







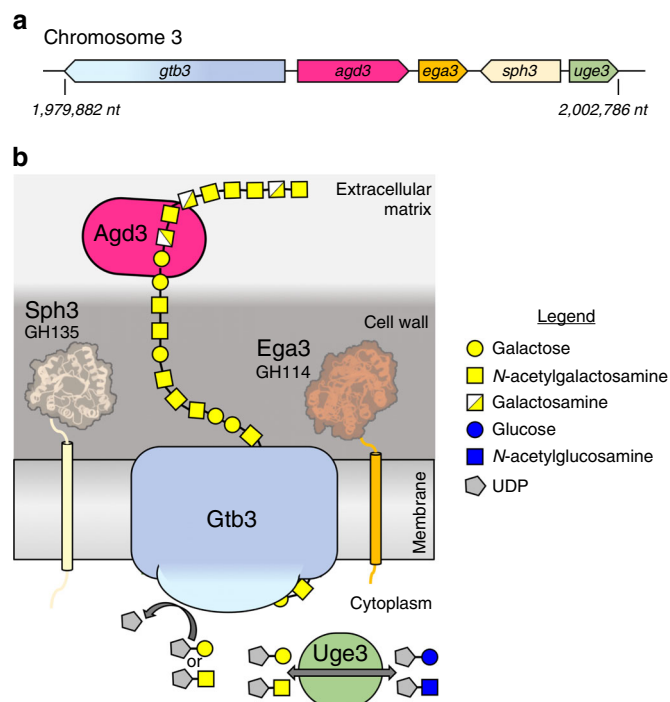
# Structural and biochemical characterization of the exopolysaccharide deacetylase Agd3 required for *Aspergillus fumigatus* biofilm formation

Natalie C. Bamford <sup>1,2,8,9</sup>, François Le Mauff<sup>3,4,5,9</sup>, Jaime C. Van Loon<sup>1,2</sup>, Hanna Ostapska<sup>3,4,5</sup>, Brendan D. Snarr<sup>3,4,5</sup>, Yongzhen Zhang <sup>6</sup>, Elena N. Kitova<sup>7</sup>, John S. Klassen<sup>7</sup>, Jeroen D. C. Codée <sup>6</sup>, Donald C. Sheppard<sup>3,4,5</sup>  & P. Lynne Howell <sup>1,2</sup> 

The exopolysaccharide galactosaminogalactan (GAG) is an important virulence factor of the fungal pathogen *Aspergillus fumigatus*. Deletion of a gene encoding a putative deacetylase, Agd3, leads to defects in GAG deacetylation, biofilm formation, and virulence. Here, we show that Agd3 deacetylates GAG in a metal-dependent manner, and is the founding member of carbohydrate esterase family CE18. The active site is formed by four catalytic motifs that are essential for activity. The structure of Agd3 includes an elongated substrate-binding cleft formed by a carbohydrate binding module (CBM) that is the founding member of CBM family 87. Agd3 homologues are encoded in previously unidentified putative bacterial exopolysaccharide biosynthetic operons and in other fungal genomes.

<sup>1</sup>Program in Molecular Medicine, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada. <sup>2</sup>Department of Biochemistry, Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada. <sup>3</sup>Department of Microbiology and Immunology, Faculty of Medicine, McGill University, Montreal, QC H3A 2B4, Canada. <sup>4</sup>Infectious Disease and Immunity in Global Health, Research Institute of McGill University Health Center, Montreal, QC H4A 3J1, Canada. <sup>5</sup>McGill Interdisciplinary Initiative in Infection and Immunity, Montreal, QC H3A 1Y2, Canada. <sup>6</sup>Leiden Institute of Chemistry, Leiden University, 2300RA Leiden, The Netherlands. <sup>7</sup>Alberta Glycomics Centre and Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2, Canada. <sup>8</sup>Present address: Division of Molecular Microbiology, School of Life Sciences, University of Dundee, DD1 5EH Dundee, UK. <sup>9</sup>These authors contributed equally: Natalie C. Bamford, François Le Mauff. ✉email: [donald.sheppard@mcgill.ca](mailto:donald.sheppard@mcgill.ca); [howell@sickkids.ca](mailto:howell@sickkids.ca)

*Aspergillus fumigatus* is a ubiquitous, biofilm-forming, filamentous fungus that causes invasive infections in immunocompromised patients<sup>1</sup>. Even with currently available antifungal agents, the mortality of invasive aspergillosis remains over 50%, highlighting the need for new therapies and better understanding of the underlying virulence factors governing *A. fumigatus* pathogenesis<sup>2</sup>. During infection *A. fumigatus* transitions to a biofilm mode of growth in which fungal hyphae are encapsulated in a self-produced matrix<sup>3</sup>. Galactosaminogalactan (GAG), an  $\alpha$ -1,4-linked linear exopolysaccharide of galactose (Gal) and *N*-acetylgalactosamine (GalNAc), is essential for *A. fumigatus* biofilm formation and a key virulence factor<sup>4</sup>. The synthesis of GAG is dependent on a cluster of genes encoding five carbohydrate-active enzymes (Fig. 1)<sup>5,6</sup>. A model of GAG biosynthesis and modification has been proposed (Fig. 1) and includes the production of activated monosaccharide building blocks by the epimerase Uge3<sup>5,7</sup> followed by synthesis and export by the predicted integral membrane glycosyltransferase Gtb3<sup>5</sup>. The GAG cluster also encodes two glycoside hydrolases: an endo- $\alpha$ -1,4-*N*-acetylgalactosaminidase, Sph3<sup>6,8</sup>; and an endo- $\alpha$ -1,4-galactosaminidase, Ega3<sup>9</sup>. Lastly, GAG is partially deacetylated by the secreted protein Agd3 and deletion of *agd3* abolishes this modification<sup>5</sup>. Although, the  $\Delta$ *agd3* mutant produces normal amounts of GAG, this strain has markedly impaired biofilm formation and lacks the cell-wall decoration associated with GAG production<sup>5</sup>. The  $\Delta$ *agd3* mutant also exhibits significantly lower virulence in a murine model of *A. fumigatus* infection compared with the WT strain, supporting Agd3 as a virulence factor<sup>5</sup>. GAG synthetic gene clusters containing Agd3 homologs have been identified in numerous fungal species including plant and animal pathogens<sup>5</sup>.



**Fig. 1** Current model of GAG biosynthesis in *A. fumigatus*. **a** Schematic of the five co-regulated genes of the GAG cluster found on chromosome 3. The location on the chromosome is denoted by nucleotide (nt) number. **b** Cartoon representation of the GAG biosynthesis system based on bioinformatics and previous studies including the two crystal structures of the glycoside hydrolase domains of Sph3 (PDB 5C5G) and Ega3 (PDB 6OJ1). The synthesized polymer is hypothetical in sequence as the GAG polymer is heterogeneous.

Deacetylation is important in exopolysaccharide processing and production in multiple bacterial biofilm systems including *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Escherichia coli*, *Bordetella bronchiseptica*, *Listeria monocytogenes*, and *Pseudomonas aeruginosa*<sup>10–17</sup>. De-*N*-acetylation of biofilm exopolysaccharide has been linked to cell aggregation, surface attachment, exopolysaccharide secretion, and biofilm maturation depending on the organism<sup>13–17</sup>. In each of these systems, de-*N*-acetylation is mediated by a protein that belongs to carbohydrate esterase (CE) family 4. Agd3 has low sequence homology to the CE4 family and does not belong to any of the currently defined Carbohydrate Active enZYme database (CAZy) CE families.

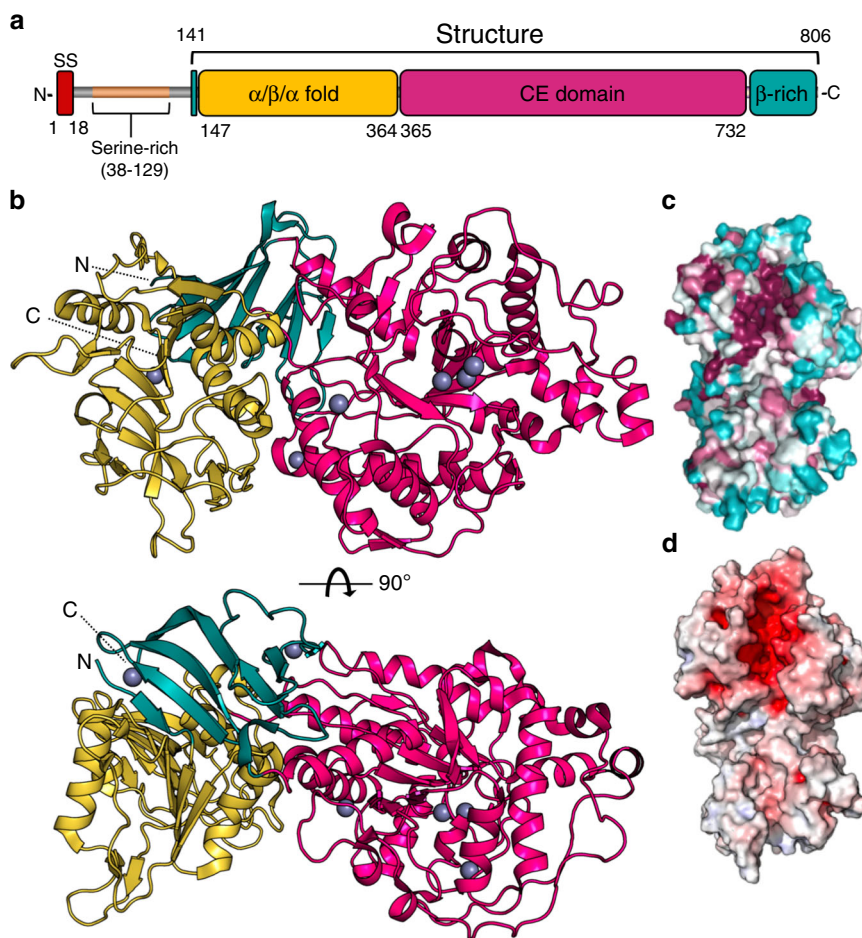
Herein, we show that *A. fumigatus* Agd3 has a unique three-domain architecture and that the enzyme specifically deacetylates  $\alpha$ -1,4-GalNAc oligosaccharides in a metal-dependent manner. Agd3 has a novel carbohydrate-binding module (CBM) that extends the substrate binding groove, increases activity on soluble GAG, and influences the location of deacetylation within oligosaccharides. Our structural and functional characterization of the CE-domain and N-terminal domain of Agd3 presented herein reveal that these domains are the founding members of CE18 and CBM87 families, respectively. Phylogenetic analysis found distant homologs in the bacterial kingdom. Although not all *agd3*-like genes are located in operons, those that were co-localize with other putative carbohydrate-active enzymes, resided in uncharacterized exopolysaccharide biosynthetic clusters. These findings suggest deacetylation of a carbohydrate polymer by an Agd3 homolog is a more prevalent occurrence than previously appreciated and will aid in further genome annotation.

## Results

**Agd3 has low homology to CE4 family members.** Bioinformatics has suggested that Agd3 is a secreted 806 amino acid multidomain protein<sup>5</sup>. The N-terminus of the protein, residues 38–129, contains a serine-rich region that is predicted to be structurally disordered by Phyre<sup>2</sup><sup>18</sup>. When submitted to NetOGlyc 4.0 server<sup>19</sup>, 69 O-glycosylation sites were identified between residues 26 and 141. Although NetOGlyc is designed for predicting sites of O-glycosylation on mammalian peptides it has been shown to accurately predict areas of hyper-O-glycosylation in fungal proteins<sup>19,20</sup>.

Agd3 was also predicted to contain a domain, residues 509–726, with structural similarity to members of the carbohydrate esterase family 4 (CE4)<sup>5</sup>. The CE4 family contains exopolysaccharide deacetylases that adopt a  $(\beta/\alpha)_7$ -fold. As previously noted, Agd3 has very low sequence similarity to the CE4 family and has a predicted CE domain encompassing at least 370 residues. This is in contrast to known CE4 family members where the CE domain is typically under 300 residues<sup>5,21</sup>. The amino acid sequence of Agd3 is too divergent to be a member of the CE4 family, thus direct evidence of carbohydrate deacetylation would allow classification of Agd3 as the founding member of a CE family. With the exception of the predicted domain, with structural similarity to CE4's, Agd3 had little predicted similarity to other biochemically or structurally characterized proteins. An 88-residue stretch, residues 147–234, was predicted to be structurally similar to amido-transferases but this region did not contain the catalytic residues required for these enzymes<sup>5</sup>. The function of this domain and the C-terminal  $\beta$ -rich domain is unknown.

**Agd3 has a compact three-domain architecture.** To gain insights into the function of Agd3, structural studies of the predicted globular protein, residues 141–806, were pursued. Agd3 crystallized in the hexagonal space group  $P6_1 2 2$ . Diffraction data were



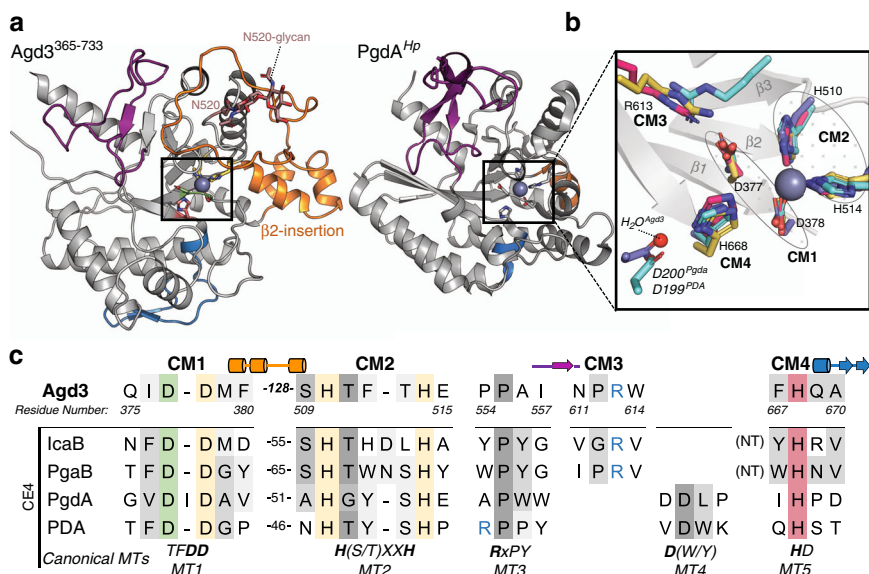
**Fig. 2 Agd3 is a multi-domain protein with a conserved electronegative cleft.** **a** Linear representation of Agd3 domains based on the structure. **b** Agd3 in cartoon representation displaying the three-domain arrangement: the N-terminal  $\alpha/\beta/\alpha$ -fold in yellow; the catalytic CE-domain in pink; and the C-terminal  $\beta$ -rich domain in teal. Zinc ions are depicted as gray spheres. **c** Surface representation of amino acid conservation colored from variable (teal) to conserved (fuchsia) as calculated by ConSurf<sup>85</sup>. **d** Electrostatics surface representation shows an electronegative cleft. Electrostatics were calculated by APBS in PyMol (v. 2.0.7) and visualized in blue to red (+10 kT/e to -10 kT/e)<sup>86</sup>.

anisotropic with diffraction limits of 2.6 Å in one dimension and 3 Å in the other two dimensions. The structure was solved by Zn single-wavelength anomalous dispersion and refined to  $R_{\text{work}}$  and  $R_{\text{free}}$  of 19% and 24%, respectively (Supplementary Table 1). Agd3 has a unique compact three-domain structure (Fig. 2a,b). The N-terminal domain, comprising residues 141–364, adopts a mixed  $\alpha/\beta/\alpha$ -fold ( $\alpha/\beta/\alpha$ -domain) with a seven-stranded  $\beta$ -sheet at its center. The central domain, residues 365–733, contains the CE4 homology region and has a distorted ( $\beta/\alpha$ )-fold (CE-domain). The C-terminal domain spanning residues 733–806 forms six  $\beta$ -strands in a seven-stranded  $\beta$ -sandwich. The seventh strand is contributed by the N-terminal region residues 146–149, residues prior to the  $\beta/\alpha/\beta$ -domain. The N- and C-terminal domains increase the stability of the CE-domain, as constructs lacking one or both of these accessory domains expressed poorly and could not be purified to homogeneity from either *E. coli* or *Pichia pastoris*. Mapping surface conservation reveals a highly conserved electronegative cleft in the CE-domain (Fig. 2c,d).

**Agd3 contains a unique CE-domain.** When the entire crystal structure of Agd3 was submitted to the DALI server, no structures with similar topological arrangements were found. The most similar structure found was the peptidoglycan deacetylase PgdA (PDB 3QBU<sup>22</sup>) from *Helicobacter pylori*, which aligned to the

putative CE-domain, with a root mean square deviation (RMSD) of 2.84 Å over 227  $\alpha$ -carbons. Agd3 and PgdA share only 11.9% sequence identity, and while the central ( $\beta/\alpha$ )-fold is similar between these proteins, there are large differences in the regions connecting the ( $\beta/\alpha$ ) motifs (Fig. 3a). As PgdA does not have all the canonical CE4 motifs, we compared Agd3 with a prototypical CE4, Bc1960 from *Bacillus cereus* (PDB 4L1G<sup>23</sup>). Bc1960 has 11.4% sequence identity to Agd3 and aligned with a RMSD of 2.61 Å over 175  $\alpha$ -carbons. A third CE4 enzyme, IcaB from *Ammonifex degensii* (PDB 4WCJ<sup>24</sup>) was also aligned to Agd3 due to its functional similarity. Both proteins are extracellular and IcaB is active on the biofilm exopolysaccharide, poly- $\beta$ -1,6-*N*-acetylglucosamine (PNAG)<sup>24</sup>. IcaB has only 5.2% sequence identity to Agd3 and aligned poorly compared with the other CE4 enzymes with a RMSD of 3.80 Å over 134  $\alpha$ -carbons. The catalytic cations of these CE4 enzymes overlapped with a Zn ion in Agd3, suggesting that Agd3 is also metal dependent.

The structure of Agd3, analysis of its conserved residues and alignment to CE4 members enabled the identification of four catalytic motifs, CM1–4. CE4 family members have five canonical active site motifs (MT1–5)<sup>21,25</sup>. The MT1–5 motifs coordinate the catalytic metal and participate in the deacetylation event<sup>21,25</sup>. In Agd3 CM1, 2, and 4 are similar to CE4 MT1, 2, and 5, respectively. Only one of these motifs (CM2) was identifiable from sequence alignments (Fig. 3b). D377 and D378 of CM1



**Fig. 3 Agd3 is a unique CE enzyme with distant homology to the CE4 family.** **a** Cartoon representation of the CE domain of Agd3<sup>365-733</sup> (left) and PgdA (right, PDB 3QBU) with the loops between the (β/α)<sub>7</sub>-fold motifs colored analogously and shown in **c**. The loop after β2 (orange), β5 (purple), and the catalytic histidine (blue) are colored. The ordered sugars of the N540 glycan are in brown. **b** Structural alignment of the catalytic residues of Agd3 (pink) with four representative CE4 enzymes: *H. pylori* PgdA (3QBU, purple); *A. degensii* IcaB (4WCJ, yellow); and *B. cereus* PDA (4L1G, blue). The β-strands of the barrel are shown in gray cartoon representation. The putative catalytic motifs are labeled from CM1–CM4 and the β-stands before CM1 and CM2 have been numbered. **c** Primary sequence alignment of Agd3 catalytic motifs (CM1–4) with the CE4 MT1–MT5 as determined by structural alignment. The putative catalytic base (D377), metal coordinating triad (D378, H510, and H514) and putative catalytic acid (H668) are highlighted in green, yellow, and red, respectively. The arginine coordinating the catalytic acid is depicted in blue font. The amino acid distance between CM1 and CM2, or MT1 and MT2 in the primary sequence is listed for each protein. The canonical CE4 motifs are summarized at the bottom of the table as defined by Aragunde et al. MT5 occurs at the N-terminal to MT1–4 for the PNAG deacetylases IcaB and PgaB as denoted by the (NT) in the table.

aligned well with CE4 MT1, which are the catalytic base and one member of the metal coordinating triad. CM2 was identified previously using sequence alignment due to its conservation with the HXXXH motif of MT2 (Fig. 3b,c). In Agd3, these histidines are H510 and H514 and each participates in metal coordination. CM1 and CM2 are separated by over 120 residues, which is more than twice the distance observed between equivalent motifs in CE4 enzymes (Fig. 3c). This difference is the result of a 64-amino acid insertion following strand β2 of the barrel, which is not present in CE4 enzymes. The 64-amino acid insertion consists of three small α-helices that constitute an extension on one side of the putative active site (β2-insertion, Fig. 3a). The β2-insertion “caps” the loops following β3 and β4 of the barrel, creating a deeper active site cleft (Supplementary Fig. 1). N520, following β3, is the location of one of the four N-glycans that could be modeled into the electron density (Fig. 3a and Supplementary Fig. 1). The β2-insertion loops around this N-glycan with T471 hydrogen bonding the O6 of the first GlcNAc, and F463 and P460 contributing to a hydrophobic pocket for the methyl of the GlcNAc (Supplementary Fig. 1).

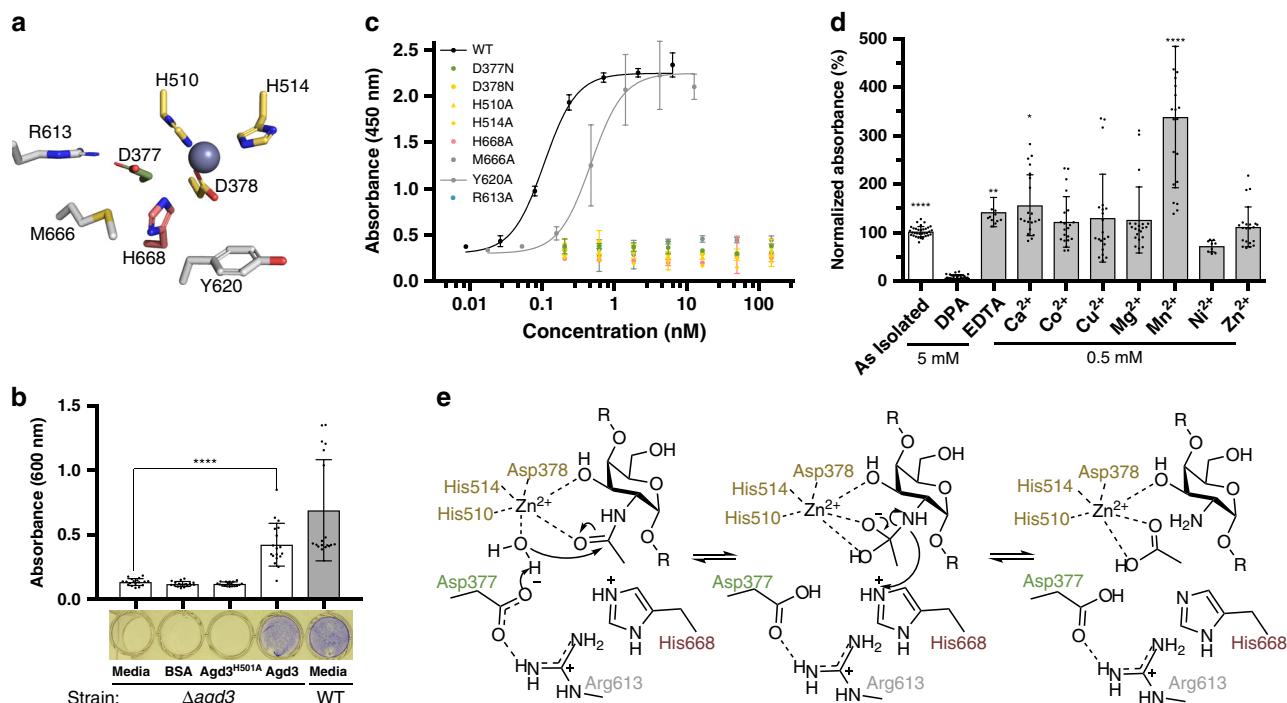
Further differences were found between Agd3 and the other CE4 motifs. CM3 is unique to Agd3 and contains a highly conserved arginine (R613) that coordinates the putative catalytic base D377. R613 aligns structurally with the base coordinating arginine in PNAG deacetylases IcaB and PgaB (PDB 4F9D<sup>26</sup>). Canonical CE4 enzymes have an equivalent arginine in MT3 but this occurs much earlier in the primary sequence. Agd3 contains the conserved proline homologous to the proline of MT3 but does not have the activating arginine in this motif (Fig. 3c). The last Agd3 family catalytic motif, CM4, contains the putative catalytic acid, H668. CM4 aligns structurally with the CE4 MT5, a motif with little sequence conservation besides the catalytic histidine<sup>21,25</sup>. Agd3 does not contain an equivalent to the CE4

MT4, and thus lacks an aspartic acid that activates the catalytic acid in prototypical CE4 enzymes<sup>21,25</sup>. An ordered water molecule was found in the place of the MT4 aspartic acid (Fig. 3b), suggesting differences in mechanism between these enzymes. Despite the low overall sequence identity, the similarity of the CM1–4 catalytic motifs of Agd3 to the MT1–5 CE4 family motifs suggests that Agd3 is an active carbohydrate deacetylase that is dependent on a divalent cation cofactor. Agd3 may have a different catalytic mechanism relative to prototypical CE4 enzymes due to the lack of the activating aspartic acid, and may be more similar to PNAG deacetylases that also lack this motif<sup>24,26</sup>.

**Agd3 induces GAG adherence and biofilm formation.** Previously we showed that the *A. fumigatus* Δ*agd3* strain could form a biofilm when co-cultured with culture supernatants from the Δ*uge3* mutant, a GAG-deficient strain that secretes Agd3. To investigate whether there were other factors involved in this complementation we sought to determine whether recombinant Agd3 was sufficient to rescue Δ*agd3* biofilm formation. Addition of 0.5 nM recombinant Agd3 to Δ*agd3* cultures rescued biofilm formation as measured by crystal violet staining (Fig. 4b). Addition of Agd3<sup>H510A</sup>, an alanine mutant that would compromise metal coordination, did not complement biofilm formation suggesting that biofilm rescue was the consequence of the activity of the enzyme and that the histidine is required for activity.

To further probe the activity of Agd3 and identify catalytically important residues, an enzyme-linked lectin assay (ELLA) to detect carbohydrate deacetylase activity was developed using the native GAG substrate. The deacetylation of GAG renders the polymer adherent to plastic surfaces. Immobilized GAG was detected and quantified using the GalNAc-specific lectin Soybean-Agglutinin (SBA). In this assay, treatment of





**Fig. 4** Agd3 deacetylates GAG in a metal-dependent manner. **a** Cartoon representation of the conserved active site residues including the metal coordinating triad (yellow), the putative catalytic base (green), and the putative catalytic acid (pink). **b** Absorbance of crystal violet stained 21 h old biofilms of  $\Delta$ agd3 or WT strains after treatment as labeled. Bars indicate mean  $\pm$  SD and represent three independent experiments of  $n = 6$ . Representative stained biofilms in a 96-well plate are aligned below. **c** Agd3 deacetylation of GAG as measured by the GAG-ELLA. The experiment was run in triplicate. **d** Agd3 deacetylates GAG in a metal-dependent mechanism. Each condition had a total number of  $n = 21$  for all metal conditions except  $\text{Ni}^{2+}$  with  $n = 9$ ,  $n = 34$  for as isolated,  $n = 12$  for EDTA, and  $n = 28$  for DPA, representing individual wells spread over nine independent experiments. **e** Metal-dependent de-*N*-acetylation reaction based on the findings of the activity assays and CE4 enzymes. The residues are color coded as in **a**. For **b** and **d** statistics were calculated one-way ANOVA with Dunnett's multiple comparison test. \* $p < 0.033$ , \*\*\* $p < 0.0002$ , \*\*\*\* $p < 0.0001$ . Error bars represent SD. Source data is available in Source Data file.

fully-acetylated GAG, from the  $\Delta$ agd3 strain, with Agd3 at concentrations as low as 80 pM resulted in detectable deacetylated adherent polymer (Fig. 4c). Mutation of the residues that coordinate the metal, H510A, H514A, or D378N, led to no detectable GAG adhesion (Fig. 4c). Similarly, mutation of the putative catalytic acid (H668A) and catalytic base (D377N) also abrogated GAG adhesion even when the enzyme concentration was increased 1875-fold, to 150 nM (Fig. 4c). The active site proximal arginine, R613 found in CM3, is also integral for activity in this assay. M666 in CM4 is highly conserved in Agd3 homologs and when mutated to alanine also abrogates adhesion suggesting the importance of this residue in the formation of the active site and/or substrate alignment.

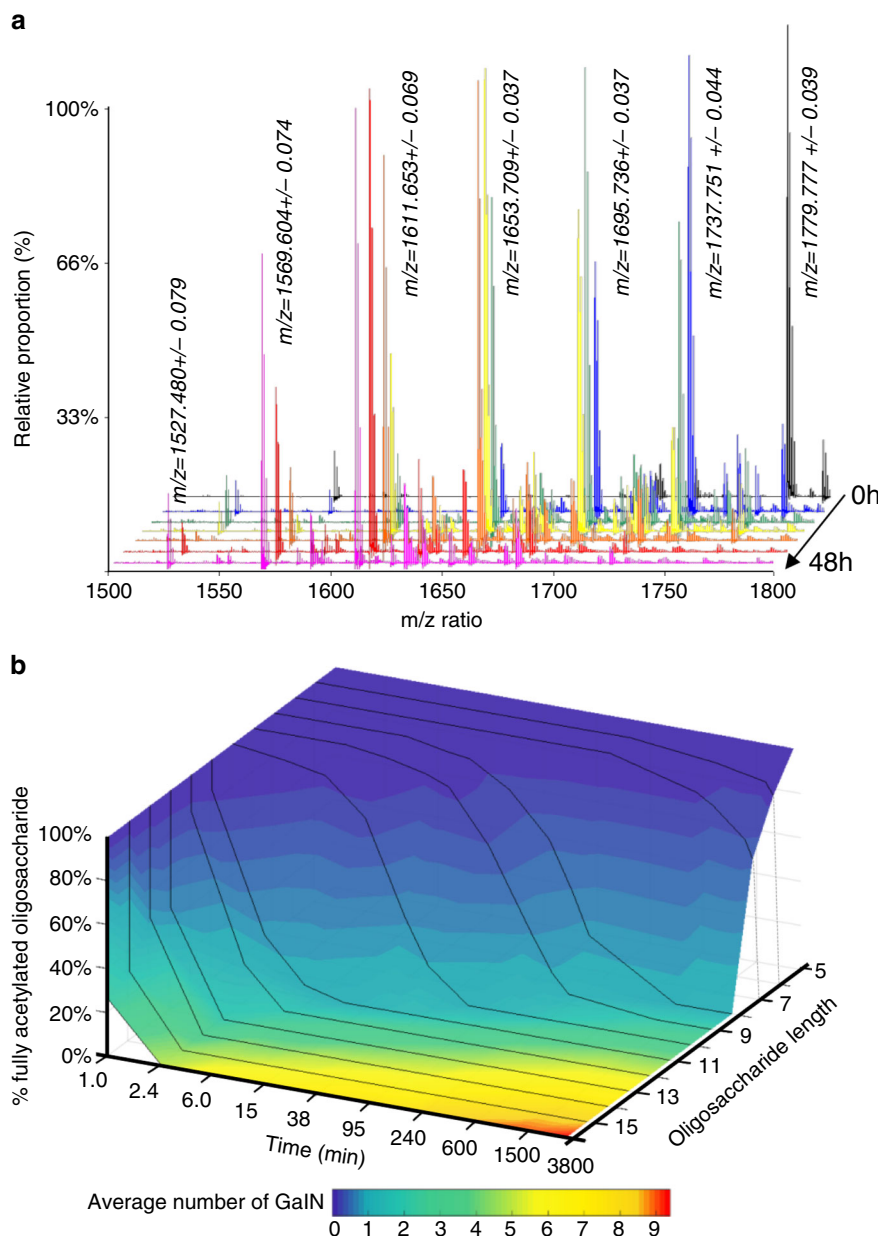
Aromatic residues are often found in and near the active sites of carbohydrate active enzymes<sup>21,24,27–29</sup>. Y620 is near the active site and could potentially be involved in substrate binding. Mutation of Y620 to alanine led to a minor decrease in activity with a 50% effective concentration ( $\text{EC}_{50}$ ) of 0.48 nM (95% confidence interval [CI] of 3.3–7.0 nM) compared with 0.11 nM (95% CI of 0.10–0.13 nM) for the wildtype protein under these conditions (Fig. 4c).

To determine whether Agd3 was metal dependent and the identity of the metal, Agd3 was incubated with chelators or the chloride salts of various divalent cations before addition of the GAG substrate. Addition of dipicolinic acid (DPA) abolished Agd3 activity whereas EDTA had no significant effect (Fig. 4d). Under the experimental conditions tested, exogenous addition of  $\text{Mn}^{2+}$  resulted in almost threefold increase in activity as compared with non-treated Agd3 (Fig. 4d). Combined, these

results suggest that Agd3, similar to CE4 family members<sup>21</sup>, uses a divalent cation to stabilize a tetrahedral intermediate (Fig. 4d).

**Agd3 de-*N*-acetylates  $\alpha$ -1,4-GalNAc oligos with length dependence.** To confirm the de-*N*-acetylation activity of Agd3 at the molecular level, the activity on synthetic  $\alpha$ -1,4-(GalNAc)<sub>8</sub> was analyzed by matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry (MS). Consistent with the results of the GAG-ELLA and biofilm assays, Agd3 treatment of the octasaccharide resulted in the progressive loss of an average  $m/z$  ratio of  $42.014 \pm 0.005$ , corresponding to the monoisotopic mass of an acetate group (Fig. 5a). Agd3 treatment at concentrations as high as 10  $\mu\text{M}$  had no effect on chitin and chitosan oligosaccharides ( $\beta$ -1,4-*N*-acetylglucosamine and partially deacetylated  $\beta$ -1,4-*N*-acetylglucosamine, respectively), suggesting that Agd3 deacetylation activity is specific for GalNAc (Supplementary Fig. 2).

To determine the effect of substrate length on enzyme activity, Agd3 was incubated with a pool of GalNAc oligosaccharides of varying lengths (5–16 units) that were isolated from *A. fumigatus* biofilms (Supplementary Fig. 3). Examination of the rate and extent of de-*N*-acetylation of the pool of oligosaccharides following Agd3 treatment revealed a direct correlation between the length of the oligosaccharides and the rate of de-*N*-acetylation (Fig. 5b). GalNAc oligosaccharides shorter than seven units were poorly deacetylated, even after 48 h of treatment. In contrast, GalNAc oligosaccharides longer than 13-mers were rapidly deacetylated by Agd3 multiple times within 5 min (Fig. 5b).



**Fig. 5** Effects of substrate length on the de-*N*-acetylation activity of Agd3. **a** Time course of  $\alpha$ -1,4-(GalNAc)<sub>8</sub> de-*N*-acetylation by 1  $\mu$ M Agd3 over 48 h. Enzymatic activity was monitored by MALDI-TOF MS of the  $m/z$  window 1500–1800. The MS traces correspond to the following time points: 0, 1, 2, 3, 6, 24, and 48 h. **b** Time course of de-*N*-acetylation of GalNAc oligosaccharides ranging from pentamer to 16-mers. Signals from each oligosaccharide were integrated and graphed as the percentage of fully acetylated oligosaccharide over time. The degree of de-*N*-acetylation of each oligosaccharide is indicated by the color of the 3D surface.

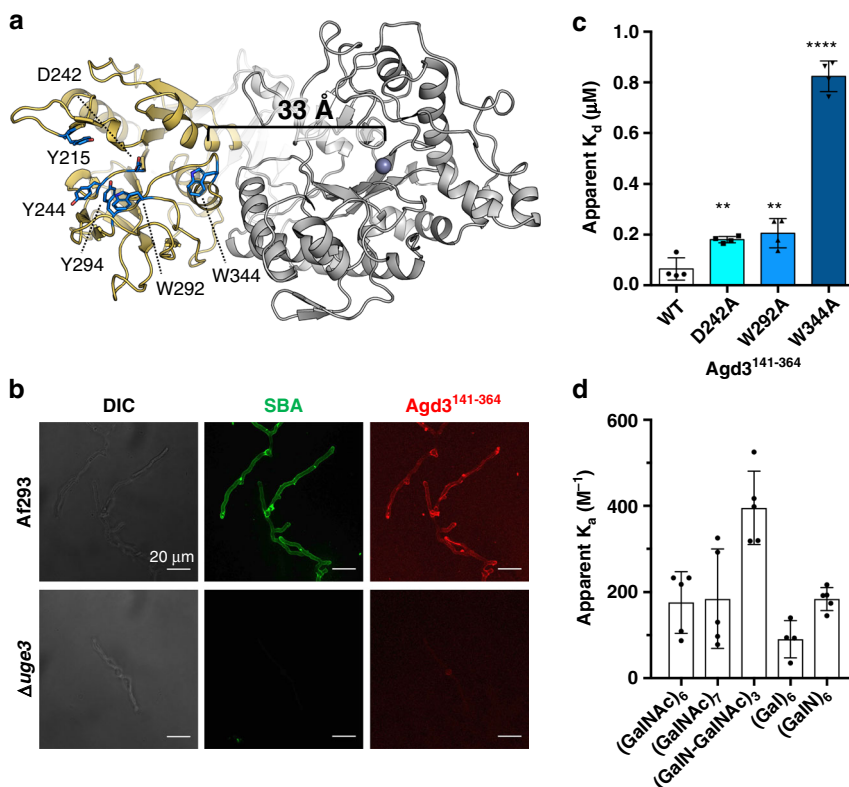
The presence of deacetylated GalN moieties within oligosaccharides did not seem to impair subsequent de-*N*-acetylation events, suggesting that Agd3 can bind oligosaccharides containing both GalNAc and GalN. The direct evidence by MS that Agd3 is an active carbohydrate deacetylase, with activity on  $\alpha$ -1,4-GalNAc substrates, and the low sequence identity to any current CE family, led to the creation of the CE18 family.

#### Agd3 contains a unique carbohydrate binding module (CBM).

The results of the oligosaccharide deacetylation studies demonstrated that Agd3 prefers longer substrates. These findings suggest a long binding cleft that provides increased affinity for the longer polymers over shorter ones. Investigation of the structure found that the cleft of the  $\alpha/\beta/\alpha$ -domain (Agd3<sup>141–364</sup>) aligns with the

cleft of the CE-domain (Fig. 6a). The  $\alpha/\beta/\alpha$ -domain cleft is lined with surface exposed conserved aromatic residues, which are commonly involved in carbohydrate binding<sup>27,29–32</sup> (Fig. 6a).

The N-terminal domain of Agd3 (141–364) forms a seven-stranded parallel  $\beta$ -sheet that is sandwiched between three  $\alpha$ -helices, one helix on one side ( $\alpha_2$ ) and two on the other ( $\alpha_1$  and  $\alpha_4$ ) (Supplementary Fig. 4a). According to the DALI server, Agd3<sup>141–364</sup> is most structurally similar to two uncharacterized class 1 glutamine synthase (GATase1)-like proteins: GATase1-like protein from *Planctomyces limnophilus* (PDB 3RHT); and the ThuA-like protein from *Bacteroides uniformis* (PDB 4JQS). Secondary structure alignment revealed the shared topology with RMSD of 2.63 and 2.73 Å over 176  $\alpha$ -carbon atoms for GATase1- and ThuA-like proteins, respectively (Supplementary Fig. 4b).



**Fig. 6 Agd3 contains a unique carbohydrate binding module.** **a** The structure of Agd3 showing the CBM in yellow with the first conserved tryptophan (W344) ~33 Å from the active site. **b** WT (Af293) and  $\Delta$ uge3 hyphae visualized by differential interference contrast (DIC, left) and fluorescent microscopy co-stained with SBA (center, green) and Alexa-568 labeled Agd3<sup>141-364</sup> (right, red). Representative images from two independent experiments each of three or more images for each condition. Recombinant Agd3<sup>141-364</sup> localizes to the hyphal surface in a GAG dependant manner similar to SBA. **c** Apparent  $K_d$  ( $\mu\text{M}$ ) for Agd3<sup>141-364</sup>, and point mutants thereof, to soluble GAG as determined using GAG-ELISA. Bars represent mean  $K_d$  values  $\pm$  one SD of four experiments run in triplicate. Statistics were calculated using one-way ANOVA with Dunnett's multiple comparison test. \* $p < 0.013$ , \*\* $p < 0.0034$ , \*\*\*\* $p < 0.0001$ . **d** Apparent  $K_a$  ( $\text{M}^{-1}$ ) for Agd3<sup>141-364</sup> for  $\alpha$ -1,4-linked carbohydrate ligands determined by a direct ESI-MS assay. Experiments were performed at 25 °C and pH 7.2 with  $n = 4$  for (Gal)<sub>6</sub> and  $n = 5$  for all others. Each data point represents the calculated  $K_a$  value from a single ESI-MS assay. Bars are the mean  $\pm$  one SD. Source data is available in Source Data file.

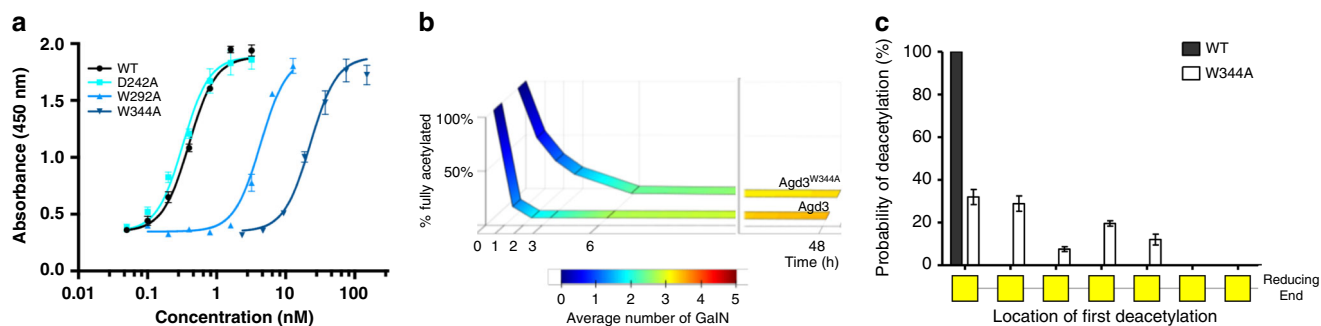
The ThuA-like protein has higher sequence identity to Agd3<sup>141-364</sup> (15.9%) than to GATase1 (10.2%). Agd3 does not share conserved surface residues with either of these proteins, nor the catalytic GATases Cys-His-Glu triad, all of which are located on the C-terminal end of the central  $\beta$ -sheet<sup>33</sup>. Instead Agd3 and homologs have conserved surface exposed aromatics in this region (Fig. 6a).

To examine whether the N-terminal Agd3<sup>141-364</sup> domain serves as a CBM in catalysis, GAG binding assays were performed. We found that recombinant fluorescently-labeled Agd3<sup>141-364</sup> co-localized with SBA on hyphae surfaces (Fig. 6b). No Agd3<sup>141-364</sup> was detected on the surface of the GAG-deficient  $\Delta$ uge3 strain (Fig. 6b), suggesting that Agd3<sup>141-364</sup> binds to cell-associated GAG. Agd3<sup>141-364</sup> also bound to soluble GAG in a dose-dependent manner. Mutation of conserved residues within the cleft, D242, W292, or W344, to alanine increased the apparent dissociation constant ( $K_d$ ) for soluble GAG binding, suggesting that this region plays a role in this process (Fig. 6c). W344A point mutation had the largest effect, with a  $K_d$  12.7-fold higher than the wild-type protein.

Ligand specificity of the CBM was explored using synthesized GAG oligosaccharides. Association constants ( $K_a$ ) were determined by direct electrospray ionization MS (ESI-MS) for each oligosaccharide. Agd3<sup>141-364</sup> bound  $\alpha$ -1,4-(GalNAc)<sub>6</sub> and  $\alpha$ -1,4-(GalN)<sub>6</sub> with  $K_a$  of  $180 \pm 60 \text{ M}^{-1}$  and  $180 \pm 30 \text{ M}^{-1}$ , respectively, (Fig. 6d). No statistically significant difference was found between

the binding of (GalNAc)<sub>6</sub> and (GalNAc)<sub>7</sub>, suggesting that the binding site spans six or fewer subsites, which is consistent with the length of the identified cleft (Fig. 6d). Agd3<sup>141-364</sup> binding of (Gal)<sub>6</sub> was negligible, suggesting that the CBM of Agd3 is specific for regions of the GAG polymer that are GalNAc/GalN rich. Interestingly, there was slightly, but significantly, higher affinity for a mixed GalN-GalNAc oligosaccharide ( $K_a$   $400 \pm 90 \text{ M}^{-1}$ , Fig. 6d). This finding suggests that partial deacetylation of the polymer could lead to higher affinity, and hence accelerated deacetylation after the initial deacetylation events have occurred.

**The CBM augments Agd3 deacetylation.** To determine the role of the CBM in modulating the activity of Agd3, the deacetylase activity of the W292A and W344A mutants were compared with that of recombinant wild-type Agd3. Agd3 constructs lacking the CBM, with and without the C-terminal domain, were unstable so activity of the CE domain alone could not be determined. Mutation of W292 or W344 to alanine resulted in a significant decrease in GAG deacetylation as measured by the GAG-ELLA. When compared with the wild-type enzyme, Agd3<sup>W292A</sup> and Agd3<sup>W344A</sup> exhibited a 13-fold ( $4.3 \pm 0.29 \text{ nM}$ ) and 58-fold ( $19.8 \pm 1.89 \text{ nM}$ ) higher  $\text{EC}_{50}$ , respectively, suggesting that these sites may influence substrate positioning and/or affinity in the context of the multidomain enzyme (Fig. 7a). As W344 is the predicted CBM binding subsite closest to the active site, and had the most dramatic effect on deacetylase activity when mutated, this mutant



**Fig. 7 Agd3 CBM increases deacetylation rate and specificity.** **a** Deacetylation of GAG by Agd3 or point mutants as determined by GAG-ELLA. Experiment performed in triplicate and EC<sub>50</sub> calculated from three independent experiments. One representative experiment is data shown as mean values  $\pm$  one SD. **b** Time course of  $\alpha$ -1,4-(GalNAc)<sub>7</sub> deacetylation by Agd3 or Agd3<sup>W344A</sup>. Products were analyzed by MALDI-TOF MS and graphed as the relative amount of unmodified substrate. The average number of deacetylation events is indicated by the color legend. **c** Analysis of the MALDI-TOF MS-MS fragmentation of mono-deacetylated heptasaccharide product produced Agd3 or Agd3<sup>W344A</sup>. The X-axis indicates the location of the deacetylation event within the heptasaccharide. Mean values are of 11 experiments and error bars represent SD. Source data is available in Source Data file.

was selected for further studies to determine the role of the CBM in substrate binding and de-*N*-acetylation.

To study the difference between the de-*N*-acetylation activity of wild-type Agd3 and Agd3<sup>W344A</sup> in detail, synthetic  $\alpha$ -1,4-GalNAc heptasaccharides and octasaccharides were treated with both enzymes, and the reaction products quantified over time by MALDI-TOF MS. Initial deacetylation of  $\alpha$ -1,4-(GalNAc)<sub>7</sub> by Agd3<sup>W344A</sup> was significantly slower than by wild-type Agd3 (Fig. 7b). Conversion of  $\alpha$ -1,4-(GalNAc)<sub>7</sub> by Agd3<sup>W344A</sup> into >97.5% (GalNAc)<sub>6</sub>GalN required over 6 h, as compared with two hours for the wild-type Agd3 (Fig. 7b). The rate of subsequent deacetylation events was not different between the two enzymes (Fig. 7b). Similar results were obtained when the synthetic GalNAc octasaccharide was used as a substrate (Supplementary Fig. 5).

The W344A mutation also resulted in a change in the position of the primary de-*N*-acetylation event within the oligosaccharide substrate (Fig. 7c and Supplementary Fig. 6). The primary deacetylation event in wild-type Agd3 was observed exclusively at the GalNAc located at the non-reducing end of the oligosaccharide. In contrast, Agd3<sup>W344A</sup> treatment resulted in deacetylation of one of the five GalNAc located at the nonreducing end of the oligosaccharide (Fig. 7c). Subsequent deacetylation events by both enzymes occurred at one of the four remaining GalNAc located at the nonreducing end of the oligosaccharide. Neither enzyme was able to modify the two GalNAc residues proximal to the reducing end (Supplementary Fig. 6). The decrease in deacetylation rate and altered specificity of the first deacetylation event in the W344A mutant suggests that the CBM aids in substrate alignment within the active site and thus more efficient, and specific, binding of oligosaccharides.

### Agd3 homologs are present in bacterial and fungal species.

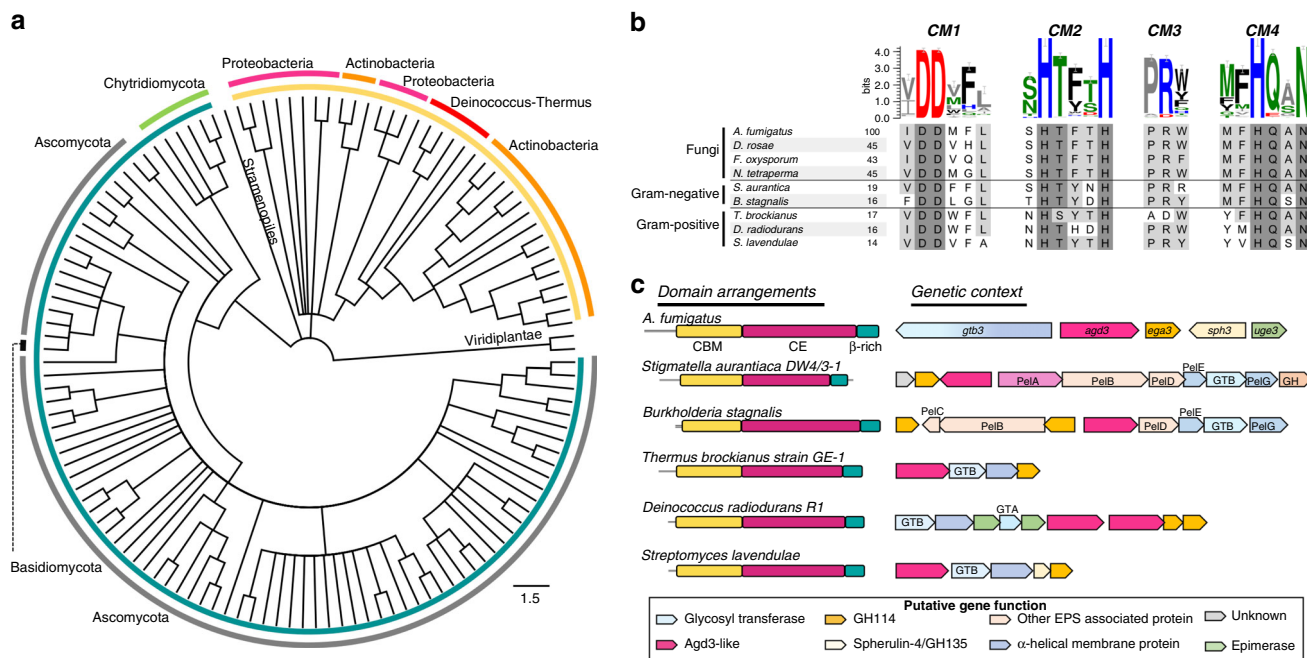
Agd3 orthologues were previously found across the *Ascomycota* phylum, located within gene clusters resembling the GAG biosynthetic gene cluster of *A. fumigatus*. Broadening this search to the nonredundant protein database using JackHMMer and BLASTP over 800 sequences with >20% identity were found. Agd3 homologs were found in 311 different species, including in 266 eukaryotic and 45 bacterial species. The majority of high confidence hits were in the *Ascomycota* phylum as previously determined, as well as in five fungal species within the *Chytridiomycota* phylum including anaerobic herbivore symbiotes and one aquatic fungi, *Gonapodya prolifera*. Agd3 homologs were found in 30 Terrabacteria, including *Actinobacter* and *Deinococcus-Thermus* spp. (Supplementary Data 1). Gram-negative hits were found in

Proteobacteria such as *Bulkholderia*, *Moritella*, and *Stigmatella*. After removing sequences with over 80% identity and those that were protein fragments, 126 sequences were selected and used to generate a maximum-likelihood tree (Fig. 8a). This tree revealed separate clustering of the bacterial and eukaryotic sequences suggesting horizontal gene transfer may have occurred between ancestors of these clades.

Alignment of the Agd3 homologs to Agd3 showed high conservation of all four CMs (Fig. 8a). CM3 is conserved in all the fungal enzymes and in most of the bacterial homologs. Bacterial genes had homology spanning all three of the globular domains of Agd3 (Fig. 8b). The C-terminal domain had the lowest sequence identity, but the secondary structure predictions were consistently  $\beta$ -rich in character. The CBM fold was found in all hits with significant conservation of the identified aromatic residues, including over 79% similarity at W334 and 88% similarity at W292. The bacterial Agd3 homologs share low sequence identity to Agd3 (13–19% MUSCLE, 20–28% BLASTP) but the presence of the catalytic motifs and the three-domain configuration suggests that they are CE enzymes related to Agd3.

In *Ascomycota*, the *agd3* gene is encoded within the putative GAG biosynthetic gene clusters of these organisms, implying a role for Agd3 in exopolysaccharide production in these fungi<sup>5</sup>. There is more variation in the genetic environment of the bacterial *agd3* genes. Of the 35 bacterial species represented in the tree, 30 contain *agd3* genes located in or near operons that encode predicted carbohydrate active enzymes (Fig. 8c). There is variation in operon synteny and content between bacterial species, but the genes have similar predicted functions to those found in the GAG gene cluster (Fig. 8c). Of the operons identified, all contain genes encoding a cytoplasmic glycosyltransferase and a multi-spanning  $\alpha$ -helical integral membrane protein. The co-occurrence of a GH114 encoding gene, homologous to *ega3*, is highly consistent throughout both Gram-negative and -positive organisms and, in some cases, multiple such genes were found in or adjacent to the operon (Fig. 8c). In *A. fumigatus*, *agd3* and *ega3* are found adjacent in the genome (Fig. 8c) and are coregulated<sup>5</sup>. Recently, *Ega3* was found to be specific for deacetylated regions of GAG supporting that *agd3* and *ega3* homologs are functionally linked<sup>9</sup>. Spherulin-4 annotated genes were also found in some of the *Actinobacteria* operons. The Gram-negative operons were more complex and contained genes with homology to the outer membrane components of the Pel polysaccharide biosynthetic machinery (Fig. 8c). These findings support that Agd3 is the first member of a larger family of carbohydrate deacetylases that span both bacterial and fungal species.





**Fig. 8 Agd3 homologs are present in bacterial and fungal kingdoms.** **a** The maximum-likelihood phylogeny tree constructed using MEGA7 containing 126 representative sequences of the BLASTP results. Visualization was done in FigTree with a bootstrap cutoff of 0.7 of 250 iterations. Bacterial sequences are highlighted in yellow and fungal in teal. **b** Sequence logo (created with WebLogo 3.6.0) of the four catalytic motifs (CM1-4) calculated from the 126 sequences used in **a**. The logo is accompanied by the amino acid sequence alignment of representative *agd3* homologs. Proteins from *Diplocarpon rosae*, *Sclerotinia sclerotiorum*, and *Neurospora tetrasperma* were used as Agd3 orthologues. Bacterial proteins were aligned and are the same as those shown in **c**. Sequence identity is shown for each as calculated by the final alignment done by MUSCLE in Geneious. **c** Linear domain arrangement and genetic location of each of the five-representative bacterial Agd3 homologs. Domains are shown to scale. Bacterial operons are to scale relative to each other but not to the GAG-cluster, due to intronic DNA. The list of organisms used in the tree is available in the Supplementary Data 2.

**Discussion**

In this study, we show that Agd3 is a carbohydrate deacetylase that defines a new family of carbohydrate esterases, CE18. Characterization of N-terminal domain also established a CBM family, CBM87. At present, there are 16 carbohydrate esterase families. Family CE10 has been withdrawn as the members of this family do not act on carbohydrate substrates. Only one new family has been created since 2008<sup>34</sup>. Agd3 has distant homology to members of the CE4 family, which includes the PNAG deacetylases, IcaB and PgaB, involved in biofilm formation in multiple Gram-positive and Gram-negative bacterial species<sup>13,14,26,35</sup>.

The crystal structure of Agd3 revealed a unique three-domain architecture not found in any structurally characterized protein to-date. The peptidoglycan deacetylase, PgdA (3QBU) has the closest structural similarity to Agd3, but this protein only aligns with the central CE domain of Agd3 (Fig. 3a). This structural alignment revealed commonality in active site motifs between Agd3 and the CE4 family despite the low sequence identity (11.9%) to PgdA. Agd3 has four highly conserved CMs that are essential for GAG deacetylation in vitro. Agd3 does not have a motif equivalent to MT4 in CE4 enzymes, which contains the catalytic acid activating aspartic acid. Instead, Agd3 has a water molecule in its place (Fig. 4b). The PNAG deacetylases IcaB and PgaB also lack this activating aspartate residue and have similarly structured water molecules<sup>26</sup>. In PgaB, the absence of the MT4 aspartic acid is proposed to be responsible for the enzyme’s lower catalytic efficiency, which correlates with the low levels of PNAG deacetylation observed in vivo<sup>26</sup>. Secreted *E. coli* PNAG is 22%<sup>15</sup> deacetylated, or less<sup>36</sup>, whereas other CE4 substrates, peptidoglycan and chitin, are typically >80% deacetylated<sup>37,38</sup>. The degree of deacetylation of *A. fumigatus* GAG is not known, however, analysis of the GAG from *Neurospora crassa* found that

the polymer has between 30 and 70% deacetylated galactosamine depending on the growth conditions. In *N. crassa*, deacetylation was correlated with the level of expression of a deacetylase<sup>39–41</sup>. Recently, it was shown that *Aspergillus oryzae* produces GAG with a degree of deacetylation near 50%<sup>42</sup>. Picomolar quantities of Agd3 are sufficient to cause soluble GAG adhesion in the GAG-ELLA assays suggesting that Agd3 is efficient, or that low levels of deacetylation are sufficient for adherence.

Agd3 has a conserved arginine, R613 in CM3, that coordinates the catalytic base, and is important for activity. This arginine is found much later in the primary sequence than the base activating arginine in MT3 of canonical CE4s<sup>43</sup>. However, R613 aligns structurally to the activating arginine found in IcaB, suggesting that it could play a similar role in GAG deacetylation. IcaB and other PNAG active CE4s, including PgaB from *E. coli*, contain an arginine separate from MT3 (Fig. 3c) which is more similar to Agd3 than canonical CE4s. Agd3 has catalytic motifs in the same order within the primary sequence as equivalent motifs of the CE4 family. The structural similarity results and equivalent order of CMs in the primary sequence suggests that Agd3 is distantly related to the CE4 family and may have diverged from a common ancestor.

Agd3 activity is metal-dependent as complete inhibition could be achieved with the addition of DPA (Fig. 4d). In contrast, addition of 12,500-fold molar excess EDTA did not inhibit Agd3 activity. The discrepancy between inhibitor effects can be explained by their differing modes of action. DPA has been reported to create a ternary complex with the bound metal, thus metal diffusion would not be a limiting step in inhibition<sup>44,45</sup>. EDTA is a larger molecule and diffusion of the metal from the metalloenzyme is required for sequestration in some cases<sup>44</sup>. The inability of EDTA, compared with DPA, to inactivate enzymes

has also been reported for PgaB and the chitin deacetylase CDA from *Collectotrichum lindemuthianum*. These results suggest a low diffusion rate of the metal ion from the Agd3 active site. The crystal structure contained a  $Zn^{2+}$  ion but addition of  $Mn^{2+}$  to the assay conditions significantly increased Agd3 activity (Fig. 4d). It is possible that Agd3 uses a different metal in vivo. Many CE4 enzymes are  $Zn^{2+}$  dependent but some have been found to be promiscuous. The biofilm state is often nutrient limited so metal promiscuity could be beneficial.

The similarity between the active site of Agd3 and that of the CE4 members supports a shared mechanism between the CE families (Fig. 4e). For Agd3, the proposed mechanism starts with the metal ion coordinating a water molecule that is deprotonated by D377 and attacks the acetate group creating a tetrahedral oxyanion intermediate. H668 can then donate a hydrogen to break the bond between the acetate leaving group and amine (Fig. 4e).

In addition to the central CE, Agd3 has two auxiliary domains. The N-terminal  $\alpha/\beta/\alpha$ -domain was found to have some structural similarity to the glutamine-transferase I family. Our findings support that Agd3<sup>141–364</sup> is a CBM capable of binding GAG. This domain does not belong to any of the current 86 families and is the first characterized CBM with a reductase-like fold<sup>30</sup>. CBMs are commonly found linked to glycoside hydrolases and increase the specificity and affinity of the enzyme for the substrate<sup>30,46</sup>. In some cases, CBMs can modulate the structure of the substrate allowing for binding by the active domain<sup>30</sup>. Some CE4 enzymes have been reported to have CBMs<sup>47</sup>. The activity of the *Podospira anserine* chitin de-*N*-acetylase on insoluble chitin was reduced by removal of one or both CBMs, although activity against soluble chitin was unaffected<sup>48</sup>. GAG is heterogeneous in composition and can be highly insoluble<sup>4–6</sup>. The Agd3 CBM may assist in binding of insoluble substrate and alignment of GAG within the active site, thus aiding in the solubilization of the acetylated polymer.

Phylogenetic analysis found that the three-domain architecture of Agd3 is present in both fungal and bacterial kingdoms. In contrast to the members of the CE4 family which do not have consistent domain arrangement, the CBM domain is present in the majority of the Agd3 homologs found<sup>21</sup>. Fungal Agd3 orthologues have been identified previously within GAG gene clusters<sup>5</sup>. The production of GAG-like polymers has been confirmed in *Aspergillus niger*, *Aspergillus parasiticus*, *A. oryzae*, *N. crassa*, *Penicillium frequentans*, *Paecilomyces* sp., and *Trichosporon asahii*<sup>5,40,42,49–53</sup>. Expression of poly-*N*-acetylgalactosamine deacetylases in both *A. parasiticus* and *N. crassa* has been confirmed and enzyme levels are linked to levels of polysaccharide produced suggesting co-regulation of polymer and deacetylase production<sup>40,54</sup>. In this study, we found Agd3 homologs in bacteria encoded within, or near, operons containing genes of analogous function to those of the GAG cluster. Some Gram-negative Agd3 homologs were located near genes with homology to Pel polysaccharide genes. Pel, like GAG, is a partially de-*N*-acetylated GalNAc-rich polymer involved in biofilm formation in some *P. aeruginosa* strains<sup>55–58</sup>. Pel operons have been found in *Proteobacteria*, including *Burkholderia* and *Moritella* species<sup>59</sup> and more recently in a wide range of Gram-positive bacteria<sup>60,61</sup>, suggesting that many bacteria produce a Pel/GAG-like polymer.

The synteny and composition of *agd3*-containing operons in Gram-positive organisms was variable. Many of the organisms identified are soil-dwelling and plant pathogens, thus inhabiting similar environments as the *Ascomycota* species. Low levels of galactosamine-containing polysaccharides have been found in the cell-walls of *Streptomyces* species<sup>62,63</sup>. Expression and secretion of an active endo- $\alpha$ -1,4-galactosaminidase has also been reported in *Streptomyces griseus*, suggesting that this species produces or encounters a GalNAc-rich polymer, similar to GAG or Pel<sup>64</sup>.

Previously, it had been suggested that the GAG gene cluster had been horizontally transferred from bacteria to a common ancestor of *Peizizomycotina*, due to the clustered nature of the genes and the similarity to bacterial synthase-dependent exopolysaccharide biosynthetic systems<sup>5</sup>. Our findings of operons containing most, or all, of the genes homologous to the GAG cluster adds further support to this hypothesis.

Agd3 is required for GAG maturation and full virulence in the mouse model of *A. fumigatus* infection. Structural and functional analysis of the domains within Agd3 have identified novel CE family, CE18, and CBM family, CBM87, domains. Our findings suggest that there are many Agd3 homologs within both fungal and bacterial species. Further experiments to determine the function and importance of Agd3 homologs will aid in our understanding of exopolysaccharide modification systems across kingdoms.

## Methods

**Aspergillus strains and growth conditions.** The wild-type *A. fumigatus* strain Af293 and Af293 with an *agd3* knockout (Af293  $\Delta$ agd3) produced by Lee et al.<sup>5</sup> were used in this study. Unless otherwise noted, strains were grown and harvested on YPD agar (Fisher Scientific) at 37 °C as previously described<sup>65</sup>. For growth in liquid medium, Brian medium<sup>4</sup>, AspMM<sup>66</sup>, and RPMI media 1640 (Wissent) were used as indicated. All strains and primers used in this study are summarized in Supplementary Table 2.

**Bioinformatics analysis.** The Agd3 sequence from *A. fumigatus* was analyzed using Phyre<sup>2</sup><sup>18</sup> and NetOGlyc 4.0 server<sup>19</sup>. Sequence alignment was done using Clustal Omega<sup>67</sup> and MUSCLE<sup>68</sup>.

**Agd3 expression and purification.** *A. fumigatus* Af293 *agd3* (Afu3g07870), was codon optimized for *P. pastoris* (Bio Basics Inc.) and received in pUC57 plasmids (pUC57-Agd3Pp). The region of pUC57-Agd3Pp encoding residues 141–806 (referred to as Agd3 herein) was cloned into a pET28a vector. This gene, with the addition of the N-terminal histidine tag and thrombin cleavage site encoded by the pET28a vector, was subsequently cloned into the pPinka-HC plasmid for expression in the PichiaPink<sup>TM</sup> secretion system (Invitrogen). Expression was induced with 1% (v/v) methanol for 16–24 h. Culture supernatants were filtered and the protein purified by precipitation with 65% (v/v) ammonium sulfate, dialysis, and size exclusion chromatography. Point mutants were expressed and purified as described above.

Agd3<sup>141–364</sup>, and point mutants thereof, were expressed and purified from *E. coli* Origami2 cells using the pET28a vector. Purification was performed using nickel affinity and size exclusion chromatography.

**Crystallization and structural determination.** Purified Agd3 was screened for crystallization conditions using sitting-drop vapor diffusion at 6 and 10 mg/mL. Crystals formed in 1.7 M ammonium sulfate, 1.7% (v/v) PEG 400, 85 mM HEPES pH 7.5, and 15% (v/v) glycerol (Rigaku TOP96 condition #8). Optimization through additives and streak seeding produced crystals in TOP96 #8 with 0.25% (v/v) PEG 3350. Two crystals were harvested and soaked in 1.36 M ammonium sulfate, 260 mM zinc sulfate, 85 mM HEPES pH 7.5, 2.36% (v/v) PEG 400, 14% (v/v) glycerol before vitrification in liquid nitrogen. Data were collected at the NSLS II using the Frontiers Macromolecular Crystallography beamline at –173 °C and a wavelength of 1.283 Å. Data sets were merged using XPrep and the structure solved using the SHELX suite<sup>69,70</sup>. SHELXE found six unique zinc ions with significant anomalous signal and was able to build poly-alanine traces for 315 amino acids. Residues 141–806 were built into initial maps using Coot through SBGrid<sup>71–74</sup>. There was no interpretable density for the hexahistidine tag. Maps were converted to .mtz format using F2MTZ through CCP4<sup>75–77</sup> and the refinement performed using PHENIX.REFINE with iterative real-space refinement in Coot. Once a model was determined with residual values in the low 20s, this model was used for molecular replacement against the highest resolution single data set. The anisotropy of this single dataset was corrected using the Diffraction Anisotropy Server<sup>78</sup>. Model refinement was done in real space in Coot and through PHENIX.REFINE. TLS groups were added with the TSLMD server in PHENIX<sup>71,72</sup>.

**Rescue of biofilm formation by addition of Agd3.** Af293  $\Delta$ agd3 and Af293 strains were cultured in Brian media with or without supplementation of BSA protein, Agd3<sup>H510A</sup>, or Agd3 wildtype for 21 h in 96-well round-bottom tissue-culture treated plates. Protein supplements were added at a final concentration of 0.5 nM. Biofilms were washed with deionized water, stained with 0.1% (w/v) crystal violet for 10 min and destained with 100% ethanol for 10 min. Absorbance of destain solution was measured at 600 nm. For biofilm imaging,

*Δagd3* and Af293 were cultured for 27 h in a 24-well flat bottom non-tissue-culture treated plate which was washed with PBS and stained with crystal violet.

**GAG deacetylation assay.** A GAG-ELLA was used to assess the activity of Agd3<sup>141–806</sup> WT or its point mutants. Briefly, 50 μL of protein, in PBS pH 7.4, was mixed with 50 μL of culture supernatants harvested from the Af293 *Δagd3* strain<sup>5</sup> in a 96-well clear flat-bottom Immulon 4HBX plate (ThermoFisher). Reactions were incubated at room temperature for 1 h on a mutator. Plates were washed with PBS-T three times before the addition of 100 μL of lectin solution (30 nM biotinylated soybean agglutinin [SBA-biotin, Vector Labs], 1/700 avidin horseradish peroxidase [Avidin-HRP, Invitrogen]). After incubation at room temperature for 1 h, plates were washed with PBS-T three times followed by addition of 100 μL of TMB Ultrasensitive Solution (Millipore). Plates were read at 370 nm or the reaction was stopped with 100 μL of 2 M H<sub>2</sub>SO<sub>4</sub> and read at 450 nm. For inhibitor and metal screening, chloride salts, EDTA, or DPA were incubated with the protein for 10 min before the addition of the culture supernatant.

**MALDI-TOF MS enzymatic fingerprint.** α-1,4-GalNAc oligosaccharides were obtained by a Sph3 partial hydrolysis. *A. fumigatus* biofilms were incubated with 5 nM Sph3 for 1 h at room temperature, solubilized oligosaccharides were then further purified on a Sep-pak C18 cartridge. In brief, cartridges were conditioned using absolute ethanol followed by water. Samples were then loaded onto the cartridge before washing and eluting using a 1% (v/v) step gradient of ACN from 1 to 4% (v/v). 3% (v/v) ACN fraction was used for the specificity study.

Purified oligosaccharide fractions were mixed with Agd3 WT or its mutants as indicated, and the samples analyzed by MS at multiple time points. Products of enzymatic reaction were diluted in 0.2% (v/v) TFA before being spotted on the MALDI-TOF plate in a ratio 1:1 (v/v) with 5 mg/mL DHB matrix reconstituted in ACN: 0.2% (v/v) TFA (70:30, v-v). Spectra were recorded on a Bruker UltrafleXtreme in positive reflector mode and an accumulation of 5000 laser shots. MALDI-TOF MS/MS experiments were performed using the same mass spectrometer. Data was collected and analyzed using Bruker Flex Control v.3.4 and Bruker Flex Analysis v.3.4.

**Investigation of Agd3 specificity.** Oligosaccharides of chitin were produced by acidic partial hydrolysis of chitin from shrimp shell (Sigma). Briefly, chitin was incubated in 0.1 M HCl for 3 h at 100 °C. Solubilized oligosaccharides of chitin were then re-N-acetylated by incubation in methanol: pyridine: anhydride acetic (10:2:3) for 1 h at room temperature. Oligosaccharides were then purified on a Hypercarb SPE cartridge (ThermoFisher) conditioned as per manufacturer instructions. After loading the sample, the cartridge was washed with water, 5% (v/v) ACN, and oligosaccharides eluted with 50% (v/v) ACN.

Purified chitin oligosaccharides were then incubated with 10 μM Agd3 in 1X PBS for 24 h. Results of the incubation were then spotted on the MALDI-TOF plate in a ratio 1:1 (v/v) with 5 mg/ml DHB matrix reconstituted in ACN: 0.2% (v/v) TFA (70:30, v-v). Spectra were recorded on a Bruker UltrafleXtreme in positive reflector mode and an accumulation of 5000 laser shots. Data were collected and analyzed with Bruker Flex Control v.3.4 and Bruker Flex Analysis v.3.4.

**Fluorescence microscopy of *A. fumigatus* hyphae.** Hyphae were grown as above on tissue culture treated cover slips, washed twice with PBS and stained with 120 μg/ml of the GalNAc specific lectin Soybean Agglutinin lectin (SBA) conjugated to fluorescein and 10 μM of Alexa Fluor<sup>®</sup>-568 labelled Agd3<sup>141–364</sup> for 2 h, at 4 °C. Samples were washed twice with PBS and fixed with 4% (w/v) paraformaldehyde for 20 min at 4 °C. Samples were washed twice with PBS and mounted with SlowFade<sup>®</sup> mounting media and sealed. Samples were imaged on a Zeiss confocal microscope with 488 and 561 nm lasers to excite fluorescein and Alexa Fluor<sup>®</sup>-568, respectively. The channels were acquired separately, with fluorescein being detected from 490–577 nm and Alexa Fluor<sup>®</sup>-568 from 577 to 691 nm.

**GAG binding assay.** Undiluted culture supernatants were incubated in Immulon 4HBX 96-well microtiter plates (ThermoFisher) overnight at 4 °C. Wells were washed 3× with wash buffer [PBS, 0.05% (v/v) Tween-20] and blocked for 1 h in blocking buffer [2% (w/v) BSA in wash buffer]. Wells were washed 1× with wash buffer and incubated with the indicated concentrations of Agd3 diluted in blocking buffer for 2 h at room temperature. Wells were washed 3× with wash buffer and incubated with mouse anti-His-Tag (Abgent) diluted 1/1000 in blocking buffer for 45 min. Wells were then washed 3× with wash buffer and incubated with donkey anti-mouse secondary antibody (Cedarlane) conjugated to horseradish peroxidase diluted 1/1000 in blocking buffer for 45 min. Wells were then washed 4× with wash buffer and incubated with TMB substrate (ThermoFisher) for 6 min at room temperature. The reaction was stopped with the addition of 1 M H<sub>2</sub>SO<sub>4</sub> and absorbance read at 450 nm.

**Synthesis of α-1,4-linked oligosaccharides.** Gal and galactosamine oligosaccharides were synthesized as described previously<sup>8</sup>. This method was also used

to generate N-acetylgalactosamine and mixed N-acetylgalactosamine/galactosamine oligosaccharides. Briefly, oligosaccharides were chemically synthesized using building blocks protected with di-*tert*-butylsilylene. These cis-galactosylating agents are extremely selective and only produced α-isomer products. After glycosylation reactions, the di-*tert*-butylsilylene group was removed with HF pyridine, and the free hydroxyl was protected with benzoyl group. Through a linear stepwise synthesis strategy, the following oligosaccharides were made: α-1,4-(Gal)<sub>6</sub>, α-1,4-(GalNAc)<sub>6–8</sub>, α-1,4-(GalN)<sub>6</sub>, and α-1,4-(GalN-α-1,4-GalNAc)<sub>3</sub>. Finally, deprotection by saponification, debenzoylation, azide reduction and acetylation of amine groups was performed to obtain the oligosaccharides. These compounds have been characterized with <sup>1</sup>H-nuclear magnetic resonance (NMR), <sup>13</sup>C-NMR, HH-correlation spectroscopy, heteronuclear single quantum coherence spectroscopy, and high resolution MS.

**Electrospray ionization mass spectrometry (ESI-MS).** Protein samples were buffer exchanged into 200 mM aqueous ammonium acetate (pH 7.2) using a 10 kDa cutoff Amicon 0.5 mL microconcentrators (EMD, Millipore, Billerica, MA). Stock solutions of oligosaccharides (10 mM) were prepared by dissolving a known amount of solid compound in milliQ water and stored at –20 °C until used. Nano-electrospray (nanoESI) mass spectrometry measurements were performed on a Synapt G2S quadrupole-ion mobility separation-time of flight (Q-IMS-TOF) mass spectrometer (Waters, Manchester, UK) equipped with nanoflow ESI source. NanoESI was performed by applying a voltage of ~1 kV to a platinum wire inserted into the nanoESI tip, which was produced from a borosilicate glass capillary (1.0 mm outside diameter [o.d.], 0.78 mm inside diameter [i.d.]) pulled to ~5 μm o.d. using a P–1000 micropipette puller (Sutter Instruments, Novato, CA). The source temperature was 60 °C and gas flow rates was 2 mL/min. The cone, trap and transfer voltages were 20 V, 3 V, and 1 V, respectively. MassLynx software (version 4.1) was used for data acquisition and processing. Mass spectra were averaged over 300 scans.

Association constants ( $K_a$ ) for the Agd3<sup>141–364</sup>-oligosaccharide interactions were quantified using the direct ESI-MS assay<sup>79,80</sup>. The binding measurements were carried out at room temperature, the reaction mixtures were prepared by mixing aliquots of the stock solutions to achieve the desired concentration of Agd3<sup>141–364</sup> (5 μM), oligosaccharide (50–350 μM), and ammonium acetate (200 mM). ESI-MS measurements were performed after 1 h incubation time period. The reported affinities are average values from three replicate measurements performed at a minimum of five different ligand concentrations (protein initial concentration was constant). All mass spectra were corrected, when needed, for the occurrence of nonspecific carbohydrate-protein binding during the ESI process using the reference protein method<sup>81</sup>.  $K_a$  values were determined from the abundance ratio ( $R$ ) of the ligand-bound ( $PL$ ) to free protein ( $P$ ) ions, after correction for nonspecific ligand binding, and the initial concentrations of protein ( $[P]_0$ ) and ligand ( $[L]_0$ ), Eq (1):

$$K_a = R / ([L]_0 - (R/[1 + R])[P]_0)$$

where  $R$  is taken to be equal<sup>79</sup> to the corresponding concentration ratio ( $[PL]/[P]$ ) in solution, Eq (2):

$$R = \frac{\sum Ab(PL)}{\sum Ab(P)} = \frac{[PL]}{[P]}$$

**Phylogenetics.** The amino acid sequence of Agd3 from *A. fumigatus* submitted to BLASTP<sup>82</sup> with a cutoff of 1e-05 identified 1560 sequences from 314 species. 744 sequences with between 20 and 95% sequence identity to *A. fumigatus* Agd3 and with >50% coverage were downloaded. Agd3 was also submitted to JackHMMER<sup>83</sup>. 121 representative sequences were selected from the BLASTP search and five from JackHMMER, by eliminating those with pairwise identity over 80%, removing protein fragments, allowing for a maximum of one sequence per species. Sequences were aligned by MUSCLE<sup>68</sup> and MEGA7<sup>84</sup> then used to construct a maximum-likelihood tree using sites with >95% coverage and 250 bootstrap replicates. Branches were retained at a bootstrap level of 0.7 and the tree was viewed using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

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

## Data availability

Crystallographic data that support the findings of this study have been deposited in the Protein Data Bank with the accession codes 6NWZ [10.2210/pdb6NWZ/pdb]. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary information files. The source data underlying Figs. 4b–d, 5b, 6b–d, 7a–c, and Supplementary Figs. 2 and 5 are provided as a Source Data file.

Received: 31 October 2019; Accepted: 16 April 2020;  
Published online: 15 May 2020



## References

- Kaur, S. & Singh, S. Biofilm formation by *Aspergillus fumigatus*. *Med. Mycol.* **52**, 2–9 (2014).
- Brown, G. D. et al. Hidden killers: human fungal infections. *Sci. Transl. Med.* **4**, 165rv13–165rv13 (2012).
- Lewis, R. E. et al. Epidemiology and sites of involvement of invasive fungal infections in patients with haematological malignancies: a 20-year autopsy study. *Mycoses* **56**, 638–645 (2013).
- Fontaine, T. et al. Galactosaminogalactan, a new immunosuppressive polysaccharide of *Aspergillus fumigatus*. *PLoS Pathog.* **7**, e1002372 (2011).
- Lee, M. J. et al. Deacetylation of fungal exopolysaccharide mediates adhesion and biofilm formation. *mBio* **7**, e00252–16 (2016).
- Bamford, N. C. et al. Sph3 is a glycoside hydrolase required for the biosynthesis of galactosaminogalactan in *Aspergillus fumigatus*. *J. Biol. Chem.* **290**, 27438–27450 (2015).
- Lee, M. J. et al. Overlapping and distinct roles of *Aspergillus fumigatus* UDP-glucose 4-epimerases in galactose metabolism and the synthesis of galactose-containing cell wall polysaccharides. *J. Biol. Chem.* **289**, 1243–1256 (2014).
- Le Mauff, F. et al. Molecular mechanism of *Aspergillus fumigatus* biofilm disruption by fungal and bacterial glycoside hydrolases. *J. Biol. Chem.* **294**, 10760–10772 (2019).
- Bamford, N. C. et al. Ega3 from the fungal pathogen is an endo- $\alpha$ -1,4-galactosaminidase that disrupts microbial biofilms. *J. Biol. Chem.* **294**, 13833–13849 (2019).
- Ostapska, H., Howell, P. L. & Sheppard, D. C. Deacetylated microbial biofilm exopolysaccharides: It pays to be positive. *PLoS Pathog.* **14**, e1007411–e1007418 (2018).
- Sheppard, D. C. & Howell, P. L. Biofilm exopolysaccharides of pathogenic fungi: lessons from bacteria. *J. Biol. Chem.* **291**, 12529–12537 (2016).
- Whitfield, G. B., Marmont, L. S. & Howell, P. L. Enzymatic modifications of exopolysaccharides enhance bacterial persistence. *Front. Microbiol.* **6**, 7350 (2015).
- Vuong, C. et al. A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence. *J. Biol. Chem.* **279**, 54881–54886 (2004).
- Little, D. J. et al. The protein BpsB is a poly- $\beta$ -1,6-N-acetyl-D-glucosamine deacetylase required for biofilm formation in *Bordetella bronchiseptica*. *J. Biol. Chem.* **290**, 22827–22840 (2015).
- Itoh, Y. et al. Roles of pgaABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine. *J. Bacteriol.* **190**, 3670–3680 (2008).
- Köseoğlu, V. K. et al. *Listeria monocytogenes* exopolysaccharide: origin, structure, biosynthetic machinery and c-di-GMP-dependent regulation. *Mol. Microbiol.* **96**, 728–743 (2015).
- Colvin, K. M. et al. PelA deacetylase activity is required for Pel polysaccharide synthesis in *Pseudomonas aeruginosa*. *J. Bacteriol.* **195**, 2329–2339 (2013).
- Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).
- Stentoft, C. et al. Precision mapping of the human O-GalNAc glycoproteome through simple cell technology. *EMBO J.* **32**, 1478–1488 (2013).
- Gonzalez, M., Brito, N. & Gonzalez, C. High abundance of serine/threonine-rich regions predicted to be hyper-O-glycosylated in the secretory proteins coded by eight fungal genomes. *BMC Microbiol.* **12**, 213 (2012).
- Aragunde, H., Biarnés, X. & Planas, A. Substrate recognition and specificity of chitin deacetylases and related family 4 carbohydrate esterases. *IJMS* **19**, 412–430 (2018).
- Shaik, M. M., Cendron, L., Percudani, R. & Zanotti, G. The structure of *Helicobacter pylori* HP0310 reveals an atypical peptidoglycan deacetylase. *PLoS One* **6**, e19207–e19208 (2011).
- Fadoulglou, V. E. et al. Unusual  $\alpha$ -carbon hydroxylation of proline promotes active-site maturation. *J. Am. Chem. Soc.* **139**, 5330–5337 (2017).
- Little, D. J. et al. Structural basis for the de-N-acetylation of poly- $\beta$ -1,6-N-acetyl-D-glucosamine in Gram-positive bacteria. *J. Biol. Chem.* **289**, 35907–35917 (2014).
- Blair, D. E. & van Aalten, D. M. F. Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with N-acetyl-glucosamine. *FEBS Lett.* **570**, 13–19 (2004).
- Little, D. J. et al. The structure- and metal-dependent activity of *Escherichia coli* PgaB provides insight into the partial de-N-acetylation of poly- $\beta$ -1,6-N-acetyl-D-glucosamine. *J. Biol. Chem.* **287**, 31126–31137 (2012).
- Jana, S. et al. Aromatic-mediated carbohydrate recognition in processive *Serratia marcescens* chitinases. *J. Phys. Chem. B* **120**, 1236–1249 (2016).
- Oda, M., Inaba, S., Kamiya, N., Bekker, G.-J. & Mikami, B. Structural and thermodynamic characterization of endo-1,3- $\beta$ -glucanase: insights into the substrate recognition mechanism. *Biochim. Biophys. Acta* **1866**, 415–425 (2018).
- Zakariassen, H. et al. aromatic residues in the catalytic center of chitinase A from *Serratia marcescens* affect processivity, enzyme activity, and biomass converting efficiency. *J. Biol. Chem.* **284**, 10610–10617 (2009).
- 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).
- Blake, A. W. et al. Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *J. Biol. Chem.* **281**, 29321–29329 (2006).
- Payne, C. M. et al. Hallmarks of processivity in glycoside hydrolases from crystallographic and computational studies of the *Serratia marcescens* chitinases. *J. Biol. Chem.* **287**, 36322–36330 (2012).
- Tesmer, J. J., Klem, T. J., Deras, M. L., Davisson, V. J. & Smith, J. L. The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families. *Nat. Struct. Biol.* **3**, 74–86 (1996).
- Li, X.-L., Skory, C. D., Cotta, M. A., Puchart, V. & Biely, P. Novel family of carbohydrate esterases, based on identification of the *Hypocrea jecorina* acetyl esterase gene. *Appl. Environ. Microbiol.* **74**, 7482–7489 (2008).
- Pokrovskaya, V. et al. Functional characterization of *Staphylococcus epidermidis* IcaB, a de-N-acetylase important for biofilm formation. *Biochemistry* **52**, 5463–5471 (2013).
- Wang, X., Preston, J. F. & Romeo, T. The pgaABCD locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J. Bacteriol.* **186**, 2724–2734 (2004).
- Nwe, N., Furuike, T. & Tamura, H. The mechanical and biological properties of chitosan scaffolds for tissue regeneration templates are significantly enhanced by chitosan from *Gongronella butleri*. *Materials* **2**, 374–398 (2009).
- Vollmer, W. & Tomasz, A. The pgaA gene encodes for a peptidoglycan N-acetylglucosamine deacetylase in *Streptococcus pneumoniae*. *J. Biol. Chem.* **275**, 20496–20501 (2000).
- Jensen, J. W., Debusk, R. M. & Debusk, A. G. Induction of cellular efflux by a galactosamine polymer from *Neurospora crassa*. *J. Gen. Microbiol.* **130**, 557–565 (1984).
- Jorge, J. A., Kinney, S. G. & Reissig, J. L. Purification and characterization of *Neurospora crassa* N-acetyl galactosaminoglycan deacetylase. *Braz. J. Med. Biol. Res.* **15**, 29–34 (1982).
- Jorge, J. A., de Almeida, E. M., de Lourdes Polizeli, M. & Terenzi, H. F. Changes in N-acetyl galactosaminoglycan deacetylase levels during growth of *Neurospora crassa*: effect of L-sorbose on enzyme production. *J. Basic Microbiol.* **39**, 337–344 (1999).
- Miyazawa, K. et al. Both galactosaminogalactan and  $\alpha$ -1,3-glucan contribute to aggregation of *Aspergillus oryzae* hyphae in liquid culture. *Front. Microbiol.* **10**, 2090 (2019).
- Blair, D. E., Schuttelkopf, A. W., MacRae, J. I. & van Aalten, D. M. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl Acad. Sci.* **102**, 15429–15434 (2005).
- Kidani, Y. & Hirose, J. Coordination chemical studies on metalloenzymes. II. kinetic behavior of various types of chelating agents towards bovine carbonic anhydrase. *J. Biochem.* **81**, 1383–1391 (1977).
- Kidani, Y., Hirose, J. & Koike, H. Coordination chemical studies on metalloenzymes. Kinetics and mechanism of the Zn(II) exchange reaction between chelating agent and apo-bovine carbonic anhydrase. *J. Biochem.* **79**, 43–51 (1976).
- Mizutani, K., Sakka, M., Kimura, T. & Sakka, K. Essential role of a family-32 carbohydrate-binding module in substrate recognition by *Clostridium thermocellum* mannanase CtMan5A. *FEBS Lett.* **588**, 1726–1730 (2014).
- Andrés, E. et al. Structural basis of chitin oligosaccharide deacetylation. *Angew. Chem. Int. Ed.* **53**, 6882–6887 (2014).
- Hofsbach, J. et al. A chitin deacetylase of *Podospora anserina* has two functional chitin binding domains and a unique mode of action. *Carbohydr. Polym.* **183**, 1–10 (2018).
- Bardalaye, P. C. & Nordin, J. H. Galactosaminogalactan from cell walls of *Aspergillus niger*. *J. Bacteriol.* **125**, 655–669 (1976).
- Ruperez, P. & Leal, J. A. Extracellular galactosaminogalactan from *Aspergillus parasiticus*. *Trans. Br. Mycol. Soc.* **77**, 621–625 (1981).
- Distler, J. J. & Roseman, S. Galactosamine polymers produced by *Aspergillus parasiticus*. *J. Biol. Chem.* **235**, 2538–2541 (1960).
- Takada, H., Araki, Y. & Ito, E. Structure of polygalactosamine produced by *Aspergillus parasiticus*. *J. Biochem.* **89**, 1265–1274 (1981).
- Guerrero, C., Prieto, A. & Leal, J. A. Extracellular galactosaminogalactan from *Penicillium frequentans*. *Microbiologia* **4**, 39–46 (1988).
- Araki, Y., Takada, H., Fujii, N. & Ito, E. A pathway of polygalactosamine formation in *Aspergillus parasiticus*: enzymatic deacetylation of N-acetylated polygalactosamine. *Eur. J. Biochem.* **102**, 35–42 (1979).
- Jennings, L. K. et al. Pel is a cationic exopolysaccharide that cross-links extracellular DNA in the *Pseudomonas aeruginosa* biofilm matrix. *Proc. Natl Acad. Sci.* **112**, 11353–11358 (2015).



56. Vasseur, P., Vallet-Gely, I., Soscia, C., Genin, S. & Filloux, A. The pel genes of the *Pseudomonas aeruginosa* PAK strain are involved at early and late stages of biofilm formation. *Microbiology* **151**, 985–997 (2005).
57. Ryder, C., Byrd, M. & Wozniak, D. Role of polysaccharides in *Pseudomonas aeruginosa* biofilm development. *Curr. Opin. Microbiol.* **10**, 644–648 (2007).
58. Ghafour, A., Hay, I. D. & Rehm, B. H. A. Role of exopolysaccharides in *Pseudomonas aeruginosa* biofilm formation and architecture. *Appl. Environ. Microbiol.* **77**, 5238–5246 (2011).
59. Marmont, L. S. et al. Oligomeric lipoprotein PelC guides Pel polysaccharide export across the outer membrane of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci.* **114**, 2892–2897 (2017).
60. Whitfield, G. B. et al. Discovery and characterization of a Gram-positive Pel polysaccharide biosynthetic gene cluster. *PLoS Pathog.* <https://doi.org/10.1371/journal.ppat.1008281> (2020).
61. Bundalovic-Torma, C., Whitfield, G. B., Marmont, L. S., Howell, P. L. & Parkinson, J. A systematic pipeline for classifying bacterial operons reveals the evolutionary landscape of biofilm machineries. *PLOS Comput. Biol.* **16**, e1007721 (2020).
62. Inoue, T., Murakawa, S. & Endo, A. Production and partial properties of an extracellular polysaccharide from *Streptomyces* sp. A-1845. *J. Ferment. Bioeng.* **73**, 440–442 (1992).
63. Cummins, C. S. & Harris, H. Studies on the cell-wall composition and taxonomy of Actinomycetales and related groups. *J. Gen. Microbiol.* **18**, 173–189 (1958).
64. Reissig, J. L., Lai, W.-H. & Glasgow, J. E. An endogalactosaminidase of *Streptomyces griseus*. *Can. J. Biochem.* **53**, 1237–1249 (1975).
65. Gravelat, F. N. et al. *Aspergillus fumigatus* MedA governs adherence, host cell interactions and virulence. *Cell Microbiol.* **12**, 473–488 (2010).
66. Gravelat, F. N., Askew, D. S. & Sheppard, D. C. Targeted gene deletion in *Aspergillus fumigatus* using the hygromycin-resistance split-marker approach. *Methods Mol. Biol.* **845**, 119–130 (2012).
67. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).
68. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
69. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr.* **64**, 112–122 (2008).
70. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 479–485 (2010).
71. 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).
72. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **58**, 1948–1954 (2002).
73. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
74. Morin, A. et al. Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
75. Pottert, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **59**, 1131–1137 (2003).
76. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 235–242 (2011).
77. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **50**, 760–763 (1994).
78. Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Nat. Acad. Sci.* **103**, 8060–8065 (2006).
79. Kitova, E. N., El-Hawiet, A., Schnier, P. D. & Klassen, J. S. Reliable determinations of protein–ligand interactions by direct ESI-MS measurements. Are we there yet? *J. Am. Soc. Mass Spectrom.* **23**, 431–441 (2012).
80. Wang, W., Kitova, E. N. & Klassen, J. S. Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nano-electrospray Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **75**, 4945–4955 (2003).
81. Sun, J., Kitova, E. N., Wang, W. & Klassen, J. S. Method for distinguishing specific from nonspecific protein-ligand complexes in nano-electrospray ionization mass spectrometry. *Anal. Chem.* **78**, 3010–3018 (2006).
82. Altschul, S. F. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
83. Potter, S. C. et al. HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
84. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
85. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533 (2010).
86. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Nat. Acad. Sci.* **98**, 10037–10041 (2001).

## Acknowledgements

We would like to thank: (i) the Drug Discovery platform of the Research Institute of the McGill University Health Center for the use of the MALDI-TOF mass spectrometer, and (ii) the Structural and Biophysical Core Facility at the Hospital for Sick Children for the use of the X-ray crystallography equipment. We appreciate the assistance of Patrick Yip in optimization of protein expression and Dr. John Parkinson for helpful discussion regarding the bioinformatics presented herein. Research described in this paper is supported by operating grants from the Canadian Institutes of Health Research (CIHR) (#81361 to P.L.H. and D.C.S., FDN-154327 and #43998 to P.L.H. and 123306 and FDN-159902 to D.C.S.), Cystic Fibrosis Canada (D.C.S. and P.L.H.) European Research Council (ERC) (J.D.C.C. #ERC-CoG-726072-GLYCONTROL). N.C.B. has been supported in part by Vanier and CGS-M graduate scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC), Mary H. Beatty, and Dr. James A. and Connie P. Dickson Scholarships from the University of Toronto, Cystic Fibrosis Canada, and The Hospital for Sick Children. P.L.H. is the recipient of a Tier I Canada Research Chair. D.C.S. is supported by a Distinguished Research Scholar award from the Fonds de Recherche Quebec Santé (FRQS). The Life Science Biomedical Technology Research resource at NSLS-II is primarily supported by the National Institute of Health, National Institute of General Medical Sciences (NIGMS) through a Biomedical Technology Research Resource P41 grant (P41GM111244), and by the Department of Energy Office of Biological and Environmental Research (KP1605010). Funds for the X-ray facilities at The Hospital for Sick Children were provided in part by the Canadian Foundation for Innovation and the Government of Ontario.

## Author contributions

All authors edited the final manuscript. N.C.B., F.L.M., D.C.S., and P.L.H. were involved with project design. N.C.B. determined the structure, characterized the activity of the protein in vitro and performed the bioinformatics analysis; F.L.M. performed the MS analysis; Y.Z. and J.D.C.C. synthesized the oligosaccharides; E.N.K. and J.S.K. performed the MS based affinity measurements with the oligosaccharides, H.O. and B.D.S. performed the ex vivo Agd3 complementation assay and in vivo *A. fumigatus* CBM binding experiment, respectively. N.C.B. and J.V.L. generated and characterized the CBM mutants.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-16144-5>.

Correspondence and requests for materials should be addressed to D.C.S. or P.L.H.

Peer review information *Nature Communications* thanks Gerlind Sulzenbacher and the other, anonymous, reviewers 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