Organization of Highly Repeated Sequences in Mouse Main-band DNA

THOMAS R. CECH† AND JOHN E. HEARST‡

Department of Chemistry
University of California
Berkeley, Calif. 94720, U.S.A.

(Received 19 May 1975, and in revised form 7 October 1975)

Renaturation analysis shows that the main CsCl density band of mouse DNA (at 400 bases mol. wt) contains a minimum of four second-order kinetic components: non-repetitive DNA; two major classes of repeated DNA, called "moderately repeated" and "highly repeated", each comprising 8 to 16% of the genome; and a minor sub-class of the highly repeated DNA, about 0.5% of the genome with a satellite-like complexity. The highly repeated main-band sequences are isolated from high molecular weight DNA (12,000 bases) by renaturation to low Cot § and hydroxyapatite chromatography. This DNA is spread for electron microscopy by the formamide isodenaturing technique. Most of the bimolecular renaturation products show interspersion of renatured and single-stranded regions. The renatured regions average about 700 base-pairs in length, most being in the range of 200 to 1000. On molecules where two or more renatured regions can be identified, they are separated by a wide range of distances averaging 1800 bases. A minority fraction (about 12% by weight) of the bimolecular renaturation products shows no evidence of interspersion, and may consist of tandemly repeated sequences.

The renaturation rate of the highly repeated DNA fraction (at initial mol. wts of 6000 and 12,000 bases) is studied before and after shearing the DNA to low molecular weight. The kinetic complexity of the interspersed, single-stranded DNA that is attached to renatured, highly repeated DNA is thereby determined. Most of the unrenatured DNA contains highly repeated sequences similar in complexity to the sequences that renatured at low Cot. Only about 20% of the interspersed DNA has a low repetition frequency, approximately non-repetitive. We conclude that about 15% of the mouse main-band DNA contains highly repeated sequences, most of which are organized as permuted tandem repeats. About 5% of the non-repetitive DNA of mouse is interspersed with this fraction of the genome. This model of sequence organization differs from those that have been demonstrated for more moderately repeated sequences of higher organisms, where much more non-repetitive DNA is interspersed with the repeated sequences.

† Present address: Department of Biology, Building 16, Room 717, Massachusetts Institute of Technology, Cambridge, Mass. 02139, U.S.A.
‡ To whom reprint requests should be sent.
§ Abbreviations used: Cot, product of initial DNA concn (mol. DNA PO₄/l) and time (s); equiv. (equivalent) Cot, Cot corrected to the standard renaturation condition of 0.12 M-phosphate buffer. This definition of equiv. Cot differs from that used in Cech et al. (1973), where Cot was corrected for both salt concentration and DNA molecular weight to give equiv. Cot; ds, double-stranded; ss, single-stranded.
1. Introduction

Largely through renaturation kinetics studies of the DNA of a wide variety of eukaryotic organisms, two general patterns of organization of repeated DNA sequences are emerging. Satellite DNAs are composed of simple (oligomeric) sequences *tandemly repeated*, with variable accuracy, hundreds or thousands of times without interruption (Waring & Britten, 1966; Bond et al., 1967; Southern, 1970; Gall & Atherton, 1974; Goldring et al., 1975; Walker, 1971 (review)). In the main CsCl density bands of all eukaryotic DNAs that have been studied, middle repetitive sequences are *interspersed* with non-repetitive sequences (see reviews by Davidson & Britten, 1973; Lewin, 1975).

Main-band DNA usually contains highly repeated as well as moderately repeated sequences. Some human and *Drosophila* main-band DNA components are simple sequence DNAs that occur in heterochromatic regions of chromosomes; these differ from satellites by base composition only (Corneo et al., 1971; Jones & Corneo, 1971; Marx, Allen & Hearst, 1976; Peacock et al., 1974). Other highly repeated main-band sequences in mouse and human DNA (Cech et al., 1973; Corneo et al., 1970; Marx et al., 1976) appear to be interspersed with other sequences rather than tandemly repeated. *In situ* hybridization shows that these highly repeated sequences are not restricted to centromeres. The average repetition frequency is intermediate between that of simple sequence DNAs and the major middle repetitive component of, for instance, *Xenopus*; to minimize confusion, they will be called “highly repeated” main-band sequences. The same sequences, when isolated in renatured form, have been called “h.a.r.r.DNA” for hydroxyapatite-isolated rapidly renaturing DNA.

In our earlier study (Cech et al., 1973), we analyzed the hydroxyapatite yield, CsCl buoyant density, S1-nuclease resistance, and hyperchromicity of renatured, highly repeated mouse main-band DNA as a function of DNA fragment size. We concluded that repeated sequence DNA of average length 1.5±0.5 kb† is interspersed with other DNA of average length 2.2±1.1 kb. In the present study, electron microscopic observation of the renatured molecules gives direct confirmation of this interspersion model for most of the highly repeated DNA. Electron microscopy gives the length distribution of the renatured segments as well as the average length. Hydroxyapatite renaturation studies result in a better estimate of the repetition frequency of the highly repeated DNA and show that it differs from more moderately repeated DNA. Most significant is the unexpected finding that the highly repeated DNA is interspersed with other highly repeated sequences, as well as with some non-repetitive DNA. Different highly repeated sequences seem to be organized in a permuted tandem repeat arrangement.

2. Materials and Methods

(a) Mouse DNA

SVT2 mouse tissue culture cells were originally obtained from Theodore Gurney, Jr (University of Utah). The cells are transformed by simian virus 40, and they have a diploid number of chromosomes. The labeling, extraction and purification of DNA from these cells has been described previously (Cech et al., 1973). Unlabeled Balb/c mouse DNA

† Abbreviations used: kb (kilobase), 1000 bases of ss or 1000 base-pairs of ds DNA; HAP, hydroxyapatite.
HIGHLY REPEATED MOUSE MAIN-BAND DNA

was extracted from combined livers, spleens, testes and brains. The procedure followed that described by Flamm et al. (1966), with the addition of a Pronase treatment before the final deproteinization. Polysaccharides were removed by a 30-min centrifugation at 45,000 revs/min in the Beckman SW50.1 rotor.

(b) Escherichia coli DNA

Frozen, ³H-labeled E. coli (strain HF4733 F⁻ thy⁻ gal⁻ Endo I⁻), originally a gift of A. John Clark (University of California, Berkeley), were obtained from David Appleby (Berkeley). The cells were thawed and treated with lysozyme, after which the DNA was extracted and purified by a method similar to the SVT2 mouse DNA method. The measured specific activity was 1.3 x 10⁸ cts/min per μg with DNA concentration determined by analytical CsCl density gradient centrifugation of the unsheared E. coli DNA with a known amount of mouse satellite DNA. This method obviated the possibility that the measured optical density of the DNA would include a contribution from scattering or contaminating RNA, resulting in an underestimated specific activity.

(c) Shearing of DNA and determination of molecular weights

Our procedures for shearing native DNA have been described previously (Cech et al., 1973). Briefly, needle shearing was used to produce DNA of 5 to 12 kb single-strand molecular weight and sonication for molecular weights of 0.3 to 0.4 kb. Molecular weights were determined by analytical boundary sedimentation in alkaline solution, using the equations of Studier (1965). These equations are based on data at molecular weights above 6 kb; so their use for the sonicated samples is an extrapolation.

In shearing renatured DNA, it is important to know whether double-stranded and single-stranded regions are being reduced to similar fragment sizes. Therefore, both native and denatured high molecular weight mouse DNA samples were sheared by two methods, and single-stranded molecular weights determined. As shown in Table 1, rapid decompression

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Shearing of DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single-stranded mol. wt (kb)</td>
</tr>
<tr>
<td></td>
<td>Native DNA</td>
</tr>
<tr>
<td>Bomb</td>
<td></td>
</tr>
<tr>
<td>400 lb/in²</td>
<td>1.2</td>
</tr>
<tr>
<td>1200 lb/in²</td>
<td>1.1</td>
</tr>
<tr>
<td>Sonicate</td>
<td></td>
</tr>
<tr>
<td>0.30 M-PO₄</td>
<td>0.33</td>
</tr>
<tr>
<td>0.05 M-PO₄</td>
<td></td>
</tr>
<tr>
<td>Driver 1</td>
<td>0.32</td>
</tr>
<tr>
<td>Driver 2</td>
<td>0.40</td>
</tr>
</tbody>
</table>

All molecular weights were determined on single-stranded DNA; “native” and “denatured” refer to the state of the DNA before shearing. In shearing DNA in the decompression bomb, each sample was equilibrated with argon for 10 min at the indicated pressure and then rapidly decompressed as it exited from the bomb through a needle valve. Sonication of nitrogen-saturated solutions was done in an ice bath in 5 30-s blasts. All samples were sheared in 0.05 M-phosphate buffer except where indicated. They were then dialyzed into 1 M-NaCl and made 0.1 M in NaOH for the sedimentation analysis. The unsheared single-stranded mol. wt was at least 40 times the sheared value in each case. “Driver 1” is the Balb/c DNA used in the driven renaturation experiment of Table 4; “driver 2” is that used in all the experiments summarized in Fig. 6. Driver 2 was sonicatod in a much larger volume than was used in the previous 2 experiments, probably accounting for its higher mol. wt. The relative resistance of denatured DNA to shear by sonication may be due to the higher ionic strength used in that experiment; denatured DNA in 0.30 M-phosphate buffer was found to be quite resistant to shear by the bomb method (data not shown).
in a bomb and sonication sheared single- and double-stranded DNA to somewhat different extents. Both techniques, however, were adequate for our purposes.

(d) Isolation of renatured DNA

DNA was alkali denatured at room temperature, warmed to 60°C, and neutralized with NaH₂PO₄ such that the final buffer was 0.12 M-phosphate buffer (Kram et al., 1972). Phosphate buffer will be used to indicate an equimolar mixture of Na₂HPO₄ and NaH₂PO₄ with a pH of 6.8. After incubation at 60°C to the desired Cot, renatured DNA was isolated by batch hydroxyapatite chromatography (Cech et al., 1973). Unless stated otherwise, the criteria for base-pairing were of “high stringency”—0.12, 0.16 and 0.45 M-phosphate buffer elutions were performed at 70°C. “Less stringent” criteria (omitting the 0.16 M-phosphate elution and eluting at 60°C) were sometimes used, especially in isolating fold-back DNA. The fraction of the DNA renatured (fraction bound to HAP) was calculated as the total radioactivity or optical density units eluted at 0.45 M-phosphate divided by the total recovered at all 3 phosphate concentrations. Recoveries averaged 100±5% of the initial cts/min.

(e) Hydroxyapatite assay of renaturation kinetics

The use of an excess of sonicated, unlabeled DNA to drive the renaturation of a trace amount of [³H]DNA, monitoring renaturation by HAP binding, has been extensively described by Britten and his co-workers (Britten & Smith, 1970; Davidson et al., 1973), and our method was taken from theirs. Each sample contained 40-60 μl total Balb/c mouse DNA (0.32 to 0.40 kb) to drive the renaturation. When this DNA was chromatographed on HAP without denaturation, 99.0% eluted in the double-stranded fraction. DNA concentrations never exceeded 50 O.D. units/ml (2.5 mg/ml), thereby avoiding high viscosities. Each sample also contained 2000 to 12,000 cts/min of ³H-labeled SVT2 tracer DNA at 4.0 × 10⁴ or 2.65 × 10⁵ cts/min per μg. The ratio of driver to tracer was 1500 in the Table 4 experiment, and ranged from 2500 to 22,000 for the rest. For renaturation times of less than 1 h the samples were denatured and renatured as described in section (d). For longer time points, samples were denatured with NaOH at room temperature, neutralized with ice cold NaH₂PO₄, sealed into glass ampoules or capillaries, and heated quickly to 60°C by submersion in a water bath. For very low Cot values (~10⁻⁶), sonicated Micrococcus lysodeikticus DNA was substituted for the Balb/c mouse DNA, with Cot based on the concentration of ³H-labeled mouse DNA only. Some very high Cot samples were renatured in 0.90 M-phosphate buffer at 70°C (to maintain the tm = 25°C criterion), in which case the Cot was multiplied by 8.06 (Britten et al., 1974) to give equiv. Cot. At the end of the renaturation time, small volume samples were diluted to give 4 ml with a final concentration of 0.12 M-phosphate. Samples were never frozen, but were either chromatographed immediately or diluted and stored on ice. HAP fractionation was done by the batch method as described in section (d), using the less stringent elution conditions. (Although 70°C and 0.16 M elutions can be useful methods of selecting for good hybrids, their use in a hybridization assay would effectively eliminate the possibility of some molecules ever being scored as renatured, because the assay criterion would be more stringent than the hybridization criterion.) The volumes of the 8 eluants were measured and the fractions were centrifuged at 2000 g for 10 min to pellet any finely suspended HAP. Optical densities were measured at 260 nm and 340 nm at room temperature in the Cary 15 spectrophotometer, using hydroxyapatite-washed phosphate buffers (which absorb less than unwashed buffers) as reference solutions. Recovery of O.D. units ranged from 111 ± 4% after 8-min incubations to 97% after 100 h incubations. The >100% recoveries are due to the hyperchromicity of both single-stranded and renatured DNA. After optical density measurements, 3.75 ml of each 0.12 M-eluant, or 1 ml of a 0.45 M-eluant diluted with 2.75 ml of water, was dissolved in 15 ml of cocktail (1 part Triton X100:2 parts PPO-POPOP in toluene) and counted for 5 to 10 min in a Beckman 200 liquid scintillation counter. Recovery of radioactivity ranged from 100±4% after short incubations to 88±4% after long times. Subsequent to the experiment of Table 4, all renaturation samples contained 2 × 10⁻⁷ mol EDTA. This gave an EDTA concentration...
of 1 mM in 0.2 ml samples, which were incubated for the longest times, and a proportionally lower concentration in larger volume samples. Because the 0.2 ml samples were diluted 20-fold before HAP fractionation, it is not surprising that no effect of the EDTA on the fractionation or the o.d. readings was noticed.

(f) Electron microscopy

Formamide spreading of DNA was done using the isodenaturing technique of Davis et al. (1971). The formamide concentration in the spreading solution was 45% or 50%. At room temperature, the denaturing ability of these spreading solutions is equivalent to 62 and 66°C, respectively, in 0-12 mM-phosphate buffer. Circular phage DNA was added to some preparations before spreading to give internal length standards. Double-stranded phage PM2 DNA was taken to have a molecular weight of 9.4 kb, and single-stranded fd DNA a molecular weight of 5.4 kb. Further details of our electron microscopic methods are given in Cech & Hearst (1975).

3. Results

(a) Isolation of highly repeated sequences from mouse main-band DNA

Our procedure for isolating highly repeated sequences from the main CsCl buoyant density class of mouse DNA is described in Cech et al. (1973). First, the simple sequence light satellite DNA is removed by Ag⁺/Cs₂SO₄ density gradient centrifugation. The main-band DNA is then sheared, to a molecular weight of 5 or 12 kb in the experiments described here. Next, foldback DNA (Walker & McLaren, 1969; Britten & Smith, 1970) is removed by denaturing the main-band DNA, renaturing it for one and a half to two minutes at low concentration (C₀ < 10⁻⁵ to 10⁻³), and removing strands containing base-paired regions by hydroxyapatite (HAP) chromatography. The foldback DNA is largely the result of intramolecular renaturation of inverted repeat sequences (Wilson & Thomas, 1974; Cech & Hearst, 1975). Our method of isolating foldback DNA removes 90% of the HAP-bindable inverted repeat sequences that can be identified in the electron microscope. Very little bimolecularly renaturing DNA is contained in this fraction.

The DNA not bound in the first HAP chromatography (i.e. the single-stranded fraction) is further incubated at 60°C in 0-12 mM-phosphate buffer to a C₀ ranging from 4 × 10⁻⁵ to 8 × 10⁻³, when bimolecular renaturation products are isolated by a second HAP chromatography. Depending on the DNA molecular weight and the C₀, 4 to 10% of the main-band DNA is collected in the second HAP chromatography. (For a more systematic study of the HAP yield of this DNA, see Cech et al., 1973.) Because the fraction of the DNA bound to HAP increases continuously with C₀ in this range, the procedure does not result in the isolation of a whole class of repeated sequences. Rather, it selects for the most highly repeated sequences. Data presented in sections (d) and (e) will demonstrate that the isolated sequences are in fact highly repeated, not simply the first moderately repeated sequences that happen to renature. This data will also show that the kinetic class or classes represented by the isolated highly repeated sequences make up 8 to 16% of the mouse genome.

The highly repeated DNA bound in the second HAP chromatography usually undergoes a further kinetic purification. The DNA is diluted to a phosphate concentration of 0.05 M, alkali denatured, neutralized and renatured in 0-12 mM-phosphate in the same volume and for the same time as the previous renaturation. Double-stranded DNA is then collected on HAP. (This is the third HAP chromatography.) This pro-
FIG. 1. Models of highly repeated DNA sequence organization and their predicted renaturation products. The left half of the Figure shows 4 ways in which repeated sequences could be organized in the native DNA. Each arrow represents random shearing of the DNA, denaturation, and renaturation to a C$_\text{o}$F sufficient for the reassociation of highly repeated sequences. The right half of the Figure shows some of the resulting types of renaturation products, with renatured regions indicated (I I I I I I). In most cases, more complex structures would form with subsequent nucleation events. (a) The repeated sequence (---) is interspersed with more complex DNA (----) and is isolated from other repeated sequences, such that a shear fragment will contain only one repeated sequence region. The expected renaturation products are (i) linear ds/ss molecules (partly double-stranded and partly single-stranded), (ii) single-forked molecules, and (iii) double-forked molecules. (b) Regions of repeated and more complex DNA alternate, such that at moderately high molecular weights many shear fragments will contain 2 or more
procedure is called "recycling". The highly repeated DNA isolated after the recycling step amounts to only 0.6 to 1.5% of the main band DNA (6.7 to 9.9% isolated as double-stranded at \(C_0t = 6 \times 10^{-2}\), and 9 to 15% of this reisolated after recycling at \(C_0t = 6 \times 10^{-3}\)). Although the recycling provides a rigorous purification of the DNA according to its renaturation kinetics, it seemed dangerous to study such a small fraction of the genome and attribute its properties to a larger class. Therefore, a portion of the unrecycled highly repeated DNA (isolated on the second HAP chromatography) was often saved and characterized along with the recycled fraction.

(b) Models of sequence organization

When single-stranded DNA containing highly repeated sequences is renatured and isolated on hydroxyapatite, the types of structures that are formed should be indicative of the way the sequences are organized in the native DNA. Figure 1 diagrams the primary (bimolecular) and secondary renaturation products expected for four possible types of sequence arrangement. In many cases the structures contain regions of highly repeated sequence that are not base-paired, so more complex structures should form upon prolonged renaturation.

Models (b) and (d) in Figure 1, in which stretches of highly repeated DNA alternate with stretches containing different sequences (either highly repeated or more complex), are consistent with our previous data for mouse main-band sequences (Cech et al., 1973). We preferred model (b) because, even when incubated to an equivalent \(C_0t\) of 50, the renatured DNA retained a considerable amount of single-strandedness. In our earlier work, however, we made no direct determination of the complexity of the interspersed, single-stranded sequences.

† There are 3 reasons for this low recovery: (1) much of the DNA isolated at \(C_0t = 6 \times 10^{-2}\) consists of sequences with \(C_0t_1 > 6 \times 10^{-2}\). Each recycling step will therefore lead to the renaturation of less than one-half of such sequences. (2) The second HAP chromatography is done under conditions of low stringency (see Materials and Methods), but the third is done at high stringency, resulting in the loss of marginally-stable duplexes. (3) Single-strand breaks occur during the alkali denaturation steps, 60°C incubations, and/or HAP chromatographies. These breaks release more-complex sequences from attachment to highly repeated sequences, preventing the former from binding in the double-stranded fraction upon recycling. Reduction of the size of the highly repeated sequences also slows their renaturation, as evidenced by data presented in section (f) of Results.

repeated sequence regions. The renaturation products include those of model (a), but in addition several types of multiple-forked structures (iv, v, vi) can form. If neighboring repeated sequence elements are similar, or if they are different but often occur in the same order, then structures (iv) and (vi) will be common. Structures like (v) give no information about the order of repeated sequence elements. They could result from a second nucleation on a single- or double-forked molecule, or from a single-strand break in structure (iv) or (vi). (c) Identical or closely related sequences are tandemly repeated. Many satellite DNAs are organized in this manner. The renaturation products do not contain structures in which a double-stranded region forks into 2 single-stranded tails, which distinguishes this model from the others. Subsequent nucleation events produce branched structures (viii) and hyperpolymers (Britten et al., 1974). (d) Repeated sequences are tandemly arranged, but the various sequence elements do not always follow each other in the same order. It is not necessary that adjacent sequence elements be unrelated, but only that they be different enough to prevent base-pairing under the hybridization criteria used. Most of the primary and secondary renaturation products predicted for this model would look identical in the electron microscope to those of model (b): (ix) looks like (i), (x) looks like (iii), (xi) looks like (iv), and so on. Although one can conceive of structures that would be unique to the permuted tandem repeated model, these involve many forks and would therefore probably be identified only as "messes"—structures that are not unambiguously traceable.
Types of molecular structures observed

Highly repeated mouse main-band DNA was isolated as described in section (a), above, and spread for electron microscopy in 45 or 50% formamide. At room temperature in these spreading solutions, base-paired regions of DNA were more stable than in the 70°C, 0.12 M-phosphate buffer–HAP elution (see Materials and Methods, section (e)). Furthermore, in the optical melting profile of this class of DNA, most of the hyperchromicity occurred above 70°C (Cech et al., 1973). Therefore, the spreading was not expected to denature the renatured regions that were responsible for the DNA eluting from HAP in the double-stranded fraction.

The highly repeated DNA fractions contained the types of molecules shown in Figures 2 and 3. Molecules were classified on the basis of the number of forks and free ends in each structure, as described in the legend to Figure 2. The double-stranded regions of renatured molecules could usually be identified by the fact that they were more densely stained and less kinky than single-stranded regions. The most difficulty arose in specifying the strandedness of short regions, because there was less distance over which the strand density could be averaged. Conversely, single-stranded phage fd DNA and double-stranded PM2 DNA, added as molecular weight standards, contained long regions of the same strandedness; these DNAs could always be positively differentiated by the criteria of strand density and kinkiness.

A summary of the types of molecules seen is given in Table 2. At initial DNA molecular weights of both 6 and 12 kb, and in the unrecycled as well as recycled preparations, the largest class of structures is the forked class, making up to 32 to 41% of the samples. In addition, many of the messes probably contain hidden bimolecular forks. The low proportion of linear (ds/ss) structures suggests that most of the mouse main-band highly repeated sequences are not arranged as simple tandem repeats (Fig. 1, model (c)). Any mouse satellite DNA contaminating the main-band DNA would not contribute significantly to either the forked or linear categories, because satellite forms distinctive mess structures even when renatured to low Cot values (Cech, 1975).

Table 3 subdivides the forked class into the types of structures diagrammed in Figure 2. Comparison of the observed molecules and those expected for the various models in Figure 1 suggests that alternating interspersion and/or permuted tandem...
HIGHLY REPEATED MOUSE MAIN-BAND DNA

Fig. 2. Examples of the types of renatured molecules seen in the highly repeated main-band DNA fraction. These molecules were on the same grids used for the measurements of Fig. 5(a), (b) and (d). The bar has a length corresponding to 1.0 kb of double-stranded or 1.2 kb of single-stranded DNA. In the interpretation to the left of each photo, presumed double-stranded regions are indicated by heavy lines, single-stranded regions by thin lines, and regions of uncertainty by dashed lines. (a) Linear ds/ss molecule, an unforked bimolecular renaturation product with single-
FIG. 3. Highly repeated main-band DNA isolated on hydroxyapatite and spread for electron microscopy in 45% formamide, as in Fig. 2. The bar again represents a length of 1-0 kb (ds). In the interpretation, double- and single-stranded regions are indicated as in Fig. 2.

(a) A molecule of the “multiple fork” category plus 3 single-stranded fd DNA circles. Although the presumed renatured region on the right is too short to appear double-stranded, the low
repetition (models (b) and/or (d)) may be the predominant types of sequence organization. The high proportion of multiple-forked and complex structures (Fig. 4) would not be expected if the repeated sequence elements were isolated in the genome (model (a)). As expected for molecules containing several segments of repeated DNA, more forked molecules are in the form of complex structures at the higher molecular weight. An increased number of forked molecules are also probably hidden in meses, which would explain the apparent decrease in the frequency of forked structures (from 40 to 32%) with higher molecular weight.

The single-stranded (ss) molecules contain no discernable base-paired regions. Because it is difficult to identify short double-stranded regions, some of these molecules could be ds/ss bimolecular renaturation products. Most ss molecules are short, however, and therefore probably result from breaks in the single-stranded tails of forked molecules or hairpins, during or after HAP fractionation.

concentration of mouse DNA on this grid makes it unlikely that it is simply a cross-over of 2 strands (see text). (b) A molecule of the “complex” category plus 2 fd DNA circles. (c) Single-stranded fd DNA (above) and double-stranded PM2 DNA (below), added to the mouse DNA as molecular weight standards. As discussed in the text, the strandedness of these DNAs was always clearly differentiable.
TABLE 2

Electron microscopic categorization of molecules in hydroxyapatite-fractionated DNA

<table>
<thead>
<tr>
<th>HAP fraction</th>
<th>% of main-band DNA</th>
<th>$n$</th>
<th>Forked</th>
<th>ds/ss</th>
<th>% of molecules</th>
<th>ss</th>
<th>Hairpin</th>
<th>Mess</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 kb DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highly repeated DNA</td>
<td>(7x10^{-4} &lt; C_{A}^{-1} &lt; 6 \times 10^{-3}, h. string.)</td>
<td>9.9</td>
<td>274</td>
<td>40±1</td>
<td>8±2</td>
<td>25±2</td>
<td>7±0</td>
<td>19±1</td>
</tr>
<tr>
<td>Same, recycled at C_{A}^{-1}</td>
<td>6x10^{-3} (h. string.)</td>
<td>1.5</td>
<td>388</td>
<td>41±2</td>
<td>38±0*</td>
<td>10±1</td>
<td>11±3</td>
<td></td>
</tr>
<tr>
<td>Unbound</td>
<td></td>
<td>77.8</td>
<td>283</td>
<td>5</td>
<td>83</td>
<td>8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>12 kb DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highly repeated DNA</td>
<td>(3x10^{-4} &lt; C_{A}^{-1} &lt; 4 \times 10^{-3}, h. string.)</td>
<td>3.9</td>
<td>550</td>
<td>32±2</td>
<td>7±1</td>
<td>29±3</td>
<td>15±1</td>
<td>17±1</td>
</tr>
<tr>
<td>Unbound</td>
<td></td>
<td>~80</td>
<td>583</td>
<td>5</td>
<td>78</td>
<td>10</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

The mol. weights are initial values determined by alkaline sedimentation. After the 1st HAP (foldback DNA) fractionation, number average mol. weights ($M_n$) were determined by electron microscopy. $M_n$ was 5-3 kb for the 6 kb DNA and 10 kb for the 12 kb DNA. $M_n$ for forked molecules in the 12 kb highly repeated DNA fraction (which has now been through 2 renaturation and HAP fractionation steps) was 7-2 kb per strand. $A < C_{A}^{-1} < B$ means that the fraction was isolated on HAP at $C_{A}^{-1} A$ after being stripped of foldback DNA by a prior HAP fractionation at $C_{A}^{-1} B$. The stringency of the HAP fractionation criteria, defined in section (d) of Materials and Methods, is designated as “h. string.” for high stringency and “l. string.” for less stringent. The “unbound” fractions contain the DNA that eluted as single-stranded after the 2nd HAP fractionation.

HAP fractions were dialyzed and spread for electron microscopy in 50% formamide. To avoid possible confusion, no circular phage DNAs were added. “$n$” is the number of random molecules scored. Small fragments (<1 kb) are excluded from the scoring. The “forked” category includes all of the molecules shown in Fig. 2 except the linear and hairpin types. “Hairpins” are defined in Fig. 2; see also Wilson & Thomas, 1974, and Cech & Hearst, 1975. If a hairpin is seen on the tail of another structure (forked or mess), it is scored in one of the latter categories only. “Meses” cannot be unambiguously traced, usually because of multiple renaturation events and crossovers. If such a structure includes an obvious bimolecular fork (Fig. 2(f)), it is scored in the forked category. Circles and lariats (Thomas et al., 1970) make up 1 to 4% of the molecules in the renatured DNA fractions, and are not listed in the Table. The “±” numbers measure reproducibility of scoring, and do not include all sources of error. They give the maximum difference found when two grids of the same DNA preparation were scored, or when one grid was scored on 2 different days. On one grid (*) the discrimination between double- and single-stranded regions was not good enough to positively differentiate ds/ss from ss molecules, so the total number of linear structures is listed.

TABLE 3

Types of forked structures in renatured highly repeated DNA

<table>
<thead>
<tr>
<th>Highly repeated DNA fraction</th>
<th>$n$</th>
<th>Single</th>
<th>% of forked structures</th>
<th>Double</th>
<th>Multiple</th>
<th>Complex</th>
<th>Branch</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 kb unrecycled</td>
<td>108</td>
<td>28±2</td>
<td>24±9</td>
<td>13±4</td>
<td>25±3</td>
<td>9±1</td>
<td></td>
</tr>
<tr>
<td>12 kb unrecycled</td>
<td>175</td>
<td>20±4</td>
<td>25±3</td>
<td>9±3</td>
<td>39±1</td>
<td>7±1</td>
<td></td>
</tr>
</tbody>
</table>

This Table subdivides the forked category of Table 2 for 2 of the DNA fractions. “$n$” is the number of forked structures scored. The types of forked structures are defined in Fig. 2. “Branch” molecules look like the diagram in Fig. 1(viii). A sample of 35 multiple-forked structures in the 12 kb DNA contained an average of 2-3 regions of renaturation per structure; 84% of these were double forks, 16% were single forks.
Fig. 4. Multiple-forked and complex structures seen in the 12 kb highly repeated DNA fraction (Tables 2 and 3). Traceable molecules containing more than one renatured region were photographed and measured. The first 17 such molecules that were encountered are interpreted here, with crossovers eliminated and the lengths of single-stranded tails averaged. These structures are the source of some of the data in Fig. 5(b) and (d). All would be scored as "multiple forks" except (k) and (o), which are "complex" structures. Structures (c), (e), (f), (h), (j), (l) and (n) contain inverted repeat sequences in the form of looped and unlooped hairpins. Structures (o), (p) and (q) contain inverted repeats that may have been distant from each other in the native DNA. Although these 17 molecules are a small sample (about $10^{-10}$ µg of DNA), they illustrate the large variety of renatured structures encountered in many hours of surveying the grids.
The finding of intramolecular renaturation products (hairpins) in the highly repeated DNA fraction was initially surprising because foldback DNA had already been removed in the first HAP chromatography. As shown in Cech & Hearst (1975), the hairpins in the highly repeated fraction contain a lower average number of base-pairs than those isolated in the foldback fraction (0.65 kb compared to 1.0 kb). Furthermore, the frequency of hairpins observed in the highly repeated fraction can be explained by HAP being 90% rather than 100% efficient in fractionating this class of molecules.

(ii) **Length measurements**

In order to determine the length distribution of the highly repeated elements of the sequence interspersion pattern, randomly encountered double-forked and multiple-forked molecules were photographed, traced, and measured. The double-stranded regions of these structures are accurately defined by the branch points, making them easy to measure. Furthermore, these structures are the most informative because the ends of the base-paired regions are determined by an end of sequence homology (i.e. insufficient sequence homology for stability under the spreading conditions used). This is not the case for linear ds/ss and single-forked molecules, in which double-stranded regions end at the point where one of the strands has been broken.

The lengths of base-paired regions in double-forked and multiple-forked molecules, given in the histograms of Figure 5(a) and (b), have similar distributions. The formation of one type of structure or the other may depend only on random processes—where the sequences happened to be sheared, and whether renaturation continued long enough to form multimolecular products.

The same sequences that formed double-forked renaturation products could have formed single-forked or linear ds/ss products if the DNA had been sheared at a different random place (see Fig. 1(a)). The expected proportion of these three products is a function of the length of the repeated sequence and the length of the whole strand (Manning et al., 1975). For the 0.8 kb renatured DNA lengths seen here (at $M_n = 7.2$ kb), the theory of Manning et al. (1975) predicts a ratio of 64 double-forked:34 single-forked:2 linear molecules. The observed ratio (from the data of Table 3, including the forked regions of multiple-forked molecules) is 65 double-forked:35 single-forked molecules. Therefore, the single-forked molecules probably result from the same size class of repeated DNA sequences that produced the double-forked

† The ending of sequence homology at a fork is only an indication that 1 of the 2 strands changes sequence at that point. For instance, a sequence element A of length 0.1 kb could be repeated within a single renatured region of length 0.4 kb. Let $A'$ be the complementary sequence to A, and X and Y any other sequences. Measurement of the renatured portions of the renaturation products

$$
\begin{array}{c}
XA & XX \\
AAAA & AAAAA & A'AAAA' & YY & YY \\
A & X & A & X & Y & Y & A' & A'Y
\end{array}
$$

would suggest a native DNA length of 0.4 kb for the repeated stretches, while the actual average lengths would be longer. Therefore, the lengths of the double-stranded regions provide a **minimum** estimate for the lengths of the repeated sequences in the native DNA. Conversely, lengths of single-stranded regions joining two double-forked regions provide a **maximum** estimate of the length of the non-homologous DNA.
Fig. 5. Electron microscopic measurements of hydroxyapatite-fractionated DNA, spread in 45% formamide with circular fd and PM2 DNAs. (a) Length “d” of base-paired regions in double-forked molecules of the 12 kb highly repeated DNA fraction scored in Tables 2 and 3. The number-average molecular weight of the base-paired regions, \( \bar{d}_n \), is 0.80 kb. \( \bar{M}_n \), the number-average single-strand molecular weight, is 6.9 kb (per strand) for these structures. This is less than the initial 12 kb molecular weight (determined by alkaline sedimentation) partly because it is an number-average value, and partly because of single-strand breaks that occur during the 2 renaturation and HAP chromatography steps. (b) Length of base-paired regions in multiple-forked molecules on the same grid as (a). Only regions terminated on each end by 2 single strands are included. The dashed line refers to measurements of renatured regions in the type of multiple-forked structure shown in Fig. 2(e), and the solid line to those of all multiple-forked structures (Fig. 2(d) and (e)). \( \bar{d}_n = 0.66 \) kb; \( \bar{M}_n = 8.0 \) kb per strand. (c) Length of base-paired regions in presumed double-forked and multiple-forked structures seen in the 12 kb unbound fraction (Table 2), the DNA that chromatographed as single-strand on HAP, \( \bar{d}_n = 0.33 \) kb; \( \bar{M}_n = 7.0 \) kb. (d) Length “s” of single-stranded DNA joining 2 base-paired regions in the multiple-forked molecules of part (b). The dashed line again refers to the type of structure shown in Fig. 2(e). \( \bar{d}_n = 1.8 \) kb; \( \bar{M}_n \) given in (b).
molecules. About 18% \([7\%/(32 + 7\%)]\) of the bimolecular structures are linear ds/ss molecules, a significantly higher fraction than the predicted 2%. Therefore, about 16% of the renaturation products probably result from highly repeated sequences occurring in longer stretches (\(> 8\) kb).

As given in Table 2, the frequency of forked molecules in the DNA unbound by HAP was about seven times less than in the highly repeated DNA fractions. The presumed base-paired regions in these molecules were measured (Fig. 5(e)). They are mostly small, an observation consistent with their being either random crossovers (not base-paired) or short base-paired regions that might be inefficiently isolated on HAP. Even if all of the double-forked molecules in the unbound fraction were really random crossovers, and there were an equal number of such crossovers in the highly repeated DNA distributions (Fig. 5(a) and (b)), the average base-paired lengths would be only slightly underestimated.

The distance between adjacent renatured regions of multiple-forked molecules was measured. As shown in Figure 5(d), these single-stranded regions are extremely heterogeneous in length. The number average length is 1.8 kb, but the distribution ranges from small gaps of \(\sim 0\text{-}1\) kb to separations of 6 to 8 kb. The distribution's maximum is limited by the fragment length, which averages 8 kb per strand.

In order to convert the number fractions of Table 2 to approximate weight fractions, measurements were made of all single- and double-stranded regions of about 200 double-forked, multiple-forked, hairpin, and ss molecules in the 12 kb highly repeated DNA. From these measurements, and excluding the mess category, the 12 kb highly repeated DNA fraction was estimated to contain 69% by weight of bimolecular products (59% forked plus 9.5% ds/ss molecules), 19% single-strands, and 12% hairpins. An average of 20% by weight of the highly repeated DNA fraction was base-paired and 80% was single-stranded. Considering the small number of molecules measured and the exclusion of the mess category from the analysis, this value is in good agreement with the \(30\pm10\%\) double-strandedness determined for the highly repeated DNA fraction by \(S_1\)-nuclease resistance, CsCl buoyant density, and hyperchromicity (Cech et al., 1973).

(d) Renaturation kinetics of mouse DNA

Kinetic complexities were determined by the technique of Britten & Davidson (e.g., Davidson et al., 1973): a tracer quantity of \(^3\text{H}\)-labeled DNA is renatured with a large excess of unlabeled, sonicated total DNA (see Materials and Methods, section (e)). \(C_{ot}\) is determined by the concentration of the unlabeled “driver” DNA, because it is in excess. Monitoring the renaturation of the driver DNA (by optical density) serves as an internal standard for renaturation of the tracer DNA (monitored by scintillation counting).

In the experiments described here, the unlabeled driver DNA was isolated from Balb/c mice and the \(^3\text{H}\)DNA from SVT2 tissue culture cells. As shown in Table 4, the renaturation of these DNAs usually differed by one per cent or less over a range of more than six orders of magnitude of \(C_{ot}\). Therefore, the \(^3\text{H}\)-labeled SVT2 DNA is interchangeable with Balb/c mouse DNA in a renaturation kinetics experiment. Britten & Rake (1969) came to the same conclusion for L-cell and mouse DNAs.

Figure 6 shows the renaturation profile of the Balb/c mouse DNA (0-4 kb), and includes all data points from the experiments done subsequently to the one described above. This \(C_{ot}\) curve is in substantial agreement with other mouse DNA \(C_{ot}\) curves.
Table 4

Renaturation of \(^3\)H-labeled SVT2 DNA driven by excess Balb/c mouse DNA

<table>
<thead>
<tr>
<th>(C_{ot})</th>
<th>(h)</th>
<th>Fraction bound</th>
<th>(\Delta)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Balb/c</td>
<td>SVT2</td>
</tr>
<tr>
<td>7.3 \times 10^{-3}</td>
<td>0-1</td>
<td>0-100</td>
<td>0-095</td>
</tr>
<tr>
<td>6.7 \times 10^{-2}</td>
<td>1-2</td>
<td>0-143</td>
<td>0-136</td>
</tr>
<tr>
<td>7.7 \times 10^{-2}</td>
<td>0-1</td>
<td>0-184</td>
<td>0-179</td>
</tr>
<tr>
<td>2.3 \times 10^{-1}</td>
<td>0-4</td>
<td>0-233</td>
<td>---</td>
</tr>
<tr>
<td>1-2</td>
<td>2-1</td>
<td>0-265</td>
<td>0-273</td>
</tr>
<tr>
<td>3-7</td>
<td>6-5</td>
<td>0-268</td>
<td>0-275</td>
</tr>
<tr>
<td>1-1 \times 10</td>
<td>1-0</td>
<td>0-321</td>
<td>0-332</td>
</tr>
<tr>
<td>1-5 \times 10</td>
<td>27</td>
<td>0-319</td>
<td>0-319</td>
</tr>
<tr>
<td>4-3 \times 10</td>
<td>76</td>
<td>0-315</td>
<td>0-299</td>
</tr>
<tr>
<td>1-1 \times 10^2</td>
<td>10</td>
<td>0-366</td>
<td>0-376</td>
</tr>
<tr>
<td>3-5 \times 10^2</td>
<td>31</td>
<td>0-441</td>
<td>0-451</td>
</tr>
<tr>
<td>9-9 \times 10^2</td>
<td>10</td>
<td>0-543</td>
<td>0-550</td>
</tr>
<tr>
<td>1-1 \times 10^3</td>
<td>103</td>
<td>0-518</td>
<td>0-536</td>
</tr>
<tr>
<td>3-0 \times 10^2</td>
<td>31</td>
<td>0-716</td>
<td>0-712</td>
</tr>
<tr>
<td>1-0 \times 10^3</td>
<td>103</td>
<td>0-801</td>
<td>0-824</td>
</tr>
</tbody>
</table>

Balb/c DNA and \(^3\)H-labeled SVT2 DNA were mixed in a ratio of 1500:1 and sonicated to a single-strand size of 0.32 kb. Portions of the mixture were alkali-denatured, renatured (0.12 m-phosphate buffer, 60°C) at different times and concentrations, and chromatographed on HAP. Each sample contained 0.2 mg unlabeled Balb/c DNA and 0.13 µg (5300 cts/min) \(^3\)H-labeled SVT2 DNA. Unlabeled DNA concentrations varied from 0-005 to 1-0 mg/ml.

† Renatured in 0.90 m-phosphate at 70°C, therefore \(C_{ot}\) was multiplied by 8-06 to give the indicated equiv. \(C_{ot}\).

‡ The renaturation time in hours = \(t/3600\).

§ The fraction of the recovered optical density units that bound to HAP, indicating the fraction of the unlabeled Balb/c mouse DNA renatured at each \(C_{ot}\).

‖ The fraction of the recovered radioactivity that bound to HAP, indicating the fraction of the \(^3\)H-labeled SVT2 mouse DNA renatured at each \(C_{ot}\).

\(\Delta = [\text{fraction bound (SVT2)}] - [\text{fraction bound (Balb/c)}]\).

This initial \(C_{ot}\) curve shows evidence of DNA degradation at long time points, and the average molecular weight of the sample incubated in 0-12 m-phosphate for 103 h was reduced to 0.24 kb. The addition of EDTA to samples in subsequent renaturation experiments prevented the degradation.

at similar fragment size and renaturation conditions (Britten & Kohne, 1966; Laird, 1971; Straus & Birnboim, 1974). The renaturation of Escherichia coli DNA, used as the standard for calculating kinetic complexities, is shown also in Figure 6; the \(C_{ot}\) of 4-0 is in close agreement with previous values (Britten & Kohne, 1966; Davidson et al., 1973; Bonner et al., 1973).

The renaturation of mouse DNA is fit with five kinetic components, summarized in Table 5. The magnitude of the most rapid transition (the foldback DNA) is known from previous work (Cech et al., 1973; Wilson & Thomas, 1974). The quantity of light satellite DNA was determined by CsCl buoyant density centrifugation, and its renaturation rate is known (Southern, 1975; Waring & Britten, 1966; Hutton & Wetmur, 1973).
The region of the $C_{st}$ curve between mouse satellite and single-copy DNA can be adequately fit only if a minimum of two components are used. The faster of these two components is called "highly repeated main-band DNA" because, as will be shown, it includes the sequences we have been characterizing. At the standard renaturation criteria, it comprises 8 to 16% of the mouse genome and has an uncorrected repetition frequency of $(1.3$ to $2.7) \times 10^4$ copies. As described in footnote (e) of Table 5, we have sufficient information about the base composition and sequence divergence of these sequences to make approximate corrections to the complexity and repetition frequency. The corrected values are $(3.1$ to $6.5) \times 10^4$ copies of a sequence with a 3000 to 12,000 base-pair complexity. (This region of the $C_{st}$ curve could of course be fit with more than one component, each of which would then have a lower complexity.) The "moderately repeated" class (~11% of the mouse genome with an average complexity of $(6$ to $12) \times 10^5$ base-pairs) has a complexity similar to that of the major repetition class in the Xenopus (Davidson et al., 1973) and rat (Rice, 1971; Holmes & Bonner, 1974) genomes, which are about the size of the mouse genome.
The 5 kinetic classes described in the text were used to fit the mouse DNA renaturation data of Fig. 6. The data was plotted and fit with a Calcomp 565 plotter and Digital PDP8e computer. Fitting was done by trial and error, after which the root-mean-square y deviations (S_y) from the fitted curves were calculated.

* The percentages are valid only for the fragment size (0-4 kb) and renaturation conditions (0-12 M-phosphate buffer, 60°C, or equivalent) used in the experiments of Fig. 6. Because 6% of the DNA failed to renature, the percentages add to 94 instead of 100%. “Fig. 6” values give the best fit (S_y = 1.8 in units of %), and a “range” of acceptable values is listed (S_y <= 2.1).

b The C_{otj} observed for each class of DNA when diluted by the remainder of the genome. Again, both the best fit to the data and a range of acceptable values are given.

c The C_{otj} estimated for the pure kinetic component, calculated as:

\[
C_{otj} = \frac{C_{otj} \text{(obs.)}}{C_{otj} \text{(pure)}} \times C_{otj} \text{(pure)}.
\]

C_{otj} (E. coli) = 4.0 from Fig. 6 and complexity (E. coli) = 4.2 x 10^6 base-pairs (Cairns, 1963).

This is a maximum value. Assuming an 8% highly repeated DNA class with C_{otj} (obs.) = 0-1, numbers which fit the data almost as well as those used, the calculated complexity is 8 x 10^6.

Also, renaturation rate and therefore kinetic complexity are a function of base composition (Wetmur & Davidson, 1968) and of sequence divergence (Bonner et al., 1973). Both factors cause the complexity to be overestimated. Given the base composition (41% G + C) and T_m (9°C) of highly repeated mouse main-band DNA (Cech et al., 1973), its corrected complexity can be estimated as [(0.8 - 2.9) x 10^4] x 0.83 x 0.55 = (0.3 - 1.2) x 10^4. Furthermore, this kinetic class could be the average of many subclasses, each of which would then have a lower complexity.

f Taking the mouse haploid genome size as 2.7 x 10^9 base-pairs (Laird, 1971), repetition frequency is calculated as 2.7 x 10^{9} x (% of genome/100)/complexity.

* In the case of an 8% highly repeated DNA class with a complexity of 8000 base-pairs (footnote e), the calculated repetition frequency is 2.7 x 10^4, or 6.5 x 10^4 when corrected for base composition and sequence divergence.

(c) Kinetic complexity of the highly repeated DNA

To determine the spectrum of repetition frequencies in the 6 to 12 kb highly repeated mouse DNA isolated on HAP, a tracer quantity of this DNA was denatured and renatured with excess sonicated Balb/c DNA. The tracer DNA was in one case the same sample as used for the electron microscopy, and in two other cases was an analogous preparation. Therefore, the 12 kb tracer was expected to include 12% by weight of hairpin-containing strands (not all of which is expected to renature after a second denaturation), 19% single-stranded molecules of unknown repetition frequency, and the remaining 69% molecules with one or more regions of highly repeated DNA.
The 6 kb tracer was expected to contain 5% hairpins, 17% ss molecules, and 79% rapidly renaturing molecules by weight.

Figure 7 shows the renaturation of 6 and 12 kb highly repeated mouse main-band DNA fractions, driven by excess 0-4 kb total DNA. Table 6(A) lists the kinetic components that best fit the data of Figure 7. Because of the small number of data points, the numbers are not as exact as those of Table 5. Nevertheless, the data is accurate enough to show that the bulk of the tracer DNAs renatured in one or two very rapid second-order transitions, with $C_{ot}^{+}$ values of $\sim 6 \times 10^{-3}$ and $\sim 0.10$. This renaturation rate corresponds to that of the highly repeated main-band components ($C_{ot}^{+} = 0.10$ to 0.20) seen in the sonicated driver DNA. Because renaturation rate is a function

![Image of Figure 7](image-url)

**Fig. 7.** The renaturation of a tracer quantity of $^3$H-labeled highly repeated mouse main-band DNA, driven by an excess of unlabeled, sonicated total mouse DNA.

- (•••••) Renaturation curve for the driver DNA, from Fig. 6.
- (×) Data points for driver DNA renaturation in the 12 kb unrecycled tracer experiment described below. The individual points for the other 2 sets of data are shown in Fig. 9.
- REFERRED TO, Range of the $C_{ot}^{+}$ of the highly repeated component of 0-4 kb mouse DNA, from Fig. 6.

Renaturation of the $^3$H-labeled highly repeated DNA, each set of points fit with 2 or 3 second-order components as given in Table 6(A):

- — ● — ● —, 12 kb unrecycled tracer (foldback DNA removed at $C_{ot} = 1 \times 10^{-3}$; 2nd HAP isolation at $C_{ot} = 8 \times 10^{-3}$, low stringency elutions, yield = 6-7% of main-band).
- — ■ — ■ —, 12 kb recycled tracer (above sample recycled at $C_{ot} = 5 \times 10^{-3}$, high stringency, yield after recycling = 0-6% of main-band).
- — ▲ — ▲ —, 6 kb recycled tracer (foldbank DNA removed at $C_{ot} = 7 \times 10^{-4}$; 2nd HAP isolation at $C_{ot} = 6 \times 10^{-2}$, low stringency, 9-9% yield; recycled at $C_{ot} = 6 \times 10^{-3}$, high stringency, yield after recycling = 1-5% of main-band). This sample was examined in the electron microscope (Table 2).

Each point was determined with 1-1 x 10$^4$, 2-2 x 10$^4$ or 2-0 x 10$^5$ cts/min of tracer DNA in the 3 experiments, respectively. The DNA specific activity was 2-6 x 10$^6$ cts/min per µg for the 12 kb DNA, 4-0 x 10$^4$ for the 6 kb DNA. The unlabeled driver DNA was present in a 3-7 x 10$^3$, 2-2 x 10$^4$, or 2-5 x 10$^5$ times excess in the 3 experiments, respectively.

- (◊) 12 kb unrecycled tracer, same DNA as (●), but incubated with carrier *M. lysodeikticus* DNA instead of with unlabeled mouse DNA driver. $C_{ot}$ for these points is calculated as if the mouse DNA driver had been present. As expected for such a rapidly renaturing DNA, tracer DNA self-reaction does occur at high $C_{ot}$ values. Because the self-reaction points are displaced by several decades of $C_{ot}$ from the driven renaturation curve, this DNA would renature with driver DNA long before significant self-reaction could occur.
HIGHLY REPEATED MOUSE MAIN-BAND DNA

TABLE 6

Kinetic components of highly repeated mouse main-band DNA, before and after shearing

<table>
<thead>
<tr>
<th>Highly repeated tracer DNA</th>
<th>FB %</th>
<th>SAT C0t1 %</th>
<th>Region of C0t4 curve</th>
<th>HR C0t4 %</th>
<th>MR C0t4 %</th>
<th>NR C0t4 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Unsheared (Fig. 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 kb recycled (▲)</td>
<td>54</td>
<td>25</td>
<td>0.004</td>
<td>62</td>
<td>0.08</td>
<td>0 −</td>
</tr>
<tr>
<td>12 kb recycled (■)</td>
<td>41</td>
<td>0 −</td>
<td>0.008</td>
<td>43</td>
<td>0.11</td>
<td>11</td>
</tr>
<tr>
<td>12 kb unrecy. (●)</td>
<td>47</td>
<td>0 −</td>
<td>0.008</td>
<td>62</td>
<td>0.13</td>
<td>15</td>
</tr>
<tr>
<td>B. Sheared (Fig. 9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 kb recy., shear to 0.9 kb (△)</td>
<td>49</td>
<td>26</td>
<td>0.008</td>
<td>54</td>
<td>0.20</td>
<td>0 −</td>
</tr>
<tr>
<td>12 kb recy., shear to 0.5 kb (□)</td>
<td>0.7</td>
<td>39</td>
<td>0.021</td>
<td>35</td>
<td>0.15</td>
<td>0 −</td>
</tr>
</tbody>
</table>

These kinetic components describe the graphs in Figs 7 and 9. They do not fit the data uniquely, but do give a good indication of the repetition frequency components that are involved. SAT, HR, MR and NR refer to regions of the C0t curve defined in Fig. 6. Here the SAT components are mostly main-band DNA with a satellite-like C0t1, rather than actual satellite sequences (see Fig. 8). The foldback (FB) values were determined by renaturation of the tracer DNA to C0t1 ≈ 10^-6 in the absence of driver DNA. The 12 kb recycled tracer values in parentheses were not experimentally determined, but were assigned the values measured for the same tracer DNA before it was recycled.

of molecular weight, the data for sheared tracer DNA (section (f)) is necessary to establish this correspondence.

In the experiment with the recycled 6 kb highly repeated tracer, there was no detectable component with a C0t1 > 0.10. In the two experiments with 12 kb DNA, however, 11 to 29% of the DNA renatured as slower components. The non-repeated DNA in these preparations must have been freed from attachment to foldback DNA or repeated DNA by single-strand breaks. The electron microscopic observation of 19% single strands in an unrecycled rapidly renaturing fraction at this molecular weight (Table 2, 12 kb DNA) is consistent with the amount of non-repeated DNA seen in the corresponding C0t4 curve. The only C0t4 curve that showed detectable amounts of moderately repeated DNA was the preparation that had not been recycled. Thus, the rigorous kinetic recycling step does seem to serve the function of purifying highly repeated from moderately repeated sequences.

Because a significant fraction (about one-third) of the recycled highly repeated DNA renatured with a C0t4 of (4 to 8) × 10^-3, similar to that of mouse satellite DNA, it was necessary to see if this component could be due to satellite DNA contaminating the original main-band DNA preparation. A 0-2 to 0-5% contamination would explain the renaturation kinetics, and such a small amount of satellite would not have been detected in the analytical CsCl centrifugation of the unFractionated, native main-band DNA. To test for satellite DNA in the tracer, portions of the low-C0t4 renatured fractions from two of the Figure 7 C0t4 curves were banded in preparative CsCl gradients. The unlabeled driver DNA served as an internal marker for the positions of renatured satellite and highly repeated main-band DNAs. As shown in Figure 8, the tracer DNAs band mostly with the main-band highly repeated DNA. (The
Fig. 8. A test for satellite DNA in the 12 kb highly repeated mouse main-band DNA tracer. (a) Unrecycled 12 kb tracer, net $^3$H cts/min (---O---O---); sonicated driver DNA, $A_{260nm}$ (---●---●---). Portions of 3 renatured fractions from the unrecycled 12 kb tracer $C_{ot}$ curve of Fig. 7 were pooled. EDTA, Sarkosyl, and solid CsCl were added, and the DNA was allowed to aggregate at 60°C to an equiv. $C_{ot}$ of 50. After adjustment to a vol. of 6 ml and a refractive index of 1.3992, the solution contained 1 mM-EDTA/0.04% Sarkosyl/0.30 M-phosphate buffer. Centrifugation was at 35,000 revs/min in the Spinco fixed-angle 65 rotor at room temperature for 36 h. The polycumberland tube was then punctured and fractions collected through a 23 G needle. The optical density of each 0.12 ml fraction was determined in a microcuvette, after which the fraction was dried on a Whatman GF/A glass filter. The filters were counted on Omnifluor (New England
experiment was repeated after shearing the tracer DNA to 0.5 kb, with the same result.) The small amount of tracer DNA that bands with the satellite is insufficient to account for the fastest-renaturing component of the tracer DNA, so it must be primarily of main-band origin. This component makes up only about 0.5% of the main-band DNA (1/3 of the 1.5% recycled highly repeated DNA), and therefore is not detected in the unrecycled C_0t curve of Figure 7 or the total DNA C_0t curve of Figure 6.

(f) Kinetic complexity of the unrenatured sequences attached to highly repeated DNA

In the experiments of Figure 7, where long DNA strands containing highly repeated sequences are renatured with total mouse DNA, the renaturation kinetics are determined by the most highly repeated sequences on each strand. In order to determine the complexity of the DNA that is adjacent to the highly repeated stretches (the single-stranded "tails" seen by electron microscopy) the C_0t analysis was repeated on the long tracer strands after shearing. The 6 kb recycled tracer of Figure 7 was sheared by rapid decompression in a bomb (850 lb/in²) to a molecular weight of 0.9 kb, and the 12 kb recycled tracer was sonicated to a molecular weight of 0.5 kb. (The sheared molecular weights were not determined for the tracer DNA itself, but in parallel experiments described in Table 1.) Random shearing is expected† to free 76±9% of the DNA in the forked structures of the 12 kb tracer DNA from attachment to renatured regions. For the 6 kb tracer DNA, 64±6% of the DNA in forked structures should be sheared from renatured regions. Therefore, if the unrenatured tails are largely non-repetitive DNA, the C_0t curve should change dramatically after the tracer is sheared.

Figure 9(a) shows the renaturation profile of the 6 kb tracer before and after shearing, and Figure 9(b) shows a similar experiment with the 12 kb tracer. The

† Let $F_{ss}$ be the fraction of the renatured DNA that would chromatograph as single-stranded on HAP after shearing, i.e. would no longer be attached to renatured regions. Then, for double-forked molecules, $F_{ss} = 1 - [(d + L - 2 S)/L_1]$, where $d$ is the length of the renatured segment, $L$ is the fragment size after shearing, $S$ is the minimum stable duplex length in HAP chromatography, and $L_1$ is the initial strand length before shearing. (This equation comes from the analysis done in the Appendix of Cech et al., 1973. It is true for $L > d$ as well as $L < d$.) For double-forked molecules in the 12 kb tracer DNA, using the electron microscopic number average values of $d$ and $L_1$, using $S = 0.06$ kb (Wilson & Thomas, 1973), and using $L = 0.5$ kb from Table 1 (although this is a weight-average value), $F_{ss} \approx 1 - [(0.7 + 0.5 - 0.12)/7] = 0.85$. Taking into account the estimated number of forked structures with more than one renatured region will increase $d$ and thereby decrease $F_{ss}$. For the whole forked category of the 12 kb tracer, $F_{ss}$ is calculated to be 0.76±0.09. The uncertainty arises from different estimates of the average number of renatured regions in more complex structures like those shown in Fig. 2(f) and (g).
Fig. 9. Renaturation of highly repeated mouse main-band DNA, before and after shearing.

(a) (▲) 6 kb recycled tracer DNA renaturation from Fig. 7.
(△) Renaturation of the same tracer after shearing to 0·9 kb.
(×) Data points for driver DNA renaturation, unsheared expt.
(+ ) Data points for driver DNA renaturation, sheared expt.
(······) Renaturation curve for the driver DNA, from Fig. 6.

(b) (■) 12 kb recycled tracer DNA renaturation from Fig. 7.
(□) Renaturation of same tracer after shearing to 0·5 kb.
(×), (+) driver DNA renaturation as in (a).

Each set of tracer DNA points is fitted with 2 or 3 second-order components as given in Table 6. The amount of 3H-tracer DNA and unlabeled driver DNA used to determine each point was the same for the sheared as for the corresponding unsheared point, and therefore is given in the legend to Fig. 7.

The kinetic components that best fit these curves are given in Table 6 B. Because of the small number of data points, other fits to the data are possible, and the listed $C_{oti}$ values must be considered only approximate. There are three important observations that can be made from the data of Figure 9.

1) Most of the interspersed, unrenatured sequences in the highly repeated DNA fraction are also highly repeated sequences, although there are some interspersed
complex sequences. The most striking fact about the Figure 9 $C_{\text{ot}}$ curves is how little they change upon shearing the tracer DNA. In both experiments, only about 10% non-repetitive DNA is detached from rapidly renaturing sequences upon shearing. (The complex sequences have $100 < C_{\text{ot}}^{-1} < 1000$ in Fig. 9(a) and $C_{\text{ot}}^{-1} \approx 300$ in (b); this component will be designated as "non-repetitive" DNA even though it seems to be slightly repetitious.) This 10% is more than can be accounted for by non-repetitive sequences sheared from foldback DNA. (The sequences near foldback DNA “hairpins” in *Xenopus*, *Drosophila*, and human DNA are largely non-repetitive sequences (Davidson et al., 1973; Schmid et al., 1975; Schmid & Deininger, unpublished observations).) Thus, there appears to be some interspersion of highly repeated and single-copy DNA sequences as in model (b), Figure 1. Most of the tracer DNA, however, continues to renature rapidly after shearing. This is the result expected for a permuted tandem repeat model of sequence organization, diagrammed in Figure 1(d). The relative amount of the non-repetitive DNA sequence interspersion is calculated as

$$I_{\text{nr}} = \frac{(\Delta NR - \Delta FB)}{(F_{ss} \times f)}.$$  

$\Delta NR$ is the increase in the non-repetitive DNA component after shearing, $\Delta FB$ is the estimated amount of non-repetitive DNA detached from foldback DNA by shearing, and $f$ is the weight fraction of forked molecules in the tracer. For the 6 kb tracer,

$$I_{\text{nr}} = \frac{[(0.09 \pm 0.02) - (0.01 \pm 0.01)]}{[(0.64 \pm 0.06) \times (0.68 \pm 0.12)]} = 0.18 \pm 0.07.$$  

For the 12 kb tracer,

$$I_{\text{nr}} = \frac{[(0.10 \pm 0.02) - (0.02 \pm 0.01)]}{[(0.76 \pm 0.09) \times (0.59 \pm 0.10)]} = 0.18 \pm 0.07.$$  

At 12 kb there is an additional 11% non-repetitive DNA that was presumably broken from attachment to rapidly renaturing regions prior to the deliberate shearing; this gives a total of $0.29 \pm 0.07$ non-repetitive interspersion at the higher starting molecular weight.

Thus, analysis of both experiments leads to the same conclusion: $18 \pm 7\%$ of the interspersed sequences within a few kilobases of highly repeated sequences are non-repetitive (or of very low repetition frequency), while $82 \pm 7\%$ contain sequences with about the same repetition frequency as the sequences initially responsible for the whole strand binding to HAP. This $82\%$ does not have to be pure highly repeated sequences—it could still contain more complex sequences, as long as each 0.5 kb fragment contained enough highly repeated DNA to give the observed rapid renaturation.

(2) The repeated sequence elements in the high molecular weight tracer DNA belong to the highly repeated class ($\sim 14\%$ of the total DNA, $C_{\text{ot}}^{-1} = 0.1$ to 0.2). This conclusion could only be made tentatively in section (e), because of the unknown effect of molecular weight on the tracer’s renaturation rate. Now (especially in the Fig. 9(b) experiment), the sheared tracer has a molecular weight about equal to that of the driver DNA, and the $C_{\text{ot}}^{-1}$ values can be directly compared. The $C_{\text{ot}}^{-1}$ of the major repeated sequence class of the sheared tracer (Table 6B) is the same as the highly repeated class in total DNA (Table 5), within experimental error.
(3) The increase in $C_{ot}$ of the highly repeated DNA component upon shearing gives additional evidence for the presence of more than one stretch of repeated DNA per high molecular weight tracer strand. In the HAP-assayed renaturation of long tracer strands and short driver strands, the rate of the tracer-driver renaturation is faster than the driver-driver renaturation in proportion to the relative number of nucleation sites on the two sizes of strands (Wetmur, 1971; Davidson et al., 1973). In our case, assuming that 70 to 80% of the sequences on each tracer strand contain highly repeated DNA, the tracer should bind to HAP with a rate about $[0.75 \times L(\text{tracer})/L(\text{driver})]$ faster than that of the driver. The tracer renaturation curve should therefore be displaced to higher $C_{ot}$ values after shearing, and the amount of the shift should be about $[0.75 \times L(\text{unsheared tracer})/L(\text{sheared tracer})]$. For the experiment of Figure 7(a) this ratio is about $(0.75 \times 5/0.9) = 4$; for Figure 9(b), about $(0.75 \times 7/0.5) = 10$. In both cases, the observed ratio is about 3—the sheared curve is displaced by half a decade of $C_{ot}$. Qualitatively, the decreased $C_{ot}$ at high molecular weight means that there are more nucleation sites on the long strands, which is consistent with the permuted tandem repeat model. Quantitatively, there is a discrepancy between the expected and observed shifts in one of the two experiments (Fig. 9(b)). This is probably due to the unsheared tracer having a lower than expected molecular weight at the time of the renaturation experiment. Fortunately, even a large decrease in molecular weight does not affect the previous conclusions about interspersion. When a starting molecular weight of 3.7 kb is substituted in the equation for $F_{ss}$, the resulting value of $I_{sr}$ (the fraction of non-repetitive sequence interspersion) is 0.24 instead of 0.18±0.07.

4. Discussion

(a) A model for the organization of highly repeated mouse main-band sequences

The electron microscopy of renatured DNA shows that, in fragments with a number-average length of 7 to 8 kb, approximately 84% by number (~88% by weight) of highly repeated mouse main-band sequences occur in an interspersion pattern. That is, renatured regions alternate with non-homologous regions. Most interspersed highly repeated DNA regions are 0.2 to 1.0 kb in length, with a number-average length of 0.7 kb. On those molecules where two or more renatured regions can be identified, the average spacing between regions varies widely, averaging about 1.8 kb. The remaining 16% (12% by weight) of the repeated DNA sequences may be tandemly repeated rather than interspersed, or may be interspersed sequences with a longer average repeated sequence length (> 8 kb). These percentages refer only to the obvious bimolecular renaturation products—intramolecular renaturation products, single-stranded DNA, and circles were also seen in the highly repeated DNA fraction. The percentages assume that the 17% messes seen in this fraction arise largely from further renaturation of molecules with the same type of sequence arrangement as the traceable molecules.

In an earlier study (Cech et al., 1973) we determined the HAP yield and per cent double-strandedness of highly repeated mouse main-band DNA at three molecular weights. We concluded that segments of renatured DNA with an average length of 1.5±0.5 kb were interspersed with stretches of unrenatured DNA averaging 2.2±1.1
HIGHLY REPEATED MOUSE MAIN-BAND DNA

kb. These numbers are in excellent agreement with the electron microscopic averages of 1.1 kb [(0.88 x 0.7 kb for forked molecules) + (0.12 x ~4 kb for ds/ss)] and ~1.8 kb, for double- and single-stranded regions, respectively. Our earlier study suggested that the interspersed, unrenatured sequences were of higher complexity than the renatured sequences. The arguments were indirect, based on the high degree of single-strandedness of the renatured molecules even after they were aggregated to a high C_d (~50). Apparently the single-strandedness was largely due to excluded volume effects that inhibited further renaturation of the large aggregates. The driven renaturation experiments described here lead us to conclude that only 18±7% of the unrenatured DNA in the 6 kb repeated DNA fraction is made up of non-repetitive sequences. At a starting molecular weight of 12 kb there is an additional 11% non-repetitive DNA, presumably broken from attachment to rapidly renaturing regions prior to the deliberate shearing. In both cases, the bulk of the single-stranded tails seem to have a repetition frequency similar to that of the renatured regions.

This suggests a permuted tandem repeat model of sequence organization, where a highly repeated sequence R_1 is sometimes adjacent to R_2, but sometimes adjacent to R_3. If the interspersed non-repetitive DNA is also adjacent to the permuted tandem repeats, the simplest model of sequence organization consistent with our data is: R_1 R_2 S R_4 R_5 . . . R_7 R_3 S R_6 R_2 . . . . S indicates a single-copy or non-repetitive sequence. Each repeated sequence element (R_n) has an average length of roughly 0.7 kb. An alternate possibly—R_1 R_2 R_3 R_4 . . . S S R_5 S S . . . . , would give a larger increase in HAP yield with molecular weight than we have observed (Cech et al., 1973). Our data does not preclude smaller stretches (<0.5 kb) of non-repetitive DNA in addition to the 18% predicted by the shearing experiments of section (f). Britten (1972), however, has degraded intermediate repetition classes of mouse DNA (including both the highly repeated and moderately repeated sequences as defined here) to lower molecular weights than we have used. He found a fine-scale interspersion of repeated and non-repeated sequences only in a low thermal stability (t_m < 68°C) fraction, a fraction excluded from our highly repeated DNA by 70°C HAP elutions. Although the sequences responsible for the HAP isolation of our highly repeated DNA fraction must be of high thermal stability, it is possible that some of the interspersed, unrenatured repeated sequences in the permuted tandem repeat arrangement are of low thermal stability (Rice, 1971).

We have several ways of estimating the amount of the mouse genome that contains highly repeated sequences (interspersed to some extent with non-repeated sequences). From the yield of highly repeated DNA at 12 kb molecular weight, a minimum of 4% of the mouse main-band contains such sequences. This is clearly an underestimate, because the C_d of isolation was less than the C_d of the highly repeated DNA. Approximately 12±4% of the sonicated driver DNA renatures in the same transition as the sheared highly repeated tracer, giving a maximum estimate for the amount of highly repeated DNA. Adding the amount of interspersed non-repetitive DNA that is no longer attached to highly repeated DNA at 0.5 kb molecular weight increases this estimate to 15±4%. The ~15% of the DNA containing highly repeated main-band sequences is probably distributed among most or all of the mouse chromosomes (Cech et al., 1973). Since only about 18% of the 6 kb highly repeated DNA fraction consists of interspersed non-repetitive DNA, the amount of the non-repetitive DNA of mouse that is so arranged is estimated as (18% x 15%)/58% = 5%.
more moderately repeated fraction of mouse DNA seems to be more widely interspersed with non-repetitive DNA (Britten, 1972), in an arrangement similar to that found in the frog, sea urchin, and human genomes (Davidson et al., 1973; Graham et al., 1974; Schmid & Deininger, unpublished observations).

We do not yet know the generality of the sequence arrangement proposed here for highly repeated mouse main-band sequences. Human homogeneous main-band DNA (Corneo et al., 1970) probably contains highly repeated sequences interspersed with other repeated sequences (Marx, Allen & Hearst, 1976). About 6% of Xenopus DNA is made up of highly repeated sequences (Davidson et al., 1973), and permuted tandem repetition has not been ruled out for their sequence organization.

(b) Biological function of the highly repeated sequences

Reasonable estimates for the reiteration frequency and complexity of the mouse main-band highly repeated DNA range as high as 64,000 repeats of a 3000 base-pair sequence (from the footnotes to Table 5). This repetition frequency, so much higher than that of the more moderately repeated class of mouse DNA (or the major repetition frequency class of Xenopus DNA), suggests that the two may have quite different origins and/or functions. Of course, the intermediate region of the mouse $C_\alpha t$ curve could be fit by more than two components, and may well represent a continuum of repetition frequency classes. In this case, it is the large range of repetition frequencies that is difficult to reconcile with a single model of sequence function.

Besides the low kinetic complexity, there are three preliminary indications that mouse main-band highly repeated sequences may have some characteristics of simple sequence (satellite) DNAs. (i) The permuted tandem repeat sequence organization of the highly repeated main-band DNA may be qualitatively similar to that of the mouse satellite. High molecular weight mouse satellite DNA shows anomalous renaturation kinetics (Hutton & Wetmur, 1973), and the renatured product has an increased buoyant density (Bond et al., 1967; Corneo et al., 1968; discussed by Flamm et al., 1969) and shows a series of single-stranded loops when viewed in the electron microscope (Cech, 1975). All of these properties might be explained by a permuted tandem repeat sequence organization. (ii) Both mouse satellite and highly repeated main-band DNA form renaturation products of high thermal stability. Rice (1971) has found that repeated mouse sequences of high thermal stability do not hybridize with the corresponding sequences of rat or hamster DNA. These sequences, unlike other repeated sequences which occur in all three rodents, have presumably arisen since the divergence of the species. (iii) Short runs of pyrimidine nucleotides are characteristic of rodent satellite DNAs (Southern, 1970; Walker, 1971). Approximately 1-3% of mouse L-cell DNA contains long pyrimidine tracts, 50 to 150 nucleotides (Straus & Birnboim, 1974). These are located in DNA with the same repetition as the highly repeated mouse main-band DNA characterized here.

Preliminary hybridization experiments have detected sequences complementary to the highly repeated main-band DNA in the whole cell RNA of SVT2 cells (Cech, 1975). In an RNA excess hybridization at an RNA $C_\alpha t$ of 820, 21% of the highly repeated DNA formed RNA:DNA hybrids, detected by HAP chromatography after RNase treatment. This value is corrected for the DNA self-reaction. The 21% is a minimum estimate, since saturation was not achieved. The $C_{\min}$ buoyant density of the hybrids after RNase treatment corresponded to an RNA/DNA ratio of 0·2/1·0. Such a ratio would be expected for the hybridization of permuted repeat RNA
transcripts to the permuted repeat DNA. More detailed studies of the transcription of these sequences are necessary to discover their biological function.

One of us (T. R. C.) is grateful to Harumi Kasamatsu for many hours spent teaching him electron microscopy techniques. We have appreciated a continual exchange of data and ideas with Carl Schmid. We thank Alan Levin for his help in computer programming. James Wang and Leroy Liu generously provided circular fd and PM2 DNAs. This work was supported in part by United States Public Health Service grant no. GM11180 and National Science Foundation grant no. GB38799 to one of us (J. E. H.). T. R. C. was supported by a National Science Foundation Graduate Fellowship.

REFERENCES


