Chromosoma (Berl.) 76, 277–294 (1980)

CHROMOSOMA

The Sequence Organization of Bovine DNA

John E. Mayfield[1], Julie F. McKenna[2] and Bruce S. Lessa

Department of Zoology, Iowa State University, Ames, Iowa 50011, U.S.A. ([1] addressee for reprint requests) and Department of Biological Sciences, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, U.S.A.; [2] Present address: Office Products Division, I.B.M. Corporation, Lexington, Kentucky 40511, U.S.A.

Abstract. The organization of repetitive DNA sequences has been investigated in bovine DNA. Repetitive sequences of all kinds constitute 25% to 30% of the total. Five density satellites constitute about 20% of the genome, and most of the remainder consists of alternating repeating and nonrepeating sequences. The nonrepeating sequences have a very broad size distribution averaging 4,000 nucleotide pairs in length, with the longest exceeding 10,000 nucleotide pairs. The interspersed repetitive sequences are much more nearly homogeneous in size, averaging 350 nucleotide pairs in length, and are divided into 8 to 14 sequence families.

Introduction

The genomes of all eukaryotes which have been studied include DNA sequences which are repeated. For vertebrates as well as many invertebrates these sequences fall into certain recognizable classes (Davidson et al., 1975). The four commonly occurring classes are: the repeated structural genes, the interspersed middle repetitive sequences, the tandomly repeated highly repetitive sequences, and the inverted repeat or foldback sequences. In addition, some genomes include sizable amounts of repeated sequences which do not obviously fall into one of these classes (Cech and Hearst, 1976; Manning et al., 1975).

The cow genome was one of the first to be studied (Britten and Smith, 1971), and it is known to include a middle repetitive sequence class which is rather highly repetitive, as well as several density satellites (Kurnit et al., 1973). This study refines and unifies these earlier data.

Materials and Methods

a) Isolation of DNA. Crude chromatin was prepared from frozen calf thymus (Bonner et al., 1968). The chromatin was then dissolved in 1 M LiCl, 0.01 M EDTA, 0.01 M Tris pH 8 and 2% sodium

0009-5915/80/0076/0277/$03.60

dodecyl sulfate. DNA was purified from the dissolved chromatin by phenol extraction, ethanol precipitation, 10 μg/ml RNAase and 100 μg/ml pronase digestion for 2 h each followed by a second round of phenol extraction and ethanol precipitation.

b) Shearing of DNA. Batches of DNA at 1 A_{260} unit/ml were variously sheared twice in the Sorvall RM1 french press, or for 90 sec in the Virtis "45" homogenizer, or by ten passages through a number 28 syringe needle at maximum thumb pressure. These methods allowed the creation of sheared DNA samples with average lengths ranging from 400 to 12,000 nucleotides. Each DNA sample was then passed through a short BioRad Chelex 100 column.

c) Hydroxyapatite Chromatography. The usual procedure was to add one-tenth volume of 0.5 M KOH to 10 A_{260} units of sheared DNA, and then, after 5 min, an equal volume of 1 M KH_2PO_4 was added to produce a solution of denatured DNA in 0.1 M potassium phosphate buffer with a pH of about 7. The solution was then incubated at 60° for the desired length of time. The effective C_0t value was taken to be: $C_0t = A_{260}/2 \times h \times 0.687$ (Britten et al., 1974), where the A_{260} was the volume corrected value before denaturation. For C_0t values of 10^{-3} the DNA solutions were diluted to 0.05 M potassium phosphate buffer and the effective C_0t taken to be: $C_0t = A_{260}/2 \times h \times 0.1021$ (Britten et al., 1974).

Following incubation the solutions were diluted with an equal volume of cold water and applied to a jacketed column of 0.6 gm of BioRad DNA grade hydroxyapatite at room temperature. The hydroxyapatite was boiled in 0.1 M potassium phosphate buffer for fifteen minutes prior to use, and was discarded after each use. All DNA, unless badly degraded, binds to hydroxyapatite in 0.05 M potassium phosphate buffer at room temperature. After all the DNA was applied to the column the temperature was raised to 60° and the column washed with 10 ml of 0.05 M potassium phosphate buffer. Under these conditions the single stranded DNA washes out of the column. Then, with the temperature still 60°, the column was washed with 0.2 M potassium phosphate buffer which removes the hybrid material. The single stranded and hybrid containing DNA fractions were then pooled separately, and dialized vs. 0.001 M Tris pH 8. The samples were heated briefly to denature them and the A_{260} and volume of each sample determined carefully.

d) Electron Microscopy. DNA for electron microscopy was denatured in 0.1 M NaOH for 10 min at room temperature and then neutralized with 2 M Tris base + 1.8 M HCl to a final cation concentration of 0.3 M. The DNA was then hybridized at 60° to the desired C_0t value determined by: $C_0t = A_{260}/2 \times h \times 2.3146$ (Britten et al., 1974). The reaction was stopped by placing the sample in an ice bath.

The DNA was visualized by the basic protein film techniques of Davis et al. (1971). The grids were stained with uranyl acetate, rotary shadowed with Au-Pd (60%-40%), and coated with a thin layer of carbon to stabilize the parlodian film. The grids were then viewed and photographed with either the JEM7 or the Phillips 300 electron microscope at a magnification of 10,000. Structures were scored and measured from complete tracings of selected negatives at a magnification of 100,000.

e) Size Standards. Eco RII restriction endonuclease "monomer" fragments of mouse satellite DNA were prepared by the method of Southern (1975). One μg of mouse satellite DNA estimated to be about 90% pure by CsCl buoyant density analysis was digested with 5 units of Eco RII restriction endonuclease purchased from Bethesda Research Laboratories in a total volume of 50 μl. The reaction was stopped by adding 1/5th volume of 0.1 M EDTA pH 7. The material was then mixed with sucrose to 8% and a small amount of bromphenol blue and electrophoresed directly on 2% agarose. The running buffer consisted of 30 mM Tris, 36 mM NaH_2PO_4, 1 mM EDTA, pH 7.8, and 0.5 μg/ml of ethidium bromide. The expected series of bands representing "monomers", "dimers", etc. (Southern, 1975) was observed with u.v. light, and the "monomer" bands were cut from 5 gels. These gel slices were then dissolved in hot (60°) 5 M $NaClO_4$ and the released DNA collected on hydroxyapatite. The DNA was eluted from the hydroxyapatite with 0.4 M sodium phosphate buffer, pH 7, was dialyzed versus 10^{-3} M Tris, 10^{-4} M EDTA, pH 7.5, and concentrated by lyophilization.

Single stranded lengths were determined by comparison to circular fd DNA which was prepared by phenol extraction of purified fd virus. The virus was the kind gift of Dr. T.-C. Lin of the

University of Pittsburgh. After dialysis to remove the phenol, the DNA was observed in the electron microscope to be about 75% circular and 25% linear.

f) Alkaline Sucrose Gradient Sedimentation. Sheared DNA was fractionated by sedimentation in a 15% to 25% linear sucrose gradient in 0.1 M NaOH for 18 h at 32,000 rpm in a Beckman SW 41 rotor.

g) CsCl and Cs_2SO_4 density gradient centrifugation. Density gradient centrifugation was performed essentially as described by Kurnit et al. (1973). Preparative gradients were centrifuged for 65 h at 26,000 rpm in a Beckman "30" rotor. Small aliquots of each preparation were simultaneously centrifuged for 24 h at 44,000 rpm in the model E analytical centrifuge. Cs_2SO_4 was purchased from the Kerr-McGee Corp. and purified by dissolving in 50% acetic acid, filtering, and precipitating with methyl alcohol. The dried salt was essentially free of chloride and was suitable for Cs_2SO_4-Ag^+ separations of satellite DNA.

Results

I. Hydroxyapatite Chromatography

Figure 1 shows the repetitive portion of a standard hydroxyapatite C_0t curve for cow DNA. Little further column binding occurs until single copy sequences begin to form hybrids at C_0t values near 500. This curve agrees closely to that previously reported by Britten and Smith (1971) and is interpreted to mean that for cow DNA sheared to this size and incubated at this criterion, about 30% of the fragments contain at least a portion of a middle repetitive sequence which is repeated approximately 60,000 times, and 20% of the fragments contain sequences which hybridize at least 10 times faster. It is likely that much, but not all, of this most rapidly reannealing material is foldback

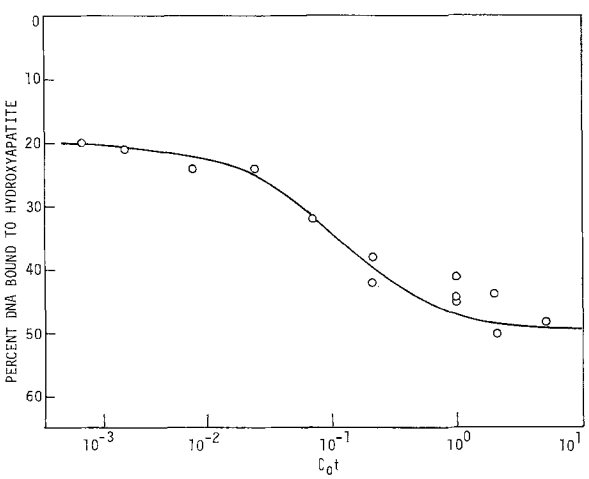


Fig. 1. The binding of bovine DNA to hydroxyapatite as a function of the log of DNA concentration times time (C_0t). The DNA was sheared to an average length of 1,000 nucleotides. The solid line represents ideal second order kinetics

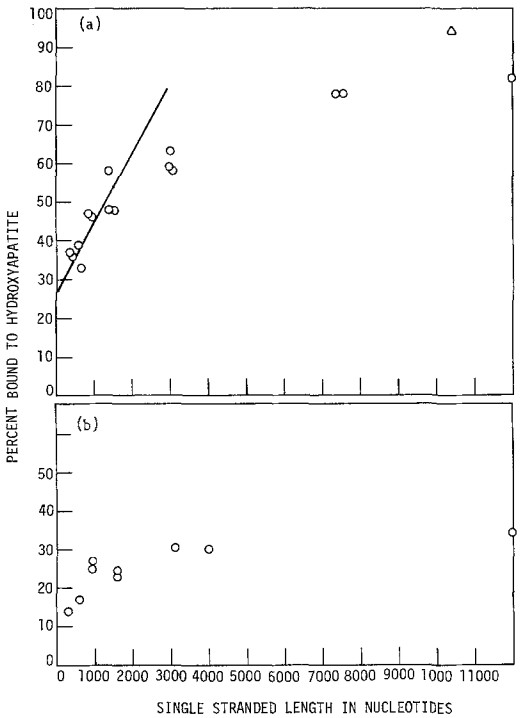


Fig. 2a and b. The binding of DNA fragments to hydroxyapatite as a function of fragment length. **a** $C_0t=1$. The point Δ represents DNA which had been sized on a sucrose gradient prior to reannealing. **b** $C_0t=10^{-3}$

in nature (Cech and Hearst, 1975; Perlman et al., 1976; Deininger and Schmidt, 1976).

Figure 2a shows the effect of DNA fragment length on the percent of the DNA which contains repeated sequences. The general rise in the curve with increasing fragment length implies that most of the genome is organized so that repeated and non-repeated sequences are interspersed (Davidson et al., 1973). The intercept at zero fragment size shows that 25 to 30% of cow DNA is repetitive. It is shown in the Appendix that the initial slope of this curve can be used to calculate the average length of one repeat plus adjacent non-repeat. For cow DNA this appears to be about 4,400 nucleotides.

Unfractionated, sheared DNA generally consists of a fairly broad distribution of sizes. Alkaline sedimentation velocity analysis (Studier, 1965) of the DNA from hybridization experiments with DNA fragments sheared to large size shows the unhybridized DNA to be much smaller than the hybridized material (data not shown). Apparently the non-hybrid DNA is selectively composed of the smaller members of the size distribution. To assess the importance of this to the total fraction of DNA in hybrid form, DNA was sheared by passage through a 28 gauge needle and fractionated on an alkaline sucrose gradient. When the fractionated DNA was incubated to a C_0t of 1 and passed over hydroxyapatite, the percent bound (Δ in Fig. 2a) was substantially higher than for unfrac-

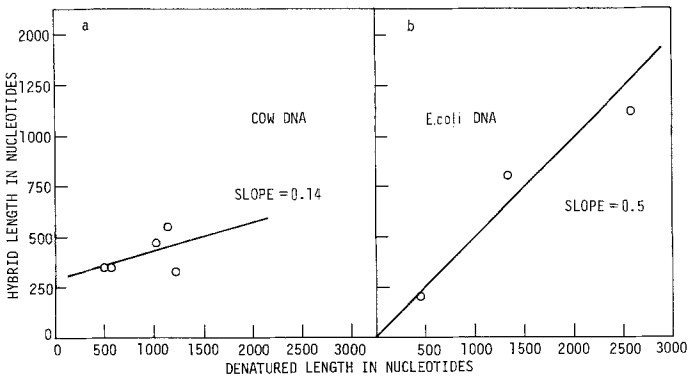


Fig. 3. The average hybrid length as a function of DNA fragment length for a bovine DNA, $C_0t=0.1$ and b *E. coli* DNA, $C_0t=1.0$. The hybrids were visualized in the electron microscope without formamide

tionated DNA. This result indicates that single copy sequences longer than 10,000 nucleotides are rare in the cow genome. In contrast to the curve at $C_0t=1$ (Fig. 2a), the data for $C_0t=10^{-3}$ (Fig. 2b) rises at small fragment sizes but then levels off implying that foldback sequences are interspersed with other DNA yet only 35% of the fragments 10,000 nucleotides in length contain such sequences.

II. Electron Microscopy of Short DNA Fragments

When DNA that has been embedded in a monolayer of cytochrome c in a suitable aqueous buffer is visualized in the electron microscope only double stranded DNA is easily seen. Single stranded DNA is collapsed into small globular structures that for short DNA fragments are generally invisible (Davis et al., 1971). This technique provides a convenient way to characterize the length of DNA hybrids which is relatively independent of the size of the input DNA.

As a control we used *E. coli* DNA. From our present understanding of bacterial DNA's we would expect that initial hybrids of sheared *E. coli* DNA would average about one-half the length of the input DNA. This is because *E. coli* has little repetitive DNA and therefore randomly sheared single stranded fragments will hybridize such that on average one-half of their lengths are complimentary (Britten and Kohne, 1966). DNAs which contain repetitive sequences on the other hand will tend to exhibit hybrid sizes of the length of the repetitive sequences provided the DNA fragments are substantially larger than the size of the repeated sequences and the hybridization does not continue so long as to allow the nonrepetitive sequences to hybridize. In practice, the technique is limited by the formation of complex structures. Nevertheless, it is clear from Figure 3 that *E. coli* and cow DNA produce quite different hybrid lengths at low C_0t values. *E. coli* DNA does indeed yield hybrids which average one-half the length of the input DNA, while cow DNA yields hybrids of more constant length implying a common repeated sequence length of 300 to 400 nucleotides.

Table 1. Number and percentage of the total length represented by molecules scored under each set of hybridization conditions. –

Average length (nucleotide pairs)	C_0t values		Hybrid shapes					Total no. scored
			Lariats and circles	Ys	Hs	Complex	Linears	
600	0.1	No. scored	3	46	4	4	257	314
		% total length	1.4	20	3.4	4.8	70	
	0.1	No. scored	2	34	3	2	264	305
		% total length	2.2	22	2.4	2.3	71	
1,500	0	No. scored	2	9	1	0	222	234
		% total length	1.6	6	1.1	–	91	
	0.01	No. scored	1	27	2	1	193	224
		% total length	0.9	17.9	1.9	1.5	77.8	
	0.1	No. scored	3	38	6	7	430	484
		% total length	1.6	16.9	3.7	8.3	69.4	
	1	No. scored	1	32	10	16	269	328
		% total length	0.5	17.1	8.8	31.7	41.9	
	10	No. scored	4	36	8	17	256	321
		% total length	1.9	17.2	6.3	31.9	42.7	
3,000	0	No. scored	0	2	0	0	167	169
		% total length	–	1.2	–	–	98	
	0.01	No. scored	10	29	8	3	323	373
		% total length	3.9	12.2	6.5	2.5	75	
	0.1	No. scored	11	16	7	19	222	275
		% total length	5	9.1	4.6	21.4	59.8	
	0.5	No. scored	6	8	2	6	53	75
		% total length	9.4	11.9	4.7	27.4	46.6	
	1	No. scored	2	13	1	14	117	147
		% total length	1.8	9.4	1.9	30.5	56.3	

When DNA has been imbedded in a monolayer of cytochrome c in the presence of high concentrations of formamide, both single and double stranded DNAs are readily visualized in the electron microscope. Under ideal conditions single stranded and double stranded segments are readily distinguished by diameter and straightness. The identification of short hybrids, however, becomes quite treacherous. For this reason, we have chosen to score the observed hybrids only on the basis of their shape which is easily determined without ambiguity.

Table 1 lists the various shapes scored along with their relative proportions in the genome for a variety of C_0t values and fragment sizes. These data can be used to form a C_0t curve which is quite similar to the hydroxyapatite binding curve (see Fig. 4) indicating that the observed hybrids are largely the same

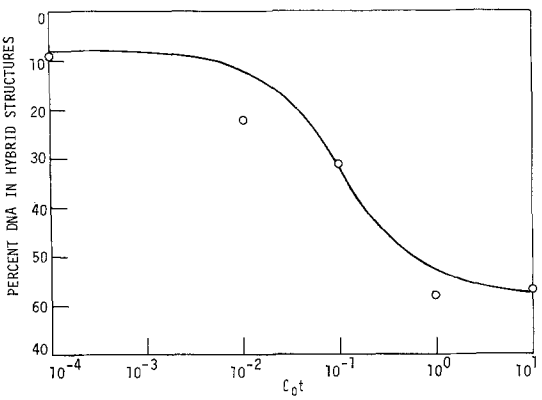


Fig. 4. C_0t curve plotted from the data in Table 1 for a DNA fragment length of 1,500 nucleotides. The solid line represents ideal second order kinetics with the same rate constant as the curve in Fig. 1

material which binds to hydroxyapatite. The electron microscope derived data has considerably more scatter but it also has the great advantage of containing information on the kinds of hybrids which are forming. This information can be used in principle to deduce the details of sequence organization.

Several features of the data are immediately apparent. First circles and lariats are present in nearly all preparations regardless of C_0t value or fragment length. The size of the loops is quite variable, but does not seem to increase with increasing fragment size, rather with short DNA there are few lariats and with larger DNA few circles. This suggests that many of these structures represent paired or tandem repeats on opposite strands.

"Y" structures are also common even at the lowest C_0t values, indicating the presence of short palindromic sequences. For these structures it is frequently possible to determine that one arm of the Y is double stranded. This double stranded arm is almost always short, averaging 620 Å (approximately 250 nucleotide pairs) in length.

"H" structures, indicative of interspersed repeated sequences (Bonner et al., 1973; Chamberlain et al., 1975) are never common, but generally increase in number with increasing C_0t values. Figure 5 shows the length distribution of the cross bar of the "H" structures. This represents the size distribution of the interspersed repeated sequences. The average length is 350 nucleotides and the distribution is substantially broader than observed for Eco RII monomer fragments of mouse satellite DNA indicating that the interspersed repetitive sequences are heterogeneous in length.

All other hybrid shapes were classified as "complex", and for DNA fragments of the larger sizes (see Table 1) this class clearly accounts for the bulk of the hybridization between a C_0t of 0.1 and 1.0. It includes a bewildering array of configurations which has thus far defied systematic analysis, but which clearly includes many examples of multiple "Hs", "Ys", and lariats (for examples see Fig. 7a).

It is of considerable interest to determine whether or not repeated DNA

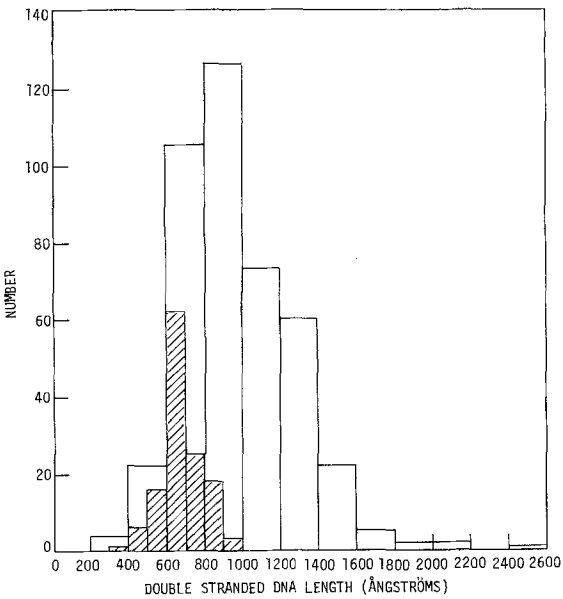


Fig. 5. The length distribution of hybrid lengths found in "H" structures. The hatched distribution is for double stranded Eco RII digestion monomers from mouse satellite DNA. The length of these fragments is known to be 245 nucleotide pairs (Southern, 1975; Bellard et al., 1976). All DNA structures were visualized in the presence of 40% formamide

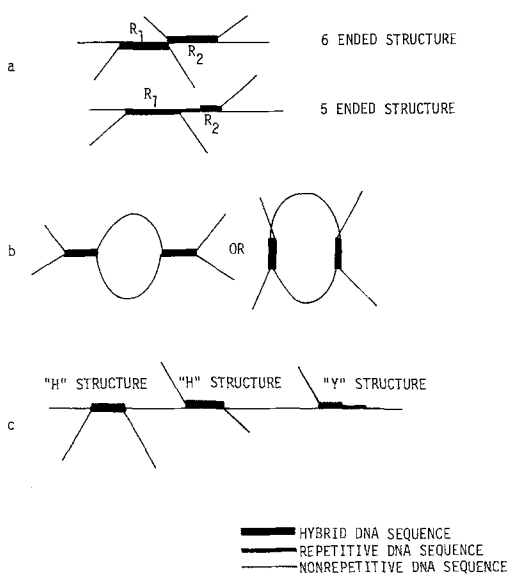


Fig. 6. a The five and six ended structures which are expected to be produced when repeated sequences occur adjacent to one another. **b** The characteristic structures which are expected if neighboring repeated sequences are often the same, and the DNA fragments are long enough to include two repeated sequences. **c** The hybrid structures which are expected when long DNA is hybridized with a much higher concentration of short DNA

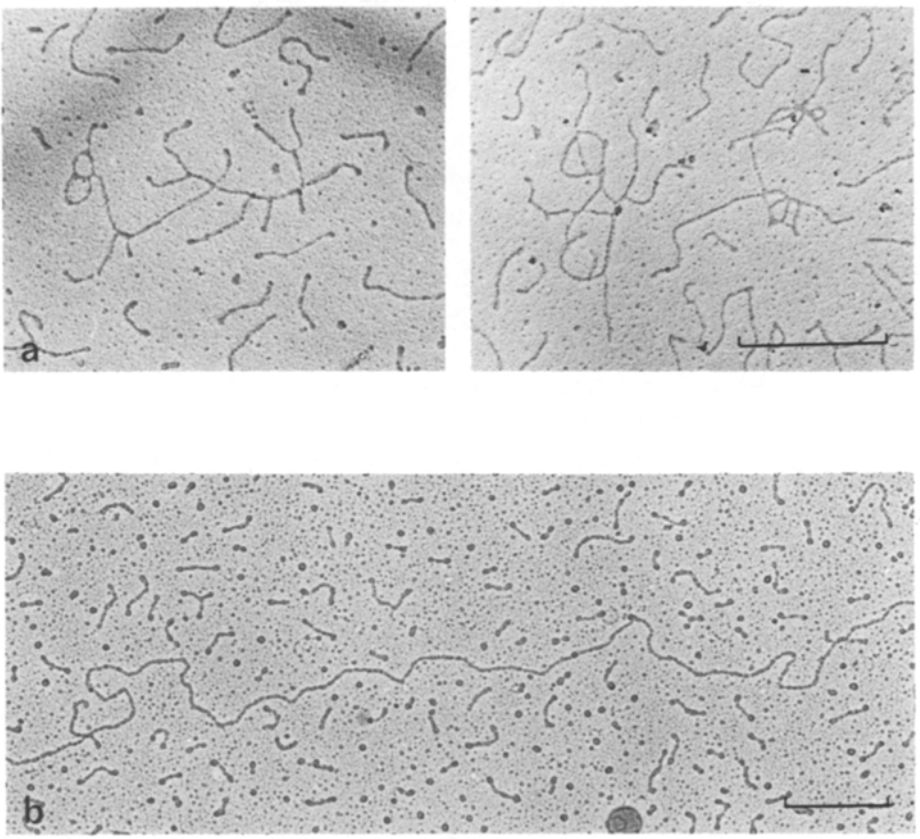


Fig. 7. **a** Electron micrographs of typical complex structures seen at high C_0t values, Bar = 0.5 micron. **b** Electron micrograph of two short "H" structures on a long single stranded DNA fragment. Bar = 0.5 μm

sequences occur physically adjacent to one another in pairs. Such an organization would tend to form hybrid structures with more than four ends. In particular, five and six ended structures would be indicative of this type of organization (see Fig. 6a). Indeed, Chamberlain et al. (1975) found some such structures in *Xenopus*, and Bonner et al. (1973) report them to be common in rat DNA. These forms are included in the "complex" category in Table 1. Most of the complex forms, however, appear to be more complicated than this, and five and six ended structures without loops represent only about 1% of the total DNA structures seen. Since there are other equally plausible explanations for five and six ended hybrid structures, we must conclude that this study provides no evidence for paired, repetitive sequences, although we cannot rule out the possibility that they occur rarely in cow DNA.

If neighboring interspersed repetitive sequences are often the same or the same two often occur near each other, then the structures shown in Fig. 6b should be readily identifiable when the DNA fragments are long enough to include two interspersed repeated sequences. Such structures are not seen, sug-

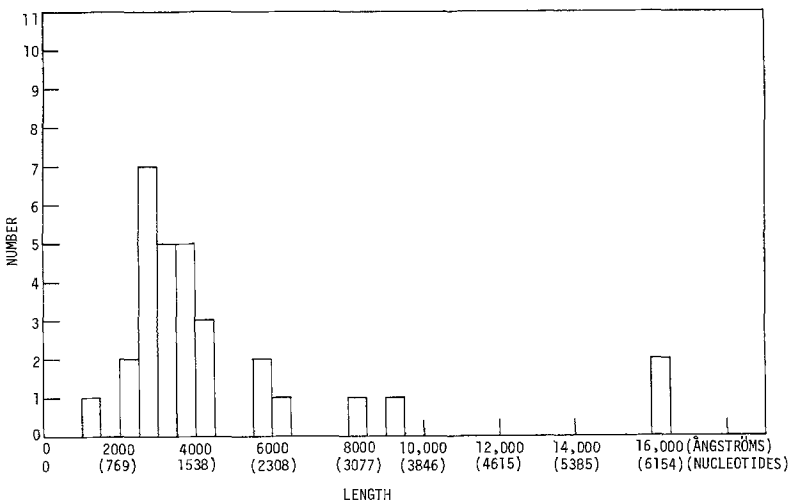


Fig. 8. The distribution of the length of the sequences separating "H" structures observed on long-short hybrid structures

gesting that neighboring repeated sequences are usually different and do not occur in a particular order.

Finally, it is clear that the complex structures grow larger with increasing C_0t values, and for moderately long DNA, complicated networks are often seen at high C_0t values (see Fig. 7a). These networks are most likely due to the presence of clusters of repeated sequences.

III. Electron Microscopy of Mixtures of Long and Short DNA

Information concerning the distance between interspersed repetitive sequences can be obtained by observing hybrids formed between long and short DNA fragments (Bonner et al., 1973). This is illustrated in Figs. 6c and 7b. Analysis is complicated by the fact that some hybrids will be of the "Y" type, some of the "H" type, and some repetitive sequences will not have hybridized at all. "Y" type hybrids are often difficult to unambiguously distinguish from foldback sequences. So we have measured only distances between "H" type hybrids which do not have branches (Ys) between them. These data are presented in Fig. 8. It is seen that the sequences separating the repetitive sequences have a very broad length distribution with a modal value of 1,100 nucleotides and an average of about 2,200 nucleotides. There are two major, uncertainties which affect this distribution. The first is that the long DNA is of finite length, and therefore very long separating sequences will not be scored; and the second is that unhybridized repeated sequences will cause two separating sequences to be counted as one. In spite of these shortcomings, this is the most direct demonstration of the interspersed nature of the repeated sequences. Figure 9 shows the length distribution of the "H" type hybrids formed between long

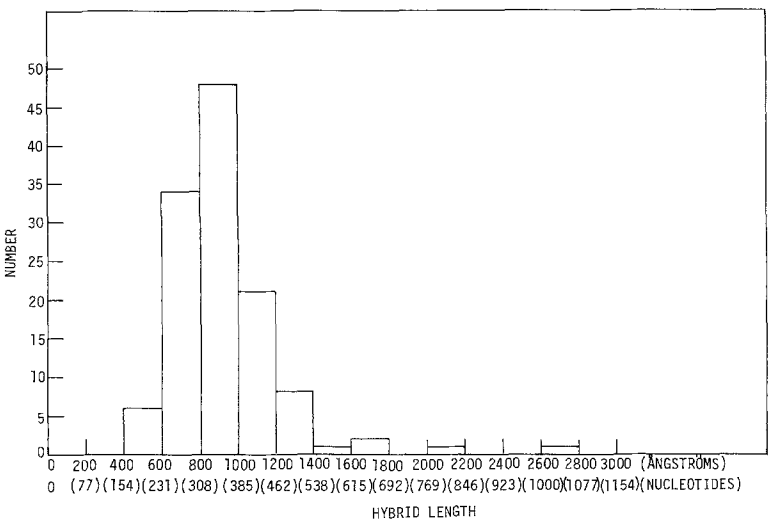


Fig. 9. The length distribution of double stranded DNA in "H" type hybrids observed to be present on long-short hybrid structures

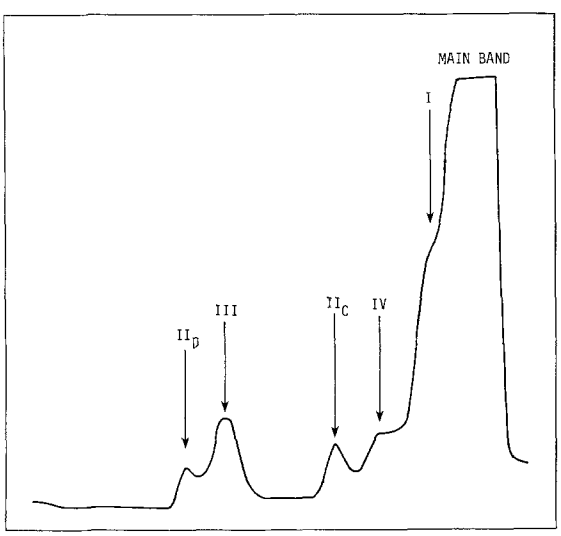


Fig. 10. Cs_2SO_4 density gradient of bovine DNA in the presence of $AgNO_3$ in a molar ratio of Ag^+ to DNA phosphorus of 0.40. The satellite identity of each peak is indicated

and short DNA. This distribution is indistinguishable from that seen in Figure 5 showing that we are in fact observing the same class of hybrids.

IV. Satellite DNA Sequences

Several workers have studied the satellite DNA's found in bovine DNA (Kurnit et al., 1973; Botchan, 1974; Mowbray et al., 1975; Macaya et al., 1978). These

Table 2. Bovine satellite DNAs. – The density and approximate proportion of the genome for each satellite DNA species which has been identified to be present in bovine DNA by the procedure outlined in the text

Satellite	Density in CsCl	% of genome
I	1.715	7
II_D	1.721	1
II_C	1.723	1
III	1.706	3.5
IV	1.709	7
Total		19.5

studies have established that cow DNA contains four to eight satellite sequences all of which are more dense than main band DNA in CsCl. The satellite most prominent in CsCl has been termed satellite I. By restriction endonuclease digestion it has been shown to have a repeat of 1,400 nucleotide pairs and constitute 6.7% of the total DNA. Figure 10 is the densitometer trace of cow DNA banded to equilibrium in a Cs_2SO_4 density gradient in the presence of $AgNO_3$ in a molar ratio of Ag^+ to DNA phosphorus of 0.40. This figure clearly shows five distinct satellite peaks. This same solution was centrifuged preparatively and the satellite containing regions divided into four fractions which were each recentrifuged to equilibrium in CsCl. Densitometer traces of the second centrifugations were then used to identify and quantitate the amount of each satellite. These data confirm that there are at least five satellite sequences in the cow genome. Table 2 lists these data, and Fig. 10 shows the location of each satellite in the Ag^+-Cs_2SO_4 density gradient.

Discussion

Cow is usually considered to be a typical example of an organism which possesses three major classes of DNA sequences. These are highly repetitive, middle repetitive, and unique sequences (Britten and Smith, 1971). A more detailed examination, however, shows that the situation is more complex. The bulk of the renaturation of repetitive sequences occurs between C_0t values of 10^{-2} and 10^0 with approximately second order kinetics. The evidence presented in this paper shows that there are several different types and many different sequence families which reanneal with the same approximate rate. Virtually all known density satellites are highly repetitive, and direct evidence that several of the cow satellites are tandomly repeated has been reported (Mowbray et al., 1975). Furthermore, it has been demonstrated that satellite I has a renaturation complexity of about 250 nucleotide pairs (Botchan, 1974). This would imply that 6 to 7% of the DNA should reanneal with a $C_0t_{1/2}$ of about 8×10^{-3}. The other satellites probably hybridize more slowly than this because they constitute a smaller proportion of the genome. The data presented in Figures 2,

3, 5 and 7 all show that a major component of the hybridization which occurs between C_0t 10^{-3} and 10^1 is due to interspersed repetitive sequences. The fraction of the genome which consists of these interspersed repetitive sequences can be calculated from the data presented in this paper in at least two ways. In the Appendix (equations 2 and 3) we show that the initial slope of the curve in Fig. 2a is equal to the fraction of the genome which is composed of alternating repeating and nonrepeating sequences (type A organization) divided by the average length of a repeat plus an adjacent nonrepeat. The best fit for the data presented in Figure 2 for DNA fragments shorter than 2,000 nucleotides yields a slope of 1/5,400 nucleotides^{-1}. [If the data are corrected for $C_0t=10^{-3}$ binding as suggested by Davidson et al. (1973) the value becomes 1/6,250 nucleotides^{-1}]. The data shown in Fig. 5 yield an average weight average length of 350 nucleotide pairs for the interspersed repetitive sequences. Therefore, estimating type A DNA to constitute 80% of the total genome, the average length of the sequences between interspersed repeated sequences is about 4,000 (corrected, 4,650) nucleotides, and the total fraction of the genome composed of these sequences is 6.5% (corrected, 5.6%). For comparison, similar data for *Xenopus* DNA yields a slope of 1/2,380 nucleotides^{-1} and for sea urchin a slope of 1/3,675 nucleotides^{-1} (Davidson et al., 1973; Graham et al., 1974). These values imply that the sequences separating repeated sequences average about 2,000 nucleotides and 3,000 nucleotides, respectively, for these two species.

A second method of calculation is based on the relative proportion of "H" shapes seen when sheared hybridized DNA is visualized in the electron microscope. The data in Table 1 shows clearly that the frequency of "H" shapes is sensitive to both fragment length and C_0t value. At a C_0t value of 0.1 there is a steady increase in the proportion of "Hs" with increasing fragment length. For DNA with an average length of 1,500 nucleotides the frequency of such structures generally increases with increasing C_0t values, whereas for fragments 3,000 nucleotides long there is a general decrease. This is what one would expect if many fragments 3,000 nucleotides long contained more than one repeated sequence. At high C_0t values these long fragments form complex structures. The data in Table 1 for 1,500 nucleotide long fragments at C_0t values of 1.0 and 10 best meet the requirements of equation 5 in the Appendix. Taking C to be 7.5% (see data for $C_0t=0$), B to be 19% (assuming B is satellite DNA), n to be 35 nucleotides, y to be 350 nucleotides, and further assuming that fraction C has the same average composition of repeated and non-repeated sequences as the rest of the genome, then the fraction of the total genome which is made up of interspersed repeated sequences calculated from equation 5 is 9.9% and 10.1%, respectively, for the $C_0t=1$ and $C_0t=10$ data. These values are in reasonable agreement with the value calculated from the hydroxyapatite binding data. The two estimates indicate that 6% to 10% of the cow genome is composed of interspersed repetitive sequences averaging 350 nucleotide pairs in length. These values are also consistent with the estimate of 25% to 30% of the genome being repetitive DNA of all kinds (DNA sequences which reanneal by C_0t value of 1) taken from the zero length intercept in Fig. 2. The implications of this calculation are quite interesting. Since on average

each of these interspersed sequences is repeated 60,000 times, there must be only 8 to 14 different families of interspersed repeated DNA sequences in the bovine genome. Thus, the number of families of interspersed repeated DNA sequences is only about twice the number of families of clustered (satellite) repeated sequences.

In summary, the bovine genome contains at least 5 density satellites comprising approximately 20% of the total DNA, and a class of interspersed repeated sequences which constitutes 6% to 10% of the total. These interspersed repeated sequences average 350 nucleotides in length, occur singly, dot not seem to be arranged in a simple order, and are separated by nonrepeated sequences of variable length which average about 4,000 nucleotides in length. The number of families of repeating sequences in the interspersed class is probably between 8 and 14. All major repeated sequences apparently reanneal between C_0t values of 10^{-3} and 10^{1} and the resultant renaturation curve closely resembles ideal second order kinetics.

The very small number of families of interspersed repetitive sequences in cow DNA contrasts with estimates of hundreds or thousands of sequence families for several other organisms (Davidson et al., 1975). In spite of this, the spacing and length of these sequences is very similar in cow, *Xenopus,* and sea urchin. This result would argue that size and spacing of these sequences is important for their function, but that repetition frequency is not. Furthermore, it seems very difficult to reconcile the very small number of middle repetitive (interspersed) sequence families in the cow genome with any of the published models which involve the recognition of particular repeated sequences as playing a major role in eucaryotic gene regulation (Davidson and Britten, 1979).

Appendix

Assume that if the DNA of a eukaryotic organism is sheared into fragments, each fragment will have one of the following types of organizations:

Type A, derived from DNA which consists of alternating repeating and nonrepeating sequences.

Type B, consisting of blocks of repeating sequences.

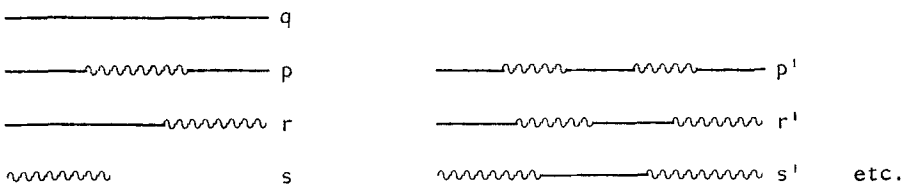
Type C, consisting of a random selection of the genome but including at least one pair of foldback sequences.

Let A be the fraction of the genome which is of the A type, B be the fraction of the genome of the B type, and C be the fraction of the genome of the C type. Then by definition

$$A+B+C=1 \tag{1}$$

Let x be the average length of the nonrepeating sequences in the type A DNA and x_i the length of the i^{th} nonrepeating sequence. Let y be the average length of the repeating sequences in the type A DNA, and y_i the length of the i^{th} repeating sequence (for simplicity y_i shall be adjacent to x_i). Let f equal the fraction of A which is composed of repetitive sequences. So that $f=y/(x+y)$.

When DNA is sheared to average fragment size z the following type A fragments will be produced with the probabilities: q, p, r, s, p′, etc.



where ——— represents a nonrepeated sequence and ∿∿∿ a repeated sequence.

Several basic equations can be derived from this model which are useful for interpreting hydroxyapatite and electron microscope shape data from reannealed DNA.

a) Hydroxyapatite Binding

When all $z_i \leqq y_i$, then

$$p = 0$$

$$s = \frac{(y-z)f}{y}$$

$$r = \frac{2f(z)}{y}$$

$$q = \frac{(x-z)(1-f)}{x}$$

and p′, r′, s′, etc. equal zero.

In this case $p+r+s=1$, and equation 1 becomes $A(q+r+s)+B+C=1$.

Let G be the fraction of DNA bound to hydroxyapatite after the repetitive DNA has completely hybridized. Then,

$$G = Ar + As + B + C = A\left(\frac{z+y}{y}\right)f + B + C$$

or

$$G = \frac{Az}{x+y} + Af + B + C. \tag{2}$$

When all $y_i < z_i \leqq x_i$ (assuming $x_i > y_i$), then

$$s = 0$$

$$r = 2f = \frac{2y}{x+y}$$

$$p = \frac{z-y}{x+y} = \frac{(z-y)}{y}f$$

$$q = \frac{x-z}{x+y} = \frac{(x-z)}{y}f$$

and p′, r′, s′, etc. equal zero.

Therefore, $G = Ap + Ar + B + C = A\left(\frac{z+y}{y}\right)f + B + C$, which is the same as equation 2.

Thus, so long as z_i is less than x_i for all i, a plot of G vs. z will yield a straight line with a slope of $A/(x+y)$, and an intercept of $Af + B + C$. When the fragment length is increased such that z_i begins to exceed x_i on frequent occasions, then p′, r′, s′, etc. begin to become significant and the slope will fall off as a function of the distributions of x_i, y_i, and z_i.

b) Correction for Short Hybrids

There will be some r type fragments which contain a portion of a repetitive sequence which is shorter than some value n, the minimum stable hybrid length. The fraction of r types which will not hybridize because the repetitive sequence fragment is too short is $\frac{2n}{y}f$. Therefore, the fraction of r types which will hybridize is $r - \frac{2nf}{y}$. By substitution, equation 2 becomes

$$G = Az/(x+y) + Af\left(1 - \frac{2n}{y}\right) + B + C. \tag{3}$$

There will also be some r type fragments which have a repetitive sequence fragment which is greater than n but less than m, where m is the minimum hybrid length required to bind to hydroxyapatite. At high C_0t values these short hybrids will not affect the analysis because in general they will hybridize further to form higher order structures which will bind to hydroxyapatite.

c) Hybrid Shapes

When $y_i < z_i \leqq x_i$, equation 1 can be written as $Ap + Ar = 1 - Aq - B - C$. Since $2nf/y$ of the r types will not hybridize we shall define $r_e = r - 2nf/y$ and $q_e = q + 2nf/y$.

Let u be the fraction of p and r type fragments which have hybridized at the C_0t value being studied. Then, $up + ur_e$ is the amount of type A DNA in hybrids, and equation 1 becomes:

$$upA + ur_eA = 1 - q_eA - B - C - (1-u)(p+r_e)A = 1 - K. \tag{4}$$

Squaring both sides equation 4 becomes:

$$\frac{[(up)^2 + 2u^2pr_e + (ur_e)^2]}{1-K}A^2 = 1 - K.$$

$\frac{(up)^2}{1-K}A^2$ is the probability of two p type fragments hybridizing at the chosen C_0t value (note that all such hybrids will be "H" shaped).

Let H be the amount (total length fraction) of H-shaped hybrids found in the sample. Then:

$$H=\frac{(up)^2}{1-K}A^2 \quad \text{or} \quad uAp=\sqrt{H(1-K)}.$$

Let L be the amount (total length fraction) of linear fragments found in the sample. Then, $L=q_eA+(1-u)(p+r_e)A+$(linear hybrid structures). At high C_0t values most hybrid linears will have hybridized further to form complex shapes (Britten et al., 1976), and $1-u\simeq 0$. Under these conditions $L\simeq q_eA$ and $K\simeq L+B+C$. Combining equations, equation 4 becomes:

$$uAr_e=1-L-B-C-\sqrt{H(1-L-B-C)}$$

$$\text{or} \quad uAr=\frac{1-L-B-C-\sqrt{H(1-L-B-C)}}{1-n/y}. \tag{5}$$

Equation 5 describes the special case of u close to 1 and fragment lengths less than x but greater than y. These special conditions are approximately met for cow DNA by Cot values greater than 1, and fragment lengths greater than 400 and less than several thousand nucleotide pairs. The relationship $r=2f$ then provides the fraction of type A DNA which is repetitive.

Acknowledgement. This work was supported in part by U.S. Public Health Service grant GM 21271.

References

Bellard, M., Oudet, P., Germond, J.-E., Chambon, P.: Subunit structure of Simian-Virus-40 minichromosome. Europ. J. Biochem. **70**, 543–553 (1976)

Bonner, J., Chalkley, G.R., Dahmus, M., Fambrough, D., Fujimura, F., Huang, R.C., Huberman, J., Marushige, K., Ohlenbush, H., Olivera, B., Widholm, J.: Isolation and characterization of chromosomal nucleoproteins. Methods in Enzymology **12B**, 3–65 (1968)

Bonner, J., Gerrard, W.T., Gottesfeld, J., Holmes, D.S., Sevall, J.S., Wilkes, M.: Functional organization of the mammalian genome. Cold Spr. Harb. Symp. quant. Biol. **38**, 303–310 (1973)

Botchan, M.R.: Bovine satellite I DNA consists of repetitive units 1,400 basepairs in length. Nature (Lond.) **251**, 288–292 (1974)

Britten, R.J., Davidson, E.H.: Studies on nucleic acid reassociation kinetics: Empirical equations describing DNA reassociation. Proc. nat. Acad. Sci. (Wash.) **73**, 415–419 (1976)

Britten, R.J., Kohne, D.E.: Nucleotide sequence repetition in DNA. Carnegie Inst. Wash. Year Book **65**, 85–87 (1966)

Britten, R.J., Smith, J.: A bovine genome. Carnegie Inst. Wash. Year Book **68**, 378–386 (1971)

Britten, R.J., Graham, D.E., Neufeld, B.R.: Analysis of repeating DNA sequences by reassociation. Methods of Enzymology **29E**, 363–418 (1974)

Cech, T.R., Hearst, J.E.: An electron microscope study of mouse foldback DNA. Cell **5**, 429–446 (1975)

Cech, T.R., Hearst, J.E.: Organization of highly repeated sequences in mouse main-band DNA. J. molec. Biol. **100**, 227–256 (1976)

Chamberlin, M.E., Britten, R.L., Davidson, E.H.: Sequence organization in Xenopus DNA studied by the electron microscope. J. molec. Biol. **96**, 317–333 (1975)

Davidson, E.H., Britten, R.J.: Regulation of gene expression: Possible role of repetitive sequences. Science **204**, 1052–1059 (1979)

Davidson, E.H., Hough, B.R., Amensen, C.S., Britten, R.J.: General interspersion of repetitive with non-repetitive sequence elements in the DNA of Xenopus. J. molec. Biol. **77**, 1–23 (1973)
Davidson, E.H., Galau, G.A., Angerer, R.C., Britten, R.J.: Comparative aspects of DNA organization in metazoa. Chromosoma (Berl.) **51**, 253–259 (1975)
Davis, R.W., Simon, M., Davidson, N.: Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. Methods of Enzymology **21**, 413–428 (1971)
Deininger, P.L., Schmid, C.W.: An electron microscope study of the DNA sequence organization of the human genome. J. molec. Biol. **106**, 773–790 (1976)
Graham, D.E., Neufeld, B.R., Davidson, E.H., Britten, R.J.: Interspersion of repetitive and nonrepetitive DNA sequences in the sea urchin genome. Cell **1**, 127–137 (1974)
Kurnit, D.M., Shafit, B.R., Maio, J.J.: Multiple satellite deoxyribonucleic acids in the calf and their relation to the sex chromosomes. J. molec. Biol. **81**, 273–284 (1973)
Macaya, G., Cortadas, J., Bernardi, G.: An analysis of the bovine genome by density-gradient centrifugation. Europ. J. Biochem. **84**, 179–188 (1978)
Manning, J.E., Schmid, C.W., Davidson, N.: Interspersion of repetitive and nonrepetitive DNA sequences in the Drosophila melanogaster genome. Cell **4**, 141–155 (1975)
Mowbray, S.L., Gerbi, S.A., Landy, A.: Interdigitated repeated sequences in bovine satellite DNA. Nature (Lond.) **253**, 367–370 (1975)
Perlman, S., Phillips, C., Bishop, J.O.: A study of foldback DNA. Cell **8**, 33–42 (1976)
Southern, E.M.: Long range periodicities in mouse satellite DNA. J. molec. Biol. **94**, 51–69 (1975)
Studier, W.F.: Sedimentation studies of the size and shape of DNA. J. molec. Biol. **11**, 373–390 (1965)

Received September 4, 1979 / Accepted September 17, 1979 by R.B. Nicklas
Ready for press October 7, 1979