Volume 13 Number 15 1985 Nucleic Acids Research

Interspersed repetitive and tandemly repetitive sequences are differentially represented in extrachromosomal covalently closed circular DNA of human diploid fibroblasts

Karl Riabowol, Robert J.Shmookler Reis and Samuel Goldstein

Departments of Medicine and Biochemistry, University of Arkansas for Medical Sciences and Geriatric Research Education and Clinical Center, Veterans Administration Medical Center, Little Rock, AR 72205, USA

Received 18 April 1985; Revised and Accepted 1 July 1985

ABSTRACT

Extrachromosomal covalently closed circular DNA (cccDNA) was isolated from human diploid fibroblasts by alkaline denaturation/renaturation and CsCl-ethidium bromide isopycnic centrifugation. Probing across these gradient fractions showed a higher proportion of cccDNA sequences homologous to the interspersed highly repetitive Alu I and Kpn I sequences than to the human tandemly-repetitive Eco RI (alphoid) DNA. Cloning of these cccDNAs was then carried out following digestion with restriction endonucleases Hind III, Bam HI or Pst I, and ligation into plasmid pBR322. Many isolated recombinant clones were unstable as seen by a high rate of loss over four cycles of antibiotic selection, and frequent plasmid modifications including deletions adjoining the site of insertion. Of 107 cloned sequences which appeared relatively stable, i.e., survived four cycles of antibiotic selection without incurring detectable deletions, 28% and 11% showed homology to Alu I and Kpn I families, respectively, while 4% contained sequences homologous to both. In contrast, less than one percent hybridized to probes for tandemly-repetitive sequences, Eco RI and Satellite III. The average insert size of cloned cccDNA derived from human fibroblasts, 2.52 Kbp, was larger than previously reported for similar clones derived from genetically less stable permanent lines, which may reflect differences in the process of cccDNA generation.

INTRODUCTION

Closed circular DNAs have been found to exist extrachromosomally in a variety of eukaryotic cell types including Drosophila (1-3), chicken bursa (4), CHO (Chinese hamster) (5), BSC-1 (African green monkey) (6,7), and a number of cell and tissue types of human origin (8-10). Electron microscopy studies have shown that these elements are quite polydisperse, ranging in size from 150 base pairs (bp) to over 9 kilobase pairs (Kbp), (4,10,11). Both the size distribution and number of these elements per cell (~$10-10^4$) vary with development, growth state and mitotic capacity (2,4,6,10). In particular, the average size of these small polydisperse circular DNAs (spcDNAs) appears to be less in established immortal cell lines than in diploid cells with a finite replicative capacity (10). Other characteristics of spcDNA include a buoyant density similar to total chromosomal DNA, a composition

involving more than one complexity class (3,6,12) and homology to chromosomal DNA.

Due to their low concentration and heterogeneous nature, spcDNAs have proven difficult to study in terms of structure, sequence, mechanism of generation and possible cellular role(s). Recent cloning of representative spcDNAs from the immortal cell lines BSC-1 (12), CHO (5) and HeLa (13), however, has shed some light on these questions. These spcDNAs contain both unique and repetitive chromosomal sequences which include Alu I interspersed repeats and the α-satellite (BSC-1 cells), and a rodent Eco RI repeat (CHO cells) (5,6). Certain cloned spcDNAs which contain Alu I elements lack short direct repeats which often flank their Alu I counterparts in the chromosome (12), including those found in tracts of α-satellite sequences (14). Recently it has been demonstrated that discrete mitochondrial DNA (mtDNA) sequences of Podospora anserina are excised and amplified extrachromosomally during senescence of this fungus (15,16), but spcDNAs studied in other organisms are clearly nonmitochondrial (5,6,12,13). It was thus of considerable interest to molecularly clone a representative cross-section of circular DNAs from diploid cells (normal human fibroblasts) in order to characterize them as individual DNA species. Because we have isolated and cloned extrachromosomal DNA molecules derived from these cells on the basis of their covalently closed circular nature and find that these DNAs are not necessarily small, we will use the term cccDNA in preference to spcDNA. This report describes the cloning and preliminary characterization of cccDNAs from a normal human fibroblast strain near the end of its limited replicative lifespan.

MATERIALS AND METHODS

Cells and Cell Culture

Fibroblast strain A2 was derived from an 11-year-old normal male donor and passaged until it approached its maximal in vitro replicative capacity of about 63 mean population doublings (MPD) (17). Cells were grown without antibiotics in 850 cm^2 plastic roller bottles in Eagle's minimal essential medium supplemented with 15% fetal calf serum. Cultures were consistently negative for the presence of mycoplasma (18).

Isolation of mtDNA

Exponentially growing K562 cells (~$4x10^8$) of an established human erythroleukemia line (19) were harvested by pelleting at 1500x g and were washed twice in phosphate-buffered saline minus glucose (PBS-Glu). Cells were swollen in hypotonic buffer (10 mM Tris, 10 mM NaCl, 1.5 mM $MgCl_2$, pH 7.4) 10 min on

ice and lysed with 10 passes of a tight ("A") pestle in a Dounce homogenizer. mtDNA was then isolated from the lysate as described previously (20).

Preparation of cccDNA

Fibroblasts (~1x10^8) grown to confluence at 63 MPD in 40 roller bottles were washed twice in PBS-Glu and detached by incubation with 0.05% trypsin (Difco) and 10 mM EDTA for 10 min at 37°C. Ice-cold growth medium was added to the cells which were then pelleted at 1500x g for 5 min at 4°C. The pellet was washed three times with PBS-Glu and thoroughly resuspended in 28 ml of 50 mM Tris, 10 mm EDTA, 50 mM glucose, pH 8.0, at a concentration of 3.5x10^6 cells/ml. Two vol of freshly made 0.2 N NaOH/1 percent SDS were added to lyse the cells. Following 5 min on ice, 1.5 vol of 3 M potassium acetate (pH 4.8) was added and the mixture placed on ice for 30 min. The cell lysate was centrifuged for 1 hr at 12000x g and the supernatant filtered through cheesecloth and precipitated with 0.6 vol of 99% isopropanol at -20°C overnight (21). The precipitated nucleic acids were then pelleted for 1 hr at 12000x g, 4°C, resuspended in 0.25 M NaOAc and extracted twice in buffer-saturated phenol (22) and once in $CHCl_3$:octanol (24:1). The aqueous phase was precipitated with 2 vol of absolute ethanol and the pellet collected and resuspended in TE buffer (10 mM Tris, 1 mM EDTA), pH 8.0. This mixture was centrifuged to equilibrium on a CsCl-ethidium bromide (EthBr) gradient (24 hr at 42,000 rpm in an SW 50.1 rotor) and the band corresponding to cccDNA isolated, diluted with CsCl-EthBr and rebanded as before. The band of supercoiled DNA, identified visually by comparison to Form I plasmid molecules run in parallel, was collected by drip fractionation (22) and confirmed to be of the correct density (1.59-1.62 gm/ml) by refractive index. Following four extractions with 5 M NaCl-saturated n-butanol, 2 vol of distilled H_2O and 6 vol of 100% EtOH were added to precipitate the cccDNAs at -20°C overnight. The pellet was collected following centrifugation, redissolved in TE, and subjected to 2 more cycles of EtOH precipitation, desiccated and resuspended in 100 µl of TE buffer on ice.

Electron Microscopy

Extrachromosomal cccDNA samples, prepared as outlined above, were relaxed in the presence of 50 µg/ml EthBr by illumination with ultraviolet light (254 nm) for one hour. Spreading of DNA on parlodion films and shadowing with platinum was then done as described (23), and molecules were visualized and photographed using a Siemens 101 electron microscope. Contour lengths of molecules were determined with an Apple IIe graphics tablet using plasmid pBR322, bacteriophage PM2 and endogenous mtDNA as size standards.

Cloning of cccDNA

Separate aliquots of total cccDNA were digested to completion with Bam HI, Hind III or Pst I using conditions recommended by Bethesda Research Labs, and heated to 65°C for 30 min. Appropriately linearized pBR322 was treated with bacterial alkaline phosphatase and then added to give an approximate ratio of 10 vector molecules per cccDNA molecule (to minimize multiple insertions) and ligated using 0.5 U T4-DNA ligase per 50 µl reaction mixture for 24 hr at 12°C. Optimal vector concentration was confirmed in parallel ligations of untreated linearized pBR322 at similar concentrations which converted approximately 50% of the material to monomeric circular species as judged by EthBr fluorescence of DNA electrophoresed on 1.0% agarose gels. Ligation mixtures were used to transform $CaCl_2$-treated competent *E. coli* HB101 as described (22). To verify its identity, host HB101 was tested and confirmed to be Rec A^-, lac^-, gal^-, ara^-, pro^- and Sm^r. Transformed colonies selected for ampicillin or tetracycline resistance were screened through 3 additional cycles of double antibiotic selection, and the clones retaining predicted antibiotic resistance were expanded and their plasmids isolated using the procedure of Birnboim and Doly (21).

Southern Blotting and Hybridization

Recombinant clone DNAs were restricted with a fivefold excess of the enzyme used for cloning, treated with RNAses A and T1, and electrophoresed in 1.5% agarose gels at 2 V/cm. Digests were visualized with EthBr and transferred to nitrocellulose filters (0.45 µ, BA-85, Schleicher and Schuell) using the method of Southern (24). Hybridization was carried out in 50% formamide, 5x SSC, 1x Denhardt's reagent, 0.5 M $NaPO_4$, pH 6.5, 0.1 mg/ml denatured herring sperm DNA, 25 µg/ml yeast tRNA, poly rC and poly rA, and 0.1 gm/ml dextran sulfate, at 42°C for approximately 20 hr. DNA probes were labeled with ^{32}P-dNTPs by nick translation (25) or by extension of random hexanucleotide primers (26) to a specific activity of $1x10^7$ to $2x10^8$ cpm/µg. Filters were washed in 0.1x SSC at 55°C for all probes except BLUR 8, which first were washed in 0.5x SSC at 55°C, exposed, rewashed in 0.1x SSC and re-exposed. Filters were autoradiographed using Kodak X-OMAT film on Cronex Lightning Plus intensifying screens at -70°C for 2 days.

Dot Blotting

To determine which clones contained repetitive genetic elements other than those specific DNA sequences probed, 0.2 µg of recombinant cccDNA-plasmid DNA was applied to nitrocellulose filters (Schleicher and Schuell) following the manufacturer's instructions using a microsample filtration manifold. Fil-

ters were hybridized with total human ^{32}P-DNA labeled by nick translation as described above and washed sequentially in 1x, 0.5x and 0.1x SSC at 55°C and compared with results obtained by probing with defined repetitive sequences.

DNA Probes

All clones obtained were probed with: 1) mtDNA isolated as above, b) the interspersed repetitive Alu I insert from BLUR 8 (27), c) a tetrameric insert (Shmookler Reis, et al., J. Mol. Biol., in press) of the human Eco RI (alphoid) tandemly repeated DNA family (28), d) BLUR 18 human tandemly repeated satellite III insert (27), e) BLUR 16, a Kpn I-like interspersed repetitive insert (27), f) Kpn I, a mixture of four inserts spanning a 6.4 kb interspersed repetitive sequence found in the ß-globin cluster (29), g) the "Inter Alu" insert, a 0.8 Kbp sequence of low copy human DNA nested within a cluster of Alu repeats (30), h) a mink retroviral LTR (31), i) the v-myc oncogene (32), j) total fibroblast DNA, and k) total cccDNA isolated as described above.

RESULTS

Isolation of cccDNA

Figure 1 shows an electron micrograph of a typical field of total fibroblast cccDNA prepared for microscopy after two rounds of CsCl-EthBr equilibrium density gradient centrifugation. The majority of molecules consisted of one size class corresponding to mtDNA; however, when multiple fields were examined (insets), other molecules of a broad size distribution were also observed.

In order to characterize repetitive, nonmitochondrial cccDNA molecules, fractions were collected from a CsCl-EthBr gradient. Refractive indices were read to determine density and the remaining material was dot blotted to nitrocellulose and probed with ^{32}P-labeled Kpn I, Eco RI and Alu I cloned repetitive sequences. As can be seen in Figure 2, cccDNA molecules containing sequences homologous to the Kpn I interspersed repetitive element were indeed found at the density corresponding to cccDNA molecules. Probing the same filter with the centromeric, tandemly repeated Eco RI sequence gave a substantially reduced fraction of the signal at this density, whereas rehybridization with a member of the Alu I interspersed repeat family as probe yielded a pattern very similar to that obtained with Kpn I (Figure 2). Separate filters of dot-blotted CsCl-EthBR gradients showed similar ratios of cccDNA to linear plus nicked DNA when probed in a different order and upon subsequent reprobing with the initial sequence used as probe. Signal was determined to be within

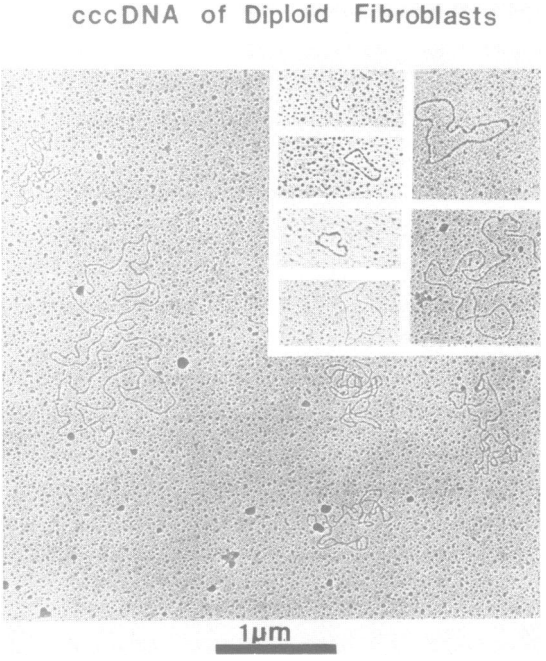


Figure 1: Electron micrographs of cccDNAs prepared from human diploid fibroblasts. Fractions of density 1.60-1.62 gm/ml from the second round of CsCl-EthBr centrifugation were pooled, precipitated, rehydrated and spread on parlodion films as described in Methods. The randomly chosen main field shows predominantly circular mtDNA (16.5 Kbp, as in lower right inset), while selected fields in the inset show cccDNAs ranging in size from 0.75 Kbp (upper left) to 10.1 Kbp (upper right). Contour length was determined as described in Methods. 1 μm = 3.1 Kbp.

the linear range of film response and independent experiments confirmed that the relative Kpn I:Eco RI:Alu I hybridization intensity in the "supercoiled DNA" portion of the gradient was 2.2 (± 0.6):1:3.7 (± 0.8) (mean ± half range), respectively, in cccDNA derived from late-passage fibroblasts. All DNA sequences used to probe this nitrocellulose filter were labeled to the same specific activity and hybridization conditions were identical.

Cloning of Fibroblast cccDNA

The heterogeneity in size and sequence of cccDNA precludes nonselective cloning of full length circular molecules by conventional procedures. Tetranucleotide-recognizing restriction enzymes, while cleaving most molecules, would generate multiple cuts in most larger cccDNAs. A cloning procedure using EthBr to limit such cleavage to a single cut per supercoiled molecule

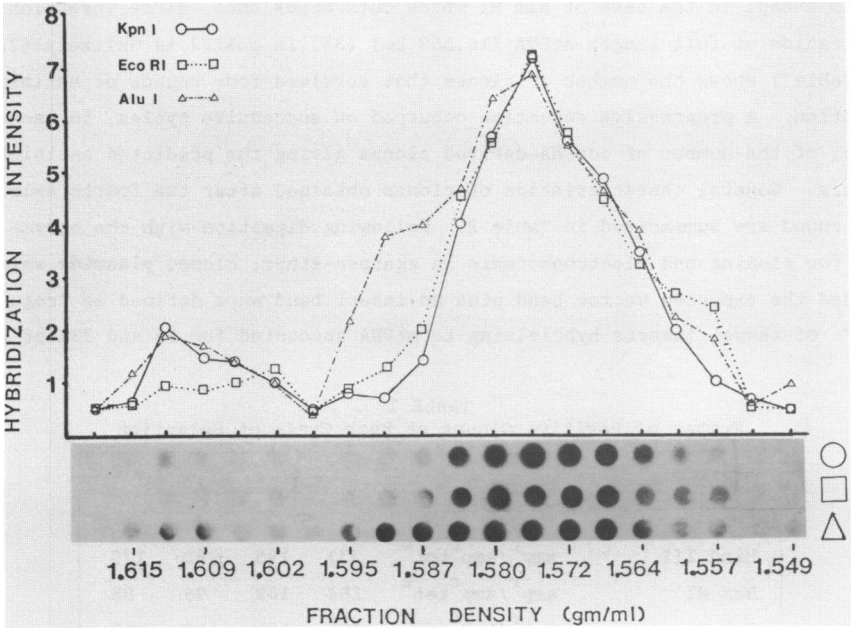


Figure 2. Hybridization of cloned interspersed repetitive (Kpn I, Alu I) and tandemly repetitive (Eco RI) DNA to CsCl-EthBr gradient fractions. Approximately 8x10^7 fibroblast cells were subjected to alkaline denaturation-renaturation (21) and the supernatant enriched for cccDNA was centrifuged once to equilibrium in the presence of 200 µg/ml ethidium bromide. Refractive indices of gradient fractions were read to determine density and 50 µl aliquots were dot blotted to a nitrocellulose filter also containing serial dilutions of total chromosomal DNA. This filter was hybridized sequentially with ^{32}P-labeled cloned Kpn I, Eco RI and Alu I repetitive sequences. Autoradiographs obtained were scanned with a Cliniscan densitometer and the areas under each peak determined using an Apple™ graphics tablet and analysis package. Values thus obtained were normalized to fractions from the major peak, (ρ = 1.572-1.580) representing linear and relaxed DNA for all three probes such that hybridization signal from cccDNAs (ρ = 1.61) is relative to signal from nonsupercoiled molecules for each probe used. Hybridization intensities are in arbitrary units.

(33) was abandoned due to poor yield. We therefore employed hexanucleotide-recognizing enzymes which should cut random sequence DNA roughly once per 4 Kbp, thus favoring large inserts, and used three different enzymes separately in order to include molecules which lack a site for one or two enzymes. Reduction of the mtDNA component, by first preparing nuclei and by preparative agarose fractionation, was initially attempted but later omitted due to substantial loss of nonmitochondrial cccDNA molecules (9). Thus, it was anticipated that a large fraction of the clones obtained would be of mitochondrial

origin except in the case of Bam HI which cuts mtDNA once, since insertion and replication of full length mtDNA (16,569 bp) (34) in pBR322 is unlikely (22).

Table 1 shows the number of clones that survived four rounds of antibiotic selection. A progressive reduction occurred on successive cycles, in each group, of the number of cccDNA-derived clones giving the predicted antibiotic signals. General characteristics of clones obtained after the fourth selection round are summarized in Table 2. Following digestion with the enzyme used for cloning and electrophoresis in agarose-EthBr, cloned plasmids which yielded the expected vector band plus an insert band were defined as "regular." Of these, inserts hybridizing to mtDNA accounted for 55 and 33% of the

TABLE 1
Number of Positive Clones at Each Cycle of Selection

Cloning Enzyme	Resistance	Selection Round[a]			
		1	2	3	4
Hind III	$amp^r/amp^r tet^s$	333	195	180	179
Bam HI	$amp^r/amp^r tet^s$	184	109	96	88
Pst I	$tet^r/tet^r amp^s$	258	181	130	122

[a]Selection in the first round was for resistance to a single antibiotic. For subsequent rounds, colonies were scored for resistance to one antibiotic and sensitivity to the other as indicated.

TABLE 2
Initial Characterization of Cloned cccDNAs[a]

			Clones with Regular Inserts			Irregular Clones		
Enzyme	A^r/T^s	T^r/A^s	MT[b]	Nuclear	Total	Modified[c]	Uncut[d]	Total
Hind III	179	---	64	53	117	58	4	62
Bam HI	88	---	0	25	25	45	18	63
Pst I	---	122	14	29	43	71	8	79

[a]Clones obtained after four rounds of antibiotic selection (Table 1).
[b]Mitochondrial clone inserts identified by Southern blot hybridization to pure ^{32}P-dNTP labeled mtDNA were the length predicted from the human mitochondrial restriction map (11) in every case.
[c]Modified recombinant plasmids were those which, when linearized, were larger or smaller than pBR322 but contained no discernible (>50 bp) inserts.
[d]Plasmids generally migrated faster than pBR322 standard but in some cases this may be due to increased supercoil density rather than deletion since these molecules could not be linearized with the enzyme used for cloning.

Hind III- and Pst I-generated inserts, respectively, but none of the Bam HI inserts, essentially as expected.

Surprisingly, among the recombinant clones which had lost the appropriate antibiotic resistance phenotype, there were also numerous "irregular" plasmids. Some appeared to have linearized vector DNA but no discernible insert, while other plasmids of varying size could not be cleaved with the restriction enzyme originally used for cloning. Three "irregular" recombinants which were smaller than pBR322, when analyzed by restriction with enzymes known to cut the vector adjacent to the cloning site, proved to have internal deletions. Further restriction mapping indicated that each deletion extended from the restriction site used for cloning, through part or all of the antibiotic resistance gene that was scored for inactivation (data not shown). In contrast, no attrition was seen for 50 clones of 1.9 kb Hind III fragments from human genomic DNA during three successive rounds of double antibiotic selection. Furthermore, no modification or failure to cut was seen on Hind III digestion of these same recombinant plasmids.

To test whether plasmid modification was due to restriction enzyme contaminants or exonuclease activity during bacterial alkaline phosphatase treatment, the following control experiment was run. Seventy-two clones obtained by transformation of E. coli HB101 with linearized pBR322 alone (with and without phosphatase treatment) were grown up and analyzed. None showed plasmids that were either modified or that could not be restricted with the enzyme initially used for linearization.

Figure 3 shows the size distribution of all nonmitochondrial inserts which resulted from cloning following linearization of cccDNAs with the three restriction enzymes. (Clones hybridizing to mtDNA probe were excluded from this and subsequent analyses.) For each of the three enzymes, a broad size distribution was apparent ranging from 0.1 to 8 kb, with no size class being particularly abundant. However, the average size (2.52 Kbp) of cloned cccDNA sequences was considerably larger than the mean, 0.65 Kbp, of clones obtained from a permanent hamster cell line, CHO (5), by comparable procedures (see Discussion).

Filter-immobilized DNA from all recombinant clones obtained was analyzed for the presence of repetitive DNA sequences. Figure 4 shows a representative autoradiograph of cccDNA clones generated using Bam HI and probed with ^{32}P-dNTP nick-translated total fibroblast DNA. The numbered dots represent clones that are homologous to sequences present at high copy number in total human DNA. Results from dot blots were corroborated by probing with total DNA

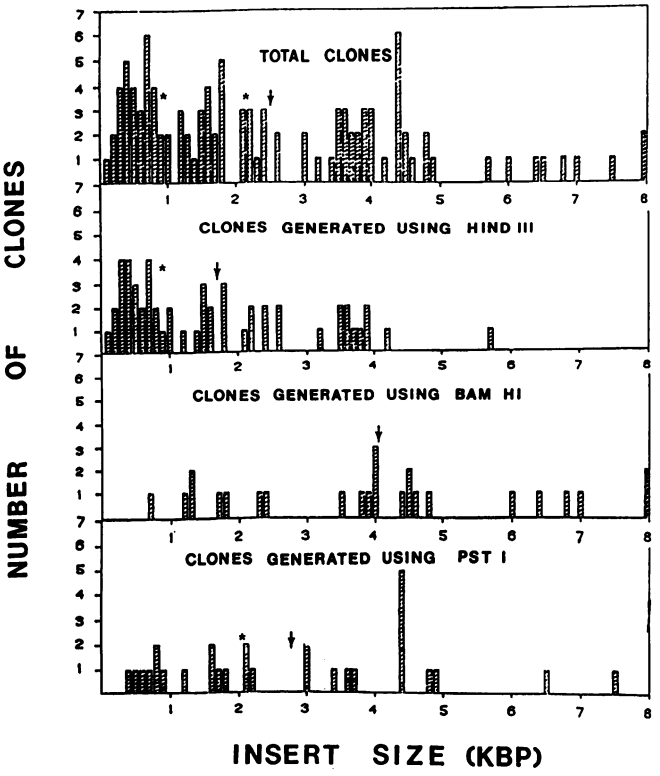


Figure 3. Size distribution of fibroblast cccDNA cloned inserts. All bacterial colonies with expected antibiotic resistance at the fourth round of selection (Table 1) were grown and their plasmids isolated, digested and visualized as in Figure 5. Insert size was determined by plotting log M.W. versus distance migrated using Hind III digested PM2 as standard. Vertical arrows indicate average M.W. in each panel and asterisks indicate size of mitochondrial DNA inserts which were isolated but not included in these graphs. Hind III digestion of mtDNA gives fragments of 0.9, 5.5 and 10.1 Kpb; Bam HI yields a 16.5 Kbp fragment, and Pst I generates 2.1 and 14.4 Kbp fragments (34).

against Southern blots of cccDNA clones, restricted with the enzyme used for cloning. For example, restriction fragments of several clones were resolved by 1.5% agarose gel electrophoresis (Figure 5A), transferred to nitrocellulose filters and probed with ^{32}P-labeled Alu I (Figure 5B) or Kpn I (Figure 5C) sequences. Four of the clones (arrow heads) included some homology to both repetitive element probes. Lanes marked Alu I and Kpn I contain nanogram amounts of the same DNA used for probing; they serve as hybridization standards and also demonstrate the absence of cross-hybridization between these repetitive elements under our conditions of stringency.

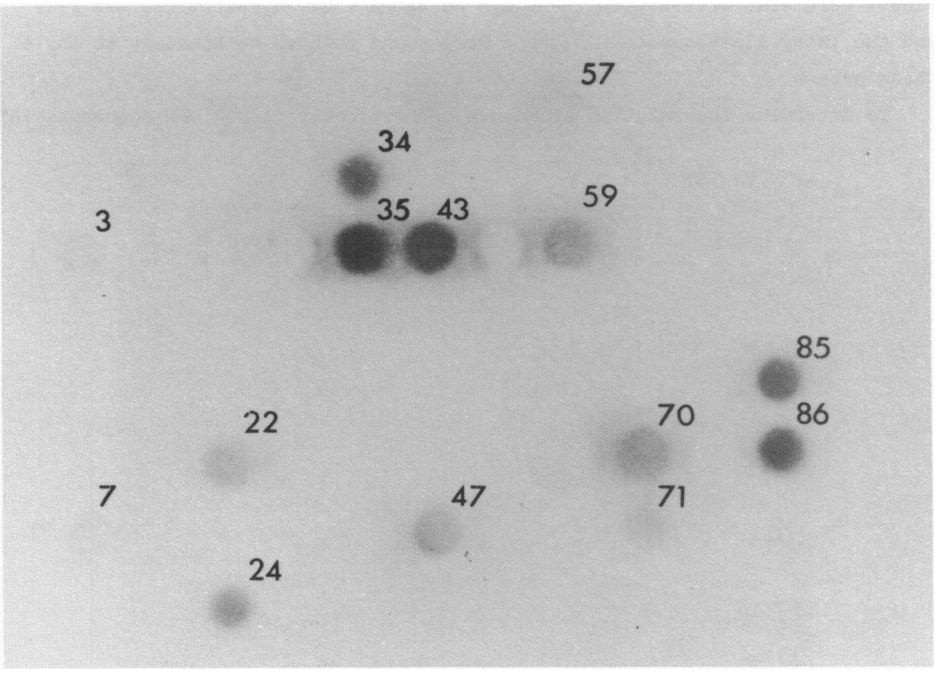
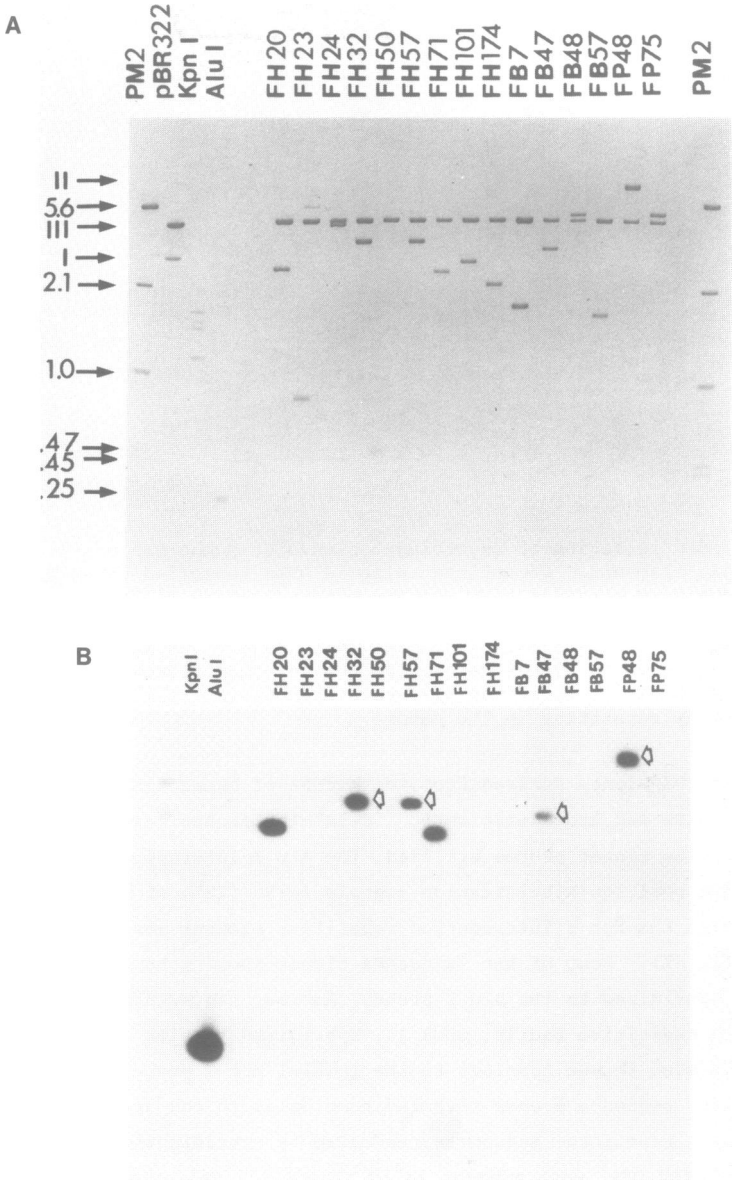


Figure 4. Identification of recombinant clones of human DNA containing repetitive genetic elements. Recombinant plasmid DNAs generated using Bam HI as linearizing enzyme were dot blotted onto nitrocellulose filters and probed with ^{32}P-dNTP nick-translated total human DNA. Colonies showing the strongest autoradiographic signals under these conditions are deemed to contain DNA homologous to highly repetitive DNA in human fibroblasts. Weaker signals (e.g., nos. 3, 7, 57 and 71) were taken to indicate sequences that are more moderately repetitive in the genome.

Data for 107 clones derived from the cccDNA of human fibroblasts, indicating insert size and hybridization to ^{32}P-labeled probes, are summarized in Table 3. Of the cloned probes utilized, the Alu I interspersed repetitive sequence gave positive hybridization signals in 30 (28%) of the 107 cloned inserts, while the Kpn I interspersed repetitive element was positive in 12 instances (11.2%). Four of the 12 cccDNA clones showing homology to the Kpn I probe also hybridized to the Alu I probe. Another representative of the Kpn I interspersed repetitive family, BLUR 16, hybridized to five of the 107 clones, all of which also showed homology to the initial Kpn I probe. Eco RI and BLUR 18 "satellite" sequence probes revealed homologies to one and zero clones, respectively. Repetitive sequences, revealed by hybridization to total genomic ^{32}P-dNTP labeled DNA, were present in 40 clones (37.4%). In addition, all

clones were also probed with the v-myc oncogene (32), a retroviral LTR (31) and the Inter Alu sequence (30), and were found to have no homology to any of these probes.

To determine the relative abundance of particular cloned DNA sequences in



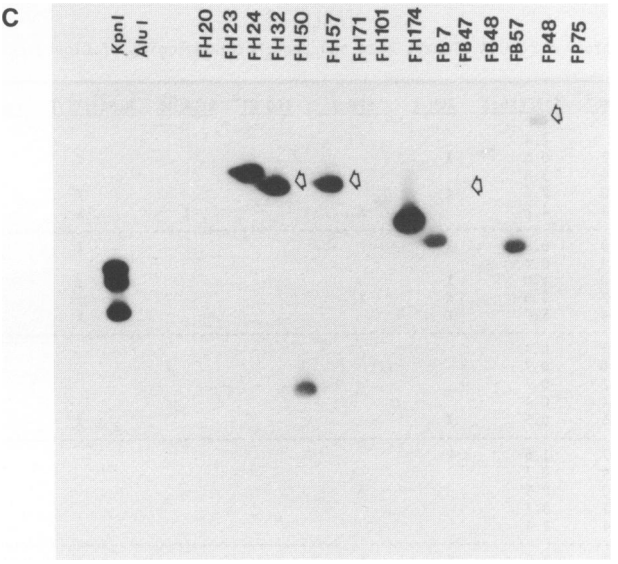


Figure 5. Hybridization of clones homologous to Alu I and/or Kpn I repetitive elements. Plasmid minipreps were prepared, digested with the restriction enzyme used for cloning and electrophoresed in a 1.2% agarose gel. (A) Ethidium bromide staining shows the presence of a single plasmid band and single insert band for each clone. DNA was then transferred from the gel to nitrocellulose and probed with (B) cloned Alu I insert and (C) cloned Kpn I inserts. Stringency of washes was 0.1x SSC at 55°C. Four clones, FH32, FH57, FB47 and FP48 (arrowheads) show varying degrees of hybridization to both probes, while others show homology to only one repetitive element. Lanes marked Kpn I and Alu I contain ng amounts of the cloned insert indicated and serve as hybridization controls, while the pBR322 lane shows covalently closed circular (I), relaxed circular (II) and linear (III) forms of the plasmid used as vector.

cccDNA, total (uncloned) fibroblast cccDNA was labeled by ^{32}P-nick translation and hybridized against arrays of individual clones. Seventy-five percent of the clones showing strong hybridization signals (and hence highly represented in the cccDNA population) were of mitochondrial origin. Among the nonmitochondrial clones, most of the hybridization was attributable to repetitive elements such as Alu (compare "Alu" and "cccDNA" columns in Table 3), although some clones showed signals even though they lacked repetitive elements for which we have probes (Table 3).

Inserts corresponding to moderate to unique copy number sequences were also isolated and used to probe Bam HI or Hind III-digested total human DNA. Figure 6 shows hybridizations of six cloned cccDNA probes to total cell DNA in the presence (A lanes) and absence (B lanes) of admixed insert DNA correspon-

TABLE 3

Characterization of cccDNA Clones: Size and Sequence Homologies of Cloned cccDNA Inserts

CLONE[a]	SIZE(Kbp)	ALU I	KPN I	ECO RI[b]	BLUR 16	REPETITIVE	cccDNA
FH-4	2.4						
FH-12	0.6	X				X	
FH-19	3.8						
FH-20	2.6	X				X	
FH-24	4.2		X		X	X	
FH-28	0.9	X				X	
FH-29	0.4						
FH-30	1.0	X				X	
FH-32	3.6	X	X			X	X
FH-34	5.7	X				X	
FH-35	0.4						
FH-38	0.7						
FH-42	0.7						
FH-44	0.4						
FH-46	0.5	X				X	
FH-47	1.6						
FH-49	0.7						
FH-50	0.5		X				
FH-51	0.3						
FH-54	2.4						
FH-57	3.6	X	X			X	
FH-69	1.4						
FH-70	3.2						
FH-71	2.6	X				X	
FH-72	1.2						
FH-74	0.4						
FH-75	1.8			X		X	
FH-76	1.8			X		X	
FH-82	3.5						
FH-89	2.2						
FH-101	3.7		X			X	
FH-102	3.5						
FH-103	0.8						
FH-117	2.2						
FH-123	0.8						
FH-124	0.3						
FH-131	1.6						
FH-137	1.8						
FH-138	1.5						
FH-140	0.3						
FH-144	3.9						
FH-147	0.6						
FH-151	0.2						
FH-153	0.2						
FH-154	1.0	X				X	X
FH-155	0.3						
FH-157	3.9	X				X	X
FH-163	0.7						
FH-165	0.5						
FH-170	1.5					X	
FH-174	2.1		X				
FH-175	0.1	X				X	X
FH-178	1.5	X				X	X
FB-1	7.0						
FB-2	3.9						X
FB-4	6.8						X
FB-7	1.8		X		X	X	X
FB-8	0.7						
FB-15	1.2						
FB-20	2.4						

CLONE	SIZE(Kbp)	ALU	KPN I	ECO RI	BLUR 16	REPETITIVE	cccDNA
FB-22	8.0	X				X	X
FB-24	1.3	X				X	X
FB-26	2.3						
FB-27	6.0						
FB-34	1.3	X				X	X
FB-35	8.0	X				X	X
FB-43	4.0	X				X	X
FB-47	3.5	X	X		X	X	X
FB-48	4.8		X				
FB-51	4.4						
FB-57	1.7		X		X	X	X
FB-59	4.0	X				X	X
FB-67	4.5						
FB-70	4.0	X				X	X
FB-71	4.5	X				X	X
FB-83	4.6						
FB-85	3.8	X				X	X
FB-86	6.4					X	X
FP-1	0.8						
FP-5	2.1						
FP-6	0.5						
FP-7	0.4						
FP-8	2.1						
FP-13	6.5	X				X	X
FP-15	0.7						
FP-20	3.7	X				X	X
FP-23	4.4						
FP-29	4.4						
FP-33	1.8					X	
FP-40	1.2						
FP-44	3.0	X				X	
FP-48	7.5	X	X		X	X	X
FP-49	4.4						
FP-50	0.9	X				X	X
FP-51	1.7						
FP-56	3.6						
FP-59	4.8	X				X	X
FP-69	0.8						
FP-75	4.9		X			X	
FP-80	4.4						
FP-81	4.4						
FP-89	1.6						
FP-90	1.6						
FP-92	2.2						
FP-109	0.6						
FP-116	3.0	X				X	X
FP-120	3.4						
TOTALS		-----	-----	-----	-----	------	------
107		30	12	(2)[2]	5	40	26

Inserts obtained following digestion of recombinant plasmids with the enzyme initially used for linearization of cccDNAs were hybridized with probes shown above, as prepared in methods.

[a]Clone code: F = fibroblast derived; H = cloned into the Hind III site of pBR322; B = cloned into the Bam HI site of pBR322; P = cloned into the Pst I site of pBR322.

[b]FH-75 and FH-76 are most likely siblings and thus represent only one clone.

ding to the probe. Arrowheads indicate self-hybridization of insert-probe. All samples were electrophoresed in the same gels, transferred and probed on the same filters, although in some cases exposure time was adjusted to obtain clear bands. Hybridization to genomic DNA was observed in all cases, with one to many bands being present. In several instances (Figure 6, FH-123 and FH-174), the fragment cloned from isolated cccDNA did not correspond in size to any of the homologous chromosomal bands. Such experiments indicate that a substantial fraction of cloned cccDNA fragments differ from their chromosomal counterparts in restriction fragment length.

DISCUSSION

Extrachromosomal Origin of cccDNA Clones

Several lines of evidence indicate that the DNA sequences, isolated here as cccDNA and cloned, in fact originated as extrachromosomal, covalently closed circular molecules of human DNA. cccDNA prepared by two cycles on CsCl-EthBr gradients was devoid of detectable linear fragments as shown by electron microscopy (Figure 1). In an initial CsCl-EthBr isopycnic gradient, fractions corresponding in density to cccDNA hybridized to probes representing interspersed highly repetitive DNA sequences (Alu I and Kpn I) but showed relatively little signal for the equally repetitive but tandemly-reiterated Eco RI family (Figure 2). This indicates that cccDNA, while containing a wide variety of chromosomal sequences, is distinctly different from chromosomal DNA in the proportions of these sequences. DNA restriction fragments cloned into pBR322 from cccDNA after two cycles of banding reflected essentially the same sequence bias: frequent occurrence of Alu I and Kpn I repeats but very infrequent incidence of Eco RI or Satellite III repeats (Table 3). Moreover, while all cloned fragments hybridized back to total human DNA, several of the inserts found in less highly repetitive clones showed sizes different from the homologous chromosomal sequences, digested with the same enzymes used for cloning (Figure 6). It should be noted that the circular form of a chromosomal sequence would often be expected to yield a restriction fragment differing in size from its chromosomal counterpart, especially if linearized by a single cleavage. Taken together, these observations demonstrate that sequences isolated and cloned were derived from extrachromosomal cccDNA.

Cloning of cccDNA Molecules

The rationale for selecting three different hexanucleotide-recognizing restriction enzymes, Hind III, Bam HI and Pst I, was to reduce the possibility of omitting tandemly repetitive sequences in cccDNA which might either be

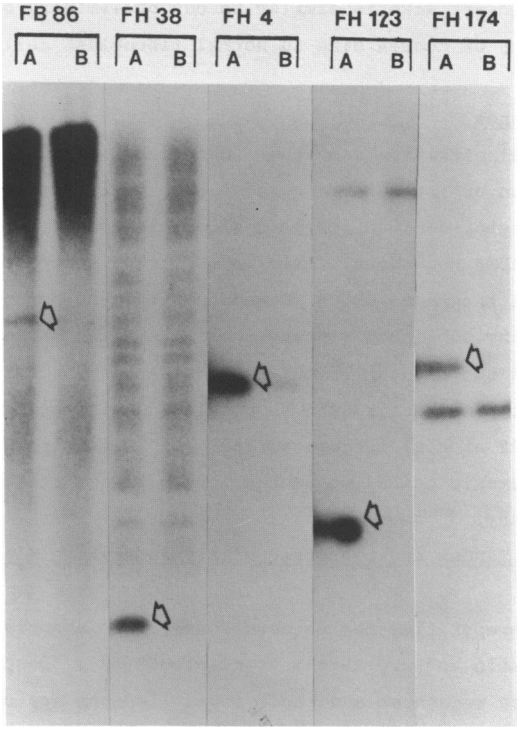


Figure 6. Comparison of insert sizes of cloned DNAs to homologous chromosomal DNA. Inserts from several clones were isolated, mixed with 3 μg of chromosomal DNA which had been digested with the same enzyme used in cloning and electrophoresed (A lanes) alongside lanes containing only 3 μg of digested chromosomal DNA (B lanes). Gels were transferred to nitrocellulose and paired lanes were probed with ^{32}P-labeled insert as indicated in the figure. Arrowheads mark the position of added insert in all A lanes.

digested into many small fragments, or not be cut at all, by any one particular enzyme. These enzymes would be expected to cut random sequence human DNA, on average, once every 5,762 bp for Pst I and Bam HI, versus once per 3,279 bp for Hind III, based on the GC content of the human genome and the recognition sequences of these enzymes (35). Clearly, this would have biased the cloning data somewhat in favor of larger, if not intact, cccDNA molecules. It is therefore likely that many clones represent full length cccDNAs since the average size of DNA inserts, severalfold larger than previously reported (5,6), approaches the average size of cccDNA visualized by electron microscopy in a variety of cultured cells with finite lifespans (30), and in tissues examined immediately after removal from their in vivo milieu (4,10,30). Fur-

thermore, cloned insert size is similar to our preliminary observations by electron microscopy of cccDNA size in normal fibroblast cells (Figure 1 and unpublished data).

Instability of Clones

The substantial attrition seen over four rounds of selection coupled with the high proportion of irregular clones, i.e., those that could not be cut or gave unexpected vector-derived fragment sizes, is consistent with instability of many cloned cccDNA sequences. This is in marked contrast to 50 clones of human Hind III DNA fragments which showed no attrition or irregularities. We are not aware of any published reports showing similar evidence of vector instability during the cloning of eukaryotic DNA sequences in Rec A^- E. coli. However, synthetic DNA inserts such as poly(dC-dG) larger than 50 bp are deleted from pBR322 at high frequency (36). It is thus possible that some of the most recombinogenic cccDNA sequences for that very reason are not included in our clone library.

Representation of Tandem and Dispersed Highly Repetitive Elements in cccDNA Clones

Tandem highly-repetitive DNA sequences are often associated with centromeric heterochromatin and are rarely transcribed (37). Despite their large copy numbers, these sequences and their repeat lengths are remarkably well conserved (28,37) and they appear only rarely to undergo translocations into other parts of the genome (38). In contrast, interspersed repetitive sequences, such as the Alu I and Kpn I families of humans, are frequently transcribed and appear highly mobile both in terms of chromosomal localization and their apparent insertion within other sequences (12,39-41). Such mobility, however, has been inferred from comparisons of sibling species or of individuals, and normally only on an evolutionary time scale (29,42,43), while ontogenetic instability within somatic tissues has not yet been demonstrated. If circular DNA molecules represent instability in the somatic cell genome, then the relative mobilities of highly repetitive DNA elements should be reflected in their occurrence within the circular DNA populations.

No clones were obtained which contained sequences homologous to human Satellite III (based on the pentanucleotide -TTCCA-), which is consistent with its simple sequence tandem repeat structure and low abundance (<1%) in the genome. The human alphoid (Eco RI) tandem repeat sequence was found in only 0.9% of independently derived clones, which appears low relative to the ~3% of the human genome composed of this repeat family. Under-representation in recombinant clones has also been observed for the homologous simian sequence,

α-satellite DNA, in BSC-1 cells (6). Although 20% of BSC-1 DNA is α-satellite, only 2-4% of cccDNA and less than 1.5% of resultant clones contained this sequence (6). Under-representation of alphoid DNAs within clone libraries may be ascribed in part to the infrequent cleavage of such sequences by the restriction enzymes used for cloning. It is clear, however, that alphoid DNA is already diminished in the initial cccDNA populations prior to cloning (Figure 2 and ref. 6). This might reflect their low frequency of genomic rearrangement (38) or their low transcriptional activity (37, and Shmookler Reis, et al., J. Mol. Biol., in press) if this is indeed an intermediate step in circle formation via reverse transcription as proposed (12,40,44-46).

Calculation of the expected number of clones containing interspersed repeat sequences, such as the Alu I and Kpn I families, is not quite as straightforward as in the case of tandem repeats since the size and spacing of the interspersed segments must also be considered. Alu I repeats (~300 bp) occur on average once per 7 kb (27,47). Since the mean cccDNA insert size was 2.5 kb, ~36% of inserts (2.5 kb/7 kb) would be expected to contain an Alu I sequence in reasonable agreement with the observed frequency, 28%.

Members of the Kpn I repeat family vary both in size, due to truncation at the 5' end (41), and in sequence (29). The 3' end of the Kpn I repeat unit occurs in about 20,000 copies per haploid human genome, i.e., once per 150 Kbp. Since the mean size of Kpn I elements is ~2.9 Kpb, with a median of ~1.4 Kbp (calculated from data in ref. 48), the expected frequency of 2.5 Kbp inserts containing Kpn I sequences would be ~2.5%. (Each 150 Kbp of genomic DNA could generate 60 inserts of mean length 2.5 Kbp, of which ~1.5 would be expected to contain some Kpn I sequences.) A similar estimate is obtained from probing of two independent clone libraries of the human genome, with two subfragments of the Kpn I repeat unit, which detected homology in 15-16% of clones containing 18 Kbp inserts (29). Kpn I elements therefore occur every 113-120 Kbp in these clones, and thus ~3% of 2.5 Kbp inserts would be expected to contain Kpn I homology. Since 12 Kpn I-hybridizing clones were observed (11.2%), we conclude that these elements are at least threefold overrepresented in our cccDNA library.

The observed concordance of Alu I and Kpn I in ~4% of clones, however, is very nearly that expected from random assortment (0.28 x 0.11 $\simeq$3%). Although detailed characterization of the irregular (likely more recombinogenic) clones has not yet been completed, the data thus far would suggest that the putatively mobile Kpn I DNA elements, but not the Alu I, repeats may promote the generation or replication of fibroblast cccDNA.

Diploid Fibroblasts vs. Established Cell Lines

Previous studies of cccDNA have focused on permanent cell lines which allow substantially higher yields of circular molecules for characterization and cloning. Because such cell lines are invariably unstable in karyotype, the possibility must be considered that this instability may also be reflected on a molecular scale, e.g., in the generation of cccDNAs. Indeed, direct visualization of cccDNA employing a mica press adsorption technique indicates that circular DNA elements are severalfold smaller in transformed cell lines than in primary cell cultures (4,5,10). This difference appears to be corroborated by a comparison of cloned insert lengths obtained here and in other cccDNA studies. The length of cloned inserts should reflect the size of endogenous cccDNAs, provided there are relatively few cuts per circle, but will also depend on the choice of vector and restriction enzymes. Our data, showing an average insert length of 2.52 Kbp, are most directly comparable to cccDNA clones from CHO immortal hamster cells which average 0.65 kb per insert (5), since both studies utilized restriction enzymes with 6 bp recognition sites and the same plasmid vector. Our clones were also slightly larger on average than those obtained from HeLa cell cccDNA using a lambda phage vector (~1.9 Kbp, ref. 13), despite the tendency of plasmid vectors to favor smaller inserts than phage vectors. Comparison of insert sizes thus argues in favor of a substantial reduction in cccDNA size accompanying cell immortalization. Interestingly, both studies of human cccDNA (13 and this report) yielded mitochondrial clones only of the 0.9 kb Hind III fragment (see Table 3 legend) although this enzyme generates 0.9 kb, 5.5 kb and 10.1 kb mtDNA fragments in equimolar amounts (34).

Origin and Significance of cccDNA

Using cccDNA clones to probe total fibroblast DNA, chromosomal counterparts were found in all cases (Figure 6), suggesting cccDNAs are derived from chromosomal DNA. Several models have been proposed for the generation of extrachromosomal cccDNAs from chromosomal locations in a variety of eukaryotic systems (11,12,45,46). These include homologous recombination and excision, perhaps between direct or inverted repetitive elements (10,12,45,46), transposition forming transient extrachromosomal cointegrate structures (49), replicon misfiring (50-53) and reverse transcription (44,45) discussed above. Replicon misfiring, leading to "onion skinning" of sequences near origins of DNA replication (51-53), is a particularly attractive model which is consistent with several features of cccDNA (52), including its amplification in cells blocked at the beginning of S phase (12,52). Since transformed (50) and

embryonic (54) cells are reported to contain severalfold more functional origins of replication than nontransformed cells, they must contain smaller replicons, which would account for the smaller cccDNAs observed (this report and ref. 10) if circular DNA species are actually generated by replicon misfiring.

In summary, we have isolated, cloned and begun to characterize cccDNA fragments from human diploid fibroblasts. We can now investigate the question of cccDNA number and size during development, aging and immortalization of cells with respect to individual extrachromosomal elements.

ACKNOWLEDGMENTS

We thank the following for generously providing probes: Drs. T. Friedmann and P. Deininger for BLUR 1, 16 and 18; Dr. G. Van De Woude for the retroviral LTR, DR. J. M. Bishop for the v-myc clone, and Dr. J Maio for Kpn I. We also thank Drs. M. Rush and J. Krolewski for stimulating discussion of this topic. Expert technical assistance by E. J. Moerman, R. Jones, B. Styles, M. Nelson and M. Morrison is gratefully acknowledged. This work was supported by grants (to KR) from the Noble Foundation and the University of Arkansas for Medical Sciences, and from the NIH (AG-03314 and AG-03787) and the Veterans Administration (to SG and RJSR).

REFERENCES

1. Flavell A.J. and D. Ish-Horowicz. Nature 292:591-595, 1981.
2. Stanfield S. and D.R. Helinski. Cell 9:333-345, 1976.
3. Stanfield S. and J.A. Lengyel. Proc. Natl Acad. Sci. U.S.A. 76:6142-6146, 1979.
4. DeLap R.J. and M.G. Rush. Proc. Natl Acad. Sci. U.S.A. 75:5855-5859, 1978.
5. Stanfield S.W. and D.R. Helinski. Molec. Cell. Biol. 4:173-180, 1984.
6. Bertelsen A.H., M.Z. Humayun, S.G. Karfopoulos and M.G. Rush. Biochem 21:2076-2085, 1982.
7. Rush M.G., R. Eason and J. Vinograd. Biochem. Biophys. Acta 228:585-594, 1971.
8. Radloff R., W. Bauer and J. Vinograd. Proc. Natl Acad. Sci. U.S.A. 57:1514-1521, 1967.
9. Smith C.A. and J. Vinograd. J. Mol. Biol. 69:163-178, 1972.
10. Yamagishi H., T. Kunisada and T. Tsuda. Plasmid 8:299-306, 1982.
11. DeLap R.J., M.G. Rush, D. Zouzias and S. Khan. Plasmid 1:508-521, 1978.
12. Krowlewski J.J., A.H. Bertelsen, M.Z. Humayun and M.G. Rush. J. Molec. Biol. 154:399-415, 1982.
13. Kunisada T. and H. Yamagishi. Gene 31:213-225, 1984.
14. Grimaldi G. and M.F. Singer. Proc. Natl Acad. Sci. U.S.A. 79:1497-1500, 1982.
15. Tudzynski P. and K. Esser. Molec. Gen. Genetics 173:71-84, 1979.
16. Wright R.M. and D.J. Cummings. Nature 302:86-88, 1983.
17. Harley C.B. and S. Goldstein. J. Cell Physiol. 97:509-515, 1978.
18. Schneider E.L. and E.J. Stanbridge. In: Methods in Cell Biology, Vol. 10, Chapt. 16, 1975.
19. Benz E.J., M.J. Murnane, B.L. Tonkonow, B.W. Berman, E.M. Mazur, C. Caulesco, T. Jenko, E.L. Snyder, B.G. Forget and R. Hoffman. Proc. Natl Acad. Sci. U.S.A. 77:3509-3513, 1980.
20. Shmookler Reis R.J. and S. Goldstein. J. Biol. Chem. 259:9078-9085, 1983.

21. Birnboim H.C. and J. Doly. Nucl. Acids Res. 7:1513-1523, 1979.
22. Maniatis T., E.F. Fritsch and J. Sambrook. Cold Spring Harbor Laboratory 1982.
23. Davis R.W. and R.W. Hyman. J. Mol. Biol. 62:277-302, 1971.
24. Southern E. J. Mol. Biol. 98:503-517, 1975.
25. Rigby P.W., M. Dieckmann, G. Rhodes and P. Berg. J. Mol. Biol. 113:237-251, 1977.
26. Feinberg A.P. and B. Vogelstein. Anal. Biochem. 132:6-13, 1983.
27. Jelinek W.R., T.B. Toomey, L. Leinwand, C. Duncan, P. Biro, P.V. Choudary, S.M. Weissman, C.M. Rubin, P.L. Deininger and C.W. Schmid. Proc. Natl Acad. Sci U.S.A. 77:1398-1402, 1980.
28. Wu J.C. And L. Manuelidis. J. Mol. Biol. 142:363-386, 1980.
29. Shafit-Zagardo B, J.J. Maio and F.L. Brown. Nucl. Acids Res. 10:3175-3193, 1982.
30. Calabretta B., D.L. Robberson, H.A. Barrera-Saldana, T.P. Lambron and G.F. Saunders. Nature 296:219-225, 1982.
31. McClements W.L., L.W. Enquist, M. Oskarsson, M. Sullivan and G.F. Vande Woude. J. Virol. 35:488-497, 1980.
32. Alitalo K., J.M. Bishop, D.H. Smith, E.Y. Chen, W.W. Colby and A.D. Levinson. Proc. Natl Acad. Sci U.S.A. 80:100-104, 1983.
33. Osterlund M., H. Luthman, S.V. Nilsson and G. Magnusson. Gene 20:121-125, 1982.
34. Drouin J. J. Mol. Biol. 140:15-34, 1980.
35. Moore G.P. and A.R. Moore. J. Theor. Biol. 98:165-169, 1982.
36. Wells R.D., R. Brenna, K.A. Chapman, T.C. Goodman, P.A. Hart, W. Hillen, D.R. Kellogg, M.W. Kilpatrick, R.D. Klein, J. Klysik, P.F. Lambert, J.E. Larson, J.J. Miglietta, S.K. Neuendorf, T.R. O'Connor, C.K. Singleton, S.M. Stirdivant, C.M. Veneziale, R.M. Wardell and W. Zacharias. C.S.H. Symp. Quant. Biol. Vol. 7, 77-84, 1982.
37. Singer M.F. Intl Rev. Cytol. 76:67-112, 1982.
38. Shmookler Reis R.J., A. Srivastava, K.T. Riabowol, J.W. Hardin and S. Goldstein. R. Sohal (ed). In: Molecular Biology of Aging: Gene Stability and Gene Expression, Raven Press, New York (in press).
39. Potter S.S. Proc. Natl Acad. Sci. U.S.A. 81:1012-1016, 1984.
40. DiGiovanni L., S.R. Haynes, R. Misra and W.R. Jelinek. Proc. Natl Acad. Sci. U.S.A. 80:6533-6537, 1983.
41. Grimaldi G. and M.F. Singer. Nucl. Acids Res. 11:321-338, 1983.
42. Gillespie D., J.W. Adams, C. Costanzi and M.J. Caranfa. Gene 20:409-414, 1982.
43. Sun L., K.E. Paulson, C.W. Schmid, L. Kadyk and L. Leinwand. Nucl. Acids Res. 12:2669-2689, 1984.
44. Flavell A.J. and D. Ish-Horowicz. Cell 34:415-419, 1983.
45. Krolewski J.J. and M.G. Rush. J. Mol. Biol. 174:31-40, 1984.
46. Krolewski J.J., C.W. Schindler and M.G. Rush. J. Mol. Biol. 174:41-54, 1984.
47. Houck C.M., F.P. Rinehart and C.W. Schmid. J. Mol. Biol. 132:289-306, 1979.
48. Grimaldi G., J. Skowronski and M.F. Singer. EMBO J 3:1753-7959, 1984.
49. Campbell A. Cold Spring Harbor Symposia on Quantitative Biology, VOl. XLV, Part. 1, pp. 1-9, 1981.
50. Martin R.G. and A. Oppenheim. Cell 11:859-869, 1977.
51. Schimke R.T. Cancer Res. 44:1735-1742, 1984.
52. Mariani B.D. and R.T. Schimke. J. Biol. Chem. 259:1901-1910, 1984.
53. Varshavsky A. Proc. Natl Acad. Sci. U.S.A. 78:3673-3677, 1981.
54. Alberts B., D. Bray, J. Lewis, M. Raff, K. Roberts and J.D. Watson. In: Molecular Biology of the Cell, Garland Publ. Inc., New York, p. 645, 1983.