Volume 13 Number 8 1985 Nucleic Acids Research

Chromosome-specific alpha satellite DNA: nucleotide sequence analysis of the 2.0 kilobasepair repeat from the human X chromosome

John S.Waye and Huntington F.Willard*

Department of Medical Genetics, University of Toronto, Toronto, Ontario, M5S 1A8, Canada

Received 8 February 1985; Accepted 29 March 1985

ABSTRACT

The pericentromeric region of the human X chromosome is characterized by a tandemly repeated family of 2.0 kilobasepair (kb) DNA fragments, initially revealed by cleavage of human DNA with the restriction enzyme BamHI. We report here the complete nucleotide sequence of a cloned member of the repeat family and establish that this X-linked DNA family consists entirely of α satellite DNA. Our data indicate that the 2.0 kb repeat consists of twelve α satellite monomers arranged in imperfect, direct repeats. Each of the αX monomers is approximately 171 basepairs (bp) in length and is 60-75% identical in sequence to previously described primate α satellite DNAs. The twelve αX monomers are 65-85% identical in sequence to each other and are organized as two adjacent, related blocks of five monomers, plus an additional two monomers also related to monomers within the pentamer blocks. Partial nucleotide sequence of a second, independent copy of the 2.0 kb BamHI fragment established that the 2.0 kb repeat is, in fact, the unit of amplification on the X. Comparison of the sequences of the twelve αX monomers allowed derivation of a 171 bp consensus sequence for α satellite DNA on the human X chromosome. These sequence data, combined with the results of filter hybridization experiments of total human DNA and X chromosome DNA, using subregions within the 2.0 kb repeat as probes, provide strong support for the hypothesis that individual human chromosomes are characterized by different α satellite families, defined both by restriction enzyme periodicity and by chromosome-specific primary sequence.

INTRODUCTION

The α satellite repetitive DNA family of primates consists of tandemly repeated units of approximately 170 basepairs (bp) located at the centromeric regions of chromosomes. In lower primates, the bulk of α satellite DNA apparently exists as tandem units of 172 bp (monomers) or 344 bp (dimers) (1-6). In contrast to the relatively simple restriction enzyme site periodicities of lower primate α satellite DNA, the genomes of lesser apes, great apes, and man appear to contain several α satellite families, as defined by different restriction enzyme patterns and degrees of relatedness to various α satellite hybridization probes (3,6). One prominent human α satellite family, the 340 bp α dimer defined by regularly spaced EcoRI sites,

has been characterized extensively by Manuelidis and her colleagues (7-10). However, the apparent increased complexity and diversity of α satellite which accompanied the evolutionary history of higher primates limits the extent to which conclusions drawn from characterization of a single molecular organization (α dimers) can be applied to human α satellite in general. Indeed, our recent investigation of the organization of α satellite DNA on individual human chromosomes, using cloned α satellite probes and rodent/human somatic cell hybrids containing single human chromosomes, suggests that much, if not most, of the α satellite DNA in the human genome is present in molecular configurations other than the prototype α dimer (11). Our data support the hypothesis that individual human chromosomes may each be characterized by specific and different subsets of α satellite DNA, defined both by restriction enzyme periodicity and by primary sequence (6,11).

One subset of human α satellite DNA, independently isolated and characterized by ourselves (12) and by Yang et al. (13), exists as a 2.0 kilobasepair (kb) repeated unit, approximately 5000 copies of which are located at the pericentromeric region of the human X chromosome. In this report we present the complete nucleotide sequence of a cloned member of this X-linked DNA family. Our data provide direct evidence that this repeat does, in fact, belong to the α satellite family, being comprised entirely of twelve tandem, but divergent monomers, each related in sequence to prototype primate α satellite sequences. A 171 bp consensus sequence of human X-linked α satellite DNA has been derived and compared to other human and primate sequences. In addition, we have identified a region within the 2.0 kb X repeat which is, in hybridization experiments, essentially specific for the human X chromosome and which may be useful as a repeated DNA probe for the human X chromosome centromere.

MATERIALS AND METHODS

DNA Sequencing

DNA fragments from pBamX7 (12) were subcloned into M13mp8 and M13mp9 (14) and sequenced by the dideoxy chain termination method of Sanger et al. (15), employing dATP[α-^{35}S] as the radioactive label in the sequencing reactions. Where possible, DNA was sequenced in both directions; otherwise, DNA was sequenced at least twice in the same direction from independent clones. All sequencing films were read independently by two individuals.

DNA Hybridization

Methods for DNA preparation, restriction endonuclease digestions, electrophoresis, transfer to nitrocellulose, prehybridization and hybridization have been described (11,12). For studies described in this report, DNA was prepared from a normal male 46,XY lymphoblast culture and the Chinese hamster/human somatic cell hybrid c12D. This hybrid, provided by Dr. S. Goss (Oxford), contains a single human X chromosome as its only human component (12,16). The probes used were pBamX7 (the complete 2.0 kb BamHI repeat) and fragments shown in figure 1D, subcloned into pUC8 or pUC9 (14). For stringent hybridization conditions, filters were hybridized at 42°C for 4-16 hours and washed as described (11,12), with a final wash at 68°C in 0.1 x SSC, 0.1% SDS.

RESULTS

Nucleotide Sequence of the 2.0 kb X Chromosome Repeat

The clone pBamX7 contains a single copy of the 2.0 kb BamHI repeat fragment which characterizes α satellite DNA on the human X chromosome

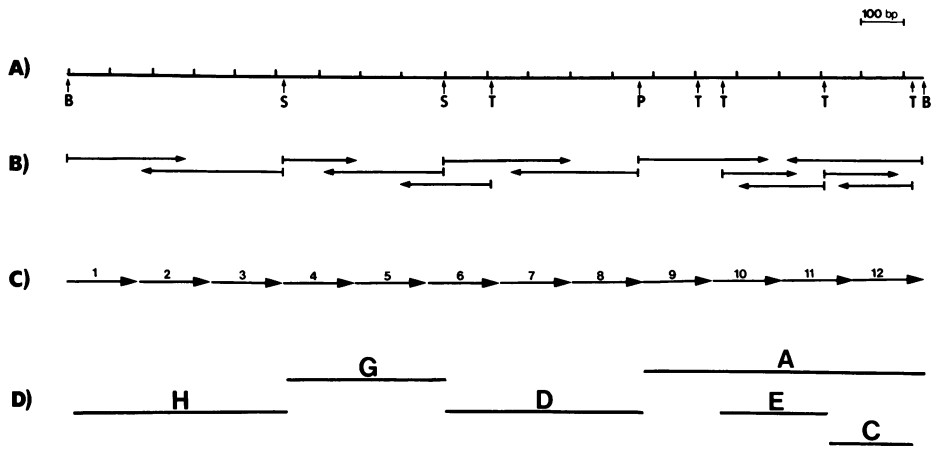


Figure 1- Restriction map, sequencing strategy, and monomer organization of 2.0 kb fragment of pBamX7.

A) restriction sites used to subclone regions of pBamX7 for M13 sequencing and Southern analysis. B, BamHI; T, TaqI; P, PstI; S, Sau3A.

B) direction and sequencing distance of the sequenced subclones. The arrows point away from the dideoxy sequencing primer.

C) positions of alpha satellite monomers (αX1 - αX12) within the 2.0 kb repeat.

D) subcloned regions of pBamX7 used for M13 dideoxy sequencing and as hybridization probes.

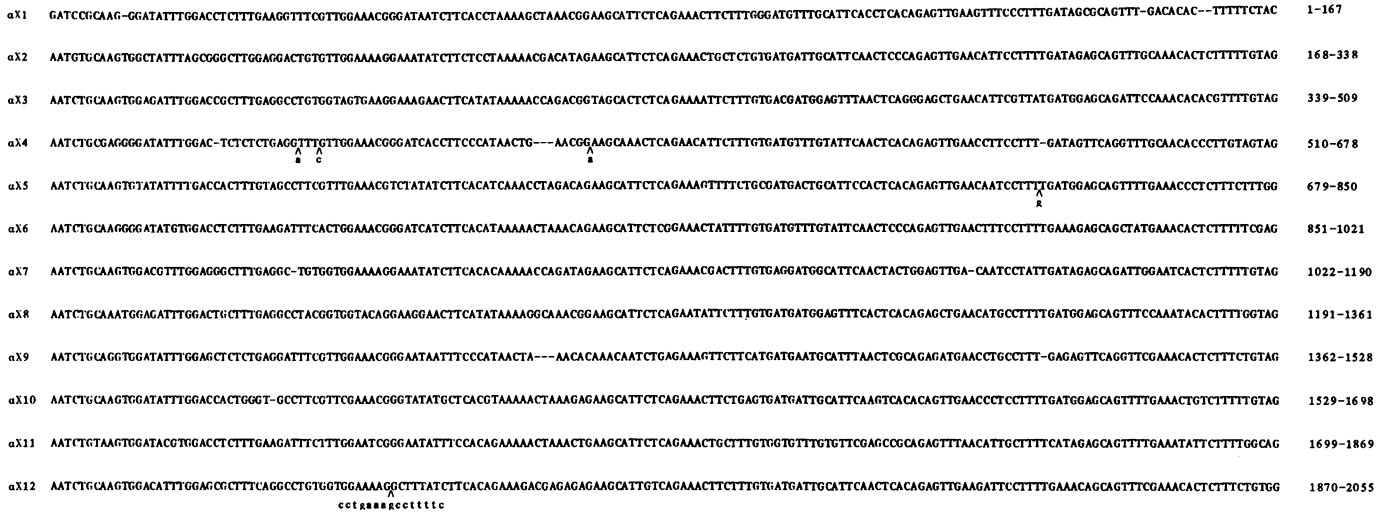


Figure 2- Complete nucleotide sequence of 2.0 kb repeat fragment. The 2055 bp sequence is arranged as monomers (αX1 - αX12) which have been aligned for maximum homology. The positions of base deletions (-) or insertions (^) are based entirely on the homology alignment and are otherwise arbitrary. Base positions of each monomer are given at the right. Base position 1 is the G after the cleavage site in the BamHI recognition sequence.

(11-13). Figure 1 shows a partial restriction map of the fragment as well as the strategy used for sequencing. We determined the complete nucleotide sequence of the repeat and found it was comprised of twelve tandemly arranged units of approximately 171 bp, designated αX1 through αX12 (figure 1C). The complete 2055 bp sequence, arranged as monomers and aligned for maximum homology, is shown in figure 2. The monomers are contiguous with respect to one another and, with the exception of a 15 bp insertion in monomer αX12, only a few base insertions and deletions interrupt the pattern. Individual monomer sequences are 68-73% identical in sequence to the reported consensus sequence of EcoRI α dimers (10) (see below).

Restriction enzyme analysis of human genomic DNA had previously suggested that the 2.0 kb fragment was the major unit of amplification on the X chromosome (11-13). To demonstrate this directly a portion of an independent 2.0 kb BamHI clone (pBamX9; see ref. 12) was sequenced. The region sequenced (the 0.7 kb PstI-BamHI fragment, analogous to αX9-αX12 of pBamX7) was >99% identical to the corresponding region in pBamX7, including the 15 bp insert in αX12. Based on this, we conclude that the 2.0 kb fragment represents the fundamental amplified unit of α satellite DNA on the X chromosome and that our sequence of pBamX7 is likely representative of the entire X-linked DNA family.

The twelve monomer sequences (figure 2) are sufficiently homologous to permit derivation of a consensus αX monomer sequence. A consensus base was assigned only if seven or more of the twelve αX monomers shared a particular base at a given position; otherwise, the position was considered ambiguous and designated 'N'. Based on this criterion, we derived a 171 bp consensus sequence. Figure 3 shows the consensus αX sequence as well as positions in each monomer which differ from the consensus. Forty-one of the 171 positions were completely conserved among the twelve monomers and 117 positions (68%) were conserved among ten or more monomers. Only five positions were considered ambiguous ('N'). The twelve monomers were each 78-89% identical in sequence to the consensus sequence (excluding 'N' positions, and the 15 bp insert in αX12). Base changes from the consensus within each monomer appeared evenly distributed.

In α satellite sequences of lower primates, Musich et al. (3) noted an apparent sequence maintenance of specific domains while other domains accumulated nucleotide changes. To extend this analysis we have compared the consensus sequences of various primate α satellites (including human EcoRI α dimer) to the human X consensus. The X consensus is 80% and 84% identical to

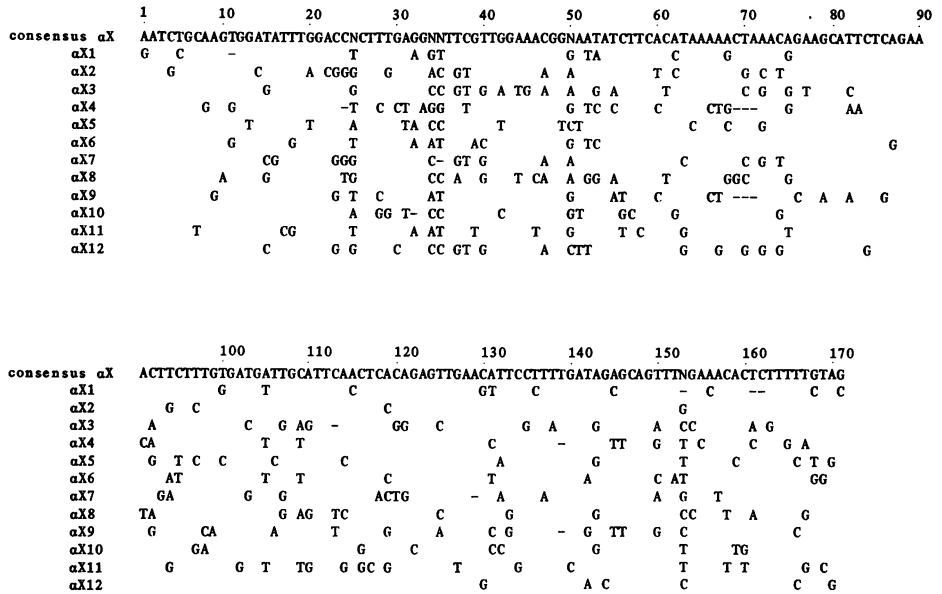


Figure 3- Consensus αX monomer sequence and homologies with the 12 αX monomers. The consensus αX monomer sequence is compared to the 12αX monomers (αX1 - αX12). Bases which differ from the consensus are indicated. Deletions relative to the consensus are indicated by (-). For clarity, the single base insertions in αX4 and αX5 and the 15 bp insertion in αX12 are not shown.

the two monomers within the human α dimer consensus sequence of Wu and Manuelidis (10) and 67% identical to the 172 bp consensus of African green monkey (AGM) α satellite (2). The results, shown in figure 4, indicate that several domains have, in fact, been conserved through evolution, including the regions from positions 13-19, 49-59, 75-100 and 116-125. Strauss and Varshavsky (17) have demonstrated preferential binding of a nuclear protein (α-protein) to three domains in AGM α-satellite DNA and proposed that α-protein binding may play a critical role in nucleosome positioning. Interestingly, the first two conserved regions noted above correspond exactly to the AGM binding sites II and III of Strauss and Varshavsky, while the fourth region overlaps their binding site I (figure 4).

Internal Organization of the X Chromosome Repeat

To gain insight into the molecular organization of monomers within the X repeat, sequence identities among individual monomers were compared. For this analysis, the single base insertions in αX4 and αX5 and the 15 bp insertion in αX12 were omitted. Deletions relative to the consensus were

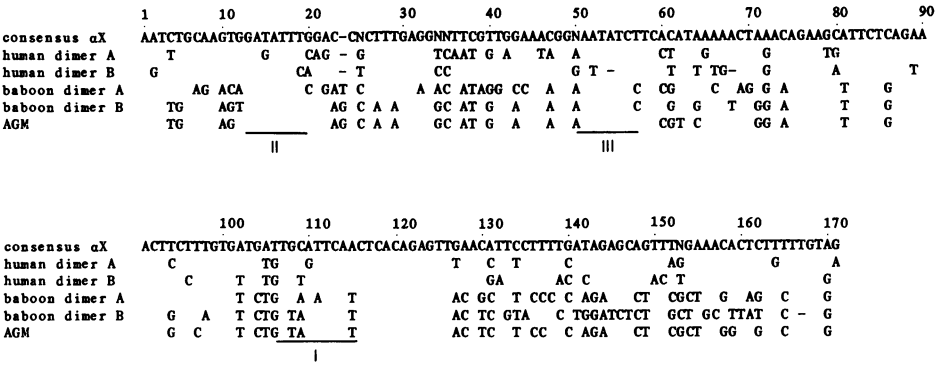


Figure 4 - Comparison of αX consensus sequence to consensus sequences from human α dimer (ref. 10), baboon α dimer (ref. 5), and AGM α monomer (ref. 2). Bases that differ from the αX consensus are indicated as are deletions (-) relative to the αX consensus. A space has been introduced between positions 23 and 24 in the αX and human dimer sequences to maximize alignment with the longer baboon and AGM sequences. The following information allows comparison with the published frames of other sequences: EcoRI cleaves human α dimer between positions 79 and 80 in 'B' (10). BamHI cleaves the baboon α dimer between bases 21 and 22 in 'A' (5). The AGM monomer is characterized by a HindIII cleavage site between positions 77 and 78. The α protein binding sites of Strauss and Varshavsky (17) are indicated beneath the AGM sequence (I,II, and III).

included. The percent sequence identity ranged from 64% to 85%, with most pairwise comparisons being in the 70-80% range (Table 1, below the diagonal). To increase the sensitivity of this analysis, we performed a comparison only of positions within monomers that differed from the consensus sequence (i.e., scoring only those positions shown in figure 3). The results of these comparisons are shown above the diagonal in Table 1, in which the numbers refer to the percentage of all non-consensus positions that are identical between any two of the monomers (the 'consensus identity index'). For example, monomers αX1 and αX2 differ from the consensus at 26 and 23 positions respectively. Only two of these changes are the same between these two monomers (A → C at position 62 in the consensus); thus a consensus identity index of 2/49 or 4% is derived (Table 1). On the other hand, monomers αX1 and αX6 differ from the consensus at 26 and 24 positions respectively, and 14 of these changes (seven positions) are identical between the two monomers, yielding an identity index of 14/50 or 28% (Table 1).

The results of both types of analysis in Table 1 reveal a striking pattern; monomers αX1 through αX5 show maximum homology to monomers αX6

Table 1 - Sequence comparisons among the twelve αX monomers. Numbers below the diagonal represent percent identity of the two compared sequences (171 bp = 100%). Numbers above the diagonal represent the consensus identity index, in wich only bases differing from the consensus are compared. See text for detailed explanation.

	αX1	αX2	αX3	αX4	αX5	αX6	αX7	αX8	αX9	αX10	αX11	αX12
αX1	--	4	3	18	4	28	4	6	10	4	18	8
αX2	75	--	20	3	8	9	38	14	7	9	7	29
αX3	68	74	--	8	9	0	30	61	6	10	3	29
αX4	74	69	66	--	6	25	3	8	41	6	14	9
αX5	74	77	70	67	--	12	11	13	6	27	7	27
αX6	82	78	69	75	77	--	4	0	14	8	33	4
αX7	70	84	78	65	72	74	--	22	3	4	0	33
αX8	71	74	85	67	72	71	73	--	11	10	9	20
αX9	71	72	65	78	69	72	67	66	--	7	15	10
αX10	76	80	68	70	79	77	73	70	72	--	11	16
αX11	74	74	65	67	71	82	67	70	69	74	--	4
αX12	75	83	76	70	79	77	82	75	74	78	78	--

through αX10. Monomer αX11 is clearly related to αX6, while αX12 seems related to a number of monomers, in particular αX2 and αX7 (also αX3 and αX5). The results are consistent with a suborganization based on two overlapping, related blocks of six monomers (represented 'abcdea'), yielding an internal repeat pattern of adjacent groups of five monomers plus two additional monomers, each related to components of the five monomer blocks but not to each other. The results do not discriminate between molecular organizations of [αX1-5][αX6-10][αX11,12] or [αX1][αX2-6][αX7-11][αX12]. However, with tandem arrays of the 2.0 kb fragment (11-13), the resulting chromosomal organization will be the same in either case.

Hybridization Specificity of the X Chromosome Repeat

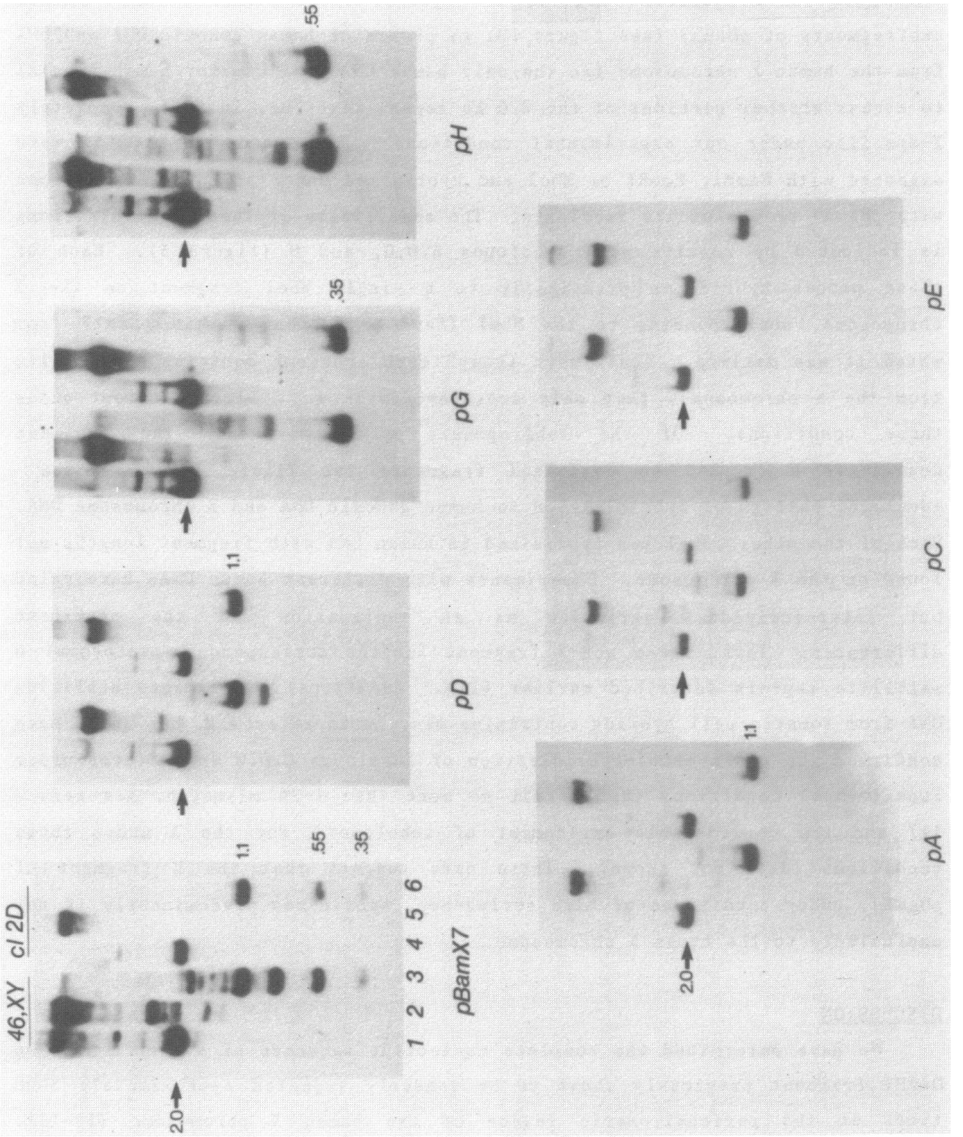
A primary objective of this work was to determine the suitability of pBamX7 as a repeated DNA probe for the human X chromosome and to provide definitive evidence of the chromosome-specificity of this α satellite family. Using pBamX7 to probe human genomic DNA and DNA from rodent/human hybrids, we

previously showed that pBamX7 hybridizes to α satellite sequences on human chromosomes other than the X, albeit to a lesser degree (11, 12).

In light of the observation above that different monomers within the 2.0 kb repeat share different degrees of sequence relatedness, we used cloned subfragments of pBamX7 (see figure 1D) as probes of human genomic DNA and DNA from the human X chromosome (as the only human DNA in a hamster/human hybrid) to assess whether portions of the 2.0 kb repeat might be, in fact, completely X-specific under our experimental conditions. Human and hybrid DNAs were digested with BamHI, EcoRI or MboI and hybridized under stringent conditions with pBamX7 or one of its subclones. The sensitivity of these hybridizations is indicated by results with subclones A,D,G, and H (figure 5). Each of these probes hybridizes principally to a single MboI fragment on the X chromosome, corresponding to the MboI (Sau3A) fragment within pBamX7 from which it was derived. Thus, even though each subclone contains α satellite from the X chromosome, they only cross-hybridize to a limited extent under these conditions. Of the subfragments evaluated, only the fragment designated E (a 242 bp TaqI-TaqI fragment; see figure 1D) gave the identical pattern of hybridization to human genomic DNA and X chromosome DNA. Each of the other subclones hybridized in human DNA with fragment lengths not found on the X chromosome. Experiments with different human DNAs have ruled out inter-individual variation as an explanation for the prominent differences. Thus, these non-X fragment lengths correspond to autosomal α satellite repeats described earlier (11). Additional experiments utilizing DNA from somatic cell hybrids containing many human autosomes, but no X, have confirmed the non-X cross-hybridization of subclones C,D,G and H (even under experimental conditions that permit no more than 1-2% mismatch; see ref. 11) and the considerable enrichment of subclone E for the X under these conditions (data not shown). These data suggest that the E fragment of pBamX7, under conditions of high stringency, hybridizes predominantly if not exclusively to the human X chromosome.

DISCUSSION

We have determined the complete nucleotide sequence of a cloned 2055 bp BamHI fragment previously shown to be tandemly repeated approximately 5000 times at the pericentromeric region of the human X chromosome (11-13). Sequence analysis has demonstrated that this repeated unit is comprised entirely of twelve tandemly repeated α satellite monomers. The molecular organization of monomers within the major repeat is apparently 5-5-1-1 or


46,XY cl 2D
2.0
1.1
.55
.35
1 2 3 4 5 6
pBamX7
pD
pG
pH
pA
pC
pE

1-5-5-1, either of which generates the same pattern in tandemly repeated 2.0 kb units. The data do not provide any support for a suborganization based on a 340 bp α dimer, and the monomers within the 2.0 kb X repeat do not appear to be more closely related to one or the other half of the EcoRI α dimer characterized in detail by Manuelidis (7-10). Wu and Manuelidis have proposed a two-step evolutionary process for the formation of tandem arrays of α satellite dimers, the first step resulting in dimer formation from divergent monomers, the second being amplification of the dimer into long tandem arrays (10). Similarly, evolution of a basic twelve-mer unit likely preceeded its tandem amplification on the human X chromosome. Genesis of the original 2.0 kb repeat may have involved unequal crossing-over between two six-monomer units as an intermediate step (Table 1). However, due to the complexity of the X repeat unit and due to the uncertain contribution of sequence rectification (conversion) (18,19) during evolution of the repeat, we refrain from proposing more detailed models delineating the molecular evolution of this α satellite family.

Comparisons of the twelve αX monomer sequences allowed us to derive a consensus sequence characteristic of X chromosome α satellite DNA. This is the first instance in which it has been possible to derive such a consensus for an individual primate chromosome. Our ability to do so depends on our having identified and sequenced the entire basic amplification unit of α satellite DNA on the X and on our assumption that the sequence of pBamX7 is representative of the entire family of 5000 members. That the sequence of a portion of a second, independent member of this α satellite family is in such close agreement with the pBamX7 sequence supports this assumption. These sequence data, together with the hybridization data shown in figure 5 and with results of Southern blotting experiments of DNA from rodent/human somatic cell hybrids containing only single human chromosomes (11), provide direct support for the hypothesis that individual primate chromosomes are each characterized by different, and specific, α satellite families, as first proposed by Maio et al.(6) and as previously demonstrated by Lee and Singer

Figure 5 - Southern blot analysis of X-specificity of pBamX7 and subclones. Each set of six lanes contains DNAs as follows: lane 1, human (46,XY) DNA digested with BamHI; lane 2, human DNA, EcoRI; lane 3, human DNA, MboI. Lane 4, c12D (human X chromosome only) DNA digested with BamHI; lane 5, c12D DNA, EcoRI; lane 6, c12D DNA, MboI. Each panel illustrates hybridization with pBamX7 or the indicated subclone (refer to Fig. 1D). Arrow to the left of each panel indicates the 2.0 kb BamHI fragment. Numbers to the right indicate size (in kb) of MboI fragments detected in c12D DNA.

for an individual AGM chromosome (20).

To test this hypothesis for the human genome, we have begun cloning and comprehensively characterizing α satellite representatives from single human chromosomes, the X chromosome being our initial focus. At present, our analysis of the αX consensus in this regard is subject to the limited availability of α satellite sequences from other known chromosomes. The consensus sequence of the EcoRI α dimer has been known for some time (10), but the chromosomal distribution of this particular molecular configuration (restriction enzyme periodicity) remains to be established, although abundant centromeric sequences most closely related to the α dimer are probably confined to only a small number of autosomes (ref. 8 and our unpublished data). Recently, partial sequence data were reported for a 5.5 kb α satellite repeat, with an underlying monomeric construction, from the human Y chromosome (21). Although remarkably similar to the X consensus, there were several nucleotide positions in the αY monomers which consistently differed from those of the X. These data tend to support the hypothesis of α satellite chromosome specificity at the level of primary sequence, although the significance of these relationships will only be apparent when the complete sequence of the entire α satellite amplified unit on the Y (i.e. 5.5 kb) is known. We suggest that comparisons between or among different α satellite families be based only on complete sequences. As evidenced by our data on the X chromosome, monomer sequences are not necessarily indicative of an overall repeat sequence and comparisons of less than complete sequences may, therefore, be misleading. We submit that comprehensive sequence analysis of α satellite DNA from specific human chromosomes may provide substantial insight into both α satellite and chromosome evolution, as well as define a collection of chromosome-specific repeated DNA hybridization probes.

As part of this study, we have identified a subfragment of the 2.0 kb X repeat which, under stringent conditions, appears to hybridize principally to the human X chromosome. This fragment encompasses approximately one and a half monomers of the X repeat (see E fragment, figure 1D). It is not apparent at the level of primary sequence why this fragment is so enriched for the X chromosome whereas other subfragments cross-hybridize with α satellite sequences elsewhere in the genome. At present, the utility of the E fragment as a centromeric X chromosome probe remains under investigation. We are encouraged by results indicating restriction fragment length polymorphisms in several human autosomal centromeric DNA families (22,23) and

by the demonstration of extensive polymorphism adjacent to α satellite DNA in the African green monkey genome (24,25). Should the X-linked α satellite repeat be polymorphic as well, the E fragment probe may be extremely useful in X chromosome linkage studies.

ACKNOWLEDGMENTS

This work was supported by grants from the Medical Research Council of Canada and from the March of Dimes Birth Defects Foundation (Grant 5-390). The authors thank S. Climie, R. Korneluk, M. Tropak and L. Rossiter for advice and assistance.

*To whom correspondence should be addressed.

REFERENCES

1. Maio, J.J. (1971) J. Mol. Biol. 56,579-595.
2. Rosenberg, H., Singer, M.F. and Rosenberg, M. (1978) Science 200, 394-402.
3. Musich, P.R., Brown, F.L. and Maio, J.J. (1980) Chromosoma 80,331-348
4. Singer, M.F. (1982) Int. Rev. Cytol. 76,67-111
5. Donehower, L. and Gillespie, D. (1979) J. Mol. Biol. 134,805-834.
6. Maio, J.J., Brown, F.L. and Musich, P.R. (1981) Chromosoma 83,103-125.
7. Manuelidis, L. (1976) Nucl. Acids Res. 3,3063-3078.
8. Manuelidis, L. (1978) Chromosoma 66,23-32.
9. Manuelidis, L. and Wu, J.C. (1978) Nature 276,92-94.
10. Manuelidis, L. and Wu,J.C. (1980) J. Mol. Biol. 142,363-386.
11. Willard, H.F. (1985) Am. J. Hum. Genet., in press.
12. Willard, H.F., Smith K.D., and Sutherland, J. (1983) Nucl. Acids Res. 11,2017-2033.
13. Yang, T.P., Hansen, S.K., Oishi, K.K., Ryder, O.A. and Hamkalo, B.A. (1982) Proc. Natl. Acad. Sci. U.S.A. 79,6593-6597.
14. Messing, J. and Vieira, J. (1982) Gene 19,269-276.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74,5463-5467.
16. Goss, S. and Harris, H. (1977) J. Cell Sci. 25,17-37.
17. Strauss, F. and Varshavsky, A. (1984) Cell 37,889-901.
18. Petes, T., and Fink, G.R. (1982) Nature 300,216-217.
19. Dover, G. (1982) Nature 299,111-117.
20. Lee, T.N.H. and Singer, M.F. (1982) J. Mol. Biol. 161, 323-342.
21. Wolfe, J., Darling, S.M., Erickson, R.P., Craig, I., Buckle, V., Rigby, R., Willard, H.F. and Goodfellow, P. (1985) J. Mol. Biol., in press.
22. Jabs, E.W., Wolf, S.F. and Migeon B.R. (1984) Proc. Natl. Acad. Sci. U.S.A. 81,4884-4888.
23. Willard, H.F. and Waye, J.S. (1984) Am. J. Hum. Genet. 36,117S.
24. Maresca, A. and Singer, M.F. (1983) J. Mol. Biol. 164,493-511.
25. Maresca, A., Singer, M.F., and Lee, T.N.H. (1984) J. Mol. Biol.179, 629-649.