Proc. Natl. Acad. Sci. USA
Vol. 75, No. 6, pp. 2679–2683, June 1978
Biochemistry

Inverted repeated DNA from Chinese hamster ovary cells studied with cloned DNA fragments

(double-stranded heterogeneous nuclear RNA/T1 fingerprints/Charon 16A λ phage/*Eco*RI/heteroduplex DNA)

WARREN R. JELINEK

The Rockefeller University, New York, New York 10021

Communicated by James E. Darnell, Jr., March 23, 1978

ABSTRACT **Fragments from the DNA of Chinese hamster ovary cells produced by restriction endonuclease *Eco*RI were cloned in Charon 16A λ bacteriophage and examined for the ability to hybridize *in situ* with ^{32}P-labeled double-stranded regions from heterogeneous nuclear RNA (hnRNA). Of 235 clones tested, 87 (37%) contained sequences that hybridized with the double-stranded hnRNA. Nine of these were examined for the presence of inverted repeat DNA structures (ir-DNA) by electron microscopy. All nine contained at least two elements of ir-DNA. Analysis of heteroduplexes formed from the DNAs of the different clones as well as T1 fingerprint analysis of the double-stranded hnRNA hybridized to each of the nine clones suggest that there is detectable nucleotide sequence homology in the various ir-DNAs. There are *ca* 3×10^5 ir-DNA pairs in the haploid Chinese hamster ovary cell genome.**

The nuclear DNA from various organisms contains inverted repeat DNA (ir-DNA) (1–6). The human haploid genome contains *ca* $1–4 \times 10^5$ such ir-DNA structures that can be isolated by treating denatured DNA with S1 nuclease (2, 6). Likewise, heterogeneous nuclear RNA (hnRNA) contains regions that can fold back to form double-stranded structures that are resistant to single strand-specific RNases (7–14). T1 fingerprints of the ds-hnRNA and of the *in vitro* transcripts of the ir-DNA from human cells are very similar, if not identical, to one another with an apparent sequence complexity of *ca* $1–5 \times 10^3$ nucleotides (13, 14); the size of these structures are *ca* 300 base pairs (5). If the ir-DNA and the DNA complementary to ds-hnRNA could be cloned, the extent of nucleotide sequence variation in these closely related sequences could be determined. Accordingly, DNA fragments produced by *Eco*RI digestion of the genome of Chinese hamster ovary cells (CHO) were cloned in Charon 16A λ bacteriophage (15) and screened for the presence of sequences that could hybridize with CHO ds-hnRNA.

This preliminary report describes nine cloned DNA fragments that contain sequences complementary to ds-hnRNA, each of which also contains ir-DNA. Electron microscopic analysis of cross hybridization between the cloned DNA fragments and fingerprint analysis of RNA hybridized to each suggest that there is considerable sequence homology, but not complete sequence identity, in the ir-DNAs in each of the nine clones.

MATERIALS AND METHODS

^{32}P-Labeled ds-hnRNA was isolated from CHO cells as described for the isolation of HeLa cell ds-hnRNA (6). Procedures for cloning *Eco*RI-generated DNA fragments in Charon 16 A λ phage were provided by F. Blattner and colleagues in an excellent laboratory "cookbook" along with the phage and bacterial strains used in its growth and testing. Phages carrying recombinant DNA molecules were screened for the presence of nucleotide sequences complementary to ds-hnRNA by the procedure of Benton and Davis (16). Heteroduplex DNA molecules were formed as described by Davis *et al.* (17) and examined in a Philips EM201 after staining in uranyl acetate and rotary shadowing with platinum/palladium. For hybridization to DNA the ds-hnRNA was denatured by boiling in distilled water, chilled, adjusted to 0.3 M in NaCl, 0.01 M in EDTA, 0.01 M in 2{[tris(hydroxy)methyl]amino}ethanesulfonic acid (pH 7.0), and 0.2% in sodium dodecyl sulfate, and incubated with DNA-bearing nitrocellulose filters at 65° for 18–20 hr. The filters were treated with RNase T1 at 37° in 0.3 M NaCl/0.03 M sodium citrate, pH 7.0, for 1 hr, the RNase was inactivated by incubating the filters in 0.15 M sodium iodoacetate/0.3 M NaCl, pH 5.3, at room temperature for 30 min and then at 55° for 45 min, and the RNA was eluted by boiling the filters in distilled water. T1 fingerprints were prepared according to Barrell (18). Phages with recombinant genomes were grown in liquid culture in *Escherichia coli* strain DP50SupF, an EK2 host/vector system, in a P3 physical containment laboratory.

RESULTS

Charon 16A λ phage DNA contains the structural gene for β-galactosidase within which is located the only *Eco*RI site in the genome (15). Phages containing a foreign DNA fragment inserted at the *Eco*RI site are unable to express β-galactosidase function and produce "nonblue" plaques when grown in *lac*$^-$ indicator bacteria on agar containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside; phages containing an uninterrupted β-galactosidase gene produce blue plaques (15).

CHO DNA was cleaved with *Eco*RI endonuclease and inserted into Chl6A λ phage DNA at the *Eco*RI site by ligation with T4 phage DNA ligase. This recombinant DNA was then used to transfect spheroplasts prepared from *E. coli* strain K802, mixed with *E. coli* strain χ7026 (*lac*$^-$), and plated on agar containing the galactoside derivative. The nonblue plaques were screened for the ability to form hybrids with ^{32}P-labeled ds-hnRNA *in situ* (16). Of 235 nonblue-plaque-producing phages tested, 87 (37%) formed detectable hybrids with the ds-hnRNA in the screening test. Nine of these 87 were chosen at random for further study.

Heteroduplexes were formed by hybridizing the DNA from each of the nine clones with that of the vector, Charon 16A, and examined by electron microscopy. Fig. 1 *A–I* shows representative molecules of each of the nine heteroduplexes, each of which contained an insertion loop that was predominantly

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

Abbreviations: ir-DNA, inverted repeat DNA; hnRNA, heterogeneous nuclear RNA; ds-hnRNA, double-stranded hnRNA; CHO, Chinese hamster ovary cells; kb, kilobases.

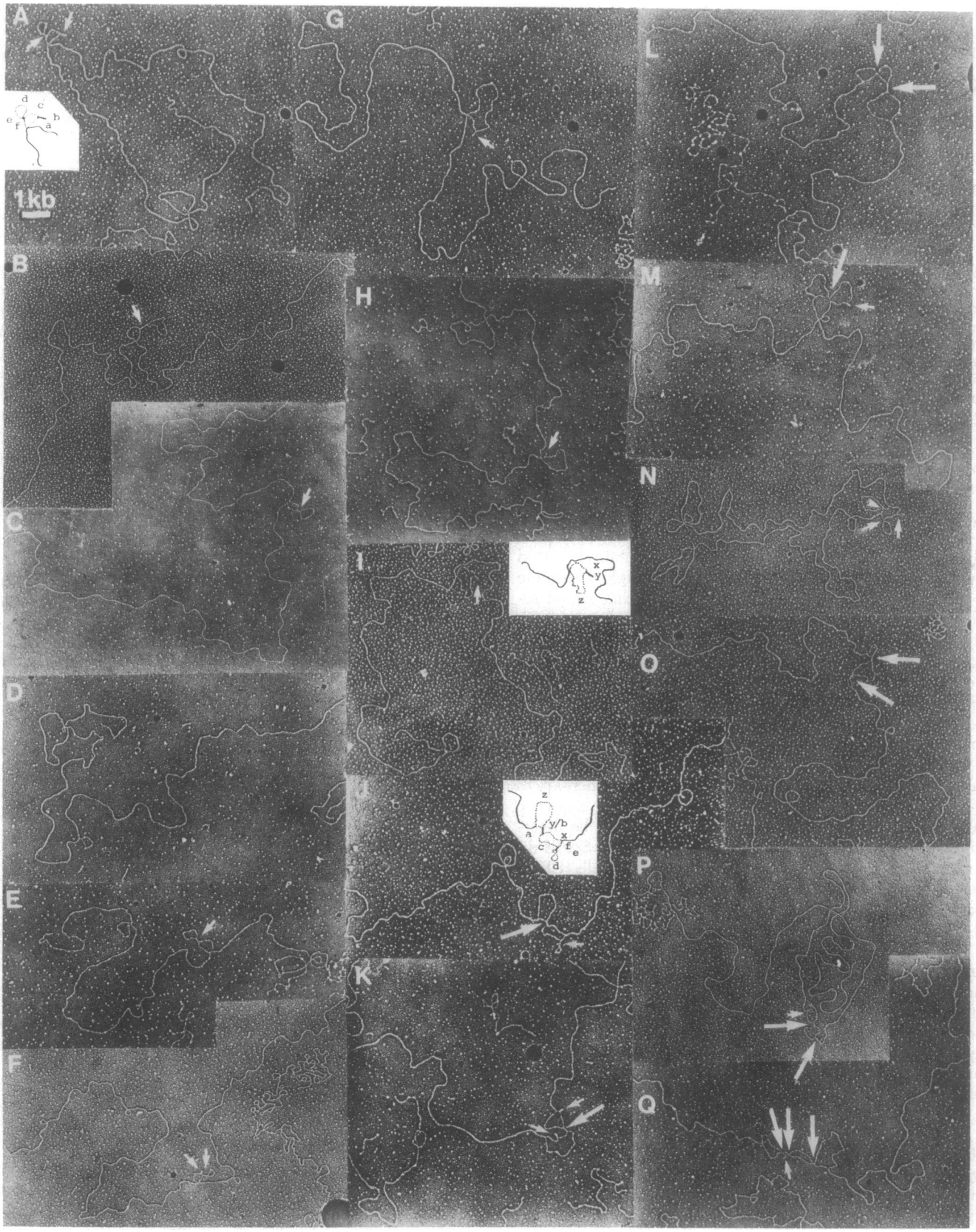

FIG. 1. Heteroduplex analysis of the nine clones described in the text. (*A–I*) Heteroduplexes formed between Ch 16A λ DNA and the recombinant DNA from each of the nine clones described in the text: (*A*) Chl6ACHOG49; (*B*) Chl6ACHOG235; (*C*) Chl6ACHOG73; (*D*) Chl6ACHOG25; (*E*) Chl6ACHOG63; (*F*) Chl6ACHOG2; (*G*) Chl6ACHOG104; (*H*) Chl6ACHOG200; (*I*) Chl6ACHOG95. (*J–Q*) The DNA from one clone (Chl6ACHOG49) was also hybridized with that from each of the eight others: (*J*) Chl6ACHOG95; (*K*) Chl6ACHOG200; (*L*) Chl6ACHOG63; (*M*) Chl6ACHOG73; (*N*) Chl6ACHOG235; (*O*) Chl6ACHOG25; (*P*) Chl6ACHOG2; (*continued at bottom of next page*)

single stranded. The nine DNA inserts ranged in size from *ca* 2.4 kilobases (kb) to 7.1 kb as verified by agarose gel electrophoresis (data not shown). For two of the nine clones, interpretive diagrams are given as *insets* in Fig. 1 (Fig. 1*A* for Chl6ACHOG49 and Fig. 1*I* for Chl6ACHOG95). The legend to Fig. 1 gives the proportional distances of the lettered regions shown in these two drawings. Eight of the insertions contained one or more regions of intramolecular double-stranded structure (ir-DNA). One (Chl6ACHOG25, Fig. 1*D*) contained what may be an ir-DNA with the two complementary DNA sequences near the CHO DNA/λ DNA junctures (not indicated by an arrow in Fig. 1*D*), but this structure never appeared large enough for a positive identification as an ir-DNA region by electron microscopy (see below for a further discussion of the structure of this clone). Two of the insertions, those in Chl6ACHOG49 (Fig. 1*A*) and in Chl6ACHOG2 (Fig. 1*F*) contained two regions of ir-DNA; in both, one ir-DNA structure had no distinguishable unpaired region between the ir-DNA elements and the other had a clearly distinguishable single-stranded region between the two ir-DNA elements. In clone 49 (for ease of reading the prefix "Chl6ACHOG" will not be used below), the two ir-DNAs were separated from one another by a single-stranded region, whereas in clone 2 the two were contiguous.

To determine whether the nucleotide sequences in the ir-DNAs of the nine clones would cross hybridize to one another in the regions of the ir-DNA, clone 49 was hybridized with the DNA of each of the other eight clones. The resulting heteroduplexes were examined by electron microscopy. Each heteroduplex showed a "substitution bubble" with one or two regions of cross-strand hybridization. Examples of the eight heteroduplexes are shown in Fig. 1 *J–Q*.

Fig. 1*J* shows a heteroduplex molecule formed from the DNAs of clone 49 and 95. To illustrate this heteroduplex and provide an example of how these molecules are interpreted, a drawing is given as an *inset*; the lettered regions correspond to those given in the *insets* to Fig. 1 *A* and *I* and the proportional distances for these regions are given in the legend. Clone 95 has an ir-DNA at "y" (see Fig. 1*I*) that is shown in Fig. 1*J* hybridized to the ir-DNA at "b" of clone 49; also visible in Fig. 1*J* is the ir-DNA of clone 49 (labeled "e") with its single-stranded "turnaround" at "d". Clone 200 (Fig. 1*H*) contains an ir-DNA that can also be seen in the heteroduplex to clone 49 (Fig. 1*K*); the ir-DNA at "b" in clone 49 is also visible. In addition, one of the elements of the "e" ir-DNA structure of clone 49 can be seen hybridized to a region of clone 200 that is not part of an ir-DNA in that cloned segment. Either this half ir-DNA sequence does not exist in close proximity to its complement in the same DNA strand or there was an *Eco*RI endonuclease site between the two complementary sequences in the original CHO DNA.

Clone 63 (Fig. 1*E*) cross-hybridized to clone 49 at two sites (Fig. 1*L*). The "b" ir-DNA of clone 49 was occasionally seen in heteroduplexes showing the two sites of cross-strand hybridization and in some instances this ir-DNA was also seen cross-hybridized with one of the ir-DNA elements of clone 63 (not shown). Clone 73 (Fig. 1*C*) cross-hybridized at least at one site with clone 49. In the heteroduplex shown in Fig. 1*M* the

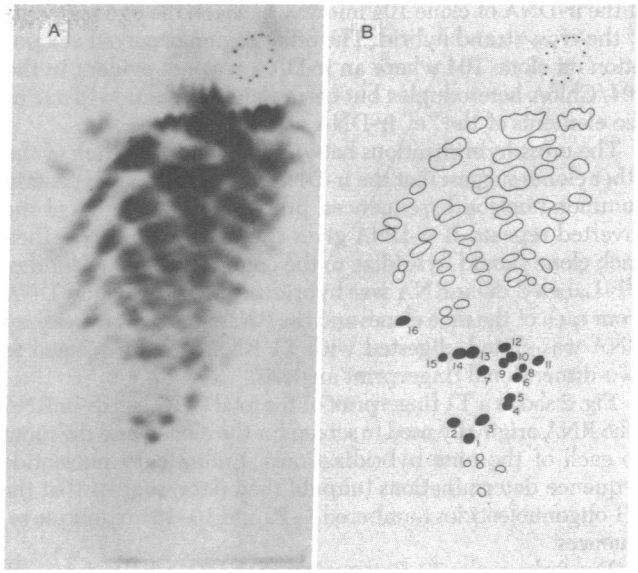

FIG. 2. T1 fingerprint of ds-hnRNA. (*A*) ^{32}P-Labeled CHO ds-hnRNA was digested with T1 RNase in 0.01 M Tris·HCl, pH 7.4/0.01 M EDTA, and the digestion products were separated in two dimensions as described by Barrell (18). Electrophoresis at pH 3.5 was from right to left and homochromatography was from bottom to top. (*B*) Tracing of *A*, indicating the positions of the major T1 products.

"b" ir-DNA of clone 49 can be identified as well as one cross-strand hybrid between one of the elements of the "e" ir-DNA of clone 49 and one of the elements of the ir-DNA of clone 73. What appeared to be a second site of cross-strand hybridization also occurred but heteroduplexes with two cross-strand hybrids appeared as "tangles" and were impossible to trace, presumably because the relative positions of the two elements of the ir-DNA in these two clones are reversed with respect to one another in the 5′-to-3′ direction. Whenever clone 235 (Fig. 1*B*) showed what could be considered as cross-strand hybridization to clone 49, the molecules appeared as tangles and could not be interpreted. Fig. 1*M* shows a heteroduplex between these two cloned DNAs in which the structures of each can be distinguished.

Although clone 25 (Fig. 1*D*) showed no clear-cut ir-DNA structure, apparently one element of each of the ir-DNAs of clone 49 formed a cross-strand hybrid with it, giving rise to a structure with three single-stranded "bubbles" separated by two regions of duplex (Fig. 1*O*). If there were two complementary ir-DNA elements, each located close to one of the two junctions between the λ DNA and the CHO DNA, they may have been missed in the heteroduplex formed between clone 25 and the Chl6A DNA because they would form a short "stem" immediately contiguous with the λ DNA. Clone 2 (Fig. 1*F*) formed two regions of cross-strand hybridization with the two elements of the "e" ir-DNA of clone 49 (Fig. 1*P*); the "b" ir-DNA of clone 49 is also visible. Clone 104 (Fig. 1*G*) showed two regions of cross-strand hybridization with clone 49 (Fig. 1*Q*). One region was broken into two parts separated by a portion

(*Q*) Ch16ACHOG104. Heteroduplex formation was accomplished as described by Davis *et al.* (17) and the heteroduplexes were spread from a hyperphase of 50% formamide/0.1 M Tris·HCl, pH 8.5/0.01 M EDTA onto a hypophase of 20% formamide/0.01 M ammonium acetate/0.01 M Tris·HCl, pH 8.5, and picked up onto Parlodion-coated copper grids. The bar in *A* represents 1 kilobase pair of double-stranded DNA. Small arrows, regions of intramolecular double-stranded DNA (ir-DNA); large arrows in *J–Q*, regions of cross-strand DNA·DNA duplexes. (*Insets*) These show the mean (±SD) proportional lengths of the indicated regions as follows. (*A*) a, 0.161 (±0.024); b, 0.155 (±0.034); c, 0.168 (±0.062); d, 0.341 (±0.051); e, 0.098 (±0.033); f, 0.096 (±0.021); 32 molecules measured. (*I*) x, 0.269 (±0.029); y, 0.096 (±0.031); z, 0.635 (±0.027); 29 molecules measured. (*J*) a, 0.151 (±0.041); b, 0.044 (±0.021); c, 0.294 (±0.042); d, 0.332 (±0.026); e, 0.098 (±0.031); f, 0.081 (±0.022); x, 0.330 (±0.044); y, 0.050 (±0.022); z, 0.629 (±0.027); 41 molecules measured.

of the ir-DNA of clone 104 midway between the two segments of the cross-strand hybrid. The other region occurred at a position on clone 104 where an ir-DNA was not evident in the 104/Chl6A heteroduplex but corresponds in position to one of the elements of the "e" ir-DNA of clone 49.

The cross-hybridizations between clone 49 and each of the other clones suggest that the ir-DNA in the nine clones contain common nucleotide sequences, probably in the region of the inverted repeats. If ir-DNA gives rise to the ds-hnRNA, then each clone should hybridize to the common oligonucleotides. ^{32}P-Labeled ds-hnRNA was hybridized to filter-bound DNA from each of the nine clones and the RNase-resistant hybridized RNA was eluted, digested with T1 RNase, and subjected to two-dimensional fingerprint analysis.

Fig. 2 shows a T1 fingerprint of the total CHO cell ds-hnRNA (the RNA originally used to screen for the clones, and the input to each of the nine hybridizations). Preliminary nucleotide sequence determinations (unpublished data) suggest that the T1 oligonucleotides numbered 1–10 and 13–15 are unique sequences.

Fig. 3 shows the T1 fingerprints of the RNA hybrids to each of the nine cloned DNAs. A number of similarities can be detected among the nine fingerprints and the total ds-hnRNA fingerprint shown in Fig. 2*A*. Seven of the nine (those of the hybrids to clones 49, 73, 200, 2, 63, 25, and 235) had T1 product 1, four (clones 49, 73, 104, and 95) had products 2 and 3, and all had products 4 or 5 (clone 73 had both). The identity of products 1, 2, and 3 in the fingerprints of each of the RNA hybrids containing them was confirmed by RNase U2 digestion and subsequent electrophoresis of the digestion products.

In addition to these T1 products, various other similarities in spot pattern could be detected in the nine fingerprints. For instance, compare the fingerprint of the hybrids to clones 73, 95, 63, and 235, all of which had a single ir-DNA structure detected by electron microscopy (Fig. 1). It should be emphasized that, in comparisons of fingerprints, similarities in spot patterns suggest similarities in nucleotide sequence, but differences do not necessarily mean great divergence in sequence. A single change in a guanosine residue will cause a profound change in the spot pattern (and other single base changes will also cause changes in spot pattern) even though all the other residues remain identical. Perfect sequence conservation of some of the larger T1 products (e.g., 1–5) in the various clones strongly suggest significant sequence similarity (but not necessarily identity) in the ir-DNAs at various sites in the CHO genome.

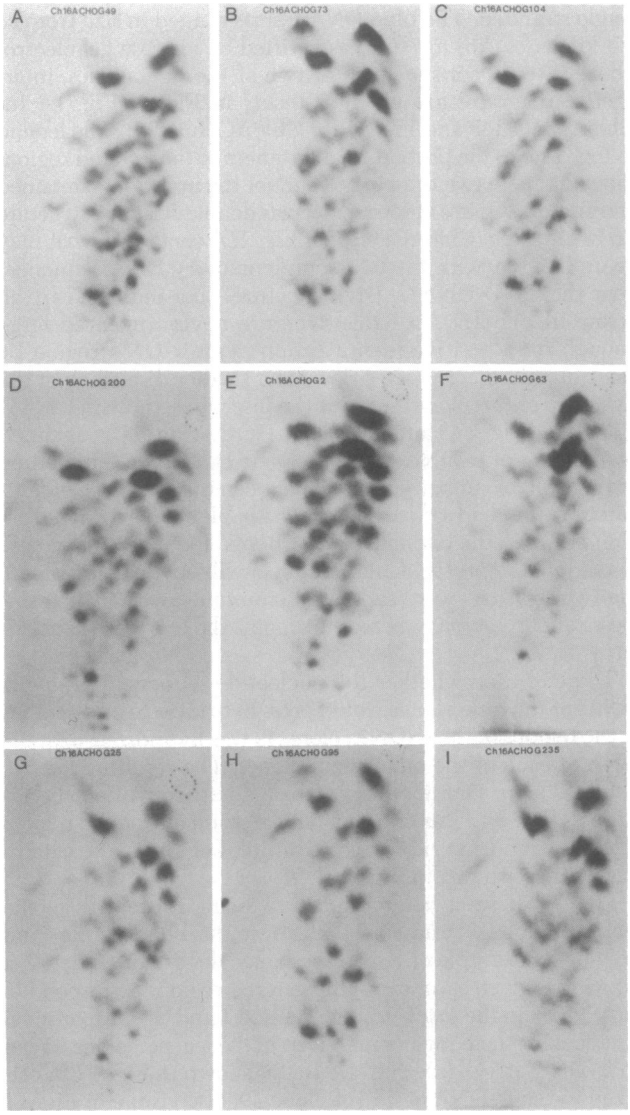

FIG. 3. T1 fingerprints of ds-hnRNA hybridized to DNA from each of the nine clones described in the text. ds-[^{32}P]hnRNA was melted by boiling in distilled water and hybridized to DNA immobilized on nitrocellulose filters from each of the nine Chl6ACHOG clones described (six filters each containing 50 μg of DNA). The filters were treated with T1 RNase (5 μg/ml) in 0.3 M NaCl, the nuclease was inactivated with sodium iodoacetate, and the RNA was eluted by boiling the filters in distilled water. The eluted RNA was digested with T1 RNase and fingerprinted as described by Barrell (18). Electrophoresis at pH 3.5 is from right to left, and homochromatography is from bottom to top. Clones: (*A*) 49; (*B*) 73; (*C*) 104; (*D*) 200; (*E*) 2; (*F*) 63; (*G*) 25; (*H*) 95; (*I*) 235.

DISCUSSION

Each of the nine cloned *Eco*RI fragments described here was chosen from among 87 phages whose recombinant genomes hybridized with ^{32}P-labeled ds-hnRNA in the Benton–Davis screening test (16). By electron microscopic analysis, eight of the nine contained at least one inverted repeat; two of the eight contained two inverted repeats (clones 49 and 2) and two others (clones 104 and 200) contained possibly one and a half such ir-DNA structures. One of the nine (clone 25) contained no unambiguous ir-DNA structure but, when hybridized with clone 49, it showed two regions of cross-strand hybridizations with two of the four ir-DNA elements identified in that clone. Perhaps it contains two direct repeats that are complementary to a portion of the ir-DNAs of clone 49. Five other clones that did not hybridize with ^{32}P-labeled ds-hnRNA *in situ* were also examined in the electron microscope but none were found to contain ir-DNAs.

The total length of the inserted DNA fragments in the nine clones described here is *ca* 40 kb. Twenty-four elements of ir-DNA were identified in the nine clones so that, on the average, there is one element of an inverted repeat for every 1.7 kb of DNA in these cloned fragments (40 kb per 24 ir-DNA elements). If the clones obtained were a random representation of all *Eco*RI fragments cut from the CHO genome, then the average distance between elements of ir-DNA in the total CHO DNA is *ca* 4.5 kb (37% of all clones hybridized with ds-hnRNA, the average number of ir-DNA elements per cloned DNA fragment was 2.7, and the average size of the cloned *Eco*RI fragments was 4.5 kb). Thus, there are *ca* 6×10^5 elements or 3×10^5 complete ir-DNA structures in one DNA strand of the CHO cell haploid genome (3×10^9 bases). There is a 2.6-fold difference in the frequency of ir-DNA elements detected in the

nine clones described and that calculated for the total genome. This may simply be due to nonrandom representation of ir-DNA elements in the nine clones, or it could mean that, throughout the CHO cell genome, when one ir-DNA element is present there is a higher-than-random probability that an average of 1.6 other ir-DNA elements exist in its vicinity; previously published data suggest that there is some clustering of ir-DNA in the human genome (2, 5).

There must be considerable nucleotide sequence similarity among the ir-DNA elements of the nine clones tested here and, therefore, in the ir-DNA at different sites in the CHO genome because many of the ir-DNA elements in various of the nine clones cross-hybridized with those of clone 49. Furthermore, the same T1 oligonucleotides that are diagnostic of CHO ds-hnRNA hybridized with the DNA of many of the nine clones. The nine fingerprints showed many similarities in T1 oligonucleotide pattern, but no two were absolutely identical. Presumably this means that there is considerable nucleotide sequence conservation but not complete sequence identity (sequence "drift") in the ir-DNA at different sites in the CHO genome. Determination of the complete nucleotide sequences for a number of ir-DNAs will allow a better estimate of the sequence drift from site to site and a more definitive conclusion concerning the number of sequence types in the ir-DNA class. It is assumed that the cross-strand hybridizations seen in the electron microscope (Fig. 1 *J–Q*) are a consequence of the similar oligonucleotides identified in these clones by the fingerprint data. However, it should be emphasized the present work does not prove this hypothesis.

The function of the ir-DNAs in eukaryotic genomes is unknown. They are transcribed as part of hnRNA molecules where they account for a substantial proportion of the RNA. In human cultured cells (HeLa cells), 27% of the hnRNA is composed of inverted repeat sequences (7). Perhaps they direct the folding of hnRNA molecules as part of the post-transcriptional processing events necessary for mRNA formation (7).

I acknowledge the excellent technical assistance of D. Wexler. I especially thank Dr. F. Blattner and colleagues who kindly provided the Charon 16A λ phage stock and the bacterial strains necessary for its growth and testing, as well as a superbly detailed laboratory "cookbook" describing the methods for cloning in this phage. This work was supported by a research grant from the National Science Foundation and by a Faculty Research Award from the American Cancer Society.

1. Davidson, E., Hough, D., Amenson, C. & Britten, R. J. (1973) *J. Mol. Biol.* **77**, 1–23.
2. Wilson, D. A. & Thomas, C. A. (1974) *J. Mol. Biol.* **84**, 115–144.
3. Schmid, C., Manning, J. & Davidson, N. (1975) *Cell* **5**, 159–172.
4. Schmid, C. W. & Deininger, D. L. (1975) *Cell* **6**, 345–358.
5. Deininger, D. L. & Schmid, C. W. (1976) *J. Mol. Biol.* **106**, 773–790.
6. Jelinek, W. R. (1977) *J. Mol. Biol.* **115**, 591–601.
7. Jelinek, W. R., Evans, R., Wilson, M., Salditt-Georgieff, M. & Darnell, J. E. (1978) *Biochemistry*, in press.
8. Montagnier, L. (1968) *C. R. H. Acad. Sci, Paris* **D267**, 1417–1420.
9. Ryskov, A. P., Farashyan, V. R. & Georgiev, G. P. (1972) *Biochim. Biophys. Acta* **262**, 568–572.
10. Kronenberg, L. H. & Humphreys, T. (1972) *Biochemistry* **22**, 2020–2026.
11. Jelinek, W. & Darnell, J. E. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2537–2541.
12. Ryskov, A. P., Saumders, G. F., Farashyan, V. R. & Georgiev, G. P. (1973) *Biochim. Biophys. Acta* **312**, 152–164.
13. Jelinek, W., Molloy, G., Fernandez-Munoz, R., Salditt-Georgieff, M. & Darnell, J. E. (1974) *J. Mol. Biol.* **82**, 361–370.
14. Robertson, H. D., Dickson, E. & Jelinek, W. (1977) *J. Mol. Biol.* **115**, 571–589.
15. Blattner, F. R., Williams, B. G., Blechl, A., Denniston-Thompson, K., Faber, H. E., Furlong, L-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L., & Smithies, O. (1977) *Science* **196**, 161–169.
16. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
17. Davis, R. W., Simon, M. & Davidson, N. (1971) in *Methods in Enzymology*, eds., Grossman, L. & Moldave, K. (Academic, New York), Vol. 21, pp. 413–428.
18. Barrell, D. (1971) in *Procedures in Nucleic Acid Research* eds., Cantoni, G. L. & Davis, D. R. Vol. 2, pps. 751–795.