J Ind Microbiol Biotechnol (2012) 39:347–357
DOI 10.1007/s10295-011-1039-z

GENETICS AND MOLECULAR BIOLOGY OF INDUSTRIAL ORGANISMS

Cloning and bioinformatics analysis of an endoglucanase gene (*Aucel12A*) from *Aspergillus usamii* and its functional expression in *Pichia pastoris*

Hongling Shi · Xin Yin · Minchen Wu · Cunduo Tang · Huimin Zhang · Jianfang Li

Received: 27 June 2011 / Accepted: 5 September 2011 / Published online: 21 September 2011
© Society for Industrial Microbiology 2011

Abstract Using 3′ and 5′ rapid amplification of cDNA ends methods, the full-length cDNA sequence encoding an endo-1,4-*β*-glucanase of *Aspergillus usamii* E001 (abbreviated as AuCel12A) was amplified from the total RNA. The clone cDNA sequence of the gene encoding the AuCel12A, named as *Aucel12A*, is 1,027 bp in length harboring 5′ and 3′ non-coding regions, as well as a 720 bp of open reading frame that encodes a 16-aa signal peptide, and a 223-aa mature AuCel12A with a theoretical M.W. of 24,294 Da, a calculated p*I* of 4.15, and one putative *N*-glycosylation site. The complete DNA sequence of the gene *Aucel12A* was amplified from the genomic DNA of *A. usamii* E001 by using the conventional PCR and pUCm-T vector-mediated PCR initially developed in our lab. The clone DNA sequence is 1,576 bp in length, consisting of a 5′ flanking regulatory region, three exons, and two introns with sizes of 50 and 66 bp. The cDNA fragment encoding the mature AuCel12A was expressed in a fully active form in *Pichia pastoris*. One *P. pastoris* transformant expressing the highest recombinant AuCel12A (rAuCel12A) activity, labeled as *P. pastoris* GSCel2-1, was chosen for subsequent studies. Integration of the *Aucel12A* into *P. pastoris* genome was confirmed by PCR analysis using 5′- and 3′-*AOX1* primers. SDS-PAGE and enzyme activity assays demonstrated that the rAuCel12A, a glycosylated protein with an apparent M.W. of 27.0 kDa and a carbohydrate content of 4.82%, was secreted into the culture medium. The purified rAuCel12A displayed the highest activity at pH 5.0 and 60°C. It was highly stable at a pH range of 3.5–7.0, and at a temperature of 55°C or below. Its activity was not significantly affected by an array of metal ions and EDTA, but inhibited by Ag^{+}, Hg^{2+} and Fe^{2+}. The K_m and V_{max} of the rAuCel12A, towards carboxymethylcellulose-Na (CMC-Na) at pH 5.0 and 50°C were 4.85 mg/ml and 160.5 U/mg, respectively.

Keywords *Aspergillus usamii* · Bioinformatics analysis · Endoglucanase gene · Cloning and expression · *Pichia pastoris* · Enzymatic properties

H. Shi · M. Wu (✉)
School of Medicine and Pharmaceutics, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu 214122
People's Republic of China
e-mail: biowmc@126.com

X. Yin · C. Tang
The Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu 214122, People's Republic of China

H. Zhang · J. Li
School of Food Science and Technology, Jiangnan University, 1800 Lihu Road, Wuxi Jiangsu 214122, People's Republic of China

Introduction

Cellulose, the most widespread and abundant non-starch polysaccharide, is a major structural component of plant cell walls and, together with hemicellulose such as xylan, mannan, heteromannan, arabinogalactan and arabinan, accounting for more than half of all renewable organic carbon resources in the biosphere [13, 20]. The complete biodegradation of cellulose requires the synergistic action of several cellulolytic enzymes, such as endo-1,4-*β*-glucanases (abbreviated as endoglucanases, EC 3.2.1.4), exo-1,4-*β*-glucanases (EC 3.2.1.91) and *β*-glucosidases (EC 3.2.1.21), which are collectively known as cellulases

[1]. Endoglucanases, which are widely distributed in various organisms such as plants, bacteria, fungi, actinomycetes and mollusks, catalyze the random cleavage of the internal β-1,4-D-glycosidic linkages within amorphous regions of cellulose chains to yield cellooligosaccharides, the latter being suitable substrates for other cellulolytic enzymes [5]. To date, almost all known cellulases have been classified into glycoside hydrolase (GH) families 5–9, 12, 26, 44, 45, 48, and 61 based on their amino acid sequence similarities and hydrophobic cluster analysis (http://www.cazy.org/). Applications of cellulases began in early 1980s in the preparation of animal feed, and later developed to the food, brewing, textile, laundry detergent, biofuel, paper and pulp, waste management and pharmaceutical industries. Since then, the demands for cellulases have increased dramatically, covering a wide range of industrial fields [2, 9, 15]. Recently, cellulases together with hemicellulases accounted for approximately 20% of the global industrial enzyme market [23], and will continue to grow as new types of cellulases are identified and characterized.

During the past decades, most studies on the industrial production, gene cloning and expression of fungal endoglucanases have focused on *Trichoderma* genera [14, 20], *Penicillium* genera [3, 5], *Volvariella volvacea* [6, 13], *Aspergillus kawachii* [17], and *Aspergillus niger* [8]. In our previous studies, a filamentous fungus of *A. usamii* E001 strain, closely related to *A. niger* in taxonomy and physiological properties, was isolated from the soil in China and preserved in our lab. This strain can produce a series of acidophilic glycoside hydrolases such as xylanases, β-mannanases and cellulases by solid-state fermentation, but the cellulolytic activity was much lower [21]. Therefore, enhancing the cellulolytic activity by means of genetic engineering is highly desirable. In this work, we reported the cloning and bioinformatics analysis of the full-length cDNA and complete genomic copy of the gene *Aucel12A* that encodes the AuCel12A of *A. usamii* E001. Meanwhile, the heterologous expression of the gene *Aucel12A* in *P. pastoris* and the enzymatic properties of the rAuCel12A were also described. As a eukaryotic expression system, the yeast *P. pastoris* expression system has been a favorite system for expressing heterologous proteins owing to its many advantages, such as protein processing, protein folding, posttranslational modification, and high expression levels. Most importantly, the yeast *P. pastoris* can be grown and induced in inexpensive media.

Our present results will provide a solid basis for further studies on the rAuCel12A production by fermentation on an industrial scale, relationship between its structure and function, and improvement of its enzymatic properties such as pH stability, thermostability, and resistibility to metal ions and reagents.

Materials and methods

Strains, vectors, and media

Aspergillus usamii E001, isolated by and preserved in the lab of biochemistry and molecular biology, School of Medicine and Pharmaceutics, Jiangnan University, China as reported previously [23], was used as the source for total RNA and genomic DNA extraction. *E. coli* JM109 and pUCm-T (Sangon, Shanghai, China) were used as host-vector system for gene cloning and DNA sequencing. *E. coli* DH5α and pPIC9 K (Invitrogen, USA) were used for construction and sequencing of the recombinant expression plasmid. The expression of the gene *Aucel12A* was in *P. pastoris* GS115 (*his4*, Mut$^+$). *A. usamii* E001 was cultured in a carboxymethylcellulose (CMC)-containing liquid medium containing 1.0% tryptone, 0.5% yeast extract, 0.1% KH_2PO_4, 0.1% Na_2HPO_4, 0.05% $MgSO_4$, 0.02% $CaCl_2$, 0.1% $(NH_4)_2SO_4$ and 1.0% CMC-Na (Sigma, St. Louis, Mo, USA), pH 6.0. *E. coli* JM109 and DH5α were grown in the Luria–Bertani medium [18], to which ampicillin was added at a concentration of 100 μg/ml as required. MD, YPD containing geneticin G418 from 0.5 to 2.0 mg/ml, BMGY, and BMMY media were prepared as described by Multi-Copy Pichia Expression Kit (Invitrogen).

Total RNA and genomic DNA extraction

Aspergillus usamii E001 was cultured in the above-mentioned medium on a rotary incubator (220 rpm) at 30°C for 36 h. The mycelia were collected through filtration, and then thoroughly washed with sterile deionized water. The total RNA was extracted from the mycelia by using one-step method according to the specification of TRIzol Kit (Sangon). Extraction of the genomic DNA from the mycelia of *A. usamii* E001 was performed according to the method as reported previously [23].

Primers for RT-PCR and nested PCR amplification

After aligning three sequences of the GH family, 12 fungal endoglucanases from the *Aspergillus* genera, *A. kawachii* (GenBank accession: D12901), *A. terreus* (GenBank accession: XM_001214697), and *A. aculeatus* (GenBank accession: X52525), we found that there is a peptide fragment located in the N-terminal region, W(T/N)WSGGE (G/N)(T/S)VKSY(S/A)NSG, with the highest identity. Two degenerate primers celF1 and celF2 were designed corresponding to W(T/N)WSGGE and VKSY(S/A)NS, respectively. Primers dT-PR and PR (original names, Oligo dT-M13 Primer M4 and M13 Primer M4, respectively) provided by the RNA PCR Kit (TaKaRa, Dalian, China), as

Table 1 The sequences of primers for PCR amplification

Fragments	Primers	Sequences (5′–3′)[a]	Size (bp)
3′-End cDNA fragment	celF1	TGGAMCTGGWSTGGYGGSGA	20
	celF2	GTSAARAGCTACKCBAACTC	20
	dT-PR	GTTTTCCCAGTCACGAC-Oligo dT	37
	PR	GTTTTCCCAGTCACGAC	17
5′-end cDNA fragment	celR1	TTGTCCTGCTTCCATTCCAC	20
	celR2	ACATCACTCACGAGCTTCTT	20
	OP	CATGGCTACATGCTGACAGCCTA	23
	IP	CGCGGATCCACAGCCTACTGATGATCAGTCGATG	34
Mature peptide cDNA	celF	CAGACGATGTGCTCTCAAT	25
	celR	GCGGCCGCCTAGTTGACACTGGCGGTC	27
Core region of the complete DNA	celF3	ATGAAGCTCCCCGTGTCAC	19
	celR	GCGGCCGCCTAGTTGACACTGGCGGTC	27
5′ Flanking regulatory region	T-PriF	ACGACTCACTATAGGGCGACA	21
	celR1	TTGTCCTGCTTCCATTCCAC	20
	celR2	ACATCACTCACGAGCTTCTT	20

[a] M = A/C, W = A/T, S = C/G, Y = C/T, R = A/G, K = G/T, and B = T/G/C

well as celF1 and celF2 were used for the cloning of 3′-end fragment of AuCel12A cDNA. Primers OP and IP (original names, 5′ RACE Outer Primer and 5′ RACE Inner Primer, respectively) provided by 5′-Full RACE Kit (TaKaRa) together with primers celR1 and celR2 were used for the cloning of 5′-end fragment of AuCel12A cDNA. A pair of specific primers celF and celR with *Eco*RI and *Not* I sites (underlined), respectively, was used directly to amplify the cDNA sequence that encodes the mature AuCel12A. All the reverse transcription reactions were performed following the kit instructions provided by the manufacturers. Primers celF3 and celR were used directly to amplify the core region of the complete DNA of the gene *Aucel12A*. Using pUCm-T vector-mediated PCR method, the 5′ flanking regulatory region of the gene *Aucel12A* was amplified with primers T-PriF (identical to the 21-bp fragment upstream the T/A clone site on pUCm-T vector), celR1 and celR2. As listed in Table 1, all PCR primers (except ones provided by kits) were synthesized by Sangon (Shanghai, China).

Cloning of the full-length cDNA encoding the AuCel12A

The 3′-end fragment of AuCel12A cDNA was amplified by using the RNA PCR Kit and nested PCR technique. The dT-PR was used as a primer for reverse transcription of the first-strand cDNA from the total RNA. Using the resulting first-strand cDNA as template, the first-round PCR was performed with primers celF1 and PR as following conditions: an initial denaturation at 94°C for 2 min; 30 cycles of at 94°C for 30 s, 52°C for 30 s, 72°C for 60 s; an extra elongation at 72°C for 10 min, and then the second-round PCR was performed with primers celF2 and PR for confirmation (nested PCR) as following conditions: an initial denaturation at 94°C for 2 min; 30 cycles of at 94°C for 30 s, 53°C for 30 s, 72°C for 50 s, and an extra elongation at 72°C for 10 min. Next, the 5′-end fragment of AuCel12A cDNA, originating from the starting point of transcription, was amplified by using 5′-Full RACE Kit and nested PCR method. The first-strand cDNA was used as template for a first round of PCR with primers OP and celR1 and then subjected to a second round of PCR with primers IP and celR2 for confirmation. Finally, the full-length cDNA sequence that encodes the AuCel12A of *A. usamii* E001 was either obtained by assembling above cloned 3′- and 5′-end cDNA fragment sequences, or directly amplified by conventional PCR using the first-strand cDNA as template.

pUCm-T vector-mediated PCR amplification

The pUCm-T vector-mediated PCR amplification, a novel method initially developed in our lab to amplify the 5′ or 3′ flanking region of a known DNA fragment sequence, was carried out by four steps as flowcharted in Fig. 1 (exemplified as the cloning of the 5′ flanking region). Firstly, the genomic DNA extracted from *A usamii* E001 was digested using two optimum restriction enzymes, which were selected by series of pre-experiments to obtain the longest sequence of an unknown region. In this work, *Hin*dIII and *Xba*I were selected for the digestion of the genomic DNA. Secondly, the cohesive end(s) and blunt end(s) were filled in and added an adenine nucleotide (A) at 3′-ends using *Ex Taq* DNA polymerase at 72°C for 10 min. The third step was to ligate the second step's products into pUCm-T

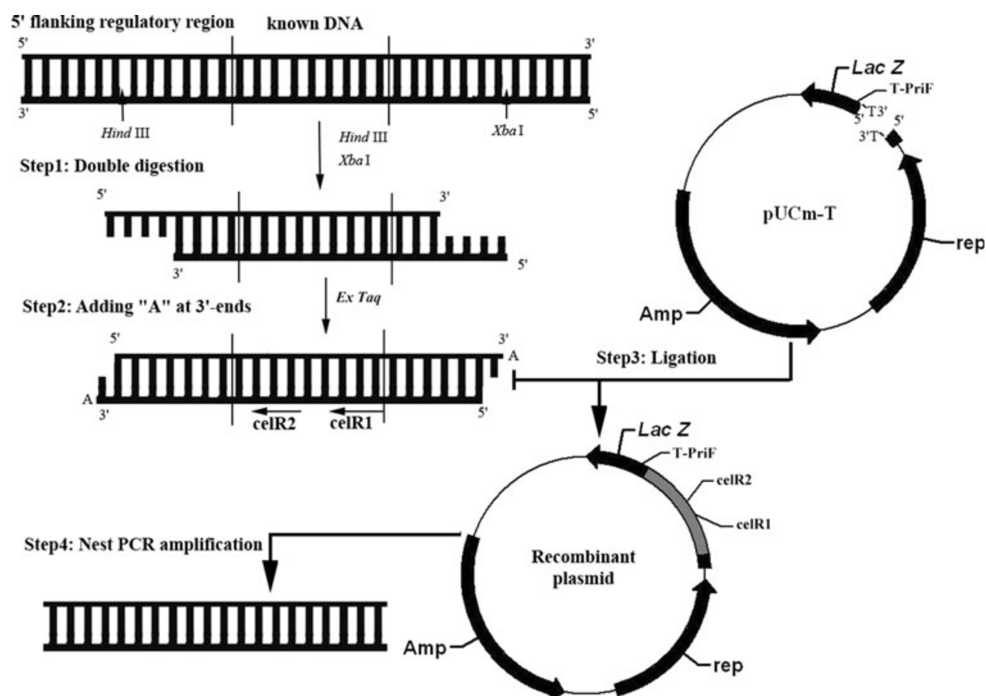

Fig. 1 Flowchart of the pUCm-T vector-mediated PCR and nested PCR amplification methods (exemplified as the cloning of the 5′ flanking region of the gene *Aucel12A*)

vector, and finally, the recombinant vectors were first amplified with T-PriF and celR1 to obtain the 5′ flanking regulatory region, and then subjected to the second-round PCR with T-PriF and celR2 for confirmation. The target PCR product was purified with EZ-10 Spin Column DNA Gel Extraction Kit (BBI, Markham, Canada), cloned into pUCm-T, and transformed to *E. coli* JM109, followed by restriction enzyme analysis and DNA sequencing.

Cloning of the complete DNA of the gene *Aucel12A*

The 5′ flanking regulatory region of the complete DNA of the gene *Aucel12A* was amplified by using the pUCm-T vector-mediated PCR method as mentioned above. While the core region of the gene *Aucel12A* was directly obtained from the genomic DNA of *A. usamii* E001 by conventional PCR with primers celF3 and celR. The complete DNA sequence was either obtained by assembling the above cloned 5′ flanking regulatory region, core region and 3′-end cDNA fragment sequences, or directly amplified by conventional PCR using the genomic DNA as template.

Analysis of amino acid and DNA sequences

The SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) was applied to predict the signal peptide sequence of the AuCel12A. Putative *N*-glycosylation site was located by using the NetNGlyc program 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/). Physicochemical properties of the AuCel12A were identified by using the Protparam (http://au.expasy.org/tools/protparam.html). Homology alignment of the primary structures between the AuCel12A and other GH family 12 endoglucanases was carried out in GenBank using the BLAST program. The ORF was determined by the program of NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The GeneMark (http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi) was used for the exact localization of the exon/intron boundaries. The prediction of the promoter region and its characterization were carried out using the Berkeley Drosophila Genome Project (http://www.fruitfly.org/seq_tools/promoter.html), together with the PLACE (http://www.dna.affrc.go.jp/PLACE/signalscan.html).

Construction of the recombinant expression plasmid

With the information of the cloned full-length cDNA sequence (GenBank accession: JN128871) and the *N*-terminal amino acid sequence alignment of the GH family 12 fungal endoglucanases, a pair of specific primers celF and celR were designed and used to amplify the cDNA fragment encoding the mature AuCel12A as following conditions: an initial denaturation at 94°C for 2 min; 30 cycles of at 94°C for 30 s, 55°C for 30 s, 72°C for 45 s; an extra elongation at 72°C for 10 min. After the target PCR product was agarose gel-purified and digested with *Eco*RI and *Not* I, the resulting cDNA fragment was inserted into pPIC9 K vector digested with the same restriction enzymes, followed by transforming into *E. coli* DH5α. The recombinant expression vector containing the correct insert, named as pPIC9 K-*Aucel12A*, was confirmed by restriction enzyme analysis and DNA sequencing.

Transformation and expression of the gene *Aucel12A*

Pichia pastoris GS115 cells were transformed with pPIC9 K-*Aucel12A* that was linearized with *Sal*I using electroporation on a Gene Pulser apparatus (Bio-Rad, Hercules, CA,USA) according to the manufacturer's instruction. The *P. pastoris* transformants were initially selected on MD plate, and then cultured on YPD plates containing geneticin G418 at increasing concentrations of 0.5, 1.0, and 2.0 mg/ml for the screening of multiple copies of the gene *Aucel12A*. The integration of the gene *Aucel12A* into the genome of *P. pastoris* was confirmed by PCR analysis using 5′- and 3′-*AOX1* primers, and yeast genomic DNA as template extracted by Yeast Genomic DNA Extract Kit (Tianwei, Beijing, China). Expression of the gene *Aucel12A* in *P. pastoris* GS115 was performed according to the instruction of Multi-Copy Pichia Expression Kit (Invitrogen) with slight modification. Each single colony of transformants was inoculated into 25 ml BMGY medium in a 250-ml flask, and grown at 30°C on a rotary incubator (220 rpm) until the OD_{600} reached 2–4. The cells were then harvested by centrifugation at 3,000 rpm and resuspended in 30 ml BMMY medium (pH 5.0) in a 250-ml flask. The rAuCel12A expression was induced by adding methanol to a final concentration of 2.0% (v/v) at 24-h intervals for 96 h. One transformant expressing the highest rAuCel12A activity was preserved and used for further studies.

Purification of the native AuCel12A and rAuCel12A

The cultured koji (10 g) of *A. usamii* E001 by solid-state fermentation was extracted with 10 volumes (w/v) of 20 mM Na_2HPO_4–NaH_2PO_4 buffer (pH 7.0) at 30°C for 30 min with shaking at 100 rpm. The extracted crude solution was brought to 70% saturation by adding solid ammonium sulfate. The precipitate was collected by centrifugation and then dissolved in 10 ml of the same phosphate buffer. Subsequent manipulations were performed according to the method reported previously [12].

After the *P. pastoris* GSCel2-1 was induced by methanol at 30°C for 96 h, the cultured broth was centrifuged at 10,000 rpm for 10 min to remove yeast cells. Twenty milliliters of resulting supernatant was brought to 70% saturation by adding solid ammonium sulfate. The precipitate was collected by centrifugation, dissolved in 4.0 ml of 20 mM Na_2HPO_4–NaH_2PO_4 buffer (pH 7.0), and dialyzed against the same phosphate buffer overnight. The dialyzed solution was concentrated to 1.0 ml by ultrafiltration at 8,000 rpm for 20 min at 4°C using a 10-kDa cut-off membrane (Millipore, Billerica, MA, USA), then applied onto a Sephadex G-75 column, followed by eluting with the same buffer at a flow rate of 0.4 ml/min. Aliquots of 2.0 ml eluent having the r-AuMan5A activity were pooled and concentrated by ultrafiltration as above.

Carbohydrate content assay and deglycosylation

The carbohydrate content of the purified native AuCel12A or rAuCel12A was assayed by the phenol sulfuric acid method [7], using D-mannose as standard. The deglycosylation was carried out by denaturing the glucoprotein at 100°C for 10 min, and then endoglycosidase H (Endo H, New England Biolabs) catalyzing the release of *N*-linked oligosaccharides was added to perform deglycosylation at 37°C for 1 h. All manipulation followed the manufacture's instructions.

rAuCel12A activity and protein assays

After the GSCel2-1 was induced by methanol at 30°C for 96 h, the cultured broth was centrifuged at 10,000 rpm for 10 min. The resulting supernatant was used for enzyme activity and protein assays. The rAuCel12A activity was assayed using the previously described method [13] with appropriate modification. In brief, 100 μl of suitably diluted enzyme was incubated with 2.4 ml of 0.5% (w/v) CMC-Na, prepared with 50 mM Na_2HPO_4–citric acid buffer (pH 5.0), at 50°C for 20 min. The reaction was terminated and color-developed by adding 2.5 ml of 3,5-dinitrosalicylic acid (DNS) reagent and boiling at 100°C for 7 min. The absorbance of colored solution was measured at 540 nm using a UV–Vis spectrophotometer (Jinghua, Shanghai, China). One unit (U) of the rAuCel12A activity was defined as the amount of enzyme liberating 1 μmol of reducing sugar equivalents per min under the assay conditions, using D-glucose as standard. Protein concentration was determined by the BCA-200 Protein Assay Kit (Pierce, Rockford, IL), using bovine serum albumin as standard. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was performed on a 12.5% gel by the method of Laemmli [10], and isolated proteins were visualized by straining with Coomassie Brilliant Blue R-250.

Characterization of the rAuCel12A

The optimum pH of the purified rAuCel12A was assayed by the standard enzyme activity assay method as described above, but 0.5% of substrate (CMC-Na) solutions, which were prepared with 20 mM Na_2HPO_4-citric acid buffer (pH 3.0–7.5) and Tris–HCl buffer (pH 8.0), respectively. To estimate the pH stability, aliquots of enzyme solution were preincubated at 40°C for 1.0 h in the same buffers over the pH range of 3.0–8.0, respectively, and the residual enzyme activity was assayed under standard conditions.

The pH stability in this work was defined as a pH range, over which the residual enzyme activity was remained over 85% of the original activity.

The optimum temperature of the purified rAuCel12A was determined at optimum pH and at temperatures ranging from 35 to 70°C, respectively. To evaluate thermal stability, aliquots of enzyme solution were mixed with Na_2HPO_4-citric acid buffer (pH 5.0) at a ratio of 1:1 (v/v), and then preincubated at various temperatures (35–70°C) for 1.0 h, respectively. The residual enzyme activity was assayed under standard conditions. The thermal stability was defined as a temperature, at and below which the residual enzyme activity was more than 85% of the original activity.

To estimate resistance of the rAuCel12A to metal ions and EDTA, aliquots of enzyme solution were preincubated, respectively, with various metal ions and EDTA at a final concentration of 5.0 mM in 20 mM Na_2HPO_4-citric acid buffer (pH 5.0) at 40°C for 1.0 h. The residual enzyme activity was assayed under standard conditions. The enzyme solution without any additive was used as the control.

Hydrolyzing reaction rate (U/ml) of the rAuCel12A was assayed under the conditions (pH 5.0, 50°C for 10 min) except the substrate (CMC-Na) concentrations ranging from 1.0 to 10 mg/ml. The hydrolyzing reaction rate versus substrate concentration was plotted to verify whether the hydrolyzing mode of the rAuCel12A conforms to Michaelis–Menten equation. Kinetic parameters, K_m and V_{max}, were graphically determined from the Lineweaver–Burk plotting.

Results and discussion

Cloning of the cDNA sequence encoding the AuCel12A

Analytical results of the extracted total RNA showed that the ratio of OD_{260} to OD_{280} was 1.96, and the 18S and 28S rRNA bands, characterized as eukaryotes, on formaldehyde denatured agarose gel electrophoresis were specific, indicating that the total RNA has high purity and is not decomposed. An approximate 780-bp band of the 3′-end fragment of AuCel12A cDNA was first amplified with primers celF1 and PR, and subjected to the second-round PCR with celF2 and PR. DNA sequencing result verified that the amplified 3′-end cDNA is 745 bp in length (except primers dT-PR and PR). An about 450-bp band of the 5′-end fragment of AuCel12A cDNA was first amplified with primers OP and celR1, and subjected to a second round of PCR with IP and celR2. DNA sequencing result showed that the first-round PCR product is 404 bp in length (except primers OP and IP), containing a 122 bp of sequence identical to that between celF1 and celR1, and a new 282 bp of sequence in which a starting point of transcription, and a cDNA fragment encoding a 16-aa signal peptide and a *N*-terminal 48-aa mature peptide were recognized (Fig. 2). The full-length cDNA sequence was obtained by assembling two cloned 3′- and 5′-end cDNA fragments. It is 1,027 bp in length (except polyA), containing a 90-bp 5′ non-coding region, a 217-bp 3′ non-coding region, and a 720-bp ORF encoding a 239-aa preAuCel12A. A 672-bp cDNA fragment was amplified with a pair of specific primers celF and celR, which encodes a 223-aa mature AuCel12A (Fig. 2).

Primary structure analysis of the AuCel12A

The SignalP 3.0 predicted an unambiguous signal peptide cleavage site between Gly^{16} and Gln^{17}, which was in agreement with the result of *N*-terminal amino acid sequence alignment of the GH family 12 fungal endoglucanases. So, a 239-aa preAuCel12A was predicted to contain a 16-aa signal peptide and a 223-aa mature peptide (Fig. 2). A theoretical M.W. of 24,294 Da and a p*I* of 4.15 are calculated from the deduced AuCel12A sequence, in which there is one putative *N*-glycosylation site (N–W–T). Amino acid homology alignment showed that the identities of the mature AuCel12A (AEL12376) with other four endoglucanases of *A. kawachii* (BAA02297), *A. terreus* (XP_001214697), *A. aculeatus* (CAA36757), and *T. reesei* (BAA20140) from GH family 12 was 87.0, 82.0, 64.3, and 50.2%, respectively (Fig. 3). Three putative catalytic amino acid residues, Asp^{99}, Glu^{116} and Glu^{204} located, respectively, in the β5, β6 and β4 strands of the mature AuCel12A, were found to be highly conserved among GH family 12 endoglucanases [19, 20]. Besides a fragment W(T/N)WSGGE(G/N) (T/S)VKSY(S/A)NSG, the other two highly conserved fragments, YDLFTAAN and SGDYELM in all reported fungal endoglucanases of the GH family 12 [5, 17], were also recognized in the AuCel12A. These features verified that the AuCel12A is a member of the GH family 12.

Cloning of the complete DNA of the gene *Aucel12A*

An approximate 910-bp 5′ flanking regulatory region of the gene *Aucel12A* was first amplified with primers T-PriF and celR1, and subjected to a second round of PCR with T-PriF and celR2 for confirmation. DNA sequencing result showed that the amplified 5′ region is 837 bp in length (except the sequence from the primer T-PriF to T/A clone site on pUCm-T vector). An about 840-bp core region of the gene *Aucel12A* was amplified by conventional PCR with primers celF3 and celR. The complete DNA sequence

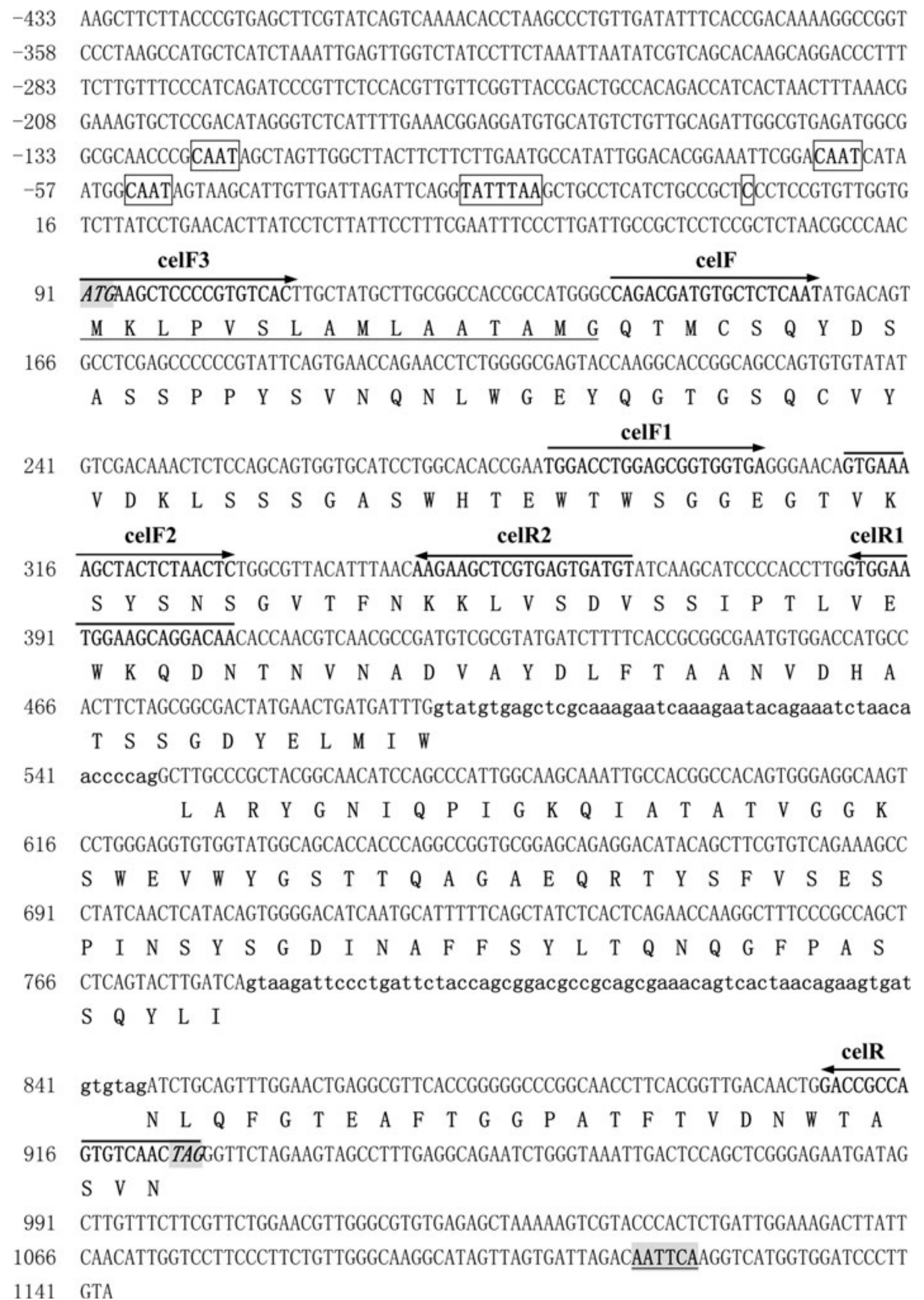

Fig. 2 Nucleotide sequence of the AuCel12A cDNA or DNA gene from *Aspergillus usamii* E001 and its deduced amino acid sequence of the AuCel12A. Nucleotide sequences of two introns with sizes of 50 and 66 bp are shown in *lowercase letters*. The signal peptide from Met[1] to Gly[16] is *underlined*. The *bold letters* of C, TATTTAA and CAAT in *boxes* indicate the starting point of transcription, the putative TATA box and CAAT box, respectively. The *grayed italic letters* of ATG and TAG represent the starting codon and stop codon, respectively. The putative polyadenylation signal, AATTCA, is shown as *grayed underlined letters*. The *bold arrows* above the letters represent the primers for PCR amplification

was obtained by assembling the above three cloned fragments: 5′ region, core region, and 3′-end cDNA. It is 1,576 bp in length (Fig. 2).

DNA sequence characterization of the gene *Aucel12A*

Compared with the full-length cDNA sequence, the complete DNA sequence of the gene *Aucel12A* is composed of 433 bp of 5′ flanking regulatory region, 90 bp of 5′ non-coding region, two introns with sizes of 50 and 66 bp, respectively, 720 bp of ORF, and 217 bp of 3′ non-coding region. Both of the exon/intron boundaries conform to the canonical GT-AG rule. It was predicted that the promoter region of the gene *Aucel12A* is located at the range from −38 to +10 bp, designating the starting point of transcription(C) as +1 bp. A TATTTAA sequence in the gene *Aucel12A* that functions as a substitute for the classical TATA box (TATAAA) which is located –25 bp upstream the transcription starting point, which was in agreement with the consensus distance generally found in eukaryote promoters. Other consensus sequences such as TTATTT could also act as substitutes for the classical TATA box [11, 16]. In eukaryotes, the functional CAAT box is typically found about −75 bp upstream from the transcription

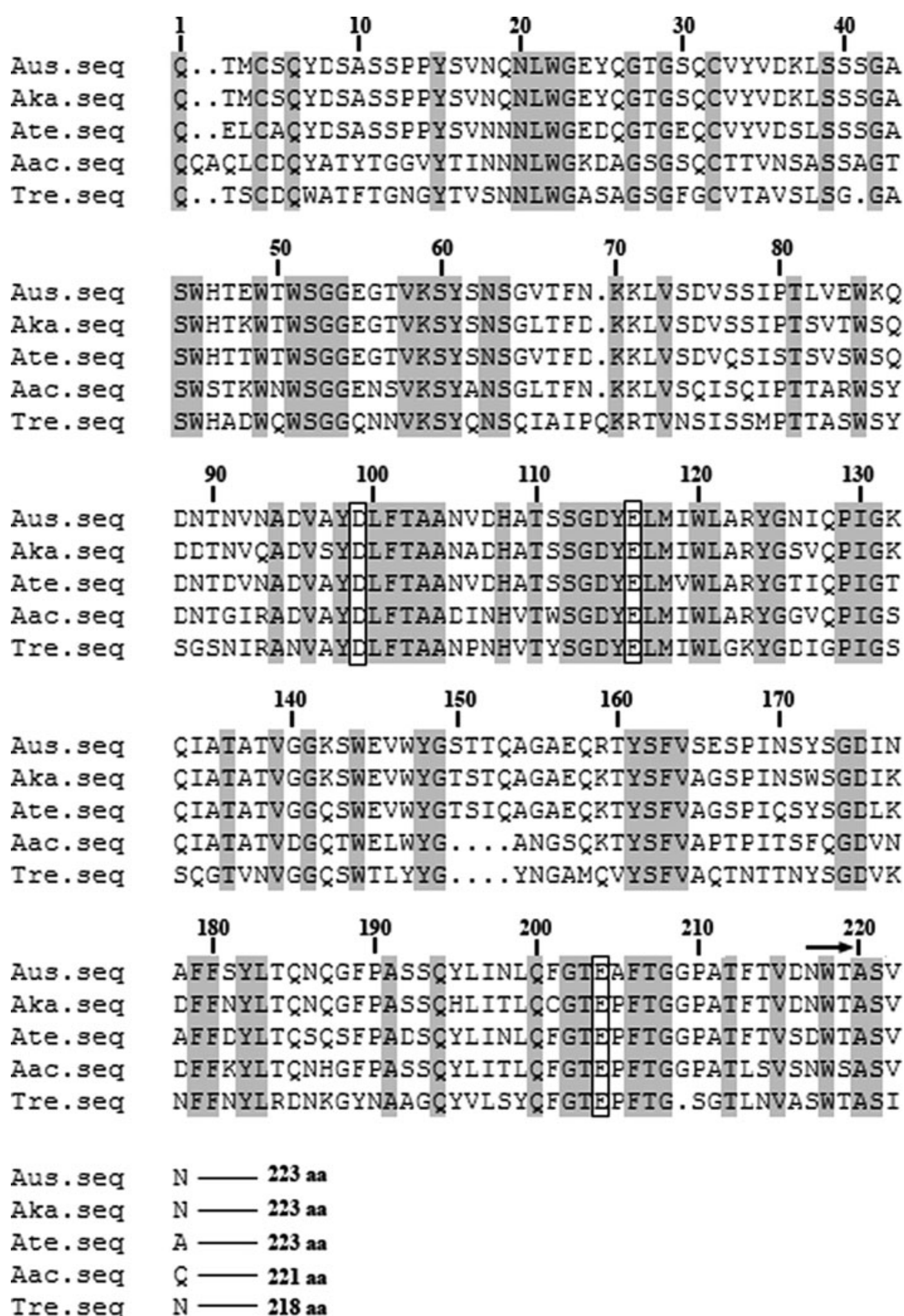

Fig. 3 Amino acid homology alignment of the AuCel12A with the other four fungal endoglucanases from the GH family 12 based on the primary structural features determined with the BLAST program in GenBank. Abbreviations: *Aus Aspergillus usamii* (AEL12376), *Aka Aspergillus kawachii* (BAA02297), *Ate Aspergillus terreus* (XP_001214697), *Aac Aspergillus aculeatus* (CAA36757), and *Tre Trichoderma reesei* (BAA20140). The identical amino acid residues in five endoglucanases are marked with a *grey background*. Three catalytic residues, strictly conserved among the GH family 12 members, are *boxed*. One putative *N*-glycosylation site is marked with a *bold arrow*

starting point. Some CAAT boxes may be located further from the starting point [22]. In this work, we found three putative CAAT boxes at −55, −67 and −123-bp positions, respectively. It was also found that an AATTCA sequence as a putative polyadenylation signal is located at +1,116 bp downstream from the transcription starting point. Bioinformatics analysis indicated that the 5′ flanking regulatory region possesses several putative transcription factor binding sites. Two HSF (heat shock factor) binding sites were found at the positions of −209 and −78 bp, respectively; one GATA-2 (GATA-binding factor 2) binding site at −383 bp.

Screening and expression of *P. pastoris* transformants

Transformants that can resist higher concentrations of geneticin G418 might have multi-copies of integration into the *P. pastoris* genome. This could potentially lead to higher level expression of heterologous proteins [4]. Therefore, we respectively chose ten transformants resistant to 0.5, 1.0, and 2.0 mg/ml of G418, numbered as *P. pastoris* GSCel0-1 to 0–10, GSCel1-1 to 1–10 and GSCel2-1 to 2–10, for shake flask expression. The *P. pastoris* GS115 transformed with pPIC9 K was used as negative control (named as *P. pastoris* GSC). The culture

Table 2 Screening and expression of *P. pastoris* GS115 transformants

G418 (mg/ml)	Representative strains	rAuCel12A activity (U/ml)
0.5	GSCel0-3	140.2 ± 6.6
	GSCel0-5	188.7 ± 6.2
	GSCel0-8	130.0 ± 6.7
1.0	GSCel1-2	192.5 ± 7.0
	GSCel1-3	136.4 ± 6.3
	GSCel1-8	215.3 ± 8.5
2.0	GSCel2-1	240.8 ± 8.0
	GSCel2-7	180.1 ± 7.5
	GSCel2-9	226.9 ± 7.4

All values are means ± SD from three independent experiments

supernatants were tested for rAuCel12A activity using the DNS assay method. The enzyme activity of the negative control could not be detected. When cultured at 30°C in BMMY with initial pH 5.0 and induced by adding 2.0% of methanol at 24-h intervals for 96 h, one transformant, numbered *P. pastoris* GSCel2-1, was obtained with the highest rAuCel12A activity of 240.8 U/ml (Table 2).

Verification of *P. pastoris* transformants

To verify whether the gene *Aucel12A* is integrated into the yeast genome, the genomic DNAs of three transformants, GSC, GSCel1-3, and GSCel2-1, were extracted and analyzed by PCR amplification using 5′- and 3′-*AOX1* primers, respectively. A complete *AOX1* gene (2.2 kb) and a partial *AOX1* gene (0.49 kb) were PCR amplified from GSC, while a complete *AOX1* gene along with a 1.2-kb fragment containing the gene *Aucel12A* (672 bp) was amplified from *P. pastoris* GSCel1-3 or GSCel2-1. These results demonstrated that in positive transformants, the gene *Aucel12A* encoding the mature AuCel12A had successfully integrated into the *P. pastoris* GS115 genome. The GSCel2-1 has good genetic stability to express the rAuCel12A, as demonstrated by genomic PCR and enzyme activity assays (data not shown).

Carbohydrate content assay and deglycosylation

SDS-PAGE analysis of the purified native AuCel12A and rAuCel12A showed a protein band of about 27.0 kDa (Fig. 4, lane 1–2), respectively, which is larger than the theoretical M.W. (24.3 kDa) of the mature AuCel12A consisted of 223 amino acids. To verify whether this difference is due to glycosylation, deglycosylation analysis was carried out. After treatment with Endo H to remove carbohydrate moieties, only a single protein band located at about 24.8 kDa was observed (Fig. 4, lane 3–4). Furthermore, carbohydrate contents of the purified native AuCel12A and rAuCel12A were assayed to be 5.25 and 4.82%, respectively. The slight difference may result from the different eukaryotic expression system, *A. usamii* and *P. pastoris*. The results indicated that the native AuCel12A and rAuCel12A are *N*-glycosylated proteins in agreement with the fact that there is one putative *N*-glycosylation site (N–W–T) in the deduced AuCel12A sequence.

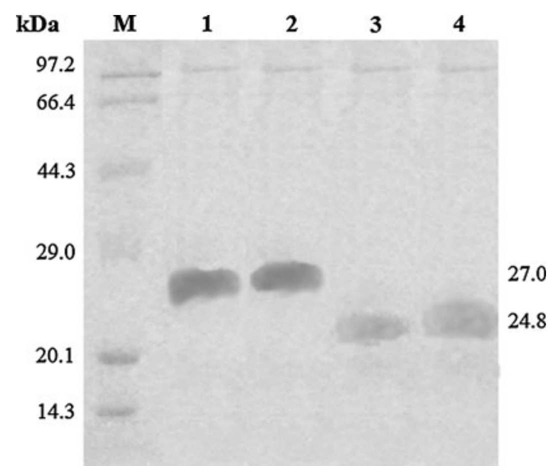

Fig. 4 SDS-PAGE analysis of the native AuCel12A purified from *A. usamii* E001 and the rAuCel12A secreted by the GSCel2-1, as well as the deglycoylated native AuCel12A and rAuCel12A with Endo H. *Lanes*: *M* protein marker (TaKaRa), *1* the purified native AuCel12A, *2* the purified rAuCel12A, *3* the deglycoylated native AuCel12A, *4* the deglycoylated rAuCel12A

Enzymatic properties of the rAuCel12A

The rAuCel12A showed higher enzyme activity at a pH range of 4.5–5.5, among which the highest enzyme activity was at pH 5.0 (measured at 50°C), and was highly stable at a pH range of 3.5–7.0 (Fig. 5a). The optimum temperature for the rAuCel12A activity, at pH 5.0, was 60°C. It remained over 85% of the original activity at a temperature of 55°C or below, but only retained 32.5% of the original activity at 70°C (Fig. 5b). The rAuCel12A activity was not significantly affected by Ca^{2+}, Co^{2+}, Cu^{2+}, Mg^{2+}, Al^{3+}, Fe^{3+}, Zn^{2+}, Ba^{2+}, Pb^{2+}, Mn^{2+} and EDTA, but inhibited by Ag^{+}, Hg^{2+} and Fe^{2+} with 36.8, 23.7, and 55.0% of the original activity, respectively. The enzymatic properties of the rAuCel12A characterized above are in good agreement with those of the native AuCel12A purified from *A. usamii* E001. The K_m and V_{max} values of the rAuCel12A, towards CMC-Na at pH 5.0 and 50°C, were graphically determined from the Lineweaver–Burk plotting to be 4.85 mg/ml and 160.5 U/mg, respectively. The K_m value of the rAuCel12A is much smaller than that (17.1 mg/ml) of the *Penicillium canescens* endoglucanase, while the V_{max} value is much

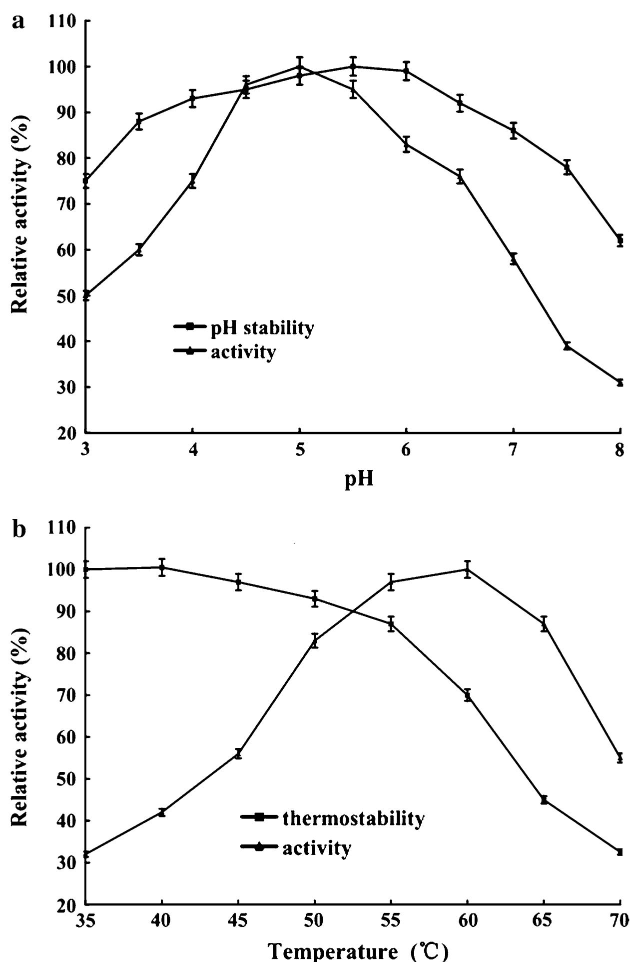

Fig. 5 Effect of temperature and pH value on the r-AuMan5A activity. **a** pH optimum and stability. The optimal pH of the rAuCel12A was measured by the DNS method using Na_2HPO_4–citric acid buffer (pH 3.0–7.5) and Tris–HCl buffer (pH 8.0). The pH stability of the rAuCel12A was determined by incubating it in the same buffers at 40°C for 1.0 h, respectively, then the residual enzyme activities were measured using the DNS method. **b** Temperature optimum and stability. The optimal temperature of the rAuCel12A was determined by using the DNS method except the reaction temperatures ranging from 35 to 70°C. The thermostability of the rAuCel12A was determined by incubating it at various temperatures (35–70°C) for 1.0 h, then the residual enzyme activities were measured using the DNS method

larger than that (18.6 U/mg) of the *P. canescens* endoglucanase [5]. The results demonstrated that the appetency for substrate and the catalytic efficiency of the rAuCel12A are significantly higher than those of the *P. canescens* endoglucanase. This superiority would make the rAuCel12A widely used in many fields.

Acknowledgments This work was supported by a grant from the National Nature Science Foundation of China (No. 20776061). We are grateful to Prof. Xianzhang Wu (School of Biotechnology, Jiangnan University) for providing technical assistance.

References

1. Adsul MG, Bastawde MG, Varma AJ, Gokhale DV (2007) Strain improvement of *Penicillium janthinellum* NCIM 1171 for increased cellulase production. Bioresour Technol 98:1467–1473
2. Bhat MK (2000) Cellulases and related enzymes in biotechnology. Biotechnol Adv 18:355–383
3. Castro AM, Caryalho MLA, Leite SGF, Pereira N Jr (2010) Cellulases from *Penicillium funiculosum*: production, properties and application to cellulose hydrolysis. J Ind Microbiol Biotechnol 37:151–158
4. Chen XL, Cao YH, Ding YH, Lu WQ, Li D (2007) Cloning, functional expression and characterization of *Aspergillus sulphureus* β-mannanase in *Pichia pastoris*. J Biotechnol 128:452–461
5. Chulkin AM, Loginov DS, Vavilova EA, Abyanova AR, Zorov IN, Kurzeev SA, Koroleva OV, Benevolensky SV (2009) Cloning of the *Penicillium canescens* endo-1, 4-β-glucanase gene egl3 and the characterization of the recombinant enzyme. Appl Biochem Miocrobiol 45:143–149
6. Ding SJ, Ge W, Buswell JA (2002) Secretion purification and characterization of a recombinant *Volvariella volvacea* endoglucanase expressed in the yeast *Pichia pastoris*. Enzyme Microb Technol 31:621–626
7. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956) Colorimetric method for determination of sugars and related substances. Anal Chem 28:350–356
8. Hong J, Tamaki H, Akiba S, Yamamoto K, Kumagai H (2001) Cloning of a gene encoding a highly stable endo-β-1, 4-glucanase from *Aspergillus niger* and its expression in yeast. J Biosci Bioeng 92:434–441
9. Kang HJ, Uegaki K, Fukada H, Ishikawa K (2007) Improvement of the enzymatic activity of the hyperthermophilic cellulose from *Pyrococcus horikoshii*. Extremophiles 11:251–256
10. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685
11. Lampidonis AD, Stravopodis DJ, Voutsinas GE, Messini-Nikolaki N, Stefos GC, Margaritis LH, Argyrokastritis A, Bizelis I, Rogdakis E (2008) Cloning and functional characterization of the $5'$ regulatory region of ovine hormone sensitive lipase (*HSL*) gene. Gene 427:65–79
12. Li JF, Ma LP, Wu MC, Xia WS (2006) Purification and some properties of an acidic β-mannanase from *Aspergillus usamii*. Food Ferment Ind 32:5–9 (in Chinese)
13. Li JF, Tang CD, Shi HL, Wu MC (2011) Cloning and optimized expression of a neutral endoglucanase gene (*ncel5A*) from *Volvariella volvacea* WX32 in *Pichia pastoris*. J Biosci Bioeng 111:537–540
14. Mach RL, Zeilinger S (2003) Regulation of gene expression in industrial fungi: *Trichoderma*. Appl Microbiol Biotechnol 60: 515–522
15. Miettinen-Oinonen A, Londesborough J, Joutsjoki V, Lantto R, Vehmaanpera J, Biotecd PL (2004) Three cellulases from *Melanocarpus albomyces* for textile treatment at neutral pH. Enzyme Microb Technol 34:332–341
16. Reed DS, Gibson JB (1994) Molecular heterogeneity of naturally occurring *sn*-glycerol-3-phosphate dehydrogenase low-activity variants in *Drosophila melanogaster*. Biochem Genet 32: 161–179
17. Sakamoto S, Tamura G, Tto K, Ishikawa T, Iwano K, Nishiya N (1995) Cloning and sequencing of cellulase cDNA from *Aspergillus kawachii* and its expression in *Saccharomyces cerevisiae*. Curr Genet 27:435–439
18. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York

19. Sandgren M, Gualfetti PJ, Shaw A, Gross LS, Saldajeno M, Day AG, Jones TA, Mitchinson C (2003) Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. Protein Sci 12:848–860
20. Sandgren M, Shaw A, Ropp TH, Wu S, Bott R, Cameron AD, Stahlberg J, Mitchinson C, Jones TA (2001) The X-ray crystal structure of the *Trichoderma reesei* family 12 endoglucanase 3, Cel12A, at 1.9 Å resolution. J Mol Biol 308:295–310
21. Wang JQ, Zhang HM, Wu MC, Tang CD (2011) Cloning and sequence analysis of a novel xylanase gene, *Auxyn10A*, from *Aspergillus usamii*. Biotechnol Lett 33:1029–1038
22. Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. Plant Cell 18:2971–2984
23. Wu MC, Wang JQ, Zhang HM, Tang CD, Gao JH, Tan ZB (2011) Cloning and sequence analysis of an acidophilic xylanase (XynI) gene from *Aspergillus usamii* E001. World J Microbiol Biotechnol 27:831–839