Protein Expression and Purification 84 (2012) 188–194

Contents lists available at SciVerse ScienceDirect

Protein Expression and Purification

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

Codon optimization for high level expression of human bone morphogenetic protein – 2 in *Escherichia coli*

Debbie S. Retnoningrum [a,*], H.T. Pramesti [a], P.Y. Santika [a], O. Valerius [b], S. Asjarie [a], T. Suciati [a]

[a] School of Pharmacy, Institut Teknologi Bandung, Jalan Ganesha No. 10, Bandung 40132, Indonesia
[b] Department of Molecular Microbiology and Genetics, Georg-August University of Göttingen, Grisebachstr. 8, Göttingen 37077, Germany

ARTICLE INFO

Article history:
Received 9 March 2012
and in revised form 28 May 2012
Available online 9 June 2012

Keywords:
Codon-optimized ORF
Escherichia coli
High level expression
Human bone morphogenetic protein-2
TBIO

ABSTRACT

Codons in the open reading frame (ORF)[1] encoding for human bone morphogenetic protein-2 (hBMP-2) were optimized to reach high level expression in *Escherichia coli*. The optimization was done by the computer programs DNA works and DNA Star according to Thermodynamically Balanced Inside Out (TBIO) approach. The ORF consisting of 342 base pairs (bp) was assembled using two-steps Polymerase Chain Reaction, cloned into a pGEM-T vector with a mutation rate of 6.38 bp per kb and transformed into *E. coli* JM109. After a DNA sequence confirmation, mutation-free ORF was subcloned into pET32b and transformed into *E. coli* BL21(DE3). The rhBMP-2 was produced as a thioredoxin-his-tag fusion protein at relatively high level, approximately 60% of total intracellular proteins as inclusion bodies (IB), with a yield of 1.39 g per liter culture. Solubilization of IB gave soluble monomer rhBMP-2 with a recovery of 13.6% and refolding of soluble rhBMP-2 produced dimeric forms with a yield of 8.7%. The size and identity of the purified rhBMP-2 was confirmed by nano-LC–MS/MS2 analysis. Our work demonstrates for the first time that by using TBIO approach, a codon-optimized ORF encoding for rhBMP-2 protein can be expressed at high level in *E. coli* expression system.

Introduction

Bone morphogenetic protein-2 (BMP-2) is member of a transforming growth factor β that includes more than 20 structurally-related bone growth factors. The protein induces bone formation and it has been used to treat fractured bones and other bone-related diseases [1,2]. Human BMP-2 (hBMP-2) is a glycoprotein and normally found as a homodimer. Dimerization is facilitated by a disulfide bridge formed between the monomer, which contains three intrachain disulfide bridges arranged in a cystine knot motif [3]. The protein is expressed as a 396 amino acids long precursor, which upon maturation is proteolytically cleaved to 114 amino acids long hBMP-2 [4].

Owing to its significance in therapeutic treatment, attempts for hBMP-2 production using heterologous expression systems have been reported. Direct hBMP-2 isolation from bone is economically unfeasible process because of very low yields (1–3 $\mu g\,kg^{-1}$), of complicated purification scheme, and most importantly of safety reason (potential risk of contaminants associated with bone donor) [5]. Heterologous production of biologically active recombinant hBMP-2 (rhBMP-2) employing mammalian systems [6,7] suffers from post-translational problems that lead to imperfect monomer processing and from low yields (nanograms amount). Expression of rhBMP-2 using bacterial systems may overcome those problems therefore it is an attractive alternative.

Overexpression of rhBMP-2 in *Escherichia coli* system has previously reported [8–16]. However, production of rhBMP-2 in *E. coli* is hampered by the formation of inclusion bodies (IB), except for being coexpressed with thioredoxin [13]. Two induction systems, IPTG or temperature shift, were generally used to regulate the gene expression. Using this expression/induction system, the level of expressions was variable with the highest yield reported to be 750 mg of rhBMP-2 dimer per liter of high density culture [8,9,12,14,16]. As rhBMP-2 is obtained as IB, is to be solubilised, and subsequently refolded to recover biologically active protein. The yield reported from various refolding attempts varies greatly [10,16]. IB formation may originate from the inability of *E. coli* to efficiently translate rare codons such as CTT, TTG, CTC and AAG, which are present in the rhBMP-2 ORF. One strategy to overcome this problem is by optimizing the codon usage, as demonstrated by Ihm et al. where twelve oligonucleotides were used to assemble three DNA fragments of the rhBMP-2 ORF construct [13].

Here we present the construction of a codon-optimized ORF encoding mature rhBMP-2 using Thermodynamically Balanced Inside Out (TBIO) principle [17], which was directed to increase

* Corresponding author. Fax: +62 022 250 4852.
E-mail address: retnoningrum@indo.net.id (D.S. Retnoningrum).

[1] *Abbreviations used*: ORF, open reading frame; TBIO, thermodynamically balanced inside out; bp, base pairs; IB, inclusion bodies; hBMP-2, human BMP-2; Tm, melting temperatures; pNPP, para nitro phenol phosphate substrate system; CAI, codon adaptation index.

1046-5928/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.pep.2012.05.010

its level of expression in *E. coli*. The hBMP-2 ORF was synthesized using 8 nucleotides and overproduced in *E. coli* BL21(DE3). The rhBMP-2 is obtained as IB and biologically active protein was recovered after solubilization and refolding steps. After one-step purification in an affinity column, the identity of rhBMP-2 was confirmed by nano-LC mass spectrometry (nano-LC–MS/MS2). Here, for the first time, the success on modifying codons in the ORF encoding hBMP-2 using TBIO method is reported. The rhBMP-2 expressed from codon-optimized ORF constituted 60% of total bacterial proteins.

Materials and methods

Bacterial strains, plasmids and culture media

E. coli strains, JM109 for the cloning, BL21(DE3) and Rosetta gami B(DE3) for protein expression are maintained at the laboratory of Pharmaceutical Biotechnology, the School of Pharmacy, Institut Teknologi Bandung. The plasmids employed for cloning and overexpression were pGEM-T (Promega) and pET32b (Novagene), respectively. The bacteria was grown at 37 °C in Luria Bertani (LB) medium containing 100 μg ml^{-1} of ampicillin and the protein expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.05 mM [18].

Construction of codon-optimized hBMP-2 ORF

The nucleotide sequences of the hBMP-2 ORF (GenBank Accession Number NM001200) was obtained from NCBI data base (http://www.ncbi.nlm.nih.gov/). Only the region that encodes 114 amino acids long mature hBMP-2 (residue 282–396) was selected for the codon optimization. The codon optimization was performed using the on-line program DNAWork ver2.4 (http://helixweb.nih.gov/dnaworks) in consultation with the table for *E. coli* codon usage (http://www.sci.sdsu.edu). For the insertion to the expression vector, *Bam*HI site was introduced to the 5′ end whilst *Xho*I site and stop codon to the 3′ end of the ORF. The design and assembly of ORF were based on TBIO bidirectional synthesis principle according to that of described by Gao et al. [17]. The criteria for oligonucleotides representing sense and antisense strands were 60–71 bases in size with overlapping sequences of about 24 bases, G + C content of ~50% and melting temperatures (*Tm*) of overlapping oligonucleotides were approximately 59 °C. The *Tm* was optimized by the program DNA Star (DNA Star Inc), which was also employed to design the two outer primers for DNA amplification. All oligonucleotides were PAGE grade and chemically synthesized (Eurogentec Ait, Singapore).

Synthesis and amplification of ORF was performed in a two-step polymerase chain reaction (PCR). The first step PCR was done according to TBIO method for ORF synthesis whilst the second step was for ORF amplification. In the ORF synthesis, the PCR was subjected to hot-start PCR with the following conditions: two minutes of pre-denaturation step at 95 °C, 25 cycles of a sequence of 30 s at 95 °C, 45 s at 54 °C, and one minutes at 72 °C, and finally ten minutes of post-elongation step at 72 °C. The PCR reaction mixture consisted of 1.5 μL of 10 mM dNTPs, 2.5 U of *Pfu* Turbo polymerase (Stratagene, La Jolla, CA, USA), 5 μL 10x *Pfu* buffer and primer pairs BMP4/BMP5, BMP3/BMP6, BMP2/BMP7 and BMP1/BMP8 to final concentrations of 40 nM, 60 nM, 120 nM dan 200 nM, respectively. The volume was adjusted to 50 μL with water. The assembled ORF obtained was immediately amplified by PCR using outer primers without prior gel purification. The second PCR step was done with the following conditions: five minutes of pre-denaturation at 94 °C, 25 cycles of a sequence of one minutes at 94 °C, one minute at 52 °C, one minute at 72 °C, and finally ten minutes of post-elongation at 72 °C. For this second step, the PCR reaction mixture consisted of 2.5 μL DNA templates from the first-step PCR, 0.3 μL of *Taq* polymerase (MD Bio), 2.5 μL 10x *Taq* buffer, 1 μL forward and 1 μL reverse primers each of 25 pmol/μL, 0.5 μL 10 mM of dNTPs, 2.5 μL 25 mM of $MgCl_2$. The total volume was adjusted to 25 μL by adding appropriate amount of water. The sequences of oligonucleotides are presented in Table 1. The final product of the PCR was analyzed in a 1.5% agarose gel electrophoresis and subsequently purified from the gel using EZ-10 Spin Column DNA Gel Extraction Kit (Biobasic, Ontario, Canada).

Construction of the recombinant plasmid

The purified synthetic rhBMP-2 ORF was inserted to pGEM-T vector and then transformed into *E. coli* JM109 according to the manufacturer's protocol (Promega, Madison, WI, USA). The recombinant plasmids were isolated from the transformants grown on an agar plate containing ampicillin and characterized by single and double digestion using the restriction enzymes *Bam*HI and *Xho*I. The integrity of the insert was evaluated by a two-direction sequencing employing T7 promotor and SP6 terminator primers (Table 1) and the nucleotide sequences were aligned against its theoretical sequences using the on-line program JustBio (http://www.justbio.com/). One recombinant pGEM-T that carries no mutation was selected as the source for further subcloning into

Table 1
Nucleotide sequences of oligonucletides used in ORF assembly and amplification.

Primers	Nucleotide sequences (5′ → 3′)
ORF assembly sense strand (1–199)	
BMP1	GGATCCGCAGGCTAAACACAAGCAGCGTAAACGCCTGAAATCTTCTTGCAAGCGTCATCCG
BMP2	AATCTTCTTGCAAGCGTCATCCGCTGTATGTTGATTTCTCCGACGTGGGCTGGAACGACTGGATTGTGG
BMP3	AATCTTCTTGCAAGCGTCATCCGCTGTATGTTGATTTCTCCGACGTGGGCTGGAACGACTGGATTGTGG
BMP4	TCTACTGCCACGGTGAATGTCCGTTCCCGCTGGCGGATCACCTGAACTCTACCAACCACGCGATCGTTCAG
Antisense (200–358)	
BMP5	AAGCCTTCGGGATCTTGGAGTTAACGGAGTTAACCAGAGTCTGAACGATCGCGTGGTTGGTAG
BMP6	GTACAGCATAGAAATCGCGGACAGCTCGGTCGGAACACAGCAAGCCTTCGGGATCTTGGAGTTAAC
BMP7	ATGTCCTGGTAGTTCTTCAGCACCACTTTTTCGTTTTCATCCAGGTACAGCATAGAAATCGCGGACAGC
BMP8	CTCGAGCTAACGACAACCGCAACCTTCCACCACCATGTCCTGGTAGTTCTTCAGCACCAC
ORF amplification	
BMPfor	GGATCCGCAGGCTAAACACAAG
BMPrev	CTCGAGCTAACGACAACCGCA
ORF sequencing	
SP6 terminator	GATTTAGGTGACACTATG
T7 promoter	AATACGACTCACTATAGG