Center for Eukaryotic Structural Genomics

Protein Structure Initiative

Evolution of Cloning Strategies at the CESG

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

University of Wisconsin-Madison Biochemistry, 433 Babcock Drive, Madison, Wisconsin, USA 53706-1549, http://www.uwstructuralgenomics.org

Abstract

The success of structural genomics initiatives is critically dependent on the incorporation of targeted open reading frames (ORFs) into vectors suitable for protein production. Among the many structural genomics groups using Gateway® recombination, CESG previously developed and used the Gateway® method to clone ~3500 eukaryotic ORFs. As part of this effort, a customized modular vector backbone was created to allow efficient swap of antibiotic resistance markers, protein tags, linker regions, and protease sites. Some of the best variants of these production vectors, available by material transfer agreement with the University of Wisconsin, are described. To facilitate expansion of this vector set to include wheat germ cell-free and other expression platforms, we evaluated Flexi®Vector, a restriction enzyme/ligation based cloning system recently developed by Promega Corporation (Madison, WI). This system offered the advantages of high-throughput cloning of PCR products directly into an expression vector and serial transfer of the sequenced verified ORFs from the first vector to others. Here we report a comparison of 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 shorter nucleotide sequences required to prepare the target ORFs 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 ORFs directly into an expression vector, the Flexi®Vector system gave time and cost savings compared to the CESG protocol originally developed 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 obtained. Detailed protocols developed for the Flexi®Vector method, the current catalog of vectors developed for this project, and opportunities for multiplexed cloning and expression studies are presented.

Project Expression Vectors

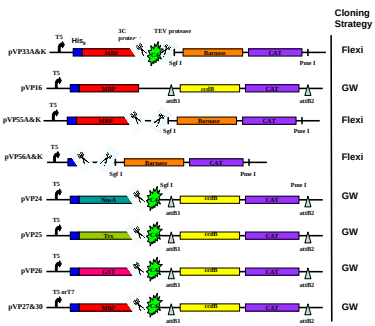

Figure 1. Expression vectors created at CESG. The Gateway® (GW) vector pVP16 and the Flexi®Vector (Flexi) pVP33K/A were used for this study. Both contain His_6 and maltose binding protein (MBP) fusion tags to aid in protein purification and solubility, respectively. For selection against non-recombinants the Gateway® system utilizes the *ccdB* gene, while the Flexi®Vector system utilizes the gene encoding the ribonuclease barnase. Flexi®Vectors were constructed in both ampicillin (A) and kanamycin (K) resistance forms to aid in the transfer of ORFs between vectors. After cloning, target proteins can be released from the N-terminal fusions by cleavage with TEV protease. To remove the residues (TSLYKKAGS) required for the *attB1* site from the N-terminus of target proteins produced by pVP16, the TEV site is incorporated between the *attB1* site and the target protein by a two-step PCR. Since fewer residues are required to incorporate the *SgfI* site in Flexi®Vector cloning (AIA), the TEV site can be incorporated into the vector, which allows a single step PCR protocol. The pVP33K/A fusion proteins contain a tetraCys motif (*C4*) for protein detection and a 3C protease cleavage site to allow an alternative to TEV cleavage. Some other Gateway® and Flexi®Vectors used successfully by the CESG are also shown. These vectors encode NusA, thioredoxin (*Trx*), or glutathione S-transferase (GST) or no solubility tag. Most CESG vectors are based on the pQE80 backbone and utilize the T5 promoter, but some vectors based on the pET32 utilize the T7 promoter.

All of these vectors can be made available by material transfer agreement with the University of Wisconsin.

Cloning Protocols

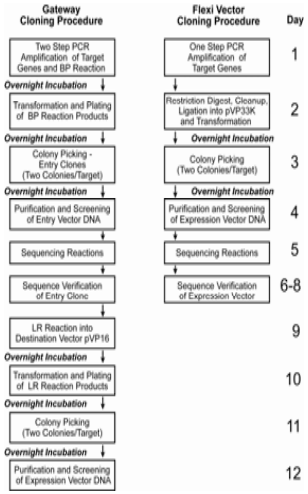

Figure 2. Comparison of Gateway® and Flexi®Vector cloning protocols. The initial cloning step for the Gateway® System places the target gene into an Entry Vector. Sequence verification is done at this point and sequence-verified. Entry Vectors are then used as "master clones" for transfer into expression vectors. The Flexi®Vector system eliminates the Entry Vector by direct cloning into an expression vector. Sequence verification takes places in the expression clone, which then may be used for serial transfer to other expression vectors as desired.

Gene Capture and Transfer

The initial cloning step for the Gateway® system places target ORFs in an Entry Vector through a BP recombination step. Entry clones serve as a consistent context for storage and sequence verification of gene sequences but are not suitable for protein expression. In the Flexi®Vector system, the vector used for initial ORF capture (pVP33K in this study) serves the same purpose as a Gateway® Entry clone, but is also suitable for protein expression. In PSI-I, we developed a Gateway® cloning system that required a two-step PCR protocol to incorporate the recom-bination sites and a TEV protease site between the attB1 site and the target ORF (**Figure 3**). However, since fewer residues are required to incorporate the *SgfI* site in Flexi®Vector cloning (AIA), the TEV site can be incorporated into the vector, which allows a single step PCR protocol (**Figure 4**). Along with shorter primers, the single step PCR resulted in fewer PCR-derived mutations in Flexi®Vector clones compared to the Gateway® clones (**Table 1**). During initial ORF capture from PCR products, similar efficiencies were found for both cloning systems (**Figure 5**).

To produce an expression clone in the Gateway® system, sequence verified entry clones are subjected to a second recombination reaction (LR reaction) that transfers the gene into a destination vector. By screening two colonies of LR reaction products, all of the target ORFs successfully incorporated into entry clones were recovered as expression clone and 97% of the total number of colonies screened contained the target ORF.

Although transfer of the ORFs is not required to produce an expression clone in the Flexi®Vector system, it is often advantageous to have the flexibility to transfer the ORF into other expression contexts. In Flexi®Vector, positive selection for transfer occurs by alternating the antibiotic resistance of the donor and acceptor plasmids. We studied the efficiency of transfers from the initial His_6-MBP expression vector pVP33K (kanamycin selection) to pVP33A (ampicillin selection) and from these two vectors to the commercially available vectors pF6A and pF6K. **Figure 5** shows the transfer efficiencies. Previous studies on palindromic DNA found that plasmids containing long, continuous palindromes could not be replicated in *E. coli*. Thus the presence of an identical sequence region adjacent to either the Sgf I or Pme I sites in the pairs of Flexi®Vectors used for transfer helps to prevent replication of faulty ligation products that do not contain the target genes. This is an important design element for construction of efficient Flexi®Vectors.

Gateway® PCR Strategy

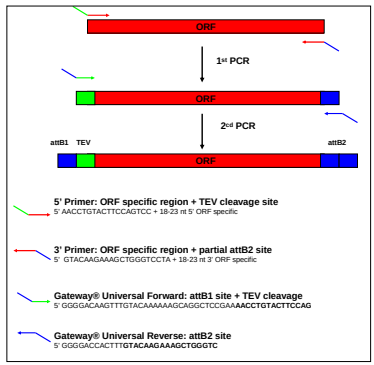

Figure 3. PCR Strategy used to clone ORFs using the Gateway® system. Diagramed is the two step amplification of a targeted ORF with the addition of the recombination sites (attB1 and attB2) and the sequence encoding the TEV protease cleavage site. The DNA sequence of the primers are shown. The region of overlap is in bold text.

Flexi®Vector PCR Strategy

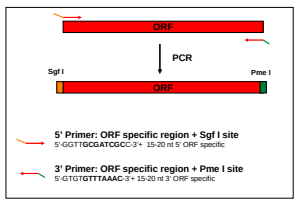

Figure 4. PCR Strategy used to clone ORFs using the Flexi®Vector system. Diagramed is the amplification of a targeted ORF with the addition of the Flexi®Vector restriction endonuclease sites Sgf I and Pme I. The DNA sequence of the primers are shown. The Sgf I and Pme I sites are in bold text.

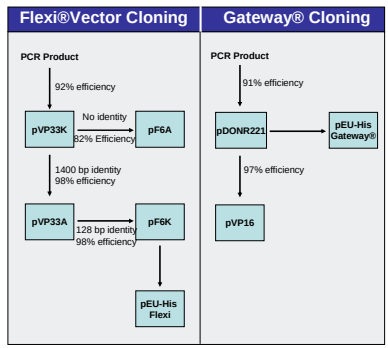

Figure 5. Percentage of colonies screened positive for target gene transfers between Flexi®Vector expression vectors or after Gateway® LR reaction transformation. The length of the region of identity between the Flexi®Vector transfer partners is shown in base pairs (bp).

Sequencing Results

Table 1. Comparison of single pass sequencing results from Gateway® and Flexi®Vector cloning protocols.

Sequencing result	Gateway® entry clone (pDONR221)	Flexi®Vector expression clone (pVP33K0)
Sequence +	71	78
Mis-sense	11	6
Silent	5	3
Sequence -	4	5
No clone	5	4
% with errors	22.0%	15.2%

[a] Sequence (+) clones matched published sequences for the target gene exactly. Missense and silent mutations differed in nucleotide sequence from the published sequences only by substitution of nucleotides. Sequence (-) clones contained fatal errors, including primer deletions causing a frameshift, nucleotide substitutions resulting in a premature stop codon, or loss of the PmeI site due to cloning artifacts.

All clones obtained by transfer from pVP33K to pVP33A were sequenced to determine the fidelity of the transfer and, importantly, the integrity of the Pme I site. Of the 184 sequences analyzed, all but one contained functional Pme I sites. Since no PCR steps were involved in the transfers, mutations in the coding regions were not expected and indeed, no changes in amino acid coding were found.

Conclusions

There are strengths and weaknesses inherent to any cloning system. The cloning steps in the Gateway® system are highly efficient, and there are a wide variety of vectors available due to the length of time this system has been in use. However, the requirement for an initial, non-productive cloning step and the long primer sequences required to encode the recombination sites are drawbacks to this system.

In the Flexi®Vector system, the initial cloning step inserts the target gene into an expression vector and the short flanking nucleotide sequences can be added in a single PCR step. As demonstrated here, the efficiency of Flexi®Vector cloning can match or exceed that of recombination cloning for both initial capture of PCR products and for transfer between different vector combinations. These advantages lead to savings in time and cost, and fewer mutations present in the expression clones. There currently fewer vector options available for the Flexi®Vector system than for Gateway® and target genes must be screened for the presence of the Sgf I and Pme I sites.

Acknowledgements

This work was supported by a sponsored research agreement from Promega Corporation to B.G.F. and by the National Institutes of Health, Protein Structure Initiative grants P50 GM-64598 and U54 GM074901 (J.L. Markley, PI; G.N. Phillips, Jr., Co-Investigator; B.G.F.; Co-Investigator). B.G.F. is a consultant to Promega Corporation.

References

1. Thao, S., Zhao, Q., Kimball, T., Steffen, E., Blommel, P., Riters, M., Newman, C.S., Fox, B.G., and Wrobel, R.L. (2004) Results from high-throughput DNA cloning of *Arabidopsis thaliana* target genes using site-specific recombination. *J of Struct. Funct. Genom.* **5**, 267-276

2. Blommel, P., Martin, P, Wrobel, R, Steffen, E, and Fox, B.G. (2005) High efficiency single step production of expression plasmids from cDNA clones using the Flexi®Vector cloning system, *Prot. Expr. Purif.* Dec 5; [Epub ahead of print].

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (1 U54 GM074901)