EXPRESSION OF CELLULOSE-DEGRADING ENDOGLUCANASE FROM *BACILLUS SUBTILIS* USING PTOLT EXPRESSION SYSTEM IN *ESCHERICHIA COLI*

HÜLYA KUDUĞ CEYLAN,[*] YAKUP ULUSU,[**] SEMA BILGIN[***] and İSA GÖKÇE[*]

[*]*Tokat Gaziosmanpaşa University, Faculty of Engineering and Architecture, Department of Genetic and Bioengineering, Tokat, Turkey*
[**]*Karamanoğlu Mehmetbey University, Faculty of Engineering, Department of Bioengineering, Karaman, Turkey*
[***]*Tokat Gaziosmanpaşa University, Faculty of Art and Science, Department of Chemistry, Tokat, Turkey*
✉ *Corresponding author: H. Kuduğ Ceylan, hulya.kudug@gop.edu.tr*

Received March 29, 2021

Endoglucanases randomly hydrolyse the cellulose chains by acting upon internal β-1,4-D-glycosidic bonds and are used extensively in industrial applications. In this study, bacterial endoglucanase gene *yhfE* was obtained by PCR, using primers based on genomic sequences of *Bacillus subtilis* strains. 1041 bp DNA fragment of *yhfE* was cloned into *Escherichia coli* DH5α through the use of pTolT expression plasmid. PCR, restriction enzyme analysis and DNA sequencing were performed in order to confirm the cloning. *E. coli* BL21-AI cells expressed the yhfE after induction at 0.04% of arabinose concentration for 4 h. The expected 38.7 kDa size yhfE protein after digestion with thrombin of the His-tagged fusion protein (yhfE-TolAIII) was visualized by SDS-PAGE. The yhfE-TolAIII production yield was approximately 82 mg/L. The recombinant yhfE was characterized by MALDI-TOF mass spectrometry and CD analysis.

Keywords: cellulase, endoglucanase, recombinant protein, pTolT, *Escherichia coli*

INTRODUCTION

Industrial enzymes are efficiently employed as biocatalysts and offer several distinct advantages over chemical processes during commercial-scale applications. The advantages that make these enzymes favourable in manufacturing products include mild reaction conditions, reduced operating costs, lower toxicity and waste generation.[1-4]

Cellulase is one of the most demanded commercial enzymes in the industrial enzyme market. Cellulase catalyzes cellulose, which is the most abundant polysaccharide on the earth and is the primary structural substance of plants. Cellulose is a fibrous and insoluble linear polymer that is made by individual glucose molecules linked by β-1,4 glycosidic linkages.[5] The hydrolysis of cellulose is mediated by a multi-enzyme system that is mainly composed of endoglucanase, exoglucanase and β-glucosidase.[6-10] Endoglucanase (EC 3.2.1.4; endoβ-1,4-D-glucanase) hydrolyzes the β-1,4 linkages of the internal region of cellulose chains in a random manner to create oligosaccharides of different sizes with new chain ends. Exoglucanase, including cellobiohydrolases (CBHs) (EC 3.2.1.91), acts on reducing or non-reducing ends of cellulose chains, in order to generate glucose or cellobiose. Lastly, β-glucosidase (BG) (EC 3.2.1.21) cleaves the soluble cellodextrin and cellobiose to glucose monomers. All these enzymes act synergistically and target the specific cleavage of β-1,4-glycosidic bonds in order to release glucose.[7-10]

Some of the primary applications of cellulases are in the following industries: textile, animal feed, baking, food, beverages, pulp and paper, pharmaceutical and agriculture, as well as for research purposes. In addition to their utilization

in the valorisation of various wastes and in the improvement of soil quality, the commercial usage of cellulases has increased considerably in the production of environmentally friendly detergents, washing and cleaning agents, fruit juice, wine making, olive oil extraction, and bioethanol.[11]

In the natural environment, cellulase is synthesized by fungi, bacteria and actinomycetes. These microorganisms naturally offer noticeable production efficiency, cellulolytic abilities, enzyme composition and secretion pathway.[12] Generally, fungi are preferred to produce these enzymes, as they are capable of secreting plentiful amounts of cellulases. However, bacterial cellulases have drawn more attention, due to their high growth speed, short generation time, utilizing cheap carbon and nitrogen sources, and their resistance to environmental stress. The high-level expression, easy manipulation and functional diversity of these complex bacterial cellulases make them more convenient for industrial production.[13]

Among bacteria, the genus *Bacillus* is known for the ability to produce and secrete extracellular enzymes that have industrial importance.[14] *Bacillus* is one of the most important gram-positive bacteria and produces hydrolytic enzymes, besides commercial enzymes. It has been studied as a model organism to understand the secretion mechanism of extracellular enzymes. Many *Bacillus* species, including *B. subtilis*, *B. cereus*, *B. polymyxa* and *B. licheniformis*, are cellulase producers.[15]

Modern biotechnology with protein engineering techniques, especially recombinant DNA technology, offers opportunities for production of industrial enzymes. The developments in enzyme technology make it possible to produce enzymes with improved qualified activity or suitability for different process conditions. Especially for industrial-scale production of bacterial enzymes, *Escherichia coli* is one of the most preferred and well-characterized hosts.[16]

Besides the *Bacillus* species, molecular cloning of endoglucanase genes has been studied from different organisms, such as *Opuntia vulgaris*,[17] *Streptococcus bovis*,[18] *Phaffia rhodozyma*[19] and *Trichoderma harzianum*.[20] In this study, bacterial cellulase gene *yhfE* endoglucanase from *Bacillus subtilis subsp. subtilis str. 168* was isolated from bacterial strains that are stored in our laboratory from soil samples from a beech forest. The gene was cloned into *E. coli* DH5α, using inducible high copy pTolT expression vector, and was expressed in *E. coli* BL21-AI host cells. The pTolT is an expression vector developed to overcome problems present in many *E. coli* expression systems, such as formation of undesired inclusion bodies. It allows heterologous proteins to be produced in fusion with the *E. coli* periplasmic region protein TolAIII.[21] The purified recombinant endoglucanase was characterized by SDS-PAGE analysis, CD spectroscopy and MALDI-TOF mass spectrometry.

EXPERIMENTAL

Bacterial strains, plasmids and culture conditions

Bacillus strains, isolated from beech forest soil samples in our laboratory (Table 1) and then identified by 16S rRNA analysis, were screened for cellulase production. These strains were cultured in PCA (Plate Count Agar) growth medium containing 0.5% peptone, 0.25% yeast extract, 1.0% glucose and 2.0% agar (pH 4.8). *E. coli* strains were grown on LB medium with 100 μg/mL ampicillin. All chemicals, media and restriction enzymes were purchased from Sigma (USA), Merck (Germany) and Takara (China), respectively.

Table 1
Strains and plasmids used in this study

	Description	Source
Strains		
E. coli DH5α	General cloning host strain	Invitrogen™
E. coli BL21-AI	Expression host strain	Invitrogen™
Bacillus subtilis	Wild-type, can degrade CMC	Isolated in our laboratory
Rahnella aquatilis	Wild-type, can degrade CMC	Isolated in our laboratory
Bacillus ginsengihumi	Wild-type, can degrade CMC	Isolated in our laboratory
Enterococcus faecalis	Wild-type, can degrade CMC	Isolated in our laboratory
Plasmids		
pTolT	Expression vector, Amp^r	Laboratory stock
pTolT-yhfE	*yhfE* gene into pTolT	Constructed in this study

*CMC (carboxymethyl cellulose)

Screening cellulase activity of strains

The cellulase activity of the strains was determined on a 1.0% carboxymethyl cellulose (CMC) (Merck, Germany) agar plate, containing 0.5% peptone (Merck), 0.25% yeast extract (Merck) and 2.0% agar (BD, United States) (pH 4.8). The glycerol stock of these cellulolytic bacterial cultures at -80 °C were incubated at 37 °C and 250 rpm for 30 h in growth medium (0.25% yeast extract, 0.5% peptone and 0.1% glucose, pH 4.8). An equal volume of each culture was inoculated in the middle of the CMC agar plates and incubated under identical conditions. The incubated plates were developed with 1.0% Congo red dye (Amresco, United States) (0.5-1 h), followed by destaining with 1 M NaCl solution for 15-20 min.[22] The cellulolytic potential of the strains was compared according to the visible clearing zones.

Investigation of the presence of cellulase genes

The sequences of four *B. subtilis* cellulase genes BSn5-00410 endo-1,4-beta-glucanase (Gene ID: 10180729), *eglS* endo-1,4-beta-glucanase (Gene ID: 9722851), *ysdC* endo-1,4-beta-glucanase (Gene ID: 936750) and *yhfE* endoglucanase (Gene ID: 936306) were obtained from the NCBI gene database for further cloning studies. Forward and reverse primers for BSn5-00410 endo-1,4-beta-glucanase (5'tttt<u>ggatcc</u>atgaaacggtcaatct3'; 5'ttt<u>acgcgt</u>tcaatttggttctgttcccc3'), *eglS* endo-1,4-beta-glucanase (5'tttt<u>ggatcc</u>atgaaacgttcagtc3'; 5'ttt<u>acgcgt</u>tcatttgggttctgttcccc3'), *ysdC* endo-1,4-beta-glucanase (5'tttt<u>ggatcc</u>atggcaaaattagat3'; 5'ttt<u>acgcgt</u>ttattggtacgtaatttc3') and *yhfE* endoglucanase (5'tttt<u>ggatcc</u>atgacgtccgtacgt3'; 5'ttt<u>acgcgt</u>ttataccattggtgactg3') genes were designed and synthesized to detect the presence of cellulase genes from the genomic DNA of positive cellulolytic isolates. Restriction enzyme sites (underlined above), *BamH*I (5' forward primer) and *Mlu*I (3' reverse primer) were incorporated into pairs of primers compatible with the expression vector pTolT for the cloning. Genomic DNA of the bacterial strains that showed the highest cellulase activity was used as DNA template for PCR, to investigate the presence of cellulase genes. PCR products were analyzed by 1.0% agarose gel electrophoresis under UV light.

Cloning of *yhfE* endoglucanase gene in pTolT

B. subtilis yhfE endoglucanase gene was amplified by the PCR method with Taq DNA polymerase (Biobasic, Canada) and purified (Qiagen QIAquick PCR Purification Kit, France). The purified PCR product of *yhfE* and plasmid DNA of the expression vector pTolT were both digested with *BamH*I (Promega, Madison, WI) and *Mlu*I (Promega, Madison, WI), and ligated with T4 DNA ligase (Promega, Madison, WI) to form recombinant *yhfE*-pTolT (Fig. 1). The recombinant plasmid DNA was transformed into *E. coli* DH5α competent cells by the heat shock method and incubated on LB agar plates containing 100 mg/mL of ampicillin at 37 °C overnight. Plasmid DNA of the colonies on the agar plate was isolated using a miniprep kit (Promega Wizard® Plus SV, Madison, WI). Positive clones were confirmed by restriction enzyme analysis, PCR and sequence analysis of plasmid DNA.

Expression of yhfE in *E. coli* BL21-AI

E. coli BL21-AI competent cells were transformed with recombinant *yhfE*-pTolT plasmid DNA and grown overnight in an LB agar plate, containing 100 µg/mL ampicillin. The single transformed colony from the plate was inoculated in 6 mL of LB medium and incubated overnight at 37 °C in a shaking incubator at 200 rpm. The overnight starter culture was used to inoculate 600 mL of the above-mentioned LB in the shake flask at a ratio of 1:100, and was cultured under the same conditions until reaching OD600 of 0.6-0.7. Then, the culture was induced by addition of L-arabinose at 0.2% final concentration. After 4 h of induction, the culture was centrifuged at 14000 g for 5 minutes. The supernatant was discarded and the pellet was stored at -20 °C until required for use in the purification steps.

Protein purification

The *E. coli* cell pellet was resuspended in lysis buffer, 100 mM Tris buffer (pH 7.5) by adding 100 mM phenylmethanesulfonylfluoride (PMSF), 100 mM benzamidine, RNAse (20 µg/mL), DNAse (20 µg/mL) and lysozyme. The mixture was lysed, subjected to high pressure in cell disruption equipment (Constant Systems, UK) in an ice/water bath. This was followed by high-speed centrifugation at 100,000 g for 1 h. The Ni^{2+}-nitrilotriacetic acid (Ni-NTA) agarose column was equilibrated with 100 mM Tris/HCl (pH 7.5) for purification of the soluble yhfE carrying N-terminal 6×histidine tag. The column was loaded with supernatant and washed with 100 mM Tris/HCl (pH 7.5) and 100 mM Tris/HCl (pH 7.5) containing 30 mM imidazole. The recombinant yhfE was eluted with 100 mM Tris/HCl (pH 7.5) containing 300 mM imidazole.

The fusion TolAIII-yhfE was dialyzed against 50 mM phosphate buffer (pH 7.5). TolAIII fusion protein in the pTolT expression system was removed from the TolAIII-yhfE fusion by using a thrombin protease cleavage site with Thrombin CleanCleave™ Kit (Aldrich, St Louis, MO, USA). Affinity chromatography was performed again to purify the yhfE from the His-tagged TolAIII after cleavage. The flow through sample, including pure recombinant yhfE, was dialyzed against 50 mM phosphate buffer (pH 7.5).

Characterization of the purified protein

The purified proteins were visualized using 12% acrylamide/bis-acrylamide gels before and after the cleavage with thrombin.[23] The gels were stained with

Coomassie Brilliant G250. The secondary structure of the pure yhfE was checked using circular dichroism (CD) spectroscopy. Far-UV circular dichroism (CD) spectra of yhfE (50 mM phosphate buffer, pH 7.5 at 25 °C) were measured in the range of 195-250 nm in a J-810 spectropolarimeter (Jasco International Co. Ltd., Tokyo, Japan), using a 0.2 cm quartz cuvette. The thermal stability was investigated at different temperatures ranging from 25 to 100 °C. Thermal unfolding was monitored by CD intensity changes at 220 nm. Identification of yhfE endoglucanase was performed by Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF).

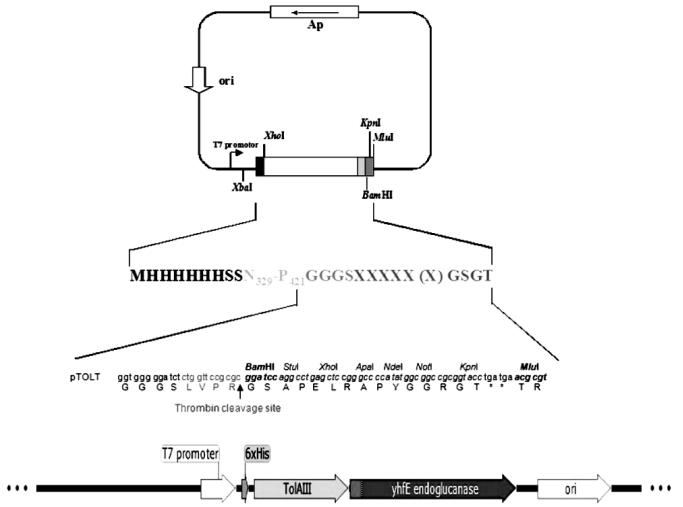

Figure 1: Multiple cloning site of pTolT expression plasmid and a schematic map of TolAIII-yhfE fusion protein (thrombin recognition site (LVPRGS) and cleavage site marked)

RESULTS AND DISCUSSION

Screening of cellulose degrading bacteria

In this study, four bacterial isolates (*Bacillus subtilis*, *Rahnella aquatilis*, *Bacillus ginsengihumi* and *Enterococcus faecalis*) that were obtained from beech forest soil and identified before in our laboratory were analyzed in terms of cellulose activity.[24] These specific four isolates were selected for activity assays as they are known to be able to degrade xylan. Strains were grown on CMC agar for selecting the cellulase producer utilizing CMC as the carbon source. After overnight incubation of the strains, the clear zone around the area of growth of each strain appeared with Congo red dyeing. The size of the decolourization zone on the plates shows the degradation of CMC and the degree of cellulase activity. Three strains (except *R. aquatilis*) showed a degradation zone. *B. subtilis* strain (a) that has the largest size of the clearing zone was considered to have the highest cellulase activity, among the strains analyzed (Fig. 2).

eglS Endo-1,4-beta-glucanase, *BSn5* endo-1,4-beta-glucanase, *yhfE* endoglucanase and *ysdC* endo-1,4-beta-glucanase are cellulase enzymes that are defined as *B. subtilis* cellulase gene. PCR was performed for amplifying the genes of these cellulase enzymes. As shown in Figure 3, the cellulase gene of *yhfE* endoglucanase (1041 bp) has been highly amplified (Lane 4). The cloning and expression experiments were performed with the *yhfE* gene.

Cloning of the *yhfE* endoglucanase gene

The pTolT expression vector system used in this study is popular due to its relative simplicity in expressing proteins. As regards *E. coli*, a vast amount of knowledge about it exists due to the low costs associated with its production. The pTolT expression vector developed to overcome some problems, such as inclusion bodies or toxic effects, allows the production of heterologous proteins as a fusion partner of the *E. coli* periplasmic region protein TolAIII. The TolAIII domain, therefore, not only drives high expression of the fusion polypeptide, but also provides an affinity tag for purification.

Several reports are available for recombinant expression studies in bacterial systems. Vadala *et al*.[25] reported cloning and expression of the cellulase gene from *B. subtilis* (natto strain) in BL21 (DE3) strain of *E. coli*. The cellulase gene

was inserted within *Bam*HI and *Hind*III sites of the T7 promoter-based vector (pET21a). In another study, randomly digested fragments of *Clostridium thermocellum* were cloned into pET28a(+) vector and expressed in *E. coli* BL21(DE3) pLysS.[26] In this study, the *Bacillus sp.* isolate was used as a source organism for the isolation of the cellulase gene. The amplified *B. subtilis yhfE* gene and pTolT plasmid DNA were digested with *BamH*I and *Mlu*I restriction enzymes. The ligation was performed using T4 DNA ligase and the recombinant plasmid named pTolT-*yhfE* was transformed into competent *E. coli* DH5α cells. Plasmid DNA of the colonies on the selective agar plate was purified to verify the cloning by colony PCR and restriction fragment analysis. Colonies that have the 1041 bp fragment as a result of colony PCR were digested with *BamH*I and *Mlu*I separately and together. The amino acid sequence of positive clones was aligned with the reference amino acid sequences of the wild type (Fig. 4).

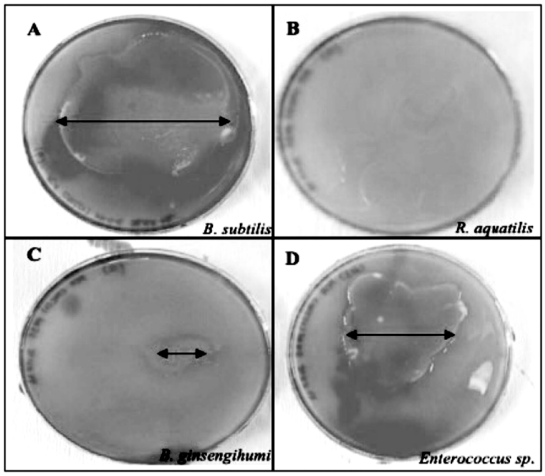

Figure 2: Plate screening of cellulase activity of different bacterial strains (A) *Bacillus subtilis*, (B) *Rahnella aquatilis*, (C) *Bacillus ginsengihumi*, and (D) *Enterococcus sp.*

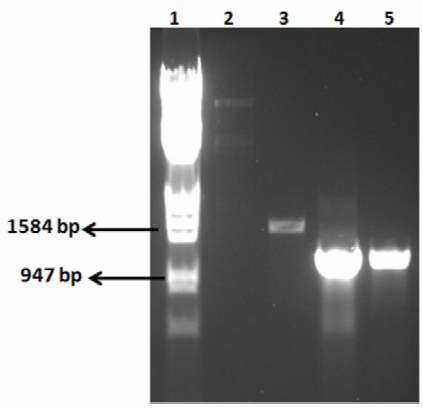

Figure 3: PCR of cellulase genes in *B. subtilis* genomic DNA; 1. λ / *Hind* III-*EcoR* I DNA Marker, 2. eglS endo-1,4-beta-glucanase, 3. BSn5 endo-1,4-beta-glucanase, 4. *yhfE* endoglucanase, 5. ysdC endo-1,4-beta-glucanase

```
wild-type      ------MTSVRKTMELIKELVSIPSPTGNTYEVINYIESLLKEWKVETVRNHKGGLIATL   54
recombinant    LVPRGSMTSVRKTMELIKELVSIPSPTGNTYEVINYIESLLKEWKVETVRNHKGGLIATL   60
               ======************************************************

wild-type      PGRDTSRHRMLTAHVDTLGAMVKEIKADGRLKIDLIGGFRYNSIEGEYCQIETASGKMYT   114
recombinant    PGRDTSRHRMLTAHVDTLGAMVKEIKADGRLKIDLIGGFRYNSIEGEYCQIETASGKTYT   120
               ************************************************************

wild-type      GTILMHQTSVHVYKDAGKAERNQENMEIRLDEPVHCRKDTEELGIGVGDFVSFDPRVEIT   174
recombinant    GTILMHQTSVHVYKDAGKAERNQENMEIRLDEPVHCRKDTEELGIGVGDFVSFDPRVEIT   180
               ************************************************************

wild-type      SSGFIKSRHLDDKASVALLLRLIHEIQTEDIELPYTTHFLISNNEEIGYGGNSNIPPETV   234
recombinant    SSGFIKSRHLDDKASVALLLRLIHEIQTEDIELPYTTHFLISNNEEIGYGGNSNIPPETV   240
               ************************************************************

wild-type      EYLAVDMGAIGDGQATDEYSVSICVKDASGPYHYQLRKHLVQLAEKHHIDYKLDIYPYYG   294
recombinant    EYLAVDMGAIGDGQATDEYSVSICVKDASGPYHYQLRKHLVQLAEKHHIDYKLDIYPYYG   300
               ************************************************************

wild-type      SDASAAIKSGHDIVHGLIGPGIDASHAFERTHKSSLRHTAKLLYYYVQSPMV   346
recombinant    SDASAAIKSGHDIVHGLIGPGIDASHAFERTHKSSLRHTAKLLYYYVQSPMV   352
               ******************************************************
```

Figure 4: Alignment of amino acid sequences of yhfE endoglucanase and related cellulase of *Bacillus subtilis* (Gene ID: 936306); thrombin recognition site (LVPRGS) is underlined

Analysis of the purified protein

SDS-PAGE analysis followed by staining with Coomassie Brilliant G250 before cleavage with thrombin shows the TolAIII-yhfE fusion protein, of expected molecular weight (Fig. 5a). The expected proteolysis products of 38.7 and 11.5 kDa (TolAIII and yhfE, respectively) were observed after cleavage of the fusion protein (Fig. 5b). In this study, the open reading frame of the *yhfE*-pTolT plasmid DNA consists of 6×His Tag, TolAIII *E. coli* periplasmic protein, thrombin enzyme recognition site (LVPRGS), protein of yhfE endoglucanase and termination codon, respectively. Further sequence analysis of the *yhfE*-pTolT construct indicated that the cloned sequence started with an ATG start codon and terminated with a TAG stop codon. The gene consists of 1041 nucleotides encoding a protein of 347 amino acids with a predicted molecular weight of 38.7 kDa. The amino acid sequences had 100% identity with the reported cellulase (Gene ID: 936306) of *B. subtilis* further confirming the identity of the cloned gene. The analysis of the *E. coli* BL21-AI cell lysate containing yhfE-pTolT plasmid DNA by SDS-PAGE showed that the yhfE was expressed in a soluble form. No inclusion bodies were revealed by visual inspection of pelleted remains of the bacteria after the lysozyme treatment, sonication and centrifugation, whereas the pTol system might also be possible for some proteins expressed as inclusion bodies. The molecular weight and molar absorption coefficient of the TolAIII-yhfE fusion protein were computed as 50.2 kDa and 39.970 $M^{-1}cm^{-1}$, respectively, by using ExPASy ProtParam Tool. The yield of the fusion protein was, on average, approximately 82 mg/L of bacterial broth, which is comparable, or even better, than other published and existing systems for production of fusion proteins in *E. coli*.

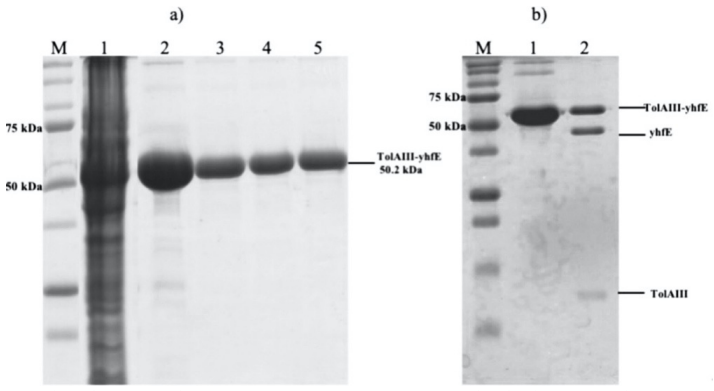

Figure 5: SDS–PAGE analysis of recombinantly expressed TolIIIA-yhfE fusion protein and its thrombin cleavage; (M. Precision Plus Protein Unstained Standards, #161-0363) a) Purification of recombinant protein, 1. *E. coli* BL21-AI cell lysate containing yhfE-pTolT plasmid DNA, 2-5. Elution fraction of TolIIIA-yhfE fusion protein obtained from the affinity column; b) Cleavage of TolIIIA-yhfE fusion protein, 1. TolAIII-yhfE fusion protein, 2. Cleavage of TolAIII-yhfE fusion protein with thrombin protease

It has been shown that the levels of expression of various fusion proteins are around 20% of total bacterial proteins and the pTolT system is able to purify 50-90 mg of fusions per liter of bacterial broth.[27] To obtain purified non-TolAIII yhfE, it was removed from the fusion protein by endopeptidase cleavage (thrombin) and purified. SDS-PAGE analysis, followed by staining with Coomassie Brilliant G250 before cleavage with thrombin, shows that the TolAIII-yhfE fusion protein had the expected molecular weight (Fig. 5a). The expected proteolysis products of 38.7 and 11.5 kDa (TolAIII and yhfE, respectively) were observed after cleavage of the fusion protein (Fig. 5b).

Characterization of yhfE endoglucanase

Circular dichroism (CD) is one of the most frequently used methods for analysing the secondary structure of proteins in solution. In order to receive a signal in the spectropolarimeter, at least one chiral center must be present in the molecule.[28] In proteins, peptide bonds are often used as chromophores that generate the CD signal. The far UV spectrum of the pure yhfE was obtained using circular dichroism (CD) spectroscopy. As can be seen from the Far UV

CD signal, the yhfE protein produced and purified in *E. coli* gained its three-dimensional conformation. The yhfE has an alpha-helical and antiparallel β-sheet structure, compared to reference spectra[28] and the secondary structure of *B. subtilis* endoglucanases described in the literature.[29] Melting temperatures were determined as described previously[30] by examining the peak minima of the first order derivative plots. These are equivalent to the midpoint transition (Tm) of the curve where 50% of the protein is unfolded. Measurement for the thermal unfolding of yhfE was performed at 220 nm (Fig. 6). Thermal unfolding temperature has also been determined by using a Jasco 810 spectropolarimeter for the yhfE protein, which was found to have acquired its secondary structure according to Far UV CD spectra results. The normalized thermal unfolding results are presented in Figure 6b, illustrating the decrease in the amount of folded fraction due to the increase in temperature. The melting temperature, which is the point when the secondary structure is lost and denaturation begins, is almost 75 °C. This result indicates that yhfE is a potential candidate as a thermostable cellulase for industrial applications.

The yhfE band was excised from the gel and was further confirmed by MALDI-TOF mass spectrometry. This method is helpful for identification of proteins by peptide mass fingerprinting (PMF). The results in Figure 7 indicate that the second match with Score 110 (Accession number: gil1607808; Mass: 38883) is described as endoglucanase (*Bacillus subtilis subsp. subtilis str. 168*).

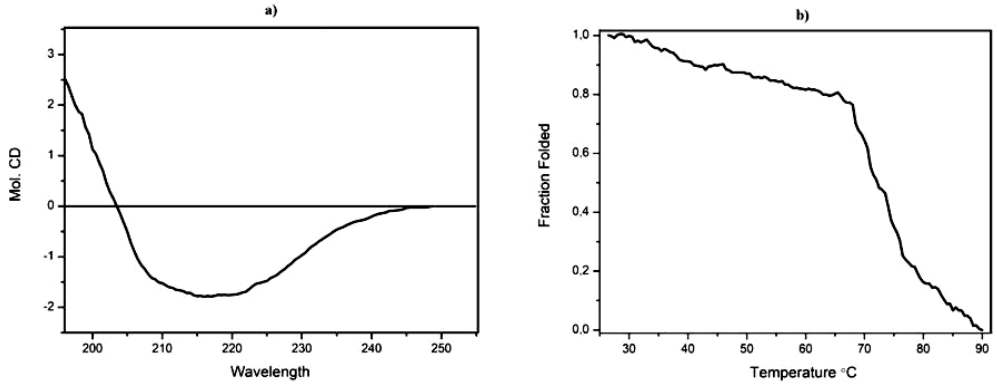

Figure 6: Characterization of yhfE biophysically; a) Far-UV CD and b) Thermal unfolding spectrum of yhfE

	Accession	Mass	Score	Description
1.	gi|381470036	38079	111	hypothetical protein BSSC8_33090 [Bacillus subtilis subsp. subtilis str. SC-8]
2.	gi|16078084	38883	110	endoglucanase [Bacillus subtilis subsp. subtilis str. 168]
3.	gi|291483495	38853	110	hypothetical protein BSNT_01736 [Bacillus subtilis subsp. natto BEST195]
4.	gi|321314744	38897	110	putative endoglucanase [Bacillus subtilis BSn5]
5.	gi|386787700	38883	110	putative endoglucanase [Bacillus sp. JG]
6.	gi|384174712	38862	99	YhfE [Bacillus subtilis subsp. subtilis str. RO-NN-1]
7.	gi|380268299	38875	99	hypothetical protein GYO_1318 [Bacillus subtilis subsp. spizizenii TU-B-10]
8.	gi|296332442	38867	90	putative endoglucanase [Bacillus subtilis subsp. spizizenii ATCC 6633]

2. gi|16078084 **Mass:** 38883 **Score:** 110 **Expect:** 0.00013 **Matches:** 15
endoglucanase [Bacillus subtilis subsp. subtilis str. 168]

Observed	Mr(expt)	Mr(calc)	ppm	Start	End	Miss	Peptide
890.5620	889.5548	889.5022	59.2	87 -	94	0	K.IDLIGGFR.Y
1033.5376	1032.5303	1032.4658	62.5	136 -	143	0	R.NQENMEIR.L
1080.6661	1079.6589	1079.5863	67.2	171 -	180	0	R.VEITSSGFIK.S
1131.7658	1130.7585	1130.6812	68.4	85 -	94	1	R.LKIDLIGGFR.Y
1306.7028	1305.6955	1305.6102	65.4	261 -	271	0	K.DASGPYHYQLR.K
1323.8156	1322.8083	1322.7194	67.2	171 -	182	1	R.VEITSSGFIKSR.H
1333.8539	1332.8466	1332.7626	63.0	45 -	57	1	R.NHKGGLIATLPGR.D
1413.8685	1412.8612	1412.7736	62.0	48 -	61	1	K.GGLIATLPGRDTSR.H
1434.8108	1433.8036	1433.7051	68.7	261 -	272	1	K.DASGPYHYQLRK.H
1450.9366	1449.9293	1449.8303	68.2	183 -	195	1	R.HLDDFASVALLLR.L
1953.0679	1952.0606	1951.9164	73.9	153 -	170	0	K.DTEELGIGVGDFVSFDPR.V
2081.1752	2080.1680	2080.0113	75.3	152 -	170	1	R.KDTEELGIGVGDFVSFDPR.V
2285.3234	2284.3161	2284.1349	79.3	303 -	324	0	K.SGHDIVHGLIGPGIDASHAFER.T
2651.5679	2650.5606	2650.3364	84.6	303 -	327	1	K.SGHDIVHGLIGPGIDASHAFERTHK.S
3014.7522	3013.7449	3013.4921	83.9	153 -	180	1	K.DTEELGIGVGDFVSFDPRVEITSSGFIK.S

Figure 7: MALDI-TOF analysis of pure yhfE endoglucanase

As a result, the MALDI-TOF MS–MS fingerprinting and sequencing results confirmed that the targeted protein sample was the endoglucanase of *Bacillus subtilis subsp. subtilis str. 168* (Accession number: gil1607808) and matched the theoretical molecular weight of 38.883 kDa. This means the target protein aimed in this study was successfully produced recombinantly.

CONCLUSION

In the current work, the pTolT expression system was used for the first time for the expression of a recombinant cellulase enzyme in a soluble fraction. The yield of 82 mg fusion protein per liter of bacterial culture was obtained. The cloning and expression of *Bacillus* cellulase enzymes has been performed in a wide range of expression levels.[31-36] We have used *E. coli* strains (DH5α and BL21-AI) as cloning and expression organism. The *yhfE* endoglucanase gene consisting of 1041 nucleotides encoding protein was cloned in the pTolT expression vector. MALDI-TOF mass spectrometry analysis revealed that the target protein obtained in this study shows the highest amino acids matches with *Bacillus subtilis subsp. subtilis str. 168* endoglucanase, as expected. The *E. coli* BL21-AI was used as expression host, which is ideal for overexpression by L-arabinose induction. SDS-PAGE results supported the yhfE expressed in high yield and purity. The expression and purification procedures reported in this study have provided a simple and efficient method to obtain pure yhfE protein. Further research will be conducted on its structure and function.

ACKNOWLEDGMENTS: This work is supported by the Scientific Research Project Fund of Tokat Gaziosmanpaşa University under the project number 2012/61. The authors are grateful to Prof. Jeremy Lakey for MALDI-TOF mass spectrometry and CD analysis of the recombinant *yhfE* endoglucanase.

REFERENCES

[1] A. S. Bommarius and M. F. Paye, *Chem. Soc. Rev.*, **42**, 6534 (2013), https://doi.org/10.1039/c3cs60137d

[2] J. M. Choi, S. S. Han and H. S. Kim, *Biotechnol. Adv.*, **33**, 1443 (2015), https://doi.org/10.1016/j.biotechadv.2015.02.014

[3] A. Madhavan, R. Sindhu, P. Binod, R. K. Sukumaran and A. Pandey, *Bioresour. Technol.*, **245**, 1304 (2017), https://doi.org/10.1016/j.biortech.2017.05.031

[4] J. Chapman, A. E. Ismail and C. Z. Dinu, *Catalysts*, **8**, 238 (2018), https://doi.org/10.3390/catal8060238

[5] D. Ciolacu, F. Ciolacu and V. I. Popa, *Cellulose Chem. Technol.*, **45**, 13 (2011), https://www.cellulosechemtechnol.ro/pdf/CCT1-2(2011)/p.13-21.pdf

[6] S. Sajith, P. Priji, S. Sreedevi and S. Benjamin, *J. Nutr. Food Sci.*, **6**, 1 (2016), https://doi.org/10.4172/2155-9600.1000461

[7] P. Béguin and J. P. Aubert, *FEMS Microbiol. Rev.*, **13**, 25 (1994), https://doi.org/10.1111/j.1574-6976.1994.tb00033.x

[8] T. T. Teeri, *Trends Biotechnol.*, **15**, 160 (1997), https://doi.org/10.1016/S0167-7799(97)01032-9

[9] L. R. Lynd, P. J. Weimer, W. H. van Zyl and I. S. Pretorius, *Microbiol. Mol. Biol. Rev.*, **66**, 506 (2002), https://doi.org/10.1128/mmbr.66.3.506-577

[10] M. Bhat and S. Bhat, *Biotechnol. Adv.*, **15**, 583 (1997), https://doi.org/10.1016/s0734-9750(97)00006-2

[11] D. Lavanya, P. K. Kulkarni, M. Dixit, P. K. Raavi and L. N. V. Krishna, *Int. J. Drug Formul. Res.*, **2**, 19 (2011)

[12] R. K. Sukumaran, R. R. Singhania and A. Pandey, *J. Sci. Ind. Res.*, **64**, 832 (2005), http://nopr.niscair.res.in/bitstream/123456789/5375/1/JSIR%2064%2811%29%20832-844.pdf

[13] W. Li, X. Huan, Y. Zhou, Q. Ma and Y. Chen, *Biochem. Biophys. Res. Commun.*, **383**, 397 (2009), https://doi.org/10.1016/j.bbrc.2009.04.027

[14] F. G. Priest, *Bacteriol. Rev.*, **41**, 711 (1977), https://doi.org/10.1128/br.41.3.711-753.1977.

[15] J. S. Han, Y. J. Yoo and H. S. Kang, *J. Biol. Chem.*, **270**, 26012 (1995), https://doi.org/10.1074/jbc.270.43.26012

[16] J. H. Choi and S. Y Lee, *Appl. Microbiol. Biotechnol.*, **64**, 625 (2004), https://doi.org/10.1007/s00253-004-1559-9

[17] S. Shyamala, S. Ravikumar, J. Vikramathithan and K. Srikumar, *Appl. Biochem. Biotechnol.*, **165**, 1597 (2011), https://doi.org/10.1007/s12010-011-9380-2

[18] M. S. Ekinci, S. I. McCrae and H. J. Flint, *Appl. Environ. Microbiol.*, **63**, 3752 (1997), https://doi.org/10.1128/aem.63.10.3752-3756.1997

[19] M. L. Bang, I. Villadsen and T. Sandal, *Appl. Microbiol. Biotechnol.*, **51**, 215 (1999), https://doi.org/10.1007/s002530051384

[20] E. F. Noronha and C. J. Ulhoa, *FEMS Microbiol. Lett.*, **183**, 119 (2000), https://doi.org/10.1111/j.1574-6968.2000.tb08944.x

[21] G. Anderluh, I. Gokce and J. H. Lakey, *Protein Expr. Purif.*, **28**, 173 (2003), https://doi.org/10.1016/s1046-5928(02)00681-2

[22] R. M. Teather and P. J. Wood, *Appl. Environ. Microbiol.*, **43**, 777 (1982), https://doi.org/10.1128/aem.43.4.777-780.1982
[23] U. K. Laemmli, *Nature*, **227**, 680 (1970), https://doi.org/10.1038/227680a0
[24] S. Bilgin, Y. Ulusu, H. Kuduğ and I. Gokce, *Sakarya Uni. J. Sci.*, **22**, 1508 (2018), https://doi.org/10.16984/saufenbilder.327153
[25] B. S. Vadala, S. Deshpande and A. Apte-Deshpande, *J. Genet. Eng. Biotechnol.*, **19**, 7 (2021), https://doi.org/ 10.1186/s43141-020-00103-0
[26] O. Jantasaeng, P. Thaenkudrua, J. S. Tan and P. Phapugrangkul, *Food Appl. Biosci. J.*, **7**, 3 (2019), https://li01.tci-thaijo.org/index.php/fabjournal/article/view/176610/125956
[27] I. Gokce, G. Anderluh and H. J. Lakey, Patent number: US 7348408 B2 (2002)
[28] S. M. Kelly, T. J. Jess and N. C. Price, *Biochim. Biophys. Acta*, **1751**, 119 (2005), https://doi.org/10.1016/j.bbapap.2005.06.005
[29] Y. Yasutake, S. Kawano, K. Tajima, M. Yao, Y. Satoh *et al.*, *Proteins Struct. Funct. Bioinform.*, **64**, 1069 (2006), https://doi.org/10.1002/prot.21052
[30] N. J. Greenfield, *Nat. Protocols*, **1**, 2527 (2006), https://doi.org/10.1038/nprot.2006.204
[31] W. Li, W. W. Zhang, M. M. Yang and Y. L. Chen, *Mol. Biotechnol.*, **40**, 195 (2008), https://doi.org/10.1007/s12033-008-9079-y
[32] J. L. You, B. K. Kima, B. H. Lee, K. I. Joa and N. K. Leea, *Bioresour. Technol.*, **99**, 378 (2008), https://doi.org/10.1016/j.biortech.2006.12.013
[33] K. K. Bo, B. H. Lee, Y. J. Leeb, H. Jina, C. H. Chunga *et al.*, *Enzyme Microb. Technol.*, **44**, 411 (2009), https://doi.org/10.1016/j.enzmictec.2009.02.005
[34] O. A. Odeniyi, A. A. Onilude and M. A. Ayodele, *Afr. J. Microbiol. Res.*, **3**, 407 (2009), https://doi.org/10.5897/AJMR.9000594
[35] G. Rastogi, A. Bhalla, A. Adhikari, K. M. Bischoff, S. R. Hughes *et al.*, *Bioresour. Technol.*, **101**, 8798 (2010), https://doi.org/10.1016/j.biortech.2010.06.001
[36] Y. Liu, H. Gou, Y. Wu and W. Qin, *Bioresour. Bioprocess*, **5**, 19 (2018), https://doi.org/10.1186/s40643-018-0204-x