Biota Vol. IX (1): 37-40, Februari 2004
ISSN 0853-8670

Construction of 798-bp Artificial Open Reading Frames (ORF) Encoding Random Sequences Proteins

Konstruksi *Open Reding Frames* (ORF) Artifisial Berukuran 798-bp yang Menyandi Protein dengan Urutan Asam Amino Acak

Irfan Dwidya Prijambada[1,2*], Khotibul Umam Al-Awally[2], Muhammad Saifur Rohman[2] and I Wayan Artama[2]

[1] *Laboratory of Microbiology, Faculty of Agriculture, Gadjah Mada University, Sekip Unit 1, Yogyakarta 55581*
[2] *Graduate School of Biotechnology, Gadjah Mada University, Jl. Teknika Utara, Barek, Yogyakarta 55581, Indonesia*
[*] *Penulis untuk korespondensi*

Abstrak

Penyusunan pustaka dari *open reading frames* (ORF) buatan yang tersusun atas 798 bp (pasangan basa), 576 di antaranya tersusun secara acak, yang mampu menyandi 266 asam amino telah berhasil dilakukan. Dalam upaya penyusunan tersebut diperoleh 32 transforman, lima di antaranya membawa ORF buatan. Dari kelima transforman yang membawa ORF buatan tersebut, hanya satu transforman yang mampu berekspresi dan menyandi suatu protein. Protein yang dihasilkan memiliki ukuran 17 kDa, berukuran lebih kecil daripada ukuran yang diharapkan yaitu 29 kDa.

Kata kunci: ORF buatan, urutan asam amino acak

Diterima: 22 April 2003, disetujui: 10 September 2003

Introduction

Knowledge of proteins has been obtained from the observation of natural proteins. As natural proteins have a long history with the influence of natural evolution, in which some of their properties must have been specialized due to evolutionary constraints, a biased knowledge with regard to the properties of proteins may be obtained. Investigation of synthetic proteins beyond the influence of natural evolution has been started. A library of artificial open reading frames encoding 141 amino acid residue proteins with random sequences has been successfully constructed (Prijambada *et al.*, 1996). The probability of the artificial open reading frames to be expressed in *Echerichia coli (E. coli)* to produce random proteins was about 50%, and the probability of soluble random proteins to be acquired was about 20% (Prijambada *et al*., 1996). Study on two soluble random protein revealed that they form an oligomeric an compact globular structure, and have low bu distinct esterase activity (Yamauchi *et al* 1998). This success challenging to furthe study on some features of proteins. In fact there are numerous natural proteins havin more amino acid residues than those encode by the artificial open reading frames. In thi work, 798-bp artificial open reading frame containing 576-bp DNA fragments wit random sequences that were expected t express 266-amino acid residues proteins hav been prepared. Expression of the artificial ope reading frames and the state of the expresse random protein in the cells of *E.coli* wer examined.

Materials and Methods

Bacterial strain and plasmids

The bacterial strain and plasmids used were *E. coli* KP3998 (F$^-$ *hsdS20* (r_B^- m_B^-) *ara-14 proA2 lacI*^{q} *galK2 rpsL20 xyl-5 mlt-1 supE44* □$^-$) (Miki *et al*., 1987), pUCIL, pEOR, a random DNA library (R1MIX) (Prijambada *et al*., 1996), and a newly constructed random DNA library (R2MIX). *Escherichia coli* KP3998 was generously given by Dr. Takeyoshi Miki (Kyushu University, Japan).

DNA manipulation

Preparation of plasmid DNA, enzyme reactions, and transformation of *E. coli* cells were carried out as described by Sambrook *et al*. (Sambrook, *et al*., 1989).

Construction of new random DNA library (R2MIX)

R1MIX, a mixture of hybrid plasmids containing 1,2 or 3 units of randomized portion of DNA, was digested with *Xho*I and *Bgl*II. The 119-bp, 227-bp, and 319-bp DNA fragments containing 1, 2, and 3 units of the randomized portion, respectively, were isolated by polyacrylamide gel electrophoresis. R1MIX was also digested with *Xho*I and *Kpn*I. The 125-bp, 233-bp, and 325-bp DNA fragments containing 1, 2, and 3 units of the randomized portion, respectively, were also isolated by polyacrylamide gel electrophoresis. The isolated *Xho*I-*Bgl*II and *Xho*I-*Kpn*I fragments were then ligated with pUCIL that had been digested with *Bam*HI and *Kpn*I. The ligated DNA was introduced into *E. coli* KP3998. All of the transformants (about 1.2 x 10^4 colonies) that grew on the plates containing 50 □g/ml ampicillin were collected. Plasmid isolation was done from the collected cells. The obtained mixture of hybrid plasmids contains 2, 3, 4, 5 or 6 units of the randomized portion DNA was named R2MIX.

Construction of hybrid plasmids for random protein expression

R2MIX was digested with *Sal*I and *Bgl*II, and the 364-bp DNA fragments containing 3 units of the randomized portion were isolated by polyacrylamide gel electrophoresis. R2MIX was also digested with *Sal*I and *Kpn*I and also subjected to polyacrylamide gel electrophoresis to obtain the 370-bp DNA fragments containing 3 units of the randomized portion. The *Sal*I-*Bgl*II and *Sal*I-*Kpn*I fragments were ligated with pEOR that had been digested with *Bam*HI and *Kpn*I. The digested pEOR vector maintains the P_{tac} promoter, an epitope tag of the first 11 codons for the T7 gene 10 protein (Studier *et al.,* 1990) and the stop codons for all the three frames. The constructed protein expression vectors containing 6 units of the randomized portion were then used to transform *E. coli* KP3998 for the production of random proteins.

Identification of transformants, expression and in vivo solubility of random proteins

E. coli KP3998 cells grew on an LB agar containing 100 μg/ml ampicillin were grown at 37^{o}C on LB broth containing 100 μg/ml ampicillin. Plasmid DNA was then isolated and digested with *Kpn*I. Plasmid having longer DNA than pEOR vector was judged to be a constructed one, which has an insertion.

E. coli KP3998 cells harboring the constructs were grown at 37^{o}C on 2xTY broth containing 100 μg/ml ampicillin. IPTG was added to a final concentration of 1 mM to the culture with OD_{600} of 0.6-0.8. IPTG induction was carried out for two hours. Proteins in the cells after IPTG induction were analyzed by SDS-PAGE (Sambrook *et al.,* 1989). Cells harboring pEOR serve as a negative control. A detected additional distinct band was judged to be the random protein produced by cells after IPTG induction.

Results and Discussion

The strategy in constructing the artificial open reading frames were designed to insert 730-bp fragment DNA's having random sequences in between the *Bam*HI and *Kpn*I sites of the expression vector resulting in 3.43 Kbp-plasmids. If no deletion occurred during the construction, the open reading frames are expected to encode random proteins with 266 amino acid residues. The proteins include fixed amino acid sequences of 17 residues at the N terminal, five residues between the first and second random sequences, two residues between the second and third random sequences, 28 residues between the third and fourth random sequences, one residue between the fourth and fifth random sequences, sixth residues between the fifth and sixth random sequences, and 17 residues at the C terminal. The first, second, fourth, and fifth random amino acid sequences were composed of 3 residues, while the third and sixth were 3 residues.

The hybrid plasmids containing th artificial open reading frames were introduce into *E. coli* KP3998 cells. While competenc of the competent *E. coli* KP3998 cells wa about 1.5 x 10^6/g intact pEOR, only 3 ampicillin-resistant transformants wer obtained on LB agar containing 100 □g ampicillin/ml. This low frequency o transformation may be caused by the multipl occurrence of direct and inverted repeats in th inserted DNA fragments that produce destabilizing effect on the hybrid plasmids Hence, the transformants were examined fo the existence of the hybrid plasmids (Fig.1)

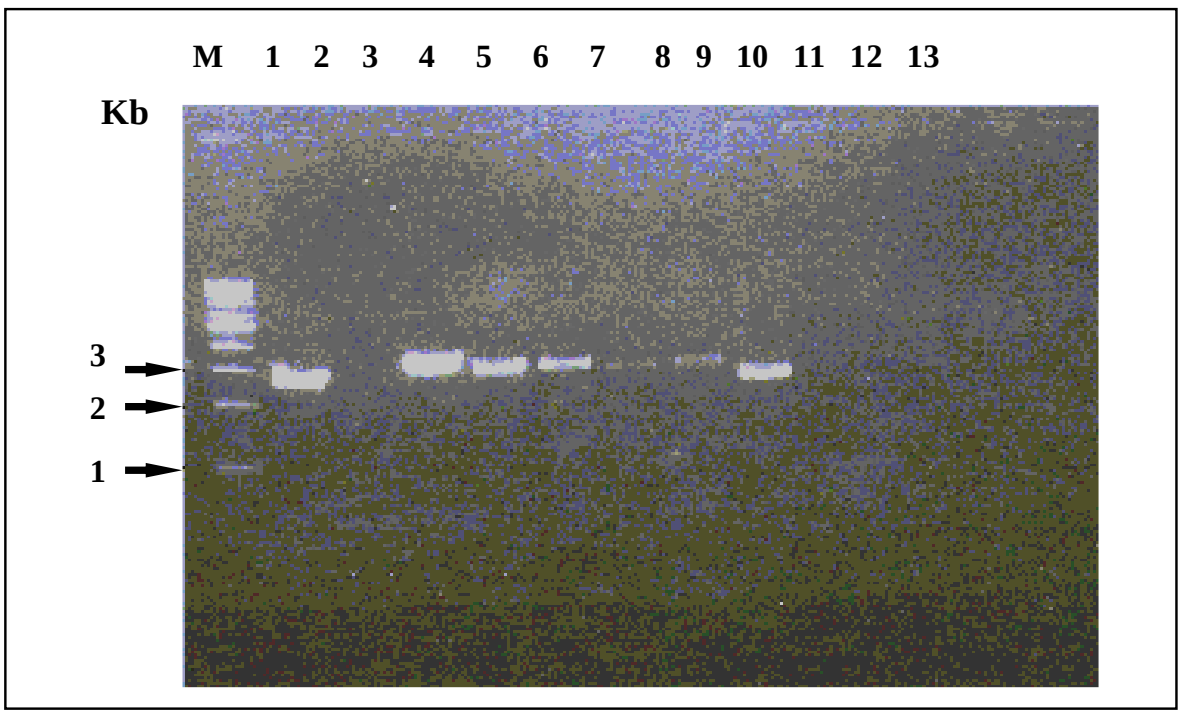

Fig.1. Plasmid DNA isolated from the transformants electrophoresed on 0.8% agarose gel. Plasmids were linearized using *Kpn*I. Lane M, 1-kbp ladder as a marker; lane 1, pEOR, and lanes 2 – 13 are transformants**.**

Out of 32 transformants, only five transformants bear the 3.43 Kbp-plasmids. The five transformants harboring these plasmids were cultivated and subjected to IPTG induction as described. Result of the SDS-PAGE analysis was shown in Fig.2. It showed that only one transformant clearly produces IPTG-inducible random protein. The IPTG-inducible protein has a molecular weight o only 17 kDa.

The solubility of this expressed random protein was then examined. Cells harveste after IPTG induction were disrupted b sonication. The supernatant and the precipitat obtained after centrifugation (12 000x*g* for 1 miniutes) of the disrupted cells were subjecte

to SDS-PAGE. Proteins detected in supernatant were evaluated as soluble proteins. The IPTG-inducible 17 kDa protein was found in the precipitate (data not shown). Therefore, the 17 kDa random protein was evaluated as an insoluble protein.

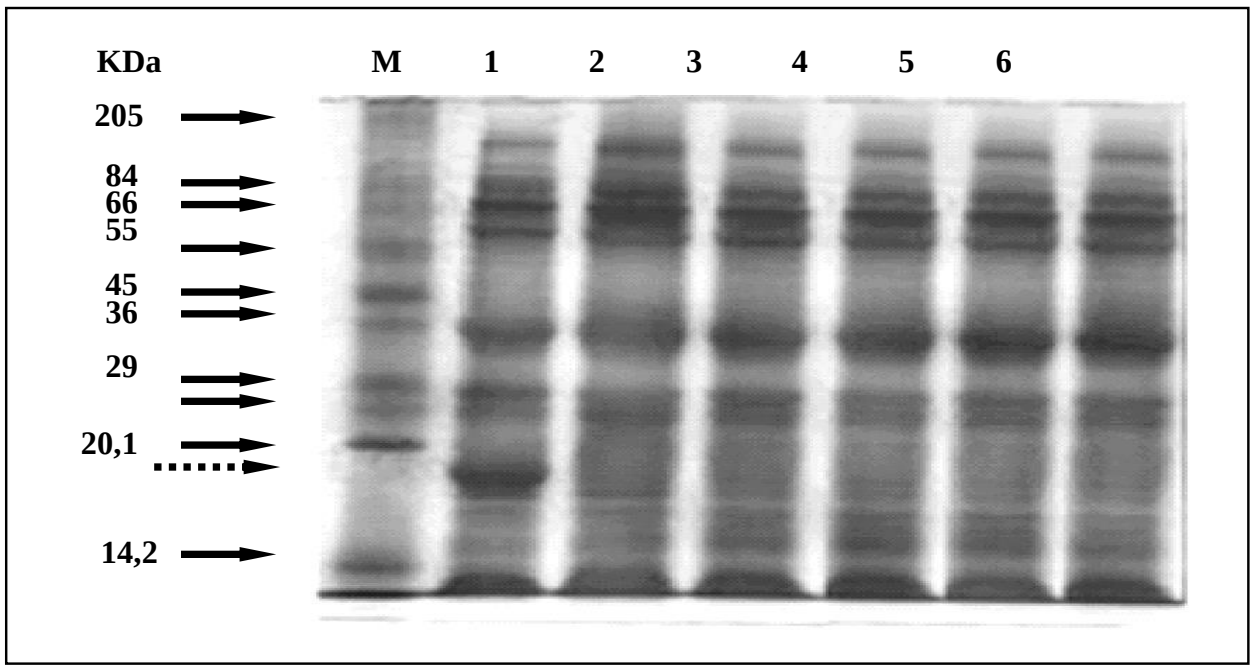

Fig. 2. Expression of artificial open reading frames in plasmids harbored by the transformants. The production of random proteins in the cells taken after IPTG induction were analyzed by SDS-PAGE. Proteins were stained with Coomassie brilliant blue R250. Lane M: marker, lanes 1 – 5: proteins from transformants cells, lane 6: proteins from cells harboring pEOR.

Results of the previous work on artificial open reading frames with random DNA sequences showed that about 50% of the introduced artificial open reading frames were expressed in *E. coli* KP3998, and about 20% of the expressed random protein were soluble (1). The 14.75 kDa and 14.73 kDa random proteins were found to form oligomeric structures (2). Despite the insertion of 730-bp fragments that was expected to encode 29 kDa proteins, the expressed random protein in this study has a molecular weight of 17 kDa. This results suggest that there is a length limitation for a protein to be expressed as a single polypeptide. Therefore, the polypeptides need to form an oligomer to produce a protein with higher density.

Acknowledgement

This work was supported in parts by a grant (012/PDB-II/URGE/1997) from the Directorate General of Higher Education, the Ministry of National Education, the Republic of Indonesia .

References

Miki, T., Yasukochi, T., Nagatani, H. *et al*., 1987. Construction of a plasmid vector for the regulatable high level expression of eukaryotic genes in *Escherichia coli*: an application to overproduction of chicken lysozyme. Protein Eng. 1: 327-332.

Prijambada, I.D., Yomo, T., Tanaka, F. *et al*.,. 1996. Solubility of artificial proteins with random sequences. FEBS Lett. 382: 21-25

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular Cloning, 2nd ed. Cold Spring Harbor. New York.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. *et al*., 1990. Use of T7 RNA polymerizes to direct expression of cloned genes. Methods Enzymol. 185: 60-89.

Yamauchi, A., Yomo, T., Tanaka, F. *et al.,* 1998. Characterization of soluble artificial proteins with random sequences. FEBS Lett. 421: 147-151