MOLECULAR AND CELLULAR BIOLOGY, Mar. 1982, p. 302-307
0270-7306/82/030302-06$02.00/0

Vol. 2, No. 3

Properties of Dispersed, Highly Repeated DNA Within and Near the Hamster CAD Gene

RICHARD A. PADGETT,† GEOFFREY M. WAHL,‡ AND GEORGE R. STARK*

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

Received 10 August 1981/Accepted 28 October 1981

Dispersed, highly repeated DNA sequences were found within and near the Syrian hamster gene coding for the multifunctional protein CAD. Most of the repeated sequences were homologous to each other and had similar properties. They hybridized to many cytoplasmic polyadenylated RNAs and to 7S and 4.5S cytoplasmic non-polyadenylated RNAs. Cloned DNA fragments containing repeated sequences were transcribed in vitro by RNA polymerase III. The repeated sequences from Syrian hamsters share many properties with the Alu family of repetitive DNA from humans. The hamster sequences were homologous to total repetitive human DNA but only very weakly homologous to two cloned members of the human Alu family.

The finding of Britten and Kohne (3) that the genomes of many eucaryotic organisms contain large amounts of repetitive DNA has stimulated much work on the properties of this class of DNA, but nothing is known yet about its function. A similar pattern of interspersion has been found for unique and moderately repetitive sequences in many species (4), and much is now known about the properties of repetitive sequences in human and rodent genomes. The dominant class of dispersed repetitive human DNA, called the Alu family because most of its members contain an Alu I restriction site (12), is comprised of sequences about 300 bases long repeated about 300,000 times. The Alu family has homology to the dominant repeated DNA family in Chinese hamsters, to repeated sequences in heterogeneous nuclear RNA, to the 4.5S RNAs of rodents, and to the 7S cytoplasmic RNA of humans (14, 20). Many members of the Alu family also contain transcription units active in vitro with RNA polymerase III (5, 6, 14).

In the course of studying the organization of the Syrian hamster gene that encodes CAD, the multifunctional enzyme catalyzing the first three steps of pyrimidine nucleotide biosynthesis (15), we found that DNA within and around the gene contained repeated sequences homologous to many polyadenylated [poly(A)$^+$] RNAs (16). We now show that these sequences are homologous to small non-polyadenylated [poly(A)$^-$] RNAs, that they have activity in in vitro transcription by RNA polymerase III, and that some of them are related to the human Alu family of repeated DNA.

MATERIALS AND METHODS

Electrophoresis of RNA and transfer from polyacrylamide gels. Total cytoplasmic RNA, isolated from SV28 Syrian hamster cells as described previously (18) but without heparin, was electrophoresed through an 8% polyacrylamide gel containing 5 M urea and 40 mM Tris-borate buffer (pH 8.3) until the bromphenol blue marker had traveled 25 cm. The gel was then washed for 30 min in water containing 1 μg of ethidium bromide per ml, followed by three 10-min washes in 25 mM sodium acetate buffer (pH 4.0). The gel was then photographed under UV light and placed in an electrophoretic destaining apparatus (E-C Apparatus Corp.) with a sheet of diazobenzyloxymethyl (DBM) paper (1) on the anode side of the gel. RNA was transferred to DBM paper by electrophoresis for 8 h at 1.5 A in 25 mM sodium acetate buffer (pH 4.0) at room temperature. Experiments with ^{32}P-labeled RNA indicated that this procedure removed from the gel all tRNA, 5S RNA, and 5.8S RNA and about 75% of the 7S RNA. Most of the RNA removed from the gel was bound to the DBM paper.

Electrophoresis, transfer from agarose gels, and detection of RNA and DNA. RNA was extracted from murine leukemia virus (MuLV) (kindly supplied by Inder Verma) with phenol and chloroform, denatured with glyoxal, run in a 1.5% agarose gel, and transferred to DBM paper (18). DNA was run in 0.7% agarose gels, stained with ethidium bromide, and transferred to DBM paper. DNA, nick-translated with [α-^{32}P]dCTP, was hybridized to RNA or DNA immobilized on DBM paper in the presence of 10% dextran sulfate 500 (18, 19).

In vitro transcription of cloned DNA. Supercoiled DNA plasmids were transcribed in reaction mixtures containing 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) buffer (pH 7.6), 0.6 mM

† Present address: Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139.

‡ Present address: Tumor Virology Laboratory, The Salk Institute, La Jolla, CA 92037.

each of GTP, ATP, and UTP, 60 μM CTP, 5 mM $MgCl_2$, 40 mM $(NH_4)_2SO_4$, 5 μCi of [α-^{32}P]CTP (400 Ci/mmol; Amersham), and 20 μg of DNA per ml. Each 50 μl contained 15 μl of a *Drosophila melanogaster* nuclear extract prepared from KC cells, kindly supplied by C. Parker (C. Parker and D. Hogness, manuscript in preparation). Reactions were carried out for 60 min at room temperature and terminated by adding EDTA to 20 mM and Sarkosyl to 0.5%. The mixtures were extracted with 1 volume of phenol-chloroform (1:1) and precipitated with 2.5 volumes of ethanol. The pellets, resuspended in 10 μl of 80% formamide–40 mM Tris-borate buffer (pH 8.3), were heated at 68°C for 5 min and loaded onto a 25-cm 6% polyacrylamide gel containing 5 M urea and 40 mM Tris-borate buffer (pH 8.3). The gel was run until the bromophenol blue ran off the end and was then dried and exposed to Dupont Cronex 4 film for 14 h at room temperature.

RESULTS

Location of repeated sequences. A 30-kilobase sequence of DNA from the region of the CAD gene has been cloned from Syrian hamster cells resistant to *N*-(phosphonacetyl)-L-aspartate, in which this region is amplified over 100-fold (16, 18). Fragments of the CAD gene which contain repeated sequences, determined by hybridization to total nick-translated Syrian hamster DNA, are shown in Fig. 1. Since the 5′ end of mature CAD mRNA is within fragment 206 and the 3′ end is within fragment 102 (16), it is clear that some repeated sequences must be located within the transcription unit, probably within intervening sequences only. For example, fragment 205 has no homology to mature CAD mRNA and must contain only intervening sequence DNA. Electron microscopy of heteroduplexes between CAD mRNA and cloned genomic DNA (16) showed that there are at least three copies of a repeated sequence in the large intervening sequence which includes fragment 205 plus portions of fragments 206 and 202. One repeat is in an inverted orientation relative to the other two since either a hairpin or a stem-and-loop structure can form (not shown). The repeated sequences in fragment 103 also appear to be in the large intervening sequence that comprises most of this fragment, since a deletion of 1.75 kilobases within the intervening sequence occurs frequently when a phage containing this DNA is grown in bacteria, presumably due to recombination between two copies of a repeated sequence in tandem orientation. Other copies of repeated sequence are found beyond the 5′ and 3′ ends of the region coding for the mature message. Most, but not all, of the repeated sequences that hybridize to total hamster DNA also hybridize to a probe prepared from fragment 101 (see below).

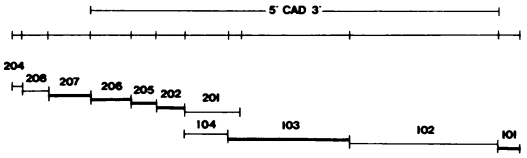
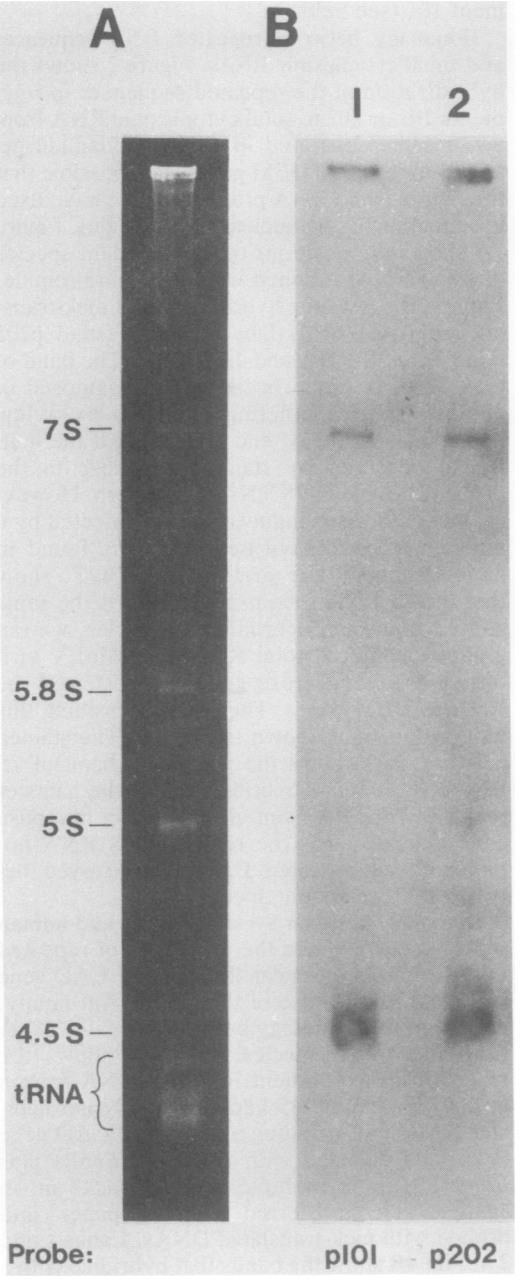


FIG. 1. Cloned fragments from the region of the CAD gene. The fragments shown as thick lines contain repeated sequences detected by hybridization to total hamster DNA, and those shown as thin lines do not. The top line shows the region from which mature CAD mRNA is derived (16).

Homology between repeated DNA sequences and small cytoplasmic RNAs. Figure 2 shows the hybridization of the repeated sequences in fragments 101 or 202 to total cytoplasmic RNA from SV28 cells, separated in a polyacrylamide gel and transferred to DBM paper. It is possible that the cytoplasmic RNA preparation we have used is contaminated by nuclear RNA species. Figure 2A shows the positions of the abundant species of small RNAs, stained with ethidium bromide. Figure 2B shows the hybridization of nick-translated plasmid p101 (lane 1) and plasmid p202 (lane 2) to 7S RNA and 4.5S RNA. The band of 4.5S RNA is broad because it is composed of multiple species differing in length by a few nucleotides at the 3′ end (10, 11). All the 4.5S RNAs observed by staining hybridize to the probe. The role of 7S RNA is unknown. However, its synthesis is induced in cells infected by a number of oncornaviruses, and it is found in large amounts in the viral particles (7). To show that the 7S RNA seen in SV28 cells is the same as that found in oncornavirus particles, we ran glyoxal-denatured total RNA from MuLV virions in a 1.5% agarose gel and transferred the RNA to DBM paper. The result of probing this blot with p101 is shown in Fig. 3B. The stained gel (Fig. 3A) shows the prominent band of 7S RNA. The intense hybridization of the hamster probe to 7S RNA from virions grown in mouse cells indicates that the region of 7S RNA homologous to repeated DNA is conserved between the two rodent species.

Homology between Syrian hamster and human repeated DNA. Since the properties of repeated DNA found in and near the hamster CAD gene were similar to those of the human Alu family, we tested for homology between repeated DNA from these two species. Figure 4 shows the result of an experiment in which DNA from a cosmid containing 45 kilobases of Syrian hamster DNA and including a functional CAD gene (16a) was digested with *Eco*RI or *Eco*RI plus *Bam*HI, run in multiple sets of tracks in an agarose gel, transferred to DBM paper, and probed with nick-translated DNAs. Lanes 1 and 2 in Fig. 4B show the bands that hybridize with a

p101 probe, lanes 3 and 4 show the bands that hybridize with a probe from total SV28 DNA, and lanes 5 and 6 show the hybridization with a probe from repetitive human DNA (kindly supplied by P. Deininger, produced by annealing the DNA to a low C_0t value and then treating it with S1 nuclease [12]). Most of the bands that hybridize with the total hamster probe also hybridize with p101. The intense bands in lanes 1 and 2 are from the pBR322 sequences present in both p101 and the cosmid, or are due to self-hybridization of p101. Hybridization of total hamster DNA to more bands indicates that there may be sequence divergence within a single family or that there may be more than one family of repeated sequences. Fragment 206 contains a repeated sequence which does not hybridize with fragment 101 under the conditions employed but does hybridize with other fragments that contain repeated sequences homologous to 101 (not shown). Hybridization of the human repeated DNA indicates that some repeated sequences are conserved between these two organisms. However, when BLUR 2 and BLUR 8 (17), two representative cloned members of the Alu family, were used to probe blots identical to those in Fig. 4, there was very little homology to sequences other than pBR322 (not shown), implying either that the hybridization seen with total repetitive human DNA is due to repeated sequences other than those of the Alu family or that only some members of the Alu family hybridize with the hamster repeated sequences. When these blots were washed under conditions more stringent than usual, with 15 mM NaCl and 1.5 mM sodium citrate at 60°C, the human probe was removed almost completely, whereas the hamster probe remained hybridized (not shown), suggesting that the human and hamster repeated sequences have diverged substantially.



Transcription in vitro of various CAD gene fragments. Since it has been shown that many members of the Alu family contain RNA polymerase III transcription units active in vitro, we tested our CAD gene fragments (as plasmids cloned in pBR322) in a system produced from nuclei of cultured *Drosophila melanogaster* cells. Figure 5 shows the transcripts produced by most of the DNA fragments from the region of the CAD gene. The plus and minus signs above the lanes indicate whether or not the DNA used in the reactions contained repeated sequences. Although this is a highly heterologous system, similar results were obtained using

FIG. 2. Homology between repeated DNA sequences and small cytoplasmic RNAs. Total cytoplasmic RNA from SV28 cells (40 μg per track) was run in multiple tracks in a 6% polyacrylamide gel and transferred electrophoretically to DBM paper. (A) Gel stained with ethidium bromide. (B) Hybridization of two plasmids containing repeated sequences to the immobilized RNA. The strips were probed with 10^7 cpm of nick-translated p101 (lane 1) or p202 (lane 2) at a specific activity of 5×10^7 cpm/μg for 12 h in the presence of 10% dextran sulfate, and washed at room temperature in 0.1× standard saline citrate. Exposure was for 5 days at −70°C with Dupont Cronex 4 film and a Dupont Lightning-Plus intensifying screen.

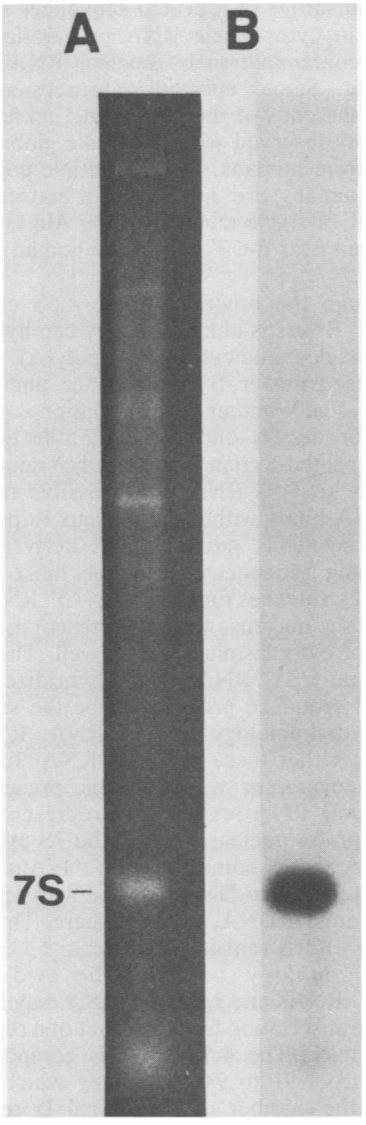


FIG. 3. Homology between repeated DNA sequences and 7S RNA from MuLV virions. Glyoxal-denatured RNA extracted from MuLV virions was run in a 1.5% agarose gel and transferred to DBM paper. (A) Gel stained with ethidium bromide before transfer. (B) Hybridization with 10^6 cpm of nick-translated p101. Washing and autoradiographic exposure for 2 days were as in Fig. 2.

a whole-cell extract from HeLa cells for in vitro transcription (not shown). As can be seen, some of the plasmids containing repeated sequences (p205, p206, and p207; lanes f, g, and h) produced strong transcripts whereas others (p202, lane e) were inactive. Failure to be transcribed in vitro could be due to an intrinsic property of a particular sequence or to inactivation caused by cutting with the restriction endonuclease used to clone the fragment. The response of plasmid p101 (lane b) is weak and variable. Faint bands are also produced by some plasmids which do not contain repeated sequences. Most of the transcripts are longer than 300 bases. We estimate that the repeated sequences are only 150 to 300 bases long, as judged by the sizes of the hairpin and stem-and-loop structures seen in electron microscopy of CAD gene fragments. Therefore, most of the transcripts must extend beyond the boundaries of the repeated region.

DISCUSSION

The repeated DNA sequences found within and near the CAD gene of Syrian hamster cells share many properties with the dominant class of dispersed repetitive sequences found in human and rodent DNAs, including homology to poly(A)$^+$ RNAs and to 4.5S and 7S poly(A)$^-$ RNAs and the ability to produce transcripts in vitro with RNA polymerase III. The functional role, if any, of these repeated sequences and

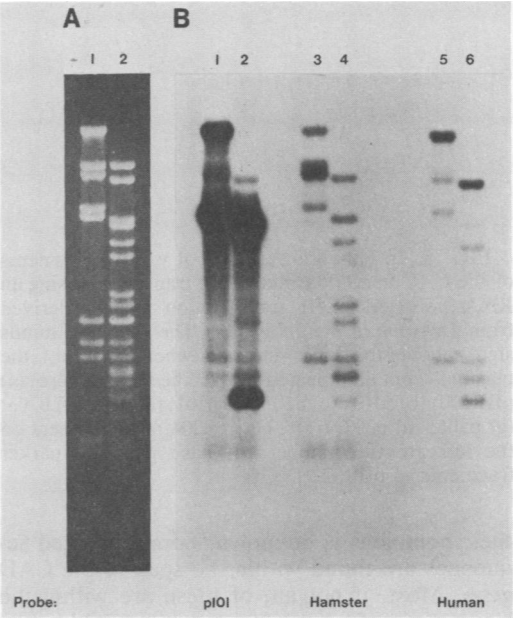


FIG. 4. Homology between human and hamster repeated DNAs. DNA (2 μg per track) from a cosmid containing a functional CAD gene was digested with *Eco*RI (lane 1) or *Eco*RI and *Bam*HI (lane 2), run in a 0.7% agarose gel, and transferred to DBM paper. (A) Gel stained with ethidium bromide before transfer. (B) Hybridization with nick-translated p101 (lanes 1 and 2), nick-translated total Syrian hamster DNA (lanes 3 and 4), or nick-translated total human highly repetitive DNA (lanes 5 and 6). Conditions were the same as for Fig. 2 except that dextran sulfate was not used. Autoradiographic exposure was for 20 h.

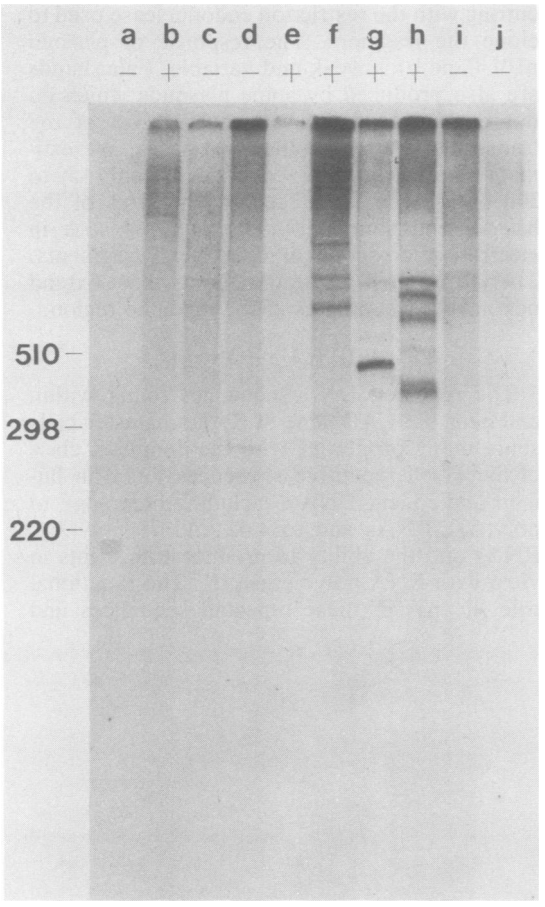


FIG. 5. In vitro transcription of various fragments of the CAD gene. Plasmids were transcribed using an RNA polymerase III transcription system derived from *Drosophila melanogaster*. The plus and minus signs above the lanes indicate whether or not the plasmids contain repeated DNA. The plasmids are: (a) pBR322; (b) p101; (c) p102; (d) p201; (e) p202; (f) p205; (g) p206; (h) p207; (i) p208; (j) p204. The numbers on the left are the numbers of base pairs in marker fragments of pBR322 DNA.

their homologs is unknown. Some repeated sequences are found within the span of the CAD gene. Most, if not all, of these are within the large intervening sequences and so would not be expected to be found in mature cytoplasmic CAD mRNA. In fact, the data of Padgett et al. (Fig. 6 of reference 16) show that p205, which contains DNA from intervening sequences and which does contain repeated DNA, does not hybridize to CAD mRNA. Many genes probably have repeated sequences within their intervening DNA since transcripts of these sequences are abundant in heterogeneous nuclear RNA but not in cytoplasmic mRNA (8). However, our hybridization experiments (16) do indicate that some transcripts of repeated sequences may be present in cytoplasmic RNA preparations, although contamination by nuclear RNA is not ruled out. Some repeated sequences are just beyond the ends of the CAD gene, an arrangement also observed with the β-like globin gene cluster from humans, where multiple copies of Alu sequences are interspersed between the genes (9). A single member of the Alu family is also found near the 3′ end of the human insulin gene (2).

Although the function of the 4.5S class of poly(A)$^-$ RNAs is unknown, they can hybridize to both nuclear and cytoplasmic poly(A)$^+$ RNAs in Chinese hamster, Syrian hamster, and mouse cells (11, 13). Whether this association is physiological or occurs during isolation is unclear. Since repeated Syrian hamster DNA sequences hybridize to 4.5S RNA, it is possible that the 4.5S RNA binds within the nucleus to portions of heterogeneous nuclear RNA derived from intervening sequences. Although most of the repeated sequences found in poly(A)$^+$ RNAs are found in the nucleus, some are present in preparations of cytoplasmic RNA as well. The cytoplasmic poly(A)$^+$ RNAs which hybridize to the repeated sequence probes may be the same as those with homology to 4.5S RNA. It is not known whether these poly(A)$^+$ RNAs function as messengers, or whether their presence in cytoplasmic RNA preparations represents contamination by nuclear RNAs. The 7S cytoplasmic RNA shares homology with the Alu family of repeated human DNA (20) and with repeated Syrian hamster DNA, as shown here. The function of 7S RNA remains obscure.

The homology between the two small poly(A)$^-$ RNAs and repeated DNA may be due to their transcription from a subset of a repeated DNA family. The 4.5S RNA is composed of species from 91 to 94 bases long which differ only in the number of 3′-terminal U residues (10), so it is possible that they could be coded in their entirety by repeated DNA segments of the size we have found. This possibility could be tested by sequence analysis, as has been done for repeated DNA sequences from Chinese hamsters (14). The 7S RNA is too large to be encoded entirely within the repeated sequences, but some members of the family may be large enough to form a complete transcription unit, or they may be associated with other functional sequences. In this regard, note that the RNA polymerase III transcripts produced in vitro from fragments containing repeated sequences are all much larger than the size of the repeat units. This is also the case for the transcripts made in vitro from Alu repeats near the β-like globin genes of humans (6). The 5′ ends of these transcripts are from an entire Alu repeat, and the

3′ ends are from non-Alu sequences that are different in the two transcripts studied. The studies of Duncan et al. (6) indicated that the Alu repeats contain the signals for RNA polymerase III to initiate transcription but lack a termination signal, so that transcription continues into the adjacent DNA until such a signal is found. A similar situation could exist for the hamster repeats studied here, as suggested by the fact that multiple bands are produced often during transcription, as though several termination sites of differing efficiency can be used.

Some repeated sequences from humans show homology to the hamster sequences, although the hybrids are not as stable as those formed between the hamster repeats themselves. In view of the similarities between the properties of the human Alu family and the hamster repeats, it is surprising that two cloned members of the Alu family which are very homologous to the consensus Alu sequence (17) do not show any significant homology with the hamster sequences. The human sequences may represent either another class of repeated DNA or divergent members of the Alu family.

Repeated sequences can cause deletions when pieces of DNA containing multiple copies of them are cloned in *Escherichia coli*. Both λ-100 and λ-200 undergo deletions with high frequency, probably between repeated sequences oriented in the same direction. Deletions are more frequent in *recA*$^+$ than in *recA E. coli* strains. Most DNA fragments subcloned into pBR322 and grown in the *recA* strain HB101 are relatively stable, but we have not been able to clone the intact 103 fragment in pBR322 in numerous attempts, indicating that deletions can occur even in *recA* hosts.

ACKNOWLEDGMENTS

We thank Carl Parker for supplying the extracts used in the RNA polymerase III transcription experiments, Inder Verma for supplying the MuLV, and Prescot Deininger and Ted Friedmann for supplying the human repeated DNA and the BLUR clones and for many helpful discussions concerning the Alu family.

This investigation was supported by Public Health Service grant CA-17287 from the National Cancer Institute to G.R.S. and by Public Health Service grant GM-27754 from the National Institutes of Health to G.M.W.

LITERATURE CITED

1. **Alwine, J. C., D. J. Kemp, and G. R. Stark.** 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U.S.A. **74**:5350–5354.
2. **Bell, G. I., R. Pictet, and W. J. Rutter.** 1980. Analysis of the regions flanking the human insulin gene and sequence of an Alu family member. Nucleic Acids Res. **8**:4091–4109.
3. **Britten, R. J., and D. E. Kohne.** 1968. Repeated sequences in DNA. Science **161**:529–540.
4. **Davidson, E. H., G. Galau, R. Angerer, and R. J. Britten.** 1975. Comparative aspects of DNA organization in metazoa. Chromosoma **51**:253–289.
5. **Duncan, C., P. A. Biro, P. V. Choudary, J. T. Elder, R. R. C. Wang, B. G. Forget, J. K. de Riel, and S. M. Weissman.** 1979. RNA polymerase III transcriptional units are interspersed among human non-α-globin genes. Proc. Natl. Acad. Sci. U.S.A. **76**:5095–5099.
6. **Duncan, C. H., P. Jagadeeswaran, R. R. C. Wang, and S. M. Weissman.** 1981. Structural analysis of templates and RNA polymerase III transcripts of Alu family sequences interspersed among the human β-like globin genes. Gene **13**:185–196.
7. **Erikson, E., R. L. Erikson, B. Henry, and N. R. Pace.** 1973. Comparison of oligonucleotides produced by RNase T1 digestion of 7S RNA from avian and murine oncornaviruses and from uninfected cells. Virology **53**:40–46.
8. **Federoff, N., P. K. Wellauer, and R. Wall.** 1977. Intermolecular duplexes in heterogeneous nuclear RNA from HeLa cells. Cell **10**:597–610.
9. **Fritsch, E. F., R. M. Lawn, and T. Maniatis.** 1980. Molecular cloning and characterization of the human β-like globin gene cluster. Cell **19**:959–972.
10. **Harada, F., and N. Kato.** 1980. Nucleotide sequences of 4.5S RNAs associated with poly (A)-containing RNAs of mouse and hamster cells. Nucleic Acids Res. **8**:1273–1285.
11. **Harada, F., N. Kato, and H. Hoshino.** 1979. Series of 4.5S RNAs associated with poly (A)-containing RNAs of rodent cells. Nucleic Acids Res. **7**:909–917.
12. **Houck, C. M., F. P. Rinehart, and C. W. Schmid.** 1979. A ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. **132**:289–306.
13. **Jelinek, W., and L. Leinwand.** 1978. Low molecular weight RNAs hydrogen-bonded to nuclear and cytoplasmic poly(A)-terminated RNA from cultured Chinese hamster ovary cells. Cell **15**:205–214.
14. **Jelinek, W. R., T. P. Toomey, L. Leinwand, C. H. Duncan, P. A. Biro, P. V. Choudary, S. M. Weissman, C. M. Rubin, C. M. Houck, P. L. Deininger, and C. W. Schmid.** 1980. Ubiquitous, interspersed repeated sequences in mammalian genomes. Proc. Natl. Acad. Sci. U.S.A. **77**:1398–1402.
15. **Padgett, R. A., G. M. Wahl, P. F. Coleman, and G. R. Stark.** 1979. N-(Phosphonacetyl)-L-aspartate-resistant hamster cells overaccumulate a single mRNA coding for the multifunctional protein that catalyzes the first steps of UMP synthesis. J. Biol. Chem. **254**:974–980.
16. **Padgett, R. A., G. M. Wahl, and G. R. Stark.** 1982. Structure of the gene for CAD, the multifunctional protein that initiates UMP synthesis in Syrian hamster cells. Mol. Cell. Biol. **2**:293–301.

16a. **Robert de Saint Vincent, B., S. Delbruck, W. Eckhart, J. Meinkoth, L. Vitto, and G. M. Wahl.** 1981. Cloning and reintroduction into animal cells of a functional CAD gene, a dominant amplifiable genetic marker. Cell **27**:267–277.

17. **Rubin, C. M., C. M. Houck, P. L. Deininger, T. Friedmann, and C. W. Schmid.** 1980. Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. Nature (London) **284**:372–374.
18. **Wahl, G. M., R. A. Padgett, and G. R. Stark.** 1979. Gene amplification causes overproduction of the first three enzymes of UMP synthesis in N-(phosphonacetyl)-L-aspartate-resistant hamster cells. J. Biol. Chem. **254**:8679–8689.
19. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. **76**:3683–3687.
20. **Weiner, A. M.** 1980. An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. Cell **22**:209–218.