



Effect of codon optimization on expression levels of human cystatin C in *Pichia pastoris*

Y.M. Li¹, D.J. Li², X.J. Xu³, M. Cui⁴, H.H. Zhen⁴ and Q. Wang⁵

¹Department of Clinical Laboratory, The Linzi Peoples' Hospital of Zibo, Shandong, China

²Department of Clinical Laboratory, Yantai Economic and Technological Development Zone Hospital, Yantai, Shandong, China

³Department of Ophthalmology, Qingdao Fifth Peoples' Hospital, Qingdao, Shandong, China

⁴Research Center, AI DE Diagnostic Co., Ltd., Qingdao, Shandong, China

⁵Department of Clinical Laboratory, The Affiliated Hospital of Medical College, Qingdao University, Qingdao, Shandong, China

Corresponding author: Q. Wang
E-mail: Wangqing0533@aliyun.com

Genet. Mol. Res. 13 (3): 4990-5000 (2014)

Received June 12, 2013

Accepted November 2, 2013

Published July 4, 2014

DOI <http://dx.doi.org/10.4238/2014.July.4.14>

ABSTRACT. Human cystatin C (CysC) is a cysteine proteinase inhibitor with many potential applications. To facilitate further studies of the functions and applications of CysC, we improved the heterologous expression of CysC using a basic codon optimization method. In this study, we cloned the high-GC content wild-type sequence of the *CysC* gene and also designed a slightly AT-biased sequence, with codons optimized for expression in the *Pichia pastoris* GS115 strain. Our results showed that the optimized coding sequence of human *CysC* increased the expression and secretion of the CysC protein by approximately 3- to 5-fold (90-96 mg CysC/L) in yeast, compared with the expression levels of the native *CysC* gene (17.9-18.4 mg CysC/L). We designed,

constructed, and applied an optimized version of the *CysC* gene for the *Pichia* expression system. Our results demonstrate that the optimized coding sequence provides a higher yield of secreted CysC than that produced using the wild-type gene. Our data also serve as a practical example demonstrating a rational design strategy for the heterologous expression of secreted proteins.

Key words: Cystatin C; *Pichia pastoris*; Codon optimization

INTRODUCTION

Cystatin C (CysC) is a small alkaline non-glycosylated protein, also known as the Y-trace protein or post-Y-globulin. CysC is composed of 122 amino acids, its isoelectric point is 9.3, and its molecular weight is 13.3 kDa (Wasen et al., 2002). CysC is a member of the cysteine proteinase inhibitor family and serves many biological functions (Abrahamson et al., 1990). Some potential therapeutic applications of CysC include its use as an antibacterial or antiviral agent (Nakamura et al., 2004), and in the inhibition of cancer metastasis (Ogawa et al., 2003). Another potential use of CysC is for the prevention of proteolysis during food processing. Additionally, endogenous human CysC may serve as a new type of marker for the glomerular filtration rate (GFR), providing a fast, accurate, and simple method for monitoring changes in the GFR in cancer, cirrhosis, rheumatic disorders, renal failure, and other diseases that affect the GFR. Preparation of calibration and control materials for the quantitative detection of CysC, the generation of a high-quality anti-CysC antibody, and an understanding of the physiological and pathological functions of CysC are all prerequisites for establishment of CysC as a marker for the GFR. One reagent that will help in all of these processes is the preparation and purification of the CysC protein. Wang et al. (2012), through the use of codon optimization, reported a dramatic increase in the amount of *cysC* expression in *Escherichia coli*; however, the CysC protein product accumulated in inclusion bodies, which would have to be subjected to complicated renaturation treatments to recover the recombinant protein. Moreover, purification difficulties, poor fusion protein solubility, and the lack of proper post-translational modification of eukaryotic proteins in bacterial cells are all factors that negatively affect the antigenicity and biological activity of the protein.

The *Pichia pastoris* yeast-based expression system is appropriate for the expression of eukaryotic genes and has been used with great success over the last decade (Cregg et al., 2000; Guo et al., 2001; Fang et al., 2004). As a eukaryote, *P. pastoris* has many advantages: improved expression of higher eukaryotic proteins due to correct processing, folding, and post-translational modification of the recombinant protein akin to that of the natural protein; retention of antigenic determinants yielding robust immune responses; and the potential for the protein to be secreted directly into liquid medium, which facilitates purification. Moreover, the expression and purification methods are comparable to the simplicity of *E. coli*-based expression systems. The *P. pastoris*-based expression system is also more efficient, simpler, less expensive, and gives higher yields than baculovirus-based or mammalian tissue culture-based eukaryotic expression systems. These characteristics make *P. pastoris* a particularly useful eukaryotic protein expression system (Corporation, 2005; Hamilton et al., 2006; Cao et al., 2007).

Many strategies have been described for improving the expression of recombinant proteins in *P. pastoris*, such as the introduction of an efficient promoter, expression vectors for the selection of the appropriate signal peptide, increased gene copy numbers, and codon optimization (Xiong et al., 2006). Yadava and Ockenhouse (2003) reported that codon optimization leads to enhanced expression of the F2 domain of the EBA-175 protein in both *E. coli* and *P. pastoris*. However, the protein produced in *P. pastoris* had superior biological properties.

Many studies have reported significant improvements in the expression of foreign genes in yeast by codon optimization (Outchkourov et al., 2002; Chang et al., 2006; Hu et al., 2006; Li et al., 2008; Shumiao et al., 2010; Bai et al., 2011; Gao et al., 2012). The significantly higher protein yields in the native folded state, higher purity, and improved function establish the value of the gene codon optimization approach.

In this study, we redesigned the *CysC* gene using codon conversion, removal of repeated sequences, and adjustment of the GC content according to the characteristics of *P. pastoris* without altering the protein sequence. Then, we cloned the wild-type (*wt-cysC*) and synthetic (*co-cysC*) *CysC* genes into the *P. pastoris* expression vector pPIC9K, with the expectation that the use of codon optimization would significantly increase the amount of recombinant protein expressed.

MATERIAL AND METHODS

E. coli DH5 α cells were from our laboratory stock. *P. pastoris* GS115 (His⁻, Mut⁺) cells, the pPIC9K vector, Platinum[®] Pfx DNA Polymerase, and Platinum[®] HIFI Taq Polymerase were obtained from Invitrogen (Carlsbad, CA, USA), and the *Eco*RI and *Not*I restriction enzymes were purchased from MBI Fermentas (Waltham, MA, USA). Mouse anti-human *CysC* monoclonal antibodies were obtained from Sigma (St. Louis, MO, USA) and the T4 DNA ligase, G418, and protein marker were obtained from Sangon Biotech (Shanghai, China). All other chemicals were of analytical grade.

Construction of the recombinant *wt-cysC* expression vector

Total RNA was extracted from cells using the RNAiso reagent (Takara, Japan) following the manufacturer protocol. The total RNA of human promyelocytic leukemia (HL-60) cells (Key Laboratory of Molecular Virology of Shandong Province, China) was analyzed using 1% agarose gel electrophoresis. Resultant bands and band intensities were observed by ultraviolet-light illumination after ethidium bromide staining. Gel images were captured and analyzed. The absorbance at 260 nm (A_{260}) value of RNA extracted from the cells was also measured using an ultraviolet spectrophotometer (Lambda 45; PerkinElmer; Waltham, MA, USA). We obtained *CysC* cDNA by reverse transcription-polymerase chain reaction (RT-PCR) using the following primers: 5'-GAATTCATGGCCGGGCCCTGCGC-3' (sense) (the *Eco*RI site is underlined) and 5'-GCGGCCCGCTAGGCGTCCTGACAGGTGGA-3' (antisense) (the *Not*I site is underlined). The PCR products were purified, digested with *Eco*RI and *Not*I, re-extracted, and inserted into the pPIC9K vector digested with the same enzymes by ligation. The ligation product was then transformed into *E. coli* DH5 α competent cells. The recombinant *wt-cysC*-pPIC9K plasmid was purified and the incorporation of the correct insert was confirmed by restriction digest and DNA sequencing (Invitrogen Shanghai Ltd., China).

Construction of a recombinant co-cysC expression vector

The mRNA sequence of *CysC* was retrieved from GenBank (Gene ID: 1471) and analyzed, and the gene sequence was optimized based on the codon preference previously reported for *P. pastoris* expression systems (Sreekrishna, 1993; Zhao et al., 2000; Sinclair and Choy, 2002; Bai et al., 2011). For gene optimization, the Leto software (version 1.0.11; Entelechon, Germany) was used, imposing the codon usage of the 30 most highly expressed genes (Bai et al., 2011), except in cases where codons were retained in order to preserve desirable restriction enzyme sites. Furthermore, extended secondary mRNA structures, long-range repeats (including AT-rich and GC-rich regions), and cryptic splice sites were removed, and the GC content was reduced from 68.6 to 43.6%.

According to the gene sequencing results, we designed and synthesized a single-strand oligonucleotide that contained the optimized sequence. *EcoRI* and *NotI* were added to the 5'- and 3'-ends of the oligonucleotide. The synthesized oligonucleotide was spliced to form an integrated gene using PCR. The synthesized sequence was inserted into the pMD-18T vector and transformed into *E. coli* as described for the *wt-cysC* expression vector, and the incorporation of the gene sequence in the recombinant plasmid was confirmed by sequencing. A fragment of approximately 400 bp was obtained from the *co-cysC*-pMD-18T plasmid after restriction with *EcoRI* and *NotI*. The recombinant *co-cysC*-pPIC9K plasmid was obtained by ligating this restriction fragment with the pPIC9K vector. The presence of the correct open reading frame (ORF) was verified by restriction and sequencing analysis of the recombinant plasmid, recovered from *E. coli* DH5 α , and transformed with the ligation mixture.

Transformation of *P. pastoris* GS115

We linearized 5-10 mg plasmid DNA with *SalI* and electrotransformed it into 115 *P. pastoris* GS cells (1.5 kV, 200 Ω , 25 mF; Gene Pulser; Bio-Rad, USA). *His+* transformants were selected on histidine-deficient MD plates. Then, PCR analysis was performed to analyze the positive transformants. A pair of primers designed to amplify the inserted fragment was used in PCR amplification. The primer sequences were as follows: *aldehyde oxidase 1 (AOX1)*-F, 5'-GACTGGTTCCAATTGACAAGC-3' and *AOX1*-R, 5'-GCAAATGGCATTCTGACATCC-3'. PCR conditions were as follows: denaturation at 94°C for 5 min; followed by 25 cycles of denaturation at 94°C for 1 min, annealing at 52°C for 1 min, and extension at 72°C for 1 min; and a final extension at 72°C for 10 min.

Screening multiple copies of positive yeast transformants

In vivo screening of multiple inserts was performed according to the pPIC9K expression manual (Invitrogen) and the G418 antibiotic was used at 4 concentrations: 0.5, 1.0, 2.0, and 3.0 mg/mL. The Mut⁺ yeast transformants were screened by replica-plating the G418-resistant transformants on MD/MM plates for 4 days.

CysC expression

We cultured *wt-cysC*-pPIC9K and *co-cysC*-pPIC9K transformants in 100 mL buffered

glycerol complex medium and grew them at 28-30°C and 250 rpm until the absorbance at 600 nm (A_{600}) reached a value between 3 and 6. The cells were collected by centrifugation at 3000 *g* for 5 min at 4°C. The cell pellets were then resuspended in 500 mL buffered methanol complex medium, pH 6.0, containing 1% casamino acids (BD Difco, San Jose, CA, USA) to produce a suspension with $A_{600} = 1$. The expression of CysC was induced by addition of 1% methanol at 24-h intervals for 5 days, and the optimal harvest time was ascertained at intervals of 0, 24, 48, 72, 96, 120, and 144 h by analysis with a 7600 Chemistry Analyzer (Hitachi, Japan). The empty pPIC9K vector lacking exogenous genes was also transformed into *P. pastoris* GS115, and the transformant was used as a negative control, which was induced under the same conditions.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot analysis

Proteins sampled from 30 mL culture supernatant that were harvested after induction from three independent *wt-cysC*-pPIC9K and *co-cysC*-pPIC9K transformants were separated by 12% (w/v) SDS-PAGE and visualized by Coomassie Brilliant Blue R-250 staining.

For Western blot analysis, the proteins separated by SDS-PAGE were transferred to a nitrocellulose membrane (Bio-Rad, Hercules, CA, USA) at 20 V for 40 min and then incubated with a mouse anti-human CysC monoclonal antibody (1:1000 dilution) (Abcam, UK) and a mouse anti-rabbit IgG antibody (1:4000 dilution) conjugated with horseradish peroxidase. The membrane was washed and developed with tetramethylbenzidine for 10 min (Maniatis et al., 1982).

RESULTS

Optimization of the *cysC* gene

As we suspected that the codon bias of the native *CST3* gene might be a possible cause of poor mRNA translation and CysC production in *P. pastoris*, we designed a codon-optimized version of the gene to overcome this problem. Several groups have proposed the preferred codons of *P. pastoris* based on different calculation methods (Bai et al., 2011; Sreekrishna, 1993). Analyses of the results of these previous studies revealed that a number of amino acids have two preferred codons. Therefore, although the most preferred codons were used during codon optimization as much as possible, the second most preferred codons were used alternatively to meet the following considerations: 1) elimination of stable secondary structures, such as hairpins exceeding 6 bp, especially in the vicinity of the 5'- and 3'-coding regions of the corresponding mRNA sequence (Outchkourov et al., 2002); 2) prevention of probable depletion and congestion of tRNAs due to the choice of the same codon consecutively for the most frequently occurring amino acids; and 3) removal of undesirable repeats and false-priming events to facilitate subsequent gene synthesis. Finally, a total of 77 non-preferred codons in the native gene were replaced by the *P. pastoris*-preferred codons in the synthetic gene, which represents 64% (77/120) of the total amino acid sequence. We eliminated 9 long-range repeats, including AT-rich and GC-rich regions, which might have affected mRNA stability and the TAG terminator conversion into TAA. As a result, codon optimization significantly reduced

the overall GC content of the full-length gene from 68.6 to 43.6%.

A comparison of the redesigned *co-cysC* gene sequence with *wt-cysC* is shown in Table 1.

Table 1. Sequence comparison between wild-type and synthetic CysC genes.

aa	S	S	P	G	K	P	P	R	L	V	G	G	P	M	D	A	S	V	E	E
co	TCT	TCT	CCA	GGT	AAG	CCA	CCA	AGA	TTG	GTT	GGT	GGT	CCA	ATG	GAC	GCT	TCT	GTT	GAA	GAA
wt	TCC	AGT	CCC	GGC	AAG	CCG	CCG	CGC	CTA	GTG	GGA	GGC	CCC	ATG	GAC	GCC	AGC	GTG	GAG	GAG
aa	E	G	V	R	R	A	L	D	F	A	V	G	E	Y	N	K	A	S	N	D
co	GAA	GGT	GTT	AGA	AGA	GCT	TTG	GAC	TTC	GCT	GTT	GGT	GAA	TAC	AAC	AAG	GCT	TCT	AAC	GAC
wt	GAG	GGT	GTG	CGG	CGT	GCA	CTG	GAC	TTT	GCC	GTC	GGC	GAG	TAC	AAC	AAA	GCC	AGC	AAC	GAC
aa	M	Y	H	S	R	A	L	Q	V	V	R	A	R	K	Q	I	V	A	G	V
co	ATG	TAC	CAC	TCT	AGA	GCT	TTG	CAA	GTT	GTT	AGA	GCT	AGA	AAG	CAA	ATC	GTT	GCT	GGT	GTT
wt	ATG	TAC	CAC	AGC	CGC	GCG	CTG	CAG	GTG	GTG	CGC	GCC	CGC	AAG	CAG	ATC	GTA	GCT	GGG	GTG
aa	N	Y	F	L	D	V	E	L	G	R	T	T	C	T	K	T	Q	P	N	L
co	AAC	TAC	TTC	TTG	GAC	GTT	GAA	TTG	GGT	AGA	ACT	ACT	TGT	ACT	AAG	ACT	CAA	CCA	AAC	TTG
wt	AAC	TAC	TTC	TTG	GAC	GTG	GAG	CTG	GGC	CGA	ACC	ACG	TGT	ACC	AAG	ACC	CAG	CCC	AAC	TTG
aa	D	N	C	P	F	H	D	Q	P	H	L	K	R	K	A	F	C	S	F	Q
co	GAC	AAC	TGT	CCA	TTC	CAC	GAC	CAA	CCA	CAC	TTG	AAG	AGA	AAG	GCT	TTC	TGT	TCT	TTC	CAA
wt	GAC	AAC	TGC	CCC	TTC	CAT	GAC	CAG	CCA	CAT	CTG	AAA	AGG	AAA	GCA	TTC	TGC	TCT	TTC	CAG
aa	I	Y	A	V	P	W	Q	G	T	M	T	L	S	K	S	T	C	Q	D	A
co	ATC	TAC	GCT	GTT	CCA	TGG	CAA	GGT	ACT	ATG	ACT	TTG	TCT	AAG	TCT	ACT	TGT	CAA	GAC	GCT
wt	ATC	TAC	GCT	GTG	CCT	TGG	CAG	GGC	ACA	ATG	ACC	TTG	TCG	AAA	TCC	ACC	TGT	CAG	GAC	GCC

aa = amino acid; co = synthetic CysC gene; wt = wild-type CysC gene. Standard single-letter notations have been used to specify the amino acids corresponding to the codons in the CysC genes.

Identification of recombinant plasmids

The recombinant *wt-cysC*-pPIC9K and *co-cysC*-pPIC9K plasmids were extracted using alkaline lysis, and then analyzed by both restriction with *EcoRI* and *NotI* and DNA sequencing. The results confirmed that both plasmids met the design requirements.

Transformation of *P. pastoris* GS115 and screening multiple copies of positive recombinant yeast transformants

His⁺Mut⁺ transformants were selected on MM and MD plates. Then, positive transformants were identified using PCR analysis of the 5'- and 3'-*AOX1* gene in the yeast vector. We used lithium chloride and electrical conversion to transform *P. pastoris*. Our results showed that the two methods were able to produce a sufficient number of transformants, but a higher efficiency of transformation was obtained with the electroschock method.

The pPIC9K plasmid, a *Pichia* vector for multi-copy integration and expression of secreted proteins, can integrate multiple copies of recombinant genes in *Pichia*, resulting in the subsequent increase in secreted protein expression. This vector contains the Tn903 kanamycin resistance (*kan^r*) gene that confers G418 resistance, which has been reported to have a tight correlation with copy number (Scorer et al., 1994). Thus, the transformants with multi-copy vectors can be isolated via a G418 growth inhibition screen. Previous studies indicated that increasing the copy number of the desired recombinant gene leads to a corresponding increase in expression (Cregg et al., 1993); therefore, in our study, when comparing the expression

level of the preoptimized gene with that of the wild-type gene in *P. pastoris*, we chose transformants with the same G418 resistance to avoid potential effects of different gene dosages. To test the effect of codon optimization on expression, both the wild-type and synthetic constructs were screened for the presence of multi-copy clones by increasing the concentration of G418. Three His⁺Mut⁺ transformants with multiple inserts of *wt-cysC* and *co-cysC* were screened on G418-containing media (by selecting the transformants that could grow at 2.0 mg/mL G418, but not at 3.0 mg/mL G418).

SDS-PAGE analysis of the purity and yield of CysC secreted into the culture supernatant

SDS-PAGE analysis indicated that a 19-kDa protein band, consistent with the theoretically calculated molecular weight of the CysC fusion protein, was common in protein lysates from all of the transformants. The highest expression of the CysC fusion protein in the culture supernatant was observed 96 h after induction (Figure 1).

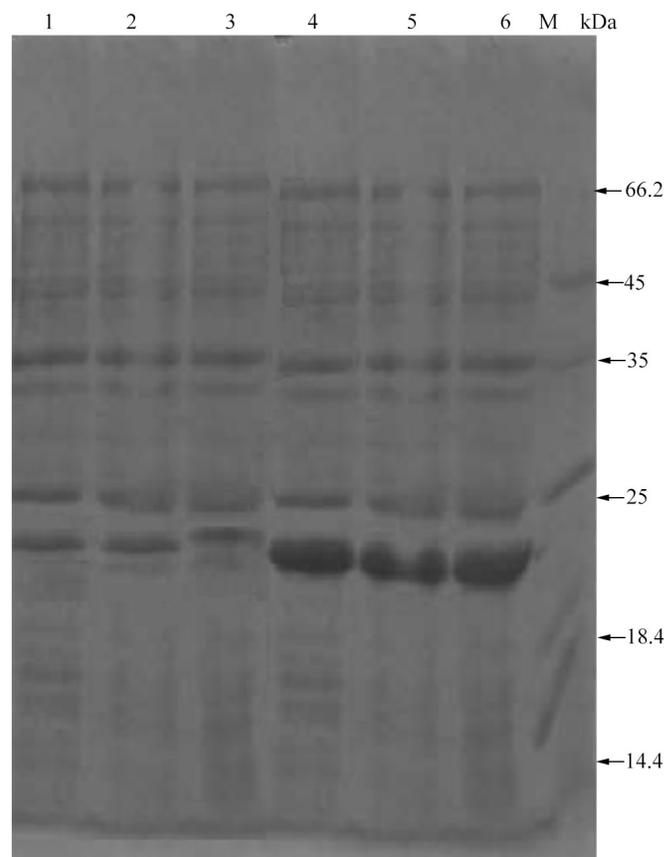


Figure 1. Lanes 1-3 = wild-type CysC protein expressed from GS115/pPIC9K-*wt-cysC*. Lanes 4-6 = modified CysC protein expressed from GS115/pPIC9K-*co-cysC*. Lane M = low-range protein molecular weight marker.

Three high-expression transformants of *wt-cysC* accounted for approximately 45, 42, and 41% of the total protein secretion, and three *co-cysC* strains accounted for secretion of 87, 83, and 80% of the total protein in the supernatant.

The mean expression levels of the *co-cysC* strains were 90, 99, and 96 mg/L. Statistical analyses revealed that there were significant differences in CysC levels between the 3 high-expression transformants of *wt-cysC* ($P < 0.05$). This yield was increased by approximately 3- to 5-fold in comparison with the CysC levels obtained in yeast transformants with the native gene, i.e., 17.9, 18.5, and 18.4 mg/L.

Characterization of recombinant protein antigenicity by Western blot analysis

The anti-human CysC monoclonal antibodies were shown to react with a single band, representing the recombinant proteins, with an approximate molecular mass of 19.0 kDa (Figure 2), suggesting that the recombinant proteins that reacted with the antibodies were human CysC.

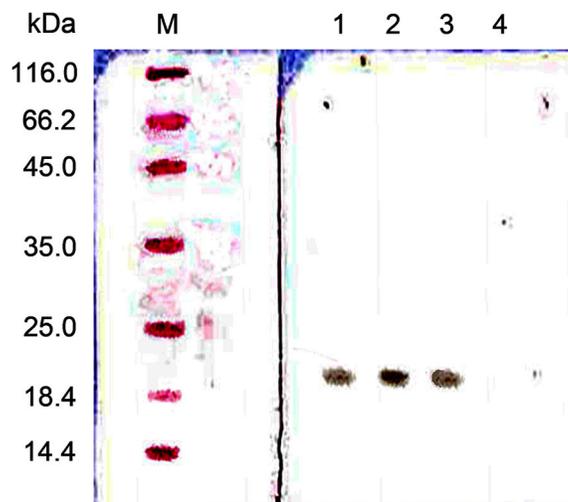


Figure 2. Lane M = low-range protein molecular weight marker. Lanes 1-3 = modified CysC protein expressed from GS115/pPIC9K-co-cysC. Lane 4 = supernatant of culture of the induced strain GS115/pPIC9K.

DISCUSSION

To obtain high-purity and high-yield expression of secreted CysC protein, we chose the pPIC9K expression vector and the *P. pastoris* GS115 strain as the host to express CysC, since a number of studies have concluded that many proteins produced in *P. pastoris* are similar in structure and bioactivity to the native proteins produced in mammalian cells (Ouchkourov et al., 2002; Hu et al., 2006; Teng et al., 2007; Fu et al., 2011; Huang et al., 2012; Yang et al., 2012). Moreover, the pPIC9K vector has an α -factor secretion signal sequence, which has been used with great success.

However, although CysC was secreted when expressed in *P. pastoris*, the very low secretion and/or expression levels of the *wt-cysC* gene restricted its yield. There are many factors that markedly influence the expression of foreign genes, including *P. pastoris* codon bias, the AT and GC contents of foreign genes, translational blocks, the mRNA secondary structure of foreign genes, and growth conditions (Qian et al., 2003).

Hu et al. (2006) speculated that the increase of gene expression by codon optimization is likely to be mainly due to the enhanced efficiency of translation elongation, rather than improvements in translation initiation. Northern blotting analysis revealed that the mRNA levels were similar between the native and the synthetic gene constructs, suggesting that the increased expression was most likely attributable to the enhancement of post-transcriptional processing. This was experimentally supported by the findings that the amount of rare codons was positively correlated with the decreased mRNA stability and translation elongation rate in the yeast *Saccharomyces cerevisiae* (Hoekema et al., 1987; Herrick et al., 1990; Liu et al., 2013). In addition to codon choice, mRNA structure can also influence translation efficiency in yeast. The specific secondary structures formed or lost near the untranslated regions and start codons of mRNA were shown to significantly influence the mRNA degradation rates and translation initiation efficiencies (Oliveira et al., 1993). However, less data are available regarding the effect of secondary structures in the coding region on mRNA stability and translation. It was reported that optimization of the GC content alone, despite the actual codon bias, could significantly improve recombinant protein expression in *P. pastoris*, as well as in mammalian cells (Kim et al., 1997; Sinclair and Choy, 2002).

In fact, since *P. pastoris* genes more commonly have A/T-ended codons, whereas mammalian genes tend toward GC-ended codons, codon optimization of a mammalian-derived gene into one that has the *P. pastoris*-preferred codons usually leads to relatively low GC content, which is accompanied by the production of mRNA with fewer stable secondary structures.

The pattern of codon usage generally dictates optimal gene expression. It was reported that 25 of the total 61 yeast codons were optimal, whereas in *Pichia*, Zhao et al. (2000) validated 19 optimal expression codons by analyzing the usage of 28 synonymous codons. Many researchers have managed to optimize these codons in *Pichia* to improve the expression of heterologous genes (Outchkourov et al., 2002; Hu et al., 2006; Chang et al., 2006; Li et al., 2008; Shumiao et al., 2010; Gao et al., 2012).

A previous study used the wild-type *CysC* gene for expression of CysC in *Pichia* (Li et al., 2002). It is possible that the codons of this gene were not the most optimal for expression in that host. Hence, we hypothesized that the expression might be improved by codon alteration. Our initial attempts at CysC expression with *wt-cycC*-pPIC9K indicated that the expression was either not successful or, if present, was weak. Therefore, after considering the characteristics of the yeast host and pattern of codon usage, taking into account removal of undesirable features, such as complex secondary structures and repeated sequences in the synthetic gene, and preventing mRNA degradation and early termination of translation, we redesigned the *CysC* gene. In the redesigned *CysC* gene, we modified the GC ratio, deleted AT-rich sequences, and removed 5 or more consecutive A/T or G/C repeats. Finally, of all of the 120 codons of the *CysC* gene, 77 were modified to obtain the redesigned version of the *CysC* gene. We altered 9 codons to improve the stability of mRNA, while the stop codon was altered from TAG to TAA, and the GC ratio was modified from 68.6 to 43.6%. According to an analysis of the coding sequence of human *CysC* with the Leto software, a high percentage of

rare codons (about 8.3%) were represented in *P. pastoris*, such as CGA (1 occurrence), CGG (1), CGC (4), CCG (2), GCG (1), and ACG (1). Meanwhile, the GC content of the *CysC* gene is high (68.6%). Thus, we redesigned and synthesized its coding sequence so as to decrease the GC content to 43.6% and chose the preferential codons of *P. pastoris*. This codon optimization moderately increased the CysC expression levels by 3- to 5-fold and up to 16 mg/L. Although this increase in magnitude is somewhat lower than many of those previously reported (5- to 10-fold) in codon optimization studies in *P. pastoris* (Brocca et al., 1998), it is still higher than the results of some other reports (Outchkourov et al., 2002; Wolff et al., 2001).

In conclusion, we optimized codons encoded in the *CysC* gene to improve its heterologous expression in *P. pastoris* GS115. This expression was significantly improved, while dispensing with other methods such as purification from inclusion bodies and cleavage with thrombin. Meanwhile, because the α -factor signal sequence in pPIC9K increased the secretion of the recombinant protein, and because *P. pastoris* secretes few extracellular proteins, the culture medium contained partially pure heterologous protein, thus facilitating its purification and improving the yield. Therefore, we conclude that codon optimization combined with the other design considerations we employed provide a better method for expressing recombinant proteins at a lower purification cost.

ACKNOWLEDGMENTS

Research supported by the Shandong Science Technology Development Project (#2011YD21030).

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

- Abrahamson M, Olafsson I, Palsdottir A, Ulvsback M, et al. (1990). Structure and expression of the human cystatin C gene. *Biochem. J.* 268: 287-294.
- Bai J, Swartz DJ, Protasevich II, Brouillette CG, et al. (2011). A gene optimization strategy that enhances production of fully functional P-glycoprotein in *Pichia pastoris*. *PLoS One* 6: e22577.
- Brocca S, Schmidt-Dannert C, Lotti M, Alberghina L, et al. (1998). Design, total synthesis, and functional overexpression of the *Candida rugosa* lip1 gene coding for a major industrial lipase. *Protein Sci.* 7: 1415-1422.
- Cao Y, Qiao J, Li Y and Lu W (2007). *De novo* synthesis, constitutive expression of *Aspergillus sulphureus* beta-xylanase gene in *Pichia pastoris* and partial enzymic characterization. *Appl. Microbiol. Biotechnol.* 76: 579-585.
- Chang SW, Lee GC and Shaw JF (2006). Codon optimization of *Candida rugosa* lip1 gene for improving expression in *Pichia pastoris* and biochemical characterization of the purified recombinant LIP1 lipase. *J. Agric. Food Chem.* 54: 815-822.
- Corporation I (2005). A Manual of Methods for Expression of Recombinant Proteins Using pPICZ and pPICZa in *Pichia pastoris*. Invitrogen Corporation, Carlsbad.
- Cregg JM, Vedvick TS and Raschke WC (1993). Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology* 11: 905-910.
- Cregg JM, Cereghino JL, Shi J and Higgins DR (2000). Recombinant protein expression in *Pichia pastoris*. *Mol. Biotechnol.* 16: 23-52.
- Fang S, Liu T, Deng P and Zhao S (2004). Expression and identification of cecropin D gene cloned in *Pichia pastoris* yeast cells. *Wei Sheng Yan. Jiu.* 33: 81-85.
- Fu XY, Zhao W, Xiong AS, Tian YS, et al. (2011). High expression of recombinant *Streptomyces* sp. S38 xylanase in *Pichia pastoris* by codon optimization and analysis of its biochemical properties. *Mol. Biol. Rep.* 38: 4991-4997.
- Gao Z, Li Z, Zhang Y, Huang H, et al. (2012). High-level expression of the *Penicillium notatum* glucose oxidase gene in

- Pichia pastoris* using codon optimization. *Biotechnol. Lett.* 34: 507-514.
- Guo RT, Chou LJ, Chen YC, Chen CY, et al. (2001). Expression in *Pichia pastoris* and characterization by circular dichroism and NMR of rhodostomin. *Proteins* 43: 499-508.
- Hamilton SR, Davidson RC, Sethuraman N, Nett JH, et al. (2006). Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* 313: 1441-1443.
- Herrick D, Parker R and Jacobson A (1990). Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 10: 2269-2284.
- Hoekema A, Kastelein RA, Vasser M and de Boer HA (1987). Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. *Mol. Cell Biol.* 7: 2914-2924.
- Hu S, Li L, Qiao J, Guo Y, et al. (2006). Codon optimization, expression, and characterization of an internalizing anti-ErbB2 single-chain antibody in *Pichia pastoris*. *Protein Expr. Purif.* 47: 249-257.
- Huang Y, Chen Y, Mo D, Cong P, et al. (2012). Attenuated secretion of the thermostable xylanase xynB from *Pichia pastoris* using synthesized sequences optimized from the preferred codon usage in yeast. *J. Microbiol. Biotechnol.* 22: 316-325.
- Kim CH, Oh Y and Lee TH (1997). Codon optimization for high-level expression of human erythropoietin (EPO) in mammalian cells. *Gene* 199: 293-301.
- Li HX, Li YL, Xu GB and Zhu LH (2002). High efficient expression of recombinant human cystatin C in *Pichia pastoris*. *Chin. J. Lab. Med.* 25: 271-273.
- Li Z, Hong G, Wu Z, Hu B, et al. (2008). Optimization of the expression of hepatitis B virus e gene in *Pichia pastoris* and immunological characterization of the product. *J. Biotechnol.* 138: 1-8.
- Liu X, Wu D, Wu J and Chen J (2013). Optimization of the production of *Aspergillus niger* alpha-glucosidase expressed in *Pichia pastoris*. *World J. Microbiol. Biotechnol.* 29: 533-540.
- Maniatis T, Fritsch EF and Sambrook J (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Cold Spring Harbor.
- Nakamura S, Hata J, Kawamukai M, Matsuda H, et al. (2004). Enhanced anti-rotavirus action of human cystatin C by site-specific glycosylation in yeast. *Bioconjug. Chem.* 15: 1289-1296.
- Ogawa M, Jing H, Kitts DD, Nakai S, et al. (2003). *In vitro* anti-cancer activities in Caco-2 and HCT-116 cells of recombinant cystatin C prepared by a *Pichia* expression system. *J. Med. Food* 6: 317-322.
- Oliveira CC, van den Heuvel JJ and McCarthy JE (1993). Inhibition of translational initiation in *Saccharomyces cerevisiae* by secondary structure: the roles of the stability and position of stem-loops in the mRNA leader. *Mol. Microbiol.* 9: 521-532.
- Outchkourov NS, Stiekema WJ and Jongsma MA (2002). Optimization of the expression of equistatin in *Pichia pastoris*. *Protein Expr. Purif.* 24: 18-24.
- Qian P, Li X, Tong G and Chen H (2003). High-level expression of the ORF6 gene of porcine reproductive and respiratory syndrome virus (PRRSV) in *Pichia pastoris*. *Virus Genes* 27: 189-196.
- Scorer CA, Clare JJ, McCombie WR, Romanos MA, et al. (1994). Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. *Biotechnology* 12: 181-184.
- Shumiao Z, Huang J, Zhang C, Deng L, et al. (2010). High-level expression of an *Aspergillus niger* endo-beta-1,4-glucanase in *Pichia pastoris* through gene codon optimization and synthesis. *J. Microbiol. Biotechnol.* 20: 467-473.
- Sinclair G and Choy FY (2002). Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. *Protein Expr. Purif.* 26: 96-105.
- Sreekrishna K (1993). *Strategies for Optimizing Protein Expression and Secretion in the Methylotrophic Yeast Pichia pastoris*. American Society for Microbiology, Washington.
- Teng D, Fan Y, Yang YL, Tian ZG, et al. (2007). Codon optimization of *Bacillus licheniformis* beta-1,3-1,4-glucanase gene and its expression in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 74: 1074-1083.
- Wang Q, Mei C, Zhen H and Zhu J (2012). Codon preference optimization increases prokaryotic cystatin C expression. *J. Biomed. Biotechnol.* 2012: 732017.
- Wasen E, Suominen P, Isoaho R, Mattila K, et al. (2002). Serum cystatin C as a marker of kidney dysfunction in an elderly population. *Clin. Chem.* 48: 1138-1140.
- Wolff AM, Hansen OC, Poulsen U, Madrid S, et al. (2001). Optimization of the production of *Chondrus crispus* hexose oxidase in *Pichia pastoris*. *Protein Expr. Purif.* 22: 189-199.
- Xiong AS, Yao QH, Peng RH, Zhang Z, et al. (2006). High level expression of a synthetic gene encoding *Peniophora lycii* phytase in methylotrophic yeast *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 72: 1039-1047.
- Yadava A and Ockenhouse CF (2003). Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. *Infect. Immun.* 71: 4961-4969.
- Yang M, Johnson SC and Murthy PP (2012). Enhancement of alkaline phytase production in *Pichia pastoris*: influence of gene dosage, sequence optimization and expression temperature. *Protein Expr. Purif.* 84: 247-254.
- Zhao X, Huo KK and Li YY (2000). Synonymous codon usage in *Pichia pastoris*. *Sheng Wu Gong Cheng Xue Bao* 16: 308-311.