Volume 14 Number 8 1986 Nucleic Acids Research

Multiple mechanisms generate extrachromosomal circular DNA in Chinese hamster ovary cells

Sharon W.Stanfield and Donald R.Helinski

Department of Biology, B-022, University of California, San Diego, La Jolla, CA 92093, USA

Received 8 January 1986; Accepted 18 March 1986

ABSTRACT

Seven cloned small circular DNA molecules from CHO cells were sequenced and examined for the presence of homologies to each other and to a number of other functional sequences present in transposable elements, retroviruses, mammalian repeat sequences, and introns. The sequences of the CHO cell circular DNA molecules did not reveal common structural features that could explain their presence in the circular DNA population. A gene bank was constructed for CHO chromosomal DNA and sequences homologous to two of the seven small circular DNA molecules were isolated and sequenced. The nucleotide sequences present at the junction of circular and chromosomal DNA suggest that a recombination process involving homologous pairing may have been involved in the generation of one, but not the other, of the two circular DNA molecules.

INTRODUCTION

Small polydisperse circular DNA (spcDNA) appears to be present in the majority, if not all, of higher eukaryotic cells in culture, as well as in a wide variety of cell and tissue types derived from intact organisms (1,2,3,4,5,6). Virtually all of the spcDNA molecules that have been investigated have been found to be homologous to, and presumably derived from, chromosomal DNA. The observation that spcDNA is homologous to all identified classes of chromosomal DNA, including unique sequences, but that the proportion of individual sequences present in spcDNA and in chromosomal DNA may differ (6,7,8), suggests that the generation of spcDNA cannot be completely random. Some of the spcDNA molecules from Drosophila and Caenorhabditis have been shown to consist entirely of transposable elements (9,10). Others, isolated from mammalian cells, were found to contain complete or partial copies of L1, Alu, or alphoid satellite DNA sequences (11,12,13,14). Given the large numbers of circular DNA molecules per cell in culture (up to several thousands in CHO and monkey cells) and the fact that their sequence complexity exceeds their total molecular mass on a per cell basis, it seems likely that a variety of different processes may be at work to generate spcDNA in a population of cells. Some genetic processes that have been proposed to produce spcDNA

include deletion of chromosomal DNA, transposition, and extrachromosomal amplification (13,14). These, in turn, may take place via different mechanisms including replicative recombination, reverse transcription, replicon misfiring, non-specific degradation, or homologous or illegitimate recombination.

Recently, sequence analysis of cloned spcDNA (13) and isolation of replicative intermediates (15) have been consistent with the involvement of homologous recombination and reverse transcription, respectively, in the generation of some spcDNAs, although other mechanisms could not be excluded. However, there has been only one report to date of a comparison between the exact chromosomal sequence from which a circular DNA molecule originated and the circular molecule itself (11). This example, involving a cloned HeLa cell spcDNA which carried a sequence homologous to a single copy chromosomal sequence, is consistent with homologous recombination as a mechanism of circle generation. The circular DNA in this case also contained a repetitive chromosomal sequence. Additional examples are needed in order to determine whether homologous recombination is a common mechanism for generating circles containing all types of sequences. For this reason, we have sequenced seven different cloned hamster spcDNA molecules, representing different repetitive sequence classes, and the chromosomal homologues for two of these molecules. By choosing two circular DNA molecules represented in the chromosome by unique and low copy number sequences, respectively, we were able to isolate and sequence the exact region from which these spcDNA molecules originated. The sequences located at the chromosome-circle junctions for one, but not the other, of the two circular molecules were consistent with the involvement of a recombination mechanism involving homologous pairing in the generation of the circular form of the molecule. The sequences of the seven individual spcDNA molecules did not provide evidence for common structural features, but suggested that spcDNAs represent the products of a diverse set of phenomena.

MATERIALS AND METHODS

Restriction Enzymes

Restriction endonucleases were obtained from Bethesda Research Laboratories (BRL), New England Biolabs, Boehringer Mannheim, and International Biotechnologies, Inc. Digestions were carried out according to the supplier's specifications.

Construction of a Chinese Hamster Ovary Cell Chromosomal Gene Bank

Chromosomal DNA from CHO-K1 cells was partially digested with Sau 3A and fragments of 17-23 kb were isolated on a sucrose gradient and cloned into the

BamH I site of ligated arms of lambda phage EMBL4 (16). The ligated mixture was packaged in vitro into lambda using BRL packaging mixes. Packaged DNA was used to infect Escherichia coli BNN45 cells. The resulting gene bank consists of a total of 8.6 x 10^5 plaques.

Isolation and Subcloning of Chromosomal Sequences Homologous to Cloned spcDNA

A total of 1 x 10^6 plaques was transferred to nitrocellulose paper and probed with nick-translated (17) purified Bgl II fragments of cloned spcDNAs B12 and B7 (1). Four clones homologous to spcB12 and five clones homologous to spcB7 were isolated. Each clone was found to contain the entire circular DNA sequence within a Hind III-Eco RI fragment. These fragments were subcloned into pUC9 (18) in order to simplify production of the deletions used for sequencing.

DNA Sequencing

Cloned CHO cell spcDNA was subcloned in M13mp9 and sequenced by the Sanger dideoxy termination method (19). Chromosomal sequences homologous to spcDNA were subcloned in pUC9 and labeled for sequencing by the denatured supercoil method (20). The entire sequence of both strands was obtained for each molecule reported here.

RESULTS AND DISCUSSION

Seven different cloned small circular DNA molecules from CHO cells (1) were selected for sequencing. These included molecules which hybridized both to low copy number (spcB12,B7,1) and to interspersed (spcB11,B9B,B2) and tandemly repeated (spcA3) sequences in the chromosome (Table 1). The results are shown in Figures 1 and 2 (brackets). The sequences of these molecules were examined for structural features that might explain their presence in the

TABLE 1. Some characteristics of the spcDNA clones chosen for sequencing.

spcDNA clone	Size (bp)	Distribution in chromosome	Homologous sequences in mouse DNA	Largest ORF (bp)
A3	269	tandemly repeated	no	186
B2	204	dispersed	no	183
B7	223	single copy	no	66
B9B	354	dispersed	not done	273
B11	354	dispersed	yes, repetitive	99
B12	391	very low copy	no	150
1	346	single copy	no	225

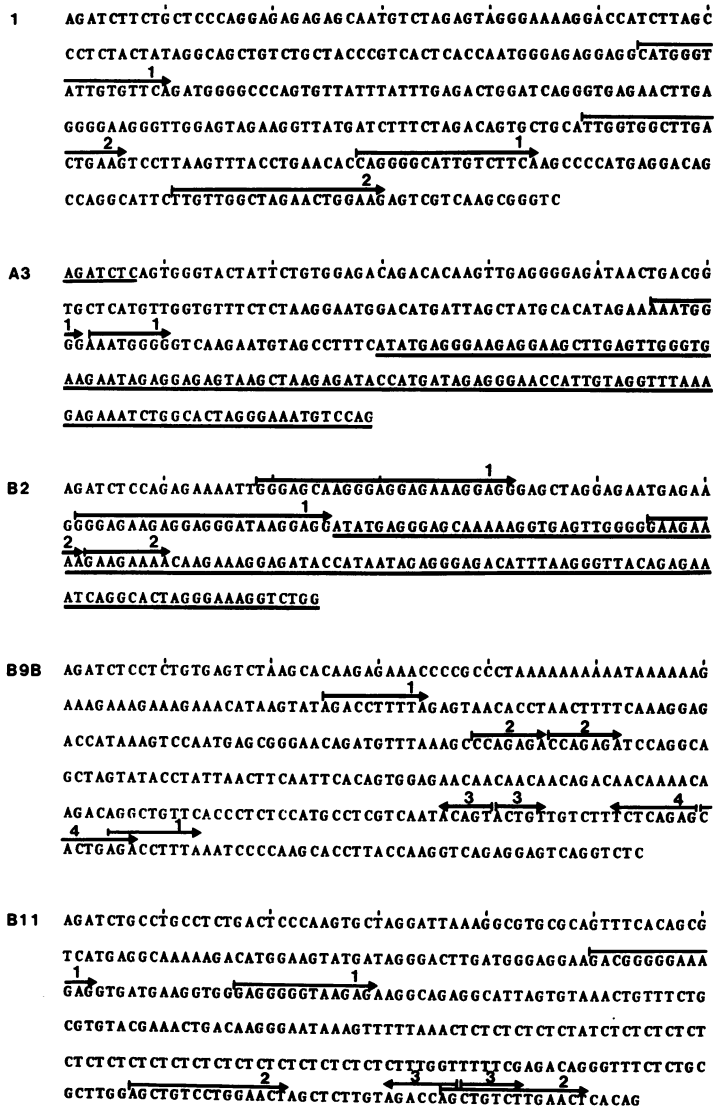

Figure 1. Complete nucleotide sequences of cloned CHO cell spcDNA. All sequences start at the Bgl II sites at which they were cloned. Underlining indicates homologous regions in spcA3 and B2. Overlining with numbered arrows indicates the location of direct and inverted repeats within each circular molecule. Additional non-adjacent repeats were found in spc1, B9B, and B11 but are not indicated.

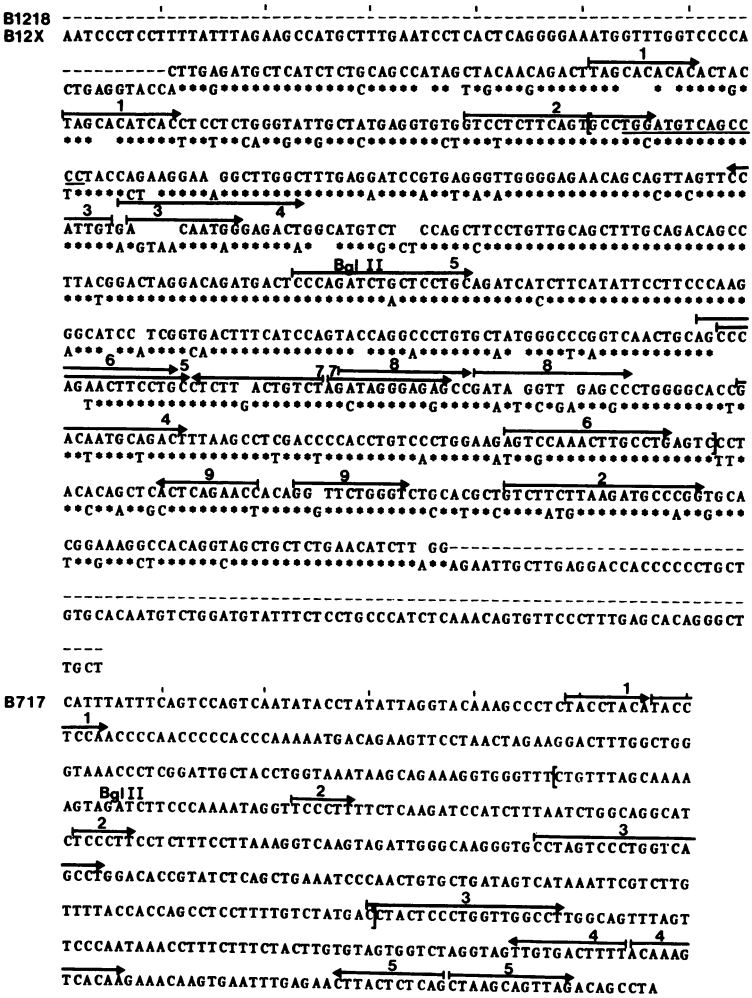

Figure 2. Nucleotide sequences of chromosomal homologues of cloned spcDNAB12 and B7. Brackets indicate the outer limits of the exact circular DNA sequences. The two homologues of B12 are aligned to facilitate sequence comparison. Dashed lines indicate the absence of sequence data. Asterisks show the location of identical base pairs. Gaps denote the absence of homologous bases. Overlining with numbered arrows indicates the location of direct and inverted repeats for cB1218 and cB717 only. Underlining delimits a region of B12 possessing homology to an LTR from a hamster IAP. The location of the Bgl II site is shown for each molecule.

population of spcDNA. These included direct and inverted repeats and homologies to each other and to known repetitive sequences, transposable elements, and retroviruses. All seven circles were found to contain one or more

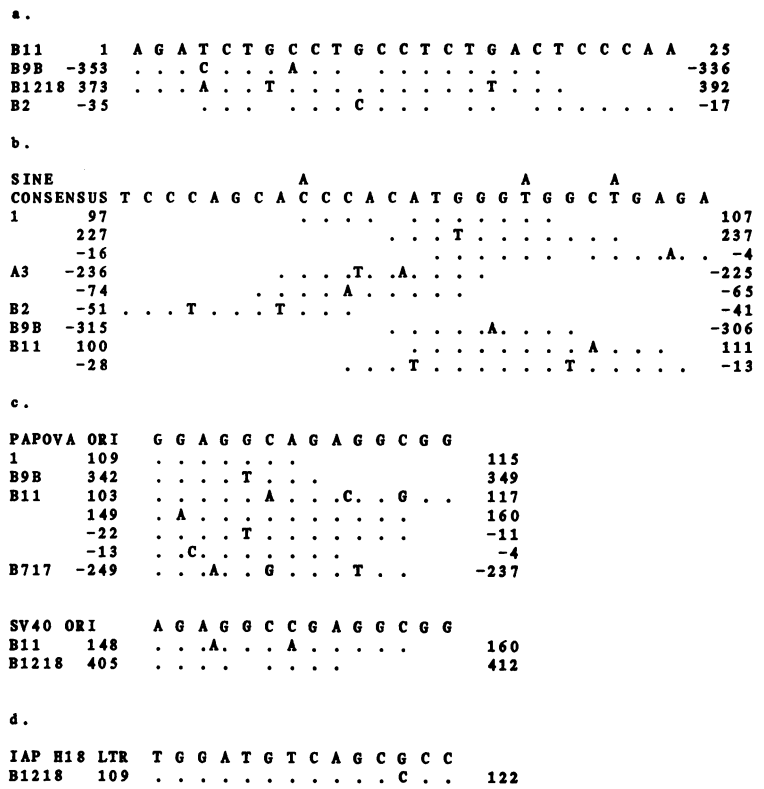

Figure 3. Alignment of portions of cloned spcDNA sequences with a variety of other sequences. a. A portion of the spcB11 sequence, b. the mammalian SINE consensus sequence, c. the SV40 and papova BK virus origin sequences, d. a portion of a Syrian hamster IAP LTR sequence. The numbers flanking each sequence indicate the location of that sequence in Figures 1 and 2. Negative numbers refer to the opposite strand.

inverted and/or direct repeats, the majority of which were imperfectly matched. Figures 1 and 2 (brackets) indicate the location of some of these internal repeats.

Short, imperfectly matched sequence homologies, ranging from 13 to 38 base pairs, were also found between many of the circles. For example, three of the seven circles sequenced had homologies to a 25 base pair region of spcB11, starting at and including the Bgl II site (Figure 3a). In addition, one long region (127 contiguous base pairs) present in spcA3 possessed 80% homology to a 126 base pair contiguous sequence in spcB2 (underlined in Figure 1). There were also numerous other homologies of polypurine rich regions

between spcB11 and spcB9B, B7, B2, and A3 (not shown). It is not clear whether any of these homologies has significance for the generation of spcDNA in the cell. All circles except spcB7 and B12, both represented by unique or low copy number sequences in the chromosome, also had some homology to a consensus sequence determined for all mammalian short interspersed repeats (21), which is itself homologous in part to the replication origins of SV40 and papova BK viruses (see Figure 3b and c).

Recent evidence has suggested a link between retrovirus-like transposable elements and introns, in which class II introns of Neurospora were found to code for a protein sharing some amino acid sequence homology with the reverse transcriptase of retroviruses (22). We therefore searched for the presence of intron-exon and exon-intron boundaries and for open reading frames (ORFs) that could code for reverse transcriptase related proteins. Two of the circles (spcB7 and B9B) contained sequences homologous to both donor and acceptor splice site consensus sequences (23). However, none had ORFs beginning with an initiation codon that could code for proteins related to reverse transcriptase.

One class of short interspersed repeats (SINES), the Alu sequences, are characterized by their ability to be transcribed by RNA polymerase III and by the presence of a variable stretch of A-rich sequence at the 3' end. Since all of the circles homologous to repetitive sequences in the chromosome also had some homology to the consensus sequence present in all mammalian SINES, we searched for RNA polymerase III recognition signals in the circles. None of the circles contained a good match to the RNA polymerase III promoter consensus sequence (24) and only two of the seven circles sequenced possessed A-rich regions (spcB2 and B9B).

The majority of cloned spcDNA molecules that have been studied by sequencing or by restriction enzyme mapping have structures consistent with an origin via homologous recombination. This includes spcDNA containing the alphoid satellite from both human (13) and monkey (8), the Alu family from monkey (14), L1 sequences from HeLa cells (11), and copia sequences from Drosophila (25). An additional mechanism for generating spcDNA may be used by copia-like transposable elements of Drosophila. Many of these copia-like sequences are present in Drosophila spcDNA (9), and RNA/DNA intermediates in reverse transcription have been detected for Drosophila mdg-1 and mdg-3 sequences (26). Based on the extensive similarities of copia-like sequences to vertebrate retroviral proviruses (27,28,29), it may be assumed that at least a portion of the intermediates would be resolved as circular molecules.

The spcDNAs selected for sequencing here were too small to possess the characteristic features of retroviral-like transposable elements. However, we did examine them for features common to the long terminal repeats (LTRs) of these sequences, such as short terminal inverted repeats, promoter-like sequences, and polyadenylation signals. None of these sequences was found. However, one circle, spcB12, did possess a short region of homology (Figure 3d) to a 14 bp sequence located 42 bp from the 5'-end of the LTR of an intracisternal A particle (IAP) from Syrian hamster (30). The Syrian hamster sequence is also identical to a sequence located in the same position in the LTR of the Mus musculus IAP rc mos (30). IAP genes are known to compose a major class of copia like elements in rodents. An spcDNA molecule containing IAP sequences has also been cloned from mouse tumor cells (15).

Two of the spcDNAs that were homologous to unique or low copy number sequences in the chromosome (spcB12 and B7) were used to probe a CHO chromosomal gene bank for homologous sequences. Four clones homologous to spcB12 and five clones homologous to spcB7 were isolated. These clones were digested with a number of restriction enzymes and run on agarose gels for rough sequence comparisons. Southern blots of these gels were probed with labeled homologous spcDNA to identify smaller homologous fragments for subcloning and to aid in determining differences between the clones. All five B7-homologous clones had very similar restriction enzyme patterns and possessed identical fragments which reacted with the labeled probe. Since B7 appeared to be a unique chromosomal sequence, based on hybridization to total chromosomal DNA (1), we assumed that all five lambda clones contained the same homologous chromosomal sequence. Therefore, only one of these clones, designated cB717, was chosen for sequencing.

Only two of the four clones homologous to spcB12 (cB12X and cB1219) had similar restriction enzyme patterns and possessed B12-homologous fragments that were identical in size. The other two clones (cB121 and cB1218) had several restriction fragments in common, but differed from each other in that one (cB1218) contained two homologous BamH I fragments, as expected (1), while the other (cB121) contained only one. Previous studies (1) had indicated that chromosomal DNA contains very few (and possibly only one) copies of the B12 sequence. To obtain as many different chromosomal homologues of spcB12 as possible, three of these clones were selected for sequencing: cB12X, cB121, and cB1218.

The sequences of the chromosomal homologues cB12X, cB1218, and cB717 are shown in figure 2. The sequence for cB121 (not shown) was identical to the

entire sequence for cB1218, with the exception of a single base change at the BamH I site in cB121 which generates the sequence 5'-AGATCC-3' and, as a result, eliminates the BamH I site in this clone. Since sequence alterations due to cloning seem to occur preferentially at symmetrical sites, such as those recognized by restriction enzymes (31), it is possible that cB121 and cB1218 are actually the same chromosomal sequence. cB12X, on the other hand, is significantly different from cB1218 (see Figure 2) and probably represents a separate sequence in the chromosome. This would be consistent with the additional bands seen when restriction enzyme digested chromosomal DNA was probed with labeled spcB12 (unpublished observations).

The chromosomal homologues cB1218 and cB717 contain complete exact copies of the B12 and B7 circle sequences, respectively. These sequences are circularly permuted with respect to the Bgl II sites at which they were cloned from the spcDNA preparation. This verifies that complete circular DNA molecules, as opposed to contaminating linear molecules or fragments of circular molecules, were cloned originally for both B12 and B7.

The isolation of two different chromosomal sequences with homology to spcB12 permits the following conclusions: spcB12 does not move independently from one site in the chromosome to another, and spcB12 DNA is part of a larger repetitive element in the chromosome. Both conclusions are based on the fact that the sequences flanking spcB12 homologous sequences in cB1218 and cB121 are homologous to each other. This would not be the case if a previously single copy of circular B12 had amplified and transposed to a new site in the chromosome. Since we were unable to isolate spcB12-homologous clones containing different flanking sequences, we must conclude that, in the hamster from which CHO-K1 cells were derived, spcB12 had not stably reentered chromosomal DNA. This is not too surprising since spcB12 was selected as having low copy representation in the chromosome. However, as previously reported (1), spcB12 homologous sequences were isolated in two separate non-sibling clones from a population of spcDNA, the second clone being designated spc20. spc20 has been sequenced (single strand only) and found to be identical to spcB12 (data not shown). Even considering preferential cloning of spcDNA containing a Bgl II site, the precise spcB12-type sequence must have been relatively abundant in the original spcDNA preparation from which it was cloned. Since spcB12 has been precisely generated at least twice, then some clue to the mechanism of its origin may be found in the sequences located at the spc-chromosomal DNA junction. These sequences and a possible mechanism for generation of a circular copy of B12 are given in Figure 4. An examination of cB1218 shows the

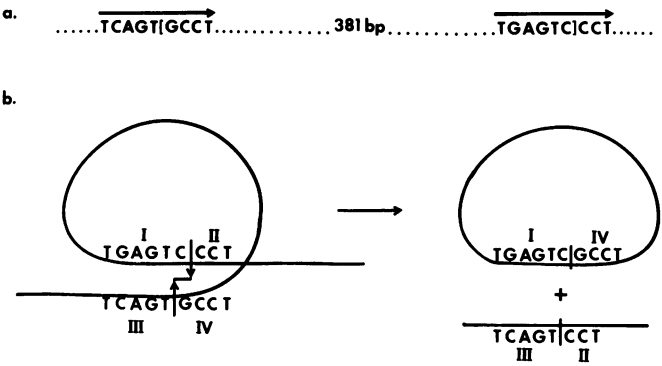

Figure 4. A model for the generation of circular DNA containing the B12 sequence. a. Imperfect direct repeats located at the chromosome-circle junction of cB1218. The brackets indicate the outer limits of the sequences present in spcB12. b. The generation of a circular copy of B12 via homologous pairing, staggered blunt end cutting, and ligation. Vertical arrows indicate the location of the staggered cuts.

presence of a 9 base pair sequence which crosses the left hand chromosomal-spcDNA junction and is repeated at the right hand junction with alterations at two of the 9 base pairs. Homologous pairing of this sequence, followed by staggered blunt end cutting and ligation of the separate strands could result in the generation of a circular copy of B12 bearing the observed sequence. It is interesting that the 14 bp region of spcB12 found to be homologous to Syrian hamster and mouse IAP LTRs is located only 3 bp away from the left hand circle-chromosomal junction shown in Fig. 2. This sequence might be recognized by enzymes that act specifically on IAP sequences and that could be involved in the type of near-homologous recombination proposed in Fig. 4.

Recombination involving homologous sequences cannot account for the generation of spcB7, however, since there are no direct or inverted repeats located at the chromosomal-circle junctions of cB717 (see Fig. 2). This is also true for a number of monkey spcDNA molecules containing sequences homologous to KpnI repeats (12). cB717 is somewhat unusual in that it is composed of a comparatively simple sequence containing a large number of short runs (3-5 base pairs) of each nucleotide. It is present as a single copy in the chromosome, but probably does not code for any protein product since its largest open reading frame with an initiation codon could code for only 22 amino acids. Since the circular copy of B7 is also infrequent (1), it is possible that spcB7 arose as a result of some random event, such as the recirculariza-

tion of degraded chromosomal DNA.

The longest possible ORF beginning with a potential initiation codon is listed for each spcDNA in Table 1. Since the longest translatable ORF present in any circle could code for a protein of only 91 amino acids, it is unlikely that any of these circular DNA's code for a complete enzyme. However, it is possible that they could participate in the formation of a functional protein when integrated in the chromosome, possibly by providing a single exon, since the average vertebrate exon consists of only 40 to 50 amino acid residues (32). This does not appear to be the case for the integrated copy of B7, where there are no possible ORFs reading into or out of the integrated circle containing more than 5 base pairs of spcDNA sequence. For the integrated B12 sequence, there is an ORF of indeterminate length reading in from the right arm, covering a total of 99 base pairs of spcDNA sequence, sufficient to code for one exon.

The results reported here support the view that spcDNAs represent the products of a diverse set of phenomena. We find no compelling evidence for common structural features among individual spcDNA molecules that could explain their presence in the circular DNA population. While it is apparent that no single mechanism is responsible for the generation of all spcDNA, a recombination mechanism which recognizes homologous sequences may be frequently involved. Also, regions possessing homology to SV40 or papova virus origins may play a role in the generation of some circular DNA molecules, possibly through replicon misfiring (33) followed by homologous recombination, but more data would be required to establish this possibility. In summary, sequence data reported here and in other reports (8,9,11,12,13,14,25,34) support the earlier kinetic and biophysical data indicating that populations of extrachromosomal circular DNA in eukaryotes are indeed very heterogeneous in nature.

ACKNOWLEDGEMENTS

This work was supported by Public Health Service grant AI-07194 from the National Institutes of Health and by National Science Foundation grant PCM77-66573.

REFERENCES

1. Stanfield, S.W. and Helinski, D.R. (1984). Molec. Cell Biol. 4, 173-180.
2. Stanfield, S. and Helinski, D.R. (1976). Cell 9, 333-345.
3. DeLap, R.J. and Rush, M.G. (1978). Proc. Natl. Acad. Sci. U.S.A. 75, 5855-5859.
4. Yamagishi, G., Kunisada, T., and Tsuda, T. (1982). Plasmid 8, 299-306.

5. Fujimoto, S., Tsuda, T., Toda, M., and Yamagishi, H. (1985). Proc. Natl. Acad. Sci. U.S.A. 82, 2072-2076.
6. Riabowol, K., Shmookler Reis, R.J., and Goldstein, S. (1985). Nucl. Acids Res. 13, 5563-5584.
7. Stanfield, S.W., and Lengyel, J.A. (1979). Proc. Natl. Acad. Sci. U.S.A. 76, 6142-6146.
8. Bertelsen, A.H., Humayun, M.Z., Karfopoulos, S.G., and Rush, M.G. (1982). Biochem. 21, 2076-2085.
9. Mossie, K.G., Young, M.W., and Varmus, H.E. (1985). J. Mol. Biol. 182, 31-43.
10. Ruan, K., and Emmons, S.W. (1984). Proc. Natl. Acad. Sci. U.S.A. 81, 4018-4022.
11. Jones, R.S. and Potter, S.S. (1985). Proc. Natl. Acad. Sci. U.S.A. 82, 1989-1993.
12. Schindler, C.W. and Rush, M.G. (1985). J. Mol. Biol. 181, 161-173.
13. Jones, R.S. and Potter, S.S. (1985). Nucl. Acids Res. 13, 1027-1042.
14. Krolewski, J.J., Bertelsen, A.H., Humayun, N.Z., and Rush, M.G. (1982). J. Mol. Biol. 154, 399-415.
15. Georgiev, G.P. (1984). Eur. J. Bioch. 145, 203-220.
16. Frischauf, A.-M., Lehrach, H., Poustka, A.-M., and Murray, N. (1982). EMBL replacement vectors. EMBL Lab Manual.
17. Rigby, P.W., Dieckmann, M., Rhodes, G., and Berg, P. (1977). J. Mol. Biol. 113, 237-251.
18. Vieira, J. and Messing, J. (1982). Gene 19, 259-268.
19. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
20. Chen, E.Y. and Seeburg, P.H. (1985). DNA 4, 165-170.
21. Hardison, R.C. and Printz, R. (1985). Nucl. Acids Res. 13, 1073-1088.
22. Michel, F. and Lang, B.F. (1985). Nature 316, 641-643.
23. Nevins, J.R. (1983). Ann. Rev. Bioch. 52, 441-466.
24. Ciliberto, G., Castagnoli L., and Cortese, R. (1983). Curr. Topics in Dev. Biol. 18, 59-88.
25. Flavell A. and Ish-Horowicz, D. (1983). Cell 34, 415-419.
26. Arkhipova, I.R., Gorelova, T.V., Ilyin, Y.V., and Schuppe, N.G. (1984). Nucl. Acids Res. 12, 7533-7548.
27. Flavell, A.J. (1984). Nature 310, 514-516.
28. Flavell, A.J. (1981). Nature 292, 591-595.
29. Shiba, T. and Saigo, K. (1983). Nature 302, 119-124.
30. Ono, M. and Ohishi, H. (1983). Nucl. Acids Res. 11, 7169-7179.
31. Vizard, D.L. and Yarsa, J. (1985). Nucl. Acids Res. 13, 473-484.
32. Gilbert, W. (1985). Sci. 228, 823-824.
33. Varshavsky, A. (1981). Proc. Natl. Acad. Sci. U.S.A. 78, 3673-3677.
34. Krolewski, J.J. and Rush, M.G. (1984). J. Mol. Biol. 174, 31-40.