DNA SEQUENCE ORGANIZATION BY ELECTRON MICROSCOPY

Thesis by
Mahlon McGregor Wilkes

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1977

(Submitted May 24, 1976)

For AriaDne

Acknowledgements

I thank my thesis advisor Dr. James Bonner for his guidance and effervescence. During all phases of this work incisive and appreciated counsel was provided by Professors Eric Davidson and Norman Davidson. I am also grateful for the help I have received from Dr. Roy Britten, Dr. R.L. Russel and Dr. Max Delbrück. Mr. William R. Pearson graciously lent me his computer skills in the analysis of much of this work.

I am happy to acknowledge the financial assistance I received through a National Institutes of Health Pre-Doctoral Traineeship and a fellowship from the Arthur McCallum Fund.

Abstract

The size and arrangement of repetitive and inverted repeat (foldback) sequences in rat DNA were studied by visualization of hybrid and heteroduplex structures in the electron microscope. The self-reassociation of repetitive sequence-bearing DNA strands often results in the formation of four-ended "H" structures, whose duplex regions equal the repetitive sequence length and can be measured in the electron microscope. In this way, it was determined that the average size of the class of numerous short repetitive sequences is 400 ± 150 nucleotides. Heteroduplex structures were prepared between long total DNA single strands and short repetitive sequence-bearing strands. The analysis of these structures confirms that the size of the repetitive sequences is 400 nucleotides on average. Length measurements between adjacent duplexes show that the average spacing between two interspersed repeats is at least 1500-1800 nucleotides. Short and long period interspersion patterns of repetitive and unique sequences can be discriminated. By examining 29.4 kb single strands after brief renaturation the size and distribution of foldback sequences were determined. There are 1.9 x 10^5 foldback pairs per rat genome, spaced an average of 9.7 kb apart throughout at least 73% of the DNA. Repetitive, inverted repeat and unique sequences are interspersed with each other in at least half the genome.

Table of Contents

Introduction

Introduction

Interest in the problem of gene regulation has stimulated extensive studies into the sequence organization of eucaryotic genomes, and these studies were pursued largely by methods employing the physical chemical properties of nucleic acid reassociation, investigated in detail by Britten and Kohne (1968) and Wetmur and Davidson (1968). The model of eucaryotic gene regulation advanced by Britten and Davidson (1969) makes several predictions about the arrangement of repetitive and non-repetitive DNA sequences: among these are that repetitive sequences be interspersed with structural genes. Also because of the relative proportions of repetitive and non-repetitive sequences in animal genomes it is expected that the repeats be shorter than their neighboring structural gene sequences.

Analysis of the kinetics of reassociation of animal DNA has revealed the existence of three sequence components present in a wide variety of organisms: repetitive sequences comprising approximately 20-40% of the genome, unique sequences accounting for 60-80%, and rapidly renaturing DNA composing the 5-15% remainder. Repetitive sequences may exercise a control function in gene regulation (Britten and Davidson, 1969). However, it is noteworthy that at least some structural gene sets, namely the histone genes (Kedes and Birnstiel, 1971), are transcribed from repetitive sequences. Campo and Bishop (1974) found that a substantial proportion of mRNA molecules in rat cells is transcribed from repeated sequences.

Most structural genes are unique sequences (Davidson and Britten, 1973; Lewin, 1975). Moreover, several isolated structural genes have been shown to be non-repeated, e.g., those coding for hemoglobin (Bishop et al., 1972), silk fibroin (Suzuki et al., 1972), and ovalbumin (Sullivan et al., 1973).

The rapidly renaturing DNA can be composed either of highly reiterated DNA (Kram et al., 1972; Hennig et al., 1970) or inverted repeat (foldback) sequences (Britten and Smith, 1970; Wilson and Thomas, 1974).

A variety of physical chemical techniques have been used to obtain detailed information about the arrangement of repetitive, foldback and single copy sequences in eucaryotic DNA. An increased rate of reassociation among DNA strand populations of increasing length assayed by hydroxyapatite chromatography indicates the physical interspersion of repetitive and unique sequences. Thus, reassociation of DNA fragments a few thousand nucleotides in length to a repetitive C_0t results in HAP-bindable structures containing partly repetitive sequences in duplex and partly single-stranded non-repeated DNA "tails." The actual amount of duplex in such reassociated preparations can be determined by S1 nuclease digestion, hyperchromicity measurements, or electron microscopic analysis.

A quantitative estimate of the lengths of single copy sequences bounded by repetitive sequences in the interspersed portion of the genome can be made by computer analysis of the corrected HAP binding of various DNA fragment sizes as a function of strand length. From such lines of study it has been shown that on the order of one half of the genomes of Xenopus (Davidson et al., 1973; Chamberlin et al., 1975), sea urchin (Graham et al., 1974) and Aplysia (Angerer et al., 1975) are organized in a "short period interspersed" pattern of single copy sequences approximately 1000 nucleotides long bounded by repetitive sequences an average of 300 nucleotides in length. In the cases of Xenopus and sea urchin an additional pattern of "long period interspersion" comprising roughly one quarter of the genome has been identified. In this case the average size of the single copy sequences interspersed with repeats is at least several thousand nucleotides.

The present study approaches the sequence organization of rat DNA using the heteroduplex method in the electron microscope (Davis, Simon and Davidson, 1971). This method consists of reassociating partly complementary and partly non-complementary nucleic acids to form "heteroduplex" structures, which can be visualized in the electron microscope. Analysis of the structures yields information about the localization of specific sequences. Wu, Hurn and Bonner (1972) used this technique in a study of Drosophila

DNA repetitive sequences; however, because the repeats in this organism are unusually long (Manning et al., 1975) and because the DNA used was much smaller than the average repetitive sequence, this study appears to have underestimated the true average repetitive sequence size. Nonetheless most animal genomes possess a numerous class of short repetitive sequences only a few hundred nucleotides long (cf. Davidson et al., 1975), and thus application of the heteroduplex method should be possible. A preliminary study provided evidence for a class of short repetitive sequences interspersed with unique DNA in the rat by this technique (Bonner et al., 1973).

References

Angerer, R.C., Davidson, E.H. and Britten, R.J. (1975). Cell 6, 29.

Bishop, J.O., Pemberton, R. and Baglioni, C. (1972). Nature New Biol. 235, 231.

Bonner, J., Garrard, W.T., Gottesfeld, J., Holmes, D.S., Sevall, J.S., and Wilkes, M. (1973). Cold Spring Harbor Symp. Quant. Biol. 38, 303.

Britten, R.J., and Davidson, E.H. (1969). Science 165, 349.

Britten, R.J., and Kohne, D.E. (1968). Science 161, 529.

Britten, R.J., and Smith, J. (1970). Carnegie Institution of Washington Year Book 68, 378.

Campo, M.S., and Bishop, J.O. (1974). J. Mol. Biol. 90, 649.

Chamberlin, M.E., Britten, R.J. and Davidson, E.H. (1975). J. Mol. Biol. 96, 317.

Davidson, E.H., and Britten, R.J. (1973). Quart. Rev. Biol. 48, 565.

Davidson, E.H., Galau, G.A., Angerer, R.C., and Britten, R.J. (1975). Chromosoma 51, 253.

Davidson, E.H., Hough, B.R., Amenson, C.S. and Britten, R.J. (1973). J. Mol. Biol. 77, 1.

Davis, R.W., Simon, M. and Davidson, N. (1971). In Methods in Enzymology (Grossman, L. and Moldave, K., eds), pp. 413-428, Academic Press, New York.

Graham, D.E., Neufeld, B.R., Davidson, E.H. and Britten, R.J. (1974). Cell 1, 127.

Hennig, W., Hennig, I., and Stein, H. (1970). Chromosoma 32, 31.

Kedes, L.H., and Birnstiel, M.L. (1971). Nature New Biol. 230, 165.

Kram, R., Botchan, M., and Hearst, J.E. (1972). J. Mol. Biol. 64, 103.

Lewin, B. (1975). Cell 4, 77.

Manning, J.E., Schmid, C.W., and Davidson, N. (1975). Cell 4, 141.

Sullivan, D., Palacios, R., Stavnezer, J. Taylor, J.M., Faras, A.J., Kiely, M.L., Summers, N.M., Bishop, J.M., and Schimke, R.T. (1973). J. Biol. Chem. 248, 7530.

Suzuki, Y., Gage, L.P., and Brown, D.D. (1972). J. Mol. Biol. 70, 637.

Wetmur, J.G., and Davidson, N. (1968). J. Mol. Biol. 31, 349.

Wilson, D.A., and Thomas, C.A. (1974). J. Mol. Biol. 84, 115.

Wu, J.-R., Hurn, J.,and Bonner, J. (1972) J. Mol. Biol. 64, 211.

Chapter 1.

The Size of Repetitive Sequences in the Rat

Summary

The size of the interspersed repetitive sequences in rat DNA was determined. DNA strands composed of repetitive sequences flanked on one or both ends by unique sequences were isolated in two fragment sizes. Foldback sequences were removed by discarding the DNA bound to HAP after a brief incubation to low C_ot; unique DNA was eliminated as HAP unbound material when a second incubation to C_ot 50 was conducted. The self-reassociation of these strands often results in the formation of "H" structures and multiple duplex structures discernible in the electron microscope; these structures consist of a repetitive duplex region terminated by four single strand regions. Statistical data were derived from the structures. From the data we conclude that a numerous class of repetitive sequences 400 ± 150 nucleotides on average is widely interspersed with unique DNA in the rat, and that a sparse class of much longer repeats is also interspersed with unique DNA.

Britten and Davidson (1969) have hypothesized that the repetitive DNA sequences found in eucaryotic genomes mediate the control of transcription of adjacent structural genes. Structural gene sequences are in most cases unique (reviewed by Lewin, 1975). Therefore, this hypothesis predicts that repetitive and unique sequences be widely interspersed. Further, since the amount of unique sequence in most animal DNAs exceeds the amount of repetitive sequences, it is expected that most interspersed repetitive sequences be shorter than unique sequences. These expectations have been confirmed by observations in numerous organisms (reviewed by Davidson et al., 1975). Most genomes possess short repetitive sequences 300-400 nucleotides long on average interspersed throughout the greater part of the DNA with unique sequences of from several hundred to several thousand nucleotides in length (Davidson, et al., 1973; Graham et al., 1974; Firtel and Kindle, 1975; Schmid and Deininger, 1975; Bonner et al.,1973). <u>Drosophila</u>, an organism whose repeat sequences are at least 5 kb on average (Manning, Schmid and Davidson, 1975), is the notable exception to this prevalent "<u>Xenopus</u> pattern" of sequence organization.

Most studies of sequence organization have employed S1 nuclease and hydroxyapatite binding experiments to measure the size of repetitive and unique sequences. The electron microscope offers a means of directly measuring the distribution of these sequence lengths (Manning, Schmid and Davidson, 1975; Chamberlin, Britten and Davidson, 1975; Bonner et al., 1973).

Herein we report a detailed measurement of repetitive sequence length in the rat. Two size classes of DNA strands, nearly all of which contain at least one repetitive sequence element, were isolated. Assuming that the DNA was sheared randomly, and that the strand size is greater than the repetitive sequence size, then many of the DNA strands should consist of a repetitive sequence flanked on each side by unique sequences. Reassociation of two such strands results in the formation of a repetitive sequence duplex terminated at each end by two noncomplementary unique DNA single strands. When visualized in the electron microscope, the structures have the appearance of an "H." Length measurement of the duplex region terminated at each end by the two single-stranded "tails" yields the length of the repetitive sequence. Chamberlin et al. (1975) have used this technique to determine the repetitive sequence distribution in *Xenopus*. The repetitive sequence length has been measured in this manner by Manning et al. (1975) in *Drosophila*.

Results

Preparation of DNA

We prepared two repetitive sequence-bearing fractions of 1.7 and 2.5 kb average length (1 kb = 1000 nucleotides or nucleotide pairs) by the protocols shown in Figures 1 and 2. It should be noted that 1.7 and 2.5 kb correspond to the number average size of the input DNA strands in the reassociation reactions described below. After completion of the reactions and upon visualization in the EM the average size of strands found in duplex structures bounded by four single-strand regions was unaltered. However, breakage during the reactions and mounting for EM reduced the mean sizes of the total strand populations to 1.2 and 1.7 kb respectively for the DNA designated 1.7 and 2.5 kb.

It has been determined by analysis of reassociation kinetics and by EM that rat DNA contains at least three major sequence components (Holmes and Bonner, 1974; Pearson, Wu, Posakony and Bonner, in preparation; Chapter 3). Unique sequences comprise 75% of this genome. Another 20% of the DNA is composed of a repetitive sequence component of overall repetition frequency 1800. However, there may be at least two frequency subcomponents in this class. Lastly, foldback sequences account for 5% of the DNA. The latter sequence elements are completely reassociable at very low C_ot ($<10^{-3}$). The repetitive component reassociates between C_ot 0.005 and C_ot 50. Single copy sequence reassociation begins at about C_ot 50. Thus, it is possible to isolate DNA strands of moderate size which contain only repetitive sequences or only repetitive and unique sequences on the same strand. This is done first by "stripping" the DNA of foldback sequences by discarding DNA bound to HAP after incubation to low C_ot. The unbound DNA is reassociated to C_ot 50 and then fractionated on HAP. The bound peak is eluted as the repetitive DNA fraction. We emphasize that this fraction contains repetitive and unique sequences on the same strand, as well as some strands composed exclusively of repetitive DNA.

After elution from HAP in single-stranded form the repetitive fraction was reassociated again to C_ot 50 to form hybrids, which were spread for electron microscopy from 50% formamide. All incubations and HAP fractionations were carried out in 0.12 M phosphate buffer at 60°C.

Details of the isolations, the average strand sizes at various stages in the experiment, and HAP binding data are presented in Figures 1 and 2. After the DNA was "stripped" of foldback sequences and reassociated to C_ot 50, the 1.7 and 2.5 kb DNA preparations experienced about 40% binding to HAP as measured by optical absorbance. In general, we would expect greater binding to HAP at this step for the larger DNA preparation. However, as indicated earlier the overall mean lengths of all strands in the two preparations were 1.2 and 1.7 kb respectively after this fractionation step, so that a relatively small increment in binding could be expected between the two preparations. That a small increment in binding was not observed may reflect the comparative inaccuracy of optical absorbance determination of binding.

Recovery of Duplex Structures

In the EM the DNA structures fell into four categories: (1) 4-ended or "H" structures, (2) 3-ended structures, (3) multiple duplex structures and (4) linear strands. The DNA on the grids was scanned at random. All the DNA observed was measured and interpreted, excluding only strands found on technically imperfect areas of the grids, e.g., those with contrast failure. Only 1% of the molecules were found in uninterpretable structures.

A composite electron micrograph of some of the "H" structures studied is shown in Figure 3. The simplest form of multiple duplex structure results from the reassociation of three strands to form a two-duplex hybrid. Figure 4 exhibits a number of these. When four or more strands participate in multiple duplex formation the resultant structures have the appearance shown in Figure 5. Statistical data on the recovery of duplex structures in the two experiments are given in Tables 1 and 2. These results show that the strands engaged in "H" and multiple duplex structures are about the same size as the input DNA.

Length of Repetitive Sequences

Table 3 shows the data pertaining to recovery of repetitive duplexes terminated by four single strand regions. The distributions of repetitive sequence lengths measured among the 1.7 and 2.5 kb DNA strands appear in Figure 6. The number average repetitive sequence length in the rat is 400 ± 150 nucleotides according to these data. The great majority of measurements fall within the limits of 100-500 nucleotides. There appears to be a rather disperse, sparse population of longer duplexes. However, since the measured size of the larger duplexes is about the same magnitude as the DNA strand length, there is the strong possibility of greatly underestimating the true average length of the larger repetitive duplex class. Further, in many cases the distinction between single- and double-stranded DNA was insufficient to rule out the possibility that some structures interpreted as long "H" structures were actually three-stranded structures. Micrographs of such structures are given in Figure 7.

For the value of 400 nucleotides to accurately reflect the average size of the short repetitive sequences we must establish that the measured duplex lengths are independent of the lengths of the single strands on which they reside. If, for example, the average duplex size measured on a particular strand length increased continuously with the resident strand length, the possibility would exist that the true average repetitive sequence size has been underestimated. Thus, the existence of a substantial population of longer duplexes could go undetected, because the reassociating strands were insufficiently long to permit formation of a long duplex with four tails. That this is not the case is shown in Figure 8. The strands containing duplexes terminated by four single strand regions were divided into intervals of 500 nucleotides in length and matched with the average length of the duplex regions measured at each strand length. Figure 8 shows that for both the 1.7 and 2.5 kb DNA the average duplex length does not vary consistently over a wide range of strand lengths (0-15 kb).

We can calculate the fraction of the genome composed of short repetitive sequences by a quantitative treatment owing to Chamberlin et al. (1975). The percent of the genome composed of short repetitive sequences is equal to the product of the percent of the total DNA found in duplex and the fraction of the genome constituted by the isolated repetitive DNA fractions. For the 1.7 and 2.5 kb DNA fractions these products were 6.0 and 8.6% respectively. The probability of formation of a duplex structure with four single strand regions at least 100 nucleotides long each is $(1 - (R + 200)/L)^2$ where R is the repetitive sequence length and L the strand length (Chamberlin et al., 1975). The mean size of the strands engaged in duplexes terminated by four single strand regions was 1.914 and 2.526 kb in the two experiments. Using these numbers for L and 400 for R, we calculate the probabilities of formation and multiply them by 15%, which is the approximate proportion of the genome comprised by short repeats (Pearson, Wu and Bonner, submitted for publication). The products represent the percent of the total DNA expected to be in duplex if recovery of repeated sequences is complete (100%). The respective values obtained for the 1.7 and 2.5 kb fractions were 7.0 and 8.7% in good agreement with observation.

Discussion

Sequence Organization in Rat DNA

The data indicate that a numerous class of short repetitive sequences comprising 15% of the genome is widely interspersed with unique DNA. The average sequence of this class is 400 ± 150 nucleotides in length. Observations on the multiple duplex structures show that some repetitive sequences may be spaced very close to each other (<0.3 kb). Whether such short spacings actually correspond to unique sequences cannot be established, since it is possible that the presence of two closely spaced duplexes may sterically hinder the hybridization of a third repetitive sequence in between the two already hybridized. A treatment of the spacings between repetitive sequences in the rat is given in Chapter 2.

Comparison to Other Organisms

The "Xenopus pattern" of interspersion of 300-400 nucleotide repeats with unique DNA prevails among eucaryotes (reviewed by Davidson et al., 1975; cf. also Firtel and Kindle, 1975; Schmid and Deininger, 1975). Our results show that rat DNA conforms to this pattern.

Experimental Procedures

Preparation of DNA

DNA was prepared from rat ascites cells by extensive protease and RNAse digestions and repeated organic extractions with 25:24:1 phenol/chloroform/isoamyl alcohol. Shearing of the DNA samples was performed in a Virtis homogenizer at 5°C in 0.05 M neutral phosphate buffer. Alkaline denaturation of the DNA was largely according to Manning, Schmid and Davidson (1975). All incubations and HAP chromatographic procedures were conducted at 60°C in 0.12 M phosphate buffer. DNA bound to HAP was eluted with 0.12 M Na_3PO_4 (pH 12.3) at 60°C.

Electron Microscopy

DNA was dialysed against 0.01 M Tris, 0.001 M EDTA (pH 8.5), made 50% in formamide and spread for microscopy by the modified Kleinschmidt technique of Davis, Simon Davidson (1971). The DNA was visualized in a Philips EM 201 electron microscope at an accelerating voltage of 60 kV. The grids contained single and double stranded PhiX174 circular DNA molecules as internal standards of known length 5.25 kb. The difference between single and double strand PhiX length determinations was negligible, and no correction factor was applied for discrepancy between single and double strand measurements. Micrographs were recorded on 35 mm film which was projected directly onto the electrosensitive platen of a Hewlett-Packard electronic digitizer. Digitizer measurements were stored in a PDP-10 computer for analysis.

Data Analysis

The photographic data were translated into a numerical code for computer analysis. Each hybrid was considered to be an array of alternating single strand and duplex regions. The data were stored as number triplets comprised of the single strand length, the double strand length and an identifying number arbitrarily assigned to each duplex in a hybrid structure. In this way the computer could reconstruct the hybrid complexes and calculate various physical parameters.

References

Britten, R.J., and Davidson, E.H. (1969). Science 165, 349-357.

Chamberlin, M.E., Britten, R.J., and Davidson, E.H. (1975). J. Mol. Biol. 96, 317-333.

Davidson, E.H., Galau, G.A., Angerer, R.C., and Britten, R.J. (1975). Chromosoma 51, 253-259.

Davidson, E.H., Hough, B.R., Amenson, C.S., and Britten, R.J. (1973). J. Mol. Biol. 77, 1-23.

Davis, R.W., Simon, M., and Davidson, N. (1971). In Methods in Enzymology (Grossman, L. and Moldave, K., eds.), pp. 413-428, Academic Press, New York.

Firtel, R.A., Neufeld, B.R., Davidson, E.H., and Britten, R.J. (1974). Cell 1, 127-137.

Holmes, D.S., and Bonner, J. (1974). Biochemistry 13, 841-848.

Lewin, B. (1975). Cell 4, 77-93.

Manning, J.E., Schmid, C.W., and Davidson, N. (1975). Cell 4, 141-155.

Schmid, C.W., and Deininger, P.L. (1975). Cell 6, 345-358.

Table 1. Disposition of 1.7 kb Repetitive DNA Strands Renatured to $C_o t$ 50

Type of Structure	Number of Strands Scored	% of All Strands Scored	Total Strand Length in Class (kb)	% of Total Length in Class	Strand Lengths (kb)	
					Number Average	Weight Average
Four-ended Structures	190	8.5	317	12.0	1.667	2.094
Three-ended Structures	364	16.3	373	14.1	1.024	1.738
Multiple Duplex Structures	259	11.7	543	20.5	2.096	3.170
Linear Molecules	1413	63.4	1413	53.3	1.000	1.424
Lariats	2	0.1	4	0.1	2.000	2.022
Uninterpretable	——	——	——	1.0	——	——
Total	2228		2650		1.188	1.907

Table 2. Disposition of 2.5 kb Repetitive DNA Strands Renatured to C_ot 50

Type of Structure	Number of Strands Scored	% of All Strands Scored	Total Strand Length in Class (kb)	% of Total Length in Class	Strand Lengths (kb)	
					Number Average	Weight Average
Four-ended Structures	248	11.4	533	14.5	2.150	2.881
Three-ended Structures	316	14.5	428	11.7	1.354	2.431
Multiple Duplex Structures	515	23.6	1394	37.9	2.707	4.008
Linear Molecules	1101	50.5	1318	35.9	1.197	1.866
Circles	1	0.04	2	0.05	1.673	1.673
Lariats	1	0.04	3	0.08	3.202	3.202
Uninterpretable	—	—	—	1.0	—	—
Total	2182		3678		1.685	2.892

Table 3. Recovery of Duplex Structures from Repetitive DNA Strands Renatured to C_ot 50

Input DNA Size (kb)	Type of Duplex-bearing Structure	Number of Duplexes Scored	Total Duplex DNA (kb)	% of Total DNA in Duplex	Duplex Lengths (kb)	
					Number Average	Weight Average
1.7	4-ended	190	154	5.8	0.406	0.781
1.7	Multiple Duplex	307	220	8.4	0.357	0.601
1.7	Total	497	374	14.2	0.376	0.675
2.5	4-ended	248	238	6.4	0.478	1.006
2.5	Multiple Duplex	715	600	16.4	0.420	0.757
2.5	Total	963	862	23.4	0.447	0.988

Figure Legends

Figure 1. Preparation of 1.7 kb Repetitive DNA

Conditions of shearing, incubation and HAP chromatography appear in the experimental procedures section. All length measurements were made by electron microscopy at the point in the isolation designated in the flowchart. The mixing in of a trace amount of long DNA to the final reaction of repetitive strands pertains to heteroduplex formation experiments reported in Chapter 2 and does not affect the results herein.

Figure 2. Preparation of 2.5 kb Repetitive DNA

See legend to Figure 1.

Figure 3. "H" Structures Formed among the 2.5 kb Repetitive DNA Strands

A marker in the micrograph gives the equivalent length of 1 kb.

Figure 4. Double Duplex Structures Formed among the 2.5 kb Repetitive DNA Strands

These six-ended structures were interpreted to be the equivalent of two "H" structures connected by a single strand.

Figure 5. Multiple Duplex Structures Formed among the 2.5 kb Repetitive DNA Strands

Scale line drawings of the interpretations placed on the structures in the composite micrograph appear below; single lines are single strands; double, cross-hatched lines are duplex regions. When an ambiguity arose, e.g., an odd number of tails in the hybrid structure, one reasonable interpretation of the complex was recorded. In some cases, this required us to designate certain tails to be single-tailed structures analogous to the three-ended molecule class; however, no duplex measurement of one-tailed structures was attempted. We draw a short duplex and dotted line to indicate the presence of a one-tailed duplex structure.

Figure 6. Repetitive Sequence Lengths Measured from the 1.7 and 2.5 kb Repetitive DNA Preparations

Frequency is equivalent to the number of measurements made; the interval size is 150 nucleotides.

(a) Duplex lengths from the 1.7 kb DNA.

(b) Duplex lengths from the 2.5 kb DNA.

Figure 7. "H" Structures from the 2.5 kb Repetitive DNA Strands apparently having long duplex lengths.

Figure 8. Relationship of Duplex Length and Strand Length

The lengths of the duplex regions of "H" and multiple duplex structures were matched with the lengths of the strands on which they resided. Only duplexes terminated by four single strand regions were scored. The strand lengths used were the average of the lengths of the two strands forming a particular duplex structure. The mean length of the duplex regions in each interval are plotted against the strand length; the interval size is 500 nucleotides. The data are derived from 449 strands of the 1.7 kb DNA preparation and 763 strands of the 2.5 kb DNA.

(a) Duplex vs. strand length plot of 1.7 kb DNA.

(b) Duplex vs. strand length plot of 2.5 kb DNA.

Preparation of 1.7 KB Repetitive DNA

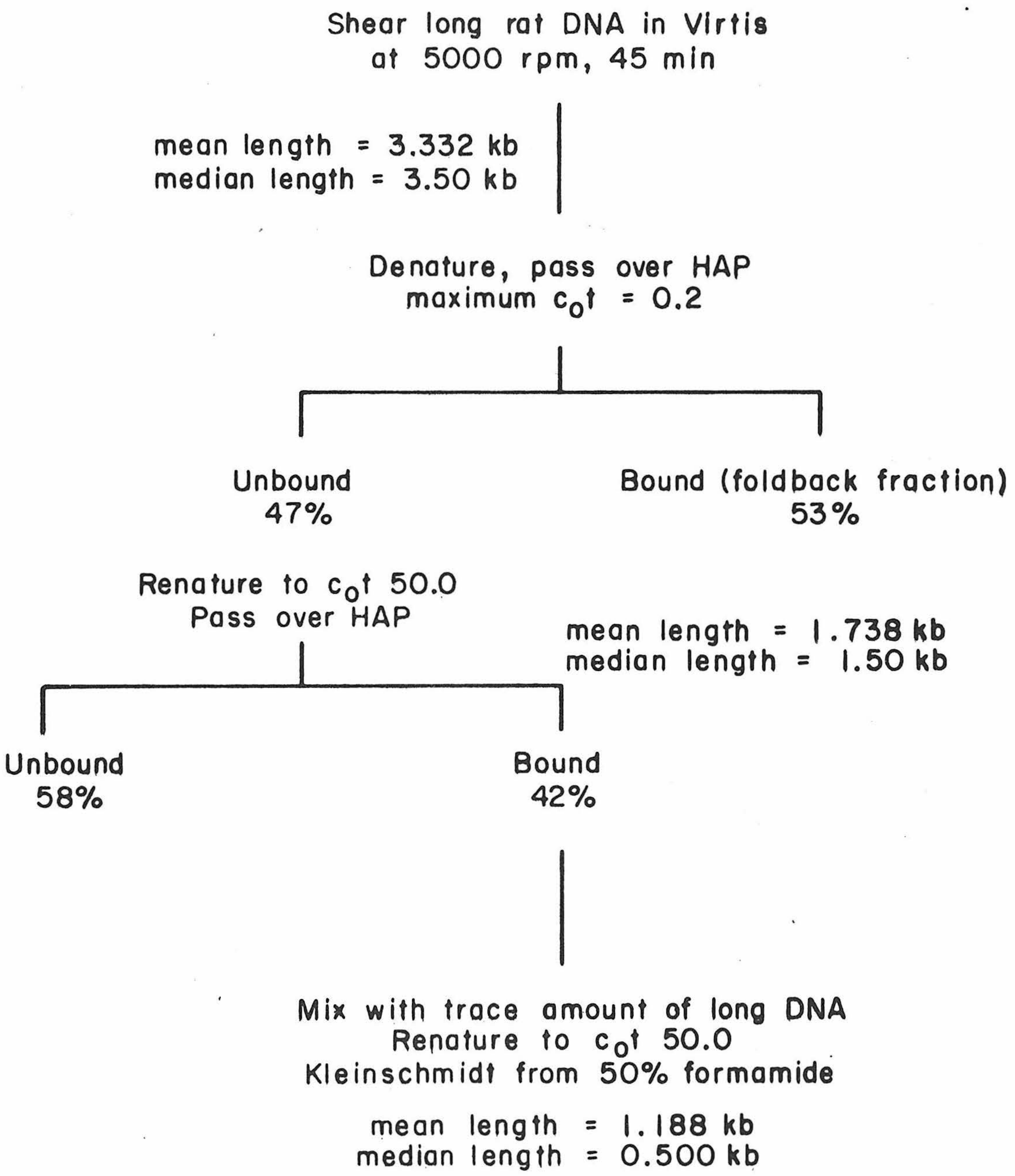


Fig. 1

Preparation of 2.5 KB Repetitive DNA

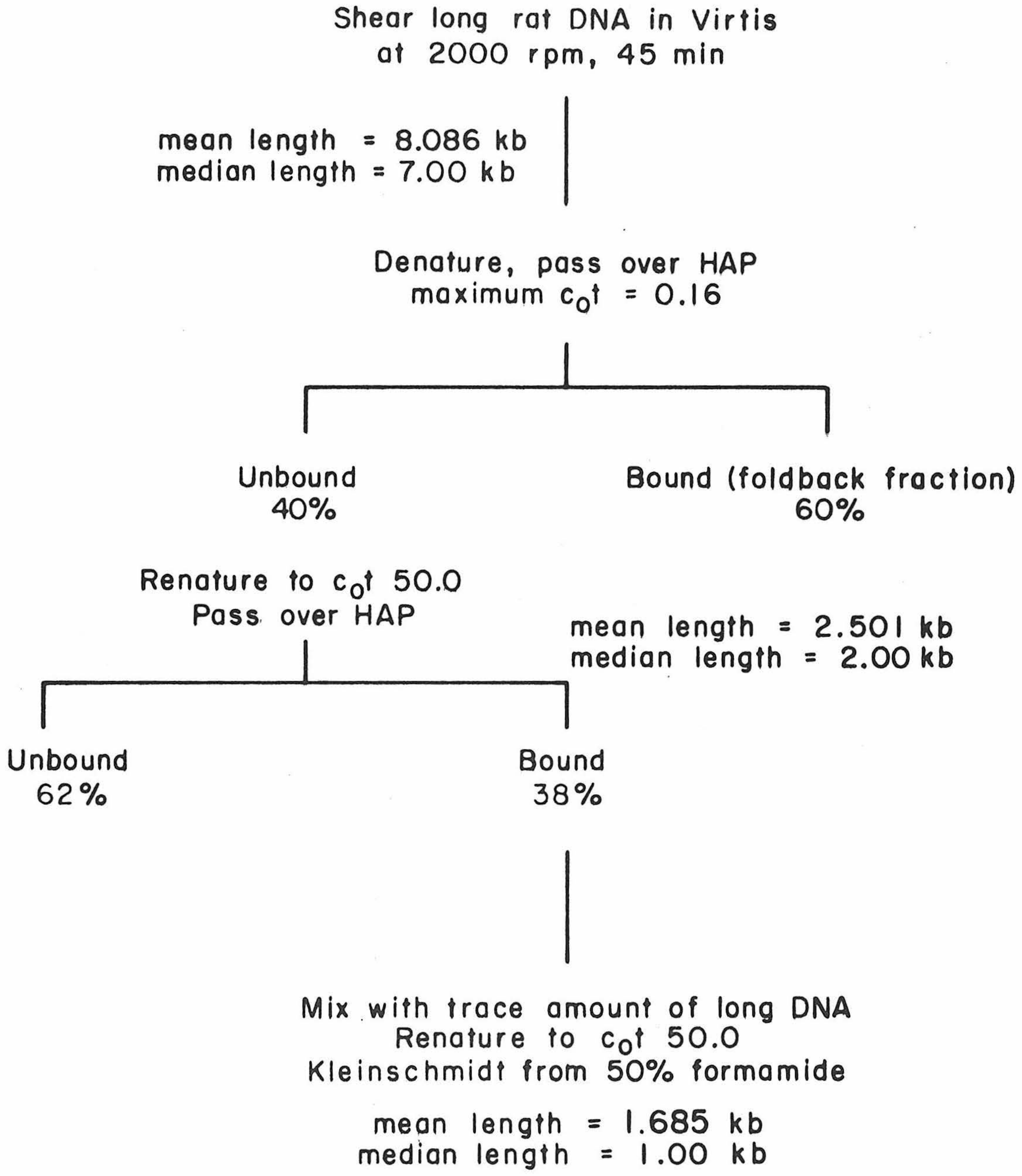


Fig. 2

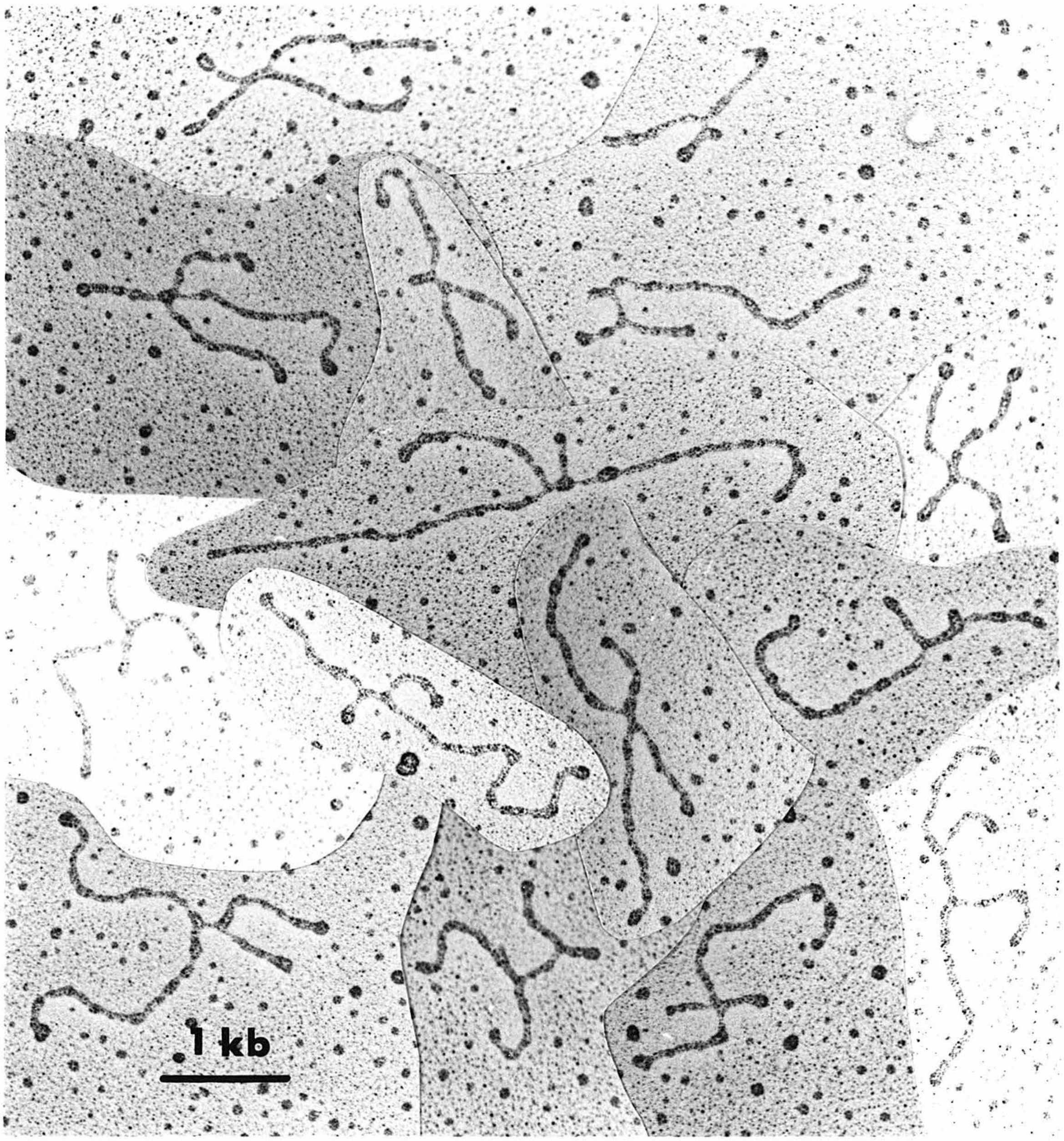


Fig. 3

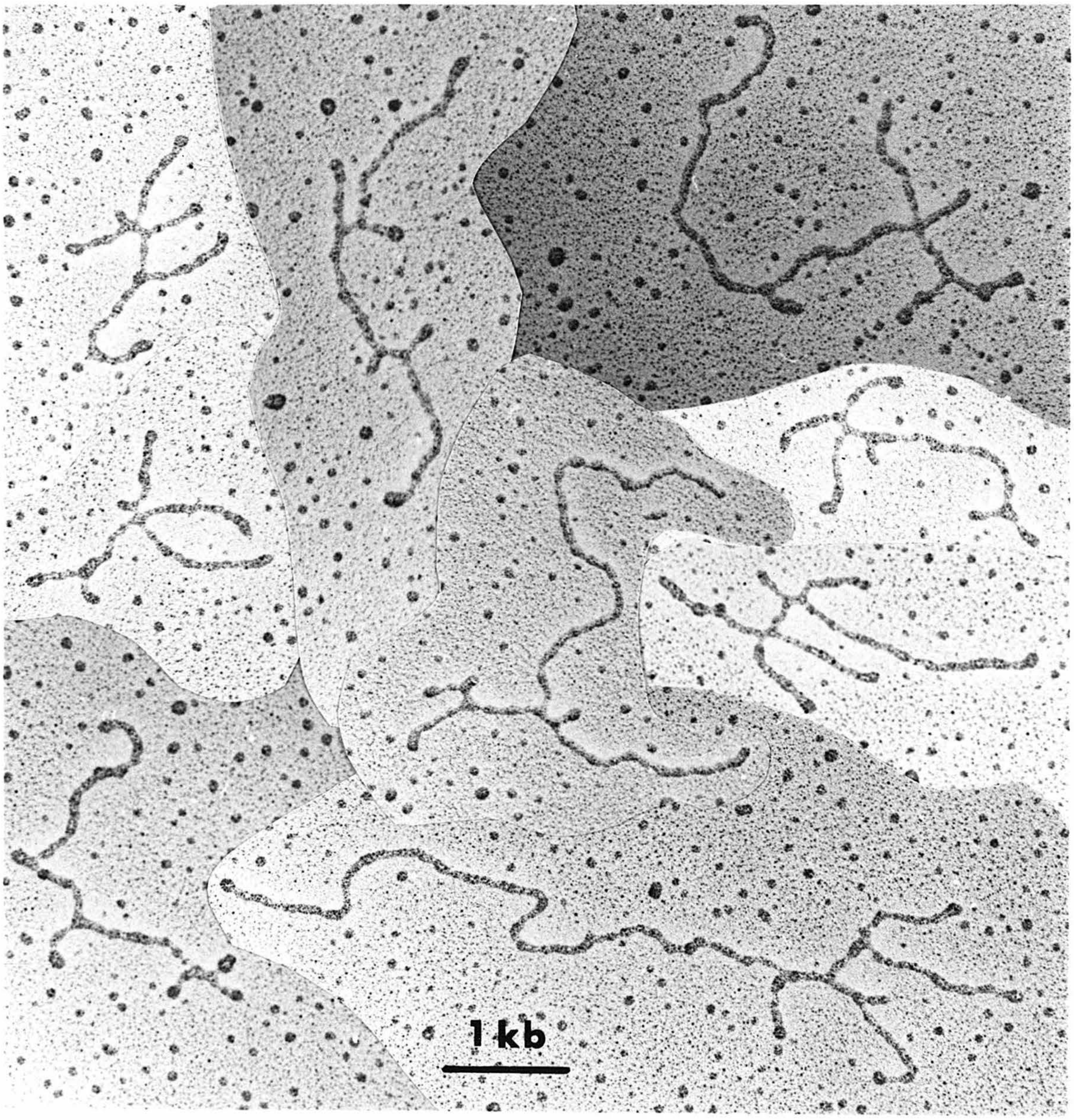


Fig. 4

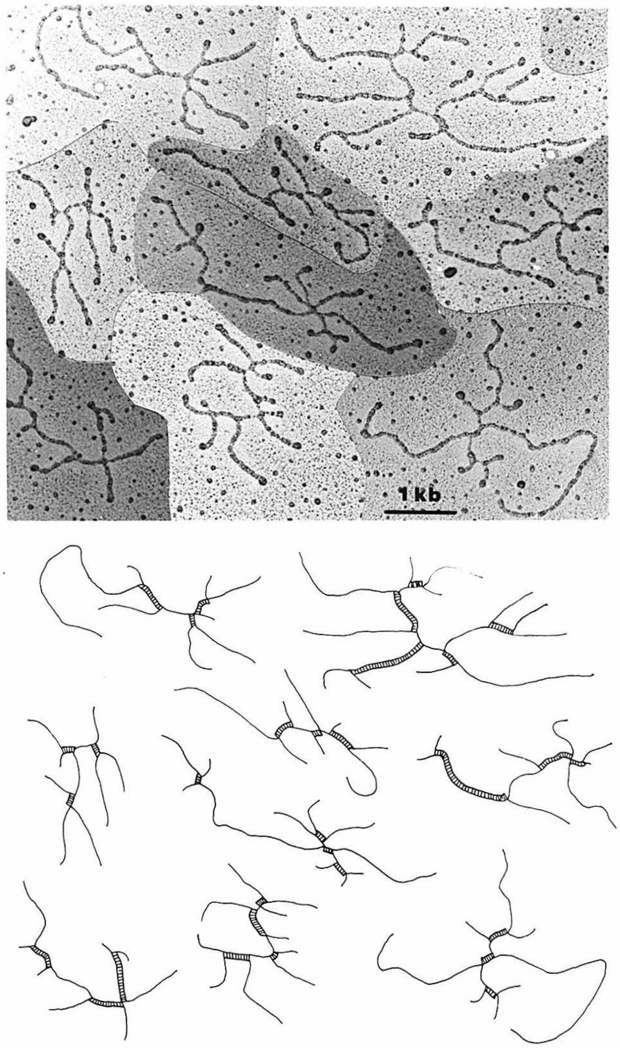


Fig. 5

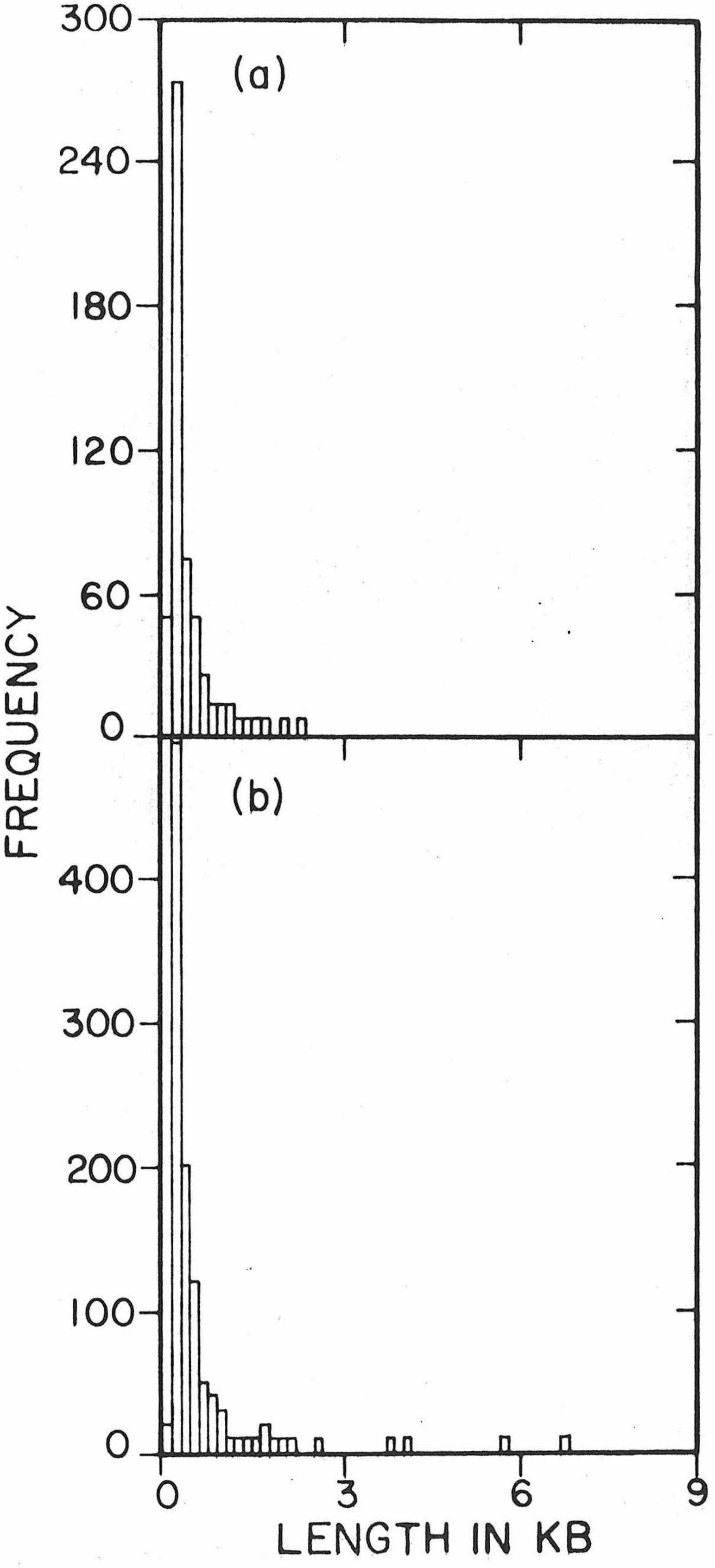

(a)
(b)
300
240
180
120
60
0
400
300
200
100
0
FREQUENCY
0
3
6
9
LENGTH IN KB

Fig. 6

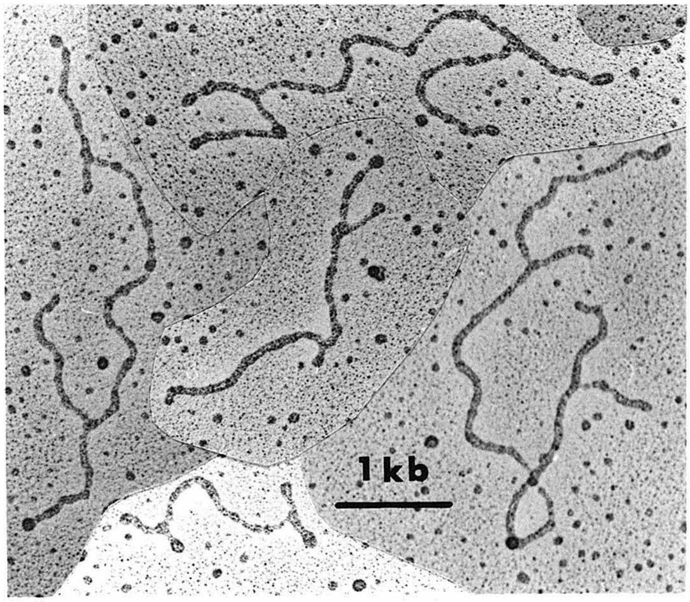


Fig. 7

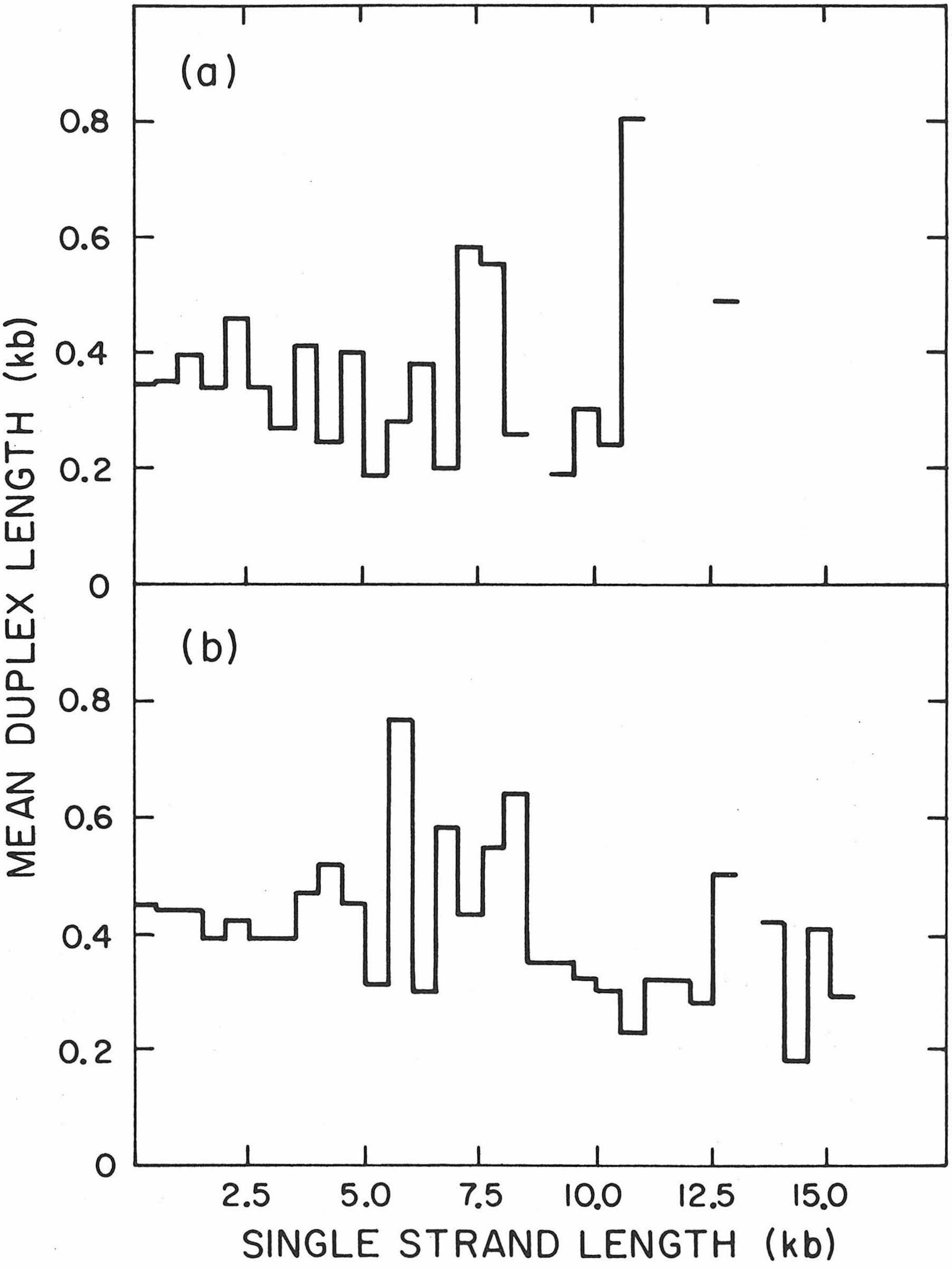

(a)
(b)
MEAN DUPLEX LENGTH (kb)
0.8
0.6
0.4
0.2
0
0.8
0.6
0.4
0.2
0
2.5
5.0
7.5
10.0
12.5
15.0
SINGLE STRAND LENGTH (kb)

Fig. 8

Chapter 2.

Sequence Organization of the Rat Genome by Electron Microscopy

Summary

Heteroduplex molecules between long total rat DNA single strands and short, repetitive sequence-bearing strands were prepared. Because most of the repetitive sequences on the short strands are flanked on both sides by unique sequences, these heteroduplexes display short unique DNA "tails" demarcating the ends of the repeat sequences in the electron microscope. In this way both the size of the repetitive sequences and the unique DNA spacings between them could be directly determined. Foldback sequences form hairpin structures in the long DNA strands, and their size and distribution can also be measured. From the data we propose this model of the rat genome. A numerous class of repetitive sequences 400 ± 150 nucleotides long on average is interspersed with unique sequences of at least 1500-1800 nucleotide number average length throughout at least 90% of the genome. Foldback sequences are interspersed with repetitive and unique sequences throughout half the genome and with unique sequences alone in approximately 5% of the DNA. 15% of the repetitive sequences are organized into doublets and a few triplets of contiguous sequences. Short and long period interspersion of repetitive sequences may be distinguished, each pattern accounting for equal fractions of the interspersed portion of the genome.

Introduction

A desire to understand the structural basis of eucaryotic gene regulation has stimulated studies into the physical organization of DNA sequences. Detailed models of the *Xenopus* (Davidson et el., 1973) and sea urchin genomes (Graham et al., 1974) have been derived from hydroxyapatite binding data by computational methods. In Xenopus for instance, half the DNA consists of repetitive sequences about 300 nucleotides long interspersed with unique sequences of 700 to 900 nucleotide average length. One quarter of the DNA is similarly arranged but with the average unique sequence length in excess of 4000 nucleotides. Regions of exclusively unique DNA and of clustered repetitive sequences comprise the remainder of the genome. In sea urchin the average unique sequence length in the short period interspersed portion of the genome is about 1 kilobase (1 kb = 1000 nucleotides or nucleotide pairs), though otherwise this genome is organized much like that of Xenopus. Aplysia DNA also generally corresponds to this pattern (Angerer et al., 1975).

An extension of the techniques of Wu, Hurn and Bonner (1972) can in principle provide direct measurements of repetitive and unique sequence length distributions. We chemically isolate short DNA strands, which are nonetheless larger than the average repetitive sequence, each containing repetitive DNA. The repetitive sequences embedded within the strands will be terminated with either one or two unique DNA flanking sequences. When this DNA is hybridized in vast excess to long, unsheared total DNA single strands, heteroduplex structures form. These structures consist of long DNA strands, whose repetitive sequences have hybridized to the complementary sequences among the driver DNA population. The unique DNA sequences flanking the driver strand repeats will in general be non-complementary to the sequences flanking the repeats to which they have hybridized. Therefore, these unique DNA "tails" will project from the long strands and be identifiable in the electron microscope. Measurement of the duplex regions terminated by two tails yields the length of repetitive sequences. Measurement of the single strand regions bounded by repetitive duplex structures gives the unique DNA sequence length.

An additional feature of this technique is that it identifies inverted repeat, or foldback, sequences in the DNA. Foldbacks have been studied previously in the electron microscope (Wilson and Thomas, 1974; Schmid, Manning and Davidson, 1975; Cech and Hearst, 1975), because they form hairpin or looped hairpin structures easily distinguishable under formamide spreading conditions. The rat genome contains 1.9 x 10^5 inverted repeat sequences of 0.71 kb number average length spaced an average of 10 kb apart throughout about three quarters of the DNA (Chapter 3). Foldback duplexes form in a rapid, concentration-independent reaction, when DNA single strands are renatured under appropriate conditions. The long DNA strands used in this study of repetitive sequence arrangement can be anticipated to form hairpin structures in addition to heteroduplexes. Since these two types of structures can occur on the same long DNA strand, an assessment of the relative spatial arrangement of repeat and inverted repeat sequences can be made. However, in the case of unlooped hairpin structures there is ambiguity in interpretation, since these may have the same appearance as repetitive hybrids with only one tail. We will not consider the unlooped hairpins further, therefore.

Results

Preparation of DNA

20% of the rat genome is comprised of repetitive sequences; 75% of unique sequences; 5% of foldback sequences (Holmes and Bonner, 1974; Pearson, Wu , Posakony and Bonner, in preparation; Chapter 3). Virtually all the repetitive sequences reassociate in the C_ot interval 0.005-50. It is possible to remove foldback sequences from the DNA by renaturation to low C_ot followed by removal of duplex-bearing DNA by hydroxyapatite chromatography. Unique DNA may be removed by renaturation to C_ot 50 followed by isolation of hydroxyapatite bound DNA. This DNA now contains only strands containing repetitive sequences. Most of these repetitive sequences have two unique DNA flanking sequences.

We prepared two repetitive DNA fractions of 0.9 and 2.5 kb number average length. These lengths represent the average size of the input strands in the reassociation reactions described below.

The details of the isolation are as follows: 1.1 and 8.1 kb DNA strands were briefly renatured and applied to hydroxyapatite (maximum $C_ot = 0.16$). Respectively, 20% and 60% of the DNA bound. The unbound strands measured 0.9 and 2.5 kb on average after this procedure. These "stripped" preparations were then renatured to C_ot 50 and fractionated on HAP. After chromatography the DNA strands were 0.9 and 1.7 kb long on average. The respective percents bound for the two fractions were 42 and 38%. These two fractions we designate the 0.9 and 2.5 kb repetitive DNA fractions. The fractions contain strands having repetitive sequences and unique DNA sequences flanking the repeats at one or both ends. We might expect the binding of the 2.5 kb fraction to be greater than that of the 0.9kb DNA at C_ot 50, since a larger proportion of single copy sequence should be bound to the column in this case. That the binding of the two fractions was essentially the same may in part reflect the inaccuracies of the optical absorbance determinations and partly the fact that the two preparations were of fairly similar size (0.9 vs. 1.7 kb) after HAP chromatography. Angerer et al. (1975) also observed similar binding of two comparably sized Aplysia DNA preparations.

The 0.9 kbrepetitive DNA fraction was hybridized to 29.4 kb long total DNA single strands in the following manner. Denatured repetitive and long DNA were mixed in a mass ratio of 100:1 and renatured to C_ot 50 in 0.12 M phosphate buffer at 60°C. Since repetitive sequences comprise 20% of the genome, (Holmes and Bonner, 1974; Pearson, Wu and Bonner, submitted) the C_ot was calculated to accommodate a 5-fold greater relative rate of reaction. This means also that the driver to tracer sequence excess was 500:1. The C_ot for tracer self-reassociation was 0.1. At the end of the reaction part of the DNA was prepared for electron microscopy, while the remainder was mixed with freshly alkali-denatured and neutralized driver DNA. The driver DNA from the first phase of the reaction was considered to be inert with respect to further reaction. In this second phase of reaction the mass and sequence excesses of fresh driver to tracer were 1700:1 and 8500:1 respectively.

The reaction was carried to C_ot 50 as before. The tracer C_ot was 0.1 . These DNA samples were mounted for electron microscopy at 25°C from 50% formamide by the modified Kleinschmidt method of Davis, Simon and Davidson (1971). This formamide concentration maintains the criterion of duplex formation equivalent to 0.12 M phosphate buffer at 60°C.

2.5 kb repetitive strands and long DNA were renatured to C_ot 50 in mass and sequence excesses of 120:1 and 600:1 respectively. The tracer C_ot was 0.05. Reaction conditions were as above. No second reaction of the repetitive and long DNA hybrids was performed.

Recovery of Hybrid Structures

The grids containing the various DNA preparations were scanned at random to the exclusion only of areas of obvious technical imperfection. All the DNA in the scanned regions was photographed and interpreted. A detailed description of the self-reassociation products of the driver DNA appears in Chapter 1.

Uninterpretable structures accounted for 1% of the total DNA and were disregarded. To collect additional repetitive/long DNA hybrid structures we scanned the grids further, photographing only the strands substantially longer than the largest of the driver molecules.

The structures found under the electron microscope fell into 3 main categories: (1) driver/long DNA hybrids with two tails, (2) duplex structures between driver strands and the tails of driver strands already hybridized to a long strand and (3) looped foldback structures. In addition, there was a substantial number of duplex structures with only one tail. These could represent either hybrids of repetitive strands flanked only at one end by a unique sequence or foldback structures with no loops. There were in the data from the 0.9 and 2.5kb fractions respectively 296 and 618 cases of two-tailed hybrid structures (Table 1) compared to 62 and 120 of the one-tailed events.

Figure 1 shows hybrid structures between 0.9 kb repetitive and long DNA and our interpretations thereof. Similar structures, which include hybrids between driver strands and previously hybridized driver strand tails, are pictured in Figure 2. Micrographs of 2.5 kb repetitive/long DNA hybrids appear in Figures 3 and 4.

The numbers of long strands scored and their molecular weights are listed in the last column of Table 5. Table 1 gives the recovery of all hybrid strands.

Recovery of Repetitive Duplexes

The length of the duplex regions demarcated by two tails equals the repetitive sequence length of rat DNA. Data in Chapter 1 have characterized a class of repetitive sequences 400 ± 150 nucleotides long in the rat from electron microscope results. A similar conclusion has been reached by means of physical chemical techniques (Pearson, Wu and Bonner, submitted). The data on duplex recovery in these experiments displayed in Table 2 are consistent with this result. Fig. 5 shows the distribution of repetitive sequence lengths for the two experiments. Among the 2.5 kb/long DNA hybrids there were apparently a number of large repetitive duplexes (>2kb). However, any measurements on this larger class of sequences may strongly underestimate the true average sequence length, since the average input strand is approximately the same size as the observed sequence length. Pearson, Wu and Bonner (submitted) present physical chemical evidence of a class of repetitive sequences in rat at least 2 kb long and comprising roughly 5% of the genome. If we eliminate such larger repeats from the 2.5 kb data, the resulting size for the short repetitive sequence class is 430 nucleotides. The 0.9 kb data indicate 340 nucleotides as the number average size of this class of repetitive sequences. Both numbers are in good agreement with the findings in Chapter 1.

Before we can proceed to the question of the average sequence length between repeats, we must establish whether the hybridization of repetitive sequences on the long strands was complete. The presence of an appreciable fraction of unhybridized repeats could result in a drastic overestimate of the actual length of unique sequences between repeats. We apply a treatment analogous to that of Chamberlin et al. (1975). The probability of formation of a two-tailed duplex structure, whose tails are at least 100 nucleotides long, i.e., sufficiently long to be identified under the electron microscope, is: $1 - (200 + R)/L$, where R and L are the repetitive sequence and the fragment lengths respectively. Setting R equal to 400 nucleotides and using the values of L from Table 1, we calculate the expected percent of DNA in duplex for the 1.3 kb and the 2.4 kb hybrid experiments to be 8.2 and 11.2 given then 15% of the genome consists of short repetitive sequences (Pearson, Wu and Bonner, submitted). The observed values 10.1 and 10.2 are in good agreement with expectation.

We conclude that recovery of repetitive duplex structures was essentially complete.

Recovery of Foldback Sequences

Chapter 3 gives evidence for two classes of foldback sequences in rat DNA: a numerous class of 0.30 average length containing 1.6% of total DNA and a disperse class of 6.1 kb long on average accounting for 2.6% of total DNA. Approximately half the foldback structures contain loops. Table 3 cites the recovery of looped foldback structures in these hybrid structure experiments. The 0.4 kb average foldback duplex size in Table 3 indicates that the recovered foldbacks were mainly from the class of short foldback sequences. Since only half the total foldbacks should have loops, we would expect to recover 1.6%/2 = 0.8% of the total DNA as foldback duplex. The actual recoveries were 1.0 and 0.9% respectively in the 0.9 and 2.5 kb experiments in good agreement with expectation.

Spacings Between Repetitive Sequences

By measuring the single-strand contour length from the proximal tail of one two-tailed repeat to the proximal tail of an adjacent two-tailed repeat, we can determine the length of the unique sequence between two repetitive sequences. The spacings measurements from the two experiments exhibit the distributions shown in Figure 6. The spacings range from the smallest observable value (∿50 nucleotides) to 30 kb. In addition 14.8% of the total repetitive sequences appeared to be located in physical contiguity (spacing = 0) to each other in groups of 2 and 3 (see below). The present data cannot establish whether spacings observed are present in rat DNA, since the average tracer strand lengths in the two experiments were 15 and 20 kb for the 0.9 and 2.5 kb data respectively.

Statistical data pertaining to the spacing measurements are found in Table 4. The spacing distributions from the 0.9 and 2.5 kb DNA (Fig. 6) display a close correspondence in the spacing frequencies less than roughly 2 kb in magnitude. However, in the range of larger spacings (>2 kb) there are far fewer measurements in the 0.9 kb than in the 2.5 kb results. This is reflected in the substantially lower average spacing values in the 0.9 kb data of Table 4. It may be due to the considerably smaller tracer strand size in this experiment (15 vs 20 kb), since the longer spacings have a decreased probability of observation in shorter molecules. The possibility exists that experiments with even longer tracer strands might yield an even larger value for the mean spacing. Nonetheless the distributions shown in Figure 6 indicate that most spacings fall in the range 0.05-3kb.

In general, we assume any spacing between two two-tailed duplexes to be equivalent to the length of that unique sequence. However, since many of the inter-repeat spacings are short (0.05-0.3 kb), we must consider a second possibility. It could be that the hybridization of a repetitive sequence located between two others, which have already hybridized, might be sterically hindered, if the distance between the two previously formed duplexes is short. Thus, while the contour length of DNA contained in all the spacings less than 0.3 kb, for example, is only 1.5% of the total DNA, replacement of all these apparent spacings with repetitive hybrids would have a very significant effect on the total spacing distributions (Figure 6 (c)-(d)). Table 4 shows that the actual average spacing range would be 12-1800 nucleotides instead of 800-1500, if this were the case. Another implication of this alternative interpretation is that the proportion of clustered repeats might be even higher than the 15% observed (see below). The 2.5 kb data more accurately reflect the total unique sequence distribution than do the 0,9 kb results for reasons previously mentioned, though they may themselves underestimate the true average spacing. Therefore, regardless of the nature of the short spacings, the average unique sequence in rat DNA is at least in the range 1500-1800 nucleotides. Using the physical chemical techniques pioneered by Britten, Davidson and coworkers (Davidson et al., 1973; Graham et al., 1974), Pearson, Wu and Bonner (submitted) have found a value of 1800-2000 nucleotides for the average unique sequence in the rat. The two results are in reasonable accord.

It is noteworthy than an appreciable fraction of the two-tailed repetitive duplexes resided contiguously to each other in pairs and infrequently also in triplets. In this regard the two experiments yielded consistent results. Using the combined data, 12.6% of the total number of duplexes scored resided in doublets and 2.2% in triplets. However, these measurements could be the spurious result of partial displacement of a two-tailed repeat by a second incoming driver strand.

This hypothesis would predict that the average size of sequences in doublets and triplets would be smaller than that of the total. This is not the case. The mean duplex length of 0.393 kb among the repeats in multuplets does not differ demonstrably from the total average.

Sequence Organization of Rat DNA

Table 5 lists the four classes into which the long strands studied fell: those possessing (1) only interspersed repetitive and unique sequences, (2) interspersed repetitive, foldback and unique sequences, (3) only foldback and unique sequences and (4) only unique sequences. We consider the 2.5 kb data to obtain a minimum estimate of the fraction of the genome containing interspersed repetitive and unique sequences. This fraction is 90%. Of this interspersed portion of the genome about 60% contains foldback sequences.

Discussion

A Model of the Rat Genome

We propose this model of the rat genome. A numerous class of repetitive sequences 400 ± 150 nucleotides long is interspersed with unique sequences averaging at least 1500-1800 nucleotides long throughout at least 90% of the genome. Foldback sequences are interspersed with repeated and unique sequences throughout at least half the genome. 15% of the repeated sequences occur in doublets and a few triplets of contiguous sequences.

Comparison to Other Organisms

Davidson et al. (1975) summarize evidence that the Xenopus pattern of sequence interspersion appears in organisms found on many of the major branches of the phylogenetic tree. Recently, the genomes of slime mold (Firtel and Kindle, 1975) and man (Schmid and Deininger, 1975) have also been shown to conform to this pattern. Only the DNA of Drosophila (Manning, Schmid and Davidson, 1975) stands as the notable exception among genomes studied until now. However, Walbot and Dure (1976) have reported that the mean length of interspersed repeated sequences in cotton is 1250 nucleotides--a value intermediate to the short repeats of most animals and the 5 kb repeated sequences of Drosophila. Angerer et al. (1975) have reviewed the issue of the regulatory significance of sequence organization. The organization of the rat genome conforms to the "Xenopus pattern" of sequence arrangement. It is noteworthy, however, that rat DNA contains a much larger proportion of contiguous repeats than does Xenopus (Chamberlin et al., 1975).

Implications of Sequence Organization for Gene Regulation

The Britten-Davidson model of gene regulation (Britten and Davidson, 1969; Davidson and Britten, 1973) proposes, in part, that the existence of a few contiguous repeats might provide the basis for coordinate control of certain structural genes. The organization of rat DNA is consistent with this hypothesis.

Comparison of Methods of Study of Sequence Organization

Chamberlin, Britten and Davidson (1975) point out that the electron microscopic determination of repetitive sequence length affords a direct measure of the distribution of sequence lengths. In contrast, the physical chemical approach offers ease and an accurate estimate of the overall population of sequences. This comparison is applicable also to the EM and physical chemical approaches to sequence organization. The former gives a direct measurement of individual repetitive and unique sequence lengths, while the latter provides a good estimate of the mean unique sequence length.

Pearson, Wu and Bonner (submitted) used nuclease and hydroxyapatite binding procedures to estimate the repetitive sequence length as 300 nucleotides and the average interrepeat spacing as 1800-2000 nucleotides. The agreement between the two measurements of repetitive sequence length in the rat is good, as Chamberlin, Britten and Davidson (1975) also found in *Xenopus*.

The applicability of the EM method appears to be quite general, though not in organisms like *Drosophila*, whose genomic organization is unsuited to this approach. The preparation of a driver repetitive fraction of sufficient length would be a great technical difficulty in this case.

Experimental Procedures

Preparation of DNA

Long DNA was isolated from rat ascites cells as described in Chapter 1. Shearing of the 0.9 and 2.5 kb drivers was performed in a Virtis homogenizer at 15000 and 2000 rpm respectively at 5°C in 0.05 M neutral phosphate buffer. Alkaline denaturation of the DNA followed the method of Manning, Schmid and Davidson (1975) generally. All incubations and HAP chromatographic procedures were conducted at 60°C in 0.12 M phosphate buffer. Elution of bound DNA from HAP was with 0.12 M Na_3PO_4 at 60°C. Details about the pedigree of the 2.5 kb driver repetitive fraction are given in Chapter 1.

Electron Microscopy

DNA was dialysed against 0.01 M Tris, 0.001 M EDTA (pH 8.5), made 50% in formamide and spread for microscopy by the modified Kleinschmidt technique of Davis, Simon and Davidson (1971). The DNA was visualized in a Philips EM 201 at an accelerating voltage of 60 kV. The grids contained single and double stranded PhiX174 circular DNA molecules as internal length standards of known length 5.25 kb. The ratio of the measured single to double strand length of a given DNA strand region was 0.97, so we treated single strand and duplex measurements without distinction. Micrographs were recorded on 35 mm film which was projected directly onto the electro-sensitive platen of a Hewlett-Packard electronic digitizer.

Digitizer measurements were stored in a PDP-10 computer for analysis.

Data Analysis

The photographic data were translated into a numerical code for computer analysis. Each molecule was considered to be a repeating array of single strand and duplex regions, repetitive hybrids having tails and foldback duplexes none. Each element of the array was measured as an ordered quadruplet consisting of the single strand region bounding the duplex on an arbitrarily chosen side, the first tail's length, the duplex length and the second tail's length. Foldbacks are considered to have tails of length zero. The computer can identify and match structural features by the particular permutation of zero and non-zero numbers in each quadruplet.

References

Angerer, R.C., Davidson, E.H., and Britten, R.J. (1975). Cell 6, 29-39.

Britten, R.J., and Davidson, E.H. (1969). Science 165, 349-357.

Cech, T.R., and Hearst, J.E. (1975). Cell 5, 429-446.

Chamberlin, M.E., Britten, R.J., and Davidson, E.H. (1975). J. Mol. Biol. 96, 317-333.

Davidson, E.H., and Britten, R.J. (1973). Quart. Rev. Biol. 48, 565-613.

Davidson, E.H., Galau, G.A., Angerer, R.C., and Britten, R.J. (1975). Chromosoma 51, 253-259.

Davidson, E.H., Hough, B.R., Amenson, C.S., and Britten, R.J. (1973). J. Mol. Biol. 77, 1-23.

Davis, R.W., Simon, M., and Davidson, N. (1971). In Methods in Enzymology (grossman, L. and Moldave, K., eds.), pp. 413-428, Academic Press, New York.

Firtel, R.A., and Kindle, K. (1975). Cell 5, 401-411.

Graham, D.E., Neufeld, B.R., Davidson, E.H., and Britten, R.J. (1974). Cell 1, 127-137.

Holmes, D.S., and Bonner, J. (1974). Biochemistry 13, 841-848.

Manning, J.E., Schmid, C.W., and Davidson, N. (1975). Cell 4, 141-155.

Schmid, C.W., and Deininger, P.L. (1975). Cell 6, 345-358.

Schmid, C.W., Manning, J.E., and Davidson, N. (1975). Cell 5, 159-172.

Walbot, V., and Dure, L.S. (1976). J. Mol. Biol. 101, 503-536.

Wilson, D.A., and Thomas, C.A. (1974). J. Mol. Biol. 84, 115-144.

Wu, J.-R., Hurn, J., and Bonner, J. (1972). J. Mol. Biol. 64, 211-220.

Table 1. Recovery of Repetitive DNA Strands Hybridized to Long DNA

Experiment	Duplexes Terminated by 2 Single Strand Tails[a]				
	Number of Strands Scored	Number Average Strand Length (kb)	Number Average Tail Length (kb)	Number[b] of Tails Proximal to Contiguous Repeats	Number Average Length of Tails Proximal to Contiguous Repeats (kb)
Long vs. 0.9 kb Repeat DNA	296	1.314	0.490	54	0.496
Long vs. 2.5 kb Repeat DNA	618	2.392	0.937	64	0.865

Footnotes to Table 1.

[a] These data apply to cases we interpreted to be repetitive hybrid duplexes consisting of a duplex region terminated at both ends by single strand unique DNA tails radiating from the long backbone strand of the heteroduplex structure. In cases wherein the duplex region between two tails could not be assigned a strandedness visually, it was assumed that this region is double-stranded provided the tails are in reasonably close adjacence to each other. In cases in which there was ambiguity of interpretation, e.g., an odd number of tails close together, one likely interpretation was recorded, assigning the odd tail to the one-tailed duplex category. In most cases no such ambiguities occur.

[b] This category includes the two tails bounding the junction between two contiguous repeats.

Table 2. Recovery of Duplex Structures Between Repetitive and Long DNA Strands

Experiment	Number Scored	Total Duplex DNA (kb)[a]	% of Long DNA in Duplex[b]	Number Average Length (kb)	Weight Average Length (kb)
Long vs. 0.9 kb Repeat DNA	327	110	10.2	0.336	0.473
Long vs. 2.5 kb Repeat DNA	652	345	12.9	0.529	1.542

Footnotes to Table 2.

[a] This is the sum of the total contour length of repetitive duplex DNA found between two single strand tails; thus, it excludes any contribution from one-tailed structures.

[b] The divisor in this computation is the total contour length of all the long tracer strands measured; this number is given in Table 5.

Table 3. Recovery of Looped Foldbacks from Long DNA Strands

Experiment	Number of Duplexes	Number Average Duplex Length (kb)	Total Fold-back DNA (kb)	% of[a] Total DNA
Long vs. 0.9 kb Repeat DNA	13	0.400	10.4	1.0
Long vs. 2.5 kb Repeat DNA	28	0.442	24.8	0.9

Footnotes to Table 3

[a] The value used for the total DNA was the combined contour length of all the long tracer strands, as cited in Table 5.

Table 4. Spacings between Duplexes Terminated by Two Single Strand Tails

Experiment	Class of Spacing[a]	Number Scored	Number Average	Weight Avearage
Long vs. 0.9 kb Repeat DNA	Smaller than 0.3 kb	82	0.192	0.214
	Larger than 0.3 kb	148	1.157	5.506
	Total	330	0.813	5.060
Long vs. 2.5 kb Repeat DNA	Smaller than 0.3 kb	105	0.204	0.223
	Larger than 0.3 kb	396	1.825	4.921
	Total	501	1.485	4.786

Footnotes to Table 4.

[a] The total spacing measurements were simply sorted into those greater and less than the value of 0.3 kb to form two separate data pools, which were analysed independently to yield the numbers cites for the less and greater than 0.3 kb categories.

Table 5. Sequence Organization of Rat DNA[a]

Experiment		Type of Organization				
		Interspersed Repeat and Unique Sequence	Interspersed Repeat, Fold-back and Unique Sequence	Interspersed Foldback and Unique Sequence	Unique Sequence	Total[b] DNA
Long vs. 0.9 kb Repeat DNA	Number Scored	19	28	8	16	71
	Mean Length (kb)	15.623	12.071	20.360	17.206	15.117
	DNA in Type (kb)	297	338	163	275	1073
	% of Total	27.7	31.4	15.2	25.7	
Long vs. 2.5 kb Repeat DNA	Number Scored	64	60	7	8	139
	Mean Length (kb)	16.560	21.967	21.047	19.834	19.308
	DNA in Type (kb)	1060	13.8	147	159	2684
	% of Total	39.5	49.1	5.5	5.9	

Footnotes to Table 5

[a] Each of the long strands analysed in the various experiments was sorted into four categories according to whether they possess (1) only repetitive hybrids, (2) repetitive structures and foldbacks, (3) only foldbacks and unique sequences or (4) no duplex structures at all. The statistics generated in the table were calculated from the separate data pools.

[b] The data in this column pertain to all the long tracer strands scored.

Figure Legends

Figure 1. Heteroduplex molecules between 20 kb total DNA and 0.9 kb repetitive sequence-bearing strands

Long total DNA strands were reacted to C_ot 50 with a 100 fold mass excess of 0.9 kb repetitive sequence-bearing DNA. Hybridization was at 60°C in 0.12 M phosphate buffer. The sample was spread for electron microscopy from 50% formamide at room temperature. Long strands were located by scanning the grids at random and scoring any strands substantially greater than the driver molecules. A bar in the composite micrograph shows the length corresponding to 1 kb. Below are scale line drawings of our interpretations of the structures. Single lines represent DNA single strands; double, cross-hatched lines show duplex DNA. Duplexes without tails are foldbacks. In the case of apparent one-tailed repetitive hybrids a short duplex is drawn ended by a dashed line. This was done for illustrative purposes, and no duplex measurements from one-tailed structures were attempted.

Figure 2. Heteroduplexes between long and 0.9 kb repetitive DNA, including hybrid structures on tails

(a) Molecules from the same grids as described in the legend to Figure 1, but including driver strand/tail hybrids.

(b) Scale interpretations of the molecules in (a) by the conventions adopted in the legend to Figure 1.

Figure 3. Heteroduplex molecules between 20 kb total DNA strands and 2.5 kb repetitive sequence-bearing DNA, including long and short repetitive duplexes

20 kb total DNA strands were reacted to C_ot 50 with a 120 fold mass excess of 2.5 kb repetitive DNA. Hybridization and microscopy procedures, as well as conventions used in drawing our interpretations are as in the legend to Figure 1.

Figure 4. Heteroduplexes between long and 2.5 kb repetitive DNA, including hybrid structures on tails.

(a) These molecules are from the same grids described in the legend to Figure 3, but include duplex structures between driver strands and tails.
(b) Scale drawings of our interpretations of the molecules in (a). Conventions followed are as in the legend to Figure 1.

Figure 5. Duplex Lengths of the repetitive sequences in rat DNA.

(a) Duplex lengths measured on long strands driven to C_ot 50 by 0.9 kb repetitive DNA are recorded. DNA reassociated to C_ot 50 as described in the text was placed at the end of reaction in a second incubation mixture containing a 1700 fold mass excess of freshly alkali-denatured and reneutralized 0.9kb repetitive DNA. Reaction was to C_ot at 60°C in 0.12 M phosphate buffer, and microscopy procedures were as in Figure 1. (b) Duplex lengths of repeats measured on the 2.5 kb repeat DNA/long DNA hybrids described in Figure 3.

Figure 6. Spacings between two repetitive sequences

(a) These data were derived from the molecules in the 0.9 kb repeat/long DNA experiment, by measuring the single strand region bounded by the proximal tails of two two-tailed hybrids. The interval size is 150 nucleotides, and the frequency is the actual number of spacing measurements made.

(b) As in (a) for the 2.5 kb repeat/long DNA hybrids. A few spacings larger than 7.5 kb ranging as high as 30 kb are not shown.

(c) This histogram displays the distribution of spacings from the 0.9 kb repeat/long DNA experiment as in (a) but excluding all spacings smaller than 0.3 kb in magnitude.

(d) As in (c) for the 2.5 kb repeat/long DNA hybrids. A few spacings larger than 7.5 kb ranging as high as 30 kb are not shown.

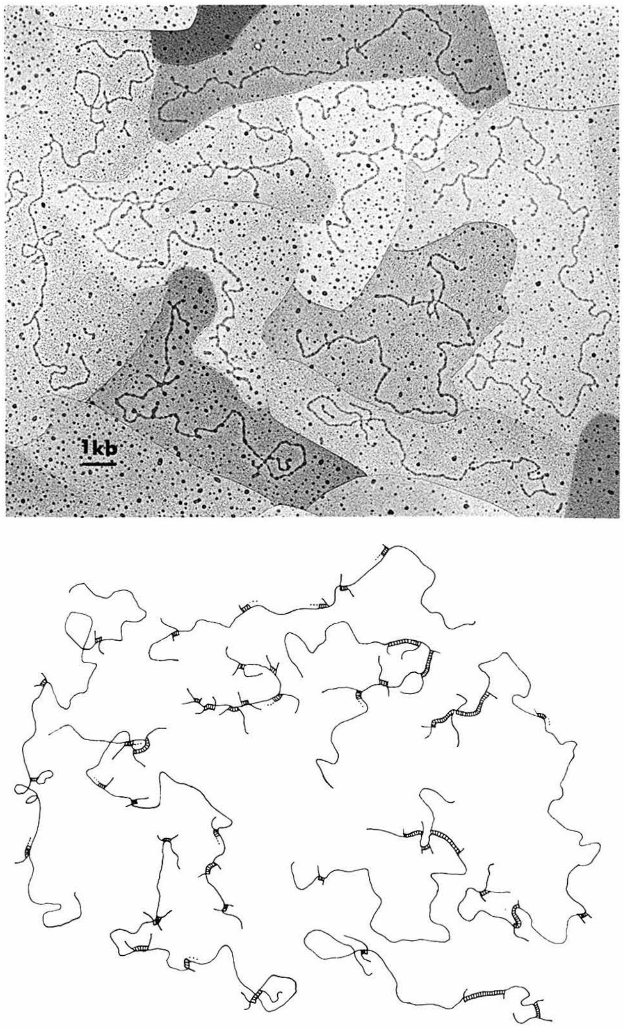


Fig. 1

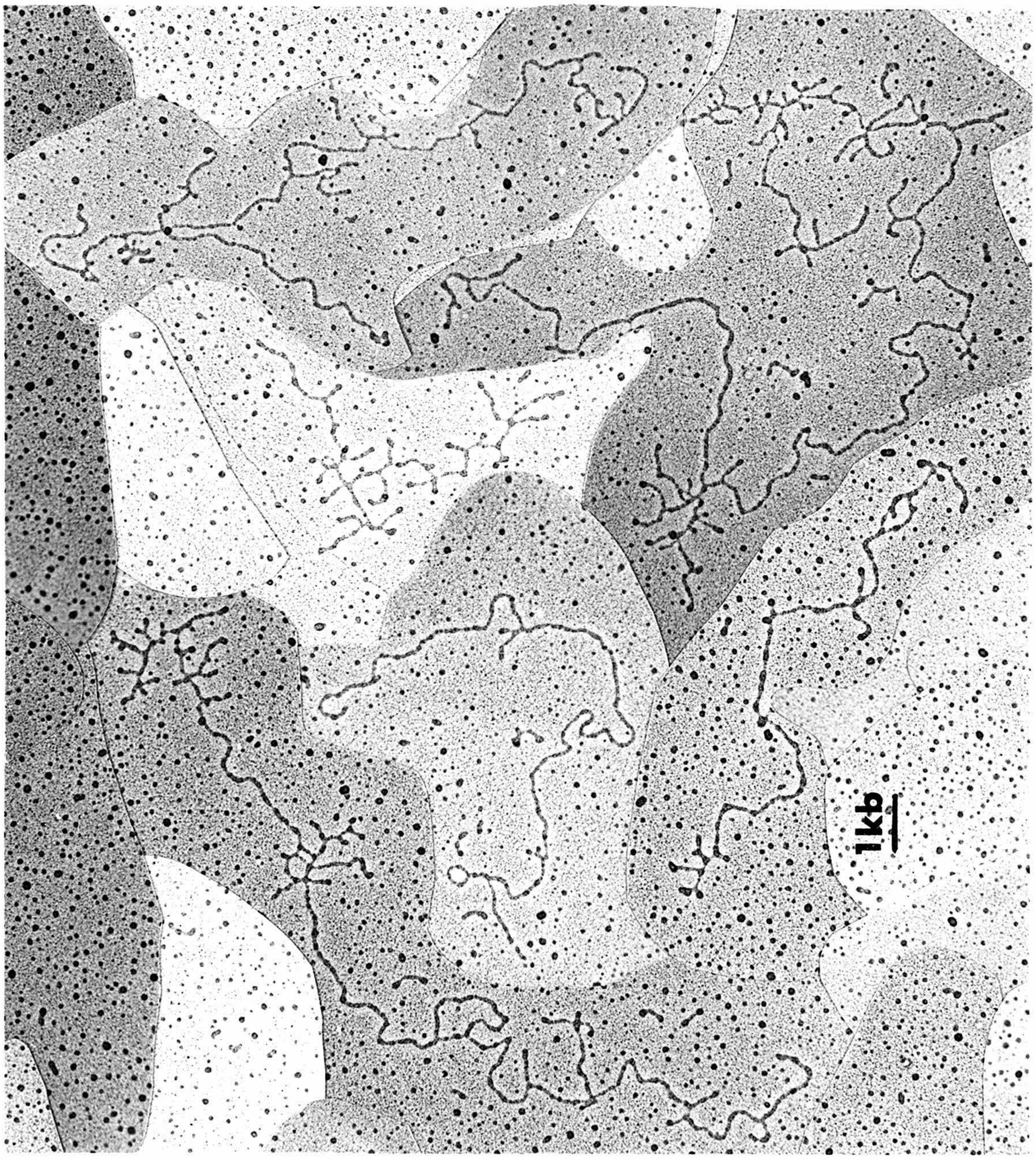
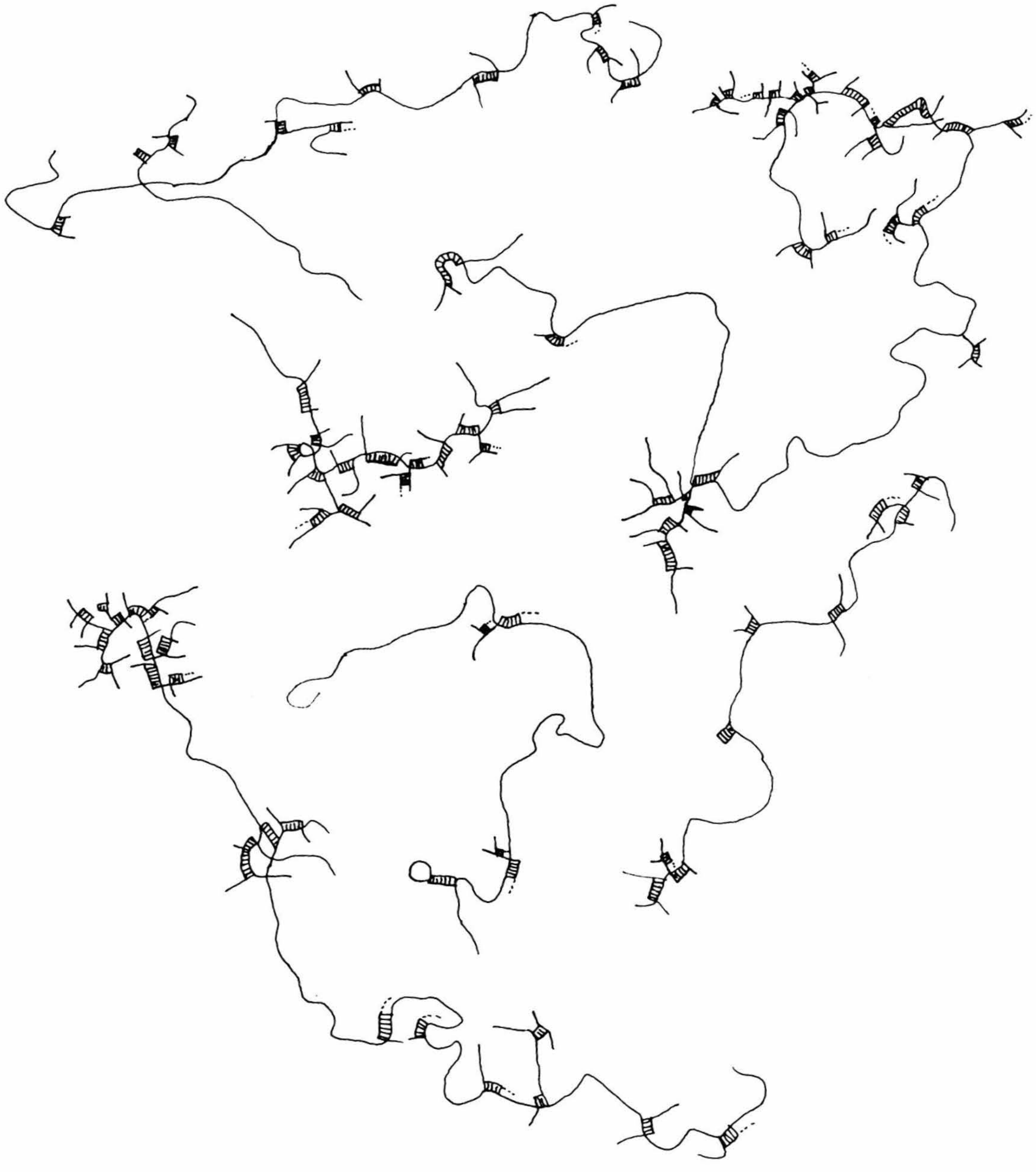


Fig. 2 (a)



(b)

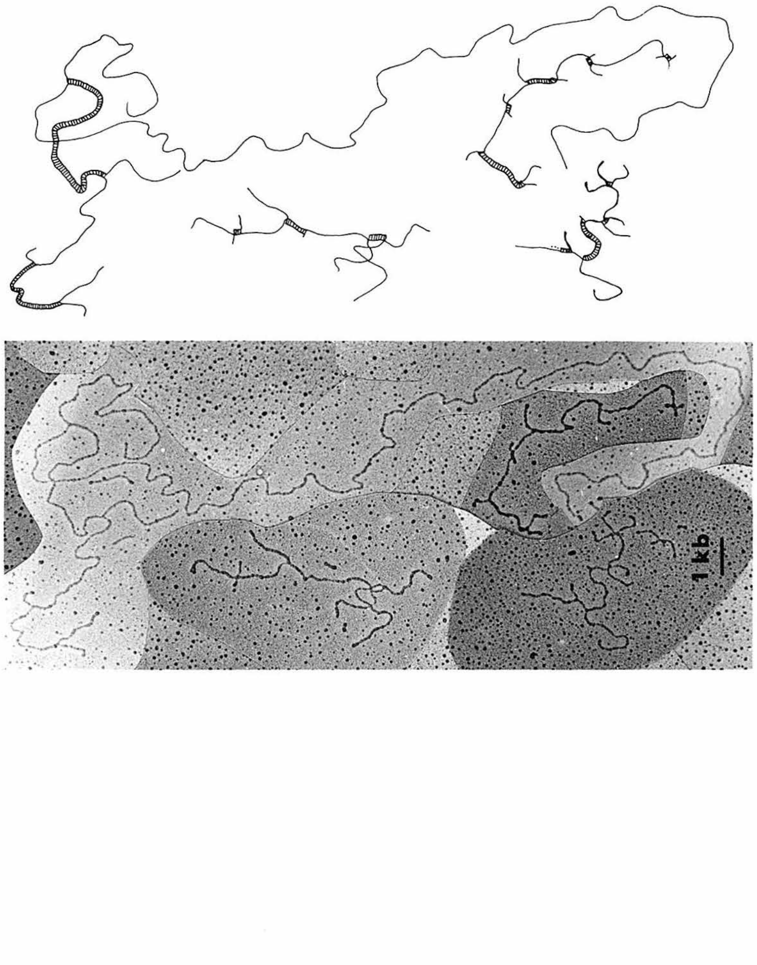


Fig. 3

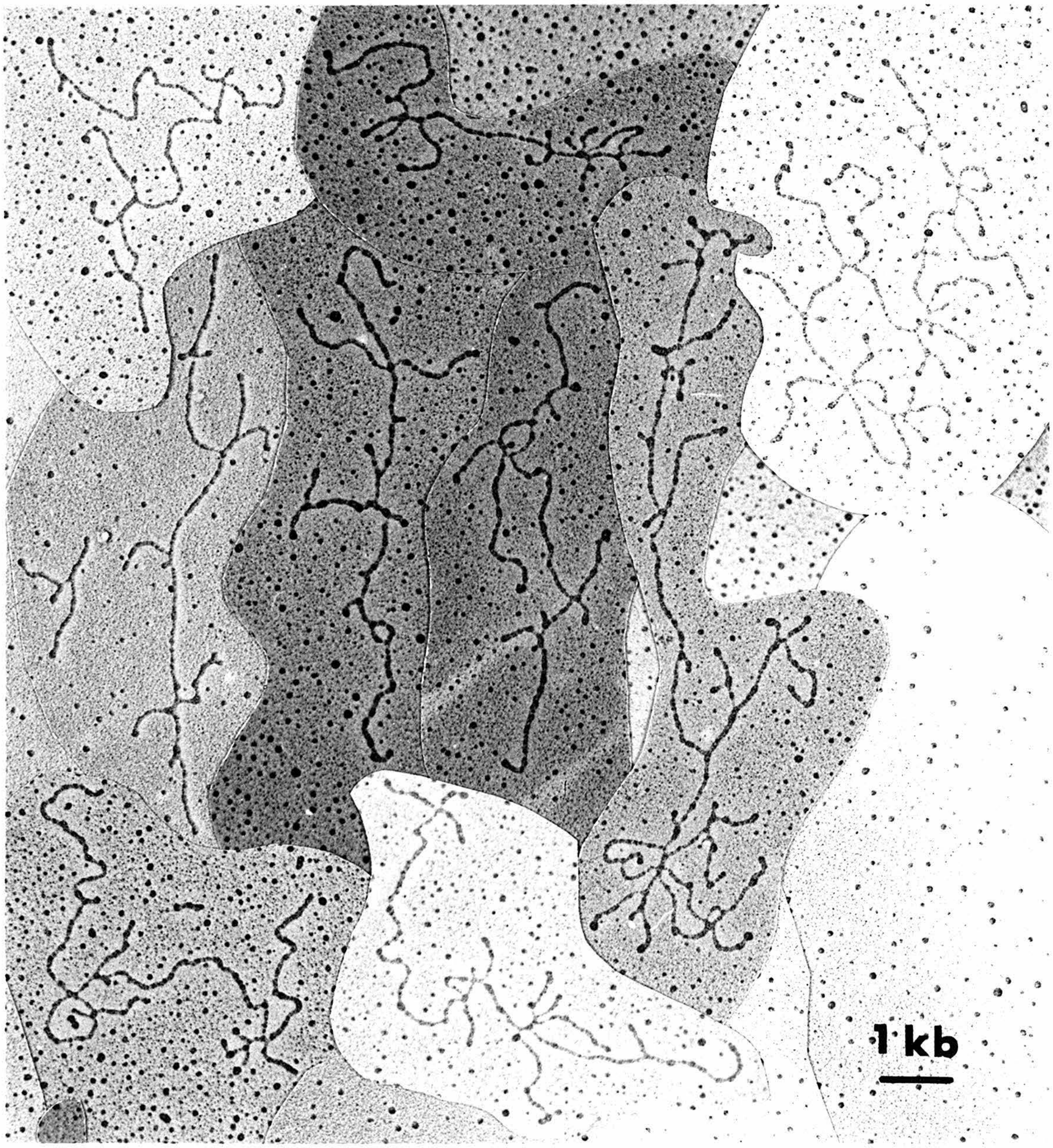
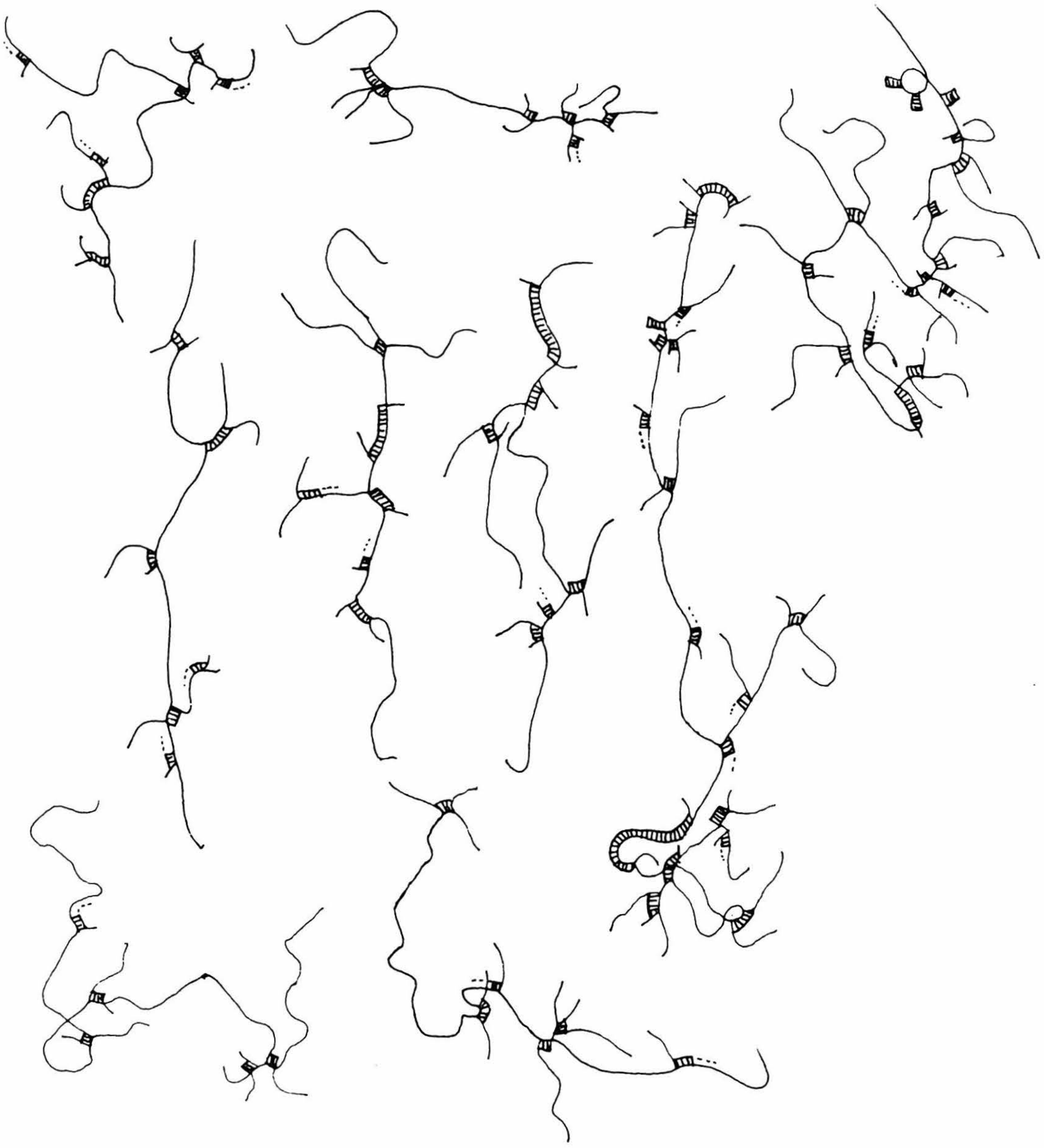


Fig. 4 (a)



(b)

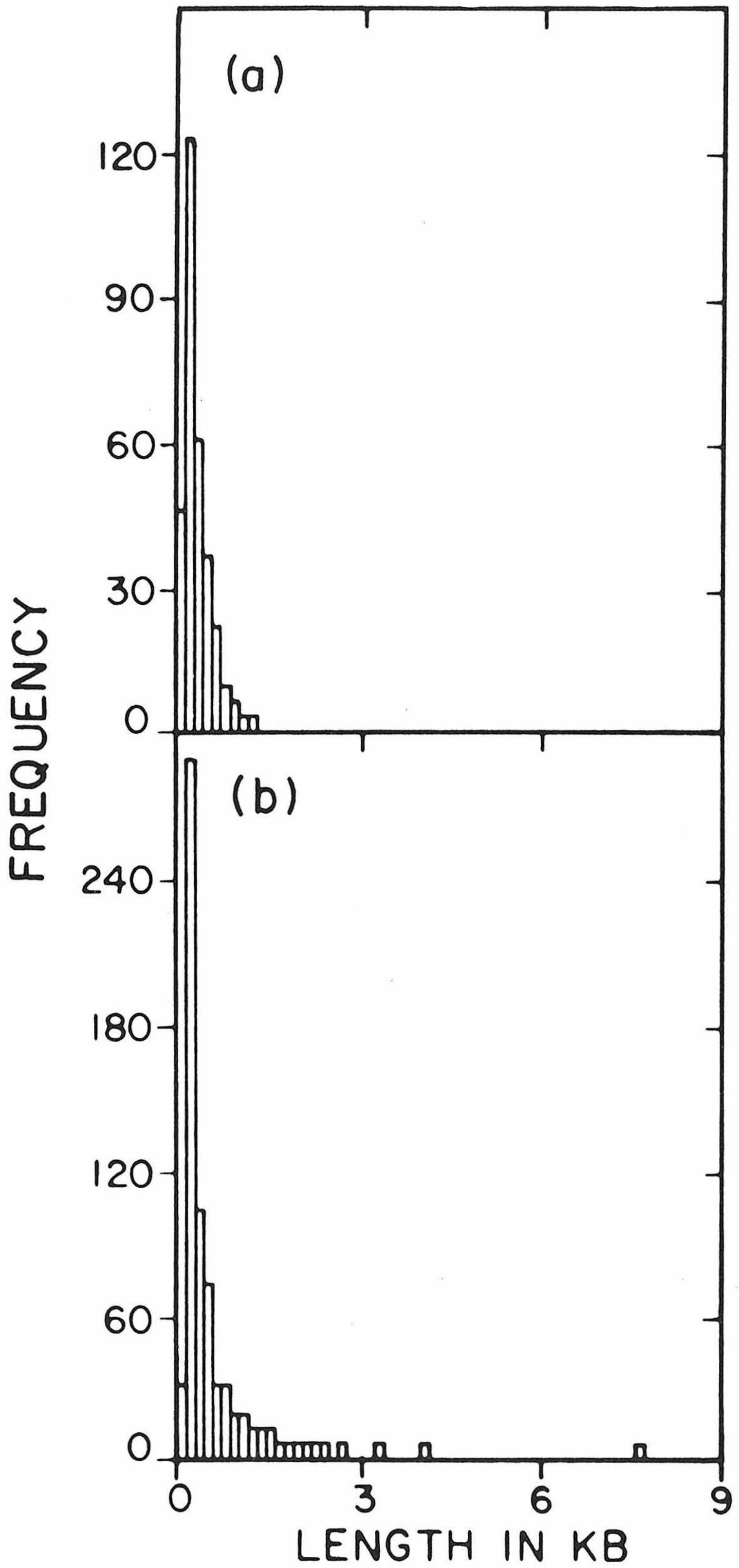


Fig. 5

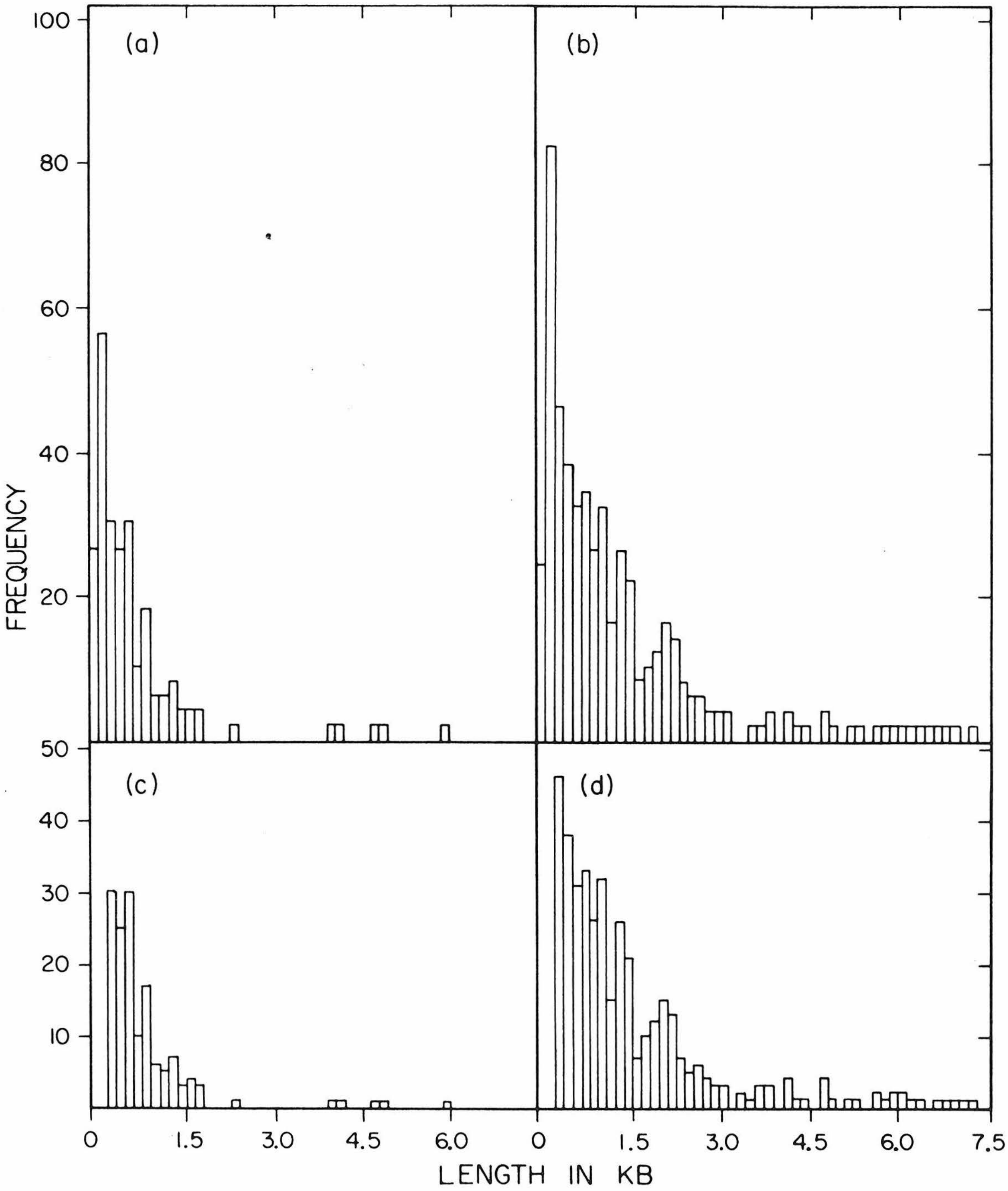

(a)
(b)
(c)
(d)
FREQUENCY
100
80
60
40
20
50
40
30
20
10
0
1.5
3.0
4.5
6.0
7.5
LENGTH IN KB

Fig. 6

Chapter 3.

Foldback Sequences in Rat DNA

Summary

The foldback sequences in rat DNA were characterized in the electron microscope. 30 kb DNA single strands were renatured for 1 min at low concentration and spread for electron microscopy from 50% formamide. Visualization of the briefly renatured molecules, whose foldback sequences have formed hairpin and looped hairpin duplex structures, permits measurement of the lengths of the foldbacks, the spacings between complementary foldback sequences, viz., the loop lengths, and the spacings between adjacent foldback duplexes. Our data indicate that 4.4% of rat DNA is comprised of foldback sequences numbering 1.9 x 10^5 per genome. Foldback duplexes are interspersed throughout at least 73% of the DNA at an average spacing of 9.7 kb. The mean loop length was 5.1 kb. While the average foldback sequence length was 0.71 kb, two size classes could be distinguished: (1) a numerous class of short foldbacks 300 nucleotides long on average comprising 39% of the total foldback DNA and (2) a sparse class of large foldbacks of 6.1 kb average length composing the other 61%.

Introduction

DNA single strands containing a sequence and the inverted repeat of that sequence form hairpin or looped hairpin structures, termed foldbacks, under renaturing conditions (Britten and Smith, 1970; Wilson and Thomas, 1974). 1.5-6% of the DNA of humans (Wilson and Thomas, 1974; Schmid and Deininger, 1975; Dott, Chuang and Saunders, submitted for publication), Drosophila (Schmid, Manning and Davidson, 1975), mouse (Cech and Hearst, 1975), Xenopus (Davidson et al., 1973; Perlman, Philips and Bishop, submitted for publication) and rat (Pearson, Wu and Bonner, submitted; herein) consists of foldback sequences. The function of these sequences is at present unknown, though some circumstantial evidence indicates a possible link to the processing of nuclear RNA.

This study seeks to discern the number, size and arrangement of the foldback sequences of rat DNA and to consider this information in relation to possible models of foldback function. We briefly renatured long DNA single strands, mounted them for electron microscopy from 50% formamide, and visualized the hairpin and looped hairpin foldback structures. We can measure the number and size of the foldback duplexes, the spacings between non-contiguous complementary inverted repeat sequences, i.e., the loop lengths, and the spacings between renatured foldback duplexes.

Results

HAP Binding Properties of Rat Foldback DNA

The extent of binding of animal DNA fragments of increasing length to hydroxyapatite after brief renaturation rises over an appreciable range of strand lengths (Wilson and Thomas, 1975; Schmid, Manning and Davidson, 1975). This follows from the interspersion of renatured foldback duplexes with unique and repetitive sequences (Schmid and Deininger, 1975; Chapter 2), such that ever greater quantities of DNA flanking the foldbacks are bound with them to HAP. With the exception of Xenopus (Davidson et al., 1973), 50-80% of animal foldbacks studied are looped; that is, the inverted complementary sequences are separated by a non-complementary spacer sequence. Thus, increasing fragment length augments the probability of formation of looped foldbacks and thereby the observed increase in HAP binding.

Holmes and Bonner (1974a) reported that 10% of 450 nucleotide rat DNA fragments bind to HAP at very low C_ot. Table 1 lists the binding properties of several larger length classes. The various DNA preparations were alkaline denatured, renatured for 1 min at DNA concentrations of 25-100 micrograms/ml and applied to HAP at 60°C. Binding climbs sharply until strand lengths exceed 3.3 kb. Thereafter, binding increments diminish so that but 2% difference in HAP retention distinguishes 8.1 and 29.4 kb fragment classes.

Hairpin Structures in High Molecular Weight Unfractionated DNA

Unsheared, unfractionated DNA single strands at a concentration of 25 micrograms/ml were renatured for 1 min, quenched by addition of 4 volumes of ice-cold deionized H_2O, and spread for electron microscopy.

The long single strands displayed numerous hairpin structures, looped and unlooped, as shown in Figures 1, 2 and 3. 396 long strands of 29.4 kb number average length were analysed. Data on foldback structures studied are tabulated in Table 2. The preponderance of duplex structures consisted of looped and unlooped foldbacks, though a small class of foldback structures formed within the loops of other foldbacks. Also, a small class of apparent intermolecular hybrids were in evidence. The latter structures consisted of four-ended "H"-type structures (cf. Chapter 1). These may have arisen in part from the breakage of foldback loops, as the maximum C_ot of 0.005 reached in this experiment is not favorable for interstrand duplex formation; or this could be due to a small amount of genuinely highly repetitive sequences.

The percent of the total DNA recovered as foldback duplex was 4.4%. The duplexes ranged in size from the smallest measurable to 30 kb as depicted in Figure 4. The distribution of foldback lengths is bimodal, in that it consists of a numerous class of sequences 100-600 nucleotides long and a sparse, disperse class encompassing the remaining foldbacks. If, for the purpose of comparison to hydroxyapatite data, we consider the foldback duplexes respectively less and greater than 2 kb independently, then the short class averaging 300 nucleotides in length contains 39% of the foldback duplex DNA. The other 61% falls to the long category of 6.1 kb mean length. There are 1.9×10^5 foldback duplexes in the genome. The overall average inverted repeat sequence size is 0.71 kb.

The distribution of spacings between proximal termini of renatured foldback duplexes appears in Figure 5. Foldbacks are interposed by single-strand regions ranging from the very short to greater than 60 kb. The mean length of the strands among which the spacings were measured, i.e., those bearing at least two foldbacks, was 45.0 kb. The average inter-foldback spacing is 9.7 kb (Table 2); therefore, it is unlikely that a strong bias toward shorter spacing has influenced the data. That there is no obvious relationship between spacings and the lengths of the strands on which they were measured can be seen in Figure 6. The relationship between spacing and strand length appears to be random.

The strands having at least 1 foldback comprise 73% of the total DNA. We conclude that, at the criterion of 29.4 kb strand length, foldbacks are scattered throughout at least 73% of the genome. However, this figure is a minimum estimate, since the strands devoid of foldback duplex averaged only 19.8 kb long by number.

54% of the foldbacks scored possessed a measurable loop. The looped duplexes are somewhat longer on average than the simple hairpin structures (Table 2). Figure 7 exhibits the distribution of loop lengths. The mean of this distribution is 5.1 kb (Table 2). There is no obvious relationship between the foldback duplex length and loop length (Figure 8). In 58% of the cases a duplex shorter than 500 nucleotides was terminated by a loop less than 4 kb.

Hairpin Structures in HAP-fractionated DNA

Further electron microscopic data were gathered from the HAP fractionation products of 29.4 kb foldback DNA. The strands were renatured as before and fractionated on HAP. The bound and unbound fractions were separately mounted for microscopy and analysed as before. The bound fraction displayed the various types of hairpin structures shown in Figures 9, 10 and 11. Table 3 shows the recoveries of duplex structures from these fractions. The product of the percent foldback duplex in the bound fraction (Table 3) and the fraction of the genome constituted by this fraction (Table 1), 4.9%, is the amount of foldback DNA recovered. This is in good agreement with the 4.4% estimate from the unfractionated DNA experiment. The average foldback size measured here, 0.61 kb (Table 3), is also comparable to that derived from the results with unfractionated DNA. The parameters of spacing and loop length from the bound DNA cannot be compared to the unfractionated DNA results, however, because the bound DNA molecules had suffered a reduction of mean strand length to 7.7 kb. This length reduction means that these results must tend to deemphasize the longer spacings and loop lengths observed among the 29.4 kb unfractionated DNA strands.

The unbound molecules, 4.1 kb long on average, also contained a few apparent foldback structures totaling 0.4% of whole DNA. These structures are substantially shorter in duplex length than the mean of the bound foldbacks and include a disproportionate number of looped foldbacks (Table 3). Possibly short duplexes with attached loops are somewhat refractory to HAP binding. Alternatively, some of these might be artifacts of spreading for electron microscopy, in that the DNA may occasionally twist in a manner spuriously similar to the configuration of a short, looped foldback. Cech and Hearst (1975) have also observed this phenomenon.

Comparison of Electron Microscopic to Physical Chemical Results

Pearson, Wu and Bonner (submitted) have found with S1 nuclease experiments that 6% of rat DNA is duplex by C_ot 0.05. This agrees with the 5% duplex estimate from the electron microscope data (Table 2). By separating the foldback duplexes shorter and longer than 2-3 kb with agarose A-50 chromatography, these authors find 32 and 68% respectively of the foldback duplex to belong to the shorter and longer categories. This can be compared to the 39-61% distribution determined by EM with the 29.4 kb unfractionated strands.

The hydroxyapatite binding data (Table 1) indicated that strands increasingly longer than 8 kb exhibited virtually no augmentation in HAP retention. Our electron microscopic measurements show that 8 kb is roughly equal to the average inter-foldback spacing and substantially longer than the 5.1 kb average loop length. It would not be anticipated that more DNA would bind among fragment lengths greater than 8 kb. The agreement between the observed and calculated binding of the 29.4 kb DNA strands (Table 1) is good in view of the knowledge that extensive breakage of the fractionated DNA probably occurred during chromatography (Table 3).

Discussion

A Model of Foldback Organization in Rat DNA

We propose this model of the arrangement of foldback sequences in the rat. The 4.4% of rat DNA constituted by foldback sequence is divided 39% among a numerous class of sequences an average of 300 nucleotides long and 61% among foldbacks at least 6.1 kb on average. However, the parameters describing the larger foldback class are somewhat uncertain, in that these sequences are few in number (Figure 4). Further, extremely long foldback sequences would appear as linear molecules and go undetected by the present technique. Therefore, the true average size and proportion of the total foldbacks of this larger class may be greater than the figures cited. Foldback pairs are interspersed throughout at least 73% of the DNA at an average spacing of 9.7 kb; the minimum average spacing between non-contiguous complementary inverted repeats is 5.1 kb. Foldbacks are interspersed with both repetitive and unique sequences in half the genome (Chapter 2).

Rat DNA may contain as little as none and as much as 0.6% highly repetitive DNA from these findings.

Comparison to Other Organisms

These data conform to an emerging pattern of foldback sequence organization in a number of animal DNAs. 1.5-6.0% of the DNA consists of foldback sequences an average of 0.5-1.4 kb long interspersed throughout 1/3-3/4 of the genome (Schmid, Manning and Davidson, 1975; Cech and Hearst, 1975; Schmid and Deininger, 1975; Dott, Chuang and Saunders, submitted for publication). In mouse, Drosophila and rat

DNA 50-80% of the foldbacks have loops an average of 3-6 kb in length. The foldbacks of Xenopus may have very few loops (Davidson et al., 1973). Throughout much of the human (Schmid and Deininger, 1975) and rat (Chapter 2) genomes foldbacks are interspersed with repetitive and unique sequences. This must be true of certain other organisms as well from the wide distributions of both interspersed repetitive and interspersed inverted repeat sequences.

Implication of Foldback Organization for Function

Foldback sequences are present in heterogeneous nuclear RNA (Ryskov, 1973; Jelinek et al., 1974; Molloy et al., 1974) and either disappear at the mature message stage (Ryskov et al., 1973) or endure in mRNA molecules as a single complement of the foldback sequence in hnRNA (Naora and Whitelam, 1975). hnRNA molecules are populated by interspersed repetitive and unique sequences (Holmes and Bonner, 1974b; Smith et al., 1974). Suggestion that the foldback sequence in hnRNA molecules might be a site for processing of the RNA has been made (Naora and Whitelam, 1975). The fact that half the rat genome consists of closely interspersed repetitive, inverted repeat and unique sequences substantiates that much of the DNA is available for transcription into RNA molecules of the construction described. Moreover, the spacings between foldbacks are on the order of tens of kilobases. This is the same size range as that of rat hnRNA molecules (Holmes and Bonner, 1974a). These results therefore are consistent with models of foldback function like that above.

Experimental procedures

Preparation of DNA

DNA was prepared from rat ascites cells by standard methods including extensive treatment with protease and RNAase and organic extraction with 25:24:1 phenol/chloroform/isoamyl alcohol. DNA was alkaline denatured by the method of Manning, Schmid and Davidson (1975). Renaturation was at 60°C in 0.12 M neutral phosphate buffer. All hydroxyapatite chromatographic procedures were carried out at this temperature. The foldback fractions were eluted from the column with 0.12 M Na_3PO_4 (pH 12.3).

Electron Microscopy

DNA samples were dialysed 4 hr against 0.01 M Tris, 0.001 M EDTA (pH 8.5), made 50% in formamide and spread for microscopy by the modified Kleinschmidt technique of Davis, Simon and Davidson (1971). Visualization was in a Philips EM 201 electron microscope at an accelerating voltage of 60 kV. Micrographs were recorded on 35 mm film; negatives were projected directly onto the electro-sensitive platen of a Hewlett-Packard electronic digitizer. Digitizer measurements of strand length were stored in a PDP 10 computer for analysis.

Data Analysis

Molecular information was translated into quadruplet code for computer analysis, as described in Chapter 2.

References

Britten, R.J., and Smith, J. (1970). Carnegie Institution of Washington Year Book 68, 378-386.

Cech, T.R., and Hearst, J.E. (1975). Cell 5, 429-446.

Davidson, E.H., Graham, D.E., Neufeld, B.R., Chamberlin, M.E., Amenson, C.S., Hough, B.R., and Britten, R.J. (1973). Cold Spring Harbor Symp. Quant. Biol. 38, 295-301.

Davis, R.W., Simon, M., and Davidson, N. (1971). In Methods in Enzymology (Grossman, L. and Moldave, K., eds), pp. 413-428, Academic Press, New York.

Holmes, D.S., and Bonnner, J. (1974a). Biochemistry 13, 841-848.

Holmes, D.S., and Bonner, J. (1974b). Proc. Nat. Acad. Sci. 71, 1108-1112.

Jelinek, W., Molloy, G., Munoz, R.F., Salditt, M., and Darnell, J.E. (1974). J. Mol. Biol. 82, 361-370.

Manning, J.E., Schmid, C.W., and Davidson, N. (1975). Cell 4, 141-155.

Mooloy, G.R., Jelinek, W., Salditt, M., and Darnell, J.E. (1974). Cell 1, 43-54.

Naora, H., and Whitelam, J.M. (1975). Nature 256, 756-759.

Ryskov, A.P., Saunders, G.F., Farashyan, V.R., and Georgiev, G.P. (1973). Biochim. Biophys. Acta 312, 152-164.

Schmid, C.W., and Deininger, P.L. (1975). Cell 6, 345-358.

Schmid, C.W., Manning, J.E., and Davidson, N. (1975) Cell 5, 159-172.

Smith, M.J., Hough, B.R., Chamberlin, M.E., and Davidson, E.H. (1974). J. Mol. Biol. 85, 103-126.

Wilson, D.A., and Thomas, C.A. (1974). J. Mol. Biol. 84, 115-144.

Table 1. Binding of DNA renatured to low C_0t to hydroxyapatite[a]

Fragment Length (kb)[b]	Fraction Bound	Maximum C_0t Reached[c]
1.1	0.20	0.15
3.3	0.53	0.20
8.1	0.60	0.16
29.4	0.62	0.06
29.4[d]	0.73	----
Calculated		

[a] DNA samples at a concentration of 25-100 micrograms/ml were alkaline denatured, renatured to a nominal C_0t of 0.005 and applied to hydroxyapatite at 60°C in 0.12 M phosphate buffer. Elution of the bound fraction was with 0.12 M Na_3PO_4 (pH 12.3) at 60°C. Yields from the columns were 95% or better as determined by optical density.

[b] This is the number average length determined by electron microscopy.

[c] This figure includes the dwell time on the column prior to absorption.

[d] This row gives the amount of the total DNA contour length measured composed of strands bearing at least one foldback structure.

Table 2. Electron microscopic analysis of 29.4 kb DNA renatured to low C_ot

Class	Number Scored	Total DNA in Class (kb)	% of Total DNA	Number Average	Weight Average
Long DNA Strands	396	11,600	100	29.421	47.630
Looped Fold-back Duplexes	176	156	2.6	0.884	7.689
Loopless Fold-back Duplexes	150	92	1.6	0.616	10.428
Total Fold-back Duplexes	326	241	4.2	0.740	8.910
Interstrand Duplexes	18	65	0.6	3.618	11.383
Loops	176	901	7.8	5.117	15.136
Foldback Duplexes in Loops	29	26	0.2	0.449	0.737
Loops of Fold-backs within Foldbacks	12	73	0.6	6.066	17.609
Spacings between Foldbacks	152	1471	12.7	9.679	21.246

Table 3. Electron microscopic analysis of DNA fractionated on hydroxyapatite at low C_ot

Experiment	No. of Duplexes	No. Ave. Duplex Length (kb)	Total Foldback DNA (kb)	% of Total DNA[a]	No. of Spacings	No. Ave. Spacing (kb)	No. of Loops	No. Ave. Loop Length (kb)	% of Foldbacks with Loops
Bound Looped Foldbacks	62	0.959	118.9	4.9	8	2.666	62	3.162	100
Bound Loopless Foldbacks	96	0.379	72.8	3.0	2	1.111	0	0	0
Bound Total Foldbacks	158	0.607	191.7	7.9	16	2.484	158	1.476	39.2
Unbound Looped Foldbacks	15	0.282	4.2	0.5	2	7.254	15	3.400	100
Unbound Loopless Foldbacks	9	0.284	5.1	0.6	0	0	0	0	0
Unbound Total Foldbacks	24	0.283	9.3	1.1	3	10.776	24	2.307	62.5

[a] This is the percent of the total DNA on the grid, not the percent of the total genomic DNA. 315 bound and 199 unbound long strands were scored.

Figure 1. Simple hairpin foldback structures among 29.4 kb DNA single strands.

After renaturation as described in the text and quenching with ice-cold H_2O, the DNA was dialyzed against 0.01 M Tris, 0.001 M EDTA (pH 8.5), made 50% in formamide and spread for electron microscopy. ØX174 circular DNA was present on the grids as a known internal length standard of 5.25 kb. Arrows on the micrograph point to the hairpin duplexes. A bar indicates the equivalent length of 1 kb.

Figure 2. Looped foldbacks among the 29.4 kb DNA strands.

These molecules containing looped foldbacks are from the same grids as described in Figure 1.

Figure 3. Multiple foldback duplexes on long single strands.

These molecules containing multiple foldback structures are from the same grids as described in Figure 1.

Figure 4. The distribution of foldback duplex lengths from the 29.4 kb single strands.

The frequency equals the actual number of measurements. The interval size is 80 nucleotides. The measurements are of the duplex length of the foldbacks, not of the combined length of the two complementary sequences.

Figure 5. Distribution of spacings between foldback duplexes among the 29.4 kb DNA strands.

The frequency is the number of spacings actually scored. Statistics concerning the distribution are given in Table 2. The interval size is 1 kb. Spacing measurements were between the proximal ends of two adjacent foldback duplexes.

Figure 6. Three-dimensional histogram of inter-foldback spacings and the lengths of strands on which they are found.

Each measurement of spacing between two foldback duplexes was matched with the length of the single strand that contains the foldbacks. The ordered pairs are separated by computer into two-dimensional intervals of strand length and spacing length in the xy plane of the figure. The height of a given box in the z space equals the number of ordered pairs falling into the interval occupied by that box. The figure suggests no trivial relationship between the spacing measurements and the strand lengths. These measurements were made among the 29.4 kb strands.

Figure 7. The distribution of loop lengths among the 29.4 kb DNA strands.

The frequency equals the number of measurements scored. The interval size is 1 kb. Statistics on this distribution appear in Table 2.

Figure 8. The relationship between foldback duplex length and attached loop length.

In each case of a looped foldback, the duplex length and the loop length were paired and plotted. The measurements were derived from the 29.4 kb DNA.

Figure 9. Simple hairpin structures from hydroxyapatite bound fraction.

29.4 kb DNA was renatured for 1 min and fractionated on hydroxyapatite, as described in the text. DNA was prepared for microscopy as described in the legend to Figure 1. The DNA pictured is from the bound fraction.

Figure 10. Looped foldbacks among the hydroxyapatite bound molecules.

See Figure 9.

Figure 11. Hydroxyapatite bound strands bearing multiple foldback structures.

See Figure 9.

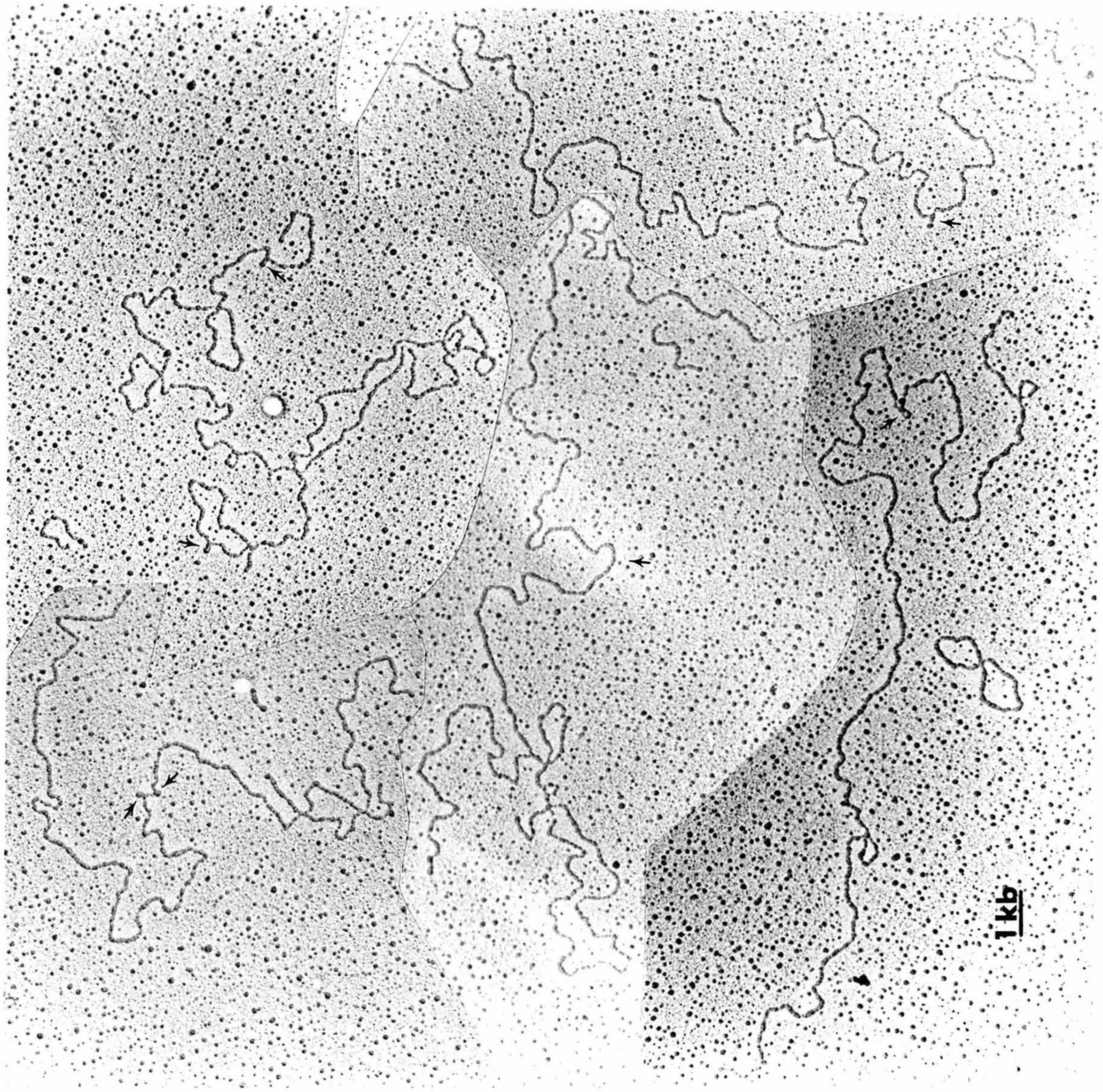


Fig. 1

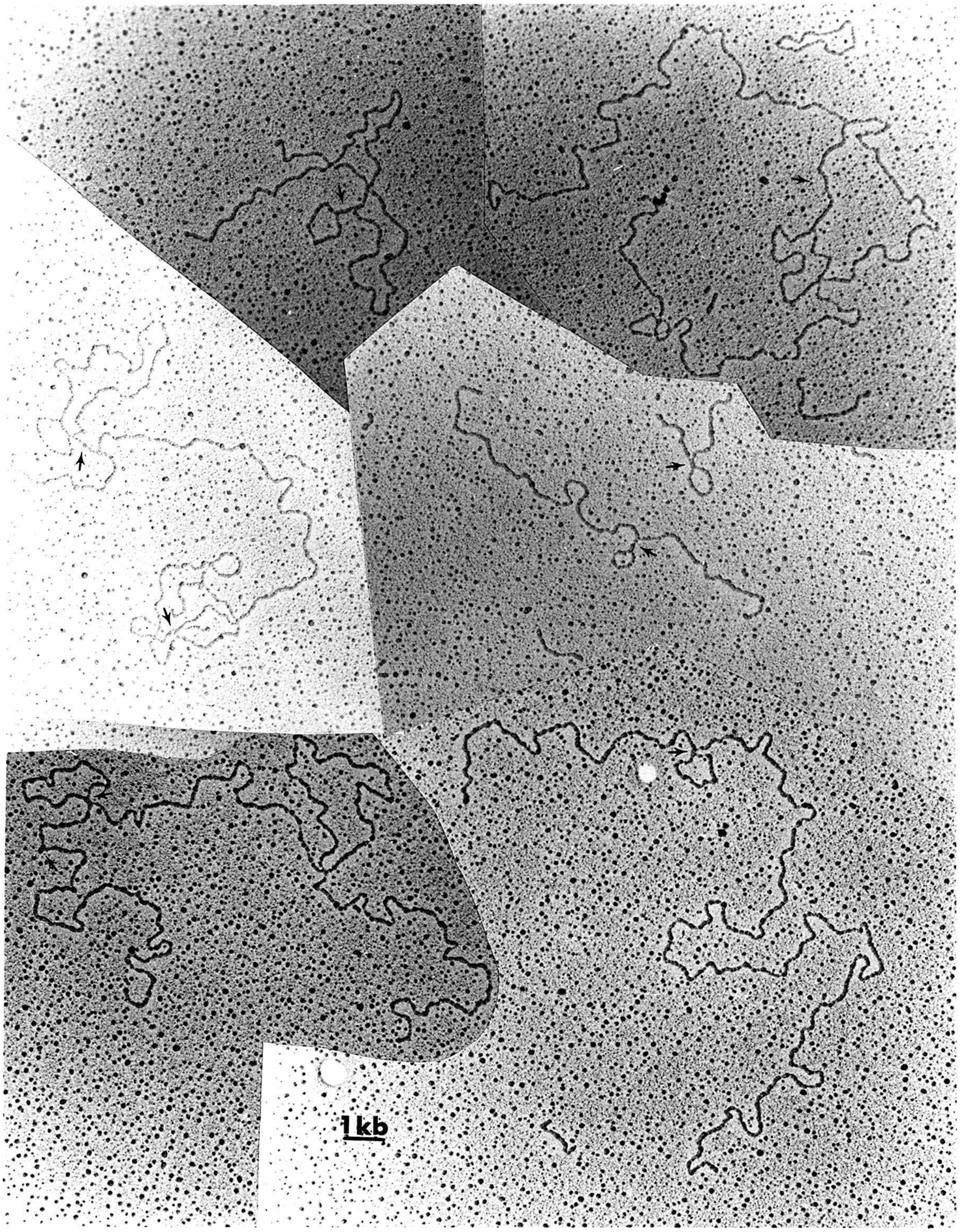


Fig. 2

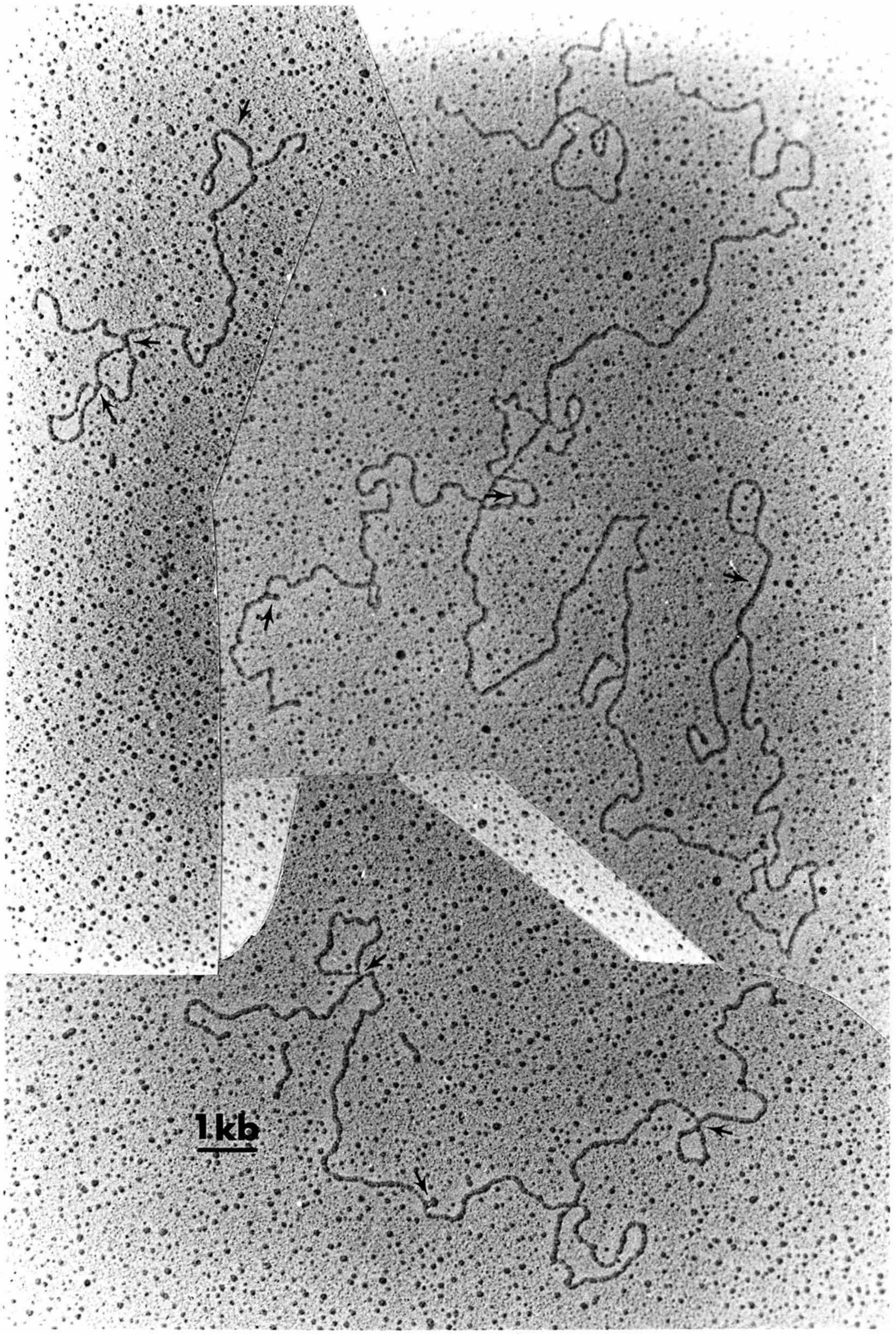


Fig. 3

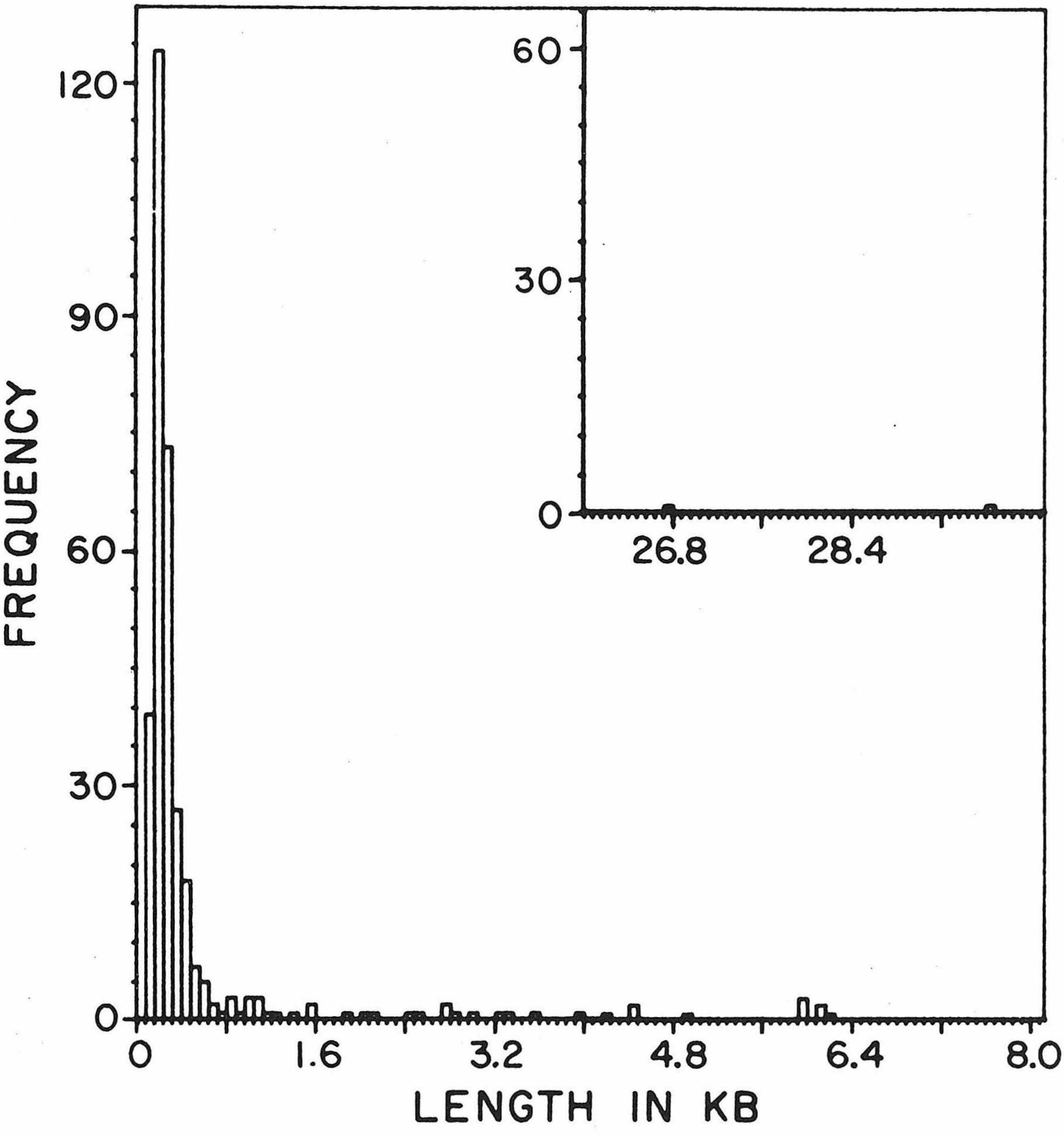

FREQUENCY
0
30
60
90
120
0
1.6
3.2
4.8
6.4
8.0
LENGTH IN KB
0
30
60
26.8
28.4

Fig.4

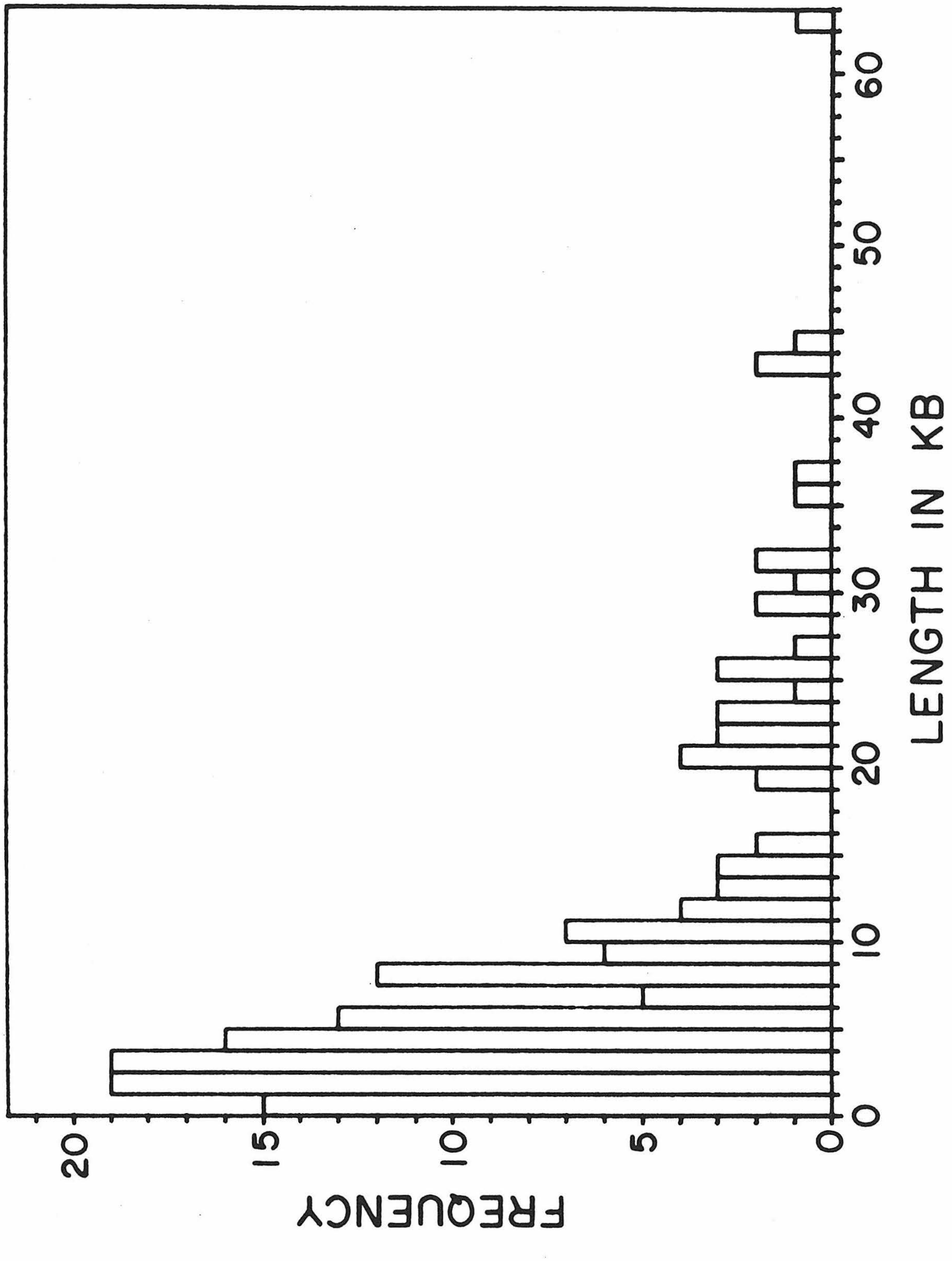


Fig.5

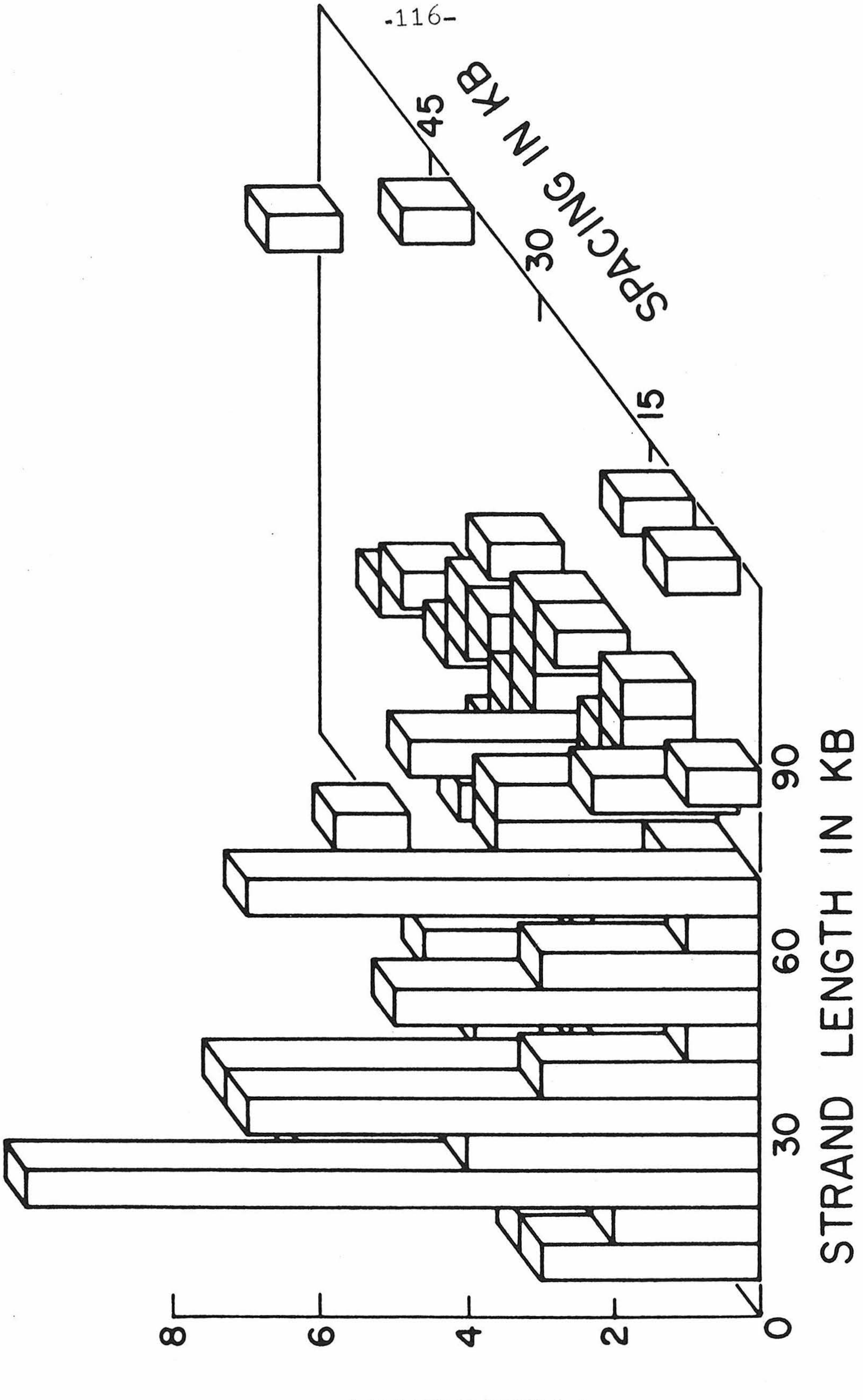

FREQUENCY
0
2
4
6
8
STRAND LENGTH IN KB
30
60
90
SPACING IN KB
15
30
45

Fig.6

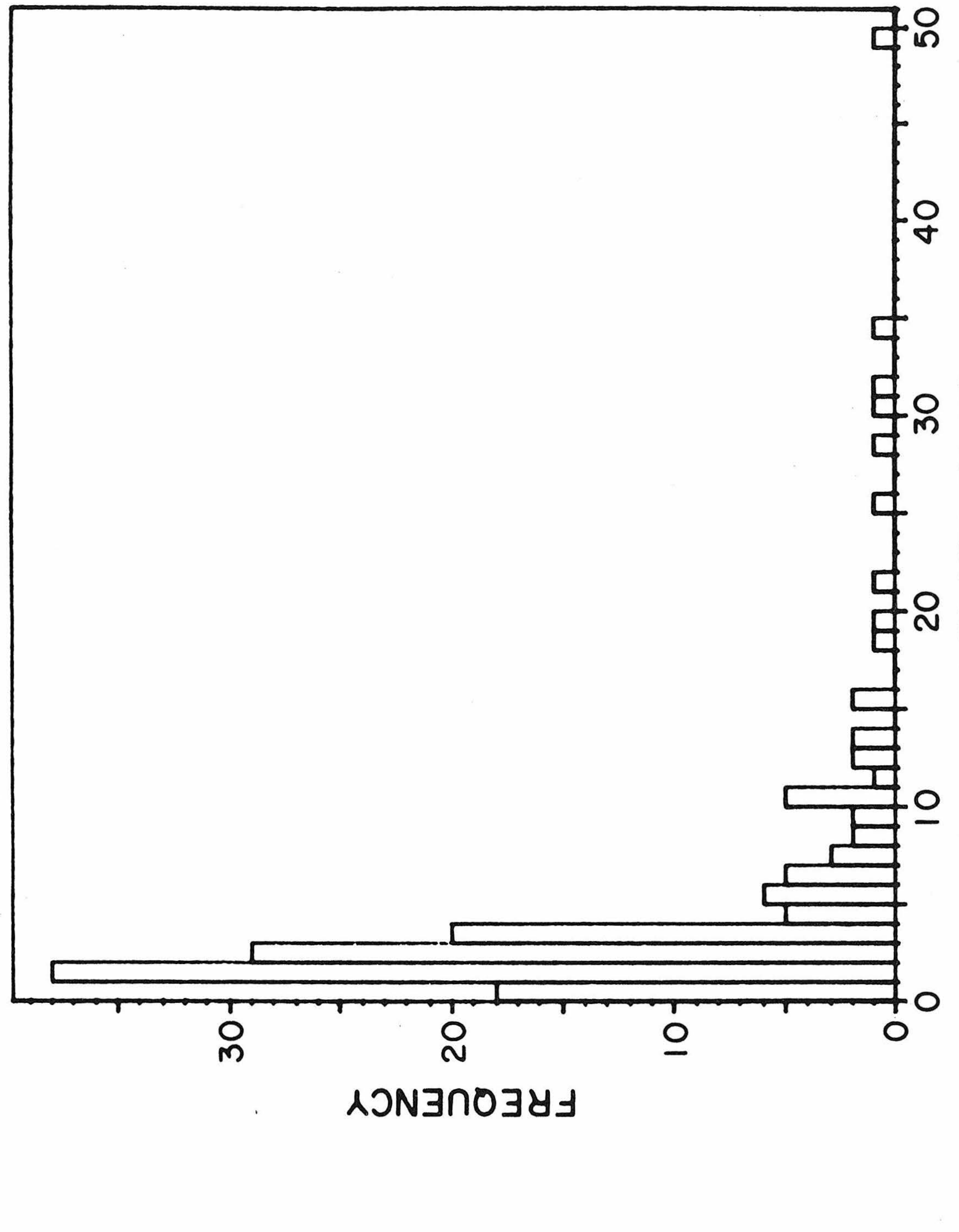


Fig.7

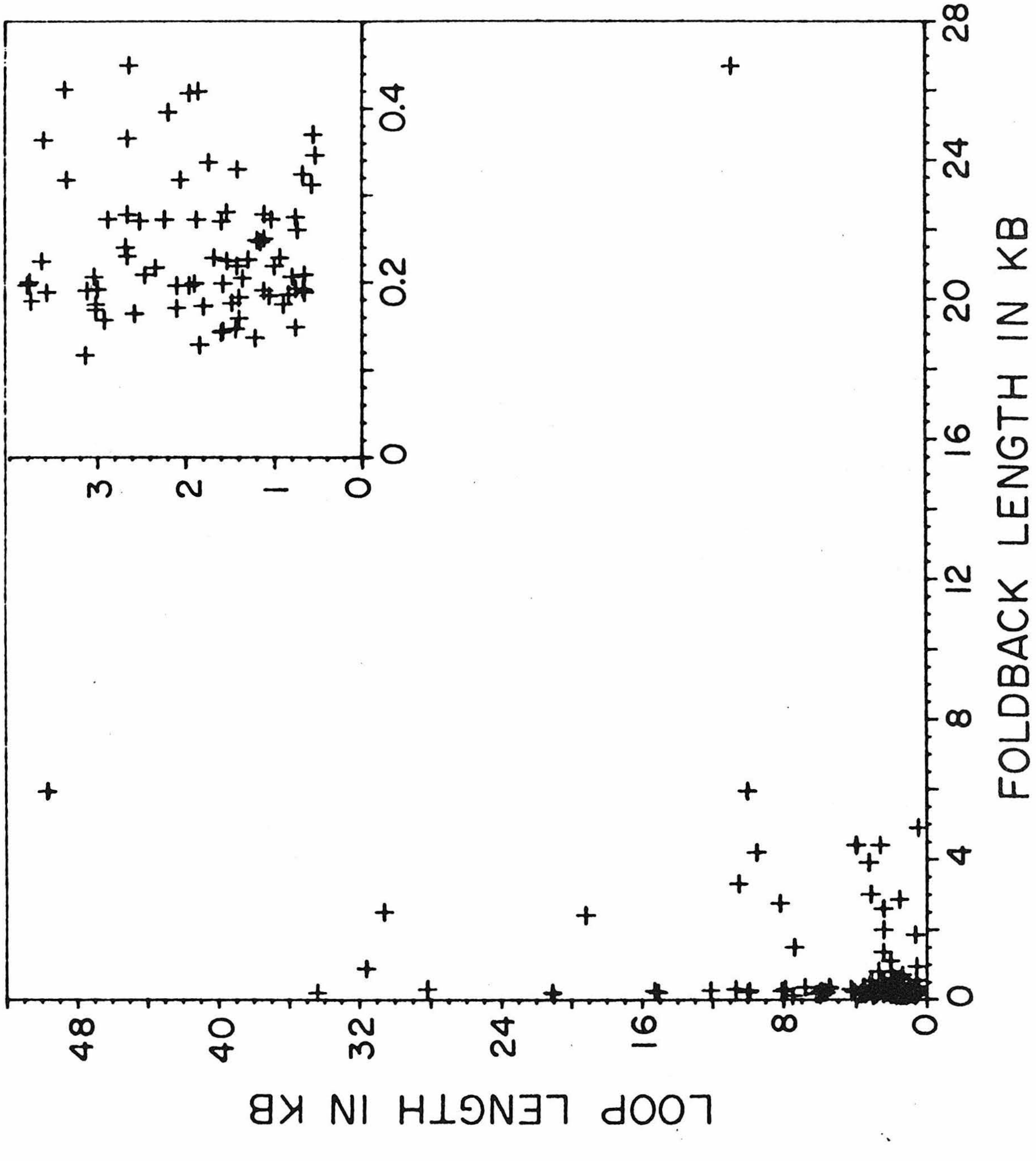

LOOP LENGTH IN KB
0
8
16
24
32
40
48
FOLDBACK LENGTH IN KB
0
4
8
12
16
20
24
28
0
1
2
3
0
0.2
0.4

Fig.8

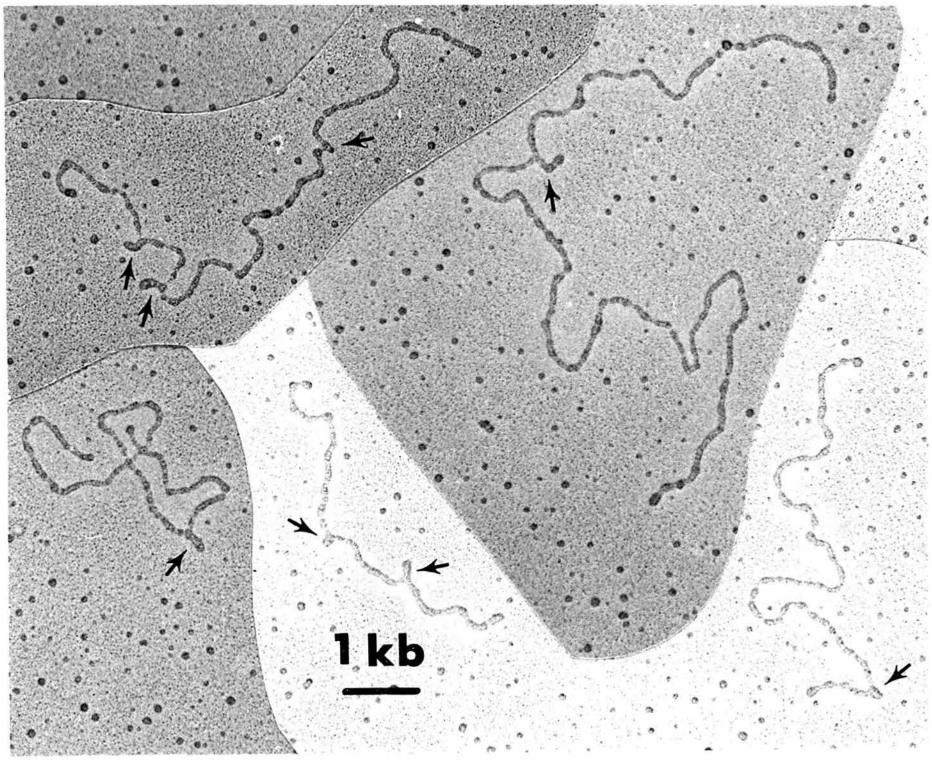


Fig. 9

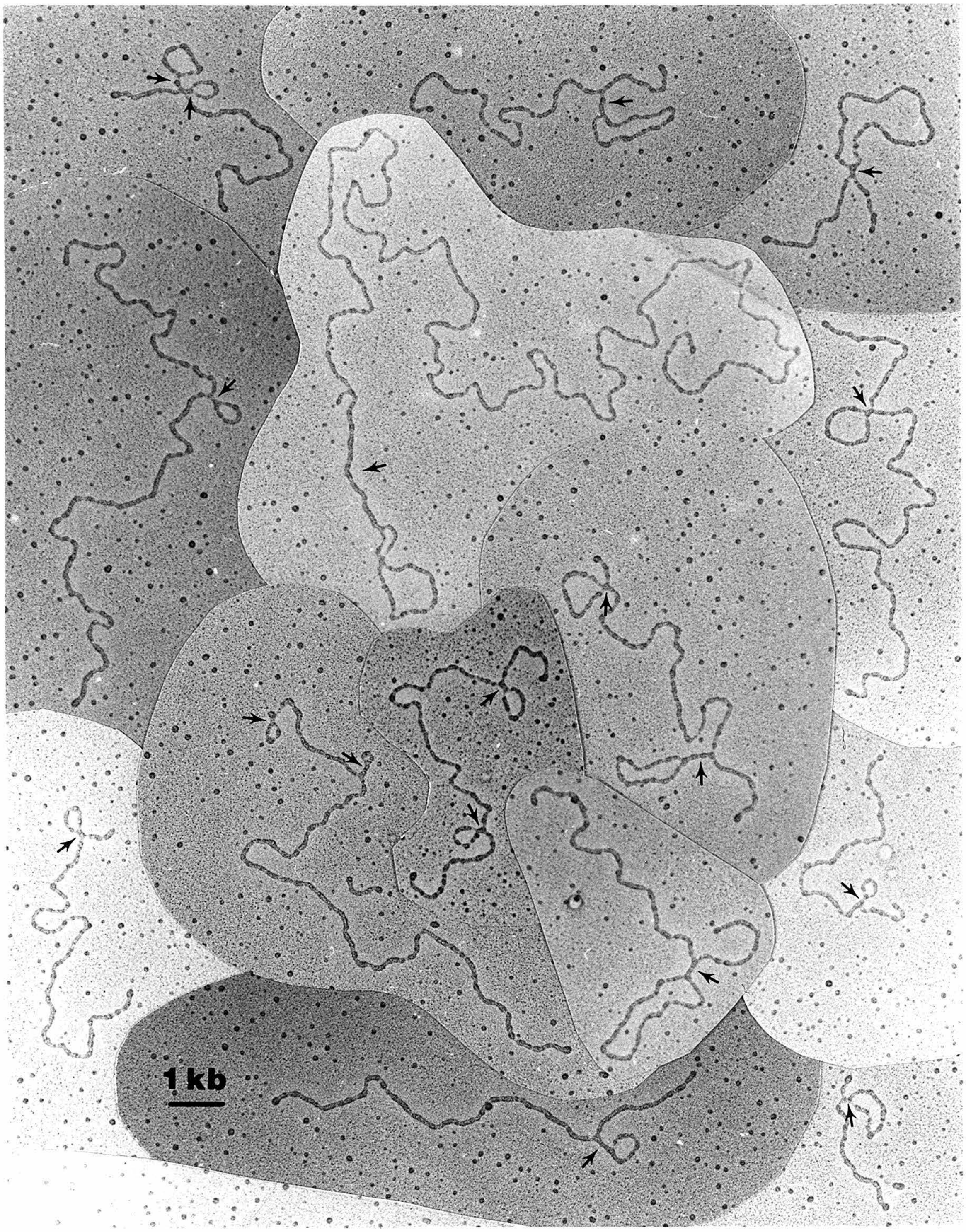


Fig. 10

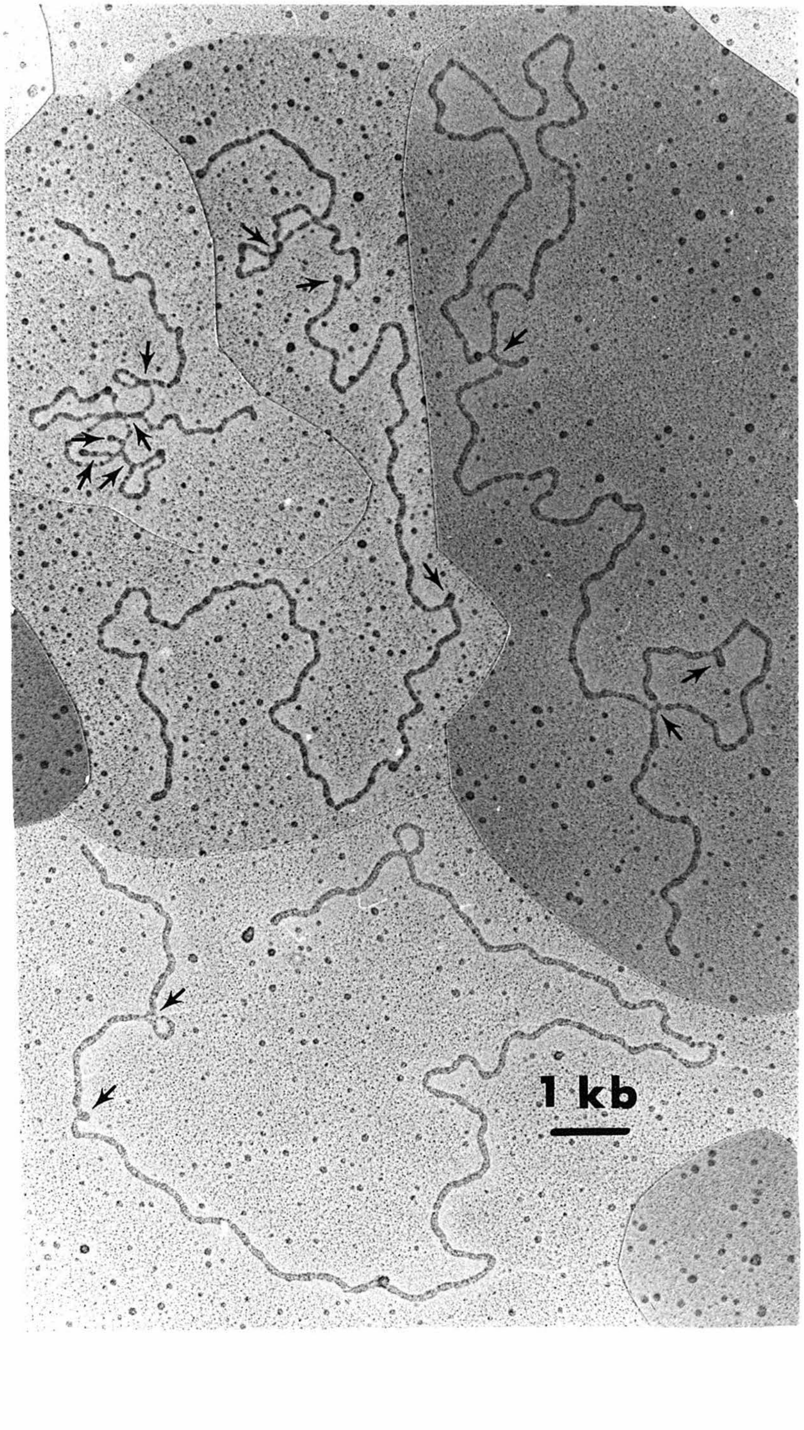


Fig. 11