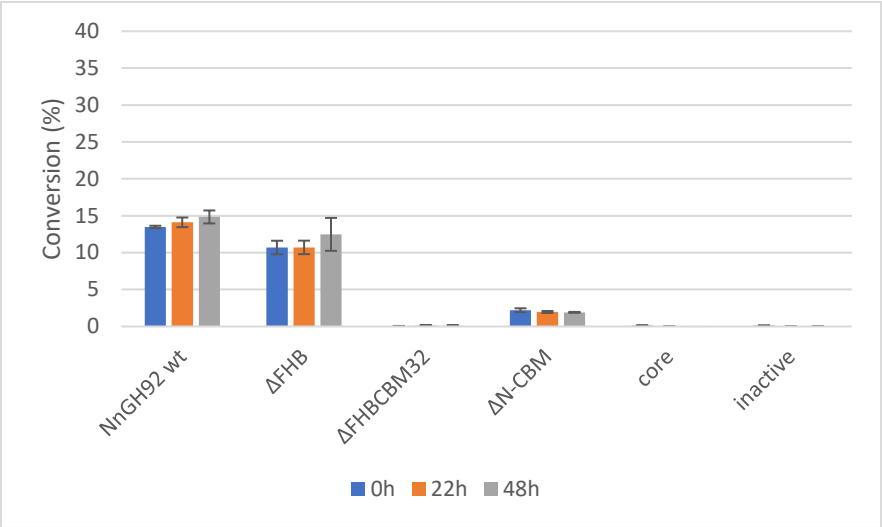

Figure S6. Residual activity of the *Nn*GH92 variants towards yeast alpha-mannan. The enzymes (0.1 µM) were incubated for 0h, 22h and 48h at 25°C. Then at each time point, enzymes were mixed with 2.5 g/L α-mannan in assay buffer and incubated for 15 min. The enzyme activity was verified using reducing sugar assay (PAHBAH). No change in the residual activity for each tested enzyme was observed.

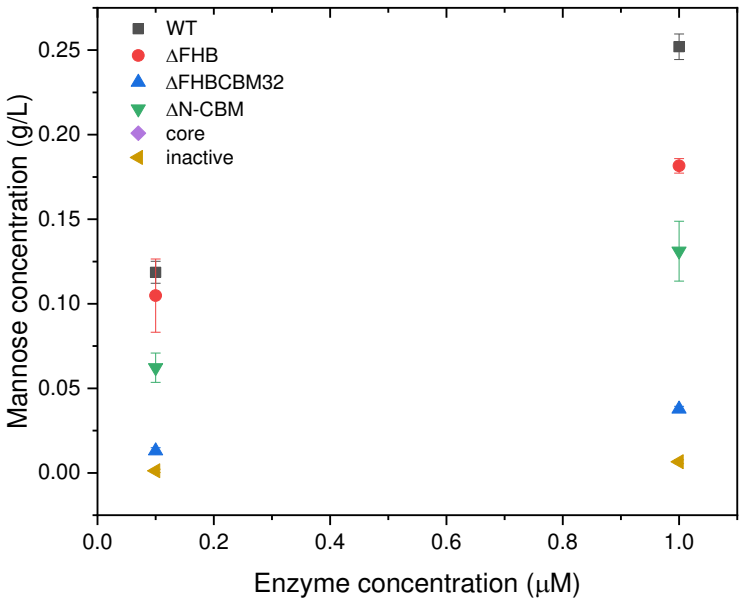

Figure S7. Activity profile of the *Nn*GH92 variants (1 h, 37°C) on the yeast cell wall (60 g/L dry weight cell in assay buffer) extracted from *S. cerevisiae*.

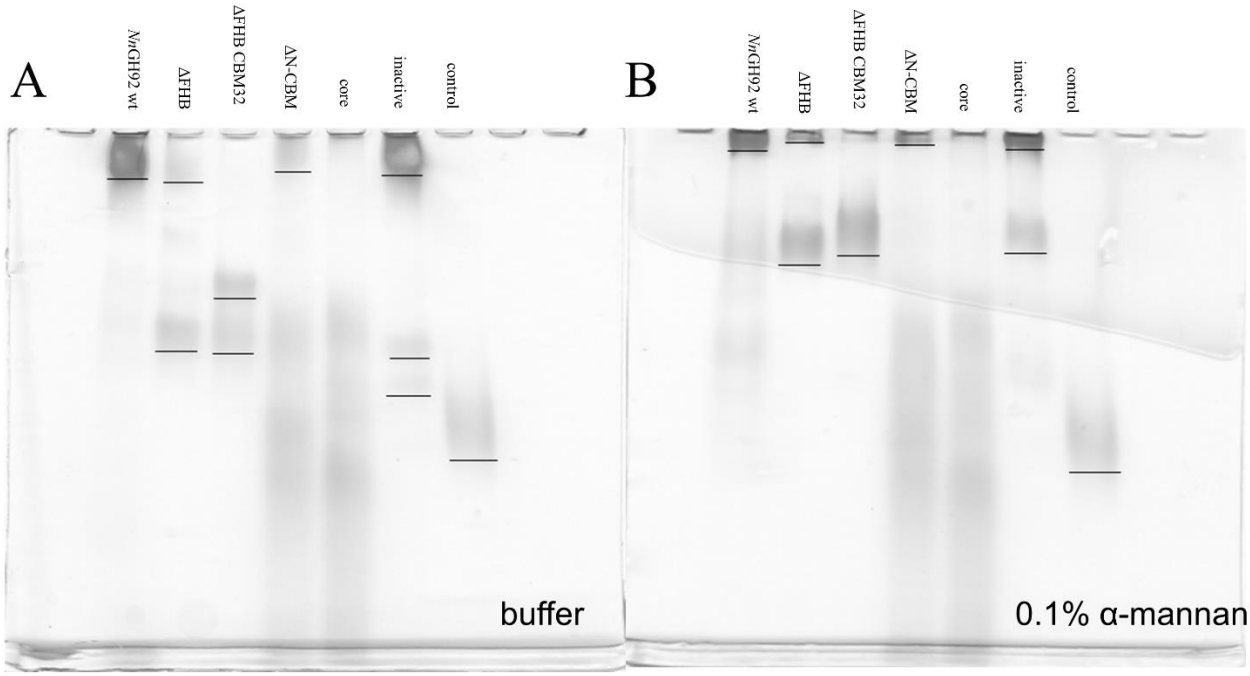

Figure S8. Native affinity gel electrophoresis of the *Nn*GH92 wild-type and variants. The enzymes were run in the control gel (A) and in the gel with yeast α-mannan (B). The migration of the bands corresponding to the enzymes were compared between two gels. The lane 'control' contained GH125 α-1,6-mannosidase used as a control protein due to it is high molecular weight.

Table S1. Derived parameters from the binding isotherm of *Nn*GH92 variants in Figure 6B. The ± values correspond to the error of non-linear fit of binding isotherm curves.

***Nn*GH92**	Γ_{max} µmol/g	K_d µM	**Statistics** Adj. R-Square
wild-type	0.29 ± 0.02	2.74 ± 0.42	0.99
ΔFHB	0.13 ± 0.01	1.16 ± 0.48	0.91
ΔFHBCBM32	0.24 ± 0.01	0.38 ± 0.10	0.97
inactive	0.28 ± 0.04	0.47 ± 0.25	0.87

References

1. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

2. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324 (2014).