Short Technical Reports

Fluorescent Protein Vector for Directional Selection of PCR Clones

BioTechniques 29:800-808 (October 2000)

ABSTRACT

Green fluorescent protein (GFP) has become a convenient and versatile tool as a reporter protein in many aspects of science. Here, we show that the enhanced yellow fluorescent protein (EYFP) variant may be used advantageously as a reporter system for directional cloning of blunt-ended PCR products. We have constructed a pUC18-derived plasmid containing a reporter gene coding EYFP cloned into the *Bam*HI/*Hind*III sites. The blunt-ended PCR product is cloned into the *Sma*I site of that plasmid. A reverse PCR primer must be designed with extra bases on the 5' end that are required to introduce a ribosome binding site (rbs) for EYFP expression. The reporter gene coding EYFP is not expressed unless an rbs is introduced in the proper orientation at the 3' end of the cloned PCR insert. The results of this cloning procedure may be analyzed by simple visual inspection using a transilluminator. In most cases, successful directional cloning results in white fluorescent colonies. The proposed procedure is a convenient method that can reduce the time- and labor-intensive analysis of the clones obtained during blunt-ended PCR product cloning.

INTRODUCTION

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is rapidly becoming an important reporter molecule (1,8,10,12). Unlike other bioluminescent reporters, the chromophore in GFP is intrinsic to the primary structure of the protein, and its fluorescence does not require a substrate or cofactor (6,9). In addition to wild-type GFP, there are several GFP variants including enhanced green fluorescent protein (EGFP) (3,13), enhanced blue fluorescent protein (EBFP) (13) and enhanced yellow fluorescent protein (EYFP) (13). EYFP has the

longest wavelength emissions of all GFP variants examined to date. This shift in the spectrum is the result of four amino acid substitutions (S65G, V68L, S72A and T203H), mutations rationally designed on the basis of the X-ray structure of GFP (11). The EYFP significantly extends the pH range over which GFPs may be employed as pH indicators in living cells.

Cloning PCR-amplified fragments has become a widely used technique in molecular biology. Common cloning strategies include sticky-end cloning, blunt-end cloning, TA cloning and ligation-free cloning. To facilitate sticky-end cloning, restriction endonuclease sites can be introduced into the amplification primers so that subsequent digestion of the reaction products with the appropriate enzymes will generate a product ready to be cloned into specific sites of the recipient vector. It is often found, however, that amplified DNA is resistant to digestion by restriction endonucleases (many enzymes have weak digestion activity at the end of the DNA fragment). Also, sequences may be lost during digestion and cloning because of the existence of unknown restriction sites within the amplified DNA. Since self-ligated and uncut vector plasmid can also transform bacteria, many colonies with no insert (false-positive) are generated as background. Many colonies have to be screened to find a colony with an insert (true-positive). PCR products may be cloned into a blunt-ended site, but this procedure is hampered by the presence of an additional nucleotide at the 3' end of the reaction product introduced by the *Taq* DNA polymerase (2,7). In addition, the orientation of the cloned DNA cannot be controlled or predetermined. To circumvent these problems, an efficient and simple method for screening the clones is described here.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, Enzymes and Reagents

The *E. coli* TOP10F[®] strain (Invitrogen, Carlsbad, CA, USA) was used for the cloning procedure. The pUC18 and pYFPrbs (5) plasmids were used

for the construction of the new cloning vector. Restriction enzymes were purchased from MBI Fermentas (Basel, Switzerland). The reagents for PCR were obtained from DNA-Gdańsk II s.c. (Poland). IPTG, X-Gal, agarose and other reagents were purchased from Sigma (St. Louis, MO, USA). Genomic DNA from *Borrelia burgdorferi* sensu lato and *Mycobacterium tuberculosis* were a gift from Department of Tropical Parasitology (Dr. P. Myjak), Institute of Maritime and Tropical Medicine, Gdynia, Poland.

Amplification and Cloning of the *yfp* Gene

The DNA template for the amplification of the *yfp* gene was from the pYFPrbs plasmid (5). Based on the known *yfp* gene sequence (GenBank[®] accession no. U76561), the specific primers for PCR amplification were synthesized. The forward primer was GFPf: 5'-GCCGGATCCG AATGGT-GAGCAGGGCGAGGAG-3' (*Bam*HI site, underlined), and the reverse primer was GFPr: 5'-CGCAAGCTTA-CTTGTAAGCTTCCATGCG-3' (*Hind*III site, underlined). The bold letters in the primer sequences are complementary to the nucleotide sequence of the *yfp* gene, and the 5' overhanging end of the reverse primer introduced a UAA stop codon.

The amplification was carried out in standard conditions using Delta2 DNA polymerase [*Pwo* DNA polymerase, DNA-Gdańsk II s.c., Poland (4)] and 30 cycles with a temperature profile of 30 s at 94°C, 1 min at 60°C and 1 min at 72°C. The 738-bp amplification product (0.2 µg), digested with *Bam*HI and *Hind*III, was isolated as a band from an agarose gel using the DNA Gel Out kit (A&A Biotechnology, Poland). The purified fragment was ligated into the pUC18 *Bam*HI-*Hind*III sites and the resulting recombinant plasmid was named pUCYFP.

Amplification and Cloning of the *M. tuberculosis mtp40* Gene Fragment

Based on the known *mtp40* gene sequences (GenBank accession no. L11868), the specific primers for PCR amplification of the *mtp40* gene frag-

Short Technical Reports

ment were designed. The forward primer was MtpB: 5'-GTCCGGTGG-CATTCGTGGTCCCG3', and the reverse primer was MtpArbs: 5'-CTC-CTTCTCCGTCGGTGGTTCCCA-ACACCACG-3'. The bold letters in the primer sequences are complementary to the nucleotide sequence of the *mtp40* gene, and the 5' overhanging end (italics) of the reverse primer contains a sequence complementary to the ribosome binding site (*rbs* found in pET expression plasmids (Novagen, Madison, WI, USA). The amplification was carried out in standard conditions using Delta2 DNA polymerase [*Pwo* DNA polymerase, DNA-Gdańsk II s.c. (4)] and 40 cycles with a temperature profile of 30 s at 94°C and 1 min at 69°C. The 327-bp amplification product (0.1 µg) was isolated from an agarose gel using DNA Gel Out kit and phosphorylated using 10 U T4 PNK (MBI Fer-

mentas). The phosphorylated fragment was ligated into the dephosphorylated pUCYFP *Sma*I site. The ligation mixture was transformed into *E. coli* TOP10F' and plated on a luria agar plate containing X-Gal (80 µL of a 40 mg/mL solution), IPTG (80 µL of a 100-mM solution) and ampicillin (100 µg/mL).

Amplification and Cloning of the *B. burgdorferi fla* Gene Fragment

Based on the known *fla* gene sequence (GenBank accession no. AE001126), the specific primers for PCR amplification of the *fla* gene fragment were designed. The forward primer was bfl1: 5'-GCTCAATATA-ACCAATGCACATG-3', and the reverse primer was bfl2rbs: 5'-CTCCT-TCTTCAAGTCTATTTTGGAAAG-CACCTAA-3'. The bold letters in the

primer sequences are complementary to the nucleotide sequence of the *fla* gene, and the 5' overhanging end (italics) of the reverse primer contains sequence complementary to the *rbs*. The amplification was carried out in standard conditions using Delta2 DNA polymerase and 35 cycles with a temperature profile of 30 s at 94°C, 1 min at 52°C and 1 min at 72°C. The 451-bp amplification product (0.1 µg) was purified and cloned in the same manner as the *mtp40* gene fragment of *M. tuberculosis*.

RESULTS AND DISCUSSION

Construction of the Cloning Vector

The pUC plasmids are one of the most efficient systems developed for cloning procedures in *E. coli*. The 726-bp *Bam*HI/*Hind*III PCR fragment

Figure 1. The nucleotide and amino acid sequences of the pUCYFP vector. The reporter gene encoding EYFP is cloned into the *Bam*HI/*Hind*III sites. The blunt-ended PCR product is cloned into the *Sma*I site of the plasmid. Abbreviations used: *P_{lac}*: lac promoter; underlined amino acid sequence: LacZ/FS-EYFP, fusion protein consisting of the LacZ peptide and frameshifted EYFP protein; *rbs_{lac}*: ribosome binding site for LacZ/FS-EYFP translation; Start_{LacZ}: N-terminal end of LacZ/FS-EYFP protein; Stop_{LacZ}: termination site for N-LacZ/FS-EYFP/C-LacZ protein translation; *rbs_{EYFP}*: ribosome binding site for EYFP protein translation; Start_{EYFP}: N-terminal end of EYFP protein; Stop_{EYFP}: Termination site for EYFP protein translation.

Short Technical Reports

of the *yfp* gene was inserted between the *Bam*HI and *Hind*III sites of pUC18 plasmid. The sequence of pUCYFP plasmid insert was verified by sequencing using the dideoxy chain termination method. The nucleotide and amino acid sequences of the pUCYFP plasmid near the site of the *yfp* gene insertion are shown in Figure 1. The obtained genetic construct is a relatively small, high copy number 3382-bp plasmid. The *LacZ* DNA encoding the α -peptide of β -galactosidase (*LacZ*) is disrupted by *yfp*, which is a +1 base frameshift relative to the initiating codon ATG of *LacZ*. Also, the UAA stop codon, introduced with primer sequence at the end

of the *yfp* gene, is not revealed. Because there is no additional stop codon, the open reading frame creates a fusion protein containing 339 amino acids (N-LacZ/FS-EYFP/C-LacZ) (Figure 2A). This protein, whose synthesis can be induced by IPTG, is capable of intra-allelic (α) complementation with a defective form of β -galactosidase encoded by host bacteria (mutation *lacZ* Δ M15) and forms blue colonies on plates containing X-Gal. Because of the *yfp* frameshift, the blue colonies do not emit any fluorescence.

Cloning a blunt-ended PCR product (with an *rbs* at the 3'-end introduced with a PCR primer) into the *Sma*I site of

pUCYFP plasmid inactivates the N-terminal fusion fragment of β -galactosidase and abolishes α -complementation. Bacteria carrying recombinant plasmids therefore give rise to white colonies and, because of the insertion of an *rbs* site before the *yfp* gene, permits the synthesis of the reporter fluorescent protein (EYFP), resulting in the fluorescence of the colony (Figure 2B). In most cases, the PCR insert also introduces a stop codon in relation to the *LacZ* start codon, and usually a short fusion protein is produced (N-LacZ /PCR fragment) without C-LacZ (Figure 2B).

This vector allows for the rapid and specific identification of correctly oriented clones by simple fluorescent visualization of white colonies (UV transilluminator, 312 nm). However, in this cloning system, we have several additional, rather rare construct possibilities, as shown in Figure 2, C–E. In Figure 2C, the full-length fusion protein is produced (N-LacZ/PCR product/FS-EYFP/C-LacZ) when the PCR insert is in frame with *LacZ* and does not introduce a stop codon. If the PCR product insertion has no influence on the N-terminal fragment of β -galactosidase, the colonies are blue. However, white colonies should also arise because of the inability of the fusion protein to undergo α complementation. In both cases,

Figure 2. Construct possibilities using the pUCYFP vector cloning system. (A) Cloning region of pUCYFP plasmid. (B) The most frequent construct possibility of recombinant plasmid with cloned blunt-ended PCR product into the *Sma*I site. (C–E) Rare construct possibilities: PCR insert is in frame with *LacZ* (C), PCR insert is +1 base frameshifted relative to the *LacZ* start codon (D); PCR insert is +2 base frameshifted in relation to the *LacZ* start codon. Abbreviation used: 5' *LacZ* and 3' *LacZ*: 5' and 3' end of *LacZ* gene, respectively; N-LacZ/FS-EYFP/C-LacZ: fusion protein consisting of *LacZ* peptide and frameshifted EYFP protein; N-LacZ/PCR product: fusion protein consisting of N-terminal *LacZ* peptide fragment and protein encoded by PCR product; N-LacZ/PCR product/FS-EYFP/C-LacZ: fusion protein consisting of *LacZ* peptide, protein encoded by PCR product and frameshifted EYFP protein; N-LacZ/PCR product/EYFP fragment: fusion protein consisting of N-terminal end of *LacZ* peptide, protein encoded by PCR product and fragment of N-terminal end of EYFP protein; N-LacZ/PCR product/EYFP: fusion protein consisting of N-terminal *LacZ* peptide, protein encoded by PCR product and EYFP protein. FS: frameshift.

Short Technical Reports

blue and white colonies should emit fluorescence (Figure 2C). In Figure 2D, the PCR insert is +1 base frame-shifted relative to the *LacZ* start codon and does not introduce a stop codon. The colonies are white because the construct produces a fusion protein without C-*LacZ* fragment (N-*LacZ*/PCR product/EYFP fragment) and the stop codon is always introduced at the beginning of the *yfp* gene sequence.

The synthesis of the reporter fluorescent protein from the *rbs* (introduced by the PCR insert) causes colony fluorescence. In Figure 2E, the PCR insert is +2 bases frame-shifted relative to the *LacZ* start codon, and does not introduce a stop codon. The colonies are white because of the active translation termination by the stop codon of the *yfp* gene (therefore, the C-*LacZ* fragment is not produced) and fluorescence is emitted as a result of EYFP synthesis. It is also possible to produce white colonies

when the cloned PCR fragment contains a strong transcriptional terminator (data not shown). In these rare cases, EYFP will not be produced, yet colonies will be white.

The Ability of pUCYFP Plasmid to Identify Correct Clones (Examples)

The cloning procedure, in which a blunt-ended PCR product is inserted into the *SmaI* site of the pUCYFP plasmid, is useful in our laboratory. This vector is available from us by request. The efficiency of the cloning scheme using pUCYFP plasmid is shown in two examples below.

We cloned the *M. tuberculosis mtp40* and *B. burgdorferi fla* gene fragments into the *SmaI* site of the pUCYFP plasmid. After transformation, we used plasmid miniprep and PCR techniques to screen for the inserts. In both cases, about 30% of white colonies revealed

fluorescence. To confirm whether the white colonies emitting fluorescence are clones with inserts in the correct orientation, we isolated the plasmid DNA from 100 transformants and PCR-amplified the insert fragment using the same primers used for cloning. We observed about 98% true-positive colonies in both cases. The same experiments were carried out with 100 transformants of white colonies without fluorescence. We found the expected insert (but in the opposite orientation) by PCR amplification in 73% and 63% transformants of *M. tuberculosis mtp40* and *B. burgdorferi fla*, respectively. These results revealed that the white fluorescing colonies, where the inserted DNA fragment was found to be in the correct orientation, were obtained with as much as 98% accuracy. In the case of white colonies without fluorescence, 30% of the recombinant plasmid products were found to be either autoligated vectors with deletions,

Short Technical Reports

unspecific products cloned or inserted fragments without an *rbs*.

Blunt ended-cloning is notoriously inefficient, with high vector backgrounds commonly as high as 80%. Several companies have devised cloning systems—for example, ZeroBlunt Vector from Invitrogen, QuanTox (QBiogene, Montreal, QC, Canada) or PCR-Script™ Cloning System (Stratagene, La Jolla, CA, USA)—that overcome this problem.

In this study, we proposed a simple, new method for screening clones obtained by blunt-ended cloning of PCR products. For this cloning, PCR products can be obtained in blunt-ended form directly after using either *Pfu* or *Pwo* DNA polymerase, or indirectly using *Taq* or *Tth* DNA polymerase for the amplification, followed by a 30-min reaction to remove overhangs from the PCR products and to phosphorylate the ends. The PCR amplification also requires a reverse primer that must contain six additional nucleotides (CTCCTT) at the

5'-end, creating an *rbs* site. The pUC-YFP vector plasmid should be prepared by *Sma*I digestion and dephosphorylated according to standard procedures.

Colonies can be screened for inserts by selecting white colonies and examining their fluorescence. The cloning system described here is a positive-selection system for the cloning of PCR products through blunt-end ligation with as much as 98% accuracy (transformants bear true recombinant plasmids—white fluorescing colonies).

The pUCYFP plasmid described here can also be used as an expression vector. Two different expression systems can be constructed. The first expression system permits the synthesis of a bicistronic mRNA encoding both a gene of interest and the gene encoding the reporter fluorescent protein (EYFP) (5). This construct allows for rapid and specific identification of protein expression in living cells by measuring fluorescence. The second expression system permits the synthesis of a full-length fusion protein when the PCR insert is in frame with *LacZ* start codon and does not introduce a stop codon. If the PCR product insertion has no effect on the N-terminal fragment of β -galactosidase (*LacZ*), the colonies are blue and fluoresce.

REFERENCES

1. Chalfie, M., Y. Tu, G. Euskirchen, W.W. Ward and D.C. Prasher. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263:802-805.
2. Clark, J.M. 1988. Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* 16:9677-9686.
3. Cormack, B.P., R.H. Valdivia and S. Falkow. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173:33-38.
4. Dabrowski, S. and J. Kur. 1998. Cloning and expression in *Escherichia coli* of the recombinant his-tagged DNA polymerases from *Pyrococcus furiosus* and *Pyrococcus woesei*. *Protein Expression Purif.* 14:131-138.
5. Dabrowski, S., A. Brillowska and J. Kur. 1999. Use of the green fluorescent protein variant (YFP) to monitor MetArg human proinsulin production in *Escherichia coli*. *Protein Expression Purif.* 16:315-323.
6. Heim, R., D.C. Prasher and R.Y. Tsien. 1994. Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* 91:12501-12504.
7. Hemsley, A., N. Arnheim, M.D. Toney, G. Cortopassi and D.J. Galas. 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* 17:6545-6551.
8. Inouye, S. and F.I. Tsuji. 1994. Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett.* 341:277-280.
9. Inouye, S. and F.I. Tsuji. 1994. Evidence for redox forms of the Aequorea green fluorescent protein. *FEBS Lett.* 351:211-214.
10. Prasher, D.C., V.K. Eckenrode, W.W. Ward, F.G. Prendergast and M.J. Cormier. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 111:229-233.
11. Wachter, R.M., M.A. Elsliger, K. Kallio, G.T. Hanson and S.J. Remington. 1998. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* 15:1267-1277.
12. Wang, S. and T. Hazelrigg. 1994. Implications for bed mRNA localization from spatial distribution of exu protein in *Drosophila oogenesis*. *Nature* 369:400-403.
13. Yang, T.T., L. Cheng and S.R. Kain. 1996. Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* 24:4592-4593.

This work was supported by KBN Grant 4 P05A 099 17 and partially by the Technical University of Gdańsk and DNA-Gdańsk II s.c to J.K. Address correspondence to Dr. Józef Kur, Technical University of Gdańsk, Department of Microbiology, ul. Narutowicza 11/12, 80-952 Gdańsk, Poland. e-mail: kur@altis.chem.pg.gda.pl

Received 17 February 2000; accepted 12 June 2000.

Slawomir Dabrowski, Anna Brillowska-Dabrowska and Józef Kur
*Technical University of Gdańsk
Gdańsk, Poland*