

This is a repository copy of *Structural and functional characterisation of three novel fungal amylases with enhanced stability and pH tolerance*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/151337/>

Version: Accepted Version

Article:

Roth, Christian orcid.org/0000-0001-5806-0987, Moroz, Olga V., Turkenburg, Johan P. et al. (8 more authors) (2019) Structural and functional characterisation of three novel fungal amylases with enhanced stability and pH tolerance. *International Journal of Molecular Sciences*. 4902. ISSN 1422-0067

<https://doi.org/10.3390/ijms20194902>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



1 Article

2 Structural and functional characterisation of three 3 novel fungal amylases with enhanced stability and 4 pH tolerance

5 Christian Roth ^{1,4#}, Olga V Moroz ^{1#}, Johan P. Turkenburg¹, Elena Blagova¹, Jitka Waterman ^{1,5},
6 Antonio Ariza^{1,6}, Li Ming², Sun Tianqi², Carsten Andersen³, Gideon J Davies¹ and Keith S
7 Wilson ^{1*}

8 ¹ York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, YO10
9 5DD, UK

10 ² Novozymes (China) Investment Co. Ltd, 14 Xinli Road, Haidian District, Beijing 100085, People's Republic
11 of China

12 ³ Novozymes (Denmark), Krogshøjvej 36, DK-2880 Bagsvaerd, Denmark

13 ⁴ Present address: Carbohydrates: Structure and Function, Biomolecular Systems, Max Planck Institute of
14 Colloids and Interfaces, 14195 Berlin, Germany

15 ⁵ Present address: Diamond Light Source, Diamond House, Harwell Science and Innovation Campus, Fermi
16 Ave, Didcot OX11 0DE

17 ⁶ Present address: Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK

18 #The first two authors contributed equally to this work.

19 * Correspondence: keith.wilson@york.ac.uk +44 1904 328262

20 Received: date; Accepted: date; Published: date

21 **Abstract:** Amylases are probably the best studied glycoside hydrolases and have a huge
22 biotechnological value for industrial processes on starch. Multiple amylases from fungi and
23 microbes are currently in use. Whereas bacterial amylases are well suited for many industrial
24 processes due to their high stability, fungal amylases are recognized as safe and are preferred in the
25 food industry, although they lack the pH tolerance and stability of their bacterial counterparts. Here,
26 we describe three amylases, two of which have a broad pH spectrum extending to pH 8 and higher
27 stability well suited for a broad set of industrial applications. These enzymes have the characteristic
28 GH13 α -amylase fold with a central (β/α)_s-domain, an insertion domain with the canonical calcium
29 binding site and a C-terminal β -sandwich domain. The active site was identified based on the
30 binding of the inhibitor acarbose in form of a transglycosylation product, in the amylases from
31 *Thamnidium elegans* and *Cordyceps farinosa*. The three amylases have shortened loops flanking the
32 nonreducing end of the substrate binding cleft, creating a more open crevice. Moreover, a potential
33 novel binding site in the C-terminal domain of the *Cordyceps* enzyme was identified, which might
34 be part of a starch interaction site. In addition, *Cordyceps farinosa* amylase presented a successful
35 example of using the microseed matrix screening technique to significantly speed-up crystallization.

36 **Keywords:** *α -amylase*; starch degradation; biotechnology; structure

38 1. Introduction

39 The use of enzymes in industrial processes is a multi-billion-dollar market. One of the first
40 enzymes discovered in 1833 was diastase, an enzyme able to hydrolyze starch [1]. Nowadays,
41 amylases, also able to hydrolyze starch, constitute up to 25% of the market for enzymes and have
42 virtually replaced chemical methods for degrading starch in the industrial sector (reviewed in [2]).
43 Amylases are the most important class of enzymes for degrading starch and can be subdivided into
44 three subclasses: α -, β -, and gluco-amylases based on their reaction specificity and product profiles.

45 α -amylases degrade the α -1,4 linkage between adjacent glucose units and are extensively used for
46 example in bioethanol production or in washing powder and detergents [3] (and reviewed in [4]).
47 One of the most widely used α -amylases is that from *Bacillus licheniformis*, known under the
48 tradename “Termamyl”. Microbial amylases are generally used in detergent applications and other
49 industrial processes, including bioethanol production, with new amylases, in particular those from
50 hyperthermophilic organisms, offering further improvement in the production process (reviewed in
51 [5]).

52 α -amylases belong to glycoside hydrolase family 13 (GH13) in the CAZy database classification
53 [6]. They have a (β/α)₈ barrel domain harbouring the active site, a subdomain which includes the
54 canonical calcium binding site inserted between the third β -strand and the third α -helix and a C-
55 terminal β -sandwich domain, thought to be important for the interaction with raw starch (reviewed
56 in [7]), [8, 9]. Amylases follow a retaining mechanism with an aspartate as nucleophile and one
57 glutamate as general acid/base [10, 11]. Up to ten consecutive sugar subsites forming the active site
58 cleft, have been identified in bacterial amylases, [12].

59 To date, recombinant fungal amylases have been isolated from mesophilic hosts such as
60 *Aspergillus oryzae* and are of particular interest to the food industry as they match the temperature
61 and pH range used in typical applications in the baking process, where they are active in the dough
62 but inactivated during baking. Due to the widespread use of fungal enzymes for the production of
63 food and food ingredients (such as citric acid), they are classified as GRAS (generally recognized as
64 safe) organisms by organizations including the FDA (US Food and Drug Administration) [13].

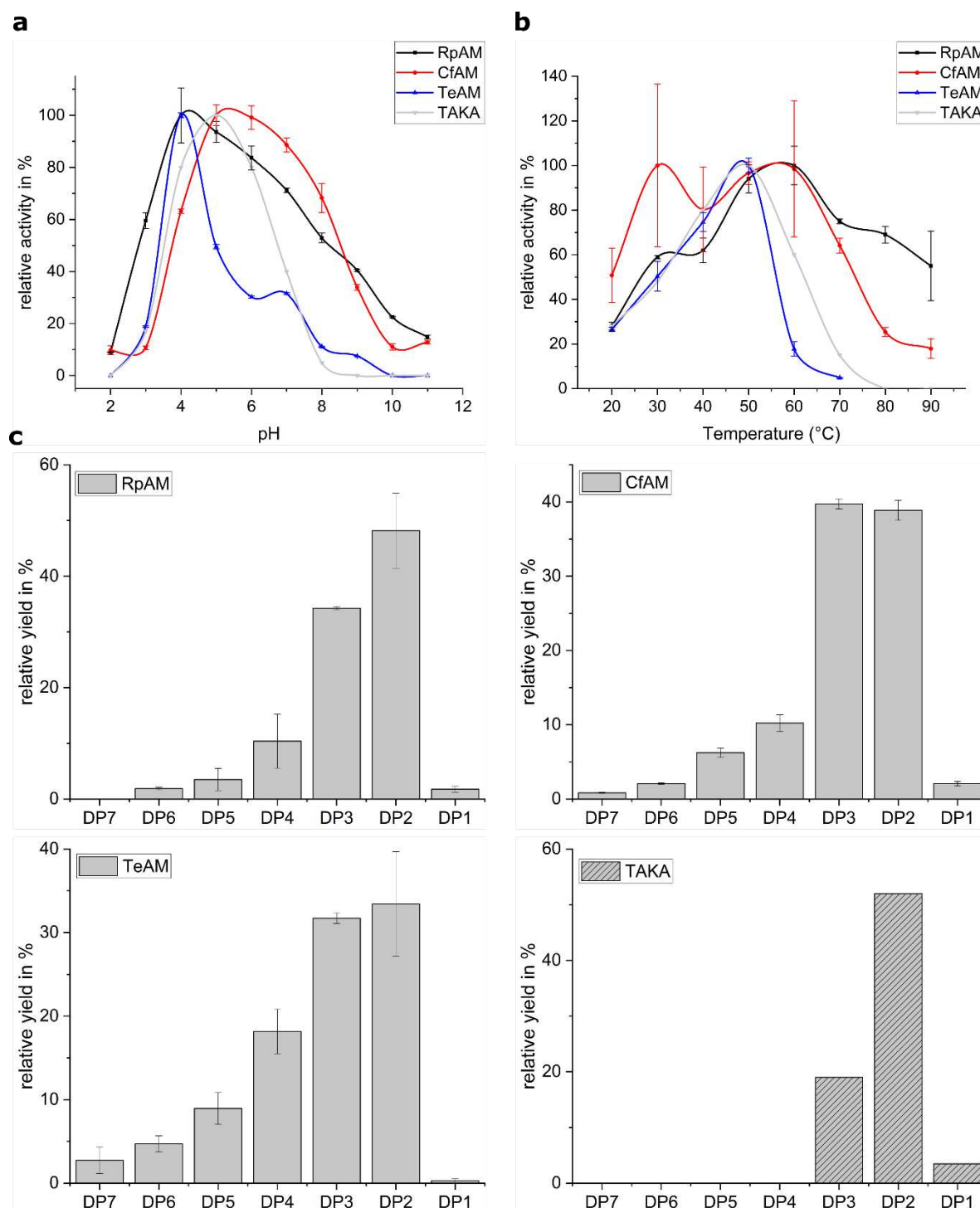
65 Up till now, fungal enzymes with a higher pH-tolerance and thermostability have not been
66 reported. Here we describe the structure and function of three novel α -amylases from *Cordyceps*
67 *farinosa* (CfAM), *Rhizomucor pusillus* (RpAM) and *Thamnidium elegans* (TeAM) with a higher stability
68 and pH-tolerance with the potential to act as novel biocatalysts for various industrial processes. The
69 sequence of all three enzymes groups them in the GH13 sub-family 1 along with, for example, the
70 amylase from *Aspergillus oryzae* (also known as TAKA amylase). However, unlike other fungal
71 amylases, the enzymes in this study have been shown to have a broad pH profile with an optimum
72 around pH 5 while retaining activity at pH 8. Furthermore, their more open crevice leads to the
73 production of longer oligomers compared to TAKA amylase.

74 The native RpAM and TeAM have a four-domain fold with a carbohydrate binding domain
75 (CBM20) at the C-terminus and a short serine-rich linker in between, while native CfAM lacks this
76 CBM20 domain. In this study, only the core of the amylases including the A, B and C domains was
77 cloned and expressed. In addition, crystallisation of *Cordyceps farinosa* amylase again demonstrates
78 the power of the microseed matrix screening technique [14].

79 2. Results

80 2.1. Biochemical characterization

81 The pH-, temperature- and product profiles were characterized for all three amylases. Of great
82 desire are amylases with a broader pH-tolerance compared to TAKA amylase. Our analysis showed
83 that all three amylases have a pH optimum around 5. Whereas TeAM has no significant activity above
84 pH 7, RpAM and CfAM retain significant activity at pH 7 extending up to a pH of 9 (Figure 1a). In
85 particular CfAM shows the highest pH-tolerance, retaining 70 % of its activity at pH 8. RpAM and
86 TeAM both show a pronounced shoulder suggesting the involvement of more titratable residues in
87 the substrate recognition and catalysis process. The temperature profiles reveal that RpAM and
88 CfAM also have a considerably higher thermotolerance compared to TAKA and TeAM (Figure 1b).
89 In particular, RpAM retains full activity even at 80°C, making it an attractive enzyme for industrial
90 high temperature starch saccharification processes. Compared to TAKA amylase, all three amylases
91 show a tendency to produce higher amounts of oligomers with a degree of polymerization (dp) of
92 three, with trace amounts of oligomers with a dp of up to seven for TeAM (Figure 1c).



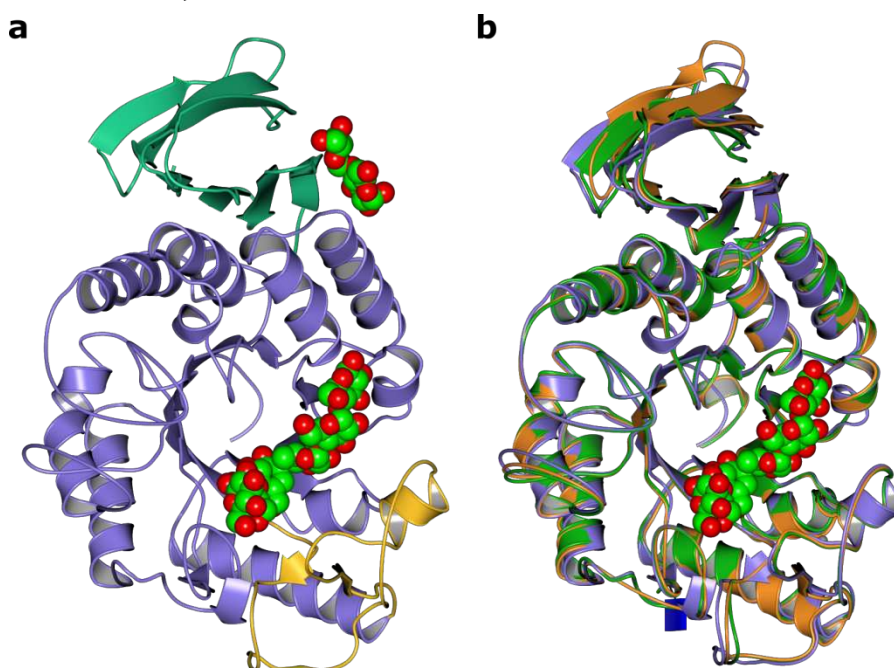
93

94 **Figure 1.** Biochemical characterization of RpAM, CfAM, TeAM and TAKA. (a) pH-profile of all three
 95 amylases in comparison with TAKA amylase. (b) Temperature profile of all three amylases in
 96 comparison with TAKA amylase. (c) Product profile of all three amylases and the abundance of
 97 oligomers with a degree of polymerization (dp) of 1 to 7 after hydrolysis of starch.

98 2.2. Overall fold

99 The structures were solved using molecular replacement starting from the *A. oryzae* amylase as
 100 template (pdb-ID: 7taa and 3vx0) to a resolution of 1.4 Å for RpAM, 1.2 Å for TeAM and 1.35 Å for
 101 CfAM respectively. The final model of RpAM includes two monomers in the asymmetric unit
 102 comprising residues 1 to 438 in both chains, which superpose on each other with an r.m.s.d. of 0.54
 103 Å. The model of TeAM contains one monomer in the asymmetric unit including residues 1 to 438.

104 For CfAM, there are two monomers in the asymmetric unit comprising residues 19 to 459 for chain
 105 A and 19 to 460 for chain B, which superpose with an r.m.s.d. of 0.3 Å. All three amylases have the
 106 classical domain structure with a central (β/α)₈-barrel with the active site located on its C-terminal
 107 face, together with a small subdomain, inserted between the third strand and helix and a C-terminal
 108 β -sandwich (Figure 2a). All three superpose with each other (Figure 2b) and with TAKA-amylase
 109 with an r.m.s.d. between 0.6 to 0.9 Å for up to 423 residues. Two conserved disulphide bridges
 110 stabilize flexible loops in subdomains A and B. There is an additional disulphide bridge in CfAM,
 111 located in the C-terminal domain. All three α -amylases have the conserved canonical calcium binding
 112 site located between the (β/α)₈-barrel and the insertion domain B.



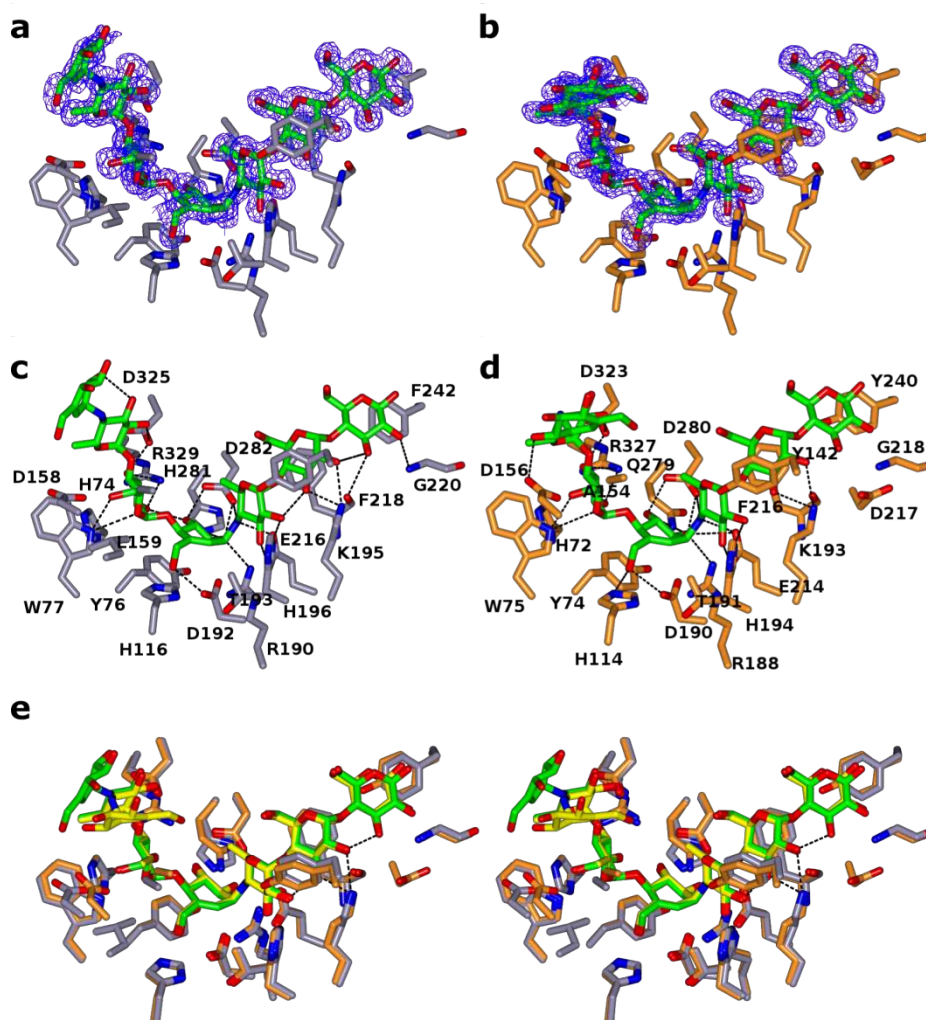
113

114 **Figure 2.** Structural overviews. (a) Ribbon representation of the structure of CfAM amylase in ribbon
 115 representation. The domains are coloured separately with the central barrel in purple. subdomain B
 116 in yellow and the C-terminal β -sandwich in green. The bound ligands acarbose transglycosylation
 117 product (ATgp) and maltose are shown as spheres. (b) Structural superposition of CfAM (purple)
 118 TeAM (orange) and RpAM (green).

119 2.3. Ligand binding site:

120 Although all three amylases were co-crystallized with acarbose, a well-known inhibitor for
 121 amylases, a complex with acarbose bound was only obtained for TeAM and CfAM. The reason why
 122 acarbose was not bound to RpAM is not clear. As expected the acarbose was found in the substrate
 123 binding cleft in each monomer of TeAM and CfAM, with the acarviosine unit sitting in subsites -1
 124 and +1, (Figure 3a-d). In both enzymes the binding mode is conserved and the ligands superpose
 125 with each other (Figure 3e), except for the monomer in subsite -4. The distorted pseudosugar
 126 valieneamine in subsite -1 with its ²H₃ half chair conformation mimics the conformation of the
 127 putative transition state along the catalytic itinerary of α -amylases. Additional density in subsites -2
 128 and -3 and -4 was modelled as a second acarbose unit, covalently attached to the first acarbose. The
 129 catalytic nucleophile D190/D192(CfAM/TeAM) is in a near attack conformation poised to react with
 130 the anomeric carbon, whilst the catalytic acid/base E214/E216(CfAM/TeAM) forms a hydrogen bond
 131 with the bridging nitrogen of the glycosidic bond with the 4-deoxyglucose in subsite +1. In addition,
 132 a hydrogen bond with H194/H196 stabilizes the 4-deoxyglucose in that subsite. The +3 subsite is
 133 formed by the sugar tong, composed of Y142/144 of subdomain B and F216/218 of the central domain,
 134 sandwiching the glucose between them. The reducing end of acarbose is stabilized by a hydrophobic
 135 platform interaction with Y240/F242 and a hydrogen bond with the main chain nitrogen of
 136 G218/G220. Interestingly, additional density at the non-reducing end was observed and was

137 modelled as an additional acarbose unit in subsites -2 and -3 and -4. The glucose in subsite -2 is
 138 stabilized by multiple hydrogen bonds with D323/325, R327/329 and W375/377. The glucose in subsite
 139 -3 is held in place by only one hydrogen bond with D323/325. The last visible part of the acarbose
 140 molecule is the acarviosine unit in subsite -4, which is not stabilized by direct interactions with the
 141 protein. Furthermore, the acarviosine unit is in two different positions in the two structures, reflecting
 142 the lack of strong stabilizing interactions between the ligand and the protein beyond subsite -3 (Figure
 143 3e).



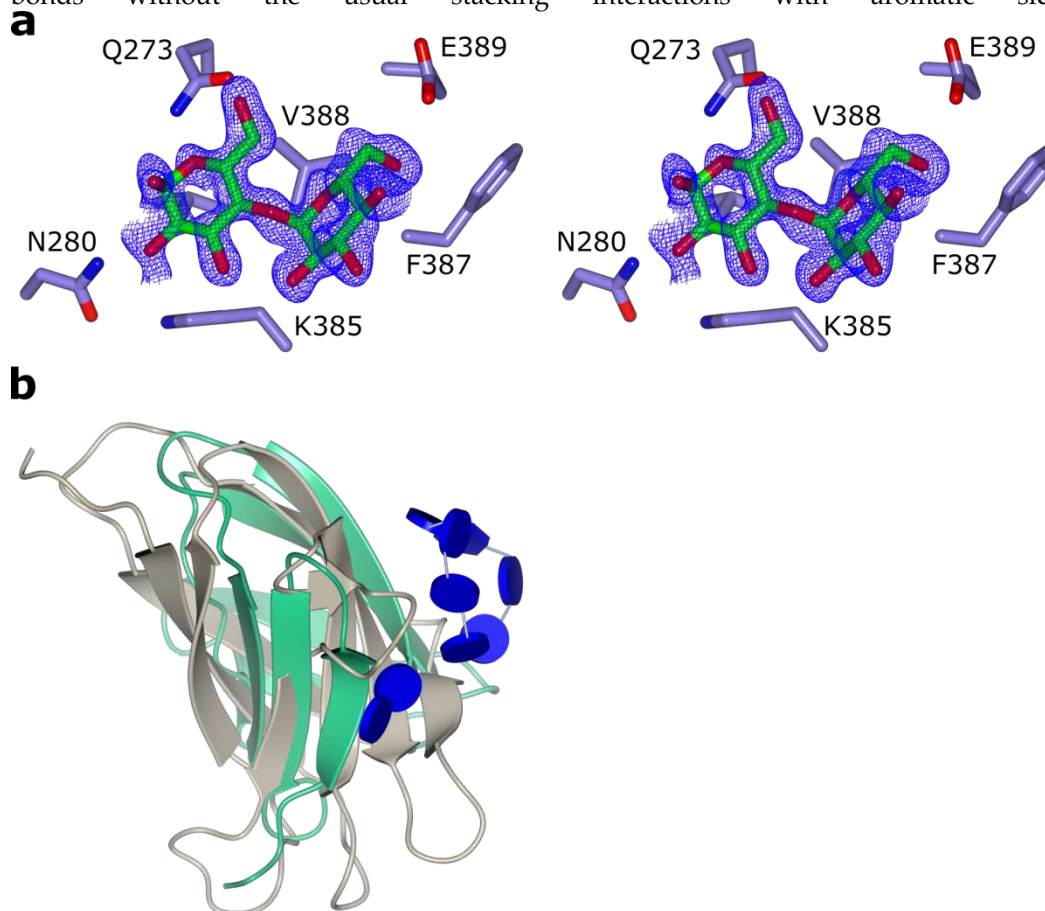
144

145 **Figure 3.** Acarbose transglycosylation product binding in CfAM and TeAM. (a) and (b) Stick
 146 representation of the acarbose derived transglycosylation product in the substrate binding crevice of
 147 CfAM and TeAM respectively. The 2Fo-Fc electron density around the ligands is contoured at
 148 $0.3 \text{ e}/\text{\AA}^3$. The interacting residues are shown as cylinders. (c) and (d) Hydrogen bonding pattern
 149 between ATgp and CfAM and TeAM in the active site. (e) Stereo view of the overlay of the binding
 150 crevice of CfAM (purple) and TeAM (orange). The residues and the ligands overlap very closely with
 151 the only major difference being the orientation of the acarviosine subunit in subsite -4.

152 2.4. Secondary glucose binding site

153 In CfAM, a secondary binding site in domain C was identified and modelled as maltose located
 154 at the edge of the β -sandwich (Figure 5a). The glucose units are held in place mainly via hydrogen

155 bonds without the usual stacking interactions with aromatic side chains.



156

157 **Figure 4.** The secondary maltose binding site in the C-terminal domain of CfAM. (a) Stereo view
 158 showing the maltose in cylinder representation with the corresponding 2Fo-Fc electron density
 159 contoured at $0.4 \text{ e}/\text{\AA}^3$. The interacting residues are shown as blue cylinders. (b) Superposition of the
 160 C-terminal domain (green) with the CBM20 domain from *A. niger* glucoamylase (pdb-ID: 1ac0) in
 161 beige. The bound β -cyclodextrin of CBM20 and the maltose unit are shown as glycoblocks [15].

162 2.6. N-glycosylation

163 There are three N-glycosylation sites, one at N144 in RpAM and two at N180 and 412 in TeAM.
 164 We observed only the core GlcNAc residue in all three enzymes. In the case of TeAM, this is due to
 165 the deglycosylation procedure with EndoH.

166 2.7. Isoaspartate formation

167 We observed the formation of an isoaspartate by succinimide formation and deamidation of
 168 N120 in chain B of RpAM. The same asparagine in chain A shows high flexibility and the resulting
 169 density suggest partial isoaspartate formation but a model could not be built with confidence.

170 3. Discussion

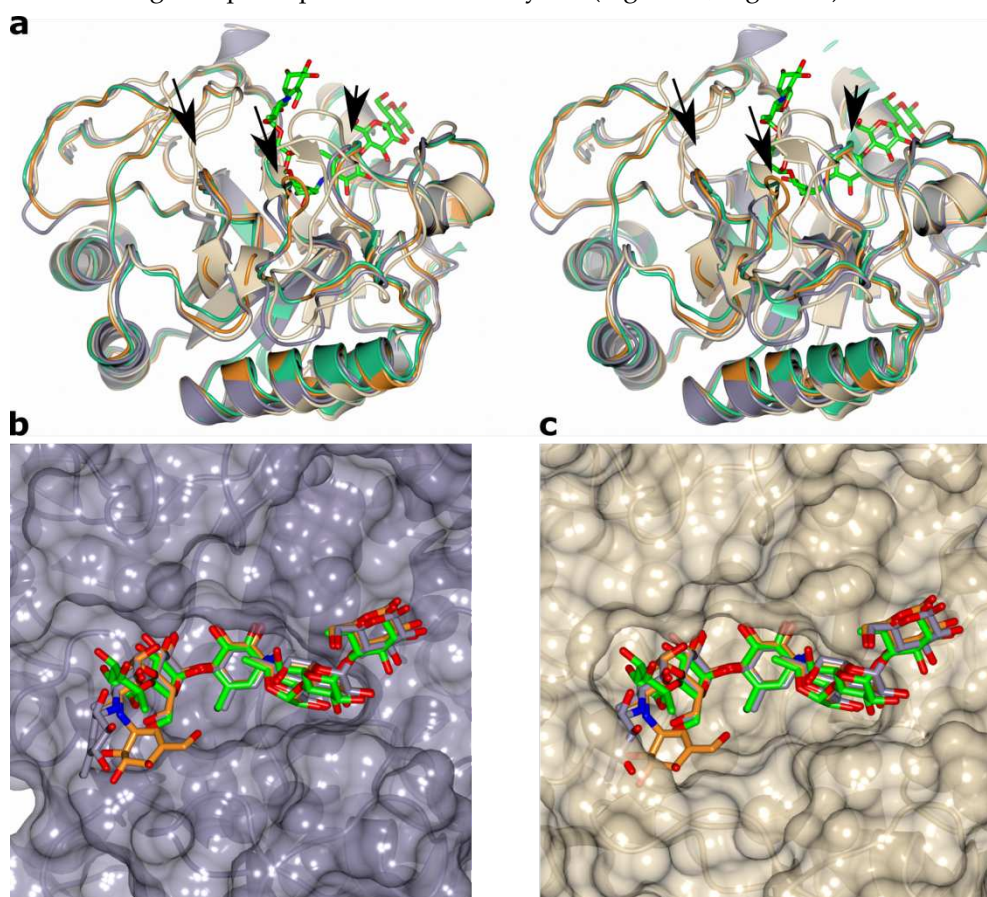
171 We have analysed structurally and functionally three novel fungal α -amylases with potential to
 172 be used in food industry and other industrial processes. All three structures determined, show the
 173 canonical amylase fold and overlap with each other with an r.m.s.d. of 0.54 \AA (Figure 1b). Further
 174 analysis of the sequence showed that both RpAM and CfAM have a slightly lower number of charged
 175 residues and a higher number of hydrophobic residues compared to TeAm and TAKA amylase,
 176 which might contribute to the higher thermostability of these two variants. Increased internal
 177 hydrophobicity while keeping external hydrophilicity, was found to correlate well with the
 178 thermostability of *Bacillus* α -amylases [16]. Furthermore, the shortened loops in these enzymes may

179 also contribute to the overall rigidity of the enzymes and therefore the thermostability as observed
180 for other enzymes as well [17, 18].

181 The substrate crevice in all three amylases, if defined on the basis of protein carbohydrate
182 interactions, spans from subsite -3 to +3. Having only three defined subsites for the non-reducing end
183 is common for amylases and is in line with the number of donor subsites described for TAKA-
184 amylase. Potentially, there could be more subsites for additional carbohydrate units at the reducing
185 end, which might connect the active site crevice with the observed second binding site (see below).

186 The observed complexes are most likely the result of limited transglycosylation, an unusual side
187 reaction previously reported *in crystallo* for several amylases, for example TAKA-amylase [19].
188 Though this reaction is common in the closely related CGTases (GH13_2) and amylomaltases (GH77),
189 it was not observed in solution for α -amylases. However, in crystals, transglycosylation products
190 with 10 or more units have been reported as a result of multiple transglycosylation events.
191 Interestingly, the final complex always has the pseudosaccharide unit, thought to mimic the
192 transition state, in the -1 subsite, rendering the enzyme inactive. Other binding modes are clearly
193 possible as evidenced by the final product and a pre-Michaelis complex observed for GH77 *Thermus*
194 *aquaticus* amylomaltase with acarbose [20].

195 All three amylases have as hallmark a shortened loop between $\beta 2/\alpha 3$ and two shorter loops in
196 subdomain B located between $\beta 3$ and $\alpha 4$ of the central (β/α)_s-barrel, compared to structures of other
197 fungal amylases, e.g. TAKA-amylase (Figure 5a). The importance of subdomain B for the
198 physicochemical properties, for example pH-stability, as well as substrate and product specificity, is
199 well known [21-24]. Indeed, the shorter loops open up the substrate crevice on the non-reducing end
200 (Figure 5b), which might explain the shift in the product profile for all three amylases towards
201 oligomers with a higher dp compared to TAKA amylase (Figure 1c, Figure 5c).



202

203 **Figure 5.** (a) Stereo view of all three amylases compared to TAKA-amylase with the three shortened
204 loops in the front marked with arrows. The ligand in CfAM is shown as sticks to identify the active
205 site. (b) Surface representation of CfAM with the bound ligand. The substrate is more open on the
206 donor subsite. (c) Surface representation of TAKA-amylase. The elongated loops create a more

207 restricted active site crevice precluding the binding mode observed in CfAM and TeAM due to steric
208 clashes. .

209 The C-terminal domain in α -amylases is implicated in starch binding and shows structural
210 similarity to classic CBM domains, based on an analysis using PDBeFOLD [25]. The additional
211 binding site in this domain in CfAM strengthen the role of this domain in substrate binding.
212 Additional carbohydrate binding sites have been observed as well for example in barley α -amylase
213 1 [26]. While none of these sites overlap with the binding site seen in CfAM, a structure of a CBM20
214 in complex with β -cyclodextrin revealed two binding sites, with the site termed SB1 in close
215 proximity to the binding site in CfAM (Figure 4b) [27]. This was confirmed to be the primary binding
216 site for the interaction with raw starch and it is likely that the observed binding site in CfAM is a
217 genuine carbohydrate binding site. Furthermore, it is intriguing to speculate about a potential path
218 from the primary substrate crevice to the secondary glucose binding site, which could be rather easily
219 thought as a simple extension of the acarbose from the reducing end.

220 Only limited information about the influence of glycosylation on amylase activity is available. It
221 was shown that for α -amylase Amy1 from the yeast *Cryptococcus flavus* N-glycosylation enhances
222 thermostability and resistance to proteolytic degradation [28]. The same effect is observed for
223 *Trichoderma reesei* Cel7a [29]. Indeed N144 is located in an extended loop and N-glycosylation might
224 help to shield the loop against proteolytic attack. The other two glycosylation sites are located in or
225 at the beginning of secondary structure elements, with N412 being located in the C- domain.

226 The observed isoaspartate formation is thought usually to be an age related side effect of protein
227 decomposition but a functional role cannot be ruled out [30]. Indeed, it was shown in GH77 enzymes
228 that such unusual posttranslational rearrangement might play a functional role in glycoside
229 hydrolases [31, 32]. The observed isoaspartate is located in one of the shortened loops in subdomain
230 B close to the substrate binding cleft, suggesting a functional role in CfAM as well.

231 4. Materials and Methods

232 4.1. Macromolecule production

233 The coding sequence of CfAM for the A, B and C domains was amplified from *Cordyceps farinosa*
234 gDNA by the polymerase chain reaction (PCR). The PCR fragment was obtained using primer pairs:
235 5'-ACACAAGTGGGGATCCACCATGAAGCTTACTGCGTCCCTC-3' and 5'-
236 GATGGTGATGGGATCCTTACTGCGCAACAAAACAATGGG-3'. The fragment was then ligated
237 in the expression vector pSUN515 using *Bam*HI and *Xho*I restriction sites. The ligation protocol was
238 performed according to the IN-FUSION™ Cloning Kit instructions. A transformation of TOP10
239 competent *E. coli* cells (Tiangen, China) with the plasmid, containing the CfAM gene, was performed
240 and positive clones confirmed by sequencing. The transformation of *Aspergillus oryzae* (strain
241 MT3568) with the expression vector comprising CfAM gene was performed according to patent
242 application WO95/002043 [33]. After incubation for 4-7 days at 37°C spores of four transformants
243 were inoculated into 3 ml of YPM medium. After 3-day cultivation at 30°C, the culture broths were
244 analysed by SDS-PAGE to identify the transformant producing the largest amount of recombinant
245 mature amylase with an estimated size of 48 kDa. Spores from the best expressing transformant were
246 cultivated in YPM medium in shake flasks for 4 days at a temperature of 30°C. The culture broth was
247 harvested by filtration using a 0.2 μ m filter device, and the filtered fermentation broth was used for
248 purification and further assays.

249 RpAM was cloned and expressed in a similar manner as CfAM while TeAM was expressed in
250 *Pichia pastoris* with a similar protocol to that described for the lipase from *Gibberella zeae* [34]. The entire
251 coding sequence of TeAM was amplified from cDNA by the polymerase chain reaction and
252 transformation into ElectroMax DH10B competent cells (Invitrogen) by electroporation. Transformed
253 cells were plated on LB plates containing 100 mM ampicillin. After overnight incubation at 27°C, a
254 positive clone was selected by colony PCR and confirmed by sequencing. The plasmid DNA of the
255 positive clone was linearized with PmeI (NEB) and transformed into *Pichia pastoris* KM71 (Invitrogen)

256 following the manufacturer's instructions. An amylase positive clone was inoculated into 3 ml BMSY
257 and incubated at 28°C for 3 days until the OD600 reached 20. Methanol was added to the culture
258 daily to a final concentration of 0.5% for the following 4 days. On day 4 of induction, the culture
259 supernatant was separated from the cells by centrifugation and the pH of the supernatant was
260 adjusted to 7.0.

261 The CfAM culture broth was precipitated with ammonium sulphate (80% saturation), then
262 dialyzed with 20 mM Na-acetate at pH 5.0. The solution was loaded on to a Q Sepharose Fast Flow
263 column (GE Healthcare) equilibrated with 20 mM Na Acetate at pH 5.0. Protein was eluted with a
264 salt gradient from zero to 1 M NaCl. Fractions were analysed for amylase activity and pooled
265 accordingly. The flow-through fraction, containing the bulk of amylase activity was supplemented
266 with ammonium sulphate to a final concentration of 1.2 M and then loaded on to Phenyl Sepharose
267 6 Fast Flow column (GE Healthcare). The activity was eluted by a linear gradient of decreasing salt
268 concentration. The fractions with activity were analysed by SDS-PAGE and then concentrated for
269 further use.

270 Amylase activity was detected by AZCL-HE-amylase (Megazyme International Ireland Ltd.) as
271 substrate. 10 µl enzyme sample and 120 µl 0.1% substrate at pH 7 were mixed in a Microtiter plate
272 and incubated at 50°C for 30 min. Then 70 µl supernatant was transferred to a new microtiter plate
273 and the absorption at 595 nm determined. All reactions were done as duplicates.

274 4.2. Biochemical characterisation

275 pH-Optimum:

276 To determine the pH-Optimum each enzyme (3 µl of a 0.5 mg/ml solution) was incubated with
277 40 µl 1% substrate (AZCL-HE-amylase (Megazyme International Ireland Ltd.)). The pH between 2
278 and 11 was adjusted using 100 µl of B&R buffer (Britton-Robinson buffer: 0.1 M boric acid, 0.1 M
279 acetic acid, and 0.1 M phosphoric acid, adjusted to pH-values 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0 and
280 11.0 with HCl or NaOH) [35]. The reactions were incubated at 30°C for 30 minutes and afterwards
281 60 µl were transferred in a new microtiter plate and the absorption was measured at 595 nm.

282 Temperature-Optimum:

283 To determine the Temperature-Optimum each enzyme was incubated with 100 µl 0.1 %
284 substrate (AZCL-HE-amylase (Megazyme International Ireland Ltd.)) in 50 mM Na Acetate pH 4.3.
285 The substrate solution was preincubated at 20-90 °C for 5 minutes and the reaction was started by
286 addition of 3 µl of enzyme solution (0.5 mg/ml). The reaction mixture was further incubated at the
287 respective temperature for 30 minutes at 950 rpm. The reaction was stopped by rapid cooling on ice.
288 Afterwards 60 µl of each reaction was transferred in a microtiterplate and the absorption was
289 measured at 595 nm. Each reaction was performed in triplicate.

290 Product profile:

291 For product profile determination each enzyme (15 µl) was incubated with 120 µl 0.1% substrate
292 (AZCL-HE-amylase (Megazyme International Ireland Ltd.)) at pH 5 and 62 °C for 14 hours. 70 µl of
293 each reaction was mixed with equal amounts of Acetonitril. The mixture was centrifuged for 30 min
294 at 16.000xg and the supernatant was analyzed using HPAEC with pulsed amperometric detection.

295 4.3. Crystallisation

296 RpAM:

297 The concentrated protein was mixed with acarbose in a molar ratio of 4:1 before the initial
298 screening in 96 well format using commercially available screens. An initial hit (0.2 M NaCl, 0.1 M
299 Na-acetate pH 4.6, 30 %MPD), was further refined in 24 well format using the initial crystals as seeds.

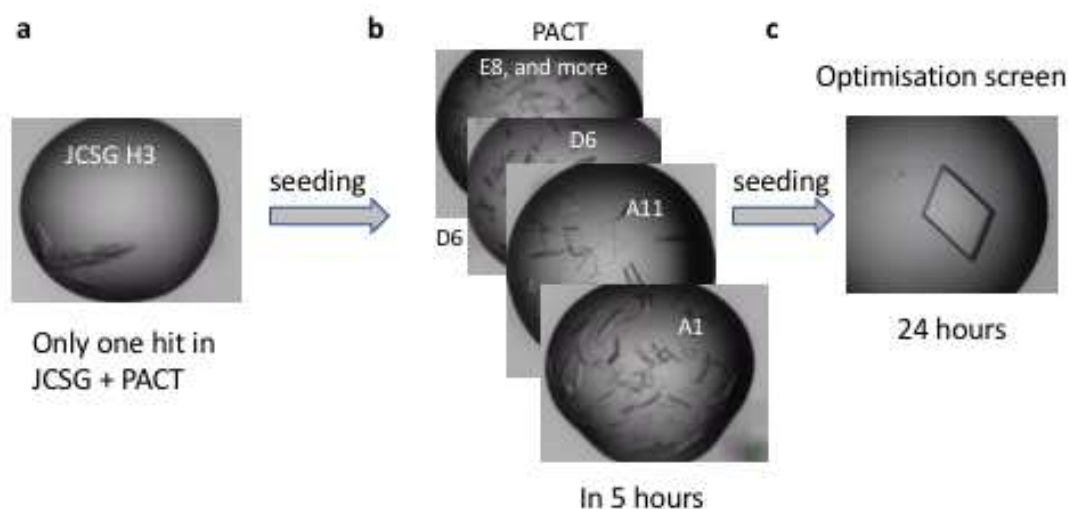
300 Crystals suitable for data collection were cryoprotected using 25% glycerol and flash frozen in liquid
301 nitrogen prior data collection.

302 TeAM:

303 Prior to providing the sample to York, the protein was deglycosylated using Endo-H treatment.
304 The protein was concentrated using Amicon filter units and stored at -80°C for later use. For the
305 crystallisation the protein was mixed with 5 mM acarbose prior to setting up the screen. Initial screens
306 were set up in 96 well sitting drop format using commercially available screens. Initial hits were
307 further refined in 24 well hanging drop format. The best crystals grew in 0.1 M di-hydrogen
308 phosphate, 1.8 M ammonium sulphate. Crystals were cryoprotected by addition of ethylene glycol to
309 a final concentration of 15 %. The crystals were flash frozen in liquid nitrogen prior to data collection.

310 CfAM:

311 Prior to crystallisation, the protein was concentrated to 22.5 mg/ml by ultrafiltration in an
312 Amicon centrifugation filter unit (Millipore), aliquoted to 50 μl ; aliquots that were not immediately
313 set up for crystallisation were flash frozen in liquid nitrogen and stored at -80°C to use later in
314 optimizations. Initial crystallisation experiments were carried out in the presence or absence of 4 mM
315 CaCl_2 and 40 mM acarbose. An initial hit was obtained for an acarbose complex, in just one condition
316 (H3, Bis-tris 5.5, 25% w/v PEG3350) of JCSG screen (Figure 6a), out of total 192 conditions in two
317 initial screens set up – JCSG (ref) and PACT premier™ HT-96 (Molecular Dimensions). The crystals
318 were imperfect and were used to make the seeding stock. The seeding stock was prepared and
319 microseed matrix screening (MMS, recent review in [14]) carried out using an Oryx robot (Douglas
320 instruments) according to the published protocols [36, 37]. Briefly, crystals were crushed, and diluted
321 with $\sim 50\mu\text{l}$ of mother liquor. The solution was transferred into a seed bead containing reaction tube
322 and vortexed for three minutes. The seeding stock was used straightaway, and the remaining seeds
323 were frozen and kept at -20°C . MMS was carried out in the PACT screen, giving a significant number
324 of hits (Figure 6b). Crystals from condition A11 were used to make a seeding stock for the next
325 seeding round. This time it was not a “classical” MMS – seeding into a random screen, but rather
326 seeding into an optimisation screen based on the initial conditions, but with different pH, salts and
327 PEGs/PEG concentrations. The crystallisation drops contained 150 nl protein + 50 nl seeding stock +
328 100 nl mother liquor from a new random screen. The final, good quality crystal was obtained in 12%
329 PEG 3350 0.2 M NaNO_3 , CAPS pH 11.0 (Figure 6c).



331 **Figure 6.** Crystal optimization using microseed matrix screening.

332 4.4. Data collection and processing

333 The data were collected at Diamond on beam line I02, processed by XDS [38], and scaled with
334 Aimless [39]. The statistics are shown in Table 1.

335 **Table 1.** Data collection and processing statistics.

	CfAM	TeAM	RpAM
Diffraction source	Diamond I02	Diamond I02	ESRF ID29
Wavelength (Å)	0.9795	0.9795	1.0004
Temperature (K)	100	100	100
Space group	P1	P2 ₁ 2 ₁ 2 ₁	P1
<i>a</i> , <i>b</i> , <i>c</i> (Å)	<i>a</i> =56.88, <i>b</i> = 61.97, <i>c</i> = 70.40	<i>a</i> =51.02, <i>b</i> = 56.63, <i>c</i> = 166.01	<i>a</i> =51.22, <i>b</i> = 62.60, <i>c</i> = 66.81
α , β , γ (°)	α =79.33, β =82.88, γ =67.99	α =90, β =90, γ =90	α =77.03, β =81.04, γ =89.62
Resolution range (Å)	33.1 – 1.35 (1.37- 1.35)	48.76 – 1.20 (1.22- 1.20)	43.21- 1.4 (1.42-1.40)
Total No. of reflections	342708	1149540	315876
No. of unique reflections	163777	150529	146177
Completeness (%)	85.3 (38.2)	99.8(96.7)	92.9(61.9)
Redundancy	2.1 (2.1)	7.6(4.6)	2.2(2.1)
$\langle I/\sigma(I) \rangle$	13.7(10.3)	14.1(1.7)	9.6(2.3)
<i>R</i> _{r.i.m.}	0.076 (0.129)	0.021(0.446)	0.030(0.225)
<i>CC</i> _{1/2}	0.983(0.970)	0.999(0.615)	0.998(0.892)
Overall <i>B</i> factor from Wilson plot (Å ²)	6.8	8.7	8.1

336 Values for the outer shell are given in parentheses.

337 4.5. Structure solution and refinement

338 The structure of RpAM was solved by molecular replacement with Molrep [40], using TAKA
339 amylase as template (pdb-ID:7taa). The structure of TeAM was solved with Molrep using the model
340 of RpAM. The CfAM structure was solved using Molrep [40] with 3vx0 α -amylase from *Aspergillus*
341 *oryzae* as a model. The final models were built using automated chain tracing with Buccaneer [41],
342 followed by manual building in Coot [42], iterated with reciprocal space refinement using Refmac
343 [43]. The statistics are summarized in Table 2.

344 **Table 2.** Structure solution and refinement.

345 Values for the outer shell are given in parentheses.

	CfAM	TeAM	RpAM
PDB-ID	6SAV	6SAO	6SAU
Resolution range (Å)	33.1-1.35 (1.385-1.35)	48.76-1.20 (1.22-1.20)	39.99-1.4 (1.42-1.40)
Completeness (%)	85.3 (39.7)	99.8(96.7)	92.8(89.2)
No. of reflections, working set	155488	143033	138848
No. of reflections, test set	8289	7574	7328
Final <i>R</i> _{cryst}	0.113 (0.09)	0.110(0.27)	0.136(0.22)
Final <i>R</i> _{free}	0.150 (0.17)	0.134(0.29)	0.164(0.26)
Cruickshank DPI	0.051	0.027	0.056
No. of subunits in the asymmetric unit	2	1	2

No. of non-H atoms	Chain A/B	Chain A	Chain A/B
Protein	3557/3609	3570	3662/3592
Ion	1/2	1	1/1
Ligand	99/120	133	14/36
Water	875	568	943
Total	8263	4272	8306
R.m.s. deviations			
Bonds (Å)	0.0191	0.0163	0.0147
Angles (°)	2.06	1.937	1.875
Average B factors (Å ²)	Chain A/B	Chain A	Chain A/B
Protein	10/8.7	12.9	12.9/12.3
Ions			
Ca ²⁺	6.7/5.8	9.5	7.39/7.5
Na ²⁺	N/A/10.9		
Ligand	19.6/18.0	22.2	19.2/21.4
Water	19.0	28.8	24.21
Ramachandran plot			
Most favoured (%)	98.6	97.7	97.2
Allowed (%)	1.4	2.3	2.7

346 5. Conclusions

347 Taken together, we describe the structural and functional characterization of three novel fungal
 348 α -amylases with enhanced stability, of which two, CfAM and RpAM, have a higher pH optimum
 349 and greater temperature tolerance, well suited for usage in the detergent or saccharification industry.
 350 The structures reveal that these amylases follow the canonical domain structure of α -amylases, and
 351 that three shortened loops between β_2/α_3 and in subdomain B are likely to be responsible for the
 352 altered enzymatic properties of the amylases compared to TAKA-amylase. For the first time we have
 353 unambiguously identified up to three different N-glycosylation sites in α -amylases in the structures.
 354 Furthermore, the observed formation of an isoaspartate from an asparagine in one of the shortened
 355 loops might play a functional role. The complexes with an acarbose derived transglycosylation
 356 products define seven subsites of the substrate binding crevice and helped to identify the catalytic
 357 residues unambiguously. In addition, a new previously unobserved carbohydrate binding site was
 358 revealed in the C-terminal β -sandwich domain of CfAM, which might be important for the initial
 359 interaction with its polymeric substrate.

360 6. Patents

361 The *Rhizomucor pusillus* amylase and the use of this amylase in various industrial
 362 applications have been claimed in patent application WO2006065579. A close homologue of
 363 the *Thamnidium elegans* amylase was claimed in patent application WO2006069290 including the use
 364 in industrial applications.

365 **Supplementary Materials:** Supplementary materials can be found at www.mdpi.com/xxx/s1.

366 **Author Contributions:** C.R., O.V.M analyzed the data, built and refined the structure and prepared the original
 367 draft. J.P.T. collected and analyzed the X-ray data. O.V.M., E.B., A.A., J.W. crystallized the amylases and solved
 368 the initial structures. L.M. and S.T. cloned, produced, purified and characterized the amylases biochemically,
 369 C.R., C.A., G.J.D. and K.S.W. wrote analyzed and reviewed all stages of the manuscript. G.J.D. C.A. and K.S.W.
 370 planned and supervised the work.

371 **Funding:** This research received no external funding. We note that ST and LM are employees of Novozymes
 372 (China) and CA of Novozymes (Denmark).

373 **Acknowledgments:** The authors are grateful for financial support by Novozymes. We thank ESRF for the access
 374 to beamline ID29 and Diamond Light Source for access to beamline I02 (proposal numbers mx-1221 and mx-

375 9948) that contributed to the results presented here. The authors also thank Sam Hart for assistance during data
376 collection.

377 **Conflicts of Interest:** The authors declare no conflict of interest, but we note that the Novozymes authors declare
378 the following competing financial interest(s): Novozymes are a commercial enzyme supplier.

379 Abbreviations

CfAM.	<i>Cordyceps farinosa</i> amylase
RpAM	<i>Rhizomucor pusillus</i> amylase
TeAM	<i>Thamnidium elegans</i> amylase
TAKA	<i>Aspergillus oryzae</i> amylase
dp	Degree of polymerisation

380 References

- 381 1. Payen, A.P.J.F., *Memoire sur la diastase, les principaux produits de ses réactions et leurs applications aux arts*
382 *industriels*" (*Memoir on diastase, the principal products of its reactions, and their applications to the industrial*
383 *arts*). *Annales de Chimie et de Physique*, 1833. **2**: p. 73-92.
- 384 2. Gurung, N., et al., *A broader view: microbial enzymes and their relevance in industries, medicine, and beyond*.
385 *Biomed Res Int*, 2013. **2013**: p. 329121.
- 386 3. Roy, J.K., et al., *Cloning and extracellular expression of a raw starch digesting alpha-amylase (Blamy-I) and its*
387 *application in bioethanol production from a non-conventional source of starch*. *J Basic Microbiol*, 2015. **55**(11):
388 p. 1287-98.
- 389 4. Gupta, R., et al., *Microbial α -amylases: a biotechnological perspective*. *Process Biochemistry*, 2003. **38**(11): p.
390 1599-1616.
- 391 5. Niehaus, F., et al., *Extremophiles as a source of novel enzymes for industrial application*. *Appl Microbiol*
392 *Biotechnol*, 1999. **51**(6): p. 711-29.
- 393 6. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013*. *Nucleic Acids Res*, 2014.
394 **42**(Database issue): p. D490-5.
- 395 7. Janecek, S., B. Svensson, and E.A. MacGregor, *Structural and evolutionary aspects of two families of non-*
396 *catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals*. *Enzyme*
397 *Microb Technol*, 2011. **49**(5): p. 429-40.
- 398 8. Liu, Y., et al., *Crystal structure of a raw-starch-degrading bacterial alpha-amylase belonging to subfamily 37 of*
399 *the glycoside hydrolase family GH13*. *Sci Rep*, 2017. **7**: p. 44067.
- 400 9. Mehta, D. and T. Satyanarayana, *Domain C of thermostable alpha-amylase of Geobacillus thermoleovorans*
401 *mediates raw starch adsorption*. *Appl Microbiol Biotechnol*, 2014. **98**(10): p. 4503-19.
- 402 10. Sogaard, M., et al., *Site-directed mutagenesis of histidine 93, aspartic acid 180, glutamic acid 205, histidine 290,*
403 *and aspartic acid 291 at the active site and tryptophan 279 at the raw starch binding site in barley alpha-amylase*
404 *1*. *J Biol Chem*, 1993. **268**(30): p. 22480-4.
- 405 11. Kadziola, A., et al., *Molecular structure of a barley alpha-amylase-inhibitor complex: implications for starch*
406 *binding and catalysis*. *J Mol Biol*, 1998. **278**(1): p. 205-17.
- 407 12. Brzozowski, A.M., et al., *Structural analysis of a chimeric bacterial alpha-amylase. High-resolution analysis of*
408 *native and ligand complexes*. *Biochemistry*, 2000. **39**(31): p. 9099-107.
- 409 13. Pritchard, P.E., *Studies on the bread-improving mechanism of fungal alpha-amylase*. *Journal of Biological*
410 *Education*, 1992. **26**(1): p. 12-18.
- 411 14. D'Arcy, A., et al., *Microseed matrix screening for optimization in protein crystallization: what have we learned?*
412 *Acta Crystallogr F Struct Biol Commun*, 2014. **70**(Pt 9): p. 1117-26.

- 413 15. McNicholas, S. and J. Agirre, *Glycoblocks: a schematic three-dimensional representation for glycans and their*
414 *interactions*. Acta Crystallogr D Struct Biol, 2017. **73**(Pt 2): p. 187-194.
- 415 16. Janeček, Š., *Does the increased hydrophobicity of the interior and hydrophilicity of the exterior of an enzyme*
416 *structure reflect its increased thermostability?* International Journal of Biological Macromolecules, 1993.
417 **15**(5): p. 317-318.
- 418 17. Mok, S.C., et al., *Crystal structure of a compact alpha-amylase from Geobacillus thermoleovorans*. Enzyme
419 Microb Technol, 2013. **53**(1): p. 46-54.
- 420 18. Mazola, Y., et al., *A comparative molecular dynamics study of thermophilic and mesophilic beta-fructosidase*
421 *enzymes*. J Mol Model, 2015. **21**(9): p. 228.
- 422 19. Brzozowski, A.M. and G.J. Davies, *Structure of the Aspergillus oryzae alpha-amylase complexed with the*
423 *inhibitor acarbose at 2.0 Å resolution*. Biochemistry, 1997. **36**(36): p. 10837-45.
- 424 20. Przylas, I., et al., *X-ray structure of acarbose bound to amyloamylase from Thermus aquaticus. Implications for*
425 *the synthesis of large cyclic glucans*. Eur J Biochem, 2000. **267**(23): p. 6903-13.
- 426 21. Juge, N., et al., *Isozyme hybrids within the protruding third loop domain of the barley alpha-amylase*
427 *(beta/alpha)8-barrel. Implication for BASI sensitivity and substrate affinity*. FEBS Lett, 1995. **363**(3): p. 299-
428 303.
- 429 22. Rodenburg, K.W., et al., *Domain B protruding at the third beta strand of the alpha/beta barrel in barley alpha-*
430 *amylase confers distinct isozyme-specific properties*. Eur J Biochem, 1994. **221**(1): p. 277-84.
- 431 23. Penninga, D., et al., *Site-directed mutations in tyrosine 195 of cyclodextrin glycosyltransferase from Bacillus*
432 *circulans strain 251 affect activity and product specificity*. Biochemistry, 1995. **34**(10): p. 3368-76.
- 433 24. Nakamura, A., K. Haga, and K. Yamane, *Four aromatic residues in the active center of cyclodextrin*
434 *glucanotransferase from alkalophilic Bacillus sp. 1011: effects of replacements on substrate binding and*
435 *cyclization characteristics*. Biochemistry, 1994. **33**(33): p. 9929-36.
- 436 25. Krissinel, E., *On the relationship between sequence and structure similarities in proteomics*. Bioinformatics,
437 2007. **23**(6): p. 717-23.
- 438 26. Robert, X., et al., *The structure of barley alpha-amylase isozyme 1 reveals a novel role of domain C in substrate*
439 *recognition and binding: a pair of sugar tongs*. Structure, 2003. **11**(8): p. 973-84.
- 440 27. Sorimachi, K., et al., *Solution structure of the granular starch binding domain of Aspergillus niger glucoamylase*
441 *bound to beta-cyclodextrin*. Structure, 1997. **5**(5): p. 647-61.
- 442 28. de Barros, M.C., et al., *The influence of N-glycosylation on biochemical properties of Amy1, an alpha-amylase*
443 *from the yeast Cryptococcus flavus*. Carbohydr Res, 2009. **344**(13): p. 1682-6.
- 444 29. Amore, A., et al., *Distinct roles of N- and O-glycans in cellulase activity and stability*. Proc Natl Acad Sci U
445 S A, 2017. **114**(52): p. 13667-13672.
- 446 30. Reissner, K.J. and D.W. Aswad, *Deamidation and isoaspartate formation in proteins: unwanted alterations or*
447 *surreptitious signals?* Cell Mol Life Sci, 2003. **60**(7): p. 1281-95.
- 448 31. Barends, T.R., et al., *Three-way stabilization of the covalent intermediate in amyloamylase, an alpha-amylase-*
449 *like transglycosylase*. J Biol Chem, 2007. **282**(23): p. 17242-9.
- 450 32. Roth, C., et al., *Amylose recognition and ring-size determination of amyloamylase*. Sci Adv, 2017. **3**(1): p.
451 e1601386.
- 452 33. DALBØGE, H., et al., *DNA encoding an enzyme with endoglucanase activity from Trichoderma harzianum*.
453 1995, Novo-Nordisk A/S.
- 454 34. Sun, Y., et al., *Crystallization and preliminary crystallographic analysis of Gibberella zeae extracellular lipase*.
455 Acta Crystallogr Sect F Struct Biol Cryst Commun, 2008. **64**(Pt 9): p. 813-5.

- 456 35. Britton, H.T.S. and R.A. Robinson, *CXCVIII.—Universal buffer solutions and the dissociation constant of*
457 *veronal*. Journal of the Chemical Society (Resumed), 1931(0): p. 1456-1462.
- 458 36. Shaw Stewart, P.D., et al., *Random Microseeding: A Theoretical and Practical Exploration of Seed Stability and*
459 *Seeding Techniques for Successful Protein Crystallization*. Crystal Growth & Design, 2011. **11**(8): p. 3432-
460 3441.
- 461 37. Shah, A.K., et al., *On increasing protein-crystallization throughput for X-ray diffraction studies*. Acta
462 Crystallographica Section D, 2005. **61**(2): p. 123-129.
- 463 38. Kabsch, W., *Xds*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
- 464 39. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D
465 Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.
- 466 40. Vagin, A. and A. Teplyakov, *Molecular replacement with MOLREP*. Acta Crystallogr D Biol Crystallogr,
467 2010. **66**(Pt 1): p. 22-5.
- 468 41. Cowtan, K., *The Buccaneer software for automated model building. 1. Tracing protein chains*. Acta Crystallogr
469 D Biol Crystallogr, 2006. **62**(Pt 9): p. 1002-11.
- 470 42. Emsley, P., et al., *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p.
471 486-501.
- 472 43. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-*
473 *likelihood method*. Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
474