Proc. Nat. Acad. Sci. USA
Vol. 71, No. 5, pp. 1920–1924, May 1974

Generation of Specific Repeated Fragments of Eukaryote DNA

(site-specific endonuclease/repeat frequency/calf DNA/satellites/gel electrophoresis)

S. L. MOWBRAY AND A. LANDY*

Division of Biological and Medical Sciences, Brown University, Providence, Rhode Island 02912

Communicated by Joseph G. Gall, January 21, 1974

ABSTRACT **Calf-thymus DNA, hydrolyzed with a site-specific endonuclease from *Haemophilus influenzae* Rd, yields 12 discrete bands on polyacrylamide-agarose gels. These range in size from 7.5×10^4 to 2×10^6 daltons, and they represent about 5% of the total DNA with individual fragments comprising 0.1–1.5%. The various DNA segments are repeated between 1500 and 220,000 times per haploid genome. Whereas the wide range of reiteration frequencies suggests different origins for some of the fragments, the bias in fragment densities in CsCl and in Ag^+-Cs_2SO_4 toward those of known satellite DNAs suggests similar origins for some of them. Models for the possible origin of the DNA fragments can be grouped into three distinct, experimentally distinguishable, classes.**

The chromosomes of most eukaryotes are organized into unique and reiterated nucleotide sequences (for review see ref. 1). Unique regions are thought to be composed mainly of structural genes; the function of the reiterated DNA remains in question. In *Drosophila* (2), guinea pig (3), and kangaroo rat (4) DNAs, at least a portion of the repeated sequences, the satellite DNAs, appear to be a simple, basic repeat of only a few nucleotides in length. Recent studies indicate that some of the reiterated DNAs are interspersed between unique sequences. In the case of *Drosophila*, *Necturus*, and mouse DNAs, these interspersed, repetitive sequences appear to be distinct from the density satellites (5–8).

We have studied repetitive sequences in calf-thymus DNA using a site-specific endonuclease (restriction enzyme) prepared from *Haemophilus influenzae* Rd (9). This enzyme fraction [abbreviated Endo R·*Hin*d (10)], which consists of two site-specific nucleases (11), makes double strand scissions in the DNA at specific recognition sites (restriction sites).

Assuming that there are at least two nuclease recognition sites in a repeated nucleotide sequence, the number of copies of the specific DNA fragments produced from that sequence will be a function of the number of times it appears in the genome. When enzymatically hydrolyzed DNA is fractionated on polyacrylamide-agarose gels, some fragments derived from repeated sequences can be identified as discrete bands, whereas fragments from unique sequences are present as a continuous distribution of various size DNA fragments (Fig. 2). The size distribution of the fragments is, of course, determined by the spacing of the restriction sites within the DNA.

METHODS

Isolation of DNA. Calf-thymus tissue was homogenized and the nuclei were isolated using a modification of the procedures of Allfrey *et al.* (12). A Teflon-glass homogenizer was used to disrupt the tissues and break the cells. Nuclei were purified by centrifugation through 2.4 M sucrose in 3 mM $CaCl_2$, 10 mM Tris·HCl (pH 7.4) and were subsequently disrupted by incubation (12–15 hr) in sodium dodecyl sulphate (0.5%), Pronase (100 μg/ml), 50 mM KCl, 10 mM Tris·HCl (pH 8.0). DNA was extracted with an equal volume of chloroform-isoamyl alcohol (24:1) at least three times, dialyzed, incubated with T_1 and pancreatic RNAase (60 units/ml each), treated with pronase, and extracted with phenol.

Enzymatic Hydrolysis of DNA. Calf DNA was hydrolyzed with Endo R·*Hin*d as previously described (11). Conditions, chosen such that the reaction was complete, were about 0.5 units of enzyme (9) per ml added at $1^1/_2$ hourly intervals for a total of three additions.

Polyacrylamide Gel Electrophoresis. The DNA fragments produced with Endo R·*Hin*d are routinely separated on composite polyacrylamide-agarose slab gels (3% acrylamide, 0.5% agarose) and stained as previously described (11). Electrophoresis was usually for $4^1/_2$ hr at 200 V in a slab gel electrophoresis cell cooled to 2–4°. The amount and position of DNA was estimated by optically scanning a gel at 590 nm followed by integration of the area in a graphic read-out of the scan.

Equilibrium Centrifugation of DNA. **CsCl:** Calf DNA was previously hydrolyzed with Endo R·*Hin*d, extracted twice with phenol, and dialyzed against 10 mM Tris·HCl (pH 7.9), 1 mM EDTA. The DNA was brought to 1.67 g/cm³ with saturated CsCl, and centrifuged at 22° for 72 hr in a Spinco 50 rotor (42,000 rpm). Buoyant density was calculated from the refractive index (13).

Ag^+-Cs_2SO_4: The method of Jensen and Davidson (14) was used as modified by Yasmineh and Yunis (15). DNA was either hydrolyzed with Endo R·*Hin*d or sheared by passage three times through a 25-gauge needle using maximum thumb pressure on a syringe. Gradients (8 ml) were made in 10 mM Borate (Na^+), pH 8.7, and contained 200–300 μg of DNA. Ag_2SO_4 was added to give a Ag^+:DNA-phosphate ratio of 0.56 and Cs_2SO_4 was added to give a density of 1.58 ± 0.01 g/cm³. At this pH, GC-rich DNA will bind more Ag^+ (14), thus accentuating the separations normally obtained in CsCl. Centrifugation was for 72 hr at 22° in the Spinco 50 rotor (42,000 rpm).

RESULTS

*Hydrolysis of Calf DNA with Endo R·Hin*d. The DNA fragments produced by hydrolysis with Endo R·*Hin*d are best visualized on polyacrylamide-agarose gels after first enriching

Abbreviation: Endo R·*Hin*d, the site-specific endonuclease from *Haemophilus influenzae* Rd (abbreviated according to Smith and Nathans, ref. 10).

* To whom reprint requests should be sent.

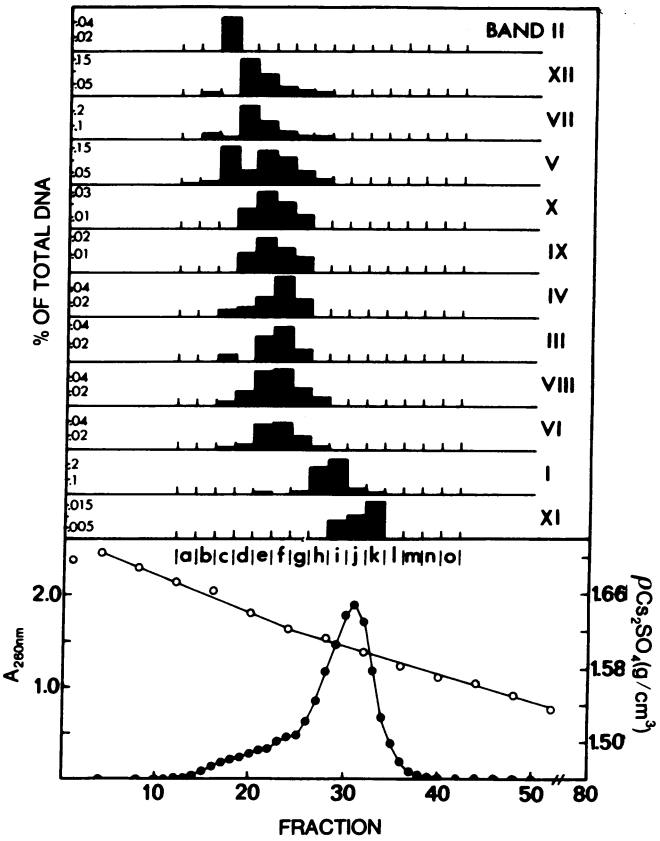

FIG. 1. Ag^+-Cs_2SO_4 equilibrium centrifugation of sheared calf thymus DNA. *Lower component of fig.:* ●——●, A_{260nm}; ○——○, density. *Upper portion of fig.:* each letter (*a* through *o*) represents one pooled fraction. After dialysis and concentration, the DNA from each pool was hydrolyzed with Endo R·*Hind*, fractionated on acrylamide gels, stained, and the gels were optically scanned (590 nm). The percentage of treated DNA is presented as a function of fraction number and density.

for these sequences by centrifuging sheared calf DNA in Ag^+-Cs_2SO_4 (Fig. 1). After enzymatic hydrolysis of fractions from the gradient, 12 bands are seen against a background of a continuous distribution of various size DNA fragments (Fig. 2).

The molecular weight of the various bands was determined by co-electrophoresis with DNAs of known size. The sizes of the fragments produced with Endo R·*Hind* from bacteriophage ϕ80h DNA were determined by Landy *et al.* (11), and these were used as standards (see Fig. 3). The repeated calf-DNA fragments have molecular weights between 7.5×10^4 (band XII) and 2×10^6 (band I).

Fraction of the Total DNA and Repeat Frequency. The repetitiousness of a sequence of nucleotides in the haploid genome is determined from the relationship:

$$Fx \cdot \frac{\text{molecular weight of haploid genome}}{\text{molecular weight of fragment Y}} = R, \qquad [1]$$

where R is the number of times sequence Y is repeated in the haploid genome, and Fx is the fraction of the total mass of DNA found in any fragment Y. Using this relationship, the approximate repeat frequencies of the various Endo R·*Hind* fragments of calf DNA have been determined. The fraction of the haploid calf genome represented by a fragment was determined by hydrolyzing calf DNA and fractionating the fragments electrophoretically on polyacrylamide-agarose gels as described. The gels were stained, optically scanned, and, finally, the areas under the tracings were integrated. The tracing of a typical gel is shown in Fig. 4. The curve consists of a number of peaks, corresponding to the visible gel bands, superimposed upon a continuous distribution of DNA fragments. The smooth curve of the interband regions was extrapolated as a baseline for purposes of calculating the total area under each peak. Dotted lines are our projections of individual peaks when two or more overlap. The areas calculated for each peak in several different experiments, including those where some peaks have been shifted to relatively different positions (compare Figs. 1, 5, and 6), are reproducible to within 30%, but the absolute error is probably considerably

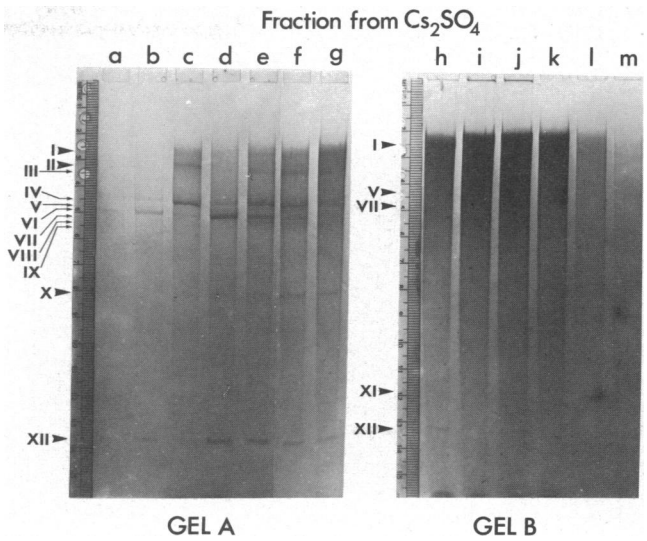

FIG. 2. Photographs of the gels containing fractions of the Ag^+-Cs_2SO_4 gradient of sheared DNA. See Fig. 1 for details.

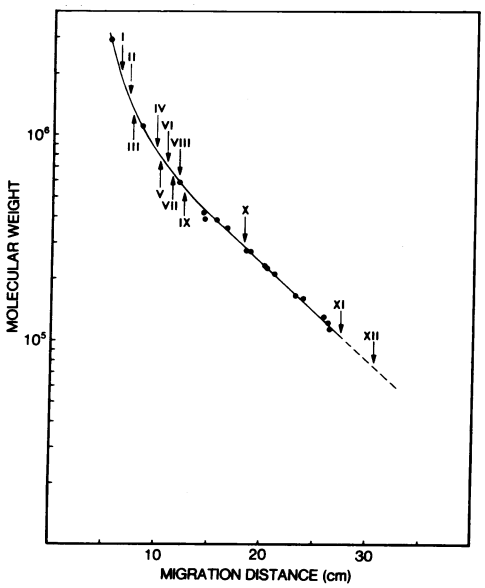

FIG. 3. Molecular weight determination of the Endo R·*Hind* fragments of calf DNA. The standard molecular weight markers, Endo R·*Hind* fragments of bacteriophage ϕ80h DNA (●——●), were electrophoresed in lanes parallel to the calf DNA fragments whose relative position is indicated by *arrows*.

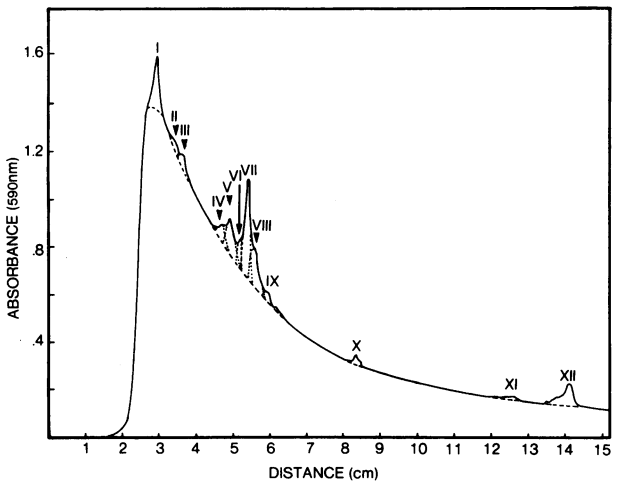

FIG. 4. Absorbance profile of a stained polyacrylamide-agarose gel containing Endo R·*Hin*d hydrolyzed calf DNA. The *solid line* is the observed absorbance (see *Methods*). The *dashed line* is the best estimate of the profile of background DNA, and the *dotted line* is a projection of the shape of individual peaks (see *text*).

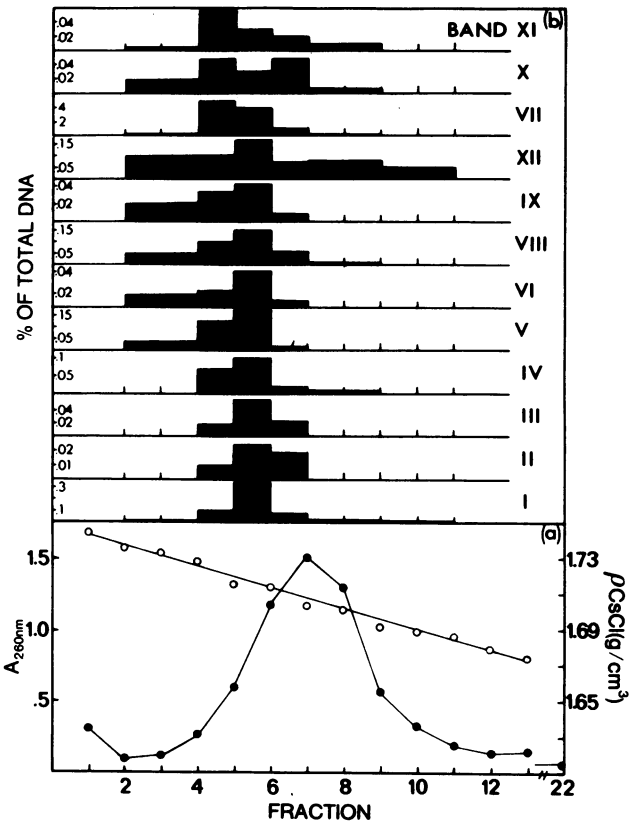

FIG. 5. CsCl centrifugation of Endo R·*Hin*d-hydrolyzed calf DNA. After centrifugation, fractions were collected and electrophoresed on acrylamide-agarose gels and subsequently stained. The *upper panel* is the distribution of the various fragments across the gradient, determined from an optical scan of the gel. See Fig. 1 for further details. A_{260nm} (●——●) and density (O——O), determined from the refractive index.

greater than this. Some nonspecific fragmentation during preparation may result in an underestimation of the amount of DNA present in the highest molecular weight fragments. Also, a considerable amount of extrapolation is necessary in projecting the shapes of the individual peaks. However, even rather gross alterations in the amount of DNA in each peak (e.g., a factor of two or more) will have little effect on the present interpretations of the relative numbers involved. In Fig. 4, band I coincides with the peak of the continuous distribution of fragments constituting the baseline. In gels containing previously fractionated DNA (e.g., on neutral CsCl, Fig. 5), the shape of band I can be determined more precisely because it is displaced from the peak of bulk DNA. The fraction of the total nucleotide present, represented by each fragment, is a simple ratio of its respective integrated area to the total area. The value for the fraction that each band represents of the total DNA (Table 1) ranges from 0.1% for band II to 1.5% for band VII, and the sum of all fragments is about 5% of the total.

The various fragments have repeat frequencies, as determined by Eq. 1, between 1,400 (band II) and 220,000 (band XII).

TABLE 1. *Molecular weight, repeat frequency, and buoyant density of repeated fragments of calf DNA generated with endo R·Hin*d

Fragment	Molecular weight ($\times 10^{-5}$)	Percent total DNA	Repeat frequency*	Buoyant density		
				CsCl (enzyme)	Ag^+-Cs_2SO_4 (enzyme)	Ag^+-Cs_2SO_4 (sheared)
I	20	0.58	5,800	1.716	1.611	1.606
II	15	0.10	1,400	1.716	1.631	1.652
III	12.8	0.17	2,700	1.716	1.642	1.626
IV	8.3	0.24	5,700	1.716	1.631	1.626
V	7.6	0.53	13,900	1.716	1.638	1.635
VI	6.9	0.11	3,200	1.716	1.669	1.619
VII	6.4	1.49	46,600	1.722	1.669	1.643
VIII	5.9	0.54	18,100	1.716	1.631	1.619
IX	5.4	0.15	5,500	1.716	1.624	1.626
X	2.9	0.13	8,600	1.722	1.650	1.626
XI	1.0	0.12	20,900	1.722	1.631	1.593
XII	0.75	0.82	220,000	1.716	1.650	1.643

* Minimum repeat frequencies were calculated from the molecular weight, percent of total DNA, and assuming 2×10^{12} daltons of DNA per haploid genome (17).

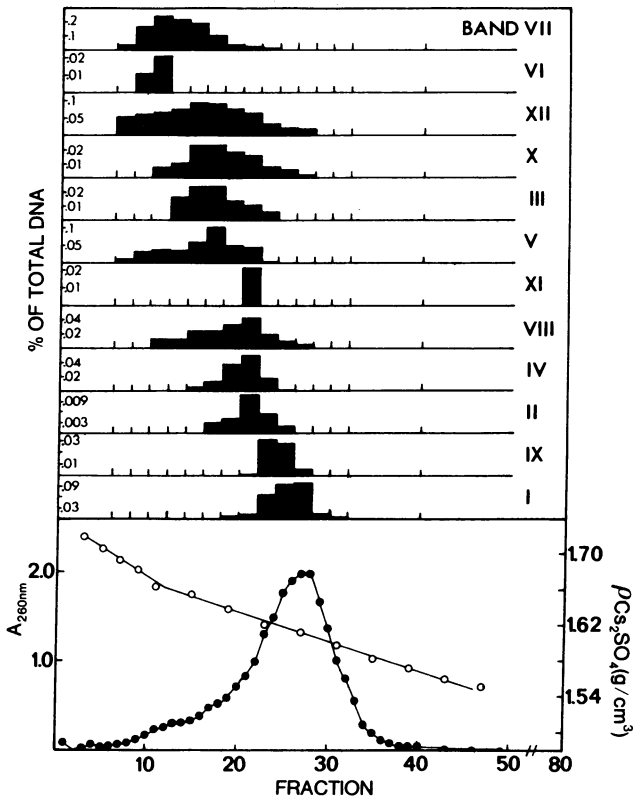

FIG. 6. Ag^+-Cs_2SO_4 equilibrium centrifugation of Endo R·*Hin*d-hydrolyzed calf-thymus DNA. The details of the experiment are the same as given in Fig. 1 except that calf DNA was enzymatically hydrolyzed before centrifugation. ●——●, A_{260nm}; O——O, density.

Densities in Neutral CsCl. Fig. 5 presents the optical density profiles of a CsCl gradient containing calf DNA hydrolyzed with Endo R·*Hin*d and the distribution of the various fragments across the gradient. The 12 bands exhibit only two density maxima: bands I through VI, VIII, IX, and XII with a mean density of about 1.716 g/cm^3, and bands VII, X, and XI having a mean density of 1.722 g/cm^3. The double peak in band X is an artifact. In other experiments, this fragment has a unimodal distribution with a peak density of 1.722 g/cm^3. These values are close to the peak densities of the two satellites found in CsCl (15).

Centrifugation in Ag^+-Cs_2SO_4. In order to obtain a different kind of fractionation, sheared calf DNA was centrifuged to equilibrium in a Ag^+-Cs_2SO_4 gradient. This also results in a better separation of satellite from bulk DNA than neutral CsCl. The main peak of DNA has a density of 1.602 g/cm^3, and there is a broad heavy region with no apparent peaks (Fig. 1). Fractions from the gradient were treated with Endo R·*Hin*d and the DNA was then fractionated on acrylamide-agarose gels. After staining, the gels were scanned and the areas were integrated. A photograph of the stained gels containing Endo R·*Hin*d hydrolyzed fractions of this gradient is presented in Fig. 2. The determination of the distribution of the 12 fragments across the gradient (Fig. 1) reveals that most fragments are more dense than main band DNA and are found in the regions of the heavy satellites (15). The 11 heavy fragments occur in six density regions of the gradient: 1.606 g/cm^3 (band I), 1.619 g/cm^3 (bands VI and VIII), 1.626 g/cm^3 (bands III, IV, IX, and X), 1.635 g/cm^3 (band V), 1.643 g/cm^3 (bands VII and XII), and 1.652 g/cm^3 (band II). Band XI is lighter than main band and possibly corresponds to a light satellite (15). Another light band may be present (Fig. 2, *gel B*, lanes 3 and 4 at about 9.1 cm); however, the band was not detected by optical scanning of the gel and it will not be discussed at the present time. Several bands (especially band V) appear to have bimodal distributions with a second heavy peak. It may be that these repetitious sequences are dispersed throughout the genome and have neighboring sequences of different base composition. Alternatively, this may reflect the presence of two different fragments of the same size.

In another Ag^+-Cs_2SO_4 gradient, the DNA was hydrolyzed with Endo R·*Hin*d before centrifugation. Fractions were collected, treated as described previously, and electrophoresed in polyacrylamide-agarose gels (Fig. 6). The broad distribution of density for some of the bands in this experiment relative to the one described previously, using sheared DNA, may be explained by the lower molecular weight of some fragments in the digested DNA. For example, band XII (molecular weight 7.5×10^4) is much more diffuse than band I (molecular weight approximately 2×10^6). The density maxima (see Table 1) of these 12 fragments are different from the previous experiment. This is to be expected, since the density of a fragment contained in a larger sheared sequence would be affected by the density of the adjacent nonrepeated DNA. An interesting result is that band XI (the light fragment in the sheared DNA-Cs_2SO_4 experiment) has a density greater than main DNA. This could be due either to the presence of neighboring sequences rich in AT, or to interactions with neighboring sequences that effect the amount of bound silver (14) (thereby resulting in a relative decrease in density).

DISCUSSION

In this report we have demonstrated the production of 12 repeated DNA fragments from the calf genome using a restriction endonuclease from *H. influenzae* Rd. They represent approximately 5% of the total DNA with individual fragments comprising between 0.1% and 1.5%. Among the 12 acrylamide bands there is a bias toward the density of calf satellite DNA on either neutral CsCl or Ag^+-Cs_2SO_4 gradients (15, 16). There are between 1,400 copies per haploid genome for the least repeated fragment and 220,000 copies for the most repeated one.

Other eukaryote DNAs hydrolyzed with the same endonuclease give different patterns and different numbers of bands (unpublished results).

A band in an acrylamide gel must contain approximately 0.05 μg of DNA in order to be detected using the procedures described here. It is easily demonstrated that in order to visualize a fragment 1000 base pairs in length, it must be repeated about 1500 times (given that the molecular weight of the calf genome is about 2×10^{12} and the amount of DNA normally loaded onto a gel is 100 μg).

Possible Origin of Fragments. Models for the possible origin of the DNA fragments described above can be grouped into three distinct classes which differ considerably in their functional implications. Although there are many possible variations, each of the three general classes depicted schematically in Fig. 7 can be distinguished experimentally.

Class A is the simplest collection of DNA fragments. These molecules are comprised of a single long sequence which occurs many times in the genome and which also contains two nuclease-sensitive sites. The essential feature of the molecules in

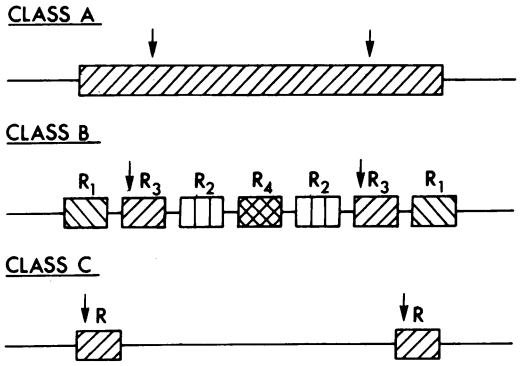

FIG. 7. Models for the origin of repeated DNA fragments. *Cross hatched boxes* indicate reiterated DNA sequences, and *arrows* indicate cleavage sites for the site-specific endonuclease. *Class A:* The reiterated sequence contains two cleavage sites. *Class B:* Reiterated sequences are tandemly arranged with one sequence (R_3 in this example) containing a cleavage site. R_1, R_2, and R_4 may or may not be identical in sequence. The number, arrangement, and spacing of various repeated sequences (R_1, R_2, etc.) are not specified except that R_3 must occur at regular intervals. *Class C:* Cleavage sites are separated at regular intervals by nonrepetitious DNA. The sequence R is minimally an endonuclease recognition site, or it may be a portion of a larger repeated sequence.

this class is that they possess relatively little internal reiteration and that they constitute a homogeneous population. In contrast to the other two classes of fragments, these molecules will reanneal with a rate close to that predicted by their size and calculated repeat frequency (Table 1).

The nucleotide sequence of the fragments in class B is comprised of a small number of blocks of relatively simple nucleotide sequences which are repeated and combined to generate the total fragment length. One of the sequence blocks contains the nuclease-sensitive site and is repeated at some constant interval which defines the fragment size. Because of the considerable internal reiteration, the renaturation kinetics of these fragments would be expected to be appreciably more rapid than predicted on the basis of the fragment repeat frequently alone. Satellite DNAs described by the sequence studies of J. Gall (2), Southern (3), and Fry *et al.* (4) are perhaps analogous to this class of DNA fragments. In this instance, it would require that short, repeating sequences are interspersed at regular intervals with different or variant sequences too infrequent to be detected by the usual sequencing procedures.

The DNA fragments in class C differ from the other two classes in that a significant portion of each molecule consists of unique, or low-multiple copy, DNA. The DNA molecules within a single acrylamide band differ from one another over a considerable portion of their length. What brings these different DNA sequences together in a single band is the fact that they are bounded at the same interval by a reiterated sequence which contains the nuclease recognition site, and hence are the same in length. This class can be distinguished from the other two in that there will be a large fraction of the total fragment which will anneal with itself or with total DNA only at a high Cot. Evidence for such a model of interspersed unique and reiterated DNA has been recently reported, and the calculations on the relative sizes and reiteration frequencies of these regions (5–8) are compatible with some of the fragments reported here (Table 1).

There is no compelling reason to assume that all of the fragments described above fall into the same class. In some instances, the broad density distributions observed for several of the fragments (compare Figs. 5 and 6) may reflect the type of heterogeneity considered in class B and (or) C. The wide range of reiteration frequencies (greater than 100-fold) suggests different origins for some of the fragments. On the other hand, the bias of the fragment densities in CsCl and Ag^+-Cs_2SO_4 toward those of known satellite DNAs suggests similar origins for some of the fragments.

Experiments are currently in progress to measure the extent of internal redundancy within the population of molecules comprising each of the reiterated fragments. This will determine which of the fragments are oligomers of simple sequences and which are more complex sequences interspersed throughout the calf genome.

We acknowledge and thank Dr. Susan Gerbi for helpful discussions and critical reading of the manuscript. This work was supported by grants from the American Cancer Society (NP-118B) and the National Institutes of Health (CA-11208). One of us (S.L.M.) was a post-doctoral fellow supported by Public Health Service Training Grant T01-A1-00418.

1. Britten, R. J. & Davidson, E. H. (1971) *Quart. Rev. Biol.* **46**, 111–133.
2. Gall, J. G. & Atherton, D. D. (1974) *J. Mol. Biol.*, in the press.
3. Southern, E. M. (1970) *Nature New Biol.* **227**, 794–798.
4. Fry, K., Poon, R., Whitcome, P., Idriss, J., Salser, W., Mazrimas, J. & Hatch, F. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2642–2646.
5. Hearst, J. E., Botchan, M., Kram, R. & Beals, E. (1973) *Biochim. Biophys. Acta* **294**, 173–183.
6. Wu, J., Hurn, J. & Bonner, J. (1972) *J. Mol. Biol.* **64**, 211–219.
7. Davidson, E. H., Hough, B. R., Amenson, C. S. & Britten, R. J. (1973) *J. Mol. Biol.* **77**, 1–24.
8. Lee, C. S. & Thomas, C. A., Jr. (1973) *J. Mol. Biol.* **77**, 25–55.
9. Smith, H. O. & Wilcox, K. W. (1970) *J. Mol. Biol.* **51**, 379–391.
10. Smith, H. O. & Nathans, D. (1973) *J. Mol. Biol.* **80**, 419–423.
11. Landy, A., Ruidisueli, E., Robinson, L., Foeller, C. & Ross, W. (1974) *Biochemistry*, in the press.
12. Allfrey, V. G., Littau, V. C. & Mirsky, A. E. (1964) *J. Cell Biol.* **21**, 213–231.
13. Vinograd, J. & Hearst, J. E. (1962) *Progr. Chem. Org. Natur. Prod.* **20**, 372–422.
14. Jensen, R. H. & Davidson, N. (1966) *Biopolymers* **4**, 17–32.
15. Yasmineh, W. G. & Yunis, J. J. (1971) *Exp. Cell Res.* **64**, 41–48.
16. Filipski, J., Thiery, J.-P. & Bernardi, G. (1973) *J. Mol. Biol.* **80**, 177–197.
17. Callan, H. G. (1972) *Proc. Roy. Soc. Ser. B.* **181**, 19–41.