Proc. Natl. Acad. Sci. USA
Vol. 78, No. 3, pp. 1508–1512, March 1981
Biochemistry

Presence of a highly repetitive and widely dispersed DNA sequence in the human genome

(genomic library/structural genes/inverted repeats)

MASARO TASHIMA, BRUNO CALABRETTA, GIUSEPPE TORELLI, MARGARET SCOFIELD, ABBY MAIZEL, AND GRADY F. SAUNDERS

Department of Biochemistry, The University of Texas System Cancer Center, M. D. Anderson Hospital and Tumor Institute, Houston, Texas 77030

Communicated by Karl Folkers, November 24, 1980

ABSTRACT **A genomic DNA library consisting of human DNA fragments about 18 kilobases long cloned in a bacteriophage λ vector was found to contain a specific repeated DNA segment. The repeated sequence is present in >95% of the genomic library, and selected clones contain at least two copies of the sequence. Our experiments indicate that this highly repetitive sequence (≈400,000 copies per haploid genome) is widely distributed in the human genome and is represented in the cytoplasmic polysomal mRNA. This sequence is homologous to the 300-base-pair Alu repeat family, the predominant repeat sequence in man.**

A sizeable portion of the eukaryotic genome is composed of similar or repetitive DNA sequences that are characterized by their ease of reassociation. The repetitive sequences can be divided into highly repetitive and middle repetitive fractions by their relative reassociation rates. The sequences that appear to occur only once in the genome are called unique or single-copy sequences. These three kinetic classes of DNA can be fractionated on hydroxyapatite after annealing denatured sheared DNA to various extents.

The human genome consists of 7×10^9 base pairs (bp) of DNA, 65% of which are single-copy sequences (1). Inverted repetitive sequences compose about 5% of the total DNA (2, 3), and about 5% of the DNA is the simple sequence (satellite) type. The remaining 25% of the DNA is in middle repetitive sequences (300 bp long), which are interspersed with single-copy sequences about 1800 bp long (3). Except for a few known repetitive genes, the biological role of repetitive sequences is unknown. Their interspersion pattern with single-copy sequences (4) and proximity to transcribed structural genes have supported the hypothesis that repetitive sequences play a role in the regulation of gene expression (5).

A repetitive sequence family has recently been identified in human DNA (6, 7). This 300-bp sequence is present in >300,000 copies per haploid genome and contains a sequence cleaved by restriction endonuclease *Alu* I. The wide distribution of this sequence may have profound influence on the plasticity of the human genome, resulting in relatively high rates of genetic rearrangement. In the work reported here, we have examined the distribution of repetitive sequences in a human genomic DNA library and find that >95% of the library, which contains human DNA segments about 18 kilobases (kb) long, contains representatives of one family of repetitive sequences. Further, this family is represented in cDNA prepared by reverse transcription of polysomal poly(A)$^+$ RNA.

MATERIALS AND METHODS

Cell Culture. CCRF-CEM cells were grown in continuous suspension culture with RPMI-1640 medium supplemented with 10–20% fetal calf serum. This cell line was derived from a child with acute lymphocytic leukemia, and the cells have been characterized as thymus derived (8). The cells used for all subsequent experiments were harvested while in the logarithmic phase of growth. The cells were routinely tested for mycoplasma contamination and were uniformly negative (9).

Isolation of mRNA. Cells were washed three times with physiological saline and lysed in buffer (10 mM NaCl/2.5 mM KCl/1.5 mM $MgCl_2$/10 mM Tris·HCl, pH 7.4) containing 0.5% Nonidet P-40 (Particle Data Laboratories, Elmhurst, IL). Nuclei were removed and the cytoplasmic RNA was purified as described (10). To prepare polysomal RNA, washed cells were suspended in buffer and 10% (vol/vol) Nonidet P-40 was added to a final concentration of 0.5% to lyse the cells. After removal of nuclei, the suspension was adjusted to 0.25 M sucrose and centrifuged at 12,000 × *g* for 10 min. The postmitochondrial supernate was adjusted to 1% Triton X-100 and layered onto a 5 ml pad of 2 M sucrose. Polysomes were pelleted by centrifugation at 49,000 rpm for 3 hr in a Beckman 60 Ti rotor. Poly(A)$^+$ RNA was obtained by two passages of total cytoplasmic or polysomal RNA over an oligo(dT)-cellulose column as described by Aviv and Leder (11). The recovery of poly(A)$^+$ RNA was 0.6–1.2% of the starting material absorbing 260-nm light.

Synthesis of cDNA. Highly purified RNA-dependent DNA polymerase from avian myeloblastosis virus was kindly provided by J. W. Beard (Life Science, St. Petersburg, FL). The specific activity of the enzyme was over 10,000 units/mg of protein.

Template RNA (5 μg) was incubated in a mixture of 100 μl containing 50 mM Tris·HCl, pH 8.3; 3 μg of $(dT)_{12-18}$ (Collaborative Research, Waltham, MA); 8 mM Mg acetate; 50 mM KCl; 20 mM dithiothreitol; 500 μM each of dGTP, dATP, and dTTP; 50 μM [α-^{32}P]dCTP (120 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels); actinomycin D at 50 μg/ml; and 45 units of RNA-dependent DNA polymerase. The reaction mixture was incubated at 46°C for 20 min, after which the cDNA was separated from unreacted substrates by chromatography on Sephadex G-50 as described (10). The cDNA fractions were pooled, concentrated, and treated with 0.1 M NaOH for 45 min at 60°C, then precipitated with ethanol and dissolved in water. The size of the cDNA determined by electrophoresis in urea/agarose gels (12) was about 1000 nucleotides, and the specific activity was $\approx 10^8$ cpm/μg.

Isolation of Repetitive DNA. Human DNA was extracted from human placenta with the phenol/proteinase K method (13). Then DNA was sonicated to an average 350-nucleotide size and dissolved in 0.12 M sodium phosphate buffer, pH 6.8. After heat denaturation for 5 min in boiling water, the DNA was allowed to reassociate at 60°C to C_0t 40 M·sec and fractionated

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); bp, base pairs(s); C_0t, initial concentration of DNA (moles of nucleotides per liter) × time (sec); IR, inverted repeat.

by hydroxyapatite column chromatography (14).

Preparation of Labeled DNA. DNA was labeled *in vitro* by nick translation with *Escherichia coli* DNA polymerase I according to the method of Maniatis *et al.* (15), using one labeled nucleotide, [α-^{32}P]dCTP (3000 Ci/mmol, New England Nuclear). After deproteinization, the labeled DNA was separated from unincorporated nucleotides on a Sephadex G-50 column. The specific activity of the DNA was 10^8 cpm/μg.

Screening of Human Genomic DNA Library. The human DNA library was obtained from Tom Maniatis and its preparation has been described (16). The library was screened by using the *in situ* plaque hybridization method of Benton and Davis (17). Recombinant phage were plated at ≈300 phage per 15-cm petri plate on a lawn of *E. coli* DP50 SupF.

Hybridization was carried out at 68°C in Denhardt's solution containing 0.5% sodium dodecyl sulfate, 0.1 mM EDTA, and poly(rA) at 10 μg/ml. After hybridization, filters were washed twice with chloroform, dried, washed three times with 300 mM NaCl/30 mM sodium citrate/0.5% sodium dodecyl sulfate at 68°C for 2 hr, and rinsed with 150 mM NaCl/15 mM sodium citrate or 75 mM NaCl/7.5 mM sodium citrate at 68°C for 15–30 min. After drying, filters were covered with Saran Wrap and exposed in x-ray cassettes at room temperature or −80°C for the appropriate time.

Agarose Gel Electrophoresis and Transfer to Nitrocellulose Filter Paper. Cloned DNA, purified in accordance with the National Institutes of Health guidelines, was digested with *Eco*RI restriction enzyme according to the procedure supplied by the vendor (Bethesda Research Laboratories, Rockville, MD). DNA fragments were electrophoresed through vertical 0.8% agarose gels in 50 mM Tris/50 mM boric acid/10 mM EDTA, pH 8.3, at room temperature. The DNA was transferred to nitrocellulose filters by the method of Southern (18).

Filters were prehybridized in Denhardt's solution at 68°C overnight and then hybridized to ^{32}P-labeled probes in hybridization solution as used in screening at 68°C for 24 hr. Then filters were washed three times in 450 mM NaCl/45 mM sodium citrate/0.5% sodium dodecyl sulfate for 2 hr and rinsed in 150 mM NaCl/15 mM sodium citrate for 15–30 min at 68°C. To extract DNA from agarose gels, the DNA bands were cut from gels after electrophoresis, put in dialysis tubes, and then electrophoresed again in 5 mM Tris·HCl at pH 7.4/2.5 mM acetic acid until all ethidium bromide disappeared from the gels. The DNA was extracted with phenol/sevag and precipitated with ethanol.

RESULTS

cDNA Sequence in the Genomic Library. The human gene library in bacteriophage λ Charon 4A described by Lawn *et al.* (16) was obtained from T. Maniatis. The library was constructed by partial digestion of human fetal liver DNA with *Hae* III and *Alu* I restriction endonucleases, size fractionation, methylation with *Eco*RI methylase, *Eco*RI oligonucleotide linker addition, and cloning in the λ vector Charon 4A. We screened the recombinant phages for those containing messenger sequences by hybridization with radiolabeled cDNA transcripts of cytoplasmic poly(A)$^+$ RNA (mRNA) of CCRF-CEM cells. A very high proportion (≈95%) of the clones were positive (Fig. 1 and Table 1). Similar results were obtained with cDNA transcripts of polysomal poly(A)$^+$ RNA. We investigated the possibility that the signal on most genomic clones was due to a highly repetitive sequence in a subset of cDNA molecules, rather than the presence of structural genes. Repetitive DNA ($C_0t \leq 40$) was labeled by nick translation and used to screen the genomic library. Almost all clones (93%) (Fig. 1 and Table 1) gave a positive signal with the repetitive DNA probe. Furthermore, all clones giving a positive signal with the repetitive DNA probe were also positive with the cDNA probe. cDNA transcripts of total poly(A)$^+$ RNA of human placenta gave less hybridization (40% positive clones), suggesting a decreased concentration of the hybridizing species in some tissues. Doubling the amount of placental [^{32}P]cDNA used in screening the library gave 65% positive clones.

Table 1. Screening of the human genomic library with DNA probes

Plate	Number of plaques hybridizing with probe (% of total plaques)		
	Nick-translated phage λ DNA	CEM cDNA	Nick-translated repetitive DNA
1	220	211 (95.9)	200 (90.9)
2	232	219 (94.3)	212 (91.3)
3	273	256 (93.7)	248 (90.8)
4	215	203 (94.4)	198 (92.0)
5	245	235 (95.9)	233 (95.1)
6	241	231 (95.8)	222 (92.1)
Average	(100)	(95)	(92)

The number of positive plaques with each probe was counted on six different plates. CEM cDNA, 2.2×10^6 cpm per filter; repetitive DNA, 1×10^6 cpm per filter; phage DNA, 1×10^6 cpm per filter.

Evidence that the hybridizing species in the cDNA population was due to repetitive sequences came from competition experiments. Addition of a 1000-fold excess of repetitive DNA abolished the hybridization with [^{32}P]cDNA (Fig. 2). Several other polynucleotides were tested as competitors in the genomic screening (Table 2). Of these, repetitive DNA was an ef-

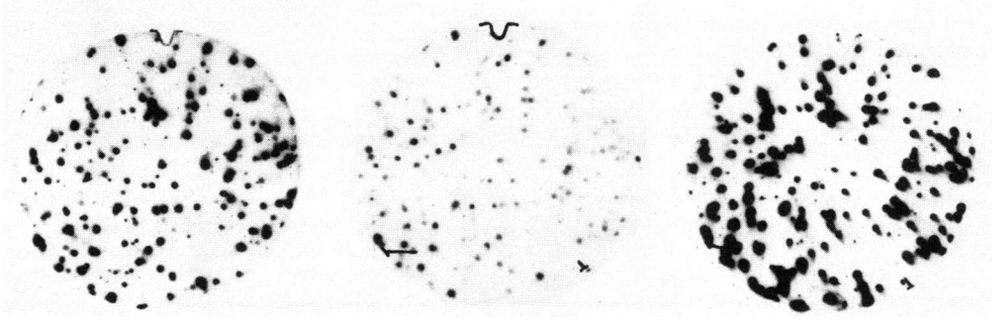

FIG. 1. Screening of the human genomic DNA library with [^{32}P]DNA probes by the Benton–Davis procedure (17). Filters were prepared in triplicate. (*Left*) [^{32}P]cDNA from cytoplasmic poly(A)$^+$ RNA (CEM cells), 2.2×10^6 cpm per filter, 81-hr exposure. (*Center*) ^{32}P-Labeled repetitive DNA, 1×10^6 cpm per filter, 36-hr exposure. (*Right*) Phage λ [^{32}P]DNA, 1×10^6 cpm per filter, 34-hr exposure.

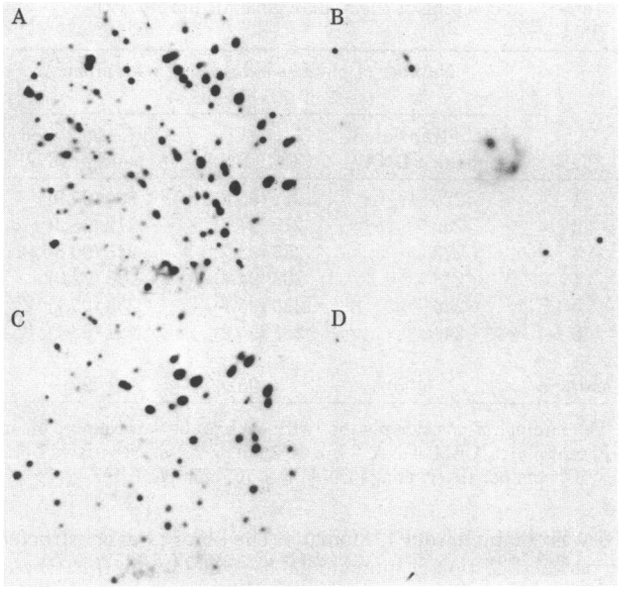

FIG. 2. Effect of unlabeled repetitive DNA competitor on hybridization of cDNA to the human genomic DNA library. The library was screened with [^{32}P]cDNA transcripts of cytoplasmic poly(A)$^+$ RNA of CCRF-CEM cells in the presence and absence of unlabeled human repetitive DNA (C_0t <40). Duplicate filters were prepared from two plates (*A* and *B*, *C* and *D*). One filter of each pair (*A* and *C*) was used in hybridization with [^{32}P]cDNA at 10^6 cpm/ml; the other filter (*B* and *D*) was hybridized with [^{32}P]cDNA at 10^6 cpm/ml and unlabeled repetitive DNA at 10 μg/ml.

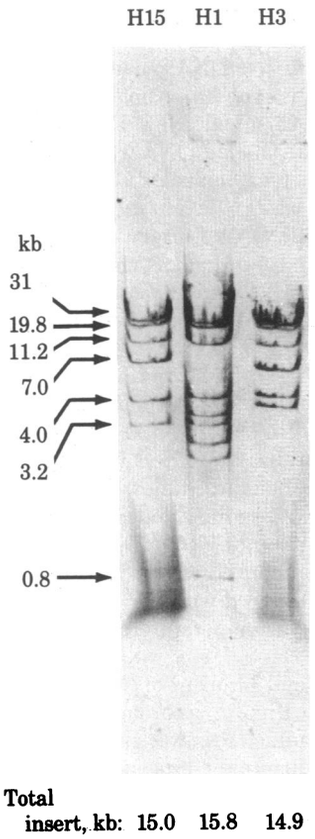

FIG. 3. Electrophoretic pattern of fragments of cloned human genomic DNA produced by *Eco*RI digestion. DNA was prepared from three recombinant phage clones, H1, H3, and H15. The digested DNA was electrophoresed on 0.8% agarose gels. After staining with ethidium bromide, the DNA bands were visualized by using ultraviolet illumination. The sizes of the fragments were determined by using Charon 4A DNA digested with *Eco*RI and phage ϕX174 DNA digested with *Hae* III as markers.

fective competitor, with partial competition observed with poly(rA), human ribosomal RNA, and calf thymus DNA. Note that poly(rA) was included at 10 μg/ml in all nucleic acid hybridization experiments reported in this paper except those in Table 2. These experiments suggest the presence in the human genome of a widely dispersed repetitive sequence that is represented in cDNA transcripts of polysomal mRNA.

Presence of the Ubiquitous Repeat in Selected Genomic Clones. Three clones were selected from the genomic library, two of which (H1 and H15) hybridize with both labeled cDNA and repetitive DNA. The size of DNA fragments resulting from digestion with *Eco*RI shows that each recombinant contains about 15 kb of human DNA (Fig. 3). Blot hybridization of the three *Eco*RI-digested clones shows that the bands in clones H15 and H1 between 7 and 3 kb contain sequences represented in repetitive DNA (Fig. 4 *A* and *C*). The bands also hybridize with [^{32}P]cDNA (Fig. 4*B*), suggesting the presence of a sequence family similar to that observed in the plaque screening of the genomic library. To examine this possibility, the DNA bands of clone H15 were excised and the DNA was labeled by nick translation. Both the 4.0- and 3.2-kb fragments hybridized to >90% of the genomic library (Fig. 5). The intensities of the positive clones differed somewhat, due to differences in specific activities of the hybridized probes. These results indicate that the ubiquitous repeat sequence present in almost all clones in the genomic library is present in at least two copies in clones H1 and H15. In contrast, clone H3 does not appear to contain repetitive sequences and did not hybridize with labeled cDNA.

Table 2. Effect of competitor nucleic acid on hybridization of [^{32}P]cDNA to human genomic DNA library

Competitor	% hybridization
—	100
Poly(rA), 100 μg	80
Human ribosomal RNA, 10 μg	87
Calf thymus DNA, 10 μg	93
Human repetitive DNA ($C_0t \leq 40$), 10 μg	1.5

Duplicate filters of the library were prepared. Filters were hybridized with the probe in the presence or absence of competitor nucleic acid; 10^6 cpm per filter of [^{32}P]cDNA was added to each fraction. % hybridization = 100 × (number of positive clones with competitor/number of positive clones without competitor).

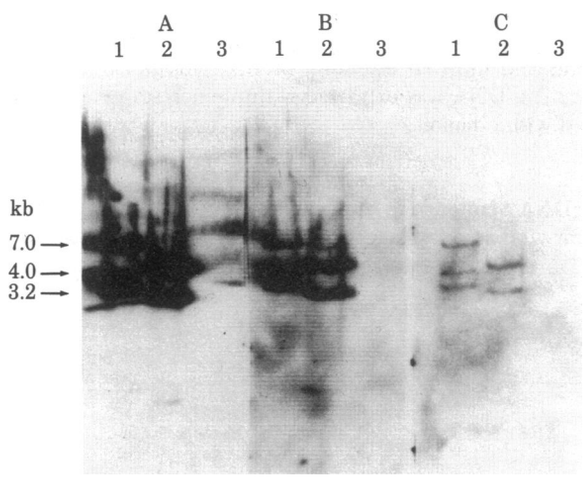

FIG. 4. Hybridization of cDNA and repetitive DNA to cloned DNA digested with *Eco*RI. Three recombinant phage DNAs were digested, electrophoresed in vertical 0.8% agarose gels, transferred to nitrocellulose filters, and hybridized. The electrophoretic pattern of each DNA is shown in Fig. 3. Gels *A*, *B*, and *C* show, respectively, the results of hybridization of ^{32}P-labeled repetitive DNA (10^6 cpm/ml), [^{32}P]cDNA (10^6 cpm/ml), and [^{32}P]cDNA (10^6 cpm/ml) with unlabeled repetitive DNA competitor (5 μg/ml). Cloned human DNAs in each gel were: lanes 1, H15; lanes 2, H1; lanes 3, H3.

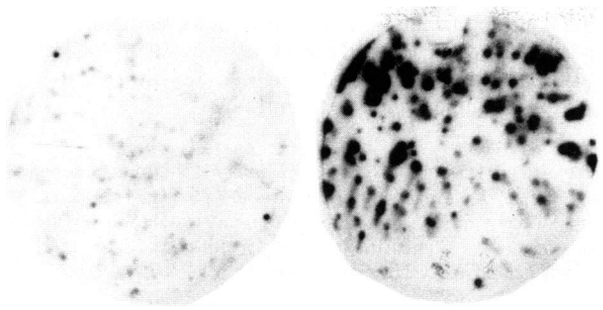

FIG. 5. Hybridization of *Eco*RI-digested clone H15 DNA with the human genomic DNA library. The 4.0- and 3.2-kb fragments of clone H15 (Fig. 3) were extracted from the agarose gel after electrophoresis, labeled by nick translation, and used in screening the genomic DNA library. (*Left*) The 4.0-kb fragment, 3×10^5 cpm, 6-day exposure. (*Right*) The 3.2-kb fragment, 6×10^5 cpm, 1-day exposure.

H3 was localized to band p36 of chromosome 1 by *in situ* hybridization experiments (19).

Sequence Relationship with the Alu Family Clone. The preceding experiments suggest the presence of a dispersed repeated sequence in human DNA that is represented in cytoplasmic poly(A)$^+$ RNA. Proof that the hybridizing species is a specific sequence or single family of sequences could be obtained by cloning an appropriate repeat sequence and examining its sequence homology with our genomic clones.

Houck *et al.* (6) have identified a repeated DNA family containing a cleavage site for the restriction endonuclease *Alu* I. This 300-bp repeat was cloned and the nucleotide sequence of some representatives of the family has been determined (7). Two members of the family, clones BLUR 2 and 8 (for *Bam*-linked ubiquitous repeat) kindly provided by P. Deininger and

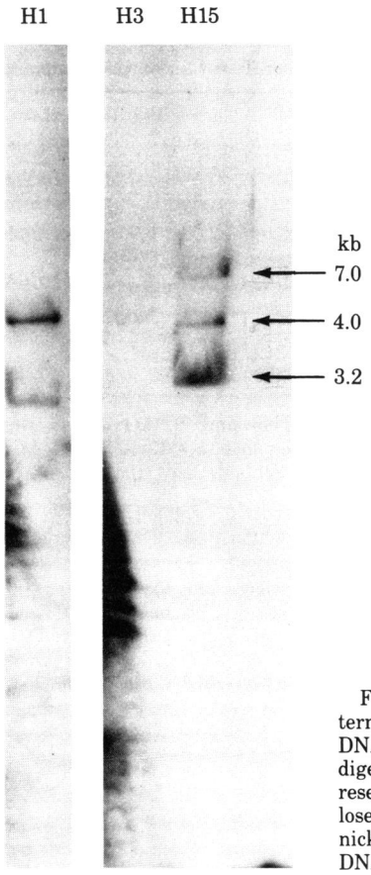

FIG. 6. Hybridization pattern of cloned human genomic DNA. Cloned human DNA was digested with *Eco*RI, electrophoresed, transferred to nitrocellulose filters, and hybridized with nick-translated BLUR 8 plasmid DNA.

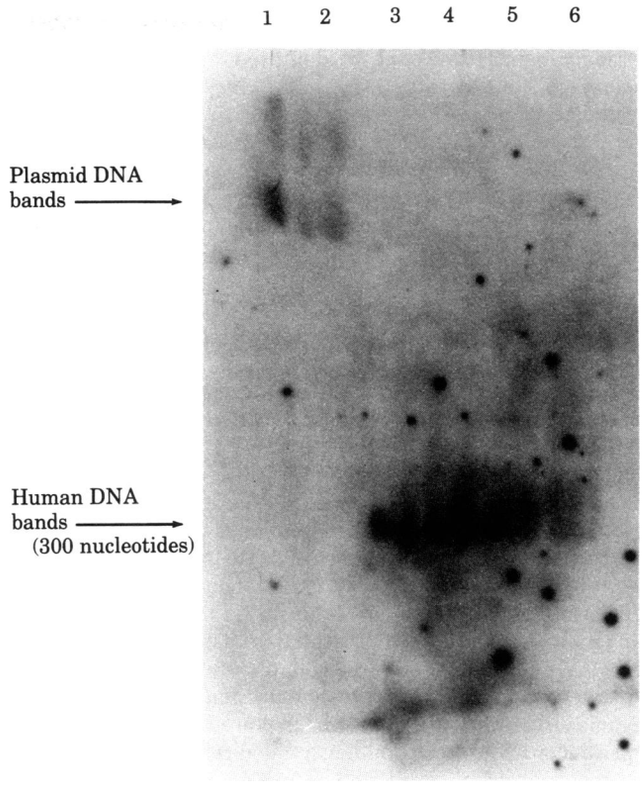

FIG. 7. Hybridization of the Alu repeat in BLUR 2 plasmid DNA with [^{32}P]cDNA. Plasmid DNA was digested with *Bam*HI, electrophoresed in a 1.5% agarose gel, transferred to a nitrocellulose filter, and hybridized with [^{32}P]cDNA (10^6 cpm/ml) transcribed from polysomal poly(A)$^+$ RNA CCRF-CEM cells. The size of the human insert DNA was determined by using pBR322 plasmid DNA digested with *Hinf* I as marker DNA. The amounts and *Bam*HI treatments of plasmid DNA were: lanes 1 and 2, 1.5 μg, undigested; lane 3, 7 μg, digested; lanes 4 and 5, 5 μg, digested; lane 6, 3 μg, digested.

C. Schmid, were used in our studies. Screening of the genomic DNA library with radiolabeled BLUR 2 or 8 resulted in >90% positive clones. Blot hybridization of ^{32}P-labeled BLUR 8 with *Eco*RI-digested genomic clones (Fig. 6) showed that H1 contains at least two Alu repeats and H15 contains at least three. Similar experiments using *Bam*HI-digested genomic clones showed six hybridization bands in the H15 digest and two in the clone H1 digest.

To confirm that the Alu family repeat is represented in polysomal mRNA, we digested clone BLUR 2 to release the human DNA insert and hybridized the plasmid digest with [^{32}P]cDNA. Fig. 7 shows that the 300-bp human sequence hybridized with cDNA and, as expected, the extent of hybridization was dependent on the plasmid DNA concentration. Furthermore, clone BLUR 2 competed with [^{32}P]cDNA in screening the genomic DNA library (data not shown). These experiments show that the Alu family repeat is present in almost all clones in the human genomic DNA library and that the Alu repeat is represented in cytoplasmic polysomal mRNA.

DISCUSSION

We have demonstrated the presence of a widely dispersed repetitive sequence in the human genome that is represented in cytoplasmic polysomal poly(A)$^+$ RNA. This sequence is present in approximately 95% of the clones in a genomic DNA library and is homologous to the Alu repeated sequence family. Clones were selected that contain two copies (H1) and at least six copies (H15) of the Alu repeat. One clone (H3) with a 15-kb insert appears devoid of repetitive DNA.

The amount of the Alu repeat in the human genome is surprisingly high. Houck *et al.* (6) identified the family containing a cleavage site for restriction enzyme *Alu* I in the 300-bp repeated sequence. They estimated that there are more than 300,000 copies per genome, making up at least 3% of the DNA, and our results support this estimate. If 95% of the clones in the genomic library contain at least one Alu repeat in their ≈18 kb of human DNA, then the minimum number of copies per human genome = (7×10^9 bp/genome ÷ 1.8×10^4 bp/clone)(0.95) = 370,000 copies per genome, and (300 bp/repeat ÷ 1.8×10^4 bp/clone)(100%) = 1.7% of the genome. Clearly these are underestimates because many clones contain several copies of the Alu repeat. One clone, H15, contains at least six copies, and the human β-globin-like cluster contains seven copies of a repeated sequence similar to the *Alu* family in a 65-kb region (20).

The distribution and in some cases clustering of Alu repeats suggest an important biological role for these sequences. Houck *et al.* (6) found that approximately one-third of the interspersed Alu repeats exist as inverted repeats (IRs) and the remaining two-thirds may exist as IRs with much longer spacer sequences separating the two halves of the inverted repeat. The human Alu family is extensively represented in the heterogeneous nuclear RNA (hnRNA) of cultured cells (21). Because hnRNA, the putative precursor of mRNA, contains long double-stranded regions of hairpin-type structures that are absent from the cytoplasmic mRNA, these double-stranded regions were thought to have some regulatory function in RNA processing (22, 23). The hairpins presumbly arise by transcription of IR sequences in DNA.

Evidence is accumulating for the involvement of IR sequences in genetic translocation. Structurally defined DNA segments carrying antibiotic-resistance genes, transposons, have been shown to be capable of translocation from plasmid to plasmid or from plasmid to the host chromosome. For example, the ampicillin-resistance DNA segment can translocate from pSC50 plasmid onto pSC101 plasmid, resulting in a recombinant plasmid (24). The translocation does not involve the conventional *recA* gene function, and the excision is terminus-site-specific. The transposable segment contains a 130-nucleotide-long IR sequence at both termini. It has been proposed that the IR helps the donor molecule to orient in configurations possessing twofold rotational symmetry that can be recognized by enzymes involved in translocation (24). This role may explain one evolutionary advantage of IR sequences.

Several other potential functions for the Alu family IR were discussed by Jelinek *et al.* (25), including (*i*) RNA polymerase III transcription initiation sites [Alu family repeats found in clones of human β-globin and β-globin-like genes may be specifically transcribed *in vitro* by RNA polymerase III into discrete-sized products (26)] and (*ii*) origins of DNA replication (several papovavirus genomes contain a 14-nucleotide sequence having high homology with a sequence found in the Alu family). This sequence constitutes a portion of a perfect IR located at or near the origin of replication in these viral DNAs (27–29).

The presence of the Alu repeat in cDNA transcripts of polysomal poly(A)$^+$ RNA is surprising, although similar results were found in mouse systems (30). Leakage of nuclear RNA during preparation of cytoplasmic RNA is always a formal possibility. We think our "polysomal" cDNA reflects Alu repeats in the mRNA because the "polysomal" cDNA gave a stronger signal with the Alu clone than did the "cytoplasmic" cDNA. It is important to determine if those messengers containing the Alu repeat represent a specific subset of the mRNA population. It is also important to establish whether the Alu repeat is contained in a coding region of the mRNA.

We are very grateful to Dr. T. Maniatis for the human genomic library, Dr. P. Deininger for the Alu repeat plasmids, and Dr. M. B. Mann for many useful suggestions. This investigation was supported by grants from the National Institutes of Health (GM 23965, CA 20214, CA 16672, and CA 21927) and the Robert A. Welch Foundation (G-267).

1. Saunders, G. F., Shirakawa, S., Saunders, P. P., Arrighi, F. E. & Hsu, T. C. (1972) *J. Mol. Biol.* **63**, 323–334.
2. Dott, P. J., Chuang, C. R. & Saunders, G. F. (1976) *Biochemistry* **15**, 4120–4125.
3. Schmid, C. W. & Deininger, P. L. (1975) *Cell* **6**, 345–358.
4. Davidson, E. H., Galau, G. A., Angerer, R. C. & Britten, R. J. (1975) *Chromosoma* **51**, 253–259.
5. Davidson, E. H., Hough, B. R., Klein, W. H. & Britten, R. J. (1975) *Cell* **4**, 217–238.
6. Houck, C. M., Rinehart, F. P. & Schmid, C. W. (1979) *J. Mol. Biol.* **132**, 289–306.
7. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedman, T. & Schmid, C. W. (1980) *Nature (London)* **284**, 372–374.
8. Kaplan, J., Shope, T. C. & Peterson, W. D. (1974) *J. Exp. Med.* **139**, 1070–1076.
9. Chen, T. R. (1975) *Tissue Culture Assoc. Man.* **1**, 229–232.
10. Kuo, M. T., Sahasrabuddhe, C. G. & Saunders, G. F. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1572–1575.
11. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
12. Woosley, J. T. & Sahasrabuddhe, C. G. (1978) *Electrophoresis 1978*, ed. Catsimpoolas, N. (Elsevier/North Holland, New York), pp. 337–342.
13. Frazier, M. L., Montagna, R. A. & Saunders, G. F. (1981) *Biochemistry* **20**, 367–371.
14. Britten, R. J., Graham, D. E. & Neufeld, B. R. (1974) *Methods Enzymol.* **29**, 363–418.
15. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
16. Lawn, R. M., Fritsch, E. H., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15**, 1157–1174.
17. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
18. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
19. Harper, M. E. & Saunders, G. F. (1981) *Chromosoma*, in press.
20. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1980) *Cell* **19**, 959–972.
21. Robertson, H. D., Dickson, E. & Jelinek, W. (1977) *J. Mol. Biol.* **115**, 571–589.
22. Jelinek, W. R. & Darnell, J. E. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2537–2541.
23. Ryskov, A. P., Saunders, G. F., Farashyan, V. R. & Georgiev, G. P. (1973) *Biochim. Biophys. Acta* **312**, 152–164.
24. Kopecko, D. J. & Cohen, S. N. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1373–1377.
25. Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. & Schmid, C. W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1398–1402.
26. Duncan, C., Biro, P. A., Choudary, P. V., Elder, J. T., Wang, R. C., Forget, B. G., De Riel, J. K. & Weissman, S. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5096–5099.
27. Yang, R. C. A. & Wu, R. (1979) *Science* **206**, 456–462.
28. Reddy, V., Thimmappaya, B., Dhar, R., Subramanian, K., Zain, S., Pan, J., Celma, M. & Weissman, S. M. (1978) *Science* **200**, 494–502.
29. Friedmann, T., LaPorte, P. & Esty, A. (1978) *J. Biol. Chem.* **253**, 6561–6567.
30. Tokarskaya, O. N., Tchurikov, N. A., Ivanov, P. L., Kramerov, D. A. & Ryskov, A. P. (1980) *Nucleic Acids Res.* **8**, 425–440.