Gene 440 (2009) 9–15

Contents lists available at ScienceDirect

Gene

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

Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in *Escherichia coli*

Christos A. Kyratsous [a,b], Saul J. Silverstein [b,*], Christine R. DeLong [b], Christos A. Panagiotidis [a]

[a] Department of Pharmaceutical Sciences, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece
[b] Department of Microbiology, College of Physicians and Surgeons, Columbia University, 701 W. 168th street, New York, NY 10032, USA

ARTICLE INFO

Article history:
Received 14 January 2009
Received in revised form 13 March 2009
Accepted 15 March 2009
Available online 26 March 2009

Received by D.L. Court

Keywords:
DnaK
GroEL
PrP
DnaJ/GrpE
GroES

ABSTRACT

The enteric bacterium *Escherichia coli* is the most extensively used prokaryotic organism for production of proteins of therapeutic or commercial interest. However, it is common that heterologous over-expressed recombinant proteins fail to properly fold resulting in formation of insoluble aggregates known as inclusion bodies. Complex systems have been developed that employ simultaneous over-expression of chaperone proteins to aid proper folding and solubility during bacterial expression. Here we describe a simple method whereby a protein of interest, when fused in frame to the *E. coli* chaperones DnaK or GroEL, is readily expressed in large amounts in a soluble form. This system was tested using expression of the mouse prion protein PrP, which is normally insoluble in bacteria. We show that while in trans over-expression of the chaperone DnaK failed to alter partitioning of PrP from the insoluble inclusion body fraction to the soluble cytosol, expression of a DnaK–PrP fusion protein yielded large amounts of soluble protein. Similar results were achieved with a fragment of insoluble Varicella Zoster virus protein ORF21p. In theory this approach could be applied to any protein that partitions with inclusion bodies to render it soluble for production in *E. coli*.

1. Introduction

Demand for proteins required for various molecular biology and biotechnology applications cannot be met from natural reservoirs. Fortunately, the universality of the genetic code enables protein expression in convenient organisms. The enterobacterium *Escherichia coli* (*E. coli*) is frequently a first choice for production of recombinant proteins because of its ease of culture, rapid growth on inexpensive carbon sources, amenability to high cell density fermentations and the extensive knowledge of its genetic and physiological processes (Gold, 1990; Hodgson, 1993; Olins and Lee, 1993; Hockney, 1994; Makrides, 1996). However, despite its popularity, not every protein can be efficiently expressed in this organism. Production of large amounts of recombinant proteins in *E. coli* is often burdened by a number of problems, including stability and translation efficiency of mRNA, degradation of the protein by host cell proteases and product-induced toxicity (Makrides, 1996). The most common problem, however, is that many recombinant proteins fail to reach and maintain their native conformation when overproduced in *E. coli* (Makrides, 1996). Accordingly, the yield of these recombinant proteins is low.

Over-expression of recombinant proteins in the crowded milieu of the *E. coli* cytoplasm, where transcription and translation are tightly coupled, makes folding an extraordinary challenge. Although small, single domain proteins often spontaneously reach their native conformation, expression of more complex and heterologous proteins frequently results in kinetically trapped, slow-folding, nonproductive intermediates that are prone to aggregate (Baneyx and Mujacic, 2004). The latter results from interaction of exposed hydrophobic surfaces within the proteins' partially unfolded structure. These regions are normally buried inside the core of a properly folded native protein. However, failure to correctly fold recombinant proteins results in their exposure and free interaction with other hydrophobic constituents of the crowded intracellular environment (Wetzel, 1994; Baneyx and Mujacic, 2004; Fahnert et al., 2004). Subsequently, these misfolded proteins are usually deposited as dense refractile particles called inclusion bodies or degraded by bacterial proteolytic systems (Wetzel, 1994; Baneyx and Mujacic, 2004; Fahnert et al., 2004). Proteins deposited into inclusions bodies require complex solubilization and refolding procedures to gain biological activity (Fahnert et al., 2004).

Bacterial chaperone proteins facilitate conformational processing of a significant fraction of *de novo* synthesized polypeptides by

Abbreviations: ORF21p, Open reading frame 21 of Varicella Zoster virus; ATP, Adenosine triphosphate; VZV, Varicella zoster virus; 2xYT, 16g tryptone, 10g yeast extract, and 5g NaCl made up to 1L in dH2O; SDS, Sodium dodecyl sulfate; TE, 10mM Tris HCl pH 8.0 and 1mM EDTA; PCR, Polymerase chain reaction; 6His, Six histidine tag; IPTG, Isopropyl B-D-1 thiogalactoside; EDTA, Disodium ethylenediamine tetracetic acid; DTT, Dithiothreitol; SDS-PAGE, Sodium dodecylsulfate polyacrylamide gel electrophoresis; GST, Glutathione S-transferase.

* Corresponding author. Tel.: +1 212 305 8149; fax: +1 212 305 5106.
E-mail addresses: cak2113@columbia.edu (C.A. Kyratsous), sjs6@columbia.edu (S.J. Silverstein), crd2105@columbia.edu (C.R. DeLong), pchristo@pharm.auth.gr (C.A. Panagiotidis).

0378-1119/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2009.03.011

preventing nonproductive hydrophobic interactions and by helping them acquire their correct tertiary conformation (Hartl, 1996; Fink, 1999; Hartl and Hayer-Hartl, 2002; Young et al., 2004). *E. coli* employs two major chaperone systems, the DnaK (Hsp70) system, composed of DnaK and its co-chaperones DnaJ and GrpE, and the GroEL (Hsp60)/GroES system (Hartl, 1996; Fink, 1999; Hartl and Hayer-Hartl, 2002; Young et al., 2004). Expression of these proteins is positively regulated under elevated temperatures or other forms of cellular stress that affect protein conformation (Lemaux et al., 1978; Hoffmann and Rinas, 2004a,b). Chaperones are protein machines that bind to nonnative proteins and rely on ATP-driven conformational changes to mediate refolding/unfolding of their substrates (Hartl, 1996; Fink, 1999; Hartl and Hayer-Hartl, 2002; Young et al., 2004).

It has been speculated that overproduction of a slow-folding recombinant protein in *E. coli* may overwhelm the cell's chaperones, leading to accumulation of aggregates as inclusion bodies (Georgiou and Valax, 1996). Expression of large amounts of chaperones can overcome this limitation and benefit folding of recombinant proteins (Georgiou and Valax, 1996; Schlieker et al., 2002). This idea was first demonstrated by increasing the solubility of Rubisco, a large oligomeric protein, in cells that over-expressed GroEL and GroES (Goloubinoff et al., 1989). Subsequently, a number of increasingly elaborate plasmid systems were developed that permit co-expression of various combinations of chaperones, or other facilitators of protein folding to allow production of large amounts of proteins of interest (Perez-Perez et al., 1995; Nishihara et al., 1998; Vonrhein et al., 1999; Yanase et al., 2002; Stevens et al., 2003; Xu et al., 2005; Schlapschy et al., 2006; de Marco et al., 2007). Although over-expression of substrate-specific chaperone and co-chaperone combinations improves folding of some recombinant proteins, this sometimes results in bacterial toxicity and leads to decreased protein yield (Blum et al., 1992).

Here we describe development of a system that utilizes bacterial chaperones to facilitate production of large amounts of soluble recombinant proteins. This system utilizes a vector where the sequence of a target polypeptide is placed in frame at the carboxy-terminus of either DnaK or GroEL. A poly-histidine carboxy-terminal tag is present for rapid affinity purification of the fusion protein and a thrombin cleavage site is placed at the fusion junction to liberate the protein of interest from its chaperone partner. The system was evaluated using the mouse prion protein (PrP), which is involved in the protein folding disease spongiform encephalopathy, and Varicella Zoster virus (VZV) ORF21p as targets. Both proteins are normally insoluble when expressed in bacteria. We demonstrate that the chaperone fusions are soluble and when these fusion proteins are co-expressed with their cognate co-chaperone solubility is further increased.

2. Materials and methods

2.1. Bacterial strains

E. coli strains used in this study were N99, Top10F' and BL21[DE3]. All strains were grown at 37 °C on 2× YT agar plates or with aeration in 2× YT broth, supplemented with 150 μg/ml Ampicillin, 50 μg/ml Kanamycin and/or 25 μg/ml Chloramphenicol as necessary.

2.2. Isolation of bacterial genomic DNA

Pellets from overnight bacterial cultures were suspended in TE buffer (10 mM Tris HCl pH8.0 and 1 mM EDTA). Sodium dodecyl sulfate (SDS) was added to a final concentration of 0.6% w/v and the solution was incubated with Proteinase K (0.12 mg/ml) for 1 h at 37 °C. Nucleic acids were extracted with phenol/chloroform and precipitated following addition of isopropanol before treatment with RNaseA. DNA was re-extracted, precipitated, dried and suspended in TE.

2.3. Isolation of Varicella Zoster virus DNA

Virus DNA was prepared using purified nucleocapsids as previously described (Kyratsous and Silverstein, 2007).

2.4. Plasmid construction

2.4.1. DnaK plasmids

The *dnaK* coding sequence was amplified from *E. coli*, strain N99 genomic DNA using the oligonucleotide primers DnaK-FW: 5′ – GGCATATGGGTAAAATAATTGGTATCG – 3′ and DnaK-RV: 5′ – GGGG-GATCCGAGCCGCGTGGGACTAATTCAAATTCAGCGTAGACAAC – 3′. The PCR product was cloned into pDrive (Qiagen) to yield pD-DnaK. The DnaK gene was released with NdeI/BamHI and cloned into NdeI/BamHI digested pET-21a(+) to produce pXCK-K.

2.4.2. GroEL plasmids

GroEL was amplified from pUCE using the oligonucleotide primers GroEL-FW: 5′ – GGCATATGGCAGCTAAAGACGTA – 3′ and GroEL-RV: 5′ – GGGAGATCTGAGCCACGAGGGACTAAGTCAGCTGCATCGTT – 3′. The PCR product was cloned into EcoRV digested pZero2.1 (Invitrogen) to yield pZ-GroEL. pXCK-EL was constructed by cloning the NdeI/BglII digestion fragment from pZ-GroEL into NdeI/BamHI digested pET-21a(+).

2.4.3. PrP plasmids

The gene encoding murine PrP was amplified from pVPmPrP14 (a gift from Dr. Sklaviadis, Aristotle University, Thessaloniki, Greece) using mPrP-FW: 5′ – GGGCTAGCAAAAAGCGGCCAAAGCCT – 3′ and mPrP-RV: 5′ – GGCGCGGCCGCGGATCTTCTCCCGTCGTA – 3′ and cloned into EcoRV digested pZero2.1 to yield pZeTA-PrP. PrP sequences were released by NheI/NotI digestion and cloned into NheI/NotI digested pET-21a(+) to construct pX-PrP. pZ-PrP was made by cloning PCR amplified mouse PrP from pVPmPrP14, using as primers mPrP-FW2: 5′ – GGCAAGCTTAAAAAGCGGCCAAAGCCT – 3′ and mPrP-RV2: 5′ – GGCGCGGCCGCGGATCTTCTCCCGTCGTA – 3′, into EcoRV digested pZero2.1. pX-DnaK–PrP and pX-GroEL–PrP were constructed by releasing PrP from pZ-PrP following NotI/HindIII digestion and cloning into NotI/HindIII digested pXCK-K and pXCK-EL respectively.

2.4.4. ORF21 plasmids

The sequence of VZV ORF21 encoding for aa878–1038 was amplified from virus DNA using ORF21for: 5′ – GGGAATTCGGGTCTTG-CAAATGTAGAGATTT – 3′ and ORF21rev: 5′ – TGCGGCCGCAGGGT-CACTCCCACTTGTAT – 3′. The PCR product was cloned into pCR2.1-TOPO to yield pTOPO-ORF21. The ORF21 coding sequence was released by digesting with EcoRI and NotI and cloned into EcoRI and NotI digested pALEX (Panagiotidis and Silverstein, 1995) to yield pALEX-ORF21. pDnaK-ORF21 was constructed by cloning the XhoI/HindIII fragment from pTOPO-ORF21 into the XhoI/HindIII digested pXCK-K.

2.4.5. Co-chaperone plasmids

GrpE and *DnaJ* were amplified from N99 genomic DNA using the oligonucleotide primers GrpE-FW: 5′ – GCCCATGGGCGGAGAAATT-CATGAGTA – 3′, GrpE-RV: 5′ – GGGCTAGCTACGAAAGCAGAAATTAAGC – 3′ and DnaJ-FW: 5′ – GGGCTAGCAAAAATAATCGCCCTATAAA – 3′, DnaJ-RV: 5′ – GGCTCGAGGGAGGTTAGCGGGTCA – 3′. The two PCR products were cloned into EcoRV digested pZero2.1 to yield pZeTA-GrpE and pZeTA-DnaJ respectively. pX-GrpE was constructed by cloning an NcoI/EcoRI fragment from pZeTA-GrpE into NcoI/EcoRI digested pET-21d(+). *DnaJ* was released from pZeTA-DnaJ with NheI/XhoI and cloned into NheI/XhoI digested pX-GrpE to yield pX-GrpE-DnaJ. The GrpE/DnaJ operon was released as an EcoRV/XhoI fragment from pX-GrpE-DnaJ and cloned into EcoRV/SalI digested pLysE to yield pXCK-E/J. *GroES* was amplified from N99 genomic DNA using the oligonucleotide primers: GroES-FW: 5′ – GGCCATGGCAAAGGAGAGTTATCAATG – 3′ and GroES-RV: 5′ –