MALAYSIAN JOURNAL OF BIOCHEMISTRY & MOLECULAR BIOLOGY

The Official Publication of The Malaysian Society For Biochemistry & Molecular Biology (MSBMB)
http://mjbmb.org

DECIPHERING THE MOLECULAR LANDSCAPE: *IN-SILICO* ANALYSIS OF CYCLODEXTRIN GLYCOSYLTRANSFERASE FOR ENHANCED ENZYME FUNCTIONALITY AND CYCLODEXTRIN SYNTHESIS

Nik Ida Mardiana Nik-Pa[1,2], Mohamad Farhan Mohamad Sobri[1,3], Nurhasliza Zolkefli[1], Suraini Abd-Aziz[1], Mohamad Faizal Ibrahim[1], Noorjahan Banu Mohamed Alitheen[4], Norhayati Ramli[1,5,*]

[1]Department of Bioprocess Technology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[2]Section of Bioengineering Technology, Universiti Kuala Lumpur Branch Campus, Malaysian Institute of Chemical & Bioengineering Technology, Taboh Naning, 78000 Alor Gajah, Melaka, Malaysia
[3]School of Bioprocess Engineering, Universiti Malaysia Perlis, Kompleks Pusat Pengajian Jejawi 3, 02600 Arau, Perlis, Malaysia
[4]Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[5]Laboratory of Biopolymer and Derivatives, Institute of Tropical Forestry and Forest Products (INTROP), Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

*Corresponding Author: Tel.: +603 9769 1945; E-mail: yatiramli@upm.edu.my

History
Received: 12 August 2023
Accepted: 23 December 2023

Keywords:

CGTase; In-silico; Structure prediction; SWISS-MODEL; Signal peptide

Abstract

Cyclodextrin glycosyltransferase (CGTase) has always played a significant role in the production of cyclodextrin through the cyclization reaction. The wide application of this valuable protein demands a better understanding, leading to a comprehensive *in-silico* analysis of CGTase. The analysis focused on the functional domain composition of the recombinant CGT-BS protein by comparing it with several other CGTase proteins from different *Bacillus* spp. A three-dimensional (3-D) model was constructed to predict the active, substrate binding and cyclization sites of the CGT-BS protein. Structural function prediction revealed the active site within domain A at the wide end of the (β/α)8-barrel, with Asp 268, Glu 296 and Asp 357 as the catalytic residues. Additionally, the reaction site for cyclization was identified in domain B at Tyr 234. In comparison to maltose binding sites (MBS) 1 and 2 which are associated with raw starch binding activity, a comparable role is deduced for the MBS identified on the surface of the Domain E protein. We additionally observed that the residues Tyr 139, Arg 266 and Asp 367, located at the substrate binding cleft of the catalytic site, exhibited heightened hydrophobicity and concurrent cyclization activity. The successful extracellular expression of the CGT-BS protein is also anticipated to be facilitated by the presence of a functional signal peptide. In conclusion, our in-depth *in-silico* analysis unveils critical insights into the structural and functional aspects of CGT-BS protein, laying the groundwork for further exploration of its catalytic mechanisms and potential applications in cyclodextrin synthesis.

INTRODUCTION

Carbohydrate-active enzymes (CAZy), involved in modification, biosynthesis, binding, and catabolism of carbohydrates, have been classified based on their catalytic roles. These roles include glycosyltransferases, glycoside hydrolases, carbohydrate esterases, polysaccharide lyases and auxiliary activities [1]. Specifically, for cyclodextrin

glycosyltransferase (CGTase, EC 2.4.1.19), it has been classified as a member of glycosyl hydrolase family 13 (http://www.cazy.org/). Its main function is to convert starch into cyclodextrins via intramolecular transglycosylation (cyclization) reaction [2]. Despite the high demand for cyclodextrin on large scales in industries such as food [3], pharmaceutical [4], environment [5] and agricultural [6], the low CGTase production by the native producer has become a bottleneck for its widespread application. Enhancing CGTase activity can be achieved by expressing the gene of interest in recombinant *Escherichia coli*. Recombinant protein production by *E. coli* is preferred due to its well-known cellular and genetic properties, which make the expression of many foreign proteins both possible and easy to carry out [7]. The recombinant *E. coli* expressing the CGTase gene was reported to enhance CGTase production with shorter cultivation time and lower cost by simplifying enzyme purification [8] due to the production of specific CGTase.

Another challenge lies in the failure of extracellular protein translocation. One of the strategies is by introducing signal peptides, which play a key role in enhancing the translocation efficiency of the recombinant proteins [9], conceptualized upon the cruciality of a positively charged amino-terminal (n-region, 1-5 residues), a central-hydrophobic region (h-region, 7-15 residues) and a polar carboxy-terminal domain (c-region, 3-7 residues) in cleaving the amino terminus of the protein once its targeting function is executed. The proper removal of the signal peptide from the mature chain has also been investigated, with significant emphasis on specifying and modifying the signal peptidase cleavage site of the c-region [10, 11]. The manipulations of the n-region [12] and h-region have been widely adopted to ensure the proper extracellular expression of CGTase [13]. Studies are also being done on the optimization of native signal peptides to enhance the extracellular recombinant protein translocation [14]. Moreover, the structure and function of various signal peptides have been precisely elucidated to evaluate the secretory production of recombinant proteins. This includes predicting functional signal peptides through *in-silico* studies [15, 16].

Furthermore, understanding the genetic information and structural properties of CGTase is of utmost priority to pioneer a suitable heterologous expression system, specifically for the synthesis of this functional protein. Various bioinformatics techniques have been developed to predict the protein interactions within and between protein complexes, drawing on prior knowledge of the corresponding protein structures. By leveraging the extensively studied structural relationships within the α-amylase family, it may be possible to elucidate the mechanisms of CGTase activity, given that these enzymes exhibit approximately 30% amino acid sequence identity [17]. In general, while the glycosidic bonds in starch molecules are being hydrolyzed by amylases, CGTases will contrarily catalyze the transglycosylation reactions instead, which shows that hydrolysis is just its minor activity [18]. CGTase structure consists of five domains; A, B, C, D and E where the first three domains are structurally similar with α-amylase. Meanwhile, domains D and E are only unique to CGTases [2]. The A domain, which is common to all α-amylase family enzymes, is the catalytic domain and comprises a (β/α)8 or triosephosphate isomerase (TIM) barrel [19; 20], while the other domains mainly consist of β-sheets and loop regions [21]. The unique domains D and E are involved in binding starch and play a role in the cyclization and transglycosylation reactions specific to CGTases [22; 23].

Cyclodextrin, typically derived from starch degradation, is a naturally occurring cyclic oligosaccharide, structured as ring-shape with non-polar interior and polar exterior [24]. Comprising seven α-D-glucose units, β-cyclodextrin forms a truncated cone with a central cavity of approximately 262 Å3 [25]. This cyclodextrin group, widely produced through enzymatic starch degradation by *Bacillus macerans* [24], is the most preferred owing to its affordability and availability [25]. Nevertheless, its derivatives are reportedly better for industrial application, as the native β-cyclodextrin tends to have lower solubility issues [26]. Therefore, structural modification, such as substituting hydrogen-bond-forming hydroxyl groups [26], offers a promising alternative to enhance its solubility in aqueous solution. However, the primary focus should be on achieving higher yield production, and the synthetic production of the CGTase enzyme, holds great promise in addressing this concern.

In our previous studies, cgt-BS gene was expressed extracellularly from *E. coli*, showing a 2.6-fold increase after optimizing glycine supplementation [27]. Furthermore, a 5.6-fold increment was achieved through combined optimization of codon and glycine induction [28], compared to the native producer, *Bacillus* sp. NR5 UPM [29]. In this study, we aimed to utilize *in-silico* approaches based on amino acid sequence correlation between distant positions in the multiple sequence alignment of several *Bacillus* spp. from previous studies. Further construction of three-dimensional (3-D) modeling of the CGT-BS protein involves comparing its sequences with those of β-CGTase from the alkalophilic *Bacillus circulans* 251. *Bacillus circulans* has been reported as a significant producer of β-CGTase in several studies focusing on its immobilization for batch fermentations [30, 31]. Specifically, for the alkalophilic strains, the remarkable specificity of its CGTase in producing 95% β-cyclodextrin [32] has further motivated us to predict of the active site, cyclization reaction site and maltose binding sites (MBS) using this bacterium as the model. The prediction of residues possible for site directed mutagenesis, aimed at increasing hydrophobicity and consequently enhancing cyclization activity, was also provided by comparing with previous studies. In addition, the signal peptide that functions in secreting the CGT-BS protein across the plasma membrane was also characterized.

MATERIALS AND METHODS

2.1. Sequence alignment and conserved motif identification

The CGT-BS protein sequences (GenBank accession number: HQ876173.1) were aligned with several CGTase sequences from *Bacillus* spp. namely *B. circulans* 251 (GenBank accession number: X78145), *B. circulans* (GenBank accession number: X68326.1) and *Bacillus* sp. KC201 (GenBank accession number: D13068.1) using protein-protein Basic Local Alignment Search Tool (BLASTP) available from National Center for Biotechnology Information (NCBI).

2.2. Secondary and tertiary structure predictions

The PSI-blast based secondary structure prediction (PSIPRED) analysis [33] was carried out to indicate the coil, strand and helix, and hydrophobic character of the protein. SWISS-MODEL (https://swissmodel.expasy.org/) was utilized in constructing the secondary and tertiary protein structures with primary protein sequence as the template. To predict the active and cyclization sites, as well as the MBS for substrate binding, the protein sequences of CGT-BS were compared with protein sequences of β-CGTase from alkalophilic *B. circulans* 251. From this analysis, Ramachandran plot was constructed as a means to indicate the quality of the protein model [34].

2.3. Prediction of physicochemical characteristics of signal peptide

Multiple physicochemical characteristics of the signal peptide were evaluated by ExPASY's ProtParam online tool at http://web.expasy.org/protparam/ [35] including number of positively and negatively charged amino acids, instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY).

2.4. Protein localization analyses

To predict the probability of the CGT-BS protein secretion into the growth media, SignalP 5.0 (http://www.cbs.dtu.dk/services/SignalP/), Phobius (https://phobius.sbc.su.se/) [36] and Philius (http://www.yeastrc.org/philius/pages/philius/runPhilius.jsp) [37] were utilized. Gpos-mPLoc software (http://www.csbio.sjtu.edu.cn/bioinf/Gpos-multi/) was also used to predict the subcellular localization of CGT-BS protein to the external medium environment.

RESULTS AND DISCUSSION

3.1. Characterization of CGT-BS by multiple sequence alignment against CGTases expressed by *Bacillus* spp.

The CGT-BS amino acid sequence from this study was compared with other CGTases sequences using the BLASTP program provided by NCBI. As shown in **Figure 1**, the five domains (labeled as domains A to E) in CGT-BS protein were identified. Domain A comprises approximately 300-400 amino acid residues which can be found in each of the enzymes coming from the α-amylase family. It comprises of eight parallel β-strands with a tremendous symmetrical fold that is organized in a barrel encompassed by eight α-helices, usually known as (β/α)8-or triosephosphate isomerase (TIM)-barrel catalytic domain [18]. Being separated into A1/B/A2, this catalytic domain is visualized as a domain B positioned between the β-sheet 3 and α-helix 3 of domain A, which primarily formed a (β/α)8 barrel with a prominent and extensive loop [38].

It is reported that A/B or A1/B/A2 structure which represents the major substrate binding site and catalytic center of the enzyme is conserved in the α-amylase family [39]. The conversion of starch to cyclodextrin is catalyzed by three conserved carboxylates located in the A2 domain, where a glucan is covalently bound and subsequently cleaved through a retaining double-displacement mechanism. Following an intramolecular transglycosylation reaction, cyclodextrin is being synthesized when the intermediate is translocated to the C4-OH group of its reducing end [40].

Meanwhile, domain B comprises 44 to 133 amino acid residues that are responsible for the binding of substrate [18]. It also has been recognized as the domain that is involved in thermostability [41], even though limited numbers of studies have been reported. The heat-labile CGTase produced from *B. circulans* 251 was previously mutated by introducing a salt bridge at domain B, which resulted in the increased activity half-life of the CGTase by 6-fold at 60°C [42].

Domain C has an antiparallel β-sandwich fold which consists of approximately 100 amino acid residues **(Figure 2)**. Lawson et al. [43] reported that domain C of β-CGTase expressed by *B. circulans* 251 assimilates one of the MBS which is observed in the structure derived from maltose dependent crystals. The detection of this MBS suggested a role for domain C in the binding of raw starch during the cyclodextrin anabolism. On the other hand, domain D which consists of 90 amino acids with an immunoglobulin fold was found only in CGTases. By comparison, domain E which is more ubiquitous in starch-degrading enzymes with an estimated 110 amino acid residues is responsible for the adsorption onto granular starch [39].

3.2. Secondary structure prediction using PSIPRED

The PSIPRED analysis showed that the protein primarily consists of coils, with 22.4% of helix and 22.5% of strand, respectively **(Figure 3(a))**. In the pursuit of enhanced cyclization activities, three additional residues have been selected as possible candidates for site directed mutagenesis, aiming to increase the hydrophobicity around the substrate binding cleft of the catalytic center. Referring to Wang et al. [44], a similar approach was done on β-CGTase from the

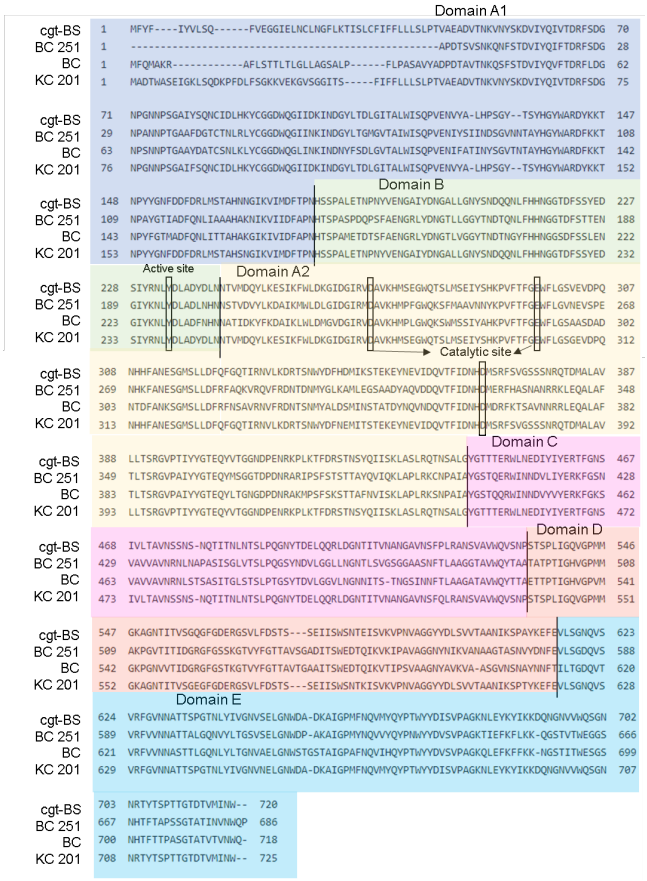

Figure 1. Amino acid sequence alignment among CGT-BS from this study (HQ876173.1), BC251: β-CGTase from *Bacillus circulans* 251 (X78145), BC: CGTase from *Bacillus circulans* (X68326.1) and KC201: CGTase from *Bacillus* sp. KC201 (D13068.1). Domains A, B, C, D and E of β-CGTase, along with the active and catalytic sites, are labeled accordingly.

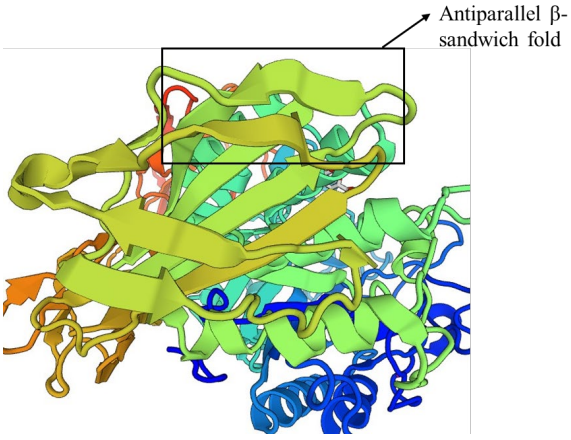

Figure 2. Theoretical 3-D structure of CGT-BS on the side of domain C, featuring an antiparallel β-sandwich fold.

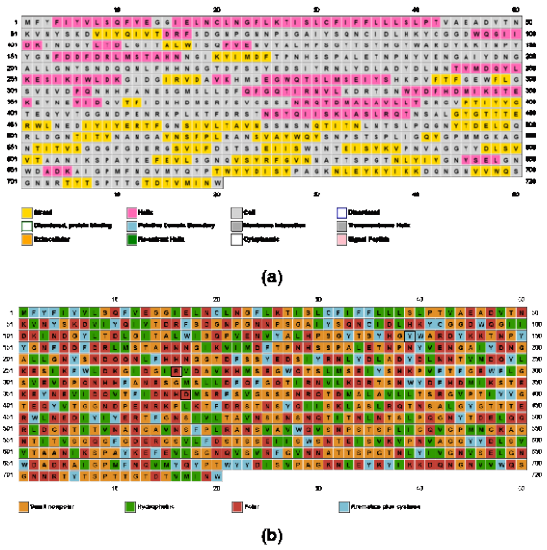

Figure 3. Secondary structure prediction determined by PSI-blast based secondary structure prediction (PSIPRED); (a) differences in coil, strand, and helix, with a higher frequency of random coil in the CGT-BS protein; (b) analysis of the protein's hydrophobic character. Regions boxed in black are suggested as possible sites for site directed mutagenesis to increase hydrophobicity around the substrate binding cleft of the catalytic center, as suggested by Wang et al. [44].

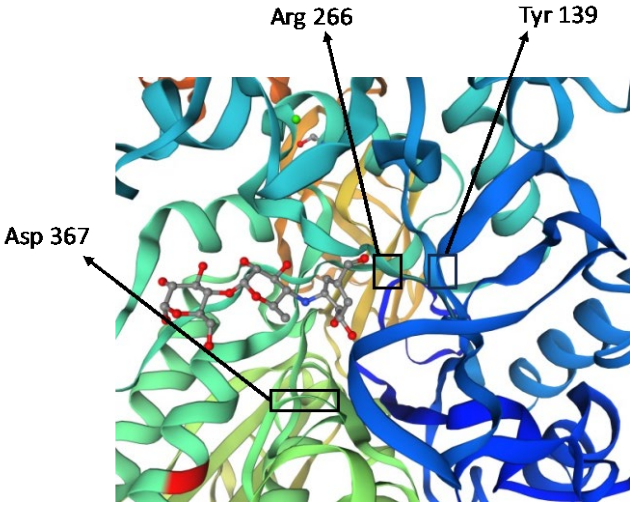

Figure 4. Theoretical 3-D structure of CGT-BS, highlighting the side of Tyr 139, Arg 266 and Asp 367.

alkalophilic *Bacillus* sp. N-227, focusing specifically on residues Tyr 127F, Arg 254F and Asp 355R. The mutant β-CGTase demonstrated the same optimal pH and temperature, as well as pH and thermal stabilities, while exhibiting an

elevated cyclization activity on corn starch substrate, ranging by 1.64 to 2.1-fold, as compared to the wild type. The corresponding residues in CGT-BS namely Tyr 139, Arg 266 and Asp 367, have been identified **(Figure 4)**. All of these residues are located within domain A, which is responsible for catalysis **(Figure 3(b))**. However, further experiments are required to validate these *in-silico* results.

3.3. Predicted protein structure using SWISS-MODEL Program

The CGT-BS protein sequence was compared to a known protein with high sequence identity and coverage to further extrapolate on the secondary and higher levels of CGT-BS protein structure. The amino acid sequences derived from cgt-BS gene of *Bacillus* sp. NR UPM (this study) showed a sequence identity of 59.26% and a coverage of 0.8 with the amino acid sequence of β-CGTase from the alkalophilic *B. circulans* strain 251 (SMTL ID: 1dtu.1) (https://swissmodel.expasy.org/templates/1dtu.1).

The construction of Ramachandran plot, as shown in **Figure 5**, allows an exposition of the energetically favored regions for the dihedral angles of backbone amino acid residues in the protein structure. Due to its capability in detecting gross error in the structure, Ramachandran plot of residue φ-ψ torsion angles is acknowledged as the best means to evaluate the quality of a protein model. The generated model for CGT-BS protein showed a calculated favorable value of 94.51%. Overall, the model generated in this research is legitimate, with favorable values exceeding the 90% threshold set for residues located within the favored regions [34].

The secondary structures and the associated residues were identified by comparing the protein sequence against amino acid sequence alignment between CGT-BS protein (HQ876173.1) and BC251: CGTase from *B. circulans* 251 (X78145) as the template **(Figure 6)**. The result elucidates the presence of the CGT-BS active site in domain A at the wide end of the (β/α)8-barrel, observed together with catalytic residues of Asp 268, Glu 296 and Asp 357 as shown in **Figure 7**, corresponding to Asp 229, Glu 257 and Asp 328 in CGTase protein from *B. circulans* strain 251 (BC251), respectively [17]. The importance of the catalytic residues was discussed by Nakamura et al. [45] in which by replacing Asp 229, Glu 257 and Asp 328 with Asn or Gln resulted in loss of their starch-degrading and β-cyclodextrin forming activities. It can be said that Asp 229, Glu 257 and Asp 328 are crucial in catalyzing the enzymatic reaction instigated by CGTase, exhibiting similar catalytic mechanisms present in both CGTases and α-amylases. In support of this finding, Dijkhuizen et al. [17] demonstrated the relevance of these carboxylates in the active site by adopting a site-directed mutagenesis, where these three residues were replaced with asparagine (Asp 229N, Asp 328N) and glutamine (Glu 257Q). These changes caused a near vanishing of the liquefying, saccharifying and cyclodextrin formation activities of CGTase, indicating the importance of these catalytic residues in CGTase reactions.

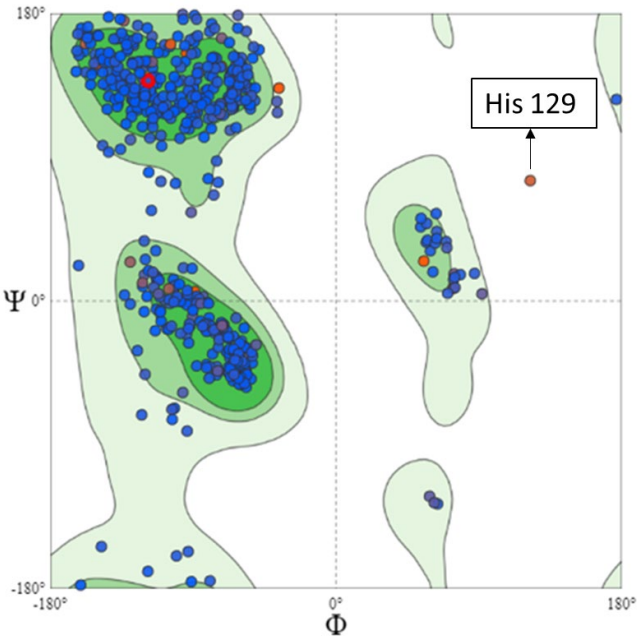

Figure 5. Ramachandran plot of CGT-BS protein. Calculated favored value: 94.51%. Outlier: Histidine (H) 129, with a value of 0.59%.

Meanwhile, the reaction site for the cyclisation reaction of CGT-BS is present at Tyr 234, corresponding with the active site of Tyr 195 in BC251, located in domain B, as shown in **Figure 8.** The cyclization of starch is exclusively carried out by CGTase, resulting in the formation of cyclic structures known as cyclodextrins. Conversely, α-amylases function by hydrolyzing starch, breaking α-(1,4) glycosidic bonds and producing linear products such as maltose and other oligosaccharides. Dijkhuizen et al. [17] identified Tyr 195 as the key factor in the cyclization of oligosaccharides. An obvious difference is observed in CGTases that possess a large aromatic amino acid (Tyr or Phe) at position 195 in contrast to α-amylases where a small residue (Gly, Ser or Val) is present at the exact same position. A previous study on site directed mutagenesis revealed that mutants with replacement of Tyr 195 with Trp, Phe, Leu and Gly led to significant reduction of cyclodextrin forming activity [46]. However, the disproportionation and saccharifying reactions of the constructed mutants were not negatively affected. Therefore, the presence of aromatic amino acid (Tyr 234) in CGT-BS is important for the efficiency of cyclization reaction, thereby potentially preventing or even minimizing the formation of linear products.

In addition, the binding of substrates is indispensable in enzyme catalysis. In affirmation to this fact, the detection of domain E in various starch-degrading enzymes was revealed when the sequences of domain E from several amylases and CGTases were compared with the raw

binding domain from glucoamylases [39]. Furthermore, three MBS were spotted on the protein surface in domain E, where MBS1 and MBS3 contribute to the intermolecular crystal contacts. It was also suggested that MBS1 and MBS2 found in domain E play a role in the raw starch binding activity **(Figure 9)**. The conserved residues of Trp 651, Trp 698, Lys 686 and Asn 703 of CGT-BS is a part of MBS1, corresponding to Trp 616, Trp 662, Lys 651 and Asn 667 in BC251, respectively. As described by Penninga et al. [47], the maltose unit binds through van der Waals contacts between the indole groups within MBS1, contributed by Trp 616 and Trp 662 with the glucose ring. The maltose binding is further enhanced by the direct hydrogen bonds with the side groups of Lys 651 and Asn 667 and water-mediated hydrogen bonds with the main chain carbonyl oxygen atoms of Trp 616 and Glu 663.

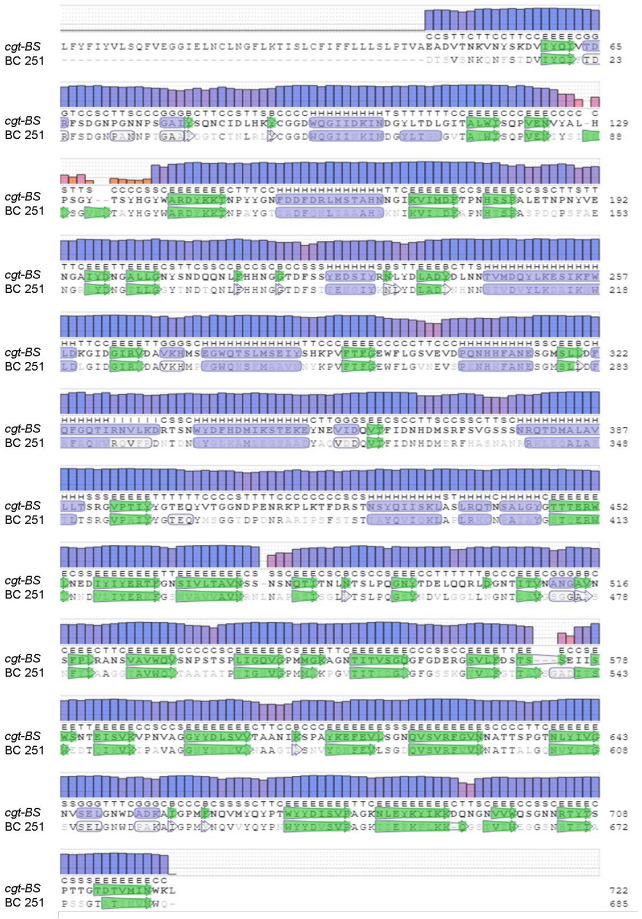

Figure 6. Alignment of CGT-BS amino acid sequences to the template, CGTase of *Bacillus circulans* 251 associated with Tris. QMEAN values indicate the modeling errors and estimate the expected model accuracy. α-Helix and β-sheets are represented by blue and green segments, respectively.

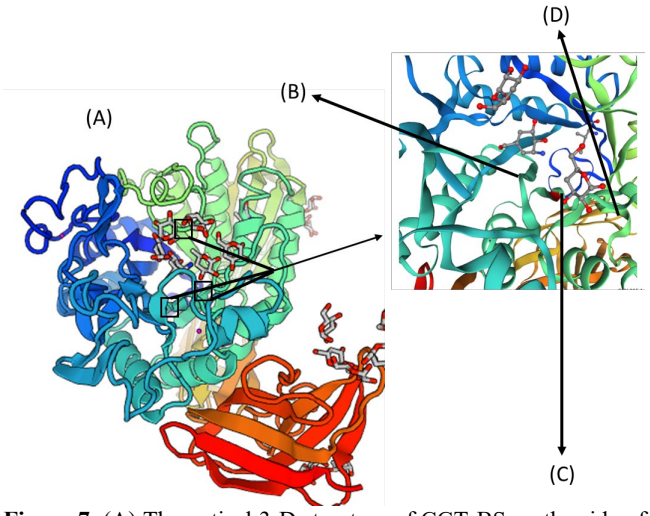

Figure 7. (A) Theoretical 3-D structure of CGT-BS on the side of the active site (boxed). The locations of the catalytic residues (Asp 268, Glu 296 and Asp 357) are as seen in (B), (C) and (D), respectively.

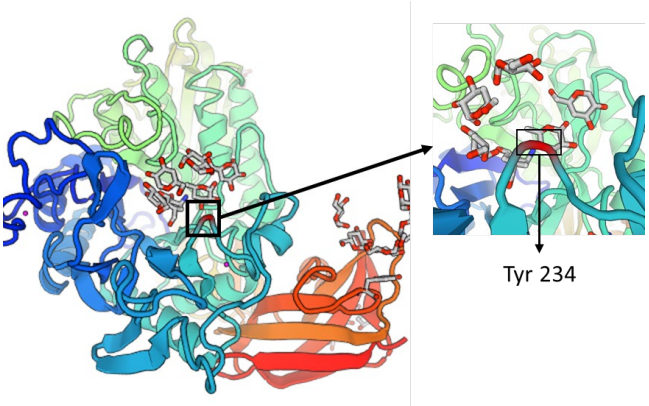

Figure 8. Theoretical 3-D structure of CGT-BS on the side of the cyclization reaction site (boxed) with the catalytic residue of Tyr 234.

In a different study, the importance of MBS1 was shown by deleting the whole MBS1 site where the conversion of starch to cyclodextrins was reduced from 28% to about 17%. In addition, the ratio of β-cyclodextrin to total cyclodextrins decreased from a level of 83% to 61% as compared to the wild-type strain [48]. The deletion of the MBS1 site has altered the catalytic functionality of β-CGTase, inhibiting the cyclodextrin formation reaction. MBS1 was also demonstrated to be crucial for raw starch binding, as proven by the mutational study of CGTase conducted by Lawson et al. [43].

Meanwhile, MBS2 in CGT-BS is present at Tyr 668, Thr 633 and Gly 636, corresponding to Tyr 633, Thr 598 and Gly 601 in BC251, respectively. Tyr 633 in MBS2 forms van der Waals contacts with a glucose residue [46]. Several elements complete the formation of the maltose binding which are the direct hydrogen bonds with the side

chains of Thr 598, Asn 627 and Gln 628, the main chain carbonyl oxygen atoms of Ala 599 and Gly 601, and one water-mediated hydrogen bond with Asn 603 [43].

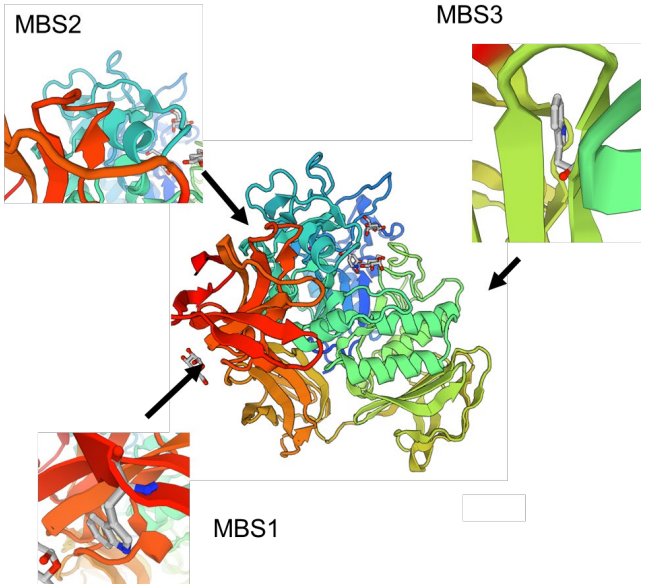

Figure 9. Theoretical 3-D structure of CGT-BS on the side of the maltose binding sites (MBS) 1, 2 and 3 (boxed).

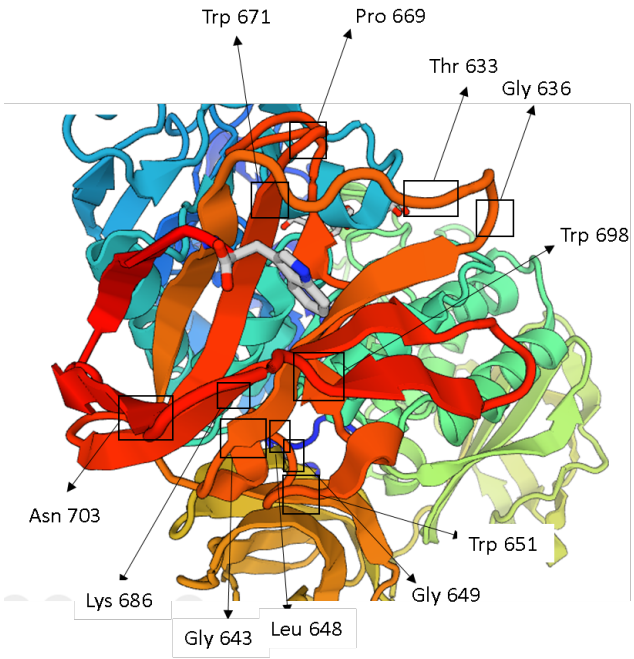

Figure 10. Theoretical 3-D structure of CGT-BS on the side of the raw-starch binding motif in domain E.

Chen et al. [49] mentioned that, as an effort to minimize product inhibition, a purposeful modification was done on the CGTase from *B. circulans* STB01, which displayed a mixed-type product inhibition. By replacing Asn 603 with aspartic acid, there was a 23.9% increase in cyclodextrin yield per gram of CGTase. This improvement resulted from the reductions in competitive and non-competitive product inhibition, as exhibited by the Asn 603D variant's activity [49]. Through a previous mutational study, it was found that MBS2 also assists in attracting linear starch chains into the active site through a groove on the surface of the CGTase protein [43].

On the other hand, MBS3 was reported to be present in domain C. Trp 452 of CGT-BS is present at MBS3, corresponding to Trp 413 in BC251. A maltose molecule is stacked with the non-polar face of the reducing sugar against the aromatic side chain of Trp 413 [17]. While three hydroxyl groups of the maltose are not involved in hydrogen bonding, another 10 hydrogen bonds work to achieve further binding to the protein. Additionally, hydrogen bonds mediating interactions between the sugar and protein were also formed in the presence of water molecules. Simultaneously, four water molecules appeared to be associated with both the protein and maltose [17].

The presence of raw-starch binding motif in domain E could indicate strong binding between CGTases and raw starch. Eleven strictly conserved residues; Thr 633-Gly 636-Gly 643-Leu 648-Gly 649-Trp 651 -Pro 669-Trp 671-Lys 686-Trp 698-Asn 703 in CGT-BS protein **(Figure 10)** corresponding to Thr 598-Gly 601-Gly 608-Leu 613-Gly 614-Trp 616-Pro 634-Trp 636-Lys 651-Trp 662-Asn 667 in CGTase protein from *B. circulans* 251 were identified to act as the raw-starch binding motif [43]. However, the other strictly conserved residues which were Gly 601, Gly 608, Leu 613, Gly 614 and Pro 634, do not directly bind to maltodextrin. Instead, they are most likely functioning as structural support for the raw-starch binding domain only [17].

3.4. Physicochemical properties of signal peptide

The physicochemical parameters of the signal peptide were analyzed using ExPASY's ProtParam tool (http://web.expasy.org/protparam/). Only one positively charged residue (Lys) was detected with the composition of 3.4%. This may contribute to the stability of the signal peptide, as the presence of a high number of positively charged amino acids (Lys and Arg) tends to destabilize the protein [50]. The instability index (II) computed using this tool also classified this signal peptide to be stable with the value of 13.83. On the other hand, the higher value of aliphatic index of 151.38 indicates the signal peptide to be thermostable. Enhanced protein secretion is associated with increased hydrophobicity levels and the length of the H-region. The aliphatic index, representing the relative volume occupied by aliphatic side chains (alanine, leucine, isoleucine, and valine), serves as a positive factor influencing both hydrophobicity and the improvement of protein thermostability [35]. In addition, the positive grand average of hydropathicity index (GRAVY) of 1.517 indicates the peptide as hydrophobic. The information on the

total hydrophobicity of the hydrophobic region of the signal peptides is crucial in determining the substrate specificity. Those with lower hydrophobicity interact with SecA, a protein that functions within complexes to "decode" targeting information in signal sequences, facilitating selective and regulated binding, while highly hydrophobic signals generally bind to ribosome-associated signal recognition particles (SRP) [51].

3.5. Prediction of extracellular protein localization

In terms of protein localization, a relatively high value (0.8174) obtained from the analysis using SignalP 5.0 **(Figure 11(a))** suggests the highest probability of the CGT-BS protein to be secreted extracellularly. The route for CGT-BS protein to be secreted out of the cytosol is predicted to be via Sec pathway, which is known as the universal general secretion pathway in bacteria [52]. This is further supported by Phobius **(Figure 11(b))** and Philius analyses data which show the detection of signal peptide and the protein localization to be non-cytoplasmic. Meanwhile, the prediction of subcellular localization of CGT-BS protein done by using Gpos-mPLoc also verifies the extracellular excretion of this protein. As indicated in the previous study, the CGTase produced by the *E. coli* harboring pQECGT-BS was translocated extracellularly with the activity of 30.371 U/mL without addition of any inducer at 12 hours of cultivation [27], indicating that the signal peptide of CGT-BS was functional in *E. coli*.

The role of the signal peptide is indispensable, particularly in ensuring the transportable folding of the protein precursor for its eventual expression into the media [53]. Ultimately, the successful enzyme translocation into the extracellular environment, aided by the signal peptide, would minimize the cost of industrial recombinant protein production due to the enhanced efficiency of the overall downstream processing [54]. Mahmud et al. [14] investigated the impact of signal peptides on CGTase secretion by *Lactococcus lactis*, employing a heterologous signal peptide (G1), an engineered signal peptide (M5), and a homologous signal peptide (Usp45). The study revealed a four-fold improvement in CGTase secretion when utilizing the G1 signal peptide compared to Usp45. **Table 1** summarizes various studies on molecular cloning of the CGTase genes aimed at improving the efficiency of extracellular recombinant protein secretion involving signal peptides. In relation to the presence of the functional signal peptide in this research, it has validated the possibility of high secretion of extracellular CGT-BS protein without excessive manipulation, making it practical for further industrial application.

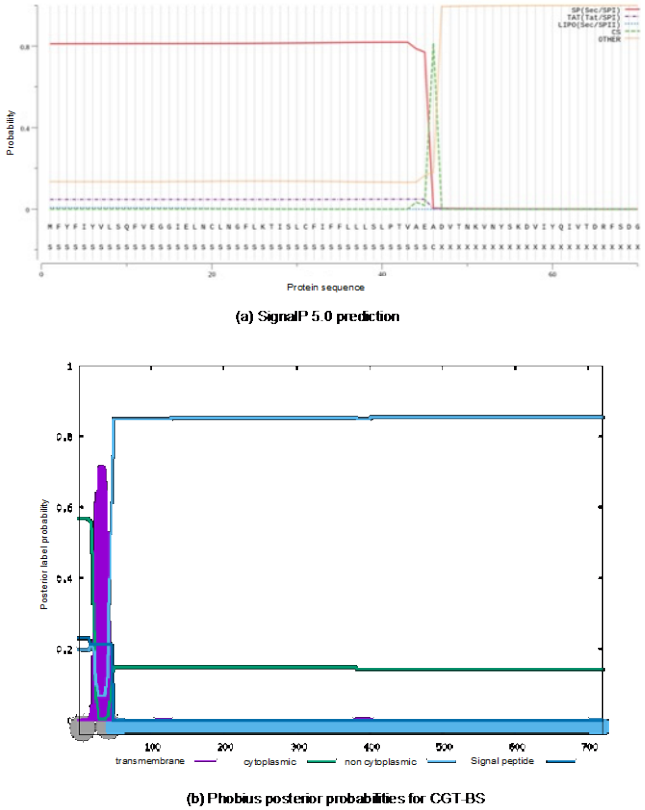

Figure 11. Localization of CGT-BS protein predicted by (a) SignalP 5.0, and (b) Phobius software.

Table 1. Translocation of recombinant CGTases into the extracellular media facilitated by functional signal peptides.

Parental strain	NCBI GenBank accession number	Cloning host	Enzyme activity (U/mL)	References
Bacillus stearothermophilus α/β-CGTase	NA	*Bacillus subtilis* WS11 and WS13	51.93 - 249.35	[12]
Bacillus sp. NR5 UPM	HQ876173.1	*Escherichia coli* JM109	38.30	[27]
Bacillus clarkii 7364	BAH14968.1	*Escherichia coli* BL21(DE3)	50.29	[54]
Paenibacillus campinasensis Hhj-1	KF143743	*Escherichia coli* BL21(DE3)	37.67	[55]
Geobacillus stearothermophilus strain No . 2	X59043.1	*Escherichia coli* BL21 (DE3)	170.60 - 213.96	[56]
Bacillus lehensis G1	AY770576	*Escherichia coli* BL21(DE3)	637.40	[57]

CONCLUSION

The thorough examination of CGT-BS protein, encompassing multiple sequence alignment and structural predictions, has elucidated its distinctive domains and key catalytic features. Secondary structure predictions revealed a balanced composition, ensuring protein stability. Notably, the catalytic domain (Domain A) and MBS in Domains B and E play crucial roles in starch hydrolysis and cyclodextrin synthesis. Additionally, the analysis of the signal peptide highlighted its stability and thermostability, essential for efficient extracellular secretion. Predictions from SignalP 5.0, Phobius, and Philius strongly support the likelihood of CGT-BS being extracellularly secreted via the Sec pathway. These insights position CGT-BS as a promising candidate for industrial applications, emphasizing its potential in

cyclodextrin production and the importance of comprehending its molecular intricacies for optimization in biotechnological processes.

ACKNOWLEDGEMENTS

This project was funded by the Fundamental Research Grant Scheme (FRGS/2/2013/SG05/UPM/02/17) provided by the Ministry of Higher Education, Malaysia.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests regarding the publication of this manuscript.

REFERENCES

1. Park, Y.J., Jeong, Y.U. and Kong, W.S. (2018) Genome sequencing and carbohydrate-active enzyme (CAZyme) repertoire of the white rot fungus *Flammulina elastica*. International Journal of Molecular Sciences, 19 (8), 2379.
2. Van der Veen, B.A., Uitdehaag, J.C., Dijkstra, B.W. and Dijkhuizen, L. (2000) The role of arginine 47 in the cyclization and coupling reactions of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251: Implications for product inhibition and product specificity. European Journal of Biochemistry, 267 (12), 3432-3441.
3. Samperio, C., Boyer, R., Eigel III, W.N., Holland, K.W., McKinney, J.S., O'Keefe, S.F., Smith, R. and Marcy, J.E. (2010) Enhancement of plant essential oils' aqueous solubility and stability using alpha and beta cyclodextrin. Journal of Agricultural and Food Chemistry, 58 (24), 12950-12956.
4. Qiu, N., Cheng, X., Wang, G., Wang, W., Wen, J., Zhang, Y., Song, H., Ma, L., Wei, Y., Peng, A. and Chen, L. (2014) Inclusion complex of barbigerone with hydroxypropyl-β-cyclodextrin: Preparation and in vitro evaluation. Carbohydrate Polymers, 101, 623-630.
5. Xu, G., Xie, X., Qin, L., Hu, X., Zhang, D., Xu, J., Li, D., Ji, X., Huang, Y., Tu, Y. and Wei, D. (2019) Simple synthesis of a swellable porous β-cyclodextrin-based polymer in the aqueous phase for the rapid removal of organic micro-pollutants from water. Green Chemistry, 21 (22), 6062-6072.
6. Campos, E.V., Proença, P.L., Oliveira, J.L., Melville, C.C., Della Vechia, J.F., De Andrade, D.J. and Fraceto, L.F. (2018) Chitosan nanoparticles functionalized with β-cyclodextrin: A promising carrier for botanical pesticides. Scientific Reports, 8 (1), 1-15.
7. Tesfai, B. T., Wu, D., Chen, S., Chen, J. and Wu, J. (2012) Strategies for enhancing extracellular secretion of recombinant cyclodextrin glucanotransferase in *E. coli*. Applied Biochemistry and Biotechnology, 167, 897-908.
8. Al-Sharawi, S.Z., Ibrahim, A.S., El-Shatoury, E.H., Gebreel, H.M. and Eldiwany, A. (2013) A new low molecular mass alkaline cyclodextrin glucanotransferase from *Amphibacillus* sp. NRC-WN isolated from an Egyptian soda lake. Electronic Journal of Biotechnology, 16 (6), 3-3.
9. Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A. and Ghasemi, Y. (2018) A comprehensive review of signal peptides: Structure, roles, and applications. European Journal of Cell Biology, 97 (6), 422-441.
10. von Heijne, G. (1990). The signal peptide. The Journal of Membrane Biology, 115, 195-201.
11. Guan, C., Cui, W., He, X., Hu, X., Xu, J., Du, G., Chen, J. and Zhou, Z. (2015) Construction and development of a novel expression system of *Streptomyces*. Protein Expression and Purification, 113, 17-22.
12. Su, L., Li, Y. and Wu, J. (2021) Efficient secretory expression of Bacillus stearothermophilus α/β-cyclodextrin glycosyltransferase in *Bacillus subtilis*. Journal of Biotechnology, 331, 74-82.
13. Ismail, A. and Illias, R.M. (2017) Site-saturation mutagenesis of mutant L-asparaginase II signal peptide hydrophobic region for improved excretion of cyclodextrin glucanotransferase. Journal of Industrial Microbiology and Biotechnology, 44 (12), 1627-1641.
14. Mahmud, H., Ismail, A., Rahim, R.A., Low, K.O. and Illias, R.M. (2019) Enhanced secretion of cyclodextrin glucanotransferase (CGTase) by *Lactococcus lactis* using heterologous signal peptides and optimization of cultivation conditions. Journal of Biotechnology, 296, 22-31.
15. Forouharmehr, A., Nassiri, M., Ghovvati, S. and Javadmanesh, A. (2018) Evaluation of different signal peptides for secretory production of recombinant bovine pancreatic ribonuclease A in Gram negative bacterial system: An in silico study. Current Proteomics, 15 (1), 24-33.
16. Kavousipour, S., Mohammadi, S., Eftekhar, E., Barazesh, M. and Morowvat, M.H. (2021) In silico investigation of signal peptide sequences to enhance secretion of CD44 nanobodies expressed in *Escherichia coli*. Current Pharmaceutical Biotechnology, 22 (9), 1192-1205.
17. Dijkhuizen, L., Penninga, D., Rozeboom, H.J., Strokopytov, B. and Dijkstra, B.W. (1995) Protein engineering of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. Progress in Biotechnology, 10, 165-174.
18. Van der Veen, B.A., Uitdehaag, J.C., Dijkstra, B.W. and Dijkhuizen, L. (2000) Engineering of cyclodextrin glycosyltransferase reaction and product specificity. Biochimica et Biophysica Acta, 1543 (336), 360.
19. Dauter, Z., Dauter, M., Brzozowski, A. M., Christensen, S., Borchert, T. V., Beier, L., Wilson, K. S., and. Davies, G. J. (1999) X-ray Structure of Novamyl, the Five-Domain "Maltogenic" α-Amylase from *Bacillus stearothermophilus*: Maltose and Acarbose Complexes at 1.7 Å Resolution. Biochemistry, 38 (26), 8385-8392.
20. Rashid, N., H. Imanaka, T. Kanai, T. Fukui, H. Atomi and T. Imanaka. (2002) A novel candidate for the true fructose-1,6-bisphosphatase in archaea. Journal of Biological Chemistry. 277:30,649–30,655.
21. Strokopytov, B., Knegtel, R. M., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L., & Dijkstra, B. W. (1996). Structure of cyclodextrin glycosyltransferase complexed with a maltononaose inhibitor at 2.6 Å resolution. Implications for product specificity. Biochemistry, 35(13), 4241-4249.
22. Han, R., Li, J., Shin, H. D., Chen, R. R., Du, G., Liu, L., & Chen, J. (2013). Carbohydrate-Binding Module–Cyclodextrin Glycosyltransferase Fusion Enables Efficient Synthesis of 2-O-D-Glucopyranosyl-L-Ascorbic Acid with Soluble Starch as the Glycosyl Donor. Applied and environmental microbiology, 79(10), 3234-3240.
23. Wind, R. D., Uitdehaag, J. C., Buitelaar, R. M., Dijkstra, B. W., & Dijkhuizen, L. (1998). Engineering of cyclodextrin product specificity and pH optima of the thermostable cyclodextrin glycosyltransferase from *Thermoanaerobacterium thermosulfurigenes* EM1. Journal of Biological Chemistry, 273(10), 5771-5779.
24. Wankar, J., Kotla, N. G., Gera, S., Rasala, S., Pandit, A., & Rochev, Y. A. (2020). Recent advances in host–guest self-assembled cyclodextrin carriers: Implications for responsive drug delivery and biomedical engineering. Advanced Functional Materials, 30(44), 1909049.
25. Crini, G., Fourmentin, S., Fenyvesi, É., Torri, G., Fourmentin, M., & Morin-Crini, N. (2018). Fundamentals and applications of cyclodextrins. Cyclodextrin fundamentals, reactivity and analysis, 1-55.
26. Bhardwaj, V. K., & Purohit, R. (2023). A comparative study on inclusion complex formation between formononetin and β-

cyclodextrin derivatives through multiscale classical and umbrella sampling simulations. Carbohydrate Polymers, 310, 120729.

27. Nik-Pa, N.I.M., Abd-Aziz, S., Ibrahim, M.F., Alitheen, N.B.M. and Ramli, N. (2019) Improved extracellular secretion of β-cyclodextrin glycosyltransferase from *Escherichia coli* by glycine supplementation without apparent cell lysis. Asia Pacific Journal of Molecular Biology and Biotechnology, 27, 93-102.
28. Nik-Pa, N.I.M., Sobri, M.F.M., Abd-Aziz, S., Ibrahim, M.F., Kamal Bahrin, E., Mohammed Alitheen, N.B. and Ramli, N. (2020) Combined optimization of codon usage and glycine supplementation enhances the extracellular production of a β-cyclodextrin glycosyltransferase from *Bacillus* sp. NR5 UPM in *Escherichia coli*. International Journal of Molecular Sciences, 21 (11), 3919.
29. Ramli, N., Abd-Aziz, S., Hassan, M.A., Mohammed Alitheen, N. and Kamaruddin, K. (2010) Potential cyclodextrin glycosyltransferase producer from locally isolated bacteria. African Journal of Biotechnology, 9 (43), 7317-7321.
30. Ravinder K, Prabhakar T, Prasanthkumar K, Venuka N (2012). Immobilization of cyclodextrin glycosyltransferase from newly isolated, mutated *Bacillus* sp. TPR71HNA6 by adsorption technique. Journal of Scientific Research in Pharmacy 1: 95–99.
31. Costa H, Gastón JR, Lara J, Martinez CO, Moriwaki C, Matioli G, Ferrarotti SA (2015). Cyclodextrin glycosyltransferase production by free cells of *Bacillus circulans* DF 9R in batch fermentation and by immobilized cells in a semi-continuous process. Bioprocess and Biosystems Engineering 38: 1055–1063
32. Stanković, S., BERIĆ, T., SIMIĆ, D., & PEŠIĆ, D. (2016). Determination of cyclodextrin production by cyclodextrin glycosyltransferase from alkalophilic *Bacillus circulans* strain B-65. Botanica Serbica, 40(1).
33. McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The PSIPRED protein structure prediction server. Bioinformatics, 16 (4), 404-405.
34. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26 (2), 283-291.
35. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. In The Proteomics Protocols Handbook, p. 571–607. Totowa, NJ, USA: Humana Press.
36. Käll, L., Krogh, A. and Sonnhammer, E. L. (2004) A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology, 338(5), 1027-1036.
37. Reynolds, S.M., Käll, L., Riffle, M.E., Bilmes, J.A. and Noble, W.S. (2008) Transmembrane topology and signal peptide prediction using dynamic bayesian networks. PLoS Computational Biology, 4 (11), e1000213.
38. Klein, C. and Schulz, G.E. (1991). Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. Journal of Molecular Biology, 217 (4), 737-750.
39. Van der Veen, B.A., van Alebeek, G.J.W., Uitdehaag, J.C., Dijkstra, B.W. and Dijkhuizen, L. (2000) The three transglycosylation reactions catalyzed by cyclodextrin glycosyltransferase from *Bacillus circulans* (strain 251) proceed via different kinetic mechanisms. European Journal of Biochemistry, 267 (3), 658-665.
40. Sonnendecker, C. and Zimmermann, W. (2019) Domain shuffling of cyclodextrin glucanotransferases for tailored product specificity and thermal stability. FEBS Open Bio, 9 (2), 384-395.
41. Goh, P.H., Illias, R.M. and Goh, K.M. (2012). Domain replacement to elucidate the role of B domain in CGTase thermostability and activity. Process Biochemistry, 47 (12), 2123-2130.
42. Leemhuis, H., Rozeboom, H.J., Dijkstra, B.W. and Dijkhuizen, L. (2004). Improved thermostability of *Bacillus circulans* cyclodextrin glycosyltransferase by the introduction of a salt bridge. Proteins: Structure, Function, and Bioinformatics, 54 (1), 128-134.
43. Lawson, C.L., van Montfort, R., Strokopytov, B., Rozeboom, H.J., Kalk, K.H., de Vries, G.E., Penninga, D., Dijkhuizen, L. and Dijkstra, B.W. (1994) Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. Journal of Molecular Biology, 236 (2), 590-600.
44. Wang, H., Zhou, W., Li, H., Rie, B. and Piao, C. (2017) Improved activity of β-cyclodextrin glycosyltransferase from *Bacillus* sp. N-227 via mutagenesis of the conserved residues. 3 Biotech, 7, 1–10.
45. Nakamura, A., Haga, K., Ogawa, S., Kuwano, K., Kimura, K. and Yamane, K. (1992) Functional relationships between cyclodextrin glucanotransferase from an alkalophilic *Bacillus* and α-amylases site-directed mutagenesis of the conserved two Asp and one Glu residues. FEBS Letters, 296 (1), 37-40.
46. Penninga, D., Strokopytov, B., Rozeboom, H.J., Lawson, C.L., Dijkstra, B.W., Bergsma, J. and Dijkhuizen, L. (1995) Site-directed mutations in tyrosine 195 of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 affect activity and product specificity. Biochemistry, 34 (10), 3368-3376.
47. Penninga, D., van der Veen, B.A., Knegtel, R.M., van Hijum, S.A., Rozeboom, H.J., Kalk, K.H., Dijkstra, B.W. and Dijkhuizen, L. (1996) The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. Journal of Biological Chemistry, 271 (51), 32777-32784.
48. Hyun-Dong, S., Park, T.H. and Lee, Y.H. (2000) Site-directed mutagenesis and functional analysis of maltose-binding site of β-cyclodextrin glucanotransferase from *Bacillus firmus* var. alkalophilus. Biotechnology Letters, 22, 115-121.
49. Chen, S., Li, Z., Gu, Z., Hong, Y., Cheng, L. and Li, C. (2019) Variants at position 603 of the CGTase from *Bacillus circulans* STB01 for reducing product inhibition. International Journal of Biological Macromolecules, 136, 460-468.
50. Anumalla, B. and Prabhu, N.P. (2019) Counteracting effect of charged amino acids against the destabilization of proteins by arginine. Applied Biochemistry and Biotechnology, 189, 541-555.
51. Clérico, E. M., Maki, J. L., & Gierasch, L. M. (2008). Use of synthetic signal sequences to explore the protein export machinery. Peptide Science, 90(3), 307-319.
52. Tsirigotaki, A., De Geyter, J., Šoštaric´, N., Economou, A. and Karamanou, S. (2017). Protein export through the bacterial Sec pathway. Nature Reviews Microbiology, 15 (1), 21-36.
53. Wu, F., Ma, J., Cha, Y., Lu, D., Li, Z., Zhuo, M., Luo, X., Li, S. and Zhu, M. (2020) Using inexpensive substrate to achieve high-level lipase A secretion by *Bacillus subtilis* through signal peptide and promoter screening. Process Biochemistry, 99, 202-210.
54. Wang, L., Chen, S. and Wu, J. (2018) Cyclodextrin enhanced the soluble expression of *Bacillus clarkii* γ-CGTase in *Escherichia coli*. BMC Biotechnology, 18, 1-9.
55. Zheng, J., Li, X. and Wu, H. (2019) High-level extracellular secretion and characterization of the thermophilic β-cyclodextrin glucanotranferase from *Paenibacillus campinasensis* in *Escherichia coli*. 3 Biotech, 9, 1-11.
56. Deng, C., Li, J., Shin, H.D., Du, G., Chen, J. and Liu, L. (2018) Efficient expression of cyclodextrin glycosyltransferase from *Geobacillus stearothermophilus* in *Escherichia coli* by promoter engineering and downstream box evolution. Journal of Biotechnology, 266, 77-83.
57. Ling, H.L., Rahmat, Z., Murad, A.M.A., Mahadi, N.M. and Illias, R.M. (2017) Proteome-based identification of signal peptides for improved secretion of recombinant cyclomaltodextrin glucanotransferase in *Escherichia coli*. Process Biochemistry, 61, 47-55.