**ORIGINAL ARTICLE** 



# A putative novel starch-binding domain revealed by in silico analysis of the N-terminal domain in bacterial amylomaltases from the family GH77

Filip Mareček<sup>1,2</sup> · Marie Sofie Møller<sup>3</sup> · Birte Svensson<sup>3</sup> · Štefan Janeček<sup>1,2</sup>

Received: 1 March 2021 / Accepted: 9 April 2021 / Published online: 21 April 2021 © King Abdulaziz City for Science and Technology 2021

### Abstract

The family GH77 contains 4- $\alpha$ -glucanotransferase acting on  $\alpha$ -1,4-glucans, known as amylomaltase in prokaryotes and disproportionating enzyme in plants. A group of bacterial GH77 members, represented by amylomaltases from *Escherichia coli* and *Corynebacterium glutamicum*, possesses an N-terminal extension that forms a distinct immunoglobulin-like fold domain, of which no function has been identified. Here, in silico analysis of 100 selected sequences of N-terminal domain homologues disclosed several well-conserved residues, among which Tyr108 (*E. coli* amylomaltase numbering) may be involved in  $\alpha$ -glucan binding. These N-terminal domains, therefore, may represent a new type of starch-binding domain and define a new CBM family. This hypothesis is supported by docking of maltooligosaccharides to the N-terminal domain in amylomaltases, representing the four clusters of the phylogenetic tree.

Keywords Amylomaltase  $\cdot$  Family GH77  $\cdot$  N-terminal domain  $\cdot$  Starch-binding domain  $\cdot$  CBM families  $\cdot$  Conserved sequence regions  $\cdot$  Evolutionary relatedness

### Abbreviations

- CBMCarbohydrate-binding moduleβ-CDβ-CyclodextrinCSRConserved sequence regionDPEDisproportionating enzyme
- G2 Maltose
- G3 Maltotriose
- G4 Maltotetraose
- GH Glycoside hydrolase
- PDB Protein Data Bank
- SBD Starch-binding domain

Štefan Janeček Stefan.Janecek@savba.sk

- <sup>1</sup> Department of Biology, Faculty of Natural Sciences, University of SS. Cyril and Methodius, Nám. J. Herdu 2, 91701 Trnava, Slovakia
- <sup>2</sup> Laboratory of Protein Evolution, Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, 84551 Bratislava, Slovakia
- <sup>3</sup> Enzyme and Protein Chemistry, Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 224, 2800 Kgs. Lyngby, Denmark

# Introduction

In the sequence-based classification system of carbohydrateactive enzymes, the CAZy database (https://www.cazy. org/; Lombard et al. 2014), glycoside hydrolase family 77 (GH77) is a monospecific family of  $4-\alpha$ -glucanotransferases (EC 2.4.1.25) (Janecek and Gabrisko 2016). In Archaea and Bacteria, these enzymes are also known as amylomaltases (Terada et al. 1999; Kaper et al. 2005; Godany et al. 2008; Mehboob et al. 2016, 2020), while they are named disproportionating enzymes (DPEs; or just D-enzymes) in plants (Takaha et al. 1993; Wattebled et al. 2003). Currently, family GH77 contains more than 12,000 sequences (Lombard et al. 2014), the vast majority (>11,800) are from bacteria; while the remaining members are distributed between Archaea (~70) and Eucarya, i.e. plants and green algae (~90). In total, so far 20 enzymes of GH77 are characterised experimentally and crystal structures are solved of 6 bacterial amylomaltases and 2 plant DPEs (Lombard et al. 2014).

At the higher hierarchical level, GH77 together with the main  $\alpha$ -amylase family GH13 and family GH70 of circularly permuted glucansucrases constitutes the socalled  $\alpha$ -amylase clan GH-H (Kuriki and Imanaka 1999; MacGregor et al. 2001; van der Maarel et al. 2002; Janecek et al. 2014; Lombard et al. 2014; Janecek and



Gabrisko 2016; Gangoiti et al. 2018). The GH77 members, such as the GH13 and GH70 enzymes, employ a retaining reaction mechanism and adopt the α-amylasetype  $(\beta/\alpha)_{\circ}$ -barrel (TIM-barrel) catalytic domain (Kuriki and Imanaka 1999; Janecek et al. 2014). Similar to TIMbarrels from the  $\alpha$ -amylase family GH13, family GH77 has a protruding domain B inserted between strand  $\beta 3$ and helix  $\alpha 3$  (Matsuura et al. 1984; Vujicic-Zagar et al. 2010). However, the GH77 TIM-barrel contains additional insertions following other  $\beta$ -strands (MacGregor et al. 2001; Janecek and Gabrisko 2016). Overall, the tertiary structure of GH77 is composed of two main domains, the catalytic TIM-barrel and domain B, the latter comprising three auxiliary subdomains known as B1, B2 and B3 (Przylas et al. 2000b). While subdomain B1 corresponds to domain B in GH13 and B3 may play the role of domain C in GH13, the subdomain B2 is unique to GH77 and has no counterpart identified in GH13 and GH70 (Matsuura et al. 1984; MacGregor et al. 2001; Vujicic-Zagar et al. 2010; Janecek and Gabrisko 2016). Notably, the lack of an antiparallel β-sandwich domain C succeeding the catalytic TIM-barrel is the main feature distinguishing GH77 4-α-glucanotransferases from the other clan GH-H members in GH13 and GH70 (Przylas et al. 2000b). The primary structures of  $4-\alpha$ -glucanotransferases of GH77 have 4-7 conserved sequence regions (CSRs), CSR-I-CSR-VII, in common with GH13 and GH70 (Janecek 2002; Janecek and Gabrisko 2016). The catalytic machinery of GH77 enzymes consists of a triad of aspartic acid, glutamic acid and aspartic acid localised at the C-terminal ends of strands  $\beta$ 4 (Asp; catalytic nucleophile),  $\beta$ 5 (Glu; proton donor) and  $\beta7$  (Asp; transition-state stabiliser), respectively. These were first observed for GH77 as Asp293, Glu340 and Asp395 in the crystal structure of amylomaltase from *Thermus aquaticus* (Przylas et al. 2000a, b). Currently, seven additional structures are available in GH77, namely of amylomaltases from Aquifex aeolicus [Protein Data Bank (PDB): 1TZ7], Thermus thermophilus (Barends et al. 2007), Thermus brockianus (Jung et al. 2011), Escherichia coli (Weiss et al. 2015) and Corynebacterium glutamicum (Joo et al. 2016) and two DPEs from Arabidopsis thaliana (O'Neill et al. 2015) and potato (Imamura et al. 2020).

In general, enzymes of GH77 catalyse transfer of a glucan chain from one  $\alpha$ -1,4-glucan to extend another  $\alpha$ -1,4-glucan or produce a cyclic  $\alpha$ -1,4-glucan from a single linear  $\alpha$ -1,4glucan chain (Takaha et al. 1993; Terada et al. 1999; Wattebled et al. 2003; Kaper et al. 2005; Godany et al. 2008; Mehboob et al. 2016). The degree of polymerization (DP) of the latter products, often referred to as cycloamylose, is 17 or higher; thus these cyclic  $\alpha$ -1,4-glucans are much larger than the  $\alpha$ -,  $\beta$ - and  $\gamma$ -cyclodextrins of DP 6, 7 and 8, respectively, produced by cyclodextrin glucanotransferases



of GH13 (Fujii et al. 2005a, b, 2007; Srisimarat et al. 2012; van der Maarel and Leemhuis 2013; Roth et al. 2017; Tumhom et al. 2017).

In addition to catalytic domains of carbohydrate-active enzymes, CAZy classifies carbohydrate-binding modules (CBMs) (Boraston et al. 2004). CBMs are functionally and structurally independent domains without enzymatic activity, which by binding carbohydrates can support the function of catalytic domains. In amylolytic enzymes, these modules are known as starch-binding domains (SBDs) or less frequently as glycogen-binding domains, and in principle, they contribute by binding  $\alpha$ -glucans, for example on starch granules (Janecek et al. 2011). Nowadays, CAZy categorises 88 CBM families (Lombard et al. 2014), among which 15 are considered as SBDs; CBM20, 21, 25, 26, 34, 41, 45, 48, 53, 58, 68, 69, 74, 82, and 83 (Janecek et al. 2019). Members belonging to these individual SBD CBM families, with the exception of family CBM74 (Valk et al. 2016), are approximately 100 amino acid residues long (Janecek et al. 2011, 2019; Carvalho et al. 2015; Armenta et al. 2017).

As it was demonstrated previously, amylomaltases from borreliae may be a rather unique group within family GH77 exhibiting unusual amino acid residues at functionally important positions in individual CSRs (Godany et al. 2008; Kuchtova and Janecek 2015). Otherwise, most of the typical GH77 4-α-glucanotransferases are covered by the prokaryotic Thermus-like amylomaltases (Przylas et al. 2000a, b; Barends et al. 2007; Jung et al. 2011) and the eukaryotic DPE1s (O'Neill et al. 2015; Imamura et al. 2020). These enzymes are usually ~ 500 residues long, however, part of the GH77 family is formed by longer prokaryotic and eukaryotic sequences (Kuchtova and Janecek 2015). Thus, in Eucarya, the DPE2s possess a ~140 residues long insertion between the catalytic nucleophile and proton donor and have two SBDs in tandem of family CBM20 preceding the TIM-barrel domain (Lloyd et al. 2004; Steichen et al. 2008; Kuchtova and Janecek 2015; Janecek et al. 2019). In addition, in Bacteria, a group of amylomaltases differs from the Thermus-like GH77 (Przylas et al. 2000a, b; Barends et al. 2007; Jung et al. 2011) by a ~190 residues long N-terminal extension (Kuchtova and Janecek 2015; Janecek et al. 2019). Currently, this group is best represented by amylomaltases from Escherichia coli (Pugsley and Dubreuil 1988; Weiss et al. 2015) and Corynebacterium glutamicum (Srisimarat et al. 2011; Joo et al. 2016) both with available crystal structures where part of the N-terminal extension adopts an immunoglobulin-like fold similar to those seen in structure-determined SBD families (Janecek et al. 2019). Although especially the C. glutamicum amylomaltase has been extensively studied from the structure/function point of view (Srisimarat et al. 2012; Rachadech et al. 2015; Nimpiboon et al. 2016a, b; Tumhom et al. 2017, 2018), no function has been assigned to the N-terminal domain from

both *E. coli* and *C. glutamicum* GH77 enzymes (Weiss et al. 2015; Joo et al. 2016; Janecek et al. 2019). Recently, the 4- $\alpha$ -glucanotransferase from *Bifidobacterium longum* was characterised (Jeong et al. 2020), which according to its sequence also belongs to the group having the N-terminal extension.

The need to expand fundamental knowledge on ancillary domains in family GH77 and clan GH-H motivates the present bioinformatics investigation of the unusual N-terminal domains, that are candidates of a novel SBD CBM family and seen in crystal structures of GH77 amylomaltases from *E. coli* and *C. glutamicum*. The approach involves search for and retrieval of a relevant group of bacterial amylomaltases containing N-terminal extensions homologous to those found in the *E. coli* and *C. glutamicum* GH77 enzymes. The comparison includes docking trials with a series of  $\alpha$ -1,4-oligoglucosides, to the two determined and the two modelled three-dimensional structures of the four phylogenetically distinguished new putative N-terminal SBDs in GH77.

# **Materials and methods**

### **Sequence collection**

There is a huge number of GH77 members in the CAZy database (currently > 12,000 sequences; Lombard et al. 2014; https://www.cazy.org/), the two GH77 amylomaltases with solved crystal structures from Escherichia coli (Pugsley and Dubreuil 1988; Weiss et al. 2015) and Corynebacterium glutamicum (Srisimarat et al. 2011; Joo et al. 2016), possessing the mutually homologous N-terminal extension, were chosen as the main representatives in the present study. All GH77 sequences exhibiting resemblance with the N-terminal module of these two enzymes were collected from CAZy, because protein BLAST searches (Altschul et al. 1990; https://blast.ncbi.nlm.nih.gov/) using these E. coli and C. glutamicum N-terminal domains as queries failed to provide meaningful results. A preliminary set of 682 sequences was obtained by browsing family GH77 data from CAZy, the main criterion being a sequence length of 650-750 residues. The sequences were aligned by the programme Clustal-Omega available at the European Bioinformatics Institute's server (Sievers et al. 2011; https://www.ebi.ac. uk/Tools/msa/clustalo/) and following an initial alignment, 194 sequences were eliminated for one or more of three reasons: (i) they did not contain a domain homologous to the N-terminal domain of amylomaltases from E. coli and C. glutamicum; (ii) they did not contain the complete catalytic machinery; and (iii) they significantly disrupted the multiple alignment. This resulted in 488 sequences of bacterial GH77 amylomaltases with a convincing N-terminal domain homologous to those found in E. coli and C. glutamicum amylomaltases. One hundred sequences (including *E. coli* and *C. glutamicum* amylomaltases as well as the third experimentally characterised amylomaltase from *B. longum* having the N-terminal extension homologous to that present in the two former enzymes) were finally selected for in-depth analysis (Table S1) ensuring the widest possible taxonomical diversity and a minimum length of ~70 residues of the predicted N-terminal domain.

All amino acid sequences were retrieved from UniProt (UniProt Consortium 2017; https://www.uniprot.org/) and/ or GenBank (Benson et al. 2018; https://www.ncbi.nlm. nih.gov/genbank/) databases. The sequence boundaries for the N-terminal domains in amylomaltases from *E. coli* and *C. glutamicum* were retrieved from their tertiary structures (Weiss et al. 2015; Joo et al. 2016), and those of other fulllength GH77 amylomaltase sequences were defined based on sequence alignment with the *E. coli* and *C. glutamicum* enzymes.

### Sequence comparison and evolutionary analysis

The alignment of the selected 100 sequences of the N-terminal module (Table S1) was performed using the programme Clustal-Omega (Sievers et al. 2011; https://www. ebi.ac.uk/Tools/msa/clustalo/). Only a subtle manual tuning was necessary to maximise similarities considering the best-conserved residues and those potentially involved in carbohydrate binding based on inspection of the two threedimensional structures (Weiss et al. 2015; Joo et al. 2016).

The evolutionary tree was calculated from the final sequence alignment of all 100 N-terminal domains including all gaps as a maximum-likelihood tree using the WAG substitution model (Whelan and Goldman 2001) and the bootstrapping procedure with 500 bootstrap trials (Felsenstein 1985) implemented in the MEGA-X package (Kumar et al. 2018). The tree was displayed with the programme iTOL (Letunic and Bork 2007; https://itol.embl.de/).

Sequence logos of five CSRs defined within the alignment were created using the WebLogo3 online server (Crooks et al. 2004; http://weblogo.threeplusone.com/). Four sequence logos were calculated—one for each of the four clusters identified in the evolutionary tree.

# Comparison of tertiary structures and molecular docking

The coordinates of the GH77 template amylomaltases from *E. coli* (Weiss et al. 2015) and *C. glutamicum* (Joo et al. 2016) were retrieved from the Protein Data Bank (PDB; Berman et al. 2000; https://www.rcsb.org/) under the PDB codes 4S3Q (4S3R) and 5B68, respectively. The structural data were modified to contain only the N-terminal domain of the two enzymes by cutting out the remaining parts of their



structures from the original PDB files based on the literature information (Weiss et al. 2015; Joo et al. 2016).

Reflecting the preliminary evolutionary distribution of all 100 GH77 amylomaltases into four groups (Table S1), structures of the N-terminal domain of two other amylomaltases from *Kushneria marisflavi* (Yun and Bae 2018; UniProt: A0A240US28) and *Pelotomaculum thermopropionicum* (Kosaka et al. 2008; UniProt: A5D1W1) were modelled using the fold recognition Phyre2 server (Kelley and Sternberg 2009; http://www.sbg.bio.ic.ac.uk/~phyre2/). The *K. marisflavi* and *P. thermopropionicum* amylomaltases represent two clusters additional to the two ones containing the experimentally characterised amylomaltases from *E. coli* and *C. glutamicum*.

All molecular docking trials were performed by the program CB-Dock (Liu et al. 2020; http://clab.labshare.cn/ cb-dock/) that utilises the AutoDock Vina (Trott and Olson 2010; http://vina.scripps.edu/) with all parameters used as default. The structures of the N-terminal domains from *E. coli* (Weiss et al. 2015) and *C. glutamicum* (Joo et al. 2016) amylomaltases as well as the structural models from homologous domains in *K. marisflavi* (Yun and Bae 2018) and *P. thermopropionicum* (Kosaka et al. 2008) amylomaltases were docked with maltose (G2), maltotriose (G3), maltotetraose (G4) and  $\beta$ -cyclodextrin ( $\beta$ -CD). Three-dimensional structures of the ligands were retrieved from the PubChem database (Kim et al. 2019; https://pubchem.ncbi.nlm.nih. gov/) and converted into PDB coordinates by the SMILES programme (Weininger et al. 1988; https://cactus.nci.nih. gov/translate/). The resulting complexes of individual structures with bound maltooligosaccharides were displayed using the UCSF Chimera programme (Pettersen et al. 2004).

# **Results and discussion**

Family GH77 has attracted a special interest not only as a member of the  $\alpha$ -amylase clan GH-H (MacGregor et al. 2001; Lombard et al. 2014; Janecek and Gabrisko 2016), but also due to the fact that GH77 amylomaltases from borreliae contain unique substitutions in their amino acid sequence, especially the presence of a lysine two residues before the catalytic nucleophile instead of an arginine otherwise invariant throughout clan GH-H (Godany et al. 2008; Kuchtova and Janecek 2015). This difference was first observed in the amylomaltase from Borrelia burgdorferi (Machovic and Janecek 2003). The present study, however, focuses on a different feature of quite a large group of bacterial GH77 amylomaltases, namely a unique N-terminal extension comprising a separate domain (Fig. 1) that adopts an immunoglobulin-like fold, which is also characteristic for SBD CBM families (Janecek et al. 2019). The first in silico analysis predicting this domain in family GH77 (Kuchtova and Janecek 2015) was later confirmed by the crystal structures of amylomaltases from E. coli (Weiss et al. 2015) and C. glutamicum (Joo et al. 2016), having this N-terminal extension as



**Fig. 1** Domain arrangement of selected amylomaltases of family GH77. The enzymes from *E. coli* (**a**) and *C. glutamicum* (**b**) represent the group of bacterial amylomaltases with N-terminal extensions, a part of which, indicated as the "N-terminal domain", constitutes a presumed CBM, i.e. a novel SBD. Individual domains or subdomains are coloured as follows: domain N1—grey; N-terminal domain (the potential novel SBD CBM family)—red; catalytic TIM-barrel domain



A—blue; subdomains B1, B2 and B3—green, cyan and yellow, respectively. The residues of the catalytic triad of family GH77 (and hence of the entire clan GH-H)—aspartic acid, glutamic acid and aspartic acid, positioned on strands  $\beta4$ ,  $\beta5$  and  $\beta7$ , respectively, of the TIM-barrel, are shown as red diamonds. For comparison, the domain arrangement of a typical bacterial *Thermus*-like GH77 amylomaltase from *Thermus aquaticus* (c) is also shown

opposed to the typical *Thermus*-like amylomaltases (Fig. 1). Notably, the N-terminal extension has two separated parts, the so-called N-terminal domain adopting an immunoglobulin-like fold and considered a potential novel SBD, which immediately precedes the catalytic TIM-barrel, and also a smaller domain N1 situated at the very N-terminus of the protein (Fig. 1).

#### Sequence analysis and evolutionary relationships

The amino acid sequence alignment of the N-terminal domains of the 100 collected GH77 amylomaltases (Fig. 2), despite the obvious overall homology, also shows two larger distinguishable groups (Table S1): (i) sequences No. 1–49 represented by the amylomaltase from *E. coli* (Weiss et al. 2015); and (ii) sequences No. 50–100 represented by the *C. glutamicum* enzyme (Joo et al. 2016). These groups clearly differ by the length of their N-terminal domain, being 70–80 residues for the *E. coli*-like and ~100 residues for the *C. glutamicum*-like group (Fig. 2).

The evolutionary tree (Fig. 3) shows that the selected 100 sequences (Fig. 2), in fact, segregate into four clusters: (i) sequences No. 1–17 (blue group in the tree); (ii) 18–49 (red); (iii) 50-64 (magenta); and (iv) 65-100 (green). In addition to the two best characterised amylomaltases, i.e. from E. coli (No. 49; Weiss et al. 2015) and C. glutamicum (No. 80; Joo et al. 2016)—representing sequences No. 18–49 (red group; covering mostly Gammaproteobacteria) and No. 65-100 (green group; mostly Actinobacteria), respectively, the hypothetical amylomaltases from K. marisflavi (No. 8; Jun and Bae 2018; UniProt: A0A240US28) and P. thermopro*pionicum* (No. 64; Kosaka et al. 2008; UniProt: A5D1W1) were chosen from the two remaining groups of sequences No. 1-17 (blue; covering mostly Proteobacteria) and No. 50-64 (magenta; mostly Firmicutes and Alphaproteobacteria). It is of note, that the third experimentally characterised amylomaltase from B. longum having the N-terminal domain (No. 69; Jeong et al. 2020) is positioned in the cluster represented by C. glutamicum amylomaltase, thus reflecting its actinobacterial origin (Fig. 3). In conclusion, each of the two large groups recognised in the multiple alignment (Fig. 2) is formed by two evolutionarily independent clusters. This division unambiguously observed in the evolutionary tree (Fig. 3) is kept throughout the present study (Table S1).

Aromatic residues, phenylalanine, tyrosine and/or tryptophan are usually responsible for binding  $\alpha$ -glucans in SBDs classified in various CBM families (Janecek et al. 2011, 2019). Therefore, a thorough inspection of the 100 aligned N-terminal domains was performed to identify the best-conserved residues with special attention to conserved aromatic residues. Indeed, aromatic residues are conserved at several positions in the four identified clusters, e.g. Tyr76 and Trp78 in the group represented by the amylomaltase from E. coli (Fig. 2 shows these as Tyr25 and Trp27 in the E. coli sequence; No. 49) and Phe133 in the group represented by the C. glutamicum amylomaltase (Fig. 2 shows this as Phe63 in the C. glutamicum sequence; No. 80). These three positions, however, are not conserved in all four groups. Overall, the only widely, albeit not invariantly conserved aromatic position corresponds to Tyr108 of the E. coli amylomaltase (Fig. 2, position Tyr57 in the E. coli sequence; No. 49). Remarkably, this tyrosine is invariant in three of the four groups, i.e. those containing E. coli, K. marisflavi and P. thermopropionicum amylomaltases. In the fourth group containing the C. glutamicum amylomaltase, it is substituted in 18 of 36 cases by tryptophan (Trp143) (Fig. 2, position Trp73 in C. glutamicum; No. 80), although, e.g. the third experimentally characterised of this set B. longum amylomaltase contains a tyrosine in that position (Fig. 2, position Tyr65 in B. longum; No. 69). The Tyr108 in E. coli amylomaltase and corresponding aromatic residues in all the sequences support the hypothesis that the N-terminal domain is involved in  $\alpha$ -glucan binding and defines a new SBD.

It is worth mentioning, however, that this Tyr108 can hardly be related to two functional tyrosines (Tyr54 and Tyr101) from the catalytic domain identified in the *Thermus aquaticus* amylomaltase contributing to the second  $\alpha$ -glucan binding site (Fujii et al. 2007). Moreover, both Tyr54 and Tyr101 do not represent invariantly conserved residues in the family GH77, as demonstrated by a previous in silico study comparing more than 400 sequences of GH77 amylomaltases (Kuchtova and Janecek 2015).

Regarding fully invariant residues, N-terminal domains of bacterial GH77 amylomaltases only have two, Gly107 and Pro128 (*E. coli* amylomaltase numbering) of which the former precedes the potential  $\alpha$ -glucan binding Tyr108, while the latter is at the C-terminal residue of the domain (Fig. 2; positions Gly56 and Pro77, respectively, in *E. coli*; No. 49). Glycine and proline residues are both significant on the so-called consensus sequence of the first known SBD (Svensson et al. 1989; Janecek and Sevcik 1999), belonging to the current family CBM20 (Janecek et al. 2011, 2019; Lombard et al. 2014).

In an effort to focus attention on segments potentially bearing residues involved in α-glucan binding, the five bestconserved short stretches were proposed to constitute conserved sequence regions (CSRs) of the N-terminal domains in bacterial GH77 amylomaltases as follows (Fig. 2; *E. coli* amylomaltase sequence and numbering): (i) CSR-1—55\_ PNVMVYTSG; (ii) CSR-2—66\_MPMVVE; (iii) CSR-3— 80\_LTTE; (iv) CSR-4—105\_PEGYHTLT; and (v) CSR-5— 121\_HCRVIVAP. Identification of CSRs has become typical for catalytic domains to emphasise key residues involved in activity and/or substrate specificity. This is particularly important for large and polyspecific GH families, such as



| 1  | A5G2P6 Acidiphilium cryptum  |
|--|--|
| 2  | A0A126NVF5 Bosea sp PAMC 26642   |
| - 2  | A0A077FE48 Pseudomonas alkylphenolica  |
| 4  | A0A0X1SYS1 Pseudomonas agarici   |
| - 5  | A0A220VS73 Azotobacter chrococcum  |
| 6  | A0A0D5LZU2 Halomonas en K0116  |
|  | ACADUSCINCE Records sp Norre   |
| 6  | * A0A240US28 Kuchnoria marieflavi  |
|  | A0A291PLT2 Acetobacter tropicalis  |
| 10   | A0A223D9M6 Herbashirillum sp meg3  |
| 11   | AOA11.7AHH8 Roseomonas gilardii  |
| 12   | OOKOX6 Cupriavidus pecator   |
| 13   | A9H855 Gluconacetobacter diazotrophicus  |
| 14   | 039N92 Burkholderia lata   |
| 15   | ASHRZ3 Azorbizobium caulinodans  |
| 16   | A0A1D8TBH6 Achromobacter ruhlandii   |
| 17   | A0A063TVJ3 Bordetella bronchisentica   |
| 18   | A0A1X9U5G7 Photobacterium damselae   |
| 19   | A0A090TF48 Moritella viscosa   |
| 20   | B6E030 Alijvibrio salmonicida  |
| 21   | A0A2A4GCI7 Salinivibrio sp YCSC6   |
| 22   | A7JXG0 Vibrio antiguarius  |
| 23   | A0A0A7MCR8 Actinobacillus equuli   |
| 24   | B8F3M3 Haemophilus parasuis  |
| 25   | A0A142FXW7 Aggregatibacter actinomycetemcomitan  |
| 26   | A0A248ZZA5 Mannheimia haemolytica  |
| 27   | C9PPA0 Pasteurella dagmatis  |
| 28   | E0SJU1 Dickeya dadantii  |
| 29   | A0A250BJQ1_Brenneria_goodwinii   |
| 30   | C6DH69_Pectobacterium_carotvorum   |
| 31   | A0A250B1T2_Gibbsiella_quercinecans   |
| 32   | A0A240CD75_Serratia_ficaria  |
| 33   | W0L855_Chania_multitudinisentens   |
| 34   | D8MLH4_Erwinia_billingiae  |
| 35   | H2ITR3_Rahnella_aquatilis  |
| 36   | E1SGW1_Pantoea_vegans  |
| 37   | A1JSF5_Yersinia_enterocolitica   |
| 38   | I2B495_Shimwellia_blattae  |
| 39   | A7ME78_Cronobacter_sakazakii   |
| 40   | B5XTS0_Klebsiella_pneumoniae   |
| 41   | A0A1T5GZA7_Kosakonia_oryzae  |
| 42   | S5NKG9_Salmonella_bongori  |
| 43   | A0A3S5XUG0_Lelliottia_amnigena   |
| 44   | AUAUH3CRR2_Enterobacter_cloacae  |
| 45   | AUA218NPXI_Leciercia_sp_LSNIHI   |
| 46   | A0A291DSB5_Cedecea_neteri  |
| 4/   | AUA223JKX6_Citrobacter_rarmer1   |
| 48   | B203M6_Shigelia_boydii   |
| <u>4</u> 4   |  |
| =0   | * PI59//_Escherichia_coli  |
| 50   | * PIS9// Escherichia_coli<br>AlSX87_Esychromonas_ingrahamii<br>000022PSP2_boromonas_modia  |
| 50<br>51<br>52   | * PI59//_ESCHERICHIa_COLI<br>AlSX87_Psychromonas_ingrahamii<br>A0A023RSB3_Aeromonas_media<br>C4UB87_Tolumonas_avancis  |
| 50<br>51<br>52<br>53   | * P159//_Escherichia_coli<br>A15X87_Psychromonas_ingrahamii<br>A0A023RSB3_Aeromonas_media<br>C4LB87_Tolumonas_auensis<br>A0A0U3A766_Lacimic_cobium_alkaliphilum  |
| 50<br>51<br>52<br>53<br>54   | * Pisy// Escherichia_coli<br>AlSX87_Psychromonas_ingrahamii<br>AOA023R5B3_Aeromonas_media<br>CALB87_Tolumonas quensis<br>AOA003AV26_Lacimicrobium_alkaliphilum<br>OISVDI_Pseudoalteromonas_atlantica   |
| 50<br>51<br>52<br>53<br>54<br>55   | * P159// Escherichia_Coll<br>AISX87 Psychromonas ingrahamii<br>AOA023RSB3 Aeromonas media<br>C4L887 Tolumonas auensis<br>AOA0U3AY26 Lacimicrobium alkaliphilum<br>Q15VD1 Pseudoalteromonas atlantica<br>G40826 Glaciceola pitratireducens  |
| 50<br>51<br>52<br>53<br>54<br>55<br>55   | * P159// ksGnerichia_Coll<br>A15X87 Psychcromonas ingrahamii<br>A0A023R5B3 Aeromonas_media<br>GCH897 Tolumonas_auensis<br>A0A003AY26 Lacimicrobium_alkaliphilum<br>Q15VD1 Pseudoalteromonas_atlantica<br>G4QH24 Glaciecola_nitratireducens<br>P52B84 Alteromonas_nabtaleniyorans   |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57   | * P159// Escherichia_Coli<br>AISX87 Psychomonas ingrahamii<br>AOA023RSB3 Aeromonas media<br>C4LB87 Tolumonas auenais<br>AOA003AYZ6 Lacimicrobium_alkaliphilum<br>Q15VD1 Pseudoalteromonas_atlantica<br>G4QH24 Glaciecola_nitratireducens<br>F5ZBB4 Alteromonas_naphtalenivorans<br>HSTEMS Phaeeogrifilum moliachianum  |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58   | * P1597/ EsGnerichia_Coll<br>A15X87 Psychcromonas ingrahamii<br>A0A023RSB3 Aeromonas_media<br>Cd1887 Tolumonas_auensis<br>A0A003AY26 Lacimicrobium_alkaliphilum<br>Q15VD1 Pseudoalteromonas_atlantica<br>G4QH24 Glaciecola_nitratireducens<br>F5ZBB4_Alteromonas_naphtaleniyoorans<br>H8FKM5_Phaeospirilum_molischianum<br>Q24208 Macmetoanirilum_macmaticum   |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59   | <ul> <li>PIS9//ESGNETICHA_COIL</li> <li>NISX87 Psychcromonas ingrahamii</li> <li>AOA023RSB3 Aeromonas media</li> <li>C4LB87 Tolumonas auensis</li> <li>AOA003RY26 Lacimicrobium alkaliphilum</li> <li>QI5VD1 Pseudoalteromonas atlantica</li> <li>G40R24 Glaciecola nitratireducens</li> <li>F5ZB84 Alteromonas inaphtalenivorans</li> <li>HBFRM3 Theospirillum molischianum</li> <li>Q2W2Q8 Magnetospirillum magneticum</li> <li>AOA00250X3 Skernanella aerolata</li> </ul>   |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60   | <ul> <li>P1997/ JsGnerichia_Coll</li> <li>A15X87 Psychcromonas ingrahamii</li> <li>A0A023R5B3 Aeromonas media</li> <li>Cd1887 Tolumonas autensis</li> <li>A0A003AY26 Lacimicrobium alkaliphilum</li> <li>Q15VD1 Pseudoalteromonas_atlantica</li> <li>G4QH24 Glaciecola_nitratireducens</li> <li>P52B84 Alteromonas inghtalenivorans</li> <li>H9FKM3 Phaeospirillum molischianum</li> <li>Q2W208 Magnetospirillum magneticum</li> <li>A0A025QX3 Skermanella_aerolata</li> <li>G7ZA14 Aeospirillum jioferum</li> </ul>   |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60<br>61   | PIS9//E3Gmetricha_Coll     MISX87 Psychcromonas ingrahamii     ADA023RSB3 Aeromonas media     CdL887 Tolumonas auensis     ADA003RY26 Lacimicrobium alkaliphilum     QISVD1 Pseudoalteromonas atlantica     G4QH24 Glaciecola nitratireducens     F5ZBB4_Alteromonas naphtalenivorans     BHFRM3 Phaeospirillum molischianum     Q2W2QB Magnetospirillum magneticum     ADA0D252QX3 Skermanella aerolata     G7ZA14 Azospirillum lipoferum     KLL90 Themoacetogenum phaeum  |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60<br>61<br>62   | <ul> <li>P1997/IsGnerichia_Coll</li> <li>AISK87 P2xchromonas ingrahamii</li> <li>AOA023RSB3 Aaromonas_media</li> <li>CALB87 Tolumonas autensis</li> <li>AOA003AY26 Lacimicrobium alkaliphilum</li> <li>Q15VD1 P2xeducalteromonas_atlantica</li> <li>G4QH24 Glaciecola_nitratireducens</li> <li>F5ZB84 Alteromonas_nphtalenivorans</li> <li>HBFKM3 Phaeospirillum molischianum</li> <li>Q2W208 Magnetospirillum ingoferum</li> <li>AOA0025QX3 Skermaella_aerolata</li> <li>G7ZA14 Acospirillum jioforum</li> <li>KALM90 Thermoactogenum pheaum</li> <li>KALK90 Thermoactogenum pheaum</li> </ul>  |
| 50<br>51<br>52<br>53<br>54<br>55<br>57<br>57<br>59<br>60<br>61<br>62<br>63   | * P159// EsGnerichia_Coll<br>AISX87 Psychcromonas ingrahamii<br>ADA023RSB3 Aeromonas media<br>GCLB87 Tolumonas auensis<br>ADA003AY26 Lacimicrobium alkaliphilum<br>Q15VD1 Pseudoalteromonas atlantica<br>G4QH24 Glaciecola nitratireducens<br>PSZBB4 Alteromonas naphtalenivorans<br>H8FRM3 Phaeospirillum molischianum<br>Q2W2Q8 Magnetospirillum magneticum<br>ADA0D25QX3 Skermanella Berolata<br>G7Zh14 Azospirillum lipoferum<br>K4LB90 Themacetogenium phaeum<br>R4KH39 Desulfallas gibsoniae<br>Q2RHD9 Mocrella thermacetica   |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60<br>61<br>62<br>63<br>64   | <ul> <li>P1997/LsGnerichla_Coll</li> <li>AlSK97 P2x96/cromonas ingrahamii</li> <li>AOAU23RSB3 Aaromonas_media</li> <li>Cd1887 Tolumonas autensis</li> <li>AOAU3AX26 Lacimicrobium alkaliphilum</li> <li>Q15VD1 Paeudoalteromonas_atlantica</li> <li>G4QH24 Glaciecola nitratireducens</li> <li>P52B84 Alteromonas inghtalenivorans</li> <li>HBFRM3 Phaeospirillum molischianum</li> <li>Q2W208 Magnetospirillum inolischianum</li> <li>Q2W203 Skermaella_aerolata</li> <li>G72A14 Acoopirillum jioferum</li> <li>K4LH30 Thermoacetogenium phaeum</li> <li>K4KN3 Desulfalla_gibsoniae</li> <li>Q2RH09 Moorella thermoacetica</li> <li>ASDIMI Pelotomaculum thermopropionicum</li> </ul>   |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60<br>61<br>62<br>63<br>64<br>65                                   | <ul> <li>P159// EsGnerichia_Coll</li> <li>A15X87 Fsychromonas ingrahamii</li> <li>A0A023RSB3 Aeromonas media</li> <li>Cd1887 Tolumonas auensis</li> <li>A0A003AY26 Lacimicrobium alkaliphilum</li> <li>Q15VD1 Pseudoalteromonas atlantica</li> <li>G4QH24 Glaciecola nitratireducens</li> <li>P52B84 Alteromonas naphtalenivorans</li> <li>H8FKM3 Phaeospirillum molischianum</li> <li>Q2K209 Magnetospirillum magneticum</li> <li>A0A0D25QX3 Skermanella aerolata</li> <li>G72A14 Azospirilum lipoferum</li> <li>R4LB90 Themacactognium phaeum</li> <li>R4KM3 Desulfallas_gibsoniae</li> <li>Q2KHD9 Moorella thermoacetica</li> <li>* A5DIM1 Pelotomaculum_thermopropionicum</li> <li>AA0A078WU76 Arthrobacter saudimassiliensis</li> </ul>   |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60<br>61<br>62<br>63<br>64<br>65<br>66                             | <ul> <li>P1997/LsGnerichla_Coll</li> <li>N15K97 /P2Krbcmonas ingrahamii</li> <li>A0A023R5B3 Aeromonas_media</li> <li>Cd1887 Tolumonas autennis</li> <li>A0A003R726 Lacimicrobium_alkaliphilum</li> <li>Q15VD1 /Paeudoalteromonas_atlantica</li> <li>G4QH24 Glaciecola nitratireducens</li> <li>P52B84 Alteromonas naphtalenivorans</li> <li>H9FKM3 Phaeospirillum_molischianum</li> <li>Q2W2Q8 Magnetospirillum_magneticum</li> <li>A0A0025QX3 Skermanella_aerolata</li> <li>G72A14 Accopirillum_lipoferum</li> <li>K4LH30 Thermoacetogenium_phaeum</li> <li>R4KHN3 Desulfalla_gibsoniae</li> <li>Q2RH09 Moorella_thermoacetica</li> <li>ADSUBTARKA Serinocous_sp_JI79</li> </ul>  |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>60<br>61<br>62<br>63<br>64<br>65<br>66<br>67                             | <ul> <li>P139//sGGPETCHA_COLI</li> <li>A15X87 / Psychcromonas ingrahamii</li> <li>A0A023RSB3 Aeromonas media</li> <li>Cd1887 Tolumonas aubensis</li> <li>A0A03AY26 Lacimicrobium alkaliphilum</li> <li>Q15VD1 Pseudoalteromonas atlantica</li> <li>G4QH24 Glaciecola nitratireducens</li> <li>F5ZB84 Alteromonas naphtalenivorans</li> <li>HBFRM5 Phaeospirillum molischianum</li> <li>Q2W208 Magnetospirillum inpofarum</li> <li>XALD0 Thermacetogenium phasum</li> <li>R4KH3 Desolfalla gibsoniae</li> <li>Q2HHD9 Moorella thermacetica</li> <li>A050/38W26 Arthrobactor saudimassiliensis</li> <li>A0A018WEN5_Microbacter saudimassiliensis</li> <li>A0A1BRUR5_Microbacterium aurum</li> </ul>  |
| 50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60<br>61<br>62<br>63<br>64<br>65<br>66<br>67<br>68                 | <ul> <li>P1997/LsGnerichla_Coll</li> <li>N15K97 P2x96rbcmonas ingrahamii</li> <li>A0A023R5B3 Aaromonas_media</li> <li>Cd1887 7.010000000000000000000000000000000000</li></ul>  |
| 50<br>51<br>52<br>53<br>55<br>55<br>57<br>58<br>50<br>61<br>62<br>63<br>65<br>66<br>768<br>69                                  | <pre>AlsW7/Escherchal_Coll<br/>AlsW7.Psychcromonas ingrahamii<br/>AOA023R5B3 Aeromonas_media<br/>Cdl887.Tolumonas_utensis<br/>AOA003AY26 Lacimicrobium_alkAliphilum<br/>(J5VD1_Pseudoalteromonas_atlantica<br/>G4QR24 Glaciecola_nitratireducens<br/>P5Z8B4 Alteromonas_naphtalenivorans<br/>H8FKM3_Phaeospirilum_molischianum<br/>Q2W2Q8 Magnetospirilum_magneticum<br/>AOA0025QX3_Skermanella_aerolata<br/>G7ZA14_Acoopirilum_lipoferum<br/>K4LL90_Thermoacetogenium_phaeum<br/>R4KHN3_Desulfallaa_gibsoniae<br/>Q2R4D9_Moorella_thermoacetica<br/>+ ADDIMT_Boltomaculum_thermopropionicum<br/>AOA01BNNR4_Serincocccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNR4_Serincoccus_sp_JTR9<br/>AOA1BNNAS_SERINCOCCUS_SP_JTR9<br/>AOA1BNNAS_SERINCOCCUS_SP_JTR9<br/>AOA1BNNAS_SERINCOCUS_SP_JTR9<br/>AOA1BNNAS_SERINCOCUS_SP_JTR9<br/>AOA1BNNAS_SERINCOCUS_SP_JTR9<br/>AOA1BNNAS_SERINCOCUS_SP_JTR9<br/>AOA1BNNAS_SERINCOCUS_SP_JTR9<br/>AOA1BNNAS_SERINCOCUS_SP_JTR9<br/>AOA1BNA</pre> |
| 50<br>51<br>52<br>53<br>55<br>55<br>55<br>57<br>58<br>50<br>61<br>62<br>63<br>64<br>65<br>66<br>67<br>68<br>97                 | <ul> <li>P1997/LsGnerichla_Coll</li> <li>N15K97 /Px9chromonas ingrahamii</li> <li>A0A023R5B3 Aaromonas_media</li> <li>Cd1887 Tolumonas automas</li> <li>Cd2887 Tolumonas</li> <li>Cd2887 Tolumonas</li> <li>Cd2887 Tolumonas</li> <li>Cd1887 Tolumonas</li> <li>Cd2887 Tol</li></ul>   |
| 501 512 512 512 512 512 512 512 512 512 51   | <ul> <li>PIS9//E3GDETCHA_COLI</li> <li>AISK87 Psychcronnas ingrahami</li> <li>AOA023R5B3 Aaromonas media</li> <li>CALB87. Tolumonas autensis</li> <li>AOA003RY26 Lacimicrobium alkaliphilum</li> <li>QISVDI Pseudoalteromonas atlantica</li> <li>GGURT Constraints</li> <li>GGURT Constant second constraints</li> <li>GGURT Constraints</li></ul>   |
| 501 512 512 512 512 512 512 512 512 512 51   | <ul> <li>PIS97/LsGnerichla_Coll</li> <li>NISK7 Psychoromas ingrahamii</li> <li>AOAU23RSB3 Aaromonas_media</li> <li>ACAB87, Tolumonas automas</li> <li>ACAU38726 Lacimicrobium alkaliphilum</li> <li>QISVDI Pseudoalteromonas atlantica</li> <li>Gégülz Glaciecola nitratireducens</li> <li>STESB4 Alteromonas naphtalenivorans</li> <li>BBFRM3 Phaeospirilum molischianum</li> <li>QXW208 Magnetospirilum magneticum</li> <li>AOAU25QX3 Skemmanla aerolata</li> <li>G72A14 Arosopririlum Jipoferum</li> <li>KALM9 Themoacetogenium pheum</li> <li>MAKN3 DeutSalis glatema</li> <li>ABDIW1 Palotomaculum thermopropionicum</li> <li>AOAUBINAMA Serinoocaterium aurum</li> <li>DZRCP2 Gardmerela vaginalis</li> <li>EMMSH3 Bitdobacterium longum</li> <li>AOAHIBINAR Friedmenla la uetola</li> <li>AOAHIBINAR Friedmenla la uetola</li> <li>AOAHIBINAR Friedmenla la uetola</li> <li>AOAHID Tutolaterium longum</li> <li>AOAHID Tutolata singuialis</li> </ul>   |
| 501<br>552<br>553<br>555<br>556<br>557<br>559<br>601<br>623<br>666<br>667<br>689<br>701<br>723<br>74                           | PIS97/IsGnerichla_Coll     AISK87 Psychromonas ingrahamii     AOA023RSB3 Aaromonas_media     CALB87 Tolumonas auensis     AOA003RY26 Lacimicrobium alkaliphilum     QISVD1 Pseudoalteromonas_atlantica     G4QH24 Glaciecola_nitratireducens     FSZB84_Alteromonas_nphtalenivorans     HBFKM3_Phaeospirillum_molischianum     QXVQ08 Magnetospirillum_magneticum     AOA002SQX3 Skermaella_aerolata     G7ZA14 Acospirillum_lipoferum     KALM90 Thermoacetogenium phaeum     KALM90 Thermoacetogenium phaeum     KALM90 Thermoacetogenium phaeum     KALM90 Thermoacetogenium thermopropionicum     AOA002SQX3 Skermaella_aerolata     G7ZA14 Acospirillum_lipoferum     KALM90 Thermoacetogenium thermopropionicum     AOA01BINRAF Serincococcus_sp_JIT99     AOA1BINRAF Serincococcus_sp_JIT9     AOA1BINRAF Serincococcus_sp_JIT9     AOA1BINRAF Priedmanniella_luteola     AOANHIT01 Luteipulveratus_mongoliensis     DOAHIFZK8F Friedmanniella_uteola     AOANHIT04 Dellafaiagidue     ZM208 Magniaes     Suma Serincoccus_sp_JIT9     AOAHIFXK8F Seriedmanniella_luteola     AOAHIFXK8F Seriedmanniella_uteola     AOAHIFXK8F Seriedmanniella     Uteola     AOAHIFXK8F Seriedmanniella     Uteola     AOAUSE Seriedmanniella     Uteola   |
| 501<br>552<br>553<br>555<br>556<br>557<br>559<br>601<br>623<br>666<br>667<br>689<br>701<br>723<br>75<br>773<br>755             | <pre>* PIS9//EsGrencha_Coll<br/>NEX87 Psy/recreting_coll<br/>NEX87 Psy/recreting_coll<br/>ADAU23REB3_Aeromonas_media<br/>C42887.Tolumonas_utensis<br/>ADAU38726_Lacimicrobium_alkaliphilum<br/>G15VD1_Pseudoalteromonas_atlantica<br/>G4QH24_Glaciecola_nitratireducens<br/>F52884_Alteromonas_naphtalenivorans<br/>BBFRM3_Phaeoopirilum_molischianum<br/>Q24XQ8_Magnetospirilum_magneticum<br/>ADAU25QX3_Skermanella_secolata<br/>G72k14_Aroopirilum_lipoferum<br/>KALM9_Thermoacetogenium_pheeum<br/>AKHN3_DeutSalsa_glasoniae<br/>Q2RUB9_Moorella_thermoacetica<br/>* ASDWR_Palotomaculum_thermopropionicum<br/>ADAU25KY3_Strimerium_hermopropionicum<br/>ADAU381NAM4_Serinoocotar_sautimassilensis<br/>ADAUB81Serinoocotar_sgi_JTF9<br/>ADAIB80815_Strimocotares_gi_JTF9<br/>ADAIB80815_Strimocotaresim_arum<br/>D2RCF2_GarGeneella_vaginalis<br/>ZBMS81_Sit4Obacterium_longum<br/>ADAU381NAM4_Fieldmantila_luteola<br/>ADAU38_INAM4_Strimocotaresim_anomes<br/>SAU32_Strimocotaresim_anomes<br/>SAU32_Strimocotaresim_anomes</pre>  |
| 551<br>553<br>555<br>555<br>555<br>555<br>555<br>555<br>555<br>555   | <ul> <li>PIS97/IsGnerichia_Coll</li> <li>AISK87 Pzycheronnas ingrahamii</li> <li>AOAU23FSB3 Aaromonas_media</li> <li>CALB87 Tolumonas auenis</li> <li>AOAU3AYZ6 Lacimicrobium alkaliphilum</li> <li>QISVD1 Pzeudoalteromonas_atlantica</li> <li>G4QH24 Glaciecola nitratireducens</li> <li>PSZB84 Alteromonas inphtalenivorans</li> <li>HBFRM3 Phaeospirillum molischianum</li> <li>QZW208 Magnetospirillum inojestimum</li> <li>QZW208 Magnetospirillum joforum</li> <li>KALH30 Thermoacetogenium phesum</li> <li>KALH30 Thermoacetogenium thermopropionicum</li> <li>AOAD181NARK Serincocces sp JLT9</li> <li>AOALBINARK Serincoccus gor JLT9</li> <li>AOALBINARK Serincoccus gor JLT9</li> <li>AOALBINARK Serincoccus gor JLT9</li> <li>AOALBINARK Friedmannila luteola</li> <li>AOANLT01 Luteipulveratus mongoliensis</li> <li>DZPKR5 Kinsella film auimas</li> <li>DZPKR5 Kinsella film auimas</li> <li>DAD4898 Projoniahacterium acmes</li> <li>CRX94 Nakamuralla multipartita</li> <li>DA0AB204 Nakamuralla multipartita</li> </ul>   |
| 501<br>551<br>553<br>555<br>556<br>559<br>601<br>663<br>666<br>666<br>670<br>772<br>775<br>777<br>777<br>777777777777777777777 | <ul> <li>PIS9//E3GDETCHA_COLI</li> <li>AISK87 Psychcronnas_ingtahami</li> <li>AOAU23R5B3 Aaromonas_media</li> <li>CAL867, Tolumonas_uenia</li> <li>AOAU23R5B3 Aaromonas_media</li> <li>CAL867, Tolumonas_uenia</li> <li>AOAU23R256 Lacimicrobium_alkaliphilum</li> <li>QISVDI Pseudosalteromonas_atlantica</li> <li>G4QH24 Glacimicrobium_alkaliphilum</li> <li>QISVDI Pseudosalteromonas_atlantica</li> <li>G4QH24 Glacimicrobium_alkaliphilum</li> <li>QIXU25QX3 Glacimicrobium_alkaliphilum</li> <li>QIXU25QX3 Skermanella_merolata</li> <li>G72H14 AAropirilum_lipoferum</li> <li>MADU25QX3 Skermanella_merolata</li> <li>G72H14 AAropirilum_lipoferum</li> <li>MADU30 Thermosostogenium_phasum</li> <li>MADU30 Thermosotoms_phasum</li> <li>MADU30 Thermosotoms_phasum</li> <li>MADU30 Thermosotoms_phasum</li> <li>MADU40 Thermosotoms_phasum</li> <li>MADU40 Thermosotoms_phasum</li> <li>MADU30 Thermosotoms_phasum</li></ul>  |
| 551235555555566123455677777777777777777777777777777777777  | <ul> <li>PISY/LsGnerichla_Coll</li> <li>NISK7 Pzychromonas ingrahamii</li> <li>AOAU23FSB3 Aaromonas_media</li> <li>CALB87, Tolumonas auensis</li> <li>AOAU3AYZ6 Lacimicrobium_alkaliphilum</li> <li>QISVD1 Pzeudoalteromonas_atlantica</li> <li>G4QH2 Glaciecola nitratireducens</li> <li>PSZB84 Alteromonas inphtalenivorans</li> <li>HBFRM3 Phaeospirillum molischianum</li> <li>QZW208 Magnetospirillum inpoferum</li> <li>KAKN3 Desnetospirillum jeoferum</li> <li>KAKN3 Desulfalla_gerolata</li> <li>G7ZA14 Acoofirilum inpoferum</li> <li>KAKN3 Desulfalla_gerolata</li> <li>G7ZA14 Acoofirilum inpoferum</li> <li>KAKN3 Desulfalla_gerolata</li> <li>G7ZA14 Acoofirilum intermorpropionicum</li> <li>AOAD250X3 Skermaella_serolata</li> <li>G7ZA14 Acoofirilum intermorpropionicum</li> <li>AOANBUNGA Scrincocces sp JLT9</li> <li>AOALBURAKA Scrincoccus sp JLT9</li> <li>AOALBURAKA Scrincoccus sp JLT9</li> <li>AOALBURAKA Friedmenila luteola</li> <li>AOAUALTICAE Friedmannila luteola</li> <li>AOAUALTICAE Friedmannila</li> <li>DEYRCK Conforella yaquinalis</li> <li>EMKGE Friedmannila luteola</li> <li>AOAUALTICAE Jurda Acousticae</li> <li>CHS94 Nakamurella multipartita</li> <li>AOAD4ZUM3 Mycobacterium aurum</li> <li>FSX094 Nicrolumatus phosphovorus</li> <li>AOAUALT/CC Corvenced cortum aurum</li> <li>FSX094 Microlumatus phosphovorus</li> <li>AOAULT/CC Corvenced cortum aurum</li> </ul>  |
| 551255555555566123455677777777777777777777777777777777777  | <ul> <li>PIS9//E3GDETCHA_COLI</li> <li>AISK87 Psychcronnas_ingtahami</li> <li>AOAU23FSB3 Aaromonas_media</li> <li>CALS87, Tolumonas_uenia</li> <li>AOAU23FSB3 Aaromonas_media</li> <li>CALS87, Tolumonas_uenia</li> <li>AOAU23F26 Lacimicrobium_alkaliphilum</li> <li>GISVDI Pseudoalteromonas_atlantica</li> <li>G4QH24 Glaciecola_nitratireducens</li> <li>F2SB4, Alteromonas_nphtalenivorans</li> <li>BSTRM5 Phaeoepirilum_moniscitum</li> <li>AOAU25CM3 Skemanella_merolata</li> <li>CPUL0_ANGENEDATILUM_Ipoferum</li> <li>KATHUN_Desnifallangeniberum</li> <li>CARD0 Skemanella_merolata</li> <li>CPUL0_ANGENEDIA</li> <li>CPUL0_ANGENEDIA</li> <li>CPUEDD Mococalla_thermapropionicum</li> <li>AOAO1800K9 Astronocces_sp_JLT9</li> <li>AOAH20K9 Friedmanniella_itecola</li> <li>AOAD1800K9 Microbacterium_any</li> <li>D2RCF2 Gardmarella_yaginalis</li> <li>EBMG81 Bifidobacterium_longum</li> <li>AOAH1FXH8 Friedmanniella_itecola</li> <li>AOAU2CXN0 Tessaracoccus_aquimaris</li> <li>DAB98 Propionibacterium_arenes</li> <li>CRSW1 NAKAmurella_multipartita</li> <li>AOAU2CXN0 Divendurestia</li> <li>ADAU2CXN0 Divendurestia</li> <li>ADAU2CX</li></ul>   |
| 551<br>551<br>555<br>555<br>555<br>555<br>555<br>555<br>555<br>555   | <ul> <li>PISM/ Excerning-Coll</li> <li>NISKO 7.PX/chromonas ingrahamii</li> <li>ADAU23RSB3 Aeromonas_media</li> <li>CALBRY Tolumonas autensis</li> <li>AOAU3XYZ6 Lacimicrobium_alkaliphilum</li> <li>GISVD1 7.Paedioalteromonas_atlantica</li> <li>GQUEYA Glaciecola nitratireducens</li> <li>PSZB84 Alteromonas naphtalenivorans</li> <li>HBFRM3 Phaeospirillum molischianum</li> <li>QXVQ28 Magnetospirillum insgneticum</li> <li>AOAU25QX3 Skermaella aerolata</li> <li>GYZA14 Accopirillum [hoferum</li> <li>KALMO Thermoacetogenium pheeum</li> <li>KALMO Thermoacetogenium pheeum</li> <li>KALMO Thermoacetogenium pheeum</li> <li>KALMO Thermoacetogenium pheeum</li> <li>KALMS Scrincoccus ego JLT9</li> <li>AOAIBINARA Scrincoccus ego JLT9</li> <li>AOAIBINARA Friedmanla la tucela</li> <li>AOAUAIM Friedmanla la serolata</li> <li>CAMBRUEY Microbacterium aurum</li> <li>DZPCF2 Gardnerela vaginalis</li> <li>EBMGBI Bifidobacterium Jongum</li> <li>AOAHIFIXER Friedmanniella luteela</li> <li>AOAUAID7 Tutela friedmaris</li> <li>DAJABINARA Myohoacterium aurum</li> <li>DZPTK3 Kribella flavida</li> <li>AOAUAIZONI Fessaracoccus aguimaris</li> <li>D4H988 Frojonialacterium acnes</li> <li>CEXYM Nakamurella multipartita</li> <li>AOAU220NIG Tessaracoccus apuimaris</li> <li>DAH988 Myoobacterium aurum</li> <li>FSXQP Microlunatus phoephovorus</li> <li>AOAUAIZONZ Corynebacterium aurum</li> <li>FSXQP Microlunatus phoephovorus</li> <li>AOAUAIZONZ Corynebacterium aurum</li> <li>FSXQP Microlunatus phoephovorus</li> <li>AOAUAIZONZ Corynebacterium dutamicum</li> </ul>   |
| 55123555555556662345566789012234556778981  | <ul> <li>PIS9//ESGDETCHA_COLL</li> <li>PISK07//ESGDETCHA_COLL</li> <li>NEKK07/PRYCheronas_ingtahamii</li> <li>ADAU25RES3_Acromonas_media</li> <li>CAL807,Tolumonas_uenia</li> <li>ADAU25RES3_Acromonas_media</li> <li>CAL807,Tolumonas_uenia</li> <li>ADAU25R256_Lacimicrobium_alkaliphilum</li> <li>GISVDI_Paeudoalteromonas_atlantica</li> <li>GGURD2_Paeudoalteromonas_atlantica</li> <li>GGURD2_FACTOREDATE</li> <li>GRACE_GENERATINE</li> <li>GRACE_GENERATINE</li> <li>GRACE_GENERATINE</li> <li>ADAU2520X3_Skermanella_merolaticum</li> <li>ADAU320X10_Microbacterium_atrum</li> <li>D2RCF2_Gardmanella_yaginalis</li> <li>EBMG81_Bifidobactarium_longum</li> <li>ADAU210X10_Messancoccus_aquimaris</li> <li>DABU884_Frojionibacterium_anenes</li> <li>CEXSW1_NakAmurella_multipartita</li> <li>ADAU26X10_Dingenbovorus</li> <li>ADAU56X05_Brevibacterium_aurimucoscum</li> <li>* QRNN7_Corynebacterium_aurimucoscum</li> <li>* QRNN7_Corynebacterium_aurimucoscum</li> </ul>   |
| 55125555555566666666755555555555555556666666   | <ul> <li>PISM/LsGnerichla_Coll</li> <li>NISK07 Psychromonas ingrahamii</li> <li>AOAU23RSB3 Aeromonas_media</li> <li>CALB87 Tolumonas autensis</li> <li>AOAU3X726 Lacimicrobium_alkaliphilum</li> <li>QISVD1 Pseudoalteromonas_atlantica</li> <li>GAURT Classification and the second and the second and and the second and and and and and and and and and a</li></ul>  |
| 55123355555556666666666777777777888283   | <ul> <li>PIS97/IsGGNETCHA_COLL</li> <li>PISK7/IsGGNETCHA_COLL</li> <li>NEKK7/PRYChromonas_ingtahamii</li> <li>ADAU25R5B3_Acromonas_media</li> <li>CALBK7_FOLMENDAS_uenais</li> <li>ADAU25R5B3_Acromonas_media</li> <li>CALBK7_FOLMENDAS_uenais</li> <li>ADAU25R256_Lacimicrobium_alkaliphilum</li> <li>GISVDI_Paeudoslteromonas_atlantica</li> <li>GGURD2_Paeudoslteromonas_atlantica</li> <li>GGURD2_FOLMENDAS_UENDAS</li> <li>GGURD2_FOLMENDAS_UENDAS</li> <li>GGURD2_FOLMENDAS</li> <li>GGURD2_FOLMENDA</li></ul>   |
| 551233555555566623455667890123777777888128388  | <ul> <li>PISM/LsGnerichla_coll</li> <li>NISKO 7 Psychromonas ingrahamii</li> <li>AOAU23RSB3 Aeromonas_media</li> <li>CALBST 7.01400000000000000000000000000000000000</li></ul>   |
| 5512345555555666266666677777777777888888888888   | <ul> <li>PIS9//ESGENTICHA_COLL</li> <li>PISK87/PSychromonas_ingtahamii</li> <li>ADAU23RSB3_Acromonas_media</li> <li>CAL887, Tolumonas_uenais</li> <li>ADAU23RSB3_Acromonas_media</li> <li>CAL887, Tolumonas_uenais</li> <li>ADAU23R25_Lacimicrobium_alkaliphilum</li> <li>GISVDI_Pseudosalteromonas_atlantica</li> <li>GGURD1_Pseudosalteromonas_atlantica</li> <li>GGURD2_Manuella_mortalica</li> <li>GGURD2_Manuella_mortalica</li> <li>GGURD2_Manuella_mortalica</li> <li>CARD0_SAUCA</li> <li>CAND0_SAUCA</li> <li>CAND0_CAUCA</li> <li>CAND0_SA</li></ul>   |
| 55123555556666666666777777777888888888888888   | <ul> <li>PISM/LsGnerichla_coll</li> <li>NISKO 7 Psychromonas ingrahamii</li> <li>AOAU23RSB3 Aeromonas_media</li> <li>CALBST, Olimonas autensis</li> <li>AOAU3XX26 Lacimicrobium_alkaliphilum</li> <li>GUSVD1 Pseudoalteromonas_atlantica</li> <li>GUSVD1 Pseudoalteromonas_atlantica</li> <li>GUSVD1 Pseudoalteromonas_atlantica</li> <li>GUZV28 Magnetospirilum molischianum</li> <li>QZW208 Magnetospirilum indirectum</li> <li>AOAU25QX3 Skermanella_aerolata</li> <li>GZAH14 Arcopirilum [hoferum</li> <li>KKHN3 Desulfalla gerolata</li> <li>GZAH4 Arcopirilum [hoferum</li> <li>KALHO9 Thermoacetogenium pheum</li> <li>KKHN3 Desulfalla gerolata</li> <li>GZAH4 Arcopirilum [hoferum</li> <li>KALHO9 Thermoacetogenium pheum</li> <li>RAKHN3 Desulfalla gerolata</li> <li>GZAH4 Arcobacter_saudimassiliensis</li> <li>AOADIBINARA Scrinococcus ego JLT9</li> <li>AOAHBYRKB Friedmanniella luteola</li> <li>AOAHBYRKB Friedmanniella guteola</li> <li>AOAHBYRKB Friedmanniella guteola</li> <li>AOAHBYRKB Friedmanniella guteola</li> <li>AOAHBYRKB Visolum avium</li> <li>FSXON9 Microlunatus phosphovorus</li> <li>AOAHSVG Corynebacterium aurim</li> <li>FSXON9 Microlunatus phosphovorus</li> <li>AOAHSVG Corynebacterium qutanticum</li> <li>AOAM20XHT5 Flavifeurum avium</li> <li>FSXON9 Microlunatus phosphovorus</li> <li>AOAHSVG Corynebacterium flavum</li> <li>AOAG30KH75 Flavifeurus aslsibiostraticola</li> <li>AOAG30KH75 Flavifeurus aslsibiostraticola</li> <li>AOAG30KH75 Flavifeurus aslsibiostraticola</li> <li>AOAG20KH55 Flavifeurus aslsibiostraticola</li> <li>AOAG20KH55 Flavifeurus aslsibiostraticola</li> <li>AOAS0KH75 Flavifeurus aslsibiostraticola</li> <li>AOAS0KH75 Flavifeurus aslsibiostraticola</li> <li>AOAS0KH75 Flavifeurus aslsibiostraticola</li> <li>AOAS0KH75 Revococcus raidotolerans</li> </ul>   |
| 55123555556666666666777777777788888888888888   | <ul> <li>PIS9//E3GENTICHA_COLI</li> <li>PISW//E3GENTICHA_COLI</li> <li>NEX87 Psychromonas ingrahamii</li> <li>ADAU23REB3 Aaromonas media</li> <li>CALB87, Tolumonas autamia</li> <li>ADAU23REB3 Aaromonas media</li> <li>CALB87, Tolumonas autamia</li> <li>CARCE Autamia</li> <li>CARCE Autamia</li> <li>CALB87, Tolumonas autamia</li></ul>   |
| 555555555666666666667777777778888888888  | <ul> <li>PISM/LsGnerichla_coll</li> <li>NISKO 7 Psychromonas ingrahamii</li> <li>AOAU23RSB3 Aeromonas_media</li> <li>CALBST, Olimonas autensis</li> <li>AOAU3XX26 Lacimicrobium_alkaliphilum</li> <li>QISVD1 Pseudoalteromonas_atlantica</li> <li>GQUXQ1 Glaciecola_nitratireducens</li> <li>PSZB84 Alteromonas naphtalenivorans</li> <li>BBFRM3 Phaeospirillum_molischianum</li> <li>QXWQ28 Magnetospirillum_molischianum</li> <li>QXWQ28 Magnetospirillum_bolesonia</li> <li>QXMQ20 Signetospirilum_polesonia</li> <li>QZMU20 Thermoacetogenium_pheum</li> <li>RAKIN3 Desulfalla_derolata</li> <li>GZR14 Arcobacter_saudimassiliensis</li> <li>AOAD18INARA Scrinococcus_sp_JIT9</li> <li>AOAIBINARA Friedmanla la etcolata</li> <li>GZR07 Androne Scrinococus_sp_JIT9</li> <li>AOAIBINARA Friedmanlal luteola</li> <li>AOAHIFIXR8 Friedmannilla_luteola</li> <li>AOAG308WT85 Friedmaterium_glutanicum</li> <li>AOAG308WT85 Friedmaterium_glutanicum</li> <li>AOAG308WT85 Friedmaterium_glutanicum</li> <li>AOASOBLT605</li></ul>   |
| 555555555666666666677777777788888888888  | <ul> <li>PIS9//E3GENTCHA_COLI</li> <li>PISW//E3GENTCHA_COLI</li> <li>NEX87 Psychromonas ingrahamii</li> <li>ADAU23REB3 Aaromonas media</li> <li>CALB87, Tolumonas autamia</li> <li>ADAU23REB3 Aaromonas media</li> <li>CALB87, Tolumonas autamia</li> <li>CALB87, Tolumonas autautamia</li> <li>CALB87</li></ul>   |
| 5555555556666666666677777777777888888888   | <ul> <li>PISM/LsGnerichle_Coll</li> <li>NISKG7 Psychromonas ingrahamii</li> <li>AOAU23RSB3 Aaromonas_media</li> <li>CALBGT ACLADATION AND AND AND AND AND AND AND AND AND AN</li></ul>   |
| 5555555555556666666666777777777788888888   | <ul> <li>PIS97/JsGnerichle_Coll</li> <li>AlSK87 Psychcronnas ingrahami</li> <li>AlSK87 Psychcronnas ingrahami</li> <li>ADAU25R53 Aarcmonas media</li> <li>CALSK7 Toumonas under the second secon</li></ul>   |
| 5555555555566666666666777777777788888888   | <ul> <li>PISM/LsGnerichle_Coll</li> <li>NISKG7 Psychromonas ingrahamii</li> <li>AOAU23RSB3 Aaromonas_media</li> <li>CALBST Olimonas autensis</li> <li>AOAU23RSB3 Aaromonas_media</li> <li>CALBST Olimonas autensis</li> <li>AOAU3XX76 Lacimicrobium alkaliphilum</li> <li>QISVD1 Pseudoalteromonas_atlantica</li> <li>GAQU2XX76 Lacimicrobium alkaliphilum</li> <li>QISVD1 Pseudoalteromonas_atlantica</li> <li>GAQU2XX76 Lacimicrobium phase</li> <li>PSEMB4 Alteromonas naphtalenivorans</li> <li>BBFRM3 Phaeospirillum molischianum</li> <li>QZW208 Magnetospirillum indirectum</li> <li>AOAU25QX3 Skermaella_aerolata</li> <li>G72A14 Azofpirillum [hoferum</li> <li>KALL90 Thermoacetozgenium phesum</li> <li>KALN90 Thermoacetozgenium phesum</li> <li>KALN94 Acthrobacter saudimassiliensis</li> <li>AOALBINARA Feriedmannella luteola</li> <li>AOALBINARA Friedmannella luteola</li> <li>AOALHISKR8 Friedmannella luteola</li> <li>AOALHISKR8 Friedmannella luteola</li> <li>AOALHIZCN80 Tessaracoccus aguimaris</li> <li>D4H988 Friedmannella luteola</li> <li>AOALHIZCN80 Tessaracoccus aguimaris</li> <li>D4H988 Friedmannella luteola</li> <li>AOALBIZN87 Tuesecens</li> <li>C3FW4 Corynebacterium flavum</li> <li>AOAL30FW75 Flaviflews asalisiostraticola</li> <li>AOAO250KH55 Flaviflews adiotolerans</li> <li>D1BW7 Sanguibacter indicus</li> <li>C7WC57 Brachybacterium flavus</li> <li>AOALSSNERF Jeneratophilus congololensis</li> </ul>  |
| 5555555555566623465667890123456777777888888888899993   | <ul> <li>PIS97/JsGnerichle_Coll</li> <li>AlSK87 Psychcronnas ingrahami</li> <li>ADAU23RSB3 Aaromonas media</li> <li>CALSK7, Psychcromonas ingrahami</li> <li>ADAU23RSB3 Aaromonas media</li> <li>CALSK7, Toumonas under the second sec</li></ul>   |
| 55555555555666666666667777777777788888888  | <ul> <li>PISM/LsGnerichie_Coll</li> <li>NISKG7 Psychromonas ingrahamii</li> <li>AOAU23RSB3 Aaromonas_media</li> <li>CALBGT ACINEMONAS ingrahamii</li> <li>AOAU23RSB3 Aaromonas_media</li> <li>CALBGT ACINEMONAS ingrahamii</li> <li>AOAU23RX56 Lacimicrobium alkaliphilum</li> <li>QISVD1 Pseudoalteromonas_atlantica</li> <li>GéQUAL Glaciecola nitratireducens</li> <li>PSZB84 Alteromonas naphtalenivorans</li> <li>BBFRM3 Phaeospirillum molischianum</li> <li>QV2RQ8 Magnetospirillum inginetum</li> <li>AOAU25QX3 Skermaella_serolata</li> <li>G7ZA14 Azofpirillum lipoferum</li> <li>KALENO Thermoacetozgenium pheum</li> <li>RKHKN3 Deusifalia girolata</li> <li>QZRL09 Moorella thermoacetica</li> <li>XBDIT Pelotomaculum thermopropionicum</li> <li>AOAU20KWV6 Arthrobacter saudimassiliensis</li> <li>AOABUENY5 Microbacterium aurum</li> <li>DZRCF2 Gardmerella vaginalis</li> <li>EBMG81 Bifidobacterium longum</li> <li>AOAHISHKRA Friedmannila luteola</li> <li>AOAHISKRAF Priedmannila luteola</li> <li>AOAHISKRAF Nyokoaterium aurum</li> <li>DZPKK3 Kribella flavida</li> <li>AOAHISKRAF Joioniabacterium aurum</li> <li>DZPKK3 Kribella flavida</li> <li>AOAHISKRAF Nyokoaterium aurum</li> <li>DZPKK3 Kribella flavida</li> <li>AOAHISKRAF Nyokoaterium aurum</li> <li>DZPKK3 Kribella flavida</li> <li>AOAHISKRAF Jriedmannika</li> <li>AOAU20KN0 Tessaracoccus aguimaris</li> <li>D4H988 Friedmannika</li> <li>AOAHISKRAF Jriedmannika</li> <li>AOAHISKRAF Jriedmannika</li> <li>AOAHISKRAF Jriedmannika</li> <li>AOAHISKRAF Jriedmannika</li> <li>AOAU30KW75 Flavifleuxeusalsihiostraticola</li> <li>AOASQBKRAF5 Jrevibacterium flavum</li> <li>AOASQBKRAF5 Jrevibacterium flavum</li> <li>AOASQBKRAF5 Jrevibacterium flavus</li> <li>AOASQBKRAF5 Jactionyces radicidentis</li> <li>AOAU35WK75 Brachbacterium flavus</li> <li>AOASQBKRAF5 Jacthopscres radicidentis</li> <li>AOAU325WK26 Janibacter indicu</li></ul>   |
| 55555555555566666666667777777777788888888  | <ul> <li>PIS97/JsGnerichla_Coll</li> <li>AlSK87 Psychcronnas ingrahami</li> <li>AlSK87 Psychcronnas ingrahami</li> <li>ADAUSTSB3 Aaromonas media</li> <li>CALSK7 Toumonas under the second secon</li></ul>   |
| 5555555555666666666677777777778888888888   | <ul> <li>PISM/LsGnerichle_Coll</li> <li>NISKG7 PRVAFCHORDAS _ INGrahamii</li> <li>ADAU23REB3 Aaromonas_media</li> <li>CALBB7 Tolumonas quensis</li> <li>AOAU23REB3 Aaromonas_media</li> <li>CALBB7 Tolumonas quensis</li> <li>AOAU23REG4 Lacimicrobium glataliphilum</li> <li>QISVD1 Preudoalteromonas atlantica</li> <li>GAU24 Glaciecola nitratireducens</li> <li>PSEMB4 Alteromonas nghtalenivorans</li> <li>MBFRM3 Phaeospirilum molischianum</li> <li>QVRVQ8 Magnetospirilium jaoferum</li> <li>KALLD0 Thermoacetorgenium pheum</li> <li>KALLD0 Thermoacetorgenium pheum</li> <li>KALMD0 Thermoacetorgenium pheum</li> <li>KALMD1 Thermoacetorgenium</li> <li>ADAIBINRAR Berichanduk</li> <li>ADAIBINRAR Priedmanniella luteola</li> <li>ADAIH TRARB Priedmanniella luteola</li> <li>ADAIH TRARB Priedmanniella guteola</li> <li>ADAID1CNR77 Corynebacterium quitamicum</li> <li>ADAID20XPT Corynebacterium quitamicum</li> <li>ADAID20XPT Silavifacus aslisiostraticola</li> <li>ADAISUMT Autonicum acaisalisiostraticola</li> <li>ADAISUMT Silavitaum autimaticum</li> <li>ADAISUMT Corynebacterium glatumicum</li> <li>ADAISUMT Actinotignum schalli</li> <li>ADAISUMT Silavitaum autimaticum</li> <li>ADAISUMT Silavitaum</li> <li>ADAISUMT Actinotignum schalli</li> <li>ADAISUMT Actinotignum facidoteris</li> <li>ADAISUMT Actinotignum facidoteris</li> <li>ADAISUMT Actinotignum facidistis</li> <li>ADAISUM</li></ul>   |
| 55555555555566666666666777777777788888888  | <ul> <li>PIS97/IsGRETCHIL_COLL</li> <li>PISK87 Psychromonas logial</li> <li>ALSK87 Psychromonas media</li> <li>ALSK87 Psychromonas media</li> <li>ACALSK7 Automonas media</li> <li>ACAUSTR53 Aaromonas media</li> <li>ACAUSTR53 Aaromonas media</li> <li>ACAUSTR55 Lacimicrobium alkaliphilum</li> <li>GISVDI Pseudoalteromonas atlantica</li> <li>GGRMC Asgretospirilum molischianum</li> <li>ACAUSTR55 Phaeospirilum temporopionicum</li> <li>ACAUSTR55 Phaeospirilum temporopionicum</li> <li>ACAUSTR55 Previbacterium anyum</li> <li>ACAUSTR55 Previbacterium atrimucosum</li> <li>ACAUSTR55 Previbacterium flavences</li> <li>CSTMM5 Actinotignum schalii</li> <li>ACAUSTR55 Plaviflexum salistiostraticola</li> <li>ACAUSTR55 Plaviflexum salistiostraticola</li> <li>ACAUSTR555 Plavibacterium flavences</li> <li>ACAUSTR55 Plaviflexum salistiostraticola</li> <li>ACAUSTR55 Plavibacterium flavences</li> <li>ACAUSTR55 Plavibacterium flavences&lt;</li></ul>  |
| 555555555666666666677777777778888888888  | <ul> <li>PISM/LsGnerichla_Coll</li> <li>NISKG7 PRVArbomonas ingrahamii</li> <li>ADAU23REB3 Aaromonas_media</li> <li>ACAU23REB3 Abromonas_media</li> <li>ACAU23REB3 Phaeospirilum_molischianum</li> <li>QXW2G8 Maqmetospirilum_molischianum</li> <li>QXW2G8 Maqmetospirilum_lispferum</li> <li>KALLB0 Thermoacetocogenium_pheum</li> <li>KALLB0 Thermoacetocogenium_pheum</li> <li>ACAU3200 Moorella thermoacetica</li> <li>ASDITA Paiotomaculum thermopropionicum</li> <li>AOA181NRAR Scrinococcus gp_UI79</li> <li>AOA18000 Astrinobacterium_longum</li> <li>AOA181NRAR Friedmanniella luteola</li> <li>AOA181NRAR Friedmanniella luteola</li> <li>AOA18100 Tessaracoccus aguimaris</li> <li>DEPRNS Kribella flavida</li> <li>AOA18200 Magmetosium_acmes</li> <li>CSWW1 Nakamurella multipartita</li> <li>AOA18200 Torynebacterium_salium</li> <li>AOA0320WRE5 Flavifexeus aglishiostraticola</li> <li>AOA320WRE5 Chernatophilus complensis</li> <li>DEWRE3 Flavifexeus aglishiostraticola</li> <li>A</li></ul>   |



**Fig. 2** Amino acid sequence alignment of the N-terminal domains of GH77 amylomaltases. The sequence order from the top corresponds to their appearance in the evolutionary tree (Fig. 3). Four proteins, representing four potential clusters in the tree (distinguished by different colours), are marked by an asterisk, i.e. amylomaltases from *Kushneria marisflavi* (blue), *Escherichia coli* (red), *Pelotomaculum thermopropionicum* (magenta) and *Corynebacterium glutamicum* (green). Each protein is labelled by the UniProt accession number and the name of the organism. The five suggested conserved sequence

regions (CSR-1–CSR-5) are indicated above the alignment. The best-conserved aromatic position corresponding to Tyr108 in amylomaltase from *E. coli* and the two invariant residues (the glycine succeeding the Tyr108 and the proline at the end of the domain) are signified by dollar and hash symbols, respectively, below the alignment. The colour code for the selected residues: W, yellow; F, Y—blue; V, L, I—green; D, E—red; R, K—cyan; H—brown; C—magenta; G, P—black





Fig. 3 Evolutionary tree of the N-terminal domains of GH77 amylomaltases. The tree displays the relatedness of 100 unique non-redundant GH77 sequences (Table S1) containing the N-terminal domain corresponding to an immunoglobulin-like fold observed in crystal structures of E. coli (Pugsley and Dubreuil 1988: Weiss et al. 2015) and C. glutamicum amylomaltases (Srisimarat et al. 2011; Joo et al. 2016). Each protein is labelled by the UniProt accession number and the name of the organism. Four proteins, representing four clusters illustrated on the tree and each distinguished by different colours, are marked by an asterisk, i.e. amylomaltases from Kushneria marisflavi (blue), Escherichia coli (red), Pelotomaculum thermopropionicum (magenta) and Corynebacterium glutamicum (green). The tree is based on the alignment of the N-terminal domains (Fig. 2)



in the individual  $\alpha$ -amylase families GH13, GH57, GH119 and GH126 (Janecek 2002; Blesak and Janecek 2012, 2013; Janecek and Kuchtova 2012; Janecek et al. 2014; Janecek and Gabrisko 2016; Kerenyiova and Janecek 2020a, b). It also makes sense, however, to establish CSRs for putative SBDs. Among all known SBDs, currently classified in 15 CBM families in CAZy (Lombard et al. 2014), residues from the best-conserved regions usually belong to one or in some cases to two binding sites (Janecek et al. 2011, 2019).

In agreement with four clusters being found in the evolutionary tree of N-terminal domains from bacterial GH77 enzymes (Figs. 2 and 3), sequence logos covering 35 positions for the five proposed CSRs were created for each cluster (Fig. 4). All logos contain the invariant Gly107 and Pro128, at CSR-4 position 22 and CSR-5 position 35, respectively (*E. coli* amylomaltase numbering), plus the potentially  $\alpha$ -glucan binding Tyr108 at CSR-4 position 23. The tripeptide GYH in CSR-4 that is invariant in two of the four evolutionary clusters, containing *E. coli* and *P. thermopropionicum* amylomaltases, respectively (Fig. 4b, c), is one of the best-conserved stretches in the N-terminal domain. The other highly conserved region is found at the last five residues, L/I/V-A/I-V/I-A/T-P, at CSR-5 positions 31–35. At the start of the logo, conserved proline residues (CSR-1; positions 1–2) are seen except in the first position in the logo of the group of *P. thermopropionicum* amylomaltase (Fig. 4c). Concerning additional positions of interest, Glu83 (*E. coli* amylomaltase numbering; CSR-3 position 19) is almost invariantly conserved and deserves attention.

Importantly, the sequence logos identify positions suitable to discriminate the four clusters by showing positions illustrating their individual uniqueness. This is a key attribute of CSRs, hence of sequence logos, which are well known and widely utilised in the  $\alpha$ -amylase families mentioned above to distinguish subfamilies and/or enzyme specificities (Janecek 2002; Blesak and Janecek 2012, 2013; Janecek and Kuchtova 2012; Janecek et al. 2014; Janecek and Gabrisko 2016; Kerenyiova and Janecek 2020a, b). Notably, it was described for a single SBD CBM family (Janecek et al. 2019), CBM41 composed of two genuine groups mutually distinguished by a characteristic sequence pattern of three essential aromatic residues, "W-W- $\sim$ 10aa-W" and





Fig. 4 Sequence logos of N-terminal domains from individual clusters of family GH77 bacterial amylomaltases. The four groups (for details, see Table S1) are represented by a *Kushneria marisflavi* (Proteobacteria; ~70 residues long; 17 sequences); b *Escherichia coli* (Gammaproteobacteria; ~80 residues long; 32 sequences); c *Pelotomaculum thermopropionicum* (Firmicutes and Alphaproteobacteria ~100 residues long; 15 sequences); and d *Corynebacterium glutamicum* (Actinobacteria; ~90 residues long; 36 sequences). CSR-1,

"W-W-~30aa-W", implying that the position of the third tryptophan in the pattern is not shared by the two groups (Janecek et al. 2017). Similarly, only the *C. glutamicum* amylomaltase group has an invariantly conserved histidine, His90 in CSR-2 position 15 (Fig. 4d). Another example of interest is the well-conserved cysteine in the *E. coli* amylomaltase group (Cys122; Fig. 4b; CSR-5 position 29).

### Tertiary structure comparison and molecular docking

The amylomaltases having the N-terminal domain are not the only family GH77 members that possess extra sequence compared to the canonical *Thermus*-like amylomaltases of just the catalytic TIM-barrel domain and a few inserted subdomains (Fig. 1). In addition to amylomaltases from borreliae that may have single unique substitutions even in



residues 1–9; CSR-2, residues 10–15; CSR-3, residues 16–19; CSR-4, residues 20–27; CSR-5, residues 28–35. The aromatic residue, marked by the red asterisk, corresponding to Tyr108 and Trp143 in *E. coli* and *C. glutamicum* amylomaltases, respectively, is potentially involved in stacking interactions with  $\alpha$ -glucans. All other residues involved in hydrogen bonding contacts with docked  $\alpha$ -glucans (G2, G3, G4 and  $\beta$ -CD; for details, see Table 1) are indicated by black asterisks

functionally important positions, but still within the otherwise conserved basic domain arrangement (Machovic and Janecek 2003; Godany et al. 2008; Kuchtova and Janecek 2015; Janecek and Gabrisko 2016), in the Eucarya, the DPE2 version exists as a typical GH77 amylomaltase with a ~140 residues insertion between the catalytic nucleophile and proton donor in the TIM-barrel which, moreover, is preceded by two SBDs of CBM20 (Lloyd et al. 2004; Steichen et al. 2008; Kuchtova and Janecek 2015; Janecek et al. 2019). Unfortunately, no three-dimensional structure is available for DPE2 and the structural fold of the ~140-residue insertion is not known, neither is its potential function (Steichen et al. 2008), nor if it is essential for activity (Ruzanski et al. 2013). Remarkably, however, the function of DPE2 with the two CBM20s in Arabidopsis thaliana was effectively retained by replacement with amylomaltase from E. coli (Ruzanski et al. 2013). In that light, it may be relevant that the N-terminal domain in the *E. coli* enzyme (Fig. 1) has an immunoglobulin-like fold typical of SBDs (Janecek et al. 2019), and can be speculated to act as an SBD.

The two crystal structures from E. coli (Weiss et al. 2015) and C. glutamicum (Joo et al. 2016) describe amylomaltases with the investigated N-terminal domain (Fig. 1). In neither case was the structure of the N-terminal domain obtained in complex with an  $\alpha$ -glucan. Since the thorough phylogenetic analysis revealed four clusters (Fig. 3), models for structural comparison were made of N-terminal domains of two hypothetical amylomaltases of K. marisflavi and P. thermopropionicum from the other two clusters. In agreement with the closer evolutionary relatedness observed on the one hand for the E. coli and K. marisflavi groups and on the other for the C. glutamicum and P. thermopropionicum groups (Fig. 3), N-terminal domains of K. marisflavi and P. thermopropionicum were modelled using those of amylomaltases from E. coli (PDB code: 4S3R; Weiss et al. 2015) and C. glutamicum (PDB: 5B68; Joo et al. 2016), respectively, as templates. Notably, none of the established SBD CBM families (Janecek et al. 2019) with available three-dimensional structures were identified as a suitable template.

To get an idea of how an  $\alpha$ -glucan binds if the N-terminal domain acted as an SBD, maltose (G2), maltotriose (G3), maltotetraose (G4) and  $\beta$ -cyclodextrin ( $\beta$ -CD) were docked. Since the SBDs from various CBM families have already been demonstrated to retain their binding abilities even if being separated from their catalytic domains, i.e. they can preserve the binding also in an isolated form (for a review, see Janecek et al. 2019), the docking was in each case performed with extracted N-terminal domains. Although it may not reflect the situation in real amylomaltases completely, for the purpose of the present study it has been considered sufficient.

Not to bias the results, docking by CB-Dock does not require to focus the ligand on a target place in the protein (Liu et al. 2020), i.e. on the presumed aromatic binding residue Tyr108 in amylomaltase from E. coli and its counterparts, Trp143, Tyr121 and Tyr146 from C. glutamicum, K. marisflavi and P. thermopropionicum, respectively. Importantly, using all four maltooligosaccharides, there should be a single  $\alpha$ -glucan-binding site in every studied N-terminal domain corresponding to each other (cf. Figures 5and 6). Nevertheless, it seems that at least Tyr121 and Tyr146 from K. marisflavi and P. thermopropionicum amylomaltases have not been recognised as making direct hydrogen bond contacts with all four ligands-Tyr121 may be involved with G2, G4 and  $\beta$ -CD, whereas Tyr 146 appears to be involved just with G2 (Table 1). Note, Table 1 summarises only the residues involved in hydrogen bond contacts. Importantly, the aromatic residues-Tyr108, Trp143, Tyr121 and Tyr146-may interact with respective ligands also via stacking interactions (cf. Figs. 5 and 6). Results of all individual



**Fig. 5** Visualisation of molecular docking experiments with real structures. The N-terminal domain of family GH77 amylomaltases from **a**, **b** *Escherichia coli* (PDB: 4S3Q; Weiss et al. 2015) and **c**, **d** *Corynebacterium glutamicum* (PDB: 5B68; Joo et al. 2016) with docked  $\beta$ -cyclodextrin (**a**, **c**) and maltose (**b**, **d**). Side-chains of residues potentially involved in carbohydrate binding are displayed (in black) and labelled accordingly (for details, see Table 1). The best-conserved aromatic residue, i.e. Tyr108 of the *E. coli* amylomaltase and Trp143 of the *C. glutamicum* counterpart (cf. Figure 2) and possibly involved in stacking interactions with glucose moieties, is specifically colour-highlighted

docking trials are summarised in Table 1. Notably, these complexes indicated an appropriate binding energy ranging from the best of -5.9 kJ/mol for maltotriose bound to the N-terminal domain from *K. marisflavi* amylomaltase to -3.7 kJ/mol for the maltose complex of the *P. thermopropionicum* N-terminal domain (Table 1). Interestingly, in addition to the aromatic position represented by Tyr108 (*E. coli* amylomaltase numbering) that might interact by stacking onto glucose moieties, a few other residues were involved in hydrogen bond formation with docked  $\alpha$ -glucans (Table 1), corresponding, e.g. to the positions of Glu83 (CSR-3), Thr110 (CSR-4), Thr112 (CSR-4) and Arg123 (CSR-5) in *E. coli* amylomaltase (Fig. 4).

There are two positions of interest—Gly107 and Pro128 (*E. coli* amylomaltase numbering), which are conserved invariantly (Fig. 2). However, with regard to their eventual involvement in the  $\alpha$ -glucan binding, neither the





**Fig. 6** Visualisation of molecular docking experiments with modelled structures. The models of the N-terminal domain of family GH77 amylomaltases from: **a**, **b** *Kushneria marisflavi* (UniProt: A0A24OU528; template: 4S3R; Weiss et al. 2015) and **c**, **d** *Pelotomaculum thermopropionicum* (UniProt: A5D1W1; template: 5B68; Joo et al. 2016) with docked  $\beta$ -cyclodextrin (**a**, **c**) and maltose (**b**, **d**). Side-chains of residues potentially involved in carbohydrate binding are displayed (in black) and labelled accordingly (for details, see Table 1). The best-conserved aromatic residue, i.e. Tyr121 of the *K. marisflavi* amylomaltase and Tyr146 of the *P. thermopropionicum* counterpart (cf. Figure 2) and possibly involved in stacking interactions with glucose moieties, is also highlighted, regardless of whether it is involved or not in hydrogen bond contacts during the docking trials (Table 1)

glycine, nor the proline here corresponding to Gly107 and Pro128, respectively, has been found as involved in binding  $\alpha$ -glucans in the docking trials (Fig. 4 and Table 1). As far as other glycine and proline residues are concerned, there are a few of them that might be interacting with the tested  $\alpha$ -glucans (Table 1). Although some of these residues are positioned within the identified CSRs (cf. Figs. 2 and 4) and some of them are located outside these regions, obviously, there seems to be no tendency in preserving their potential binding function.

Each N-terminal domain thus exhibits potential to act as an SBD with at least one starch-binding site as seen to be the case in the currently established SBD CBM families (Janecek et al. 2019). Admittedly, docking of a given α-glucan resulted in slightly different arrangements of binding residues within the same indicated single binding site (Table 1). It should be noted that to see different residues in a binding site of a potential SBD (a CBM in general) may not be too surprising since in an SBD (CBM) there typically should be one or two main binding residues; the overall binding being then helped by different surrounding residues depending on a given particular case (Penninga et al. 1996; Sorimachi et al. 1997; Janecek et al. 2017, 2019). In the present case of the N-terminal domain of GH77 amylomaltases, this residue is suggested to be the Tyr108 of E. coli amylomaltase (and its counterparts in homologous amylomaltases). What is, however, most important is the fact that each of the four different maltooligosaccharides (G2, G3, G4 and  $\beta$ -CD) was docked in each N-terminal domain (4 different bacterial origins) within the mutually corresponding single potential binding site (cf. Figs. 5 and 6; Table 1).

### Conclusions

The present study provides an in silico analysis of the N-terminal domain from 100 selected bacterial amylomaltases classified in family GH77. This domain is predicted to function as a type of SBD that would define a novel CBM family. From the evolutionary point of view, these GH77 amylomaltases are divided into four clusters, roughly reflecting bacterial phyla and classes as follows: (i) Gammaproteobacteria; (ii) Proteobacteria; (iii) Firmicutes and Alphaproteobacteria; and (iv) Actinobacteria, illustrated by amylomaltases from E. coli, K. marisflavi, P. thermopropionicum and C. glutamicum, respectively. The conserved Tyr108 of E. coli amylomaltase and its counterparts throughout the four phylogenetic clusters are proposed as the key residue responsible for  $\alpha$ -glucan binding. Based on a careful sequence comparison including definition of CSRs coupled with docking of linear maltooligosaccharides and  $\beta$ -CD, a few additional residues are predicted to belong to the starchbinding site. All candidate residues identified in the present study should be among the first targets for future mutational analysis. The experimental work to confirm the starch-binding role of this N-terminal domain has been initiated.



| N-terminal domain    | G2                             | G3   | G4  | β-CD                                      |
|----------------------|--------------------------------|--|---|---|
| E. coli              | Vina Score: – 4.6              | Vina Score: – 5.1                                      | Vina Score: – 5.8                               | Vina Score: – 5.0                         |
|                      | T81, T82, E83, Y108,<br>T110×2 | T82, E83, E84, Y108,<br>T110×2, R123×2                 | T81×2, E83, Q87,<br>Y108×2, H109, T110,<br>T112 | N56×2, Y108, T110×2,<br>R123×3            |
| C. glutamicum        | Vina Score: - 4.3              | Vina Score: - 4.6                                      | Vina Score: - 5.1                               | Vina Score: - 4.7                         |
|                      | E102×2, L103×2, E104,<br>W143  | E102×2, L103, E104×2,<br>W143, H144×2, K145×3,<br>G158 | E102, L103×2, E104,<br>W143, H144, E156, G158   | W143, H144×2, K145×2,<br>E156×3, G158     |
| K. marisflavi        | Vina Score: - 4.7              | Vina Score: - 5.9                                      | Vina Score: - 4.8                               | Vina Score: - 4.3                         |
|                      | D97, E98, Y121, Y122×2         | P71, I96, D97, Y122×2,<br>R130×3                       | I96, D97, E98×3, Y121,<br>R130×2,               | E98×2, S99, G100×2,<br>Y121, Q123×3, R130 |
| P. thermopropionicum | Vina Score: - 3.7              | Vina Score: - 3.8                                      | Vina Score: - 4.6                               | Vina Score: - 4.2                         |
|                      | L103, E104×3, Y146, Q148       | S78×2, K107×2, G141,<br>L142                           | S78, E104, N105×2, K107,<br>L142, P143, W144×2  | K150×3, Q157, E159                        |

Table 1 Characteristics of docking trials of the N-terminal domain of four selected amylomaltases<sup>a</sup>

<sup>a</sup>The results of docking maltose (G2), maltotriose (G3), maltotetraose (G4) and  $\beta$ -cyclodextrin ( $\beta$ -CD) on the structure of the hypothesised CBM and putative novel SBD of four representative family GH77 amylomaltases. The structures of potential CBMs in amylomaltases from *Escherichia coli* (Weiss et al. 2015; PDB: 4S3Q) and *Corynebacterium glutamicum* (Joo et al. 2016; PDB: 5B68) were extracted from their structural coordinates, whereas the counterparts from *Kushneria marisflavi* (UniProt: A0A240US28) and *Pelotomaculum thermopropionicum* (UniProt: A5D1W1) were obtained by homology modelling using the fold recognition approach performed on the Phyre2 server (Kelley and Sternberg 2009). Vina Score means the binding energy (kJ/mol) for the individual complexes. The residues represent the residues with numbering from the N-terminus of a given amylomaltase involved in the hydrogen bond contact with the respective oligosaccharide; for some of the residues, there are more bond contacts (indicated by "×2" or "×3"). Note, not all of them belong to the five CSRs (Figs. 2 and 4). The complexes with maltose and  $\beta$ -CD are displayed in Fig. 5 for the two experimentally determined structures and in Fig. 6 for two modelled structures

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13205-021-02787-8.

Acknowledgements This work was financially supported by the Grant No. 2/0146/21 from the Slovak Grant Agency VEGA and by Grant No. 6108-00476B from Independent Research Fund Denmark | Natural Sciences (FNU).

Author contributions FM collected data, analysed results, prepared figures and contributed to writing the manuscript; MSM and BS contributed to interpreting results and writing the manuscript; SJ designed the study, contributed to collecting data, analysed and interpreted results, prepared figures and wrote the manuscript. All the authors contributed to discussion of the research and approved the final version of the manuscript.

### Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

### References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi. org/10.1016/S0022-2836(05)80360-2
- Armenta S, Moreno-Mendieta S, Sanchez-Cuapio Z, Sanchez S, Rodriguez-Sanoja R (2017) Advances in molecular engineering of carbohydrate-binding modules. Proteins 85:1602–1617. https:// doi.org/10.1002/prot.25327

- Barends TR, Bultema JB, Kaper T, van der Maarel MJEC, Dijkhuizen L, Dijkstra BW (2007) Three-way stabilization of the covalent intermediate in amylomaltase, an α-amylase-like transglycosy-lase. J Biol Chem 282:17242–17249. https://doi.org/10.1074/jbc. M701444200
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW (2018) GenBank. Nucleic Acids Res 46:D41– D47. https://doi.org/10.1093/nar/gkx1094
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235
- Blesak K, Janecek S (2012) Sequence fingerprints of enzyme specificities from the glycoside hydrolase family GH57. Extremophiles 16:497–506. https://doi.org/10.1007/s00792-012-0449-9
- Blesak K, Janecek S (2013) Two potentially novel amylolytic enzyme specificities in the prokaryotic glycoside hydrolase α-amylase family GH57. Microbiology 159:2584–2593. https://doi.org/10. 1099/mic.0.071084-0
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781. https://doi.org/10.1042/BJ20040892
- Carvalho CC, Phan NN, Chen Y, Reilly PJ (2015) Carbohydratebinding module tribes. Biopolymers 103:203–214. https://doi. org/10.1002/bip.22584
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190. https://doi.org/10.1101/gr.849004
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10. 1111/j.1558-5646.1985.tb00420.x
- Fujii K, Minagawa H, Terada Y, Takaha T, Kuriki T, Shimada J, Kaneko H (2005a) Protein engineering of amylomaltase from *Thermus aquaticus* with random and saturation mutageneses. Biologia 60(Suppl. 16):97–102



- Fujii K, Minagawa H, Terada Y, Takaha T, Kuriki T, Shimada J, Kaneko H (2005b) Use of random and saturation mutageneses to improve the properties of *Thermus aquaticus* amylomaltase for efficient production of cycloamyloses. Appl Environ Microbiol 71:5823–5827. https://doi.org/10.1128/AEM.71.10.5823-5827.2005
- Fujii K, Minagawa H, Terada Y, Takaha T, Kuriki T, Shimada J, Kaneko H (2007) Function of second glucan binding site including tyrosines 54 and 101 in *Thermus aquaticus* amylomaltase. J Biosci Bioeng 103:167–173. https://doi.org/10.1263/jbb.103. 167
- Gangoiti J, Pijning T, Dijkhuizen L (2018) Biotechnological potential of novel glycoside hydrolase family 70 enzymes synthesizing α-glucans from starch and sucrose. Biotechnol Adv 36:196–207. https://doi.org/10.1016/j.biotechadv.2017.11.001
- Godany A, Vidova B, Janecek S (2008) The unique glycoside hydrolase family 77 amylomaltase from *Borrelia burgdorferi* with only catalytic triad conserved. FEMS Microbiol Lett 284:84–91. https://doi.org/10.1111/j.1574-6968.2008.01191.x
- Imamura K, Matsuura T, Nakagawa A, Kitamura S, Kusunoki M, Takaha T, Unno H (2020) Structural analysis and reaction mechanism of the disproportionating enzyme (D-enzyme) from potato. Protein Sci 29:2085–2100. https://doi.org/10.1002/pro. 3932
- Janecek S (2002) How many conserved sequence regions are there in the  $\alpha$ -amylase family? Biologia 57(Suppl. 11):29–41
- Janecek S, Gabrisko M (2016) Remarkable evolutionary relatedness among the enzymes and proteins from the α-amylase family. Cell Mol Life Sci 73:2707–2725. https://doi.org/10.1007/ s00018-016-2246-6
- Janecek S, Kuchtova A (2012) *In silico* identification of catalytic residues and domain fold of the family GH119 sharing the catalytic machinery with the α-amylase family GH57. FEBS Lett 586:3360–3366. https://doi.org/10.1016/j.febslet.2012.07.020
- Janecek S, Sevcik J (1999) The evolution of starch-binding domain. FEBS Lett 456:119–125. https://doi.org/10.1016/s0014-5793(99) 00919-9
- Janecek S, Svensson B, MacGregor EA (2011) Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. Enzyme Microb Technol 49:429–440. https://doi.org/10. 1016/j.enzmictec.2011.07.002
- Janecek S, Svensson B, MacGregor EA (2014) α-Amylase: an enzyme specificity found in various families of glycoside hydrolases. Cell Mol Life Sci 71:1149–1170. https://doi.org/10.1007/ s00018-013-1388-z
- Janecek S, Majzlova K, Svensson B, MacGregor EA (2017) The starchbinding domain family CBM41—an *in silico* analysis of evolutionary relationships. Proteins 85:1480–1492. https://doi.org/10. 1002/prot.25309
- Janecek S, Marecek F, MacGregor EA, Svensson B (2019) Starchbinding domains as CBM families—history, occurrence, structure, function and evolution. Biotechnol Adv 37:107451. https://doi. org/10.1016/j.biotechadv.2019.107451
- Jeong DW, Jeong HM, Shin YJ, Woo SH, Shim JH (2020) Properties of recombinant 4-α-glucanotransferase from *Bifidobacterium longum* subsp. *longum* JCM 1217 and its application. Food Sci Biotechnol 29:667–674. https://doi.org/10.1007/s10068-019-00707-4
- Joo S, Kim S, Seo H, Kim KJ (2016) Crystal structure of amylomaltase from *Corynebacterium glutamicum*. J Agric Food Chem 64:5662– 5670. https://doi.org/10.1021/acs.jafc.6b02296
- Jung JH, Jung TY, Seo DH, Yoon SM, Choi HC, Park BC, Park CS, Woo EJ (2011) Structural and functional analysis of substrate recognition by the 250s loop in amylomaltase from *Thermus brockianus*. Proteins 79:633–644. https://doi.org/10.1002/prot.22911

- Kaper T, Talik B, Ettema TJ, Bos H, van der Maarel MJEC, Dijkhuizen L (2005) Amylomaltase of *Pyrobaculum aerophilum* IM2 produces thermoreversible starch gels. Appl Environ Microbiol 71:5098–5106. https://doi.org/10.1128/AEM.71.9.5098-5106. 2005
- Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4:363–371. https://doi.org/10.1038/nprot.2009.2
- Kerenyiova L, Janecek S (2020a) A detailed *in silico* analysis of the amylolytic family GH126 and its possible relatedness to family GH76. Carbohydr Res 495:108082. https://doi.org/10.1016/j. carres.2020.108082
- Kerenyiova L, Janecek S (2020b) Extension of the taxonomic coverage of the family GH126 outside Firmicutes and in silico characterization of its non-catalytic terminal domains. 3 Biotech 10:420. https://doi.org/10.1007/s13205-020-02415-x
- Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE (2019) PubChem 2019 update: improved access to chemical data. Nucleic Acids Res 47:D1102–D1109. https://doi.org/10.1093/nar/gky1033
- Kosaka T, Kato S, Shimoyama T, Ishii S, Abe T, Watanabe K (2008) The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota. Genome Res 18:442–448. https://doi.org/10.1101/gr.7136508
- Kuchtova A, Janecek S (2015) *In silico* analysis of family GH77 with focus on amylomaltases from borreliae and disproportionating enzymes DPE2 from plants and bacteria. Biochim Biophys Acta 1854:1260–1268. https://doi.org/10.1016/j.bbapap.2015.05.009
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/ molbev/msy096
- Kuriki T, Imanaka T (1999) The concept of the α-amylase family: structural similarity and common catalytic mechanism. J Biosci Bioeng 87:557–565. https://doi.org/10.1016/s1389-1723(99) 80114-5
- Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128. https://doi.org/10.1093/bioinformatics/btl529
- Liu Y, Grimm M, Dai WT, Hou MC, Xiao ZX, Cao Y (2020) CB-Dock: a web server for cavity detection-guided protein-ligand blind docking. Acta Pharmacol Sin 41:138–144. https://doi.org/ 10.1038/s41401-019-0228-6
- Lloyd JR, Blennow A, Burhenne K, Kossmann J (2004) Repression of a novel isoform of disproportionating enzyme (stDPE2) in potato leads to inhibition of starch degradation in leaves but not tubers stored at low temperature. Plant Physiol 134:1347–1354. https:// doi.org/10.1104/pp.103.038026
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/ nar/gkt1178
- MacGregor EA, Janecek S, Svensson B (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim Biophys Acta 1546:1–20. https://doi.org/10. 1016/s0167-4838(00)00302-2
- Machovic M, Janecek S (2003) The invariant residues in the α-amylase family: just the catalytic triad. Biologia 58:1127–1132
- Matsuura Y, Kusunoki M, Harada W, Kakudo M (1984) Structure and possible catalytic residues of Taka-amylase A. J Biochem 95:697– 702. https://doi.org/10.1093/oxfordjournals.jbchem.a134659
- Mehboob S, Ahmad N, Rashid N, Imanaka T, Akhtar M (2016) Pcal\_0768, a hyperactive 4-α-glucanotransferase from *Pyrobacculum calidifontis*. Extremophiles 20:559–566. https://doi.org/10. 1007/s00792-016-0850-x



- Mehboob S, Ahmad N, Munir S, Ali R, Younas H, Rashid N (2020) Gene cloning, expression enhancement in *Escherichia coli* and biochemical characterization of a highly thermostable amylomaltase from *Pyrobaculum calidifontis*. Int J Biol Macromol 165:645–653. https://doi.org/10.1016/j.ijbiomac.2020.09.071
- Nimpiboon P, Kaulpiboon J, Krusong K, Nakamura S, Kidokoro S, Pongsawasdi P (2016a) Mutagenesis for improvement of activity and thermostability of amylomaltase from *Corynebacterium glutamicum*. Int J Biol Macromol 86:820–828. https://doi.org/10. 1016/j.ijbiomac.2016.02.022
- Nimpiboon P, Krusong K, Kaulpiboon J, Kidokoro S, Pongsawasdi P (2016b) Roles of N287 in catalysis and product formation of amylomaltase from *Corynebacterium glutamicum*. Biochem Biophys Res Commun 478:759–764. https://doi.org/10.1016/j.bbrc. 2016.08.021
- O'Neill EC, Stevenson CE, Tantanarat K, Latousakis D, Donaldson MI, Rejzek M, Nepogodiev SA, Limpaseni T, Field RA, Lawson DM (2015) Structural dissection of the maltodextrin disproportionation cycle of the *Arabidopsis* plastidial disproportionating enzyme 1 (DPE1). J Biol Chem 290:29834–29853. https://doi. org/10.1074/jbc.M115.682245
- Penninga D, van der Veen BA, Knegtel RM, van Hijum SA, Rozeboom HJ, Kalk KH, Dijkstra BW, Dijkhuizen L (1996) The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. J Biol Chem 271:32777–32784. https://doi. org/10.1074/jbc.271.51.32777
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 13:1605–1612. https://doi.org/10.1002/jcc.20084
- Przylas I, Terada Y, Fujii K, Takaha T, Saenger W, Sträter N (2000a) X-ray structure of acarbose bound to amylomaltase from *Thermus* aquaticus. Implications for the synthesis of large cyclic glucans. Eur J Biochem 267:6903–6913. https://doi.org/10.1046/j.1432-1033.2000.01790.x
- Przylas I, Tomoo K, Terada Y, Takaha T, Fujii K, Saenger W, Sträter N (2000b) Crystal structure of amylomaltase from *Thermus aquaticus*, a glycosyltransferase catalysing the production of large cyclic glucans. J Mol Biol 296:873–886. https://doi.org/10.1006/jmbi. 1999.3503
- Pugsley AP, Dubreuil C (1988) Molecular characterization of malQ, the structural gene for the *Escherichia coli* enzyme amylomaltase. Mol Microbiol 2:473–479. https://doi.org/10.1111/j.1365-2958. 1988.tb00053.x
- Rachadech W, Nimpiboon P, Naumthong W, Nakapong S, Krusong K, Pongsawasdi P (2015) Identification of essential tryptophan in amylomaltase from *Corynebacterium glutamicum*. Int J Biol Macromol 76:230–235. https://doi.org/10.1016/j.ijbiomac.2015.02.035
- Roth C, Weizenmann N, Bexten N, Saenger W, Zimmermann W, Maier T, Sträter N (2017) Amylose recognition and ring-size determination of amylomaltase. Sci Adv 3:e1601386. https://doi.org/10. 1126/sciadv.1601386
- Ruzanski C, Smirnova J, Rejzek M, Cockburn D, Pedersen HL, Pike M, Willats WG, Svensson B, Steup M, Ebenhöh O, Smith AM, Field RA (2013) A bacterial glucanotransferase can replace the complex maltose metabolism required for starch to sucrose conversion in leaves at night. J Biol Chem 288:28581–28598. https://doi.org/ 10.1074/jbc.M113.497867
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75
- Sorimachi K, Le Gal-Coëffet MF, Williamson G, Archer DB, Williamson MP (1997) Solution structure of the granular starch

binding domain of *Aspergillus niger* glucoamylase bound to  $\beta$ -cyclodextrin. Structure 5:647–661. https://doi.org/10.1016/s0969-2126(97)00220-7

- Srisimarat W, Powviriyakul A, Kaulpiboon J, Krusong K, Zimmermann W, Pongsawasdi P (2011) A novel amylomaltase from *Corynebacterium glutamicum* and analysis of the large-ring cyclodextrin products. J Incl Phenom Macrocycl Chem 70:369–375. https://doi.org/10.1007/s10847-010-9890-5
- Srisimarat W, Kaulpiboon J, Krusong K, Zimmermann W, Pongsawasdi P (2012) Altered large-ring cyclodextrin product profile due to a mutation at Tyr-172 in the amylomaltase of *Corynebacterium glutamicum*. Appl Environ Microbiol 78:7223–7228. https://doi. org/10.1128/AEM.01366-12
- Steichen JM, Petty RV, Sharkey TD (2008) Domain characterization of a 4-α-glucanotransferase essential for maltose metabolism in photosynthetic leaves. J Biol Chem 283:20797–20804. https://doi. org/10.1074/jbc.M803051200
- Svensson B, Jespersen H, Sierks MR, MacGregor EA (1989) Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem J 264:309–311. https:// doi.org/10.1042/bj2640309
- Takaha T, Yanase M, Okada S, Smith SM (1993) Disproportionating enzyme (4-α-glucanotransferase; EC 2.4.1.25) of potato. Purification, molecular cloning, and potential role in starch metabolism. J Biol Chem 268:1391–1396
- Terada T, Fujii K, Takaha T, Okada S (1999) Thermus aquaticus ATCC 33923 amylomaltase gene cloning and expression and enzyme characterization: production of cycloamylose. Appl Environ Microbiol 65:910–915. https://doi.org/10.1128/AEM.65.3.910-915.1999
- Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function efficient optimization and multithreading. J Comput Chem 31:455–461. https:// doi.org/10.1002/jcc.21334
- Tumhom S, Krusong K, Pongsawasdi P (2017) Y418 in 410s loop is required for high transglucosylation activity and large-ring cyclodextrin production of amylomaltase from *Corynebacterium glutamicum*. Biochem Biophys Res Commun 488:516–521. https:// doi.org/10.1016/j.bbrc.2017.05.078
- Tumhom S, Krusong K, Kidokoro SI, Katoh E, Pongsawasdi P (2018) Significance of H461 at subsite +1 in substrate binding and transglucosylation activity of amylomaltase from *Corynebacterium glutamicum*. Arch Biochem Biophys 652:3–8. https://doi.org/10. 1016/j.abb.2018.06.002
- UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169. https://doi.org/10.1093/ nar/gkw1099
- Valk V, Lammerts van Bueren A, van der Kaaij RM, Dijkhuizen L (2016) Carbohydrate-binding module 74 is a novel starch-binding domain associated with large and multidomain α-amylase enzymes. FEBS J 283:2354–2368. https://doi.org/10.1111/febs. 13745
- van der Maarel MJEC, Leemhuis H (2013) Starch modification with microbial α-glucanotransferase enzymes. Carbohydr Polym 93:116–121. https://doi.org/10.1016/j.carbpol.2012.01.065
- van der Maarel MJEC, van der Veen B, Uitdehaag JC, Leemhuis H, Dijkhuizen L (2002) Properties and applications of starch-converting enzymes of the α-amylase family. J Biotechnol 94:137–155. https://doi.org/10.1016/s0168-1656(01)00407-2
- Vujicic-Zagar A, Pijning T, Kralj S, Lopez CA, Eeuwema W, Dijkhuizen L, Dijkstra BW (2010) Crystal structure of a 117 kDa glucansucrase fragment provides insight into evolution and product specificity of GH70 enzymes. Proc Natl Acad Sci USA 107:21406–21411. https://doi.org/10.1073/pnas.1007531107
- Wattebled F, Ral JP, Dauvillee D, Myers AM, James MG, Schlichting R, Giersch C, Ball SG, D'Hulst C (2003) STA11, a



3 Biotech (2021) 11:229

*Chlamydomonas reinhardtii* locus required for normal starch granule biogenesis, encodes disproportionating enzyme. Further evidence for a function of  $\alpha$ -1,4 glucanotransferases during starch granule biosynthesis in green algae. Plant Physiol 132:137–145. https://doi.org/10.1104/pp.102.016527

- Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36. https://doi.org/10.1021/ci000 57a005
- Weiss SC, Skerra A, Schiefner A (2015) Structural basis for the interconversion of maltodextrins by MalQ, the amylomaltase of

Escherichia coli. J Biol Chem 290:21352–21364. https://doi.org/ 10.1074/jbc.M115.667337

- Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. Mol Biol Evol 18:691–699. https://doi.org/ 10.1093/oxfordjournals.molbev.a003851
- Yun JH, Bae JW (2018) Complete genome sequence of the halophile bacterium Kushneria marisflavi KCCM 80003<sup>T</sup>, isolated from seawater in Korea. Mar Genomics 37:35–38. https://doi.org/10. 1016/j.margen.2017.11.002

