Article:
Pickles, Isabelle B., Chen, Yurong, Moroz, Olga et al. (9 more authors) (2024) Precision activity-based α-amylase probes for dissection and annotation of linear and branched-chain starch degrading enzymes. Angewandte Chemie International Edition. e202415219. ISSN 1433-7851

https://doi.org/10.1002/anie.202415219

GDCh Research Article Angewandte International Edition Chemie www.angewandte.org

How to cite: Angew. Chem. Int. Ed. 2024, e202415219
doi.org/10.1002/anie.202415219

Activity-Based Probes

Precision Activity-Based α-Amylase Probes for Dissection and Annotation of Linear and Branched-Chain Starch-Degrading Enzymes

Isabelle B. Pickles[+], Yurong Chen[+], Olga Moroz, Haley A. Brown, Casper de Boer, Zachary Armstrong, Nicholas G. S. McGregor, Marta Artola, Jeroen D. C. Codée, Nicole M. Koropatkin, Herman S. Overkleeft, and Gideon J. Davies**

Abstract: α-Amylases are the workhorse enzymes of starch degradation. They are central to human health, including as targets for anti-diabetic compounds, but are also the key enzymes in the industrial processing of starch for biofuels, corn syrups, brewing and detergents. Dissection of the activity, specificity and stability of α-amylases is crucial to understanding their biology and allowing their exploitation. Yet, functional characterization lags behind DNA sequencing and genomics; and new tools are required for rapid analysis of α-amylase function. Here, we design, synthesize and apply new branched α-amylase activity-based probes. Using both α-1,6 branched and unbranched α-1,4 maltobiose activity-based probes we were able to explore the stability and substrate specificity of both a panel of human gut microbial α-amylases and a panel of industrially relevant α-amylases. We also demonstrate how we can detect and annotate the substrate specificity of α-amylases in the complex cell lysate of both a prominent gut microbe and a diverse compost sample by in-gel fluorescence and mass spectrometry. A toolbox of starch-active activity-based probes will enable rapid functional dissection of α-amylases. We envisage activity-based probes contributing to better selection and engineering of enzymes for industrial application as well as fundamental analysis of enzymes in human health.

Introduction

Retaining α-amylases, primarily found in CAZY family GH13 (www.cazy.org),[1] are starch degrading enzymes with myriad industrial and health implications. α-Amylases hydrolyse α-1,4 D-glucosidic linkages to yield maltooligosaccharides and maltose. The enzymes play central roles in dietary starch degradation and glucose assimilation, and are thus key targets for treatments for diabetes (recently reviewed in Kaur et al., 2021).[2] α-Amylases are also the key players in industrial starch processing,[3] such as in the breakdown of starch to ethanol and high-fructose corn syrups, and also in societal processes such as brewing, baking and in household detergents.[4] Starch itself consists of an α-1,4 D-glucosidic backbone with varying extents of α-1,6 branch points. These branches impact on the digestibility of starch,[5] both its rate (impacting on whether starches are resistant to hydrolysis and thus low glycaemic index (GI) foods)[6] and the structure of the "limit dextrins" (the final oligosaccharide products that can be digested no further by the given enzymes). If, how and where α-amylases can accommodate / harness branch points (for review see MacGregor et al., 2001)[7] is therefore crucial to their biological function and their deployment in industry and for societal applications.

Starch is a main foodstuff for humans and other animals, and in humans, salivary and pancreatic α-amylases breakdown dietary starch to shorter oligosaccharides, ultimately to glucose which enters the bloodstream. α-Amylases are therefore known therapeutic targets for commercial drugs; an area still receiving major interest.[2,8] Furthermore, lower down the intestinal tract, members of the gut microbiota ferment any starch that reaches the colon into short-chain fatty acids, including acetate, propionate, and butyrate, all of

[*] Dr. I. B. Pickles,[+] Dr. O. Moroz, Dr. N. G. S. McGregor, Prof. Dr. G. J. Davies
York Structural Biology Laboratory, Department of Chemistry, University of York,
York, North Yorkshire, YO10 5DD, UK
E-mail: gideon.davies@york.ac.uk

Dr. Y. Chen,[+] Dr. C. de Boer, Dr. Z. Armstrong, Dr. M. Artola, Prof. Dr. J. D. C. Codée, Prof. Dr. H. S. Overkleeft
Department of Bio-organic Synthesis, Leiden Institute of Chemistry, Leiden University,
Einsteinweg 55, 2333 CC Leiden, The Netherlands
E-mail: h.s.overkleeft@lic.leidenuniv.nl

Dr. H. A. Brown, Dr. N. M. Koropatkin
University of Michigan Medical School
6605D Med Sci II, 1150W Medical Center Drive, Ann Arbor, MI 48109-5620 USA

[+] These authors contributed equally

which are believed to help maintain gut health.[9] The study of polysaccharide utilization *loci* (PULs), discrete gene clusters that encode the requisite proteins for starch assimilation, in gut bacteria,[10] has therefore revealed how starch is degraded by key members of the human microbiota. In this community, α-amylases are the most predominant enzymes,[9a] which intensifies emerging interest in α-amylase study.

α-Amylases, according to commercial estimates, are a market currently worth ~USD 1900 M per annum, approaching 2700 M dollars by 2032.[11] Key market areas include baking, household detergent applications[4] and the synthesis of high glucose or fructose corn syrups;[5] the latter a process which demands conditions of high temperature, extreme pH, and (to aid the other enzymes required) low [Ca^{2+}]. The main industrial α-amylases are those from *Aspergillus oryzae* (TAKA) and various heat stable bacillus species such as *Bacillus licheniformis* (BLA).[12] Further progress, either through enzyme engineering or natural discovery of enzymes better adapted to their application, would benefit from rapid biochemical tools. All these emerging applied processes, as well as dissecting the roles of α-amylases in human disease and microbiotal health, demand better knowledge of α-amylases, notably their specificity and stability. But, such insights are rarely yielded from primary sequence analysis or 3D structure comparison, alone. Precision chemical tools are required to allow more rapid functional analyses. Biochemical assessments lag many years, arguably decades, behind the ease of genomic sequencing. One set of tools that has been developed in recent years, with the idea of speeding up functional dissection (and enzyme discovery), are activity-based probes (ABPs).[13] Activity-based protein profiling (ABPP) has emerged as a technique for the rapid, initially in-gel, assessments of enzyme action, specificity, sensitivity to inhibitors, and stability.[14] The electrophilic warheads of these probes target nucleophilic side chains in the enzymatic reaction mechanism (Figure 1). In the case of retaining glycoside hydrolases, the vast majority are known to use a double displacement reaction mechanism, in which a covalent intermediate is formed and subsequently hydrolysed via oxocarbenium-ion-like transition states. This "Koshland" mechanism[15] demands two key catalytic residues, an enzymatic nucleophile and an acid/base to activate the leaving-group through protonation and the incoming water in the second step by deprotonation.[16] It is the nucleophile (typically Asp or Glu, occasionally Tyr or Cys) that is targeted by the ABPP approach. Foundational work on cyclophellitol[17] ABPs for the lysosomal storage disease (LSD) Gaucher's has extended,[18] not only to a wider range of precision tools for Pompe,[16] Fabry[19] and other LSDs, but also as anti-cancer compounds in the context of heparanase inhibition and imaging.[14,20] More recently, cyclophellitol-derived ABPs have been designed and applied in the biotechnology sector notably, for example, for the analysis of xylanases,[21] arabinofuranosidases,[22] cellulases,[23] amylases[24] and xyloglucanases.[25] With these developments, it is clear that ABPs offer great potential for rapid and dynamic analysis of glycoside hydrolases, and that tuning them to focus on current biological and industrial challenges is a logical next step.

Here, to identify α-amylases which can cut close to α-1,6 branches, the synthesis of branched α-amylase probes **2a–d** and **3a–d** was undertaken. α-Amylases which can cleave close to these branches are of major importance, both for human digestion of recalcitrant starches, but especially in industrial applications where subsite specificity, in respect of branch accommodation, defines the structure of subsequent limit dextrins contributing to the recalcitrance of starch and dietary and "mouthfeel" properties.

Whereas AlphaFold2[26] now provides rapid insight into approximate 3D structure, information about substrate specificity in the context of branch points, stability, and expression profiles is impossible to determine from genetic and sequence information alone. We therefore apply these ABPs to study, first, pure α-amylases from prominent human gut microbes, and then whole *Bacteroides thetaiotaomicron* lysates grown on different carbon sources, with gene knockouts in their polysaccharide utilization *loci*, identifying labelled α-amylases using a pull-down proteomics experiment. We then extend this analysis, in tandem with the previously described maltobiosyl probes **1a–c**,[24] to study a panel of industrial enzymes, their pH and [Ca^{2+}] activities and their acceptance of α-1,6 branches in the −2 and −1 subsites. Finally, we apply this technique to the discovery of novel α-amylases for the industrial processing of starch from a commercial composter, using a pull-down proteomics experiment to detect enzymes from a diverse collection of thermophilic bacteria.

Results and Discussion

Synthesis of Branched α-Amylase Activity-Based Inhibitors and Probes

We have previously shown that pseudo-disaccharide epoxide compounds **1a–c** are potent and selective inhibitors of, and probes for, retaining α-amylases in complex biological mixtures.[24] We sought to extend this methodology by synthesizing branch point mimics of natural substrates, as tools to dissect substrate specificity of these enzymes. To this end, we synthesized epoxide inhibitors and ABPs incorporating an α-1,6 glucose −2 subsite or −1 subsite branch. Accommodation of an α-1,6 branch in the −2 subsite has not been extensively studied but there are structural examples of branch point accommodation (AliC[27] and SusG[28]). There are no known α-amylases which have been shown to cleave next to a −1 subsite branch and so these probes may act as negative controls for non-specific labelling but also may identify enzymes with novel activity. The reporter tags were incorporated via the O4' of the non-reducing end of the α-1,4 glucose in the −2 subsite position.

Scheme 1 provides a representative synthesis of free amine **2a**, with the α-1,6 branch in the −2 subsite, and showcasing a newly developed strategy to install, late stage and in a stereospecific manner on glycosylated cyclohexene precursors, the epoxide. Briefly, orthogonally and partially

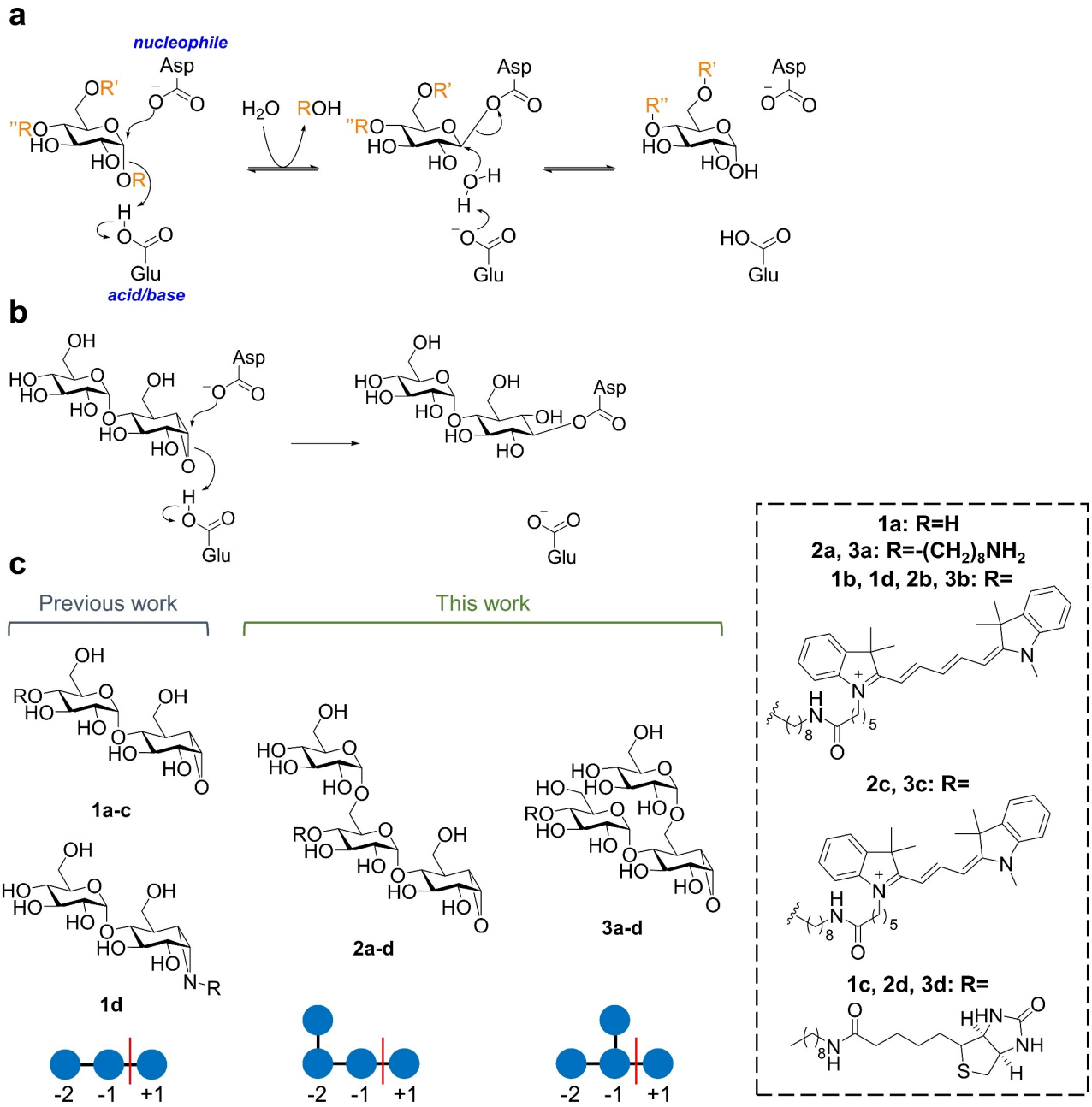
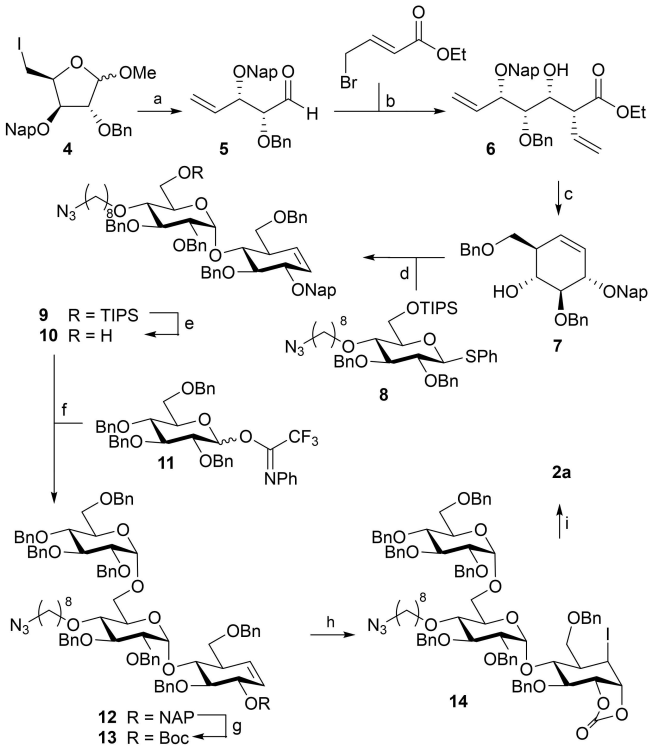


Figure 1. Design and mechanism of covalent inhibitors and activity-based probes targeting GH13 α-amylases, mimicking different substrate specificities. (**a**) Mechanism of action of GH13 α-amylases including nucleophilic attack by the active site aspartic acid and the resulting covalent intermediate. (**b**) Mechanism of epoxide activity-based probes to target the active site nucleophile and trap the bound enzyme. (**c**) Inhibitors and activity-based probes described previously and those synthesized and used in this work to further understand substrate specificity of these enzymes.

protected glucose-configured cyclohexene **7**, obtained by adaptation of existing routes,[29] was subsequently glycosylated at O4 and the O6 to arrive at trisaccharidic cyclohexene **12**. The naphthalene protective group in **12** was then exchanged for the tert-butyloxycarbonyl (**12** to **13**) setting the stage for a stereospecific Bartlett carbonate iodocyclization event (**13** to **14**).[30] Hydrolysis of the carbonate in **14** under basic conditions led to intramolecular iodine displacement, overall delivering the epoxide stereospecifically from the α-face of the alkene, and global deprotection then yielded free amine **2a**. Adaptation of this Scheme yielded free amine **3a**, and the amines in **2a** and **3a** were acylated to give Cy5 fluorophores **2b** and **3b**, Cy3 fluorophores **2c** and **3c** and biotin derivatives **2d** and **3d** (Scheme S1–3).



Scheme 1. Representative route of synthesis. Reagents and conditions: a) Zn, THF/H_2O, ultrasound, 40 °C, 85 %; b) ethyl 4-bromocrotonate, indium powder, $La(OTf)_3$, H_2O, rt, 61 %; c) *i*) Grubb's II catalyst, DCM, 40 °C, 85 %; *ii*) DIBAL-H, THF, 0 °C to rt; *iii*) $NaBH_4$, H_2O, EtOAc, rt, 87 % over 2 steps; *iv*) BnBr, 2-aminoethyl diphenylborinate, KI, K_2CO_3, MeCN, 60 °C, 95 %; d) **8**, NIS, TMSOTf, DMF, DCM, 3 Å MS, 0 °C, **9** 52 %, **10** 32 %; e) TBAF, THF, rt, 87 %; f) **11**, TMSI, $OPPh_3$, 3 Å MS, DCM, rt, 82 %; g) *i*) DDQ, DCM/H_2O (10/1), rt; *ii*) Boc_2O, DMAP, THF, 0 °C to rt, 75 % over 2 steps; h) NIS, AcOH, DCM, rt, 61 %; i) *i*) NaOMe, MeOH, DCM, rt, 81 %; *ii*) polymer-bound PPh_3, MeCN, H_2O, 65 °C, 88 %; *iii*) $Pd(OH)_2$/C, TFA, H_2, dioxane, H_2O, rt, quant.

Substrate Specificity of Human Gut Bacterial α-Amylases

The human gut microbiome contains many bacteria which express α-amylases to digest complex carbohydrates. The substrate specificity of many of these enzymes, and how they aid human digestion, is unknown. We hypothesized using our suite of α-amylase ABPs to probe the substrate specificity of these enzymes. To this end, we selected six α-amylases from prominent gut bacteria which themselves are known to grow on pullulan (a substrate consisting of maltotriosyl moieties linked by α-1–6 bonds, and hence a surrogate for branch point accommodation). $GH13B_{Sus}$ from *Bacteroides ovatus*,[31] SusA and SusG from *Bacteroides thetaiotaomicron*,[10a] Amy13B and Amy13 K from *Eubacterium rectale* and Amy5 from *Ruminococcus bromii*[32] were all expressed in *E. coli* and purified to homogeneity. The enzymes were first reacted with the original disaccharide probes, either epoxide **1b**, aziridine **1d** or monosaccharide α-glucosidase probe **15**[33] for 1 hour at 37 °C, and a fluorescence gel scan of the SDS-PAGE performed (Figure S1a). Both *Bo*$GH13B_{Sus}$ and *Bt*SusA are equally labelled by **1b** and **1d**, whilst *Er*Amy13B is labelled more strongly by aziridine **1d**. In contrast *Bt*SusG,

*Er*Amy13 K and *Rb*Amy5 are either not or only weakly labelled by **1d**. This preference between **1b** and **1d** has been observed previously with the industrial α-amylase, TAKA, with epoxide **1b** labelling much more effectively than aziridine **1d**.[24] This was hypothesized to be because of favourable interactions between the −3 subsite in the α-amylase and the O4' substituents of **1b** and unfavourable alkyl substitution of the aziridine warhead of **1d**. These subtle differences in probe preference of the human gut bacterial α-amylases could be an influence of either the type of electrophilic warhead or, more likely, a result of differing interactions between the α-amylase substrate groove and the O4' or N' reporter handles, thus also informing on the architecture of the binding subsites. All six human gut bacterial α-amylases were then reacted with **1b**, **2b** or **3b** at pH 6 (Figure 2a; the reaction pH optimum of all six enzymes with both **1b** and **2b** was 6 (Figure S1b + c)). A stoichiometric ratio of protein to probe was sufficient to give strong signal in the Cy5 fluorescence gel scan of the SDS-PAGE. All enzymes, except for *Er*Amy13 K, were labelled by both disaccharide and −2 branched probes **1b** and **2b** but not −1 branched probe **3b** (indeed, there are no known α-amylases which accept a −1 α-1,6 branch). In fact, for these five enzymes, stronger labelling occurred with **2b** than **1b**, suggesting the pseudo-trisaccharide may have higher affinity for the active site than the pseudo-disaccharide. *Er*Amy13 K is faintly labelled by **1b** but not **2b**, indicating a −2α-1,6 branch is not accepted by this enzyme.

To confirm preference of the human gut bacterial α-amylases for the −2 branch over the linear, a competition experiment was performed. The enzymes were treated with either 1 μM **1b** or Cy3 −2 branched ABP **2c** or 1 μM **1b** in combination with increasing concentrations of **2c**. Cy5 and Cy3 fluorescence gel scans were overlaid (Figure 2b). Except for *Er*Amy13 K, **1b** was outcompeted by **2c** at concentrations as low as 100 nM, confirming the suspected preference of these α-amylases for more complex substrates.

As a measure of the efficacy of the labelling reaction and to confirm the in-gel results, intact mass spectrometry of the human gut bacterial panel before and after reaction with inhibitors **1a** and **2a** was performed (Figure 2c and S2). Conversion to the covalent product was observed in all cases, with complete conversion observed with **2a** for all enzymes. Complete conversion with **1a** was only observed with *Er*Amy13B and *Rb*Amy5.

Crystal Structures of Bacteroides thetaiotaomicron SusA and SusG and *Ruminococcus bromii* Amy5 in Complex with Branched ABP 2a

Branch point accommodation in α-amylases, whilst extremely important, has rarely been studied or observed at the structural level (rare examples include AliC[27] and SusG[28,34]). Whether, and how well, branch points are accommodated defines the structure of limit dextrins (the oligosaccharide(s) that remain after hydrolysis) which

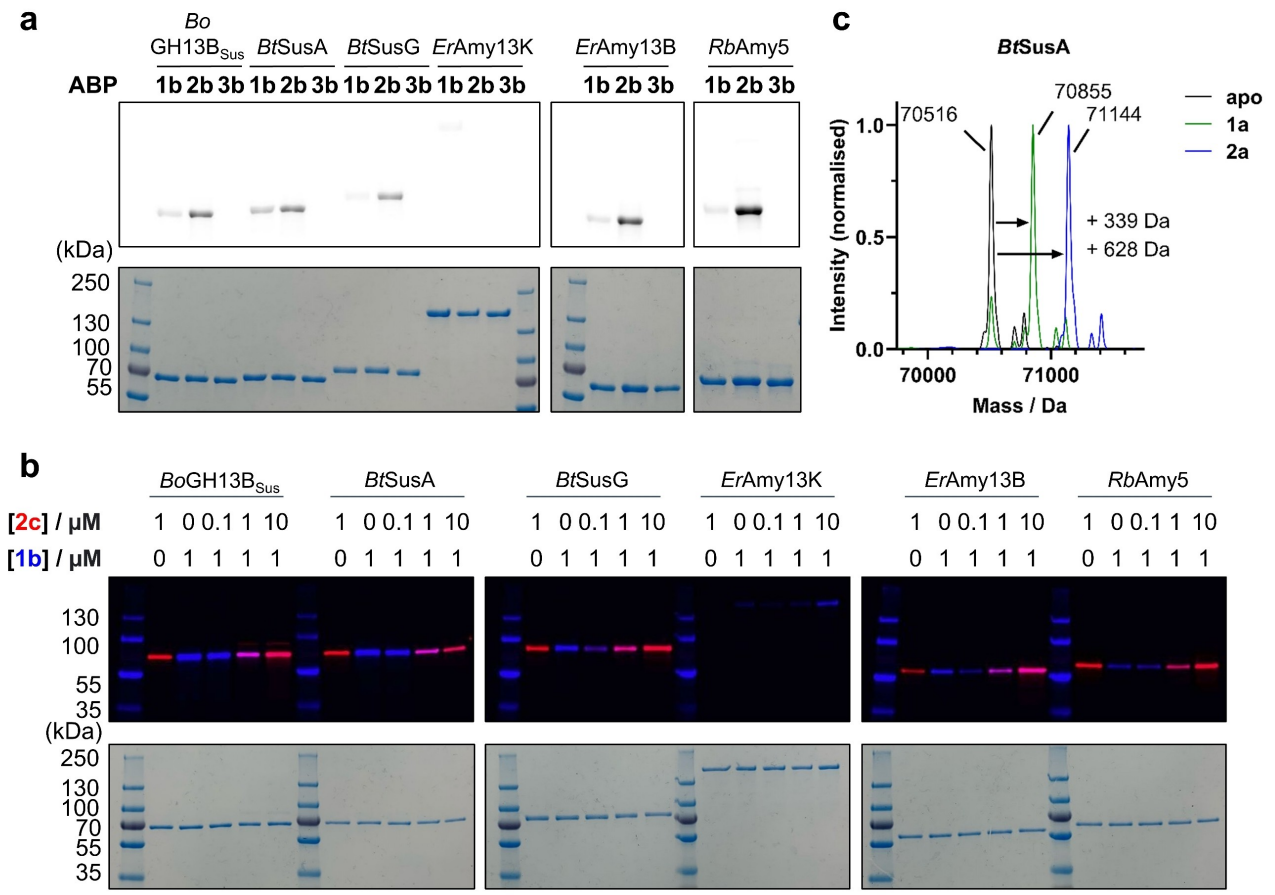


Figure 2. Substrate specificity of human gut bacterial α-amylases. (**a**) Treatment of human gut bacterial α-amylases with stoichiometric amounts of disaccharide probe **1b**, −2 branch point mimic **2b** or −1 branch point mimic **3b**. Cy5 fluorescence gel scans are the top panel and total protein Coomassie stained gels on the bottom. (**b**) Competitive ABPP of human gut bacterial α-amylases with Cy5 disaccharide probe **1b** and Cy3 branch point probe **2c**. Cy3 (red) and Cy5 (blue) fluorescence gel scans are overlaid in the top panel and total protein Coomassie stained gels are on the bottom. (**c**) Intact mass spectrum of *Bt*SusA in the absence and presence of disaccharide probe **1a** and −2 branch point probe **2a**.

impacts on diverse processes such as the completeness of starch degradation (and thus the need for, and identity of, other enzymes), dietary glucose uptake and the mouthfeel of beer. To validate both these branch point preferences, and the use of our ABPs to predict substrate specificity, the crystal structures of *Bt*SusA, *Bt*SusG and *Rb*Amy5 were solved in complex with **2a** to 2.4, 2.6 and 1.4 Å resolution respectively. The inhibitor reacted to form a covalent bond with the nucleophilic aspartic acid in each structure and the ligand was modelled at an occupancy of 1 (notably, this also represents the first structure of *Bt*SusA; Figure 3a). There are six chains in the unit cell of *Bt*SusA, with PISA analysis[35] indicating the interface of chains D and F most energetically favourable. Size-exclusion chromatography-multi angle light scattering (SEC-MALS) of *Bt*SusA in solution revealed a dimer as the only species present (Figure S3). The ligand makes many key hydrogen bonding interactions with neighbouring residues of *Bt*SusA, including O3 of the original epi-cyclophellitol ring with Asp440 (2.3 Å), O2 and O3 of the glucose in the −2 subsite with Arg493 (2.7 Å) and Asp489 (2.4 Å) respectively, and weak interactions between the α-1,6 linked glucose and Thr295 (~3.2 Å). A small flexible region was observed in chain B (residues 332–336) adjacent to the active site nucleophile, but this change did not affect the binding pose of **2a**.

2a makes similar hydrogen bonding interactions in *Bt*SusG as those observed in *Bt*SusA, apart from interactions observed between the pseudo-O6 (CH_2OH) and Met298 and His243 in *Bt*SusA which are not observed (Phe345 in this position in *Bt*SusG; Figure 3b). An additional un-reacted molecule of **2a** was observed bound to chain A confirming a known[28] surface starch-binding site. SEC-MALS analysis demonstrated the protein is a monomer in solution (Figure S3), in agreement with previous studies. SEC-MALS analysis also confirmed *Rb*Amy5 is a monomer in solution (Figure S3). Similar hydrogen bonding interactions were observed in *Rb*Amy5 as *Bt*SusA, with Ser213 in the place of Thr295 (Figure 3c).

The structure of the GH13 domain of *Er*Amy13 K is currently experimentally undetermined but in the Alpha-Fold model of *Er*Amy13 K (AF-A0A450VHT8-F1),[26] Trp820 clashes with the α-1,6 linked glucose of **2a** in the

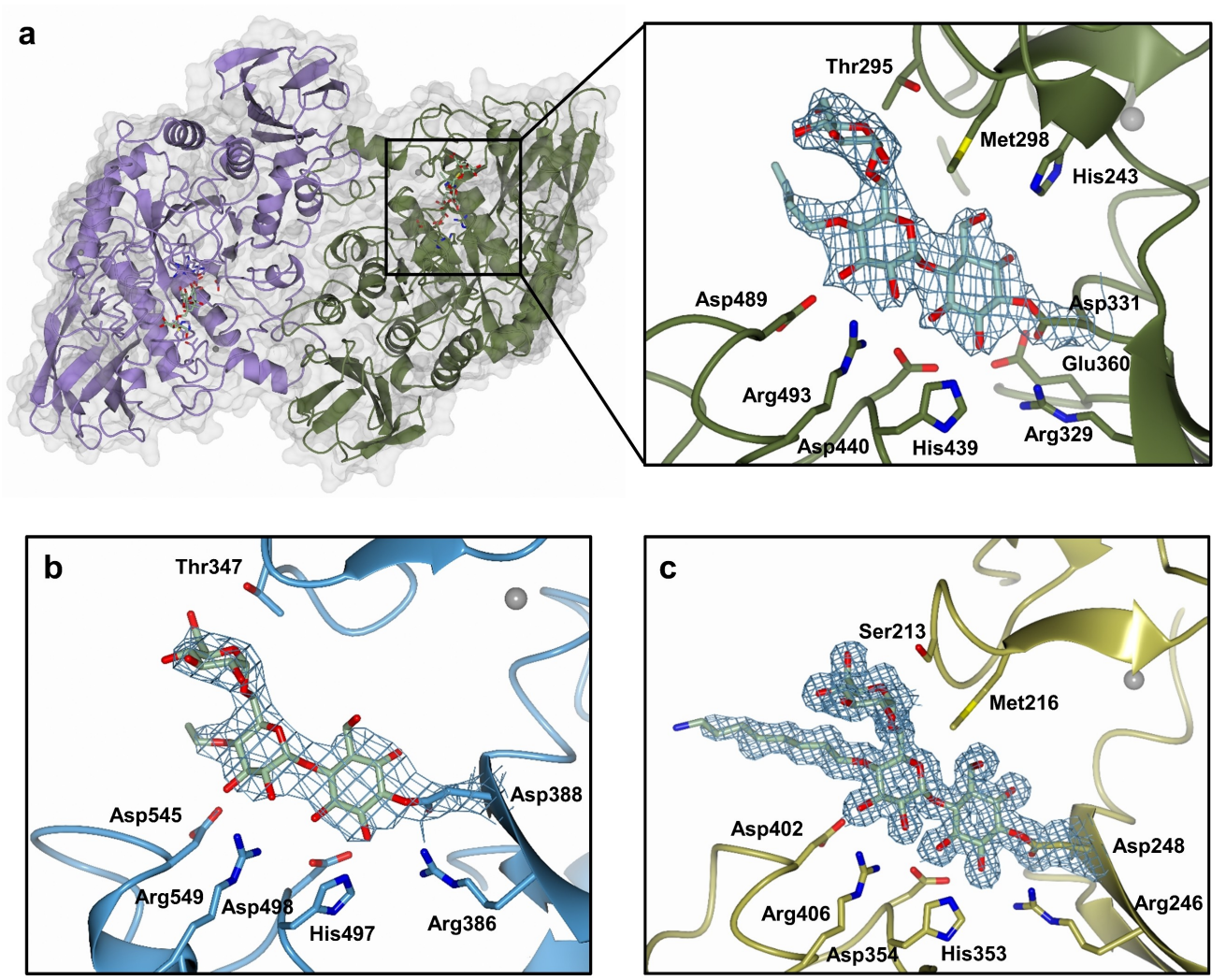


Figure 3. Crystal structures of human gut bacterial α-amylases *Bt*SusA, *Bt*SusG and *Rb*Amy5 covalently bound to branched inhibitor **2a**. (**a**) Overall fold of the *Bt*SusA dimer and one active site showing interactions of the protein with covalently bound ligand **2a**. The two chains (D + F) of the homodimer are depicted in purple and dark green, the bound ligand in sea green and the $2F_O–F_C$ map around the ligand in blue at levels of 1 σ. (**b**) The active site interactions of *Bt*SusG with covalently bound inhibitor **2a**. The protein chain is depicted in blue, the bound ligand in sea green and the $2F_O–F_C$ map around the ligand in blue at levels of 1 σ. (**c**) The active site interactions of *Rb*Amy5 with covalently bound inhibitor **2a**. The protein chain is depicted in gold, the bound ligand in sea green and the $2F_O–F_C$ map around the ligand in blue at levels of 1 σ.

overlaid *Rb*Amy5 structure (Figure S4). This clash in the in silico determined structure, may explain the lack of labelling of *Er*Amy13 K by branched probe **2b**.

Selective Labelling of Bacteroides thetaiotaomicron α-Amylases in Whole Cell Lysates

To test the probes on a more complex system and to study the prominent gut bacteria's α-amylases further, *Bacteroides thetaiotaomicron* cells were grown on glucose, maltose, pullulan, autoclaved soluble potato starch or autoclaved corn starch. The whole cell lysates were harvested and treated with either 1 μM **1b**, **2c**, or **1b** and **2c** in combination, and their protein content resolved by SDS-PAGE. The Cy5 and Cy3 fluorescence gel scans were overlaid (Figure 4a). Two bands corresponding to *Bt*SusA (71 kDa) and *Bt*SusG (78 kDa) were observed when grown on maltose or higher order oligosaccharides. Detection of maltose or a higher order oligosaccharide by the transcriptional regulator, SusR, is required for expression of the Sus operon.[10a,36] Regardless of the carbon source, labelling of both α-amylases by **1b** is completely abolished by competition with branched probe **2b**. An additional band of around 50 kDa is observed for potato starch lysates when reacted with **2b**.

The *Bacteroides thetaiotaomicron* VPI-5482 genome encodes seven GH13 (potential α-amylase) enzymes (see www.cazy.org/b134.html). Of these, only SusA and SusG show a fitness defect when grown on starch and the roles of the five remaining enzymes (four of which are predicted to be intracellular and thus may play a role in glycogen

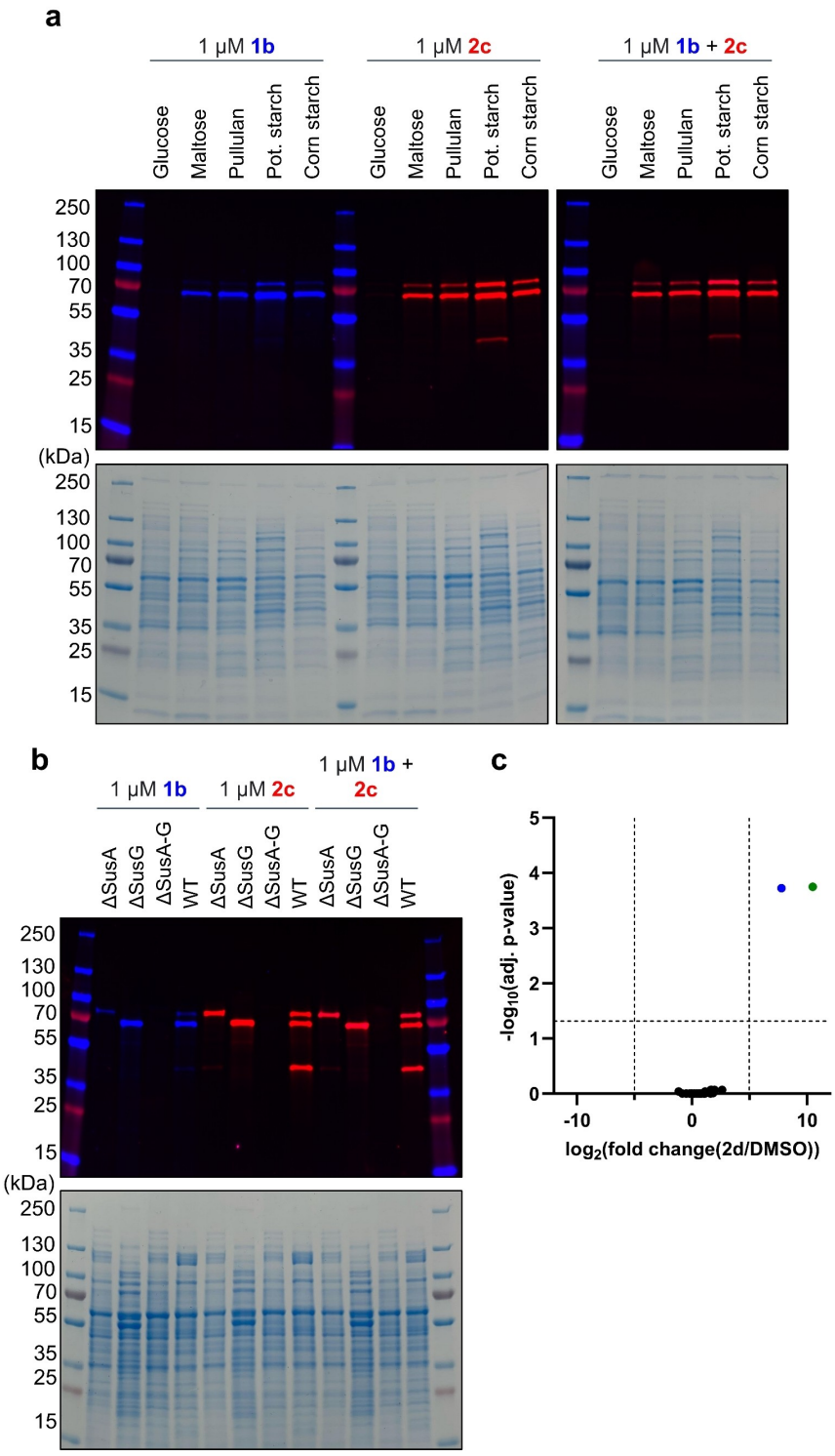


Figure 4. Detection of SusA and SusG in *Bacteroides thetaiotaomicron* whole cell lysates. (**a**) *Bacteroides thetaiotaomicron* was grown on minimal media with either glucose, maltose, pullulan, autoclaved potato starch or autoclaved corn starch, the cells lysed and the lysates reacted with 1 μM of either Cy5 disaccharide probe **1b** (blue), Cy3 branched probe **2c** (red) or both in competition. (**b**) *Bacteroides thetaiotaomicron* with the genes encoding SusA, SusG or SusA–G removed and wild type (WT) were grown on maltose and the lysates reacted with 1 μM of either Cy5 disaccharide probe **1b** (blue), Cy3 branched probe **2c** (red) or both in competition. In (**a**) and (**b**) an overlay of Cy5 (blue) and Cy3 (red) fluorescence is shown in the top panel and the Coomassie stained gel is shown below. (**c**) Volcano plot of the identified proteins from the *Bacteroides thetaiotaomicron* **2d** pull-down analysis by LC–MS/MS. SusA is indicated in green and SusG in blue on the graph. A minimum $\log_2$ fold change of 5 over the DMSO control and a Benjamini-Hochberg adjusted p value of lower than 0.05 (equivalent to a $\log_{10}$ value of 1.3) was required for a hit. All identified proteins are described in Supplementary File 2. The Benjamini-Hochberg adjusted p values are plotted here and non-adjusted p values are plotted in Figure S6.

metabolism) are unknown; they do not lie in PULs.[37] To determine the identity of the labelled α-amylases, deletion mutants of *Bacteroides thetaiotaomicron* were obtained in which either SusA, SusG or the whole Sus operon removed.[38] These mutants were grown on maltose and their lysates treated with either 1 μM **1b**, **2c** or **1b** and **2c** in combination (Figure 4b). When SusA was deleted the band at ~71 kDa disappeared and when SusG was removed, the band at ~78 kDa was not visible. When the whole Sus operon was removed, no labelled proteins were observed with either probe. The unknown band at ~50 kDa is only visible in the wild type and ΔSusA cells. This unknown protein is therefore likely to be a SusG degradation product, with the mass corresponding roughly to the larger subunit of the structure.

A pull down of *Bacteroides thetaiotaomicron* lysates with biotinylated probe **2d** was performed, to confirm the selectivity of −2 branched probes for SusA and SusG. First, lysates were treated with Cy5 probe **2b** and biotin probe **2d** in a competitive ABPP experiment (Figure S5). The Cy5 fluorescence gel scan and streptavidin-HRP western blot showed that **2d** can label at least one α-amylase and can outcompete **2b** when present in excess. The lysates were then treated with 1 μM **2d** before pull-down using magnetic streptavidin beads. An on-bead digest was performed, and the tryptic peptides analysed using a TIMS-TOF HT LC–MS/MS. Peptides identified as belonging to SusA and SusG were only present in the probe-treated samples and not in the negative control and these were the only proteins which were significantly enriched by the probe (Figure 4c).

Profiling pH and Temperature Dependence of a Panel of Industrial α-Amylases

We then aimed to apply our suite of probes to the industrial processing of starch. We envisaged using the probes to quickly profile panels of enzymes to select the correct α-amylase for the correct condition, but to also identify new α-amylases with novel properties. With the previously described maltobiose inhibitors and ABPs in hand,[24] we used them to profile known industrial α-amylases rapidly. Five α-amylases commonly used in industry were either bought in powder form ('TAKA' from *Aspergillus oryzae*)[39] or expressed and purified from *E. coli* ('BHA' from *Bacillus halmapalus*,[40] 'BLA' from *Bacillus licheniformis*,[41] 'K38' from *Bacillus sp. KSM-K38*[42] and 'AliC' from *Alicyclobacillus sp. 18711*[27]). All five purified enzymes were labelled by original Cy5 maltobiose epoxide probe **1b** at pH 6 after incubation with a 10× molar excess of ABP (Figure S7a). TAKA was labelled most strongly, likely reflecting its shorter active centre with fewer subsites contributing to binding. The "industrial panel" was then reacted at pH 3 to 9, with different pH values giving optimum activity for different enzymes (Figure 5a and S7b). TAKA was labelled most strongly at pH 5, compared to a literature optimum for catalysis of pH 5.6 (activity measured using the Blue Value method, which is a simple colorimetric assay for the presence of starch amylose).[43] Both BLA (reported maximal activity between pH 5 and 8, also using the Blue Value method)[41] and AliC (optimum pH not reported) also labelled best at pH 4.5, K38 at pH 5.5 (reported optimum of 8.0 to 9.5, using dinitrosalicylic acid (DNSA) reducing sugar assay)[42] but BHA labelled consistently over a range of pHs (noted to demonstrate activity above pH 9;[40] increasing activity from pH 6 to 10 using a PNP-glucose assay[44]). The optimum pH for labelling with ABP **1b** is consistently 1–2 pH units below the pH activity optima reported using other methods.

The enzyme panel was next profiled for thermal stability at temperatures ranging from 15 to 95 °C (Figure 5b). Each enzyme was incubated at the indicated temperature at pH 6 for 1 hour before cooling to room temperature and subsequent reaction with **1b** at 37 °C. The different thermal stabilities of each enzyme measured using the ABPP in-gel method were compared to those observed by nano-differential scanning fluorimetry (nano-DSF) of the unliganded α-amylase (Figure S7c). Nano-DSF is a measure of the intrinsic fluorescence of a protein's tryptophan residues, changes in which can be correlated to protein unfolding. TAKA and K38 were the least thermostable in the conditions used, reflected in their nano-DSF melting temperatures of 61 and 59 °C respectively, whereas BHA and BLA, which are prized for their greater thermostability, are indeed active up to 75 °C (nano-DSF melting temperatures 67 and 80 °C respectively).

Calcium Dependence of Industrial α-Amylases

Given the importance of amylase activity at low [Ca^{2+}] for high-fructose corn syrup manufacture, we then set out to investigate whether our probes could be used to monitor calcium sensitivity of α-amylases. We took the reported calcium-independent α-amylase, K38, and compared directly with well-studied calcium-dependent α-amylase, BLA. K38 and BLA were incubated at 40 °C for 2 hours at pH 10 with chelating agents EDTA or more potent Ca^{2+} chelator EGTA, in order to directly compare with experiments performed by Hagihara et al.[42] The proteins were then buffer exchanged to pH 6, with concurrent removal of the chelating agents, before treatment with **1b** (Figure 5c). To validate the results, a reducing sugar assay with colorimetric dye DNSA was also performed, where both K38 and BLA were incubated at pH 10 with EDTA or EGTA for 2 hours before the α-amylase activity on 1 % potato starch was monitored (Figure 5d). K38 activity was consistent between the no chelating agent control and the EDTA and EGTA reactions, as also observed from the reducing sugar assay with DNSA. Conversely, BLA activity proved sensitive to EDTA and EGTA in both the ABPP assay with **1b** and the reducing sugar assay.

Substrate Specificity of Industrial α-Amylases

To study industrial α-amylase substrate specificity, we again used branch point mimetic probes **2b** and **3b** in combination

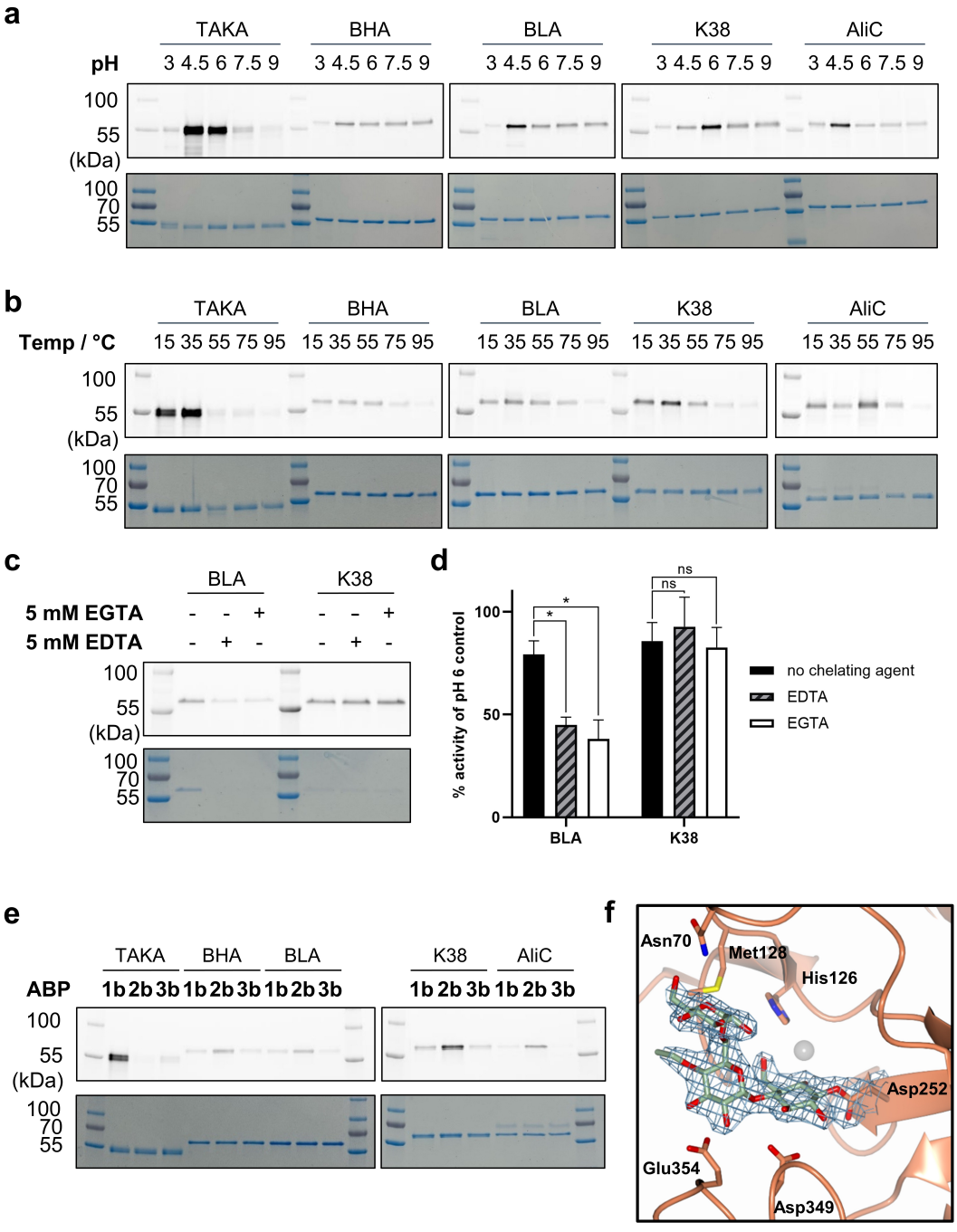


Figure 5. Profiling of industrial α-amylases for condition and substrate specificity. (**a**) The panel of industrial α-amylases was reacted at pH 3, 4.5, 6, 7.5 and 9 with a 10× excess of ABP **1 b**. (**b**) The enzymes were first incubated at the indicated temperature for 1 hour at pH 6 before treatment with a 10× excess of **1 b** at 37 °C. (**c**) The enzymes were incubated at 40 °C, pH 10 for 2 hours with either 5 mM EDTA, 5 mM EGTA or water control, before treatment with **1 b** at 37 °C at pH 6. (**d**) Activity of each enzyme after incubation at 40 °C, pH 10 for 2 hours, measured using the 3,5 dinitrosalycylic acid (DNSA) assay, normalized to the activity at pH 6 without chelating agent. Plotted as the average of three repeats (each the average of two technical repeats) ± SD. An unpaired t test was performed to declare significance (for ns (non-significant), $p > 0.05$, for *, $p \leq 0.05$ and for **, $p \leq 0.01$). (**e**) Treatment of the industrial enzymes with 10× excess disaccharide probe **1 b**, −2 branch point mimic **2 b** or −1 branch point mimic **3 b**. (**f**) Crystal structure of K 38 α-amylase depicting the active site interactions with covalently bound inhibitor **2 a**. The protein chain is depicted in orange, the bound ligand in sea green and the $2F_O$–F_C map around the ligand in blue at levels of 1 σ. In (**a**–**c**) and (**e**) Cy5 fluorescence gel scans are the top panel and total protein Coomassie stained gels on the bottom.

with disaccharide probe **1b**. Each enzyme was incubated with a 10× excess of either **1b**, **2b** or **3b** at pH 6 (Figure 5e). TAKA was most strongly labelled by the maltobiose probe **1b**, whereas AliC, which is known to accommodate a −2 branch point (as well as +1 and +2 branch points),[27] was labelled most strongly by **2b**. The substrate specificity of K38 has not yet been described and the finding that K38 reacts most effectively with the −2 branched ABP **2b** strongly points towards its preference to process branched substrates (this finding is supported by structural analysis, see below). BHA and BLA are weakly labelled by both **1b** and **2b** but TAKA is only labelled by **1b**. There is unexpected, but very weak labelling of all five enzymes by the −1 branched probe **3b** suggesting a very small degree of accommodation (which may demand structural rearrangements), or perhaps a low level of non-specific binding at the high probe concentrations.

To investigate further the unexpected strong labelling of K38 by branched probe **2b**, the crystal structure of the enzyme in complex with **2a** was solved to 2.0 Å resolution through co-crystallisation (Figure 5f). Clear density of reacted **2a** was observed protruding from Asp252, allowing the ligand to be modelled at an occupancy of 1. The probe reacted to form a covalent adduct with Asp252 with the unhindered −2 α-1,6 branch making key hydrogen bonding interactions, such as O3 of the α-1,6 linked glucose and the backbone nitrogen of Val75 (2.9 Å), O3 and O4 of the cyclitol with His348 (2.9 Å) and Asp 349 (2.8 Å), and O3 of the α-1,4 linked glucose and Ser355 (2.6 Å).

Overlay of the new K38 structure with published structures of TAKA, BHA and BLA suggest the α-1,6 branch in the −2 position may be sterically hindered in TAKA by Trp83 (alanine in the three other enzymes) and His80 (threonine in the three other enzymes; Figure S8). The branch may be accepted in BHA and BLA, as indicated by labelling with **2b**, as there are no clashes with the branch when overlaid with the K38 structure. Although weaker overall binding of BHA and BLA with the disaccharide ABPs, reflecting the comparatively extended binding clefts of these enzymes compared to TAKA, makes this harder to quantify without structural evidence.

Identification of Thermophilic α-Amylases from Commercial Compost

To assess the utility of the ABPs to find new α-amylases with favourable properties, such as high thermostability, the probes were used with a species-rich commercial composter sample treated with high temperatures. A sample was taken from four-day old material in an in-vessel composting process, with domestic waste as starting material including leaves, twigs, grass and food waste, which had been subjected to an initial 48-hour period at 70 °C. The sample was cultured in minimal media plus glucose overnight and this pre-culture was used to then inoculate minimal media supplemented with autoclaved soluble potato starch for growth at 55 °C. The cells were lysed at pH 6 and treated with 10 μM **1b**, **2b** or **3b** (Figure 6a) and subjected to the in gel ABPP protocol. A strong band was observed around 70 kDa with both **1b** and **2b**, and a less intense one when using **3b**. To rule out degradation of the branched probes by endogenous α-glucosidase enzymes in the lysate mixture, the samples were also pre-incubated with covalent α-glucosidase inhibitor **16**[33] and non-covalent inhibitor 1-deoxynojirimycin (DNJ) before treatment with the ABPs (Figure S9a). Labelling was not inhibited by pre-treatment with these probes, suggesting the probes are not degraded in the lysate environment. ABPP was then performed on samples exposed to varying pH and temperature (Figure 6b + c). The strong band at ~70 kDa is observed between pH 6 and 9, suggesting tolerance of some, or all, of the ABP-active enzymes to high pH. The lysate was next pre-incubated at temperatures from 15 to 75 °C before treatment with **1b** or **2b** as a test of thermal stability. Consistent activity was observed after incubation at 55, but not 75 °C.

To identify the species that make up the commercial composter sample, the total genomic DNA was then extracted from the cultured cells and sequenced. DNA from *Aeribacillus pallidus*, *Geobacillus thermodenitrificans*, *Parageobacillus toebii*, *Paenibacillus sp 32OW*, *Thermobacillus composti*, *Hydrogenophilus thermoluteolus*, *Tepidiphilus thermophilus*, *Pseudoxanthomonas suwonensis* and *Tepidiphilus margaritifer* was detected and the proteomes of these bacteria were used to generate a library which was used as a proteomics database. Both whole cell lysates and concentrated secretomes after growth on autoclaved potato starch were reacted with **1c**, **2d** or **3d**, pulled down with streptavidin beads, on-bead digested, and the resultant tryptic peptides analysed by LC–MS/MS (Figure 6d–f). The results were searched against the sequenced proteomics database using MSFragger[45] in FragPipe,[45] with IonQuant[46] used to reduce the false discovery rate. Three proteins were pulled down by disaccharide probe **1c** with a $\log_2$ fold change of greater than 5 compared to the DMSO control and a Benjamini-Hochberg adjusted p value of lower than 0.05 (the Benjamini-Hochberg adjusted p values are plotted in Figure 6d–f and the non-adjusted values in Figure S10). These were annotated as neopullulanases from *Parageobacillus toebii* (A0A150MUK2 and A0A150MUL0) and *Geobacillus thermodenitrificans* (A4IKY8). Surprisingly, one of these annotated neopullulanases (A0A150MUK2) was not also captured by −2 branched probe **2d**. Study of the AlphaFold structures of the three putative neopullulanases overlaid with the structure of *R. bromii* Amy5 reacted with **2a** (Figure S11), reveals A0A150MUK2 may not accept a −2 branch point due to blockage of this subsite by Trp127 (Asp in *R. bromii* Amy5, A0A150MUL0 and A4IKY8) and Tyr169 (Ser in *R. bromii* Amy5 and Tyr in A0 A150MUL0 and A4IKY8 but with an alternate predicted loop configuration). No proteins were captured by −1 branched probe **3d**. A pull-down experiment was also performed on the concentrated secretome after growth on autoclaved potato starch, but this yielded no significant hits over three biological repeats. These results highlight the use of ABPs to quickly and easily discover novel α-amylases from complex, crude environmental samples, whilst also predict-

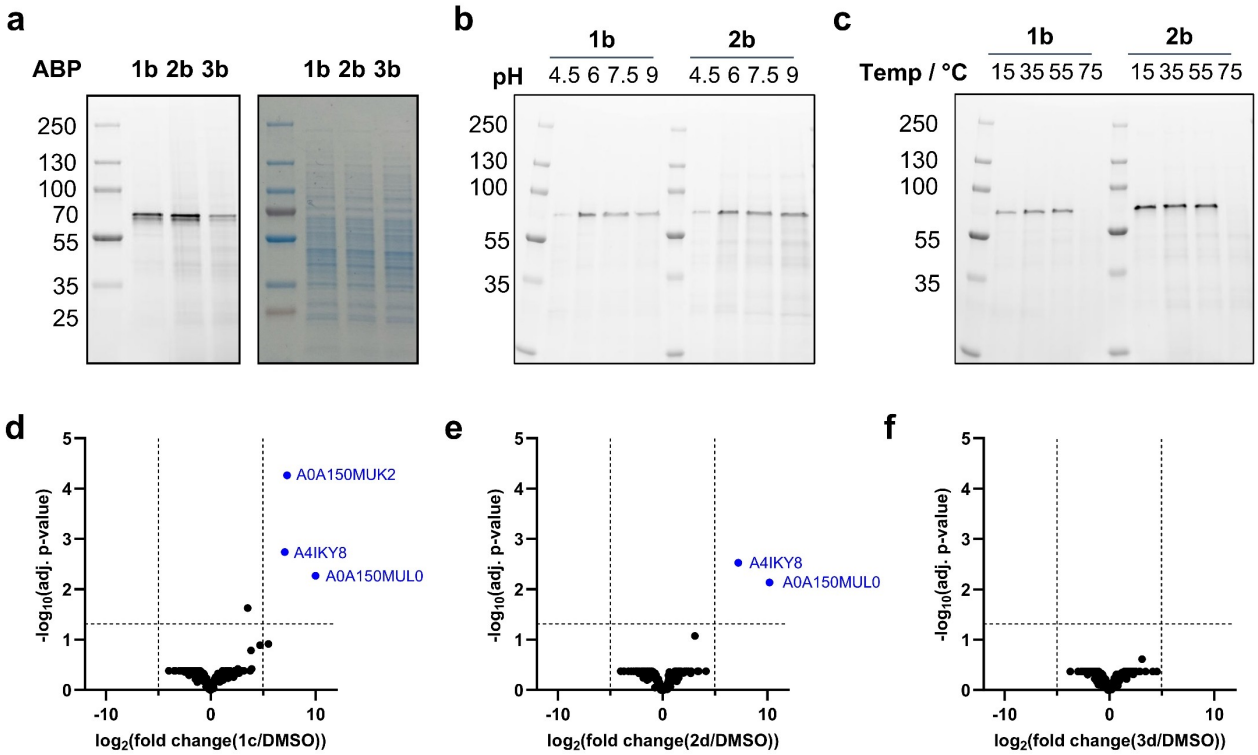


Figure 6. Discovery of novel GH13 enzymes from a complex commercial composter sample. (**a**) Whole cell lysates from growth of a commercial composter sample on minimal media plus potato starch were treated with probes **1b**, **2b** and **3b**. The Cy5 fluorescence gel scan is shown on the left and the coomassie stained gel on the right. (**b**) Cell lysate was treated with **1b** or **2b** at different pH points. The Cy5 fluorescence gel scan is shown here and the Coomassie stained gel in Figure S9b. (**c**) Cell lysate was incubated at the indicated temperature for 1 hour before treatment with **1b** or **2b** at 37 °C. The Cy5 fluorescence gel scan is shown here and the Coomassie stained gel in Figure S9c. (**d–f**) Volcano plot of observed proteins in LC–MS/MS analysis of the cell lysate treated with **1c** (**d**), **2d** (**e**) or **3d** (**f**) and pulled down using streptavidin beads. A minimum $\log_2$ fold change of 5 over the DMSO control and a Benjamini-Hochberg adjusted p value of lower than 0.05 (equivalent to a $\log_{10}$ value of 1.3) was required for a hit. The Benjamini-Hochberg adjusted p values are plotted here and the significant values are highlighted in blue and labelled. Non-adjusted p values are plotted in Figure S10. All identified proteins are described in Supplementary File 2.

ing their substrate specificity without the need for structural information.

Conclusions

Branched amylase activity-based probes allow the detailed and rapid dissection of starch degrading enzymes, individually and as parts of ecosystems. Bacterial α-amylases are amongst the most prominent enzymes expressed in the gut environment and have evolved to digest the complex α-glucans provided by the host. A recent study of 354 different human and animal Bacteroidota strains demonstrated that α-glucans are the most widely utilized complex polysaccharide.[47] Of a panel of several dozen different plant, animal, and microbial carbohydrates, pullulan was the most widely utilized by these Bacteroidota, suggesting some specialization by these bacteria to accommodate or tolerate α-1,6 linkages within an α-1,4 polymer. The substrate specificity of many of these gut bacterial enzymes has not been well studied, however the structure of SusG from *Bacteroides thetaiotaomicron* has previously been solved with an α-1,6 branch in the −2 position.[28]

Six α-amylases from prominent gut bacteria, which can grow on α-1,4-α-1,6 linked pullulan in isolation, were expressed and purified to homogeneity. All the representative gut bacterial α-amylases demonstrated a preference for −2 branched probe **2b** over pseudo-disaccharide **1b**, with the exception of *Er*Amy13 K, suggesting specificity and higher affinity for branched substrates. Structural studies with **2a** confirmed accommodation of the −2α-1,6 branch, highlighting the power of the ABPP approach to rapidly understand substrate specificity. As an illustration of this, *Bacteroides thetaiotaomicron* was grown on several carbohydrate sources and the resultant lysates treated with **1b** and **2b**, to probe for α-amylase substrate specificity in a complex bacterial cell lysate. Bands suspected to correspond to SusA and SusG were detected when grown on maltose or a higher order oligosaccharide, consistent with the fact that detection of maltose by SusR is required for expression of the Sus operon.[10a,36] Both proteins reacted more readily with **2b** than with **1b**, in agreement with the results from the pure proteins studies and ΔSusA, ΔSusG and ΔSusA–G knockout strains allowed identification of the two labelled bands. Pull-down of the SusA and SusG using biotinylated −2 branched probe **2d** both confirmed the identity of the in-gel labelled

enzymes and demonstrated the use of the ABPs to detect α-amylases in complex environments. We next used disaccharide probe **1b** to profile industrially useful bacterial α-amylases for their resilience against a range of conditions. These pH and temperature optimums and calcium-sensitivities of the industrial α-amylases correlated with known values measured by alternative methods, validating the use of the probes to quickly and accurately identify the correct enzyme for a required condition. The substrate specificities were also annotated, with known α-1,6 accommodating α-amylase AliC,[27] but also K38, demonstrating a strong preference for −2 branched **2b** over **1b**. Comparison of the active sites of these five α-amylases corroborated the lack of labelling of TAKA by **2b**. This agreement between the labelling and structural studies, validates our more straightforward method to assign substrate specificity.

Finally, to demonstrate how our ABPs could be used to find new commercially advantageous α-amylases, the biotinylated probes were used to identify enzymes from a complex industrial composter sample. Novel putative neopullulanases were detected which had not previously been characterized and their substrate specificities assigned. These enzymes were active with both **1b** and **2b** from pH 6 to 9 and after incubation at temperatures as high as 55 °C, characteristics which are advantageous for the industrial processing of starch. Work to further profile these enzymes is needed, but their discovery highlights the usefulness of this technique to find societally useful α-amylases from less well-characterized bacteria.

In conclusion, our suite of α-amylase ABPs and inhibitors allow quick evaluation of optimum conditions for activity and substrate specificity for both biotechnological applications and study of starch digestion in the human gut. These tool compounds could facilitate understanding of how starch is degraded by the human gut microbiome, including the use of the probes to identify expression of these enzymes in complex gut microbiome samples. Levels of specific α-amylases, in response to different dietary factors, could be identified using the biotinylated probe, pull-down method, linking diet to bacterial starch degradation. The probes enable highly specific identification of novel GH13 enzymes from crude environmental samples with advantageous reaction optimums. We envisage using the probes to quickly identify ideal α-amylases, selected for activity under specific conditions in the industrial processing of starch. The here presented activity-based α-amylase probes and ABPP technology may find use, both to identify new industrially relevant α-amylases and their resilience towards varying (extreme) conditions, and to directly monitor levels of α-amylases in human gut microbiomes.

Supporting Information

The authors have cited additional references within the Supporting Information.[24,26,29–30,32a,34,38,42b,45–46,48]

Acknowledgements

We thank the Royal Society (Ken Murray Research Professorship to G.J.D.), the Netherlands Organization for Scientific Research (NWO TOP grant 2018–714.018.002 to H.S.O.) and the European Research Council (ERC-2011-AdG-290836 "Chembiosphing" to H.S.O. and ERC-2020-SyG-951231 "Carbocentre" to G.J.D. and H.S.O.). We thank Diamond Light Source for beamtime (proposals 24948 and 32736), and the staff of beamline I03 for assistance with crystal testing and data collection. We thank Johan Turkenburg and Sam Hart for coordinating X-ray data collection, the Bioscience Technology Facility and Andrew Leech for assistance with SEC-MALLS analysis, and Adam Dowle, Chris Taylor and Chloë Baldreki for assistance with mass spectrometry data collection (all University of York). Mass spectrometry was performed at the York Centre of Excellence in Mass Spectrometry, which benefited from capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent support from EPSRC (EP/K039660/1; EP/M028127/1). We also thank Ryedale Organics Ltd. for donation of a compost sample.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Atomic coordinate files and structure factors have been deposited in the Protein DataBank (PDB) with accession code 9FYZ (BtSusA), 9FZ0 (BtSusG), 9FZ2 (RbAmy5) and 9FZ3 (K38). Data collection and refinement statistics are presented in Supporting Information Tables S1+2. Proteomics data has been deposited to the ProteomeXchange Consortium via the PRIDE partner repository and can be accessed with the data set identifiers PXD053850 (B. theta. pull-down) and PXD053848 (compost culture pull-down).

Keywords: activity-based probe · carbohydrates · enzymes · glycoside hydrolase · structural biology

[1] E. Drula, M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat, N. Terrapon, *Nucleic Acids Res.* **2021**, *50*, D571–D577.
[2] N. Kaur, V. Kumar, S. K. Nayak, P. Wadhwa, P. Kaur, S. K. Sahu, *Chem. Biol. Drug Des.* **2021**, *98*, 539–560.
[3] R. Gupta, P. Gigras, H. Mohapatra, V. K. Goswami, B. Chauhan, *Process Biochemistry* **2003**, *38*, 1599–1616.
[4] O. Kirk, T. V. Borchert, C. C. Fuglsang, *Curr. Opin. Biotechnol.* **2002**, *13*, 345–351.
[5] P. M. de Souza, P. de Oliveira Magalhães, *Braz. J. Microbiol.* **2010**, *41*, 850–861.
[6] B.-H. Lee, L. Yan, R. J. Phillips, B. L. Reuhs, K. Jones, D. R. Rose, B. L. Nichols, R. Quezada-Calvillo, S.-H. Yoo, B. R. Hamaker, *PLoS One* **2013**, *8*, e59745.

[7] E. A. MacGregor, Š. Janeček, B. Svensson, *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **2001**, *1546*, 1–20.

[8] a) C. A. Tarling, K. Woods, R. Zhang, H. C. Brastianos, G. D. Brayer, R. J. Andersen, S. G. Withers, *ChemBioChem* **2008**, *9*, 433–438; b) C. Tysoe, L. K. Williams, R. Keyzers, N. T. Nguyen, C. Tarling, J. Wicki, E. D. Goddard-Borger, A. H. Aguda, S. Perry, L. J. Foster, R. J. Andersen, G. D. Brayer, S. G. Withers, *ACS Cent. Sci.* **2016**, *2*, 154–161; c) C. Li, A. Begum, S. Numao, K. H. Park, S. G. Withers, G. D. Brayer, *Biochemistry* **2005**, *44*, 3347–3357.

[9] a) A. E. Kaoutari, F. Armougom, J. I. Gordon, D. Raoult, B. Henrissat, *Nat. Rev. Microbiol.* **2013**, *11*, 497–504; b) F. M. Cerqueira, A. L. Photenhauer, R. M. Pollet, H. A. Brown, N. M. Koropatkin, *Trends Microbiol.* **2020**, *28*, 95–108; c) Y. Furusawa, Y. Obata, S. Fukuda, T. A. Endo, G. Nakato, D. Takahashi, Y. Nakanishi, C. Uetake, K. Kato, T. Kato, M. Takahashi, N. N. Fukuda, S. Murakami, E. Miyauchi, S. Hino, K. Atarashi, S. Onawa, Y. Fujimura, T. Lockett, J. M. Clarke, D. L. Topping, M. Tomita, S. Hori, O. Ohara, T. Morita, H. Koseki, J. Kikuchi, K. Honda, K. Hase, H. Ohno, *Nature* **2013**, *504*, 446–450; d) A. L. McOrist, R. B. Miller, A. R. Bird, J. B. Keogh, M. Noakes, D. L. Topping, M. A. Conlon, *The Journal of Nutrition* **2011**, *141*, 883–889; e) M. N I, *The American Journal of Clinical Nutrition* **1984**, *39*, 338–342.

[10] a) M. H. Foley, D. W. Cockburn, N. M. Koropatkin, *Cell and Molecular Life Sciences* **2016**, *73*, 2603–2617; b) H. J. Flint, E. A. Bayer, M. T. Rincon, R. Lamed, B. A. White, *Nat. Rev. Microbiol.* **2008**, *6*, 121–131.

[11] N. Rana, A. Walia, A. Gaur, *National Academy Science Letters* **2013**, *36*, 9–17.

[12] B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383–414.

[13] L. Wu, Z. Armstrong, S. P. Schröder, C. de Boer, M. Artola, J. M. F. G. Aerts, H. S. Overkleeft, G. J. Davies, *Curr. Opin. Chem. Biol.* **2019**, *53*, 25–36.

[14] D. E. Koshland Jr, *Biological Reviews* **1953**, *28*, 416–436.

[15] D. L. Zechel, S. G. Withers, *Acc. Chem. Res.* **2000**, *33*, 11–18.

[16] S. Atsumi, K. Umezawa, H. Iinuma, H. Naganawa, H. Nakamura, Y. Iitaka, T. Takeuchi, *The Journal of Antibiotics* **1990**, *43*, 49–53.

[17] M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. C. H. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, B. Hooibrink, C. E. M. Hollak, R. Ottenhoff, R. G. Boot, G. A. van der Marel, H. S. Overkleeft, J. M. F. G. Aerts, *Nat. Chem. Biol.* **2010**, *6*, 907–913.

[18] L. I. Willems, T. J. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. van Rijssel, B. I. Florea, J. D. C. Codée, G. A. van der Marel, J. M. F. G. Aerts, H. S. Overkleeft, *J. Am. Chem. Soc.* **2014**, *136*, 11622–11625.

[19] L. Wu, J. Jiang, Y. Jin, W. W. Kallemeijn, C.-L. Kuo, M. Artola, W. Dai, C. van Elk, M. van Eijk, G. A. van der Marel, J. D. C. Codée, B. I. Florea, J. M. F. G. Aerts, H. S. Overkleeft, G. J. Davies, *Nat. Chem. Biol.* **2017**, *13*, 867–873.

[20] S. P. Schröder, C. de Boer, N. G. S. McGregor, R. J. Rowland, O. Moroz, E. Blagova, J. Reijngoud, M. Arentshorst, D. Osborn, M. D. Morant, E. Abbate, M. A. Stringer, K. B. R. M. Krogh, L. Raich, C. Rovira, J.-G. Berrin, G. P. van Wezel, A. F. J. Ram, B. I. Florea, G. A. van der Marel, J. D. C. Codée, K. S. Wilson, L. Wu, G. J. Davies, H. S. Overkleeft, *ACS Cent. Sci.* **2019**, *5*, 1067–1078.

[21] N. G. S. McGregor, M. Artola, A. Nin-Hill, D. Linzel, M. Haon, J. Reijngoud, A. Ram, M.-N. Rosso, G. A. van der Marel, J. D. C. Codée, G. P. van Wezel, J.-G. Berrin, C. Rovira, H. S. Overkleeft, G. J. Davies, *J. Am. Chem. Soc.* **2020**, *142*, 4648–4662.

[22] C. de Boer, N. G. S. McGregor, E. Peterse, S. P. Schröder, B. I. Florea, J. Jiang, J. Reijngoud, A. F. J. Ram, G. P. van Wezel, G. A. van der Marel, J. D. C. Codée, H. S. Overkleeft, G. J. Davies, *RSC Chemical Biology* **2020**, *1*, 148–155.

[23] Y. Chen, Z. Armstrong, M. Artola, B. I. Florea, C.-L. Kuo, C. de Boer, M. S. Rasmussen, M. Abou Hachem, G. A. van der Marel, J. D. C. Codée, J. M. F. G. Aerts, G. J. Davies, H. S. Overkleeft, *J. Am. Chem. Soc.* **2021**, *143*, 2423–2432.

[24] N. G. S. McGregor, C. de Boer, Q. P. O. Foucart, T. Beenakker, W. A. Offen, J. D. C. Codée, L. I. Willems, H. S. Overkleeft, G. J. Davies, *ACS Cent. Sci.* **2023**, *9*, 2306–2314.

[25] a) J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589; b) M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, *Nucleic Acids Res.* **2021**, *50*, D439–D444.

[26] J. Agirre, O. Moroz, S. Meier, J. Brask, A. Munch, T. Hoff, C. Andersen, K. S. Wilson, G. J. Davies, *Acta Crystallographica Section D* **2019**, *75*, 1–7.

[27] G. Arnal, D. W. Cockburn, H. Brumer, N. M. Koropatkin, *Protein Sci.* **2018**, *27*, 1093–1101.

[28] a) K.-Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, H. van den Elst, C.-S. Wong, S. D. Chander, S. Hoogendoorn, T. J. M. Beenakker, J. D. C. Codée, J. M. F. G. Aerts, G. A. van der Marel, H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, *2014*, 6030–6043; b) F. G. Hansen, E. Bundgaard, R. Madsen, *J. Org. Chem.* **2005**, *70*, 10139–10142.

[29] P. A. Bartlett, J. D. Meadows, E. G. Brown, A. Morimoto, K. K. Jernstedt, *J. Org. Chem.* **1982**, *47*, 4013–4018.

[30] H. A. Brown, A. L. DeVeaux, B. R. Juliano, A. L. Photenhauer, M. Boulinguiez, R. E. Bornschein, Z. Wawrzak, B. T. Ruotolo, N. Terrapon, N. M. Koropatkin, *Cell. Mol. Life Sci.* **2023**, *80*, 232.

[31] a) D. W. Cockburn, N. I. Orlovsky, M. H. Foley, K. J. Kwiatkowski, C. M. Bahr, M. Maynard, B. Demeler, N. M. Koropatkin, *Mol. Microbiol.* **2015**, *95*, 209–230; b) D. W. Cockburn, F. M. Cerqueira, C. Bahr, N. M. Koropatkin, *Amylase* **2020**, *4*, 24–44.

[32] J. Jiang, C. L. Kuo, L. Wu, C. Franke, W. W. Kallemeijn, B. I. Florea, E. van Meel, G. A. van der Marel, J. D. Codee, R. G. Boot, G. J. Davies, H. S. Overkleeft, J. M. Aerts, *ACS Cent. Sci.* **2016**, *2*, 351–358.

[33] N. M. Koropatkin, T. J. Smith, *Structure* **2010**, *18*, 200–215.

[34] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774–797.

[35] J. N. D'Elia, A. A. Salyers, *J. Bacteriol.* **1996**, *178*, 7180–7186.

[36] M. Wetmore Kelly, N. Price Morgan, J. Waters Robert, S. Lamson Jacob, J. He, A. Hoover Cindi, J. Blow Matthew, J. Bristow, G. Butland, P. Arkin Adam, A. Deutschbauer, *mBio* **2015**, *6*, DOI:10.1128/mbio.00306-00315.

[37] a) H. A. Brown, A. L. Morris, N. A. Pudlo, A. E. Hopkins, E. C. Martens, J. L. Golob, N. M. Koropatkin, *bioRxiv* **2024**, 2024.2005.2020.595031; b) E. A. Cameron, M. M. A. C. J. Smith, T. J. Smith, N. M. Koropatkin, E. C. Martens, *J. Biol. Chem.* **2012**, *287*, 34614–34625.

[38] a) J. Takamine, *J. Soc. Chem. Ind.* **1898**, *17*, 118–120; b) H. C. Sherman, E. C. Kendall, E. D. Clark, *J. Am. Chem. Soc.* **1910**, *32*, 1073–1086.
[39] G. J. Davies, A. M. Brzozowski, Z. Dauter, M. D. Rasmussen, T. V. Borchert, K. S. Wilson, *Acta Crystallographica Section D* **2005**, *61*, 190–193.
[40] N. Saito, *Arch. Biochem. Biophys.* **1973**, *155*, 290–298.
[41] a) H. Hagihara, K. Igarashi, Y. Hayashi, K. Endo, K. Ikawa-Kitayama, K. Ozaki, S. Kawai, S. Ito, *Appl. Environ. Microbiol.* **2001**, *67*, 1744–1750; b) T. Nonaka, M. Fujihashi, A. Kita, H. Hagihara, K. Ozaki, S. Ito, K. Miki, *J. Biol. Chem.* **2003**, *278*, 24818–24824.
[42] G. Tomita, S. S. Kim, *Zeitschrift für Naturforschung B* **1967**, *22*, 294–300.
[43] C. Faber, T. J. Hobley, J. Mollerup, O. R. T. Thomas, S. G. Kaasgaard, *Chem. Eng. Process.* **2008**, *47*, 1007–1017.
[44] A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, A. I. Nesvizhskii, *Nat. Methods* **2017**, *14*, 513–520.
[45] F. Yu, S. E. Haynes, A. I. Nesvizhskii, *Mol. Cell. Proteomics* **2021**, *20*, 100077.
[46] N. A. Pudlo, K. Urs, R. Crawford, A. Pirani, T. Atherly, R. Jimenez, N. Terrapon, B. Henrissat, D. Peterson, C. Ziemer, E. Snitkin, C. Martens Eric, *mSystems* **2022**, *7*, e00947–00921.
[47] a) M. K. Bacic, C. J. Smith, *Current Protocols in Microbiology* **2008**, *9*, 13C.11.11–13C.11.21; b) V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallographica Section D* **2010**, *66*, 12–21; c) J. J. W. Duan, A. B. Smith III, *J. Org. Chem.* **1993**, *58*, 3703–3711; d) P. Emsley, K. Cowtan, *Acta Crystallographica Section D* **2004**, *60*, 2126–2132; e) P. Evans, *Acta Crystallographica Section D* **2006**, *62*, 72–82; f) Y. Hsiao, H. Zhang, G. X. Li, Y. Deng, F. Yu, H. V. Kahrood, J. R. Steele, R. B. Schittenhelm, A. I. Nesvizhskii, *bioRxiv* **2024**, 2024.2003.2005.583643; g) W. Kabsch, *Acta Crystallographica Section D* **2010**, *66*, 125–132; h) N. M. Koropatkin, E. C. Martens, J. I. Gordon, T. J. Smith, *Structure* **2008**, *16*, 1105–1115; i) M. V. Kvach, A. V. Ustinov, I. A. Stepanova, A. D. Malakhov, M. V. Skorobogatyi, V. V. Shmanai, V. A. Korshun, *Eur. J. Org. Chem.* **2008**, *2008*, 2107–2117; j) J. Larsbrink, T. E. Rogers, G. R. Hemsworth, L. S. McKee, A. S. Tauzin, O. Spadiut, S. Klinter, N. A. Pudlo, K. Urs, N. M. Koropatkin, A. L. Creagh, C. A. Haynes, A. G. Kelly, S. N. Cederholm, G. J. Davies, E. C. Martens, H. Brumer, *Nature* **2014**, *506*, 498–502; k) A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin, G. N. Murshudov, *Acta Crystallographica Section D* **2012**, *68*, 431–440; l) F. Long, R. A. Nicholls, P. Emsley, S. Grazulis, A. Merkys, A. Vaitkus, G. N. Murshudov, *Acta Crystallographica Section D* **2017**, *73*, 112–122; m) E. C. Martens, H. C. Chiang, J. I. Gordon, *Cell Host Microbe.* **2008**, *4*, 447–457; n) N. G. S. McGregor, H. S. Overkleeft, G. J. Davies, *Methods in Enzymology Vol. 664* (Ed.: E. E. Carlson), Academic Press, **2022**, pp. 103–134; o) S. McNicholas, E. Potterton, K. S. Wilson, M. E. M. Noble, *Acta Crystallographica Section D* **2011**, *67*, 386–394; p) G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallographica Section D* **1997**, *53*, 240–255; q) M. Shan, E. U. Sharif, G. A. O'Doherty, *Angew. Chem. Int. Ed.* **2010**, *49*, 9492–9495; r) A. J. Simpkin, J. M. H. Thomas, R. M. Keegan, D. J. Rigden, *Acta Crystallographica Section D* **2022**, *78*, 553–559; s) A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022–1025; t) L. Wang, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *J. Am. Chem. Soc.* **2018**, *140*, 4632–4638; u) G. Winter, *J. Appl. Crystallogr.* **2010**, *43*, 186–190.

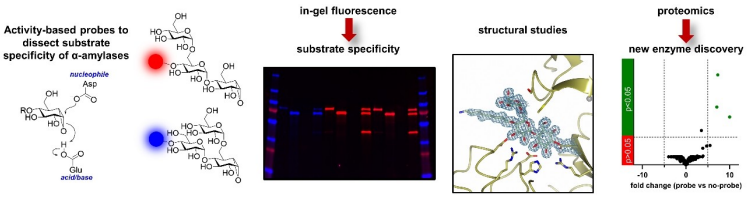
Manuscript received: August 9, 2024
Accepted manuscript online: November 27, 2024
Version of record online: ■■, ■■

Research Article

Activity-Based Probes

I. B. Pickles, Y. Chen, O. Moroz, H. A. Brown, C. de Boer, Z. Armstrong, N. G. S. McGregor, M. Artola, J. D. C. Codée, N. M. Koropatkin, H. S. Overkleeft,* G. J. Davies* e202415219

Precision Activity-Based α-Amylase Probes for Dissection and Annotation of Linear and Branched-Chain Starch-Degrading Enzymes



Activity-based probes were designed as a method to annotate substrate specificity of retaining α-amylases. Fluorescent probes were used in activity-based protein profiling experiments with pure protein, a prominent gut microbe and in a species-rich composter sample. Proteomics experiments were performed to identify, and annotate the substrate specificity of, novel α-amylases with favourable properties for the industrial processing of starch.