Efficient selection of 3'-terminal exons from vertebrate DNA

David B. Krizman and Susan M. Berget*

Verna and Marrs McClean Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030, USA

Received July 8, 1993; Revised and Accepted September 13, 1993

ABSTRACT

Identification of expressed sequences within genomic DNA is a hurdle in the characterization of complex genomes. We developed an exon trapping scheme that provides a positive selection for vertebrate 3'-terminal exons. A copy of the trapped exon sequence is obtained by RT/PCR amplification. The technique detects valid terminal exons without interference from partial exons or non-specific sequences, including simple human repeated sequences. Application to random human cosmid yielded one unique trapped terminal exon per cosmid on average. Because vertebrate terminal exons average 600–700 nucleotides in length, the technique provides transcribed sequences of sufficient length to assist further mapping efforts.

INTRODUCTION

The goal of positional cloning methods of gene identification is the elucidation of those regions of DNA that harbor transcribed sequences. A major problem facing these efforts arises from the fact that most vertebrate genes are split into a number of very small exons separated by long introns. Theoretically, exons could be located either by virtue of their open reading frames or by the RNA processing signals that border them. 5' splice sites are reasonably well conserved; however, the consensus sequence is relatively short and only the two nucleotides at the splice junction are invariant. 3' splice sites are poorly conserved with the exception of the invariant dinucleotide at the splice site (1). Thus, finding exons by either sequence analysis or PCR amplification from conserved splicing sequences is problematic, although recent informatic approaches to exon identification in sequenced DNA are encouraging (2–5).

Alternative approaches to finding transcribed sequences rely on the ability to detect a cDNA complementary to a designed genomic probe (6–8). cDNA-based strategies suffer from both the difficulty of creating a unique, contiguous probe from a piece of uncharacterized genomic DNA and the inherent sequence bias in most cDNA libraries. Because of these problems, technologies aimed at detection of expressed sequences within stretches of genomic DNA, regardless of their pattern of expression, are

useful. To this end, exon trapping schemes have been developed. By using the cellular transcription and processing machinery to convert genomic DNA into mRNA, exon identification can gain the specificity of the cellular exon identification process. Current exon trapping protocols are designed to trap either internal exons or isolated 3 splice sites (9–11). Recent work from our laboratory has suggested that 3'-terminal exons are units of recognition in vertebrate splicing (12,13). From this perspective, we developed a trapping protocol based on the requirement for an intact 3'-terminal exon within any transcription unit.

MATERIALS AND METHODS

Plasmid construction

pTAG4 was constructed in the vector pSL301 by insertion of a 710 bp transcription unit in the NcoI site of the superlinker. The transcription unit consists of a 410 bp NcoI/HindIII fragment of the SV40 early region promoter obtained from pSV-Bgal (nucleotides 10 to 420, Promega) inserted upstream of a 300 bp BamHI fragment from pIVBX that consists of 35 bp of SP6 promoter, 40 bp adenovirus leader exon 1, 230 bp truncated natural adenovirus intron 1, 72 bp adenovirus leader exon 2, followed by the first 40 bp of natural adenovirus intron 2. The polylinker of pSL301 begins 40 bp downstream of the adenovirus leader exon 2 5' splice site.

Cell culture and transfections

Cos7 cells were kindly provided by Dr John Wilson and were maintained in DMEM (10% FCS). Transfectase reagent (Life Technologies Inc.) was used to introduce 1 µg of recombinant pTAG4 into 6-well tissue culture plates of Cos7 cells according to manufacturers recommendations.

RT/PCR assay

Total cellular RNA was harvested 24hrs after transfection by the RNazol B method (CINNA/BIOTECX). The reverse transcription reaction used 5 µg of total RNA mixed with 1 µg oligo(dT) 3RACE adapter primer in 1mM EDTA to a final volume of 20 µl. 50 µl of mineral oil was overlaid and the mixture was heated to 94°C for 3 min., cooled to 42°C for 20 min., at which time 30 µl of the reverse transcription mixture was added.

* To whom correspondence should be addressed

The RT mixture consisted of 10 μ l of 5 $\times$ RT buffer (250mM Tris-HCl; pH7.5; 250mM KCl; 50mM DTT; 50mM MgCl₂), 1 μ l of 25mM dNTPs, 1 μ l of actinomycin D (500 μ g/ml in H₂O), 2 μ l of AMV-RT (15 units), and 16 μ l of sterile H₂O. The RT reaction was incubated at 42°C for 60 min. followed by the addition of 2.5 μ l stop solution (100mM EDTA, 800mM NaOH) and heated to 94°C for 3 min. The cDNA was precipitated and resuspended in 50 μ l TE.

Polymerase chain reaction was performed by standard methods. Trapping of 3'-terminal exons from plasmid inserts (HPRT and ADA) used a single round of PCR. 5 μ l of the cDNA was amplified in 100 μ l volume of PCR reaction consisting of 10mM Tris-Cl; pH8.4, 50mM KCl, 1.2mM MgCl₂, 0.01% gelatin, 250 μ M dNTPs, 50 pmoles of each primer, and 2.5 units of Taq DNA polymerase. Reactions were performed in a Genetic Thermal Cycler 2 (Precision Scientific) according to the following parameters. Reactions were initially incubated at 94°C for 5 min, 55°C for 2 min., and 72°C for 40 min. to ensure complete second strand cDNA synthesis. The reactions were then cycled 30 times following a protocol of 94°C (45sec.), 55°C (45 sec.), and 72°C (2min.) with a final extension at 72°C for 10 min.

A nested polymerase chain reaction was performed for the cosmid trapping experiments using primers from vector sequences and the tail sequence of the RT adaptor oligo(dT) primer. 5 μ l of cDNA was amplified in 100 μ l volume of PCR #1 reaction consisting of 10mM Tris-HCl (pH8.4), 50mM KCl, 1.2mM MgCl₂, 0.01% gelatin, 250 μ M dNTPs, 50 pmoles of primer SV40P, 5 pmoles of primer 48-O, and 2.5 units of Taq DNA polymerase according to the following parameters; 94°C for 5 min., 55°C for 2 min., and 72°C for 40 min. to ensure complete second strand cDNA synthesis followed by 30 cycles of 94°C (45 sec.), 55°C (45 sec.), and 72°C (2 minutes) with a final extension at 72°C for 10 minutes. 0.1 μ l from PCR #1 was used for PCR #2 using the same volume and reagent concentrations except for the use of a nested set of primers, Ad2 and 48-I at 50 pmoles each. Primer SV40P is directed to the portion of the SV40 promoter that is transcribed while Ad2 is specific for exon 2 from the adenovirus. 48-O and 48-I are both specific for the tail sequence of the oligo(dT) RT adapter primer. Products were analyzed by 1.2% agarose gel electrophoresis. Polaroid photos or negatives of photos of ethidium bromide-stained gels are displayed.

For the gene expression studies pairs of GTS primers were synthesized in a sequence-specific manner for each potential trapped 3'-terminal exon. Five micrograms of total HeLa RNA was reverse transcribed as described above with the following exceptions: 1) 3' GTS-specific primers to prime the RT reaction instead of the 3 RACE oligo(dT) RT adapter primer, 2) 55°C incubation for 60 min. instead of 42°C. A single round of PCR (58°C annealing for 35 cycles) was performed on 5 μ l of GTS-specific primed HeLa cDNA.

RESULTS

Trapping strategy

Our general approach to trapping 3'-terminal exons from genomic DNA is diagrammed in Figure 1A. The strategy is designed to trap 3'-terminal exons in a mammalian expression vector containing a partial transcription unit consisting of a promoter and two viral exons, but lacking a terminal exon. Such a transcription unit does not produce stable polyadenylated mRNA upon transfection of cultured mammalian cells unless a terminal

exon is provided by the insert DNA to be tested. Note that an entire 3'-terminal exon, including all intronic sequences required to recognize the 3' splice site beginning the exon and all sequences upstream and downstream of the poly(A) site required for polyadenylation must be present to successfully complete the transcription unit. The trapped 3'-terminal exon is detected and subsequently obtained by RT/PCR amplification of RNA from cells transfected with the recombinant plasmid(s) by performing 3 rapid amplification of cDNA ends (3 RACE) (14). The RNA is reverse transcribed using an oligo(dT) primer specific for the poly(A) tail and containing a unique sequence X, and subsequently subjected to PCR amplification utilizing primers complementary to vector-specific viral exons and the sequence X. Use of an oligo(dT)-based primer for RT ensures that all polyadenylated mRNA species will be substrates for PCR analysis. Use of a viral-specific primer for PCR ensures that no endogenous cellular mRNAs will contribute to signal.

Trapping of known last exons

We used the technique in Figure 1A to isolate 3'-terminal exons from previously cloned segments of the human hypoxanthine phosphoribosyltransferase (HPRT) gene (Figure 1B) and the human adenosine deaminase (ADA) gene (data not shown) (15,16). Observation of signal from RT/PCR amplification required the insertion of foreign genomic DNA into the trapping vector. Little to no assayable mRNA was detected in cells transfected with vector lacking insert (Figure 1B, lanes 1-4), using either of two primer sets. Correctly spliced and polyadenylated mRNA containing vector-derived exon 2 and the HPRT terminal exon was detected from transfections using vector containing either the purified terminal exon (lanes 5 and 6) or a shotgunned mixture of restriction fragments from the HPRT gene (lanes 7 and 8). Note there are two PCR products (485bp and 725bp) resulting from amplification using primer set b/d, due to the fact that there are two polyadenylation sites, a₁ and a₂, within the terminal exon of the HPRT gene. The two bands represent the usage of both of these poly(A) sites in the population of vector-derived mRNA. Only the a₂ site is represented in known HPRT cDNAs; it is unclear if site a₁ is used in HPRT mRNAs from other sources (17). It appears that the reported poly(A) site is utilized at a much greater rate than the alternative site based on the ethidium bromide intensity of the PCR products. Diagnostic restriction analysis of the PCR products confirmed the identity of the PCR products (data not shown). These data indicate that the vector design provides a positive selection for 3'-terminal exons.

Results from the ADA trapping experiments were similar. A mixture of restriction fragments from an 18.5kb segment of the human ADA gene were subjected to trapping. Correctly spliced and polyadenylated mRNA containing the ADA 3'-terminal exon was detected following transfection. Thus, even from a source of DNA of 18.5kb in length, the technique is specific for 3'-terminal exons.

Partial exons are not selected

To determine if partial exons containing isolated 3' splice sites or poly(A) sites would score as false positives in the trapping protocol, specific sub-exonic DNA fragments from the HPRT gene were tested (Figure 1C). Neither the first half of the HPRT 3'-terminal exon (nucleotides 41155 to 41590, including all elements of the 3' splice site) (lane 2) nor the second half of the exon (nucleotides 41590 to 42110, including all elements of the

A

B

C

poly(A) site) (lane 3) directed the synthesis of stable RNA upon transfection. Both elements of a terminal exon, therefore, are required to define a terminal exon and produce a positive trapping signal in our protocol (lane 1). Thus, the designed vector provides a positive selection for 3'-terminal exons with low potential for false positives arising from activation of cryptic 3' splice sites or poly(A) sites within genomic DNA.

As a further control, total *E. coli* DNA was digested to completion with *RsaI* and shotgun cloned into the trapping vector. Over 500 recombinant plasmids were grown in a 10ml LB culture overnight and an alkaline lysis miniprep was performed to isolate plasmid DNA. Results indicate that no polyadenylated chimeric mRNA resulted from transfection of these recombinants (Figure 1C, lane 4). The lack of signal from this experiment indicates that simple sequence DNA does not present a false positive in the assay.

It should be noted that the human HPRT gene tested in Figure 1B contains both ALU repeats as well as the hexanucleotide AATAAA present throughout the introns. These sequences did not generate a positive signal, again suggesting that the technique will not be hampered by high levels of false positives from common human repeat sequences and random hexanucleotides that may appear as poly(A) sites. Our hope is that the need for a 3' splice site in conjunction with a valid poly(A) signal (ie, a defined last exon) will keep false positive scoring to a minimum.

Trapping from complex DNA sources

To assess the ability of the technique to trap exons from complex sources of genomic DNA, the approach was used to detect 3-terminal exons in randomly-chosen cosmids from chromosome X (Lawrence Livermore National Laboratory X cosmid library clones U133: A6, A7, A8, A9, and A10). ADACos2.2, a cosmid clone encompassing the entire mouse adenosine deaminase gene, was used as a positive control (18). Following restriction nuclease digestion with either *EcoRI* or *PstI* and ligation into the trapping

Figure 1. Trapping of 3'-terminal exons from plasmids. (A) Outline of the 3'-terminal exon trapping strategy. DNA fragments are inserted into the pTAG4 mammalian expression vector which lacks a natural 3'-terminal exon. Insertion is downstream of exons 1 and 2 derived from the human adenovirus 2 late transcription unit (white boxes). If the insert DNA contains a last exon (hatched box), stable mRNA is made upon transfection of mammalian cells. A copy of the mRNA is generated by 3' RACE (RT/PCR) as diagrammed. (B) RT/PCR analysis of a trapped 3'-terminal exon from human HPRT. Lanes 1–2, RNA from untransfected Cos 7 cells; lanes 3–4, RNA from cells transfected with vector lacking insert; lanes 5–6, RNA from cells transfected with vector containing a restriction fragment containing just the 3'-terminal HPRT exon; lanes 7–8, RNA from cells transfected with vector containing a mixture of *SmaI*-digested fragments generated from a 4.5kb portion of the human HPRT gene. The sizes of the expected RT/PCR products are shown to the right of the gel. Primers b, c, and d are specific for adenovirus exon 2, the HPRT 3'-terminal exon, and the X tail sequence on the RT primer, respectively. Lanes 1,3,5,7, primer set b/c; lanes 2,4,6,8, primer set b/d. The HPRT terminal exon contains two poly(A) sites denoted a1 and a2. Only PCR primer d detects usage of both sites. Size markers are pUC BM21 digested with *HpaII*, *DraI*, and *HindIII* (Boehringer Mannheim Biochemicals). (C) RT/PCR analysis of partial exons. RNA from transfections of lane 1, vector containing a complete 3'-terminal exon; lane 2, the first half of the human HPRT 3'-terminal exon including all of the elements of the 3' splice site; lane 3, the second half of the 3'-terminal HPRT exon including all elements of the polyadenylation signal; and lane 4, *E. coli* DNA digested with *RsaI*. Primers are set b/d from (B). Lanes 5–7 are positive controls for the reverse transcription product from lanes 2–4, respectively, in which primers for cellular GAPDH mRNA were used.

Figure 2. Trapping of 3'-terminal exons from random human cosmids. (A) RT/PCR products from 3'-terminal exon trapping assay of cosmids from the human X chromosome. Cosmid DNA was digested with *EcoRI* for cloning into pTAG4 by the direct ligation/transfection approach described. mRNA was analyzed by nested PCR following reverse transcription. Lanes 1–5, inserted DNA from cosmids U133 A6, A7, A8, A9, and A10, respectively. Lane 6, positive control using plasmid shown in Fig. 1B. Trapping of the 3-terminal exon of the mouse ADA gene from ADAcos2.2 is shown to the right. The dominant PCR products from lanes 1–4 were gel purified, cloned, and multiple clones were sequenced. (B) Sequence analysis of the positive strand of the candidate terminal exons. Boxed sequences at the 5' end of each show the viral exon sequences beginning each RT/PCR product. Underlined sequences at the 3' end indicate the AAUAAA polyadenylation signal. (C) RT/PCR analysis of HeLa cell RNA. GTS primers were created from the A6.1 and A7 trapped terminal exons and used to amplify HeLa cell RNA. Lanes 1,3, amplification with reverse transcription; lanes 2,4, without reverse transcription. GTS primers from cosmids A6.2, A8, and A9 gave no detectable signal.

vector, the ligation reactions were transiently transfected into Cos7 cells; the total-cell RNA produced was assayed using nested PCR to amplify the RT product (Figure 2A). The ADAcos2.2 control produced a PCR product of correct size and restriction nuclease digestion pattern to indicate trapping of the ADA 3'-terminal exon (Figure 2A, unlabelled lane on right). The random cosmids also produced signals. Three cosmids yielded one major band; one yielded two bands; and one yielded no major band (Figure 2B, lanes 1–5). Lane 6 is a positive control for transfection and RT/PCR. The dominant PCR products (as judged by ethidium staining intensities) were gel-purified, cloned, and multiple clones were sequenced. Other PCR products appeared in each lane that were not purified and sequenced. These bands were not as bright by ethidium inspection and thus were not selected. However, this does not rule out the possibility that these products represent other potential trapped 3'-terminal exons or false positive scoring from cryptic processing signals. The sequences of the potential trapped terminal exons are shown in Figure 2B. Three of the five cosmid clones (A7, A8, A9) yielded one potential 3'-terminal exon per cosmid of sizes 110, 275, and 300bp, respectively. A6 yielded two distinctly different products of the same general size (A6.1=177 bp and A6.2=167 bp).

Sequence analyses (Figure 2B) indicated that all five potential terminal exons contained a consensus AATAAA or ATATAA polyadenylation signal located 15–24 bp upstream of the poly(A) tract. Each sequence began with adenovirus sequence and diverged into new sequence exactly after the last nucleotide of vector exon 2 indicating that a correct splicing event had occurred at the 5' splice site of exon 2. Termination codons within all three reading frames were found in each potential exon. Thus, the trapped sequences had all of the sequence hallmarks characteristic of valid 3'-terminal exons. GenBank searches revealed no significant homology between the trapped exons and any previously-discovered sequences. Thus, these five exons potentially represent five new human genes.

Sequence analysis of the potential last exons was used to generate a gene tagged site (GTS) for each and to synthesize oligonucleotide primers specific to each newly-acquired sequence for further characterization. The GTS primers were used in PCR reactions to map each GTS back to both the cosmid of origin and genomic DNA (data not shown). Each GTS was successfully mapped to its cosmid of origin, indicating that the trapped exon did derive from the tested cosmid. Each 3'-terminal exon was also detected in human genomic DNA. The trapped fragments were the same size when amplified from either cDNA or genomic DNA indicating that no internal upstream exons were co-trapped with the 3'-terminal exon. We have, however, observed co-trapping of internal exons with other sources of genomic DNA (data not shown). The ability of internal exons to be co-trapped depends upon their distance from the 3'-terminal exon (and, therefore, their presence in the same restriction fragment) and the relative strength of the splice sites of the exon as compared to those of the vector and 3'-terminal exons.

Expression studies of each trapped exon was accomplished via GTS-oligo primed RT/PCR amplification of HeLa cell RNA (Figure 2C). Two of the gene tagged sites, identified from cosmids A6.1 and A7, identified a gene that is expressed in HeLa cells. All gene tagged sites were assayed for expression in HeLa both with (+) and without (–) the addition of reverse transcriptase to the RT reaction to exclude the possibility of genomic DNA contamination of the RNA preparation. We are

presently testing other sources of RNA to ascertain if the other cosmids code for exons that are transcribed in other cell types.

A major concern when wishing to analyze more complex sources of target DNA is the need to amplify recombinant plasmids in *E. coli* prior to the transfection of tissue culture cells. To circumvent any bias in sequence maintenance by passage through bacteria, a direct ligation/transfection protocol was devised that takes advantage of the ability of cells to transiently express linear DNA molecules. Briefly, the trapping vector is digested with *MscI* and one of 20 sites within the multiple cloning site to leave a blunt end 4 kb upstream of the SV40 promoter and a complementary protruding end within intron 2. Target DNA is digested to produce the same complementary ends. The two DNAs are ligated in equal molar ratios. The ligation mixture is used to directly transfect tissue culture cells without an intermediate amplification in bacteria. This approach was used to generate the data in Figure 2.

DISCUSSION

The vector strategy employed was developed on the basis of earlier work suggesting that the exon is the unit of splice site selection in vertebrates (19). From this perspective, recognition of 3'-terminal exons involves cooperation between the splicing factors that recognize 3' splice sites and polyadenylation factors that recognize poly(A) sites. The results presented here are consistent with the concept that terminal exons are units of recognition during pre-mRNA processing and that in genes with introns, neither 3' splice sites nor poly(A) sites are capable of being recognized by the processing machinery without the presence of the other.

Other laboratories have used the cellular processing machinery to identify exons within stretches of genomic DNA. Two current approaches use trapping of either isolated 3' splice sites or intact internal exons. The first approach suffers from a high background of selection of false positives as might be expected from a technique requiring the presence of only one of a pair of required sites. Our trapping protocol failed to detect false positives thus demonstrating a high degree of specificity for intact 3'-terminal exons. The second approach is directed at trapping internal exons and has successfully been used to select new human genes (20,21).

We suggest that trapping of 3'-terminal exons offers the following advantages over previously-described exon trapping protocols: 1) last exons are 627 bp on average (22), thus providing a sequence long enough for physical gene mapping approaches, 2) this trapping scheme involves a positive selection with a low background and thereby eliminates sorting through false positives, 3) the technique is detecting 3'-terminal exons from random cosmids at close to the theoretical estimate, suggesting that it does not suffer from a high rate of false negatives, and 4) the vast majority of vertebrate genes contain a single 3'-terminal exon, thus minimizing multiple scoring of the same gene. Once a PCR product representing the 3'-terminal exon of a given gene is obtained, sequencing provides a gene tagged site (GTS) for further use in mapping the gene.

A potential limitation of this approach is in the selection of large 3'-terminal exons. Some last exons can be on the order of thousands of nucleotides in length and these may be missed in the process of PCR. While the average length of last exons is 627 bp the vast majority fall in the 200–400 bp range, thus

residing within a reasonable size for the PCR reaction. A second potential limitation is the rate of background of the system. A more exhaustive analysis will be necessary to understand what artifacts might arise, such as cryptic poly(A) signals scoring positive, as well as how to identify what is positive scoring and what is false positive scoring. This technology has the potential to aid the human genome effort by providing a means to identify transcribed sequences within genomic DNA that has yet to be sequenced.

ACKNOWLEDGEMENTS

We thank Dr Craig Chinault for pHPT3.7, Dr Rodney Kellems for pADA18.5 and ADACos 2.2, and Drs Gilbert Cote, Louis Zumstein, and Mike Frohman for many discussions. DNA sequence analysis and oligonucleotide synthesis was carried out by the Nucleic Acids Core facility in the Institute for Molecular Genetics at Baylor College of Medicine. The X chromosome-specific cosmid library LLOXNC01 was constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory under the auspices of the National Laboratory Gene Library Project sponsored by the US Department of Energy. This work was supported by Grant # USPH RO1 HG-00220 to S.B. and NIH postdoctoral fellowship # HG-000074 to D.B.K.

REFERENCES

- Moore, M., Query, C.C., and Sharp, P.A. (1992) In, *The RNA World*. Cold Spring Harbor Press, Cold Spring Harbor, N.Y., pp. 1–30.
- Konopka, A.K., and Owens, J. (1990) *Genet. Anal. Tech. & Appl.* **7**, 35–38.
- Uebacher, E.C., and Mural, R.J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1261–1265.
- Brunak, S., Engelbrecht, J., and Knudsen, S. (1991) *J. Mol. Biol.* **220**, 49–65.
- Jackson, I.J. (1991) *Nucleic Acids Res.* **19**, 3795–3798.
- Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D., and Weissman, S.M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9623–9627.
- Lovett, M., Kere, J., and Hinton, L.M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9628–9632.
- Hochgeschwender, U. (1992) *Trends in Genet.* **8**, 41–44.
- Duyk, G.M., Kim, S., Myers, R.M., and Cox, D.R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8995–8999.
- Auch, D., and Reth, M. (1991) *Nucleic Acids Res.* **18**, 6743–6744.
- Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A., and Housman, D.E. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4005–4009.
- Niwa, M., Rose, S.D., and Berget, S.M. (1990) *Genes & Dev.* **4**, 1552–1559.
- Niwa, M., and Berget, S.M. (1991) *Genes & Dev.* **5**, 2086–2095.
- Frohman, M.A., Dush, M.A., and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
- Edwards, A., Voss, H., Rice, P., Civitello, A., Stegemann, J., Schwager, C., Zimmerman, J., Erfle, H., Caskey, C.T., and Ansorge, W. (1990) *Genomics* **6**, 593–608.
- Wiginton, D.A., Kaplan, D.J., States, J.C., Akeson, A.L., Perme, C.M., Bilyk, I.J., Vaughn, A.J., Lattier, D.L., and Hutton, J.J. (1986) *Biochem.* **25**, 8234–8244.
- Jolly, D., Okayama, H., Berg, P., Esty, A.C., Filpula, D., Bohlen, P., Johnson, G.G., Shively, J.E., Hunkapillar, T., and Friedmann, T. (1982) *Proc. Natl. Acad. Sci. USA* **80**, 477–481.
- Al-Ubaidi, M.R., Ramamurthy, V., Ming-Chei, M., Ingolia, D.E., Chinsky, J.M., Martin, B., and Kellems, R.E. (1990) *Genomics* **7**, 476–485.
- Robberson, B.L., Cote, G.J., and Berget, S.M. (1990) *Mol. Cell. Biol.* **10**, 84–94.
- Trofatter, J.A., MacCollin, M.M., Rutter, J.L., Murrell, J.R., Duyao, M.P., Parry, D.M., Elridge, R., Kley, N., Menon, A.G., Pulaski, K., Haase, V.H., Ambrose, C.M., Munroe, D., Bove, C., Haines, J.L., Martuza, R.L., MacDonald, M.E., Seizinger, B.R., Short, M.P., Buckler, A.J., and Gusella, J.F. (1993) *Cell* **72**, 791–800.
- Huntington Disease Collaborative Research Group. (1993) *Cell* **72**, 971–983.
- Hawkins, J.D. (1988) *Nucleic Acids Res.* **16**, 9893–9908.