Highly Repeated DNA Families in the Rat*

Frank R. Witney‡ and Anthony V. Furano
From the Section on Genomic Structure and Function, Laboratory of Biochemical Pharmacology, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20205

We have analyzed the repeated DNA fraction of the rat by characterizing approximately 500 repeat DNA-containing clones using hybridization to a variety of rodent nucleic acids. To facilitate this analysis we devised a method whereby the cloned DNA is transferred to nitrocellulose paper by blotting directly out of colonies of the bacterial clones. In addition to identifying repeated sequences of potential interest (e.g. those transcribed in a tissue-specific manner, or those that are highly conserved in non-rat genomes), we found that, in contrast to what is revealed by the reassocation of rat DNA (e.g. Pearson, W. R., Wu, J. R., and Bonner, J. (1978) Biochemistry 17, 51–59), the rat genome contains a number of different highly repeated (>50,000 copies) sequences. We distinguished the different highly repeated sequences both by their hybridization to different nucleic acids as well as by DNA sequence determination. The highly repeated sequences shared three characteristics that distinguished each of them from the 100,000-member rat satellite I family: (i) they were recovered less often in the cloned repeat DNA library than expected from their copy number in the rat genome; (ii) they reannealed abnormally slowly for their copy number even though they are not significantly divergent; and (iii) they are transcribed in one or more rat tissues. The implications of these findings for the organization of repeated sequences in the rat genome are discussed.

Families of repeated DNA sequences make up from 10 to 50% of all metazoan genomes (1–3). Only 1–2% of the repeated DNA corresponds to families of known function, such as the genes for ribosomal RNA, histones, etc., and although some mammalian repeated DNA sequences contain recognition sites for RNA polymerase III (4), the function of these sequences and of essentially all the other repeated DNA is unknown (5, 6).

Prior to the availability of recombinant DNA techniques, most of the information on repeated DNA was obtained by analyzing the reannealing of denatured DNA (7–11). From these analyses can be deduced such things as the fraction of the genome that is repeated DNA, the length of the archtypical repeated DNA sequences, the approximate number of different families of repeated sequences, and the average number of copies (members) in these families. Emerging from this body work is the currently held consensus (summarized in Ref. 12) that most of the copies of a given repeated DNA family are either clustered together in long tandem arrays (i.e. the satellite DNA families) or are interspersed among single copy (nonrepeated) DNA. The interspersion format has received considerable attention regarding possible functions for repeated DNA (2, 9–18).

In order to study the repeated DNA of the rat genome, we prepared and characterized a collection of repeated DNA-containing bacterial clones. Prior to cloning by the "G-C tailing" method (19) into the Pst I site of pBR322 (20), the repeated DNA sequences were purified from randomly sheared total rat DNA by two cycles of denaturation and reannealing to $C_{ot}^1$ 200 (8). At this point 90% or so of the rat DNA repeated 150 or more times will be in duplex, and contamination by reannealed single-copy sequences should be less than 0.5% of the mass of reannealed duplexes (11).

To select repeated DNA sequences of potential interest, we characterized the collection of cloned repeated DNA sequences with respect to such properties as their approximate copy number in the rat genome, their transcription in various rat cell types, and their presence or absence in other rodent genomes. We did this by hybridizing the DNA from about 500 bacterial clones to a variety of radioactive nucleic acids. To facilitate these experiments, we devised a method whereby the DNA is blotted directly out of colonies of the bacterial clones onto nitrocellulose paper to produce "clone blots."

Using this method we found that the rat genome contains many more highly repeated families (i.e. 50,000 or more copies) than the 1 or 2 that are revealed as a rapidly reannealing fraction by reassociation studies with rat DNA (Ref. 11; also see Discussion in Ref. 21). Much of this rapidly reannealing fraction can be accounted for by the 100,000-member rat satellite I family (22). The highly repeated families that we found share three characteristics that distinguish them from the satellite I family: (i) they are represented far fewer times in the cloned repeated DNA population than predicted from their copy number. (ii) The members of these families, although not very divergent, reanneal abnormally slowly for their copy number when total rat DNA is reannealed. (iii) They are transcribed into RNA in various rat cells. The implications of these findings for the organization of repeated DNA in the rat genome are discussed.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported in part by Public Health Service Grant 5SF2CA06886 awarded by the National Cancer Institute, Department of Health and Human Services. Present address, Bio-Rad Laboratories, Richmond, CA 94804.

The abbreviations used are: $C_{ot}$, the product of initial DNA concentration (in moles of nucleotide per liter) and time (in seconds) (see Ref. 8); kb, kilobase pairs; PB, Na phosphate buffer, pH 7.0.
Repeat DNA Clones—Rat repeated DNA clones were prepared by following in general the method used by Scheller et al. (23) for cloning PC12 nuclei treated DNA. Briefly, the method consists of the following: rat DNA that had been sheared by sonication to about 0.8 kb (range, 0.4-1.4 kb) was denatured by boiling and allowed to reanneal in 0.1 M PE at 60°C until a Cdt of 200 was attained. The duplexes were purified by chromatography on hydroxyapatite, dialyzed, and allowed to reanneal to Cdt 200 at 50°C. The duplexes were digested with HindIII and treated with 30 units of S1 nuclease (24) at 37°C for 1 h to remove single-stranded regions. dCMP residues were added to the 3’ termini (19) using deoxynucleotidyl terminal transferase (P-LBiochemicals), and the "T"-tailed" duplexes were annealed to the P-tailed Psi site of pBR322 (22). The labeled DNA molecules were used directly to transform Escherichia coli L1392 (a lac derivative of ED8645 (25)) as described by Morrison (26). Sixteen hundred randomly selected tetracycline-resistant, ampicillin-sensitive colonies were stored in 96-well microtiter dishes at -70°C.

DNA from each of the clones of 5 of the microtiter dishes, as well as from 12 clones of a sixth microtiter dish, was characterized by hybridization after the cloned DNA was transferred directly out of the bacterial colonies onto nitrocellulose by blotting (27). We call these blot clone blots. The method is described in general under "Results" and in detail in the Miniprint.

Preparation and Radiolabeling of DNA—Plasmid DNA was prepared either from 1-2 ml of overnight cultures (in L broth without antibiotic) of E. coli LE392 (a lac- derivative of ED8645 (25)) as described by Tata (35) and Fleischer and Kervina (36) but slightly modified by our inclusion of the final volume of the ethidium bromide and CsCl-containing clear lysate into the procedure of Ish-Horowicz and Burke (28), modified by our inclusion of the clear lysate method of Clewell and Helinski (30) modified to our procedure. The DNA was precipitated from 0.3 M Na acetate using 2 volumes of ethanol. The labeled DNA was radiolabeled in vitro with deoxynucleotide [alpha-32P]triphosphates (31). The radiolabeled DNA was precipitated and redissolved in 1 mM Na EDTA, 1 mg/ml of polyvinyl pyrrolidone, 100 μg/ml of bovine serum albumin, and 5 units of T4 polynucleotide kinase (New England Nuclear) and held at 37°C for 60 min. The reaction was adjusted to 50 mM NaEDTA, pH 7.0, 6% glycerol, and 2% (w/v) sodium dodecyl sulfate. After extraction with 1/10 volume of phenol, the separated aqueous phase was dried and labeled DNA was incorporated per μg of DNA.

Preparation and Radiolabeling of RNA—Nuclear RNA was prepared from nuclei of the PC12 neural rat cell line (33) or from rat organs. The PC12 nuclei were prepared by the method of Zieve and Fleischer (37, 38). Total rat DNA that had been randomly fragmented by sonication and treated with 0.8-kb fragments was denatured, sonicated salmon sperm DNA (see Ref. 40). The DNA was then incubated with the appropriate radioactive nucleic acids in the amounts, for the times, and at the temperatures indicated in the figure legends. The DNA-RNA hybridizations were purified by chromatography and dialyzed against several changes of a lysis solution and 100 μg/ml of sonicated salmon sperm DNA for 10 or 5 min, respectively, quick chilled in an ice bath, and added to the "prehybridized" filters. After hybridization, the filters were washed at room temperature twice with several hundred milliliters of 2 X SSC (1 X SSC = 0.15 M Na citrate, pH 7.4, 0.15 M NaCl) and twice with several hundred milliliters of 0.3 M Tris base (41), dried, and exposed to x-ray film (Kodak X-AR) at -70°C using DuPont, Cronex Lightning Plus intensifying screens unless stated otherwise.

Repeat DNA Clones—Rat repeated DNA clones were prepared by following in general the method used by Scheller et al. (23) for cloning PC12 nuclei treated DNA. Briefly, the method consists of the following: rat DNA that had been sheared by sonication to about 0.8 kb (range, 0.4-1.4 kb) was denatured by boiling and allowed to reanneal in 0.1 M PE at 60°C until a Cdt of 200 was attained. The duplexes were purified by chromatography on hydroxyapatite, dialyzed, and allowed to reanneal to Cdt 200 at 50°C. The duplexes were digested with HindIII and treated with 30 units of S1 nuclease (24) at 37°C for 1 h to remove single-stranded regions. dCMP residues were added to the 3’ termini (19) using deoxynucleotidyl terminal transferase (P-LBiochemicals), and the "T"-tailed" duplexes were annealed to the P-tailed Psi site of pBR322 (22). The labeled DNA molecules were used directly to transform Escherichia coli L1392 (a lac derivative of ED8645 (25)) as described by Morrison (26). Sixteen hundred randomly selected tetracycline-resistant, ampicillin-sensitive colonies were stored in 96-well microtiter dishes at -70°C.

DNA from each of the clones of 5 of the microtiter dishes, as well as from 12 clones of a sixth microtiter dish, was characterized by hybridization after the cloned DNA was transferred directly out of the bacterial colonies onto nitrocellulose by blotting (27). We call these blot clone blots. The method is described in general under "Results" and in detail in the Miniprint.

Preparation and Radiolabeling of RNA—Nuclear RNA was prepared from nuclei of the PC12 neural rat cell line (33) or from rat organs. The PC12 nuclei were prepared by the method of Zieve and Fleischer (37, 38). Total rat DNA that had been randomly fragmented by sonication and treated with 0.8-kb fragments was denatured, sonicated salmon sperm DNA (see Ref. 40). The DNA was then incubated with the appropriate radioactive nucleic acids in the amounts, for the times, and at the temperatures indicated in the figure legends. The DNA-RNA hybridizations were purified by chromatography and dialyzed against several changes of a lysis solution and 100 μg/ml of sonicated salmon sperm DNA for 10 or 5 min, respectively, quick chilled in an ice bath, and added to the "prehybridized" filters. After hybridization, the filters were washed at room temperature twice with several hundred milliliters of 2 X SSC (1 X SSC = 0.15 M Na citrate, pH 7.4, 0.15 M NaCl) and twice with several hundred milliliters of 0.3 M Tris base (41), dried, and exposed to x-ray film (Kodak X-AR) at -70°C using DuPont, Cronex Lightning Plus intensifying screens unless stated otherwise.

Preparation and Radiolabeling of Rat DNA of Different Repeat Classes—Total rat DNA that had been randomly fragmented by sonication and treated with 0.8-kb fragments was denatured, sonicated salmon sperm DNA (see Ref. 40). The DNA had then been treated with S1 nuclease. They were reincubated with 100 μg/ml of sonicated salmon sperm DNA for 10 or 5 min, respectively, quick chilled in an ice bath, and added to the "prehybridized" filters. After hybridization, the filters were washed at room temperature twice with several hundred milliliters of 2 X SSC (1 X SSC = 0.15 M Na citrate, pH 7.4, 0.15 M NaCl) and twice with several hundred milliliters of 0.3 M Tris base (41), dried, and exposed to x-ray film (Kodak X-AR) at -70°C using DuPont, Cronex Lightning Plus intensifying screens unless stated otherwise.

Preparation and Radiolabeling of Rat DNA of Different Repeat Classes—Total rat DNA that had been randomly fragmented by sonication and treated with 0.8-kb fragments was denatured, sonicated salmon sperm DNA (see Ref. 40). The DNA had then been treated with S1 nuclease. They were reincubated with 100 μg/ml of sonicated salmon sperm DNA for 10 or 5 min, respectively, quick chilled in an ice bath, and added to the "prehybridized" filters. After hybridization, the filters were washed at room temperature twice with several hundred milliliters of 2 X SSC (1 X SSC = 0.15 M Na citrate, pH 7.4, 0.15 M NaCl) and twice with several hundred milliliters of 0.3 M Tris base (41), dried, and exposed to x-ray film (Kodak X-AR) at -70°C using DuPont, Cronex Lightning Plus intensifying screens unless stated otherwise.

Preparation and Radiolabeling of Rat DNA of Different Repeat Classes—Total rat DNA that had been randomly fragmented by sonication and treated with 0.8-kb fragments was denatured, sonicated salmon sperm DNA (see Ref. 40). The DNA had then been treated with S1 nuclease. They were reincubated with 100 μg/ml of sonicated salmon sperm DNA for 10 or 5 min, respectively, quick chilled in an ice bath, and added to the "prehybridized" filters. After hybridization, the filters were washed at room temperature twice with several hundred milliliters of 2 X SSC (1 X SSC = 0.15 M Na citrate, pH 7.4, 0.15 M NaCl) and twice with several hundred milliliters of 0.3 M Tris base (41), dried, and exposed to x-ray film (Kodak X-AR) at -70°C using DuPont, Cronex Lightning Plus intensifying screens unless stated otherwise.

Preparation and Radiolabeling of Rat DNA of Different Repeat Classes—Total rat DNA that had been randomly fragmented by sonication and treated with 0.8-kb fragments was denatured, sonicated salmon sperm DNA (see Ref. 40). The DNA had then been treated with S1 nuclease. They were reincubated with 100 μg/ml of sonicated salmon sperm DNA for 10 or 5 min, respectively, quick chilled in an ice bath, and added to the "prehybridized" filters. After hybridization, the filters were washed at room temperature twice with several hundred milliliters of 2 X SSC (1 X SSC = 0.15 M Na citrate, pH 7.4, 0.15 M NaCl) and twice with several hundred milliliters of 0.3 M Tris base (41), dried, and exposed to x-ray film (Kodak X-AR) at -70°C using DuPont, Cronex Lightning Plus intensifying screens unless stated otherwise.

Preparation and Radiolabeling of Rat DNA of Different Repeat Classes—Total rat DNA that had been randomly fragmented by sonication and treated with 0.8-kb fragments was denatured, sonicated salmon sperm DNA (see Ref. 40). The DNA had then been treated with S1 nuclease. They were reincubated with 100 μg/ml of sonicated salmon sperm DNA for 10 or 5 min, respectively, quick chilled in an ice bath, and added to the "prehybridized" filters. After hybridization, the filters were washed at room temperature twice with several hundred milliliters of 2 X SSC (1 X SSC = 0.15 M Na citrate, pH 7.4, 0.15 M NaCl) and twice with several hundred milliliters of 0.3 M Tris base (41), dried, and exposed to x-ray film (Kodak X-AR) at -70°C using DuPont, Cronex Lightning Plus intensifying screens unless stated otherwise.
DNA as driver and are shown in Fig. 5 (inset). The reassociation reactions were carried out in solution at 60 °C in 0.12 M FB, and duplex formation at the indicated *Cot* was monitored by hydroxyapatite chromatography as described by Britten et al. (8). The radioactive "Cot-fractionated" DNAs were hybridized to clone blots as described in the legend to Fig. 5.

**DNA Sequences**—The DNA sequence of the Pst inserts of various clones was determined by the Maxam-Gilbert technique (42) or, after subcloning portions of them into M13mp9 (43), by the chain termination technique (44). The restriction endonuclease fragments sequenced by the Maxam-Gilbert technique were radiolabeled at their 3' ends with the Klenow fragment of DNA polymerase I (45) in a 25- or 50-μl reaction containing 10 mM Tris-Cl, pH 7.5, 7 mM MgCl₂, 50 μM dNTP, 100 μCi of the appropriate [α-32P]dNTP (1500-3000 CPM/nmol), 10 mM β-mercaptoethanol, and 1-2 units of the Klenow fragment (New England Biolabs). After a 30-min incubation, the uniquely labeled fragments were purified by polyacrylamide gel electrophoresis after digestion of the labeled fragment with a second restriction endonuclease. To facilitate the Maxam-Gilbert sequencing, the Pst inserts of some clones were recloned into the plasmid pUC9 (46). Sequence homology searches were carried out using the "SEQ" program (47) or the NUCALN program (48).

**RESULTS**

**Clone Blots**—In order to screen a large number of repeat DNA-containing bacterial clones for repeated sequences of potential interest, we sought to hybridize the cloned DNA with a variety of radioactive nucleic acids. However, our initial attempts to transfer DNA from the bacterial clones to nitrocellulose using versions of the originally described techniques for lysing colonies directly on nitrocellulose paper (e.g. Refs. 49-51) were not satisfactory. The amount of DNA that stably adhered to the nitrocellulose varied greatly from colony to colony, and variable nonspecific hybridization signals occurred at the points of contact between the bacteria and the nitrocellulose paper. We therefore could not reliably compare the hybridization signals generated by the different radioactive nucleic acids.

We largely eliminated this problem by using a technique that we devised for first embedding colonies of the bacterial clones in agarose and then blotting the DNA from them to nitrocellulose (27). As explained in full detail in the Miniprint, bacteria from five 96-well (12 columns x 8 rows) microtiter dishes are grown on nitrocellulose paper (3 x 9.5 cm) in the same grid pattern as the microtiter dish, but one-fifth the size. The nitrocellulose "master" grid is replicated onto agarose "growth sheets," and after colony growth and plasmid amplification in the presence of chloramphenicol, each agarose growth sheet is embedded in a block of agarose. The colonies are lysed in situ and the DNA is transferred from the colonies to nitrocellulose by blotting, and the nitrocellulose filters, which we call clone blots, are then treated like any DNA blot. The "ghosts" of the bacterial colonies remain in the agarose block and are in exact register with the DNA blotted out of them. As is shown by the experiments presented in the Miniprint, the hybridization signals produced on clone blots with either RNA or DNA probes are reliable and are reproduced when the hybridization is carried out using purified cloned DNA.

Fig. 1c shows the ghosts of a grid of bacterial colonies from which the DNA was blotted. Numbers 1-5 refer to the microtiter dish number in which the clones were stored. For each microtiter dish, A through H refer to the rows; 1 through 12 refer to the columns. pR refers to a single top row of 12 clones selected from a sixth microtiter dish.

**Hybridization of Clone Blots with Rat DNA**—Fig. 1, b and c, show the hybridization of radioactive total rat DNA to replicate clone blots at 65 and 55 °C, respectively. A comparison of these results with the corresponding ones shown in Fig. 1, B and C, of the Miniprint shows that, in general, the relative hybridization signals with individual clones is reproducible among replicate clone blots.

Since 90% of the clones (38 of 42 tested) contain inserts between 0.1 and 0.4 kb (the other 4 were 0.4-0.8 kb) and since each of these clones yielded about the same amount of plasmid DNA, as determined from mini-clear lysates (28) (data not shown), then the hybridization signals should be related to

![Fig. 1. Hybridization of total rat DNA to clone blots at 65 and 55 °C. a, the ghosts of the bacterial colonies from which the DNA was blotted. b and c, 17-day exposures of clone blots that had different autoradiograms as b but with the satellite DNA-containing clones indicated by arrowheads.](image)
Repeat DNA

the concentration of sequences complementary to the clone in the radioactive rat DNA. To estimate this relationship, we identified, as described in the Miniprint, those clones that contain satellite I DNA. This family contains about 100,000 0.37 kb long repeat units clustered in long tandem arrays (22), and as Fig. 1d shows (arrowheads), we recovered 11 satellite I clones, a number that is consistent with the copy number of this family. Of the 8 clones tested, each contained less than a full length repeat unit, and most of these clones give among the strongest hybridization signals (e.g. pR10).

Less strong clone blot signals should correspond to clones that contain less highly repeated DNA, and although this assay would only be semiquantitative, this seems to be so. For example, clone pR6, which contains an 0.2-kb sequence repeated 5,000-10,000 times (as determined by reassociation assay would only be semiquantitative, this seems to be so. For example, clone pR6, which contains an 0.2-kb sequence repeated 5,000-10,000 times (as determined by reassociation, data not shown), hybridizes less strongly to total rat DNA than do most of the satellite I clones (Fig. 1, text and Miniprint). Furthermore, a clone that contains a portion of the 200-member rat ribosomal RNA gene family gives considerably weaker but reproducible hybridization signals with radioactive rodent DNA (see the next two sections below and Fig. 2, e-h).

The similarity between the hybridization signals at 65 and 55 °C indicates that most of the repetitive DNA sequences in these clones are not very divergent, with certain obvious exceptions, such as clones G1 and H10. These results were also seen on less exposed autoradiograms of these blots (data not shown), and similar results were also seen with another pair of replicate clone blots (see Fig. 1, B and C, in the Miniprint) as well as with DNA purified from selected clones (see below and Fig. 3 in the Miniprint).

Hybridization of Clone Blots with Rat RNA—Hybridizations of radioactive total liver RNA or the nuclear RNA from rat liver, kidney, and a neural cell line, PC12 (33), are shown in Fig. 2, a-d, respectively. Only one clone (4A5) gave a strong signal with total RNA which is dominated by ribosomal RNA. Clone 4A5 hybridizes specifically to 18 S ribosomal RNA (see Miniprint Fig. 2A).

Comparison of the clone blots hybridized with the nuclear RNAs allowed us to distinguish three classes of repeat DNA clones: (i) those that are not represented in any nuclear RNA (e.g. pR10, pR6) (ii) those that are represented in all nuclear RNA populations tested, including adult muscle nuclear RNA (data not shown) (e.g. clones 1B12, 4A1, and 4D12) and (iii) those cloned sequences that are transcribed in a cell-specific manner (e.g. clone 3B5 in kidney and liver but not PC12 nuclear RNA, clone 2F6 in kidney nuclear RNA). Many of the transcribed clones give a strong hybridization signal with rat DNA and therefore contain sequences highly repeated in rat (e.g. 1B12, 4A1, 3B5, 4D12; Fig. 1b). The hybridization, or lack of it, by some clones with liver nuclear RNA was verified by hybridizing liver nuclear RNA to “dots” of purified plasmid DNA (see Miniprint, Fig. 2B).

4The 100,000 0.37-kb members of the rat satellite I family represent 37,000 kb of DNA or about 8% of the approximately 750,000 kb of rat repeat DNA (i.e. the DNA that reanneals by C22 200 (11)). If the reannealed rat satellite I duplexes and the total C22 200 duplexes that were used for cloning have the same size distribution, then 0.05 of the 487 clones on a clone blot, or 24, should contain rat satellite I sequences. Hybridization of a satellite I DNA clone or total rat DNA to a blot of electrophoretically fractionated C22 200 duplexes that had been treated with nuclease S1 (see “Materials and Methods”) showed that the size distribution of satellite-containing reannealed duplexes was about the same as that of total repeat DNA duplexes (data not shown). Furthermore, 8 of the satellite I clones that were tested contained inserts of between 0.1 and 0.4 kb, which is about the same as the size distribution of inserts in the nonsatellite clones.

Fig. 2e shows that only about 12 or so rat clones are sufficiently homologous to mouse DNA to react with it at 65 °C. Eight of these clones, including 1B12, 4A1, and 5A1, are transcribed in the rat (white arrows, Fig. 2b). Also indicated in Fig. 2e (open squares) are those rat sequences that are more highly repeated in the mouse genome than in the rat genome.

Many of the clones that gave weak hybridization signals with rat DNA at 65 °C do not react at all with mouse DNA at 65 °C. This result indicates that the weak hybridizations of clones with rat DNA (Fig. 1, b and c) are not due to nonspecific binding of radioactivity to clone blots. These weakly hybridizing clones do not contain abnormally short inserts since those from 17 of such clones tested have the same size distribution as those that give good hybridization signals (data not shown). We have not investigated these clones further.

Lowering the stringency of the hybridization to 55 °C (Fig. 2f) increased not only the number of rat clones that can hybridize to mouse DNA, but also the extent of hybridization by some, but not all, of the clones that hybridized to mouse DNA at 65 °C. This would be expected since many, but not necessarily all, of the DNA sequences present in the ancestor of present day rat and mouse would have been expected to diverge since the time that these animals shared a common ancestor estimated to be about 10 million years ago (52). Since mouse and rat do not share satellite DNA sequences (22), each of the satellite I-containing clones (circled in Fig. 2f) reacted poorly, if at all, with mouse DNA, even at 55 °C.

Comparison of Figs. 1c with 2f shows that two rat clones (5G1, 2H10) contain repeated DNA sequences that are divergent in both the rat and mouse genomes. We verified this result, as well as several other findings obtained from hybridizations of mouse or rat DNA to clone blots using plasmid DNA purified from several clones (see Miniprint, Figs. 2, C-F, and 3).

Hybridization of radioactive gerbil or guinea pig DNA is shown in Fig. 2, g and h, respectively. Since rodents share about 200 very homologous ribosomal RNA genes (53), all of the rodent DNAs hybridize to about the same extent to the rat 18 S ribosomal DNA clone, 4A5 (open squares). Furthermore, taken together, the rat repeat sequences present in the mouse, gerbil, and guinea pig genomes account for many of the clones transcribed in rat cells (see Fig. 2b, white arrowheads). Also note that gerbil DNA hybridizes to largely the same subset of clones as does mouse DNA at 65 °C.

Many of the clones that contain highly repeated sequences can be distinguished from each other by their hybridization with the different rodent DNAs. For example, the 3B5 sequence which we distinguished from other highly repeated, transcribed sequences such as 4D12, 1B12, 4A1, and 5A1 by hybridization with nuclear RNA (see above) can also be distinguished from them by hybridization with rodent DNA. The 3B5 sequence is present only in rat DNA; the others are present in mouse, gerbil, or guinea pig DNA. On the other hand, the 4D12 sequence, which like the 1B12, 4A1, and 5A1 sequences is transcribed in all cell types examined, is, in contrast to these sequences, barely detectable with mouse DNA even at 55 °C. However, clone 4D12 is readily detectable with guinea pig DNA at 55 °C (Fig. 2h). And as a final example, both clones 1B12 and 4A1 contain sequences highly repeated in mouse DNA, but they can also be distinguished from each other. The 4A1 sequence is somewhat divergent in the mouse genome, whereas the 1B12 family is not (compare the hybridization of mouse DNA at 65 and 55 °C (Fig. 2, e and f) to clones 1B12 and 4A1).

The Rat Genome Contains a Number of Different Highly

...
Repeat DNA

Total RNA

Liver

Kidney

PC12

Repeat Families—About 30 of the clones produced hybridization signals with total rat DNA that were equal to or greater than the strong signals given by most of the satellite I clones (see Fig. 1, text, and Fig. 1, A and B, Miniprint). Since 9 of the 30 strong signal clones include representatives of the 100,000-member rat satellite I family, we initially assumed that most of the remaining 20 or so strong signal clones contain copies of 1 or 2 other highly repeated sequences. However, as was just discussed, many of the non-satellite strong signal clones can be distinguished from each other by their hybridization characteristics. To verify these distinctions, we determined the DNA sequence of 5 such clones (1B12, 3B5, A14, 5A1, and 4D12), and as Fig. 3A shows, each is different, except that about 80 bp of the 1B12 insert corresponds to a slightly divergent version of the 4D12 repeat. The remainder of the 1B12 insert is also highly repeated in the rat genome and can be readily distinguished from the 4D12 part because, unlike 4D12, it hybridizes well to mouse DNA (Figs. 2 and 4, Table I) and hybridizes with a distinct pattern to restriction enzyme digests of total rat DNA (Fig. 4 and experiments not shown).

To verify that these clones contained highly repeated se-
quences, we determined their frequency in a Charon 4A λ library of rat genomic fragments. As Table I shows, each cloned sequence is highly repeated in the rat genome. Although we did not determine the frequency of the 5A1 sequence in the λ library, hybridization experiments (Fig. 1, text and Miniprint; Fig. 3, Miniprint) as well as other experiments (see “Discussion”) indicate that the 5A1 sequence is repeated at least as often as the 4A1 sequence. Furthermore, as is consistent with the hybridization of mouse DNA to clone blots at 65 °C, the sequences in clones 1B12 and 4A1 are also highly represented in the mouse genome (Table I).

Our detection of a number of highly repeated DNA sequences in the rat genome was not expected from the reassociation studies of rat DNA (11). Just the 5 repeated sequences shown in Fig. 3A would account for about 10^6 kb of rapidly reannealing DNA (Cot 1/2 of less than 0.1); this should be revealed upon reassociation of rat DNA in addition to the satellite I sequences, but is not (Ref. 11; also see “Discussion” in Ref. 21).

Furthermore, it appears that most of the highly repeated, transcribed sequences were not recovered in the repeat DNA library in the numbers expected from their repetition frequency. For example, if clones of each of the sequences shown in Fig. 3A were present on clone blots at the same frequency as the satellite I clones (11 clones per 100,000 genomic copies; see Footnote 4), then the clone blot should contain at least 40 strong signal clones corresponding to just these 5 sequences. This is more than all of the other strong signal clones detected, including the satellite I clones. Furthermore, the hybridizations with mouse or guinea pig DNA show directly that the sequences in Fig. 3A were present fewer times than expected. For example, at 65 °C mouse DNA hybridizes well to 1B12, 4A1, and 5A1, but only to 5 other transcribed clones (Fig. 2e). Even if all 5 cloned sequences contained other copies of 1B12, 4A1, or 5A1 sequences, it is clear that at most only a few clones containing each of these sequences are present on the clone blot. A similar conclusion concerning the frequency of 4D12 clones can be made from the results using guinea pig DNA (Fig. 2h) which hybridized well to 4D12 but only to 3 other transcribed clones.

One explanation for both the low yield of clones of each of the above highly repeated sequences and the reassociation results is provided by the results of Cech and Hearst (54) and Moyzis et al. (55, 56). These workers showed that most of the reannealed duplexes of repeat DNA from mouse and Syrian hamster, respectively, contained single-stranded regions which were not single copy DNA but other repeated DNA sequences. These unpaired repeated sequences anneal to other copies of their families slowly, if at all, presumably due to “steric” hindrance (54–56). Together, their results mean that most, but not necessarily all, of the copies of different repeated DNA sequences are interspersed in clusters of other repeated DNA families (and not in single copy DNA) and that these clusters are scrambled or permuted. Otherwise, once duplex formation started anywhere in the cluster, it would proceed unhindered as in a nonpermuted tandem array, such as satellite DNA. If members of certain rat repeated families are so arranged, then the S1 nuclease step of our cloning procedure

<table>
<thead>
<tr>
<th>Table I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimated copy number of several repeat families</strong></td>
</tr>
<tr>
<td>Clone</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1B12</td>
</tr>
<tr>
<td>3B5</td>
</tr>
<tr>
<td>4A1</td>
</tr>
<tr>
<td>4D12</td>
</tr>
</tbody>
</table>

* Charon 4A λ library (72) of approximately 15-kb rat genomic fragments produced by partial HaeIII digestion of rat liver DNA (kindly supplied by Dr. T. Sargent, National Institutes of Health, Bethesda, MD) was screened by hybridization at 65 °C with about 10^6 cpm of [32P]DNA purified from the indicated clone as described by Benton and Davis (73). The Charon 4A λ mouse library containing similar sized genomic fragments was produced by partial HaeIII digestion of BALB/c mouse embryo DNA by Dr. Jon Seidman (Department of Genetics, Harvard Medical School, Boston, MA). The approximate copy number was calculated as follows: (percent positive plaques) 2.7 million kb rat genome/15-kb genomic fragment. Similar values for the 3B5 and 4D12 sequences were obtained using another (partial EcoRI) library of rat DNA.

May be a low estimate, since four randomly selected genomic clones contain more than one copy of the indicated cloned repeat element.

---

Fig. 3. Opposite page. Partial DNA sequences of several rat repeat DNA clones.

Part A, the DNA sequence that we determined for the indicated clones. (The G-C tails are not shown.) Since these sequences were determined to confirm the differences among these repeated DNAs that were revealed by hybridization experiments (i.e. Figs. 1 and 2), several ambiguities in the sequences (indicated by P, Q, or N for purines, pyrimidines, or any nucleotide, respectively) have not been resolved and gaps remain in some of the sequences. The homology between 4D12 and 1B12 is indicated by the solid line connecting the asterisks on the 1B12 sequence. Dots indicate identity between the sequences, and dashes indicate gaps. The numbering on the solid line corresponds to that shown for the 4D12 sequence. This was the only significant homology detected among these sequences.

Part B, the homologies found between 1B12 and the repeat DNA cluster of Barta et al. (57). The complete Barta et al. (57) sequence (568 bp) is given and is indicated by the up arrows. The comparison with 1B12 is shown above the lines containing the Barta et al. (57) sequence. Dots indicate identity between the sequences; dashes in the Barta et al. (57) sequence correspond to regions of insertions in the 1B12 sequence. The complement of the first 82 bp of 1B12 is homologous to the non-4D12 portion of the Barta et al. (57) cluster, indicating that relative to the 4D12 sequence this repeated DNA sequence is in the opposite orientation, as is its homolog in the Barta et al. cluster. The last 56 base pairs of the 1B12 sequence are not homologous to any of the Barta et al. (57) cluster.

Part C, the homology between 5A1 or 4A1 sequences with a concatenate of two independently sequenced portions of the mouse long interspersed repeat (LINE) family called MIF-1 (67–70). This concatenate was taken from Ref. 68 and consists of the sequence called E24 (nucleotides 1–274 (68)) and Bam5 (nucleotides 275–781 (69, 70)); the 5' and 3' ends are indicated by the up arrows. The homology between the sequenced portion of the 5A1 clone begins as position 147 of the mouse sequence and extends, as is indicated by the solid line below the numbers of the 5A1 sequence, to position 367 of the mouse sequence. The homology therefore extends across the junction of the E24 and Bam5 sequences. The homology between the 4A1 insert extends from position 359 of the mouse sequence (asterisk) and extends, as is indicated by the solid line above the numbers of the 4A1 sequence, to approximately position 550 of the mouse sequence, at which point the sequences begin to diverge. None of the sequences shown in A were homologous to any other sequenced portions of the mouse MIF-1 family (see Ref. 68), nor were any homologous to the mouse B1 (4), or E sequences (71), or any of the “Alu” sequences of human, primate, or Chinese hamster (4) (results not shown).
was exposed for CPM of plasmid ['P]DNA. The autoradiogram of 1B12 with rat DNA enzyme digests of rat or mouse DNA revealed ethidium bromide staining (data not shown). Revealed by the hybridization between 1B12 and rat DNA was not revealed by ethidium bromide staining (data not shown).

A further prediction would be that the reannealing of these repeated sequences is not revealed by reassociation, and this prediction was verified by the experiment shown in Fig. 5, which shows the hybridization of clone blots with total rat DNA and 2 "Cot fractions" of rat DNA: one that reannealed by Cot 1 and a second fraction that reannealed more slowly, between Cot 1 and 200 (Fig. 5c). Fig. 5 also shows the kinetic properties of these DNA fractions. Note that the pattern of hybridization of total DNA to the clone blots in Fig. 1 differs somewhat from these clone blots, which were prepared with a different lot of nitrocellulose than that used for the others (also see the legend to Fig. 5).

As expected, many moderately and highly repeated sequences (e.g. pR6, all of the satellite I clones) are represented in Cot 1 DNA but not Cot 1-200 DNA. On the other hand, a number of highly repeated sequences (e.g. 1B12, 3B5, 4A1, 4D12, and 5A1) are represented in both Cot 1 and Cot 1-200 DNA. This means that copies of these highly repeated sequences are reannealing abnormally slowly. As mentioned above, slow reannealing would be expected if many copies of these repeated sequences are organized as permuted clusters. The alternate explanation for slow reannealing is that many copies of these are divergent; this is not supported by the effect of temperature on the hybridization by these clones to rat DNA (see Figs. 1 and 3, Miniprint). Furthermore, the repeated sequences represented in clones 5G1 and 5H10 which are demonstratively divergent in the rat genome (see Figs. 1 and 3 in Miniprint) are found in the Cot 1-200 DNA fraction (cf. Figs. 1, and 5, b and c).

Fig. 5 also shows that, in addition to the highly repeated sequences, many less highly repeated sequences are also pres-

(see "Materials and Methods") would digest the repeated sequences that remain unpaired. This would account for their underrepresentation in the repeat DNA library. A further prediction would be that the reannealing of these repeated DNA sequences would be slower than expected for their copy number (54-56). This could explain why the highly repeated sequences are not revealed by reassociation, and this prediction was verified by the experiment shown in Fig. 5, which shows the hybridization of clone blots with total rat DNA and 2 "Cot fractions" of rat DNA: one that reannealed by Cot 1 (Fig. 5b) and a second fraction that reannealed more slowly, between Cot 1 and 200 (Fig. 5c). Fig. 5 also shows the kinetic properties of these DNA fractions. Note that the pattern of hybridization of total DNA to the clone blots in Fig. 1 differs somewhat from these clone blots, which were prepared with a different lot of nitrocellulose than that used for the others (also see the legend to Fig. 5).
ent in Cot 1–200 DNA. Comparison of Fig. 5, c and d (which is a copy of Fig. 2b), shows that most of the repeat clones that are represented in Cot 1–200 DNA are transcribed into RNA.

**DISCUSSION**

In addition to identifying a number of different repeated DNA sequences of potential interest (e.g. those that are transcribed in a tissue-specific manner, or are evolutionarily conserved), the major conclusion of this paper is that rat DNA contains a number of different highly repeated DNA sequences. These sequences reanneal abnormally slowly, which would explain why they are not revealed when rat DNA is reannealed (Ref. 11; also see "Discussion" in Ref. 21). Since most of the copies of these highly repeated families are not divergent (Figs. 1, 3, Miniprint), the most likely explanation for their underrepresentation in the repeat DNA library and their retarded reannealing (Fig. 5) is that a significant number (but not necessarily all) of the copies of these different repeated sequences are not interspersed in single copy DNA but are organized as permuted clusters. Direct evidence for this idea comes from the sequence of the repeat DNA in the 1B12 clone, which showed that it is a permuted version of the two rat repeat DNA sequences in a cluster previously identified by Barta et al. (57) (see Fig. 3B and below). The failure of the reassociation studies (11) to reveal highly repeated rat sequences has implications beyond the organization of repeated DNA in certain rodent genomes (Refs. 54–56 and this study). As has been thoroughly discussed by Moyzis et al. (55, 56), conclusions about the organization of interspersed repeat sequences based on much of the earlier reassociation experiments may require re-evaluation.

Along with the evidence for permuted clusters that was obtained by reassociation studies of mouse and Syrian hamster DNA (54–56), specific examples of permuted clusters have been found in the sea urchin (58), chicken (59), Droso-

**ACKNOWLEDGMENTS**—We thank David Lin and Susan Waitzkin for their advice and encouragement, and F. R. Witney and A. V. Furano, unpublished observations.

---

10489
their help in determining some of the DNA sequences, Dr. Francine Eden for screening the Charon 4A λ rat library with the pR4D12 and pR3B5 clones, and Dr. Ron Lieberman for screening the Charon 4A λ mouse library with clones 1B12 and 4A1.

REFERENCES

Procedure for Making Clone Blots

Clones from 5 microtiter dishes are transferred to a 96-prong colony stamper to each of 5 fresh 1-agar (L-broth) plates without glucose with 25 g/l of agarose (80°C). The master grid is incubated at 37°C, colony side up, on 1-agar plates containing 20 mg/ml tetracycline for 24 hours. After incubation for 3 to 7 hours the growth sheet is removed and the nitrogen cellulose master grid is imprinted by pressing with a sterile forceps. At this time. At least 1 cm border of agarose overlay and the growth sheet. After blotting (see below), the master grid is transferred to a fresh L-agar plate until the clones have grown to their original size and anodized set of agarose growth sheet replicas can be made from them. At any point the master grid can also be replicated by allowing colony growth on agar containing 25 g/l of agarose and 0.5% glucose with 1% tetracycline. After incubation for 20 to 24 hours the growth sheet is rehydrated for about 10 minutes in 0.25 M NaOH, 1.0 M NaCl, to lyse the colonies. The gel is then treated exactly as described by Wahl et al. (17) for blotting DNA fragments out of agarose. Colony clusters are excised from the agarose and placed on a blotting sheet until imbedding the clones. Whatman No. 1 filter paper that has been placed on a scintillated glass funnel. When frozen into a 2.0 cm disk of Whatman No. 1 filter paper that had been placed on a scintillated glass funnel, Whatman No. 1 paper is dry, not overly. The growth sheet is transferred colony side up to a flat surface and slowly overlaid with about 6 ml of 1a agarose which is at 65°C. After the agarose gels, the agarose overlay and the growth sheet. After the overlay is completed, the growth sheet is placed on agarose sheets containing 20 g/l of tetracycline and after suitable growth of the bacteria the colonies are excised from the agarose sheet. In this way suspensions of freshly grown clones each containing about 10^8 bacterial cells are prepared. Colonies from 5 microtiter dishes are transferred with a %-prong colony stamper to fresh 1-agar plates containing 25 mg/ml tetracycline and after suitable growth of the colonies the growth sheet is removed. In this way suspensions of freshly grown clones each containing about 10^8 bacterial cells are prepared. Colony clusters are excised and then immediately plunged into 0.5 M Tris-C1 buffer (pH 8.0) or 0.25 M NaOH for 10 minutes and then immediately plunged into 0.5 M Tris-C1 buffer (pH 8.0).

The nitrocellulose "master grid" is incubated at 37°C, colony side up, on 1-agar containing 20 mg/ml tetracycline. After incubation for 20 to 24 hours the growth sheet can be removed and the nitrocellulose master grid is imprinted by pressing with a sterile piece of stiff plastic (e.g., the top of a plastic petri dish). The master grid can then be transferred to a fresh 1-agar plate until the colonies have grown to their original size and anodized set of agarose growth sheet replicas can be made from them. At any point the master grid can also be replicated by allowing colony growth on agar containing 25 g/l of agarose and 0.5% glucose with 1% tetracycline. After incubation for 20 to 24 hours the growth sheet is rehydrated for about 10 minutes in 0.25 M NaOH, 1.0 M NaCl, to lyse the colonies. The gel is then treated exactly as described by Wahl et al. (17) for blotting DNA fragments out of agarose. Colony clusters are excised from the agarose and placed on a blotting sheet until imbedding the clones. Whatman No. 1 filter paper that has been placed on a scintillated glass funnel. When frozen into a 2.0 cm disk of Whatman No. 1 filter paper that had been placed on a scintillated glass funnel, Whatman No. 1 paper is dry, not overly. The growth sheet is transferred colony side up to a flat surface and slowly overlaid with about 6 ml of 1a agarose which is at 65°C. After the agarose gels, the agarose overlay and the growth sheet. After the overlay is completed, the growth sheet is placed on agarose sheets containing 20 g/l of tetracycline and after suitable growth of the bacteria the colonies are excised from the agarose sheet. In this way suspensions of freshly grown clones each containing about 10^8 bacterial cells are prepared. Colony clusters are excised and then immediately plunged into 0.5 M Tris-C1 buffer (pH 8.0) or 0.25 M NaOH for 10 minutes and then immediately plunged into 0.5 M Tris-C1 buffer (pH 8.0).
Verification of clone blot signals by hybridizing purified cloned repeat DNA to various nucleic acids. Panel A: Approximately 10^6 cpm of clone 5H10 DNA was hybridized in 10 ml at 65°C for 24 hours to a nitrocellulose blot (76) of glyoxal-treated (77) rat RNA that was fractionated on a 1.2% (w/v) agarose gel (electrode buffer, 10 mM Tris, pH 8.0). Ethidium bromide-stained bands of parallel tracks are shown at the right side of panel A. The x-ray was exposed to the blot for 18 hours. Panel B: The 2 nitrocellulose strips of "dots" (76) loaded with 0.15 ml of the indicated plasmids were hybridized to 500,000 cpm of liver nuclear DNA at 65°C for 18 hours at 0.5 M NaPO₄, pH 6.5. The exposure to x-ray film was for 10 days. Panel C: The x-ray was exposed to the blot for 18 hours. Panel D: The 2 nitrocellulose strips of "dots" (76) loaded with 0.15 ml of the indicated plasmids were hybridized to 500,000 cpm of liver nuclear DNA at 65°C for 18 hours at 0.5 M NaPO₄, pH 6.5. The exposure to x-ray film was for 10 days. The numbers 1 and 2 refer to the position of the insert in clones 5H10 and 5G1, respectively.