Available online at www.sciencedirect.com
SCIENCE DIRECT
Protein Expression and Purification 47 (2006) 562–570
Protein Expression & Purification
www.elsevier.com/locate/yprep

High efficiency single step production of expression plasmids from cDNA clones using the Flexi Vector cloning system ☆

Paul G. Blommel, Peter A. Martin, Russell L. Wrobel, Eric Steffen, Brian G. Fox *

Department of Biochemistry, Center for Eukaryotic Structural Genomics, University of Wisconsin, Madison, WI 53706-1549, USA

Received 21 October 2005, and in revised form 10 November 2005
Available online 5 December 2005

Abstract

The success of structural genomics and proteomics initiatives is dependent on the availability of target genes in vectors suitable for protein production. Here, we compare two high-throughput methods for producing expression vectors from plasmid-derived cDNA fragments. Expression vectors were constructed for compatibility with the Gateway recombination cloning system and the Flexi Vector restriction-based cloning system. Cloning protocols for each system were conducted in parallel for 96 different target genes from PCR through the production of sequence-verified expression clones. The short nucleotide sequences required to prepare the target open reading frames for Flexi Vector cloning allowed a single-step PCR protocol, resulting in fewer mutations relative to the Gateway protocol. Furthermore, through initial cloning of the target open reading frames directly into an expression vector, the Flexi Vector system gave time and cost savings compared to the protocol required for the Gateway system. Within the Flexi Vector system, genes were transferred between four different expression vectors. The efficiency of gene transfer between Flexi Vectors depended on including a region of sequence identity adjacent to one of the restriction sites. With the proper construction in the flanking sequence of the vector, gene transfer efficiencies of 95–98% were demonstrated.

Keywords: High-throughput cloning; Proteomics; Expression vectors

The large number of gene sequences now available has led to proteome-scale projects to evaluate protein structure and function [1–4]. Cloning of target genes into expression vectors is an essential part of these efforts [5]. Many structural genomics and proteomic groups use the recombination-based Gateway system [6–9] or ligation-independent cloning to produce expression clones [10,11]. As part of the Protein Structure Initiative, the Center for Eukaryotic Structural Genomics (CESG)[1] previously developed and applied a two-step Gateway protocol to evaluate ~3500 target genes [6].

As part of ongoing efforts, we are interested in evaluating other cloning systems. Flexi Vector, a restriction enzyme cloning system, apparently offered the advantages of high-throughput cloning of PCR products directly into an expression vector and serial transfer from the first, sequence-verified expression vector to others. In this work, we compare the sub-cloning of 96 human target genes from a plasmid cDNA source using these two cloning systems and the efficiency of transfer of the coding sequences between different Flexi Vectors.

Materials and methods

Materials

Flexi Vector reagents, high concentration T4 DNA ligase, PCR clean-up kits (SV96 and Magnesil), Select 96 competent cells, PCR Master Mix, Magnebot II magnetic bead separation block, expression vectors pF1K, pF6A, and pF6K, and DNA molecular weight markers were from

☆ This work was supported by the National Institutes of Health, Protein Structure Initiative Grant P50 GM-64598.
* Corresponding author. Fax: +1 608 262 3453.
E-mail address: bgfox@biochem.wisc.edu (B.G. Fox).
[1] *Abbreviation used:* CESG, Center for Eukaryotic Structural Genomics.

1046-5928/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2005.11.007

Promega (Madison, WI). Gateway reagents, *Escherichia coli* Top 10 competent cells, and 2% Agarose Egel 96 s were from Invitrogen (Carlsbad, CA). Oligonucleotide primers were from Integrated DNA Technologies (Coralville, IA). The ExSite PCR mutagenesis kit and Yield Ace polymerase were from Stratagene (La Jolla, CA). CircleGrow medium was from Q-Biogene (Irvine, CA). Aeraseal gas permeable sealing tape was from Excel Scientific (Wrightwood, CA). Big Dye Version 3.1 sequencing reagents were from Applied Biosystems (Foster City, CA). Miscellaneous reagents were as previously reported [6].

Target genes

A set of 96 target genes selected by CESG for Gateway cloning was used for the comparative study. The genes were obtained as full-length cDNA clones from Open Biosystems (Huntsville, AL). This set consisted of human ORFs 600–1200 nucleotides in length. The target genes are listed in the Supplementary data.

Vectors

Features of CESG expression vectors are reported elsewhere [6]. Standard molecular cloning techniques [12] were used unless specified. pVP33K and pVP33A were produced from pVP16 by several steps. These included: addition of *Nsi*I and *Pac*I sites flanking the solubility domain by ExSite PCR mutagenesis; modification of the linker to include cleavage sites for the 3C and TEV proteases, the tetraCys motif, and the *Sgf*I and *Pme*I cloning sites also by using the ExSite system; creation of kanamycin (pVP33K) and ampicillin (pVP33A) resistance cassettes flanked by *Avr*II and *Bsi*WI, and insertion of these cassettes by digestion and ligation into a PCR amplified vector backbone modified to contain the *Avr*II and *Bsi*WI sites flanking the positions of the resistance genes; reintroduction of the MBP solubility domain by digestion and ligation; production of the Bar-CAT cassette by overlap extension PCR [13]; insertion of the Bar-CAT cassette into the *Sgf*I and *Pme*I sites of pF1K by digestion and ligation; and transfer of the Bar-CAT cassette to pVP33A/K using PCR, restriction digest, and ligation. The coding region from promoter through terminator was sequenced after each step to verify mutagenesis and the lack of second site mutations. pVP33A shares 128 bp of sequence identity with pF6A/K 3′ to the *Pme*I site while pVP33K does not. This region of identity was used to reduce the frequency of double backbone ligation products through the formation of a non-replicating extensive DNA palindrome [14]. To include the region of identity, the Bar-CAT cassette was amplified out of pF1K using the T7 terminator primer and ligated into pVP33A without digesting the cassette with *Pme*I (Fig. 1). pEU-His-Flexi was produced by modifying pEU-His [15] to include the *Sgf*I and *Pme*I sites followed by insertion of the Bar-CAT cassette including the 128 bp of 3′ identity. The parental pVP33K, pVP33A, and pEU-His-Flexi plasmids must be propagated in a barnase resistant strain (*E. coli* BR610, Promega).

PCR primers

For Flexi Vector cloning, target genes were amplified using a single-step PCR. The forward and reverse primers

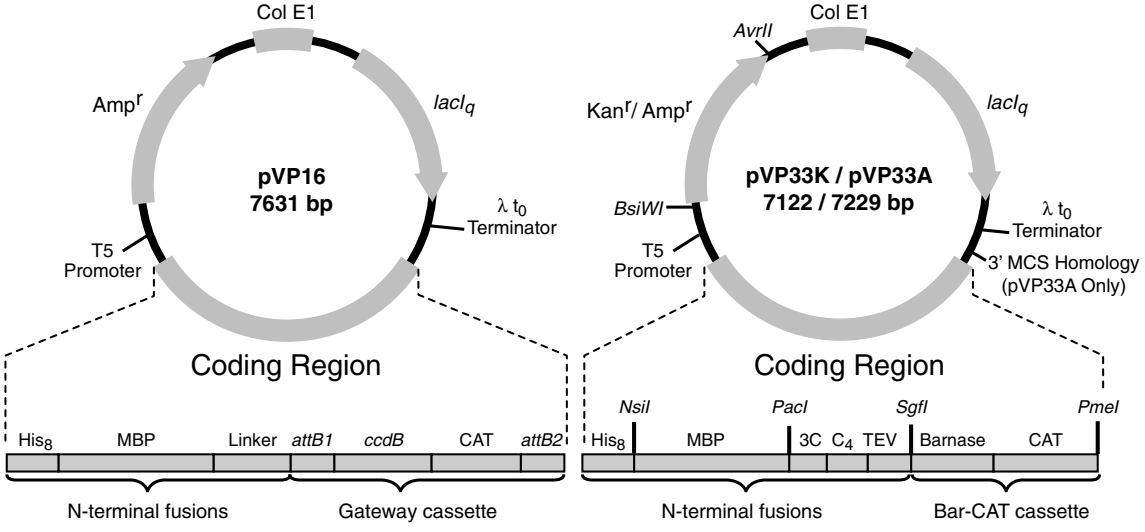

Fig. 1. Vectors used in this study. Vectors were derived from pQE80 (Qiagen, Valencia, CA) and contain a ColE1 origin. After cloning, the target gene sequence replaces the Gateway or Bar-CAT cassette for pVP16 and pVP33K or pVP33A, respectively. Both vectors contain N-terminal fusions to allow standardized purification (His_8 tag) and increased target protein solubility (MBP). Target proteins may be liberated from the N-terminal fusions using TEV protease. For pVP16, the TEV protease recognition site is incorporated during amplification of the target gene using the two-step PCR protocol. pVP33K and pVP33A contain a vector encoded TEV recognition site. pVP33K and pVP33A also contain a tetraCys motif (labeled C4 in the figure) for site-specific fluorescent labeling of fusion protein and a 3C protease recognition site. Restriction sites in pVP33K and pVP33A allow switching of the solubility enhancing tag (*Nsi*I and *Pac*I) or antibiotic resistance (*Avr*II and *Bsi*WI). The differences in total sizes of pVP33K and pVP33A arise from the sum of differences in the region designated 3′ MCS homology (present in pVP33A only) and in the sizes of the genes encoding the respective antibiotic resistances.