Molecular Cloning and Expression of Cellulase and Polygalacturonase Genes in *E. coli* as a Promising Application for Biofuel Production

Eman Ibrahim1,2, Kim D. Jones3, Ebtesam N. Hosseny4 and Jean Escudero4*  

1Faculty of Science, Nasr City, Cairo, Egypt  
2Department of Botany and Microbiology, Al-Azhar University, Nasr City, Cairo 11884, Egypt  
3Department of Environmental Engineering, Texas A&M University-Kingsville, Texas 78363, USA  
4Department of Biological science, Texas A&M University-Kingsville, Texas 78363, USA

**Abstract**  
Lignocellulosic biomass has potential for bioethanol, a renewable fuel. A limitation is that bioconversion of the complex lignocellulosic material to simple sugars and then to bioethanol is a challenging process. Recent work has focused on the genetic engineering of a biocatalyst that may play a critical role in biofuel production. *Escherichia coli* have been considered a convenient host for biocatalysts in biofuel production for its fermentation of glucose into a wide range of short-chain alcohols and production of highly deoxygenated hydrocarbon. The bacterium *Pectobacterium carotovorum* subsp. *carotovorum* (P. *carotovorum*) is notorious for its maceration of the plant cell wall causing soft rot. The ability to destroy plants is due to the expression and secretion of a wide range of hydrolytic enzymes that include cellulases and polygalacturonases. *P. carotovorum* ATCC™ no. 15359 was used as a source of DNA for the amplification of *celB*, *celC* and *peh*. These genes encode 2 cellulases and a polygalacturonase, respectively. Primers were designed based on published gene sequences and used to amplify the open reading frames from the genomic DNA of *P. carotovorum*. The individual PCR products were cloned into the pTAC-MAT-2 expression vector and transformed into *Escherichia coli*. The deduced amino acid sequences of the cloned genes have been analyzed for their catalytically active domains. Estimation of the molecular weights of the expressed proteins was performed using SDS-PAGE analysis and *celB*, *celC* and *peh* products were approximately 29.5 kDa, 40 kDa and 41.5 kDa, respectively. Qualitative determination of the cellulase and polygalacturonase activities of the cloned genes was carried out using agar diffusion assays.

**Keywords:** *P. carotovorum* subsp. *carotovorum*; Cellulase; Polygalacturonase; Expression; Lignocellulosic material; Bioethanol; Biofuel

**Introduction**  
The massive usage of petroleum and petroleum products in the last decade, with the consequent reverse effect on minimizing consumption of these unsustainable resources, has increased the demand for the development of renewable sources [1,2]. Currently, based on the carbon neutrality concept, two sources of biofuels have entered the marketplace: ethanol from cellulosic materials and biodiesel from soybean or palm oil [3]. The bioconversion of lignocellulosic materials is a challenging process which requires two steps. During the bioconversion process, the lignin and the hemicellulosic parts are first degraded into simpler sugars and/or organic acids, followed by a deoxygenating step to produce a liquid fuel [4]. Design of a genetically modified microorganism for direct lignocellulosic biomass conversion purposes has recently been taken into consideration [5]. The production of several types of fuel through direct lignocellulosic biomass conversions has been demonstrated by various studies [6,7]. A genetically engineered *E. coli* capable of degrading pectin-rich lignocellulosic biomass by cellulolytic and pectinolytic activities has been developed [8]. *E. coli* has been considered a convenient biocatalyst in biofuel production for its fermentation of glucose into a wide range of short-chain alcohols [9,10], and production of highly deoxygenated hydrocarbon through fatty acid metabolism [11,12]. Moreover, the ability to ferment several pentoses and hexoses makes *E. coli* an ideal ethanologen for biofuel production [5,13].

The recalcitrant nature of lignocellulosic biomass was attributed to its complex constituents of cellulose and hemicellulose in addition to variable amounts of lignin [14,15]. Lignin has been reported to impede the enzymatic hydrolysis of plant cell wall polysaccharides [16,17]. However, a pectin-rich substance such as citrus waste residue has been proposed as an ideal substrate for ethanol production due to its low lignin content and high soluble sugar concentration in comparison to other lignocellulosic feedstocks [18,19]. Enzymatic hydrolysis of the insoluble carbohydrate parts of the citrus waste residue, such as cellulose, hemicellulose and pectin, can be performed using an enzyme cocktail of cellulase and pectinase [20].

Hydrolysis of the insoluble cellulotic biomass in citrus waste residue can be carried out by the action of a group of cellulase enzymes collectively known as the cellulosome. Typically, endoglucanases (EGs), and cellobiohydrolases (CBH) are the main cellulase enzymes that act synergistically in hydrolyzing a cellulotic substrate [21]. EGs are endo-acting enzymes that function in the hydrolysis of glycosidic bonds, making free ends available for the exo-action of CBH to produce cellobiose and some glucose molecules [22]. However, β-glucosidase is also necessary for reducing the end-product inhibitory effect of cellobiose on CBH via its hydrolytic conversion into glucose [23]. On the other hand, pectin is a homo-poly saccharide located in the middle lamella of the cell walls of the plant tissues and represents...
one-third of their dry weights [24]. Pectin typically consists of long chains of galacturonic acid with carboxyl groups and methyl ester residues [25]. Production of the polygalacturonase enzyme is an area of utmost concern for its role in releasing cell wall cellulose fibrils, which are tightly cemented and embedded into the pectin matrix. The characterization of polygalacturonases has been reported by many researchers [26,27]. Polygalacturonases are a group of enzymes that collectively function in the hydrolysis of non-esterified polygalacturonide chains of pectin. Hydrolysis of these chains can be accomplished by either the random breakdown of the internal C1-1, C-glycosidic, linkages (endopolygalacturonases) or by the release of the digalacturonic acid (exo-poly-a-D-galacturonosidase) from the free ends of the pectin chains. Endo-polygalacturonases have been found to have variable forms which ranged in molecular weights from 30 to 80 kDa, with an acidic optimum pH range of 2.5-6.0 and temperature optima of 30-50°C [28,29]. Exo-polygalacturonases are commonly found in Pectobacterium sp. and some other bacterial, fungal and plant species with molecular weight ranges of 30 and 50 kDa [30,31].

Pectobacterium carotovorum subsp. carotovorum (P. carotovorum), formerly referred to as Erwinia carotovora subsp. carotovora, is a phytopathogenic bacterium that causes soft rot. Plant cell wall maceration by P. carotovorum has been shown to occur through the action of cellulases, β-glucosidases and proteases. Moreover, a group of pectinolytic enzymes including polygalacturonase (PG), pectin lyase (PNL), and pectin methyl esterase (PME) have been recognized to have critical roles in plant cell wall penetration leading to maceration by P. carotovorum [32]. These enzymes have been found in various isozymes that are believed to bring about more effective hydrolysis of the plant cell wall polysaccharides [33]. Pectinases of P. carotovorum, along with a few other bacterial species, have the specific property of being alkaline tolerant, which is useful in many industrial applications [33]. celA, celB and celC encoded cellulases have been previously isolated and characterized from P. carotovorum LY43 [34]. The role of these cellulases in sugar uptake systems inside the cell was described in the Bacillus stearothermophilus B, Bacillus carotovorum encoding cellulases and polygalacturonase into E. coli B, cells. Specifically, celA, celB and celC cloned into E. coli carotovorum subsp. carotovorum (PNL), and pectin methyl esterase (PME) have been recognized to have variable forms which ranged in molecular weights from 30 to 50 kDa [30,31].

The aim of this study was to clone and express the genes of P. carotovorum encoding cellulases and polygalactanase in E. coli cells. Specifically, celB, celC and peh of P. carotovorum were amplified by polymerase chain reaction and cloned in an expression vector to be expressed in E. coli. The cloned genes were sequenced and catalytically active residues were identified in the deduced amino acid sequences based on homologies with known degradative enzymes. Characterization of the enzymes and evaluation of their ability to degrade recalcitrant products found in citrus waste are also part of this study. This work could lead to a low cost system using a genetically recombinant E. coli for the direct citrus waste bioconversions into bioethanol. A genetically recombinant E. coli with cellulase and pectinase activities would be used as an ideal candidate for biofuel production.

Materials and Methods

Bacterial strains, plasmids and media

Pectobacterium carotovorum subsp P. carotovorum (P. carotovorum), ATCC™ no. 15359, was used as a source of DNA in this study. Typically, -20°C stored cells were revived by streaking on Luria Bertani (LB) agar, and incubated overnight at 26°C. A single colony was inoculated into 3 ml of LB broth (5 g/ml yeast extract, 10 g/l tryptone, 0.5g/l NaCl) and propagated overnight in an orbital shaking incubator (220 rpm) at 37°C. Approximately, 1.5 ml of the overnight culture was centrifuged at 4,000 x g for 10 min at room temperature and the genomic DNA was isolated from the pellet using the E.Z.N.A.™ bacterial isolation kit protocol (Omega Bio-tek, cat. no. D3350-02, Norcross, GA). Isolated DNA was eluted in deionized, nuclease-free, distilled water and quantitated spectrophotometrically using an Eppendorf Biophotometer (AG. 22331, Hamburg, Germany). E. coli DH5α chemically competent cells (Lucigen, cat. no. 95040-456, Middleton, WI) were used for cloning of genes and expression of the enzymes. Plasmid-containing E. coli was grown in LB broth containing 100 μg/ml ampicillin or plated on antibiotic containing LB agar supplemented with 0.1 mM of isosopropyl β-D-1-thiogalactopyranoside (IPTG) and 40μg/ml of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) for the selection and differentiation between empty-plasmid transformed E. coli and clone-transformed E. coli (blue compared to white, respectively).

pGEM-Teasy vector with a molecular size of 3015bp (Promega, Madison, WI) facilitates cloning of PCR products without prior restriction digestion and was used for general DNA manipulation and DNA sequencing of cloned genes. The expression vector pTAC-MAT-TAG-2 with a 5178 bp molecular size (Sigma Aldrich, cat. no. E5405, St. Louis, MO) was used for heterologous expression of celB, celC and peh in E. coli DH5 α.

Molecular biological techniques

Primer design: The sequences of celB, celC and peh were found on GenBank®, the NIH genetic sequence database (accession numbers AF025769.2, AY188753 and BAA74431.1, respectively). In order to amplify the 3 genes of interest, primers were designed corresponding to the open reading frames (ORFs) of the 3 genes. To facilitate cloning into pTACMAT, restriction enzyme sites were incorporated at the 5’ end of each primer. The designed primers, as well as their corresponding restriction sites indicated by the underscore, are shown in Table 1.

PCR amplification: One hundred micrograms of genomic DNA was used as a template for PCR amplification using a PCR Sprint Thermocycler (Thermo Electron Corporation, Milford MA) under the following conditions: 1 cycle of 95°C for 5 min; 30 cycles of 95°C for 30s, 55°C for 30s, and 72°C for 1-2 min, according to the length of the fragment. The PCR products with sizes of 795 bp (celB), 1105 bp (celC), and 1200 bp (peh) were purified following agarose gel electrophoresis, desalted and concentrated using QIAEX II® Gel extraction Kit (Qiagen, USA).

Cloning and sequencing: The cleaned products were ligated into the pGEM-Teasy vector, transformed into E. coli DH5α chemically

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Restriction sites</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>celBF</td>
<td>Xhol</td>
<td>5’GGCCTCGAGATGCTAGCAGTGAATAGAAAG3’</td>
</tr>
<tr>
<td>celBR</td>
<td>Smal</td>
<td>5’GCGGCCGCGGTTATTTATGTCCTAGCCTC3’</td>
</tr>
<tr>
<td>celCF</td>
<td>EcoRI</td>
<td>5’GGGGAATTCGATGCGAGGTCGACCTAC3’</td>
</tr>
<tr>
<td>celCR</td>
<td>Bgl II</td>
<td>5’GCGGACATCTTTACGCTGTTGCTGTTGCTG3’</td>
</tr>
<tr>
<td>pehF</td>
<td>EcoRI</td>
<td>5’GCGGAAGCTTACGATGCGAGGTCGACCTAC3’</td>
</tr>
<tr>
<td>pehR</td>
<td>Bgl II</td>
<td>5’GCGGACATCTTTACGCTGTTGCTGTTGCTG3’</td>
</tr>
</tbody>
</table>

Table 1: The designed oligonucleotide primers and their corresponding restriction enzyme sites.
competent cells and plated on antibiotic containing LB agar supplemented with 0.1 mM IPTG and 40 µg/ml X-gal. Following overnight incubation at 37°C, white colonies were picked and propagated in LB broth containing 100 µg/ml ampicillin. Plasmid DNA was isolated using the E.Z.N.A.* Plasmid Midi Kit (Omega Bio-tek, USA) and resulting DNA was digested with EcoRI to confirm the presence of the gene of interest and the DNA was sequenced by MCLAB (San Francisco, CA, USA). Nucleotide sequence translation, nucleotide alignments, and the deduced amino acid sequences were performed online using the bioinformatics tools available at http://www.ustjohn.com. The alignment of the obtained sequences were also carried out using basic local alignment search tool (BLAST) available through National Center for Biotechnology information (NCBI; http://www.ncbi.nlm.nih.gov). Confirmed cloned sequences were digested with their respective restriction enzymes (Table 1) and ligated into pTAC-MAT vector and transformed into E. coli DH5α cells. Plasmid DNA was isolated as described above and inserts were confirmed by PCR with the primers used for the original amplification as well as with restriction digests.

Gene expression, enzyme extraction and purification

Freshly inoculated E. coli cells harboring celB and celC and peh were grown in LB broth containing 100 µg/ml ampicillin to an optical density of 0.5 at 595 nm. Gene expression was then induced by the addition of 0.1 mM IPTG and cells were harvested by centrifugation 4 hours later. The empty vector strain was propagated and induced in the same manner as a negative control. Over-expressed soluble proteins were extracted and partially purified using the B-PER bacterial protein extraction kit (Pierce Scientific, cat. no. 90078) with DNAse (1,2,500 unit/ml), lysozyme (50 mg/ml) and nonionic detergent in 20mM Tris-HCl buffer, pH 7.5. An approximate equivalent amount of 100 µg of protein was denatured by incubation at 95ºC for 5 min in loading buffer (0.05% bromophenol blue, 5% β-mercaptoethanol, 10% glycerol and 1% SDS in 0.25M Tris-HCl buffer, pH 6.8). Proteins were separated by electrophoresis through a 10% polyacrylamide gel at 50 mA for approximately 1 h and protein bands were visualized following staining with Gel Code blue stain reagent (Thermo Scientific, cat. no. 24590) for 1 h followed by destaining for another 1 h in deionized distilled water. The molecular weights of the detected protein bands were determined using a molecular weight standard protein marker kit (ProSieve*, cat. no. 50550, USA).

CMC-agar diffusion method for detection of cellulase activities

Carboxymethyl cellulose (CMC) agar was prepared using LB agar containing ampicillin (100 µg/ml), 0.1 mM IPTG, 40 µg/ml X-gal and 1% (w/v) of CMC. Expression of cellulase was confirmed by inoculating the celC or celB plasmid containing E. coli on the inducing medium for 3 days at 37°C and then staining with 0.5% Congo red for 30 min followed by 2 washings with 1M NaCl [43]. E. coli containing an empty pTAC-MAT vector was included as a negative control. Enzymatically active cellulases were indicated by a yellow halo against a red background where the cellulose had been digested and was no longer available to bind Congo red. Sterilized 1% CMC agar based medium was prepared for cellulolytic activity determination of the crude extracts. Twenty ml of the prepared medium was transferred to 100×15 mm Petri plates and 5 mm holes were made in the solidified medium. Thirty-five microliters of the crude E. coli expressed cellulases as well as the extract from the negative control were loaded separately in the corresponding well areas. The plates were incubated 24 h at 37°C and the cellulase activity was determined by staining with 0.5% Congo red as described.

Detection of polygalacturonase activity

The determination of polygalacturonase activity was done using a modified method described by [44]. Polygalacturonic acid-based substrate agar was prepared by dissolving 1% polygalacturonic acid (Sigma, cat. no. 9049-37-0) in 50 mM sodium acetate pH 4.6 containing 0.8% agarose (Sigma-Aldrich, cat. no. A9918). The medium was then heated to dissolve the polygalacturonic acid and agarose, and 0.2% of sodium azide was added after cooling to 60°C. Nine mm diameter wells were made in the agar and 100 µl of the supernatant of recombinant E. coli was loaded in the corresponding wells. Supernatant of empty vector transformed E. coli was used as a negative control. The plates were incubated 24 h and polygalacturonase activity was detected by appearance of a clear white halo after the addition of 6 M HCl.

Results

Cloning and restriction analysis of the recombinant clones

Primers were designed based on published DNA sequences for celB, celC and peh encoding two cellulases and a polygalacturonase enzyme from P. carotovorum. The genes were amplified by PCR and the products were separated by gel electrophoresis in 0.8% agarose and visualized following staining with ethidium bromide. The predicted 795 bp, 1105 bp and 1200 bp fragments were cloned into pGEM-Teasy, sequenced and then expressed in pTAC-MAT-TAG 2. DNA sequence analysis confirmed the identity of the genes and comparison with published sequences resulted in up to 95% similarities with the corresponding genes in P. carotovorum. Sequenced plasmids were digested using the restriction enzymes XhoI and Smal for celB clones, and EcoRI, and Bgl II restriction enzymes for celC and peh clones. The isolated fragments were cloned into pTAC-MAT-TAG®-2 expression vector and recombinant clones were picked based on blue/white selection, and propagated overnight at 37°C in LB broth containing ampicillin (100 µg/ml) for DNA plasmid isolation. The isolated plasmids were digested and the products were analyzed on 0.8% agarose gel (data not shown). The inserts were confirmed by comparing PCR products amplified from the clones with products amplified using genomic DNA from P. carotovorum (Figure 1).

Nucleotide sequence analysis

The translated sequences of the cloned genes were compared with published sequences using the BLAST program available through NCBI. The celB, celC and peh translated ORFs were similar in sequences to those of several endoglucanases, glucan-glucanohydrolases and...
polygalacturonases, respectively. The nucleotide sequences along with their corresponding amino acids were found to exhibit a high degree of similarities with *P. carotovorum* and other organisms possessing cellulase and polygalacturonase activities (Table 2). Based on the available conserved domain sequences, the deduced amino acid sequences of the 3 tested ORFs were found to belong to glycosyl hydrolase families 12, 8 and 28 for celB, celC and peh, respectively [45,46]. The nucleotide length of the tested *celB*, *celC* and *peh* ORFs were determined to be 795 bp, 1105 bp and 1209 bp, with putatively encoding proteins of 266, 369 and 404 amino acids, and with estimated molecular weights of 29.5 kDa, 41.3 kDa and 42.5 kDa, respectively. The nucleotide sequences of the ORFs of the 3 cloned genes and their corresponding amino acids are shown in Figure 2.

**CMC-agar diffusion method for detection of cellulase activity**

Three transformants each of *celC* and *celB* were inoculated on CMC-agar including the negative control, *E. coli* containing an empty pTAC-MAT vector. The agar plates were incubated 3 days at 37°C and then stained with 0.5% Congo red for 30 min followed by 2 washings with 1M NaCl [43]. Enzymatically active cellulases were indicated by a yellow halo against a red background where the cellulose had been digested and was no longer available to bind Congo red. All of the tested strains were positive for cellulases by visual inspection (data not shown) so one representative transformant of each of the cloned cellulases was used for gene expression and characterization. The transformed *E. coli* strains, including the empty-vector negative control strain, were grown and induced for gene expression for 4 hours. The cells were harvested and 35 µl of the supernatants were loaded separately in the corresponding wells in the CMC-agar. The plates were incubated 24 h at 37°C and the cellulase activity was determined by staining with 0.5% Congo red as described before. As shown in Figure 3, the crude supernatants of the *celB* and *celC* transformants contained cellulase activity as indicated by the appearance of the yellow halo against the red background. However, the CMC-cellulase activity of the *celC* protein is much higher than was detected by *celB* encoded protein as indicated by the larger diameter of the yellow halo. Furthermore, no activity was detected in the well containing the negative control supernatant. These results confirm the expression and cellulase activity of the cloned *celB* and *celC* products.

**Detection of polygalacturonase activity**

The *peh*-pTAC-MAT transformed *E. coli* was grown in LB broth containing ampicillin (100µg/ml), 0.1 mM IPTG, 40µg/ml X-gal to induce the expression of the cloned *peh* product, polygalacturonase. After

---

### Table 2: Comparisons of *celB*, *celC* and *peh* translated amino acid sequences with various species using NCBI’s BLAST search.

<table>
<thead>
<tr>
<th>Tested clone</th>
<th>Aligned organism</th>
<th>Enzyme coded</th>
<th>% Identity</th>
<th>% positivity</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>celB</em></td>
<td><em>P. carotovorum</em> subsp. <em>carotovorum</em></td>
<td>beta(1,4)-glucan glucanohydrolase precursor</td>
<td>97%</td>
<td>98%</td>
<td>AAC02965.2</td>
</tr>
<tr>
<td><em>ceBP</em></td>
<td><em>P. carotovorum</em> subsp. <em>carotovorum</em> PC1</td>
<td>Cellulase</td>
<td>96%</td>
<td>97%</td>
<td>YP_003017082.1</td>
</tr>
<tr>
<td><em>B. licheniformis</em> WX-02</td>
<td>glycoside hydrolase protein family</td>
<td>84%</td>
<td>90%</td>
<td>YP_003260319.1</td>
<td></td>
</tr>
<tr>
<td><em>P. wasabiae</em> WPP163</td>
<td>Cellulase</td>
<td>66%</td>
<td>81%</td>
<td>ZP_1765456.1</td>
<td></td>
</tr>
<tr>
<td><em>P. wasabiae</em> WPP14</td>
<td>endo-1,4-D-glucanase</td>
<td>53%</td>
<td>74%</td>
<td>YP_00314236.1</td>
<td></td>
</tr>
<tr>
<td><em>cePC</em></td>
<td><em>P. carotovorum</em> subsp. <em>carotovorum</em> PC1</td>
<td>Cellulase</td>
<td>92%</td>
<td>94%</td>
<td>refYP_003015672.1</td>
</tr>
<tr>
<td><em>Serratia marcescens</em> FGI94</td>
<td>endoglucanase Y</td>
<td>87%</td>
<td>91%</td>
<td>YP_00257546.1</td>
<td></td>
</tr>
<tr>
<td><em>Yersinia intermedia</em> ATCC 29909</td>
<td>Endoglucanase</td>
<td>70%</td>
<td>79%</td>
<td>YP_007342699.1</td>
<td></td>
</tr>
<tr>
<td><em>peh</em></td>
<td><em>P. carotovorum</em></td>
<td>polygalacturonase</td>
<td>99%</td>
<td>99%</td>
<td>AAA03624.1</td>
</tr>
<tr>
<td><em>P. carotovorum</em></td>
<td>polygalacturonase</td>
<td>99%</td>
<td>99%</td>
<td>BAA74431.1</td>
<td></td>
</tr>
<tr>
<td><em>P. atrosepticum</em> SCR11043</td>
<td>endo-polygalacturonase</td>
<td>95%</td>
<td>97%</td>
<td>YP_049201.1</td>
<td></td>
</tr>
<tr>
<td><em>Erwinia amylovora</em> DSM 12163</td>
<td>polygalacturonase</td>
<td>60%</td>
<td>73%</td>
<td>YP_005802329.1</td>
<td></td>
</tr>
<tr>
<td><em>Erwinia amylovora</em> ATCC BAA-2158</td>
<td>polygalacturonase</td>
<td>59%</td>
<td>73%</td>
<td>CBX81050.1</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Nucleotide sequences of the ORFs of the cloned genes A) celB, B) celC and C) peh from P. carotovorum, ATCC no. 15359. The amino acid residues are placed below the nucleotides of the corresponding codon. The catalytic active site residues of each gene are underlined.
4 hours of incubation, the cells were harvested and the supernatant was collected and tested for enzymatic activity. The negative control strain of *E. coli* transformed with the empty pTAC-MAT vector was treated in the same manner. One hundred µl of the respective supernatants were incubated in the wells incorporated in the polygalacturonic acid agar plates and incubated for 24 h. Polygalacturonase activity was indicated by the appearance of a white halo around the wells after the addition of 6 M HCl. As shown in Figure 4, polygalacturonase activity was detected in the supernatant of the *E. coli* transformed with peh-pTAC-MAT but it was not detected in the well of the negative control (panels A and B, respectively). These results clearly indicate expression and activity of polygalacturonase from the cloned *peh* gene.

**SDS-PAGE for molecular weight determination**

The molecular weights of the expressed *celB*, *celC* and *peh* products were determined using the SDS-PAGE method. The crude extracts of the expressed *E. coli* transformed with the cloned *celB*, *celC* and *peh* genes, as well as the corresponding semi-purified fractions with maximum cellulosytic and pectinolytic activities were used for molecular weight determination. The molecular weights of the protein products found in the supernatant of the negative control *E. coli* were also analyzed. As shown in Figure 5, protein bands of approximate molecular weights of 29.5, 40, and 41.5 kDa were detected in the lanes loaded with the crude supernatants of the expressed *celB*, *celC* and *peh* along with their representative semi-purified fractions, respectively. Similar molecular weights were previously detected in the protein products of the cloned genes of *P. carotovorum* [34,47,48]. In contrast, the lane loaded with the crude supernatant of the negative control of *E. coli* had no bands in the aforementioned molecular weight ranges. These results indicate the putative expression of the 3 cloned genes.

**Discussion**

Glycoside hydrolases (GHs) are comprised of a wide range of enzymes that are capable of hydrolyzing the glycosidic bonds in glycosides, glycans and glycoconjugates [49]. GHs are playing a critical role in biofuel production through their wide application in production of reducing sugars from pre-treated biomass materials. The reducing sugars formed are useful substrates for ethanol and butanol production which, indeed, can be used as renewable sources for gasoline [50].

Based on the available conserved domain sequences in the NCBI website, the deduced amino acid sequences of the 3 cloned ORFs were found to belong to GHs-12, 8, 28 for the cloned *celB*, *celC* and *peh*, respectively [45,46]. The enzymes belonging to GH-28 family are classified into several categories in accordance with their catalytic hydrolysis mechanism and are given specific E.C. numbers based on the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (http://www.brenda-enzymes.org/). Comparison of the deduced amino acid sequences encoded by *peh* with those GH amino acid sequences of various species using NCBI’s BLAST search revealed that the cloned gene belongs to the polygalacturonase category of the GH-28 enzymes. Polygalacturonases (PG) are a group of enzymes that function in the hydrolysis of the α-linkage of the galacturonic acid (GalA) monomer residues in pectin (E.C. 3.2.1.15 [endo-PG] and 3.2.1.67 [exo-PG]) [51]. Genes encoding GH-28 enzymes have been identified in a number of plant pathogens and fungal species [52]. These degradative enzymes found in pathogenic microorganisms may act as virulence factors and play a critical role in plant cell wall maceration [53,54]. *P. carotovorum* is a known plant pathogen and the presence of the GH-28 polygalacturonase gene in the genome supports its pathogenic role in the hydrolysis of plant cell wall pectin. The role of the *peh* product in plant tissue maceration along with the other expressed enzymes has been investigated by many researchers [55,56]. The GH-28 polygalacturonases are characterized by the presence of 4 conserved amino acid regions (NTD, DD, HG, and RIK) which were thought to be implicated in the catalytic mechanism [57]. As indicated in Figure 2, the aforementioned conserved amino acid regions were found to be present in the deduced amino acid sequence of the *peh* encoded protein. Comparison of these results with those of others [58] reveals replacement of a histidine residue with an arginine residue in their identified *pehN* gene of *E. chrysanthemi* 3937. The possible replacement of the *pehN* gene was inferred to suggest a mechanism adapted to different substrate specificities [58]. Histidine (H) was also recognized as a conserved active site residue among fungal and bacterial GH-28 polygalacturonases with an identified motif of ([G/S]/D/E/N/K/R/H*-][x2( V/M/F/C)-x2( G/S)]H*-G-(L/I/V/M/A/G)-x(1,2)-(L/I/V/M)-G-S) [59]. The deduced amino acid sequence of the cloned *peh* was found to include the histidine active site residue in the following motif ([H*]- N-E- (F) -G-T- (G)- H*-G-(L/I/V/M/A/G)-x(1,2)-(L/I/V/M)-G-S) [59].

Rye et al. noted that the cleavage of glycosidic bonds by several glycoside hydrolases can be achieved by either a single- or double-displacement mechanism, which gives rise to inversion or retention of anomeric configuration, respectively [60]. Hydrolysis of the homogalacturonan and the rhamnogalacturonan components of the pectin chain by GH-28 polygalacturonases has been revealed to be a single-displacement inverting mechanism [59,61]. Pickersgill et al. have mapped the active site of the *pehA* encoded GH-28 polygalacturonase of *P. carotovorum*. Three aspartate residues: D202, D223 and D224 have been found to be involved in the catalytic mechanism of this polygalacturonase [62]. The 3 catalytic aspartate residues have also been recognized to be functionally conserved between exo-and endo-acting polygalacturonases [59]. Site-directed mutagenesis studies
of endopolypgalacturonase II from *Aspergillus niger* revealed that these residues have a role in protonation and deprotonation actions throughout the hydrolysis process [63]. Mutation in any of these residues was found to have a significant negative influence in the catalytic activity of the enzyme. As shown in Figure 2, the 3 putative catalytically active aspartate residues D228, D249 and D250, are present and have the same spatial alignment in the amino acid sequences of the cloned *peh* protein as in the *pehA* protein. These results along with those obtained from NCBI’s BLAST-x alignments of the other bacterial and fungal species are evidence for identifying the cloned *peh* gene as an encoding gene for a GH-28 polygalacturonase enzyme.

The NCBI BLAST results of the deduced amino acid sequences of cloned *celB* and *celC* with the other available conserved sequences showed a high degree of homology with the cellulases of *P. carotovorum* and other bacterial species. These results indicate that the cloned aforementioned genes are a part of a cellulose cluster encoding cellulases that may be transcribed by a promoter placed directly upstream of the genes in the chromosome of *E. coli*. SDS-PAGE for molecular determination of expressed proteins from *E. coli* showed a high degree of homology with the cellulases of *P. carotovorum* and other bacterial species. These results indicate that the cloned aforementioned genes are a part of a cellulose cluster encoding cellulases that may be transcribed by a promoter placed directly upstream of the genes in the chromosome of *E. coli*.

Analysis of the three-dimensional structures of the GH-families’ proteins is considered the best approach for identifying the enzymatic nucleophilic residues. These residues are generally located in the active site clefts with notable characteristics of being conserved, polar and hydrogen bonded [67]. The GH-12 family was among many other families whose members were found to have a glycosidase mechanism with net retention of anomeric configuration [68,69]. In the retention mechanism, hydrolyses are generally achieved through a double-displacement mechanism in which a glycosyl-enzyme intermediate is initially formed followed by its hydrolysis through oxocarbenium-ion transition states [70]. On the other hand, the GH-8 family includes enzymes that catalyze the glycosidic hydrolysis through a single-displacement inverting mechanism with α-configuration products [71,72]. In the inverting mechanism, a carboxyl group (acting as the general acid) and a carboxylate group (acting as the general base) are considered as the two main functional groups. These two functional groups were found to have an important role in the hydrolysis process and are generally conserved among a particular family. GH-12 family members were found to have a generally conserved glutamic acid (E) residue with the following motif: [E*-(I/L)-M-(I/V)-W], [73], which was believed to play a key catalytic role as a nucleophile [74]. Another conserved glutamic acid residue identified in a conserved motif of [G-(T/F)-E*] was also recognized as acid-base residue catalysts in the BLAST search comparing twelve known GH-12 cellulase sequences performed [73]. Comparisons with other retaining glycosidases revealed that a general acid-base catalytic mechanism is less consistent compared with a nucleophile catalytic mechanism [75,76]. Both the nucleophile conserved sequence (E*-(L)-M-(I)-W) and the acid-base residues sequence (G-(T)-E*) have been detected in cloned *celB* (amino acid residues 178-182 and 244-246, respectively). This investigation along with the data obtained from NCBI’s BLAST-x search strongly confirm that the cloned *celB* is an encoding gene for a GH-12 group member.

Concerning the GH-8 group, amino acid residues with the following pattern [A-(S/T)-D-(A/G)-D-X(2)-(I/M)-A-X-(S/A)-L/I/V/M)-(L/I/V/M/G)-X-A-X(3)-(F/W)] have been identified as a conserved region [77]. The deduced amino acid sequences of the cloned *celC* have been shown to involve similar GH-8 conserved residues with the following pattern: [A-(S)-D-(A)-D-(L)-W)-(I)-A-(Y)-(N)-(L)-(L)-A-(G-R-L)-(W)], (Figure 2b). Glutamic acid and aspartic acid have been recognized as conserved catalytically active amino acid residues in GH-8 [77,78] with the following motif [(T/V)-S-E*- (G/A)-(Q/H/LM)] and [D-(G/A)- D*- (L/M/E)], respectively [79]. These residues have been observed in the amino acid sequence of the cloned *celC* with similar motif: [(T)-S-E*- (G)-(Q)] and [D-(A)- D*- (L)]. The conserved

Figure 5: SDS-PAGE for molecular determination of expressed proteins from *E. coli*. Panel A – *peh*; Panels B and C – *celB*; Panel D – *celC*. Lanes 1, 2, 7, 9, semi-purified proteins; Lanes 4, 6, 8, 12. Prosieve protein marker mixture; Lanes 3, 6, 10 crude *peh*, *celB* and *celC* protein extract, respectively, and Lane 12, crude extract of the negative control. Bands with molecular weights of approximately 29.5, 40 and 41.5 kDa are indicated in the representative lanes of *peh*, *celB* and *celC*, respectively. No bands in these ranges were detected in Lane 12 confirming the expression of the tested cloned genes.
that has been developed by researchers at TAMUK will be a helpful tool in understanding the catalytic action of the expressed enzymes in the bioconversion process.

Acknowledgements

The authors are grateful to Prof. Keith E. Taylor, Department of Chemistry and Biochemistry, University of Windsor, for his helpful comments on the manuscript.

References


36. Cote CK, Honeyman AL (2003) The LiCt protein acts as both a positive and a negative regulator of loci within the bi1 regulon of Streptococcus mutans. Microbiology 149: 1333-1340.


