Introducing the Flexi® Vector System

A New System for Cloning and Expressing Protein-Coding Regions

By Michael Slater, Ph.D., Jim Hartnett, M.S., Natalie Betz, Ph.D., Jami English, M.S., Ethan Strauss, Ph.D., Becky Pferdehirt, B.S., and Elaine Schenborn, Ph.D., Promega Corporation

Abstract

The Flexi® Vector Systems provide a rapid, efficient and high-fidelity method to transfer protein-coding regions between vectors with various expression or peptide tag options. Inserts are efficiently transferred between Flexi® Vectors following restriction enzyme digestion, maintaining insert orientation and reading frame. During the initial cloning step, a lethal gene in the Flexi® Vector cloning region allows positive selection for ligation of the insert. Antibiotic resistance genes simplify selection of the desired clone during subsequent transfer to other Flexi® Vectors.

The Flexi® Vector cloning strategy maintains insert orientation and reading frame, eliminating the need to resequence the insert after each transfer.

Introduction

The Flexi® Vector Systems[a,b] use two rare-cutting restriction enzymes, *Sgf* I[c] and *Pme* I[d], in a simple yet powerful directional cloning method for protein-coding sequences. Inserts are efficiently transferred to other Flexi® Vectors[a,b,e,f] following digestion with *Sgf* I and *Pme* I, maintaining insert orientation and reading frame and eliminating the need to resequence the insert after each transfer (Figure 1). This approach is easily adapted to high-throughput formats.

Unlike site-specific recombination vector systems, the Flexi® Vector Systems do not require appending multiple amino acids to the protein of interest. The systems do not require an archival entry vector and, for most applications, allow direct entry into the vector best suited to the experimental design (e.g., mammalian expression or N-terminal GST-fusion vectors).

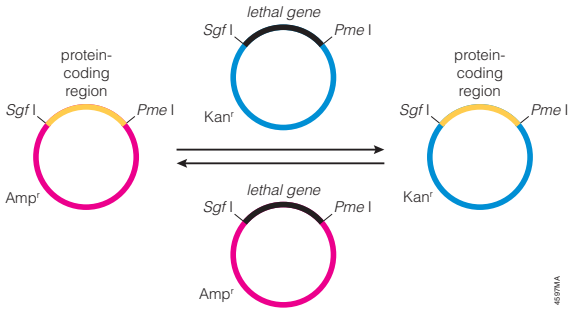

Figure 1. Transferring protein-coding regions in the Flexi® Vector Systems. The Flexi® Vector Systems use a flexible, directional cloning method for expressing protein-coding regions with or without peptide fusion tags. The features necessary for expression and any protein fusion tag are carried on the vector backbone, and the protein-coding region can be shuttled between vectors using two rare-cutting restriction endonucleases, *Sgf* I and *Pme* I. The Flexi® Vectors contain a lethal gene, barnase, for positive selection of the protein-coding sequence and an antibiotic resistance marker for selection of colonies containing the Flexi® Vector.

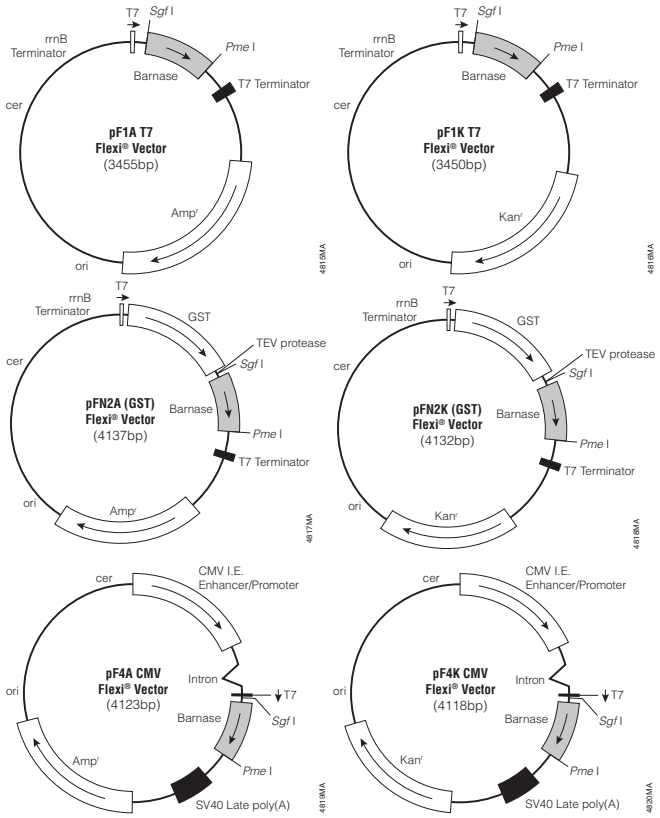

Figure 2. The Flexi® Vectors. The pF1A and pF1K T7 Flexi® Vectors[a,b] are designed for bacterial or in vitro protein expression via the T7 RNA polymerase promoter. The pFN2A and pFN2K (GST) Flexi® Vectors[a,b,e] append an N-terminal, glutathione-S-transferase (GST) peptide tag to a protein. The GST peptide tag can be removed by cleavage with TEV protease. The pF4A and pF4K CMV Flexi® Vectors[a,b,f] allow constitutive protein expression in mammalian cells using the human cytomegalovirus (CMV) intermediate-early enhancer/promoter. All of the Flexi® Vectors also contain the T7 RNA polymerase promoter for in vitro protein expression and a *cer* site, a 200bp site for the *E. coli* XerCD recombinase, which resolves plasmid multimers into monomers.

Any Flexi® Vector can act as an acceptor of a protein-coding region flanked by *Sgf* I and *Pme* I sites. All Flexi® Vectors carry the lethal barnase gene, which when replaced by the DNA fragment of interest, allows positive selection for the insert in bacterial strains commonly used for plasmid propagation and protein expression. Antibiotic resistance genes carried on the Flexi® Vectors facilitate transfer of protein-coding regions between vectors. The Flexi® Vectors contain various

Table 1. Frequencies of Restriction Enzyme Sites in Open Reading Frames.

Species	Number of ORFs	Mean ORF Size (bp)	ORF % GC	*Sgf* I	*Pme* I	*Pme* I or *Sgf* I	*Pac* I	*Swa* I	*Asc* I	*Bsi*W I	*Nru* I	*Sna*B I	*Pvu* I	*Fsp*A I	*Mlu* I	*Rsr* II	*Fse* I	*Not* I	*Srf* I
Homo sapiens	27,974	1503.8	52.42	0.42	0.70	1.12	0.87	2.01	2.15	3.23	3.36	3.56	3.63	3.85	4.04	4.46	4.53	5.73	6.02
Mus musculus	26,538	1370.5	51.65	0.32	0.81	1.13	0.59	1.43	1.10	3.02	3.04	3.34	4.04	3.88	3.36	3.15	2.49	3.37	3.38
Rattus norvegicus	22,845	1535.9	51.73	0.35	0.81	1.15	0.63	1.48	1.08	3.58	3.51	3.86	4.67	3.85	4.04	3.99	2.65	3.47	3.45
Caenorhabditis elegans	21,124	1312.1	42.80	0.81	0.75	1.54	0.77	2.86	0.23	7.44	15.72	7.27	20.51	1.94	9.54	3.37	0.34	0.89	0.18
Arabidopsis thaliana	28,951	1253.8	44.20	0.64	1.76	2.39	0.50	1.13	0.08	6.16	8.80	7.71	10.99	1.38	6.59	3.01	0.25	0.37	0.10
Saccharomyces cerevisiae	5,869	1492.4	39.62	0.46	2.52	2.96	3.27	6.13	0.22	8.96	8.43	14.26	8.86	2.42	5.40	1.43	0.15	0.49	0.17
Drosophila melanogaster	18,748	1692.7	53.82	5.03	0.71	5.66	0.62	1.86	2.02	14.90	28.02	8.81	33.37	11.33	15.45	10.65	1.43	5.51	1.27
Escherichia coli K124255	4,255	961.5	51.85	4.82	1.67	6.35	1.55	1.29	3.62	11.66	25.10	10.20	23.43	8.04	22.91	7.97	0.09	0.45	0.94

The 15 least frequent restriction sites in human open reading frames (ORF) are listed. Frequencies are expressed as the percent of ORFs that contain the indicated restriction site. Data was obtained from RefSeq Release 6; this full release incorporates genomic, transcript and protein data available as of July 5, 2004, and includes 1,050,975 proteins and sequences from 2,467 different organisms. The current RefSeq release is available at: ftp://ftp.ncbi.nih.gov/refseq/release/

expression or peptide tag options that enable expression of native or fusion proteins to study protein structure and function, or protein-protein interactions (Figure 2).

ORF Capture by Flexi® Vectors Using PCR

The Flexi® Vector Systems use two infrequently cutting restriction endonucleases: *Sgf* I and *Pme* I. *Sgf* I has the fewest restriction sites in human open reading frame (ORF) sequences, and *Pme* I has the second fewest. These enzymes also cut infrequently in the open reading frames of many other organisms (Table 1). Most (99%) of the annotated human open reading frames are not impacted by the use of these restriction enzymes for directional cloning. However, we recommend scanning your protein-coding region for *Sgf* I and *Pme* I sites. If your protein-coding region contains these sites, consider cloning a portion of the protein-coding region or using RecA protein to protect *Sgf* I or *Pme* I sites within the protein-coding region from digestion (1). Alternatively, PCR-based, site-directed mutagenesis methods (2,3) can be used to mutate restriction enzyme sites without changing the amino acid sequence.

Sgf I and *Pme* I sites are appended to the protein-coding region by PCR. To facilitate cloning, the 5′ (forward) primer used to amplify the protein-coding region has an *Sgf* I site appended, and the 3′ (reverse) primer has a *Pme* I site appended (Figure 3). The *Sgf* I site is placed one base upstream of the start codon. This allows de novo initiation at the native translation start site in Flexi® Vectors that do not produce fusion proteins and allows readthrough of the *Sgf* I site in Flexi® Vectors that produce N-terminal fusion proteins. The *Pme* I site is placed at the carboxy terminus, appending a single valine residue to the last amino acid of the protein-coding region. The valine codon, GTT, is immediately followed by an ochre stop codon, TAA. When a protein-coding region, flanked by *Sgf* I and *Pme* I sites, is cloned into a vector cut with *Sgf* I and *Pme* I, the translation stop codon is recreated. You can design your primers to place the *Pme* I site downstream of the native stop codon, so translation terminates at the native stop codon and a valine residue is not appended to the protein. However, in doing so you lose the ability to express the protein with a carboxy-terminal fusion tag in the future. A carboxy-terminal fusion protein can be created by fusing the blunt *Pme* I end of the protein-coding region with a blunt end generated by a different restriction enzyme (e.g., *Eco*ICR I). Primer design software is provided at: **www.promega.com/techserv/tools/flexivector/**

Cloning *Renilla* Luciferase into the pF1A T7 Flexi® Vector

The synthetic *Renilla* luciferase (hRL) protein-coding region, which has codons optimized for mammalian expression (933bp), was amplified by PCR[g] using *Pfu* DNA polymerase. An overview of the cloning strategy is shown in Figure 4; protocol details can be found in the *Flexi® Vector Systems Technical Manual* #TM254. Briefly, the 958bp *Renilla* luciferase PCR product was purified and digested with the Flexi® Enzyme Blend (*Sgf* I and *Pme* I)[c,d] to yield a 941bp fragment, which was then cleaned up to remove the small oligonucleotides released by the restriction enzyme digestion. Postamplification and postdigestion purifications were performed using the Wizard® SV Gel and PCR Clean-Up System (Cat.# A9281). The 941bp hRL product was ligated into a pF1A T7 Flexi® Vector that had been digested with the Flexi® Enzyme Blend (*Sgf* I and *Pme* I). The ligation mixture was transformed into competent JM109(DE3) cells, and the cells were selected on LB plates supplemented with ampicillin.

The resultant clones were screened for *Renilla* luciferase activity and the presence of *Sgf* I and *Pme* I sites. Most (93.7%, 267/285) of the ampicillin-resistant colonies

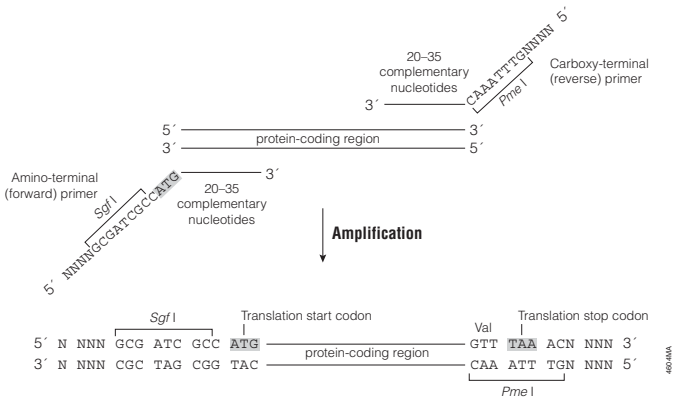

Figure 3. Amplification to append *Sgf* I and *Pme* I sites. The *Sgf* I site is upstream of the start codon. This allows de novo initiation at the start site or readthrough to append an amino-terminal peptide, depending on the vector backbone. Addition of a *Pme* I site appends a single valine codon at the 3′ end of the protein-coding region and allows either termination or readthrough to append a carboxy-terminal peptide, depending on the vector backbone. To amplify the *Renilla* luciferase gene, the amino (5′) primer sequence was GTCAGCGATCGCCATGGCTTCCAAGGTGTACGAC, and the carboxy (3′) primer sequence was CTTCGTTTAAACCTGCTCGTTCTTCAGCACGC.

contained the hRL protein-coding region, and of those hRL positive clones, 94.5% (86/91) could be cut with both *Sgf* I and *Pme* I, as determined by agarose gel electrophoresis. To calculate the number of colonies to screen (N), we assumed a 99% accuracy rate for the high-fidelity amplification (4) and set the probability of finding the desired clone at 99.9%.

$$N = \ln(1–P) / \ln(1–f)$$

where P is the desired probability and f is the fractional portion of the desired clone, which is the product of amplification accuracy and insertion and recut frequencies. In this case, P = 0.999,
f = (0.990)(0.937)(0.945) = 0.8766 and the resulting N = 3.3.

Thus, the desired clone should be identified by screening only 3 or 4 colonies. Three pF1A-hRL clones were sequenced to confirm the insert sequence, and no unwanted variations were noted. One clone was chosen for further studies.

High-Efficiency Transfer Between Flexi® Vectors

We measured the transfer frequency between various Flexi® Vectors by screening plasmid DNA for the presence of the correctly sized *Sgf* I / *Pme* I DNA fragment encoding the synthetic *Renilla* luciferase gene. Briefly, the donor hRL-Flexi® Vectors were paired with acceptor Flexi® Vectors with the opposite drug resistance. Donor and acceptor plasmid DNAs were combined, digested simultaneously with the Flexi® Enzyme Blend (*Sgf* I and *Pme* I) and ligated (Figure 5). Ligated products were transformed into JM109(DE3), and cells were selected on LB plates supplemented with the appropriate drug for selection of the acceptor vector. The resultant clones were screened for *Renilla* luciferase activity and the presence of *Sgf* I and *Pme* I sites. Transfer frequencies between Flexi® Vectors ranged from 89.5% to 100%, with an average transfer frequency of 95.2% (Table 2). For the average and worst-performing transfer reactions, N = 2.3 and 3.1, respectively. Therefore, 2 to 3 clones should be adequate for screening in these experiments.

Expression of *Renilla* Luciferase in JM109(DE3)

Expression levels of synthetic *Renilla* luciferase were measured for the pF1K-hRL and pFN2K-hRL vectors in JM109(DE3), a strain that expresses the T7 RNA polymerase gene under the control of an IPTG-inducible promoter. Following IPTG induction, each culture was diluted 1:1,000 with LB medium, then 15µl of diluted culture was mixed with 15µl of 2X *Renilla* Luciferase Assay Lysis Buffer. This mixture was frozen at –70°C and thawed once to maximize lysis. *Renilla* Luciferase Assay Reagent (100µl, Cat.# E2810) was added, and light output was measured with an EG&G plate-reading luminometer. Both vectors expressed *Renilla* luciferase (Table 3). Coomassie®-stained SDS-polyacrylamide gels showed native protein (pF1K-hRL) or GST-fusion protein (pFN2K-hRL) of the correct size with no detectable

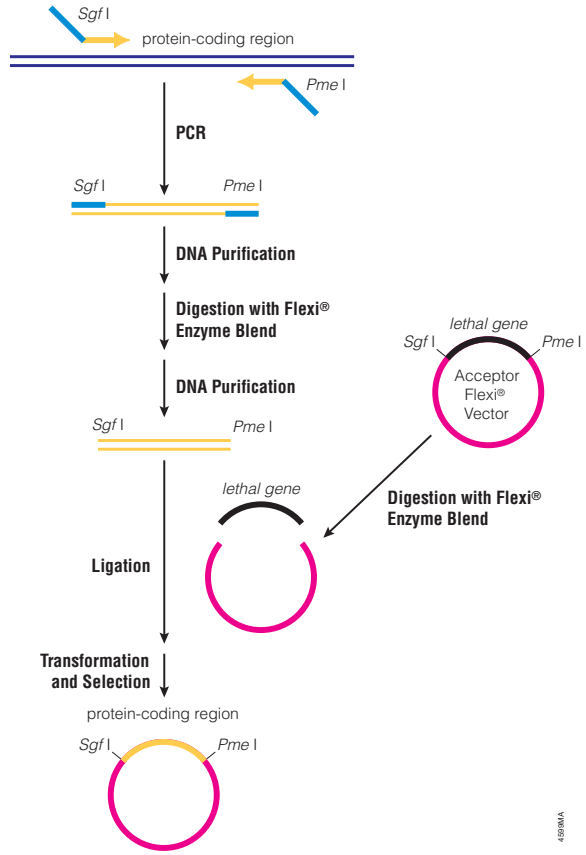

Figure 4. Cloning a protein-coding region into the Flexi® Vectors. PCR primers append *Sgf* I and *Pme* I sites onto the protein-coding region. After amplification, the PCR product is purified to remove the DNA polymerase and primers then digested with *Sgf* I and *Pme* I. The DNA is purified again to remove small oligonucleotides released by the restriction enzymes. The digested PCR product is ligated into an acceptor Flexi® Vector that has been digested with *Sgf* I and *Pme* I. Following transformation, the cells are selected with the appropriate antibiotic for the acceptor Flexi® Vector used.

Table 2. Transfer Efficiencies Between Flexi® Vectors.

Donor	Acceptor Flexi® Vector					
Flexi® Vector	pF1A	pF1K	pFN2A	pFN2K	pF4A	pF4K
pF1A-hRL		95.3%		97.8%		89.5%
pF1K-hRL	95.6%		91.0%		96.9%	
pF2A-hRL		100%		93.3%		96.8%
pF4K-hRL	96.7%		93.8%		95.8%	

For transfer of the hRL sequence from pF1A-hRL to the F1K Vector, 190 colonies were analyzed. For all other transfers, the number of colonies analyzed ranged from 89 to 96 with an average of 93 colonies.

Table 3. *Renilla* Luciferase Activity in JM109(DE3).

Vector	Average Luminescence (RLU)*
Untransformed JM109(DE3)	100 ± 10
pF1K-hRL	42,487 ± 1,775
pFN2K-hRL	151,329 ± 17,444

*RLU factor = 1.0; delay time = 2.0 seconds; read time = 2.0 seconds. Results are expressed as the mean ± standard deviation, n = 4.

degradation (data not shown). Thus, the T7 RNA polymerase promoter in the pF1K and pFN2K Flexi® Vectors is functional to drive expression of native and GST-fusion proteins, respectively, in vivo in *E. coli*.

Expression of *Renilla* Luciferase in Mammalian Cells

Chinese hamster ovary (CHO) cells were transfected with the pF4K-hRL or phRL-CMV Vector (Cat.# E6271) using the TransFast™ Transfection Reagent (Cat.# E2431). The pGL4.13 [*luc2*/SV40] Vector (Cat.# E6681), which constitutively expresses firefly luciferase, was cotransfected to normalize for differences in transfection efficiency. Luciferase levels were measured using the Dual-Glo™ Luciferase Assay System (Cat.# E2920). *Renilla* luciferase levels are reported in Table 4. Firefly luciferase levels and the ratio of *Renilla* and firefly luciferase levels were similar for both pF4K-hRL and phRL-CMV vectors. Thus, protein expression levels from the pF4 Flexi® Vectors are comparable to those from other CMV-driven expression vectors.

Expression of *Renilla* Luciferase in TnT® Lysates

In vitro expression of synthetic *Renilla* luciferase for the Flexi® Vector clones was assayed with the TnT® T7 Quick Coupled Transcription/Translation System (rabbit reticulocyte lysate-based, Cat.# L1170). Plasmid DNA (500ng) was added to a 25µl TnT® reaction and incubated at 30°C for 90 minutes. Two positive controls were included: pRL-SV40 (Cat.# E2231) and phRL-null (Cat.# E6231). Synthesized protein was labeled with

Table 4. *Renilla* Luciferase Activity in CHO Cells Transfected with the pF4K-hRL or phRL-CMV Vector.

Vector	Average Luminescence (RLU)*
pF4K-hRL	2,806,000 ± 377,754
phRL-CMV	3,456,183 ± 768,855

*Results are expressed as the mean ± standard deviation, n = 3.

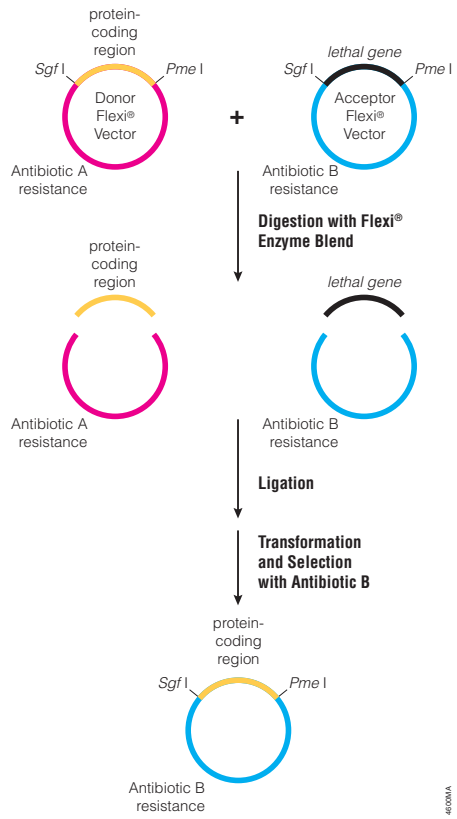

Figure 5. Transfer of a protein-coding region between Flexi® Vectors. The donor Flexi® Vector containing the protein-coding region is mixed with an acceptor Flexi® Vector that has a different antibiotic resistance. The two plasmids are digested with Flexi® Enzyme Blend (*Sgf* I and *Pme* I), and the mixture is ligated and transformed into *E. coli*. The cells are plated on the appropriate selective media for the acceptor Flexi® Vector.

the FluoroTect™ $Green_{Lys}$ in vitro Translation Labeling System (Cat.# L5001). An aliquot of the TnT® reaction (2.5µl) was added to 100µl of *Renilla* Luciferase Assay Reagent, and luminescence was measured using an Orion Microplate Luminometer (Berthold Detection Systems). Results are shown in Table 5. Proteins were separated by SDS-PAGE. The synthetic *Renilla* luciferase and *Renilla* luciferase GST-fusion protein were of the expected sizes (data not shown). These results confirm the functionality of the T7 RNA polymerase promoter in the Flexi® Vectors for protein expression in TnT® Systems. Expression levels from different vectors were not identical; differences may be attributed to fusion protein production (e.g., pFN2K-hRL) or differing 5′ and 3′ sequences flanking the protein-coding region.

Conclusions

We have developed a new directional cloning system for protein-coding sequences based on two rare-cutting restriction enzymes, *Sgf* I and *Pme* I. The Flexi® Vector Systems provide a rapid, efficient and high-fidelity method to transfer protein-coding regions between a variety of vectors. Unlike site-specific recombination

Table 5. *Renilla* Luciferase Activity in the TnT® Quick Coupled Transcription/Translation System.

Vector	Average Luminescence (RLU)
No-DNA control	167.5 ± 12.0
pF1K-hRL	1,106,418 ± 155,552
pFN2K-hRL	292,691 ± 38,055
pF4K-hRL	838,061 ± 61,998
pRL-SV40 (Cat.# E2231)	190,797 ± 54,347
phRL-null (Cat.# E6231)	462,695 ± 7,525

Delay time = 2 seconds, read time = 5 seconds. Results are expressed as the mean ± standard deviation, n = 9 for the pF1K-hRL, pFN2K-hRL and pF4K-hRL vectors, n = 3 for the pRL-SV40 and phRL-null Vectors, n = 2 for the no-DNA control.

vector systems, the Flexi® Vector Systems do not require appending multiple amino acids to the amino- or carboxy-terminus of the protein or domain of interest. In addition, the systems use routine and robust molecular biology techniques and reagents.

We demonstrate efficient capture of protein-coding regions as inserts in the Flexi® Vectors following PCR. Inserts were easily transferred with >90% efficiency to other Flexi® Vectors that were digested with *Sgf* I and *Pme* I. This cloning strategy maintains insert orientation and reading frame and eliminates the need to resequence the insert after each transfer. We also show that the Flexi® Vectors enable expression of native or fusion proteins in *E. coli*, mammalian cells and in vitro translation systems. In the future, additional vectors with new expression options, such as the production of C-terminal fusion proteins, will be available.

References

1. Schoenfeld, T., Harper, T. and Slater, M. (1995) *Promega Notes* **50**, 9–13.
2. Higuchi, R., Krummel, B. and Saiki, R.K. (1988) *Nucl. Acids Res.* **16**, 7351–67.
3. Ho, S.N. *et al.* (1989) *Gene* **77**, 51–9.
4. Slater, M. *et al.* (1998) *Promega Notes* **68**, 7–10.

Protocol

◆ *Flexi® Vector Systems Technical Manual* #TM254. Promega Corporation. **www.promega.com/tbs/tm254/tm254.html**

Ordering Information

Product	Size	Cat.#
pF1A T7 Flexi® Vector	20µg	C8441
pF1K T7 Flexi® Vector	20µg	C8451
pFN2A (GST) Flexi® Vector	20µg	C8461
pFN2K (GST) Flexi® Vector	20µg	C8471
pF4A CMV Flexi® Vector	20µg	C8481
pF4K CMV Flexi® Vector	20µg	C8491
Flexi® System, Entry/Transfer	1 system	C8640
Flexi® Vector System, Transfer	1 system	C8820
10X Flexi® Enzyme Blend (*Sgf* I and *Pme* I)	25µl	R1851
	100µl	R1852

Michael Slater, Ph.D.
Project Leader

Jim Hartnett, M.S.
Research Scientist

Natalie Betz, Ph.D.
Applications Scientist

Jami English, M.S.
Research Scientist

Elaine Schenborn, Ph.D.
Project Manager

Ethan Strauss, Ph.D.
Technical Services Scientist

Not Pictured
Becky Pferdehirt, B.S. *Research Scientist*

[a] Patent Pending.

[b] For research use only. Persons wishing to use this product or its derivatives in other fields of use, including without limitation, commercial sale, diagnostics or therapeutics, should contact Promega Corporation for licensing information.

[c] U.S. Pat. No. 5,391,487.

[d] Licensed under U.S. Pat. No. 5,945,288.

[e] This product or portions thereof is manufactured under license from Amrad Corporation Limited. For non-commercial research use only. All other uses require a license from Amrad Corporation Limited, Unit 6, 663 Victoria Street, Abbotsford, Victoria 3067, Australia, under U.S. Pat. No. 5,654,176, Australian Pat. No. 607511, Canadian Pat. No. 1338903 and other issued patents.

[f] The CMV promoter and its use are covered under U.S. Pat. Nos. 5,168,062 and 5,385,839 owned by the University of Iowa Research Foundation, Iowa City, Iowa, and licensed FOR RESEARCH USE ONLY. Commercial users must obtain a license to these patents directly from the University of Iowa Research Foundation.

[g] The PCR process is covered by patents issued and applicable in certain countries*. Promega does not encourage or support the unauthorized or unlicensed use of the PCR process.
*In the U.S., effective March 29, 2005, U.S. Pat. Nos. 4,683,195, 4,965,188 and 4,683,202 will expire. In Europe, effective March 28, 2006, European Pat. Nos. 201,184 and 200,362 will expire.

Products may be covered by pending or issued patents or may have certain limitations. Please visit our Web site for more information.