Volume 10 Number 10 1982 Nucleic Acids Research

Genomic representation of the Hind II 1.9 kb repeated DNA

L.Manuelidis and P.A.Biro

Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA

Received 3 February 1982; Revised and Accepted 30 April 1982

ABSTRACT

The genomic representation and organization of sequences homologous to a cloned Hind III 1.9 kb repeated DNA fragment were studied. Approximately 80% of homologous repeated DNA was contained in a genomic Hind III cleavage band of 1.9 kb. Double digestion studies indicated that the genomic family, in the majority, followed the arrangement of the sequenced clone, with minor restriction cleavage variations compatible with a few base changes. Common restriction sites external to the 1.9 kb sequence were mapped, and hybridization of segments of the cloned sequence indicated the 1.9 kb DNA was itself not tandemly repeated. Kpn I bands which were homologous to the sequence contained specific regions of the repeat, and the molecular weight of these larger fragments could be simply explained. Mapping of common external restriction sites indicated that in some but not all cases the repeat could be organized in larger defined blocks of $\geq$ 5.5 kb. In some instances, flanking regions adjacent to the repeat may contain common DNA elements such as other repeated DNA sequences, or possibly rearranged segments of the 1.9 kb sequence. It is suggested that although the 1.9 kb sequence is not strictly contiguous, at least some of these repeated sequences in the human genome are arranged in clustered or intercalary arrays. A region of the 1.9 kb sequence hybidized to a mouse repeated DNA, indicating homology beyond the primates.

INTRODUCTION

In human nuclear DNA digested with Hind III a band of 1.9 kb can be visualized after staining with ethidium bromide (1,2,3). We previously utilized a series of human repeated DNA probes to ascertain if the DNA sequences in this band were homologous to previously described human repeated DNAs. The Eco R1 340 bp tandem repeat (4) that is highly represented on specific human chromosomes at the centromeres (5), and probes from the "Alu family" of short interspersed repeated sequences (6,7) both failed to highlight the 1.9 kb band; other minor human centromeric DNA satellites were also distinguished from this 1.9 kb sequence (3,8). A cloned member of the Hind III 1.9 kb family was sequenced (8). In accord with hybridization studies, it is a member of a novel repeated DNA sequence family.

0305-1048/82/1010-3221$2.00/0

Using this cloned 1.9 kb DNA, or defined segments of it, we were able to study the restriction sites of related family members in whole genomic digests in greater detail. The interpretation of the data was simplified because ~80% of the hybridizing family members in the genome are defined by the Hind III restriction sites, yielding a single band of 1.9 kb. First, we show in double digestion experiments that most genomic members of this family contain the same restriction sites and internal order of subsegments as that of the cloned DNA. Minor variants, consistent with a few base changes, are also detected in some members of the genomic family. Second, we have used the clone to determine major restriction sites external to the 1.9 kb sequence. These latter studies indicated that the 1.9 kb sequence is not tandemly repeated. Several other restriction enzyme bands detected by hybridization are specifically related to regions of the 1.9 kb repeat. Genomic maps of common external restriction sites are presented; these maps explain the interrelationships of the more numerous Kpn I bands. A simple explanation of the relationships of these Kpn I bands by analysis of DNA fragment size was previously not possible.

MATERIALS AND METHODS

Isolation of nuclear DNA, restriction enzyme digestion, gel electrophoresis, and Southern hybridization were as described with minor modifications (4,9). Gels were depurinated in 0.1 N HCl for 10 min prior to alkaline denaturation and blotting to ensure transfer of higher molecular weight DNA fragments. Several gels analyzed contained ^{32}P markers of 4-6 kb and these labelled DNAs were efficiently transferred. The Hind III 1.9 kb band was isolated from a preparative gel and cloned as described (10) in λ Charon 27 (Blatter, personal communication).

Some DNA clones derived from the impure preparative band contained DNA other than the 1.9 kb repeat (3). After finding a clone that contained the 1.9 kb repeat, the insert was subcloned in pBR322. Plasmid DNA was prepared, and the insert excised and end labelled with ^{32}P as described (8). Since there was little homology between different subsegments of the 1.9 kb sequence (8), it was possible to determine the orientation of fragments in longer arrays in the genome. End labelled subfragments were prepared as described (8).

Fibers were hybridized in 4xSSC at 60°C without dextran sulfate for 16-18 hours as described (9), and washed at 60°C in 2xSSC. These conditions were considered to be of moderate stringency. Blots that were reused were

generally stored for $\geq$ 6 weeks, and the initial labelled probe was then eluted by washing the filter (prewetted in 2xSSC) for 5min. in 450 ml of H_2O containing 0.7mM EDTA pH 8.0 at 86°C (starting) to 76°C (final temperature), as suggested by S. Dube (personal communication). Hybridization results were confirmed on virgin blots. Autoradiography was generally for 1-5 days, with or without intensifying screens at -70°C except for studies of mouse DNA blots, where exposures of 1-3 weeks were necessary to detect hybridization.

For molecular weight determinations, a small amount of Eco RII digested cloned DNA in plasmid was run in parallel lanes in gels used for Southern hybridizations. Thus exact fragment lengths as determined by sequencing were detected at the same magnification on the same hybridized blot as the lanes containing genomic fragments. These hybridization studies also allowed confirmation of the purity of the probes utilized, since only certain restriction fragments of the cloned plasmid should hybridize to probes from a given segment of the insert (8). By these hybridization criteria probes were judged to be >90-95% pure. Internal controls on whole genomic digests also indicated clear differences between different probe segments (see Results). When the genomic fragment migrated with the cloned sequence fragment, the molecular weight is given as the exact sequence length. For other fragment length determinations, several gels were analyzed on standard semi-log plots. In all experiments DNA digestion was considered complete since additional digestions with 10 fold excess of enzyme for 16hr failed to change the position of hybridizing bands.

RESULTS

Figure 1 shows the Hind III 1.9 kb ethidium bromide stained band in addition to other minor bands. A densitometer trace of a typical Hind III digest of human placental DNA is also shown, with the peak of the 1.9 kb fragments of interest on top of a smear of other DNA. The area under this peak was calculated in several traces as described (11). Analysis of most traces indicated this band comprised 0.24-0.35% of the total DNA of the human genome, although one trace showed ~1% of the total. Study of a number of cloned neuroectodermal tumors, and male and female DNAs showed comparable amounts of this band in each preparation (3). Other minor bands seen in ethidium bromide stained Hind III digests (Fig. 1) were homologous with the Eco Rl 340 b.p. human tandem repeats (3) and were not related to the 1.9 kb band (8). When labelled 1.9 kb cloned DNA, or segments of the cloned insert,

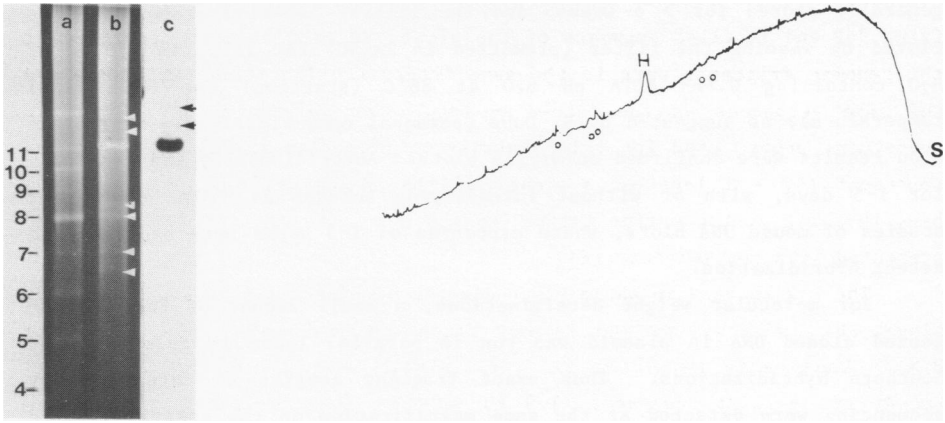

Figure 1. Whole nuclear human DNA cleaved with Xba I (lane a) and Hind III (lane b) show ethidium bromide visualized bands. The Xba I bands represent the Eco RI 340 bp dimer family (4), indicated as molecular weight multimers of 170 bp. Note that the major genomic Hind III band migrates just above the 11-mer of 1870 bp. Some other faint ethidium bromide bands (white arrowheads) strongly hybridized to the Eco RI dimer family (3), but not to the prominent 1.9 kb Hind III band. Lane c shows hybridization of the 1.9 kb cloned DNA to the 1.9 kb genomic band. Two very minor bands (black arrows) were also visualized. Densitometry and analysis of this overexposed film (where the Hind III band was outside the linear film density) indicated that it accounted for ~70% of the total radioactivity, and the minor bands each represented ~4% of the total. In less exposed films the background radioactivity was essentially not detectable, suggesting most homologous members were contained in the Hind III 1.9 kb band. The densitometer trace is from a typical digest of whole nuclear DNA digested with Hind III. From this and other traces, the amount of DNA in the Hind III 1.9 kb band (H) was calculated to contain 0.25%-0.35% of the total DNA. S is slot for orientation. Circles beneath minor ethidium bromide band indicate positive hybridization to the Eco RI dimer family of sequences.

were hybridized to whole human DNA digested with Hind III, essentially only the 1.9 kb band was detected (Fig. 1). Densitometry of overexposed autoradiograms indicated ~70% of the hybridizing family members were contained in the 1.9 kb band; exposures in which the 1.9 kb band was within the linear film density showed virtually no other regions of hybridization. Thus we estimated that ~80% of the family members are contained in the 1.9 kb band. In very overexposed autoradiographs a smear of higher molecular weight material also showed homology to this sequence, (e.g., Fig. 4A, lane 5) with minor additional bands of 2.5 and 3.0 kb.

The cloned 1.9 kb fragment fairly represents the major genomic family

Using double digests of whole human DNA we were able to analyze

restriction sites present in the genomic 1.9 kb band. Furthermore, utilizing ^{32}P end labelled segments of the clone, we were able to ascertain if the genomic fragments were in the same internal orientation, and contained the same sequences as the cloned member. For reference, a map of the subsegment probes used for this analysis is presented in Fig. 2, where A-G represent the end labelled segments of the clone used as probes, shown beneath a map of the entire insert (Cp); a more complete restriction map is presented with the sequence (8).

The 1.9 kb band in whole human digests of Hind III plus Xba I showed a single Xba I site was present in most of the genomic members (Fig. 3B, lane H/X). Thus bands of 402 bp and 1492 bp were seen in addition to a minor amount of residual undigested 1.9 kb material (see Methods for molecular weight determinations). Digestion with a 10-fold excess of enzyme for 24 hours still left a small amount of residual 1.9 kb material (as in other double digests described). The major pattern of Xba I digestion in genomic DNA was compatible with the single Xba I restriction site seen in the cloned member of this family. In order to further test this homology, segments of the clone were used for hybridization. Segments G or H from the 3' end of the clone (representing insert positions 1682-1894) hybridized as predicted only to the 402 bp fragment (Fig. 3C, lane H/X) while other segments up to

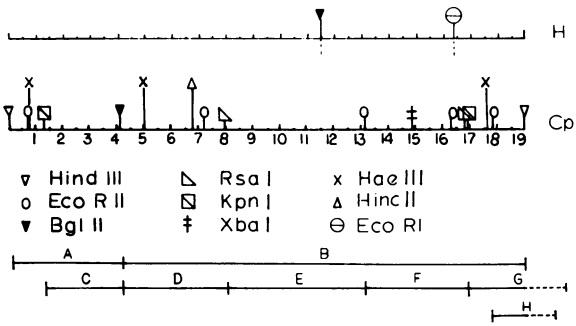

Figure 2. A simplified restriction map of the cloned insert is shown (Cp). Segments A-H are segments of the clone that were used in hybridization studies. The dotted line indicates segments of pBR322 included in two fragments (pBR322 sequences do not hydridize to these bands). Xba I, Hae III, Rsa I, Hinc II, and Eco RII sites were represented in the majority of genomic members as well as in the clone. One Kpn I site in the cloned DNA (at position 1685) was present only in ~1/2 of the genomic family (see Fig. 3B, lane H/K). Restriction variants in the human genomic family are shown in the human map H, with a dotted line indicating cleavage of some family members. A common Eco RI fragment was mapped in most genomic members. The Bgl II site at position 403 was also not apparent in all genomic members, and an additional Bgl II site is suggested at position 1142 for some members of this family (see Results).

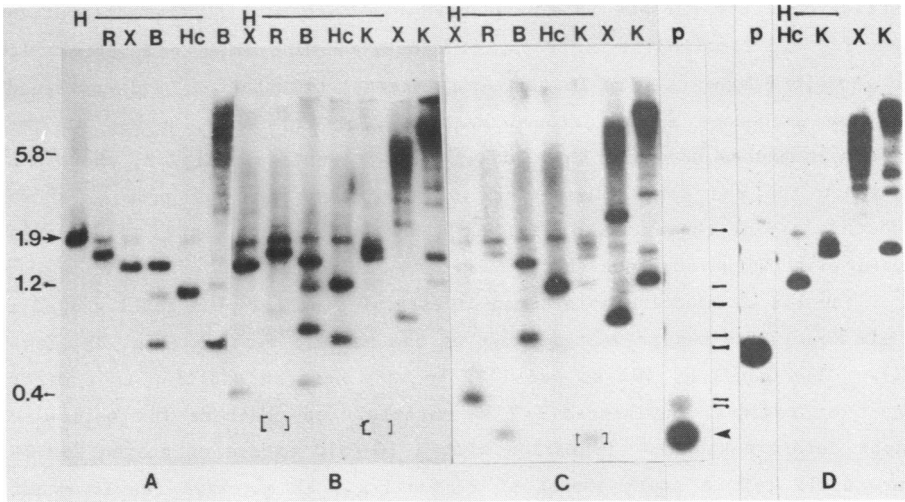

Figure 3. Digests of total nuclear DNA (5 μg), hybridized to the 1.9 kb cloned DNA or to segments of the clone. H is Hind III digestion. A line indicates double digest with Hind III plus Eco RI (R), Xba I (X), Bgl II (B), Hinc II (Hc), and Kpn I (K). Single digests with Bgl II, Xba I and Kpn I and 0.25 μg plasmid digested with Eco RII (p) were included as controls. Panel A shows hybridization with segment E of the clone (position 792-1303). In panel B the whole 1.9 kb insert was used for hybridization. Panel C shows hybridization with 3' terminal segment H (position 1779-1894). Definite but faint bands that do not reproduce well are in brackets. Molecular weight markers (in kb) are indicated, with the 1.9 kb position (arrow). Panel C, lane p shows the predicted overexposed 215 bp insert band (◄) even though it is not well retained on filter. This probe also contained minor contamination with an adjacent short 3' terminal sequence that accounts for faint bands of 1.6 kb seen in lanes H/R and H/K. Lines indicate positions of all other plasmid fragments. Panels A,B,C were not previously hybridized with another probe. Panel D shows part of panel C rehybridized with segment E, 792-1303. Intense bands are at same positions as seen in A and overexposed plasmid lane (p) shows only a single appropriate hybridizing band, not contaminated by other clone segments.

the Xba I site, e.g., segment E, from the center of the cloned insert (Fig. 2) hybridized to the 1492 bp fragment (Fig. 3A, lane H/X). Thus the Xba I site in the whole genomic family showed the same orientation as that described for the cloned member.

Other restriction sites were similarly studied. For example, a single site for Hinc II is seen in the cloned insert, which yields a 672 and a 1222 bp fragment on double digestion with Hind III plus Hinc II. The majority of the genomic family members in similar double digests also yielded two bands of 1222 and 672 bp; segments from position 1-672 of the cloned DNA showed hybridization to the 672 bp fragment (data not shown), while more 3' seg-

ments E and G hybridized only to the 1222 bp fragment (Fig. 3A,C,D, lanes H/Hc). Eco RII plus Hind III digests also yielded identical and appropriately hybridizing bands as the cloned member run in a parallel lane. Furthermore, these studies showed that under the hybridization conditions employed, there was faithful homology only to the expected segment, with no significant labelling of other regions of the 1.9 kb sequence. The lack of internal homology between different segments was independently predicted by sequence analysis (8). Purity of probes was judged not only by sequence, but by hybridization criteria. Since there were roughly equal proportions of each segment of the sequence in the genome, the intensely hybridizing bands could be interpreted. Very faint other bands could represent minor restriction site variants, or reflect a small amount of probe cross-contamination (e.g., Fig. 3C).

Although the pattern of most of the digests studied showed major similarities to the cloned member of this family, several minor differences in some double digests of genomic DNA were observed. These indicated variation at several restriction sites. The genomic 1.9 kb family is largely cut once by Eco RI, into fragments of 260 bp and 1640 bp, yet the cloned DNA has no Eco RI site. Using labelled segments of the clone we were able to determine that the Eco RI site in the genome was 260 bp 5' to position 1894, since the genomic 260 bp fragment hybridized only with terminal segments G and H (Fig. 3C, Lane H/R), and not significantly with other segments (e.g., segment E, Fig. 3A, lane H/R). In several experiments, more residual 1.9 kb material was seen in Hind III plus Eco RI digests of whole human DNA than in other double digests; the clone we have sequenced is a less representative member of this family with respect to the Eco RI site. There are two Eco RI* sequences at positions 1641 and 1652 in the cloned sequence; a single base change at one of these sites would yield an Eco Rl site, and is likely to occur in the major genomic members of this family. Additional Eco RI sites may also occur in a few members of the genomic family; the cloned sequence has many additional Eco RI* sites (8).

Additional restriction sites, or lack of sites in some members of this family were also seen in Kpn I and other digests. Two Kpn I sites were present in the clone, yielding a central 1.55 kb fragment; this band was seen in $\sim$1/2 of the genomic family (Fig. 3B, lane H/K). The other half of the genomic DNA family contained only one Kpn I site, and migrated between the 1.55 kb and 1.9 kb bands. Variation in one Kpn I site (at sequence position 1685) helps to explain two of the major hybridizing bands observed

in Kpn I single digests of whole genomic DNA (vide infra). Hae III digests showed a major band of 1247 bp hybridizing to appropriate central segments of the clone (See Fig. 6, lane b). Again, the whole genomic array generally reflected the Hae III sites in the clone at positions 504 and 1751 with a lack of other central Hae III sites; a few fainter higher molecular weight bands were also obtained which could be consistent with single base alterations deleting these Hae III sites in some family members. Principal Eco RII and Rsa I cleavage fragments also migrated at identical positions as the cloned repeat, and each hybridized to the same segments as the clone, again with a few minor variations consistent with absence of these sites in a few members of the family. Bgl II restrictions yielded a more complex set of bands than the other digests. An additional Bgl II site at position 1142 and a loss of the 403 site in some genomic members was inferred from comparison of single and double digestion experiments (vide infra).

In gels loaded with equal amounts of DNA it could also be appreciated, especially on shorter autoradiographic exposures, that the 1.9 kb band was darker than the cleaved fragments containing two or more hybridizing subsegments. Band intensities also indicated that there were roughly equivalent representations of each subsegment of the repeat, further suggesting that the 1.9 kb unit may define the most common members of this family.

Mapping common restriction sites external to the 1.9 kb sequence in genomic DNA

In hybridization studies of whole genomic DNA digested with a single enzyme, several prominent higher molecular weight bands were observed in addition to a smear of other hybridizing material. When gels were loaded with the same amount of DNA in each lane, it was apparent especially on shorter exposures, that the 1.9 kb band was considerably more intense than the bands seen in other restriction digests (e.g. Fig. 4B,C,E,F). Although we have not strictly quantitated these studies, the results are consistent with the view that the 1.9 kb repeat represented most members of this family. A second feature of interest was that although there was a smear of material (indicating "interspersion" with many other variable restriction fragments), the bands detected indicated there were common external restriction sites shared by some members of the 1.9 kb repeat. Our purpose in these studies was to define and orient some of these common sites. Using segments of the 1.9 kb clone we were also able to show that the segments external to the 1.9 kb fragment did not contain exact tandem repeats.

The 1.9 kb clone contained no Bam HI, Bst EII or Pst I restriction sites. The whole genomic representation of this family likewise showed only

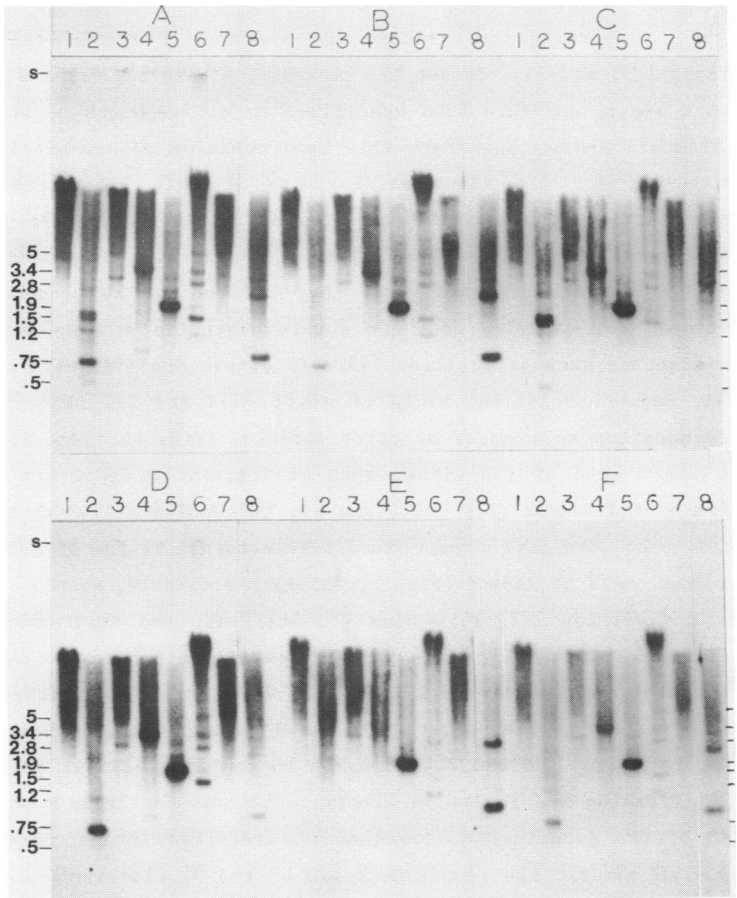

Figure 4. Single digests of whole genomic DNA (5 μg) from 0.7% agarose gels hybridized with 1.9 kb probe or segments of it. Restriction enzyme cleavages are as follows: 1=Bam Hl, 2=Bgl II, 3= Bst EII, 4=Eco RI, 5=Hind III, 6=Kpn I, 7=Pst I, 8=Xba I. Panel A was hybridized with the whole 1.9 kb in plasmid. Panel B shows hybridization with segment B (insert positions 403-1894). Panel C shows hybridization with segment A from the 5' end of the insert (positions 1-403), Panel D shows hybridization with segment E segment E (positions 792-1303), Panel E hybridized with segment H from the 3' end of the insert (positions 1779-1894), Panel F hybridized with segment D (position 403-792). Two parallel blots were used in these experiments (panels hybridized in the following order: A→ B → D→ F and C → E). Other experiments on virgin blots gave comparable results to these blots that had been eluted and rehybridized except for panel D lane 8 which shows a spurious 0.88kb Xba I band which had not been completely eluted (see Fig. 3D, lane X). All exposures here were for 3-5 days, and in some lanes are overexposed in order to detect fainter bands in adjacent lanes (e.g., lanes 6). S is slot and molecular weight markers are shown on both sides of the figure.

a smear of larger molecular weight fragments (> 1.9 kb) after cleavage with these enzymes, suggesting these sites were external to the 1.9 kb repeat (Fig. 4, lanes 1,3, and 7). In Bst EII cleavage, a band of 3 kb was seen in addition to a smear, and this band hybridized to all segments of the clone (Fig. 2, segments A-H). Therefore this band contains the complete 1.9 kb segment with two Bst EII sites external to it that are shared by a detectable proportion of this family (e.g., see Fig. 4A-F, lanes 3).

Other common restriction sites external to the 1.9 kb sequence could be precisely mapped and oriented since they cut the 1.9 kb sequence at a known position. For example, from the double digestion studies Xba I cut 90% of the sequence once at position 1492; in single Xba I digests of whole genomic DNA, major hybridizing bands of $\sim$0.88, 2.4 and 3.3 kb were highlighted, in addition to a smear of other material (Fig. 4A, lane 8). Segments from the 5' end of the clone, such as fragment A (position 1-403), were used in hybridization (Fig. 4C, lane 8); here only the 3.3 kb fragment was detected. Thus the most common Xba I site external to the 5' end of the 1.9 kb fragment could be mapped (Fig. 5, Xba I). Similarly, when 3' terminal segment H (position 1779-1894 plus plasmid) was used in hybridization (Fig. 4E, lane 8), the 0.88 and 2.4 kb bands were highlighted, in this case with little other background hybridization. These results indicate there are two principal Xba I sites external to the 3' end of the 1.9 kb sequence. In order for both the 0.88 and 2.4 kb bands to contain terminal segment H, the first 3' flanking Xba I site is likely to be only partially represented (in 1/2-2/3 of the genomic DNA, based on the relative intensities of the 0.88 and 2.4 kb bands); the remaining $\geq$ 1/3 of the 3' flanking DNA regions lack the Xba I sites nearest to the 1.9 kb 3' end, yielding a second 2.4 kb band as proposed in the map in Fig. 5.

Eco RI and Hinc II digests were similarly analyzed, and some of the hybridizations for the Eco RI studies are shown in Fig. 4, lanes 4. The common 3.3 kb fragment extended from the genomic Eco RI site from position 1642 backwards as mapped in Fig. 5. In studies where segments 5' to position 1640 were used as a probe, the 3.3 kb fragment was clearly highlighted (Fig. 4C and F, lanes 4). The opposite 3' end of the cloned sequence (probe segment H) showed only a smear of hybridizing DNA, which indicated there were variable Eco RI sites flanking the 3' end (Fig 4E, lane 4). A major common Hinc II site at 1.45 kb 5' to the internal Hinc II position 675 (present in 90% of family members) was similarly mapped (Fig. 5). At the 3' end of the repeat, Hinc II digestion revealed a set of discrete faintly

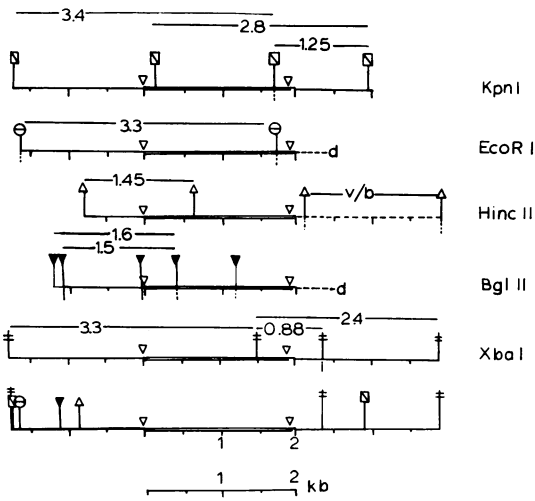

Figure 5. Maps of common restriction sites external to or flanking the 1.9 kb Hind III sequence. The orientation of the repeat is shown as it was sequenced in pBR322 (8). The 1.9 kb fragment is indicated by double lines limited by an open triangle. d = diffuse pattern, no discrete band visualized; v/b indicates multiple weak bands. Dotted lines beneath restriction symbol indicates sites present only in some family members. A composite of external sites as determined in each restriction analysis is shown above the scale; the $\geq$ 5.5 kb DNA, reflecting common restriction sites, may define a subset of the 1.9 kb repeats. The maps depict restriction sites for some, but not all genomic family members.

hybridizing bands, which appeared in a ladder-like fashion. Molecular weight analysis however, revealed no obvious exact or simple numerical relationships between the position of these sites which were 200 bp-2 kb external to the 3' end. It is possible that some alternating common element could define these ladder-like sequences, e.g., as those seen adjacent to some ribosomal genes (12). Further cloning experiments may clarify the nature of these flanking sites.

Kpn I digests of whole genomic DNA show a series of ethidium bromide stained bands that in hybridization studies were shown to be homologous to the 1.9 kb repeat. Analysis of Kpn I digests using segments of the clone as probes (vide supra), clearly indicated the origin of these bands and helped to explain their exact relationship. The Kpn I 1.55 kb fragment represents the segment from Kpn I positions 130-1685 in the repeated sequence. Kpn I bands of 1.25 kb and 2.8 kb preferentially hybridized to 3' terminal segment H, (Fig. 3C, lane K and Fig. 4E, lane 6). This 1.25 kb band did not hybrid-

ize to segment A from the 5' end of the 1.9 kb sequence; segment A did hybridize to the 1.55 kb fragment as predicted (Fig. 4C, lane 6). From double digests we knew that ~1/2 of this family lacked the Kpn I site at position 1685. The 1.25 and 2.8 kb fragments were most consistently and simply explained as shown in the Kpn I map in Fig. 5; they have a common Kpn I site external to the 3' end of the 1.9 kb sequence, and some of the family members lack the Kpn I site at position 1685. Hybridization studies of the 3.4 kb fragment indicated it contained segments up to position 1685, but not more terminal 3' segments (e.g., Fig. 4D and E, lanes 6). Thus there was a common Kpn I site external to the 5' end as mapped in Fig. 5.

Bgl II digests of genomic DNA gave a complex series of bands upon hybridization with the cloned 1.9 kb repeat (Fig. 4A, lane 2). There were clear differences in the hybridization pattern using 5' or 3' terminal segment probes. Segment A from the 5' end clearly highlighed bands of ~1.5 and 1.6 kb, and a minor band of ~500 bp (Fig 4C, lane 2). These bands were not seen when probes 3' to position 403 were used (e.g. Fig. 4B,D,E,F, lanes 2). An infrequent site ~100 bp 5' to the 1.9 kb repeat, and two frequent Bgl II sites 1.5 and 1.6 kb 5' to position 403 (internal Bgl II site) were therefore suggested (Fig. 5); alternatively, a single external Bgl II site with a single base change at internal position 484 could yield the 1.5 and 1.6 kb bands. Hybridization with the 3' terminal segment H showed only a diffuse hybridizing pattern of bands >1.9 kb, without hybridization to the 0.5, 1.5 and 1.6 kb bands (Fig. 4E, lane 2).

Single Bgl II genomic digests also showed a prominent ~750 bp band that hybridized to internal segments of the 1.9 kb repeat, (3' to position 403 but not to 3' terminal segment H). In contrast, using this same 3' terminal probe in double Hind III/Bgl II genomic digests, hybridization to the 750 bp band was seen (Fig. 3C, lane H/B compared with Fig. 4E, lane 2). An additional restriction site in some genomic members at position 1142 (requiring a 2 bp insertion) is therefore postulated; the Hind III/Bgl II digest would thus yield two segments (403-1142 and 1142-1894) that both contribute to the 750 bp band whereas the single Bgl II digest would be derived only from positions 403-1142, as the sequence 3' to 1142 is not cut with Hind III. Furthermore, absence of Bgl II sites in some family members at positions 403 and 1142 would yield bands of ~1.2 kb (1-1142) and 1.5 kb (403-1894) compatible with the hybridizations obtained in Hind III/Bgl II digests (Fig. 3B, lane B).

The 1.9 kb repeat is itself not tandemly repeated

If the 1.9 kb sequence were tandemly repeated, the initial 5' segment should be preceded by sequences from the 3' end. Similarly, the 3' end of the sequence should be followed by the 5' initial segment. Several mapped fragments, including clearly delineated ends of the 1.9 kb sequence plus flanking DNA external to the ends were studied using hybridization of cloned DNA segments, to determine if tandem repeats occurred. For example, the XbaI fragment of 0.88 kb (Fig. 5) in the case of a tandem repetition should contain sequences homologous to segment A from the 5' end of the sequence. No hybridization was detected between fragment A and this 0.88 kb DNA (Fig. 4C, lane 8). Controls indicated that segments B and H which included the 3' end of the sequence (1492-1894) showed clear hybridization to this 0.88 kb band as predicted (Fig. 4B and E, lane 8). Similarly, Bgl II digests of whole human DNA contained 1.5 and 1.6 kb fragments including the beginning of the 1.9 kb sequence plus flanking DNA 5' to it (Fig. 5). Again hybridization of segment H from the opposite 3' end of the 1.9 kb sequence showed no homology to either of these bands (Fig. 4E, lane 2); controls such as fragment A from the 5' end did show hybridization to both bands (Fig. 4C, lane 2), and hybridization with more central segments of the molecule confirmed that the Bgl II 1.5 and 1.6 kb sequences did not include the 1.9 kb sequence 3' to position 403 (e.g., hybridization with fragments B, D, and E, Fig. 4B,F, and D, lanes 2). Study of Kpn I genomic digests also revealed no tandem repetition. The 3.4 kb fragment which included the 5' initial segment and flanking DNA (Fig. 5) did not hybridize with 3' terminal segment H (Fig 4E, lane 6), and the 1.25 kb fragment from the 3' end did not hybridize with 5' segment A (Fig. 4C, lane 6). Studies with Hinc II (data not shown) also demonstrated no tandem repetition of the sequence; there was no hybridization of the 1.45 kb 5' end containing flanking DNA (Fig. 5) with fragment H from the 3' end.

An unexpected finding was that other segments of the 1.9 kb fragment might be included in the regions flanking the sequence ends. One example of this is shown in Fig. 4F, lane 8. Some members of the 0.88 kb Xba I band hybridized with segment D from internal clone positions 403-792. Trivial explanations of this result could include some contamination of the probe with 3' terminal sequences, or additional restriction sites not mapped here. However, it is also possible that some subsegments of the 1.9 kb repeat might be permutated in some genomic members. "Scrambling" of repeated elements has been noted in other systems (13) and further cloning will help to

clarify this possibility.

The above results demonstrate that a fair proportion of the 1.9 kb genomic family is relatively homogeneous in its internal organization. Some members furthermore contain common restriction sites external to the 1.9 kb sequence, whereas others appear more diffusely associated with many other heterogeneous sequences.

An homologous sequence element is detectable in mouse genomic DNA

We previously identified two distinct ethidium bromide stained bands of ~1.5 and 1.7 kb in mouse DNA digested with Bst Nl that did not hybridize to mouse satellite DNA; each of these bands represented distinct repeats that appeared to be "interspersed", or present on chromosome arms (14). We also found that the human 1.9 kb cloned sequence, in hybridization studies highlighted the mouse Bst Nl band of ~1.5 kb, although considerably longer film exposures were necessary to detect this homology than were used to detect the human genomic sequence (3). In order to explore the possibility that only a portion of the mouse sequence was homologous to the human 1.9 kb repeat, we utilized segments of the clone for hybridization. Segment D (position 403-792) hybridized to the mouse 1.5 kb Bst Nl band (Fig. 6 lane e); but there was no homology between this human segment and the mouse 1.7 kb Bst Nl band or its counterpart Eco RI 1.3 kb band. In digests of mouse DNA with Eco RI or Xba I, a smear of hybridizing higher molecular weight material was observed, and a few faint bands were visible amidst this smear in the Xba I digests (Fig. 6, lane c). For comparison, adjacent lanes run with human genomic DNA served as controls, and indicated that in order to detect the mouse homology, $\geq$ 10-fold exposure times were necessary even though the ethidium bromide visualized bands of the repeated DNAs were of equivalent intensities in the mouse and human digests. Segments E (position 792-1303) and C (130-403) hybridized only extremely faintly to the mouse 1.5 kb band, and a more 3' segment, F, showed no detectable hybridization. Thus we suspect that when the mouse 1.5 kb repeat is cloned and sequenced, it is likely to contain homologies to the human repeat, at least in the region between position 380 and 820 in the 1.9 kb sequence. Controls including nick translated pBR322 and other repeated human DNA inserts showed no hybridization to this band on equivalent or longer exposures. Furthermore the 1.5 kb mouse band (uncloned) in reverse hybridization experiments, hybridized to the 1.9 kb human genomic band (3). Such homologies could easily be overlooked especially in digests where homologous mouse sequences were not present as a single band, since human components run in the same

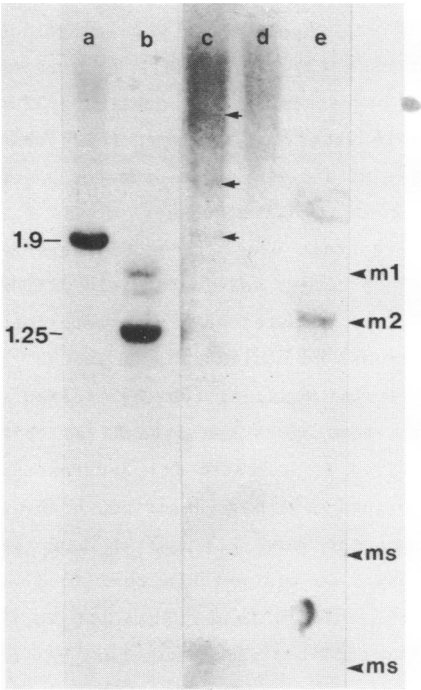

Figure 6. Human and mouse genomic DNAs (~8μg) electrophoresed on the same gel were blotted and probed with segments of the 1.9 kb cloned insert. Lanes a and b are human DNA cleaved with Hind III and Hae III respectively, hybridized with segment E (792-1303), exposed for 1 day. The Hind III 1.9 kb band and Hae III 1.25 kb band are already overexposed. This segment showed only extremely faint hybridization to mouse DNA on adjacent lanes with 2 week exposures. Lanes c,d and e are adjacent mouse DNA lanes hybridized with human cloned segment D (403-792) exposed for 8.5 days. A smear of higher molecular weight material with faint bands is detected in Xba I mouse digests (lane c); a smear of larger DNA fragments, but no discrete bands are detected in mouse Eco RI digests (lane d); lane e (Bst NI digestion) shows a discrete mouse band (m2) hybridizing to this human cloned segment. In lane e, mS are positions of mouse satellite ethidium bromide stained DNA bands of 234 and 468 bp, that also show no detectable hybridization with this probe, and ml and m2 are positions of ethidium bromide stained bands of other repeated DNAs previously estimated to be ~1.7 and 1.5 kb respectively (14). The adjacent human DNA lanes in this segment hybrid are not shown because they were grossly overexposed at times necessary to detect the homologous mouse sequences; shorter exposures with this probe were equivalent to intensities seen in a and b.

lanes could obliterate regions of mouse DNA hybridization on long exposures (see Discussion).

DISCUSSION

We have shown that 80% of the genomic family members of a repeated sequence related to a cloned copy are defined by 1.9 kb Hind III boundaries. Furthermore, most of the repeats in the genome followed the same sequence arrangement as the cloned copy as determined by segment hybridization studies; a few base changes in the cloned sequence could account for most of the genomic variants observed. From direct measurement of stained gel profiles, the 1.9 kb repeat was found to represent in the order of 0.25%-0.35%

of the total genomic DNA, which would account for roughly 3-4,000 copies of this sequence per haploid cell. These direct measurements were consistent with in-situ chromosome hybridization studies (3) where the 1.9 kb repeat signal was 10 fold less than that of Alu sequences, which constitute 3% of the genome. These estimates are compatible with, but somewhat lower than, other estimates of copy number from studies of a 6.4kb clone from the 3' end of globin DNA (15), which is likely to contain DNA sequences of the same family. The 6.4 kb cloned fragment could contain other repeated sequences that may not always be adjacent to the 1.9 kb repeat and could yield different estimates of copy number. In fact, other repeated DNAs have been found in this region 3' to the globin gene (personal communication, D. Tuan).

It is of interest that a clone containing a repeat that hybridized to the Hind III 1.9 kb band, and to the same homologous Kpn I bands as those observed here, was also independently derived in a search for X-chromosome specific markers (1). Unlike the cloned sequence we have described however, a smear of other material in addition to the Hind III 1.9 kb band was strongly detected in some autoradiographs. It was notable that in these other studies, the cloned DNA was also considerably longer than 1.9 kb (16 kb), and therefore might also contain other repeated sequences that may not always be contiguous with the 1.9 kb sequence. At least some of the 1.9 kb sequence is "interspersed" with many other DNAs, yielding a smear of larger hybridizing DNA segments in genomic digests, suggesting multiple regions of integration into diverse DNAs. Long repeated elements of >1000 bp as well as shorter ones have been shown to be interspersed with single copy sequences in a number of eukaryotic genomes (16).

In addition to the smear of DNA, common bands of >1.9 kb were also observed in hybridization studies e.g., as those obtained with Kpn I digestion. This indicates to us that the "interspersion" with other DNAs is not entirely diffuse, that is, some copies are non-randomly arranged in longer repeats. We have been able to partially map restriction sites flanking the 1.9 kb sequence in these longer arrays. The hybridizing bands in the Kpn I and other digests were also clearly less intense than the Hind III 1.9 kb band, also suggesting other DNA sequences contributed to a proportion of the higher molecular weight bands. It is therefore likely that other repeated elements may flank a fair proportion of the members of the 1.9 kb family. Our mapping studies indicate a total length of $\geq$ 5.5 kb for such longer elements. Such flanking sequences could possibly define a more specific subset of the 1.9 kb repeats with respect to their "function" or mode of

integration in the genome.

In comparing our external map sites with those of a presumably homologous sequence from the 3' end of the β globin gene (17) we find there are comparable Eco RI and Bgl II sites external to the same side of the 1.9 kb fragment. Also, as in that repeat, no internal Bam HI or Pst I sites were seen. One difference between the two maps is that the 3' globin element contains an Eco RI site at the oposite end of the 1.9 kb sequence to the one we have mapped for the predominant genomic Eco RI site here. In accord with our studies (3) these investigators have also failed to detect a major RNA transcript (15) from these sequences.

Several studies have indicated that the Hind III 1.9 kb sequence and the related Kpn I bands are conserved in primate evolution (2,15,18). In previous work we demonstrated that the cloned 1.9 kb human sequence also hybridized to a repeated DNA band in mouse DNA, and visa versa (3). The clone isolated by Schmeckpeper et al. (1), which is likely to contain sequences homologous to the repeat described here, was reported not to hybridize to mouse DNA. There are several possible explanations for this discrepancy: 1) differences in stringency of hybridization could affect the results; in this context other investigators have pointed out that high stringency conditions may fail to uncover homologous sequences in other organisms (19), 2) long autoradiographic exposures were needed to detect the hybridizing mouse Bst NI band, 3) without using Bst NI digestion (i.e., using other restriction digests), the hybridizing mouse DNA is detected only as a faint smear that could easily be overlooked without very long exposures, or 4) the clone used in these other studies does not include the part of the 1.9 kb sequence that is homologous to the mouse sequence. In our segment hybridization studies we considered the last possibility in more detail. We have shown that there is at least one region of the 1.9 kb sequence that by hybridization is homologous to the mouse Bst NI 1.5 kb band. This mouse repeat has not yet been sequenced; indeed detailed sequence comparisons of other repeats have indicated previously unsuspected homologies between rodents and man (20). The homologous segment of the 1.9 kb sequence shown here may be of greater importance in defining an essential unit of this repeat, possibly a sequence having a specific protein binding or other "functional" capacity. We have also identified a different human repeated DNA that hybridizes to the mouse 1.7 kb Bst NI band or Eco RI 1.3 kb band (3). We do not know if such sequences arose independently during the evolution of both species, or are "conserved", but the fact that homologies occur

in such divergent species suggests that these sequences may have some meaningful or common roles in both genomes.

We have shown there is no strict tandem repetition of the 1.9 kb sequence itself in the genome although there may be rearrangements within minor members of the family. Longer repeated elements of $\geq$ 5.5 kb depicted here, and also previously indicated by Adams et al. (15) apparently contain other non-identical reiterated sequences. These longer repeats could conceivably be tandemly repeated, or partially reduplicated. Elements within such long repeats could also be "scrambled" as described in other genomes (13). Excision or insertion of selected DNA segments, or segment permutations within such long repeats, could have partial site specificity, creating files of specific repeat length or of specific sequence clusters. Such files could define specific chromosomal domains.

Chromosome hybridization studies of the 1.9 kb repeat using high resolution immunological detection methods (3, and in preparation) indicate nonrandom "hot spots" of hybridization, or banding, on chromosome arms. These results are most consistent with clustering of at least some of these repeated elements, as these immunological methods do not yet allow detection of only 2 kb of DNA. In such a clustered arrangement, the 1.9 kb repeat need not be contiguously followed by itself, but could be repeated following flanking common DNAs, possibly such as those seen at common external restriction sites. We therefore suggest, based on the above data, and on chromosome hybridization studies, that at least some of the 1.9 kb repeated sequences may be represented together in an "en bloc" (3) or "intercalary" arrangement on chromosome arms, rather than in a diffuse or randomly interspersed pattern. The Hind III 1.9 kb sequence may represent a more conserved portion of such longer repeats. In contrast, the "Alu family" of shorter interspersed repeats (see 21) has a more diffuse pattern of hybridization seen on chromosome arms (3, and unpublished data). Further cloning studies of flanking segments, combined with high resolution in-situ chromosome hybridization of specific segments of each repeated element are needed to define which elements are diffusely interspersed, and which ones contibute to an "en bloc" organization.

One possibility we have considered is that the sequence described here may participate in the definition of three-dimensional chromosome order in the nucleus, and the indexing, or alignment, of specific chromosome regions (3). Unraveling the history and organization of such repeated DNAs is likely to give us more insight into the essentials of eukaryotic chromosome

architecture and the C-value paradox.

ACKNOWLEDGEMENTS

This work was supported by N.I.H. grant CA15044.

REFERENCES

1. Schmeckpeper, B.J., Willard, H.F. and Smith, K.D. (1981) Nucleic Acids Res 9, 1853-1872.
2. Maio, J.J., Brown, F.L. and Musich, P.R. (1981) Chromosoma (Berl.) 83, 103-125.
3. Manuelidis, L. (1982) In: "Genome Evolution and Phenotypic Variation" G.A. Dover and R.B. Flavell, eds., pp 263-285. Academic Press, New York.
4. Wu, J.C. and Manuelidis, L. (1980) J. Mol. Biol., 142, 363-386.
5. Manuelidis, L. (1978) Chromosoma (Berl.) 66, 23-32.
6. Rubin, C.M., Houck, C.M., Deininger, P.L., Friedmann, T. and Schmid, C.W. (1980) Nature 284, 372-374.
7. Pan, J., Elder, J.T., Duncan, C.H. and Weissman, S.M. (1981) Nucleic Acids Res. 9, 1151-1170.
8. Manuelidis, L. (1982) Nucleic Acids Res (in press).
9. Manuelidis, L. (1976) Nucleic Acids Res. 3, 3063-3076.
10. Grosveld, S., Dahl, H-K., deBoer, E. and Flavell, R.A. (1981) Gene 13, 227-237.
11. Manuelidis, L. (1979) Chromosoma (Berl.) 72, 257-269.
12. Boseley, P., Moss, T., Machler, M., Portermann, R. and Birnstiel, M. (1979) Cell 17, 19-31.
13. Wensink, P.C., Tabata, S. and Pachl C. (1979) Cell 18, 1231-1246.
14. Manuelidis, L. (1980) Nucleic Acids Res. 8, 3247-3258.
15. Adams, J.W., Kaufman, R.E., Kretschmer, P.J., Harrison, M. and Nienhuis, A.W. (1980) Nucleic Acids Res. 8, 6113-6128.
16. Christie, N.T. and Skinner, D.M. (1978) Nucleic Acids Res. 6, 781-794.
17. Kaufman, R.E., Kretschmer, P.J., Adams, J.W., Coon, H.C. and Nienhuis, A.W. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 4229-4233.
18. Maio, J.J., Brown, F.L., McKenna, W.G. and Musich, P.R. (1981) Chromosoma (Berl.) 83, 127-144.
19. Miesfeld, R., Krystal, M. and Arnheim, N (1981) Nucleic Acids Res. 9, 5931-5947.
20. Pech, M., Igo-Kemenes, T., and Zachau, H.G. (1979) Nucleic Acids Res. 7, 417-432.
21. Singer, M.F. (1981) International Review of Cytology (in press).