Structural elements determining the transglycosylating activity of glycoside hydrolase family 57 glycogen branching enzymes

Gang Xiang[1], Hans Leemhuis[2], and Marc van der Maarel[1]

[1]University of Groningen
[2]Avebe Group

May 4, 2021

Abstract

Glycoside hydrolase family 57 glycogen branching enzymes (GH57GBE) catalyze the formation of an α-1,6 glycosidic bond between α-1,4 linked glucooliogosaccharides. As an atypical family, a limited number of GH57GBEs have been biochemically characterized so far. This study aimed at acquiring a better understanding of the GH57GBE family by a systematic sequence-based bioinformatics analysis of almost 2,500 gene sequences and determining the branching activity of several native and mutant GH57GBEs. A correlation was found between a very low or even no branching activity with the absence of a flexible loop, a tyrosine at the loop tip, and two β-sheets.

Structural elements determining the transglycosylating activity of glycoside hydrolase family 57 glycogen branching enzymes

Gang Xiang[1], Hans Leemhuis[1,2], and Marc J.E.C. van der Maarel[1]

1. Bioproduct Engineering, Engineering and Technology institute Groningen (ENTEG), University of Groningen, Groningen, 9747 AG, the Netherlands

2. Avebe Innovation Center, Groningen, 9747 AA, the Netherlands

Email address: Gang Xiang: g.xiang@rug.nl

Hans Leemhuis: hans.leemhuis@avebe.com

Marc J.E.C. van der Maarel: m.j.e.c.van.der.maarel@rug.nl

Corresponding author: Marc J.E.C. van der Maarel

Abstract

Glycoside hydrolase family 57 glycogen branching enzymes (GH57GBE) catalyze the formation of an α-1,6 glycosidic bond between α-1,4 linked glucooliogosaccharides. As an atypical family, a limited number of GH57GBEs have been biochemically characterized so far. This study aimed at acquiring a better understanding of the GH57GBE family by a systematic sequence-based bioinformatics analysis of almost 2,500 gene sequences and determining the branching activity of several native and mutant GH57GBEs. A correlation was found between a very low or even no branching activity with the absence of a flexible loop, a tyrosine at the loop tip, and two β-sheets.

Key words: Glycogen branching enzymes, Glycoside hydrolase Family 57, Glycogen, Amylose

Introduction

Glycogen branching enzymes (GBE) (EC 2.4.1.18) play a key role in the biosynthesis of glycogen, a dendrimeric polyglucose carbon and energy storage molecule present in many prokaryotic microorganisms, fungi, yeast, and animals[1-4]. GBE catalyze the formation of α-1,6 glycosidic linkages in glycogen by transglycosylating a cleaved-off α-glucan chain from a growing α-1,4 linked anhydroglucopyranose chain to the same or a different chain[1, 5-7]. In addition to the transglycosylating or branching reaction, GBEs also catalyze the hydrolysis of an α-1-4 glycosidic linkage using water as an acceptors[8], or perform a cyclization reaction resulting in branched cyclic glucans[9].

Based on the primary amino acid sequence and conserved motifs, all known GBEs are categorized in either the glycoside hydrolase (GH) family 13 or 57[10-13]. The GBEs from GH13 have a substantially high activity on the model substrate amylose, a typical linear α-glucan[14-17]. GH13 GBEs are involved in the classical glycogen biosynthetic pathway by the tandem action of glucose-1-phosphate adenylyltransferase (*glgC*) - glycogen synthase (*glgA*) - glycogen branching enzyme (*glgB*)[2, 18]. In contrast to GH13 GBEs, the role of GH57 GBEs is much less clear. So far, the biochemical properties of only five GH57 GBEs have been reported, while the crystal structure of only four of these five GH57 GBEs has been solved[6, 10, 19, 20]. The activity of these five GH57 GBEs on amylose is relatively low to almost zero, ranging from a few mU/mg (*Thermotoga maritima*SMB8) to a few hundred mU/mg (*Thermus thermophilus* HB8, *Thermococcus kodakarensis* KOD1)[8, 21]. The GH57 GBE of *Mycobacterium tuberculosis* , having all the key features of GH57 GBEs, was reported to have no detectable activity on a range of α-glucans[22]. Based on the genomic organization, it is assumed that this GBE branches glucosylglycerate and plays a role in the biosynthesis of polymethylated polysaccharides[22-24].

GH57 GBEs have a triangular three-dimensional shape consisting of three domains including a catalytic $(\beta/\alpha)_7$ barrel containing the two catalytic residues, a glutamate nucleophile and an aspartate acid/base catalyst[6, 10, 19, 20]. All GH 57 GBEs have five conserved sequence regions (CSR), with the nucleophile located in CSRIII and the acid/base catalyst located in CSRIV[25]. Similar to GH13 GBEs, GH57 GBEs employ a double-displacement reaction mechanism resulting in retention of the α-configuration in the products[6]. In the crystal structures of the *P. horikoshii* , *T. kodakarensis,* and *T. thermophilus* GBE a flexible loop with a conserved tyrosine at the tip was identified[6, 10, 19]. Mutational analysis showed that the flexible loop and the tyrosine play a key role in the branching activity; replacing the tyrosine with an alanine in the *T. thermophilus* GBE or shortening the loop in *P. horokoshii* GBE resulted in a loss of the branching activity[6, 19].

The simultaneous presence of a gene encoding a GH13 GBE and one encoding a GH57 GBE in the genomes of a large number of bacteria makes the physiological role of GH57 GBEs even more puzzling[21]. To gain more insight into the enzymatic activity and possible physiological role of GH57 GBEs, an in-depth sequence-based bioinformatics analysis of almost 2,500 GH57 GBE sequences and a basic biochemical characterization of a number of carefully selected GH57 GBEs overexpressed in *Escherichia coli* was conducted. Surprisingly, the flexible loop covering the active site was absent in the vast majority of the GH57GBE sequences analyzed in this study. Several of these loop-deficient GBEs displayed very low to no activity on amylose. Besides the flexible loop an additional structural element, two adjacent beta sheets, was identified to play a key role in the branching activity. It is proposed that GH57 GBEs devoid of the flexible loop and the two beta sheets are not glycogen branching enzymes and do not play a role in glycogen biosynthesis but in one or more yet to be identified metabolic pathways.

2. Materials and methods

2.1 Sequence collection, alignment and phylogenetic analysis

Sequences were collected by searching the key word "DUF1957" at the National Center for Biotechnology Information (NCBI) on 04th February 2020. To filter redundant and short sequences, the source database and the sequence length were set at RefSeq and 400-1,200 amino acid residues, respectively. Partial and hypothetical sequences were deleted manually. Finally, 2,497 sequences with the title "DUF1957 domain-containing protein" or "Glycoside hydrolase family 57 protein" were used in the sequence alignment and phylogenetic analysis. ClustalW multiple alignment was conducted with the software package MEGA-X[26]

by setting gap opening penalty at 10, the gap extension penalty at 0.2, the negative matrix at off, and a 30 % delay divergent cutoff. A phylogenetic tree was constructed using the Jones-Taylor-Thornton (JTT) model with the maximum likelihood method in default setting in the software package MEGA-X[26, 27]. Five conserved sequence regions were analyzed by calculating possibilities of residues within conserved sequence regions based on the multiple sequence alignment, as described previously[28].

2.2 Identification of flexible loop and two beta sheets

Protein structures were visualized using the Pymol software[29]. The three-dimensional structures of *T. thermophilus* (PDB ID: 3P0B)[6], *T. kodakarensis* (PDB ID: 3N8T)[10], *P. horokoshii*(PDB ID: 5WU7)[19] and *T. maritima* GBE (AmyC) (PDB ID: 2B5D)[20] were downloaded from the Protein Data Bank (PDB). The distances between two residues was measured with the measurement command in Pymol. The flexible loop and the two beta sheet regions were defined based on the crystal structures.

2.3 Overexpression and enzyme production

Various putative GH57 GBEs were overexpressed in *Escherichia coli*(Table 1), and several mutants were also constructed (Table 2). Genes with optimized codons for overexpression in *Escherichia coli* were synthesized and then ligated in the pRSET B (*T. kodakarensis* GBE, *T. thermophilus* GBE and *T. maritima* GBE) or pET 28a(+) (the remaining GBEs used in this study) vector by Genscript (USA). Plasmids were transformed into *E. coli* BL21 (DE3) by subjecting competent cells to a heat shock at 42°C for 90 seconds. A single colony was selected from ampicillin or kanamycin-containing agar plate for further cultivation. *E. coli*strains carrying the target gene were stored in 25% glycerol stock at -80°C. *E. coli* BL 21(DE3) carrying the native or mutant *gbe* was grown in Luria-Bertani (LB) medium (Becton, Dickinson and Company, USA) at 37°C and 150 rpm. When the optical density reached 0.6-0.8 (600 nm), protein expression was induced at 20°C for 12 hours by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Thermo Fisher Scientific, USA) to a final concentration of 0.2 mM. In order to enhance the solubility of the target protein, the chaperone plasmid pG-KJE8 (Takara, Japan) was co-expressed in *E. coli* by following the manufacture's protocol. Cells were harvested by centrifugation at 8,000×g for 10 min at 4°C, washed cells twice with 50 mM phosphate buffered saline (PBS) buffer (pH 7.0), and finally resuspended in 5 mM PBS (pH7.0). Cells were disrupted by high pressure homogenizer (Avestin, Canada) at room temperature, with an air pressure of 8 bar and a valve pressure of 6 bar. Heat treatment was performed at 50°C (*Calidithermus timidus* GBE,*Meiothermus silvanus* GBE, *Petrotoga mexicana* GBE) or 65°C (the remaining GBEs used in this study) for 15 minutes to denature *E. coli* host proteins. Then the supernatant was incubated with nickel metal resin, followed by removal of non-specifically bound host protein by applying a washing buffer containing 25 mM imidazole (Sigma-Aldrich, USA) and elution of the target protein with 250 mM imidazole. Imidazole was removed using a desalting column (Thermo Fisher Scientific, USA) and a storage buffer to 5 mM PBS (pH7.0). Purity of the target protein was checked by SDS-PAGE. Protein concentration was assayed by Bradford (Bio-Rad, USA) using bovine serum albumin as standard.

2.4 Determination of catalytic activity

The branching reaction was performed at 50°C in 5 mM PBS buffer (pH7.0) using 0.125% amylose V (Avebe, The Netherlands) as substrate, and a final concentration of enzyme ranging from 0.03 mg/ml to 0.13 mg/ml was used depending on the total activity of enzyme. Samples were taken at time intervals to monitor the reaction progress. The reaction was stopped by boiling samples at 100°C for 10 min. Branched samples were treated with *Pseudomonas* sp isoamylase (0.4 U/ml) and *Klebsiella planticola* pulullanase M1 (1.4 U/ml) (Megazyme, Ireland) at 40°C for 24 h with constant mild shaking. Reducing ends were quantified by the 2,2'-bicinchoninic acid (BCA) method using glucose as standard[30]. The activity unit was calculated based on reducing end profiles of samples before and after debranching. One unit of activity is defined as 1 μmol reducing ends released or transferred per minute under aforementioned reaction conditions.

3. Results and discussion

3 .1 Sequence-based analysis of putative GH57 GBE sequences

Using the key words "DUF1957 domain-containing protein" or "Glycoside hydrolase family 57 protein", 2,497 amino acid sequences were retrieved from the NCIB database. These sequences varied in length between 418 and 1,184 amino acids. Except for 50, all sequences had the nucleophile and acid/base catalyst and contained the five conserved sequence regions typical for GH57 members (Fig. 1). The exception were sequences that missed one or both catalytic residues and showed a large variation in four of the five conserved sequence regions. These sequences were excluded from further analysis as it was assumed that they are not active. The first four conserved sequence regions are positioned within the A-domain containing the $(\beta/\alpha)_7$ barrel. Conserved region 5 is located in the C domain on the second α helix.

When comparing the sequence logo for all the GH57 GBEs of this study with the logos recently published on 1,602 GH57 sequences[25], two GBE specific fingerprints become clear; the first is a quintet of amino acids with the combination HxHLP, with x being A, S or T, found in CSRI of almost all GBE sequences; in a small number of GBE sequences the L at position 4 is replaced by an I or M. In all other GH57 enzymes a Q is present instead of an L at position 4 whereas in α-galactosidases and -related proteins there is an L at position 4 but this is followed by a Q or A/M and not a P as is the case for GH57 GBEs. The second GBE fingerprint is the sextet ELF(Y)GHW present in CSRIV. The first position of this fingerprint, the E, is conserved among all proteins assigned to a functional GH57 enzyme subfamily. This E is not conserved in the proteins that are categorized as -like proteins. These proteins miss one or both of the catalytic residues and are very likely not active. In the 4-α-glucanotransferase of *Thermococcus litoralis* the E is only 5.1 Å from the acid-base catalyst D (Fig. 2A) and is involved in binding the -1 subsite residue through a water molecule[31]. In the other GH57 crystal structures, the conserved E is 4 Å (*T. maritima* AmyC) to 7 Å (*T. thermophilus* GBE) from the acid-base catalyst (Fig. 2B and 2C). In GH13 enzymes, a catalytic triad of a catalytic nucleophile D, a general acid base catalyst E, and a transition state stabilizer D play a key role in catalysis [32-34]. As the CSRIV E is completely conserved in all GH57 proteins assigned to a functional subfamily and is positioned close to the acid-base catalyst in all available GH57 crystal structures, it is not unlikely to assume that this E plays a similar role as the transition state stabilizer D in GH13.

The other positions of the sextet are completely conserved in all GBE proteins analyzed in this study, with the exception of the third position, the F which is in *K. pacifica* GBE replaced by another hydrophobic side chain containing amino acid, an Y. Whereas previously it was reported that the C at position 16 (CSRIII) is conserved among GH57 GBEs[6], this position is not absolutely invariant, as 72 out of the 2,497 (2.9%) sequences have a different amino acid in this position, a feature also noticed by Martinovičová and Janeček[25]; the majority of these have an M (56; 2.2%), nine have an S, five an L and two an F. This almost fully conserved C can still be seen as a fingerprint as none of all the other GH57 enzymes and -like proteins have a C at this position.

In addition, five residues in the vicinity of the active sites were identified to be fully conserved in all sequences; three tryptophans (W274, W404 and W413), one histidine (H146), and one arginine (R265) (*T. thermophilus* numbering). In *T. kodakarensis* , the three tryptophans and the one at position 28 have been defined as the aromatic gate keepers[8]. This group of four aromatic gate keeper tryptophans is highly conserved in all GH57 GBEs except W28, which is a threonine in the GBEs of *Thermus* and *Meiothermus* species. In the *T. thermophilus* GBE the W274 is positioned to the side and the W404 at the bottom of the positive subsites[6]. Both are involved in substrate binding by aromatic stacking (W274) and hydrogen bonding (W404). In the *T. maritima* GBE, the W274 equivalent (W246) is buried such that aromatic stacking is very unlikely to occur while the position of the W413 equivalent (W411) is difficult to predict[6]. The role of the H146 and the R265 is not clear.

In the *P. horokoshii* GBE, a tryptophan (W22) at the bottom of the active site groove is involved in substrate recognition. Changing this W into an A resulted in almost complete loss of activity[19]. This W is also found at the same position in the crystal structure of *T. thermophilus* and *T. kodakarensis* . In GH57GBEs from *T. maritima* , *P. mexicana* ,*P. mobilis* and *K. pacifica* , this W is replaced by D or E or P. Besides the bottom W four other aromatic amino acids are found in close vicinity of the active sites of *T. kodakarensis* , *T. thermophilus* , *P. horokoshii* , or *T. maritima* ; F23, F289, W360 and F461 (*T. thermophilus* numbering).

In all the other sequences of this study, three of these four aromatic amino acids are functionally conserved while the F23 is not conserved. Zhang et al.[8] reported another three important amino acids near the active site, H11, S462, and D463 (*T. thermophilus* numbering). These are all conserved at the respective positions in all the 2,497 sequences.

3.2 The flexible loop and tyrosine

From the sequence alignment, two major gaps became visible. The first was a sequence of 19 amino acids from position 229 to 247 (*T. thermophilus* GBE numbering) in between CSR III and IV. In the crystal structures of *T. thermophilus* [6], *T. kodakarensis* [10], and *P. horokoshii* [19] GBE, these amino acids make up a flexible loop covering the catalytic cleft (Fig. 3). In 201 other amino acid sequences, a flexible loop of varying length is present (Fig. 4). The average loop size is 26 amino acids with the shortest being 13 amino acids, while the longest is 50 amino acids. At the tip of the flexible loop of *T. thermophilus* , *T. kodakarensis* , and *P. horokoshii* GBE, a tyrosine is present[6, 10, 19]. The *T. thermophilus* tyrosine was shown to play a prominent role in the branching activity. Replacing this tyrosine with an alanine resulted in a loss of branching activity[6]. A minority of GBEs with a flexible loop do not have a tyrosine at the tip (35; 17.2%). Instead an alanine, serine, or threonine are found. The *Thermoanaerobaculum aquaticum* GBE, which has a medium sized loop of 22 residues with an alanine at the tip, has a relatively high activity towards amylose, dominated by the branching activity of 168 mU/mg and a ratio of branching over hydrolytic activity of 38.1 (Table 3). On the contrary, the *Calidithermus timidus* GBE with a flexible loop of 23 amino acids and a tyrosine at the tip, has a branching activity of 356.9 mU/mg, being twice that *T. aquaticum* GBE. This result indicates that the tyrosine at the tip of the flexible loop does not play a determining role in the branching activity. This was confirmed by a tyrosine to alanine mutation in the *T. kodakarensis* GBE, which has a flexible loop of 19 amino acids, and the same size and configuration as that of the *T. thermophilus* GBE (Table 3; Fig. 3). Whereas the wild type *T. kodakarensis* GBE has a dominant branching activity of 480 mU/mg and a branching over hydrolysis ratio of 41, the Y/A mutant retains the dominant branching activity (426 mU/mg). The hydrolytic activity of the Y/A mutant doubled from 12.6 mU/mg to 26.3 mU/mg, resulting in a branching over hydrolysis ratio of 16.2. These results suggest that not only the tyrosine at the tip of the loop but also the size and configuration of the flexible loop play a role in the branching activity.

In contrast to the *T. thermophilus* , *T. kodakarensis* and *P. horokoshii* GBE, the flexible loop is absent in the *T. maritima* GBE (Fig. 3 and 5), as was already noted by Zhang and coworkers[8]. It has been reported that AmyC, in spite of the lack of a flexible loop, has a low but reproducible branching activity towards amylose[8]. Shortening of the flexible loop in the *P. horokoshii* GBE from 19 to 9 amino acids resulted in a twofold increase in total activity and a considerable reduction of the branching activity[19]. The flexible loop is not only absent in AmyC but also in 2,296 of the 2,497 sequences (92%) used in this study (Table 4). A comparable low branching activity was found for the GBE of *P. mexicana* and *Kosmotoga pacifica* , of 14.1 mU/mg and 9.1 mU/mg respectively (Table 3). Although having a low activity, both enzymes introduced 6% branches in the final product when incubated with amylose V (data not shown), confirming that both enzymes clearly have branching activity. Introduction of the full or partial flexible loop of *T. kodakarensis* GBE, including the tyrosine, in the *T. maritima* GBE resulted in a two-fold increase of the branching activity and a three-fold increase of the hydrolytic activity (Tables 3), being substantially lower than the activity of the wild type *T. kodakarensis* GBE. The flexible loop appears not to be the only structural element that determines the overall activity of the GH57 GBEs.

The length of about 19 residues and configuration of the *T. thermophilus* and *T. kodakarensis* GBE flexible loop seems to be optimal with respect to the branching activity. Hundred and three of the 201 sequences (51%) have a flexible loop of 17 to 22 amino acids (Fig. 4). A smaller, but still significant number of proteins (81; 40%) have a flexible loop of 24 or more amino acids while 17 proteins (9%) have a flexible loop of 13 to 15 amino acids. The models of the *T. hugenholzii* and *T. lipolytica* GBE, with 36 and 28 amino acids loop resp., show that only the first part of the flexible loop folds into the active site cleft while the rest folds next to or behind the part that covers the cleft (Fig. 5). The *T. hugenholzii* flexible loop contains a tyrosine at position 245 which is turned away from the cleft being far away from the two active site residues (E191

and D382). This GBE has a very low branching activity of only 13.9 mU/mg, while the *T. lypolytica* GBE did not show any activity (Table 3). Thus, for a GH57 GBE to act as a "true" glycogen branching enzyme, a flexible loop of 17-22 amino acids covering the active cleft and, when present, a tyrosine positioned deep into the active site close to the catalytic E and D is required. The absence of the flexible loop or a loop smaller or larger than 17-22 amino acids results in a significant reduction to complete loss of activity towards amylose. What the *in-vivo*substrate for these loop-deficient and long-looped GH57 GBEs is, remains to be established. It could be that these GBEs do show activity towards a growing α-glucan chain, the *in-vivo* substrate of most GBEs[1].

3.3. The two beta sheets

The second gap in the sequence alignment is in the β5-β6 region, where in the *T. thermophilus* GBE two prominent beta sheets (β5-β6; 188-205, *T. thermophilus* GBE numbering), are localized[6]. Without exception, all 2,296 GBEs (92 %) that miss the flexible loop also do not have the β5 nor β6 sheet (Table 4). In contrast, the majority of the 201 sequences that have a flexible loop also possess the β5-β6 region (159 or 79.1%). Four of the GBEs with a flexible loop and the β5-β6 region were included in this study and showed considerable branching activity ranging from 168 mU/mg to 550 mU/mg. Of those GBEs with a flexible loop but without the β5-β6 region, one, that from *Meiothermus* sp. PNK-Is4, was overproduced and analysed. It was active towards amylose but as other GBEs without the two beta sheets, it only had a very low branching activity (29.1 mU/mg). Also, the branching over hydrolysis ratio was very similar to those GBEs that missed the β5-β6 region; e.g. *K. pacifica* and AmyC with a branching over hydrolysis ratio of 3.8 and 5.7 respectively. A similar result was found for the *T. hugenholzii* GBE, having a long flexible loop of 36 amino acids with a tyrosine and a very short β5-β6 region (Table 3). No activity could be detected for the *T. lipolytica* GBE, which has a long flexible loop of 28 residues without a tyrosine and no β5-β6 region. Removing the β5-β6 region from the*T. kodakarensis* GBE resulted in a complete loss of activity (Table 3).

Interestingly, the *Mycobacterium tuberculosis* H37Rv GH57 GBE also misses the flexible loop as well as the β5-β6 region. Gusthart overexpressed the *M. tuberculosis* GH57GBE but did not find any activity[22]. When incubating the *M. tuberculosis*GH57GBE with amylose, no activity was detected (Gang and van der Maarel, unpublished results). The β5-β6 region thus plays a determining role in the activity of GH57 GBEs, in particular in the branching activity. How exactly this beta sheet region exerts its influence on the branching activity remains unclear.

4. Conclusions

Bioinformatics analysis reveals the sequence-based features of putative GH57GBE sequences, that are two GH57GBE-specific sequence logos, variations in flexible loop and presence/absence of two β-sheets, were noted in 2,497 sequences, with a majority of them (92%) having no flexible loop and absence of two β-sheets; Biochemical data verified that putative GH57GBE sequences can be defined by the former feature and true GH57GBEs can be designated by the latter two structural features which were assigned as optimal flexible loop and presence of two β-sheets. These results guide us to further investigate GH57GBEs or –like sequences in an explicit direction.

Acknowledgments

We thank China Scholarship Council (GX) and University of Groningen for financial support. Special appreciation to Zhenhua Zhang for assistance with sequence collection.

Conflict of interest

The authors declare that they have no conflict of interests.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

References

1. Suzuki R, Suzuki E. Structure and function of branching enzymes in eukaryotes. *Trends Glycosci Glycotechnol.* 2020; 32:21-30.

2. Wilson WA, Roach PJ, Montero M, et al. Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiol Rev* . 2010; 34:952-985.

3. Adeva-Andany MM, González-Lucán M, Donapetry-García C, Fernández-Fernández C, Ameneiros-Rodríguez E. Glycogen metabolism in humans. *BBA Clin.* 2016; 5:85-100.

4. Wu S, Liu Y, Yan Q, Jiang Z Gene cloning, functional expression and characterisation of a novel glycogen branching enzyme from *Rhizomucor miehei* and its application in wheat breadmaking. *Food Chem.* 2014; 159:85-94.

5. Li L, Su L, Hu F, Chen S, Wu J. Recombinant expression and characterization of the glycogen branching enzyme from *Vibrio vulnificus* and its application in starch modification. Int J Biol Macromol. 2020; 155:987-994.

6. Palomo M, Pijning T, Booiman T, et al. *Thermus thermophilus* glycoside hydrolase family 57 branching enzyme crystal structure, mechanism of action, and products formed.*J Biol Chem.* 2011; 286:3520-3530.

7. Van der Maarel, MJEC, Leemhuis H. Starch modification with microbial α-glucanotransferase enzymes. *Carb Pol.* 2013; 93:116-121.

8. Zhang XW, Leemhuis H, Janeček Š, et al. Identification of*Thermotoga maritima* MSB8 GH57 α-amylase AmyC as a glycogen-branching enzyme with high hydrolytic activity. *Appl Microbiol Biotechnol.* 2019; 103:6141-6151.

9. Takata H, Ohdan K, Takaha T, Kuriki T, Okada S. Properties of branching enzyme from hyperthermophilic bacterium, *Aquifex aeolicus* , and its potential for production of highly-branched cyclic dextrin. *J Appl Glycosci.* 2003; 50: 15-20.

10. Santos CR, Tonoli CCC, Trindade DM, et al. Structural basis for branching-enzyme activity of glycoside hydrolase family 57: Structure and stability studies of a novel branching enzyme from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *Proteins.* 2011; 79:547-557.

11. Stam MR, Danchin EG, Rancurel C, Coutinho PM, Henrissat B. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Eng Des Sel.* 2006; 19:555-562.

12. Blesak K, Janeček Š. Sequence fingerprints of enzyme specificities from the glycoside hydrolase family GH57. *Extremophiles* . 2012; 16:497–506.

13. Martinovičová M, Janeček Š. In silico analysis of the α-amylase family GH57: eventual subfamilies reflecting enzyme specificities.*Biotech.* 2018; 8:307.

14. Van Der Maarel MJEC, Vos A, Sanders P, Dijkhuizen L. Properties of the glucan

branching enzyme of the hyperthermophilic bacterium *Aquifex aeolicus.* Biocatal Biotransfor. 2003; 21:199–207.

15. Palomo M, Kralj S, van der Maarel MJEC, Dijkhuizen, L. The unique branching patterns of *Deinococcus* glycogen branching enzymes are determined by their N-terminal domains. Appl Environm Microbiol. 2009; 75:1355–1362.

16. Zhang X, Leemhuis H, Maarel MJEC. Synthesis of highly branched α-glucans with different structures using GH13 and GH57 glycogen branching enzymes. *Carb Pol.* 2019; 216:231-237.

17. Ban X, Dhoble AS, Li CM, et al. Bacterial 1, 4-α-glucan branching enzymes: characteristics, preparation and commercial applications. *Crit Rev Biotechnol.* 2020; 40:380-396.

18. Roach PJ, Depaoli-Roach AA, Hurley TD, Tagliabracci VS. Glycogen and its metabolism: some new developments and old themes. *Biochem J.* 2012; 441:763-787.

19. Na S, Park M, Jo I, Cha J, Ha N-C. Structural basis for the transglycosylase activity of a GH57-type glycogen branching enzyme from *Pyrococcus horikoshii .Biochem Biophys Res Commun.* 2017; 484:850-856.

20. Dickmanns A, Ballschmiter M, Liebl W, Ficner R. Structure of the novel α-amylase AmyC from*Thermotoga maritima* . *Acta Crystallogr D.* 2006; 62:262-270.

21. Zhang X, Leemhuis H, Maarel MJEC Characterization of the GH13 and GH57 glycogen branching enzymes from *Petrotoga mobilis* SJ95 and potential role in glycogen biosynthesis. *PloS One.* 2019; 14: e0219844

22. Gusthart JS An investigation into an unusual glycan branching enzyme from

Mycobacterium tuberculosis: PhD thesis University of Southampton. 2018. https://eprints.soton.ac.uk/ 425868/. Accessed April 28, 2021

23. Mendes V, Maranha A, Alarico S, Empadinhas N. Biosynthesis of mycobacterial methylglucose lipopolysaccharides. *Nat Prod Rep.*2012; 29:834-844.

24. Stadthagen G, Sambou T, Guerin M, et al. Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in *Mycobacterium tuberculosis* . *J Biol Chem.* 2007; 282:27270-27276.

25. Janeček Š, Martinovičová M. New groups of protein homologues in the α-amylase family GH57 closely related to α-glucan branching enzymes and 4-α-glucanotransferases. Genetica. 2020; 148:77-86.

26. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018; 35:1547-1549.

27. Jones DT, Taylor WR, Thornton JM The rapid generation of mutation data matrices from protein sequences.*Bioinform.* 1992; 8:275-282.

28. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14:1188-1190.

29. DeLano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsletter on protein crystallography 2002; 40:82-92. http://legacy.ccp4.ac.uk/newsletters/ newsletter40.pdf. Accessed April 28 2021

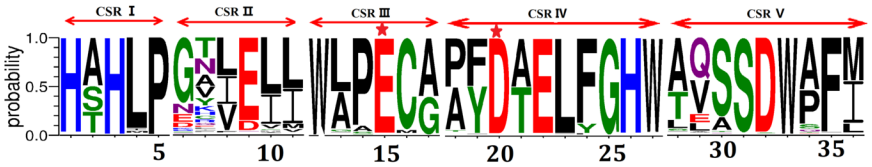
30. Utsumi Y, Yoshida M, Francisco PB, et al. Quantitative assay method for starch branching enzyme with bicinchoninic acid by measuring the reducing terminals of glucans. J *Appl Glycosci.* 2009; 56:215-222.

31. Imamura H, Fushinobu S, Yamamoto M, et al. Crystal structures of 4-α-glucanotransferase from*Thermococcus litoralis* and its complex with an inhibitor. *J Biol Chem.* 2003; 278:19378-19386.

32. Matsuura Y, Kusunoki M, Harada W, Kakudo M Structure and possible catalytic residues of Taka-amylase A.*J Biochem.* 1984; 95:697-702.

33. Buisson G, Duee E, Haser R, Payan F. Three dimensional structure of porcine pancreatic alpha-amylase at 2.9 A resolution. Role of calcium in structure and activity. EMBO J. 1987; 6:3909-3916.

34. Uitdehaag JC, Mosi R, Kalk KH, et al. X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family.*Nature Struct Biol.* 1999; 6:432-436.



* Two catalytic residues are indicated by a red star symbol.

Figure 1. Fingerprint logos of five CSRs among 2,447 glycoside hydrolase family 57 glycogen branching enzyme homologues

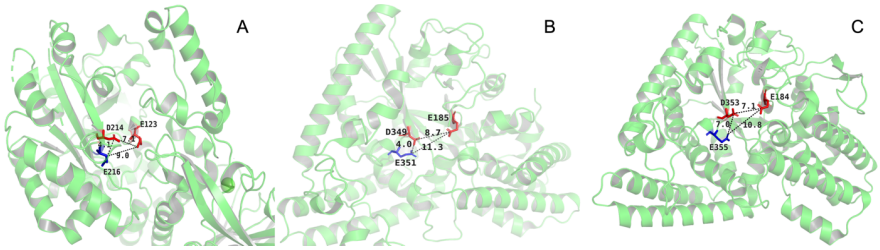


Figure 2. Distance in Å between the conserved E of CSRIV (blue) and catalytic residues E in CSR III and D in CSR IV (red). A. GH57 4-α-glucanotransferase of *Thermococcus litoralis* ; B. GH57 AmyC of *Thermotoga maritima* ; C. GH57 Glycogen branching enzyme of *Thermus thermophilus* .

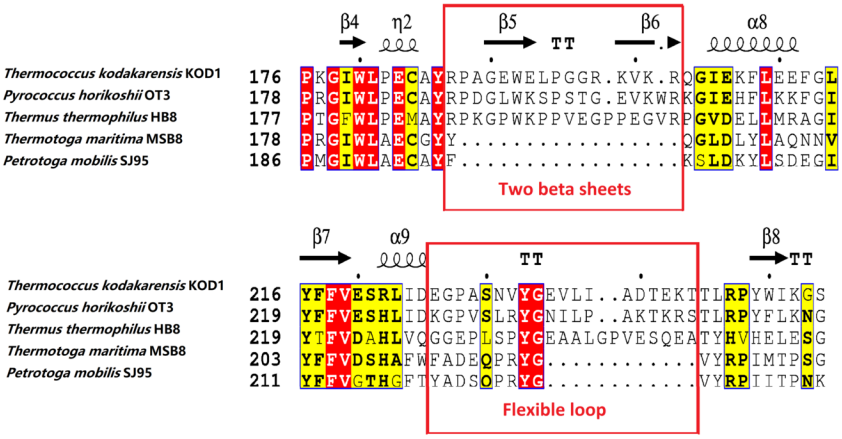


Figure 3. Partial alignment of the regions containing the β5 and β6 sheets and the flexible loop plus tyrosine of the five biochemically characterized glycoside hydrolase family 57 glycogen branching enzymes.

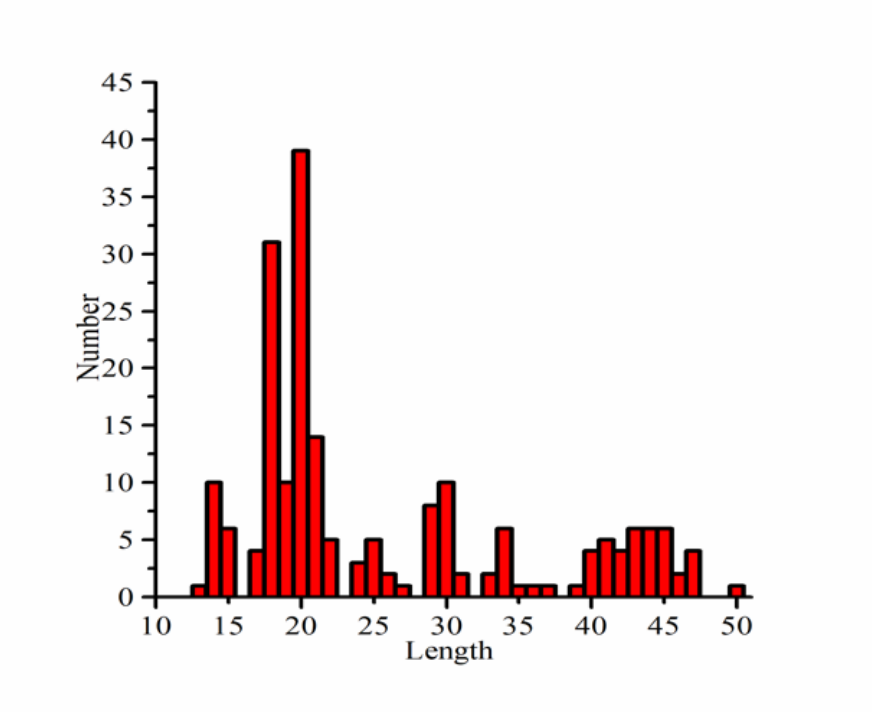


Figure 4. Length distribution of the flexible loop among the 201 loop-containing glycoside hydrolase family 57 glycogen branching enzymes

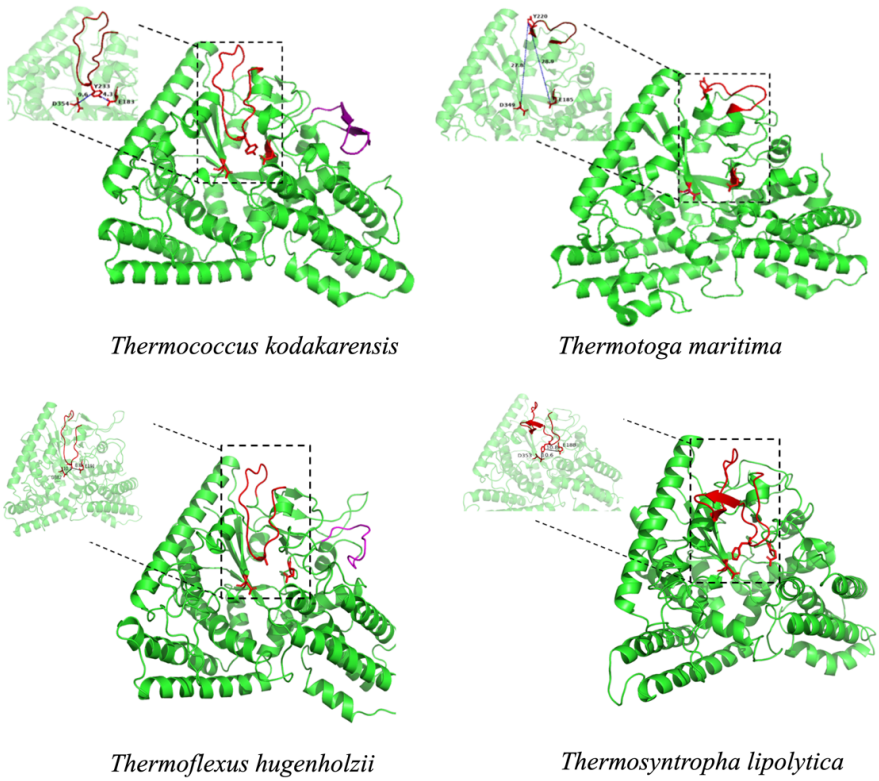


Figure 5. Configuration of the several glycoside hydrolase family 57 glycogen branching enzyme showing the position of the flexible loop, the tyrosine at the tip of the flexible loop, and the presence/absence of the β5 and β6 sheets (purple). The insert shows the position and distance of the tyrosine at the tip of the flexible loop and the two catalytic residues.

Table 1. Information and sequence similarity of various putative GH57GBEs used in this study.

Organism	Abbreviation	Accession number	Length (aa)	Sequence identity (%)
Thermus thermophilus HB8	Tt GBE	WP_011228999.1	520	100.00
Calidithermus timidus DSM 17022	Ct GBE	WP_018467494.1	708	66.2
Meiothermus sp. PNK-Is4	Ms GBE	WP_129865543.1	520	62.9
Thermoflexus hugenholtzii JAD2	Th GBE	WP_088572346.1	558	46.3
Thermococcus kodakaraensis KOD1	Tk GBE	WP_011250387.1	675	44.6
Thermoanaerobaculum aquaticum MP-01	Ta GBE	WP_053334947.1	522	39.4
Thermosyntropha lipolytica DSM 11003	Tl GBE	WP_073092191.1	519	36.4
Kosmotoga pacifica SLHLJ1	Kp GBE	WP_047754759.1	531	36.0
Petrotoga mexicana DSM 14811	Pme GBE	WP_103077822.1	538	33.6
Thermotoga maritima SMB8	Tm GBE	WP_004081707.1	528	32.8

Table 2. Mutants of tyrosine residue, flexible loop and two beta sheets used in this study

Mutant enzyme code	Mutant type	Sequence
Tk GBE Y233A	Single site mutation	221 SRLID EGPASNV*A*GEVLIADTEKT TLRP248
Tm GBE LS	Loop swap	208 SHAFW *EGPASNVYGEVLIADTEKT* GVYR236
Tm GBE LE	Loop extension	208 SHAFW FADEQPRY*GEVLIADTEKT* GVYR236
Tk GBE β sheet deletion	β sheet deletion	182 PECAY \soutRPAGEWELPGGRKVKR QGIE190

Table 3. Branching and hydrolytic activity of various glycoside hydrolase family 57 glycogen branching enzymes. The activity was determined using amylose as substrate.

	Structural element Loop	**Structural element Tyrosine**	**Structural element β σηεετς**	**Branching (mU/mg)**	**Branching (mU/mg)**	**Hydrolytic (mU/mg)**	**B/H[&]**

Wild-type enzyme							
Thermus thermophilus HB8	+	+	+	480.0 ± 3.5[#]	480.0 ± 3.5[#]	11.7 ± 1.5	41.0 ± 2.1
Thermococcus kodakaraensis KOD1	+	+	+	550.0 ± 5.5	550.0 ± 5.5	12.6 ± 1.7	43.6 ± 3.3
Calidithermus timidus DSM 17022	+	+	+	356.9 ± 10.6	356.9 ± 10.6	15.3 ± 2.5	24.1 ± 4.6
Thermoanaerobaculum aquaticum MP-01	+	-	+	168.0 ± 9.8	168.0 ± 9.8	4.4 ± 0.2	38.3 ± 4.1
Meiothermus sp. PNK-Is4	+	+	-	29.1 ± 3.5	29.1 ± 3.5	6.5 ± 1.5	4.7 ± 1.7
Thermosyntropha lipolytica DSM 11003	+	+	-	NA[$]	NA	NA	/
Thermoflexus hugenholtzii JAD2	+	+	+	13.9 ± 0.1	13.9 ± 0.1	0.5 ± 0.1	30.6 ± 3.5
Petrotoga mexicana DSM 14811	-	+	-	14.1 ± 3.4	14.1 ± 3.4	1.4 ± 0.3	10.0 ± 0.5
Kosmotoga pacifica SLHLJ1	-	+	-	9.1 ± 0.8	9.1 ± 0.8	2.4 ± 0.1	3.8 ± 0.8
Thermotoga maritima SMB8	-	+	-	10.5 ± 0.3	10.5 ± 0.3	1.8 ± 0.5	5.7 ± 0.1
Mutant enzyme							
Tk GBE Y233A				426.2	26.3	26.3	16.2
Tm GBE LS				24.4 ± 3.9	6.2 ± 0.1	6.2 ± 0.1	4.0 ± 0.6
Tm GBE LE				22.6 ± 3.1	5.6 ± 0.4	5.6 ± 0.4	4.0 ± 0.8
Tk GBE β sheet deletion				NA	NA	NA	/

[&] ratio branching activity to hydrolysis activity

[#] average of three independent measurements with standard deviation

[$] NA = no activity

Table 4. Distribution of flexible loop and β5-β6 sheets among 2.497 glycoside hydrolase family 57 glycogen branching enzymes

Structural elements	# sequences	% total
No loop and no β5-β6 sheets	2,296	92
Loop and β5-β6 sheets	159	5.3
Loop and no β5-β6 sheets	42	2.7