Protein Expression and Purification 63 (2009) 40–49

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Mocr: A novel fusion tag for enhancing solubility that is compatible with structural biology applications

James DelProposto [a], Chinmay Y. Majmudar [b], Janet L. Smith [c], William Clay Brown [a,*]

[a] High-throughput Protein Lab, Center for Structural Biology, Life Sciences Institute, University of Michigan, 210 Washtenaw Avenue, Ann Arbor, Michigan 48109, USA
[b] Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA
[c] Life Sciences Institute and Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA

ARTICLE INFO

Article history:
Received 11 June 2008
and in revised form 18 August 2008
Available online 12 September 2008

Keywords:
Fusion protein
Solubility enhancement
Purification handle
Passenger protein

ABSTRACT

A persistent problem in heterologous protein production is insolubility of the target protein when expressed to high level in the host cell. A widely employed strategy for overcoming this problem is the use of fusion tags. The best fusion tags promote solubility, may function as purification handles and either do not interfere with downstream applications or may be removed from the passenger protein preparation. A novel fusion tag is identified that meets these criteria. This fusion tag is a monomeric mutant of the Ocr protein (0.3 gene product) of bacteriophage T7. This fusion tag displays solubilizing activity with a variety of different passenger proteins. We show that it may be used as a purification handle similar to other fusion tags. Its small size and compact structure are compatible with its use in downstream applications of the passenger protein or it may be removed and purified away from the passenger protein. The use of monomeric Ocr (Mocr) as a complement to other fusion tags such as maltose-binding protein will provide greater flexibility in protein production and processing for a wide variety of protein applications.

Due to its low cost, compatibility with automation and ease of scale-up, *Escherichia coli* remains the most widely used host for high-throughput protein production [1–3]. A major hurdle for heterologous protein production in *E. coli* is the formation of insoluble aggregates. This problem is commonly addressed through the use of fusion tags to enhance solubility [4–6]. Comparative studies of the effectiveness of fusion tags have shown the maltose-binding protein (MBP)[1] to be one of the best at solubilizing passenger proteins [7,8]. The properties that make a fusion tag capable of enhancing solubility are not fully understood, although the acidity of the fusion tag is often correlated with this capability [9,10]. Due to MBP's solubilizing capability and its affinity for amylose, which allow it to be used as an affinity handle, vectors containing MBP fusion tags have been developed for use in high-throughput cloning and expression [11].

Although MBP is quite effective in solubilizing its passenger proteins during expression, a number of problems have been identified with its use that occurs during purification and processing of the fusion. MBP fusions do not always bind to amylose resin and so a His tag is commonly added to facilitate affinity purification using a metal chelating resin [12]. Many proteins that are soluble when fused to MBP have been observed to precipitate when the MBP-fusion is cleaved [11]. Additionally, the incomplete removal of MBP from the passenger protein after cleavage of the fusion [11] may interfere with downstream applications such as NMR or crystallization.

We sought to develop a new fusion partner with solubilizing capabilities similar to MBP while avoiding or reducing the problems found in MBP fusion purification and processing. The bacteriophage T7 0.3 protein (Ocr for ability to *o*vercome *c*lassical *r*estriction) is a 13.8 kDa, highly charged, very acidic (p*I* = 3.8) protein [13]. It has an efficiently translated transcript and is tolerated at high levels in the cell. There are no cysteines in the Ocr protein [13]. The protein is completely soluble even when expressed to high level and is soluble in 95% ethanol [14]. The Ocr protein may be separated to high levels of purity from *E. coli* using DEAE resins [14]. These properties suggested it could make a robust fusion partner to promote target protein production and solubility.

In its native state, the Ocr protein forms a dimer, which could potentially foster aggregation of a fused partner. The crystal structure of the Ocr dimer revealed a small, hydrophobic subunit interface [15]. Here we report two amino acid substitutions that disrupt dimer formation and that stabilize the monomeric form of the protein, which we call Mocr, for *m*onomeric Ocr. We further characterize Mocr as a fusion partner that displays solubilizing activity with problematic passenger proteins.

* Corresponding author.
E-mail address: wclayb@umich.edu (W.C. Brown).

[1] *Abbreviations used:* Mocr, monomeric Ocr; MBP, maltose-binding protein; LIC, ligation-independent cloning; TEV, tobacco etch virus; GST, glutathione-*S*-transferase; MAV-1, mouse adenovirus type 1; MNV, mouse norovirus.

1046-5928/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2008.08.011

Materials and methods

Construction and cloning of mocr

A synthetic sequence of the *ocr* gene was created to yield oligos with minimal dimerization potential and self-annealing. The only purpose of the silent mutations contained in the synthetic sequence was to improve the efficiency of the synthetic gene construction. This sequence also incorporated mutations that change the amino acid sequence. Codon 53 has been changed from TTT to CGT and codon 77 has been changed from GTA to GAC (Fig. 1A). These codon changes gave rise to the following amino acid changes: F53R and V77D. The amino acid sequence of Mocr is shown in Fig. 1B.

The gene was then constructed through a combination of sequential oligo pair annealing and ligation and PCR amplification. For sequential oligo pair annealing and ligation, the oligos were resuspended in water to 20 μM. Each oligo was the phosphorylated in a 20 μl reaction containing 2 μl ATP (10 mM), 2 μl 10× ligation buffer (NEB), 12.5 μl of the oligo, 2.5 μl H_2O and 1 μl of polynucleotide kinase. Reactions were incubated at 37 °C for 1 h, then at 95 °C for 10 min. Complementary oligo pairs were mixed and slow cooled to 37 °C to create double-stranded fragments. Adjacent fragments were then mixed in equal volume and slow cooled to 4 °C. At 4 °C additional ATP was added along with 2 U of T4 DNA ligase and incubation was continued at 4 °C for 1 h. Mixing of adjacent fragments at 37 °C followed by slow cooling and ligation was continued until all the fragments had been mixed. The resulting fragment was then PCR amplified using outside primers to add a BglII restriction site to the 5′-end and a KpnI site to the 3′-end. The resulting fragment was gel purified and cloned into pMCSG7 [16] digested with BglII and KpnI.

Cloning wild-type ocr

Bacteriophage T7 DNA was obtained from ATCC (BAA-1025-B2). Phage were produced by transfection of strain HMS174 (ATCC 47011) with resuspended DNA. The cells were grown until lysis. This broth was used to infect two 5 ml cultures of HMS174. After lysis the cell debris was spun out. The phage was precipitated by the addition of PEG with incubation at 4 °C overnight followed by centrifugation. The pellets were resuspended in a total of 1 ml of PBS. A 200 μl aliquot was extracted with phenol/chloroform and the DNA was then precipitated by addition of ammonium acetate and ethanol. This DNA was used as PCR template in reactions using outside primers to add a BglII restriction site to the 5′-end and a KpnI site to the 3′-end. The resulting fragment was gel purified and ligated with pMCSG7. Positive clones were identified by PCR and confirmed by DNA sequencing.

Gel filtration of Ocr and Mocr

Cultures (250 ml) in terrific broth (TB: 6 g tryptone, 12 g yeast extract, 4% (20 ml) glycerol, 1.15 g KH_2PO_4, 6.25 g K_2HPO_4 in 500 ml distilled water) were grown in 1 L flasks at 37 °C, 250 rpm to an OD at wavelength 600 of approximately 1. The temperature was reduced to 20 °C and after equilibration at this temperature for 1 h the cultures were induced by addition of 200 μM IPTG. Incubation was continued at 20 °C overnight (18 h). Cultures were centrifuged and pellets were frozen at −80 °C. Cell pellets (5–6 g) were resuspended in 40 ml PBS with 0.1 mg/ml lysozyme and benzonase and lysed by sonication. Lysate was centrifuged at 20,000g for 1 h. Soluble fraction was batch bound overnight to 2 ml Ni-NTA agarose from Qiagen. Resin was washed with 20 mM imidazole in PBS and eluted off with 250 mM imidazole in PBS. The 10–15 ml eluate was dialyzed in 50 mM Tris pH 8.0, 150 mM NaCl, 0.1 mM EDTA, and 1 mM DTT. Gel filtration was performed using a 120 ml Superdex 75 column on an Akta Explorer FPLC. The running buffer composition was the same as the dialysis buffer.

DEAE chromatography

Soluble fraction was batch bound to 2 ml DEAE resin for 2 h. DEAE chromatography was performed using DEAE Ceramic HyperD F resin from Pall Life Sciences. The resin was washed with 50 mM sodium phosphate and varying concentrations of NH_4Cl, from 50 to 400 mM, and then eluted with 50 mM sodium phosphate and varying concentrations of NH_4Cl, from 400 mM to 1 M.

High-throughput construct design, cloning, expression and purification

Design

Protein sequences were analyzed for ordered and disordered regions using server based programs: Foldindex (http://bioportal.weizmann.ac.il/fldbin/findex) and DisEMBL (http://

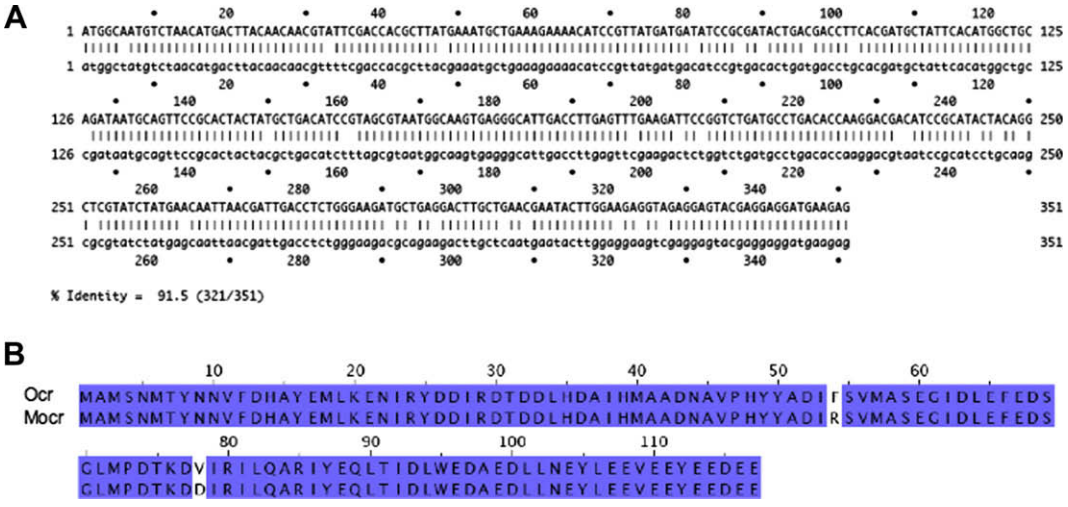

Fig. 1. (A) Strider alignment of the DNA sequence of the synthetic *mocr* gene (top) and the native *ocr* gene (bottom). The silent mutations in the synthetic sequence were introduced to reduce self-annealing or dimerization in the oligos used to construct the synthetic gene. (B) A Jalview (www.jalview.org) alignment of the Ocr amino acid sequence (top) with the mutant Ocr (Mocr) amino acid sequence (bottom) is shown. Identical amino acids are highlighted. The Mocr sequence contains two mutations, F53R and V77D. These were made to disrupt the dimer interface.