Isolation and characterization of a major tandem repeat family from the human X chromosome

Huntington F. Willard¹*, Kirby D. Smith² and Joanne Sutherland¹

¹Dept. Medical Genetics, Univ. Toronto, Toronto, Ontario, Canada, M5S 1A8 and ²Howard Hughes Medical Institute, Laboratory for Human Biochemical Genetics, Division of Medical Genetics, Dept. Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Received 7 January 1983; Revised and Accepted 28 February 1983

ABSTRACT

We report the identification and characterization of a family of repeated restriction fragments whose molecular organization is apparently specific to the human X chromosome. This fragment, identified as an ethidium bromide-staining 2.0 kilobase (kb) band in BamHI-digested DNA from a Chinese hamster-human somatic cell hybrid containing a human X chromosome, has been cloned into pBR325 and characterized. The 2.0 kb repeated family has been assigned to the Xp11→Xq12 region on the X by Southern blot analysis of somatic cell hybrids and is predominantly arranged in tandem clusters of up to seven 2.0 kb monomers. Homologous DNA sequences, not organized as 2.0 kb BamHI fragments, are found elsewhere on the X chromosome and on at least some autosomes, but are not found on the Y chromosome. From a dosing experiment using various amounts of the cloned repeat, we estimate that there are 5,000 - 7,500 copies of the 2.0 kb BamHI repeat per haploid genome. Since the vast majority, if not all, of these are confined to the X chromosome, this repeated DNA family must account for 5-10% of all X chromosome DNA and must constitute the major sequence component of the pericentromeric region of the X.

INTRODUCTION

Recent advances in the molecular biology of mammalian X chromosomes have provided a new dimension for understanding the genetic content of the X and for analyzing factors responsible for the unique genetic, cellular, and developmental regulation of this chromosome. In the last three years, a large number of DNA sequences from X chromosomes, in particular from the human X, have been isolated and characterized. Specific genomic DNA and cDNA probes for at least four mammalian X-linked genes have been described (1-6). In addition, numerous single copy DNA sequences of unknown function have been isolated from the human (7-11) and mouse (12) X chromosomes. These probes will be valuable tools for the analysis of the molecular basis for X chromosome inactivation (13) and should provide insight into the molecular organization of the chromosome itself.

Understanding of the structure of any higher eukaryotic chromosome is complicated by a complex array of repeated DNA sequences, whose role in genome structure or genetic regulation remains unclear. Nonetheless, the

description of families of reiterated DNA classes localized to specific regions of the human genome or to specific human chromosomes (14-18) promises to significantly aid our understanding of the molecular organization of the human genome and, in particular, the role of interchromosomal relationships and intragenomic dynamics in determining the content of specific chromosomes through evolution (19). This latter approach has been successful in illuminating certain aspects of the organization and evolutionary history of the human Y chromosome (15-17, 20, 21).

Analogous studies of DNA sequences on the human X chromosome should be equally informative. Such analyses should be particularly enlightening with respect to the possible evolutionary confinement of the X, as suggested by the marked conservation of a primordial X gene linkage group (22) and, to a lesser degree, of X chromosome structure (23). Both reiterated as well as single-copy DNA probes from the X should be extremely useful in determining whether, or the extent to which, such conservation of gene content and cytogenetic morphology can be extended to the DNA level and to non-coding X DNA sequences (24). To date, several classes of X-specific and non-X-specific reiterated DNA have been described (14, 25-28). In this report, we describe the isolation, molecular cloning, and characterization of a family of 2.0 kilobasepair (kb) repeated fragments released from the X chromosome by cleavage of genomic DNA with the restriction endonuclease BamHI. The existence of this family of repeated fragments was originally suggested by the data of Beauchamp et al. (14), and, since the completion of this work, the cloning of a member of this family has been reported by Hamkalo and her colleagues (29). They suggested by dosage analysis that the 2.0 kb spacing of BamHI sites within the repeat was specific to the human X chromosome and used in situ hybridization to assign the majority of these sequences to the pericentromeric region on the X (29).

We have identified this family of repeated sequences using a slightly different approach from Yang et al. (29). In this report, we describe the isolation of this family and its direct assignment to the human X chromosome by analysis of somatic cell hybrids. Using a cloned repeat from this family as probe, we have regionally mapped the 2.0 kb BamHI repeats to the pericentromeric or proximal long arm (Xq) region of the X by analysis of DNA from human or hybrid cell lines containing structurally abnormal human X chromosomes. Further, using the cloned repeat, we estimate that there are 5,000 - 7,500 copies of this repeat on the human X chromosome, arranged in tandem clusters of up to at least seven 2.0 kb monomers. And lastly, our data indicate considerable restriction site heterogeneity among these repeated

fragments and suggest that some members of this family may exist as solitary elements dispersed from the tandem clusters. A preliminary account of this work has appeared (28).

MATERIALS AND METHODS

Cell Lines

Human skin fibroblasts from normal males and females and from individuals with abnormal karyotypes (45,X; 47,XXX; 49,XXXXX (from S. Funderburk); and 46,Xi(Xq) (from W.R. Breg)) were maintained as described (27). A Chinese hamster-human somatic cell hybrid (designated c12D) containing only a single human chromosome, the X, and the hamster parental fibroblast line wg3H were provided by S. Goss (30). A mouse-human somatic cell hybrid (designated AHA-11a) containing only a single human chromosome, the X, and the mouse parental line A9 were provided by F. Ruddle (31). AHA-11aAz is an azaguanine-resistant derivative of AHA-11a which no longer contains a human X chromosome (27). Y15/1 is a mouse-human somatic cell hybrid containing multiple human Y chromosomes, but no other human chromosomes, and was kindly provided by M. Marcus (32). c12DAz1a is an azaguanine-resistant subclone of c12D, selected in our laboratory, which is missing human phosphoglycerate kinase activity, human α -galactosidase A activity, and human glucose-6-phosphate dehydrogenase activity, but which continues to express human steroid sulfatase activity and contains the Xpter $\rightarrow$ Xq12 portion of the human X translocated to a hamster chromosome (H. Willard and M. Tzetis Holmes, unpublished observations). DNA from two additional mouse-human somatic cell hybrids, FA9-B2 and FA9-C1T3, was provided by R. Worton. These hybrids were formed between mouse A9 cells and human fibroblasts carrying a balanced X;21 translocation (46,X,t(X;21)(p21;p12)). FA9-B2 contains the human der(X) chromosome with Xp11 $\rightarrow$ Xqter. FA9-C1T3 contains the human der(21) with Xpter $\rightarrow$ Xp21 (R. Worton, personal communication).

Purification and Radiolabelling of DNA

Nuclear DNA from cultured cells or from male and female peripheral blood was prepared as described (27). Where required, fragments of DNA were purified from Seaplaque agarose gels (Marine Colloids Division) by minor modification of the procedure of Higuchi et al. (33). DNA was radiolabelled with [α -³²P]dNTP's by nick-translation with E. coli DNA polymerase I (34). Specific activities of $1-2 \times 10^8$ cpm/ μ g were routinely obtained.

Enzyme Digestions, Gel Electrophoresis, and Analysis of DNA

Restriction endonuclease digestions of genomic DNA or plasmid DNA were

performed essentially as described (27). DNA fragments were separated by agarose gel electrophoresis, stained, photographed, and transferred to nitrocellulose filters (27,35). Molecular weight standards were HindIII digests of λ DNA and HaeIII digests of ϕ X 174 DNA.

After transfer, filters were prehybridized at 37° as described (27) and hybridized under one of two sets of conditions. When gel-purified fragments from BamHI digests of genomic DNA were used as probe, the conditions were: 2 x SSC; 70% formamide; 0.2% each polyvinylpyrrolidone, Ficoll, and bovine serum albumin; 50mM Na-PO₄ buffer (pH 6.8); 1% glycine; 250 μ g/ml salmon sperm DNA; and probe at $5 \times 10^5 - 10^6$ cpm/ml; for 48-96 hr. at 37°. When cloned DNA was used as probe, the conditions were: 3 x SSC; 50% formamide; 10% dextran sulfate (36); 0.2% each polyvinylpyrrolidone, Ficoll, and bovine serum albumin; 50mM Na-PO₄ (pH 6.8); 1% glycine; 250 μ g/ml salmon sperm DNA; and probe at $1-3 \times 10^6$ cpm/ml; for 12-16 hr. at 37°. After hybridization, filters were washed at 65° as described (27, 37). In most experiments, additional washing was performed in 0.1 x SSC, 0.1% SDS at 65°. Autoradiography was performed by exposure to Kodak XAR-5 film with a Dupont Cronex intensifying screen.

Molecular Cloning and Identification of Recombinant Plasmids

DNA from a 49,XXXXX fibroblast line was used as source of 2.0 kb fragment for cloning. After BamHI digestion and electrophoresis, the ethidium bromide-staining band at 2.0 kb was visualized by long-wave ultraviolet illumination, cut out, and recovered (33). The DNA fragments were ligated to BamHI-digested pBR325 DNA, and used to transform E. coli strain HB101 by the calcium chloride procedure (38). Chloramphenicol-resistant, tetracycline-sensitive recombinants were selected and rapid DNA preparations made by modification of the method of Crouse et al. (38). Plasmid DNA was digested with BamHI to confirm the presence of a 2.0 kb insert. Plasmids were then screened in a blot experiment for the presence of an insert from the X by hybridizing in the presence of excess unlabelled hamster DNA with a ³²P-labelled crude probe prepared from BamHI-digested c12D DNA. Previous experiments (see Fig. 1) had demonstrated that this crude genomic probe could be used to detect specifically 2.0 kb repeated fragments from the human X chromosome. Approximately 20% of the recombinants were positive by this screen; two, designated pBamX5 and pBamX7, were selected for further study and were shown in subsequent experiments (see Results) to be representative of this X-linked DNA family. Large-scale plasmid DNA preparations were made by modification of the method of Pulleyblank et al. (39).

Estimated Copy Number of the 2.0 kb Repeat

To estimate the number of copies of the 2.0 kb BamHI repeat on the X, various female human DNAs and plasmid DNAs were digested with BamHI, and a Southern blot experiment performed with dilutions of plasmid DNA. In order to accurately determine DNA content, samples were first extensively digested with RNAase T1 and precipitated with ethanol. Gel analysis confirmed absence of RNA in the plasmid DNAs and DNA concentrations were determined from the absorbance at 260 nm. The amount of human DNA used was 5 µg. Dilutions of plasmid DNA corresponded to the amount expected for approximately 500–10,000 copies of the repeat per human haploid genome (3×10^9 bp), corrected for the fact that only 26% of the plasmid is human insert (2.0 kb/7.6 kb). For example, the amount of BamHI-digested plasmid DNA corresponding to 5000 copies of a 2000 bp fragment per haploid genome in a genomic DNA sample of 5 µg is:

$$(5000 \text{ copies} \times 2000 \text{ bp} / 3 \times 10^9 \text{ bp}) \times 5000 \text{ ng} \times (7.6 \text{ kb} / 2.0 \text{ kb}) = 63 \text{ ng}.$$

As an internal standard, each gel contained commercially obtained HindIII-digested λ DNA of known amount and concentration. The 1.96 kb λ band was conveniently used for this comparison. In a control experiment, 5 µg of BamHI-digested mouse A9 DNA was added to plasmid DNA dilutions before electrophoresis to evaluate potential differences in transfer efficiency between genomic DNA samples and purified plasmid DNA preparations. No differences were noted for the 2.0 kb band under our conditions.

RESULTS

Identification and Characterization of the Restriction Fragment

The reiterated DNA fragment was initially detected as a distinct ethidium bromide-staining band after digestion of genomic DNA from a Chinese hamster-human somatic cell hybrid, c12D, with BamHI; the 2.0 kb band was absent from a similar digest of Chinese hamster DNA. Since the only human chromosome in c12D is the X, we concluded that this 2.0 kb repeated fragment was probably from the human X chromosome.

To demonstrate this directly, this fragment was cut out of a gel and the DNA recovered. The resulting crude probe, containing Chinese hamster DNA in addition to the fragment of interest, was radiolabelled and hybridized to a Southern blot of various genomic DNAs digested with BamHI. As shown in Fig. 1A, the probe, as expected, hybridized very strongly to both Chinese hamster DNA and c12D DNA, but hybridized differentially to a 2.0 kb band in c12D. This probe also hybridized with human genomic DNAs from a normal male (46,XY) and a 47,XXX female. Although there was some hybridization to higher molecular

Figure 1 - Identification of a reiterated DNA fragment from the human X chromosome. A: Approximately equal amounts of Chinese hamster wg3H DNA (CH), c12D DNA (2D), human 46,XY DNA (X), and human 47,XXX DNA (XXX) were digested with BamHI and a Southern blot experiment performed with ^{32}P -labelled crude probe prepared from c12D. B: Same as A, except that hybridization conditions included 150 $\mu\text{g}/\text{ml}$ sheared unlabelled Chinese hamster DNA as competitor.

weight DNA of approximately equal intensity in the two digests, hybridization at 2.0 kb was significantly more intense in the sample with three X chromosomes than in the sample with a single X, consistent with a portion of the crude probe being homologous to human X chromosomal DNA (Fig. 1A).

Because the probe in this experiment contained contaminating hamster DNA, the hybridization was repeated in the presence of a large excess of unlabelled genomic hamster DNA. As shown in Fig. 1B, the presence of competing unlabelled DNA prevented virtually all of the hybridization to BamHI-digested hamster DNA, while a strong signal at 2.0 kb remained in the c12D and human digests. Additional, less intense bands were observed at 4.0 kb and occasionally at 3.0 kb. From these data, we conclude that the probe DNA contained reiterated DNA sequences from the human X chromosome organized primarily as repeating 2.0 kb BamHI fragments.

The increased hybridization to 47,XXX human DNA indicated that these repeated fragments might be restricted to the human X chromosome. Densitometric measurements provided support for this (Fig. 2A,B), since the ratio of the peak heights as well as peak areas was precisely proportional to the number of X chromosomes in each sample. This result suggests that the large majority, if not all, of the 2.0 kb BamHI repeats in the human genome

Figure 2 - X-specificity of 2.0 kb BamHI repeated fragment. A: Equal amounts of human 46,XY (lane 1) and 47,XXX (lane 2) DNAs were digested with BamHI and a Southern blot experiment performed as in Fig. 1B. Arrow indicates 2.0 kb band. B: Densitometric tracings of 2.0 kb region of lanes 1 and 2 in A. Tracings were performed on a Joyce-Loebl densitometer. C: As in A, except that DNAs are 45,X (lane 1), 47,XXX (lane 2), and 46,Xi(Xq) (lane 3). Only a portion of the autoradiogram is shown. Arrow indicates 2.0 kb band.

are, in fact, located on the X chromosome.

To determine whether these repeated fragments are distributed uniformly along the length of the X, equal amounts of DNA from 45,X, 47,XXX and 46,Xi(Xq) individuals were digested with BamHI, transferred after electrophoresis, and hybridized with the crude probe. The last sample contains one copy of the X short arm (Xp) and three copies of the X long arm (Xq) in each cell. Hybridization of the probe to the 46,Xi(Xq) DNA was approximately equivalent to the 47,XXX sample and significantly greater than to the 45,X DNA (Fig. 2C). These data suggest that the majority of these 2.0 kb fragments must be restricted to the duplicated portion (Xcen → Xqter) of the isochromosome.

X-linkage of the Cloned Repeat and Regional Localization on the X

The 2.0 kb repeated fragment was gel-purified from a BamHI digest of human 49,XXXXX DNA and cloned into the BamHI site of pBR325 (see Materials and Methods). Two of the resulting plasmids, pBamX5 and pBamX7, were selected for use in this study. A partial restriction map of the 2.0 kb BamHI insert in pBamX7 is given in Fig. 3. The plasmid insert is not cleaved by EcoRI, HindIII, KpnI, BglII, HpaII, or SalI. Of note are single sites for PstI and SstI, either of which can be used instead of BamHI to define this class of 2.0 kb repeats in genomic digests (see below). In addition to the sites shown, there are multiple HinfI sites at 170 bp and 340 bp intervals and HaeIII sites at 340 bp and 680 bp intervals (in addition to end fragments); the precise location of these sites has not been determined. The restriction map of the other pBamX plasmid examined in detail is similar, except for the absence of

Figure 3 - Partial restriction map of the insert in pBamX7, as determined by single, double, and partial enzyme digests of the isolated 2.0 kb insert. B=BamHI. H=HhaI. P=PstI. R=RsaI. Sa=SauIII A. S=SstI. T=TaqI.

one HinfI site, the middle RsaI site, and the righthand (as shown) HhaI site.

When either pBamX5 or pBamX7 was radiolabelled and hybridized to Southern blots of BamHI-digested human genomic DNAs with different numbers of X chromosomes, the data were similar to those obtained with the crude probe (see Fig. 4 for pBamX5; data for pBamX7 not shown). Distinct bands of hybridization were noted at 2.0, 3.0, 4.0 and 6.0 kb in addition to a higher molecular weight smear. The intensity of hybridization at the bands, but not the smear, increased proportionately with the number of X chromosomes (Fig. 4). (In some experiments, the 3.0 kb band did not appear to dose, but in such digests, the band at 6.0 kb (a presumptive dimer of the 3.0 kb fragment; see below) was accentuated.) On longer gels (Fig. 4, right), an additional band of hybridization was noted at 1.65 kb, which also showed X dose dependence. To confirm the X-linkage of the insert in pBamX5, the plasmid was radiolabelled and hybridized to a Southern blot of BamHI digests of genomic DNA from several human, rodent, and somatic cell hybrid lines (Fig. 5). The probe did not hybridize to mouse DNA or hamster DNA, but did hybridize to DNA from hybrids

Figure 4 - Southern blot analysis of human DNAs using the cloned X repeat as probe. Equal amounts of human 46,XY and 49,XXXXX (left) and of human 45,X and 47,XXX (right) DNA were digested with BamHI and a blot experiment performed with ³²P-labelled pBamX5 as probe. Arrows indicate specific bands of hybridization that dose in an X-dependent fashion. Estimated band sizes are given in kb on the right.

Figure 5 - Southern blot analysis of somatic cell hybrid DNAs. DNAs were digested with BamHI and a blot experiment performed with 32 P-labelled pBamX5 as probe. DNAs are human 49,XXXXX (lane 1), mouse A9 (lane 2), hybrid AHA-11a (contains complete human X) (lane 3), hybrid AHA-11aAz (no human X) (lane 4), hybrid c12D (complete human X) (lane 5), hamster wg3H (lane 6), hybrid c12DAz1a (Xpter $\rightarrow$ Xq12) (lane 7), hybrid Y15/1 (no X, but multiple Y's) (lane 8), hybrid FA9-C1T3 (Xpter $\rightarrow$ Xp21 only) (lane 9), and hybrid FA9-B2 (Xp11 $\rightarrow$ Xqter only) (lane 10). Arrow indicates band at 2.0 kb. At the right is a schematic representation of the human X chromosome, indicating the three band regions involved in translocations. 'Xpter' refers to the terminus of the short arm. 'Xqter' refers to the terminus of the long arm. The total length of the human X chromosome is 1.5×10^8 bp.

AHA-11a and c12D, both of which contain a human X. DNA from AHA-11aAz, which no longer contains a human X, did not hybridize the probe (Fig. 5). Other somatic cell hybrids, containing various human autosomes, but no human X (data not shown), or multiple human Y chromosomes, but no X (Fig. 5), also didn't bind the probe. Thus, the data indicate that the cloned repeat in pBamX5 (and pBamX7) is indeed a member of the family of X-linked sequences characterized above with an uncloned genomic probe and strengthen the conclusion based on experiments with the genomic probe that the 2.0 kb repeat family (including the related 3.0, 4.0, and 1.65 kb fragments; data derived from longer exposures of the experiment in Fig. 5) is confined largely, if not entirely, to the X chromosome.

The data in Fig. 2C indicated that the 2.0 kb repeated fragments were located predominantly at the X centromere or on Xq. To further regionally localize these sequences, pBamX5 was hybridized to Southern blots of BamHI-digested DNAs from hybrids containing only a portion of the X. As shown in Fig. 5, the probe hybridized to DNAs containing Xpter $\rightarrow$ Xq12 and Xp11 $\rightarrow$ Xqter,

Figure 6 - Partial BamHI digests of c12D DNA. DNA was digested with 1 unit enzyme/ μ g DNA for 10-60 mins (right to left). A Southern blot experiment was performed with 32 P-labelled pBamX5 as probe. Band sizes are indicated in kb.

but not to DNA containing Xpter $\rightarrow$ Xp21. Therefore, by this approach we can assign at least the majority of the fragments to Xp11 $\rightarrow$ Xq12. Combined, the data in Figs. 2 and 5 indicate that the organization of these DNA sequences is largely X-specific and that they can be localized primarily to the pericentromeric or proximal long arm region of the X (Xcen $\rightarrow$ Xq12). The data do not exclude the possibility of smaller numbers of repeats elsewhere on the long arm of the X.

Tandem Clustering of the 2.0 kb Repeats on the X

The appearance of a specific band of hybridization at 4.0 kb after BamHI digestion of either human DNAs or hybrid DNAs (Figs. 1, 4, 5) suggested the possibility that some of the 2.0 kb repeats on the X might be tandemly arranged. To examine this possibility directly for the X, c12D DNA was subjected to a series of partial BamHI digestions. The resulting fragments were separated by electrophoresis, transferred to nitrocellulose, and hybridized to 32 P-labelled pBamX5, and the autoradiogram was examined for the generation of a ladder of fragments corresponding to multimers of the 2.0 kb fragment (Fig. 6). Indeed, hybridizing fragments from 4.0 kb up to 14 kb in size at 2 kb increments were clearly detected, superimposed on a background smear of hybridization. In addition, the distinct band at approximately 3.0 kb was seen; however, no band at 5 kb (or at 7 or 9 kb etc.) was detected. The data therefore indicate that the 2.0 kb fragments are, in part, organized in clusters of tandem arrays of up to, and perhaps more than, seven repeating

Figure 7 - Estimation of number of 2.0 kb BamHI fragments per human haploid genome. Two different human 46,XX DNAs and pBamX7 DNA were digested with BamHI and a Southern blot experiment performed with ^{32}P -labelled pBamX7 as probe. Only portion of autoradiogram at 2.0 kb is shown. Amounts of DNA are: 46,XX, 5 μg (lanes 1 and 2); and pBamX7, 6.3 ng, 12.6 ng, 31.5 ng, 63.0 ng, 94.5 ng, and 126.0 ng (lanes 3 - 8, respectively). DNA in lanes 3 - 8 corresponds to approximately 500, 1000, 2500, 5000, 7500, and 10,000 copies of 2.0 kb fragment per haploid genome. Similar results were obtained when pBamX5 was used. See Materials and Methods for additional details.

units on the X and show further that the 3.0 kb fragments are not, in any major way, interspersed with the homologous 2.0 kb repeats. Dimers of the 3.0 kb fragment, however, may contribute to the band seen at 6.0 kb. We have no information on the genomic organization of the 1.65 kb band.

Estimated Copy Number of the 2.0 kb Repeated Fragment

To estimate the number of copies of the 2.0 kb BamHI repeat on the X, female human DNA and pBamX7 plasmid DNA were both digested with BamHI, and a Southern blot experiment was performed with dilutions of plasmid DNA corresponding to the amount of 2.0 kb insert expected for approximately 500 to 10,000 copies of the fragment per haploid genome (i.e., per human X) (see Materials and Methods). From the experiment shown in Fig. 7, we estimate that there are approximately 5,000 to 7,500 copies of a 2.0 kb repeat homologous to pBamX7 per haploid genome.

Organizational Heterogeneity Among Members of the Family

To gain insight into the organization of these repeated sequences in the human genome, female DNA was digested with a variety of enzymes and a blotting experiment performed with ^{32}P -labelled pBamX5 or pBamX7. Results with Dde I, HaeIII, Mbo I, and HinfI indicated the existence of a ladder of hybridizing fragments based on multiples of 170 basepairs, in agreement with the restriction map in Fig. 3. Sequences in this 2.0 kb repeat, therefore, are probably related to other families of reiterated DNA described previously in the human and other primate genomes. (See discussion in ref. 29).

With some enzymes, however, hybridization to bands or areas other than the predominant ones was detected after longer exposure, indicating a subset of the sequences in the human genome with a different organization. In BamHI digests,

Figure 8 - Southern blot analyses of genomic DNAs with pBamX5. Sizes marked are indicated in kb. A: 46,XX DNA digested with BamHI (lane 1); c12D DNA with BamHI (lane 2); 46,XX DNA with PstI (lane 3). 1.5% agarose gel. B: 46,XY DNA with BamHI (lane 4); 46,XX DNA with BamHI (lane 5); c12D DNA with BamHI (lane 6), EcoRI (lane 7), and both BamHI and EcoRI (lane 8); c12D DNA with PstI (lane 9). 1% agarose gel. Arrow in B indicates the 1.6 kb band in lanes 4 and 6.

for example, much of the hybridization was to an ill-defined high molecular weight smear in addition to the previously noted bands at 2.0, 3.0 and 4.0 kb. This high molecular weight subset was particularly noticeable on higher percentage agarose gels (e.g. Fig. 8A). Interestingly, this pattern was not detected in BamHI-digested c12D DNA (Fig. 8A, lane 2), indicating a difference in the major location of BamHI sites in and around this family of DNA sequences on the X and on autosomes. As noted above, a smaller band of approximately 1.65 kb was also detected in some BamHI-digested DNAs (Fig. 8A, lane 2; Fig. 8B, lanes 4, 6). This band must represent an alternate arrangement of some BamHI sites on the X. This 1.65 kb band has been seen in a number of males and females and in the two hybrids, c12D and AHA-11a. This band has been absent from only a single female thus far (Fig. 8A, lane 1 and Fig. 8B, lane 5).

The enzymes PstI and SstI also recognize a fundamental 2.0 kb repeat on the X chromosome. However, in addition to bands at 2.0, 3.0 and 4.0 kb, and to a high molecular smear in PstI-digested human DNA, smaller minor bands are also

clearly evident, again indicating a subset of the members of this repeated DNA family with different restriction site organization (Fig. 8A, lane 3). It is perhaps of note that the four additional bands generated after PstI digestion are all multimers of 170 basepairs (0.68, 1.02, 1.19, and 1.36 kb). At least some of these bands can be localized to the X chromosome, since they are detected in PstI-digested c12D DNA (Fig. 8B, lane 9).

Finally, the data in Fig. 8B indicate that EcoRI primarily cleaves outside of the tandem repeat on the X, since most of the hybridization in digested c12D DNA is to high molecular weight DNA. However, a ladder of minor bands is generated above 2.0 kb, indicating regular spacing of EcoRI sites within a minority of this repeated family on the X. This conclusion is confirmed by the generation of new, smaller bands after double digestion with both BamHI and EcoRI (Fig. 8B, lane 8).

DISCUSSION

We have reported the isolation, molecular cloning, and initial characterization of a family of repeated DNA sequences whose organization, as defined by the presence of regularly spaced BamHI sites 2.0 kb apart, is largely specific to the human X chromosome. This repeated family has also recently been described by Hamkalo and her colleagues (29). Both our fragment and theirs are presumably related to a similar sequence first described by Beauchamp et al. as a component of satellite III DNA homologous to the human X chromosome in a somatic cell hybrid (14). In addition to confirming all of the essential findings of Yang et al. (29), we have demonstrated directly that these repeated sequences are, in fact, located on the human X chromosome, that they are tandemly repeated on the X in long arrays, and that the sequences map to the pericentromeric or proximal long arm region of the X. In addition, we have been able to estimate directly the number of copies of this repeat on the X chromosome. The repetitive segment of DNA on the X homologous to our cloned probes has been formally designated DXZ1 (Sixth International Workshop on Human Gene Mapping; see ref. 40).

The 2.0 kb, 3.0 kb, and 4.0 kb BamHI fragments are for the most part confined to the human X, based both on dosage experiments and analysis of DNA from somatic cell hybrids. The sequences are, however, apparently homologous to other autosomal sequences, although there is no detectable homology with the human Y chromosome. The organization of these homologous autosomal sequences is thus different from the X-linked fragments, although we cannot rule out the possibility that a small proportion of the 2.0 kb fragments, perhaps accounting

for 10% of the total, is located on one or more autosome.

In a dosage experiment, using the cloned repeat, we estimated that there are between 5,000 and 7,500 copies of the 2.0 kb repeat per haploid human genome (Fig. 7). This number is a minimum estimate, since it doesn't include related fragments at 3.0 kb or 4.0 kb and wouldn't detect any 2.0 kb fragments having significantly diverged DNA sequences. However, this estimate is in good agreement with that based on ethidium bromide staining. Thus, since we estimate that at least 90% of these fragments are located on the X and since the X chromosome contains approximately 1.5×10^8 bp of DNA, this family of repeated 2.0 kb fragments must account for between 5% and 10% of all human X chromosome DNA.

A class of DNA sequences repeated many thousand times on the human X chromosome was not identified in an earlier study of Schmeckpeper et al. (26). That study identified two classes of human-specific reiterated DNA on the X, one repeated 20-30 times per haploid genome and confined to the X and one repeated several thousand times per haploid genome, with proportional distribution on the X and autosomes (26). That the family of 2.0 kb repeated fragments presented here does not apparently fall into either of these previously described classes of sequence elements can likely be explained by the different hybridization criteria and methods used. Under the fairly stringent filter hybridization and washing procedures employed here, the 2.0 kb BamHI repeat does not hybridize to rodent DNA (e.g., Fig. 5). Under reduced stringency, however, this family of sequences is not human-specific (unpublished observations) and thus would have been excluded from the analysis of Schmeckpeper et al. (26). Additionally, sequences in the 2.0 kb BamHI family may self-reassociate at a rapid rate and thus may have been eliminated from that study.

We have been able to localize at least the majority of the members of the DXZ1 repeated family to a small region on the X. Our assignment to the pericentromeric or proximal Xq portion of the X, based on analysis of somatic cell hybrids containing portions of the X and on dosage analysis of a cell line with an X isochromosome (Figs. 2C, 5), is fully consistent with the in situ hybridization data of Yang et al. (29), but provides direct demonstration that the repeated sequences clustered at the X centromere are, in fact, members of the family characterized by the 2.0 kb repeat unit and not homologous sequences with a different organization (i.e., the high molecular weight smear). Neither our data nor theirs, however, exclude the possibility that smaller numbers of this repeat might be localized elsewhere on the X. An important conclusion

from our study is that the DXZ1 family, since it comprises 5 - 10% of all X chromosome DNA, must constitute the major sequence component of the centromeric and/or proximal long arm portion of the X chromosome. This finding has significant implications for consideration of gene function in this region of the X and for models regarding the putative X chromosome 'inactivation center', which has provisionally been mapped to this region (41).

Yang et al. (29) suggested the possibility that the 2.0 kb fragments were tandemly repeated in the human genome. Our data demonstrate this directly for the X specifically and indicate tandem arrays of up to, and perhaps more than, seven copies of the 2.0 kb monomer (Fig. 6). However, since not all hybridization occurred in multimers of the 2.0 kb unit, we conclude that at least some members of this repeated family are not so clustered and must be adjacent to non-homologous DNA on the X (i.e., adjacent to BamHI sites located at heterogeneous intervals). The minor population of 3.0 kb repeats does not appear to be interspersed with 2.0 kb repeats, at least on the X (Fig. 6), although the two fragments do map to the same pericentromeric region (Fig. 5). Further indication of such organizational heterogeneity was provided by our demonstration of distinct subsets of this family whose spacing of restriction sites (BamHI, PstI, or EcoRI; Fig. 8) was different from the majority. Such minority members of this family are formally analogous to the 'orphons' described by Childs et al. as solitary genetic elements derived from tandem multigene families (42). Whether these variant family members show other properties attributed to 'orphons' (e.g., physical dispersion elsewhere on the chromosome or in the genome and heterogeneity among individuals) remains to be determined. Further consideration of the degree of sequence and restriction site divergence within this highly repeated family will require analysis of isolated copies of the repeat from genomic libraries of cloned X DNA. Such studies are in progress and should provide insight into both the molecular organization and evolutionary history of the human X chromosome.

ACKNOWLEDGEMENTS

We thank B. Schmeckpeper for helpful comments and various colleagues mentioned in the text for providing cell lines. H.F.W. thanks M. Tzetis Holmes and V. E. Powers for expert technical assistance, R. Worton for providing DNAs, and E. Chan-Erickson for typing the manuscript. This work was supported in part by grants from the National Institutes of Health (K.D.S.) and the Medical Research Council of Canada (H.F.W.). During a portion of this work, H.F.W. was funded by National Institutes of Health Postdoctoral Fellowship GM07471 at Johns Hopkins University.

*To whom reprint requests should be sent

REFERENCES

1. Persico, M.G., Toniolo, D., Nobile, C., d'Urso, M., and Luzzatto, L. (1981) *Nature* 294, 778-780.
2. Brennand, J. Chinault, A.C., Konecki, D.S., Melton, D.W., and Caskey, C.T. (1982) *Proc. Natl. Acad. Sci.* 79, 1950-1954.
3. Jolly, D.J., Esty, A.C., Bernard, H.U., and Friedmann, T. (1982) *Proc. Natl. Acad. Sci.* 79, 5038-5041.
4. Choo, K.H., Gould, K.G., Rees, D.J.G., and Brownlee, G.G. (1982) *Nature* 299, 178-180.
5. Kurachi, K., and Davie, E.W. (1982) *Proc. Natl. Acad. Sci.* 79, 6461-6464.
6. Michelson, A.M., Markham, A., and Orkin, S.H. (1983) *Proc. Natl. Acad. Sci.* 80, 472-476.
7. Wolf, S.F., Marenzi, C.E., and Migeon, B.R. (1980) *Cell* 21, 95-102.
8. Davies, K.E., Young, B.D., Elles, R.G., Hill, M. E. and Williamson, R. (1981) *Nature* 293, 374-376.
9. Kunkel, L.M., Tantravahi, U., Eisenhard, M., and Latt, S.A. (1982) *Nucl. Acids Res.* 10, 1557-1578.
10. Schmeckpeper, B.J., Willard, H.F., and Smith, K.D. (1983), in preparation.
11. Bruns, G.A.P., Gusella, J.E., Keys, C., Leary, A.C., Housman, D., and Gerald, P.S. (1983), in preparation.
12. Disteché, C.M., Kunkel, L.M., Lojewski, A., Orkin, S.H., Eisenhard, M., Sahar, E., Travis, B., and Latt, S.A. (1982) *Cytometry* 2, 282-286.
13. Lyon, M.F. (1972) *Biol. Rev.* 47, 1-35.
14. Beauchamp, R.S., Mitchell, A.R., Buckland, R.A., and Bostock, C.J. (1979) *Chromosoma* 71, 153-166.
15. Cooke, H. (1976) *Nature* 262, 182-186.
16. Kunkel, L.M., Smith, K.D., and Boyer, S.H. (1976) *Science* 191, 1189-1190.
17. Cooke, H.J., and Hindley, J. (1979) *Nucl. Acids Res.* 6, 3177-3197.
18. Manuelidis, L. (1978) *Chromosoma* 66, 23-32.
19. Dover, G.A., and Flavell, R.B. (eds.) (1982) *Genome Evolution*, Academic Press, London.
20. Cooke, H.J., and McKay, R.D.G. (1978) *Cell* 13, 453-460.
21. Kunkel, L.M., and Smith, K.D. (1982) *Chromosoma* 86, 209-228.
22. Ohno, S. (1967) *Sex Chromosomes and Sex-Linked Genes*, Springer-Verlag, New York.
23. Ohno, S., Becak, W., and Becak, M.L. (1964) *Chromosoma* 15, 14-22.
24. Willard, H.F. (1983) in *Cytogenetics of the Mammalian X Chromosome*, Sandberg, A.A., ed., Alan R. Liss, Inc., New York, in press.
25. Balazs, I., Szabo, P., and Siniscalco, M. (1978) *Somat. Cell Genet.* 4, 617-631.
26. Schmeckpeper, B.J., Smith, K.D., Dorman, B.P., and Ruddle, F.H. (1979) *Proc. Natl. Acad. Sci.* 76, 6525-6528.
27. Schmeckpeper, B.J., Willard, H.F., Smith, K.D. (1981) *Nucl. Acids Res.* 9, 1853-1872.
28. Willard, H.F., and Smith, K.D. (1981) in *Human Gene Mapping 6: Sixth International Workshop on Human Gene Mapping*, pp.327-328, Birth Defects: Original Article Series 18, 2, 1982, March of Dimes Birth Defects Foundation, White Plains, New York.
29. Yang, T.P., Hansen, S.K., Oishi, K.K., Ryder, O.A., and Hamkalo, B.A. (1982) *Proc. Natl. Acad. Sci.* 79, 6593-6597.
30. Goss, S.J., and Harris, H. (1977) *J. Cell Sci.* 25, 17-37.
31. Dorman, B.P., Shimizu, N., and Ruddle, F.H. (1978) *Proc. Natl. Acad. Sci.* 75, 2363-2367.
32. Marcus, M., Tantravahi, R., Dev, V.G., Miller, D.A. and Miller, O.J. (1976) *Nature* 262, 63-65.

-
33. Higuchi, R., Stang, H.D., Browne, J.K., Martin, M.O., Huot, M., Lipeles, J., and Salser, W. (1981) *Gene* 15, 177-186.
 34. Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
 35. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
 36. Wahl, G.M., Stern, M., and Stark, G.R. (1979) *Proc. Natl. Acad. Sci.* 76, 3683-3687.
 37. Peden, K., Mounts, P., and Hayward, S.H. (1982) *Cell* 31, 71-80.
 38. Crouse, G.F., Frischauf, A., and Lehrach, H. (1983) *Meth. in Enzym.*, in press.
 39. Pulleyblank, D., Michalek, M., Daisley, L., and Glick, B. (1983) *Mol. Biol. Rep.*, in press.
 40. Report of the committee on human gene mapping by recombinant DNA techniques (1981) in *Human Gene Mapping 6: Sixth International Workshop on Human Gene Mapping*, pp. 194-204, *Birth Defects: Original Article Series* 18, 2, 1982, March of Dimes Birth Defects Foundation, White Plains, New York.
 41. Therman, E., Sarto, G.E., and Patau, K. (1974) *Chromosoma* 44, 361-366.
 42. Childs, G., Maxson, R., Cohn, R.H., and Kedes, L. (1981) *Cell* 23, 651-663.