Protein Expression and Purification 62 (2008) 116–119

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Soluble expression of archaeal proteins in *Escherichia coli* by using fusion-partners

Seonghun Kim [a], Sun Bok Lee [a,b,*]

[a] School of Environmental Science and Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Pohang 790-784, Republic of Korea
[b] Department of Chemical Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Pohang 790-784, Republic of Korea

ARTICLE INFO

Article history:
Received 4 June 2008
and in revised form 24 June 2008
Available online 10 July 2008

Keywords:
Archaeal proteins
Soluble fusion tag
D-Gluconate dehydratase
2-Keto-3-deoxy-D-gluconate kinase
Sulfolobus solfataricus

ABSTRACT

Expression of archaeal proteins in soluble form is of importance because archaeal proteins are usually produced as insoluble inclusion bodies in *Escherichia coli*. In this study, we investigated the use of soluble fusion tags to enhance the solubility of two archaeal proteins, D-gluconate dehydratase (GNAD) and 2-keto-3-deoxy-D-gluconate kinase (KDGK), key enzymes in the glycolytic pathway of the thermoacidophilic archaeon *Sulfolobus solfataricus*. These two proteins were produced as inclusion bodies in *E. coli* when polyhistidine was used as a fusion tag. To reduce inclusion body formation in *E. coli*, GNAD and KDGK were fused with three partners, thioredoxin (Trx), glutathione-*S*-transferase (GST), and N-utilization substance A (NusA). With the use of fusion-partners, the solubility of the archaeal proteins was remarkably enhanced, and the soluble fraction of the recombinant proteins was increased in this order: Trx > GST > NusA. Furthermore, In the case of recombinant KDGKs, the enzyme activity of the Trx-fused proteins was 200-fold higher than that of the polyhistidine-fusion protein. The strategy presented in this work may contribute to the production of other valuable proteins from hyperthermophilic archaea in *E. coli*.

Proteins derived from hyperthermophiles, microorganisms that thrive at temperatures above 80 °C, are very stable and typically remain folded under harsh conditions due to their adaptation to extreme environments. Thus, these hyperstable enzymes are considered as novel biocatalysts which have significant potential for industrial application in the chemical, food, textiles, and pulp and paper industries [1–4]. However, despite the biotechnological potential of these hyperstable enzymes, massive production from hyperthermophilic microorganisms has been a bottleneck to their use because of low specific growth rates and their very low biomass productivity [3]. For several years, extensive research has been performed to mass-produce these enzymes for industrial applications and scientific studies [1–4]. The most promising approach has been heterologous expression of the genes encoding hyperthermophilic enzymes in mesophilic hosts, including bacteria (*Escherichia coli* and *Bacillus* sp.) or eukarya (yeast, *Aspergillus* sp., and fungi) [3–8]. Among these mesophilic host systems, the *E. coli* expression systems have been the most extensively studied in the last decade as they offer certain advantages [3,4]. First, the purification of target proteins is very simple, in that heat treatment removes a large number of the host proteins. Second, the genes encoding hyperthermophilic proteins can be modified by molecular biology techniques, such as mutation or fusion tags, to improve their biochemical properties. Third, the productivity of hyperthermophilic enzymes can be enhanced by using strong promoters or controlling the copy number of plasmids. In addition, sufficient productivity of enzymes can be obtained from recombinant *E. coli* by the high density cultivation techniques that are well established in the field of recombinant protein production.

Although *E. coli* has a remarkable capacity to produce large quantities of hyperthermophilic proteins, recombinant proteins are often expressed at low levels when the codon usage of the mRNA encoding the foreign protein differs from that of the *E. coli* host [9]. Occasionally, recombinant proteins expressed in *E. coli* are also produced in an inactive form due to misfolding and unfolding. One study indicated that, when hyperthermophilic archaea's proteins were expressed in *E. coli*, over 50% of the recombinant products were found in the insoluble fraction in cell lysates [10]. Insoluble fractions consist of unordered aggregates called inclusion bodies [11,12] due to incorrect folding of the polypeptide. Inclusion bodies could result from the failure to control misfolded or unfolded proteins depending on their propensity to trigger a cellular stress response, when recombinant protein is expressed at high rates, or when the protein is expressed under unfavorable protein-folding conditions [13–15].

Inclusion body formation can be minimized by several methods: co-expression of molecular chaperones, decreased cultivation temperature, and reducing the recombinant gene expression rate.

* Corresponding author. Address: Department of Chemical Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Pohang 790-784, Republic of Korea. Fax: +82 54 279 2699.
E-mail address: sblee@postech.ac.kr (S.B. Lee).

1046-5928/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2008.06.015

Table 1
Oligonucleotides used in this study

Oligonucleotides	Sequences	Use
gnaD sense	5′-GGAATTCCATATGCGAATCAGCGAAATAGAA-3′ (NdeI)	pET21-gnaD clone
gnaD sense	5′-CGGGATCCATGCGCATCCGCGAAATTGAACCG-3′ (BamHI)	pET32-gnaD, pET42-gnaD clone
gnaD antisense	5′-CCGCTCGAGAACACCATAATTCTTCCAGG-3′ (*Xho*I)	pET21-gnaD, pET32-gnaD, pET42-gnaD clone
kdgK sense	5′-GGAATTCCATATGGTTGATGTAATAGCTTTG-3′ (NdeI)	pET21-kdgK clone
kdgK sense	5′-GGAATTCATGGTTGATGTAATAGCTTTG-3′ (EcoRI)	pET32-kdgK, pET42-kdgK, pET-43.1-kdgK clone
kdgK antisense	5′-CGGCTCGAGCGTTTTAAACTCATT-3′ (*Xho*I)	pET21-kdgK, pET32-kdgK, pET42-kdgK, pET43.1-kdgK clone

Co-expression of molecular chaperones enhances the solubility of target proteins by facilitating proper folding and assembly of the protein [16]. Decreased cultivation temperature [17] can be beneficial by permitting slower production; in some cases, this has had positive effects on solubility, especially when using the appropriate promoters/inducers, such as the T7 promoter, the arabinose promoter, the chloroplast psbA promoter or the tac promoter [18–20]. Reducing the recombinant gene expression rate has also enhanced the solubility of recombinant protein in some cases [21]. However, these strategies do not give successful results for all recombinant proteins. Consequently, the production of soluble proteins remains a trial-and-error process [22,23]. Another preferable strategy for reducing inclusion body formation is to create a fusion by introducing a unique fusion tag to a target protein [24,25]. This strategy is attractive for producing a soluble, active, properly folded target protein. Commercially available fusion tags for expressing target proteins include thioredoxin (Trx), glutathione-*S*-transferase (GST), N-utilization substance A (NusA) and maltose binding protein (MBP) [26]. These fusion tags could enhance solubility and folding in two ways: the tags provide for fusion to a polypeptide that is itself highly soluble (Trx, GST, NusA, and MBP) and tags can provide for fusion to an enzyme that catalyzes disulfide bond formation (Trx).

In a previous study, we reported the study of the effect of rare codons on the expression of hyperthermophilic archaeal proteins in *E. coli* by examining the heterologous production of *Sulfolobus solfataricus* D-gluconate dehydratase (*Sso_gnaD*) [27]. By using modified *E. coli* cells that contain extra copies of the argU, ileY, and leuW tRNA genes and mutation of the rare codons (AGA and ATA), we were able to express *Sso_gnaD* in *E. coli*. Unfortunately, however, a large portion of recombinant proteins were obtained as biologically inactive inclusion bodies. In this work, we studied the soluble expression of fusion-tagged archaeal D-gluconate dehydratase (GNAD[1]) and 2-keto-3-deoxy-D-gluconate kinase (KDGK), two enzymes involved in the modified Entner–Doudoroff pathways of thermoacidophilic archaea [28,29], in *E. coli*. We also showed that the use of fusion tags could enhance the enzyme activity significantly presumably due to the formation of soluble and biologically active proteins. It is believed that the strategy presented in this work may contribute to the production of other valuable proteins from hyperthermophilic archaea in *E. coli*.

Materials and methods

Cloning and expression of gnaD and kdgK genes

To construct plasmids encoding the target proteins combined with the soluble fusion tags, commercially available pET vectors (Novagen, Madison, WI, USA) were used: pET-21a (6× His), pET-32a (Trx), pET-42a (GST), and pET-43.1a (NusA). The genes encoding D-gluconate dehydratase (*gnaD*) and 2-keto-3-deoxy-D-gluconate kinase (*kdgK*) were directly amplified from the genomic DNA of *S. solfataricus* as the template, by polymerase chain reaction (PCR) with *Taq* DNA polymerase (Bioneer, Daejeon, Korea) using the sense and the antisense primers listed in Table 1. The PCR reaction conditions were as follows: one cycle at 99.9 °C for 10 min; 30 cycles of 94 °C for 45 s, 48 °C for 45 s, and 74 °C for 1 min; and finally one cycle at 74 °C for 7 min. The amplified PCR products of about 1188-bp for *gnaD* and 942-bp for *kdgK* were purified with PCR purification kit (*AccPrep*®, Bioneer, Daejeon, Korea) and then digested with the appropriate restriction enzymes. These coding fragments were inserted between the corresponding predigested restriction sites to generate the plasmids pET21-gnaD, pET32-gnaD, and pET42-gnaD for recombinant GNAD and pET21-kdgK, pET32-kdgK, pET42-kdgK, and pET43.1-kdgK for the recombinant KDGK.

The ligation mixtures were transformed into *E. coli* DH5α. Positive transformants were verified by restriction analysis and sequencing of the plasmid DNA. Some of the correct *gnaD* clones (pET21-gnaD and pET32-gnaD) and *kdg*K clones (pET21-kdgK, pET32-kdgK, and pET43.1-kdgK) were then transformed into two *E. coli* strains, BL21(DE3)pLysS and Origami(DE3)pLysS, each containing a chromosomal copy of the T7 RNA polymerase gene under the control of the *lac*UV5 promoter, which can be induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG). The other plasmids, pET42-gnaD and pET42-kdgK, were transformed only into *E. coli* BL21(DE3)pLysS for recombinant protein expression. The transformants were grown in Luria–Bertani (LB) medium supplemented with 100 μg mL^{-1} of ampicillin at 37 °C. The *E. coli* strains harboring the recombinant plasmids with pET42 origin were cultivated in LB medium supplemented with 30 μg mL^{-1} of kanamycin. To induce expression of the recombinant proteins, IPTG was added to a final concentration of 0.5 mM to culture broths that had grown at 37 °C to OD_{600} 0.4–0.6, and the incubation was continued for 12 h at 20 °C. The cells were then harvested by centrifugation, and frozen at −80 °C until used for further studies.

Enzyme activity measurements

GNAD activity was determined using thiobarbituric acid (TBA) method [28] as follows: the reaction mixture of 50 μl was oxidized using 125 μl of 25 mM periodic acid in 0.25 N H_2SO_4 at room temperature for 20 min. To terminate the oxidation, 250 μl of 2% (w/v) sodium arsenite dissolved in 0.5 N HCl was added. Finally, 1 ml of 0.3% TBA was added and the reaction mixture heated at 100 °C for 10 min. The red chromophore produced was monitored at 549 nm ($\varepsilon = 6.78 \times 10^4\,M^{-1}\,cm^{-1}$) after adding an equal volume of DMSO. One unit of D-gluconate dehydratase was defined as the amount of enzyme producing 1 μmol of 2-keto-3-deoxy-D-gluconate (KDG) per min from D-gluconate under standard assay conditions (30 °C and pH 7.5). All enzyme activities were determined in triplicate.

KDGK activity was determined by enzyme-coupled assay, measuring the amount of ADP generated by the kinase as follows [29]. The reaction mixture of total volume 100 μl was incubated at 60 °C in 50 mM Tris–HCl buffer (pH 7.2), with 1 mM KDG and enzyme solution. After 30 min, a 400 μl solution containing 1 mM phosphoenolpyruvate, 0.5 mM freshly prepared NADH, lactate dehydrogenase (1.5 U)/pyruvate kinase (1.0 U) in 50 mM Tris–HCl buffer (pH 7.2) was added and then the mixture incubated at 37 °C

[1] *Abbreviations used:* GNAD, D-gluconate dehydratase; KDGK, 2-keto-3-deoxy-D-gluconate kinase; Trx, thioredoxin; GST, glutathione-*S*-transferase; NusA, N-utilization substance A; MBP, maltose binding protein.