Front. Agric. China 2011, 5(1): 72–76
DOI 10.1007/s11703-011-1060-z

RESEARCH ARTICLE

Construction of the expression vector and location analysis of thermotolerant endoglucanase in *E. coli*

Runfang GUO, Kexue GAO, Hongwei YU, Yingmin JIA (✉)

College of Food Science and Technology, Agricultural University of Hebei, Baoding 071001, China

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2011

Abstract **To obtain the secreting expression vector, the signal peptide sequence and mature peptide sequence of endoglucanase from *Streptomyces xylophagus* KX6 were cloned into the pET28a plasmid. The recombinant vector pET28a/KX6 was transformed *Escherichia coli* Rosetta (DE3), and the transformant was induced by IPTG. The expression products were primarily distributed in the medium fluid of host cell in a soluble form and the activity was higher than that of other fractions. Both location analysis of targeting protein and activity analysis showed that the signal peptide of endoglucanase from *S. xylophagus* KX6 had played a very important role in the secret expression and activity of foreign proteins in the *E. coli* host cell.**

Keywords thermotolerant endoglucanase, expression vector, signal peptide, location analysis

Introduction

The endocellulase (EC 3.2.1.4), the major component of the cellulolytic enzyme system, first attacks on the cellulose polymer by breaking down the internal β-1,4 glucosidic linkages within amorphous regions of cellulose chains (Wulff et al., 2003). Therefore, it plays an important role in degradation of cellulosic biomass (Solingen et al., 2001). A number of cellulase genes from bacteria and fungus have been cloned and also expressed in yeast or *Escherichia coli* (Qiao et al., 2004; Kim and Lee, 2008). However, no information on endoglucanase gene from *Streptomyces xylophagus* was found.

Pure, soluble, and functional proteins are of high demand in modern biotechnology. Natural protein sources rarely meet the requirements for quantity and ease of isolation or price; hence, recombinant technology is often the method of choice. Recombinant cell factories are constantly employed for the production of protein preparations for downstream purification and processing (Sørensen and Mortensen, 2005). *E. coli* is a frequently used host, but expression of recombinant proteins with *E. coli* as the host often results in insoluble and/or nonfunctional proteins (Huo et al., 2008).

Here, we attempted a new approach to overcome these obstacles above by a strategy that focuses on controlled expression of target protein in a soluble form through target signal sequence. The main purpose of our study was to obtain a high degree of accumulation of soluble product in the bacterial cell cultures for simplified purification and large-scale production and thus to provide a basis for understanding the secreting mechanism of recombinant proteins in *E. coli*.

Received February 20, 2010; accepted May 3, 2010
Correspondence: Yingmin JIA
E-mail: ymjia@mail.hebau.edu.cn

Materials and methods

Strains and plasmids

The vectors pET-28a and pMD18-T-cel (containing the target gene *cel*) and host strains *E. coli* JM109 and Rosetta (DE3) were conserved in the Enzyme Engineering Laboratory of Agriculture University of Hebei, Baoding, China.

Construction of recombinant expression vectors

The pair of specific primers was designed according to the cellulase gene sequences from *S. xylophagus* KX6 (FJ441063): P1: 5′-TCGAATTCGATGCCACGGTTAC-3′ (underlined is the site of *Eco*RI) and P2: 5′-TTGCGGCCGCTCACGCGGTGCT-3′ (underlined is the

site of *Not*I).

The full-length DNA sequence of endoglucanase gene *cel* was obtained by PCR amplification using pMD18-T-cel as a template. PCR amplification was done using a Thermal Cycler 100 (USA) in 50 μL reaction volumes for 30 cycles of 94°C for 1 min, 63°C for 1 min, and 72°C for 1 min. The PCR product of about 1200 bp was purified by agarose gel electrophoresis. The PCR products and the expression vector pET-28a were digested by the *Eco*RI/*Not*I, respectively, and then ligated *in vitro* to construct the recombinant plasmid pET-28a/KX6. To verify the gene *cel* and its signal sequence under the transcriptional control of the T7 promoter, the recombinant plasmid pET-28a/KX6 was sequenced. After the construct was verified, plasmid pET-28a/KX6 was transformed into an expression host for protein production.

Induced expression of recombinant proteins

The recombinant vector pET-28a/KX6 was transformed *E. coli* Rosetta (DE3) competent cells and the transformants were screened at LB plate containing kanamycin. The different resistance transformants were picked and inoculated into 10 mL NZYCM medium and grown at 30°C in a shaking incubator (250 $r \cdot min^{-1}$) for 12 to 16 h. Then, 1% inoculation was conducted into 100 mL NZYCM medium. When the culture grew at 30°C and reached an OD600 = 0.4–0.6, IPTG was added to a final concentration of 1 $mmol \cdot L^{-1}$ to induce expression. The culture and cell pellet from the different locations of the cell were used to analyze the cellulase activity and solubility.

SDS-PAGE of the TCP

The expression of target genes may be assessed by analysis of total cell protein (TCP) on a SDS-polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie blue staining. SDS-PAGE was performed in gel containing 12% (w/v) acrylamide and 0.1% SDS (w/v) using a Tris/glycine buffer system. The protein was stained with 0.1% Coomassie brilliant blue R250 in 7% (v/v) acetic acid in 50% (v/v) methanol solution. Destaining was carried out with 7% acetic acid in 50% methanol.

Location of endoglucanase in the *E. coli* cells

The location of target protein was performed according to the Manual to pET system. The extracellular fluid from the *E. coli* engineering strain was extracted by centrifugation at 10000×g for analyzing the solubility and activity. The cell pellet was saved and allowed to drain. After removing as much liquid as possible, the pellet was resuspended thoroughly in 30 mL of 30 $mmol \cdot L^{-1}$ Tris-HCl at pH 8.0 and 20% sucrose followed by addition of 60 L of 0.5 $mol \cdot L^{-1}$ EDTA at pH 8.0 into the suspension to reach the final concentration of 1 $mmol \cdot L^{-1}$, which was stirred slowly using a magnetic stirring bar at room temperature for 10 min. The cells were collected by centrifugation at 10000×g at 4°C for 10 min. After the pellet was thoroughly resuspended in 30 mL of ice-cold 5 $mmol \cdot L^{-1}$ $MgSO_4$, the cell suspension was stirred slowly for 10 min on ice. During this step, the periplasmic proteins were released into a buffer and the cell pellet was ice-stored for further processing of the soluble and insoluble cytoplasmic fractions. Through adding 1 L (25 units) lysozyme per 1 mL BugBuster reagent used for resuspension, rotating mixer at a slow setting for 20 min at room temperature, and centrifuging the mixture at 16000×g for 20 min, the supernatant was gained and placed into a fresh tube for analysis. The pellet may be used to isolate the insoluble cytoplasmic fraction. The final pellet of inclusion bodies was compatible with resuspension in denaturing buffer.

Determination of endoglucanase from various fractions

Enzyme activity was assayed with CM-cellulose as a substrate by measuring the amount of reducing sugar in various fractions according to the dinitrosalicylic acid method of Miller (1959). The reaction mixture contained 0.8 mL of 1% CM-cellulose solution of 200 $mmol \cdot L^{-1}$ citric acid-Na_2HPO_4 buffer at pH 5.5 and 0.2 mL appropriately diluted enzyme solution, which was incubated at 50°C for 20 min, and the reaction was stopped by addition of 0.5 mL DNS reagent. The reducing sugar generated was quantified at A540 by adding 3 mL water and using D-glucose as a standard for spectrophotometric estimation. One unit of endoglucanase activity was defined as the amount of enzyme that releases 1 μg reducing sugar per minute under the assay conditions. The cogon red–CMCNa plate (containing 1% CMCNa, 0.05% cogon red, and 2% agar) was also used to determine the endoglucanase activities of expressing proteins. By adding 50 μL sample into the hole (0.6 mm in diameter), which was incubated at 50°C for overnight, the enzyme activity was determined and described by the size of discoloring zone around the hole.

Results

The construction of recombinant vector

The target sequence about 1200 bp with signal sequence was cloned via PCR and then inserted into the pET-28a vector to construct the secreted recombinant expression vector pET28a/KX6. The recombinant plasmid DNA was isolated using standard preparation procedure and then digested with restriction enzymes *Eco*RI and *Not*I. The result showed two bands about 5.9 kb and 1.2 kb were found on agarose gel electrophoresis, respectively, which conforms to the size of prediction (Figs. 1 and 2). The sequencing result also suggested the insert sequence and open reading frame are fully accurate.

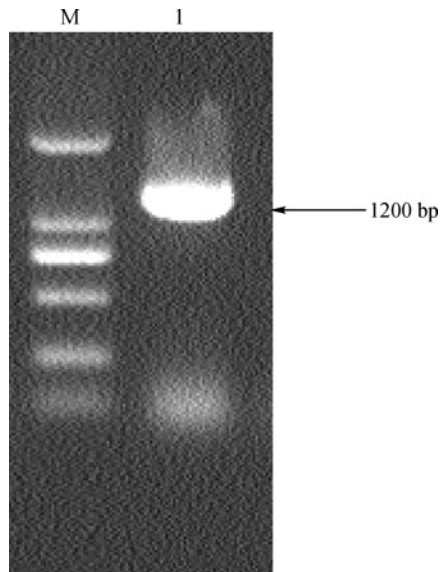

Figure 1 Agarose gel electrophoresis of PCR product
Note: M is the marker DL 2000 and 1 is PCR product.

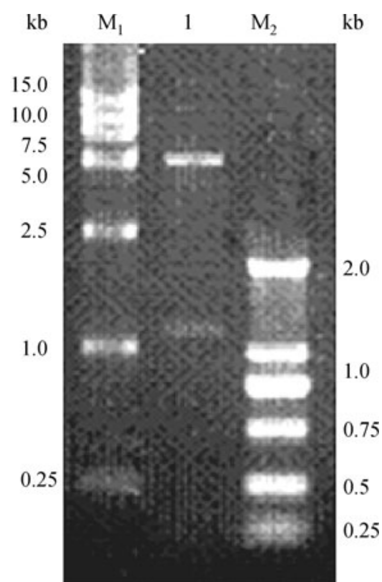

Figure 2 Identification of recombinant plasmid digested by *Eco*RI/*Not*I
Note: M_1 is the marker DL15000, 1 is recombinant plasmid pET-28a/KX6, and M_2 is the marker DL2000.

SDS-PAGE analysis of TCP

After a target plasmid was transformed into *E. coli* Rosetta (DE3), the expression of the target DNA was induced by the addition of IPTG to a growing culture. The induced cultures were examined for the production of the target protein. The expression of target gene was assessed quickly by analysis of TCP on SDS-PAGE followed by Coomassie blue staining. SDS-PAGE analysis of TCP could be seen in Fig. 3. From Lanes 3 to 9, the target protein was increasingly accumulated along with IPTG induction. The size of expression protein was about 43 KD, the same as putative molecular weight.

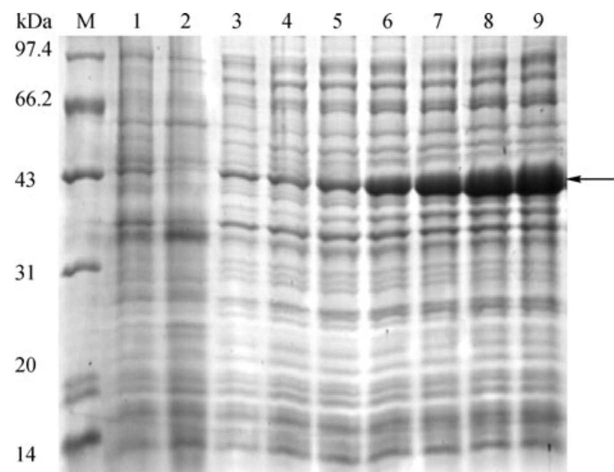

Figure 3 SDS-PAGE analysis of different induced time of pET-28a/KX6
Note: M represents the protein marker, 1 represents pET-28a vector, 2 represents pET-28a vector induced by IPTG for 6 h, and 3 to 9 represent recombinant vector pET-28a/KX6 induced by IPTG for 0 to 6 h, respectively.

Location of endoglucanase in *E. coli* cells

The target protein may be distributed to different locations of *E. coli* cells. A small-scale sample of the four fractions in the following medium, periplasm, soluble cytoplasm, and insoluble cytoplasm is analyzed by SDS-PAGE. The results showed the target product was mainly distributed the extracellular fluid as soluble form (Fig. 4, Lane 5) and cytoplasmic fraction as inclusion bodies (Fig. 4, Lane 2). However, a small part of the target protein also was found in the cytoplasm as soluble fractions (Fig. 4, Lane 3) and hardly determined target protein in periplasm (Fig. 4, Lane 4).

The distribution of the target protein suggested the expression product of target gene secreted into the medium fluid; moreover, expression levels right in the culture medium were higher than that in other ones, indicating its contribution to recovering and purifying the target products from the extracellular fluid. In addition, amounts of target proteins also aggregated in the cytoplasmic fraction known as inclusion bodies, and the formation of insoluble inclusion bodies suggested that the high rate of protein synthesis could not export the target protein or folded properly, which led to aggregation in the cytoplasm in a soluble form. Results of this analysis may lead to further optimization of the induction conditions and decreasing the rate of protein synthesis, such as low induction temperatures or growth in the minimal medium tended to increase the percentage of target protein in

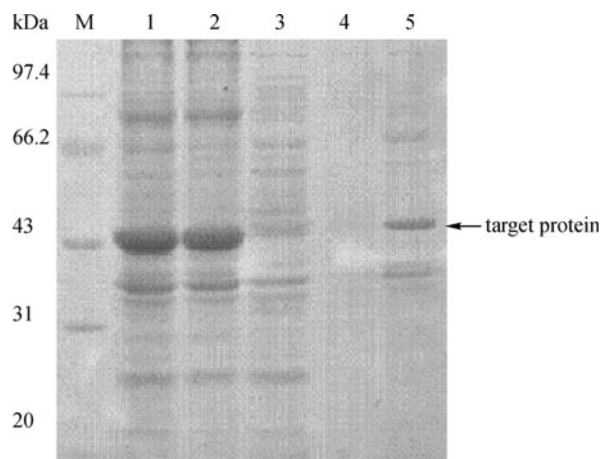

Figure 4 Localization analysis of recombinant plasmid pET-28a/KX6-expressing protein
Note: M represents the protein marker. Lanes 1 to 5 represent TCP fraction, insoluble cytoplasmic fraction, soluble cytoplasmic fraction, periplasmic fraction, and medium fraction, respectively.

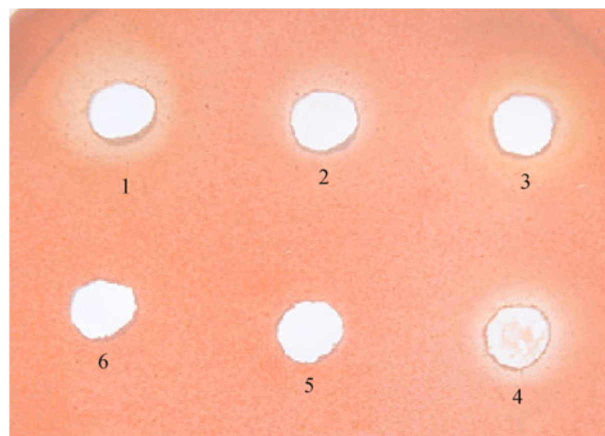

Figure 5 Cogon red–CMCNa plate determination
Note: 1 to 4 represent medium fraction, periplasmic fraction, soluble cytoplasmic fraction, and insoluble cytoplasmic fraction, respectively. 5 is blank and 6 is control.

soluble form. A TCP sample was also analyzed in parallel with the various fractions serving as a control for recovery of the target protein (Fig. 4, Lane 1).

Activity analysis of expression product

The samples from the medium fluid and soluble cytoplasmic fraction showed high specific endoglucanase activities with 534.24 and 127.44 U, respectively, whereas the inclusion bodies could only present 3.78 U, which might have resulted from the retaining soluble cytoplasmic fraction (Table 1). In this study, the cogon red–CMCNa plate was also used to determine the specific endoglucanase activities of expressing protein; the results conformed to that from the spectrophotometry method. The supernatant from medium fraction revealed the largest transparency zone among four fractions, followed soluble cytoplasmic fraction and insoluble cytoplasmic fraction, and could hardly show the bleaching circle around the hole No. 2 poured by the periplasmic fraction (Fig. 5). These results suggested that vector carrying signal peptide sequence was used to export significant amounts of active and soluble target proteins to the extracellular medium for simplifying purification.

Discussion

Recombinant proteins expressed in *E. coli* are often produced as aggregates called inclusion bodies. In many applications, however, it is desirable to express target proteins in their soluble and active forms.

To obtain active and soluble extracellular proteins, secreted expression vector was constructed to allow the target protein export into the periplasm and even extracellular medium (Lei et al., 2008). For this purpose, many vectors, such as pET-20, pET-22, and pET-40, are designed to carry the signal sequence, but the target proteins will not be desirable for periplasmic localization, and recombinant proteins expressed in *E. coli* are often produced as aggregates called inclusion bodies. Ikemura et al. (1987) found that the pro-sequence of subtilisin plays an important role in the formation of enzymatically active subtilisin. We assumed the signal peptide sequence of target genes may be contributive to the expression product transportation of the *E. coli* cells through across-membrane. Based on this idea, we selected pET-28a as expression vector without signal sequence and made the heterologous gene *cel* and its signal sequence insert the pET-28a vector to form a complete opening reading frame. The recombinant vector pET-28a/KX6 was transformed host cell *E. coli* Rosetta (DE3) and high-level gene expression products were produced into medium fluid. Obviously, the self-signal sequence plays an important role in secreting the target protein into medium in a soluble and active form.

With each target protein coming from engineering strains, it is important to verify the localization and estimate the yield in the culture medium or cell. In general, with the high

Table 1 Endoglucanase activity of different fractions

location of target proteins	endoglucanase activity/U	total activity/%
insoluble cytoplasmic fraction	3.78	0.57
soluble cytoplasmic fraction	127.44	19.15
periplasmic fraction	0	0
medium fraction	534.24	80.28

expression levels of the pET system, there may exist part of soluble material in the periplasm and most of the insoluble target protein mass in aggregates. Takeshi Yamamoto et al. (1991) reported that chimeric ras protein with OmpF signal peptide expressed in *E. coli* and OmpF fusion protein was localized in the inner membrane. The target signal sequence may leak into the medium to export proteins to the periplasm with the CBD cenA simplified purification (Weickert et al., 1996). In our experiment, the most target proteins were distributed in medium fluid instead of periplasmic space, which is not consistent with the results from other researches that many recombinant proteins are directed to the periplasm and often end up in the medium through a poorly understand "leakage" phenomenon. Stader and Silhavy (1990) thought that target proteins, in most cases, leaked into the medium due to a damage to the cell envelope rather than true secretion. However, in our study, it is incredible for the mass of target proteins to leak to the extracellular medium unless target proteins secrete into the medium. Therefore, the function of signal peptide sequence of target genes needs to be further studied, which will provide a basis for understanding the target protein secreting mechanism in *E. coli*.

Acknowledgements

This study was supported by the project of Sic-Tech Plan of Hebei Province, China (No. 07225553) and the Key Science and Technology Project of the 11th Five-Year-Plan of Hebei Province, China (No. 06220106D).

References

Huo Y M, He Q W, Zhao S Y, Xu Y F (2008). Prokaryotic expression of *OC- I D86* (Oryzacystatin –I D86) gene and analysis of its activity. Chin J Biotechnol, 24(7): 1194–1198 (in Chinese)

Ikemura H, Takagi H, Inouye M (1987). Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli.* J Biol Chem, 262(16): 7859–7864

Kim S, Lee S B (2008). Soluble expression of archaeal proteins in *Escherichia coli* by using fusion-partners. Protein Expr Purif, 62(1): 116–119

Lei R Y, Qiao Y H, Yan J D, Yang S, Zhu T H (2008). Soluble expression of recombinant human BMP6 in *Escherichia coli* and its purification and bioassay *in vitro.* Chin J Biotechnol, 24(3): 452–459 (in Chinese)

Miller G L (1959). Use of dinitrosalicyclic acid regent for determination of reducing sugar. Anal Chem, 31(3): 426–428

Qiao Y, Mao A J, He Y Z, Liu W F, Dong Z Y (2004). Secreted expression of *Trichoderma reesei* endo-glucanase II gene in *Pichia pastoris* and analysis of enzymatic properties. Mycosystema, 23(3): 388–396

Sørensen H P, Mortensen K K (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli.* Microb Cell Fact, 4(1): 1–8

Stader J A, Silhavy T J (1990). Engineering *Escherichia coli* to secrete heterologous gene products. Methods Enzymol, 185: 166–187

Solingen P, Meijer D, Kleij W A, Barnett C, Bolle R, Power S D, Jones B E (2001). Cloning and expression of an endocellulase gene from a novel streptomycete isolated from an East African soda lake. Extremophiles, 5(5): 333–341

Weickert M J, Doherty D H, Best E A, Olins P O (1996). Optimization of heterologous protein production in *Escherichia coli.* Curr Opin Biotechnol, 7(5): 494–499

Wulff N A, Carrer H, Pascholati S F (2003). Cloning and expression of cellulase Xf818 from *Xylella fastidiosa* in *Escherichia coli*. Scientia Agricola (Piracicaba, Braz.), 60 (4): 715–721

Yamamoto T, Okawa N, Endo T, Kaji A (1991). Expression of chimeric ras protein with OmpF signal peptide in *Escherichia coli*: localization of OmpF fusion protein in the inner membrane. Appl Microbiol Biotechnol, 35(5): 615–621