Synthesis of genes with multiple identical domains

Clemens Achmüller, Florian Werther, Philipp Wechner, and Bernhard Auer
University of Innsbruck, Innsbruck, Austria

BioTechniques 42:43-46 (January 2007)
doi 10.2144/000112313

Overexpression and purification of recombinant proteins in heterologous systems are widely used for protein interaction studies, structure and function determination, and production of biopharmaceuticals. Beside the isolation of genes from cDNA libraries, synthesis of genes by PCR techniques has become a common strategy because of the additional possibility to optimize codon usage for high expression yields in heterologous organisms. A set of overlapping primers will reconstitute the desired gene in a two-step PCR method known as gene assembly or gene synthesis (1,2). However, for genes with multiple identical domains or internal sequence repeats, which have increased interest as models for repeat proteins (3,4) and as designed proteins (5), this method of gene synthesis may fail due to mispriming. This drawback might be avoided by synthesis of alternative primers for each domain due to the degenerated genetic code (6), but with increasing numbers of domains, the variability of the oligonucleotides will be exhausted. Alternatively, such genes may be assembled by random ligation of monomeric fragments (6,7), by shotgun ligation of oligonucleotides (7), one by one from the monomeric domain by complex time-consuming directed cloning strategies in the desired orientation using compatible overhangs that cannot be digested after ligation (8) or by seamless cloning (9).

Here we present a simple PCR-based method to produce synthetic genes encoding multiple identical protein domains (Figure 1). The desired monomeric domain was synthesized using gene assembly techniques (1,2). In a first step (Figure 1, PCR1) the single domains were linked together by using forward and reverse link-primers, which contain the reverse complementary sequence of the 3’ and 5’ ends of the amplified fragment, respectively (Figure 1). Thereby the newly created ends of each fragment are able to anneal with each other resulting in formation of fragments containing multiple domains. However, using the standard primer-template ratio, amplification of the single domain was the only result. Linkage only occurred in substantial amounts when the link-primer concentration was reduced to 0.5 pmol, while the amount of template (single domain) was elevated to 10–25 pmol (approximately 1.5–3 μg for a 200-bp gene). Ideally, all link-primers should be incorporated during the first cycles of the PCR1, and the domains will link together by overlap-extension. However, this was theoretically impossible, since as a side reaction, annealing of the reverse complementary link-primers had to be taken into account, which further reduced the actual initial primer concentration but left a surplus of primers masked for monomer amplification in later PCR steps. PCR1 resulted in a typical DNA ladder, which showed DNA fragments with up to 10 units of the monomeric domain on a 1% agarose gel (Figure 2A). To allow subsequent isolation, cloning of the desired number of domains and to get rid of the 5’ and 3’ ends attached during PCR1, adaptor-primer containing anchor and restriction sites were incorporated by a second PCR (Figure 1, PCR2). In this step, annealing of the adaptor primers to the link-primer sequences at the domain border would result in amplification of monomers. To avoid this,

Figure 1. A method to synthesize genes with multiple identical domains. Single domains are linked together by PCR1 using link-primers (wide and narrow dashed lines) which attach complementary sequences to allow overlap-extension. Adaptors with anchor (green) and restriction enzyme (blue) sequences are incorporated by PCR2 to allow subsequent isolation and cloning of the genes with the desired number of domains. Finally, isolation of the synthetic multidomain genes was carried out by PCR3 with anchor-primers after a gel extraction step.
low primer concentrations had to be used again. One-tenth of PCR1 was directly added to the PCR2 mixture containing only 0.5 pmol adaptor-primer. The DNA ladder of PCR2 showed the same number of fragments as PCR1 as verified by agarose gel electrophoresis (Figure 2A). The fragments containing the desired number of domains were extracted from the gel using the QIAquick® Gel Extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturer’s recommendations and amplified using 50 pmol anchor-primers to allow subsequent cloning (Figure 1, PCR3). The PCR product of PCR3 was again purified by agarose gel electrophoresis (not shown), digested with restriction enzymes, and ligated into a suitable vector (Figure 2B).

We have successfully constructed synthetic codon optimized versions of streptococcal protein G domain B1 (SpG-B1) (10,11) with one, two, and three domains (Figure 2) and staphylococcal protein A domain D (SpA-D) (12,13) with one and two domains (data not shown) using this novel approach. The accuracy of this method could be estimated by the result that with (SpG-B1)2 two out of four and with (SpG-B1)3 and (SpA-D)2 one out of four clones showed the correct sequence using a standard Taq DNA polymerase. For genes containing multiple units, usage of a proofreading thermostable DNA polymerase would be advisable.

There are various examples of time-consuming complex construction of genes for model proteins with multiple identical domains or internal sequence repeats to produce these proteins for experiments to further understand the relationships between molecular structure, function, and evolution of protein repeats. For instance, eight identical repeats of an immunoglobulin domain gene of the muscle protein titin had to be cloned one by one using different restriction sites into an expression vector to produce a model protein for mechanical unfolding studies of titin (14). As another example, up to six designed ankyrin repeat genes had to be cloned one by one using the seamless cloning technique (9) into an expression vector to produce designed ankyrin repeat proteins for high affinity target binding (15). Construction of such genes as well as the genes for biotechnologically relevant proteins like combinations of silk and elastin peptides (5) and for protein libraries with various numbers of such repeats would be greatly facilitated by use of the described method.

ACKNOWLEDGMENTS

Financial support was provided by the Austrian Center of Biopharmaceutical Technology (ACBT).

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES


Benchmarks


Address correspondence to Bernhard Auer, Institute of Biochemistry, University of Innsbruck, Peter-Mayr-Str. 1a, A-6020 Innsbruck, Austria. e-mail: bernhard.auer@uibk.ac.at

To purchase reprints of this article, contact: Reprints@BioTechniques.com

Received 1 August 2006; accepted 26 September 2006.