Easy Gene Walking


Rapid sequencing of genomic DNA by “gene walking” has become an important procedure in many research projects (1-4,9-11,13). The strategies used frequently require restriction enzyme digestion or ligation and enzymatic amplification of the DNA (1,2,9). However, DNA amplification can be done successfully using random oligomers as one or both primers (3,5,11,12). In one report, unknown DNA could be amplified using a primer from a region of known DNA paired with a primer comprised of a short degenerate sequence followed by a specific 4-6-base sequence on the 3′ end of the primer (11). It was reasoned that the short, specific sequence would find several matches in the genomic DNA and that sufficient additional matches could be obtained from the degenerate region to allow adequate priming. We have used a combination of random and specific primers and cycle sequencing to obtain about 300 bp of unknown, uncloned sequence per “walk” within 48 h.

Figure 1 illustrates the amplification strategy and Table 1 contains the actual protocol. Primers are chosen from within the known region of DNA (nested primers 1-3, Figure 1) adjacent to the target region. A random primer was used that contained 10 bases of degenerate sequence followed by 4-6 bases of a specific sequence. In other words, each base position preceding the 4-6-base specific sequence could contain A, T, G or C. The 4-6 specific bases at the 3′ end ensure that the end of the primer most critical to its functionality as primer will find an exact match. Additional matches will be selected from the degenerate sequence that follows in the 5′ direction. Thus, in the illustration, the random primer found a complementary site and, together with primer 1, produced a product (Round 1; Figure 1). Re-amplification of the product using the same random primer and a second (and then third) nested primer from the known sequences would lead to a progressive increase in product quantity (Rounds 2 and 3; Figure 1). The sequencing reaction would then be performed with the nested primer (number

Table 1. Gene Walking Protocol

1. Using 200 ng genomic DNA in a 20-µL reaction, perform 25 cycles using nested primer 1 and one or more random primers.
2. Using nested primers 2 and 3 in succession, repeat the amplification. For each random primer, use 1 µL of its previous reaction.
3. Run electrophoresis on the products from Reaction 3 on a 1.5% agarose gel and stain with ethidium bromide. Cut out bands in the range of 100-800 bp and isolate the DNA using a MicroPur™ Separator and a Gel Nebulizer™ (both from Amicon, Beverly, MA, USA) according to the manufacturer’s directions. Resuspend the product in 15 µL water.

The genomic DNA can be isolated by procedures found in standard manuals or by using 1 µL of a lysate made by suspending 10⁶ cultured cells in 1 mL of water and heating at 94°C for 15 min. The cell lysate can be frozen and thawed at least five times.

All amplifications were performed on a Hybaid Model TR1 Thermal Cycler (National Labnet, Edison, NJ, USA) using PE Applied Biosystems reagents at the recommended buffer concentrations. Nested and random primers were used at concentrations of 4 pM/10 µL and 10 pM/10 µL, respectively. The amplification cycle parameters were 30 cycles: 94°C for 1 min, 50°C for 2 min and 72°C for 3 min (10 min for last cycle).

The random primers we used contained the recognition sites for the restriction enzymes Sau3AI, EcoRI, BamHI and TaqI at the 3′ end preceded by ten degenerate bases, so that 5′-NNNNNNNNNNNNNN-3′ would be the primer containing the Sau3AI site (GATC), where N = A, T, G or C.

The cycle sequencing parameters were 25 cycles: 96°C for 30 s, 50°C for 15 s and 60°C for 4 min.

Figure 1. Schematic of random primer PCR. In this diagram, an unknown DNA sequence lies adjacent and 5′ to a known sequence. The illustration assumes a single site within the unknown sequence, to which the random primer hybridizes. Other random primers would have the same general construction, i.e., sequence corresponding to a restriction site at the 3′ end, preceded by ten degenerate bases. The figure illustrates the results of three amplification rounds performed using three nested primers (Step 1). In practice, 1/10 of Reaction 1 is used in Reaction 2, etc. The final product (from Round 3) is then purified and cycle-sequenced (Step 2) using nested primer 3. Spacing of primer 3 from the 5′ end of the known DNA allows the sequence to be read through from known DNA to unknown DNA, confirming that the sequence obtained is from DNA abutting the known region.
Figure 2. Results of sequential random primer PCR of genomic DNA and a first and second walk sequence. Top panel: ethidium bromide-stained gels of successive amplifications using nested primers from within the transgene and random primers containing the restriction sites for EcoRI (E), BamHI (B), Sau3AI (S) and TaqI (T). The source of DNA was a lysate of cultured AtT-20 cells from two clonal lines that had been stably transfected. Middle panel: printout of sequence from first walk. Boxed sequences (yellow) were used as the nested primers for Walk Two. Bottom panel: printout of sequence from second walk.
3) used in the third amplification round.

We have used this technique to sequence genomic DNA adjacent to an enhancer trap—a selectable transgene used to identify genomic regulatory sequences and their associated genes (6–8). The composite Figure 2 contains, in the upper section, ethidium bromide-stained gels of the products obtained in three amplification rounds performed using DNA from two cell lines and four different random primers. The ethidium bromide-stained gels show, for two separate DNA samples, the progressive development of bands, as sequential amplification is performed using three nested primers. Interestingly, the bands observed in Round 3 may not have been observed in preceding rounds, and bands present in preceding rounds may have lost prominence in subsequent rounds. The products in each round vary in abundance, and an abundant product in an early round may be less prominent in a later one. We believe that this results from subtle changes in the amplification conditions that favor one random site over another, even in sequential reactions.

As an illustration of sequence quality, the middle and lower sections of Figure 2 show the actual sequence obtained from two sequential walks. Walk One was the result obtained from sequencing a single isolated band (see Table 1). The final, nested primer used in this experiment was within 15 bp of the known/unknown DNA junction, and the known DNA sequence was not clearly seen. However, reactions using this primer or one farther 3′ usually showed the known DNA sequence abutting the unknown sequence. Three new nested primers were selected from areas with approximately 50% GC content (see yellow regions in Walk One) and a 3′-terminal G or C. The amplification was then repeated. Isolation of a band from the second walk and cycle sequencing yielded the results shown as Walk Two. Some sequencing ambiguity was seen within 10–20 bp of the sequencing primer, but thereafter, less than one per hundred bases was ambiguous. Nevertheless, the accuracy of this type of sequencing method would be further bolstered by sequencing of the opposite strand.

The procedure we used (Table 1) consistently (80%) yielded results using a variety of input DNA, including cultured cells lysed in water. Amplification of 32 consecutive nonidentical DNA samples resulted in 26 bands that yielded an average of 293 bases/walk. Turnaround was rapid since the amplification was performed in a day, followed by band isolation, cycle sequencing and gel analysis the following day. Band isolation resulted in a more reliable sequence of longer products.
However, excellent sequence can be obtained without band isolation, since multiple bands will have the same origins at the nested primer end of the reaction, and the final nested primer is also used in the cycle sequencing reaction. Therefore, when sequencing reactions were run with products containing several bands, the distant sequences were unaffected and were very readable. Forty-one samples sequenced without band isolation yielded an average of 212 bases/walk. Since the sequence farthest away from the primer is unaffected and can be used to select primers for an additional walk, band isolation is optional.

This procedure compares favorably with others. Advantages include its simplicity (since it involves no DNA digestion or ligation), consistent results and speed. Cost is low since the expense of three primers per walk is balanced by a significant savings of technical time and in the cost of other reagents such as restriction enzymes and ligases. Using this procedure and a modification of an “enhancer trap” (8), we have rapidly identified several glucocorticoid-regulated genes in AtT-20 mouse pituitary tumor cells.

REFERENCES


Supported by NCI Grant No. CA 42091. The authors thank Ms. Andrea Weinstein for excellent secretarial assistance. Address correspondence to Robert W. Harrison, University of Rochester, 601 Elmwood Avenue, Box 693, Rochester, NY 14642, USA. Internet: rharri@medicine.rochester.edu

Received 20 May 1996; accepted 4 October 1996.

Robert W. Harrison, Janice C. Miller, Mary J. D’Souza and George Kampo
University of Rochester School of Medicine and Dentistry
Rochester, NY, USA

Automated Fluorescent DNA Sequencing by a Simplified Solid-Phase Chemical Sequencing Method


DNA sequencing is now done principally by two methods, the chemical degradation method of Maxam and Gilbert (2) and the enzymatic chain termination method of Sanger et al. (9). Although the enzymatic method is overwhelmingly preferred in large-scale DNA sequencing projects, the