Benchmarks

replacement" illustrated here may be useful in distinguishing expression levels of both transgenic and endogenous products of a wide variety of bioeffectors in which the transgenic impacts expression of its endogenous counterpart.

REFERENCES

1. **Feng, S., F. Wang, A. Matsubara, M. Kan and W.L. McKeehan.** 1998. Fibroblast growth factor receptor 2 limits and receptor 1 accelerates tumorigenicity of prostate epithelial cells. *Cancer Res.* 57:5369-5378.
2. **Klint, P. and L. Claesson-Welsh.** 1999. Signal transduction by fibroblast growth factor receptors. *Front. Biosci.* 4:d165-177.
3. **McKeehan, W.L., F. Wang and M. Kan.** 1998. The heparan sulfate-fibroblast growth factor family: diversity of structure and function, p. 135-176. In K. Moldave (Ed.), *Prog. Nucleic Acid Res. & Mol. Biol.* Academic Press 59(5), San Diego, CA.
4. **Naski, C.M., Q. Wang, J. Xu and D.M. Ornitz.** 1996. Graded activation of fibroblast growth factor receptor 3 by mutations causing achondroplasia and thanatophoric dysplasia. *Nat. Genet.* 13:233-237.
5. **Wang, F., M. Kan, G. Yan, J. Xu and W.L. McKeehan.** 1995. Alternately-spliced NH2-terminal immunoglobulin-like loop I in the ectodomain of the fibroblast growth factor (FGF) receptor 1 lowers affinity for both heparin and FGF-1. *J. Biol. Chem.* 270:10231-10235.
6. **Wang, F., W. Lu, K. McKeehan, K. Mohamedali, J.L. Gabriel, M. Kan and W.L. McKeehan.** 1999. Common and specific determinants for fibroblast growth factors (FGF) in the kinase ectodomain of the receptor complex. *Biochemistry* 38:160-171.
7. **Xu, J., M. Nakahara, J.W. Crabb, E. Shi, Y. Matuo, M. Fraser, M. Kan, J. Hou and W.L. McKeehan.** 1992. Expression and immunochemical analysis of rat and human fibroblast growth factor receptor (flg) isoforms. *J. Biol. Chem.* 267:17792-17803.

I thank Dr. W.L. McKeehan for support and encouragement. This work was supported by National Institutes of Health grant nos. DK35310 and CA59971 to W.L.M. Address correspondence to Dr. Fen Wang, Center for Cancer Biology and Nutrition, Institute of Biosciences and Technology, Texas A&M University System Health Science Center, 2121 W. Holcombe Blvd., Houston, TX 77030-3303, USA. Internet: fwang@ibt.tamu.edu

Received 17 September 1999; accepted 23 December 1999.

Fen Wang
*Texas A&M University System
Health Science Center
Houston, TX, USA*

Characterization of Flanking Sequences Using Long Inverse PCR

BioTechniques 28:838-844 (May 2000)

Characterizing genomic regions that flank specified sequences is important for promoter analysis and the identification of gene sequences disrupted by insertional mutagenesis and gene-marking strategies. However, traditional cloning and plasmid rescue methods that are used to identify such flanking sequences are cumbersome. Alternative PCR-based techniques such as inverse (3,7) partial inverse (4), ligation-linker

(5) and panhandle PCR (1) are limited by both the number of steps required before DNA sequencing and the frequent need for prior restriction mapping analysis. While long-distance, inverse and panhandle PCR have recently been used to characterize fragments up to 5.4 kb (8) and 9.4 kb (2), respectively, these techniques are still limited by additional cloning or amplification steps.

We have adapted the long-distance inverse PCR method to enable the direct amplification and sequencing of relatively large-sized flanking fragments. The method does not require knowledge of restriction sites surrounding the insertion sequence nor does it involve a bacterial cloning step. In addition, it facilitates the rapid analysis of large numbers of flanking sequences up to 10 kb adjacent to an insertion tag.

The strategy designated long inverse PCR (LI-PCR) is shown in Figure 1A. Digestion with two restriction endonucleases that produce compatible cohesive ends and the use of a thermostable polymerase is central to the LI-PCR approach. The use of two 6-base recognition enzymes, with one flanking the insertion tag, ensures that the plasmids generated following ligation will be reasonably sized (about 10 kb) for LI-PCR amplification. Depending on which sites are available in the vector being used, a number of other restriction endonuclease combinations that produce compatible ends can be used. This approach also avoids the fastidious procedure of blunt-ending the PCR products before ligation.

LI-PCR was initially validated in a microbial system and used to identify the genomic location of a single copy *lacZ* transgene in *Schizosaccharomyces pombe* transformants. The vector copy number was confirmed for each strain by Southern analysis (data not shown). For these experiments, the restriction enzymes *Bam*HI (used to subclone the *lacZ* gene) and *Bgl*II (absent from the integrating vector) were used (Figure 1A). Approximately 200 ng *S. pombe* genomic DNA were digested with 20 U *Bam*HI and *Bgl*II for 3 h, heat inactivated and then diluted with 1× T4 DNA ligase buffer [50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 10 mM dithiothreitol, 1 mM ATP, 25 µg/mL bovine serum albumin] (New England Biolabs,

Benchmarks

Beverly, MA, USA) to a final DNA concentration of 2.5 ng/μL. This low concentration of DNA is important for achieving intramolecular ligation. T4 DNA ligase was added to a final concentration of 4 U/μL before ligation at 15°C for 16 h. The ligation reaction was extracted with phenol/chloroform/isoamylalcohol and precipitated with 0.1 volume 3 M sodium acetate, 2 volumes 100% ethanol and 20 μg glycogen carrier (Roche Molecular Biochemicals, Mannheim, Germany) at -70°C for 1 h. The sample was then centrifuged at 12000× *g* for 10 min, washed with 70% ethanol, dried and resuspended at a concentration of 20 ng/μL. A total of 40 ng ligated DNA was used as a template for amplification using primers that specifically amplify the upstream genomic flanking sequence (Figure 1A). *PfuTurbo*TM DNA polymerase (Stratagene, La Jolla, CA, USA) was used for its ability to amplify genomic targets up to 10 kb in length and for its inherent thermostable properties, which enhance PCR prod-

uct yield. AmpliWax[®] PCR Gem 100 wax beads (PE Biosystems, Foster City, CA, USA) were used to increase specificity and product yield.

If intermolecular ligation had occurred, multiple PCR products would have been generated. However, single products were generated for all transformants containing the *lacZ* transgene, indicating that LI-PCR was specific and

that intramolecular ligation had been facilitated (Figure 1B). These PCR products contain both an unknown length of flanking sequence and a portion of vector that is dependent on the site of vector recombination. Here, the size of the PCR products does not reflect the distance of the flanking restriction site to the site of the recombination junction.

PCR products were purified and

Figure 1. LI-PCR in *S. pombe*. (A) The transgene (hatched) was cloned between the promoter and terminator sequences (filled) as a *Bam*HI fragment. After integration, digestion of genomic DNA with *Bam*HI and *Bgl*II cuts both the upstream and downstream flanking sequences (bold lines) at unknown distances from the transgene. Self-ligation of genomic DNA forms multiple circular molecules, two of which contain the upstream or downstream flanking sequences of the integrated vector. Thin lines indicate vector backbone, and bold lines indicate genomic DNA. Arrows A and B represent primers specific for vector sequence. (B) *S. pombe* strains were subjected to LI-PCR. PCR buffer consisted of 20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM (NH₄)₂SO₄, 0.1% Triton[®] X-100, 100 μg/mL nuclease-free BSA, 250 μM of each dNTP and a total of 3.5 mM free Mg²⁺. Reactions were performed at 94°C for 3 min and cycled 30 times at 94°C for 30 s, 55°C for 1 min and 72°C for 20 min. Product sizes ranged from 6–10 kb. The integrating vector (pG14-6) was amplified as a 12-kb positive control. The low molecular weight marker is *Eco*RI restricted SPP-1 bacteriophage DNA with bands ranging from 0.36–8.51 kb. The high molecular weight marker is *Hind*III restricted DNA with bands ranging from 0.56–23.1 kb.

Benchmarks

Table 1. Mapping of Integrated Vectors in *S. pombe*

Strain	Junction Sequence ^a	Cosmid	P(N) ^b	Site ^c
J45-3	TTACCCGGGGATC <u>CT</u> TATCAGATAATTGAAATCTA	unknown	-	<i>Bgl</i> II
I10-9	TTACCCGGGGATC <u>CT</u> GCAAGGTCGAGATCGCTTTG	SPCC1620	5.50E-42	<i>Bgl</i> II
J16-22	TTACCCGGGGATC <u>CT</u> TATTGGGTGACACCGCAACT	SPAC23D3	1.30E-82	<i>Bam</i> HI
J16-32	TTACCCGGGGATC <u>CT</u> GCTGATGCGCAGCCCAAAC	SPBC646	4.60E-82	<i>Bgl</i> II
J16-35	TTACCCGGGGATC <u>CT</u> GTAGACCACTTTATGTTTAT	SPBC3B8	5.40E-96	<i>Bgl</i> II
J16-42	TTACCCGGGGATC <u>CT</u> CAACCCGGGAGCCGCTGCATG	SPCC1840	1.30E-98	<i>Bam</i> HI
J16-45	TTACCCGGGGATC <u>CT</u> GCTTCGTGCTACTTTCATT	SPCC1235	3.20E-93	<i>Bam</i> HI
J45-10	TTACCCGGGGATC <u>CT</u> GTTAGGGTGGGACCAACTTA	SPAC17G6	1.50E-54	<i>Bgl</i> II
J45-14	TTACCCGGGGATC <u>CT</u> GCTCGGGTGCGATGGACTTG	SPAC23A1	9.10E-33	<i>Bgl</i> II
K40-7	TTACCCGGGGATC <u>CT</u> GATTGGGTGCCACAAATTAA	SPCC1223	2.90E-76	<i>Bgl</i> II

^aVector sequence is indicated in bold.
^bThe probability [P(N)] of genomic sequence having homology to sequenced cosmids is also indicated. Scores of less than 10⁻¹⁰ were considered highly significant.
^cReligation to either a *Bam*HI or *Bgl*II site was determined from the junction sequence (underlined).

concentrated with QIAquick™ PCR purification columns (Qiagen, Valencia, CA, USA) and then sequenced directly using an ABI Prism™ automated sequencer (PE Biosystems, Foster City, CA, USA). The specific amplification of a single product allows the circumvention of gel purification and cloning procedures inherent in standard approaches. However, a cloning step can be introduced by using a TA cloning protocol if required. A blastn search was performed with sequenced products against the available EMBL *S. pombe* genomic database (6) (Table 1). Nine gave matches against cosmids from the database, enabling the precise mapping of the integrated vector. No

Figure 2. Long inverse PCR in human cells. (A) The structure of the integrated *Afl*III linearized vector is shown. The multiple cloning site (MCS) contains *Bgl*II, *Xho*I, *Sal*I and *Bam*HI restriction sites. The size of the *gfp* cassette is 1.1 kb. Following double digestion with compatible enzymes and self-ligation, outward-facing primers specific for the 5' and 3' regions of the *gfp* gene were used to amplify the downstream genomic DNA (half arrows). (B) Southern analysis identified four clones with a single-copy integrated vector. Genomic DNA was digested with *Bam*HI and the Southern blot was probed with a *gfp* fragment. The sizes of the *Bam*HI fragments are indicated. (C) Clonal DNA (500 ng) was digested with *Bam*HI, *Bgl*II or both. Clone 1 was also digested with a second pair of compatible enzymes, *Sal*I and *Xho*I. Then, 100 ng of ligated DNA were subjected to LI-PCR as described in the *S. pombe* system. Product sizes ranged from 0.7–7.0 kb.

Benchmarks

homology was found for one sequence, probably because of incomplete sequencing of the *S. pombe* genome map.

Adaptation of this method to higher organisms is essential for the general utility of LI-PCR. We have shown this is feasible with complex (10^9 bp) human genomic material by amplifying flanking sequence from a single copy integrated *gfp* transgene from human 293 cell-line clones. In this experiment, the vector was linearized before transfection with *Afl*III and, as a result, clones that contained functional *gfp* cassettes had integrated in a predetermined fashion (Figure 2A). Clones were then screened by Southern analysis for the identification of single-copy integrants and four clones were chosen with *Bam*HI flanking regions of 1.8, 2.6, 4.0 and 9.0 kb (Figure 2B).

A total of 500 ng genomic DNA from the clones were digested with various restriction enzyme combinations, ligated and subjected to LI-PCR as described for the *S. pombe* system. The human genomic templates generated

specific PCR products up to 7 kb (Figure 2C). The fourth clone, in which the downstream *Bam*HI site was 9 kb away from the vector's internal *Bam*HI site, failed to produce any amplicons (data not shown). This suggests that the upper size limit of amplification in this system is around 7 kb.

In contrast to the *S. pombe* experiments, the internal restriction sites were at a known distance from the integration junction (1.1 kb). Therefore, the LI-PCR product sizes were indicative of the distance between the site of integration and the flanking restriction site. This is demonstrated by comparing the size of Southern blot bands and *Bam*HI LI-PCR products (Figure 2, B and C). The Southern blot fragments are approximately 1.1 kb larger than the LI-PCR products because of the inclusion of the *gfp* fragment. Additionally, the value of using enzymes that produce compatible ends was confirmed by using two enzyme combinations.

In some instances, PCR products could not be generated with one of the enzymes being used, presumably because the size of the ligated fragment was larger than 7 kb. However, depending on the site of vector integration, one of the compatible enzymes did produce a PCR product. For example, for clone 1, digestion with *Bam*HI and *Sal*I generated a PCR product while no product was produced for their respective compatible restriction enzymes, *Bgl*II and *Xho*I. Therefore, digestion with both enzymes greatly enhances the probability of generating a PCR product and should be used in any approach of this type.

A simple, rapid method for the characterization of long regions of DNA flanking known sequences has been described for both simple and complex genomes. By making use of compatible 6-base restriction endonucleases before ligation and a thermostable polymerase for long PCR, specific PCR products were generated, which allowed for their direct sequencing. The benefit of this method for mapping sites of integration was demonstrated in a fission yeast model, while the robustness of this strategy was shown by amplifying flanking regions up to 7 kb in more complex human genomic DNA. The simplicity of this technique should complement strategies that are currently used for

characterization of flanking sequences. In addition, the application of the technique for mapping insertion tags should become more general as the complete sequence of more organisms, including humans, becomes available.

REFERENCES

1. Jones, D.H. and S.C. Winistorfer. 1992. Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. *BioTechniques* 15:894-904.
2. Jones, D.H. and S.C. Winistorfer. 1997. Amplification of 4-9 kb human genomic DNA flanking a known site using a panhandle PCR variant. *BioTechniques* 23:132-138.
3. Ochman, H., A.S. Gerber and D.L. Hartl. 1988. Genetic applications of inverse polymerase chain reaction. *Genetics* 120:621-623.
4. Pang, K.M. and D.A. Kneet. 1997. Partial inverse PCR: a technique for cloning flanking sequences. *BioTechniques* 22:1046-1048.
5. Shyamala, V. and G.F.-L. Ames. 1989. Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. *Gene* 84:1-8.
6. Stoesser, G., M.A. Moseley, J. Sleep, M. McGowran, M. Garcia-Pastor and P. Sterk. 1998. The EMBL nucleotide sequence database. *Nucleic Acids Res.* 26:8-15.
7. Triglia, T., M.G. Peterson and D.J. Kemp. 1988. A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res.* 16:8186.
8. Willis, T.G., D.M. Jadavay, L.J.A. Coignet, M. Abdul-Rauf, J.G. Treleaven, D. Catovsky and M.J.S. Dyer. 1997. Rapid molecular cloning of rearrangements of the IGHJ locus using long-distance inverse polymerase chain reaction. *Blood* 90:2456-2462.

We thank Dr. Wayne Gerlach for helpful comments on the manuscript and Margaret Patrikakis for technical assistance. Address correspondence to Dr. Greg M. Arndt, Johnson and Johnson Research Pty Ltd, Locked Bag 4555, Strawberry Hills, NSW 2012, Australia. Internet: garndt@medau.jnj.com

Received 16 July 1999; accepted 5 January 2000.

Mitch Raponi^{1,2}, Ian W. Dawes¹ and Greg M. Arndt²

¹University of New South Wales Sydney, NSW, Australia

²Johnson and Johnson Research Pty Ltd
Eveleigh, NSW, Australia