0273-2289/1559–0291 (Online)/07/136–140/703–710/$30.00

Comparison of Multiple Gene Assembly Methods for Metabolic Engineering

CHENFENG LU,[1] KAREN MANSOORABADI,[2] AND THOMAS JEFFRIES*,[1–3]

[1]*Department of Food Science;* [2]*Department of Bacteriology, University of Wisconsin, Madison, WI 53706, E-mail: twjeffri@wisc.edu; and* [3]*USDA Forest Service, Forest Products Laboratory, Madison, WI 53726*

Abstract

A universal, rapid DNA assembly method for efficient multigene plasmid construction is important for biological research and for optimizing gene expression in industrial microbes. Three different approaches to achieve this goal were evaluated. These included creating long complementary extensions using a uracil-DNA glycosylase technique, overlap extension polymerase chain reaction, and a *Sfi*I-based ligation method. *Sfi*I ligation was the only successful approach for assembling large DNA fragments that contained repeated homologous regions. In addition, the *Sfi*I method has been improved over a similar, previous published technique so that it is more flexible and does not require polymerase chain reaction to incorporate adaptors. In the present study, *Saccharomyces cerevisiae* genes *TAL1*, *TKL1*, and *PYK1* under control of the 6-phosphogluconate dehydrogenase promoter were successfully ligated together using multiple unique *Sfi*I restriction sites. The desired construct was obtained 65% of the time during vector construction using four-piece ligations. The *Sfi*I method consists of three steps: first a *Sfi*I linker vector is constructed, whose multiple cloning site is flanked by two three-base linkers matching the neighboring *Sfi*I linkers on *Sfi*I digestion; second, the linkers are attached to the desired genes by cloning them into *Sfi*I linker vectors; third, the genes flanked by the three-base linkers, are released by *Sfi*I digestion. In the final step, genes of interest are joined together in a simple one-step ligation.

Index Entries: Gene expression; cloning; ligation; optimization; xylose; yeast.

Introduction

Most metabolic engineering approaches use genetic tools to alter a target metabolic pathway. This is usually done by sequential gene overexpression, deletion, or promoter replacement to alter gene activity *(1)*. To implement these tools, traditional cloning methods based on restriction sites are often used. This is sufficient in the construction of single insert or knockout vectors. Tweaking the activity of a single gene, however; usually

*Author to whom all correspondence and reprint requests should be addressed.

has little effect on the metabolite flux *(2,3)*, because flux control is usually exerted at multiple steps. Also changing expression level of a single gene is not usually enough to perturb fluxes in a tightly regulated pathway, such as the glycolytic pathway *(4)*. In addition, engineering a *de novo* metabolic pathway for a more complex product, such as polyketide, usually requires multiple steps *(5)*. This makes the identification of target genes for strain improvement difficult. If one wishes to approach the optimal expression or activity of several different genes in a pathway by altering multiple promoters or by using genes for enzymes with various kinetic or regulatory properties, the problem spirals out of control. Simultaneous multiple-gene overexpression could help overcome such problems, so we have developed a multigene vector construction approach as a tool for strain improvement *(6)*.

Construction of multigene vectors, especially those in which multiple variants are incorporated, calls for a more efficient vector assembly method than what can be achieved with single gene inserts. Conventional cloning methods based on restriction sites are too time-consuming and awkward for this purpose. As the number and length of gene inserts increases, the efficiency of vector assembly decreases. If multiple genes and multiple variants of each gene are required in each vector, the problem becomes similar to that encountered in assembling a library, and the efficiency of target vector construction decreases rapidly. Although, direct construction from overlapping oligonucleotides is feasible with smaller genes, the error frequency inherent with oligonucleotide synthesis *(7)* prohibits direct construction of larger genes.

Several techniques that are potentially suitable for multiple gene assembly, including methods involving exonuclease *(8–11)*, uracil-DNA glycosylase (UDG) *(12,13)*, polymerase chain reaction (PCR) *(14)*, and enzymes with degenerate recognition sequence such as *Sfi*I *(15)*. These methods, although diverse, are based on two mechanisms: either create compatible but different overhangs for each DNA fragment, or piece together unique fragments using overlap extension PCR. The exonuclease, UDG, and *Sfi*I methods belong to the first category. In the exonuclease method, a mild exonuclease, such as T4 DNA polymerase, is used to create a 5′ overhang by removing bases from one DNA strand from 3′ to 5′. The 5′ region of each DNA fragment is homologous to the 3′ end of the neighboring DNA fragment, thus creating sticky ends *(10)*. Multiple DNA fragments can be joined together by exonuclease treatment and ligation. The UDG method also generates sticky ends. Primers containing uracil bases are used to amplify the genes of interest. The three prime (3′) overhangs are created by releasing the uracil bases from the dsDNA fragments with UDG treatment *(12)*. The *Sfi*I method takes advantage of the degenerate *Sfi*I restriction site, and produces 3′ overhangs by *Sfi*I digestion. Most PCR methods are based on the overlap extension mechanism. Although these

methods are able to construct a multigene assembly, they require several PCR amplifications of each DNA fragment and are not versatile enough for assembling large multicistronic vectors. In this study, multigene assembly by joint PCR, UDG, and *Sfi*I was evaluated and a new PCR-free *Sfi*I method was developed, which we have shown to be efficient, flexible, and able to join relatively large DNA fragments (>1 kb).

Materials and Methods

DNA Manipulation

Escherichia coli DH5 (Gibco-BRL, Gaithersburg, MD) was routinely used for transformation of ligation products. The transformants obtained were plated on ampicillin containing Luria Bertani medium. PCR amplifications were performed in 50 µL volumes containing primers (0.5 m*M* each) custom-made by Invitrogen (Carlsbad, CA), deoxynucleotide triphospates in varying concentrations, chromosomal DNA (0.5 mg), and Taq polymerase (1 U) in the buffer recommended by the manufacturer (Promega, Madison, WI). Temperature cycling was performed by a programmable thermocycler (PTC-200 thermal cycler; MJ Research Inc., Watertown, MA) following standard protocols with minor modifications based on specific primers and amplification results.

UDG Method

UDG method was performed with minor modification to the published protocol *(12)*. After UDG treatment, the reaction mix was incubated at 70°C for 10 min to disassociate the uracil-containing single stranded DNA. The DNA fragments were purified with a Qiaquick PCR purification kit (Qiagen, Hilden, Germany) and ligated in equal molar using T4 DNA ligase (New England Lab, Ipswich, MA) overnight.

SfiI Method

The DNA inserts consisting of a 2.3 kb 6-phosphogluconate dehydrogenase (*GND2*) promoter-controlled transaldolase (*TAL1*) gene, 3.4 kb *GND*2 promoter-controlled transketolase (*TKL1*) gene, and 2.9 kb *GND2* promoter-controlled pyruvate kinase (*PYK1*) gene, were prepared by PCR from genomic *Saccharomyces cerevisiae* DNA. The ligation strategy is shown in Fig. 1. The vector p*Sfi*-linker was constructed by subcloning the modified multiple cloning site (MCS) into pBluescript (Stratagene, La Jolla, CA) through *Sac*I and *Kpn*I sites. This modified MCS was flanked by three different *Sfi*I sites, each of which contained the desired three base overhang. Each modified MCS was synthesized by phosphorylation of two pairs of o*Sfi*-linker oligos using T4 polynucleotidekinase (New England Lab) and annealing at room temperature. The adaptors for the three *Sfi*I-linkers

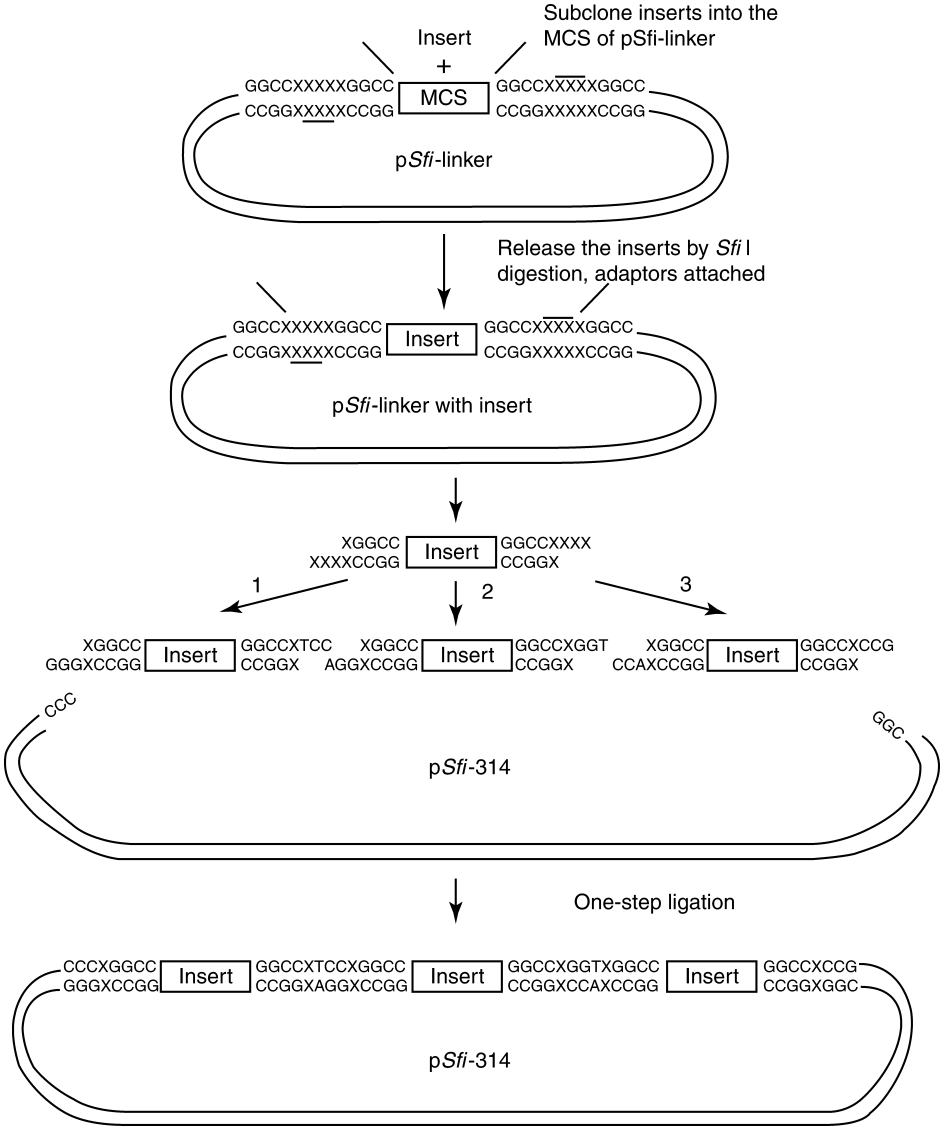

Fig. 1. Experimental strategy for PCR free *Sfi*I assembly method. In the first step, desired inserts were subcloned to the p*Sfi*-linker vector through *Apa*I and *Bam*HI. In the second step, inserts were released from the linker vector by *Sfi*I digestion, three base adaptors were attached to 5′ ends of the insertion on digestion. Finally inserts were joined together through these compatible adaptors.

were CCC-TCC, TCC-CCT, and CCT-CCG. For one-step assembly of the three inserts and p*Sfi*-314 *(16)*, which has the adaptor CCC-CCG, each promoter-gene cassette was subcloned to the p*Sfi*-linker through *Apa*I and *Bam*HI sites, digested with *Sfi*I, gel purified using Geneclean kit (Qbiogene, Morgan Irvine, CA), and ligated together with p*Sfi*-314 through *Sfi*I sites in equal molar concentration using T4 DNA ligase (Table 1).

Table 1
Oligos Used in This Study[a]

Oligo name	Description	Sequence
ocfl 113	Amplify *GND*2 promoter region for UDG method	AGAACTAGTGGATCCCC CUCCGTCATAAC TTTGA ATCCT GTCAT
ocfl 114	Amplify *RPE*1 open reading frame (ORF) downstream with 23-mer linker to *GND*2 promoter upstream for UDG method	ATTCAAAGTTATGACGGC GAAAUGGA TATT GATCTAGATGGC
ocfl 115	Amplify *GND*2 promoter region and 23-mer linker with cfl114 for UDG method	ATTTCGCCGTCATAA CTTTGAAUCCTGTCAT
ocfl 116	Amplify *TAL1* ORF downstream with 22-mer linker to *GND*2 promoter for UDG method	AGTTATGACGGGGGACG TTGAUTTAAGGTGGTTCC
oGND2p-for	Forward primer to amplify *GND*2 promoter	GGGCCCCCGTCATAACT TTGAATCCTGTCAT
oGND2p-rev	Reverse primer to amplify *GND*2 promoter	GTCGACTCTGTTCCTCG TGTTTTTTTAATTGTAG
oTAL1-for	Forward primer to amplify *TAL1* ORF	CTCGAGATGTCTGAAC CA GCTCAAAAGAAAC
oTAL1-rev	Reverse primer to amplify *TAL1* ORF	GGATCCGGGACGTTGA TTTAAGGTGGTTCC
oTKL1-for	Forward primer to amplify *TKL1* ORF	CTCGAGATGACTCAATTCA CTGACATTGATAAGC
oTKL1-rev	Reverse primer to amplify *TKL1* ORF	GGATCCTTCTTTATTGGCT TTATACTTGAATGGTG
oPYK1-for	Forward primer to amplify *PYK1* ORF	CTCGAGATGTCTAGATTA GAAAGATT GACCTCAT TAAACG
oPYK1-rev	Reverse primer to amplify *PYK1* ORF	GGATCCGAATTTTTAGC GTATCCTTTCGC C
oSfi-linker1	One of the four oligos to make *Sfi*I-linker vector	TGGCCXXXXXGGCCCAC CGCGGTGGCGGCCGCTC TAGAACTAGTGGATCCC CCGGGCTGCAGGAA[b]
oSfi-linker2	One of the four oligos to make *Sfi*I-linker vector	TTCGATATCAAGCTTATC GATACCGTCGACCTC GAGGGGGGGCCCGGC CXXXXXGGCCTGTAC[b]
oSfi-linker3	One of the four oligos to make *Sfi*I-linker vector	AGGCCXXXXXGGCCGGGC CCCCCCTCGAGGT CGACGGTATCGATAA GCTTGATATCGAA TTCCTG[b]

(Continued)

Table 1 *(Continued)*

Oligo name	Description	Sequence
oSfi-linker4	One of the four oligos to make *Sfi*I-linker vector	CAGCCCGGGGGATCCAC TAGTTCTAGAGCGGC CGCCACCGCGGTGGGC CXXXXXGGCCAAGCT[b]

[a]*XYL3* orf was amplified from *Pichia stipitis* wild type strain CBS6054, all other promoters and genes are amplified from *S. cerevisiae* strain YSX3. All PCR products were subcloned to TOPO vector (TOPO PCR 2.1 kit, Invitrogen) and sequenced before further cloning.

[b]XXXXX is CCCCC, ATCCA, TGGAT, GGGGG for insert 1 (TAL1 cassette); ATCCA, ACCTA, TAGGT, TGGAT for insert 2 (*TKL1* cassette); and ACCTA, ACCGA, TCGGT, TAGGT for insert 3 (*PYK1* cassette).

Results and Discussion

We were unsuccessful in using PCR method to assemble DNA fragments with homologous regions. In our preliminary experiments, the joint PCR method worked well for joining DNA fragments without significant homology, but failed to join DNA fragments with significant homology, such as two genes driven by the same promoter (data not shown). This is probably because of the binding of homologous regions. The UDG method is based on the use of primers that contain dUTP at the 5′ end. A 3′ sticky end is formed by treatment of the PCR products with UDG, which releases the uracil bases *(12)*. Pairwise ligation was used to determine the quality of the ligation products formed with different fragments. The pairwise ligation pattern at different time-points is shown in Fig. 2. Some smearing was observed after the 30 min incubation. This became stronger with prolonged incubation, indicating residual exonuclease activity in the enzyme UDG. Hence, there could be some nonspecific activity resulting in degradation of the DNA. This may have contributed to the unsuccessful multiple DNA assembly of longer inserts (2 kb or larger) (data not shown).

Experimental Strategy and Results of PCR-Free SfiI Method

A previous published *Sfi*I method *(15)* was modified by incorporating linker vectors into the assembly process (*see* Fig. 2). The linker vector was constructed by subcloning a 127 bp fragment into the pBluescript. This 127 bp fragment is generated by annealing four oligomers, which contain polylinkers, two *Sfi*I sites, and all the restriction sites in the pBluescript MCS, except *Kpn*I and *Sac*I. It adds versatility to the existing *Sfi*I method *(15)* and does not require PCR to incorporate the *Sfi*I sites. Adaptors were attached to the 5′ and 3′ ends of the DNA fragment by subcloning it into the linker vector and subsequent *Sfi*I digestion. After the fragment was attached to the adaptors onto both ends, each DNA fragment was gel purified and ligated using T4 DNA ligase in a final volume of 50 μL in

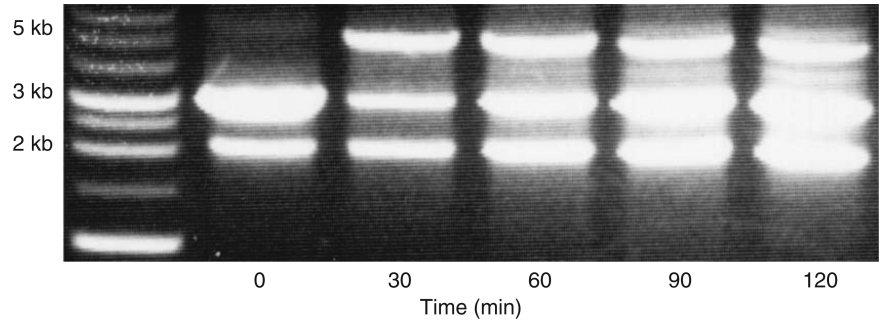

Fig. 2. Effect of prolonged UDG incubation on pairwise ligation. *Rpe*1 cassette (2.0 kb) and *TAL1* (2.4 kb) cassette were ligated in vitro at varied incubation time. At each time-point, the reaction was stopped by heating the samples at 70°C for 10 min. Subsequently samples were purified and ligated overnight at room temperature. Ligated products were loaded onto a 0.8% agarose gel.

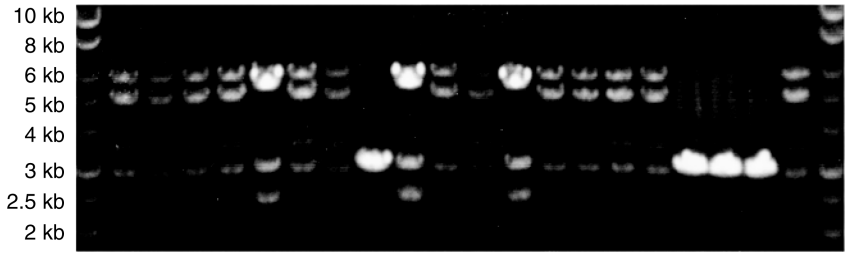

Fig. 3. Restriction digestion of the four piece ligation clones. About 0.5 µg of plasmid DNA were digested by 20 U *Sfi*I enzyme for 1 h in a total volume of 20 µL, and 5 µL samples were loaded onto a 0.8% agarose gel. The first and last lane were loaded with 0.2 µg of 1 kb DNA ladder (Promega). The rest of the lanes were loaded with *Sfi*I digested four piece ligation products.

equal molar. The ligation was performed at room temperature overnight. About 1.0–2.0 ng of ligation product was transformed into *E. coli* DH5α and plated on Luria Bertani-ampicillin. Plasmid DNA was isolated from 20 of the 325 ampicillin-resistant colonies. Restriction analysis showed 65% of the recovered plasmids to be correct. Results from the restriction analysis are shown in Fig. 3. Bands are present at 4.9 kb (vector), 2.3 kb (*TAL1*), 3.4 kb (*TKL1*), 2.9 kb (*PYK1*), and 5.7 kb (dimeric product of *TAL1* and *TKL1*) in Fig. 3. This dimeric digestion product likely caused the higher band intensity of *PYK1*, and lower intensities for *TAL1* and *TKL1*. Four of the correct construct were subject to additional PCR verification, and proved to be bonafide constructs (data not shown). Among the 20 plasmids subject to restriction digestion, three were nonspecific ligation products, showing band sizes of 2.3 kb (*TAL1*) and 2.9 kb (*PYK1*), missing the middle DNA *TKL1* cassette. The other three incorrect constructs were beyond our understanding. These results demonstrated that our PCR-free *Sfi*I method successfully ligated four DNA fragments together in a very flexible way: each

insert is interchangeable and can be easily replaced. Among the five nucleotide degenerate nucleotides in the *Sfi*I site, three are capable of forming "sticky" overlaps *(17)* and can be used as an adaptor. This gives $4^3 = 64$ choices of adaptors. In practice, adaptors consisting of three A/T bases do not work well *(15)*, probably because of less hydrogen bonding than G/C rich adaptors. We observed that adaptors containing at least two C/G worked much better. There are P(2 G/C) + P(3 G/C) = 32 possible adaptors, which is plenty for multiple DNA fragment assembly.

Conclusions

Three multiple DNA assembly methods were tested. The joint PCR method and the UDG method were not successful for assembling DNA fragments. A novel assembly method based on *Sfi*I was developed and found to be efficient for joining multiple DNA fragments. Sixty five percentage of the four-piece ligation clones were correct after restriction digestion and PCR confirmation. In addition, it is versatile and does not require PCR to incorporate adaptors. Adaptors can be attached to any insert by subcloning it into a *Sfi*I-linker vector and releasing by *Sfi*I digestion.

Acknowledgment

This work was supported by NIH grant GM67933-03 to TWJ.

References

1. Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005), *Proc. Natl. Acad. Sci. USA* **102,** 12,678–12,683.
2. Cornish-Bowden, A. and Hofmeyr, J. H. (1994), *Biochem. J.* **298 (Pt 2),** 367–375.
3. van Heeswijk, W. C., Bakker, B. M., Teusink, B., et al. (1999), *Biochem. Soc. Trans.* **27,** 261–264.
4. Schaaff, I., Heinisch, J., and Zimmermann, F. K. (1989), *Yeast* **5,** 285–290.
5. Palaniappan, N., Kim, B. S., Sekiyama, Y., Osada, H., and Reynolds, K. A. (2003), *J. Biol. Chem.* **278,** 35,552–35,557.
6. Bailey, J. E. (1999), *Nat. Biotechnol.* **17,** 616–618.
7. Gao, X., Yo, P., Keith, A., Ragan, T. J., and Harris, T. K. (2003), *Nucleic Acids Res.* **31,** E143.
8. Yang, Y. S., Watson, W. J., Tucker, P. W., and Capra, J. D. (1993), *Nucleic Acids Res.* **21,** 1889–1893.
9. Kuijper, J. L., Wiren, K. M., Mathies, L. D., Gray, C. L., and Hagen, F. S. (1992), *Gene* **112,** 147–155.
10. Hsiao, K. (1993), *Nucleic Acids Res.* **21,** 5528–5529.
11. Lohff, C. J. and Cease, K. B. (1992), *Nucleic Acids Res.* **20,** 144.
12. Rashtchian, A., Buchman, G. W., Schuster, D. M., and Berninger, M. S. (1992), *Anal. Biochem.* **206,** 91–97.
13. Booth, P. M., Buchman, G. W., and Rashtchian, A. (1994), *Gene* **146,** 303–308.
14. Pachuk, C. J., Samuel, M., Zurawski, J. A., Snyder, L., Phillips, P., and Satishchandran, C. (2000), *Gene* **243,** 19–25.
15. Tsuge, K., Matsui, K., and Itaya, M. (2003), *Nucleic Acids Res.* **31,** E133.
16. Ni, H. Y. (2004), *Master Thesis,* University of Wisconsin, Madison, USA.
17. Qiang, B. Q. and Schildkraut, I. (1984), *Nucleic Acids Res.* **12,** 4507–4516.