Biotechnology Reports 15 (2017) 55–62

Contents lists available at ScienceDirect

Biotechnology Reports

journal homepage: www.elsevier.com/locate/btre

In vitro and *in silico* characterization of metagenomic soil-derived cellulases capable of hydrolyzing oil palm empty fruit bunch

Laura Marcela Palma Medina[a,1,2], Diana Catalina Ardila[a,1,3], María Mercedes Zambrano[b,d], Silvia Restrepo[c], Andrés Fernando González Barrios[a,*]

[a] Grupo de diseño de Productos y Procesos (GDPP), Department of Chemical Engineering, Universidad de los Andes, Carrera 1E N 18A-10, Bogotá, Colombia
[b] Molecular Genetics, Corporación Corpogen, Cra 5 N° 66A-34, Bogotá, Colombia
[c] Department of Biological Sciences, Universidad de los Andes, Carrera 1E N 18A-10, Bogotá, Colombia
[d] Gebix Center for Genomics and Bioinformatics of Extreme Environments, Cra 5 N° 66A-34, Bogotá, Colombia

ARTICLE INFO

Keywords:
Cellulases
Metagenomics
Oil palm empty fruit bunch
Fractal-like kinetic model
Molecular docking

ABSTRACT

Diversification of raw material for biofuel production is of interest to both academia and industry. One attractive substrate is a renewable lignocellulosic material such as oil palm (*Elaeis guineensis Jacq.*) empty fruit bunch (OPEFB), which is a byproduct of the palm oil industry. This study aimed to characterize cellulases active against this substrate. Cellulases with activity against OPEFB were identified from a metagenomic library obtained from DNA extracted from a high-Andean forest ecosystem. Our findings show that the highest cellulolytic activities were obtained at pH and temperature ranges of 4–10 and 30 °C–60 °C, respectively. Due to the heterogeneous character of the system, degradation profiles were fitted to a fractal-like kinetic model, evidencing transport mass transfer limitations. The sequence analysis of the metagenomic library inserts revealed three glycosyl hydrolase families. Finally, molecular docking simulations of the cellulases were carried out corroborating possible exoglucanase and β-glucosidase activity.

1. Introduction

In recent years there has been an increase in the use, trade, and production of biofuels, due to the need to replace fossil fuels with renewable energy sources. However, biofuel production has generated an ethical debate because the main raw materials can also be basic food crops (e.g., sugarcane, sugar beet and corn). In consequence, the search for alternative sources such as lignocellulosic material from industrial processes has increased [1–4]. One of the industries that currently generates higher amounts of suitable material for bioethanol production is the palm oil extraction. The palm oil industry generates lignocellulosic by-products from fresh fruit bunches, which in terms of mass contain 21% of palm oil, 27% of water and 52% of byproducts. A total of 44.2% of these byproducts constitutes empty fruit bunch (OPEFB) [5]. The utilization of OPEFB to obtain reducing sugars for bioethanol synthesis can, therefore, generate added value to the oil palm industry. The bioethanol produced from this alternative raw material is potentially competitive and has environmental advantages as it helps in reducing greenhouse gas emissions. Nevertheless, the industrialization of this process requires optimization at different stages, including the enzymatic hydrolysis by cellulases, which is a limiting step for the utilization of any cellulosic biomass [4].

The degradation of cellulose through enzymatic hydrolysis [6] is carried out by three kinds of cellulases: (1) endoglucanases, that can randomly hydrolyze internal glycoside linkages of the amorphous region of cellulose; (2) exoglucanases that progressively attack cellulose molecules at non-reducing ends of the chain, producing cellobiose molecules; and (3) β-glucosidases that hydrolyze cellobiose into glucose [7]. Although there are several sources of commercially available cellulases produced by microorganisms, their effectiveness depends on their affinity for the substrate [8]. It is expected that cellulases highly specific for OPEFB can produce fermentable sugars from this substrate in an optimized manner, making the process of saccharification easier and reducing costs. In consequence, the identification of cellulases capable of degrading OPEFB with high affinity and good reaction rates is a priority to optimize bioethanol production from this material.

* Corresponding author.
E-mail address: andgonza@uniandes.edu.co (A.F.G. Barrios).
[1] Both authors contributed equally to this work.
[2] Current Affiliation: Department of Medical Microbiology, University of Groningen. University Medical Center Groningen. Hanzeplein1, P.O. box 30001, 9700 RB Groningen, The Netherlands.
[3] Current Affiliation: Department of Bioengineering, The University of Pittsburgh, 302 Benedum Hall, Pittsburgh, PA 15219, USA.

http://dx.doi.org/10.1016/j.btre.2017.06.003
Received 2 August 2016; Received in revised form 17 May 2017; Accepted 6 June 2017
Available online 10 June 2017

Bioprospecting for microorganisms and their metabolic potential is increasingly used as a strategy to identify novel enzymes that may foster the biofuel industry [9,10]. However, the majority of microorganisms available in an environmental sample are unculturable, and therefore their study and assessment of their full biotechnological potential are difficult [11]. Metagenomics, a culture-independent strategy, has been used for discovering products from DNA isolated directly from the environment [11,12]. This approach has led to the discovery and characterization of a wide range of biocatalysts [11–13], which roused interest for the search of novel cellulases for biofuel production [14,15]. Previous metagenomics studies have reported as many as 105 new cellulases, 60 of which were obtained from soil samples [16]. This work reports the identification and characterization of cellulases from a metagenomic library of a high-Andean forest ecosystem, as part of the studies on metagenomics, microbial diversity and bioprospecting done by the GEBIX Center (*Colombian Center for Genomics and Bioinformatics of Extreme Environments*) in the Colombian National Natural Park "*Los Nevados*". The identified enzymes showed activity against OPEFB and are potentially useful in second-generation biofuel production.

2. Materials and methods

2.1. Metagenomic library

The metagenomic library, which consists of 18,432 clones, was constructed using DNA extracted from high-Andean forests soils from the National Natural Park "Los Nevados [17]. DNA was purified using the Ultra Clean Mega Soil DNA kit (MoBio) and fragments of approximately 30 kb were ligated to the pCC2FOS vector (Epicentre) and used to transform *Escherichia coli* EPI300™, following the manufacturer's indications (Epicentre). The identification of positive clones for cellulose degradation was done as published [17]. Briefly, clones capable of growing on minimum salt medium (MM; 0.2% $NaNO_3$, 0.1% K_2HPO_4, 0.05% $MgSO_4$, 0.05% KCL, 0.02% Peptone, 12.5 μg/ml chloramphenicol) with pretreated OPEFB as only source of carbon for 35 days at 30 °C and 200 rpm, were presumed to be cellulases carriers. Then, these colonies were enriched in Luria-Bertani (LB) medium for 2 days and transferred to a solid MM containing carboxymethyl cellulose. To detect cellulose hydrolysis, Congo red staining was used and the presence of a hydrolysis halo surrounding a colony was taken as a positive clone for cellulose degradation.

2.2. Oil palm empty fruit bunch pretreatment

OPEFB was shredded by grinding to a diameter of 5 mm in a hammer mill and pre-hydrolyzed by soaking in 1% (w/v) H_2SO_4 (100 mL for each 5 g of OPEFB) for 1 h, followed by autoclaving the acid-treated material at 121 °C for 15 min. The lignocellulosic material was washed with deionized water until the pH was close to 6.5, and then dried in an oven at 45 °C for at least 48 h [5]. Prior to each inoculation, the lignocellulosic material was autoclaved with the media.

2.3. Determination of cellulase activity

Cells harboring clones with cellulolytic activity [17], were grown in 2.5 L of MM [18], containing 12.5 μg/ml chloramphenicol, with pretreated OPEFB (1% w/v) as carbon source to the middle of the exponential growth phase (OD_{600} of 0.4), collected by centrifugation for 30 min at 4500 rpm (2,000 × *g*), and resuspended in 5 mL of buffer (Tris-HCl 50 mM, NaCl 100 mM, EDTA 1 mM, 0.15% Triton X-100, pH 8). Cell membranes were disrupted using a Beadbeater (Biospec Products, Bartlesville, OK, USA). After centrifugation for 10 min at 13,000 rpm (17,000 × *g*), supernatants (crude extracts) were used for enzymatic assays.

Enzymatic reactions were performed in buffer (different buffers were used depending on the pH condition required for the reaction) containing OPEFB (2.5% w/v) and crude extract (1.25 mg/ml), and incubating at varying conditions for two hours with agitation (250 rpm). All assays were done in duplicate. After incubation, samples were centrifuged for 1 min at 13,000 rpm (17,000 × *g*) in order to eliminate OPEFB. The phenol-sulfuric acid assay [6] was used for quantification of sugars: 50 μL of sample were mixed with 30 μL of 5% phenol and 180 μL of 96% sulfuric acid, incubated for 5 min, and the absorbance was measured at 480 nm. The concentration of sugar was determined using a standard curve generated with eight glucose solutions of varying concentrations ranging between 10 and 400 μg/ml.

To determine the effect of metal ion addition, reactions were carried out in the presence of 10 mM each of $MgCl_2$, $CuSO_4$, $ZnSO_4$ and KCl. These reactions were performed in buffer McIlvaine [19] at pH 5.5 and 50 °C.

The pH effect on reaction efficiency was evaluated by performing hydrolysis experiments under different pH conditions using seven different buffers (KCl-HCl for pH 1 and 2.5, McIlvaine for pH 4, 5.5 and 7, Tris-HCl for pH 8.5 and Borax-NaOH for pH 10). The reactions were carried out at 50 °C in the presence or absence of metal ions. Temperature (10 °C to 70 °C) was analyzed at the pH that displayed the highest activity.

2.4. Cellulase kinetics evaluation

Each clone with evidence of cellulase activity was evaluated by obtaining a time profile of fermentable sugar concentration using different initial OPEFB substrate concentrations (5%, 6.25%, and 7.5%) under the specific reaction conditions of metal ions, temperature, and pH for each one. The crude extracts were added when reactions were started. Samples were collected every three minutes up to twenty-seven minutes. Reducing sugars were quantified using the phenol-sulfuric acid method [6]. The experiments were performed by quintuplicate. The results were adjusted to a semi-empirical fractal-like kinetic model (Eq. (1)) [20,21], where $[S]_0$ is the initial concentration of substrate, k is a kinetic constant and h is the fractal dimension. The kinetic constant represents the affinity and velocity of the reaction and the fractal dimension represents the influence of the transport phenomena on the reaction kinetics. The fractal dimension value represents the effect of the surrounding factors such as the average diffusion distance of the protein in the 3D space, the 2D diffusion over the cellulase surface and the adsorption rate. The parameter k is mostly related to the efficiency and affinity of the proteins for the substrate and also the ability of the different kinds of cellulases to interact synergistically. To find the values of each constant, a regression was performed by the least squares method and corroborated by the open fitting curve toolbox (cftool) on Matlab™ (http://www.mathworks.com/products/curvefitting/).

$$P(t) = [S]_0(1 - \exp(-kt^{(1-h)})) \quad (1)$$

2.5. Fosmid DNA extraction and sequencing

Fosmid DNA was purified using PureLink™ Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad), according to the manufacturer's protocol, and resuspended in a final volume of 75 μL. The DNA quality was verified by 2% agarose gel electrophoresis in 1× TAE buffer. The quantity and quality were analyzed by measuring absorbance 260 nm/280 nm in a Nanodrop™ ND-1000 (Thermo Scientific). The isolated DNA was sequenced by Ion Torrent (314 chip) (Life Technologies, Carlsbad) at the Huck Institutes of Life Sciences, Pennsylvania State University.

2.6. Determination of sequences coding for cellulases

The quality of the reads was checked with the FastQC package [22] and trimmed and filtered by quality using FASTX-ToolKit [23].

Sequences that mapped to the *E. coli* K12 genome and the pCC2FOS™ sequence (EU140752.1) were removed using Bowtie™ [24]. The *de novo* assembly of the reads was carried out with CLC™ Genomics Workbench software [25]. Cellulase sequences were identified by running a Basic Local Alignment Search Tool (Blastx) [26] analysis against the non-redundant (nr) protein sequences database (parameters by default). The reads that showed identity over 75% to cellulase sequences were evaluated using Pfam [27] to determine the family to which each hypothetical cellulase belongs.

2.7. Determination of amino acid sequence conservation

Multiple alignments were carried out using ten cellulase sequences (Supplemental Material A1) from other organisms identified using PSI − Blast [26] with three iterations. The alignments were done using MUSCLE [28].

2.8. Structure modeling and molecular docking

The three-dimensional structures of the cellulase sequences were obtained by homology modeling using SWISS-MODEL™ Workspace [29]. The geometric optimization was performed in Hyperchem™ [30], *in vacuo* using the polak-ribiere algorithm, and RMS gradient of 0.1 kcal/(Å mol) as the termination condition. Each protein model was assessed with the Procheck™ analysis in Swissmodel™ [29,31].

Molecular docking was performed with AutoDock™ 4.2 [32] using a rigid model for the macromolecule. Since the active site was not previously known, the grid was set with a spacing of 0.375 Å and 126 points of evaluation in each direction; thus, the complete surface of the protein was evaluated (blind docking). The genetic algorithm method was used with a number of 10000 evaluations. Fifteen possible binding sites were proposed by AutoDock and those that showed lower binding energy were selected for further analysis.

In order to evaluate exoglucanase capability, all the cellulases were tested with a polysaccharide of five glucoses. The identified β-glucosidases were tested with cellobiose. Molecular docking simulations were additionally performed with crystal structures of reference cellulases (gi|327200721 [33], gi|339717359 [34], gi|364506202 [35], gi|16761109 [36], gi|336122540) in order to corroborate the energy magnitude and the location of the substrate. Determination of the amino acids that interact with the ligands was performed with the 'Receptor-Ligand interaction' toolbox by Discovery Studio [37].

3. Results

3.1. Identification and kinetic characterization of cellulases

Four fosmid library clones were identified as possible producers of cellulases with activity with OPEFB (clones 4, 8, 12 and 13). These clones were then purified and selected for further characterization. To assay for cellulolytic activity, crude extracts were prepared from cultures derived from each clone. The enzymatic activity present in all crude extracts was affected, either enhanced or reduced, by the addition of ions (Fig. 1). All metals, with the exception of Zn^{+2}, enhanced the activity of clone 4 while for clone 8 extract activity was only slightly increased by copper and potassium. In the case of clone 12, the addition of magnesium has a strong inhibitory effect on the degradation in comparison with the other metal additions. For clone 13, which had very low activity in the absence of ions, all metal additions stimulated the hydrolytic activity, being Zn^{+2} the ion with the strongest effect. Based on these results, potassium was selected for further assays using clones 4 and 8, and $Zn^{+2}$for clones 12 and 13.

Reactions were then carried out at different conditions of pH and temperature. The optimum for clone 4 was at the temperature of 50 °C and pH 4 in the presence of KCl, while for clone 8 the best condition was 40 °C and pH 8.5 in the absence of KCl. The activity of clone 12 produced more simple sugars at pH 8.5 and 40 °C, without Zn^{+2}, while clone 13 worked best with this metal at pH 7 and 40 °C.

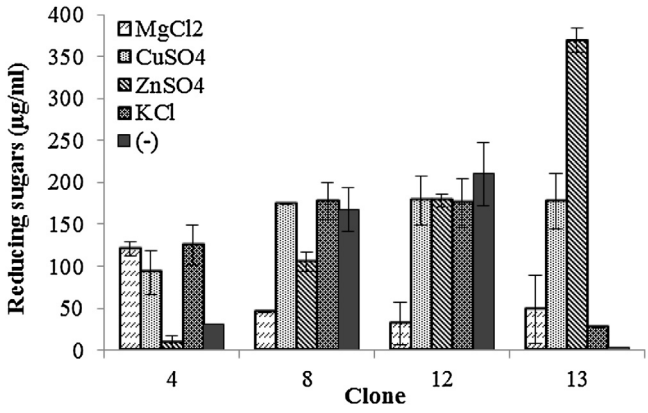

Fig. 1. Effect of different metal ions on reducing sugars production by cellulases of clones 4, 8, 12, and 13 against OPEFB after two hours of reaction. Error bars shown are standard deviation, n = 2.

The time profiles of reducing sugars obtained from hydrolysis were then analyzed by running assays at these optimal conditions of pH, temperature, and metal ion presence or absence. The enzymatic reaction hydrolysis profiles did not adjust to the Michaelis-Menten model because the results showed that the kinetic constant is dependent on the reaction time (Fig. 2).

The observed hydrolytic profiles (Fig. 2) indicated that the increment in substrate concentration had an inverse effect on the fractal dimension (*h*) and on the quantity of reducing sugars obtained for clones 4, 8, and 13. These values both decreased as the substrate concentration increased; the opposite happened, however, for the *h* value when clone 12 was analyzed. It is noticeable that the value of *k*, which could represent factors directly related to the interaction of the molecules, is in all cases in the same order of magnitude.

3.2. Identification of the cellulases via genomic analysis

To identify the possible cellulases present in each clone, each fosmid DNA was purified and sequenced, generating 100,000 reads per clone with a maximum and average lengths of 326 bp and 240 bp, respectively. For each fosmid, approximately 70,000 reads were assembled and the resultant 7000 contigs were analyzed using Blastx [26]. Among the sequences that showed identity with cellulases, two contigs of each clone 4, 8, and 13 and one contig from clone 12 were selected for Pfam characterization. Three glycosyl hydrolase families were found on seven of the contigs (Table 1): β-glucosidases (families 1 and 3), Endoglucanases (family 8) and Exoglucanases (family 3). Contig 13I, derived from clone 13, contained a sequence belonging to the peptidase family and it showed a characteristic Zn-dependent domain. This sequence also showed a M42 glucanase-like (cd05657) subdomain that was detected with the Blastx analysis. In summary, the metagenomic inserts from clones 4 and 13 could code for endoglucanase and β-glucosidase cellulases, the insert from clone 8 could produce endoglucanase and exoglucanase cellulases, while clone 12 contained a possible exoglucanase. Based on this analysis it was predicted that amino acid 373 of contig 4II, a hypothetical β-glucosidase, could be in the active site of the protein. The analysis of the putative cellulase sequences using the Pfam platform gave low e-values, suggesting that the alignments were not due to randomness. The previously described protein sequences were submitted to UniProt with accession numbers from LT853709 to LT853715.

3.3. Structure modeling and molecular docking assays

Further characterization of the cellulases through biochemical

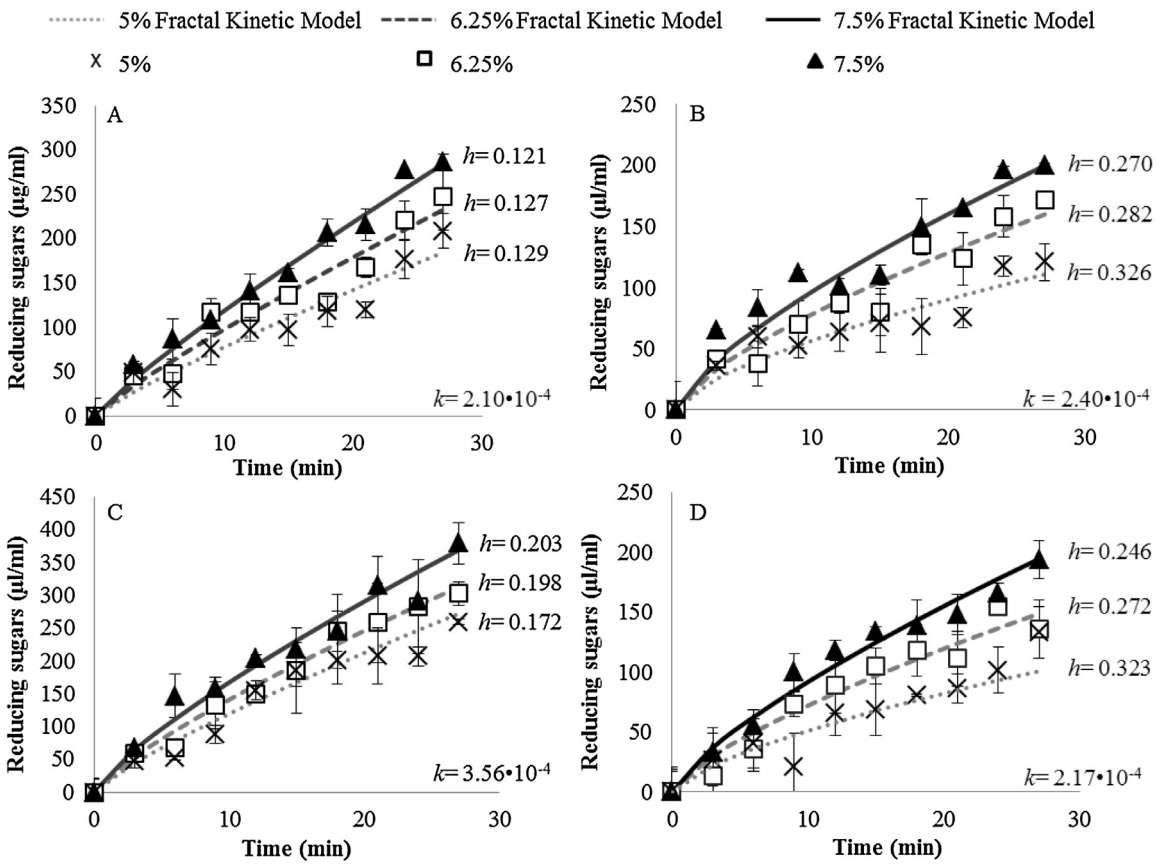

Fig. 2. Time profile for the hydrolytic activity of each metagenomic clone using different concentrations of OPEFB (5%, 6.25%, and 7.5%). (A) Clone 4 with the addition of KCl, (B) Clone 8 without any metal ion addition, (C) Clone 12 without any metal ion addition, (D) Clone 13 with the addition of $ZnSO_4$. All the curves were adjusted to a fractal kinetics model (continuous lines). Error bars shown are standard deviation, n = 5.

analyses would be possible after extraction and purification of the proteins. However, bioinformatics tools can be used as a first approach to the functionality of the novel cellulases. The proteins' sequences were modeled by homology based on crystal structures of reported cellulases (reference cellulases) and then geometrically optimized. The quality of the resulting structures was checked with structure assessment (Table 2). More than 96% of the amino acids' dihedral angles for all structures were in allowed regions. Nevertheless, the 3D structures derived from the sequences of contigs 8II and 13I had a higher percentage of amino acids in disallowed regions of the Ramachandran plot, indicating that some zones should be more carefully refined. The Qmean scores for all structures lied within the range of 0.561 and 0.747. These results indicated that the models had an acceptable quality for simulations.

Molecular docking analyses showed that the endoglucanases

Table 2
Qmean scores and Ramachandran Plot statistics from 3D modeled structure quality check of each individual contig. The results were obtained on the Swiss Model platform.

Contig ID	Qmean6 Score	Residues in allowed regions	Residues in disallowed regions
4I	0.659	99.6%	0.4%
4II	0.711	99.5%	0.5%
8I	0.747	99.3%	0.7%
8II	0.561	97.4%	2.6%
12I	0.63	99.7%	0.3%
13I	0.61	98.3%	1.7%
13II	0.63	99.8%	0.2%

Table 1
Identification of Pfam domains in each contig that had Blast results related with cellulases.

Clone	Contig ID	Enzyme expected	Domains	Length	Alignment		E-value
					Start	End	
4	4I	Endoglucanase	Glycosyl hydrolases family 8	271	2	257	2.8E-100
	4II	β-glucosidase	Glycosyl hydrolase family 1	475	4	472	1.7E-127
8	8I	Endoglucanase	Glycosyl hydrolases family 8	370	3	348	8.5E-144
	8II	Exo-1,3-1,4-glucanase	Glycosyl hydrolase family 3 N-terminal domain	631	25	227	1.4E-63
					269	331	4.6E-09
			Glycosyl hydrolase family 3 C-terminal domain		374	627	3.2E-61
12	12I	Exo-1,3-1,4-glucanase	Glycosyl hydrolase family 3 N-terminal domain	411	51	338	7.4E-91
13	13I	Endoglucanase	M42 glutamyl aminopeptidase	373	46	346	3.0E-77
	13II	β-glucosidase	Glycosyl hydrolase family 1	476	4	472	1.0E-122

Table 3
Results of molecular docking simulations of the cellulases derived of each contig. A chain of five glucose residues (referred to as Polysaccharide) and a cellobiose molecule were used as ligands. Binding energy and important amino interacting with the ligand are reported. Beside conventional hydrogen bonds other interactions were considered for the analysis as carbon-hydrogen bonds, Pi-Sigma and Pi-Donor Bonds.

Contig ID	Ligand	Binding Energy	Ligand interactions	
			Conventional H-Bonds	Other interactions
4I	Polysaccharide	−5.72	Ala3-Trp34	Gly25-Tyr48
4II	Cellobiose	−7.04	Gln19-His130-Asn175-Glu176-Tyr313-Glu373-Phe429-Thr430	Trp421
	Polysaccharide	−4.28	Arg44-Asp344-Trp345	Cys43-Pro186-Gln433
8I	Polysaccharide	−6.04	Lys57-Tyr244-Asp245-Tyr333-Tyr334	Trp98-Asp331
8II	Cellobiose	−6	Ile36-Ala88-Asp90-Trp134-Asp299	Tyr89
	Polysaccharide	−6.12	Arg148-Lys199-Arg239-Leu244-Leu272	Tyr201
12I	Cellobiose	−5.46	Gly15-Leu19-Phe345-Asn346	Glu22-Tyr340-Asp341
	Polysaccharide	−3.96	Val71-Arg104-Ile109	Thr70
13I	Polysaccharide	−3.18	Cys240-Thr242-Pro270-Tyr327-Thr328	
13II	Cellobiose	−6.54	Leu32-Phe55-Pro101-Thr145-Glu146-	Gly31-Gln102
	Polysaccharide	−3.78	Glu260-Asn323-Asn331-Val332-Val333	Gly259-Gly263
gi|327200721–Ref 4I and 8I	Polysaccharide	−5.94	Ser46-Trp96-Asp110-Tyr182-Tyr331-Tyr332	Asp44-Ala330
gi|339717359 − Ref 4II and 13II	Cellobiose	−6.55	Gln22-His134-Asn179-Glu180-Tyr317-Glu377-Phe433	
	Polysaccharide	−6.32	Glu180-Asn183-Asn320-Phe335-Ser338-Gly350-Trp351	
gi|364506202 − Ref 8II	Cellobiose	−6.98	Arg172-His215-Glu493	Gly65-Asp293-Trp294-Trp436
	Polysaccharide	−5.03	Gly66-Asn67-Lys74-Asn116-Trp294-Asn295-Glu323-Val434	Ser435
gi|16761109 - Ref 12I	Cellobiose	−6.07	Arg163-Glu176-Thr178-Ser611-His612	Asp177-Asn234
	Polysaccharide	−3.78	Val84-Gly428-Ala433	Ser429-Gly434-Glu518
gi|336122540 − Ref13I	Polysaccharide	−4.13	Met69-Lys83-Gly85-Lys212	Glu208-Arg263

interacted with a polysaccharide of five glucoses and that the β-glucosidases interacted with both ligands, the polysaccharide and cellobiose, as expected. These ligand-protein interactions involved conventional hydrogen bonds, carbon-hydrogen bonds, Pi-Sigma and Pi-donor bonds which are located in putative active sites (Supplemental material A2). The relevant residues for the interaction and binding energies are listed in Table 3.

Simulations carried out with putative endoglucanases (derived from contigs 4I, 8I and 13I) showed that the 8I resultant protein had less binding energy among them, and therefore more affinity for the polysaccharide substrate than 4I- and 13I-derived cellulases and the reference cellulases used for comparison (Table 3). It is also notable that the ligand was located in the same cavity on 8I and 13I, resulting in cellulases like their respective protein reference. On the contrary, the ligand is located at a different position in the 4I-derived protein. Equal amounts of hydrogen bonds were found in cellulases 8I and 13I, but the first also displayed another type of bonds with the ligand. The protein derived from sequence 4I had the lower amount of bonds.

In all molecular docking results using family 1 glycosyl hydrolases (from 4II and 13II sequences) the binding energies for simulations with polysaccharide as ligand were higher than those carried out with cellobiose. For these proteins, both ligands attached to the same cavity and the number of hydrogen bonds formed was either the same or greater with cellobiose than with the polysaccharide substrate. The hydrogen bond interactions with cellobiose involved the same homologous residues for the cellulases derived from 4II and its reference protein. Nevertheless, the putative cellulase also displayed other types of bonds that interact with the ligand. The lowest binding energy was shown by the putative cellulase derived from 4II, which was also predicted to interact with the ligand through amino acid glu373, a result consistent with the analysis using Pfam.

The cellulase derived from sequence 8II, which putatively belonged to family 3 of glycosyl hydrolases, showed lower binding energy for the polysaccharide than for cellobiose, indicating a higher exoglucanase activity. On the contrary, its reference protein showed more affinity for cellobiose (Table 3). Both analyzed ligands were located in the same cavity for the putative cellulase and for its reference protein, but the involved amino acids were different.

Finally, the putative cellulase derived from 12I showed two main cavities where the ligands were located depending on their length. Interestingly, more bonds were formed with the polysaccharide than with the cellobiose. We found that the ligands were located in the same cavity for the reference and the novel putative cellulase.

Multiple alignments were carried out in order to corroborate the conservation of the amino acids that showed interaction with the ligands using ten cellulase sequences from other organisms (Supplemental material A3). This analysis showed that most of the residues with ligand interactions in the novel putative cellulases were conserved in the homologous sequences. The overall conservation of the protein sequence was more evident for the proteins derived from 12I and 4II.

4. Discussion

In this work, putative cellulases identified in metagenomic clones with enzymatic activity on OPEFB were analyzed. The effect of various metals on enzymatic activity was first evaluated and it was shown that metals can have either a positive or negative effect on hydrolysis. The presence of a metallic element could influence the transport of the cellulase along the fiber as well as the affinity of the enzyme for the substrate. Subsequent bioinformatics analysis of the putative cellulases from clone 13 corroborated the experimental results and showed that the produced endoglucanase had a motif for recognition of Zn^{+2}, which could be a cofactor for the protein. The temperatures and pH conditions that allowed the best production are within ranges previously reported for other cellulases found in similar studies [38–41]. The conditions that we are reporting were the optimal found for the degradation of OPEFB by the crude protein extracts from the selected isolates and therefore could be related to the activity of more than one cellulase present in the fosmid inserts. Currently, there exist commercially available enzymes for production of sugars, such as the ACCELLERASE® (Dupont, Wilmington, DE, USA) product line. The offered products have a wide range of temperatures for activity, from 30 °C to 75 °C, and pH ranges from 3.5 to 7, depending on the product. Although the temperature range of our protein extracts is similar to that of this commercially available product, the optimum pH is higher for the enzymes identified in this study. In consequence, the cellulases available in our extracts could be used in processes where higher pH is needed for specific degradation of OPEFB.

The change with time of the kinetic constant (Hydrolysis profile

gradient, Fig. 2) could be caused by the transport phenomena (diffusion rate over the cellulose surface, diffusion time in 3D space, adsorption rate) effect over the reaction and the autoregulation process, where the depolymerization ceases due to the excess in product concentration [41]. The influence of the transport phenomena, substrate and protein locations, and the autoregulation over the reaction is mathematically represented by the fractal dimension (h) and the parameter k. Previous studies [21] have shown that fractal dimension changes with the variation of the initial concentration of the substrate. Nevertheless, this behavior depends on the kind of carbon source and the cellulase itself. The experimental data fit with the kinetic model, allowing us to infer that in all cases the enzymatic reaction may be limited by mass transport. For the profiles of clones 4, 8 and 13 the h value decreased as the substrate concentration increased, indicating a possible reduction in adsorption time or an increment in diffusion rate in the cellulose surface or in the 3D space. For clone 12, however, it can be inferred that the reaction was limited by transport phenomena because the fractal dimension increased simultaneously with the substrate concentration.

Other studies on the kinetics of cellulases have previously included the influence of transport phenomena in their models [42]. In the case of cellulases hydrolyzing OPEFB, the reaction is heterogeneous and biphasic because of the solid nature of the substrate, which limits the reaction-diffusion rate [42–44]. In consequence, the kinetic reaction was adjusted to a semi-empirical fractal-like kinetic model (Eq. (1)), which takes into account the mentioned physical limitations [20,21].

These assays carried out with OPEFB as carbon source showed that the cellulases produced by the clones could potentially be used for industrial application of sugar production from this lignocellulosic material. The capacity of these cellulases for specific degradation could imply an improvement in the process of bioethanol production. Although the general kinetics of the sugar production by each clone was analyzed, the efficiencies of each novel cellulase should be further studied. Likewise, the comparison with other reported cellulases that are highly specific for OPEFB could only be made once these enzymes are purified and concentrated.

We found predicted proteins that belong to the three types of cellulases required for complete cellulose degradation (Table 1). The contigs chosen for analysis for each clone were the longest and the ones containing the most complete amino acid sequences that represented a cellulase. Even though other cellulases domains were found, they were not further analyzed in this study because they were not long enough to be modeled. Qmean values and the Ramachandran plot of the modeled proteins indicated good quality structures and absence of energetic or steric problems. The Qmean scores are consistent with those previously reported for proteins of these sizes [45] and most of the amino acids' dihedral angles are in allowed regions of the Ramachandran plot where they do not result in steric hindrance. Therefore, the models were reliable and suitable to perform simulations of molecular docking. The model of the resultant protein from contig 8II was based on a template with low identity percentage, which is was consistent with the low amount of reference structures for the family 3 glycosyl hydrolases. In consequence, the homology modeling and later minimization were more challenging.

The experimental results indicated low glucose production, which can be explained by the fact that β-glucosidases degrade glucose dimers and thus it is possible for these cellulases to have exoglucanase activity, giving them the capability to simplify glucose polysaccharides as well. This raises the possibility that the obtained domains could have an additional exoglucanase activity in any of the predicted proteins given that cellulases hydrolyze the same kind of bond [7]. In order to verify each predicted protein's affinity, they were tested with two ligands, cellobiose and a five-glucose polysaccharide, through molecular docking. This characterization was performed *in silico* and it provided a first description of the cellulases activity. Further confirmation of the obtained results should be confirmed in the future by biochemical analyses.

The glycosyl-hydrolase family domain found in each contig corresponded to the most probable complex formed by the putative cellulases. Therefore, the binding energy was measured to confirm the proteins' affinity to the substrates. For the putative endoglucanases, spontaneous complex formation with the polysaccharide was observed during the simulation. However, these results should be compared with dockings performed using a sugar with higher complexity, in order to define the substrate with the highest affinity for these cellulases. For this ligand, it is expected to obtain lower binding energies than the ones obtained from current molecular docking simulations. It should be noted that family 1 glycosyl hydrolases also include 6-phospho-beta-glucosidases that degrade 6-phospho-beta-D-glucosyl-(1,4)-D-glucose instead of cellobiose. Therefore, further simulations should be performed for this family of enzymes in order to determine the exact type of predicted cellulase present in each clone and confirmed by biochemical analyses of purified protein. The family 3 glycosyl hydrolases domain, which includes both β-glucosidases and exoglucanases, was recognized in contigs 8II and 12I. Thus, it is reasonable that these putative cellulases differ in substrate affinity based on binding energy values.

Polar amino acids such as Asparagine, Threonine, Proline, Glutamine, Arginine, Glutamic acid and Aspartic acid are present in all complexes and form more than one hydrogen bond with the ligand (Table 3). Most of them have been previously reported as common amino acids present in active sites of cellulases [46], which makes them important for the interaction and thus for complex formation. The other hydrogen bonds are formed with the hydrophilic side chain of the aromatic and aliphatic amino acids such as Tyrosine, Leucine, Glycine, Valine and Phenylalanine.

The amino acids with relevant interactions were located in cavities, which in most of the cases, were the same for the predicted cellulases and their respective reference sequences. In consequence, these pockets could be presumed to be active sites. Moreover, the structure of these cavities agrees with those found in the literature: endoglucanases bind to the cellulose matrix at planar surfaces; exoglucanases and β-glucosidases have both hollows, being the first ones shallower [47]. It is also important to mention that the number of hydrogen bonds and other types of interactions in each simulation represents the stability of the complex and thus could indicate the efficiency of the hydrolysis reactions.

Finally, the comparison of the sequences with other cellulases (Supplemental material A3) revealed the probable origin of the DNA present in of each clone analyzed. Based on the alignment results, it is highly probable that sequences of these inserts belong to organisms of the domain Bacteria. We are aware that this result could be biased by the selection of *E. coli* as the host for the metagenomic clones. Most of the amino acids that are important for ligand interactions were found to be highly conserved, especially Glutamic acid and Aspartic acid, which reinforces the probable location of the active site previously predicted by the molecular docking simulations. The high percentage of preservation of these amino acids could indicate the evolutionary importance in cellulase functionality. Nevertheless, those amino acids with low conservation percentage could be involved in specific interaction with OPEFB and should be taken into account for further protein engineering studies, since the putative cellulases identified here are presumed to have a high affinity for this substrate.

The *in silico* characterization performed over the putative novel cellulases sequences predicted the behavior of the found proteins in the metagenomics library. Although these analyses are of major importance for the exploration of new cellulases that are suitable for the degradation of OPEFB, further biochemical characterization should be performed in the future to confirm the obtained results.

5. Conclusions

The production of bioethanol from lignocellulosic material has been

a topic of research and discussion in recent years. Nevertheless, the viability of this process depends on the optimization of the different operations, being one of them the efficiency of the saccharification process. In this regard, the identification of cellulases specific for lignocellulosic material could be a significant improvement. In this study, metagenomics was used as a tool to identify novel cellulases with specific cellulolytic activity against OPEFB. Four clones were identified as producers of cellulases active against this substrate. The kinetics of the hydrolysis performed by each clone were adjusted to a semi-empirical fractal-like kinetic model that was previously reported in mass transfer limited reactions.

Three families of cellulases were identified within the resulting contigs of the clones analyzed. Molecular docking allowed the recognition of the active sites and of important amino acids in each cellulase. Through multiple alignments, it was identified that the metagenomics sequences were most probably derived from bacteria. In all cases, the majority of the amino acids that interact with the substrate were highly conserved in homologous sequences, indicating evolutionary preservation.

Acknowledgement

This work was performed under MAVDT contract number 15, 2008 that regulates access to genetic resources in Colombia and UAESPNN research permit number DTNO-*N*-20/2007.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.btre.2017.06.003.

References

[1] A. Limayem, S.C. Ricke, Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects, Prog. Energy Combust. Sci. 38 (2012) 449–467, http://dx.doi.org/10.1016/j.pecs.2012.03.002.
[2] C.W. Reeb, T. Hays, R.A. Venditti, R. Gonzalez, S. Kelley, Supply chain analysis, delivered cost, and life cycle assessment of oil palm empty fruit bunch biomass for green chemical production in Malaysia, BioResources 9 (2014) 5385–5416, http://dx.doi.org/10.15376/biores.9.3.5385-5416.
[3] S. Chiesa, E. Gnansounou, Use of empty fruit bunches from the oil palm for bioethanol production: a thorough comparison between dilute acid and dilute alkali pretreatment, Bioresour. Technol. 159 (2014) 355–364, http://dx.doi.org/10.1016/j.biortech.2014.02.122.
[4] C.E. Wyman, What is (and is not) vital to advancing cellulosic ethanol, Trends Biotechnol. 25 (2007) 153–157, http://dx.doi.org/10.1016/j.tibtech.2007.02.009.
[5] M.S. Umikalsom, A.B. Ariff, H.S. Zulkifli, C.C. Tong, M.A. Hassan, M.I.A. Karim, The treatment of oil palm empty fruit bunch fibre for subsequent use as substrate for cellulase production by Chaetomium globosum Kunze, Bioresour. Technol. 62 (1997) 1–9, http://dx.doi.org/10.1016/S0960-8524(97)00132-6.
[6] Y.H.P. Zhang, J. Hong, X. Ye, Cellulase assays, in: J.R. Mielenz (Ed.), Biofuels, Humana Press, 2009, pp. 213–231, , http://dx.doi.org/10.1007/978-1-60761-214-8_14 (accessed February 6, 2017).
[7] A. Hernández-Santoyo, E. García-Hernández, A. Rodríguez, Celulosomas: sistemas multienzimáticos, J. Mex. Chem. Soc. 43 (1999) 137–142.
[8] M. Balat, Production of bioethanol from lignocellulosic materials via the biochemical pathway: a review, Energy Convers. Manage. 52 (2011) 858–875, http://dx.doi.org/10.1016/j.enconman.2010.08.013.
[9] F. Niehaus, C. Bertoldo, M. Kähler, G. Antranikian, Extremophiles as a source of novel enzymes for industrial application, Appl. Microbiol. Biotechnol. 51 (1999) 711–729.
[10] B. van den Burg, Extremophiles as a source for novel enzymes, Curr. Opin. Microbiol. 6 (2003) 213–218, http://dx.doi.org/10.1016/S1369-5274(03)00060-2.
[11] C.S. Riesenfeld, P.D. Schloss, J. Handelsman, Metagenomics: genomic analysis of microbial communities, Annu. Rev. Genet. 38 (2004) 525–552, http://dx.doi.org/10.1146/annurev.genet.38.072902.091216.
[12] W.R. Streit, R. Daniel, K.-E. Jaeger, Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms, Curr. Opin. Biotechnol. 15 (2004) 285–290, http://dx.doi.org/10.1016/j.copbio.2004.05.006.
[13] C. Schmeisser, H. Steele, W.R. Streit, Metagenomics, biotechnology with non-culturable microbes, Appl. Microbiol. Biotechnol. 75 (2007) 955–962, http://dx.doi.org/10.1007/s00253-007-0945-5.
[14] P. Lorenz, K. Liebeton, F. Niehaus, J. Eck, Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space, Curr. Opin. Biotechnol. 13 (2002) 572–577, http://dx.doi.org/10.1016/S0958-1669(02)00345-2.
[15] S. Voget, H.L. Steele, W.R. Streit, Characterization of a metagenome-derived halotolerant cellulase, J. Biotechnol. 126 (2006) 26–36, http://dx.doi.org/10.1016/j.jbiotec.2006.02.011.
[16] C.-J. Duan, J.-X. Feng, Mining metagenomes for novel cellulase genes, Biotechnol. Lett. 32 (2010) 1765–1775, http://dx.doi.org/10.1007/s10529-010-0356-z.
[17] D.C. Ardila Montoya, L.M. Palma Medina, M.M. Zambrano, S. Restrepo, A.F. González-Barrios, Bioprospección de estudios de celulasas en ambientes extremos colombianos, Aprovech. Biomasa Lignocelulósica Algunas Exp. Investig. En Colomb., 1 st Ed. Universidad Jorge Tadeo Lozano, Bogotá, Colombia, 2015, pp. 117–144.
[18] R.C. Kasana, R. Salwan, H. Dhar, S. Dutt, A. Gulati, A rapid and easy method for the detection of microbial cellulases on agar plates using gram's iodine, Curr. Microbiol. 57 (2008) 503–507, http://dx.doi.org/10.1007/s00284-008-9276-8.
[19] T.C. McIlvaine, A buffer solution for colorimetric comparison, J. Biol. Chem. 49 (1921) 183–186.
[20] R. Kopelman, Fractal reaction kinetics, Science 241 (1988) 1620–1626, http://dx.doi.org/10.1126/science.241.4873.1620.
[21] P. Väljamäe, K. Kipper, G. Pettersson, G. Johansson, Synergistic cellulose hydrolysis can be described in terms of fractal-like kinetics, Biotechnol. Bioeng. 84 (2003) 254–257, http://dx.doi.org/10.1002/bit.10775.
[22] Simon Andrews, FastQC: a Quality Control Tool for High Throughput Sequence Data, (2012) Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
[23] G.J. Hannon, FASTX-Toolkit, (2012) Available at: http://hannonlab.cshl.edu/fastx_toolkit.
[24] B. Langmead, Aligning short sequencing reads with Bowtie, Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al. CHAPTER (2010) Unit-11.7. 10.1002/0471250953. bi1107s32.
[25] Qiagen, CLC Genomics Workbench, (2012) Available at https://www.qiagenbioinformatics.com/.
[26] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[27] M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn, The Pfam protein families database, Nucleic Acids Res. 40 (2012) D290–D301, http://dx.doi.org/10.1093/nar/gkr1065.
[28] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, http://dx.doi.org/10.1093/nar/gkh340.
[29] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, SWISS-MODEL: an automated protein homology-modeling server, Nucleic Acids Res. 31 (2003) 3381–3385.
[30] HyperChem(TM), Hypercube, Inc. 1115 NW 4th Street, Gainesville, Florida 32601, USA, n.d.
[31] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291, http://dx.doi.org/10.1107/S0021889892009944.
[32] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4. Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791, http://dx.doi.org/10.1002/jcc.21256.
[33] O. Mazur, J. Zimmer, Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ, J. Biol. Chem. 286 (2011) 17601–17606, http://dx.doi.org/10.1074/jbc.M111.227660.
[34] M. Totir, N. Echols, M. Nanao, C.L. Gee, A. Moskaleva, S. Gradia, A.T. Iavarone, J.M. Berger, A.P. May, C. Zubieta, T. Alber, Macro-to-micro structural proteomics: native source proteins for high-throughput crystallization, PLoS One 7 (2012), http://dx.doi.org/10.1371/journal.pone.0032498.
[35] Y. Nakatani, S.M. Cutfield, N.P. Cowieson, J.F. Cutfield, Structure and activity of exo-1,3/1,4-β-glucanase from marine bacterium Pseudoalteromonas sp. BB1 showing a novel C-terminal domain, FEBS J. 279 (2011) 464–478, http://dx.doi.org/10.1111/j.1742-4658.2011.08439.x.
[36] J. Parkhill, G. Dougan, K.D. James, N.R. Thomson, D. Pickard, J. Wain, C. Churcher, K.L. Mungall, S.D. Bentley, M.T.G. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R.M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T.T. Hien, S. Holroyd, K. Jagels, A. Krogh, T.S. Larsen, S. Leather, S. Moule, P. Ó'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, B.G. Barrell, Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18, Nature 413 (2001) 848–885, http://dx.doi.org/10.1038/35101607.
[37] Discovery Studio Modeling Environment, Dassault Systèmes BIOVIA, (2016).
[38] D.W. Cockburn, A.J. Clarke, Modulating the pH-activity profile of cellulase A from Cellulomonas fimi by replacement of surface residues, Protein Eng. Des. Sel. Print. 24 (2011) 429–437, http://dx.doi.org/10.1093/protein/gzr004.
[39] S.-J. Kim, C.-M. Lee, B.-R. Han, M.-Y. Kim, Y.-S. Yeo, S.-H. Yoon, B.-S. Koo, H.-K. Jun, Characterization of a gene encoding cellulase from uncultured soil bacteria, FEMS Microbiol. Lett. 282 (2008) 44–51, http://dx.doi.org/10.1111/j.1574-6968.2008.01097.x.
[40] A. Beloqui, T.Y. Nechitaylo, N. López-Cortés, A. Ghazi, M.-E. Guazzaroni, J. Polaina, A.W. Strittmatter, O. Reva, A. Waliczek, M.M. Yakimov, O.V. Golyshina, M. Ferrer, P.N. Golyshin, Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species, Appl. Environ. Microbiol. 76 (2010) 5934–5946, http://dx.doi.org/10.1128/AEM.00902-10.
[41] Y. Sun, J. Cheng, Hydrolysis of lignocellulosic materials for ethanol production: a

review, Bioresour. Technol. 83 (2002) 1–11, http://dx.doi.org/10.1016/S0960-8524(01)00212-7.
[42] P. Bansal, M. Hall, M.J. Realff, J.H. Lee, A.S. Bommarius, Modeling cellulase kinetics on lignocellulosic substrates, Biotechnol. Adv. 27 (2009) 833–848, http://dx.doi.org/10.1016/j.biotechadv.2009.06.005.
[43] F. Xu, H. Ding, A new kinetic model for heterogeneous (or spatially confined) enzymatic catalysis: contributions from the fractal and jamming (overcrowding) effects, Appl. Catal. Gen. 317 (2007) 70–81, http://dx.doi.org/10.1016/j.apcata.2006.10.014.
[44] M. Yao, Z. Wang, Z. Wu, H. Qi, Evaluating kinetics of enzymatic saccharification of lignocellulose by fractal kinetic analysis, Biotechnol. Bioprocess Eng. 16 (2011) 1240–1247, http://dx.doi.org/10.1007/s12257-011-0283-4.
[45] P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, Bioinformatics 27 (2011) 343–350, http://dx.doi.org/10.1093/bioinformatics/btq662.
[46] N. Todaka, T. Inoue, K. Saita, M. Ohkuma, C.A. Nalepa, M. Lenz, T. Kudo, S. Moriya, Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach, PLoS One 5 (2010) e8636, http://dx.doi.org/10.1371/journal.pone.0008636.
[47] V. Notenboom, A.B. Boraston, D.G. Kilburn, D.R. Rose, Crystal structures of the family 9 carbohydrate-binding module from thermotoga maritima xylanase 10A in native and ligand-bound forms, Biochemistry (Mosc.) 40 (2001) 6248–6256, http://dx.doi.org/10.1021/bi0101704.