MOLECULAR AND CELLULAR BIOLOGY, July 1985, p. 1621–1629
0270-7306/85/071621-09$02.00/0

Vol. 5, No. 7

Properties of Some Monkey DNA Sequences Obtained by a Procedure That Enriches for DNA Replication Origins

MARIA ZANNIS-HADJOPOULOS,[1] GABRIEL KAUFMANN,[2]† SHAW-SHYAN WANG,[2] ROBERT L. LECHNER,[2] ESSAM KARAWYA,[2] JOANNE HESSE,[2] AND ROBERT G. MARTIN[2]*

Laboratory of Molecular Biology, National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases, Bethesda, Maryland 20205,[2] and McGill Cancer Center, McGill University, Montreal, Canada H3G1Y6[1]

Received 27 December 1984/Accepted 5 April 1985

Twelve clones of monkey DNA obtained by a procedure that enriches 10^3- to 10^4-fold for nascent sequences activated early in S phase (G. Kaufmann, M. Zannis-Hadjopoulos, and R. G. Martin, Mol. Cell. Biol. 5:721–727, 1985) have been examined. Only 2 of the 12 *ors* sequences (origin-enriched sequences) are unique (*ors1* and *ors8*). Three contain the highly reiterated *Alu* family (*ors3*, *ors9*, and *ors11*). One contains the highly reiterated α-satellite family (*ors12*), but none contain the *Kpn* family. Those remaining contain middle repetitive sequences. Two examples of the same middle repetitive sequence were found (*ors2* and *ors6*). Three of the middle repetitive sequences (the *ors2-ors6* pair, *ors5*, and *ors10*) are moderately dispersed; one (*ors4*) is highly dispersed. The last, *ors7*, has been mapped to the bona fide replication origin of the D loop of mitochondrial DNA. Of the nine *ors* sequences tested, half possess snapback (intrachain reannealing) properties.

Very little is known of the structural features of nuclear mammalian origins of replication. If DNA synthesis is precisely programmed (12, 15), then one might expect some common feature(s) of structure or sequence among those origins activated at each point in S phase (3, 5, 13, 29). Indeed, Radford et al. (27) observed an accumulation of repetitive sequences after release from a hydroxyurea block and therefore suggested that origin regions might be enriched for dispersed repetitive sequences. Gronostajski et al. (16a) have purified a cellular protein required for the initiation of adenovirus DNA replication, nuclear factor I, and have shown that HeLa cell sequences can be isolated that preferentially bind to this protein. They have proposed that these sequences might represent HeLa replication origins. On the other hand, it has been observed (41) that short nascent DNA strands are enriched for snapback sequences and has been suggested therefore that origins may be enriched for inverted repeats. Finally, Roth et al. (30) and Montiel et al. (25) have shown that a number of mouse and human DNA sequences, respectively, contain AT-rich regions homologous to the presumptive nuclear replication origins of yeast (*ars* sequences) and can serve as replication origins for yeast plasmids. They have proposed that these sequences may be mouse and human DNA replication origins.

Unlike the situation in mammalian cells, considerable progress toward the characterization of replication origins has been made for *Saccharomyces cerevisiae*. The *ars* sequences have been isolated on the basis of their ability to serve as origins for plasmid replication (7, 33, 34). Curiously, the yeast mitochondrial replication origin also serves as an *ars* element (18). Both repetitive and unique nuclear *ars* sequences have been found (8, 9, 22). Of the former, three classes of repetitive sequences have been identified. One family is found in the tandemly arranged ribosomal gene units (35); the other two families, *ARS-x* and *ARS-y*, are tandemly repeated at the telomeres of yeast chromosomes (8, 9). The relevance of these findings to mammalian origins is unknown.

In mammalian cells, mitochondrial DNA (mtDNA) is known to commence replication at a few closely linked sites with the L strand as the template. Most replication ceases after the synthesis of a ~700-base-pair (bp) fragment, so that most mtDNA molecules contain a D loop (10, 16, 20). Elongation of the H strand in the D loop proceeds approximately two-thirds of the way around the circular mtDNA before L-strand replication starts at a fixed (palindromic) site (20). mtDNA replicates throughout S phase.

The isolation of monkey (CV-1) sequences enriched 10^3- to 10^4-fold for nascent DNA that replicated early in S phase (and hence are believed to contain origins of DNA replication) has been reported (21). The isolation of these *ors* sequences (origin enriched sequences) entailed the following: (i) synchronization of CV-1 cells by serum starvation, followed by serum stimulation and aphidicolin block; (ii) release for 1 min from the aphidicolin block to allow the activation of replication origins that commence replication at the beginning of S phase; (iii) preparation of cell extracts (nuclei and cytosol); (iv) brief elongation in the presence of 5Hg-dCTP of those nascent strands that had been activated during the first minute of S phase; (v) purification of high-molecular-weight DNA; (vi) extrusion of the nascent strands as double-stranded DNA (42); (vii) sucrose gradient fractionation of those nascent strands with molecular weights corresponding to ~500 to 2,000 bp; (viii) chromatography on thiol agarose; (ix) concentration by Elutip-d column chromatography; (x) digestion briefly with mung bean nuclease to generate flush ends; and (xi) cloning in pBR322 at the *Nru*I site.

It has been shown that four of the first five *ors* fragments analyzed (*ors1*, *ors2*, *ors4*, and *ors5*) preferentially replicate early in S phase (21). In this communication, we describe the genomic frequency, the genomic distribution, and the snapback properties of 12 *ors* sequences. It has not been determined whether any of these sequences contain bona

* Corresponding author.
† Present address: Department of Biochemistry, University of Tel Aviv, Ramat Aviv, Israel 69978.

fide nuclear origins of DNA replication, but one, *ors7*, contains the mtDNA origin sequence.

MATERIALS AND METHODS

The isolation of the *ors* sequences has been described (21). Restriction maps were made with commercial restriction enzymes (Bethesda Research Laboratories, Rockville, Md., or New England Biolabs, Beverly, Mass.) and analyses by acrylamide or agarose gel electrophoresis. Restriction fragments were purified by acrylamide gel electrophoresis, eluted in 0.2 M NaCl–1 mM EDTA–20 mM Tris-hydrochloride (pH 7.4), concentrated by passage through Elutip-d columns (Schleicher and Schuell, Inc., Keene, N.H.), and precipitated with 3 volumes of alcohol. Nick translation was carried out by the method of Rigby et al. (28) as modified by Bethesda Research Laboratories.

Dot blots were carried out with Gene-Screen Plus membranes (New England Nuclear Corp., Boston, Mass.) and under conditions recommended by the manufacturer for hybridization (68°C) and washing. Salmon sperm DNA was used as the carrier. Radioactivity was assayed by densitometer measurements of the radioautograms. Blot hybridizations were carried out by the method of Southern (32), except that Gene-Screen Plus membranes were used. Electrophoresis was at 35 V overnight in 0.8 or 1.2% agarose gels.

Snapback assays were performed as previously described (41). The CV-1 DNA used as controls was digested with either *Hin*dIII and *Bgl*I or a combination of *Ava*II, *Bcl*I, *Bgl*I, *Hin*fI, *Taq*I, and *Bst*NI. The DNA fragments were end labeled with ^{32}P by exchange labeling with polynucleotide kinase (2). The M_ns were determined by agarose or acrylamide gel electrophoresis of the labeled fragments with molecular weight markers and densitometer tracings of the radioautograms. Sequencing was performed by the method of Maxam and Gilbert (23).

RESULTS

Genomic frequency of the *ors* sequences. Twelve pBR322 clones containing monkey *ors* sequences were chosen for analysis. As previously reported (21), ~4 × 10^3 colonies resistant to ampicillin but sensitive to tetracycline were obtained after transfection with pBR322 DNA that had been linearized with the restriction endonuclease *Nru*I and flush end ligated to origin-enriched monkey DNA. Of the first 600 examined, 43 (~7%) contained inserts. Although we had fractionated the monkey DNA to obtain material of ~500 to 2,000 bp, most of the inserts were ≤500 bp in length. This may have resulted (at least in part) from our use of mung bean nuclease to generate flush ends for ligation. We have therefore concentrated on those monkey sequences that are 500 bp or greater in length. As reported, probes for the first five sequences chosen at random (*ors1* to *ors5*) were prepared and tested against the initial 43 clones that contained inserts. Only one showed cross-reactivity. *ors2* cross-reacted with a clone containing an insert of ~200 bp, and this insert was called *ors6*. An additional 6 sequences (not from the initial 43)—5 of ≥500 bp and 1 of ~200 bp—were chosen at random.

Restriction maps were made for all 12 *ors* sequences (9 of which are illustrated in Fig. 1). As can be seen, there are no discernible similarities between or among the sequences, with two exceptions. The map for *ors6* is identical to an internal portion of *ors2*, and *ors7* appears to contain a tandem repeat.

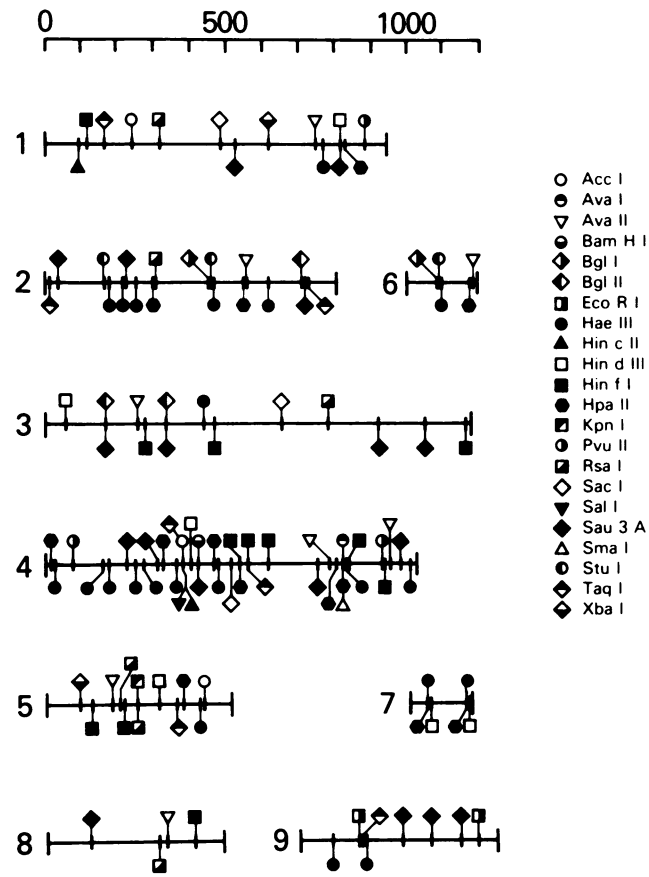

FIG. 1. Restriction maps of *ors1* through *ors9*.

To ascertain whether or not these sequences could represent a random sample of the monkey genome, 60% unique and 40% moderately to highly reiterated sequences (19), quantitative dot blots were made for each sequence. Serial dilutions of CV-1 DNA and of the plasmid DNAs containing the inserts (as well as controls of pBR322 DNA lacking any monkey insert) were adsorbed to Gene-Screen Plus paper and probed with the purified insert. Sample blots are illustrated in Fig. 2.

Of the 12 *ors* sequences, 10 contained moderately or highly reiterated sequences. From analyses similar to those in Fig. 2, estimates of the reiteration frequencies were made. Only two unique sequences were found. *ors1* appeared to be represented fewer than five times per haploid genome, as did the right-hand portion of *ors4*. From this analysis, *ors8* appeared to be represented ~300 times. However, it was impossible to obtain a probe from *ors8* that was completely free of contaminating pBR322 sequences, and the control lanes (data not shown) indicated that a significant proportion of our probe cross-reacted with pBR322 sequences. We therefore cloned *ors8* into M13 and, using the technique of Hu and Messing (17) to prepare a partially (M13) double-stranded probe, again performed quantitative dot hybridizations. The results (Table 1) suggested that *ors8* is also a unique sequence. *ors3*, *ors9*, *ors11*, and *ors12* contain highly reiterated sequences, whereas *ors2*, *ors4*, *ors5*, *ors6*, *ors7*, and *ors10* were all moderately reiterated.

Could these sequences have been derived from a random sample of CV-1 DNA? First, the probability of finding a second example of *ors2* among 43 inserts, even though *ors2*

(~800 bp) is reiterated ~2 × 10^3 times per haploid genome, is only 0.019, i.e., 1 − [1 − (800 × 2,000/3.5 × 10^9)][43]. Second, it is improbable that only 2 unique sequences of 11 (or of 12) sequences would be randomly selected from a population (CV-1 DNA) in which 60% of the sequences are unique. By χ^2 analysis, the probability that the selected sequences represent a random sample of CV-1 DNA is <0.03. This second calculation is only an approximation, since the expected frequency depends on the size and dispersion of all reiterated sequences. If many reiterated sequences are not highly dispersed, then our sample becomes more unlikely. Thus, statistical analyses suggest that these 12 *ors* sequences do not represent a random sample of CV-1 DNA, but that a unique subset of CV-1 sequences has been cloned.

Three of the four highly reiterated *ors* sequences contain *Alu* sequences. Each of the sequences were cross hybridized to plasmids containing the following sequences: *Alu* and *Kpn* (both kindly provided by M. Singer), mouse ribosomal DNA (kindly provided by G. Vogeli), and yeast *ARS-x* and *ARS-y* (kindly provided by C. Chan and B. Tye). The latter two were also tested at reduced stringency (50°C). Only *ors3* (the right-hand portion), *ors9*, and *ors11* hybridized to *Alu* sequences. No other hybridization was apparent. The remaining highly reiterated sequence (*ors12*) hybridized to nearly 15% of the entire CV-1 genome (Table 1) and was therefore tentatively identified as containing an α-satellite sequence (24, 36).

Genomic distribution of *ors* sequences. To examine the nature of the reiterated sequences, CV-1 DNA was purified, and samples were digested with one of nine different enzymes, fractionated by agarose gel electrophoresis, and blotted onto Gene-Screen Plus paper. The blots were then hybridized with probes to each of the *ors* sequences. Sample patterns are illustrated in Fig. 3. The sizes of the various fragments detected are listed in Table 2.

The genomic DNA blots hybridized with *ors3* (both the right-hand and left-hand portions), *ors4* (the left-hand portion), *ors9*, and *ors11* each showed smears of radioactivity

TABLE 1. Molecular abundance of the *ors* sequences[a]

Sequence	Probe (bp)[b]	Copy no.[c]
ors1	*Hin*cII-*Hin*dIII (716)	4
ors2	*Taq*I (708)	2,000
ors3	*Hin*fI (700)	20,000
ors3 R	*Rsa*I-*Hin*fI (379)	60,000
ors3 L	*Bgl*II-*Sac*I (322)	2,000
ors4	*Pvu*II (855)	1,000
ors4 R	*Bam*HI-*Pvu*II (512)	5
ors4 L	*Pvu*II-*Bam*HI (343)	200
ors5	*Acc*I-*Xba*I (340)	3,000
ors7	*Hin*dIII (108)	2,000
ors8	*Rsa*I-*Dde*I (212)	300
	*Rsa*I-*Dde*I (212)	5
ors9	*Eco*RI (324)	70,000
ors10	*Xba*I-*Eco*RI (658)	100
ors11	*Eco*RI-*Xba*I (328)	70,000
ors12	*Dde*I (56, 58, 101, 109, 135)	>10^6

[a] All probes were prepared by nick translation; the *ors8* probe was also prepared with M13 by the method of Hu and Messing (17). Restriction sites correspond to those in Fig. 1, except that *Dde*I sites are not shown. R, Right-hand portion; L, left-hand portion.

[b] The numbers in parentheses refer to the sizes of the probe used (unpublished results).

[c] Copy number per haploid quantity of DNA, assuming 3.3 × 10^9 bp per haploid genome (31).

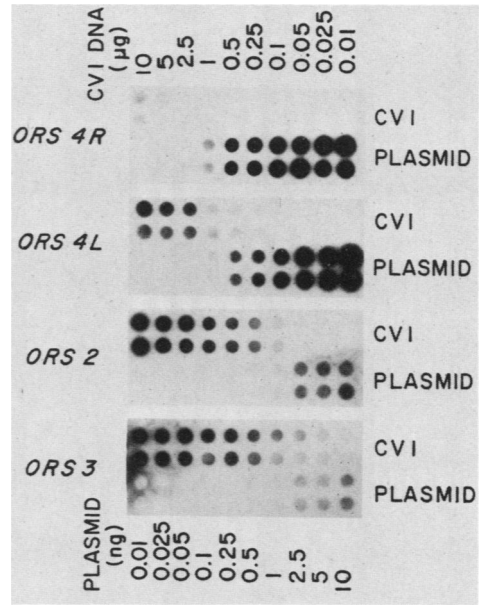

FIG. 2. Dot blots of monkey DNA and plasmids carrying the various *ors* segments hybridized with nick-translated *ors* probes.

extending from >23 kilobases (kb) to the ends of the gels (~1.5 kb). (Two or more probes were used in each case with different blots of cellular DNA.) We therefore conclude that the reiterated sequences in *ors4* (and the left-hand portion of *ors3*) are highly dispersed. Since *ors9*, *ors11*, and the right-hand portion of *ors3* contain *Alu* sequences, our results with these probes merely confirm that *Alu* sequences are highly dispersed (19). The genomic DNA blot hybridized with a probe from *ors12* showed a ladder of bands separated by ~170 bp as expected for α-satellite sequences (24).

The sequences in *ors2* and *ors5* are (at least in part) present as tandem repeats (Fig. 3 and Table 2), since in each case a limited number of prominent bands were observed after extensive digestion of the monkey DNA (even though these sequences are moderately reiterated, as shown above).

For the *ors2* probe, multiple bands were seen on short exposure times only with the restriction enzymes *Bgl*I and *Bgl*II which generate internal cuts (Fig. 1). Whereas this could indicate that all ~2 × 10^3 copies of *ors2* are present as a simple, tandemly repeated region, we do not believe this to be the case. On longer exposure, each of the lanes showed extensive smears of radioactivity, suggesting that the flanking regions were highly dispersed.

With the *ors5* probe, multiple bands with a "no-cut" enzyme, *Sac*I, were evident on short exposure, suggesting multiple chromosomal locations of this sequence. Again, smears were observed on prolonged exposure of the lanes in which other restriction enzymes had been used. We tentatively conclude, therefore, that *ors2* and *ors5* sequences are tandemly arranged as oligomeric units with the oligomers located at diverse chromosomal sites. With the *ors10* probe, multiple bands were found with several no-cut enzymes (*Bam*HI, *Bgl*I, *Hin*dIII, and *Pvu*II), suggesting that perhaps all of these sequences are located at different chromosomal sites.

Even on prolonged exposure with an *ors7* probe (Fig. 4), very little smearing of the label was apparent. As noted above, the *ors7* sequence itself contains a tandem repeat (Fig. 1; and confirmed by sequence analysis below) delimited by two *Hin*dIII sites. (The 108-bp sequence liberated by

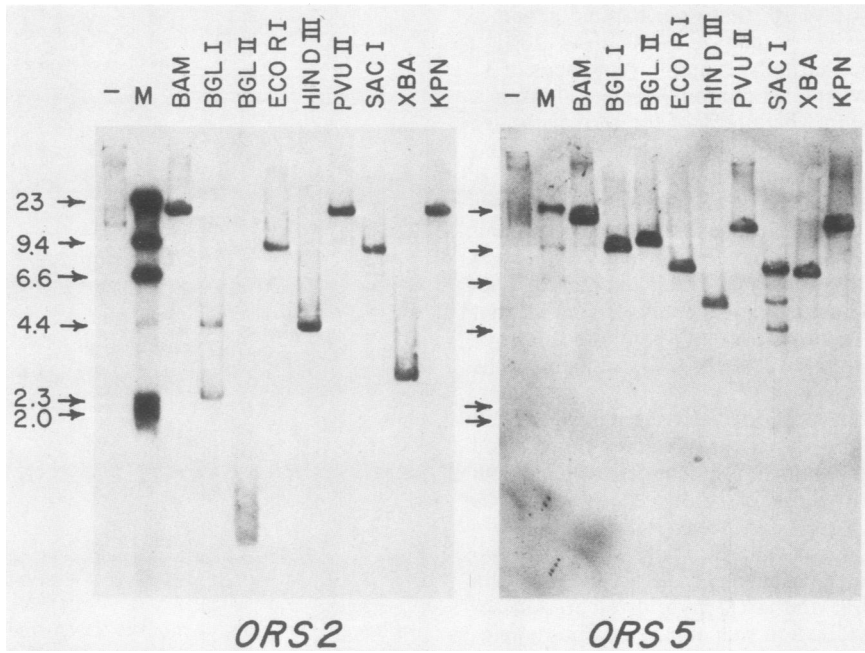

FIG. 3. Genomic CV-1 DNA digested with the indicated restriction enzymes were electrophoresed overnight at 35 to 40 V in 0.8% agarose gels and blotted to Gene-Screen paper. The blots were hybridized with nick-translated probes to *ors2* or *ors5*. M, *Hin*dIII λ marker DNA; −, undigested high-molecular-weight CV-1 DNA. Numbers represent sizes of the markers in kilobases.

*Hin*dIII digestion of the plasmid was nick translated to make the probe used in these experiments.) The fact (Fig. 4) that, even after digestion of CV-1 DNA with *Hin*dIII, two prominent bands of 4.7 and 5.6 kb can be seen (the 108-bp sequence having eluted from the gels) suggests that the 108-bp repeat unit in *ors7* is present as a very short oligomer within a larger repeating unit. From the fact that *Bam*HI, *Pvu*II, and *Kpn*I each generate a 17-kb fragment, we inferred that the larger unit was 17 kb in length. To confirm this hypothesis, pairwise and three enzyme digestions of CV-1 DNA were carried out and blotted (a portion of these data is shown in the right-hand panel of Fig. 4). The blots were then probed with the *ors7* probe, and the fragment sizes are given in Table 3. It is possible to generate a unique restriction map for the cellular DNA from these results (Fig. 5). The restriction map is very similar to that about the origin of the H strand (D loop) in human mtDNA (1, 11). Consequently, CV-1 mtDNA was purified (4), immobilized on Gene-Screen Plus paper, and hybridized with the *ors7* probe. A strong signal indicated that *ors7* was homologous to mtDNA. The sequence of *ors7* and its similarity to human mtDNA are illustrated in Fig. 6.

Snapback of *ors* DNA. At least half of the *ors* sequences tested were retained by hydroxyapatite immediately after denaturation (snapback DNA) as if they contained double-stranded regions. When insufficient time is allowed for interchain reannealing to occur, denatured DNA can still be partially retained by hydroxyapatite if intrachain annealing occurs due to the presence of inverted repeats (39, 40, 42). (Simple *Alu* sequences have very low C_0t values by virtue of their great abundance in monkey DNA but do not exhibit snapback properties [19; unpublished data].) As it has been reported (41) that origin-enriched mammalian DNA is also enriched for snapback sequences, we tested the first nine of these *ors* sequences for their retention on hydroxyapatite both before and immediately after denaturation in 95% formamide (Table 4). The control of CV-1 DNA of 33% retention at M_n of ~2 kb is somewhat higher than we or

TABLE 2. Molecular size of CV-1 restriction fragments hybridizing to the *ors* sequences[a]

Sequence	Molecular size (kb) of:								
	*Bam*HI	*Bgl*I	*Bgl*II	*Eco*RI	*Hin*dIII	*Pvu*II	*Sac*I	*Xba*I	*Kpn*I
ors1	20	—	16	18	—	7.5	6.6	1.4	—
ors2	18	4.4, 2.4	0.9	8.8	4.4	19	8.7	2.8	19
ors3	S	S	S	S	S	S	S	S	S
ors4 R	18	—	—	14	23	—	1.5	3.0	18
ors5	18.5	10	11.5	8.0	5.8	16.5	7.8, 6.0, 4.7	7.6	17
ors7	17	4.5	5.0	9.0	5.6, 4.4	17	8.8	6.0	17
ors8	>23	—	>23	>23	>23	>23	>23	>23	>23
ors9	S	S	S	S	S	S	S	S	S
ors10	12.5, 5.3, 3.8	20, 7.4, 5.7	13.5	11.6	17.5, 5.5	6.8, 5.4	6.2	—	18.5
ors11	S	S	S	S	S	S	S	S	S
ors12	1	1	1	1	1	1	1	1	—

[a] S, Smears of label; —, no distinct band observed; 1, 172-bp spaced ladder of bands; R, right-hand portion of *ors4* sequence.

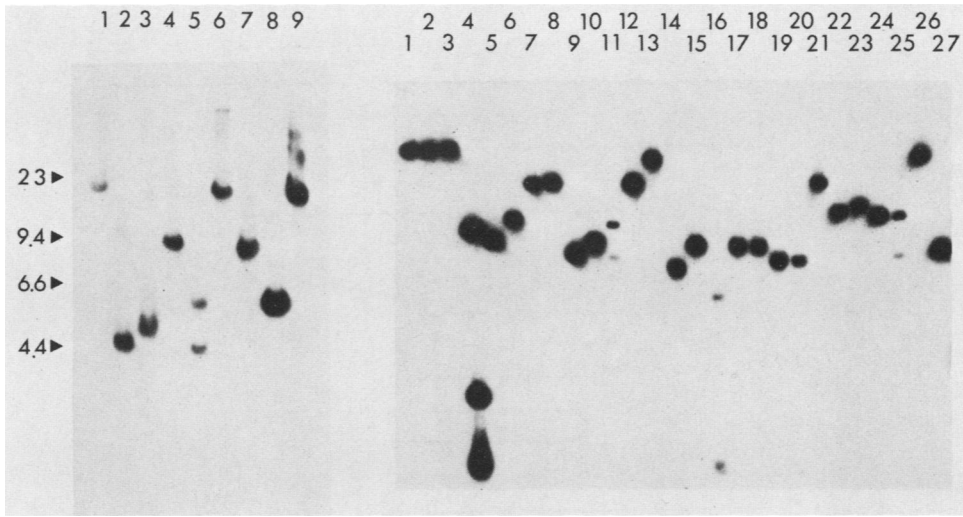

FIG. 4. Genomic blots of CV-1 DNA hybridized with an *ors7* probe. The left-hand panel is of a 0.8% gel, and the right-hand panel is of a 1.2% gel. The size markers for the left-hand panel are given in kilobase pairs. The CV-1 DNA was digested with the following enzymes: *Bam*HI (lane 1), *Bgl*I (lane 2), *Bgl*II (lane 3), *Eco*RI (lane 4), *Hin*dIII (lane 5), *Pvu*II (lane 6), *Sac*I (lane 7), *Xba*I (lane 8), and *Kpn*I (lane 9). The right-hand panel, lane 4, contains markers of 5.4, 1.76, and 1.13 kb. The other lanes in this panel contain CV-1 DNA digested with the following: *Kpn*I (lane 1), *Bam*HI (lane 2), *Pvu*II (lane 3), *Bgl*II (lane 5), *Xba*I (lane 6), *Sac*I (lane 7), *Eco*RI (lane 8), *Bam*HI-*Bgl*I (lane 9), *Bam*HI-*Bgl*II (lane 10), *Bam*HI-*Hin*dIII (lane 11), *Bam*HI-*Eco*RI (lane 12), *Bam*HI-*Pvu*II (lane 13), *Bgl*II-*Bgl*I (lane 14), *Bgl*II-*Eco*RI (lane 15), *Bgl*II-*Hin*dIII (lane 16), *Bgl*II-*Pvu*II (lane 17), *Bgl*II-*Sac*I (lane 18), *Eco*RI-*Bgl*I (lane 19), *Eco*RI-*Hin*dIII (lane 20), *Eco*RI-*Pvu*II (lane 21), *Eco*RI-*Sac*I (lane 22), *Eco*RI-*Kpn*I (lane 23), *Pvu*II-*Xba*I (lane 24), *Pvu*II-*Hin*dIII (lane 25), *Pvu*II-*Kpn*I (lane 26), and *Sac*I-*Bgl*I (lane 27).

others have obtained previously (6, 39, 41). The value of 13% retention for the control of CV-1 DNA (M_n of ~150 bp) is more in line with previous results. Whichever control is used, however, it is clear that *ors1* to *ors4* and *ors9* sequences are significantly retained by hydroxyapatite. Since these sequences are not quantitatively retained, we assume that the presumptive inverted repeats in these sequences are not extensive. (We have previously shown that the perfect inverted repeat [13 by 13 bp] at the simian virus 40 origin is sufficient to cause 94% retention of single-stranded fragments of 850 bp containing the inverted repeat, but only 74% retention when the single-stranded DNA is 5,243 bp [41].)

DISCUSSION

Because we and others (unpublished data) had not obtained plasmids that could replicate autonomously in mammalian cells when chromosomal mammalian DNA was ligated to DNA fragments containing selective markers, we devised an alternative approach to the isolation of mammalian origins of DNA replication. Nascent DNA enriched 10^3- to 10^4-fold for sequences replicating early in S phase were obtained by purely physical and chemical methods and used to generate clones. Four clones of the first five we examined, *ors1*, *ors2*, *ors4*, and *ors5*, were shown to replicate preferentially early in S phase (21).

Although our sample is small, and we cannot be certain that any of the segments contain true origins of replication (except *ors7*), certain common features seem to emerge. First, two independent isolates of the same sequence (*ors2* and *ors6*) were found in a sample of 43 plasmids derived from the origin-enriched DNA. Second, 9 of 11 of the sequences are reiterated. Of these, four are highly reiterated; three contain the *Alu* sequence, and one contains the α-satellite sequence. The remainder contain moderately reiterated sequences. Three, *ors5*, the *ors2-ors6* pair, and *ors10*, are

TABLE 3. Molecular sizes of *ors7*-containing fragments on multiple digestion of CV-1 DNA

Fragment	Molecular size (kb) of:									
	*Bam*HI	*Bgl*I	*Bgl*II	*Kpn*I	*Eco*RI	*Hin*dIII-A	*Hin*dIII-B[a]	*Pvu*II	*Sac*I	*Xba*I
*Bam*HI	17	4.5	5.0	9.4	9.0	5.6	4.5	14	8.8	6.0
*Bgl*I		4.5	4.1	4.5	4.4	<1	3.8	4.5	4.5	4.0
*Bgl*II			5.0	5.0	5.0	1.1	3.7	5.0	5.0	4.7
*Kpn*I				17	6.4	2.2	4.5	13	8.8	5.5
*Eco*RI					9.0	4.4	4.4	8.2	6.0	6.0
*Hin*dIII						5.6	4.5			
*Pvu*II						5.6	4.4	17	8.8	6.0
*Sac*I						1.8	4.5		8.8	5.1
*Xba*I						2.6	3.2			6.0
*Hin*dIII-*Bgl*I A	0.6		0.6	0.6	0.6			0.6	0.6	0.6
*Hin*dIII-*Bgl*I B	3.8		3.5	3.8	3.8			3.8	3.8	3.2

[a] The two fragments could be distinguished by their intensities. See, for example, lanes 11, 16, and 25 of the right-hand panel of Fig. 4.

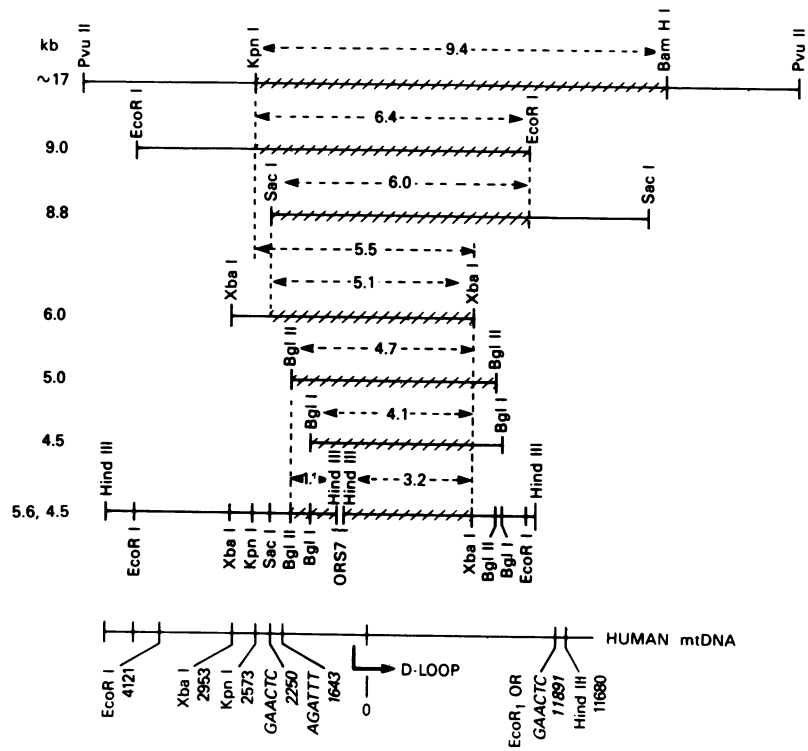

FIG. 5. A genomic map of the monomeric 17-kb unit detected with the *ors7* sequence as a probe. The hatched lines refer to the fragments detected on hybridization after multiple enzyme digestions (see the legend to Fig. 4 and Table 3). It was on the basis of the similarity of this map to that of human mtDNA (bottom portion of the figure) that we first suspected that *ors7* might be mitochondrial in origin. The restriction map of human mtDNA and the numbering system are those of Anderson et al. (1). A discrepancy exists between the map of human mtDNA as determined by Drouin (11) and the sequence of Anderson et al. (1). Drouin located an *Eco*RI site near the *Hin*dIII site at 11680, whereas no corresponding sequence (GAATTC) exists in Anderson et al. (1). However, the sequence GAACTC is present at approximately the same position that Drouin (11) found an *Eco*RI site.

dispersed throughout the genome, and all but perhaps *ors10* appear to be present as tandem repeats.

Our results are consistent with the predictions that mammalian origins occur in regions enriched for reiterated sequences (27) and snapback sequences (41). On the other hand, none appear to bind tightly to nuclear factor I (R. M. Gronostajski, personal communication).

It may be worth noting similarities of some of the reiterated monkey *ors* to some yeast *ars* sequences. The *ors2* (*ors6*) and *ors5* sequences are distributed in a manner similar to those of the yeast telometric *ARS-x* and *ARS-y* sequences, although they are more abundant than the yeast sequences. However, no cross hybridization between any *ors* sequence and the *ARS-x* or *ARS-y* sequences was found, and there is no other reason for supposing that these *ors* sequences are present near the telomeres.

Since we have no direct biological evidence that any of these *ors* sequences are bona fide replication origins (except *ors7*; see below), we cannot exclude the possibility that the true nuclear replication origins are only those containing the highly reiterated *Alu* sequences. On the other hand, it may be significant that no examples of the highly dispersed *Kpn* family were found.

ors7 requires special comment. It is clearly homologous to mtDNA in that region which corresponds to the region immediately upstream of the major 5′ termini of D loops in replicating mouse and human mtDNA (10, 16) (Fig. 5 and 6). This region of mtDNA is the most evolutionarily divergent portion of the molecule (1, 38), and subtle differences between the locations of the D-loop termini of mouse and human mtDNA have been demonstrated. Although considerable detailed information is available for both human and mouse mtDNA, no precise mapping of monkey D loops has been reported (4). Thus, whereas we cannot state unequivocally that *ors7* includes the monkey mtDNA origin, it does include an 18-bp origin consensus sequence (38).

How is it that our purification procedure could have selected a mitochondrial origin even though mtDNA replication does not proceed simultaneously on both DNA strands? There are at least three explanations. First, the isolation of *ors7* could be entirely fortuitous. It is clear that the nuclear and cytosol preparations we used for the incorporation of mercurated dCTP were heavily contaminated with mitochondria. It is possible that the mtDNA was sheared and inadvertently ended up in our final DNA for cloning. This explanation seems remote because unmercurated DNA is very efficiently eliminated during thiol-agarose chromatography (21) and because it offers no explanation for our isolation of the mtDNA origin.

A second class of explanations would suppose that, in making our nuclear and cytosol preparations, we somehow permeabilized the mitochondrial membrane, thus allowing mercurated nucleotides to enter the mitochondria and be incorporated into the continuously growing strands of the replicative intermediates. Alternatively, permeabilization might have allowed nuclear primase to enter the mitochondria, permitting Okazaki fragment formation on the single-stranded portion of the D loop. Again, there are difficulties

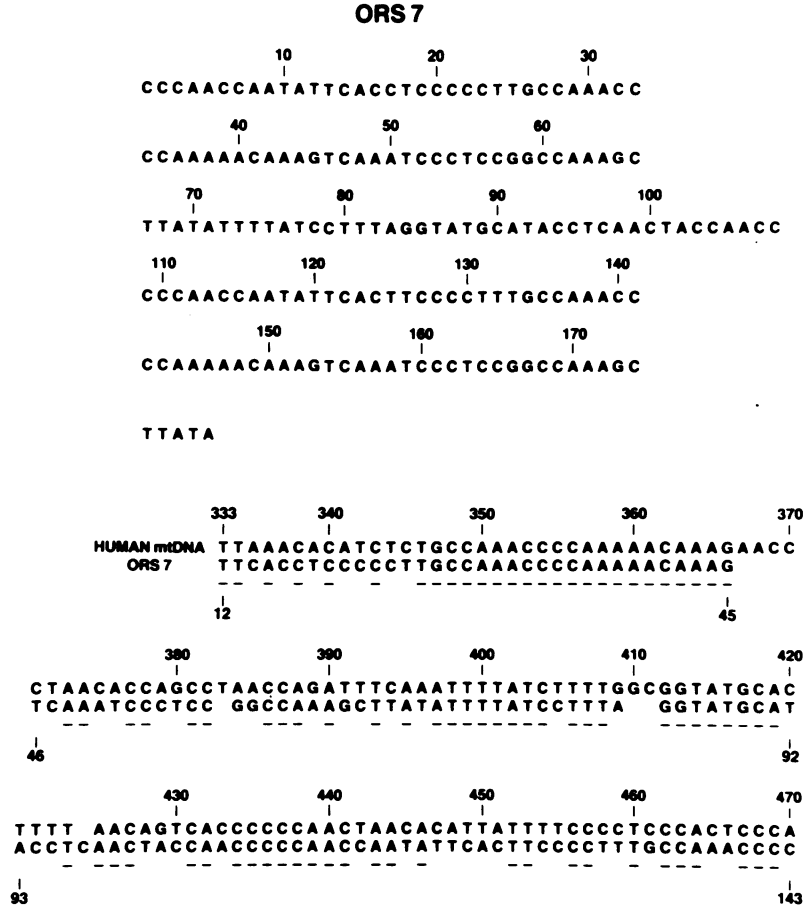

FIG. 6. The sequence of *ors7* as determined by the technique of Maxam and Gilbert (23). Both strands were sequenced. The sequence shows a number of unusual features. First, the sequence from 1 to 71 is almost perfectly repeated from position 109 to 179. Second, there is 64% homology between the first and second 33 bp and 61% homology between the bases from 109 to 141 and 142 to 174. In the lower portion of the figure, the sequence of *ors7* is compared with that of human mtDNA (1). A dash below the sequence indicates the identity of human and *ors7* sequences. The 18-bp sequence from position 346 to 363 of human mtDNA (25 to 42 of *ors7*) is conserved without a single base change in rat mtDNA and differs by only 1 bp from mouse mtDNA (38).

with this class of explanations. (i) *ors7* is just upstream of the major 5′ termini of the D loop (by analogy to mouse and human mtDNA) and is thus outside of the region in which Okazaki fragments would form. Also (ii), there is no point in the normal replication of mtDNA at which only this portion of the H and L strands have been replicated, i.e., *ors7* could not have arisen from a normal mtDNA replicative intermediate. The only explanation of this class that would seem to account for the isolation of *ors7* would be if permeabilization allowed the entry of Hg-dCTP into the mitochondria and if a γ-polymerase molecule which had just begun replication on the L strand then jumped to the H strand during the incubation in vitro.

TABLE 4. Snap-back properties of *ors* sequences

ors probe	Size (bp)	% Eluting at 0.3 M PO_4		
		Native (N)	Denatured (D)	D/N ($\times 10^2$)
ors1	716	92.0 ± 0.3	38.6 ± 1.5	42.0 ± 1.9
ors2	708	88.0 ± 0.9	51.8 ± 3.1	58.9 ± 4.0
ors3	700	85.6 ± 1.6	35.3 ± 3.6	41.2 ± 5.2
ors4	855	62.5 ± 2.5	37.0 ± 7.0	59.2 ± 9.5
ors5	340	87.7 ± 2.6	21.8 ± 2.4	24.9 ± 5.0
ors7	108	73.8 ± 0.8	23.1 ± 2.9	31.3 ± 3.7
ors8	212	94.7 ± 0.7	24.3 ± 1.3	25.7 ± 2.0
ors9	324	90.9 ± 4.2	36.7 ± 4.4	40.4 ± 8.6
CV-1	~2,000	96.4 ± 1.1	31.7 ± 1.5	32.9 ± 2.6
CV-1	~150	81.1 ± 0.9	10.1 ± 3.2	12.5 ± 4.2

We find a third explanation more appealing. It has recently been shown that copies of portions of mtDNA are also present in the nuclear genome (14, 26, 37). In *S. cerevisiae*, the segment of mtDNA also present in the nuclear genome clearly includes the mtDNA origin (14). Thus, it is possible that *ors7* was derived not from a normal mitochondrion but rather from a copy of the mitochondrial origin present in the CV-1 nuclear genome. Of course, such an explanation requires that the mtDNA D-loop origin function as an origin with either the mitochondrial replication protein complexes or the nuclear replication machinery. It is highly pertinent that the yeast mtDNA origin functions as an *ars* sequence (18) since the *ars* function is generally regarded as being dependent on the nuclear replication machinery. If this latter explanation were correct, then the isolation of *ors7* would provide the first evidence that mtDNA origins integrated into the nuclear genome still function as origins.

Whatever the true reason for our having isolated *ors7*, the isolation of a bona fide origin of replication, albeit of mtDNA, encourages us to suppose that other *ors* sequences contain nuclear origins.

ACKNOWLEDGMENTS

We thank W. M. Brown, C. S. M. Chan, D. A. Clayton, D. Livingston, J. Nichol, G. B. Price, M. Simpson, B.-K. Tye, and especially M. F. Singer for their comments and suggestions and R. M. Gronostajski for testing the binding of the *ors* sequences to nuclear factor I.

Parts of this research were supported by grant MA-7965 (to M.Z.-H.) from the Medical Research Council of Canada. M.Z.-H. is a recipient of the University Researcher Award of the Secrétariat à la Science et à la Technologie du Gouvernement du Québec.

LITERATURE CITED

1. **Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden, and I. G. Young.** 1981. Sequence and organization of the human mitochondrial genome. Nature (London) **290**:457–465.
2. **Berkner, K. L., and W. R. Folk.** 1977. Polynucleotide kinase exchange reaction Eco R I cleavage and methylation of DNAs containing modified pyrimidines in the recognition sequence. J. Biol. Chem. **252**:3176–3184.
3. **Braustein, J. D., D. Schultze, T. DelGiudice, A. Furst, and C. L. Schildkraut.** 1982. The temporal order of replication of murine immunoglobulin heavy chain constant region sequences corresponds to their linear order in the genome. Nucleic Acids Res. **10**:6887–6902.
4. **Brown, W. M., J. Shine, and H. M. Goodman.** 1978. Human mitochondrial DNA: analysis of 7S DNA from the origin of replication. Proc. Natl. Acad. Sci. U.S.A. **75**:735–739.
5. **Calza, R. E., L. A. Eckhardt, T. DelGiudice, and C. L. Schildkraut.** 1984. Changes in gene position are accompanied by a change in time of replication. Cell **36**:689–696.
6. **Cech, T. R., A. Rosenfeld, and J. E. Hearst.** 1973. Characterization of the most rapidly renaturing sequences in mouse mainband DNA. J. Mol. Biol. **81**:299–325.
7. **Chan, C. S. M., and B.-K. Tye.** 1980. The chromosomal replication origins in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. **77**:6233–6239.
8. **Chan, C. S. M., and B.-K. Tye.** 1983. Organization of DNA sequences and replication origins at yeast telomeres. Cell **33**:563–573.
9. **Chan, C. S. M., and B.-K. Tye.** 1983. A family of Saccharomyces cerevisiae repetitive autonomously replicating sequences that have very similar genomic environments. J. Mol. Biol. **168**:505–523.
10. **Chang, D. D., and D. A. Clayton.** 1985. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. U.S.A. **82**:351–355.
11. **Drouin, J.** 1980. Cloning of human mitochondrial DNA in Escherichia coli. J. Mol. Biol. **140**:15–34.
12. **Edenberg, H. J., and J. A. Huberman.** 1975. Eukaryotic chromosome replication. Annu. Rev. Genet. **9**:245–284.
13. **Fangman, W. L., R. H. Hice, and E. Chlebowitz-Siedziewska.** 1983. *ARS* replication during the yeast S-phase. Cell **32**: 831–838.
14. **Farrelly, F., and R. A. Butow.** 1983. Rearranged mitochondrial genes in the yeast nuclear genome. Nature (London) **301**: 296–301.
15. **Furst, A., E. H. Brown, J. D. Braunstein, and C. L. Schildkraut.** 1981. Globin sequences are located in a region of early replicating DNA in murine erythroleukemia cells. Proc. Natl. Acad. Sci. U.S.A. **78**:1023–1027.
16. **Gillum, A. M., and D. A. Clayton.** 1979. Mechanism of mitochondrial DNA replication in mouse L-cells: RNA priming during the initiation of heavy-strand synthesis. J. Mol. Biol. **135**:353–368.

16a. **Gronostajski, R. M., K. Nagata, and J. Hurwitz.** 1984. Isolation of human DNA sequences that bind to nuclear factor I, a host protein involved in adenovirus DNA replication. Proc. Natl. Acad. Sci. U.S.A. **81**:4013–4017.

17. **Hu, N., and J. Messing.** 1982. The making of strand-specific M13 probes. Gene **17**:271–277.
18. **Hyman, B. C., J. H. Cramer, and R. H. Rownd.** 1982. Properties of a Saccharomyces cerevisiae mtDNA segment conferring high-frequency yeast transformation. Proc. Natl. Acad. Sci. U.S.A. **79**:1578–1582.
19. **Jelinek, W. R., and C. W. Schmid.** 1982. Repetitive sequences in eukaryotic DNA and their expression. Annu. Rev. Biochem. **51**:813–844.
20. **Kasamatsu, H., and J. Vinograd.** 1973. Unidirectionality of replication in mouse mitochondrial DNA. Nature (London) New Biol. **241**:103–105.
21. **Kaufmann, G., M. Zannis-Hadjopoulos, and R. G. Martin.** 1985. Cloning of monkey DNA synthesized early in the cell cycle. Mol. Cell. Biol. **5**:721–727.
22. **Maine, G. T., P. Sinha, and B.-K. Tye.** 1984. Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes. Genetics **106**:365–385.
23. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–580.
24. **McCutchan, T. H., H. Hsu, R. E. Thayer, and M. F. Singer.** 1982. Organization of African green monkey DNA at junctions between α-satellite and other DNA sequences. J. Mol. Biol. **157**:195–211.
25. **Montiel, J. F., C. J. Norbury, M. F. Tuite, M. J. Dobson, J. S. Mills, A. J. Kingsman, and S. M. Kingsman.** 1984. Characterization of human chromosomal DNA sequences which replicate autonomously in Saccharomyces cerevisiae. Nucleic Acids Res. **12**:1049–1068.
26. **Nomiyama, H., M. Fukuda, S. Wakasugi, T. Tsuzuki, and K. Shimada.** 1985. Molecular structures of mitochondrial-DNA-like sequences in human nuclear DNA. Nucleic Acids Res. **13**:1649–1658.
27. **Radford, I. R., R. F. Martin, L. R. Finch, and G. S. Hodgson.** 1982. Inhibition of DNA synthesis and cell death. Biochim. Biophys. Acta **696**:154–162.
28. **Rigby, D. W. J., M. Duckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
29. **Rivin, C. J., and W. L. Fangman.** 1980. Replication fork rate and origin activation during the S-phase of Saccharomyces cerevisiae. J. Cell Biol. **85**:108–115.
30. **Roth, G. E., H. M. Blanton, L. J. Hager, and V. A. Zakian.** 1983. Isolation and characterization of sequences from mouse chromosomal DNA with *ARS* function in yeasts. Mol. Cell. Biol. **3**:1898–1908.
31. **Singer, D.** 1979. Arrangement of a highly repeated DNA sequence in the genome and chromatin of the African green monkey. J. Biol. Chem. **254**:5506–5514.
32. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
33. **Stinchcomb, D. T., K. Struhl, and R. W. Davis.** 1979. Isolation and characterization of a yeast chromosomal replicator. Nature (London) **282**:39–43.
34. **Struhl, K., D. T. Stinchcomb, S. Scherer, and R. W. Davis.** 1979. Eukaryotic DNA segments capable of autonomous replication in yeast. Proc. Natl. Acad. Sci. U.S.A. **77**:4559–4563.
35. **Szostak, J. W., and R. Wu.** 1979. Insertion of a genetic marker into the ribosomal RNA gene of yeast. Plasmid **2**:536–554.
36. **Thayer, R. E., and M. F. Singer.** 1983. Interruption of an α-satellite array by a short member of the KpnI family of interspersed, highly repeated monkey DNA sequences. Mol. Cell. Biol. **3**:967–973.
37. **Tsuzuki, T., H. Nomiyama, C. Setoyama, S. Maeda, and K. Shimada.** 1983. Presence of mitochondrial-DNA-like sequences in the human nuclear DNA. Gene **25**:223–229.

38. **Walberg, M. W., and D. A. Clayton.** 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. **9**:5411–5421.
39. **Wilson, D. A., and C. A. Thomas, Jr.** 1973. Hydroxyapatite chromatography of short double-helical DNA. Biochim. Biophys. Acta **331**:330–340.
40. **Wilson, D. A., and C. A. Thomas, Jr.** 1974. Palindromes in chromosomes. J. Mol. Biol. **84**:115–144.
41. **Zannis-Hadjopoulos, M., G. Kaufmann, and R. G. Martin.** 1984. Mammalian DNA enriched for replication origins is enriched for snap-back sequences. J. Mol. Biol. **179**:577–586.
42. **Zannis-Hadjopoulos, M., M. Persico, and R. G. Martin.** 1981. The remarkable instability of replication loops provide a general method for the isolation of origins of DNA replication. Cell **27**:155–163.