Research Article

Advanced Techniques of Heterologous Gene Expression for Preparing Pharmaceutical and Clinical Proteins

Wang H[1*], Ma K[2*], Shao W[1], Wang Q[1], He Y[3]
1. School of Environment, Jiangsu University, Zhenjiang, China.
2. Department of Biology, University of Waterloo, Waterloo, ON, N2L 3G1 Canada.
3. Shine E Biotech, Nanjing, China.

ABSTRACT
Recombinant gene expression has been the most commonly used method to obtain large quantity of proteins such as enzymes and antigens for pharmaceutical and clinical applications. Although many genes have been efficiently expressed in *Escherichia coli*, low level of expression and inclusion body formation are still a bottle-neck problem in desired protein production. The expression vectors pHsh and pEXC are newly developed by using elements regulating the expression of a group of genes encoding heat-shock or cold-shock chaperones to overcome the low expression level and the formation of inclusion body. Thermostable enzymes are obtained using the pHsh system for applications in the biosynthesis or modification of pharmaceutical chemicals, while unstable antigens are prepared in soluble form using the T-vector of the pEXC expression system. The successful development of these new vectors offers a convenient technique for the preparation of recombinant proteins being used in enzymatic drug production, as diagnostic markers and medicinal proteins.

Keywords: Diagnostic markers, drug production, gene expression, new expression vectors, soluble expression

Received 1 Feb 2016 Received in revised form 10 Aug 2016 Accepted 13 Aug 2016

***Address for correspondence**
Wang H,
School of Environment, Jiangsu University, 301 Xuefu Road, Zhenjiang, Jiangsu 212013, PR China.
E-mail: hcwang@ujs.edu.cn
Ma K,
Department of Biology, University of Waterloo, Waterloo, ON, N2L 3G1 Canada.
E-mail: kma@uwaterloo.ca

INTRODUCTION

An increasing number of proteins including enzymes and antigenic proteins are employed for biosynthesis and bio-modification of drugs, or as diagnostic tools and medicinal proteins. For the modern pharmaceutical and clinical studies, recombinant gene expression has been the mostly used method to obtain enzymes and antigenic proteins. *Escherichia coli* is often the first choice for being a heterologous host of gene expression because it is easily accessible, less time consuming and inexpensive to cultivate, and its expression vector systems have been well developed.

To increase the heterologous expression levels in *E. coli*, the foreign gene is placed under the control of a regulatory promoter, which plays important roles in reducing the adverse effects of recombinant protein on host cells. Using promoters of different sources, many *E. coli* expression vector systems have been developed and the best known vectors are those containing the *lac* promoter and its hybrids, *tac*/*trc*, the bacteriophage λ p_L promoter and T7 promoter, which are, respectively, identified as the *lac*/*tac*/*trc* system, the p_L system and the T7 system [1]. These expression systems have been widely used for the production of valuable proteins.

A high expression level is desired for every target gene because it affects not only the production but also the purification of the target protein. However, most genes encoding those valuable thermostable enzymes are often expressed at very low levels in *E. coli*; proteins encoded by genes from animal, plant or mesophilic microbes often lose activities or become denatured within a few hours at the normal growth temperatures of *E. coli* [2]; some other target proteins are toxic to host cells and therefore difficult to be over-expressed [1,2]. Hence, advanced techniques are to be developed for heterologous gene expression in *E. coli*

system. It is extremely important and useful to develop new techniques and protocols to avoid formation of inclusion body and the inactivation of soluble recombinant proteins.

Chaperone is a type of protein that assists non-covalent folding/unfolding *in vivo* [3,4]. It has been shown that co-expression of chaperone genes can greatly increase the soluble proportion of the resulting recombinant proteins [2]. *E. coli* possesses a group of chaperone genes to overcome stresses such as a heat shock, cold shock or starving during growth phase. In recent years, a gene expression system, pHsh, is developed by using the conserved promoter for gene expression of a group of heat shock proteins which include some chaperones [5]. The expression of foreign genes in pHsh is controlled by an alternative sigma factor, σ^{32} of *E. coli*, while that in the commonly used vectors is controlled by a repressor protein [5]. The presence of hundreds of copies of the heat shock promoter in a cell strengthens the heat shock system of *E. coli*, and favors the high level of production of the target protein in soluble form [6]. A similar gene expression system has also been generated using a promoter of a gene encoding a cold shock protein [7].

DEVELOPMENT OF HEAT AND COLD SHOCK VECTORS

When the growing *E. coli* is subjected to a sudden rise of temperature, the sigma factor σ^{32} recognizes and activates the heat shock promoters that control a group of genes, resulting in the expression of the genes and the quick production of heat shock proteins. The DNA sequences of heat shock promoters are known, and their consensus sequences are different from those of the general promoters recognized by σ^{70}. Although the physiological functions and regulatory mechanisms of the heat shock system in *E. coli* are well understood, heat shock promoters were not used effectively as promoters to regulate the expression of foreign genes in plasmid vectors until the Hsh system is constructed recently [5].

The typical Hsh system vectors such as pHsh contain an origin of replication from pUC18/19 and a synthesized promoter recognized and regulated by σ^{32} encoded by the *rpoH* gene for heat shock responses in *E. coli*. It has been shown that gene expression under the control of the Hsh promoter in the vector pHsh can continue up to 10 h after a heat shock induction [5]. Meanwhile, a procedure named “flow-in-heat” has been designed and has been successfully used to perform temperature up-shift in fermenters so that heat shock induction of gene expression is no longer difficult to perform in large-scale applications [5]. By using pHsh vectors, recombinant proteins can be produced in bigger fermenters at high expression levels and low costs.

Gene expression using the plasmid vectors of the Hsh system employs the regulation mechanism of an alternative sigma factor of *E. coli*, while those in other expression systems are directly or indirectly regulated by repressors [5]. Although its transcription may not be as strong as bacteriophage promoters, the Hsh promoter allows its plasmids to have a very high copy number, which have a pUC18/19 origin. In return, this many copies of Hsh promoter significantly increase the concentration of σ^{32} protein in the host cells by protecting σ^{32} from degradation, resulting in the highest level of the expression of the target gene as well as the expression of the chaperone genes of the heat shock system [6]. Therefore, the overexpression of difficult genes and the soluble expression for aggregation-prone enzymes have been achieved in *E. coli* by using the Hsh expression vector system [5–7].

Cold shock response and cold shock proteins of *E. coli* have been extensively investigated [4], and cold shock promoter has been employed to control recombinant protein expression at low temperatures since 1996 [8,9]. In comparison with the gene expression at low temperatures using T7 or other expression system, the heterologous gene expression controlled by cold shock promoter has significant advantages: a) a fast growth at 37°C before the expression of target gene is induced by a cold shock, and b) the expression of chaperone genes of the cold shock system in *E. coli* is activated for their chaperoning functions. The mechanism for the regulation of the cold shock system is somehow more complicate than that for the heat shock system of *E. coli*. However, the high copy number and small vector size of pHsh may favor the heterologous gene expression controlled by the cold shock promoter. Qian et al. use the regulatory elements of the cold shock system in *E. coli* to replace the heat shock promoter of pHsh and

generate pEXC for the gene expression at low temperatures [10,11].

Thymine-Adenine (TA) cloning is a simple and efficient method for a direct cloning of PCR products into vectors [12]. To reduce steps of cloning into a vector and then sub-cloning into a gene expression vector, we've further constructed the T-vectors of pHsh and pEXC, resulting in pHsh-T and pEXC-T as shown in (**Figure 1**) [11,13]. These T-vectors not only combine the advantages of TA cloning and gene expression functions, but also optimizes the upstream and downstream sequences of a target gene.

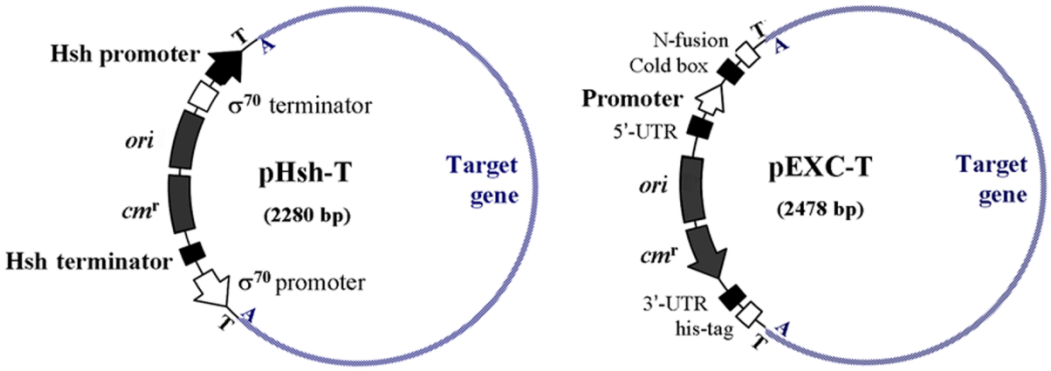

Figure 1: The structures of the expression T-vectors for the pHsh and pEXC Systems. pHsh-T contains a pair of Hsh (heat shock) promoter/terminator, and a pair of those in the reverse direction for the expression of the house keeping genes.

TECHNIQUES FOR CONTROLLING RECOMBINANT GENE EXPRESSION

In reality, the existence and expression of a foreign gene largely stress living cells; expression plasmids are continuously subjected to mutation or modification in cells and the subsequent result is that the recombinant cells harboring modified plasmids suffer less stress and grow faster with less activity or less expression of a target product. Even stored at −70°C, an expression plasmid should not be stored in transformed cells because the cells may modify plasmids by their life activities in their maintenance phase. Therefore, the expression plasmids of the correct sequence must be purified from a fresh culture and dissolved in TE buffer (10 mM Tris-HCl and 1 mM EDTA) for long-term usage and storage [7]. The competent cells of *E.coli* should be freshly transformed to obtain the maximal expression of a recombinant protein.

Growth temperature is an important factor for the application of the expression vectors regulated by temperature up- or down-shift, including those gene expression controlled by σ^{32}, the temperature-sensitive repressors such as *cI*(ts)857, and the cold shock promoters. For example, recombinant cells should not be cultivated at temperatures above 30°C during the procedures before inducing gene expression, which includes cultivation during molecular cloning, sequence improving, gene transformation, inoculum scale-up and so on. Although a sharp up-shift instead of a slow rise of temperature can increase gene-expression levels by about 50%, the basal expression of a target gene is increasing as the growth temperature is raised [7]. Therefore, 30°C and lower temperatures are favorable for reducing basal expression level, maintaining healthy growth of the recombinant cells, and avoiding mutations of the target gene or the vector components in an expression plasmid. This is especially true when a target gene is expressed at very high levels, or a target protein is toxic to the host cells.

The vectors pHsh-T and pEXC-T provide a convenience and a great potential to get a higher overexpression level of the target genes. However, most natural genes are not significantly expressed in an expression plasmid. When a target protein is very important for a research or application purpose, we would modify gene sequence by site-directed mutagenesis and random mutation for improving the expression levels and the enzyme activities. For example, the gene *lcs* of *Thermun thermophiles* HB27 encodes a useful thermostable laccase, however, the wild-type gene is difficult to be expressed in *E. coli*. We mutated *lcs* to reduce

its rear codons and the formation of a potentially secondary structure of mRNA, and modified the promoter sequence to adapt *lcs* sequence, and performed an *in situ* error-prone PCR upon *lcs* open reading frame [14,15]. The expression level of *lcs* is significantly improved step by step (**Figure 2**).

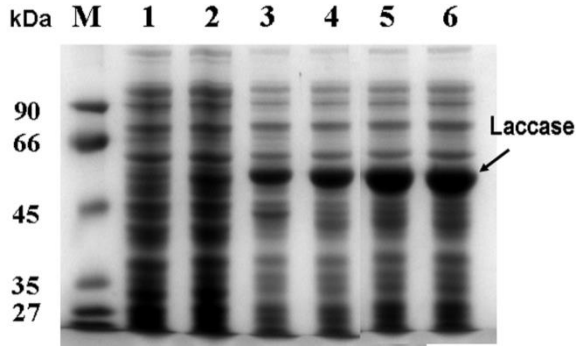

Figure 2: The improvement of the expression levels of a laccase gene (*lcs*) using the pHsh system. Lanes: M, protein molecular mass markers (kDa); 1, wild-type of *lcs*; 2, mutation at the rear codons of *lcs*; 3, mutagenesis of the 2nd structure of mRNA; 4, N-terminal fusion; 5, mutation of the promoter sequence; and 6, after *in situ* error-prone PCR upon the *lcs* gene.

Thermostable enzymes from extremophiles are useful tools for catalyzing the biosynthesis and modification of pharmaceutical chemicals, and some examples are β-glucuronidase, α-arabinosidase and S-adenosyl homocysteine hydrolase [16–18]. These enzymes are efficiently produced using the pHsh vector in *E. coli* (**Figure 3**), and this overcomes the difficulties that their native microbes require extreme conditions to grow. It is also known that some other proteins are present in limited amount and unstable during preparation. While solving the problems of the formation of the inclusion body, we found that many recombinant proteins became precipitated because they were unstable and denatured during the heterologous gene expression or the protein purification from the cell free extracts. We take the advantageous features of the pHsh plasmid elements to construct pEXC, by which the genes encoding P1 protein of *Mycoplasma pneumonia* and streptolysin O of *Streptococcus pyogenes* are overexpressed in soluble form at high levels and the recombinant proteins are purified (**Figure 3**) [11]. The T-vectors pHsh-T and pEXC-T are better systems for being used in the gene cloning and expression for obtaining a large quantity of the functional pharmaceutical and clinical proteins.

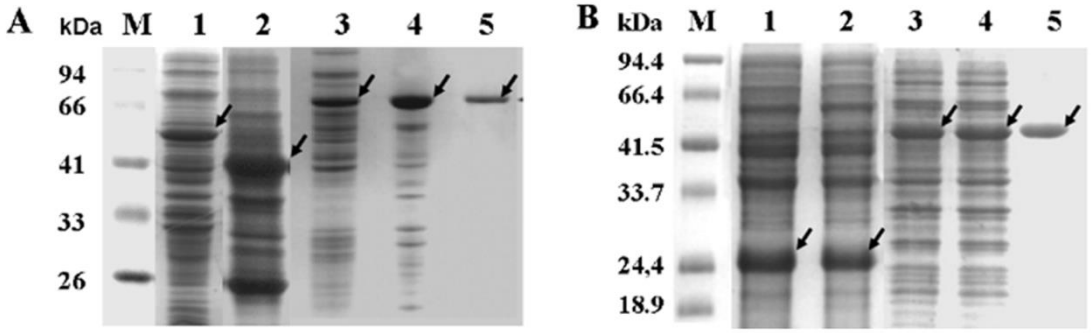

Figure 3: The overexpression of the genes encoding the proteins for pharmaceutical and clinical studies. A. Thermostable enzymes expressed by using the vector pHsh. Lanes: M, protein molecular mass markers; 1 and 2, cell protein of *E. coli* expressing α-arabinosidase and S-adenosyl homocysteine hydrolase; 3-5, expression, purification of and purified β-glucuronidase, respectively. B. Antigens expressed in *E. coli* by using the vector pEXC. Lanes: M, protein molecular mass markers; 1 and 2, whole cell and soluble protein of *E. coli* expressing P1 protein of *Mycoplasma pneumonia*; 3-5, streptolysin O of *Streptococcus pyogenes* from the whole cell, cell-free extract and purified form, respectively.

CONCLUSION

Many foreign genes can be efficiently expressed in *E. coli*, however, most of those unmodified genes are expressed only at low levels and often associated with the formation of the inclusion bodies. The pHsh

and pEXC systems are newly developed by using elements regulating the expression of a group of genes encoding heat-shock or cold-shock chaperones. These vectors exhibit significant advantages in overcoming the low level of the expression and the formation of the inclusion body. Thermostable enzymes can be obtained using the pHsh system for applications in the biosynthesis or modification of pharmaceutical chemicals, while unstable antigens can be prepared in soluble form using the T-vector of the pEXC system. Furthermore, the small sizes of the expression T-vectors, pHsh-T and pEXC-T, are constructed for a one-step constructing of the expression plasmid, an efficient transformation, and a site-directed mutagenesis by inverse PCR. These advanced techniques offer a great convenience for the preparation of recombinant proteins being used in an enzymatic drug production, as a diagnostic marker, and a medicinal protein.

ACKNOWLEDGEMENTS

This work was supported by grants 30970062, 31300088 from the National Science Foundation of China, and BE2016353 from the Science and Technology Program of Jiangsu Province, China.

REFERENCES

1. Baneyx F. Recombinant protein expression in *Escherichia coli*. Curr Opin Biotechnol. 1999; 10: 411–421.
2. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Frontiers in Microbiol. 2014; 5: 172.
3. Skelly S, Coleman T, Fu C, Brot N, Weissbach H. Correlation between the 32-kDa a factor levels and in vitro expression of *Escherichia coli* heat shock genes. Proc Natl Acad Sci. 1987; 84: 8365–8369.
4. Phadtare S, Alsina J, Inouye M. Cold-shock response and cold-shock proteins. Curr Opin Microbiol. 1999; 2: 2: 175-180.
5. Wu H, Pei J, Jiang Y, Song X, Shao W. pHsh vectors, a novel expression system of *Escherichia coli* for the large-scale production of recombinant enzymes. Biotechnol Lett. 2010; 32: 795–801.
6. Le Y, Peng J, Wu H, Sun J, Shao W. A technique for production of soluble protein from a fungal gene encoding an aggregation-prone xylanase in *Escherichia coli*. PLoS One; 6: 4: e18489.
7. Shao W. Cloning, Mutation and over-expression of lignocellulase genes, *in* Biological Conversion of Biomass for Fuels and Chemicals. Cambridge (UK). The Royal Society of Chemistry. 2014. p. 298–317.
8. Vasina JA, Baneyx F. Recombinant protein expression at low temperatures under the transcriptional control of the major *Escherichia coli* cold shock promoter cspA. Appl Environ Microbiol. 1996; 62; 4: 1444-1447.
9. Qing G, Ma LC, Khorchid A, Swapna GVT, Mal TK, Takayama MM et al. Cold-shock induced high-yield protein production in *Escherichia coli*. Nat Biotechnol 2004; 22: 7: 877–882.
10. Yamanaka K, Fang L, Inouye M. The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. Mol Microbiol. 1998; 27: 2: 247–255.
11. He Y, Qi Y, Huang L, Zhou R, Shao W. An expression T-vector and its application at low temperatures. Chin J Biotechnol. 2015; 31: 12: 1773–1783.
12. Holton TA, Graham MW. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res. 1991; 19: 5: 1156.
13. Zhou R. Construction and application of an expression T-vector, pHsh-T (Thesis). Nanjing Normal University. 2015; p: 13–26.
14. Le Y, Chen H, Zagursky R, Wu JHD, Shao W. Thermostable DNA ligase-mediated PCR production of circular plasmid (PPCP) and its application in directed evolution via *in situ* error-prone PCR. DNA Res. 2013; 20: 375–382.
15. Zhang Y. High-level expression of extremely thermostable laccase and xylanase and application in biobleaching of wheat straw pulp (Dissertation). Nanjing Normal University. 2014; p: 27–64.
16. Wang Z, Pei J, Li H, Shao W. Expression, characterization and application of thermostable β-glucuronidase from *thermotoga maritime*. Chin J Biotech. 2008; 24: 8: 1407–1412
17. Pei J, Shao W. Purification and characterization of an extracellular α-L-arabinosidase from a novel isolate *Bacillus pumilus* ARA and its over-

expression in *Escherichia coli*. Appl Microbiol Biotechnol. 2008; 78: 115–121.
18. Qian G, Chen C, Zhou R, He Y, Shao W. A thermostable S-adenosylhomocysteine hydrolase from *Thermotoga maritima*: Properties and its application on S-adenosylhomocysteine production with enzymatic cofactor regeneration. Enzyme Microb Technol. 2014; 64: 65: 33–37.