Proc. Natl. Acad. Sci. USA
Vol. 77, No. 3, pp. 1398–1402, March 1980
Biochemistry

Ubiquitous, interspersed repeated sequences in mammalian genomes

(inverted repeats/double-stranded heterogeneous nuclear RNA/RNA polymerase III templates/low molecular weight RNA/DNA replication origins)

WARREN R. JELINEK*, THOMAS P. TOOMEY*, LESLIE LEINWAND*, CRAIG H. DUNCAN†, PAUL A. BIRO†, PRABHAKARA V. CHOUDARY†, SHERMAN M. WEISSMAN†, CAROL M. RUBIN‡, CATHERINE M. HOUCK‡, PRESCOTT L. DEININGER§, AND CARL W. SCHMID‡

*The Rockefeller University, New York, New York 10021; †Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06410; ‡Department of Chemistry, University of California, Davis, California 95616; and §Department of Pediatrics, University of California, San Diego, La Jolla, California 92093

Communicated by James E. Darnell, Jr., December 17, 1979

ABSTRACT **DNA base sequence comparisons demonstrate that the principal family of 300-nucleotide interspersed human DNA sequences, the repetitive double-strand regions of HeLa cell heterogeneous nuclear RNA, and specific RNA polymerase III *in vitro* transcripts of cloned human DNA sequences are all representatives of a closely related family of sequences. A segment of approximately 30 residues of these sequences is highly conserved in mammalian evolution because it is also present in the interspersed repeated DNA sequences of Chinese hamsters. Further DNA sequence comparisons demonstrate that a portion of this highly conserved segment of repetitive mammalian DNA sequence is similar to a sequence found within a low molecular weight RNA that hydrogen-bonds to poly(A)-terminated RNA molecules of Chinese hamsters and a sequence that forms half of a perfect inverted repeat near the origin of DNA replication in papovaviruses.**

A large portion of the DNA from most eukaryotic organisms is organized with alternating regions of unique and repeated sequences; the interspersed repeated sequences are characteristically described as having a size of about 300 base pairs (1–3). Approximately one-third of these interspersed repeated sequences in humans exist as inverted repeates of the form a b c . . . t . . . c′ b′ a′, in which a b c represents a sequence complementary to c′ b′ a′, and t, which can vary in length, represents the "turnaround", or non-base-paired sequence between the two intramolecular self-complementary sequences [inverted repeated DNA (ir-DNA)] (2, 3). Likewise heterogenous nuclear RNA (hnRNA) contains 300-nucleotide-long interspersed repeated sequences, some of which can exist as intramolecular pairs of inverted repeats (4).

These repeated sequences can be isolated from DNA or hnRNA as either intramolecular or intermolecular base-paired duplexes that are resistant to single-strand-specific nucleases (5–8). A fingerprint of *in vitro* RNA transcripts from human ir-DNA is found to be indistinguishable from a fingerprint of the RNase-resistant hnRNA isolated from growing cells, and thus it appears that the ir-DNA can serve as transcription templates for the RNase-resistant hnRNA (7). Prolonged self-annealing of the hnRNA allows a substantial fraction of its mass to become RNase resistant, and a fingerprint of the resistant RNA suggests that it is composed of only a few families of sequences (8).

Renaturation of human RNA to moderate C_0t values (C_0t = 68 mol·sec/liter) followed by S1 nuclease treatment reveals that approximately 5% of the genome can be isolated as 300-nucleotide base-paired fragments (5). When treated with the restriction endonuclease *Alu* I at least one-half of these repeated sequences gives rise to two fragments, one of approximately 170 base pairs and one of approximately 120 base pairs; the 300-nucleotide ir-DNA gives a similar cleavage pattern when treated in the same manner. Therefore, at least one-half of the 300-nucleotide C_0t 68 DNA fragments and the 300-nucleotide ir-DNA belong to a single sequence family (termed the *Alu* family) (5). Similar conclusions are suggested for the interspersed repetitive sequences in Chinese hamster cells (9). These repetitive sequences are estimated to be present approximately 300,000 times, or on the average once for every few thousand DNA base pairs throughout the human and Chinese hamster haploid genomes (5, 7, 9).

These observations imply that most DNA clones prepared from the genomes of mammalian cells will contain such sequences. A report by Duncan *et al.* suggests that DNA clones from the human β and β-like globin genes contain them (10). These authors describe two discrete RNA polymerase III transcription products synthesized *in vitro*, one from a DNA sequence approximately 1500 base pairs to the 5′ side of the $^G\gamma$-globin gene and the other from a DNA sequence approximately 1000 base pairs to the 5′ side of the δ-globin gene (10). The two DNA regions from which these RNAs are transcribed show nucleotide sequence homology to one another and to a third region of DNA located in inverted orientation downstream from the 3′ side of the β-globin gene. Thus, the region of the human genome bearing the β and β-like globin genes contains interspersed repeated sequences in both direct and inverted orientation that are potential examples of the *Alu* family of interspersed repeated DNA.

Jelinek and Leinwand observed a low molecular weight RNA approximately 100 nucleotides long that was isolated from other low molecular weight RNAs because it was base paired to nuclear and cytoplasmic poly(A)-terminated RNA molecules in Chinese hamster cells (11). On the average, each poly(A)-terminated nuclear RNA has one molecule of this low molecular weight RNA hydrogen bonded to it (11). The high abundance of this low molecular weight RNA in poly(A)-terminated hnRNA suggests that it might interact with sequences transcribed from the *Alu* family of interspersed repeated DNA because they are also present in hnRNA molecules in high abundance (8).

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

Abbreviations: hnRNA, heterogeneous nuclear RNA; ir-DNA, inverted repeated DNA; C_0t, initial concentration of DNA (moles of nucleotide/liter) × time (sec); BLUR, Bam Linked Ubiquitous Repeat.

As a result of nucleic acid sequence studies in our laboratories it has recently become apparent that there are nucleotide sequence similarities among (*i*) the *Alu* family of interspersed repeated sequences, (*ii*) regions of cloned DNA fragments containing mRNA coding sequences from both humans and rodents, (*iii*) regions of cloned DNA fragments that can serve as templates in the *in vitro* RNA polymerase III transcription system, (*iv*) regions in a low molecular weight RNA found hydrogen bonded to hnRNA from rodent cells, (*v*) the inverted repeated hnRNA from cultured human cells, and (*vi*) a sequence located at one of the junctions between the large intervening sequence and the coding sequence in the human β-globin gene. Furthermore, a comparison of these sequences with the DNA sequences of BK virus, simian virus 40, and polyoma virus reveals that they share homology with a sequence at or near the origins of DNA replication in these viral genomes. This homology suggests that the *Alu* family of interspersed repeated sequences might function as origins of DNA replication in mammalian cells. We describe here these nucleotide sequence similarities. It is not our intent to present the complete sequences of the fragments of interest to us nor the data that allow their derivation. These will be reported elsewhere in complete form by our respective laboratories.

RESULTS

The *Alu* Family of Interspersed Repeated Sequences. As reported (5), restriction fragments can be isolated from the *Alu* family of 300-nucleotide interspersed repeated DNA. This enabled us (C.M.R., C.M.H., P.L.D., and C.W.S) to attempt to determine the base sequence of this family of repeated DNA by use of the Maxam/Gilbert procedure (12). A possible problem in interpreting this sequence is that repeated DNA sequences such as the *Alu* family are not precise copies of a single sequence but instead are related copies of a "consensus" sequence. As measured by DNA duplex stability, it is established that the repeated human DNA sequences have diverged an average of 12% from their common ancestral sequence (13). This sequence divergence causes the appearance of a background heterogeneity in the Maxam/Gilbert sequencing ladders. In spite of this background it is possible to recognize the most prominant base at any position in the sequence. We assume that the most prominant sequence corresponds to an average sequence of the different members of the whole *Alu* family. A portion of this sequence is shown in Fig. 1.

The BLUR Clones. The reliability of this direct method of determining an average sequence of the *Alu* family may be questioned, and thus, as a control, we (C.M.R., C.M.H., P.L.D., and C.W.S.) also determined the base sequence of clones of individual members of the *Alu* family. These clones were prepared by ligating the 300-base pair interspersed repeated human DNA sequences with *Bam* I linkers and inserting these DNA fragments at the *Bam* site in plasmid pBR322 [Bam Linked Ubiquitous Repeat (BLUR)]. The nucleotide sequences of the BLUR clones agree closely with the consensus base sequence determination for the entire *Alu* family. Portions of the base sequence of one such clone, BLUR8, is shown in Fig. 1.

The existence of a consensus base sequence for the *Alu* family, which can be accurately determined by the Maxam/Gilbert method, confirms the previous conclusion that most of the 300-nucleotide interspersed repeated DNA sequences in humans belongs to a highly conserved, single family of sequences. This point will be considered below in light of additional base sequence comparisons to be made.

Repeated Sequences in Mammalian hnRNAs. Mammalian hnRNAs contain sequences that can form intramolecular or

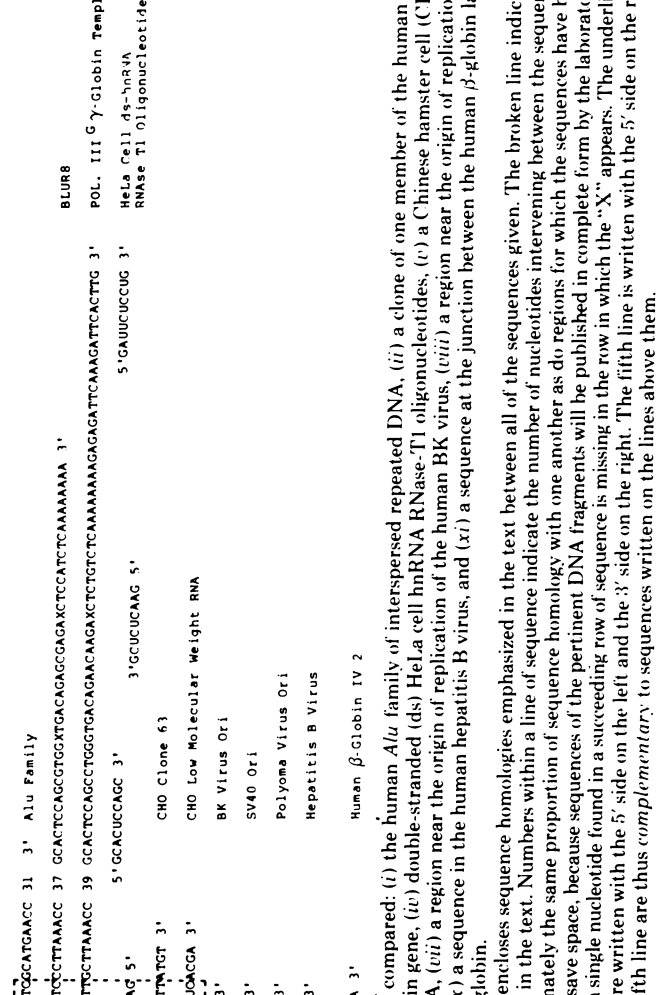

FIG. 1. Sequence comparisons. Portions of the following nucleotide sequences are compared: (*i*) the human *Alu* family of interspersed repeated DNA, (*ii*) a clone of one member of the human *Alu* family, BLUR8, (*iii*) a repeated sequence located at the 5′ side of the human $^{G}\gamma$-globin gene, (*iv*) double-stranded (ds) HeLa cell hnRNA RNase-T1 oligonucleotides, (*v*) a Chinese hamster cell (CHO) cloned genomic DNA fragment, (*vi*) a Chinese hamster cell low molecular weight RNA, (*vii*) a region near the origin of replication of the human BK virus, (*viii*) a region near the origin of replication of simian virus 40 (SV40), (*ix*) a region near the origin of replication of polyoma virus, (*x*) a sequence in the human hepatitis B virus, and (*xi*) a sequence at the junction between the human β-globin large intervening sequence and the sequence encoding amino acid residue 105 of human β-globin.

The sources for these sequence determinations are given in the text. The solid line encloses sequence homologies emphasized in the text between all of the sequences given. The broken line indicates sequence homologies between human and Chinese hamster DNAs and RNAs described in the text. Numbers within a line of sequence indicate the number of nucleotides intervening between the sequences given. Successive rows of sequences with numbers at the same position show approximately the same proportion of sequence homology with one another as do regions for which the sequences have been written out. The numbers were substituted for the actual sequences at these regions to save space, because sequences of the pertinent DNA fragments will be published in complete form by the laboratories that derived them. Occasionally an "X" is included in a row of sequence to indicate that a single nucleotide found in a succeeding row of sequence is missing in the row in which the "X" appears. The underlined sequences are described in the text. All sequences with the exception of the fifth line are written with the 5′ side on the left and the 3′ side on the right. The fifth line is written with the 5′ side on the right and the 3′ side on the left. The double-stranded hnRNA sequences indicated on the fifth line are thus *complementary* to sequences written on the lines above them.

intermolecular base-paired duplexes (4, 8). These sequences, which resist digestion by single-strand-specific RNases, account for as much as 27% of the mass of human hnRNA, but have an estimated sequence "complexity" of only a few thousand residues (8). The nucleotide sequences of six of the larger RNase-T1-generated oligonucleotides from this duplex hnRNA that have been determined are given in Fig. 1. The *Alu* family of repeated sequences contains all six. From this comparison we conclude that the *Alu* family of interspersed repeated DNA sequences is largely, if not exclusively, responsible for the simple repeated sequences observed in the double-stranded hnRNA. The agreement between six-oligonucleotide sequences of repetitive hnRNA and interspersed repeated DNA also supports the previous conclusions that most of the interspersed repeated DNA sequences and repetitive double-stranded hnRNA sequences are dominated by a single family of sequences (5–8).

Clone 63 from Chinese Hamster Cells. So far the discussion has centered on sequence comparisons from human DNA and RNA. The hnRNA from rodent cells also contains a simple family of repeated sequences that can form intramolecular and intermolecular base-paired structures (9). These sequences are also broadly distributed throughout the Chinese hamster genome. Jelinek (9) found that 37% of all Chinese hamster cloned genomic DNA fragments contain them. A number of these cloned DNA fragments contain ir-DNA, and one, clone 63, was also identified as containing coding regions for two poly(A)-terminated, polyribosome-associated RNAs in cultured Chinese hamster ovary cells (11). These RNAs map to a position on the cloned DNA between the ir-DNA elements (L.L. and W.R.J., unpublished results). A portion of the nucleotide sequence of this clone is given in Fig. 1. A sequence of approximately 30 nucleotides shares homology with the human *Alu* family of repeated sequences (sequences enclosed by the broken line in Fig. 1). Adjacent to this region there is little homology between the Chinese hamster and human sequences. The conservation of this limited sequence since the divergence of primates and rodents suggests its primary structure is functionally important. We also note that the sequence common to humans and Chinese hamsters has within it a region with homology to a sequence located at one of the junctions between the large intervening sequence and the coding sequence in the human β-globin gene (14) (last line of Fig. 1 should be compared with sequences enclosed by the broken line).

Low Molecular Weight RNA. A discrete low molecular weight RNA has been found hydrogen-bonded to Chinese hamster ovary cell poly(A)-terminated nuclear and cytoplasmic RNA (11); a similar RNA has been found hydrogen-bonded to mouse and rat cell poly(A)-terminated RNA (ref. 15; T.P.T. and L.L., unpublished results). In cultured Chinese hamster cells on the average one molecule of this low molecular weight RNA is hydrogen-bonded to each poly(A)-terminated hnRNA molecule (11). The high frequency of occurrence of this RNA in poly(A)-terminated molecules suggests that it might have sequences in common with ir-DNA. As evidence in favor of this suggestion, we (T.P.T., unpublished results) found by Southern blot analysis (16) that this low molecular weight RNA maps within or near the ir-DNA sequences in nine out of nine clones tested that contain ir-DNA sequences. Clone 63, described above, is one of this group of nine clones analyzed. The nucleotide sequence of this low molecular weight RNA has been determined (T.P.T) and a portion of it is given in Fig. 1. There is close homology between this RNA sequence and the region of homology between the human *Alu* family and the Chinese hamster DNA in clone 63. Outside this conserved region the sequence of the low molecular weight Chinese hamster RNA diverges from both the human and Chinese hamster repeated DNA sequences. We also note here that an RNA similar, if not identical, to the Chinese hamster low molecular weight RNA is found hydrogen-bonded to the genomes of spleen focus-forming virus and Moloney murine leukemia virus (ref. 17; unpublished results).

As has been emphasized (5, 8, 11), the observations reported so far are consistant with the idea that the *Alu* family of interspersed repeated sequences and its rodent equivalent function in RNA processing. However, two additional observations suggest other functions for these sequences.

***In Vitro* RNA Polymerase III Transcription Products from Cloned DNAs.** Duncan *et al.* (10) have recently demonstrated that interspersed repeated DNA sequences found in clones of human β and β-like globin genes may be specifically transcribed *in vitro* by RNA polymerase III into discrete products. We (C.H.D., R.C. Wang, and S.M.W.) have determined the DNA base sequence of one of these repeated sequences, which is located approximately 1500 nucleotides upstream from the 5′ end of the $^{G}\gamma$-globin gene. Portions of this sequence are shown in Fig. 1. It can be seen that there is extensive sequence homology between it and the *Alu* family of repeated sequences. The exact position of the 5′ end of the RNA transcribed from this DNA in the *in vitro* reaction is not known, but preliminary experiments indicate that it is within 30 base pairs of the short imperfect inverted repeat 5′-TCCCAGAATTTTGGGA-3′ (indicated by dotted underlining in Fig. 1). The sequence at this position is within the *Alu* family of 300-nucleotide repeats and corresponds to sequences previously determined by Jelinek and coworkers as major RNase T1 products of HeLa cell double-stranded hnRNA (6, 7). From this position the *in vitro* RNA polymerase III transcription product extends about 570 residues in the 3′ direction, and thus it also has extensive sequence homology to the *Alu* family. Although we compare only selected regions of these sequences in Fig. 1, we note that they share extensive homology that spans the 300 nucleotides we estimate as the length of the interspersed repeated sequences. We therefore regard each of this sequences as individual members of a related family of sequences, the *Alu* family.

Homology Between the Sequences Described Above and Viral Sequences. A search of the nucleotide sequences of the genomes of BK virus (18), simian virus 40 (19), and polyoma virus (20) reveals that these viral genomes contain a 14-nucleotide sequence having considerable homology with a nucleotide sequence found in the *Alu* family, the Chinese hamster cloned DNA fragment, and the low molecular weight RNA described above. (For ease of reference, these sequences will be called the 14-nucleotide common sequences.) It is particularly interesting to note that this sequence constitutes a portion of a perfect inverted repeat located at or near the origin of DNA replication in these viral DNAs (18–20). We also note that this sequence is present, with one inserted nucleotide, in the hepatitis B viral genome, but its inverted complement is not present (21). The origin of DNA replication for the hepatitis virus has not yet been located. These viral DNA sequences are included in Fig. 1 where they are enclosed by a continuous solid line. It is also of interest to note that the BK viral genome contains a 7-nucleotide sequence (CTACTTG) that precedes this common 14-nucleotide sequence by 11 nucleotides. This 7-nucleotide sequence is also found in the RNA polymerase III transcribed sequence and an almost identical sequence (CTACTTA) is found in the BLUR8 sequence immediately preceding the 14-nucleotide common sequence. In the simian virus 40 genome two nucleotide sequences of interest, one seven residues long (CTGGAAT) and one six residues long (CTACTT), precede the 14-nucleotide common sequence. These two oligonucleotide sequences are found in the region preceding the 14-nucleo-

tide-long common sequence in the BLUR8 sequence and in the RNA polymerase III transcribed sequence. They are underlined in Fig. 1.

DISCUSSION

A number of issues are raised by the sequence comparisons described here. It was implied (5) that the interspersed 300-nucleotide repeated sequences in humans are largely dominated by a single sequence family, the *Alu* family. The partial sequence of this repeated DNA reported here and the more complete sequence to be reported elsewhere support this conclusion. The *Alu* family contains only one major "consensus" sequence. This sequence is present at hundreds of thousands of different sites throughout the human genome; approximately one-third of these repeats occur as pairs of inverted repeats (3). Likewise, the Chinese hamster genome contains 300,000 copies of a sequence that also occurs as pairs of inverted repeats and appears to be the equivalent of the human *Alu* family (9). The human and Chinese hamster repeats share nucleotide sequences with one another (Fig. 1), and thus it seems likely that they evolved from a common ancestral sequence. We assume that these sequences serve the same function in both organisms.

The human *Alu* family and the Chinese hamster equivalent are heavily transcribed as part of hnRNA molecules in cultured cells (6, 8, 9). Whether these sequences have a functional role as part of hnRNA molecules is unknown. They may function only as DNA sequences, for instance, as RNA polymerase III transcription initiation sites, or as origins of DNA replication. Their presence within hnRNA molecules would then be a consequence of their presence within RNA polymerase II transcription units, which might traverse these other DNA functional units. On the other hand, they might have a function in hnRNA, perhaps as sequences involved in RNA processing. Whatever their function, it must be common to many sites in DNA, hnRNA, or both, because they are represented frequently in both of these molecules.

The human *Alu* family, its Chinese hamster equivalent, and the Chinese hamster low molecular weight RNA share a common sequence. The specific species of low molecular weight RNA described by Jelinek and Leinwand (11) cannot be transcribed from within the Chinese hamster *Alu* family equivalent because only a portion of its sequence is represented there, but must be transcribed elsewhere in the Chinese hamster genome. Because it contains sequences frequently represented in the Chinese hamster poly(A)-terminated RNA (i.e., transcripts of the Chinese hamster *Alu* family equivalent), it can interact with them in high molar proportions (11). It remains unknown whether this RNA·RNA interaction occurs within cells, or whether an RNA·DNA interaction occurs between members of the *Alu* family or its equivalent and the low molecular weight RNA. However, it is especially noteworthy that this low molecular weight RNA sequence resembles that region of the repeated sequence that has been so highly conserved in the mammalian genome.

The high frequency of occurrence of the human *Alu* family or its equivalent in the mammalian genome dictates its presence in the vicinity of structural genes. In the human β and β-like globin gene region a member of the *Alu* family is present at least once at the 5′ sides of the $^{G}\gamma$ and the δ genes and probably also at the 3′ sides of the $^{A}\gamma$ and the β genes (10). In the Chinese hamster genome it is present on either side of a structural gene of unknown function (11). Whether RNA polymerase III initiates transcription within the *Alu* family inside growing cells as it does in the *in vitro* transcription reaction is unknown (10). If so, it can potentially initiate within the hundreds of thousands of different *Alu* family members distributed throughout the mammalian genome. Perhaps either transcription of these repeated sequences by RNA polymerase III, or the interaction of these sequences with the low molecular weight RNA, or both, play a role in the production of structural gene transcripts. There is currently no available data to decisively test these intriguing possibilities.

The comparisons made in Fig. 1 suggest the 300-nucleotide interspersed *Alu* family of repeated sequences might be origins of DNA replication. There is excellent agreement among the 14-nucleotide sequence common to the human *Alu* family, the Chinese hamster clone 63, the Chinese hamster low molecular weight RNA, and the nucleotide sequence that comprises one-half of the perfect inverted repeat at or near the origin of DNA replication in the papovaviruses for which nucleotide sequence data is available. Furthermore, as indicated in *Results* this homology extends into a region immediately preceding this 14-nucleotide sequence. It seems unlikely that these homologies between the cellular repeated DNA sequences and the viral sequences near the origins of DNA replication arose by chance.

It is particularly noteworthy that the sequences in question here reside in the simian virus 40 genome at the major binding site for T antigen (22–24). This virus-encoded protein has been implicated in a number of functions including the initiation of viral DNA replication (25), the stimulation of cellular DNA synthesis (26), the regulation of its own synthesis (27, 28), the expression of late simian virus 40 genes (29), and the transformation and maintenance of the transformed phenotype in cultured cells (27, 30). The homology between the nucleotide sequence at the viral T antigen binding site and the *Alu* family of interspersed repeated DNA sequences suggests that T antigen may interact with cellular DNA at these repeated sequences. This interaction could in turn be responsible for some of the effects of T antigen on cell metabolism, including its stimulation of cellular DNA synthesis and its necessity for the initiation and maintainance of the transformed state.

Members of the *Alu* family of interspersed repeated sequences and its rodent equivalents may be the normal cellular DNA replication initiation sites. In mammalian cells DNA replication proceeds bidirectionally simultaneously from many sites (31), and thus the initiation sites for replication might be expected to be interspersed repeated sequences with two-fold rotational symmetry (32). The inverted repeated examples of the *Alu* family of interspersed repeated sequences and their Chinese hamster equivalents show these attributes. These considerations raise the question of whether the transcription of these repeated sequences by RNA polymerase III, or the interaction of these sequences with the low molecular weight RNA, or both, may play a role in the initiation of DNA replication.

P.D. thanks Prof. T. Friedmann (University of California, San Diego) for his generosity in allowing the construction of the BLUR clones in his laboratory. Some of the work reported here was supported in part by grants from the National Institutes of Health [GM 26704 (W.J.), GM 20121-07 (S.W.), and GM 21346 (C.S.)] and the National Science Foundation [PCM77 1729 (W.J.)]. C.H.D. was supported by fellowship F32CA06532-01 from the National Institutes of Health. P.D. was supported by National Institutes of Health Training Grant GM07199. C.H. is an American Association of University Women predoctoral fellow. W.J. is a recipient of a Faculty Research Award from the American Cancer Society.

1. Davidson, E. H., Galau, G., Angerer, R. & Britten, R. J. (1975) *Chromosoma* **51**, 253–289.
2. Schmid, C. W. & Deininger, P. L. (1975) *Cell* **6**, 345–358.
3. Deininger, P. L. & Schmid, C. W. (1976) *J. Mol. Biol.* **106**, 773–790.

4. Federoff, N., Wellauer, P. K. & Wall, R. (1977) *Cell* **10,** 597–610.
5. Houck, C. M., Rinehart, F. P. & Schmid, C. W. (1979) *J. Mol. Biol.* **132,** 289–306.
6. Robertson, H. D., Dickson, E. & Jelinek, W. (1977) *J. Mol. Biol.* **155,** 571–589.
7. Jelinek, W. R. (1977) *J. Mol. Biol.* **115,** 591–600.
8. Jelinek, W. R., Evans, R., Wilson, M., Salditt-Georgieff, M. & Darnell, J. E. (1978) *Biochemistry* **17,** 2776–2783.
9. Jelinek, W. R. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2679–2683.
10. Duncan, C., Biro, P. A., Choudary, P. V., Elder, J. T., Wang, R. C., Forget, B. G., DeRiel, J. K. & Weissman, S. M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5095–5099.
11. Jelinek, W. & Leinwand, L. (1978) *Cell* **15,** 205–214.
12. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
13. Deininger, P. L. & Schmid, C. W. (1979) *J. Mol. Biol.* **127,** 437–460.
14. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15,** 1157–1174.
15. Harada, F., Kato, N. & Hoshino, H. (1979) *Nucleic Acids Res.* **7,** 909–917.
16. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
17. Harada, F. & Ikawa, Y. (1979) *Nucleic Acids Res.* **7,** 895–908.
18. Yang, R. C. A. & Wu, R., (1979) *Science* **206,** 456–462.
19. Reddy, V., Thimmappaya, B., Dhar, R., Subramanian, K., Zain, S., Pan, J., Celma, M. & Weissman, S. (1978) *Science* **200,** 494–502.
20. Friedmann, T., La Porte, P. & Esty, A. (1978) *J. Biol. Chem.* **253,** 6561–6567.
21. Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. (1979) *Nature (London)* **281,** 646–650.
22. Reed, S. I., Ferguson, J., Davis, R. W. & Stark, G. R. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1605–1609.
23. Jessel, D., Landau, T., Hudson, J., Lalor, T., Tenen, D. & Livingston, D. (1976) *Cell* **8,** 535–545.
24. Tjian, R. (1978) *Cell* **13,** 165–179.
25. Tegtmeyer, P. (1972) *J. Virol.* **10,** 591–598.
26. Graessmann, M. & Graessman, A. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 366–370.
27. Tegtmeyer, P. (1975) *J. Virol.* **15,** 613–618.
28. Reed, S. I., Stark, G. & Alwine, J. C. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3083–3087.
29. Cowan, K., Tegtmeyer, P. & Anthony, D. D. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 1927–1930.
30. Kimura, G. & Dulbecco, R. (1972) *Virology* **52,** 529–534.
31. Huberman, J. A. & Riggs, A. D. (1968) *J. Mol. Biol.* **32,** 327–341.
32. Tamm, I., Jasny, B. R. & Cohen, J. E. (1979) in *Specific Eukaryotic Genes*, eds. Engberg, J., Klenow, H. & Leick, V. (Munksgaard, Copenhagen), pp. 221–226.