

Modification and Expression of Beta-1,4-Endoglucanase encoding sequences of fungal origin in *Escherichia coli* BL21

Modificación y expresión de secuencias codificantes de Beta-1,4-Endoglucanasa de origen fúngico en *Escherichia coli* BL21

Narmer Fernando Galeano Vanegas*, **Natalia Gutiérrez Calle****, **Gloria María Restrepo Franco*****

DOI: 10.15446/rev.colomb.biote.v22n2.79448

ABSTRACT

Lignocellulose is the main and most abundant component of biomass. Annually, 200 million tons are generated in the world. Colombia has a high production of lignocellulosic residues that can be used in many industrial processes such as bioethanol production, promoting the bioeconomy. The objective of the present work was to express lignocellulolytic enzymes of eukaryotic origin in *Escherichia coli* BL21 (DE3). Initially, endoglucanase eukaryotic genes were selected and modified using bioinformatics methods for their production in *E. coli* BL21 (DE3) and saccharification of pure cellulose substrates. The gene selected for its modification and expression was *eglB* from the fungus *Aspergillus nidulans*. Subsequently the enzyme integrity was tested by 3D modeling and molecular docking, as well as the conformation of its active site and its affinity for substrates of interest. Finally, cloning of the modified gene in plasmid pET151 TOPO was made and transformed in the strain *E. coli* BL21 (DE3) where several lignocellulose degradation tests were carried out using semiquantitative methods for the enzyme activity in carboxymethylcellulose. The presence of the three genes of interest within the plasmid pET151 TOPO and within the transformed cells of *E. coli* TOP10 and *E. coli* BL21 (DE3) was verified by colony PCRs performed. The presence of this gen was corroborated by sequencing. Expression of the modified endoglucanase enzyme was achieved in *E. coli* BL21 (DE3) expression cells, in soluble and functional form, demonstrated by the hydrolysis of the CMC substrate.

Key words: Cellulose, Heterologous expression, Cellulolytic enzymes.

RESUMEN

La lignocelulosa es el componente principal y más abundante de la biomasa. Anualmente se generan 200 millones de toneladas en el mundo. Colombia tiene una alta producción de residuos lignocelulósicos que pueden ser utilizados en muchos procesos industriales como la producción de bioetanol, promoviendo la bioeconomía. El objetivo del presente trabajo fue expresar enzimas lignocelulolíticas de origen eucariota en *Escherichia coli* BL21 (DE3). Inicialmente, los genes eucariotas de endoglucanasa se seleccionaron y modificaron mediante métodos bioinformáticos para su producción en *E. coli* BL21 (DE3) y la sacarificación de sustratos de celulosa. El gen seleccionado para su modificación y expresión fue *eglB* del hongo *Aspergillus nidulans*. Posteriormente se evaluó la integridad de la enzima mediante modelado 3D y acoplamiento molecular, así como la conformación de su sitio activo y su afinidad por sustratos de interés. Finalmente, se realizó la clonación del gen modificado en el plásmido pET151 TOPO y se transformó en la cepa *E. coli* BL21 (DE3) donde se realizaron varios ensayos de degradación de lignocelulosa utilizando métodos semicuantitativos para la actividad enzimática en carboximetilcelulosa. La presencia del gen de interés dentro del plásmido pET151 TOPO y

* Universidad Católica de Manizales. ORCID 0000-0002-8337-2824. Colombia. ngaleano@ucm.edu.co

** Universidad Católica de Manizales. Colombia. natalia.gutierrez1@ucm.edu.co

*** Universidad Católica de Manizales. ORCID 0000-0003-0443-0369. Colombia. grestrepo@ucm.edu.co

dentro de las células transformadas de *E. coli* TOP10 y *E. coli* BL21 (DE3) se verificó mediante PCR de colonia. La presencia de este gene se corroboró por secuenciación *egfB*. La expresión de la enzima endoglucanasa modificada se logró en células de *E. coli* BL21 (DE3), en forma soluble y funcional, demostrada por la hidrólisis del sustrato de CMC.

Palabras claves: Celulosa, Expresión heteróloga, Enzimas celulolíticas.

Recibido: mayo 2 de 2019 **Aprobado:** noviembre 20 de 2020

INTRODUCCIÓN

Bioethanol can replace part of the consumption of traditional fuels such as oil or coal. It is possible to mix gasoline with bioethanol to offer savings in CO₂ emissions of at least 50% compared to fossil fuels (Hernández Rodríguez & Hernández Zárate, 2008). However, because bioethanol is traditionally manufactured with lignocellulose from crops such as corn, the price of food tends to increase, which becomes a reason to focus efforts on the use of lignocellulose from lignocellulosic waste for its production.

The lignocellulosic waste is composed of cellulose, hemicellulose and lignin and around 200,000 million tons are produced annually (Medina-Morales, Lara-Fernández, Aguilar, & de la Garza-Toledo, 2011; Saini, Saini, & Tewari, 2015). Colombia has a high production of lignocellulose, which can be used to produce bioethanol. If waste from Colombian industries such as citrus juice and rice husks were used, 136,045,500 L and 12,200,000 L of bioethanol would be obtained annually (Sánchez Riaño *et al.*, 2010). Taking into account the price of the gallon of ethanol of U\$ 3.07 Colombian pesos (COP) (August 2018), approximately U\$ 120,240,897 would be obtained (Sánchez Riaño *et al.*, 2010).

Lignocellulose is currently hydrolyzed by mechanical breakdown followed by acid, alkaline and / or enzymatic hydrolysis (Persson, Tjerneld, & Hahn-Hägerdal, 1991; Cuervo, Folch, & Quiroz, 2009; Castillo *et al.*, 2012). Of these three methods, the most promising by specificity and high recovery of glucose is enzymatic hydrolysis, with the disadvantage that the production of enzymes has high costs (Gao *et al.*, 2008, G. Liu, Qin, Li, & Qu, 2013), which can be reduced if genetically modified microorganisms are used, to produce lignocellulolytic enzymes more efficiently in large quantities and in shorter times (Gao *et al.*, 2008; Zelena, Eisele, & Berger, 2014).

Degradation of lignocellulose is carried out for a cellulolytic system composed of endoglucanase, exoglucanase and cellobiosidase (Munjal, Jawed, Wajid, & Yazdani, 2015).

According to the CAZy database, enzymes that degrade cellulose and other polysaccharides from plant cell wall,

are classified into the glycosyl hydrolase families. In these GH families enzymes with the same enzymatic activity are not necessarily gathered, this classification is based on the sequence and structure of the protein, therefore families will understand different enzymatic activities (Busk, Lange, Pilgaard, & Lange, 2014).

The objective of this work was to evaluate the expression of endoglucanase fungal gene in *E. coli* BL21 (DE3) bacterium for the saccharification of lignocellulose residues, through the modification and cloning of eukaryotic genes of lignocellulolytic enzymes.

MATERIALS AND METHODS

Selection of eukaryotic genes encoding lignocellulolytic enzymes

The identifiers of the GH family were downloaded from the CAZy database containing the endoglucanase enzymes. Using these identifiers, the sequences of all the GH enzymes were downloaded from the GenBank, PDB and Uniprot databases (Berman *et al.*, 2000, Magrane & Consortium, 2011).

Subsequently, a search was made of similar sequences to GH families enzymes using the BLAST program (Altschul *et al.*, 1990), identifying functional protein domains annotated in Pfam (<http://pfam.xfam.org/>) (Finn *et al.*, 2016) and Prosite (<http://prosite.expasy.org/>) (Sigrist *et al.*, 2013). Later, domains such alignments were downloaded from Uniprot and each domain function were consulted and downloaded. Redundant sequences at 100% were eliminated with the CD-Hit software 4.6 (W. Li & Godzik, 2006).

Eukaryotic endoglucanase genes from fungi were selected. Selection criteria included presence of the specific domains for each group of enzymes, curation degree, presence of complete sequences in the databases, the size of the enzymes equal or greater than 60 kDa, ability to act extracellularly and the hydrolysis effectiveness of enzymes on specific substrates. The genes of the selected enzymes were annotated using InterproScan (Jones *et al.*, 2014) to check their molecular function and biological process.

Modification of selected genes

Selected genes CDS were obtained using the FGENESH software (Salamov & Solovyev, 2000). A codon preference adjustment was made for the standard organism *Escherichia coli* W3110, genetically similar to *E. coli* BL21 (DE3) (Daegelen *et al.*, 2009), using the Visual Gene Developer software (Gvritishvili *et al.*, 2010; Jung & McDonald, 2011). Each of the modified sequences were checked using BLAST to verify that there was no loss of information and to corroborate that there was no loss of molecular function or biological process. Each sequence was annotated using Interproscan (Jones *et al.*, 2014). Each of the genes was synthesized by GENEWIZ USA, in plasmid pUC57 for preservation, propagation and subsequent cloning in the expression vector pET151 TOPO.

3D modeling of modified proteins

3D modeling of the recombinant modified proteins was carried out with the I-Tasser software (J Yang *et al.*, 2015), using as template the amino acid sequences of *eglB* modified genes in order to check their protein structure, its affinity with specific ligands and its molecular function. To perform 3D modeling quality were used Ramachandran plot generated by the Swiss-PdbViewer software (Guex & Peitsch, 1997).

Molecular docking of 3d modeling versus pure substrates

Molecular docking was performed using SwissDock program (Schwede *et al.*, 2003; Grosdidier *et al.*, 2011), selecting the best 3D model with the C-score closest to 1. In the same way, the ligands were selected with the C-score closest to 1: (i) The ligand selected for docking with the 3D model of the endoglucanase enzyme was the cellobiose (CBI) in the PDB database (<https://www3.rcsb.org/ligand/CBI>).

The UCSF CHIMERA software (Pettersen *et al.*, 2004) was used to elaborate the molecular analyzes and the docking graphs. OpenBabel software (Guha *et al.*, 2006) were run to prepare the selected ligands according to the 3D topology to meet requirements of the SwissDock server.

Transformation of modified genes in *E. coli* Top10 cells for their conservation and propagation

The cloned and synthesized *eglB* gene in the plasmid pUC57 were transformed into *E. coli* Top10 competent cells for their stabilization and propagation, following manufacturer transformation protocol (FE Young & Spizizen, 1961; Hanahan, 1983; Inoue, Nojima, & Okayama, 1990, BD Hanahan *et al.*, 1991). Then, positive transformants were selected and conserved by cryopreservation at -80 °C in 15% glycerol.

Positive transformants were inoculated in LB broth with ampicillin (100 µg / ml), to grow for 8 hours at 37 °C. Plasmid DNA extraction was performed following the standard procedure of alkaline lysis (Birboim & Doly, 1979; Engbrecht, 1990). The results of the extraction were visualized in 1% agarose gel. Subsequently, concentration and quality of extracted DNA was made using a nanodrop spectrophotometer (NanoDrop ND-2000 from Thermo Scientific).

Directional cloning of modified genes in expression vector pET151 TOPO

A standard PCR was run, preparing a reaction of 25 µl for each pUC57 plasmid DNA (Polymerase Platinum taq; DNA constructs of 1-10 ng per reaction; Primers 0.5 mM; dNTPs 0.2 mM, MgCl₂ 25 mM and 10X PCR Buffer). For the sequence of the *eglB* gene, the forward initiator Eg_CS_F and the reverse primer Eg_CS_R were used. Primers were designed according to the indications of the Champion™ pET151 Directional TOPO® Expression kit from Invitrogen (Invitrogen™, 2010).

The thermal profile in the thermal cycler was constructed as described below: initial denaturation at 95° C for 5 minutes, followed by 35 cycles, each cycle with the cycle denaturation at 92° C for 2 minutes, then primer alignment to 57° C for 20 seconds and the extension at 72° C for 2 minutes a final extension of 72° C for 10 minutes was performed. PCR fragments was cloned following manufacturer cloning protocol in the TOPO vector (Invitrogen™, 2010).

Constructs transformation in *E. coli* BL21 (DE3) for enzyme expression

The constructs were transformed into *E. coli* BL21 (DE3) cells (BD Hanahan *et al.*, 1991; Invitrogen, 2013). Presence of genes were checked in colonies selected as positive transformants using standard colony PCR (Invitrogen™, 2010) with appropriate primers.

Plasmid DNA extractions were performed for gene constructs in the pET151 TOPO vector, transformed in *E. coli* BL21 (DE3), and sequenced with the T7 promoter and T7 terminator primers. The integrity of the genes was corroborated using BLAST2seq and its orientation within the expression plasmid with the CLC program Main Workbench (www.qiagenbioinformatics.com/products/clc-main-workbench/).

Expression of the endoglucanase recombinant enzyme

Twenty microliters of cells transformed with *eglB* gene in *E. coli* BL21 (DE3) were inoculated on surface of CMC agar (10 g of CMC, 1 g of KH₂PO₄, 0.5 g of MgSO₄·7H₂O, 0.01 g of FeSO₄·7H₂O, 0.01 g of

Table 1. Selected CAZy families containing enzymes Endoglucanases.

CAZy Family	Mechanism	Prosite ID	CAZy NOTE	Members Number
GH5 (56 subfamilies)	Retaining	PS00659	Once known as cellulase family A; many members have been assigned to subfamilies as described by Aspeborg et al. (2012) BMC Evol Biol. 12(1):186 (PMID: 22992189).	12233 Total 93 Archaea 9874 Bacteria 2139 Eukaryote 127 others
GH6	Inverting	PS00655	formerly known as cellulase family B. The cellobiohydrolases of this family are widely believed to act processively from the non-reducing ends of cellulose chains to generate cellobiose.	1334 Total 1144 Bacteria 187 Eukaryota 16 others
GH7	Retaining	NA	formerly known as cellulase family C. The cellobiohydrolases of this family act processively from the reducing ends of cellulose chains to generate cellobiose. This is markedly different from the IUBMB definition of cellobiohydrolases (EC 3.2.1.91), which act from the non-reducing ends of cellulose.	5134 Total 10 Bacteria 5120 Eukaryota 4 others
GH8	Inverting	PS00812	formerly known as cellulase family D	3838 Total Bacteria 3809 11 Eukaryota 18 others
GH9	Inverting	PS00592	formerly known as cellulase family E	2771 Total 7 Archaea 1598 Bacteria 954 Eukaryota 212 others
GH12	Retaining	NA	formerly known as cellulase family H	882 Total 90 Archaea 627 Bacteria 164 Eukaryota 1 Others
GH44	Retaining	NA	formerly known as cellulase family J. Note that the stereochemistry of the reaction has recently been corrected: Kitago et al, <i>J. Biol. Chem.</i> 282(2007)35703-11 (PMID: 17905739)	154 Total 149 Bacteria 3 Eukaryota 2 Others
GH45	Inverting	PS01140	formerly known as cellulase family K; distantly related to plant expansins	385 Total 23 Bacteria 354 Eukaryota 8 Others
GH48	Inverting	NA	formerly known as cellulase family L. Some cellobiohydrolases of this family have been reported to act from the reducing ends of cellulose (EC 3.2.1.-), while others have been reported to operate from the non-reducing ends to liberate cellobiose or cellotriose or cellotetraose (EC 3.2.1.-). This family also contains endo-processive cellulases (EC 3.2.1.-), whose activity is hard to distinguish from that of cellobiohydrolases.	1032 Total 995 Bacteria 34 Eukaryota 3 others
GH51	Retaining	NA	NA	2712 Total 20 Archaea 2486 Bacteria 165 Eukaryota 41 others

Table 2. Description of the selected endoglucanase enzyme.

Identifier	CAZy family	Enzyme description	Fungi	Interproscan GO Term Prediction
GenBank: AAM54071 Uniprot: Q5B7R2	GH7	Endo-beta-1,4-glucanasa: actividad extracelular. Su sustrato preferido es la Carboximetilcelulosa soluble. Tamaño 48 kDa. pH óptimo 5,5. Temperatura óptima 42 °C. (Lockington et al, 2002; Galagan et al., 2005; Bauer et al., 2006)	<i>Aspergillus nidulans</i> Ascomycota	Función molecular: actividad hidrolasa, hidroliza compuestos O-glicosilados

MnSO₄.H₂O, 0.3 g of NH₄NO₃ and 12 g of bacteriological agar, in one liter of distilled water with pH 7.0) (Ariffin et al., 2006).

Then plates were incubated at 37 ° C for 48-72 hours. Then, agar was stained with 1% congo red and 1M NaCl. The hydrolysis rings were quantified using a colorimeter. Measurement was taken from the outer end of the bacterial colony to the end of the hydrolysis halo on CMC agar (Meddeb-Mouelhi, Moisan, & Beauregard, 2014; Montoya et al., 2014).

Statistical analysis

To check which of the evaluated treatments had the highest degrading activity, an ANOVA analysis was performed, and a comparison of means was made using the Duncan method. Three technical replicas were made for each measurement. For each of four treatments evaluated (i) Treatment 1: negative control, *E. coli* BL21 (DE3) without transformation. (ii) Treatment 2: experimental control, *Bacillus xiamenensis*. (iii) Treatment 3: positive control, *E. coli* BL21 (DE3) transformed with plasmid pET151 TOPO without insert. (iv) *E. coli* BL21 (DE3) transformed with plasmid pET151 TOPO with the modified *eglB* gene.

RESULTS AND DISCUSSION

Ten families sequences belonging to CAZY database were downloaded from corresponding database (Figure 1), the identification catalytic regions from endoglucanase enzymes from literature search was done, as a result, an eukaryotic gene was selected from *Aspergillus nidulans* (Table 2).

Endoglucanase presented only one intron and the modification of the genes resulted in a sequence with different length from the original (Berg et al., 2002).

I-Tasser program predicted ten template structures, all of which are cellulase enzymes with a z-score greater than 1, which indicates that the alignments have a high quality (Roy et al., 2011). The program predicted 3 models

for this enzyme, suggesting that the models used as a mold are remarkably similar to each other (Roy et al., 2011; J Yang & Zhang, 2016). The Ramachandran plot showed that the best model (Figure 1) presents most of the amino acids (which are not glycine or proline) in allowed areas, which indicates a good conformation (Mathews et al., 2002).

The models structurally similar to the problem protein (modified endoglucanase) were ten, which overlapped to the problem structure coincide almost in their entirety in their conformation, with TM-scores greater than 0.5, indicating the reliability of the prediction (J Yang & Zhang, 2016). The prediction of the function using the COACH server (Jianyi Yang et al., 2013) yielded 2 possible ligands: (i) Cellobiose. Identifier in the ChEBI, PubChem and PDB databases: CBI (<https://www3.rcsb.org/ligand/CBI>). (ii) Beta-D-Glucose. Identifier in the ChEBI, PubChem and PDB databases: BGC (<https://www3.rcsb.org/ligand/BGC>).

The first of these substrates (ligands) corresponds to the cellulose monomer, which is the preferred substrate for endoglucanase enzymes and the second corresponds to the molecule that makes up the same cellobiose, so it also has affinity for this enzyme (Bohórquez Saval, 2012). This shows that the active site of the protein was correctly modeled (Roy et al., 2011). The residues in which the two similar ligands are located since both ligands share the amino acids located at positions 194, 218, 221, 232 and 376, which, as shown in Figure 2, are located in the central region of the protein.

The identifiers for the biological function or Enzyme Commission Numbers (E.C) found by the I-Tasser were two: (i) 3.2.1.4 which corresponds to the family of the cellulase enzymes or endoglucanases. (ii) 3.2.1.91, which is the identifier of cellulose 1,4-beta-cellobiosidases (non-reducing end). Both identifiers correspond to enzymes that degrade cellulose. The consensus of the Gene Ontology prediction yielded as result for the molecular function the term with the number

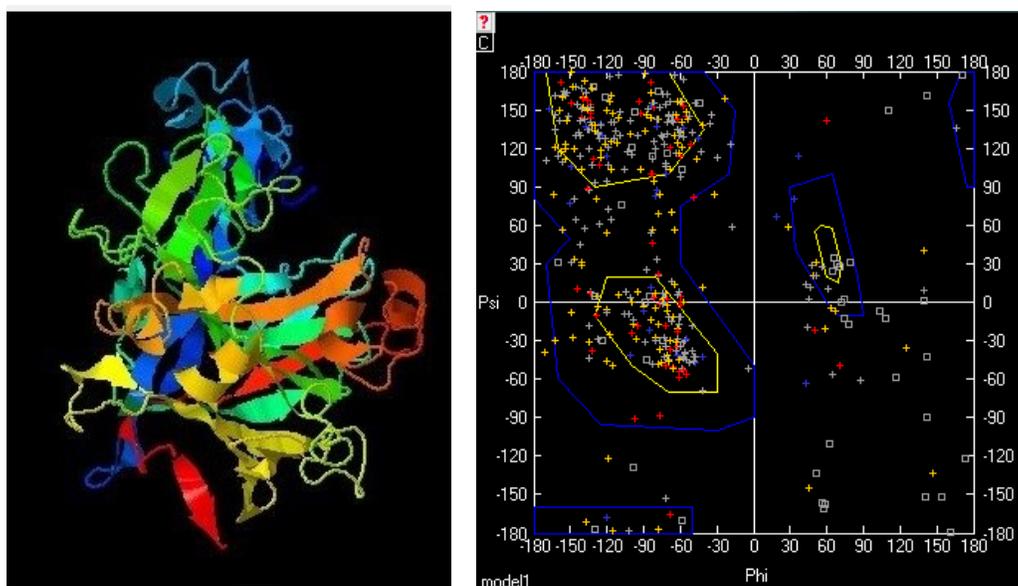


Figure 1. 3D modeling for endoglucanase. (Left) I-TASSER best model for modified endoglucanase. C-score = 0.52 TM-score = 0.78 ± 0.09 RMSD = 5.9 ± 3.7 Å. (Right) Ramachandran graphic generated for the model.

Ranking	C-Score	Cluster Size	PDB Hit	Ligand Name	Binding Site residues
1	0.56	34	2ovwA	CBI	124,164,166,190,192,194,216,218,221,232,374,376
2	0.52	40	1ojkB	BGC	194,218,221,232,267,268,376,385
3	0.32	17	2rg0D	CBI	52,124,166,190,374,376,377
4	0.04	4	3pfzA	CTT	51,52,53,65,96,97,120,121,124,164,166,190,192,199,216,374,376,377
5	0.02	5	2rfzA	CTR	51,52,54,97,120,121,124,376,377

Figure 2. Ligands identified for the modified endoglucanase protein. (Left) Docking site of modified endoglucanase. In blue, amino acids that share the three best ligands. In green, location of the CBI ligand. (Right) Binding sites of the best ligands selected by the I-Tasser for the endoglucanase 3D model.

GO: 0016162 which corresponds to the cellulose activity 1,4-beta-cellobiosidase and the GO term: 0008810, which identifies the cellulase activity. For the biological process, Gene Ontology gave the term GO: 0030245 that corresponds to the catabolic process of cellulose. For the cellular component, the term GO: 0005576 was obtained, which indicates that the protein functions in the extracellular region. The predictions of Gene Ontology yielded results that prove the biological process and molecular function of cellulose degradation. The above corroborates that the modified endoglucanase enzyme

that was modeled by the I-Tasser can perform the degradation function of the cellulose polymer.

The molecular docking of the best model obtained from the I-Tasser against the specific ligands selected yielded mixed results, being able to compare the positions obtained in the SwissDock with the actual location of the ligands obtained by crystallography (G. Wu, Robertson, Brooks III, & Vieth, 2003) for endoglucanase enzyme, with a good result (Figure 3).

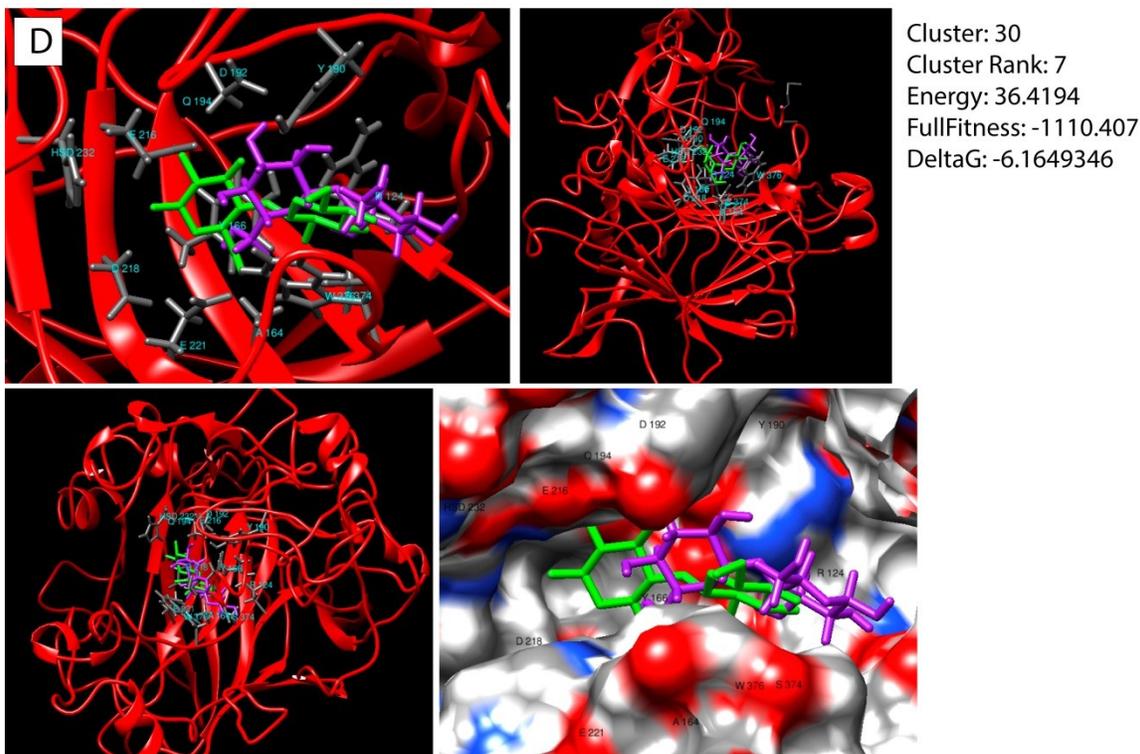


Figure 3. Best predicted docking for cellobiose ligand with the modified endoglucanase enzyme. Green colored molecule is the cellobiose ligand. Violet molecule is docking predicted by the SwissDock. In dark gray color the residues involved in the interaction with the ligand are appreciated.

Table 3. *eglB* gene sequence results obtained from nucleotide BLAST.

Description	Max score	Total score	Query Coverage	E-val.	Identity	Accession
<i>Aspergillus nidulans</i> endo-1,4-beta-glucanase (<i>eglB</i>) gene, complete cds	1905	1905	93%	0.0	98%	AF42002 1.1
TPA <i>Aspergillus nidulans</i> FGSC A4 chromosome IV	1853	1853	93%	0.0	97%	BN00130 6.1
<i>Aspergillus nidulans</i> FGSC A4 hypothetical protein AN3418.2 partial mRNA	1853	1853	93%	0.0	97%	XM.6559 30.1
Synthetic construct 6H-V5-geloin gene, complete cds	154	154	7%	5e - 33	100%	JX301693 .1

PCR of the genes modified for cloning in pET151 TOPO vector, in the agarose gel, yielded bands of DNA fragments. *eglB* gene a band of \approx 1300 bp was observed as expected. (Belancic *et al.*, 1995, Lockington *et al.*, 2002, Chávez *et al.*, 2002, Ruhl *et al.*, 2013).

From each of the colonies obtained in this transformation, the colony PCR was conducted, verifying that the colonies contained the respective amplicon with the expected sizes. The sequencing of PCR products resulted in the sequencing of the genes of interest for one *eglB* gene in pET151 in *E. coli* Top10 identified as E4T3.

The other samples sent for sequencing showed exceptionally low quality of the sequencing and were not analyzed. The only sample obtained for the *eglB* gene coincided with the sequence of the *eglB* gene from the organism *Aspergillus nidulans* by 99% (Table 2).

The analysis of sample E4T3, using the CLC Main Workbench program, to check the orientation of the gene, showed that it was in the correct orientation within the expression vector pET151 TOPO.

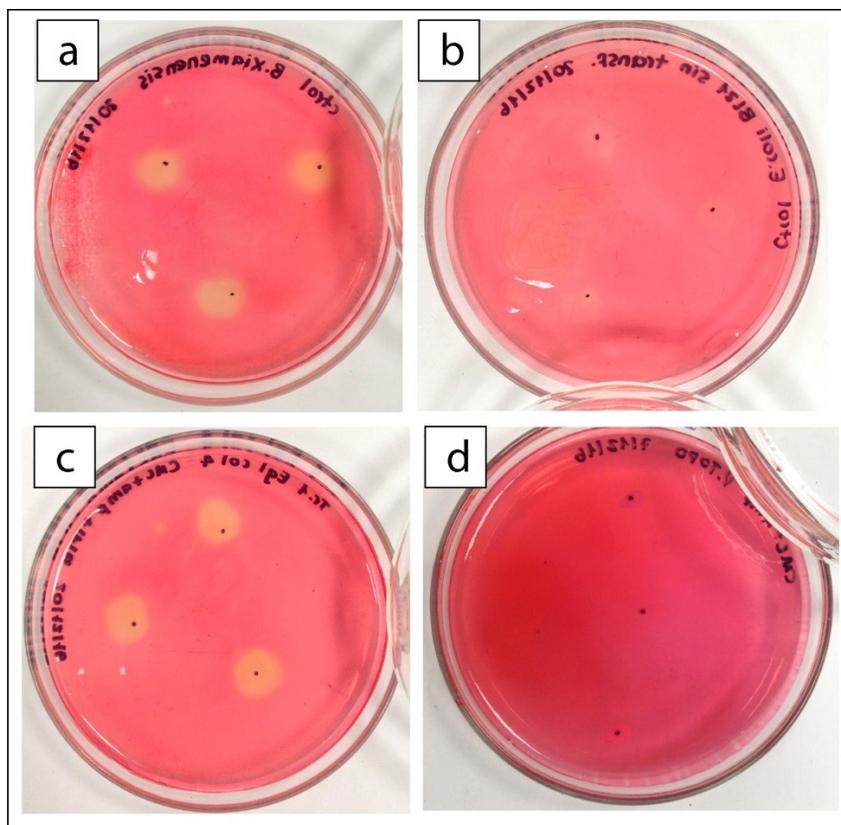


Figure 4. hydrolysis haloes obtained from the third round of enzymatic activity of the endoglucanase recombinant enzyme. (a) Replies 1, 2 and 3. Experimental control *Bacillus xiamenensis*. (b) Replicas 1, 2 and 3. Negative control. *E. coli* BL21 (DE3) untransformed. (c) Replies 1, 2 and 3. *E. coli* BL21 (DE3) transformed with *eglB* gene in pET151 TOPO. (d) Replies 1, 2 and 3. Positive control. *E. coli* BL21 (DE3) transformed with pET151 TOPO vector.

In the three tests of enzymatic activity endoglucanase in *E. coli* BL21 (DE3), the eight replicas evaluated from the experimental control of *Bacillus xiamenensis* (positive control) presented halos of hydrolysis of the medium, between 2.45 and 4 mm, after development with Congo red (Meddeb-Mouelhi *et al.*, 2014; Montoya *et al.*, 2014), which was expected since their endoglucanase activity has already been proven (Lai *et al.*, 2014). (Figure 4). The eight replicates of the *E. coli* BL21 (DE3) transformed with the *eglB* gene in pET151 TOPO presented hydrolysis halos of between 0.24 and 0.46 cm on the CMC agar in the presence of the IPTG expression inducer (Studier & Moffatt, 1986), similar to the hydrolysis obtained for experimental control.

On the other hand, the eight replicas of the negative control, of *E. coli* BL21 (DE3) without transformation, presented faint hydrolysis haloes that were almost imperceptible, showing little growth of colonies on the agar, which may indicate contamination with another bacterium that is able to grow in the agar taking advantage of

the sugars of the CMC, since this is a semi-selective agar, whose only carbon source is the CMC (Meddeb-Mouelhi *et al.*, 2014), which cannot be metabolized by the *E. coli* strain BL21 (DE3) untransformed. As regards the eight replicates of the positive control of *E. coli* BL21 (DE3) transformed with the plasmid pET151 TOPO without insert, they did not present any hydrolysis halo. (Figure 16). These results suggest that the endoglucanase enzyme was expressed correctly in *E. coli* BL21 (DE3) cells after the first transformation.

The statistical analysis was performed for the semiquantitative measurement of the hydrolysis of the recombinant endoglucanase enzyme. The ANOVA analysis resulted in statistical significant differences (p value <0.0001). Duncan's analysis showed the four treatments, grouped into three different groups. The values obtained by the comparison for each treatment were the following: (i) Treatment 1: 0 cm. (Group C) (ii) Treatment 2: 0.29625 cm. (Group B) (iii) Treatment 3: 0 cm. (Group C) (iv) Treatment 4: 0.39375 cm. (Group A). The treatment iv

with the highest hydrolysis halo was *E. coli* BL21 (DE3) transformed with the modified *eglB* gene in the plasmid pET151 TOPO (Group A), followed by the experimental control *Bacillus xiamenensis* (Treatment 2, Group B), which presented the second largest hydrolysis halos.

CONCLUSIONS

The 3D structures predicted by the I-Tasser proved to be very close to the existing structures of similar enzymes in the PDB database, being able to demonstrate that they were structures that were correctly folded that did not undergo major changes in their conformation due to the elimination of introns and modification of codons, which shows that this process did not affect *in silico* the functionality of the enzymes. Regarding the ligation sites, predicted with the SwissDock, were highly conserved sites modeled correctly in the modified enzymes. The results of 3D modeling and docking allowed to corroborate in a preliminary way that these did not suffer important changes that affected their functionality during the process of gene modification.

The sequencing also allowed us to conclude that *eglB* genes cloned within the *E. coli* Top10 and *E. coli* BL21 (DE3) cells underwent very little alteration and remained almost identical to the originally modified genes. T7 promoter and the restriction analysis carried out for each of the constructs, allow to elucidate whether the genes were in the correct orientation within the plasmid for the *eglB* gene, which was in the correct orientation.

The expression of the modified endoglucanase enzyme was achieved in the expression cells *E. coli* BL21 (DE3), in soluble and functional form, demonstrated by hydrolysis of the CMC substrate. Cells from *E. coli* BL21 (DE3) transformed with the modified *eglB* gene in the pET151 TOPO vector were shown to form larger hydrolysis rings than the experimental control *Bacillus xiamenensis*.

The expression and activity of enzymes sequences have to overcome many troubles such as possible incorrect orientation within the expression plasmid; the possible formation of inclusion bodies that could have prevented the secretion to the outside of the cell due to its size and recombinant nature; the natural defense mechanisms of *E. coli* to avoid toxic levels of enzymatic cofactors within its cytoplasm, which may prevent the correct folding of enzymes; the possibility that there has been an incorrect folding enzyme due to its recombinant nature. All these problems can be partially avoided if bioinformatics is applied to change amino acids without alter functionality and affinity for substrate. This is a step forward to produce

endoglucanase *in vitro*, with a high efficiency because bacteria low cost and high producing system expression.

REFERENCES

- Ariffin, H., Abdullah, N., Umi Kalsom, M. S., Shirai, Y., & Hassan, M.. (2006). Production and characterization of cellulase by *Bacillus pumilus* EB3. *International Journal of Engineering and Technology*, 3(1), 47–53. Retrieved from <http://www.ijet.feiiic.org/journals/J-2006-V1005.pdf>
- Bauer, S., Vasu, P., Persson, S., Mort, A. J., & Somerville, C. R. (2006). Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proceedings of the National Academy of Sciences of the United States of America*, 103(30), 11417–11422. <http://doi.org/10.1073/pnas.0604632103>.
- Belancic, A., Scarpa, J., Peirano, A., Díaz, R., Steiner, J., & Eyzaguirre, J. (1995). *Penicillium purpurogenum* produces several xylanases: Purification and properties of two of the enzymes. *Journal of Biotechnology*, 41(1), 71–79. [http://doi.org/10.1016/0168-1656\(95\)00057-W](http://doi.org/10.1016/0168-1656(95)00057-W).
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242. <http://doi.org/10.1093/nar/28.1.235>.
- Bohórquez Saval, S. (2012). Aprovechamiento de residuos agroindustriales: pasado, presente y futuro. *Revista de La Sociedad Mexicana de Biotecnología Y Bioingeniería*, 16(2), 14–47.
- Busk, P. K., Lange, M., Pilgaard, B., & Lange, L. (2014). Several genes encoding enzymes with the same activity are necessary for aerobic fungal degradation of cellulose in nature. *PLoS One*, 9(12). <https://doi.org/10.1371/journal.pone.0114138>.
- Chávez, R., Navarro, C., Caldeón, I., Peirano, A., Bull, P., & Eyzaguirre, J. (2002). Secretion of endoxylanase A from *Penicillium purpurogenum* by *Saccharomyces cerevisiae* transformed with genomic fungal DNA. *FEMS Microbiology Letters*, 212(2), 237–241. [http://doi.org/10.1016/S0378-1097\(02\)00750-4](http://doi.org/10.1016/S0378-1097(02)00750-4).
- Cuervo, L., Folch, J., & Quiroz, R. (2009). Lignocelulosa como fuente de azúcares para la producción de etanol. *Bio Tecnología*, 13(3), 11–25. Retrieved from

- http://www.smbb.com.mx/revista/Revista_2009_3/Lignocelulosa.pdf.
- Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S., & Kim, J. F. (2009). Tracing Ancestors and Relatives of *Escherichia coli* B, and the Derivation of B Strains REL606 and BL21(DE3). *Journal of Molecular Biology*, 394(4), 634–643. <http://doi.org/10.1016/j.jmb.2009.09.022>.
- Engbrecht, J., Brent, R., & Kaderbhai, M. A. (1990). Minipreps of Plasmid DNA. *Current Protocols in Molecular Biology*, 1(1991), 1.6.1-1.6.10. <http://doi.org/10.1002/0471142727>.
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., ... Punta, M. (2014). Pfam: The protein families database. *Nucleic Acids Research*, 42(D1), 1–9. <http://doi.org/10.1093/nar/gkt1223>.
- Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L., Wortman, J. R., Batzoglou, S., ... Birren, B. W. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, 438(December), 1105–1115. <http://doi.org/10.1038/nature04341>.
- Gao, J., Weng, H., Zhu, D., Yuan, M., Guan, F., & Xi, Y. (2008). Production and characterization of cellulolytic enzymes from the thermoacidophilic fungal *Aspergillus terreus* M11 under solid-state cultivation of corn stover. *Bioresource Technology*, 99(16), 7623–7629. <http://doi.org/10.1016/j.biortech.2008.02.005>.
- Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, 39(SUPPL. 2), 270–277. <http://doi.org/10.1093/nar/gkr366>.
- Gvritshvili, A. G., Leung, K. W., & Tombran-Tink, J. (2010). Codon preference optimization increases heterologous PEDF expression. *PLoS ONE*, 5(11), 1–13. <http://doi.org/10.1371/journal.pone.0015056>.
- Hanahan, B. D., Jessee, J., & Bloom, F. R. (1991). Plasmid Transformation of *Escherichia coli* and Other Bacteria. *Methods in Enzymology*, 204(1970), 63–113.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology*, 166(4), 557–580. [http://doi.org/10.1016/S0022-2836\(83\)80284-8](http://doi.org/10.1016/S0022-2836(83)80284-8).
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240. <http://doi.org/10.1093/bioinformatics/btu031>.
- Jung, S.-K., & McDonald, K. (2011). Visual gene developer: a fully programmable bioinformatics software for synthetic gene optimization. *BMC Bioinformatics*, 12(1), 340. <http://doi.org/10.1186/1471-2105-12-340>.
- Kovalenko, V. I. (2010). Crystalline cellulose: structure and hydrogen bonds. *Russian Chemical Reviews*, 79(3), 231–241. <http://doi.org/10.1070/RC2010v079n03ABEH004065>.
- Kumar, R., Singh, S., & Singh, O. V. (2008). Bioconversion of lignocellulosic biomass: Biochemical and molecular perspectives. *Journal of Industrial Microbiology and Biotechnology*, 35(5), 377–391. <http://doi.org/10.1007/s10295-008-0327-8>.
- Lai, Q., Liu, Y., & Shao, Z. (2014). *Bacillus xiamenensis* sp. Nov., isolated from intestinal tract contents of a flathead mullet (*Mugil cephalus*). *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 105(1), 99–107. <http://doi.org/10.1007/s10482-013-0057-4>.
- Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. <http://doi.org/10.1093/bioinformatics/btl158>.
- Liu, G., Qin, Y., Li, Z., & Qu, Y. (2013). Development of highly efficient, low-cost lignocellulolytic enzyme systems in the post-genomic era. *Biotechnology Advances*, 31(6), 962–975. <http://doi.org/10.1016/j.biotechadv.2013.03.001>.
- Magrane, M., & Consortium, U. P. (2011). UniProt Knowledgebase: A hub of integrated protein data. *Database*, 2011, 1–13. <http://doi.org/10.1093/database/bar009>.
- Meddeb-Mouelhi, F., Moisan, J. K., & Beauregard, M. (2014). A comparison of plate assay methods for detecting extracellular cellulase and xylanase activity. *Enzyme and Microbial Technology*, 66, 16–19. <http://doi.org/10.1016/j.enzmictec.2014.07.004>.
- Montoya, S., Sánchez, O. J., & Levin, L. (2014). Evaluación de actividades endoglucanasa, exoglucanasa, lacasa y lignina peroxidasa en diez hongos de

- putridi3n blanca. *Bioteconol3gía En El Sector Agropecuario Y Agroindustria*, 12(2), 115–124.
- Munjal, N., Jawed, K., Wajid, S., & Yazdani, S. S. (2015). A constitutive expression system for cellulase secretion in *Escherichia coli* and its use in bioethanol production. *PLoS One*, 10(3). <https://doi.org/10.1371/journal.pone.0119917>.
- Persson, I., Tjerneld, F., & Hahn-Hägerdal, B. (1991). Fungal cellulolytic enzyme production: A review. *Process Biochemistry*, 26(2), 65–74. [http://doi.org/10.1016/0032-9592\(91\)80019-L](http://doi.org/10.1016/0032-9592(91)80019-L).
- Reiss, R., Ihssen, J., & Thöny-Meyer, L. (2011). *Bacillus pumilus* laccase: a heat stable enzyme with a wide substrate spectrum. *BMC Biotechnology*, 11(1), 9. <http://doi.org/10.1186/1472-6750-11-9>.
- Rivera Bustamante, R. F., Trejo Saavedra, D. L., & Rodríguez Negrete, E. A. (2015). Detection of transgenes in genetically modified organisms and their subproducts. *Acta Universitaria*, 25(NE-3), 24–39. <http://doi.org/10.15174/au.2015.906>.
- Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2011). A Protocol for Computer-Based Protein Structure and Function Prediction. *Journal of Visualized Experiments*, (November), 1–11. <http://doi.org/10.3791/3259>.
- Ruhl, M., Majcherczyk, A., & Kues, U. (2013). Lcc1 and Lcc5 are the main laccases secreted in liquid cultures of *Coprinopsis cinerea* strains. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 103(5), 1029–1039. <http://doi.org/10.1007/s10482-013-9883-7>.
- Sánchez Riaño, A. M., Gutiérrez Morales, A. I., Muñoz Hernández, J. A., & Rivera Barrero, C. A. (2010). Producci3n de bioetanol a partir de subproductos agroindustriales lignocelul3sicos. *Tumbaga*, 5, 61–91. Retrieved from <http://revistas.ut.edu.co/index.php/tumbaga/article/view/194/163>.
- Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*, 31(13), 3381–3385. <http://doi.org/10.1093/nar/gkg520>.
- Sigrist, C. J. A., De Castro, E., Cerutti, L., CuChe, B. A., Hulo, N., Bridge, A., ... Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41(D1), 1–4. <http://doi.org/10.1093/nar/gks1067>.
- Studier, F. W., Daegelen, P., Lenski, R. E., Maslov, S., & Kim, J. F. (2009). Understanding the Differences between Genome Sequences of *Escherichia coli* B Strains REL606 and BL21(DE3) and Comparison of the *E. coli* B and K-12 Genomes. *Journal of Molecular Biology*, 394(4), 653–680. <http://doi.org/10.1016/j.jmb.2009.09.021>.
- Studier, F. W., & Moffatt, B. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189(1), 113–130. [http://doi.org/10.1016/0022-2836\(86\)90385-2](http://doi.org/10.1016/0022-2836(86)90385-2).
- Wu, G., Robertson, D. H., Brooks III, C. L., & Vieth, M. (2003). Detailed analysis of grid-based molecular docking: A case study of CDOCKER - A CHARMM based MD docking program. *Journal of Computational Chemistry*, 24, 1549–1562. <http://doi.org/10.1002/jcc.10306>.
- Yang, J., Yan, R., Roy, A., Xu, D., J, P., & Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nat Methods*, 12(1), 7–8. <http://doi.org/10.1038/nmeth.3213>.
- Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, 43(W1), W174–W181. <http://doi.org/10.1093/nar/gkv342>
- Yang, J., & Zhang, Y. (2016). Protein structure and function prediction using I-Tasser. *Current Protocols in Bioinformatics*, 8(5), 583–592. <http://doi.org/10.1002/aur.1474.Replication>.
- Young, F. E., & Spizizen, J. (1961). Physiological and genetic factors affecting transformation of *Bacillus subtilis*. *Journal of Bacteriology*, 81(5), 823–829. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=13787458.
- Zelena, K., Eisele, N., & Berger, R. G. (2014). *Escherichia coli* as a production host for novel enzymes from basidiomycota. *Biotechnology Advances*, 32(8), 1382–1395. <http://doi.org/10.1016/j.biotechadv.2014.08.006>.