Korean J. Chem. Eng., **25**(5), 1125-1130 (2008)

Analysis and comparison of cloning methods for the preparation of repetitive polypeptides

Hun Su Chu*, Cha Yong Choi*, and Jong In Won,†**

*School of Chemical and Biological Engineering, Seoul National University, Gwanak-gu, Seoul 151-742, Korea
**Department of Chemical Engineering, Hongik University, Mapo-gu, Seoul 121-791, Korea
(*Received 3 January 2008 • accepted 27 February 2008*)

Abstract−Repetitive polypeptides, defined as a protein consisting primarily of tandemly repeated blocks of amino acid sequence, are widely used biomaterials. These repetitive polypeptides can be used in diverse biological fields including tissue engineering scaffolds, drug delivery systems, biomaterials, and DNA separation systems. The physical/chemical properties of the repetitive polypeptides can be improved by changing the composition of the repeated amino acid sequence. In this study, we introduced genetic methods for the production of repetitive polypeptides. By using recursive directional ligation (RDL) and controlled cloning method (CCM), multimerized genes were cloned and identified. Also, we compared the characteristics of recursive directional ligation (RDL) with those of controlled cloning method (CCM).

Key words: Repetitive Polypeptide, ELP, Recursive Directional Ligation (RDL), Controlled Cloning Method (CCM)

INTRODUCTION

Many natural proteins, such as the silks, elastin, and collagen, have excellent material properties, and, in principle, both these and entirely new protein materials can be produced from artificial genes [1]. In the case of spider silk, which can be easily seen in nature, it is so remarkably strong that its tensile strength is comparable to that of high-grade steel [2]. These proteins are composed of numerous repetitive amino acids, and sometimes they are called protein polymers. Because of the specific physical/chemical properties depending on the repetitive amino acid structures and patterns, these repetitive polypeptides are applied to many different biological fields [2-5].

When repetitive polypeptides are compared with natural common proteins, they have a remarkable advantage in that their charges, molecular weights, and hydrophilic properties can be strictly controlled by changing the composition of the repeated amino acid sequence. This concept can be applied to the design of new polymeric biomaterials with improved properties, such as better defined biocompatibility, biorecognition, biodegradation, and stimuli sensitivity. Considering biomedical applications such as drug delivery and tissue engineering, these genetically engineered biomaterials are particularly attractive for their ability to precisely place specific amino acids within a protein polymer [6].

Repetitive polypeptides, synthesized by genetic engineering, have several significant advantages over chemically synthesized polypeptides. First, it is possible to obtain specific materials with monodispersity. In the case of chemical synthesis, most of the synthetic polypeptides are inevitably polydisperse and, as a result, show molecular weight distribution. Because the molecular weight of the biomaterials influences pharmacokinetics and transport phenomena, most of the polypeptides used for biomedical applications are required to be monodisperse. Second, current methodologies of polymer synthesis by chemical process result in only the production of polymers with short oligo peptides. The longer the peptide chain length is, the less yield synthetic amino acids have. However, repetitive polypeptides synthesized by genetic engineering are less affected by these limitations.

Repetitive polypeptides synthesized by genetic engineering have a short history, as it was only in the 1990s that DNA which encodes $(Ala\text{-}Gly)_3$-Pro-Glu-Gly was first successful in multimerization by self-ligation [3]. Hence, Prince et al. investigated the methods of controlling repetitive length by modulating self-ligation level [7]. And in 1999, McMillan et al. used a specific restriction enzyme to produce polypeptides [8]. But all these methods involved some problems in that there were expressions of unnecessary polypeptides, limit of monomer selection, or low reproducibility. In 2002, two cloning methods, recursive directional ligation (RDL) and controlled cloning method (CCM), were introduced to solve these problems [9,10]. In the recursive directional ligation (RDL) method, a monomer DNA segment encoding a polypeptide sequence of interest is seamlessly self-ligated [9]. To make a dimer, a vector containing the DNA monomer is ligated with the isolated DNA monomer, and this process is repeated until the gene of a desired length is obtained. In controlled cloning method (CCM), a vector in which a specific oligomer is inserted is amplified with proper primers. And then the amplified monomer is doubled by the digestion and the ligation based on two restriction enzymes, *Eam*1104I, *Lgu*I (or *Sap*I). These enzymes have a very similar recognition site, but the only difference is that *Lgu*I needs one more base to recognize [10]. As a result, a DNA, which was digested with *Lgu*I, can also be digested with *Eam*1104I, but not vice versa. This is the core reason that these two endonucleases are used.

Here, we made several multimerized genes with various lengths for the production of repetitive polypeptides by genetic engineering. In detail, for the production of elastin-like polypeptides (ELPs) and neutral polypeptides which will be used in DNA separation [11-

†To whom correspondence should be addressed.
E-mail: jiwon@hongik.ac.kr
‡This work was presented at 13*th* YABEC symposium held at Seoul, Korea, October 20-22, 2007.

17], oligonucleotides were multimerized by recursive directional ligation (RDL) and controlled cloning method (CCM), respectively. We also compared the characteristics of recursive directional ligation (RDL) with those of controlled cloning method (CCM) as multimerized gene cloning methods [18].

MATERIALS AND METHODS

1. Materials

E. coli strains DH5α, DH10b, and plasmid pUC18 were purchased from Novagen (Madison, WI). *Taq* DNA polymerase, T4 DNA ligase, T4 DNA kinase, Alk.phosphatase (CIP), restriction endonuclease *Eco*RI, *Bam*HI were obtained from New England Biolabs (Beverly, MA). Restriction endonuclease *Lgu*I, *Bve*I, *Bpi*I, *Eam*1104I and all other restriction enzymes were purchased from Fermentas (Ontario, Canada). PCR purification kit and colony PCR premix were purchased from Core biotech (Seoul, Korea) and Genotech corp (Seoul, Korea). Synthetic oligonucleotides were supplied by Bioneer corp (Seoul, Korea).

2. General Methods

The procedures for growth of bacterial culture, DNA manipulation and transformation conditions were adopted from published literature or from instructions provided by manufacturers. DNA sequence analyses were performed by Macrogen corp. DNA electrophoresis was performed on 0.8%, 1.5%, 2% agarose gels with appropriate DNA Ladder (GeneRuler™, Fementas).

3. Preparation of Synthetic DNA Monomers for RDL

A single-stranded synthetic oligonucleotide (104 bases, Gene 1, see Fig. 1), which encodes four tandem repeats of -(VPGVG)-, was amplified via PCR with two oligonucleotide primers [TP1 (21nts): 5′-ATATAGAATTCACCTGCAATA-3′, TP3 (19nts): 5′-ATATTGGATCCGA AGACAT-3′] (see the underline in Fig. 1). TP1 (21nts) primer contains *Eco*RI recognition site (-GAATTC-), and TP3 (19nts) primer contains *Bam*HI recognition site (-GGATCC-). As a result, the PCR products can be digested by *Eco*RI and *Bam*HI restriction enzymes. Another single-stranded synthetic oligonucleotide (104 bases, Gene 2, see Fig. 1), which encodes -[VPGVG (VPGAG)$_2$VPGGG]-, was amplified via PCR with the same primers above. PCR was performed using a Mycycler™ (BioLAD) with 30 cycles at 95 °C for 1 min (denaturing), 58 °C for 1 min (annealing), and 72 °C for 2 min (elongation). PCR products were digested with *Eco*RI, *Bam*HI endonuclease (37 °C) to produce DNA monomers with cohesive ends. The DNA monomers were recovered via 2% agarose gel electrophoresis by using QIAquick Spin Columns (Qiagen). The isolated DNAs were purified by Spin Columns (Core biotech).

Fig. 1. A single-stranded synthetic oligonucleotide of gene 1, gene 2, and gene 3.

4. Preparation of a Synthetic DNA Monomer for CCM

A single-stranded synthetic oligonucleotide (99 bases, Gene 3, see Fig. 1), which encodes -(GA)$_2$S(GA)$_2$T(GA)$_2$S(GA)$_2$T-, was amplified via PCR with two oligonucleotide primers [FP (21nts): 5′-ATATAGAATTCCTCTTCAGGT-3′, BP (21nts): 5′-TATATGAATTCCTCTTCTA CC-3′] (see the underline in Fig. 1). These two primers contain *Eam*1104I recognition site (-CTCTTC-). Therefore the PCR products can be digested by *Eam*1104I restriction enzyme. PCR was performed by the same procedure described above. The PCR product was digested with *Eam*1104I endonuclease (37 °C) to produce DNA monomers with cohesive ends. The DNA monomers were recovered via 2% agarose gel electrophoresis and purified.

5. Preparation of a Modified-pUC18 Recipient Vector

To clone the DNA monomer by controlled cloning method (CCM),

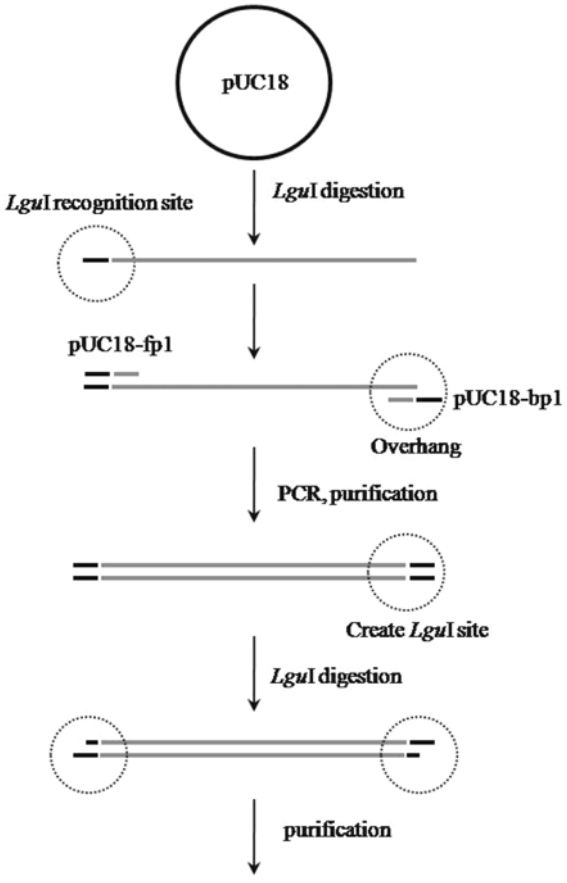

Fig. 2. Preparation of a modified-pUC18 recipient vector.

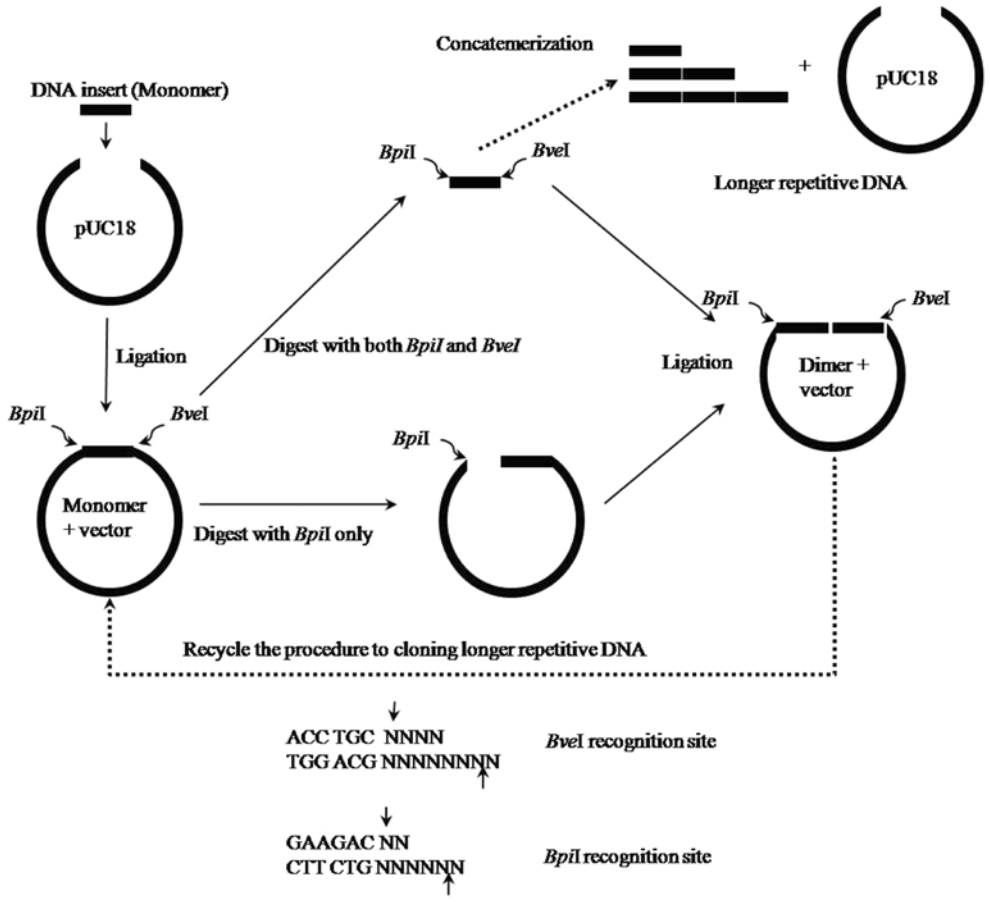

Fig. 3. Synthesis of DNA multimers by recursive directional ligation.

pUC18 cloning vector was digested with *Lgu*I endonuclease (37 °C). To generate two *Lgu*I recognition sites at the ends of pCU18 vector, the linearized pUC18 vector was amplified with two primers [pUC18-FP1: 5′-AGTTAGGTGGAAGAGCGCCCAATACGCAA-3′, pUC18-BP1: 5′-ATTACCTG AAGAGCTCCTCGCTCACT-GACTCGCTG-3′]. The PCR product was digested with *Lgu*I restriction enzyme (37 °C), followed by identified via 1% gel electrophoresis. Through these procedures, the modified-pUC18 recipient vector was prepared (see Fig. 2).

RESULTS AND DISCUSSION

1. Synthesis of DNA Multimers by Recursive Directional Ligation (RDL)

As stated above, two single stranded oligonucleotides were prepared for the multimerization by RDL. Gene 1, which encodes four tandem repeats of (VPGVG), and gene 2, which encodes [VPGVG $(VPGAG)_2$VPGGG], were designed and synthesized by chemical methods (Fig. 1). Each was composed of 104 bases, but the sequences were different to analyze the effects of sequence and length on the transition temperature and physicochemical properties of the expressed proteins (Gene 1 and gene 2 encode elastin-like polypeptides (ELPs)) [11,12]. The oligonucleotides were amplified via PCR with TP1 & TP3 primers, and PCR products were recovered via 2% agarose gel electrophoresis. Then PCR products were digested with *Eco*RI and *Bam*HI. A plasmid pUC18 was also digested with the same enzymes. After ligation and purification, the cloned monomer genes (gene 1 and gene 2) were identified by enzyme digestion (data not shown). For multimerization, gene 1, 2 monomers were digested again with *Bpi*I, *Bve*I, and inserts encoding only tan-

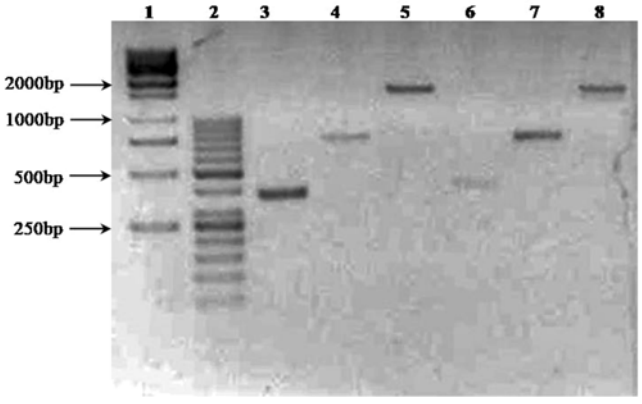

Fig. 4. Synthesis of repetitive oligonucleotides by RDL: Lane 1, 1 kb DNA ladder; Lane 2, 50 bp DNA ladder; Lane 3, 7mer of Gene 1 (420 bp); Lane 4, 14mer of Gene 1 (840 bp); Lane 5, 28mer of Gene 1 (1680 bp); Lane 6, 7mer of Gene 2 (420 bp); Lane 7, 14mer of Gene 2 (840 bp); Lane 8, 28mer of Gene 2 (1,680 bp).

dem repeats of specific amino acids (60 base pair each) were isolated (Fig. 3). The recombinant pUC18 plasmids containing gene 1, 2 monomers were also digested with *Bpi*I to make a linearized vector. The linearized vector and isolated monomer genes were ligated with T4 DNA ligase. Through these processes, DNA dimers were synthesized, and the same process was repeated until the gene of a desired length was obtained (Fig. 3). Several multimers with various lengths and different sequences were produced. These DNA multimers are shown in Fig. 4. Each multimer was digested with *Bpi*I, *Bve*I for identification of exact size, and analyzed via 0.8-2% agarose gel electrophoresis. Finally, DNA sequencing analyses were performed to verify the correct sequence. 7mer (140 amino acids), 14mer (280 amino acids), 28mer (560 amino acids) of each gene 1, 2 were synthesized and identified (Fig. 4). These multimer genes are going to be used to analyze ELP properties after protein expression.

2. Synthesis of DNA Multimers by Controlled Cloning Method (CCM)

Gene 3 encoding -[$(GA)_2S(GA)_2T(GA)_2S(GA)_2T$]- was used as a starting material to synthesize DNA multimers by CCM. The synthetic template was PCR-amplified with two primers to create a large amount of dsDNA. The PCR product was digested with *Eam*1104I to generate a DNA monomer. This monomer gene was isolated by 2% agarose gel electrophoresis and ligated with a prepared modified-pUC18 vector (4, room temperature for 3 h). Then the mixture was transformed into *E. coli* strain DH5α or DH10*b*. The transformed plasmids containing a DNA monomer were identified by colony-PCR or digestion with *Eam*1104I. The identified recombinant plasmids including a DNA monomer were PCR-amplified with two primers [pUC3: 5′-TTAATGAATCGGCCAACGCGC-3′ pUC4: 5′-TGAGCGAGGAACTCTTCAGGT-3′] to exchange one *Lgu*I restriction site to the *Eam*1104I restriction site (Fig. 5). This exchange of the recognition site is important for the synthesis of multimers by CCM. After the reaction was over, the PCR products had both *Lgu*I recognition site at one end and *Eam*1104I recognition site at the other end. Then half of the PCR products were digested with *Lgu*I endonuclease, and the others were digested with *Eam*1104I endonuclease (37 °C, Fig. 5). In the case of digestion with *Eam*1104I, both ends of the monomer gene become sticky ends. However, in

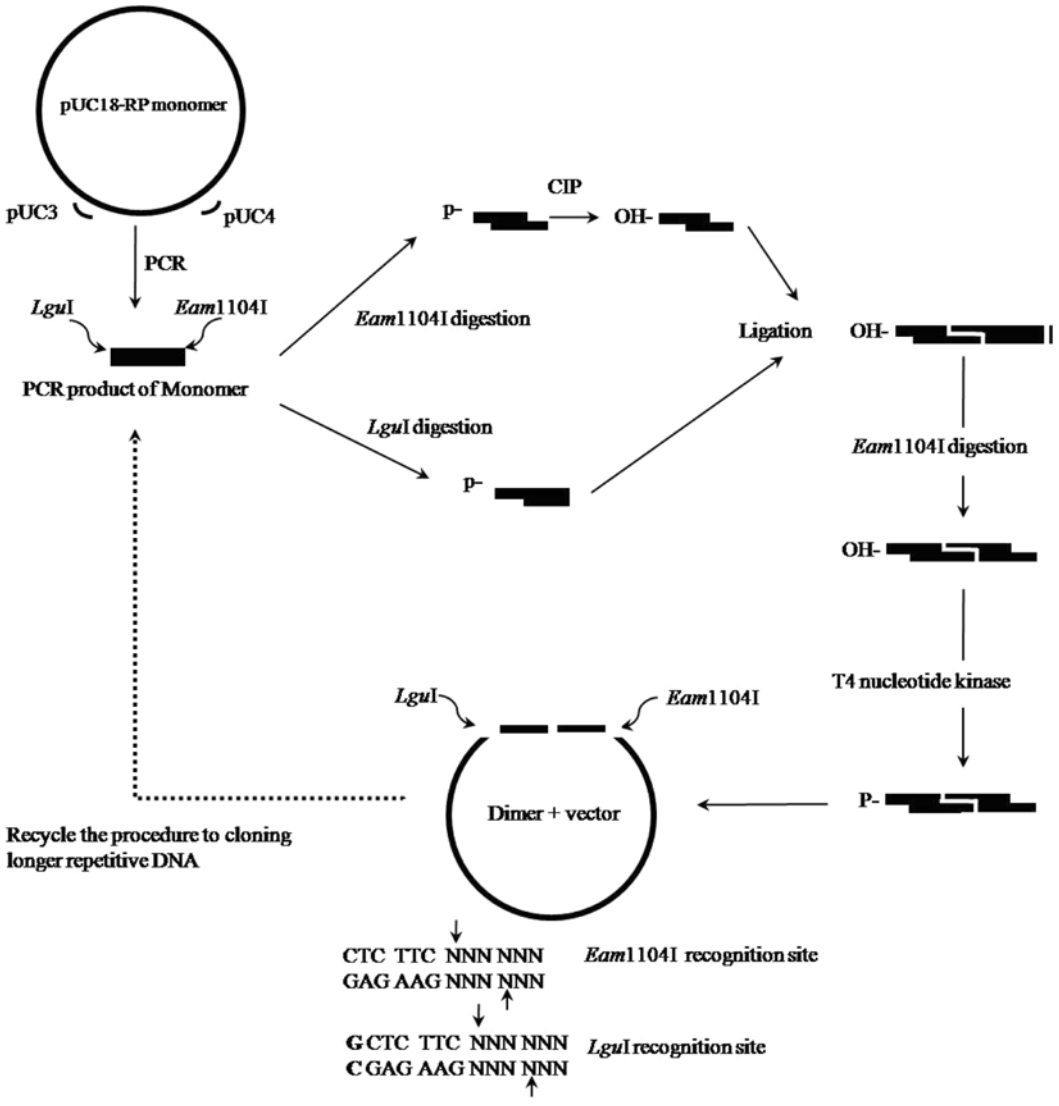

Fig. 5. Synthesis of DNA multimers by CCM.

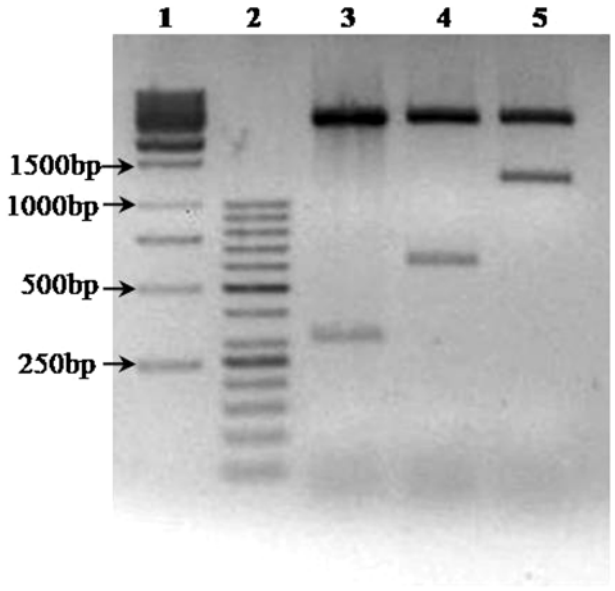

Fig. 6. Synthesis of repetitive oligonucleotides by CCM: Lane 1, 1 kb DNA ladder; Lane 2, 50 bp DNA ladder; Lane 3, 5mer of Gene 3 (300 bp); Lane 4, 10mer of Gene 3 (600 bp); Lane 5, 20mer of Gene 3 (1,200 bp).

the case of digestion with *Lgu*I, only one side of the monomer becomes sticky end. Then the monomer, digested with *Eam*1104I, was reacted with Alk.phosphatase (CIP, 37 °C, 2 h) to prevent self-ligation. After purification, the product was ligated with the prepared monomer, which was digested with *Lgu*I endonuclease (37 °C, 2 h). After purification, the double-sized product (a dimer) was incubated with T4 nucleotide kinase (37 °C, 2 h) to exchange OH^- end to P^-. This dimer was ligated with the modified-pUC18 recipient vector, and reaction mixture was transformed into the competent cell. Transformants were identified via colony-PCR or agarose gel electrophoresis. These processes were repeated until longer multimers were obtained. The results of oligomerization are shown in Fig. 6. Each DNA multimer, which was isolated by *Lgu*I, was verified by DNA sequence analysis (data not shown). 5mer (100 amino acids), 10mer (200 amino acids), 20mer (400 amino acids) of gene 3 multimers were synthesized (Fig. 6).

3. Comparison of Recursive Directional Ligation (RDL) with Controlled Cloning Method (CCM) as Cloning Methods

The method of obtaining recombinant DNA multimers is classified iinto three categories: simple concatemerization, recursive directional ligation (RDL), and controlled cloning method (CCM). In simple concatemerization, a DNA monomer encoding tandem repeats of specific motif is amplified via PCR with proper primers (as TP1 & TP3, FP & BP in this report), and these amplified DNAs are self-ligated. As a result, multimers with different lengths can be generated. In actuality, we could synthesize multimers from monomer up to tetramer by this simple concatemerization (data not shown). However, this method is ineffective for synthesizing precisely length-controlled multimers because the results are not reproducible. (For example, in some cases, up to dimers are generated only, and other cases, up to tetramers are generated.) Hence, even though this method has the merit that the process is simple and fast, it is difficult to synthesize the long DNA multimers that we want to get exactly in size by this method. Maybe this method can be effectively used in the synthesis of dimer or trimer, which is a basic starting material for multimerization.

Recursive directional ligation (RDL) is a method of recursive multimerization of a DNA monomer until a gene of a desired length is obtained. By using RDL method, we obtained DNA multimers with different lengths from monomer to 20mer. The merits of RDL are as follows. First, because RDL is not a PCR-based method, this method can reduce base-mismatch errors, which can be created in the process of PCR reaction. In RDL, there is no need for PCR reaction for the extension of the insert length. The synthesis of a dimer from a monomer or a tetramer from a dimer can be accomplished, not by the PCR reaction of the insert but by the digestion with endonucleases. That is, the insert, digested with two different endonucleases, is ligated with a plasmid containing DNA multimers for the production of longer multimers. For this reason, some problems based on PCR, such as base mutation, do not happen in RDL. In addition, multimerization by RDL is simpler than CCM for obtaining DNA multimers. Once DNA monomer is synthesized, multimers can be obtained only by the isolation of the DNA monomer and ligation of this DNA monomer with a plasmid containing the DNA monomer, without other complicated enzyme reactions.

Using controlled cloning method (CCM), we could also precisely control the length of multimers, but this process is a little more complicated than recursive directional ligation (RDL) in cloning steps. However, since a DNA monomer (or a multimerized gene) is isolated by PCR amplification and only the isolated DNA is treated with enzymes, this method can be used even though there are several enzyme recognition sites in a cloning vector. On the other hand, if there are several recognition sites for a specific enzyme in a cloning vector, this enzyme cannot be used for multimerization in the RDL method, because it can digest not only the insert region but also the other region in the vector. Thus, CCM has the merit that the scope of enzyme selection is broadened. In spite of these differences between RDL and CCM, by using these two methods, length control of the multimers can be easily accomplished and block copolymers (for example, A-B-A-B-A-B- or A-A-B-A-A-B-A-A-B-) can be generated unlike simple concatemerization. In conclusion, if simple concatemerization based on a self-ligation, RDL, and CCM is effectively used, repetitive polypeptides comprising from several amino acids up to thousands of amino acids can be easily generated.

ACKNOWLEDGMENTS

This work was supported by the KOSEF through the Center for Advanced Bioseparation Technology at Inha University. And this work was also supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2006-D00456).

REFERENCES

1. M. J. Dougherty, S. Kothakota, T. L. Mason, D. A. Tirrell and M. J. Fournier, *Macromolecules*, **26**, 1779 (1993).
2. C. J. Trendss, *Biotechnol.*, **8**, 309 (1990).
3. K. P. McGrath, D. A. Tirrell, M. Kawai, T. L. Mason and M. J.

Fournier, *Biotechnol. Prog*., **6**, 188 (1990).
4. A. E. Barron and R. N. Zuckermann, *Curr. Opin. Chem. Biol*., **3**, 681 (1999).
5. A. Nagarsekar and H. J. Ghandehari, *J. Drug Target*., **7**, 11 (1999).
6. M. S. Park, C. Y. Choi and J. I. Won, *Korean J. Biotechnol*., **22**, 179 (2007).
7. J. T. Prince, K. P. McGrath, C. M. Digirolamo and D. L. Kaplan, *Biochemistry*, **94**, 10879 (1995).
8. R. A. Mcmillan and V. P. Conticello, *Macromolecules*, **32**, 3643 (1999).
9. D. E.Meyer and A. Chilkoti, *Biomacromolecules*, **3**, 357 (2001).
10. J. I. Won and A. E. Barron, *Macromolecules*, **35**, 8281 (2002).
11. J. Y. Kim, S. O'Malley, A. Mulchandani and W. Chen, *Anal. Chem.*, **77**, 2318 (2005).
12. D. T. McPherson, J. Xu and D. W. Urry, *Protein Expression Purif*., **7**, 51 (1996).
13. J. Noolandi, *Electrophoresis*, **13**, 394 (1992).
14. J. Noolandi, *Electrophoresis*, **14**, 680 (1993).
15. P. Mayer, G. W. Slater and D. Drouin, *Analytical Chemistry*, **66**, 1777 (1994).
16. W. N. Vreeland, C. Destruisseaux and A. E. Karger, *Analytical Chemistry*, **73**, 1795 (2001).
17. Z. Sun and J. Cai, *Korean J. Chem. Eng.*, **23**, 607 (2006).
18. A. Chilkoti, R. D. Matthew and D. E. Meyer, *Advanced Drug Delivery Reviews*, **54**, 1093 (2002).