Curr Microbiol (2009) 58:404–408
DOI 10.1007/s00284-008-9346-y

Identification of Cellulase Genes from the Metagenomes of Compost Soils and Functional Characterization of One Novel Endoglucanase

Hao Pang · Peng Zhang · Cheng-Jie Duan · Xin-Chun Mo · Ji-Liang Tang · Jia-Xun Feng

Received: 9 September 2008 / Accepted: 10 December 2008 / Published online: 22 January 2009

Abstract Metagenomics, a new research field developed over the past decade, aims to identify potential enzymes from nonculturable microbes. In this study, genes encoding three glycoside hydrolase family (GHF) 9 endoglucanases and one GHF 5 endoglucanase were cloned and identified from the metagenome of the compost soils. The shared identities between the predicted amino acid sequences of these genes and their closest homologues in the database were less than 70%. One GHF 9 endoglucanase, Umcel9B, was further characterized. The recombinant protein, Umcel9B, showed activity against carboxymethyl cellulose, indicating that Umcel9B is an endoactive enzyme. Enzymatic activity occurs optimally at a pH of 7.0 and a temperature of 25°C.

Introduction

Cellulases have attracted enormous research and commercial interests due to their potential values in biotechnology (e.g., conversion from biomass to biofuels). Cellulases comprise three classes of enzymes, namely, endo-1,4-β-glucanase (EC3.2.1.4), cellobiohydrolase (EC3.2.1.91), and β-glucosidase (EC3.2.1.21) [10]. These cellulolytic enzymes catalyze the hydrolysis of the β-1,4-glucosidically linked cellulose chain.

Endoglucanase has broad applications in industry [2]. For example, in the biopolishing process, endoglucanase is more efficient than any other cellulases in removing color from denim to produce a good stonewashing effect [12]. As a detergent product, endoglucanase is an additive efficient for color brightening, softening, and removal of soil particulate [17]. In addition, endoglucanase is important for its ability to cleave the glycosidic bond of β-glucan by a random action that decreases the viscosity of the β-glucan solution, benefits beer brewing, and improves the quality of animal feed [2]. Specifications of these applications suggested the necessity to identify a wide range of endoglucanases with varying optimal pH and temperature.

Fungi and bacteria are the main sources of cellulases [10]. However, because the majority of the microorganisms are unculturable with traditional techniques [14], obtaining novel genes from those unculturable environmental microorganisms became necessary [3]. The metagenomic approach, a culture-independent method, has been used to clone a number of genes, including cellulase genes [19]. To date, only three metagenome-derived cellulases genes have been identified, with biochemical characterization of the protein products [5, 9, 20]. The potential use in industry for the cellulases cloned from metagenome has not been fully explored.

In this study, a cosmid metagenomic library was constructed using DNA isolated from compost soils. Through activity-based screening of the library, four novel cellulase genes were isolated. One of these genes, *umcel9B*, was

Electronic supplementary material The online version of this article (doi:10.1007/s00284-008-9346-y) contains supplementary material, which is available to authorized users.

H. Pang · P. Zhang · C.-J. Duan · X.-C. Mo · J.-L. Tang · J.-X. Feng (✉)
Guangxi Key Laboratory of Subtropical Bioresources Conservation and Utilization, The Key Laboratory of Ministry of Education for Microbial and Plant Genetic Engineering, College of Life Science and Technology, Guangxi University, 75 Xiuling Road, Nanning, Guangxi 530005, People's Republic of China
e-mail: jiaxunfeng@sohu.com

H. Pang · P. Zhang · C.-J. Duan · X.-C. Mo · J.-L. Tang · J.-X. Feng
Guangxi Key Laboratory of BioIndustry Technology, 98 Daling Road, Nanning, Guangxi 530004, People's Republic of China

overexpressed in *Escherichia coli*, and the purified recombinant enzyme was biochemically characterized.

Materials and Methods

Metagenomic DNA Preparation and Purification

Metagenomic DNA was extracted from aerated compost soil (~50 g). The microbial cells were recovered from the soil samples by differential centrifugation, as described by Patrick et al. [13]. After treatment with lysozyme, microbial cells were resuspended in 600 µl of the cell lysis solution from the Puregene Genomic DNA Purification Kit (Gentra Systems Inc., Minneapolis MN, USA). This was incubated for 5 min at 80°C, and then 200 µl of the protein precipitation solution from the kit was added. After centrifugation, the DNA in the supernatant was precipitated by isopropanol and dissolved in 500 µl of tris-EDTA (TE) buffer.

Next, the crude DNA solution was loaded onto a column (200 × 10 mm) containing Sephadex G200 and acid-washed polyvinylpolypyrrolidone (PVPP) at a ratio of 50:1. Finally, DNA was eluted from the column with TE buffer and further purified by electroelution as described [16].

Construction and Screening of the Cosmid Library

A metagenomic library was constructed by using the pWEB::TNC Cosmid Cloning Kit (Epicentre, Madison, WI, USA) according to the manufacturer's instruction. Briefly, the purified DNA was end repaired with T4 DNA polymerase and T4 polynucleotide kinase to generate blunt ends. End-repaired DNA was separated on a low-melting-point agarose gel (FMC Corporation, Philadelphia, PA, USA). The DNA fragments between 40 and 50 kb were recovered and ligated into the cosmid vector pWEB::TNC that had been linearized at the unique *Sma*I site and dephosphorylated. The ligated products were packaged, and infected *E. coli* EPI100.

The colonies within the library were screened for endo-1,4-β-glucanase (carboxymethylcellulase) activity according to protocols described by Teather and Wood [18].

Identification of Cellulase Genes

To identify the cellulase genes carried on recombinant plasmids pGXN1027, pGXN1030, pGXN1042, and pGXN1044, these plasmids were nested-deleted from one end by *in vitro* Tn5 transposition (pWEB::TNC Deletion Cosmid Transposition Kit; Epicentre) according to the manufacturer's protocol. The plasmid with the smallest insertion expressing carboxymethylcellulase (CMCase) activity was sequenced from priming sites on the vector adjacent to the deletion start sites using primer 5′-TGTGAAATTTGTGATGCTATTGCT-3′ provided by the manufacturer. Complete coverage of the sequences for the target gene was obtained by primer walking sequencing from both strands.

Overexpression and Purification of the Recombinant Endoglucanase

The sequence of *umcel9B*, excluding the sequence encoding the N-terminal signal peptide, was amplified by polymerase chain reaction (PCR) using pGXN1030 as the template and the following primers: sense primer 5′- ATAGCATGCC CGCAGCTGCGAAGATCC-3′ (containing a *Sph*I site at the 5′ end) and antisense primer 5′-ATAGTCGACCTA CTCCCGCAATGCCGC-3′ (containing an *Sal*I site at the 5′ end). Amplified DNA was digested with *Sph*I and *Sal*I before its ligation into the vector pQE31 digested with the same enzymes (Qiagen, Valencia, CA, USA), resulting in the plasmid pGXN9B. The recombinant plasmid, pGXN9B, then was transformed into *E. coli* strain M15 (pREP4) (Qiagen). His-tagged Umcel9B was expressed and purified using nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen) according to the manufacturer's instruction.

Enzymatic Assays

Enzymatic reaction was carried out in 500 µl of mixtures (pH 7.0) containing 1% substrate at 25°C, and the enzymatic activity was quantified by measuring the quantity of generated reducing sugars (as D-glucose equivalents) with the 3′, 5′-dinitrosalicylic acid reagent (DNS reagent). One international unit (U) corresponds to the production of glucose at 1 µmol/min.

Results and Discussion

Construction of the Metagenomic Library and Screening of the Library for Clones Expressing CMCase Activity

A cosmid library of *ca.* 100,000 clones (library NNLC1) was constructed using the metagenomic DNA isolated from the compost. A *Bam*HI restriction analysis of 24 randomly selected clones from NNLC1 showed that all the clones harbored insertion DNAs ranging from 20 to 50 kb (data not shown). The average insert size of the library was approximately 33 kb. The library possessed a capacity approximating 3.3×10^9 bp of insertion DNA. Four clones expressing CMCase activity were isolated after all the clones in the NNLC1 library had been screened.

The positive rate of CMCase activity in the library was approximately 1/25,000. Because metagenomic DNA originated from tens of thousands of microorganisms [11], it becomes difficult to screen the positive clones from metagenomic libraries [6]. In an early report, Rondon et al. [15] identified clones with the activity of DNAse (one clone), antibacteria (one clone), lipase (two clones), and amylase (eight clones) from a library consisting of 24,000 clones (a bacterial artificial chromosome (BAC) library constructed with metagenomic DNA from soil). A recent article reported that the positive rate of cellulase in a cosmid library constructed from soil samples was approximately 1/70,000 [9], which is slightly lower than the results obtained in this study.

Sequence Analyses of the Cloned Cellulase Genes

The positive clones (pGXN1027, pGXN1030, pGXN1042, and pGXN1044) were sequenced, and the genes in each clone were named as *umcel9A*, *umcel9B*, *umcel9C*, and *umcel5A,* respectively (Table 1). Overall, the shared identities between the deduced amino acid sequences of these four genes and the known cellulases in the database are less than 70%, indicating that these four cloned genes are novel (Table 1). Genes *umcel9A*, *umcel9B,* and *umcel9C* encoded an endoglucanase consisting of an immunoglobulin (Ig)-like domain and a glycoside hydrolase family (GHF) 9 catalytic domain. Gene *umcel5A* encoded a GHF 5 endoglucanase that contains only a catalytic domain. Phylogenetic analysis showed that these three GHF 9 cellulases were grouped in one branch (Supplementary Fig. S1), indicating that they probably evolved from the same origin and shared similar enzymatic functions.

The amino acid sequences of Umcel9A, Umcel9B, and Umcel9C also were compared using the AlignX Module of Vector NTI suite 10 (Informax, North Bethesda, Maryland, USA). The results showed that Umcel9A and Umcel9B shared a similarity of 58.8% (identities of 46.9%) and that Umcel9A and Umcel9C shared a similarity of 62.8% (identities of 49.6%). The similarity shared by Umcel9B and Umcel9C was 56.7% (identities of 45.3%).

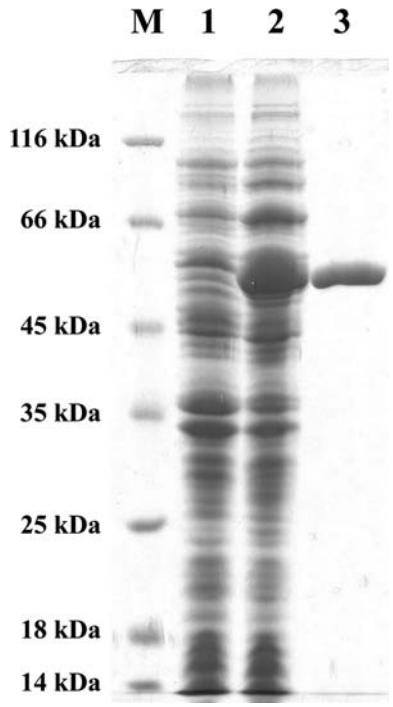

Fig. 1 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of recombinant Umcel9B protein stained with Coomassie blue. Lane M: protein molecular weight marker. Lane 1: crude extract of control *Escherichia coli*. Lane 2: crude extract of *E. coli* harboring the expression plasmid with the *umcel9B* gene. Lane 3: purified Umcel9B. Recombinant protein was purified with nickel-nitrilotriacetic acid (Ni-NTA) chromatography

Overexpression of the Gene *umcel9B* in *Escherichia coli* and Purification of the Recombinant Umcel9B

Umcel9B without the N-terminal signal peptide was expressed in *E. coli* M15 (pREP4) with a six-histidine tag at the N terminus of the fused protein. Overexpression was achieved by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mmol/l and incubating it at 37°C for 5 h after OD_{600} of the culture reached 0.5.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the crude extract showed abundant expression of 6xHis tagged proteins, as observed

Table 1 Sequence analyses of four cellulase genes from the metagenomic library

Clone name	Gene name	Gene accession number	ORF (bp)	Most similar enzyme, accession number, and original strain	Identity/positive (%)
pGXN1027	*umcel9A*	DQ235085	1,851	Cel9B, ACE85719, *Cellvibrio japonicus*	69/81
pGXN1030	*umcel9B*	DQ235086	1,740	Umcel9A, ABB51609, uncultured bacterium	52/67
				Cel9B, ACE85719, *Cellvibrio japonicus*	51/67
pGXN1042	*umcel9C*	DQ235087	1,761	Cellulase, AAM41665, *Xanthomonas campestris* pv. *campestris* str. ATCC 33913	58/71
pGXN1044	*umcel5A*	DQ235088	1,047	Cel5H, ACE83988, *Cellvibrio japonicus*	60/72

ORF open reading frame

Table 2 Substrate specificity of Umcel9B

Substrates	Specific activity[a] (U/mg)	Relative activity (%)
Carboxymethyl cellulose	5.56 ± 0.27	100
β-D-glucan from barley	3.98 ± 0.07	71.6
Methyl cellulose	1.00 ± 0.07	18.0
2-Hydroxyethyl cellulose	0.61 ± 0.02	10.8
Lichenan	0.09 ± 0.02	1.6
Oat spelt xylan	UD	
Laminarin	UD	

UD, undetectable

[a] The incubation time for activity measurement was 10 min. The measurements were performed in triplicate, and the standard deviations were calculated

by the appearance of an extra protein band migrating at about 60 kDa upon induction (Fig. 1, lanes 1 and 2). The size of the expressed Umcel9B was similar to the molecular mass calculated from the amino acid sequences (60 kDa). After purification with the Ni-NTA column, a single band was shown on the SDS-PAGE gel correlating with the size of the enzyme, indicating that the enzyme was purified to homogeneity (Fig. 1, lane 3).

Functional Characterization of Umcel9B

The enzymatic activity of the recombinant Umcel9B on substrates was determined (Table 2). The results demonstrated that Umcel9B possessed high activity toward soluble cellulosic derivatives such as carboxymethyl cellulose (CMC), 2-hydroxyethyl cellulose, and methyl cellulose. This enzyme also hydrolyzed β-D-glucan from barley and lichenan. Umcel9B showed a weak activity toward Avicel even after a long incubation (data not shown). It was not able to hydrolyze xylan or laminarin. The substrate specificity of this enzyme indicated that it belongs to the family 9 cellulases [1].

Metal ions (10 mmol/l) were added to the optimal reaction systems to determine their effects on the enzymatic activity. The presence of $CoCl_2$ and $CaCl_2$ enhanced the enzymatic activity to 140% and 143%, respectively, whereas $AgNO_3$, $CuCl_2$, and $FeCl_3$ dramatically reduced the enzymatic activity to 54%, 53%, and 62%, respectively.

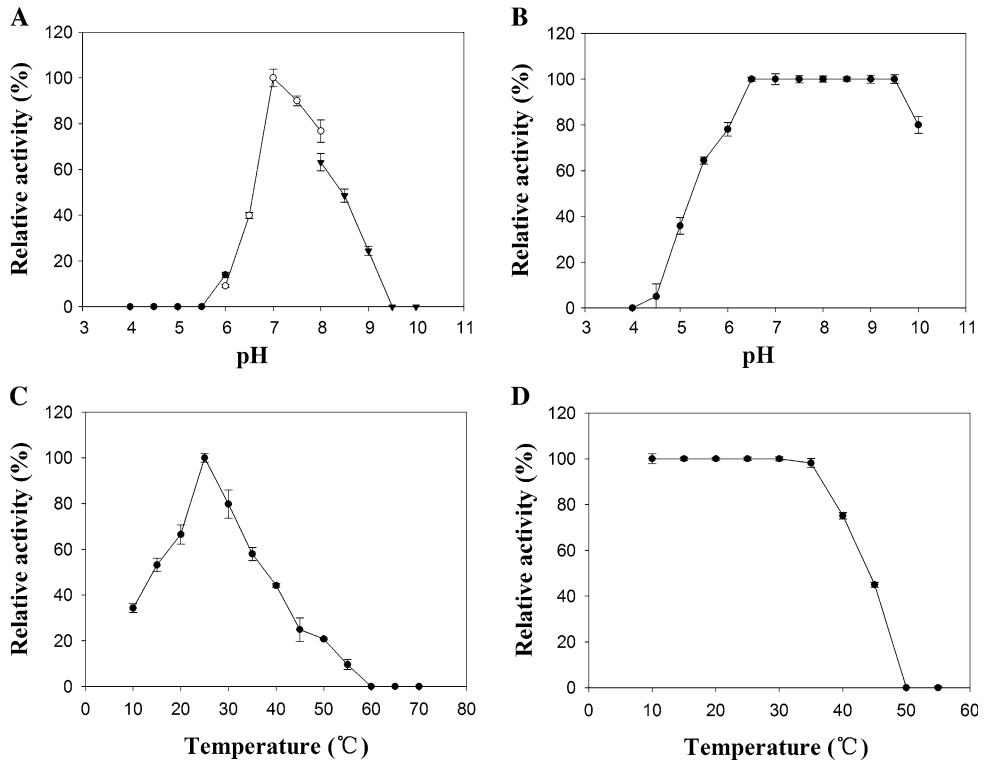

Fig. 2 Effects of pH and temperature on the activity and stability of the enzyme. **a** Effect of pH on enzymatic activity measured in 100 mmol/l of succinic acid-NaOH buffer (pH 4–6), sodium phosphate buffer (pH 6–8), and glycine-NaOH buffer (pH 8–10). **b** Stability of the enzyme in different pH conditions. Enzyme activity was measured under the optimal condition (pH 7.0, 25°C, 20 min) after the enzyme was incubated in the buffers described earlier at 4°C for 24 h. **c** The effect of temperature on the activity of Umcel9B. Experiments were conducted in a temperature range of 10°C to 70°C at pH 7.0. **d** Stability of Umcel9B in different temperature conditions. Enzyme activity was measured under the optimal condition (pH 7.0, 25°C, 20 min) after the enzyme had been incubated at the indicated temperature for 1 h. The error bars represent the standard deviation of triplicate measurements

According to the findings, KCl, NaCl, LiCl, KNO_3, and $MgCl_2$ had no significant effect on the enzymatic activity. Chelating agent ethylenediaminetetraacetic acid (EDTA) (1 mmol/l) and anionic surfactant SDS (2%) were effective inhibitors, reducing the enzyme activity to 52% and 32%, respectively.

Temperature and pH Effect on Activity of Umcel9B

This purified enzyme Umcel9B was active at a pH range of 6 to 9. The optimal pH of the enzyme is 7.0 because it achieves maximal enzymatic activity at this condition (Fig. 2a). Enzymatic activity remained at a high level when the pH was between 6.5 and 9.0 (Fig. 2b). However, when the pH was out of the 5.5 to 9.5 range, the activity of Umcel9B was completely lost. The pH range of the recombinant enzyme was consistent with the property of alkaline endoglucanase reported previously [7].

This enzyme Umcel9B showed its main activity in the temperature range of 10 to 40°C, and its maximum activity was observed when Umcel9B was incubated at 25°C (Fig. 2c). When the temperature was higher than 45°C, enzymatic activity was quickly lost. After 1 h of incubation at different temperatures, Umcel9B retained its activity only below 40°C (Fig. 2d).

The enzymatic property of Umcel9B at low temperatures is similar to the property of cold-active endoglucanase [8, 21]. Cold-active enzymes are attractive for their values in biotechnological applications. They also are useful tools for protein-folding studies because of their high activity and stability at low temperatures [4].

Acknowledgment This work was supported by Grants for New Century Excellent Talents in Universities of China (NCET-05-0752), the National Natural Science Foundation of China (30560003) and Hi-tech Research and Development Program of China (863 Program 2007AA021307).

References

1. Arai T, Kosugi A, Chan H et al (2006) Properties of cellulosomal family 9 cellulases from *Clostridium cellulovorans*. Appl Microbiol Biotechnol 71:654–660
2. Bhat MK (2000) Cellulases and related enzymes in biotechnology. Biotechnol Adv 18:355–383
3. Daniel R (2004) The soil metagenome—a rich resource for the discovery of novel natural products. Curr Opin Biotechnol 15:199–204
4. Feller G, Gerday C (2003) Psychrophilic enzymes: hot topics in cold adaptation. Nat Rev Microbiol 1:200–208
5. Feng Y, Duan CJ, Pang H et al (2007) Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. Appl Microbiol Biotechnol 75:319–328
6. Ferrer M, Golyshina O, Beloqui A et al (2007) Mining enzymes from extreme environments. Curr Opin Microbiol 10:207–214
7. Hakamada Y, Endo K, Takizawa S et al (2002) Enzymatic properties, crystallization, and deduced amino acid sequence of an alkaline endoglucanase from *Bacillus circulans*. Biochim Biophys Acta 1570:174–180
8. Iyo AH, Forsberg CW (1999) A cold-active glucanase from the ruminal bacterium *Fibrobacter succinogenes* S85. Appl Environ Microbiol 65:995–998
9. Kim SJ, Lee CM, Han BR et al (2008) Characterization of a gene encoding cellulase from uncultured soil bacteria. FEMS Microbiol Lett 282:44–51
10. Lynd LR, Weimer PJ, van Zyl WH et al (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577
11. McHardy AC, Rigoutsos I (2007) What's in the mix: phylogenetic classification of metagenome sequence samples. Curr Opin Microbiol 10:499–503
12. Miettinen-Oinonen A, Suominen P (2002) Enhanced production of *Trichoderma reesei* endoglucanases and use of the new cellulase preparations in producing the stonewashed effect on denim fabric. Appl Environ Microbiol 68:3956–3964
13. Patrick W, Joey W, Alison R et al (2003) A rapid, cost-effective procedure for the extraction of microbial DNA from soil. World J Microb Biot 19:85–91
14. Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394
15. Rondon MR, August PR, Bettermann AD et al (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66:2541–2547
16. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York
17. Shimonaka A, Koga J, Baba Y et al (2006) Specific characteristics of family 45 endoglucanases from *Mucorales* in the use of textiles and laundry. Biosci Biotechnol Biochem 70:1013–1016
18. Teather RM, Wood PJ (1982) Use of Congo red-polysaccharide interactions in enumeration and characterisation of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol 43:777–780
19. Voget S, Leggewie C, Uesbeck A et al (2003) Prospecting for novel biocatalysts in a soil metagenome. Appl Environ Microbiol 69:6235–6242
20. Voget S, Steele HL, Streit WR (2006) Characterization of a metagenome-derived halotolerant cellulase. J Biotechnol 126:26–36
21. Zeng R, Xiong P, Wen J (2006) Characterization and gene cloning of a cold-active cellulase from a deep-sea psychrotrophic bacterium *Pseudoalteromonas* sp. DY3. Extremophiles 10:79–82