Research Article

# A Role of Cellulose Binding Module of the Thermophilic Endoglucanase TbCel12A

Teerachai Kuntothom*
Chemistry Department, Faculty of Science, Mahasarakham University, Khamriang Campus, Mahasarakham, Thailand

Jirapa Phetsom
Department of Biology, Faculty of Science, Mahasarakham University, Maha Sarakham, Thailand

Sunanta Chuncher
Department of Biology, Faculty of Science, Mahasarakham University, Maha Sarakham, Thailand

James R. Ketudat Cairns
Center for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima, Thailand
School of Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand

* Corresponding author. E-mail: teerachai.k@msu.ac.th DOI: 10.14416/j.asep.2023.09.006
Received: 8 May 2023; Revised: 13 June 2023; Accepted: 20 July 2023; Published online: 20 September 2023


**Abstract**
Endoglucanases are enzymes that play an important role in hydrolysis of lignocellulose by attacking glycosidic linkages in cellulose fibers and other glucans. The cellulose binding module (CBM) is responsible for binding the enzyme to the substrate. However, CBMs in certain enzymes interfere with substrate hydrolysis resulting in moderate or low activity. In a previous study, the processive endoglucanase TbCel12A including its CBM had low activity towards carboxymethyl cellulose (CMC). To assess the effect of the CBM, the catalytic domain of TbCel12A was produced without the CBM. The TbCel12A catalytic domain without the CBM hydrolyzed CMC 23 times more rapidly, while the pH and temperature optima and thermotolerance remained unchanged compared to full-length TbCel12A. Therefore, TbCel12A does not require the CBM for CMC hydrolysis and its application may be improved without it.

**Keywords**: Bio circular green economic (BCG), Biorefinery, Cellulase, Cellulose binding module, Endoglucanase, Enzyme engineering

## 1 Introduction

Climate change is a global community problem, which is caused by emissions of greenhouse gases (GHGs). The main sources of GHGs emissions are industries and transportation [1]. Much of the world relies on petroleum as the primary source for energy production to serve manufacturing and transportation which are the main drivers of the economy [2]. The use of petroleum releases key GHGs, which are carbon dioxide, methane, and nitrous oxide. In the Conference of the Parties in 2021 (COP26), many countries announced the plan to achieve carbon neutrality by 2050 and net-zero greenhouse gas emissions by 2060 [2]. In order to achieve the goals, bio circular green economic (BCG) model was introduced [3]. The BCG model is designed to reduce inappropriate management of agricultural waste and efficiently utilize it for the sustainable production of biofuels, platform chemicals and high value-added products, which is in accordance

with the sustainable development goals (SDG) of United Nations [4]. This is an ideal circumstance for countries, such as Thailand which is a key agricultural supplier and thus produces tons of agricultural wastes annually. Traditionally, biomass is burned to produce GHGs so converting it to biofuel is desirable, but it is difficult to compete with petroleum fuel since production cost of biofuel is higher and has required government subsidization [5]. Besides converting to biofuel, biomass can be converted into higher value products via biorefinery [6]. The products from biorefinery such as furfural and lactic acid can be used as platform chemicals in downstream industries to produce higher value products [7].

Lignocellulose, the most abundant biomass, is a complex structure of cellulose, hemicellulose, and lignin. The cellulose and hemicellulose are wrapped in layers of the aromatic polymer lignin, which acts as a protective layer [8]. Lignocellulose digestion by cellulases or microorganisms suffers from lignin recalcitrance which reduces glucose yield [7]. Therefore, chemical pretreatment is introduced to efficiently remove lignin and hemicellulose while keeping cellulose intact for enzymatic hydrolysis [7].

Cellulases are a group of enzymes responsible for converting cellulose into monosaccharides. β-1,4-Endoglucanase randomly hydrolyzes cellulose into short chain oligosaccharides, β-1,4-exoglucanase or cellobiohydrolase attacks cellulose ends and releases cellobiose, and β-glucosidase is responsible for breaking down cellobiose and other oligosaccharides into glucose [9]. Monosaccharides such as glucose are not only used for biofuel production but can also be submitted to biorefinery to produce a wide collection of marketable chemicals [6].

Cellulases with thermotolerance or alkaline pH optima can be applied in major industries, such as food and beverages, textiles, animal feed, and biofuels. Moreover, cellulases are widely used in natural product extraction to digest cell walls and facilitate the release of active compounds [10]. The use of cellulases in combination with proteases has been shown to increase extraction yield compared to chemical extraction [11], [12].

Endoglucanases which play parts in hydrolyzing cellulose can be categorized into non-processive and processive types [13]. The processivity involves endoglucanase absorbing onto the substrate during hydrolysis, which is crucial for hydrolyzing crystalline cellulose [14]. Most endoglucanases consist of a catalytic domain, which is responsible for biomass hydrolysis, linked by a flexible peptide to a cellulose binding module (CBM) [15]. Numerous CBMs have been identified and classified into 97 families based on amino acid sequence similarities [16]. The functional role of the CBM is to assist in binding and hydrolyzing substrates, which comprise of homogenous and heterogenous polysaccharides, especially insoluble substrates, by holding the enzyme in proximity to the substrate, targeting specific substrate regions and sometimes disrupting the substrate structure [17], [18]. However, the CBM may not be crucial for the hydrolysis of soluble substrates by endoglucanases. For instance, Cel5L, EG5C-1 and GsCelA are glycoside hydrolase family 5 (GH5) endoglucanases that demonstrated increased carboxymethylcellulose (CMC) hydrolysis after removal of CBMs at their C-termini [19]–[21].

A previous study on TbCel12A, a thermophilic processive endoglucanase with a CBM, showed a low hydrolysis rate towards CMC [22]. For that reason, the catalytic domain of TbCel12A ($CAT_{TbCel}$) was produced without the CBM to assess whether removing the CBM increases substrate hydrolysis and how it affects thermophilicity and function of TbCel12A in the current study.

## 2 Materials and Methods

### 2.1 *Protein sequence analysis and homology modeling*

The amino acid sequence of TbCel12A (Genbank accession number QPB77271.1) was aligned to those of thermophilic endoglucanases from *Acidothermus cellulolyticus* (PDB code 7MKS), *Aspergillus fischeri* (PDB code 6K98), *Bacillus lichenformis* (PDB code 2JEM), *Streptomyces lividans* (PDB code 2NLR), and *Streptomyces* sp. (PDB code 1OA4) by the Clustal Omega multiple sequence alignment program on the European Bioinformatics Institute (EBI) website to identify the catalytic domain and CBM [23], [24]. The amino acid sequence of the putative CBM was further analyzed by SMART to determine the catalytic domain and CBM [25].

The cellulose binding module of TbCel12A ($CBM_{TbCel}$) was modeled with CBM2a of *C. fimi* as

a template (PDB code 1EXG) in Alphafold2 at the Alpha-Fold Google Colaboratory [26], [27]. Pymol [28] was used to superimpose the homology model of $CBM_{TbCel}$ onto the structure of CBM2a (PDB code 1EXG) and CBM2b (PDB codes 1XBD) of *C. fimi* to analyze and compare the substrate binding amino acids.

### 2.2 *Gene synthesis and cloning*

The amino acid sequence of $CAT_{TbCel}$ was submitted to Genscript (Piscataway, New Jersey, USA.) for gene synthesis with *Escherichia coli* codon optimization and cloning into pET32a (Novagen, Merck KGa, Damstadt, Germany) between the NcoI and XhoI restriction sites.

### 2.3 *Recombinant protein production and purification*

The pET32a/$CAT_{TbCel}$ expression plasmid was introduced into Origami B(DE3) for expression and purification of recombinant $CAT_{TbCel}$ as previously reported for the full-length TbCel12A protein [29]. Full-length TbCel12A was produced and purified as previously described [22], [29]. Briefly, the expression construct was transformed into competent Origami B(DE3) cells (Novagen, Merck KGa) and selected on LB agar containing 50 µg/mL ampicillin, 15 µg/mL kanamycin, and 12.5 µg/mL tetracyclin. A single colony was selected, grown, and induced for expression with 0.2 mM IPTG. After overnight expression, *E. coli* cells were collected by centrifugation. Collected cells were lysed by lysis buffer [29]. The cell suspension was centrifuged to separate the supernatant from the cell debris. The supernatant containing recombinant $CAT_{TbCel}$ was transferred to a new tube and incubated at 60 °C to precipitate *E. coli* proteins and centrifuged. After centrifugation, recombinant $CAT_{TbCel}$ was purified by immobilized metal affinity chromatography (IMAC) using IMAC Sepharose 6 Fast Flow resin (Cytiva, Marlborough, MA, USA). The purity of purified TbCel12A was assessed by 12% SDS-PAGE.

### 2.4 *Characterization of $CAT_{TbCel}$*

For convenience, a colorimetric assay was used for studying general enzymatic properties that are not specific to polysaccharide hydrolysis. Enzyme activity was determined by hydrolysis of 1 mM *p*-nitrophenyl (*p*NP) -cellobioside (*p*NPC) in 50 mM sodium acetate pH 5.5 at 60 °C for 30 min, unless otherwise stated. The enzyme activity was stopped by adding 1 volume of 2 M sodium carbonate and absorbance at 405 nm was determined. The optimum pH of recombinant $CAT_{TbCel}$ was determined by incubating 10 µg enzyme and 1 mM *p*NPC in McIlvaine buffers from pH 3–8 for 30 min before stopping [30]. The optimum temperature of $CAT_{TbCel}$ was determined by incubating 10 µg enzyme and 1 mM *p*NPC in McIlvaine buffer pH 5.5 at 35–85 °C in 5 °C increments for 30 min before stopping the reaction as described above. The enzyme activity assays were performed in triplicate. The average enzyme activity and standard deviations were determined and plotted on a graph using Prism (GraphPad Software, Boston, MA, USA).

### 2.5 *Thermostability of $CAT_{TbCel}$*

The thermostability of $CAT_{TbCel}$ was determined by incubating 10 µg enzyme in McIlvaine buffer pH 5.5 at 50–80 °C for 30 min and determining enzyme activity at 60 °C, as described in section 2.4.

### 2.6 *Tolerance to heavy metals*

The tolerance of $CAT_{TbCel}$ to heavy metals was evaluated by incubating 10 µg enzyme in 1 mM *p*NPC reaction containing 5 mM of heavy metals $Co^{2+}$($CoCl_2$), $Cu^{2+}$($CuSO_4$), $Hg^{2+}$ ($HgCl_2$), $K^+$ (KCl), $Fe^{2+}$ ($FeSO_4$), $Mg^{2+}$ ($MgCl_2$), $Mn^{2+}$ ($MnCl_2$), $Ni^{2+}$ ($NiSO_4$), or $Zn^{2+}$ ($ZnSO_4$) for 30 min. The reactions were stopped by adding 1 volume of 2 M sodium carbonate and absorbance at 405 nm was determined.

### 2.7 *Tolerance to alcohols*

The tolerance of $CAT_{TbCel}$ to ethanol and methanol was determined by incubating 10 µg enzyme in 1 mM pNPC reactions containing 0–50% (v/v) alcohol for 30 min. The reactions were stopped with 2 M sodium carbonate and the absorbance at 405 nm was determined.

### 2.8 *Carboxymethyl cellulose (CMC) hydrolysis*

The activities of $CAT_{TbCel}$ and full-length TbCel12A towards carboxymethyl cellulose were determined by

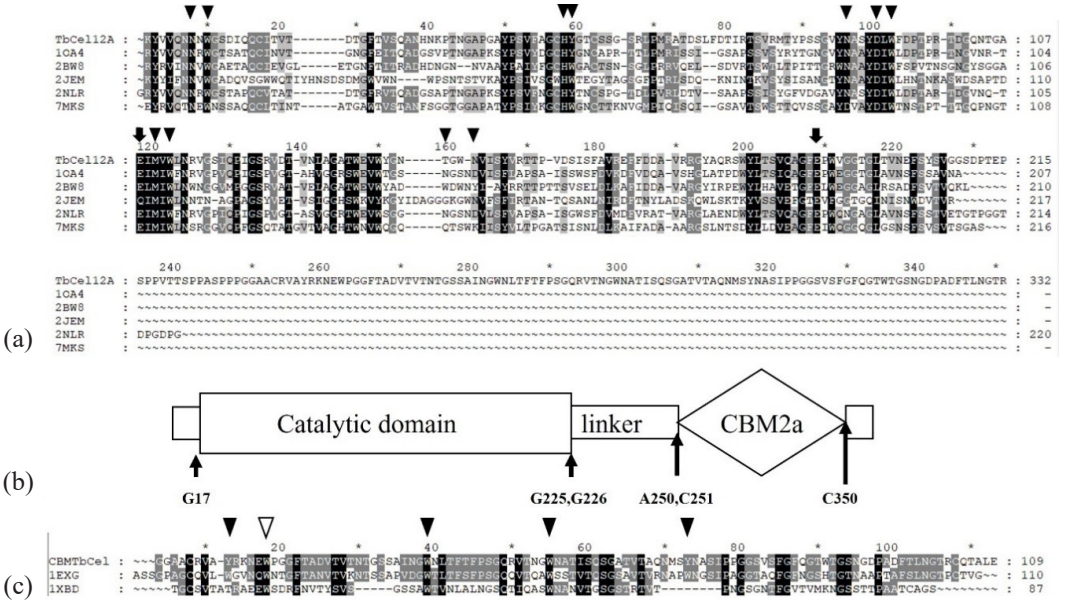


**Figure 1**: Sequence analysis of TbCel12A. (a) Multiple sequence alignment of TbCel12A and thermotolerant endoglucanases from *Acidothermus cellulolyticus* (PDB code 7MKS), *Aspergillus fischeri* (PDB code 6K98), *Bacillus lichenformis* (PDB code 2JEM), Rhodothermus marinus (PDB codes 2BW8), *Streptomyces lividans* (PDB codes 2NLR), and *Streptomyces* sp. (PDB code 1OA4). The black arrows indicate catalytic amino acids and filled triangles indicate substrate binding amino acids. (b) Domain arrangement of TbCel12A by SMART analysis. The amino terminus of TbCel12A contains the catalytic domain which is linked to a CBM2a domain at the carboxylic terminus by a flexible peptide linker. (c) The amino acid sequence of $CBM_{TbCel}$ was aligned to $CBD_{cex}$ from *C. fimi* (PDB code 1EXG) and CBM2b of Xyn10A (PDB code 1XBD). Black triangles indicate aromatic amino acids involved in substrate binding. The white triangle indicates the tryptophan of the CBM2a cellulose binding module that lies parallel on the CBMa surface, but in CBM2b xylan binding module is perpendicular to the CBMb surface.

incubating 10 μg enzyme in 1% (w/v) CMC in 50 mM sodium acetate buffer pH 5.5 for 60 min. The amount of reducing sugar was determined by 3,5-dinitrosalicylic acid (DNS) assay [31].

## 3 Results and Discussion

### 3.1 *Protein sequence analysis*

Protein sequence analysis showed that $CAT_{TbCel}$ is closely related to endoglucanases of *S. lividans* (PDB code 2NLR) and *Streptomyces* sp. (PDB code 1OA4) with 54.94% and 54.75% amino acid sequence identity, respectively (Figure 1). Glutamates 109 and 138 were conserved with the catalytic amino acids among glycoside hydrolase family 12 endoglucanases. The amino acids that are responsible for substrate binding are also conserved among the enzymes.

Analysis of TbCel12A amino acid sequence using SMART showed that full length TbCel12A comprises a catalytic domain, from Gly17 to Gly225, a flexible peptide linker from Gly226 to Ala250, and a CBM2a domain from Cys251 to Cys350 (Figure 1). The domain arrangement of TbCel12A coincides with a small number of GH12 enzymes containing a CBM in which the catalytic domain is connected to a single CBM2 domain at the carboxyl terminus by a short peptide linker. Multiple sequence alignment of the $CBM_{TbCel}$, and CBM2a, and CBM2b of *Cellulomonas fimi* (PDB code 1EXG) showed that they share approximately 55% identity (Figure 1). The amino acid sequence of $CBM_{TbCel}$ showed roughly 50% or less similarity with other CBM2, which is normal because CBMs have diverse amino acid sequences [18].

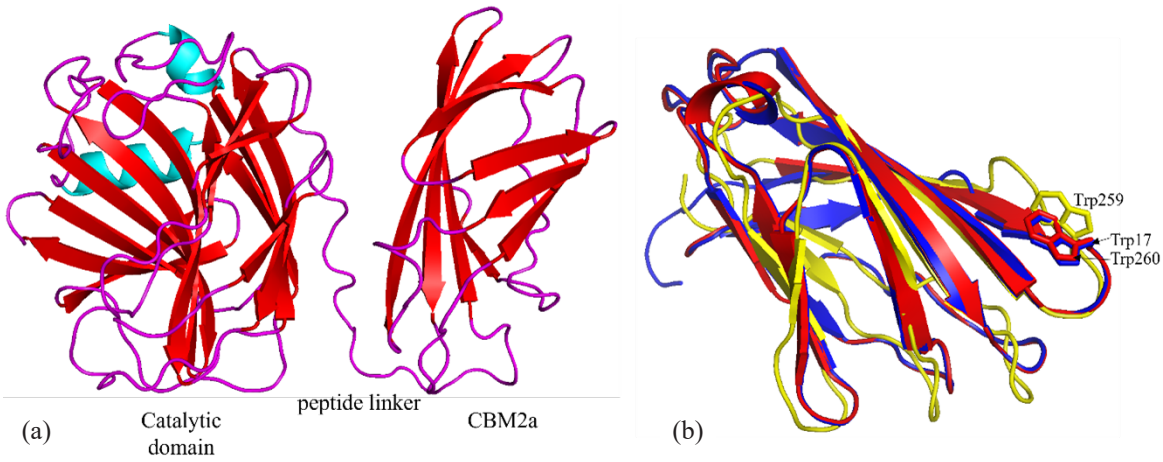


**Figure 2**: Homology model of full-length TbCel12A containing a catalytic domain which is linked to a CBM2a domain by a flexible peptide linker (a). Superposition of $CBM_{TbCel}$ homology model (red) onto the *C. fimi* CBM2a (PDB code 1EXG) and *C. fimi* CBM2b (PDB code 1XBD) crystal structures, which are shown in blue and yellow, respectively (b). Trp17 of *C. fimi* CBM2a and Trp260 of $CBM_{TbCel}$ is parallel while the corresponding Trp259 of *C. fimi* xylanase is perpendicular on the CBM surface.

### 3.2 *Analysis and comparison of TbCel12A CBM homology model*

There are several structural folds of CBM, which can be broken into 3 types according to their binding mode: type A presents a flat surface to bind to insoluble polysaccharide surfaces, type B has a groove to bind to soluble fibers, and type C binds specifically to a small group of monosaccharides [18]. The structure of $CBM_{TbCel}$ is canonical β-jelly-roll as found in all CBM2 family members (Figure 2). The CBM2 family contains both type A (CBM2a) and type B (CBM2b) structures. Amino acid sequence alignment of $CBM_{TbCel}$ with CBM2 domains from *C. fimi* endoglucanase and xylanase showed that $CBM_{TbCel}$ has conserved aromatic amino acids (Tyr255, Trp260, Trp281, Trp297 and Tyr315), which structurally correspond to five tryptophan residues of the insoluble cellulose-binding type A CBM CBDcex [32], and tryptophan and tyrosine residues in CBM2a of endoglucanase D (EngD) [33], as shown in Figure 1C. Aromatic amino acids play major roles in the binding of CBM to insoluble substrates, as their dispersed pi electrons interact favorably with the weakly polarized axial hydrogens on the faces of sugar rings [18]. A structural comparison of CBM2a and CBM2b shows that Trp260 of $CBM_{TbCel}$ and Trp17 of $CBD_{cex}$ of CBM2a are specific to cellulose because they lie flat on the CBM surface to interact with the flat surface of crystalline cellulose [Figure 2(b)]. The corresponding Trp259 of Xyl10A xylanase CBM2b, which is specific to xylan, is perpendicular to the CBM surface, where it helps to form a groove for xylan fiber binding [34]. This is due to the presence of arginine at residue 262 in Xyl10A vs. glycine in this position in CBM2a proteins since the R262G mutant of CBM2b lost xylan binding activity and gained cellulose binding [34]. As seen in Figure 1(c), $CBM_{TbCel}$ has glycine (G) in the corresponding position in addition to the conserved aromatic residues found in CBM2a, it appears to be a type A CBM. Since type A CBMs enhance binding to insoluble polysaccharide but not to soluble polysaccharides [18], the effect of its removal on the hydrolysis of soluble CMC polysaccharide was explored.

### 3.3 *Enzyme expression and extraction*

Recombinant $CAT_{TbCel}$ was produced in *E. coli* strain Origami B(DE3) with an N-terminal thioredoxin and $His_6$ fusion tag. The expressed TbCel12A catalytic domain was purified from other proteins and had a molecular weight of 45 kDa (Figure 3). Inducing protein expression with 0.2–0.6 mM IPTG produced

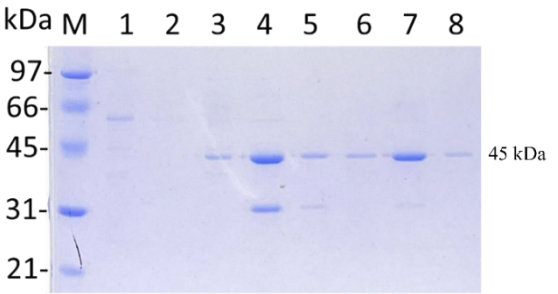


**Figure 3**: Recombinant $CAT_{TbCel}$ that was purified by immobilized metal ion affinity chromatography (IMAC) and separated on 12% SDS-PAGE. Lane M : low molecular weight protein standards, lane 1 : soluble proteins from bacteria expressing original TbCel12A, lane 2 : full-length TbCel12A purified by IMAC, lane 3 : soluble proteins of bacteria expressing $CAT_{TbCel}$ induced with 0.2 mM IPTG, lane 4 : purified $CAT_{TbCel}$ from bacteria induced with 0.2 mM IPTG, lane 5 : soluble proteins of bacteria expressing the $CAT_{TbCel}$ induced with 0.4 mM IPTG, lane 6 : purified $CAT_{TbCel}$ from induction with 0.4 mM IPTG, lane 7 : soluble proteins of bacteria expressing $CAT_{TbCel}$ induced with 0.6 mM IPTG, lane 8 : purified $CAT_{TbCel}$ from induction with 0.6 mM IPTG.

an active enzyme. The optimal IPTG concentration for protein expression induction was 0.2 mM, which produced an enzyme that showed the highest specific activity towards *p*NPC. An obvious contaminant was seen in the purified protein at approx. 31 kDa, which does not correspond to the mass of any of the common *E. coli* proteins that co-purify by IMAC [35] and was too large for protein that had lost its thioredoxin-fusion tag or for the tag itself. The 31 kDa protein could likely be inactive $CAT_{TbCel}$ that was cleaved in the middle by a protease during protein extraction, which would lack one of its catalytic residues.

### 3.4 *Characterization of recombinant TbCel12A*

The $CAT_{TbCel}$ had optimal conditions of pH 5.5 at 65 °C (Figure 4). The optimal temperature of the catalytic domain remained the same as the full-length enzyme [22]. The $CAT_{TbCel}$'s optimal pH of 5.0–5.5 of is a common range seen in glycoside hydrolases [15]. The optimal temperature of $CAT_{TbCel}$ at 65 °C is higher than certain fungal endoglucanases at 40 °C [15] and it

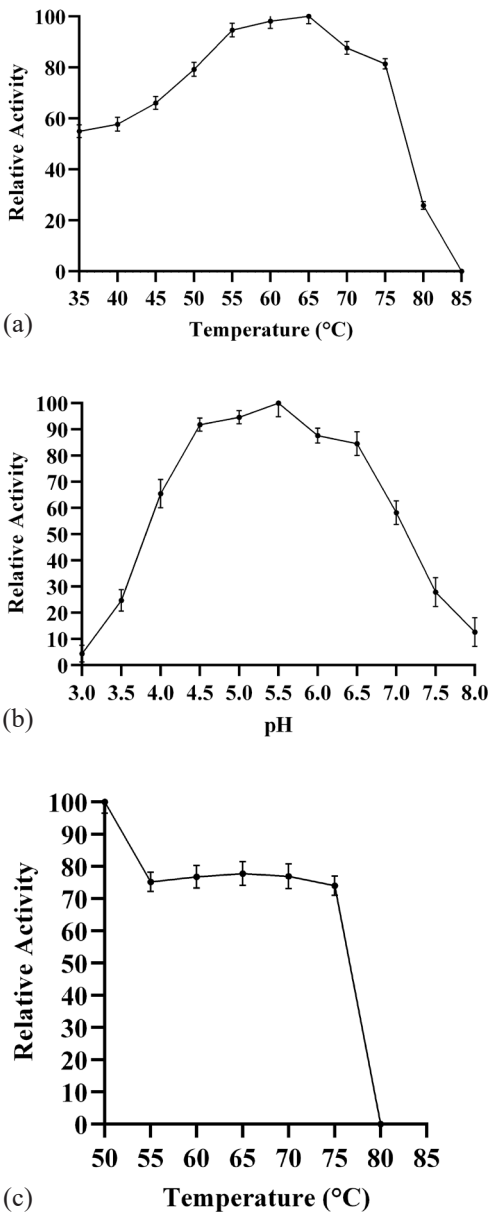


**Figure 4:** Determination of the optimum working temperature (a), pH (b), and thermotolerance (c) of $CAT_{TbCel}$. Temperatures were varied for reactions in McIlvaine buffer, pH 5.5, $CAT_{TbCel}$ was incubated for 30 min at the specified temperature. Thermotolerance of $CAT_{TbCel}$ was assayed by incubating the enzyme at the specified temperatures for 30 min followed by determining activity towards *p*NPC in 30 min reactions under optimum conditions.

coincided with that of GuxA, a GH12 endoglucanase from the thermophilic bacterium A. cellulolyticus, with optimal temperatures at 75 °C [36]. Clearly, the CBM does not play a role in determining the pH and temperature optima of the catalytic domain, consistent with its independent role in binding insoluble substrate [18].

### 3.5 *Thermostability of recombinant $CAT_{TbCel}$*

The thermostability of $CAT_{TbCel}$ showed that it retained 80% activity toward *p*NPC between 50–70 °C and the activity drastically decreased at 80 °C (Figure 4). The thermostability of the catalytic domain alone was observed in Cel5A and Cel5L studies. The native Cel5a and Cel5a-CBM6 fusion protein remained fully active at 80 °C for 16 h [37]. Furthermore, Cel5L-p35, a catalytic domain lacking CBM, retained 20% relative activity at 70 °C after 1 h, while Cel5L-p50 containing a CBM3 domain lost nearly all activity after 40 min of incubation [21]. Therefore, CBMs may not increase thermostability of certain endoglucanases.

### 3.6 *Tolerance to heavy metals*

Previously, it was shown that full-length TbCel12A could tolerate certain metal ions, but was not tolerant of $Hg^{2+}$, $Fe^{2+}$ and $Cu^{2+}$ [22]. When incubating with the same metal ions, $CAT_{TbCel}$ showed the same activity profile as full-length TbCel12A (Figure 5). Manganese enhanced enzyme activity, $Co^{2+}$, $K^{+}$, $Mg^{2+}$, $Ni^{2+}$, and $Zn^{2+}$ slightly decreased enzyme activity by less than one-fifth, and $Cu^{2+}$ and $Fe^{2+}$ inhibited enzyme by one-third and two-thirds, respectively. $Hg^{2+}$ completely inhibited enzyme activity, which is common in glycoside hydrolase enzymes [38]. The effects of metal ions were studied in the GH5 endoglucanase Cel5L, in which the full-length enzyme and the catalytic domain (Cel5L-p35) had the same activity profile in 5 mM metal ions [21]. Divalent ions can either promote or inhibit endoglucanase activity by redox effects on amino acids [39]. Cobalt slightly inhibited $CAT_{TbCel}$ but drastically inhibited Cel5L-p35 activity to 60%. The activity of $CAT_{TbCel}$ was slightly inhibited by $K^{+}$, $Mg^{2+}$, and $Zn^{2+}$. This was the opposite of the effects on Cel5L-p35 in which enzyme activity was increased up to 117% by $K^{+}$ and $Mg^{2+}$ but was not affected by $Zn^{2+}$. Copper ion greatly enhanced Cel5L-p35 activity

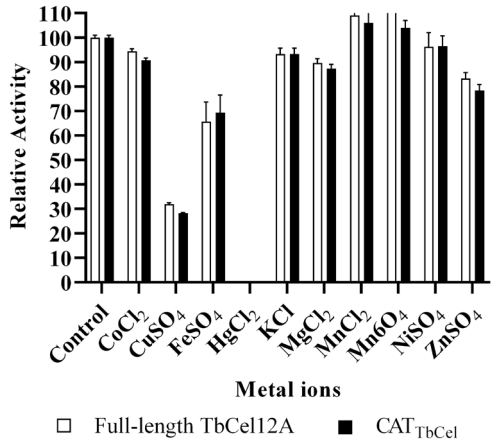


**Figure 5**: The comparison of relative activities between full-length TbCel12A and $CAT_{TbCel}$ in the presence of metal ions. The full-length and $CAT_{TbCel}$ activities towards *p*NPC were determined in the presence of metal ions at optimum conditions.

to 140% but drastically reduced $CAT_{TbCel}$ activity to 30%. Manganese promoted activity of $CAT_{TbCel}$ to 105% but inhibited Cel5L-p35 by 50%. It is clear that metal ion effects on activity in this assay only reflect the catalytic domain and attachment to the linker and CBM did not affect them.

### 3.7 *Tolerance to alcohols*

Cellulolytic enzymes are used in saccharification of biomass, in which solvents may be left from pretreatment, while simultaneous saccharification and fermentation is another application in which alcohol builds up over time [40]. So, it was of interest to see whether the CBM affected tolerance to alcohol.

The activities of $CAT_{TbCel}$ and full-length TbCel12A were not affected by methanol and ethanol at 10% (v/v). At 20% (v/v), ethanol inhibited the enzyme activity by 50%, while methanol did not. At 30% (v/v), ethanol almost completely inhibited $CAT_{TbCel}$ activity. This was superior to endoglcanase from *A. niger*, which relative activity sharply dropped to 60% at 10% (v/v) ethanol and 45% at 16% (v/v) ethanol [41]. Methanol at 30% (v/v) sharply suppressed $CAT_{TbCel}$ activity to one-fifth and enzyme activity was completely inhibited at 40% (v/v), as shown in Figure 6. Increasing concentrations of alcohols can form interactions with hydrophobic side chains of amino

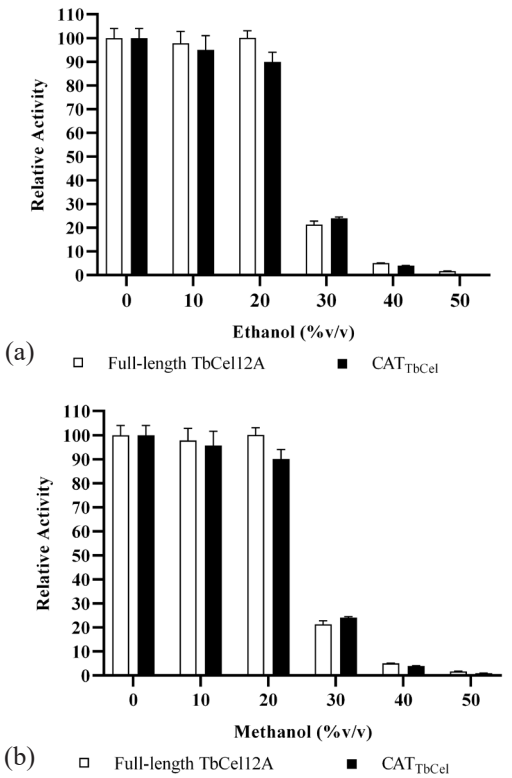


**Figure 6**: Relative activity comparison of full-length TbCel12A and $CAT_{TbCel}$ in the presence of various concentrations of ethanol and methanol. The $CAT_{TbCel}$ activity for *p*NPC was determined in sodium acetate buffer pH 5.5 containing 0–50% (v/v) ethanol or methanol at 60 °C for 30 min.

acid causing disruption of protein tertiary structure, precipitation, and loss of enzyme activity [42].

### 3.8 *Carboxymethyl cellulose (CMC) hydrolysis*

The activities of full-length TbCel12A and its catalytic domain towards CMC were determined. It was found that $CAT_{TbCel}$ had a CMC hydrolysis rate 23 times higher than the full-length enzyme (Figure 7). Increased enzyme activity resulting from removal of the CBM was also found in Cel5L, EG5C, and GsCelA endoglucanases, for which the truncated enzyme lacking the CBM had increased activity towards substrate without losing thermostability [16]–[18]. For GsCelA, truncation of the C-terminus by 60 amino acids resulted in an increase of about 2.5-fold in activity

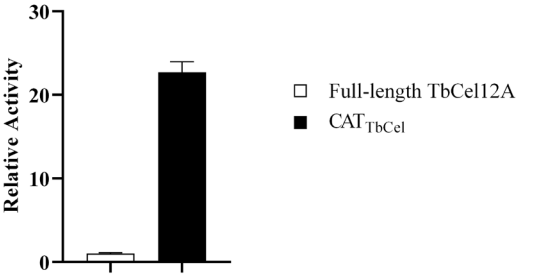


**Figure 7**: Relative activity comparison of full-length TbCel12A and $CAT_{TbCel}$ towards carboxymethyl cellulose (CMC). The enzymes were incubated with 1 (%w/v) CMC at pH 5.5 and 60 °C for 60 min and then the amount of reducing sugar was determined by DNS assay.

on CMC and about 2-fold in activity on phosphoric acid swollen cellulose (PASC). This part of the protein was lost automatically upon storage and the authors considered that release of the C-terminal CBM might be part of the biological function of the domain. After acting on an area of insoluble cellulose for some time, the catalytic domain could be released in a more active form to act on generated soluble fragments. However, the authors only observed this auto-truncation in GH5 enzymes and not the GH12 endoglucanase that they tested. The current report is the first to show that a GH12 enzyme can be activated by removing its CBM.

Most of GH12 endoglucanase do not have CBMs but are active towards substrates, therefore, removing CBM from $CAT_{TbCel}$ does not decrease its activity on soluble substrates [43]. Since the $CAT_{TbCel}$ did not lose thermostability or show other changes in chemical tolerance or enzymatic properties, it is clearly an improved enzyme for a breakdown of soluble biomass. Future studies are needed to see whether the C-terminal CBM is useful for breaking down insoluble cellulosic biomass or whether a mixture of TbCel12A with and without this domain would increase saccharification of such insoluble biomass substrates.

## 4 Conclusions

TbCel12A, a thermophilic endoglucanase, was analyzed for the cellulose binding module by NCBI BLAST and sequence comparison to thermophilic endoglucanases with experimentally resolved structures. The catalytic domain of TbCel12A, $CAT_{TbCel}$, was

identified as 221 amino acid residues at the amino-terminal end and an expression construct encoding this domain of TbCel12A without the CBM was synthesized. The $CAT_{TbCel}$ was recombinantly expressed to produce the active enzyme, which $CAT_{TbCel}$ retained the full-length protein's optimum temperature at 65 °C and optimum pH at 5.5. The removal of the cellulose binding module increased activity on CMC by twenty-three-fold. Compared to full-length TbCel12A, and the $CAT_{TbCel}$ had similar thermostability and tolerance to metals and alcohols, compared to full-length TbCel12A. In this study, it can be concluded that the CBM of TbCel12A was not required for enzyme function on soluble substrates, as shown by the activity of the catalytic domain alone. Moreover, removing the CBM can increase enzyme activity on CMC.

**Acknowledgments**

We would like to thank supporting staff at Mahasarakham University, Suranaree University of Technology and JKC lab members for helping with equipment usage. This research project was financially supported by Thailand Science Research and Innovation (TSRI).

**Author Contributions**

T.K.: funding acquisition, research design, analysis, data experiments, data analysis, writing original draft; J.P.: research idea, data analysis; S.C.: research idea, data analysis; J.K.C.: conceptualization, reviewing and editing, funding acquisition, project administration. All authors have read and agreed to the published version of the manuscript.

**Conflicts of Interest**

The authors declare no conflict of interest.

**References**

[1] M. Crippa, D. Guizzardi, E. Solazzo, M. Muntean, E. Schaaf, F. Monforti-Ferrario, M. Banja, J. G. J. Olivier, G. Grassi, S. Rossi, E. Vignati, "GHG emissions of all world countries," 2021. [Online]. Available: https://edgar.jrc.ec.europa.eu/report_2021

[2] IEA, "Outlook for energy demand," in *World Energy Outlook 2022*. Paris, France: OECD Publishing, 2022.

[3] F. G. Santeramo, "Circular and green economy: The state-of-the-art," *Heliyon*, vol. 8, no. 4, Apr. 2022, doi: 10.1016/j.heliyon.2022.e09297.

[4] T. Ruensodsai and M. Sriariyanun, "Sustainable development and progress of lignocellulose conversion to platform chemicals," *The Journal of King Mongkut's University of Technology North Bangkok*, vol. 32, no. 4, pp. 815–818, Oct. 2022, doi: 10.14416/j.kmutnb.2022.03.001

[5] A. P. Saravanan, A. Pugazhendhi, and T. Mathimani, "A comprehensive assessment of biofuel policies in the BRICS nations: Implementation, blending target and gaps," *Fuel*, vol. 272, Apr. 2020, Art. no. 117635, doi: 10.1016/j.fuel.2020.117635.

[6] M. Sriariyanun and K. Kitsubthawee, "Trend in lignocellosic biorefinery for production of value-added biochrmicals," *Applied Science and Engineering Progress*, vol. 13, no. 4, pp. 283–284, Oct. 2020, doi: 10.14416/j.asep.2020.02.005.

[7] M. P. Gundupalli and M. Sriariyanun, "Recent trends and updates for chemical pretreatment of lignocellulosic biomass," *Applied Science and Engineering Progress*, vol. 16, no. 1, pp. 1–4, Jan. 2023, doi: 10.14416/j.asep.2022.03.002.

[8] G. Brodeur, E. Yau, K. Badal, J. Collier, K. B. Ramachandran, and S. Ramakrishnan, "Chemical and physicochemical pretreatment of lignocellulosic biomass: A review," *Enzyme Research*, vol. 2011, no. 1, May 2011, doi: 10.4061/2011/787532.

[9] Y. H. Lee and L. T. Fan, "Properties and mode of action of cellulase," in *Advances in Biochemical Engineering, vol. 17*, Berlin, Heidelberg: Springer, pp. 101–129, 1980, doi: 10.1007/3-540-09955-7_9.

[10] Y. Hu, G. Kang, L. Wang, M. Gao, P. Wang, D. Yang, and H. Huang, "Current status of mining, modification, and application of cellulases in bioactive substance extraction," *Current Issues in Molecular Biology*, vol. 43, pp. 687–703, Sep. 2021, doi: 10.3390/cimb43020050.

[11] K. H. Chang, H. S. Jee, N. K. Lee, S. H. Park, N. W. Lee, and H. D. Paik, "Optimization of

the enzymatic production of 20(S)-ginsenoside Rg3 from white ginseng extract using response surface methodology," *New Biotechnology*, vol. 26, no. 3, pp. 181–186, Oct. 2009, doi: 10.1016/j.nbt.2009.08.011.

[12] W. Winotapun, P. Opanasopit, T. Ngawhirunpat, and T. Rojanarata, "One-enzyme catalyzed simultaneous plant cell disruption and conversion of released glycoside to aglycone combined with in situ product separation as green one-pot production of genipin from gardenia fruit," *Enzyme and Microbial Technology*, vol. 53, no. 2, pp. 92–96, Jul. 2013, doi: 10.1016/j.enzmictec.2013.05.001.

[13] S. Wu and S. Wu, "Processivity and the mechanisms of processive endoglucanases," *Applied Biochemistry and Biotechnology*, vol. 190, no. 2, pp. 448–463, Feb. 2020, doi: 10.1007/s12010-019-03096-w.

[14] Z. Wang, T. Zhang, L. Long, and S. Ding, "Altering the linker in processive GH5 endoglucanase 1 modulates lignin binding and catalytic properties," *Biotechnology for Biofuels*, vol. 11, no. 1, pp. 1–11, Dec. 2018, doi: 10.1186/s13068-018-1333-3.

[15] M. Sandgren, J. Ståhlberg, and C. Mitchinson, "Structural and biochemical studies of GH family 12 cellulases: improved thermal stability, and ligand complexes," *Progress in Biophysics and Molecular Biology*, vol. 89, no. 3, pp. 246–291, doi: 10.1016/j.pbiomolbio.2004.11.002.

[16] E. Drula, M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat, and N. Terrapon, "The carbohydrate-active enzyme database: Functions and literature" *Nucleic Acids Research*, vol. 50, no. D1, pp. D571–D577, Nov. 2021, doi: 10.1093/nar/gkab1045.

[17] G. Carrard, A. Koivula, H. Söderlund, and P. Béguin, "Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 19, pp. 10342–10347, Sep. 2000, doi: 10.1073/pnas.160216697.

[18] A. B. Boraston, D. N. Bolam, H. J. Gilbert, and G. J. Davies, "Carbohydrate-binding modules: Fine-tuning polysaccharide recognition," *Biochemical Journal*, vol. 382, no. Pt 3, pp. 769–781, Sep. 2004, doi: 10.1042/BJ20040892.

[19] B. Wu, S. Zheng, M. M. Pedroso, L. W. Guddat, S. Chang, B. He, and G. Schenk, "Processivity and enzymatic mechanism of a multifunctional family 5 endoglucanase from Bacillus subtilis BS-5 with potential applications in the saccharification of cellulosic substrates," *Biotechnology for Biofuels and Bioproducts*, vol. 11, no. 1, Jan. 2018, doi: 10.1186/s13068-018-1022-2.

[20] M. H. Wu, M. R. Kao, C. W. Li, S. M. Yu, and T. H. D. Ho, "A unique self-truncation of bacterial GH5 endoglucanases leads to enhanced activity and thermostability" *BMC Biology*, vol. 20, no.1, Dec. 2022, doi: 10.1186/s12915-022-01334-y.

[21] J. P. Lee, E. S. Shin, M. Y. Cho, K. D. Lee, and H. Kim, "Roles of Carbohydrate-Binding Module (CBM) of an Endo-β-1,4-Glucanase (Cel5L) from Bacillus sp. KD1014 in Thermostability and Small-Substrate Hydrolyzing Activity," *Journal of Microbiology and Biotechnology*, vol. 28, no. 12, pp. 2036–2045, Dec. 2018, doi: 10.4014/JMB.1810.10001.

[22] T. Kuntothom and J. K. Cairns, "Expression and characterization of TbCel12A, a thermophilic endoglucanase with potential in biomass hydrolysis," *Biocatalysis and Agricultural Biotechnology*, vol. 30, Nov. 2020, doi: 10.1016/j.bcab.2020.101835.

[23] F. Sievers, A. Wilm, D. G. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Molecular Systems Biology*, vol. 7, Oct. 2011, doi: 10.1038/msb.2011.75.

[24] M. Goujon, H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, and R. Lopez, "A new bioinformatics analysis tools framework at EMBL-EBI," *Nucleic Acids Research*, vol. 38, pp. W695–W699, May 2010, doi: 10.1093/nar/gkq313.

[25] I. Letunic, S. Khedkar, and P. Bork, "SMART: recent updates, new developments and status in 2020," *Nucleic Acids Research*, vol. 49, no. D1, pp. D458–D460, Jan. 2021, doi.org/10.1093/nar/gkaa937.

[26] J. Jumper, R. Evans, A. Pritzel, T. Green,

M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis , "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/S41586-021-03819-2.

[27] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, "ColabFold: Making protein folding accessible to all," *Nature Methods*, vol. 19, no. 6, pp. 679–682, Jun. 2022, doi: 10.1038/S41592-022-01488-1.

[28] L. Schrödinger and W. DeLanoz, "PyMOL," 2020. [Online]. Available: http://www.pymol.org/pymol

[29] T. Kuntothom, S. Luang, A.J. Harvey, G.B. Fincher, R. Opassiri, M. Hrmova, and J. K. Cairns, "Rice family GH1 glycoside hydrolases with β-d-glucosidase and β-d-mannosidase activities," *Archives of Biochemistry and Biophysics*, vol. 491, pp. 85–95, Nov. 2009, doi: 10.1016/j.abb.2009.09.004.

[30] T. C. McIlvaine, "A buffer solution for colorimetric comparison," *Journal of Biological Chemistry*, vol. 49, no. 1, pp. 183–186, Nov. 1921, doi: 10.1016/S0021-9258(18)86000-8.

[31] G. L. Miller, "Use of dinitrosalicylic acid reagent for determination of reducing sugar," *Analytical Chemistry*, vol. 31, pp. 426–428, Mar. 1959, doi: 10.1021/ac60147a030.

[32] G. Y. Xu, E. Ong, N. R. Gilkes, D. G. Kilburn, D. R. Muhandiram, M. Harris-Brandts, J. P. Carver, L. E. Kay, and T. S. Harvey, "Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy," *Biochemistry*, vol. 34, no. 21, pp. 6993–7009, May 1995, doi: 10.1021/bi00021a011.

[33] C. M. Bianchetti, P. Brumm, R. W. Smith, K. Dyer, G. L. Hura, T. J. Rutkoski, and G. N. Phillips Jr, "Structure, dynamics, and specificity of endoglucanase D from Clostridium cellulovorans," *Journal of Molecular Biology*, vol. 425, no. 22, pp. 4267–4285, Nov. 2013, doi: 10.1016/J.JMB.2013.05.030.

[34] P. J. Simpson, H. Xie, D. N. Bolam, H. J. Gilbert, and M. P. Williamson, "The structural basis for the ligand specificity of family 2 carbohydrate-binding modules," *Journal of Biological Chemistry*, vol. 275, no. 52, pp. 41137–41142, Dec. 2000, doi: 10.1074/JBC.M006948200.

[35] Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, S. Gräslund, P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H. W. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S. H. Kim, Z. Rao, Y. Shi, T. C. Terwilliger, C. Y. Kim, L. W. Hung, G. S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, R. C. Stevens, S. A. Lesley, I. A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S. K. Burley, J. S. Emtage, J. M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, D. I. Stuart, R. J. Owens, S. Daenke, A. Schütz, U. Heinemann, S. Yokoyama, K. Büssow, and K. C. Gunsalus, "Protein production and purification," *Nature Methods*, vol. 5, no. 2, pp. 135–146, Feb. 2008, doi: 10.1038/nmeth.f.202.

[36] N. N. Hengge, S. J. B. Mallinson, P. Pason, V. V. Lunin, M. Alahuhta, D. Chung, M. E. Himmel, J. Westpheling, and Y. J. Bomble, "Characterization of the biomass degrading enzyme GuxA from *Acidothermus cellulolyticus*," *International Journal of Molecular Sciences*, vol. 23, no. 11, Jun. 2022, doi: 10.3390/IJMS23116070.

[37] S. A. Mahadevan, S. G. Wi, D. S. Lee, and H. J. Bae, "Site-directed mutagenesis and CBM engineering of Cel5A (*Thermotoga maritima*)," *FEMS Microbiology Letters*, vol. 287, no. 2, pp. 205–211, Oct. 2008, doi: 10.1111/J.1574-6968.2008.01324.X.

[38] G. Geiger, G. Furrer, F. Funk, H. Brandl,

and R. Schulin, "Heavy metal effects on β-glucosidase activity influenced by pH and buffer systems," *Journal of Enzyme Inhibition*, vol. 14, no. 5, pp. 365–379, 1999, doi: 10.3109/14756369909030329.

[39] J. de C. Pereira, E. C. Giese, M. M. de S. Moretti, A. C. dos S. Gomes, O. M. Perrone, M. Boscolo, R. da Silva, E. Gomes, and D. A. B. Martins, "Effect of metal ions, chemical agents and organic compounds on lignocellulolytic enzymes activities," in *Enzyme Inhibitors and Activators*, M. Senturk, Ed., Rijeka: IntechOpen, 2017, doi: 10.5772/65934.

[40] M. Hans, S. Kumar, A. K. Chandel, and I. Polikarpov, "A review on bioprocessing of paddy straw to ethanol using simultaneous saccharification and fermentation," *Process Biochemistry*, vol. 85, pp. 125–134, 2019, doi: 10.1016/j.procbio.2019.06.019.

[41] D. S. Xue, X. Zeng, D. Lin, and S. Yao, "Ethanol tolerant endoglucanase from Aspergillus niger isolated from wine fermentation cellar", *Biocatalysis and Agricultural Biotechnology*, vol. 15, pp. 19–24, 2018, doi: 10.1016/j.bcab.2018.04.016.

[42] H. Yoshikawa, A. Hirano, T. Arakawa, and K. Shiraki, "Mechanistic insights into protein precipitation by alcohol," *International Journal of Biological Macromolecules*, vol. 50, no. 3, pp. 865–871, Apr. 2012, doi: 10.1016/J.IJBIOMAC.2011.11.005.

[43] Z. Zhu, J. Qu, L. Yu, X. Jiang, G. Liu, L. Wang, Y. Qu, and Y. Qin, "Three glycoside hydrolase family 12 enzymes display diversity in substrate specificities and synergistic action between each other," *Molecular Biology Reports*, vol. 46, no. 5, pp. 5443–5454, Oct. 2019, doi: 10.1007/s11033-019-04999-x.