Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4382–4386, October 1977
Biochemistry

Sequence relationship between long and short repetitive DNA of the rat: A preliminary report

(long repetitive DNA/sequence homology/rat genome)

JUNG-RUNG WU, WILLIAM R. PEARSON, JAMES W. POSAKONY, AND JAMES BONNER*

Division of Biology, California Institute of Technology, Pasadena, California 91125

Contributed by James Bonner, August 8, 1977

ABSTRACT **Long and short repetitive sequences of rat DNA can be isolated and characterized. Long [>1.5 kilobases (kb)] sequences can be separated from short (0.2–0.4 kb) sequences by exclusion chromatography after renaturation of 4-kb DNA fragments to a repetitive C_0t and digestion with the single-strand-specific S1 nuclease. (C_0t is the initial concentration of DNA in mol of nucleotides/liter multiplied by time in sec.) Long repetitive DNA can be driven by an excess of whole rat DNA to measure its repetitive frequency. Excess long repetitive DNA can also be used to drive tracer quantities of either long (self-renaturation) or short repetitive DNA. Both the extent and the rate of the renaturations are found to be similar, suggesting that long and short DNA fragments share sequences. When long repetitive DNA is used to drive whole DNA tracers of various lengths, a 3.2-kb interspersion period is found. These data are consistent with the concept that short repetitive sequences are present within long repetitive DNA sequences in the rat genome.**

Recent studies on repetitive DNA sequence organization (1–4, †) have revealed that there are two size classes of repeated DNA sequences. In many organisms, interspersed sequences 0.2–0.4 kb long comprise 50–70% of the repeated DNA while the remainder of the repeated sequences are substantially longer, more than 1.5 kb (1 kb = 1000 base pairs or nucleotides) in length. The existence of two size classes of repeated sequences poses a number of interesting questions. First, are long and short sequences kinetically different? This question is answered by measuring the repetition frequency and complexity of purified long and short repeated sequences. A second question is perhaps more interesting: Do sequences that appear in short DNA sequences also appear in long sequences? This question is relevant to the "integrator gene" hypothesis (5). Long "integrator" sequences also present throughout the genome as short interspersed sequences suggest batteries of control elements as proposed by Britten and Davidson (5, 6).

In this paper we present data concerning the kinetic parameters of long and short repeated DNA sequences and we examine the sequence relationships between the two classes of sequences. Long and short repeated DNAs have been isolated by nuclease digestion and agarose A-50 fractionation. Long repeated DNA has been driven by whole DNA to determine repetition frequency, self-renatured to determine its complexity, and used to drive short repeated DNA to examine cross-renaturation. We do not find any significant kinetic difference between the long repeated DNA elements and all repeated rat DNA sequences by these criteria. In addition, there is evidence for some sequence homology between long and short repeated DNA sequences. A more sensitive experiment has been used to look for short sequences internal to long repeated DNA fragments. Long repeated DNA has been used to drive whole DNA tracers of various lengths to determine the interspersion period of these sequences in long whole DNA. Our data are consistent with the model that short repeated sequences are present in the long repeated sequence elements in the rat genome.

MATERIALS AND METHODS

Preparation of DNA. Unlabeled DNA was extracted from rat ascites cells and labeled DNA was extracted from Novikoff hepatoma cells.† DNA fragments 3–4 kb long were prepared by shearing DNA for 45 min at 7500 rpm in a Virtis 60 homogenizer (7) in 30 mM NaOAc (pH 6.8). DNA was sheared to 0.35 kb at 50,000 rpm in the Virtis 60 and in 66% glycerol (7). The DNA was then passed over Chelex 100 (Bio-Rad), filtered, and precipitated with EtOH.

Sizing DNA Fragments. Single-stranded DNA fragment lengths were determined by sedimentation through alkaline sucrose gradients. Isokinetic sucrose gradients (8) were formed in SW41 tubes in 0.1 M NaOH using a V_{mix} of 10.4 ml, C_{flask} = 16.0% (wt/vol), and C_{res} = 43% (wt/vol). Gradients were centrifuged from 16 to 24 hr at 40,000 rpm. All tubes contained two markers of known molecular weight and samples were run at least two times. Molecular weights were calculated from sedimentation rates by the Studier (9) equations.

Preparation of Long Repeated DNA Fragments. DNA sheared to an average length of 4 kb was denatured and incubated at 65° in 0.3 M NaCl/10 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (Pipes) at pH 6.8 to an equivalent C_0t of 5 using a factor of 2.31 for the reaction rate increase due to the Na^+ concentration. (C_0t is the initial concentration of DNA in mol of nucleotide/liter multiplied by time in sec.) After incubation, samples were diluted with an equal volume of 50 mM NaOAc/0.2 mM $ZnSO_4$ at pH 4.2, and dithiothreitol was added to a final concentration of 5 mM. The final reaction mixture was 0.15 M NaCl/5 mM Pipes/25 mM NaOAc/0.1 mM $ZnSO_4$/5 mM dithiothreitol at pH 4.4.

DNA samples were incubated with S1 nuclease at 37° for 45 min and the reaction mixture was chilled on ice and made 0.12 M in phosphate buffer. Duplex DNA strands were separated by hydroxyapatite chromatography, eluted with 0.5 M phosphate buffer, and chromatographed on Bio-Gel agarose A-50 (Bio-Rad) as described by Britten *et al.* (7).

The long unlabeled DNA used in the self-reaction and long/short cross-renaturation and interspersion experiments

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: kb, kilobase (1000 base pairs); C_0t, initial concentration of DNA in mol of nucleotide/liter multiplied by time in sec; Pipes, piperazine-*N*,*N*′-bis(2-ethanesulfonic acid).

* To whom reprint requests should be addressed.

† Pearson, W. R., Wu, J. R., and Bonner, J. (1977) *Biochemistry*, in press.

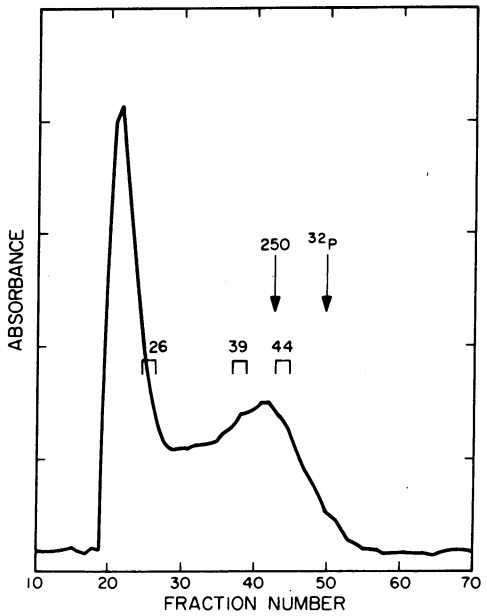

FIG. 1. Profile of rat repetitive DNA duplexes on agarose A-50. DNA sheared to 4 kb was denatured and renatured to C_0t of 5, digested with S1 nuclease, and bound to hydroxyapatite. The double-strand fraction (15%) was eluted with 0.5 M phosphate buffer and chromatographed on Bio-Gel agarose A-50. The size of the fraction indicated was determined by alkaline sucrose sedimentation. The fractions marked were used for the melting experiment in Fig. 2.

was isolated from 10 mg of 4-kb DNA. The DNA was denatured for 5 min at 100°, incubated to C_0t of 5 (13 min), and digested with 250 μl of an S1 nuclease preparation (the gift of Francine Eden). This nuclease preparation has been extensively characterized (10). The concentration used (25 μl/mg) corresponds to a digestion estimate of 0.85 which is 1.7 times the standard incubation. After hydroxyapatite chromatography, 17% of the DNA was found duplexed (30 A_{260} units), and this duplex DNA was passed over agarose A-50. Fifty-five percent of the DNA was excluded.

The ^{3}H-labeled DNA used as tracer in the repetition frequency and cross-renaturation experiments was prepared as above. In this preparation, the enzyme-to-DNA ratio was 25 μl/mg; 19% of the DNA was bound to hydroxyapatite, and 47% of the repetitive duplexes were excluded from agarose.

Melting. DNA samples were melted in 0.12 M phosphate buffer in a Gilford model 2400 spectrophotometer equipped with a model 2527 thermal cuvette. Samples were melted at a rate of 0.5°/min and the A_{260} was automatically sampled at 0.5° intervals. Hyperchromicity was calculated from the formula

$$H = \frac{A_{260}(98°) - A_{260}(60°)}{A_{260}(98°)}$$

after subtraction of the buffer absorbance at each temperature.

DNA·DNA Renaturation. Samples that were not to be digested by S1 nuclease were incubated in 0.12 M phosphate buffer at 60° or in 0.48 M phosphate buffer at 70°. After incubation, samples were frozen in dry ice/ethanol. Samples were thawed and diluted to 0.12 or 0.14 M phosphate buffer and passed over hydroxyapatite at 60°. The fraction bound was eluted after thermal denaturation at 100°. The fraction and rate parameters for the renaturation curves were calculated from a nonlinear least-squares program (11).

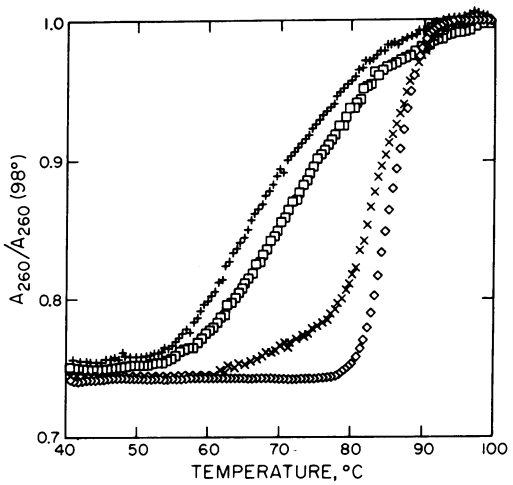

FIG. 2. Melting experiment of repetitive DNA duplexes of different fragment lengths. Repetitive DNA duplexes were isolated and fractionated as described in the legend of Fig. 1. Fractions 26, 39, and 44 were melted in a spectrophotometer equipped with a thermal cuvette. The temperature was raised at a rate of 0.5°/min to 98°. +, Fraction 44; □, fraction 39; ×, fraction 26; ◇, native DNA reference.

RESULTS

When 4-kb DNA is renatured to C_0t of 5 and digested with S1 nuclease and the duplexes are sized on agarose A-50, repetitive sequences are fractionated into two classes.† A typical column profile is shown in Fig. 1. Long (4 kb) fragments were used to minimize creation of short fragments by random shear and overlap of long repeated sequences. The incubation was carried out to C_0t of 5 to prevent renaturation of single copy sequences. (Long fragments have a higher effective C_0t because of the fragment length.) The long repetitive DNA excluded from agarose A-50 is an average of 1.5–2.0 kb long. The short included repetitive DNA is 0.2–0.5 kb long.

To make certain that long duplexes did not contain single-strand tails, we melted duplexes fractionated on A-50. A sample melting experiment of three fractions indicated in Fig. 1 is shown in Fig. 2. All fractions show more than 90% of native hyperchromicity. The melting temperatures range from 2.5° below native for the long (excluded) material to 15° below native for the short (0.25 kb) fragments.

Renaturation of Selected Repetitive Sequences with Excess of Whole DNA. ^{3}H-Labeled 0.35-kb fragments were renatured to C_0t of 100 and the duplexed repetitive sequences were separated on hydroxyapatite. This total repetitive fraction was then renatured with a 100- to 1000-fold excess of 0.35 kb of whole genomal DNA (repetitive plus single copy). The data are shown in Fig. 3*A*. The line drawn through the data displays the best least-squares fit. The dashed line above the data is the fit for whole genomal rat DNA.† Table 1 (part A) shows the results of a least-squares fit to the data. About half of the repetitive reaction takes place with a $C_0t_{1/2}$ of 0.337, corresponding to a repetition frequency of 7400. The remainder renatures with a $C_0t_{1/2}$ of 242, corresponding to a repetition frequency of 10.

Labeled long repetitive DNA elements (those excluded from A-50 agarose) were sheared to 0.35 kb and driven by 0.35-kb fragments of whole DNA to determine the repetition frequency of the long repetitive sequences. Fig. 3*B* shows the renaturation curve for the long repetitive sequences. Table 1 (part B) shows the results of a least-squares fit to the data. A large fraction of the DNA (60%) renatures by C_0t of 10 with a $C_0t_{1/2}$ of 0.153.

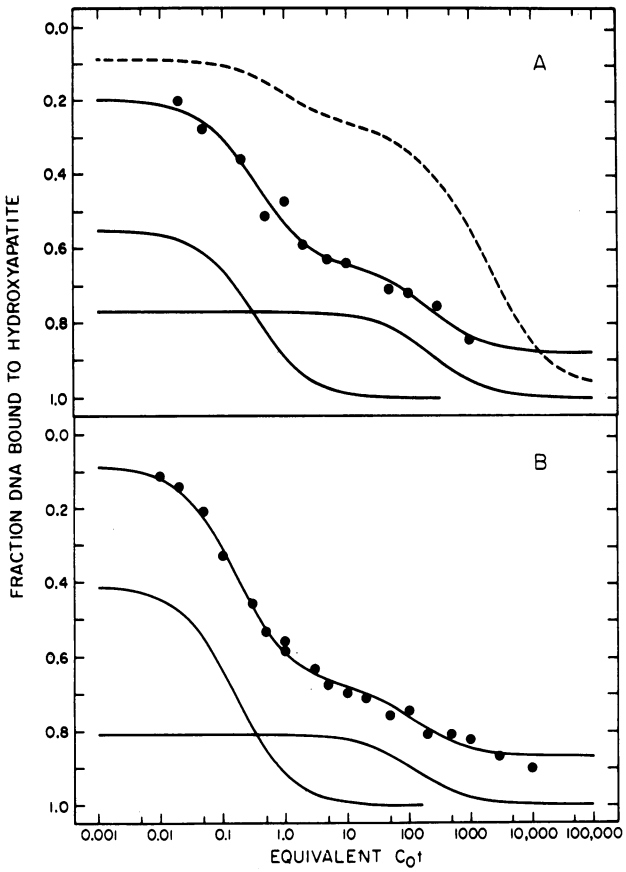

FIG. 3. Renaturation of selected repetitive sequences. (*A*) ^{3}H-Labeled rat DNA (350 nucleotides) was renatured to C_0t of 100 and the double-strand-containing fraction (30%) was separated on hydroxyapatite. This fraction was driven by unfractionated 350-nucleotide unlabeled rat DNA. The mass ratio of driver to tracer DNA was 100 from C_0t of 0.001 to C_0t of 0.5 and 1000 from C_0t of 1.0 to C_0t of 1000. The line drawn through the data displays the best least-squares fit. The two repetitive components are also shown. The dashed line above the data is the whole rat DNA fit. (*B*) ^{3}H-Labeled long repeated DNA fragments were isolated after fractionation on Bio-Gel agarose A-50 and sheared to 0.35 kb. These fragments were driven by a 100-fold excess from C_0t of 0.01 to C_0t of 0.1, by a 200- to 500-fold excess from C_0t of 0.2 to C_0t of 0.5, by a 1000-fold excess from C_0t of 1.0 to C_0t of 1000, and a 10,000-fold excess of unlabeled 0.35 whole rat DNA at C_0t greater than 1000 in 0.12 M phosphate buffer at 60°.

The remaining 20% of the DNA, which renatures by C_0t of 1000, contains less frequently repeated sequences, thus a lesser repeat class. This part of the reaction may also include some single-copy DNA sequences.

The measurements we have made on the repetition frequency and complexity of long repetitive DNA in the rat show little difference between the long repetitive sequences isolated from agarose A-50 and total (long plus short) repetitive sequences in whole DNA. The repetition frequency of long sequences is quite close to the repetition frequency of total repetitive sequences in whole DNA and the complexity of the long size class is also about the complexity expected for the fraction of sequences purified from whole DNA.

Renaturation of Selected Repetitive Sequences with Excess of Long Repetitive DNA. The complexity of the long repetitive DNA elements was determined by their self-renaturation kinetics, and the C_0t curve is shown in Fig. 4*A*. Again the DNA preparation was sheared to 0.35 kb to exclude length effects. Table 2 (part A) presents the fit of the data. Our best

Table 1. Renaturation of isolated repetitive sequences with a large excess of whole DNA

Component	Fraction	Rate	$C_0t_{1/2}$	Approximate repetition frequency
		A. Total repetitive DNA		
1	0.45	2.97	0.337	7,400
2	0.23	0.00413	242	10
Final fraction unreacted: 0.12				
Goodness of fit: 3.4%				
		B. Long repetitive DNA		
1	0.62	6.52	0.153	16,000
2	0.21	0.0140	71.4	35
Final fration unreacted: 0.15				
Goodness of fit: 2.2%				

estimate for the true fraction that repetitive sequences constitute at C_0t of 5 is 16%, so the long sequences (45%), whether found only with a class of elements, should represent a $1/(0.45 \times 0.16) = 14$-fold enrichment of sequences over whole DNA. However, the enrichment found [Table 2 (part A)] is only 8-fold [51.4 compared to 6.52 (Table 1 part B)]. Thus, the complexity found in the long repetitive DNA elements is twice that expected on the basis of its fraction of the genome. This suggests

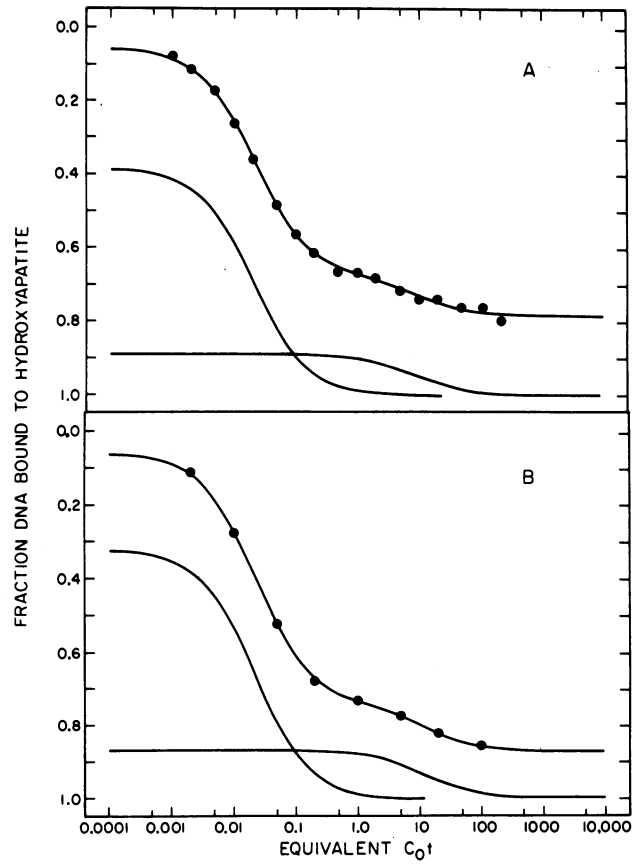

FIG. 4. Renaturation of long and short repetitive sequences with sheared long repetitive DNA fragments. (*A*) ^{3}H-Labeled and unlabeled long repetitive DNA was isolated as described in the legend of Fig. 3. The two DNA fractions were mixed and incubated to the C_0t values shown. (*B*) The unlabeled long repetitive fraction used in *A* was used to drive sheared short DNA fragments from the included peak in Fig. 1. The driver to tracer ratio was 10 for C_0t of 0.002, 20 for C_0t of 0.01, and 100 for C_0t greater than 0.5.

Table 2. Renaturation of long and short repetitive sequences by a large excess of long repetitive DNA

Component	Fraction	Rate	$C_0t_{1/2}$
A. Long repetitive sequences			
1	0.64	51.4	0.019
2	0.13	0.161	6.2
Final fraction unreacted: 0.23			
Goodness of fit: 1.0%			
B. Short repetitive sequences			
1	0.67	40.4	0.025
2	0.13	0.0768	13.0
Final fraction unreacted: 0.13			
Goodness of fit: 0.5%			

that some complexity of the long repeats may be shared with that of the short repeats.

A more direct way to look for cross homology between long and short repetitive sequences is to drive one preparation by the other. The results of an experiment in which sheared long repetitive DNA drove sheared short repetitive DNA are shown in Fig. 4*B*. The kinetics of the reaction are summarized in Table 2 (part B). The rates of renaturation of long and of short repetitive sequences driven by a vast excess of long repetitive DNA are the same within a factor of 2. Many sequences may be shared.

Short sequences were not used to drive long sequences because of expected cross-homology. An unknown fraction of the short DNA may be derived from the long elements by random mechanical shear. The kinetic complexity and crossreaction experiments are in fact sensitive to contamination of short repeated DNA with long repeated elements that have been sheared to short fragments. Such shear could happen in the initial preparation of the DNA or it could occur during the hydroxyapatite separation of repeated from single-copy sequences. Melting experiments (1, 12,†) suggest that some short repeated sequences are different from some sequences contained in long repeated elements. Short sequences exhibit more mismatch than do long sequences. But some short sequences must be derived from long sequences, and the melt of the short repetitive fragments shows a high precision component.†

We cannot estimate how many of the short sequences are derived from long sequences without knowing the *in vivo* length of the long repeated sequences in the genome. Some may be longer than the 4-kb fragments used in our experiments. The length of the long duplexes excluded from agarose A-50 is about 1.5–2.0 kb. These may include some molecules that were mechanically sheared during hydroxyapatite fractionation.

It is possible to put an upper limit on the amount of contamination by placing a lower limit on the number of short sequences using the rat interspersed data.† At C_0t of 5, 57% of 2.5-kb DNA fragments contain a short repeated DNA sequence. If the short repeated sequences are 0.25 kb long, 0.25/2.5 = 10% of the bound DNA is repetitive, so 5.7% of the DNA contains short repetitive sequences renaturing by C_0t of 5. At this C_0t, 0.16 of the genome is in true duplex, so 0.057/0.16 = 0.356 of duplex sequences must be short. We find 50–60% of duplex DNA included on agarose A-50, so 15–25% of that DNA or 15%/50% = 30% to 25%/60% = 42% of the short duplexes may be derived from long sequences. This is an upper limit. Many short sequences are much longer than 0.25 kb and others may be interspersed at a longer period. If 33% of short sequences are derived from long sequences, short sequences would drive all long sequences in a cross-reaction experiment. Conversely,

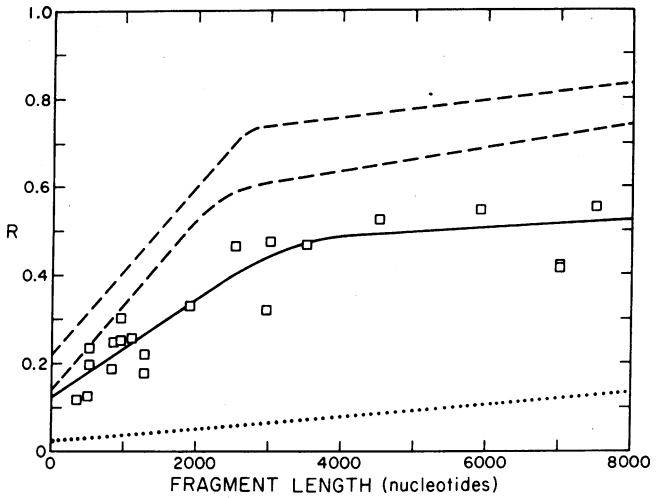

FIG. 5. The fraction of rat DNA (R) containing long repetitive DNA elements as a function of whole DNA fragment length. Labeled whole rat DNA fragments of various sizes were prepared and driven to C_0t of 0.5 (the lower C_0t was used to adjust for the 10-fold lower complexity of the repetitive driver) by the sheared long repeat DNA fraction used in Figs. 3 and 4. The solid line drawn to the data represents the best least-squares fit to the data. The dashed lines are interspersion curves with the labeled tracers driven by whole rat DNA at C_0t of 5 (middle curve) and at C_0t of 50 (upper curve). The dotted line at bottom represents the background hydroxyapatite binding of long tracers in the absence of drivers. The fraction of the fragments (F) containing duplexes was measured by binding to hydroxyapatite. The value plotted (R) is the value of F corrected for the amount of zero time binding in the long tracers (Z). The values plotted are: $R = (F - Z)/(1.0 - Z)$. The values of Z were calculated from a linear fit of the zero time binding data [Pearson, W. R., Wu, J. R., and Bonner, J. (1977) *Biochemistry*, in press].

30–40% of the short fragments would be driven by long sequences. According to the results of Table 2 (part B), the sharing of sequences between short and long is therefore more than simple contamination.

Interspersion of Long Repeated DNA Sequences. Long repetitive sequences should be relatively free of short repetitive sequence contamination. It is difficult to imagine a mechanical shear, renaturation, or nuclease artifact that would create long repeated sequences from short ones. We have used long repeated DNA to drive various lengths of whole DNA tracers to determine whether the long repeated DNA is able to renature the interspersed short repeats of whole DNA. Fig. 5 shows the reaction of sheared long repeated DNA as driver at C_0t of 0.5 with whole DNA tracers as a function of tracer length, corrected for zero time binding in the tracer. This experiment is similar to experiments done to determine the short repetitive sequence interspersion period in other organisms (2, 3, 13, 14,†). The dashed lines plotted in Fig. 5 are interspersion curves which can be drawn through data from whole DNA tracers of different lengths driven by whole rat DNA at C_0t of 5 and at C_0t of 50,† but we have used a particular class of repeated sequences (Fig. 5, solid line) instead of whole DNA as driver.

The interspersion of long repeated elements is similar to that of all repeated DNA sequences in the rat. The increase in hydroxyapatite-bound DNA from 12% at zero length to 32% at 1.8 kb cannot be due to a reaction between long repeated elements (sheared driver) and the complementary long segments in whole DNA tracer. The difference in the single-stranded tail length would only account for a 20% change (from 12.0 to 14.4%) if only long sequences (i.e., 1.5 kb) in a 1.8-kb whole DNA tracers had renatured. Some of the sequences in the long

repeated DNA preparation are apparently able to react with short sequences interspersed throughout the genome.

The 3.2-kb best-fit interspersion period is slightly longer than the 2.5-kb value found for whole short repeats in whole rat DNA.† It suggests that some long repetitive DNA elements are interspersed with a longer period than is found for the period of the short repeats.

DISCUSSION

The renaturation experiments we have presented all suggest that there are no significant differences in complexity or repetition frequency between most of the long and short repeated DNA sequences in the rat. Eden *et al.* (16) have measured the repetition frequency, complexity, and sequence overlap between long and short repeated sequences in the sea urchin. They found that the repetition frequencies of the long and short repetitive sea urchin sequences in whole DNA are similar, with the long sequences repeated slightly less frequently. They also found the kinetic complexity of the short sequences to be about 3 times that of the long DNA elements, reflecting the fraction of long sequences in the whole repetitive sequence population of the sea urchin (about 30%).

The cross-renaturation results—using long repetitive sequences to drive short repetitive sequences—in the sea urchin are quite different from our results with rat DNA. In the sea urchin, short repetitive tracer was driven at a rate 1/10 that of the long repetitive driver. This result suggests that 10% or fewer of the sequences in long repeated DNA elements can renature with short repeated DNA sequences. This result contrasts with the results presented for the rat; we find no difference between short repetitive tracer renaturation and the self-reaction of the long repetitive driver.

There are two possible explanations for these conflicting results. First, the higher fraction of long repeated element in rat DNA may cause more cross-contamination problems. Second, the internal structure of the long repeated elements of the rat DNA may be different from those of sea urchin.

Under virtually identical digestion and fractionation conditions, we find that 40% to more than 50% of rat repeated DNA sequences from 4-kb fragments are excluded on agarose A-50, while Eden *et al.* (16) found with sea urchin that about 30% of the sequences from 2-kb fragments were excluded. The calculations we presented earlier suggest that as much as 70% of rat repeated DNA may be longer than 1 kb. Britten *et al.* (10) showed that almost 47% of isolated repeated sequence duplexes of sea urchin DNA are longer than 1 kb (mild S1 nuclease digestion). The smaller the fraction of true short repeats in a genome, the more difficult will be the isolation of the short repeat fraction free of long repeat contamination. While it may be simple to show that two repetitive sequence populations are distinct, it is difficult to demonstrate unambiguously the degree of true sequence overlap between them.

Only a small fraction of the repeated DNA in higher organisms can be accounted for by sequences of known function. Some of the long repeated sequences must include the transcribed multigene families, such as ribosomal and histone genes (17–19), and must be distinct from short repeated sequences. Long repeated DNA sequences also form higher precision duplexes than do most of the short repeated DNA (1, 10, 12, †).

In sum, our evidence suggests that long and short repeated DNA fragments share sequences in the rat. Possible contaminant artifacts contribute to ambiguity as to the extent of this sharing, but we have not been able to provide any strong evidence that long and short repetitive sequence sets are different from one another in the rat.

If long and short repetitive sequences are shared in the rat, a number of structural and organizational questions are raised. Some short repeated sequences may be present as a single sequence within a long repeated sequence or long repeated sequences may be arranged as tandem arrays of short repeats. In addition, some long repeated sequences may be made up of a number of different repeated sequences that are also found in the genome as short interspersed repeated sequences. Either structural model could provide the "integrator gene" function suggested by Britten and Davidson (5).

We thank Dr. Francine Eden and Ms. Denise Painchaud for the gift of the S1 nuclease used in this study, Dr. Anthony Bakke for valuable discussion, and Mr. Bill Buchanan for his technical assistance. This research was supported in part by U.S. Public Health Service Training Grant GM00086 and in part by U.S. Public Health Service Research Grants GM13762 and GM20927.

1. Davidson, E. H., Graham, D. E., Neufeld, B. R., Chamberlin, M. E., Amenson, C. S., Hough, B. R. & Britten, R. J. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 295–301.
2. Angerer, R. C., Davidson, E. H. & Britten, R. J. (1975) *Cell* **6**, 29–39.
3. Efstratiadis, A., Crain, W. R., Britten, R. J., Davidson, E. H. & Kafatos, F. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2289–2293.
4. Wu, J. R., Pearson, W. R., Wilkes, M. & Bonner, J. (1977) in *The Molecular Biology of the Mammalian Genetic Apparatus*, ed. Ts'o, P.O.P. (Elsevier/North-Holland Biomedical Press, Amsterdam, The Netherlands), Vol. 2, pp. 51–62.
5. Britten, R. J. & Davidson, E. H. (1969) *Science* **165**, 349–357.
6. Davidson, E. H. & Britten, R. J. (1973) *Q. Rev. Biol.* **48**, 565–613.
7. Britten, R. J., Graham, D. E. & Neufeld, B. R. (1974) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K., (Academic Press, New York), Vol. 29, pp. 363–418.
8. Noll, H. (1967) *Nature* **215**, 360–363.
9. Studier, F. W. (1965) *J. Mol. Biol.* **11**, 373–390.
10. Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M. & Davidson, E. H. (1976) *J. Mol. Evol.* **9**, 1–23.
11. Pearson, W. R., Davidson, E. H. & Britten, R. J. (1977) *Nucleic Acids Res.* **4**, 1727–1735.
12. Goldberg, R. B., Crain, W. R., Ruderman, J. V., More, G. P., Barnett, J. R., Higgins, R. C., Gelfand, R. A., Galau, G. A., Britten, R. J. & Davidson, E. H. (1975) *Chromosoma* **51**, 225–251.
13. Davidson, E. H., Hough, B. R., Amenson, C. S. & Britten, R. J. (1973) *J. Mol. Biol.* **77**, 1–23.
14. Graham, D. E., Neufeld, B. R., Davidson, E. H. & Britten, R. J. (1974) *Cell* **1**, 127–137.
15. Schmid, C. W. & Deininger, P. L. (1975) *Cell* **6**, 345–358.
16. Eden, F. E., Graham, D. E., Painchaud, D. M., Davidson, E. H. & Britten, R. J. (1977) *Nucleic Acids Res.* **4**, 1553–1567.
17. Brown, D. D. & Sugimoto, K. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 501–505.
18. Birnstiel, M., Telford, J., Weinberg, E. & Stafford, D. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2900–2904.
19. Galau, G. A., Chamberlin, M. E., Hough, B. R., Britten, R. J. & Davidson, E. H. (1976) *J. Mol. Evol.* 200–224.