Biotechnol Lett (2021) 43:691–700
https://doi.org/10.1007/s10529-020-03056-z

ORIGINAL RESEARCH PAPER

Characterization and engineering of two new GH9 and GH48 cellulases from a *Bacillus pumilus* isolated from Lake Bogoria

Lydia A. Ogonda · Amélie Saumonneau · Michel Dion · Edward K. Muge · Benson M. Wamalwa · Francis J. Mulaa · Charles Tellier

Received: 30 September 2020 / Accepted: 10 December 2020 / Published online: 1 January 2021

Abstract
Objectives To search for new alkaliphilic cellulases and to improve their efficiency on crystalline cellulose through molecular engineering
Results Two novel cellulases, *Bp*GH9 and *Bp*GH48, from a *Bacillus pumilus* strain were identified, cloned and biochemically characterized. *Bp*GH9 is a modular endocellulase belonging to the glycoside hydrolase 9 family (GH9), which contains a catalytic module (GH) and a carbohydrate-binding module belonging to class 3 and subclass c (CBM3c). This enzyme is extremely tolerant to high alkali pH and remains significantly active at pH 10. *Bp*GH48 is an exocellulase, belonging to the glycoside hydrolase 48 family (GH48) and acts on the reducing end of oligo-β1,4 glucanes. A truncated form of *Bp*GH9 and a chimeric fusion with an additional CBM3a module was constructed. The deletion of the CBM3c module results in a significant decline in the catalytic activity. However, fusion of CBM3a, although in a non native position, enhanced the activity of *Bp*GH9 on crystalline cellulose.
Conclusions A new alkaliphilic endocellulase *Bp*GH9, was cloned and engineered as a fusion protein

Supplementary Information The online version of this article (https://doi.org/10.1007/s10529-020-03056-z) contains supplementary material, which is available to authorized users.

L. A. Ogonda · A. Saumonneau · C. Tellier (✉)
Université de Nantes, CNRS, UFIP, UMR6286, 2, rue de la Houssinière, 44322 Nantes, France
e-mail: charles.tellier@univ-nantes.fr

L. A. Ogonda · E. K. Muge · F. J. Mulaa
Department of Biochemistry, School of Medicine, College of Health Sciences, University of Nairobi, P.O BOX 30197-00100, Nairobi, Kenya

Present Address:
L. A. Ogonda
Department of Medical Biochemistry, School of Medicine, Masinde Muliro University of Science and Technology, P.O BOX 190-50100, Kakamega, Kenya

M. Dion
Université de Nantes, IRS2, 44000, Nantes, France

B. M. Wamalwa
Department of Chemistry, School of Physical Sciences, College of Biological and Physical Sciences, University of Nairobi, P.O BOX 30197-00100, Nairobi, Kenya

(CBM3a-*Bp*GH9), which led to an improved activity on crystalline cellulose.

Keywords Cellulose · Endoglucanase · Cellobiohydrolase · Alkaliphilic · CBM · *Bacillus pumilus*

Introduction

Cellulose is the major component of the plant cell wall and thus, the most abundant renewable biomass for the production of biofuels and biochemicals. It is a linear unbranched homopolymer of glucose residues linked by β-1,4 glycosidic bonds. Despite its simple structure, efficient degradation of cellulose is prevented by the complex structure of the plant cell wall wherein cellulose is embedded in a network of hemicellulose and lignin. Consequently, although cellulose is a potential feedstock for several industrial applications in food, textile, beverage, biofuel, pharmaceutical, paper, pulp and detergent industries (Jahangeer et al. 2005), it remains largely underutilized (Klemm et al. 2005).

Various microorganisms, including bacteria and fungi, have evolved enzymes to tackle the cellulose breakdown problem (Uchiyama et al. 2020). Aerobic microorganisms secrete to the milieu, a cocktail of cellulases and ancillary enzymes, which act synergistically to breakdown cellulose. Anaerobic organisms, on the other hand, are more efficient cellulose users/degraders. They possess an inherent multiple enzyme complex consisting of different catalytic units known as a cellulosome (Bayer et al. 2004; Doi and Kosugi 2004; Fontes and Gilbert 2010). In both mechanisms, the complete degradation of cellulose is a concerted effort between three different cellulase enzyme types, namely: β-1,4 endoglucanase, cellobiohydrolase/cellobiosidase, and β-glucosidase (Zhang and Zhang 2013). Analysis of genome sequences of cellulolytic microorganisms has allowed the identification of numerous cellulases that are involved in cellulose degradation (Lombard et al. 2014). However, the biochemical characterization of these new cellulases remains a hurdle to the identification of their functional properties.

In this study, we used functional screening methods to directly identify alkaliphilic cellulases secreted by strains isolated from Lake Bogoria in Kenya. This ecosystem is interesting since its water pH is alkaline (pH 9.3–10.3) and its water temperature varies from 25 to 60 °C due to sub-aquatic hydrothermal activity (Duckworth et al. 1996). The water is also particularly rich in organic matter due to the decomposition of terrestrial plants, microalgae and cyanobacteria. This study aimed at improving cellulose utilization by screening for novel, native, extremophilic cellulases with improved saccharifying capacity. Further, we sought to bioengineer the said cellulases using protein engineering tools in order to improve cellulose breakdown.

Materials and methods

Chemicals

Majority of the chemical materials used in this study were purchased from Merck-Sigma-Aldrich®: Sodium carboxymethylcellulose (Na-CMC), Avicel PH-101, Bovine serum albumin (BSA), 2,2′,2′′,2′′′-(ethane-1,2-diyldinitrilo)tetra acetic acid (EDTA), Isopropyl β-D-1-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), *para*-Nitro-Phenyl-β-D-cellobiose (*p*NP-β-D-cellobiose), aminobenzonitrile, $NaCNBH_3$. Protein and DNA reagents for molecular mass determination and chemicals for polyacrylamide gel electrophoresis (PAGE) were ordered from Thermoscientific® and Bio-Rad®. Oligosaccharides, cellobiose, maltotriose and cellopentaose, were purchased from Megazyme®. All other chemicals and reagents used were of analytical grade.

Screening Bogoria strains for cellulase producers

A microbial library (95 strains) previously isolated from the hot springs of Lake Bogoria (Kenya) was used for screening (Moallic et al. 2006). Each strain previously stored at − 80 °C, was enriched on a minimal salt media (supplemented and non-supplemented) at pH 7 (Supplementary Table 1) with 1% CMC. After growth of the strains (24 h at 37 °C), the plates were flooded with 0.1% (w/v) Congo red solution and left to stand for 30 min with intermittent shaking followed by a single rinse with H_2O. The plates were then flooded with 1 M NaCl with

intermittent shaking to destain and reveal the clear zones (halos) of cellulolytic activity (Supplementary Fig. 1). Strains that were positive for cellulolytic activity at pH 7 were further screened by inoculation on solid minimal salt media (supplemented or non-supplemented) at pH 9 and 10 to identify alkalophilic cellulase producers. Potency of cellulolytic activity was determined using the halo to colony diameter ratios. From the screening activity, the highest ranked microbial strain was selected for downstream work. The genomic DNA (gDNA) was extracted using phenol–chloroform extraction method. 16srRNA was amplified by PCR using 534R and 27F primers (Supplementary Table 2) and sequenced to identify the microbial strains.

Cloning of *BpGH9* and *BpGH48* cellulase genes

Wild type endoglucanase (*Bp*GH9) and wild type cellobiohydrolase (*Bp*GH48) genes were then amplified by PCR from gDNA of the selected isolate, A8, using primers: *Bp*GH9F, *Bp*GH9R, *Bp*GH48F *and* *Bp*GH48R (Supplementary Table 2). These primers were designed from the corresponding sequences of a closely related strain, *Bacillus pumilus* WP8, whose genome (CP010075.1) has been fully sequenced (Kang et al. 2015). *Bp*GH9 and *Bp*GH48 genes were then cloned into linearized pECa (corresponding to $pET21_{(a)}$ plasmid with insertion into *Sph*I-*Bgl*II restriction sites of Tac2 constitutive promoter, 28pb) by *BamHI*/*Xho*I restriction enzyme in order to insert the poly-histidine tagged gene at the *C*-terminal of the coding region.

Assembly of *BpGH9* gene with *CBM3a* sequence

In order to construct a recombinant endoglucanase gene (BpGH9), the wild type endoglucanase gene was fused with the sequence of type 3a Cellulose Binding Module (*CBM3a*) from *Clostridium thermocellum* optimized for *E. coli* expression (synthesized by Genscript®). These two genes are bound by a long linker sequence made of 123 nucleotides. This construct was assembled using Gibson Assembly® protocol (New England BioLabs®Inc) with recombinant primers (Supplementary Table 2). This recombinant enzyme construct (CBM3a-*Bp*GH9; Fig. 2) was designed to contain in fusion, a poly-histidine tag at the *C*-terminal.

BpGH9 gene with a truncated *CBM3c* module

In order to remove the CBM3c sequence from wild type *BpGH9* and *CBM3a-BpGH9*, 2 primer couples (Table S2) were used to amplify the DNA fragments corresponding to *Bp*GH9 without endogenous CBM3c (*Bp*GH9t) and CBM3a-*Bp*GH9 without CBM3c (CBM3a-*Bp*GH9t). The purified PCR products were introduced using Gibson Assembly® protocol (New England BioLabs®Inc.) into pECa plasmid digested by *Bam*HI and *Xho*I restriction enzymes and purified.

Production and purification of recombinant enzymes

Recombinant strains (*E.coli* BL21 (DE3)) expressing the entire assembly or truncated *Bp*GH9 and *Bp*GH48 genes were grown in 1 L of LB media with ampicillin antibiotic (100 mg/mL), and supplemented with calcium at 37 °C to 0.6 of $O.D_{600nm}$ (Optical Density measured at λ = 600 nm). IPTG at 0.5 mM was added to induce the expression at 30 °C overnight under agitation (200 rpm). The culture was then centrifuged, the pellet frozen at − 20 °C and then resuspended in 20 mL of lysis buffer (20 mM Tris–HCl pH 8.0) containing 10 mM Imidazole, 5 mM DNaseI, 1 mg/mL lysozyme and a cocktail of protease inhibitors (Sigma-Aldrich®). After sonication and centrifugation, the recombinant proteins with His-tag in the supernatant (soluble fraction), were purified using immobilized nickel ion—affinity chromatography Ni–NTA Superflow (Qiagen). Protein expression and purity was then checked using 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (Supplementary Fig. 3). Enzyme concentrations were determined by UV absorbance at 280 nm using a NanoDrop 1000 (ThermoScientific) (Supplementary Table 3). This data was confirmed using Agilent 2100 Bioanalyzer® (Supplementary Fig. 2).

Enzymatic characterization

The enzymatic activities of *Bp*GH9 (β-1,4-endoglucanase) and that of its variants (*Bp*GH9t, CBM3a-*Bp*GH9, CBM3a-*Bp*GH9t) and of *Bp*GH48 (β-1,4-cellobiosidase), were conducted using different cellulosic substrates: Na-CMC, microcrystalline cellulose substrate (Avicel PH-101), amorphous cellulosic

substrate phosphoric acid swollen cellulose (PASC) (Stern et al. 2015) and pNP-β-D-cellobiose.

The activities of the enzymes (0.5 µM or 1 µM) were tested in 50 mM citrate buffer pH 6.0 and with 1% of cellulosic substrate (Na-CMC, Avicel or PASC) or 5 mM and pNP β-D-cellobiose. This was done at 37 °C for 1 h for Na-CMC, PASC and pNP β-D-cellobiose and at 50 °C for 24 h using Avicel. For the assays with Na-CMC, Avicel or PASC, 70 µL of reaction mixture was withdrawn at regular time intervals and added to 70 µL of dinitro salicylic acid solution (DNS). The mixture was heated at 95 °C for 10 min (Miller 1959). The optical density (O.D.) was then read at 540 nm by spectrophotometry (Magellan, Tecan®), to quantify the reducing sugars released. For the assay on pNP β-D-cellobiose, 50 µL aliquots of the reaction mixture was withdrawn at regular time intervals and added to 100 µL of 1 M Na_2CO_3. The O.D. was then read at 405 nm to quantify the pNP released.

The mode of action of the *Bp*GH48 was determined using a derivatized cellopentaose substrate with aminobenzonitrile (ABN) (Kipper et al. 2005). The cellopentaose was derivatized by reductive amination for 2 h at 70 °C in a buffer reaction prepared with 100 mg of $NaCNBH_3$, 0.6 g of 4-aminobenzonitrile (ABN) and 0.5 mL acetic acid dissolved in 10 mL of methanol. The purified *Bp*GH48 enzyme was incubated with 20 mM of derivatized and non derivatized cellopentaose in 50 mM of citrate buffer (pH 6.0) at 37 °C for 8 h. At the start of the reaction (t0) and after at 1, 2, 4, 6 and 8 h, aliquots (1 µL) of the reaction mixture were loaded onto Thin Layer Chromatography (TLC) plate with 1 µL (20 mM) each of the reference molecules (glucose, cellobiose, maltotriose, cellopentaose and derivatized cellopentaose). The TLC plate was eluted in butanol/ethanol/H_2O (5/5/3) and visualized using orcinol solution (0.1% orcinol (w/v), 10% (v/v) sulfuric acid in absolute ethanol.

To determine the pH profile of β-1,4-endoglucanase (*Bp*GH9) and cellobiohydrolase (*Bp*GH48), enzyme assays were conducted at 37 °C within pH ranges 2–12. The pH was adjusted in a buffer reaction consisting of 50 mM NaH_2PO_4, 50 mM boric acid and 40 mM acetic acid. The enzyme was added in a pH buffer containing 1.25% (w/v) Na-CMC. Aliquots (70 µL) of reactions were withdrawn at t = 0, 15 min and 30 min and added to 70 µL of DNS solution (Miller 1959). The initial reaction velocity was measured by the reported O.D. difference per time.

Similarly, the optimum temperature for *Bp*GH9 was determined by conducting enzyme assay between 20 to 75 °C at pH 8.0 in 50 mM of citrate buffer with 1.25% (w/v) Na-CMC. In the same way, the pH stability for *Bp*GH9 was also determined by incubating the enzymes in a buffer reaction consisting of 50 mM NaH_2PO_4, 50 mM boric acid and 40 mM acetic acid at pH 10.0 and at 37 °C for t = 0 min, 30 min, 60 min, 90 min and 120 min. The initial reaction velocity was then measured at pH 8.0 in 50 mM of citrate buffer with 1.25% (w/v) CMC at 37 °C.

Affinity Pull-Down assays

The functionality of the CBM3a module following its fusion to the bi-modular wild type endoglucanase was ascertained by the binding of the CBM to Avicel PH-101 (Stern et al. 2015). The enzymes (*Bp*GH9, CBM3a-*Bp*GH9, *Bp*GH9t and CBM3a-*Bp*GH9t), 3.5 µM each in 50 mM acetate buffer (pH 5.0), 12 mM CaCl2, 2 mM EDTA, were incubated at 4 °C for 2 h in the presence of 10% insoluble cellulose (Avicel PH-101). The mixture was then centrifuged for 2 min, the supernatant (S) recovered and the pellet washed in 10 mM Tris/HCl buffer, pH 8 containing 0.05% (v/v) Tween®20 in order to remove the non-specifically bound proteins. The mixture was again centrifuged for two minutes and the pellet (P) resuspended in 10 mM Tris/HCl buffer, pH 8. The fractions containing the pellets (P) and the supernatants (S) were denatured for 3 min at 95 °C in a loading buffer (2% (w/v) SDS, 0.08 M Tris–HCl, pH 6.8, 10% (v/v) glycerol, 0.06% (v/v) ß-mercaptoethanol) then loaded onto an SDS-PAGE gel containing 12% acrylamide under denaturing conditions.

Results

Screening Bogoria strains for alkaliphilic cellulases

Among the 95 isolates collected from Lake Bogoria (Kenya), 42 isolates were identified as neutrophilic CMCase producers (pH 7) on supplemented minimal salt media (MSM) using the Congo red plate assay.

Further, isolates were screened on supplemented, and non supplemented MSM (Supplementary Table 1) at higher pHs (pH 9 and 10) in order to identify isolates that had alkaliphilic activity. Among the 95 isolates screened, 45 and 23 isolates were identified as alkaliphilic CMCase producers at pH 9 and pH10 respectively. These isolates were then ranked based on their cellulolytic activity using the halo to colony diameter ratios in order to identify the best and highest cellulase producer (Fig. 1, Supplementary Fig. 1). The highest alkaliphilic producer A8 was then identified and used for further downstream processes. This isolate A8 was previously identified as a *Bacillus pumilus with* 99% rRNA 16S gene sequence identity (Moallic et al. 2006).

Cloning for wild type *Bp*GH9 and *Bp*GH48 protein expression

Based on the sequencing of rRNA 16S gene from A8 strain, *Bacillus pumilus* WP8 (CP010075.1) (Kang et al. 2015) was the closest strain to A8 (99.79%) for which the whole genome was sequenced. Its genome contains two cellulase genes (QR42_07950; QR42_07955) in the same gene cluster that was annotated, but not biochemically characterized, to cellulases from GH9 and GH48 families, respectively (Lombard et al. 2014). From their sequence, primers were designed for PCR amplification of *BpGH9* and *BpGH48* from strain A8 (Supplementary Table 2). After amplification and sequencing, the *BpGH9* and *BpGH48* genes obtained, coded for protein sequences with 623 and 710 aa with a molecular mass of 70.79 kDa and 79.59 kDa, respectively (Supplementary Table 3 and Supplementary Fig. 2). The translated *Bp*GH9 and *Bp*GH48 proteins showed more than 99% identity to several glycoside hydrolases from *Bacillus pumilus*. The nucleotide sequence of these two cellulases have been deposited in the NCBI Genbank database under the accession numbers KY441475.1 and KY441476.1 respectively. The closest biochemically characterized cellulases to *Bp*GH9 and *Bp*GH48 were the endoglucanase EglA (99% identity) (Lima et al. 2005) and a β1,4-beta-cellobiosidase from *Bacillus pumilus* (SAFR-032, 95% identity), respectively. A 3D structure was recently released for the GH48 cellobiosidase (PDB 5BV9), while the closest structural data to *Bp*GH9 was the 1,4β-endoglucanase CelA from *Caldicellulosiruptor bescii* (PDB 4DOE, 52% identity with the catalytic module) (Brunecky et al. 2013) and Cel9G from *C. cellulolyticum* (PDB 1G87, 49% identity) (Mandelman et al. 2003). These sequence homologies suggest that *Bp*GH9 has a modular structure containing a catalytic domain (residues 1–439) and a cellulose-binding domain (CBM3c) (Fig. 2 and Supplementary Fig. 7). On the other hand, *Bp*GH48 appears to be non-modular with only a catalytic domain.

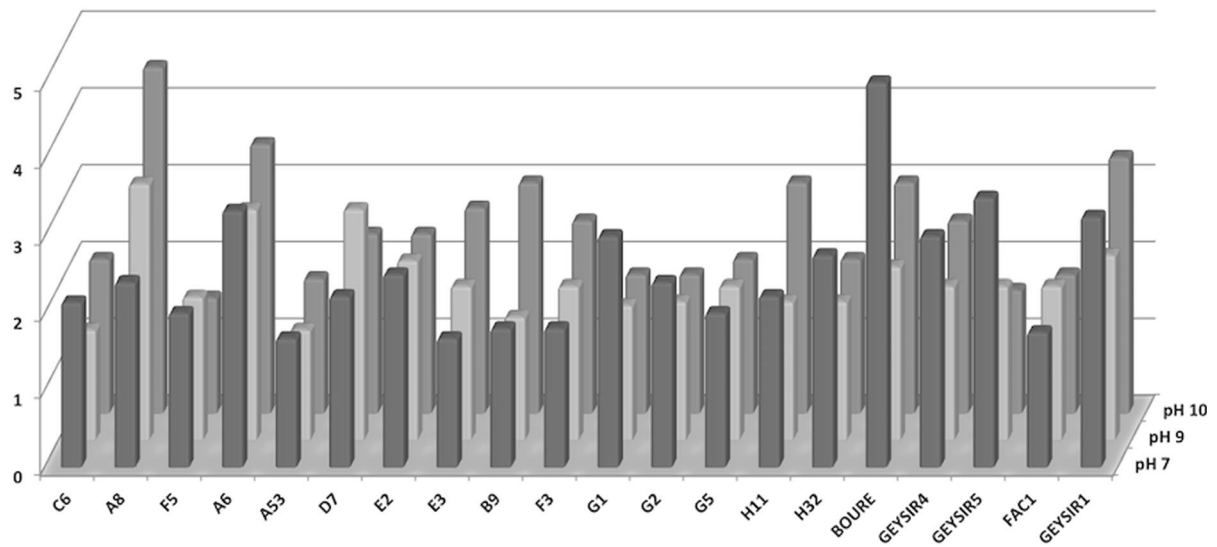

Fig. 1 Histogram of halo over colony size ratio at pH 7, 9 and 10 of cellulase producers, measured by Congo red plate screening to observe cellulase activity on minimal medium supplemented with yeast extract (0.5%) and CMC (1%)

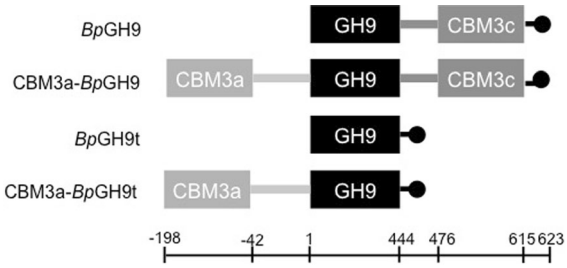

Fig. 2 Schematic diagram of the recombinants *Bp*GH9 proteins with a *C*-terminal His-tag added to all fusion proteins. Scales show the number of amino acid and the boundaries between the different domains

Enzymatic characterizations of the WT *Bp*GH9 and *Bp*GH48

The activities of both cellulases were tested on different cellulose substrates and results are summarized in Table 1. As predicted by the activity of its homologous enzymes, *Bp*GH9 is active on both crystalline cellulose substrates (Avicel) as well as on derivatized and amorphous pre-treated cellulose substrates (Na-CMC and PASC, respectively). In contrast, *Bp*GH48 has almost no activity on crystalline cellulose substrates (Avicel) as well as derivatized cellulose (Na-CMC) and pNP-cellobiose, but acts on amorphous cellulose (PASC). The major product of PASC hydrolysis by *Bp*GH48 was cellobiose, as revealed by TLC (data not shown), suggesting that it is an exo-cellobiohydrolase. Combination of *Bp*GH9 and *Bp*GH48 showed a significant synergistic effect of this pair of cellulases (Table 1).

To confirm the mode of action of *Bp*GH48, its activity was tested on cellopentaose and cellopentaose modified by reductive amination with aminobenzonitrile (Fig. 3). On free cellopentaose, the action of *Bp*GH48 led to the production of a mixture of cellobiose and cellotriose, and no further hydrolysis of cellotriose was observed since no glucose was detected (Fig. 3a). Minimal hydrolytic activity was detected on cellopentaose modified at the reductive end, which led to the liberation of cellotriose and a derivatized oligosaccharide (Fig. 3b). These results are in agreement with cellobiohydrolase activity, and further distinguishes *Bp*GH48 as a reducing end cellobiohydrolase (Cellobiohydrolase II). Consequently, there is a diminished activity when this reductive end is modified by reductive amination, and further explains why *Bp*GH48 does not hydrolyse pNP-cellobioside.

Screening of cellulolytic producers was done at high pH (7–10), this conferred an alkaliphilic environment for the extracted cellulases. We then determined the optimum pH for enzymatic activity of *Bp*GH9 and *Bp*GH48 cellulases (Fig. 4). The optimum pH of BpGH9 was determined at 6.0, where it retained $> 70\%$ of its activity until pH 10 thus validating the screening process. Furthermore, its stability upon exposure to high pH is excellent since no decrease in activity was observed after 2 h incubation at pH 10 (Supplementary Fig. 5). This pH dependence activity is quite different from that of EglA, a highly homologous enzyme (99%) to BpGH9, which showed much lower relative activity at high pH. In addition, *Bp*GH9 shows optimal activity at a temperature of 55 °C (Supplementary Fig. 6). On the other hand, *Bp*GH48 exhibited optimal activity at a pH of 5 and a low activity at high pH suggesting that this activity did not contribute to the screening process. This is in line with previous research that screening of

Table 1 Activities (μmol/min/mg) of BpGH9 and BpGH48 on different cellulosic substrates

Enzyme	CMC (1%)	PASC (1%)	Avicel (1%)	PNP-cellobiose (5 mM)
BpGH9	[a](3.0 ± 0.5) × 10^5	(5.6 ± 1.0) × 10^4	8.1 ± 1.0	nd
BpGH48	nd	36.8 ± 4.4	nd	nd
BpGH9 + BpGH48	[b](5.3 ± 0.4) × 10^5	(6.2 ± 0.8) × 10^4	15.5 ± 0.4	nd

nd Not detected

[a]The activities were determined at pH 6 in a citrate buffer 50 mM at 37 °C, excepting with the Avicel substrate where the activities were measured at 50 °C

[b]The activities are expressed per mg of GH9. Both enzymes are in equimolar ratio

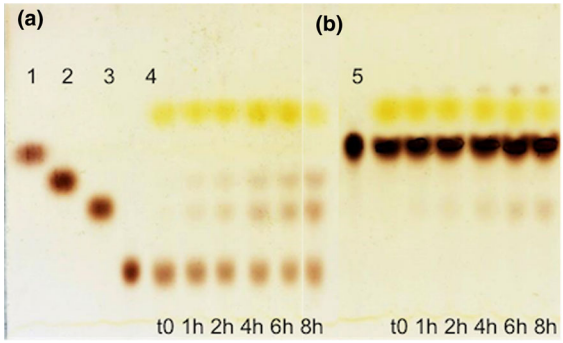

Fig. 3 Kinetics of hydrolysis of cellopentaose (4) (5 mM) hydrolysis (**a**) and ABN-cellopentaose (5) (**b**) by *Bp*GH48 at pH 8 as followed by TLC. Standards: glucose (1), cellobiose (2), cellotriose (3)

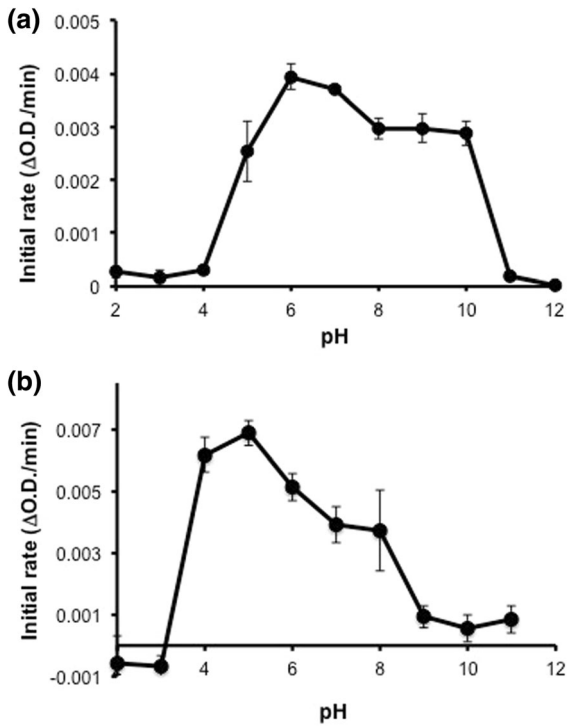

Fig. 4 pH activities profile (ΔO.D./min) of *Bp*GH9 (**a**) and *Bp*GH48 (**b**) on 1% CMC at 37 °C. Data points with standard deviations derive from triplicates

cellulase activity on CR plates exploits the endoglucanase activity.

Construction and characterization of a modular *BpGH9*

In order to improve the catalytic efficiency of *Bp*GH9 on crystalline cellulose, we added an exogenous cellulose binding domain to this endocellulase. The *BpGH9* gene was fused upstream with the IIIa cellulose binding module (CBM3a) gene from *C. thermocellum*, with a peptide linker sequence of 41 amino-acids (Fig. 2). This recombinant fusion protein (CBM3-*Bp*GH9) was expressed and purified by affinity chromatography using a His-tag sequence at the *C*-terminal end of *Bp*GH9. The resulting protein is 821 aa long with a calculated molecular weight of 90.5 kDa, which corresponds to the size of the expressed protein as revealed by SDS-PAGE (Fig. 5).

The binding activity of the CBM3a in the fusion protein was examined by a pull down experiment on Avicel. Figure 5 shows that the CBM3a-*Bp*GH9 fusion was detected in the Avicel pellet, while *Bp*GH9 alone was not retained. The improved binding of the CBM3a-*Bp*GH9 to crystalline cellulose also had a positive effect on the cellulose hydrolysis since a twofold improvement in the level of degradation of Avicel was observed compared to the native *Bp*GH9 (Fig. 6a).

Two other constructs were tested to investigate the role of CBM3c and CBM3a on the catalytic activity of *Bp*GH9 (Fig. 2). A truncated form, *Bp*GH9t, composed only of the catalytic module and a modular construct with the CBM3a fused with the *Bp*GH9t, CBM3a-*Bp*GH9t, were expressed and purified as a single band in SDS-PAGE (Fig. 5).

The catalytic activity of these constructs was tested on Na-CMC substrate (Fig. 6b). Results revealed that *Bp*GH9t had reduced activity on the Na-CMC compared to the native enzyme *Bp*GH9. The fusion of CBM3a to the truncated form of *Bp*GH9t without the CBM3c however, resulted in a recovery of the activity. Furthermore, truncated constructs exhibited rapid inactivation at 37 °C (data not shown). This low thermal stability prevented us from assaying the activity on crystalline cellulose, which requires longer incubation time at 50 °C. The pull-down experiments with Avicel (Fig. 5) revealed that the truncated form of *Bp*GH9 exhibited no binding affinity to crystalline cellulose. However, adding the CBM3a to this truncated form enhanced the binding to cellulose.

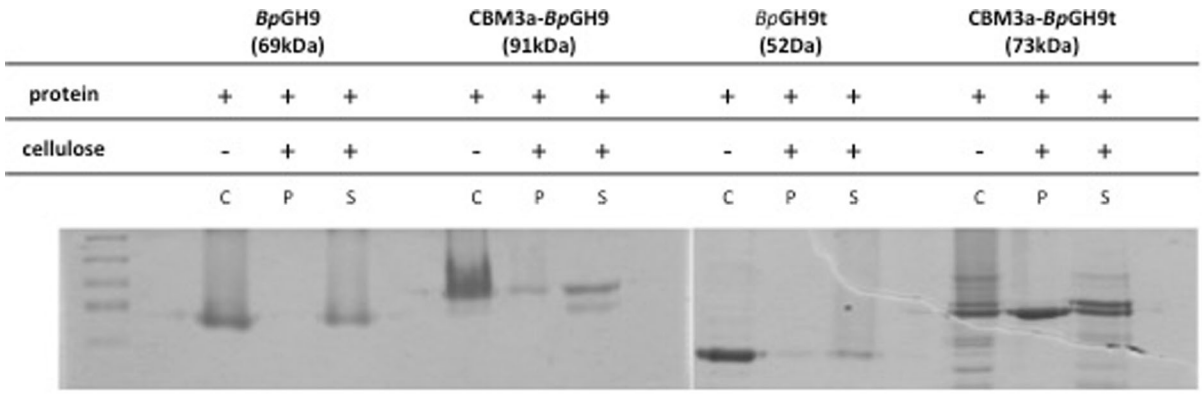

Fig. 5 Cellulose based affinity pull down assay for *Bp*GH9, CBM3a-*Bp*GH9, *Bp*GH9t and CBM3a-*Bp*GH9t over microcrystalline cellulose (Avicel PH-101), observed by SDS-PAGE with Coomassie blue staining (*C*: control protein fraction without Avicel, *P*: pelleted fractions with Avicel, *S*: supernatant soluble fraction; MW; protein molecular weight marker).). 3.5 µM of each protein was incubated at 4 °C for 2 h in presence of 10% Avicel PH-101 and centrifuged before being separated by SDS-PAGE

Discussion

Microbial strains isolated from the Kenyan soda lakes, such lakes Bogoria, Magadi, Turkana and Nakuru, have attracted attention since they have the potential to produce alkaliphilic enzymes. This is due to the high alkalinity (pH > 9) of their waters (Mwirichia et al. 2010). Indeed, several interesting alkaliphilic enzymes have been already characterized. Amylases active at pH 10 were extracted from a strain of *B. halodurans* (Hashim et al. 2004), lipases (Vargas et al. 2004), transpeptidases from a strain of *B. pumilus* (Moallic et al. 2006) and cellulases (Taylor et al. 2012). Some of these have found biotechnological applications in stain removal and colour preservation in textiles (Puradax HA, Genencor), or in the production of stonewashed effect on jeans (Indiage™, Genencor).

To further explore the diversity of the alkaliphilic cellulases, we identified a Bogoria strain of Bacillus sp., which expresses cellulolytic enzymes active at high pH (9–10). These cellulases were biochemically characterized after their cloning, which led to the identification of a new endo-cellulase from family GH9 (Cazy) and a new cellobiohydrolase from family GH48. Only the endocellulase BpGH9 was alkaliphilic and was shown to remain active up to pH 10. However, both enzymes are intracellular based on analysis with SignalP-5.0 (Almagro Armenteros et al. 2019), which suggests that the activity observed on the agar plates is probably a result of partial cell lysis.

The closest GH9 sequence to *Bp*GH9 that was biochemically characterized was EglA (99%) (Lima et al. 2005). This endo-cellulase was not alkaliphilic in contrast to *Bp*GH9. In addition, its translated sequence differs by only 8 amino acid residues (K40E, E86K, R191K, S371P, K511N, V553A, E604D, D605N). Based on the 3D model of *Bp*GH9 constructed by homology modelling (Supplementary Fig. 7), it appears that none of these mutations are located in the vicinity of the catalytic residues (D58 and E423), but they are located on the molecular surface. Therefore, it is difficult to infer the possible factors implicated in the alkaline adaptation of *Bp*GH9. Curiously, these mutations lead to a decreased charge residue content (K511N, D605N), which is in apparent contradiction with previous observations on alkaliphilic glycosidases (Zhao et al. 2011; Bai et al. 2015). However, the pH dependent activity curve of *Bp*GH9 compared to that of EglA, does not exhibit a significant shift in the optimum pH, but a better residual activity at high pH (> 8). Such small effects may involve a subtle change in the structure, which would require crystal structure analysis in order to be detected.

In nature, most cellulolytic enzymes are, either non-complexed cellulases such as those produced by aerobic fungi and most bacteria, or complexed cellulases organized in cellulosomes found in anaerobic bacteria and fungi. Most cellulases display a modular structure comprising a catalytic domain (GH) and accessory domains such as carbohydrate-binding

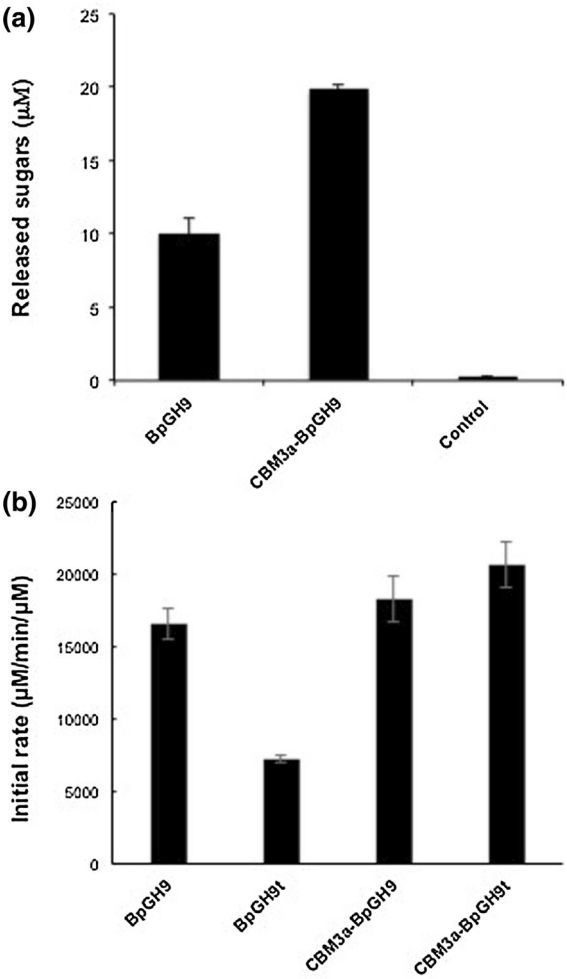

Fig. 6 Comparison of the enzymatic activities of *Bp*GH9, CBM3a-*Bp*GH9 on Avicel at 50 °C (**a**) and of *Bp*GH9t, CBM3a-*Bp*GH9t, *Bp*GH9, CBM3a-*Bp*GH9 on 1% Na-CMC at 20 °C (**b**). Enzymatic activities are defined as µM total reducing sugars liberated after 48 h on Avicel and µM total reducing sugars liberated per min and per µM enzyme on Na-CMC. Bar errors correspond to the standard deviations of triplicates

modules (CBMs). *Bp*GH9 from *B. pumilus* is a modular endoglucanase, that contains a catalytic domain and a *C*-terminal family 3c cellulose binding domain (CBM3c). This domain is known to be important for processivity by acting cooperatively with the catalytic domain (Chiriac et al. 2010).

In order to understand the role of the CBMs in the cellulose hydrolysis by *Bp*GH9, several truncated constructions of *Bp*GH9 without CBM3c and with an additional CBM3a from *C. thermocellum* were designed. As previously demonstrated in other GH9 endocellulases (Gilad et al. 2003; Petkun et al. 2015), removal of the CBM3c results in a strong reduction of the endoglucanase activity on soluble cellulose substrate (CMC). Regrettably, the effect of this truncation could not be explored on crystalline cellulose due to the lack of thermostability by the truncated enzyme (Kim et al. 2016). Moreover, the fusion of a CBM3a from *C. thermocellum* to *Bp*GH9 improves its activity on crystalline cellulose (Fig. 6a). Interestingly, in most natural endocellulases, CBMs are fused at the *C*-terminal position, but, here, the addition of a CBM3 on the *N*-terminus slightly enhances the enzyme activities on the soluble substrate CMC (Fig. 6b), in addition to increasing the activities on the insoluble substrates such as microcrystalline cellulose (Avicel). It is also interesting to note that the addition of the CBM3a module on the truncated *Bp*GH9t restores the catalytic activity to a level above to that of the native enzyme on amorphous cellulose (Na-CMC) (Fig. 6b). These data suggest that addition of CBM modules could modulate cellulolytic activity, whatever their positioning at either the *N*-terminus or the *C*-terminus of the catalytic domain. This fusion approach provides new routes to improve activity of endocellulases, which are naturally devoid of CBMs (Walker et al. 2015).

Acknowledgements LO was a recipient of a PhD scholarship (2014–2018) from the Embassy of France in Kenya, and of a stipendium (2015–2019), from the University of Nairobi. This work was supported by the Swedish International Development Cooperation Agency (Sida) sponsored Bioinnovate-Africa Program and UNEP GEF Agency Project No. 00876 on developing the microbial biotechnology industry from Kenya's Soda Lakes in line with the Nagoya Protocol.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK et al (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423

Bai W, Zhou C, Zhao Y, Wang Q, Ma Y, Permyakov EA (2015) Structural insight into and mutational analysis of family 11

xylanases: implications for mechanisms of higher pH catalytic adaptation. PLoS ONE 10:e0132834

Bayer EA, Belaich J-P, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58:521–554

Brunecky R, Alahuhta M, Xu Q et al (2013) Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science 342:1513–1516

Chiriac AI, Cadena EM, Vidal T, Torres AL, Diaz P, Javier Pastor FI (2010) Engineering a family 9 processive endoglucanase from *Paenibacillus barcinonensis* displaying a novel architecture. Appl Microbiol Biotechnol 86:1125–1134

Doi RH, Kosugi A (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol 2:541–551

Duckworth AW, Grant WD, Jones BE, Van Steenbergen R (1996) Phylogenetic diversity of soda lake alkaliphiles. FEMS Microbiol Ecol 19:181–191

Fontes CMGA, Gilbert HJ (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem 79:655–681

Gilad R, Rabinovich L, Yaron S et al (2003) CelI, a noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. J Bacteriol 185:391–398

Hashim SO, Delgado O, Hatti-Kaul R, Mulaa FJ, Mattiasson B (2004) Starch hydrolysing *Bacillus halodurans* isolates from a Kenyan soda lake. Biotechnol Lett 26:823–828

Jahangeer S, Khan N, Jahangeer S et al (2005) Screening and characterization of fungal cellulases isolated from the native environmental source. Pak J Bot 37:739

Kang Y, Shen M, Wang H, Zhao Q (2015) Complete genome sequence of *Bacillus pumilus* strain WP8, an efficient plant growth-promoting rhizobacterium. Genome Announc 3:e01452-e1514

Kim S-J, Kim SH, Shin SK, Hyeon JE, Han SO (2016) Mutation of a conserved tryptophan residue in the CBM3c of a GH9 endoglucanase inhibits activity. Int J Biol Macromol 92:159–166

Kipper K, Väljamäe P, Johansson G (2005) Processive action of cellobiohydrolase Cel7A from *Trichoderma reesei* is revealed as "burst"kinetics on fluorescent polymeric model substrates. Biochem J 385:527–535

Klemm D, Heublein B, Fink H-P, Bohn A (2005) Cellulose: fascinating biopolymer and sustainable raw material. Angew Chem Int Ed 44:3358–3393

Lima AOS, Quecine MC, Fungaro MHP et al (2005) Molecular characterization of a β-1,4-endoglucanase from an endophytic *Bacillus pumilus* strain. Appl Microbiol Biotechnol 68:57–65

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495

Mandelman D, Belaich A, Belaich JP, Aghajari N, Driguez H, Haser R (2003) X-Ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. J Bacteriol 185:4127–4135

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428

Moallic C, Dabonné S, Colas B, Sine J-P (2006) Identification and characterization of a gamma-glutamyl transpeptidase from a thermo-alcalophile strain of *Bacillus pumilus*. Protein J 25:391–397

Mwirichia R, Muigai AW, Tindall B, Boga HI, Stackebrandt E (2010) Isolation and characterisation of bacteria from the haloalkaline Lake Elmenteita, Kenya. Extremophiles 14:339–348

Petkun S, Rozman Grinberg I, Lamed R et al (2015) Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: structural and functional significance of the intermodular linker. PeerJ 3:e1126

Stern J, Kahn A, Vazana Y et al (2015) Significance of relative position of cellulases in designer cellulosomes for optimized cellulolysis. PLoS ONE 10:e0127326

Taylor MP, van Zyl L, Tuffin M, Cowan D (2012) Extremophiles and biotechnology: how far have we come? In: Anitori RP (ed) Extremophiles: microbiology and biotechnology. Horizon Scientific Press, Hethersett, pp 1–24

Uchiyama T, Uchihashi T, Nakamura A, Watanabe H, Kaneko D, Samejima M, Igarashi K (2020) Convergent evolution of processivity in bacterial and fungal cellulases. Proc Natl Acad Sci 117:19896–19903

Vargas VA, Delgado OD, Hatti-Kaul R, Mattiasson B (2004) Lipase-producing microorganisms from a Kenyan alkaline soda lake. Biotechnol Lett 26:81–86

Walker JA, Takasuka TE, Deng K et al (2015) Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. Biotechnol Biofuels 8:220

Zhang X-Z, Zhang Y-HP (2013) Cellulases: characteristics, sources, production, and applications. Bioprocess Technol Biorefinery Sustain Prod Fuels Chem Polym 1:131–146

Zhao Y, Zhang Y, Cao Y et al (2011) Structural analysis of alkaline β-mannanase from alkaliphilic *Bacillus* sp. N16–5: implications for adaptation to alkaline conditions. PLoS ONE 6:e14608