GATA 11(2): 34–38, 1994

A Consensus *Alu* Repeat Probe for Physical Mapping

MARK A. BATZER,
MICHELLE ALEGRIA-HARTMAN,
and PRESCOTT L. DEININGER

*Physical mapping of the human genome involves a variety of complex hybridization-based procedures, some of which rely upon the ability to separate human clones derived from human–rodent hybrid cell lines from those that contain background rodent-derived DNA sequences. The ability to block the repetitive element (*Alu *repeat) portion of inter-*Alu *PCR products derived from a variety of complex sources is also crucial for the isolation of unique DNA sequences. Here we report the construction and characterization of a new consensus* Alu *repeat probe (pPD39) designed for these purposes.*

Introduction

The isolation and analysis of specific DNA sequences from high-density recombinant library filter grids derived from cosmids [1], bacterial artificial chromosomes (BACs) [2], yeast artificial chromosomes (YACs) [3], and bacteriophage P1-derived artificial chromosomes (PACs) [4] is crucial to mapping the human genome. High-density filter grids may be screened using a variety of complex probes including radiolabeled chromosome-specific flow-sorted material, polymerase chain reaction (PCR) products derived from any number of popular substrates, inter-*Alu* PCR products derived from monochromosomal hybrid cell lines, or other large insert recombinant clones, plasmids, or highly specific oligonucleotide probes. The vast majority of these procedures are limited only by the specificity of the hybridization probe. The application of radiolabeled inter-*Alu* PCR products to this procedure facilitates the rapid preparation of human specific DNA probes derived from complex hybrid cell lines and large insert clones [5]. However, the amplification of exogenous DNA sequences using inter-*Alu* PCR also leaves a portion of the *Alu* repeat sequence itself at the 5′ and 3′ termini of the derived products and the potential for background hybridization to the *Alu* repeats. To mitigate this background, inter-*Alu* PCR-derived hybridization probes or mixes must be preannealed to blocking DNA. The blocking DNA commonly used in these procedures is either Cot-1 DNA fraction, or *Alu* repeat plasmid DNA from the clone BLUR 8 [6]. Cot-1 DNA fraction is time consuming to make, but is commercially available. BLUR 8 plasmid DNA can be isolated within the laboratory but suffers from a small truncation in the 5′ region of the BLUR 8 *Alu* repeat sequence [6]. The truncation in BLUR 8 limits the blocking potential for 5′ portion of *Alu* repeats, resulting in an increase in background hybridization to *Alu* sequences.

The unique identification of human clones derived from complex sources such as human–rodent hybrid cell lines is commonly facilitated based upon hybridization screening of the recombinant clones. This type of screening is generally carried out using radiolabeled material derived from species-specific repeated DNA sequence probes such as the *Alu* clone BLUR 8, Cot-1 DNA fraction, or total genomic DNA for humans. Each of these procedures suffers as a result of the expense involved in generating the material (Cot-1), decrease in hybridization signal intensity from single-copy DNA sequences (total human DNA), or mismatch with the majority of *Alu* repeats (BLUR 8). To mitigate these shortcomings, we have constructed a new *Alu* consensus clone (pPD39) for the preannealing of complex probe mixtures and the identification of human derived clones. This probe has the advantage of representing a perfect match to the human specific (HS) *Alu* subfamily consensus sequence [7, 8], minimizing the mismatch to the majority of *Alu* repeats dispersed throughout the genome.

Materials and Methods

DNA Samples

DNA from PAC 24G6 was isolated from a 50-ml LB broth overnight culture supplemented with 20

From the Human Genome Center, L-452 (M.A.B., M.A.-H.), Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California; and the Department of Biochemistry and Molecular Biology (P.L.D.), Louisiana State University Medical Center; and the Laboratory of Molecular Genetics (P.L.D.), Alton Ochsner Medical Foundation, New Orleans, Louisiana, USA.

Address correspondence to Dr. M.A. Batzer, Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, PO Box 808, Livermore, CA 94551, USA.

Received 3 January 1994; accepted 15 March 1994.

1050-3862/94/$7.00

μg/ml kanamycin, as previously described, using a modified alkaline lysis procedure [4]. Plasmid DNA samples from the clones pPD39 and BLUR 8 were also prepared from 50-ml LB broth overnight cultures that contained 50 μg/ml ampicillin, using Qiagen tips as recommended by the manufacturer. HS C3N1 DNA was isolated from a 5-ml LB broth culture supplemented with 50 μg/ml ampicillin by using a standard alkaline lysis miniprep procedure [9].

Construction of pPD39

The plasmid pPD39 was constructed with the PCR products derived from the amplification of clone HS C3N1 as outlined below using the primers

PD1 5′-GCCGGATCCGAAAAAAGAGCAGGGCAGT-3′
and
PD2 5′-GCCGGATCCGGCCGGGCGCGGTGGCTCACGCC-3′

which contain BamHI restriction sites. The amplified PCR products were digested to completion with BamHI and ligated to BamHI-digested dephosphorylated pBluescript KS+ (Stratagene) for 16 h at 14°C using standard conditions [9]. The resultant ligation mixture was used to transform competent *Escherichia coli* XL1 Blue cells (Stratagene) and plated on LB agar supplemented with 50 μg/ml ampicillin [9]. Small toothpick transfers were then made from several colonies and subjected to PCR amplification using T3 and T7 promoter-specific primers. A single colony that amplified a DNA fragment of ~500 bp was isolated and designated pPD39. The nucleotide sequence of clone pPD39 was subsequently determined by standard dideoxy procedures using Sequenase (US Biochemicals) and [α-^{35}S]dATP on plasmid templates with T3 and T7 promoter primers as well as internal HS *Alu*-specific primers [7, 8] and has been assigned Genbank accession number U 02043. The clone (pPD39) has been submitted to the American Type Culture Collection (ATCC 79806 for bacterial cells and 79807 for DNA). DNA sequences were compared by using the FSTNSCAN program from the PC/Gene (Intelligenetics) suite. The FSTNSCAN program is based on the FASTN algorithm of Lipman and Pearson [10].

Polymerase Chain Reaction Amplification

PCR amplification of plasmid DNA samples was carried out in 100-μl reactions by using 100 ng of target DNA, 750 ng of each primer, 200 μM dNTPs, and ampliTaq DNA polymerase (3 U) according to the supplier's instructions (Roche Biomedical). Each sample was subjected to the following amplification profile: 5 min at 94°C, and 30 of the following cycles: 45 s at 94°C (denature), 45 s at 50°C, and 45 s at 72°C (extension). A total of 20 μl of each sample was then fractionated on a 2% agarose gel with 0.5 μg/ml ethidium bromide. PCR products were directly visualized using UV fluorescence.

Hybridization to High-Density Cosmid Arrays

High-density grid filters in a duplicate 6 × 6 array derived from a chromosome-19-specific cosmid library were prepared as previously described [11]. DNA (100 ng) from the chromosome-19-specific PAC (24G6) clone was subjected to inter-*Alu* PCR using the primers *Alu*-5′ and *Alu*-3′ as previously described [12]. The PCR products were purified using column chromatography (Bio-Rad Prep-A-Gene kit) and radiolabeled using a Megaprime DNA-labeling system (Amersham) and [α-^{32}P]-dCTP by the random priming method [13] to a specific activity of $>10^8$ cpm/μg. The resultant radiolabeled inter-*Alu* PCR products were purified by column chromatography using minispin columns (Worthington) and equal amounts of the probes were preannealed with 500 μg/ml of BLUR 8, Cot-1 (Gibco BRL), or pPD39 DNA by boiling and incubation at 65°C for 20 min prior to hybridization. DNA from the plasmid pUC 119 was also radiolabeled as described above and used to hybridize to the offset positions of the high-density cosmid filter grids. The filters were prehybridized for 2 h at 65°C in 0.6 M NaCl, 0.2 M Tris pH 8.0, 2% SDS, and 0.1% sodium pyrophosphate, followed by hybridization in 50 ml of fresh solution that contained 1×10^7 cpm/ml of one of the preannealed inter-*Alu* PCR-derived probes for 16 h. The membranes were subjected to a stringent wash in 0.1× SSC–0.1% SDS for 30 min at 65°C and exposed to a PhosphorImager (Molecular Dynamics) plate overnight followed by quantitative analysis.

Results and Discussion

The nucleotide sequence of pPD39 as well as BLUR 8 are compared with the HS [7, 8] or predicted variant (PV) [14, 15] *Alu* subfamily consensus sequence in Figure 1. The clone pPD39 is an

```
HS CON      GGCCGGGCGC  GGTGGCTCAC  GCCTGTAATC  CCAGCACTTT  GGGAGGCCGA      50
pPD39       ..........  ..........  ..........  ..........  ..........
BLUR 8      xxxxxxxxxx  xxxxxxxxxx  xxxx......  ..........  ........A.

HS CON      GGCGGGCGGA  TCACGAGGTC  AGGAGATCGA  GACCATCCCG  GCTAAAACGG     100
pPD39       ..........  ..........  ..........  ..........  ..........
BLUR 8      ..A....A..  ....CT.AAGTC.....T.T..  .....G..T.  ..C..C.T..

HS CON      TGAAACCCCG  TCTCTACTAA  AAATACAAAA  AATTAGCCGG  GCGTAGTGGC     150
pPD39       ..........  ..........  ..........  ..........  ..........
BLUR 8      ......T..A  ........G.  ..........  .x......A.  ..A.G...AT

HS CON      GGGCGCCTGT  AGTCCCAGCT  ACTTGGGAGG  CTGAGGCAGG  AGAATGGCGT     200
pPD39       ..........  ..........  ..........  ..........  ..........
BLUR 8      .C.T.....G  .A........  ....A.....  .....A...A  .....CC.T.

HS CON      GAACCCGGGA  GGCGGAGCTT  GCAGTGAGCC  GAGATCCCGC  CACTGCACTC     250
pPD39       ..........  ..........  ..........  ..........  ..........
BLUR 8      A....AAx..  ..T....G..  ..........  ......G.A.  GG........

HS CON      CAGCCTGGGC  GACAGAGCGA  GACTCCGTCT  CAAAAAAAAA        290
pPD39       ..........  ..........  ..........  .......... A12
BLUR 8      ........Tx  ..........  ......A...  .........x
```

Figure 1. Alignment of pPD39 and BLUR 8. The alignment of pPD39 and BLUR 8 as compared with the human specific (HS) *Alu* subfamily consensus sequence [7, 8]. Nucleotide substitutions and insertions are denoted with the appropriate base. Deletions are marked with an *x*.

exact match to the HS subfamily consensus sequence with BLUR 8 being ~10% divergent from this subfamily. The HS subfamily of *Alu* repeats represents one of the most recently inserted groups of *Alu* repeats located within the human genome, and is comprised of 500 [7, 8]–2000 [14] nearly identical members. BLUR 8 is derived from the older primate specific (PS) subfamily of *Alu* repeats that amplified within the genome 40–60 million years ago [16], but is still diverged from that subfamily consensus by ~15%. The majority of *Alu* repeats dispersed throughout the genome differ from each other by 20%–30% [6]. Therefore, a consensus *Alu* repeat sequence should be the most effective hybridization probe to detect human DNA sequences. To ascertain the amount of nucleotide sequence similarity among BLUR 8, pPD39, and a random sample of *Alu* repeats from the human genome, we aligned each of the probes with the *Alu* repeats located within the human thymidine kinase (*tk*) gene [17]. The results of these alignments are shown in Table 1. These data show that pPD39 is on average a 3% better match with the *Alu* repeats located within the introns of the *tk* gene than is BLUR 8. This implies that pPD39 will serve as a slightly better hybridization probe for the identification of human DNA sequences than will BLUR 8 through hybridization to a broader spectrum of *Alu* repeats. Similar arguments apply to the mixture of diverse *Alu* repeats in total or Cot-1 DNA. In hybridization experiments using Southern blots derived from human–rodent hybrid cell line DNA to detect human *Alu* sequences, we find that pPD39 hybridized effectively at 5°C higher temperature than did BLUR 8 (data not shown). Thus, pPD39 is also useful as a probe to detect human DNA sequences (*Alu* repeats) under higher-stringency hybridizations.

Table 1. Similarity of Selected *Alu* Repeats and Probes[a]

	BLUR 8	BLUR 8 opt	pPD39	pPD39 opt
tkA	66.4	73.8	68.7	77.6
tkB	79.4	80.2	79.7	81.7
tkD	77.4	79.3	73.5	81.2
tkE	78.5	79.4	78.7	79.4
tkF	75.4	76.9	75.5	78.0
tkG	80.5	82.6	81.5	84.9
tkH	78.1	78.4	79.4	80.0
tkI	79.8	80.3	82.7	85.0
tkJ	81.8	81.8	84.7	85.5
tkK	84.2	84.8	85.6	86.4
tkL	82.1	82.6	93.0	93.0
Average		79.1		81.9

[a]*Alu* repeats taken from the human thymidine kinase gene [17]. Initial alignment using FSTNSCAN (PC/Gene) and manually optimized (opt). Scores are listed as the percentage of nucleotide overlap.

Another important criterion for the application of any hybridization probe is the amount of effort required to isolate a sufficient amount of material. The HS consensus *Alu* repeat located within pPD39 is flanked on either side by a variety of potential oligonucleotide-priming sites. Therefore, the consensus *Alu* repeat sequence within the pPD39 plasmid may be amplified using PCR from a combination of primers including the Bluescript forward and reverse primers as well as the T3 and T7 promoter primers as shown in Figure 2. This facilitates the rapid production of a large amount of *Alu* repeat-derived material from a small amount of template DNA. The ability to generate an *Alu* containing PCR product mitigates the necessity for large-scale plasmid DNA preparations. This also greatly increases the potential methods for radiolabeling the consensus *Alu* repeat sequence prior to use.

To ascertain the efficiency of pPD39 as a blocking (preannealing) reagent for inter-*Alu* PCR products, we made a direct comparison with two other common blocking reagents (Cot-1 DNA and BLUR 8). Inter-*Alu* PCR products derived from a human chromosome-19-specific PAC clone were radiolabeled and used as probes on high-density chromosome-19-specific cosmid filter arrays as shown in Figure 3. A single cosmid displayed a positive hybridization signal on this filter (in duplicate) from a set of seven total high-density chromosome-19-specific cosmid filters. In this approach, we were already aware of which cosmid clones should in fact hybridize with the PAC inter-*Alu* PCR products. This allowed us to determine

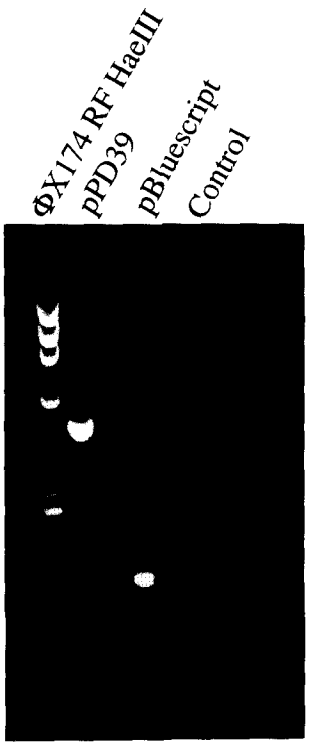

Figure 2. Polymerase chain reaction (PCR) amplification of pPD39. An agarose gel (2%) chromatograph of PCR products derived from the amplification of pPD39 using T3 and T7 primers (*Materials and Methods*): (*lane 1*) PCR products derived from the amplification of pPD39, (*lane 2*) pBluescript KS+ without the pPD39 *Alu* repeat; and (*lane 3*) water-negative control. The marker was ϕX174 RF HaeIII digest.

whether any of the blocking reagents had any type of detrimental effect upon the hybridization process. Following exposure to a PhosphorImager, we were also able to compare quantitatively the background hybridization intensity and positive hybridization signals to determine the relative effectiveness of pPD39 as a blocking reagent. A printout from the PhosphorImager shows that there is little if any visual difference in the effectiveness of the various blocking substrates. Several other chromosome-19-specific PAC and BAC clones gave similar results (data not shown). In general, after quantification and normalization of the data, pPD39 was 30%–40% more efficient as a blocking probe, resulting in approximately the same visual appearance. In each instance, all of the cognate-positive chromosome-19-specific cosmids were detected, indicating that there were in fact no detrimental effects introduced by any of the blocking reagents. The application of pPD39 as a blocking reagent is also enhanced by the ability to generate material using the PCR.

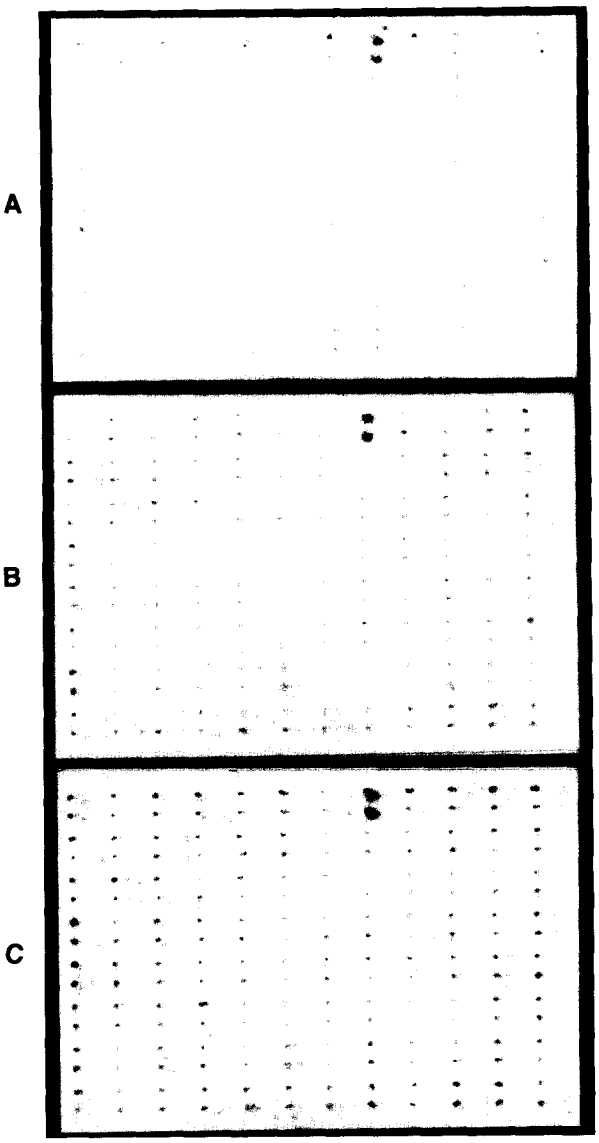

Figure 3. Application of pPD39 as a blocking reagent. Photograph of the hybridization of inter-*Alu* polymerase chain reaction (PCR) products derived from a chromosome-19-specific bacteriophage P1-derived artificial chromosome (PAC) clone (24G6) to a high-density chromosome-19-specific cosmid array after preannealing (blocking) with different blocking reagents and exposure to a Molecular Dynamics PhosphorImager. Radiolabeled inter-*Alu* PCR products derived from PAC 24G6 (*Materials and Methods*) were preannealed to either (**A**) Cot-1, (**B**) BLUR 8, or (**C**) pPD39 DNA prior to hybridization.

Summary

All human *Alu* family members can be roughly considered to be randomly mutated from the *Alu* consensus sequence by ~14%–15%. Thus, any two *Alu* sequences will show roughly twice that divergence from each other. Any individual *Alu* sequence, or even mixture of *Alu* sequences, will almost always show more divergence from another *Alu* repeat than from the *Alu* consensus sequence. Therefore, using a consensus *Alu* repeat as a blocking agent, or hybridization probe, has theoretical advantages. We have demonstrated that although all of the *Alu* blocking and hybridization strategies work well, there is some advantage to the use of pPD39. In addition, we demonstrate the simple preparation of pPD39 as a blocking and hybridization sequence by using a PCR approach.

This work was supported by NIH grants HG 00340 and HG 00770 (P.L.D.), US Department of Energy LDRD 94-LW-103 (M.A.B.) and CMC 07 from the University of California Center for Molecular Cytometry (M.A.B.). Work at Lawrence Livermore National Laboratory was performed under the auspices of the US Department of Energy contract W-7405-ENG-48.

References

1. De Jong PJ, Yokabata K, Chen C, Lohman F, Pedersen L, McNinch J, Van Dilla M: Cytogenet Cell Genet 51:985, 1989
2. Shizuya H, Birren B, Kim U-J, Mancino V, Slepak T, Tachiiri Y, Simon M: Proc Natl Acad Sci USA 89:8794–8797, 1992
3. Burke DT, Carle GF, Olson MV: Science 236:806–812, 1987
4. Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ: Nature Genet 6:84–89, 1994
5. Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Caskey CT: Proc Natl Acad Sci USA 86:6668–6690, 1989
6. Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW: J Mol Biol 151:17–33, 1981
7. Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL: Nuceic Acids Res 18: 6793–6798, 1990
8. Batzer MA, Deininger PL: Genomics 9:481–487, 1991
9. Ausabel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K: Current Protocols in Molecular Biology. New York, John Wiley and Sons, 1987
10. Lipman DJ, Pearson WR: Science 227:1435–1441, 1985
11. Olsen AS, Combs J, Garcia E, Elliot J, Amemiya C, de Jong P, Threadgill G: BioTechniques 14:116–123, 1993
12. Tagle DA, Collins FS: Hum Mol Gen 1:121–122, 1992
13. Feinberg AP, Vogelstein B: Anal Biochem 132:6–13, 1983
14. Matera AG, Hellmann U, Schmid CW: Mol Cell Biol 10: 5425–5432, 1990
15. Matera AG, Hellmann U, Hintz MF, Schmid CW: Nucleic Acids Res 18:6019–6023, 1990
16. Shen MR, Batzer MA, Deininger PL: J Mol Evol 33:311–320, 1991
17. Flemington E, Bradshaw HD Jr, Traina-Dorge V, Slagel V, Deininger PL: Gene 52:267–277, 1987