Research Report

General Method for HPLC Purification and Sequencing of Selected dsDNA Gene Fragments from Complex PCRs Generated during Gene Expression Profiling

BioTechniques 28:776-783 (April 2000)

L.Y. Wong, V. Belonogoff, V.L. Boyd, N.M. Hunkapiller, P.M. Casey, S.-N. Liew, K.D. Lazaruk and S. Baumhueter
Celera Genomics, Foster City, CA, USA

ABSTRACT

Gene expression profiling using an AFLP-based technique generates a large number of gene fragments that require identification by sequencing. The DNA fragments vary in length from about 50–500 bp. Ion-pair reversed-phase HPLC can be used to purify selected double-stranded DNA fragments that represent differentially expressed genes. The gene fragments are sequenced directly after vacuum drying of the collected HPLC fractions.

INTRODUCTION

The purification of double-stranded DNA (dsDNA) products from PCR using HPLC as an alternative to gel electrophoresis has been demonstrated with both ion exchange and ion-pair reversed-phase methods (2,5,6,7,10,11). The HPLC columns, depending on the size and packing material, can resolve a wide range of dsDNA sizes, with a capacity for 1–10 µg of total DNA. The columns can last for several hundred injections, and samples can be injected sequentially after a brief equilibration time. A major advantage of HPLC purification is the ability to automate the purification process, which is essential for laboratories requiring high-throughput sample processing.

The GeneTag™ process is a gene expression profiling technique based on AFLP (1,9) that generates no more than one dsDNA fragment from each actively expressed gene. Since an estimated 10 000 or more genes may be expressed in a cell at any given time, the complete GeneTag process generates many unique fragments. For expression profiling, the relative abundance of each gene fragment is determined. To identify a differentially expressed gene, the dsDNA fragments must be isolated and sequenced.

HPLC was investigated as a method for isolating single gene fragments, before direct sequencing, from the complex mixture of PCR-generated AFLP fragments. Figure 1 shows that the number of gene fragments in a PCR is systematically reduced by the use of selective primers. Reducing the complexity of the PCR simplifies the analysis and HPLC purification steps. By adding a fluorescent label, 6-carboxyfluorescein (FAM), to one of the primers, the PCR product can be visualized during denaturing capillary electrophoresis, which separates the gene fragments by length. Computer programs are used to compare the electropherogram profiles from different tissue sources and to quantitate peak-height differences. When a differentially expressed gene fragment is found, the fragment is isolated and sequenced to determine or confirm the identity of the gene. Here, we report the application of the new Zorbax Eclipse™ dsDNA analysis column (Hewlett Packard, Palo Alto, CA, USA) (3,4) to separate fragments generated by an AFLP-based process and directly determine their DNA sequence.

MATERIALS AND METHODS

Sample Preparation

Rat brain and liver total RNA samples were acquired from Clontech Laboratories (Palo Alto, CA, USA). PolyA+ mRNA was isolated using Magnesphere® paramagnetic particles (Promega, Madison, WI, USA) and biotinylated oligo-dT following the manufacturer's protocol with slight modifications. cDNA was prepared using the SuperScript™ Choice system cDNA synthesis kit from Life Technologies

(Rockville, MD, USA). The GeneTag process was performed similarly to AFLP as previously described (9) with three exceptions: first, cDNA was used instead of genomic DNA; second, FAM-labeled primers were used in place of radioactively labeled primers; and third, the two restriction cuts were made using *Bst*YI and *Mse*I (New England Biolabs, Beverly, MA, USA). AmpliTaq Gold®, dNTPs, AmpliTaq Gold PCR buffer (pH 8.0) and $MgCl_2$ were purchased from PE Biosystems (Foster City, CA, USA). Thermocycling was performed on a Model 9600 GeneAmp® PCR system (PE Biosystems). Dilutions were made with DNase and RNase free water (Sigma, St. Louis, MO, USA). PCR products were analyzed on a Model 310 ABI Prism® genetic analyzer (PE Biosystems) using run module GS STRA with the run time extended to 30 min. One microliter of each PCR was mixed with 12 µL of formamide containing ROX-tagged size standard from PE Biosystems. Samples were electrokinetically injected into 47 cm × 50 µm capillaries filled with POP-4™ polymer (both from PE Biosystems). A 1 µL aliquot of the PCR, which contained a differentially expressed gene fragment selected to be sequenced, was amplified in a 100 µL PCR to provide enough DNA for HPLC purification and sequencing. The 100 µL PCR in preparation for HPLC was 2 mM $MgCl_2$, 3.75 U AmpliTaq Gold, 200 µM of each dNTP, 250 nM of the forward and reverse primer (unlabeled), prepared in 1× AmpliTaq Gold PCR buffer. Following an 11 min hold at 95°C to activate the enzyme, 15 cycles of thermocyling were performed at 95°C for 15 s, 60°C for 2 min and 75°C for 2 min.

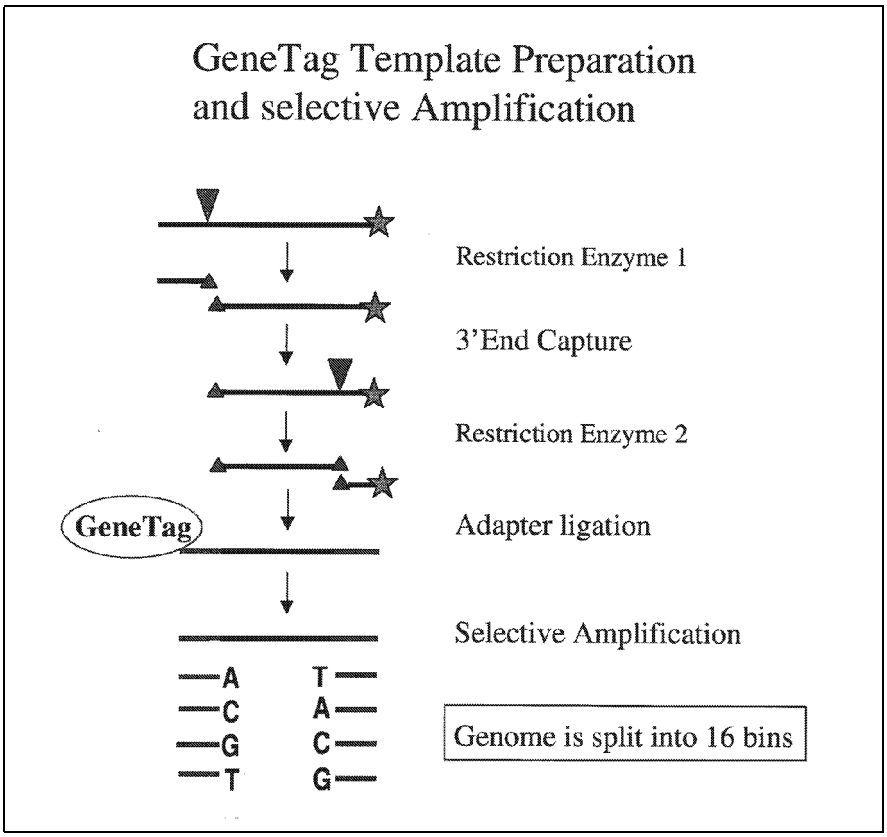

Figure 1. The steps of the GeneTag process for determining differences in the expression level for both known and unknown genes. mRNA, isolated from the cells or tissues of interest, is reverse transcribed into cDNA using biotinylated oligo-dT as a primer. Therefore, the 3′ end of the cDNA, represented by the star, has a biotin group used for subsequent 3′ end capture. The cDNA is sequentially cut with two restriction enzymes. First, the cDNA is restricted with *Bst*YI, adapters are ligated and the reaction mixture is incubated with streptavidin-coated beads to capture the 3′ end fragments. The bound cDNA is subsequently restricted with *Mse*I, which is predicted to cut closer to the 3′ end than *Bst*YI based on the statistical probability of finding the sequence recognized by the enzyme. The second restriction cut releases the cDNA from the support, and an *Mse*I adapter is ligated to the end. The adapters (5′ *Bst*YI and 3′ *Mse*I), and the residual bases of the restriction site then serve as primer-binding sites for simultaneous (nonselective) amplification of all cDNA fragments. The number of fragments in the nonselective PCR is too great to permit the identification of differentially expressed gene fragments. The complexity of the adapter-tagged cDNA pool is reduced by adding one to three additional bases to the 3′ ends of each of the primers. There are 16 primer pair combinations created by the addition of one extra base to each nonselective primer and 256 possible primer pair combinations with two additional bases added to the ends of the nonselective primers. Only those fragments with complementary bases matching the extended primers will amplify. The addition of two or three selective bases added to the 3′ end of the primers therefore distributes the fragments into primer pair "bins" reducing the average number of fragments per reaction. Typically, about 150–200 fragments are present in the PCR where two selective bases were added; the number of peaks is reduced by a factor of approximately 16 when a third selective base is added to each primer. However, the frequency of the recognition site of a given restriction enzyme pair determines the number and length of the cDNA fragments generated. Differentially expressed genes are identified by comparison of the gene fragments present in the selective PCR generated from different cDNA sources.

HPLC

Two Series 200 micropumps (PE Biosystems) were connected with a low-volume tee mixer (2 µL dead volume; Alltech, San Jose, CA, USA). The mobile phase was passed through a static mixer and a 75 µL dynamic mixer (both from PE Biosystems) before reaching the injector. A 2.1 mm ID × 75 mm Zorbax Eclipse dsDNA analysis column (Hewlett Packard) was heated to 50°C with a wraparound column heater from Cera (Baldwin Park, CA, USA). A guard column holder with insert (Hewlett Packard) and a 0.5 µm inline filter, housed in a pre-column filter holder (both from Upchurch Scientific, Oak Harbor, WA, USA), were placed immediately before the analytical column. A Model 785 UV detector set at 260 nm was equipped with a 2.4 µL flow cell (both from PE Biosystems). Triethylammonium acetate (TEAA) as a 2 M solution and acetonitrile (CH_3CN; flammable and an irritant) were purchased from PE Biosystems. A solution of 0.5 M EDTA, pH 8.0, was purchased from Life Technologies. The gradient was controlled by programming the flow rate vs. time for each pump and holding the combined flow at 200 µL/min. The time field is per seg-

ment and not cumulative when programming the Series 200 pumps.

Data were collected using the Turbochrom Workstation (Version 6.1.0.2; PE Biosystems). The column was equilibrated at 40% B for 15 min between injections. Each day, 5 µL of a 0.25 µg/µL solution of a *Hae*III digest of pBR322 (Roche Molecular Biochemicals, Indianapolis, IN, USA) were injected just before the selected PCR for quantitation and retention time calibration. Evaporation of acetonitrile resulted in a shift in retention time to larger values, so that the standard did not elute within the 50 min gradient. To prevent daily fluctuations in the retention time, the solvent reservoirs need to be capped. A 100 µL PCR containing the differential fragment of interest was injected onto the HPLC. One-hundred-microliter fractions (every 30 s) were collected over a 50 min gradient as the DNA eluted from the column. Each fraction was concentrated to dryness in a SpeedVac® (Savant Instruments, Holbrook, NY, USA) and re-dissolved in 25 µL of TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0; Teknova, Half Moon Bay, CA, USA) for sequencing.

Table 1. Gradient and Solvents Used for the HPLC Isolation of dsDNA Fragments

Time	%B	Flow (mL/min) Pump A	Flow (mL/min) Pump B
0	40	0.120	0.08
10	53	0.094	0.106
40	80	0.040	0.160
2	40	0.120	0.08

Solvent A: 0.1 M TEAA, 0.1 mM EDTA
Solvent B: 0.1 M TEAA, 25% CH_3CN, 0.1 mM EDTA

Sequencing

Ten microliters of each reconstituted HPLC fraction were sequenced with the ABI Prism dRhodamine Terminator Cycle Sequencing Ready Reaction kit (PE Biosystems) following the manufacturer's instructions (8). Cycle sequencing was performed on a Model 9600 GeneAmp PCR system using the same nonselective forward and reverse primers used in AFLP profiling. No adjustment to the sequencing protocol was made to compensate for fluctuations in the template concentration. After precipitation with 3 M sodium acetate (PE Biosystems) and ethanol (Aaper Alcohol and Chemical, Shelbyville, KY, USA), the pelleted DNA was dissolved in template suppression reagent (PE Biosystems) and analyzed on a Model 310 ABI Prism Genetic Analyzer with POP-6 polymer (PE Biosystems). Sequence data were analyzed using ABI Prism Sequencing Analysis Software version 3.0 (PE Biosystems).

RESULTS AND DISCUSSION

Differentially expressed gene fragments were identified during expression

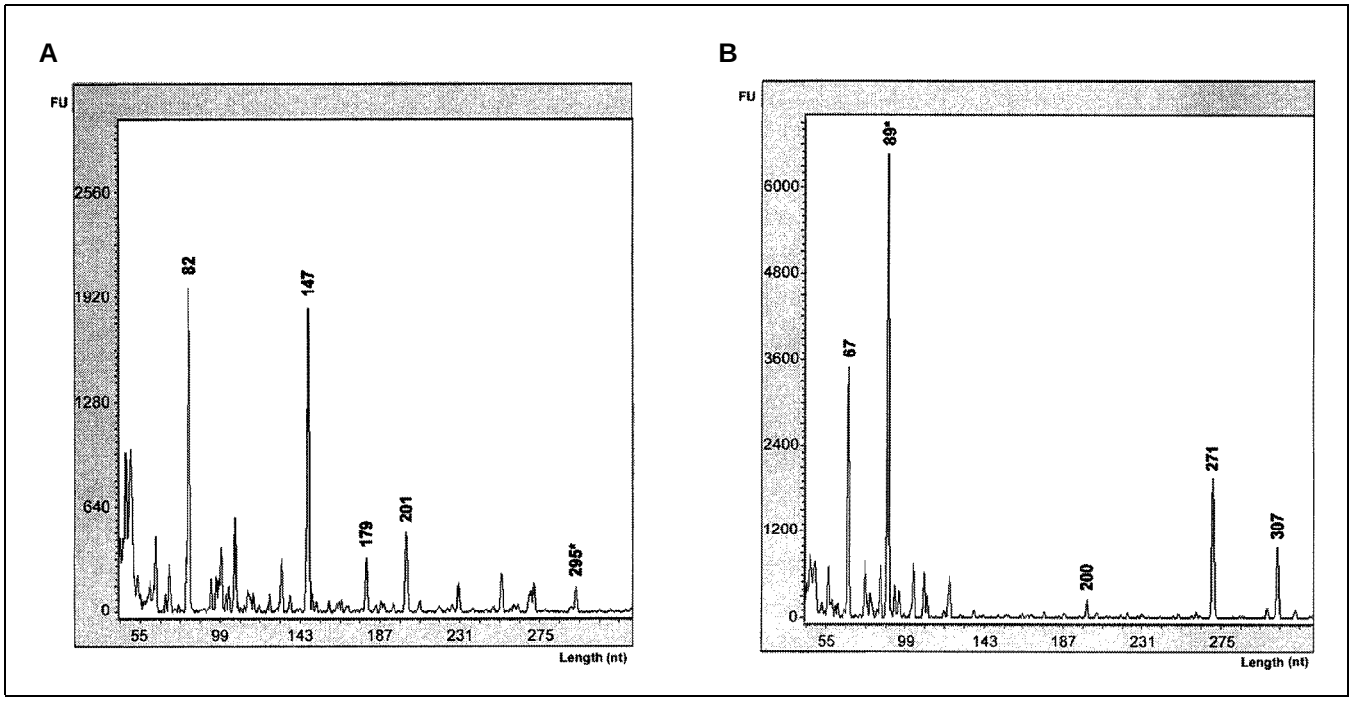

Figure 2. Electropherograms of differentially expressed gene fragments at 295 bp (A) and 89 bp (B) that were present in brain, but not in liver tissue. The differential peaks were identified by comparing the brain expression profiling electropherograms to the corresponding electropherograms obtained during profiling of liver with the same primer pair. One FAM label is present for each fragment, so the fluorescent signal should be proportional to the number of copies of each gene fragment present in the PCR.

profiling when comparing commercially available rat brain and liver tissue. Four examples of gene fragments that were expressed in brain but not in liver tissue were selected to be sequenced. The electrophoretically estimated lengths of these fragments were 89, 209, 257 and 295 nucleotides. Figure 2 presents the electropherograms containing two of the differential fragments. The number of other gene fragments present in each of the PCR containing the selected differentially expressed fragments was variable. In Figure 2A, the 295 bp fragment is part of a complex PCR, and the 89 bp fragment (Figure 2B) is part of a somewhat less complex PCR.

HPLC Isolation of Differentially Expressed Gene Fragments

HPLC was used here as a time-saving, cost-effective and automated alternative to PAGE for the purification of the selected PCR products identified in the electropherograms (Figure 2). Before HPLC purification of the fragments, a *Hae*III digest of pBR322 was

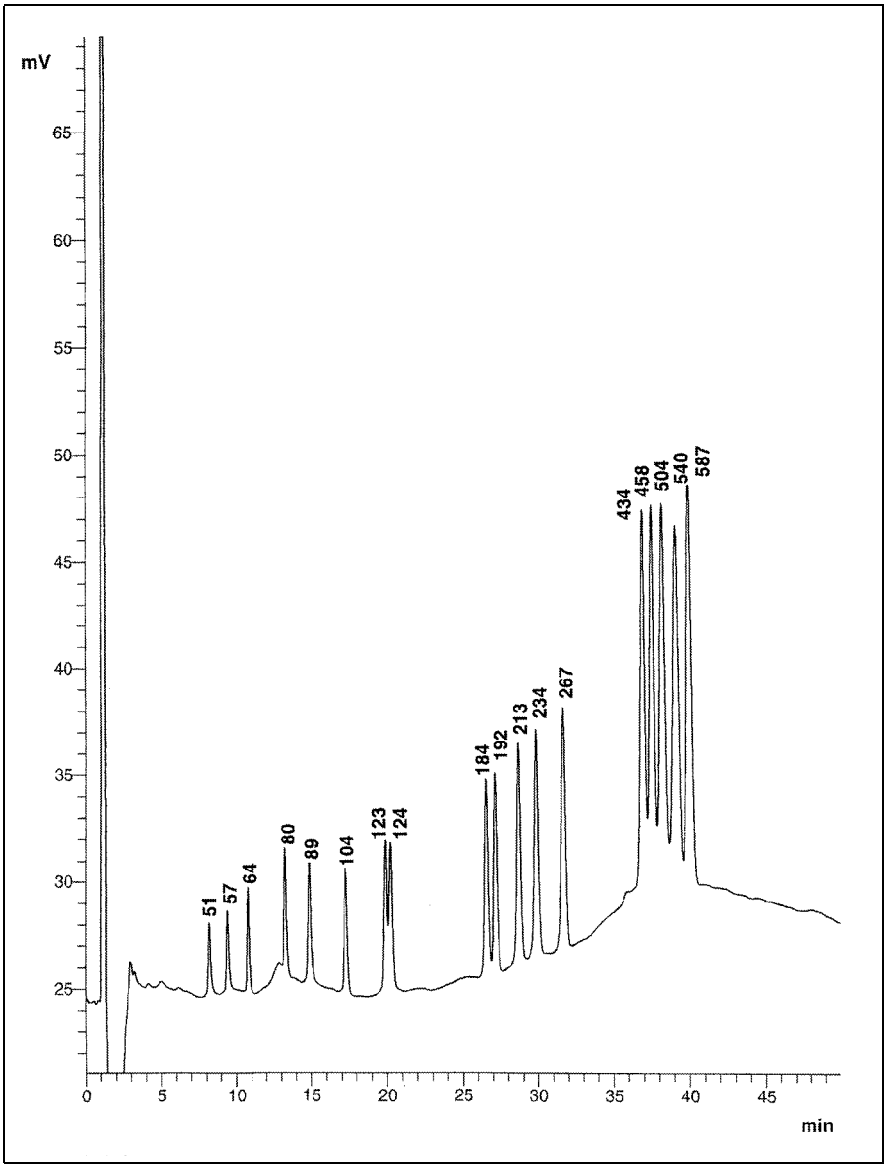

Figure 3. HPLC chromatogram of 1.25 µg of the *Hae*III digest of pBR322, which is used as a size and quantitation standard. Because the digested fragments are equimolar, peak area is proportional to fragment length. A linear regression analysis of peak area vs. nanograms of DNA present in each peak of the standard yields an equation that can be applied to quantitate unknowns.

injected and served as a size standard (51–587 bp; Figure 3). The length of the fragments for the standard is similar to the length of the fragments generated by the expression profiling experiment. Following the manufacturer's recommendation for the Zorbax Eclipse dsDNA analysis column, a 30 min gradient for the separation of the *Hae*III digest was readily reproduced in our laboratory. Because of the large number of fragments generated by the AFLP-based, expression-profiling PCR, many of similar lengths, a 50 min gradient was used to improve fragment resolution.

A 1 µL aliquot from each of the four PCR containing the differentially expressed gene fragments as determined from the AFLP-profiling experiment, served as template in large-scale, 100 µL PCR using unlabeled primers. The entire PCR was injected onto the HPLC, without prior clean-up steps, and timed fractions were collected. The retention time of any fragment of known length can be predicted from the retention time of the size standard. As the fractions are collected based on time, the fraction(s) containing the fragment of interest can be readily determined. The HPLC chromatograms of the PCR containing the 295 and 89 bp fragments are reproduced in Figure 4 and can be compared to the electropherograms of the PCR containing the same two gene fragments shown in Figure 2. The latter half of the chromatogram is compressed relative to the more evenly distributed nucleotide/time display of the electropherogram. However, a similarity of the peak profiles was seen when comparing the chromatograms to the electropherograms.

The HPLC fraction(s) containing the fragment(s) of interest were concentrated to dryness under vacuum, which also removed the TEAA and acetonitrile used during HPLC chromatography. The isolated PCR product was then sequenced. A complete sequence was obtained for each of the four differentially expressed gene fragments. For three of the four gene fragments, the full-length rat gene was identified through a sequence match found in GenBank®. The fourth gene fragment gave a novel sequence. Figure 5A (295 bp) and 5B (89 bp) show part of the sequencing traces obtained for the 295 and 89 bp fragments, indicating that the recovery of DNA was in sufficient quantity and quality to give a sequence with very low background. The sequence length matched the fragment length obtained in the original electropherogram profile.

Quantitation of dsDNA

The strength of the UV signal in the HPLC chromatogram can be used to determine the amount of dsDNA that was injected. By comparison to the standard,

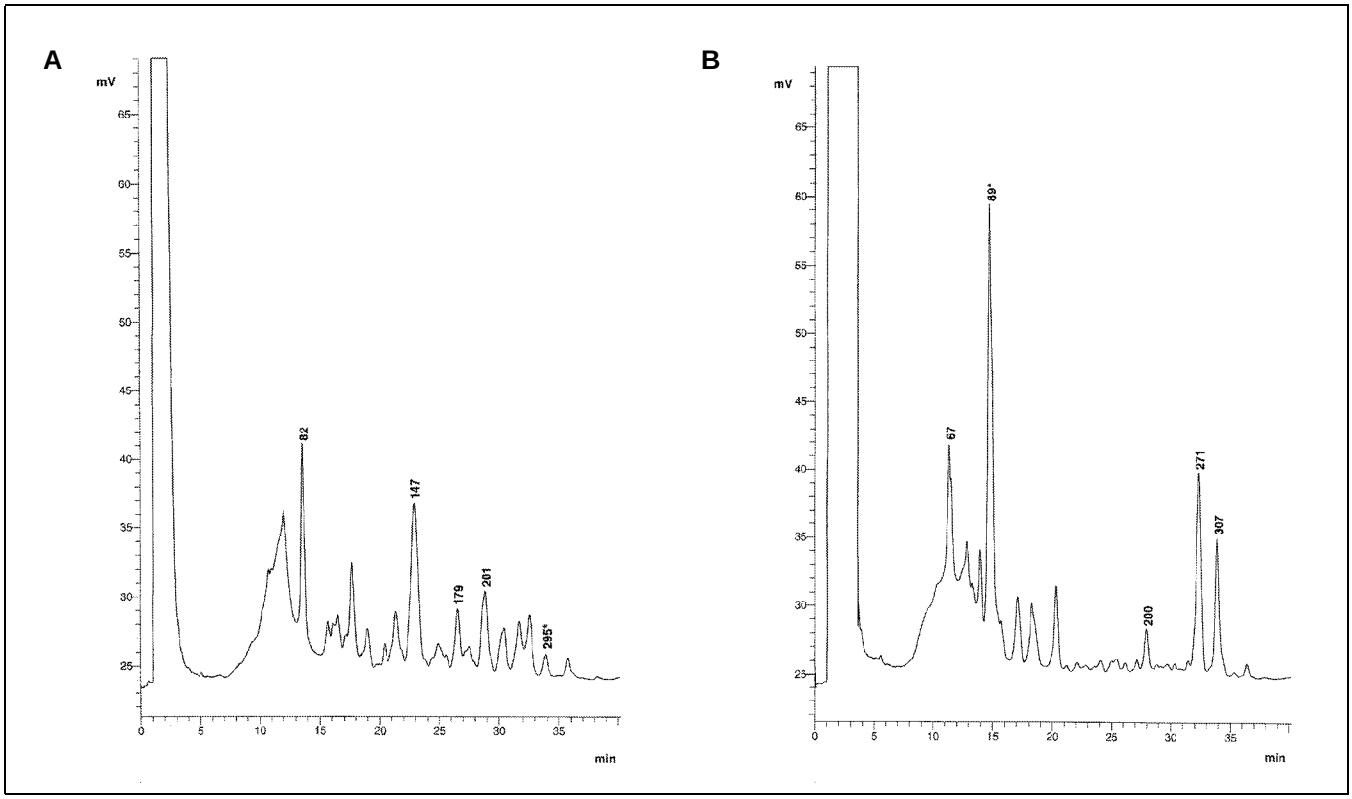

Figure 4. HPLC chromatogram of the scaled-up PCR containing the 295 bp (A) and the 89 bp (B) fragments. The source of the template for the scaled-up PCR was the same as the profiling PCR analyzed by capillary electrophoresis and shown in the electropherograms in Figure 2. The resulting peak profile of the HPLC chromatogram is comparable to the peak profile seen in the electropherogram containing the same fragment. A broader peak width for the PCR products, relative to those for the size standard in Figure 3, is due to the much greater amount of dsDNA injected. Independent experiments support that the broad peak seen at 10–12 min is single-stranded DNA. The electropherogram analysis of fractions collected in this region, amplified with one FAM-labeled primer, contains small quantities of fragments of many different lengths. Additionally, when linear amplification is used to produce single-stranded DNA, the product elutes in the 10–12 min region. The large peak that elutes within the first 5 min is from dNTPs and unlabeled PCR primer.

the yield for the 89 bp fragment (Figure 4B) was 253 ng, and the yield for the 295 bp fragment (Figure 4A) was 15 ng.

The capacity of the RP HPLC column (10 μg) (3,4) allowed the entire 100 μL PCR to be injected. For the four gene fragments described, a single injection provided more than sufficient sample for sequencing. The quantity of DNA collected was often less than the amount of DNA estimated from the strength of the UV signal. In independent experiments, we observed as much as a 50% loss of DNA whenever DNA, dissolved in 12.5% acetonitrile and 0.1 M TEAA, was evaporated to dryness and resuspended in TE buffer. This loss was observed, both with siliconized and standard tubes. Presumably, the DNA adhered to the walls of the microcentrifuge tubes. During automated fraction collection based on time, the fragment of interest could also be collected over two or even three fractions, further reducing the final concentration of purified dsDNA fragment in a given fraction submitted for sequencing. Sequencing was nevertheless successful because of the combined ultrahigh sensitivity of the current cycle-sequencing methodology and fluorescence-based sequencing instruments.

Sequencing of PCR Products Isolated by HPLC

The dye terminator sequencing protocol using dRhodamine dyes (PE Biosystems) recommends between 30 and 90 ng of PCR product for sequencing (8). However, direct sequencing of one of the HPLC-isolated gene fragments was successful with a lower amount of DNA (15 ng of the 295 bp differential fragment). Assuming that a 50% loss of DNA occurred during vacuum drying and resuspension into 25 μL of TE buffer and using a 10 μL aliquot for sequencing in each direction, the estimated amount of DNA sequenced was about 2.8 ng or 0.01 ng/bp. Figure 5 shows that the sequencing data were unambiguous and that the genes could be readily identified.

Additional Observations

The HPLC gradient used in this study was optimized for shorter fragments (no longer than 200 bp). The gradient can be adjusted easily so that resolution is optimized for a particular size range (3,4). The leading and tailing edges of a poorly resolved peak can be selectively collected based on the UV response on the detector. Enough purified DNA may be collected so that the fragment still can be sequenced. Alternatively, a partially purified fragment can be reamplified and repurified by HPLC using an optimized gradient for the fragment size of interest. Purifica-

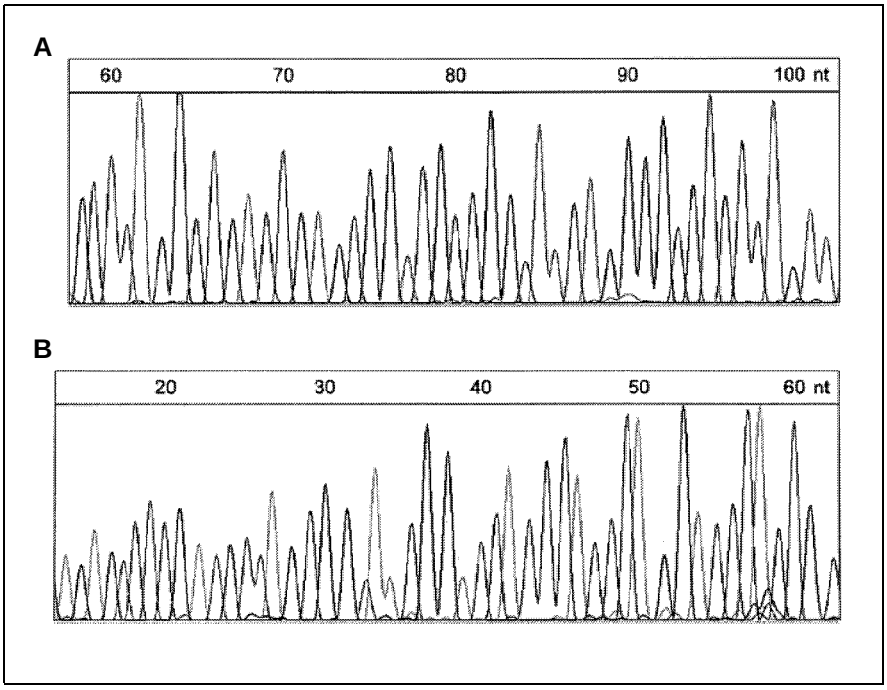

Figure 5. Partial electropherograms obtained from dRhodamine dye terminator sequencing of the HPLC purified fragments seen at 295 bp (A) and 89 bp (B). The raw data signal intensity for the sequencing of the 295 bp fragment was lower than that for the 89 bp fragment as expected, based on the UV signal seen in the corresponding HPLC chromatograms in Figure 4. The signal values provided by the ABI Prism sequencing analysis software were as follows: G = 479, A = 230, T = 234, C = 121 (295 bp fragment); G = 2719, A = 2498, T = 2188, C = 1030 (89 bp fragment). Peak color corresponds to the nucleotide base called in sequencing: A = green, C = blue, G = black, T = red.

tion after reamplification would be necessary because PCR reamplification generally causes neighboring fragments, not necessarily detected in the chromatogram, to be amplified.

HPLC was useful for the isolation of a single PCR fragment out of the many products generated by an AFLP-based technique for gene expression profiling. In addition to the four differentially expressed gene fragments presented here, a large number of other gene fragments has been purified to date by HPLC and identified by sequencing. The different sizing techniques, capillary electrophoresis and HPLC, produced similar peak profiles when analyzing a sample from the same source (one primer pair bin). The elution order and approximate signal intensity were preserved in spite of differences in the sampling (injection) technique, analyte conformation (native vs. denatured), mode of separation and detection method. Our data support the proposition that HPLC can be used in place of gel electrophoresis as a reliable and efficient purification step for resolving a multicomponent PCR because the process can be automated readily and the collected fractions can be sequenced directly after vacuum drying.

ACKNOWLEDGMENTS

The authors wish to express thanks to Will Bloch (PE Biosystems), who shared his knowledge and provided encouragement and advice for the HPLC application described in this manuscript. We also thank Dave Dupont, Steve O'Neil, John Wages (all from PE Biosystems) and Frank Vandemark (Perkin-Elmer) for their help in making this work possible. In addition, we thank the entire GeneTag team, for contributing to the success of the profiling experiment.

REFERENCES

1. **Bachem, C.W.B., R.S. van der Hoeven, S.M. deBruijn, D. Vreugdenhil, M. Zabeau and R.G.F. Visser.** 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. Plant J. *9*:745-753.
2. **Bloch, W.** 1992. Anion-exchange separation of nucleic acids. European patent application No. 92302930.0. 1991. U.S. Patent Application No. 679736.
3. **Hewlett Packard.** 1999. Product Insert for the Eclipse dsDNA Analysis Column. Hewlett Packard, Palo Alto, CA.
4. **Hewlett Packard.** 1998. Technical Report for the Zorbax Eclipse dsDNA Analysis Column. Hewlett Packard, Palo Alto, CA.
5. **Huber, C.G., P.J. Oefner and G.K. Bonn.** 1995. Rapid and accurate sizing of DNA fragments by ion-pairing chromatography on alkylated nonporous poly(styrene-divinylbenzene) particles. Anal. Chem. *67*:578-585.
6. **Huber, C.G., P.J. Oefner, E. Preuss and G.K. Bonn.** 1993. High-resolution liquid chromatography of DNA fragments on nonporous poly(styrene-divinylbenzene) particles. Nucleic Acids Res. *21*:1061-1066.
7. **Katz, E.D. and M.W. Dong.** 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. BioTechniques *8*:546-555.
8. **PE Biosystems.** 1997. Protocol for the ABI Prism dRhodamine Dye Terminator Cycle Sequencing Ready Reaction Kit. PE Biosystems, Foster City, CA, USA.
9. **Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot et al.** 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. *23*:4407-4414.
10. **Wages, J.M., X. Zhao and E.D. Katz.** 1995. High-performance liquid chromatography analysis of PCR products, p.140-153. *In* M.A. Innis, D.H. Gelfand and J.J. Sninsky (Eds.), PCR Strategies. Academic Press, San Diego, CA.
11. **Zeillinger, R., C. Schneeberger, P. Speiser and F. Kury.** 1993. Rapid quantitative analysis of differential PCR products by high-performance liquid chromatography. BioTechniques *15*:89-95.

Received 26 July 1999; accepted 1 November 1999.

Address correspondence to:

Dr. Victoria L. Boyd
Celera Genomics
850 Lincoln Centre Drive
Foster City, CA 94404, USA
Internet: boydvl@fc.celera.com