Comparative Sequence Analysis of Human Minisatellites Showing Meiotic Repeat Instability

John Murray, Jérôme Buard, David L. Neil, et al.

Genome Res. 1999 9: 130-136
Access the most recent version at doi:10.1101/gr.9.2.130

Letter

Comparative Sequence Analysis of Human Minisatellites Showing Meiotic Repeat Instability

John Murray, Jérôme Buard, David L. Neil, Edouard Yeramian,[2] Keiji Tamaki, Caroline Hollies, and Alec J. Jeffreys[1]

Department of Genetics, University of Leicester, Leicester LE1 7RH, UK; [2]Unité de Physico-Chimie, Institut Pasteur, 75015 Paris, France

The highly variable human minisatellites MS32 (*D1S8*), MS31A (*D7S21*), and CEB1 (*D2S90*) all show recombination-based repeat instability restricted to the germline. Mutation usually results in polar interallelic conversion or occasionally in crossovers, which, at MS32 at least, extend into DNA flanking the repeat array, defining a localized recombination hotspot and suggesting that *cis*-acting elements in flanking DNA can influence repeat instability. Therefore, comparative sequence analysis was performed to search for common flanking elements associated with these unstable loci. All three minisatellites are located in GC-rich DNA abundant in dispersed and tandem repetitive elements. There were no significant sequence similarities between different loci upstream of the unstable end of the repeat array. Only one of the three loci showed clear evidence for putative coding sequences near the minisatellite. No consistent patterns of thermal stability or DNA secondary structure were shared by DNA flanking these loci. This work extends previous data on the genomic environment of minisatellites. In addition, this work suggests that recombinational activity is not controlled by primary or secondary characteristics of the DNA sequence flanking the repeat array and is not obviously associated with gene promoters as seen in yeast.

[The sequence data described in this paper have been submitted to the GenBank data library under accession nos. AF048727 (CEB1), AF048728 (MS31A), and AF048729 (MS32).]

Minisatellites include some of the most unstable loci in the human genome and provide highly informative systems for dissecting processes of tandem repeat instability in the germline. Analysis of de novo mutant alleles identified in families or in single sperm has shown that repeat instability most likely arises at meiosis and can result in complex allelic rearrangements (Jeffreys et al. 1994; Jeffreys and Neumann 1997). For minisatellites MS31A, MS32, and MS205, these rearrangements are dominated by gene conversion-like transfers of repeat units between alleles that are largely restricted to one end of the repeat array and do not result in exchange of flanking markers (Armour et al. 1993; Jeffreys et al. 1994; May et al. 1996). The highly unstable minisatellite CEB1 similarly shows polar interallelic conversion but also complex and nonpolar intraallelic rearrangements that nevertheless appear to be meiotic in origin, possibly arising by alternative processing of a recombination initiation complex (Buard and Vergnaud 1994; Buard et al. 1998).

Analysis of MS31A and MS32 has shown that meiotic crossovers with exchange of flanking markers do occur in sperm within the repeat array, though at a much lower frequency than for conversion (Jeffreys et al. 1998b). Analysis of crossovers in DNA flanking the unstable end of the MS32 locus has shown that this minisatellite is located at the boundary of an intense and highly localized meiotic recombination hotspot ~1.5 kb long that reaches maximal activity about 200 bp upstream of MS32 and extends into the beginning of the repeat array, where it results in unequal and equal (in register) crossover between allelic arrays (Jeffreys et al. 1998a). Preliminary evidence at MS31A and CEB1 also suggests that crossover activity is not restricted to the repeat array but extends into flanking DNA (C. Hollies and J. Buard, unpubl.).

At minisatellite MS32, a single base transversion (G → C) at the O1 site 48 bp outside the repeat array and within the flanking recombination hotspot is strongly associated with, and probably directly causes, cosuppression of array conversion, array crossover, and possibly flanking crossover in sperm (Monckton et al. 1994; Jeffreys et al. 1998a,b). This cosuppression provides strong evidence that minisatellite conversion and crossover arise by a common mechanism such as alternative processing of a meiotic recombination initiation complex and, further, suggests that repeat instability is not an intrinsic property of the repeat array but is instead controlled by recombination initiation elements located outside the array and presumably in or near the flanking hotspot (Jeffreys et al. 1998a). The stabilizing MS32 O1C variant appears to

[1]Corresponding author.
E-MAIL ajj@le.ac.uk; FAX (0116) 252-3378.

9:130–136 ©1999 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/99 $5.00; www.genome.org

function by blocking the initiation of recombination. This could occur by altering the binding site for a recombinogenic protein; alternatives include the adventitious creation of a binding site for a protein that interferes with recombination, or alteration of a local open chromatin domain that normally provides accessibility to the meiotic recombination machinery.

Human minisatellites preferentially cluster in subtelomeric regions (Royle et al. 1988; Amarger et al. 1998), which are involved in initiating meiotic chromosome pairing and are proficient in meiotic recombination. MS32 occupies an atypical interstitial position on the long arm of chromosome 1 (1q42-43), whereas both MS31A (7p22-pter) and CEB1 (2q37.3) are in subtelomeric locations (Royle et al. 1988; Vergnaud et al. 1991). Previous evidence suggested that human hypervariable minisatellites tend to occur in regions rich in dispersed repeats and often clustered with other minisatellites (Armour et al. 1989; Vergnaud et al. 1993), but little information exists on long-range sequence organization around unstable minisatellites. Therefore, cosmid clones containing MS31A, MS32, and CEB1 were isolated and sequenced to search for features such as primary sequence similarity, secondary DNA structure, and gene elements that might be associated with recombinational proficiency at these loci.

RESULTS

Sequencing Minisatellites

Cosmids containing minisatellites MS31A, MS32, and CEB1 were isolated and fully sequenced on both strands, yielding 14, 19, and 24 kb of sequence upstream of the unstable end of each locus, respectively, and 12, 10, and 13 kb of downstream flanking sequence. Long PCR amplification of overlapping genomic DNA fragments from regions flanking MS32 and CEB1 and Southern blot analysis of MS31A flanking sequences showed that all cosmid inserts were bona fide copies of genomic DNA without obvious rearrangement (data not shown). MS31A and CEB1 are located in highly GC-rich DNA (~60% and ~64% GC, respectively) as predicted for GC-rich terminal isochores in human chromosomes. In contrast, the interstitial minisatellite MS32 is located in moderately GC-rich DNA (45% GC).

Repetitive DNA

MS31A and CEB1 are closely linked to other minisatellites, particularly for CEB1 where there are six other minisatellites clustered within 12 kb of the 3′ end of the locus; none of these additional loci shows high levels of variability (data not shown; Fig. 1). In contrast, there are no other minisatellites near MS32, per-

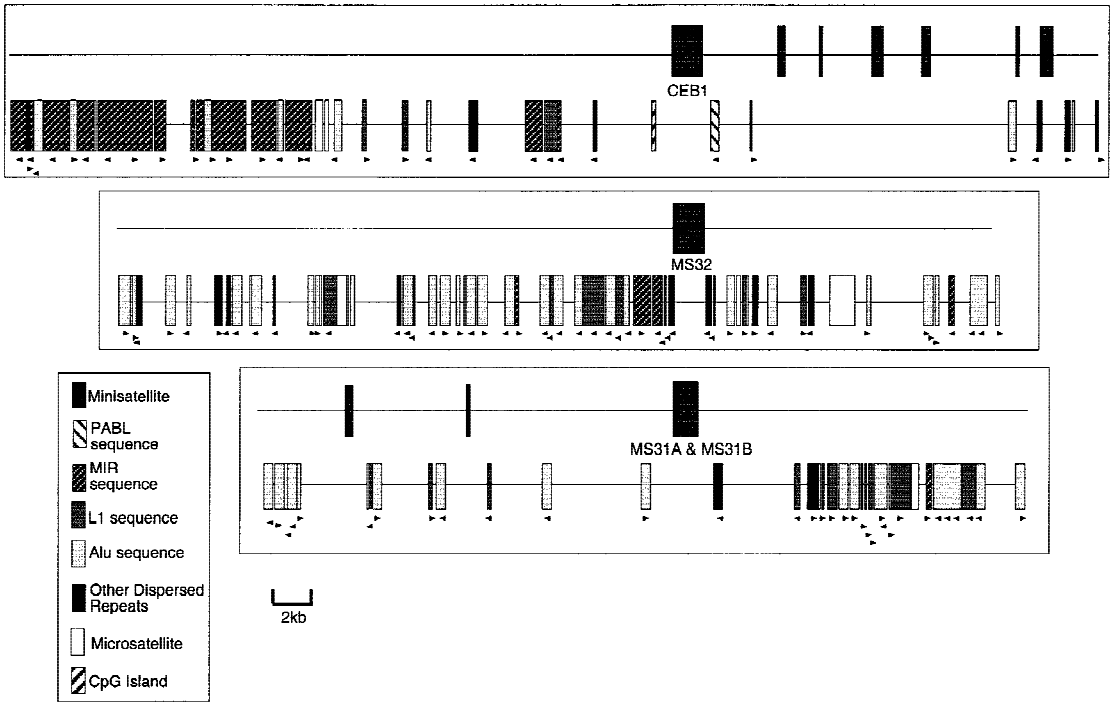

Figure 1 Repeat elements surrounding human minisatellites CEB1, MS32 and MS31. Minisatellites (*top*) and other tandem and dispersed repetitive elements (*bottom*) are shown for each locus. The highly variable minisatellites CEB1, MS32 and MS31A are orientated with the unstable end of the repeat array to the left. The closely linked minisatellite MS31B shows only modest variability. The orientation of dispersed repeats is indicated by arrows, with multiple dispersed repeats per box indicated by additional arrows.

haps reflecting its interstitial location. MS31A, MS32, and CEB1 are all surrounded by an abundance of simple tandem repeat DNA and dispersed repeats, though dispersed repeats are under-represented in the minisatellite-rich region of the CEB1 cosmid. Alu repeats occur at a frequency four times the genome average (one every 1 kb instead of one per ~4 kb; Schmid and Jelinek 1982), particularly surrounding MS32, where they make up 31% of the sequence. Alu repeats are less numerous near CEB1, and the 5′ flanking DNA is instead dominated by MER repeats (Donehower et al. 1989) that make up 20% of the upstream DNA. The Alu repeats around all three loci vary in the age of their subfamilies, but none belongs to the currently active group in humans (Matera et al. 1990).

The 3′ flanking sequence of CEB1 showed strong homology beginning 129 bp downstream of the minisatellite to a diverged, but complete pseudo-autosomal boundary-like sequence (PABL-B subfamily), with 87% sequence similarity over 662 bp. The PABL sequence was initially found at the boundary of the pseudo-autosomal region of the sex chromosomes and subsequently at boundaries between megabase-level GC-rich and AT-rich isochores defined by chromosome walking and base compositional analysis (Fukagawa et al. 1995, 1996). At CEB1, the GC content gradually increases from ~40% GC in the dispersed repeat-rich region upstream of the PABL to ~63% GC in the downstream minisatellite-rich region. Much more extensive sequence analysis around CEB1 would be required to establish whether the CEB1-associated PABL genuinely constitutes an isochore boundary.

DNA Sequence Comparisons

To determine whether the three minisatellite loci share sequence similarities outside the repeat arrays, nonrepetitive sequences representing 46% of the total DNA were compared between loci by dot matrix analysis. No regions of significant similarity were found. Similarly, BESTFIT analysis yielded only small segments of similarity of marginal significance; these segments showed no consistent location common to all three loci (data not shown). Therefore, there is no evidence for primary sequence similarities shared by these loci, in particular over the DNA flanking the unstable end of the repeat array.

Thermal Stability Analysis

Thermal stability profiles were analyzed around each minisatellite (Fig. 2), using published models (Yeramian et al. 1990) to determine whether there were consistent patterns of stability shared by the three loci. All three minisatellites are extremely refractory to thermal denaturation, reflecting their high GC content. The unstable end of MS32 is flanked by a 4-kb domain of easily opened DNA (defined for three given temperatures, Fig. 2); however, this domain starts 240 bp upstream of the array and does not coincide in location with the 1.5-kb MS32-associated recombination

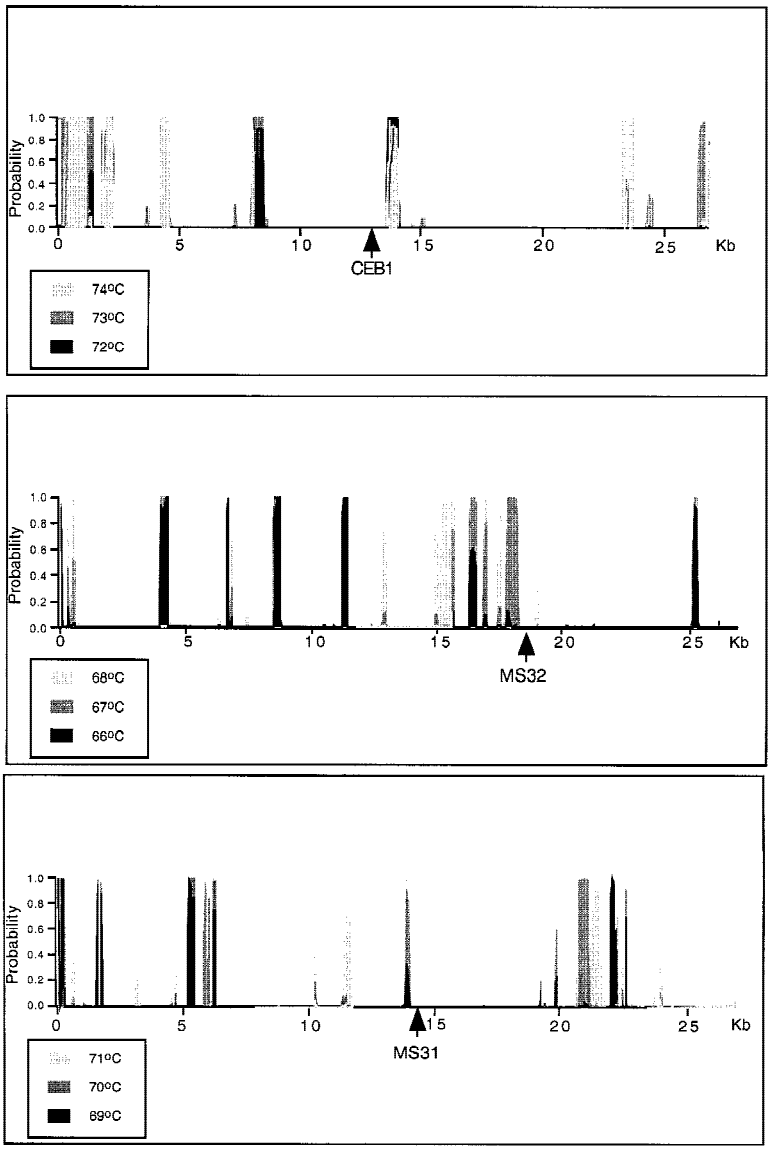

Figure 2 Thermal stability profiles of sequences surrounding minisatellite loci. The local probability of DNA helix opening was determined for three different temperatures per locus selected on the basis of overall base composition for each locus. The highly variable minisatellites are orientated with the unstable end of the repeat array to the left.

hotspot that extends into the beginning of the repeat array (Jeffreys et al. 1998b). Similarly, the stabilizing O1C variant near MS32 lies outside the easily opened domain and has no effect on thermal stability. The upstream flanking region of MS31A also shows a low stability domain; however, this region is short (390 bp) and does not immediately abut the minisatellite; no data currently exist concerning recombination profiles surrounding MS31A. Finally, the 3.5 kb of DNA immediately upstream of CEB1 shows high thermal stability, though there is an openable domain close to the 3′ end of the minisatellite. Therefore, there is no consistent pattern of thermal instability shared by the upstream flanking DNA of these loci.

DNA Secondary Structure

Secondary structure in DNA can be described by two parameters, namely the intrinsic flexibility of the DNA helix (bendability) and the tendency of the helix to form a bent structure (curvature propensity) as a result for example of purine/pyrimidine strand asymmetry. Data from Dnase I digestion and nucleosome positioning (Gabrielian and Pongor 1996) were used to predict bendability and curvature propensity around all three minisatellites. No consistent patterns shared by the flanking DNA of these loci were observed (data not shown). This analysis was extended to the immediate vicinity of the stabilizing O1C variant near minisatellite MS32 (Fig. 3) to determine whether this variant might influence secondary structure in this region. The O1 site is located within a few base pairs of a region of relatively low bendability, but the O1 G → C transversion has no effect on flexibility. This region also shows very low curvature propensity (<0.06 compared with the genome average of 0.3; Gabrielian and Pongor, 1996). The O1C variant enhances curvature over a small domain 10–20 bp upstream of the O1 site; however, this enhancement is also seen if similar G → C switches are made in the vicinity of O1, indicating that this effect is not specific to the O1 site (data not shown).

Putative Genes

The Nix search engine (see Methods) was used to search for putative genes in the vicinity of minisatellites MS31A, MS32, and CEB1. There was no evidence for coding regions or CpG islands near MS32. The MS31A flanking sequence yielded a 391-bp region 99.5% identical in sequence to a phaeochromocytoma IMAGE cDNA clone (dbEST identification number 1311250). This expressed sequence tag corresponds to the 3′ end of a clone from an oligo(dT)-primed cDNA library and therefore, should correspond to a 3′-untranslated transcribed region (UTR). However, this region, situated 1.8–2.2 kb downstream of the minisatellite, is not located near any predicted exons, there is no continuous open reading frame near the region of homology, nor are there any polyadenylation signals, although about half of true 3′ UTRs show no obvious polyadenylation signal. Therefore, it remains to be established whether this region contains a genuine expressed sequence, or whether it is part of a noncoding RNA gene, which cannot be detected by current exon prediction programs (Claverie 1997). Alternatively, the IMAGE clone may be a genomic contaminant.

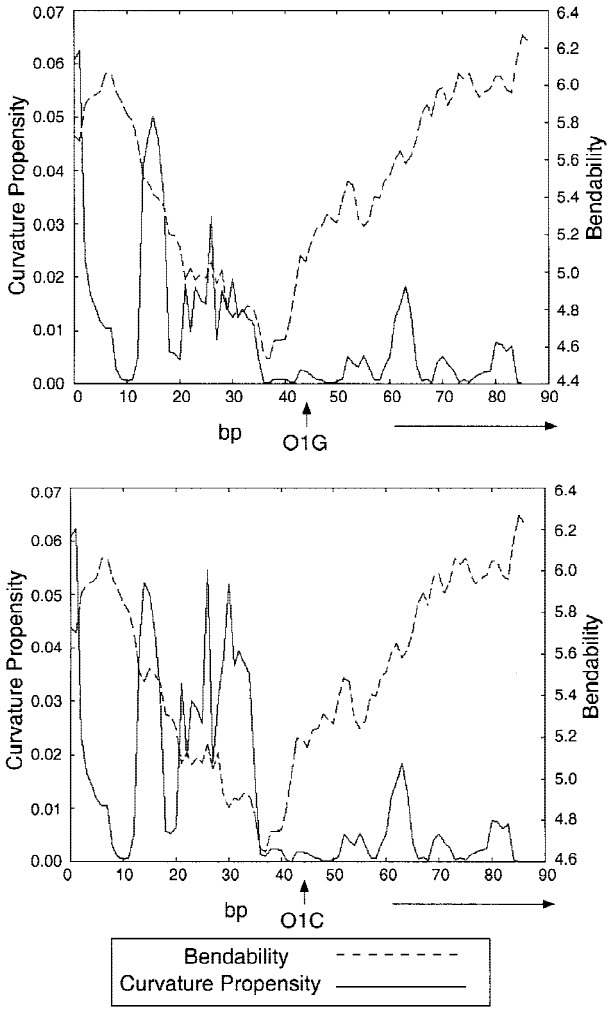

Figure 3 Influence of the O1 G → C polymorphism on bendability and curvature propensity of DNA adjacent to the unstable end of the minisatellite MS32 repeat array. Secondary structural analysis with the Bend.It Server was performed on 90 bp of MS32 flanking DNA plus the diverged repeat unit (horizontal arrow) that precedes the repeat array proper, with analyses repeated for the O1G and O1C alleles that differ by a single base transversion.

In contrast, the CEB1 cosmid yielded two candidate genes (Fig. 4). The 5′-flanking region produced a highly significant BLAST/trembl match (59% sequence similarity) to cerebroside sulphotransferase mRNA (accession no. d88667). The matching regions coincided with strong exon predictions from GRAIL/exons (prob-

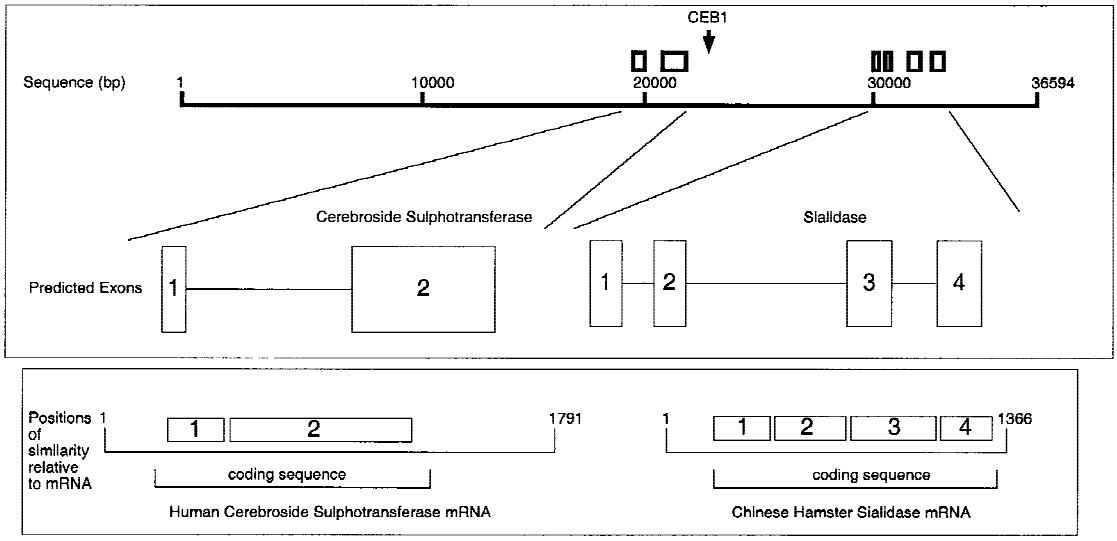

Figure 4 Putative genes near minisatellite CEB1. Predicted exons (bold boxes) related to cerebroside sulphotransferase and sialidase coding sequences are shown relative to the location of CEB1 (arrow) and to mRNA sequences (*below*).

ability score, 0.84) and GENEMARK (protein-coding exon probability, 0.94). Dot matrix comparisons defined two exons with open reading frames that together cover 60% (1076 bp) of human cerebroside sulphotransferase mRNA. The second exon terminates 1175 bp upstream of the unstable end of CEB1. No homologies to the remaining mRNA sequence were detected in the CEB1 cosmid, suggesting that this putative gene is a highly diverged member of a sulphotransferase gene family or possibly a pseudogene. Similar analysis of DNA flanking the 3′ end of CEB1 revealed four exons with homology (53% sequence similarity) to Chinese hamster sialidase mRNA (accession no. u06143), with the first exon commencing 6530 kb downstream of CEB1. No homologies to the 5′-untranslated region of sialidase mRNA were detected, preventing localization of the gene promoter, assuming that this gene is functional. There is a possible CpG island just upstream of CEB1 that could mark the promoter, but it is only 440 bp long, considerably shorter than the bulk of gene-associated CpG islands (Bird 1987).

DISCUSSION

Human minisatellites preferentially cluster near human telomeres in GC-rich isochores abundant in dispersed repeats from which some VNTR loci such as MS32 have amplified (Royle et al. 1988; Armour et al. 1989; Vergnaud et al. 1993; Amarger et al. 1998). Sequence analysis of DNA around the proterminal minisatellites CEB1 and MS31A confirms their location in GC-rich domains containing minisatellite clusters and abundant dispersed repeats, reminiscent of the well-characterized VNTR-rich domain at 16pter that includes the α-globin gene cluster (Flint et al. 1997). The interstitial minisatellite MS32 is located in a less GC-rich domain and appears not to be a component of a VNTR cluster. Minisatellites MS31A, MS32, and CEB1 all show common features of tandem repeat instability, which is largely restricted to the germline, appears to be meiotic in origin, and frequently involves polar recombinational interactions between alleles resulting in interallelic conversion or occasionally crossover (Buard and Vergnaud 1994; Jeffreys et al. 1994; Jeffreys and Neumann 1997; Buard et al. 1998). Detailed analysis of MS32 has shown that it is located at the boundary of a localized recombination hotspot (Jeffreys et al. 1998a,b); preliminary data at MS31A and CEB1 suggest that crossover activity similarly extends into DNA flanking the unstable end of the locus, although it remains to be established whether this flanking recombination activity is restricted to a localized hotspot as seen at MS32. Further evidence at MS32 strongly suggests that repeat instability is controlled in *cis* by as-yet-unidentified flanking DNA elements in the recombination hotspot (Jeffreys et al. 1998a).

The lack of significant sequence similarity between DNA flanking MS32, MS31A, and CEB1 suggests that these flanking mutation/recombination initiator elements will not be definable by primary DNA sequence. This is concordant with data in yeast, which again show lack of sequence similarity between different recombination hotspots (Rocco et al. 1992; deMassy and Nicolas 1993; Goyon and Lichten 1993).

In *Saccharomyces cerevisiae*, there is a strong tendency for recombination hotspot-associated double-strand breaks to localize to transcriptional promoters (Baudat and Nicolas 1997). The lack of obvious gene sequences near MS32 and MS31A suggests that this

correlation may not necessarily apply to human recombination hotspots, though direct analysis of germline transcriptional activity near these minisatellites is required to clarify this issue. For CEB1, there is a putative sialidase-related gene downstream of the locus whose promoter could be near the end of the minisatellite showing maximal recombination activity; however, the location of the promoter and the functional status of the gene in the germline remain unknown.

Double-strand breaks that initiate meiotic recombination in fusion yeast tend to occur at or near sites of open chromatin (Ohta et al. 1994; Wu and Lichten 1994; Fan and Petes 1996), suggesting that as-yet unidentified DNA elements that control chromatin architecture are responsible for creating the open chromatin domains necessary for generating recombination hotspots by providing accessibility to the recombination machinery. If minisatellite-associated recombination hotspots in humans are associated with open chromatin domains, then such domains may be detectable in primary DNA sequences as regions showing unusual secondary structure or propensity to open the helix. Thermal stability modelling shows the presence of regions near all three minisatellites that are intrinsically more able to open the helix; however, these regions do not colocalize with the recombination hotspot near MS32, nor do they show a consistent pattern shared by all three loci, suggesting that the hotspot activity is not defined by relatively AT-rich domains (although the possibility remains of some role in initiation of recombination).

A common feature of GC-rich minisatellites is purine/pyrimidine strand asymmetry (Jeffreys et al. 1985). GC-rich repeats from herpes simplex virus 1 (HSV1) with a pronounced strand asymmetry have been shown to adopt unusual DNA conformations in plasmids in vivo (Wohlrab et al. 1991). However, curvature analysis on MS31A, MS32, and CEB1 as naked DNA (reflecting static molecular properties) shows that all three minisatellites will tend to adopt straight rigid structures because their repeat unit periodicity prevents long-range curvature (E. Yeramian, in prep.). DNA flanking the minisatellites does show regions of significant curvature and bendability, but again there is no consistent pattern shared by all three loci. The O1 G → C transversion near MS32, which appears to be directly responsible for suppressing in *cis* the initiation of crossovers and conversion in and near the minisatellite (Monckton et al. 1994; Jeffreys et al. 1998a,b), does influence curvature propensity nearby, and it is possible that this effect could affect nucleosome positioning, and thus chromatin structure, as intrinsically curved DNA is more readily wrapped around a histone octamer (Fitzgerald and Anderson 1998). Further analysis of the chromatin structure of human meiotic recombination hotspots will require the definition of additional hotspots in the human genome and the development of methods for probing chromatin architecture in human meiotic cells.

METHODS

Cosmid Cloning and Sequencing

The isolation of cosmids containing minisatellites MS32 and CEB1 has been described previously (Vergnaud et al. 1991; Bois et al. 1997). MS31A was isolated from a human cosmid library constructed with the vector pAVCV007 (Choo et al. 1986) by hybridization screening with an MS31A probe (Wong et al. 1987). Three cosmids were isolated, two extending upstream from the minisatellite and one extending downstream. Appropriate cosmids were sonicated and 1.0- to 1.5-kb DNA fragments cloned into pBlueScriptII SK$^+$. Ordered array shotgun clones were screened by hybridization with appropriate cosmid restriction fragments covering regions to be sequenced. Relevant phagemid DNAs were subjected to automated sequencing with the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit and sequence data were assembled by use of ABI AutoAssembler software. Single-stranded regions were reanalyzed by reverse sequencing of PCR-amplified phagemid inserts and gaps closed by sequencing appropriate cosmid PCR products. Finished cosmid sequences were complete on both strands.

Sequence Analysis

Known dispersed repeats were identified by use of the CENSOR software (Jurka et al. 1996). The Nix search engine provided by the MRC Human Genome Mapping Programme (http://www.hgmp.mrc.ac.uk) was used to screen for putative coding sequences and exons. DNA sequence comparisons by dot matrix and BESTFIT analysis were performed with the GCG package of software. Thermal stability curves were produced by software written by E.Y. DNA secondary structural analysis was carried out by the Bend.It Server (http://www.icgeb.trieste.it/dna/bend_it.html).

ACKNOWLEDGMENTS

We are grateful to Kathryn Lilley and Stuart Bayliss for oligonucleotide synthesis and assistance with automated sequencing, and to colleagues for helpful discussions. This work was supported in part by an International Research Scholars Award to A.J.J. from the Howard Hughes Medical Institute and in part by grants from the Wellcome Trust, Medical Research Council and Royal Society. E.Y. acknowledges generous support from the Pasteur Institute.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked ''advertisement'' in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Amarger, V., D. Gauguier, M. Yerle, F. Apiou, P. Pinton, F. Giraudeau, S. Monfouilloux, M. Lathrop, B. Dutrillaux, and G. Vergnaud. 1998. Analysis of distribution in the human, pig, and rat genomes points toward a general subtelomeric origin of minisatellite structures. *Genomics* **52:** 62–71.

Armour, J.A.L., Z. Wong, V. Wilson, N.J. Royle, and A.J. Jeffreys.

1989. Sequences flanking the repeat arrays of human minisatellites: Association with tandem and dispersed repeat elements. *Nucleic Acids Res.* **17:** 4925–4935.

Armour, J.A.L., P.C. Harris, and A.J. Jeffreys. 1993. Allelic diversity at minisatellite MS205 (16S309): Evidence for polarized variability. *Hum. Mol. Genet.* **2:** 1137–1145.

Baudat, F. and A. Nicolas. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc. Natl. Acad. Sci.* **94:** 5213–5218.

Bird, A.P. 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3:** 342–347.

Bois, P., A. Collick, J. Brown, and A.J. Jeffreys. 1997. Human minisatellite MS32 (*D1S8*) displays somatic but not germline instability in transgenic mice. *Hum. Mol. Genet.* **6:** 1565–1571.

Buard, J. and G. Vergnaud. 1994. Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *EMBO J.* **13:** 3203–3210.

Buard, J., A. Bourdet, J. Yardley, Y. Dubrova, and A.J. Jeffreys. 1998. Influences of array size and homogeneity on minisatellite mutation. *EMBO J.* **17:** 3495–3502.

Choo, K.H., G. Filby, S. Greco, Y.F. Lau, and Y.W. Kan. 1986. Cosmid vectors for high-efficiency DNA-mediated transformation and gene amplification in mammalian-cells—studies with the human growth-hormone gene. *Gene* **46:** 277–286.

Claverie, J.M. 1997. Computational methods for the identification of genes in vertebrate genomic sequences. *Hum. Mol. Genet.* **6:** 1735–1744.

deMassy, B. and A. Nicolas. 1993. The control in cis of the position and the amount of the ARG4 meiotic double-strand break of *Saccharomyces cerevisiae*. *EMBO J.* **12:** 1459–1466.

Donehower L.A., B.L. Slagle, M. Wilde, G. Darlington, and J.S Butel. 1989. Identification of a conserved sequence in the noncoding regions of many human genes. *Nucleic Acids Res.* **17:** 699–710.

Fan, Q.Q. and T.D. Petes. 1996. Relationship between nuclease-hypersensitive sites and meiotic recombination hot-spot activity at the HIS4 locus of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16:** 2037–2043.

Fitzgerald, D.J. and J.N. Anderson. 1998. Unique translational positioning of nucleosomes on synthetic DNAs. *Nucleic Acids Res.* **26:** 2526–2535.

Flint, J., K. Thomas, G. Micklem, H. Raynham, K. Clark, N.A. Doggett, A. King, and D.R. Higgs. 1997. The relationship between chromosome structure and function at a human telomeric region. *Nature Genet.* **15:** 252–257.

Fukagawa, T., Y. Nakamura, K. Okumura, M. Nogami, A. Ando, H. Inoko, N. Saitou, and T. Ikemura. 1996. Human pseudoautosomal boundary-like sequences—expression and involvement in evolutionary formation of the present-day pseudoautosomal boundary of human sex-chromosomes. *Hum. Mol. Genet.* **5:** 23–32.

Fukagawa, T., K. Sugaya, K.-I. Matsumoto, K. Okumura, A. Ando, H. Inoko, and T. Ikemura. 1995. A boundary of long-range G+C% mosaic domains in the human MHC locus: Pseudoautosomal boundary-like sequences exists near the boundary. *Genomics* **25:** 184–191.

Gabrielian, A. and S. Pongor. 1996. Correlation of intrinsic DNA curvature with DNA property periodicity. *FEBS Lett.***393:** 65–68.

Goyon, C. and M. Lichten. 1993. Timing of molecular events in meiosis in *Saccharomyces cerevisiae*: Stable heteroduplex DNA is formed late in meiotic prophase. *Mol. Cell. Biol.* **13:** 373–382.

Jeffreys, A.J. and R. Neumann. 1997. Somatic mutation processes at a human minisatellite. *Hum. Mol. Genet.* **6:** 129–136.

Jeffreys, A.J., K. Tamaki, A. MacLeod, D.G. Monckton, D.L. Neil, and J.A.L. Armour. 1994. Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.* **6:** 136–145.

Jeffreys, A.J., V. Wilson, and S.L. Thein. 1985. Hypervariable ''minisatellite'' regions in human DNA. *Nature* **314:** 67–73.

Jeffreys, A.J., J. Murray, and R. Neumann. 1998a. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell* **2:** 267–273.

Jeffreys, A.J., D.L. Neil, and R. Neumann. 1998b. Analysis of repeat instability at human minisatellites arising from meiotic recombination. *EMBO J.* **17:** 4147–4157.

Jurka, J., P. Klonowski, V. Dagman, and P. Pelton. 1996. CENSOR—a program for identification and elimination of repetitive elements from DNA-sequences. *Comput. Chem.* **20:** 119–121.

Matera, A.G., U. Hellmann, and C.W. Schmid. 1990. A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.* **10:** 5424–5432.

May, C.A., A.J. Jeffreys, and J.A.L. Armour. 1996. Mutation-rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (D16S309). *Hum. Mol. Genet.* **5:** 1823–1833.

Monckton, D.G., R. Neumann, T. Guram, N. Fretwell, K. Tamaki, A. MacLeod, and A.J. Jeffreys. 1994. Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nature Genet.* **8:** 162–170.

Ohta, K., T. Shibata, and A. Nicolas. 1994. Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J.* **13:** 5754–5763.

Rocco, V., B. deMassy, and A. Nicolas. 1992. The *Saccharomyces cerevisiae ARG4* initiator of meiotic gene conversion and its associated double-strand DNA breaks can be inhibited by transcriptional interference. *Proc. Natl. Acad. Sci.* **89:** 12068–12072.

Royle, N.J., R.E. Clarkson, Z. Wong, and A.J. Jeffreys. 1988. Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3:** 352–360.

Schmid, C.W. and W.R. Jelinek. 1982. The Alu family of dispersed repetitive sequences. *Science* **216:** 1065–1070.

Vergnaud, G., D. Mariat, F. Apiou, A. Aurias, M. Lathrop, and V. Lauthier. 1991. The use of synthetic tandem repeats to isolate new VNTR loci: Cloning of a human hypermutable sequence. *Genomics* **11:** 135–144.

Vergnaud, G., D. Gauguier, J.-J. Schott, D. Lepetit, V. Lauthier, D. Mariat, and J. Buard. 1993. Detection, cloning, and distribution of minisatellites in some mammalian genomes. In *DNA fingerprinting: State of the science* (ed. S.D.J. Pena, R. Chakraborty, J.T. Epplen, and A.J. Jeffreys), pp. 47–57. Birkhäuser Verlag, Basel, Switzerland.

Wohlrab, F., S. Chatterjee, and R.D. Wells. 1991. The herpes-simplex virus-1 segment inversion site is specifically cleaved by a virus-induced nuclear endonuclease. *Proc. Natl. Acad. Sci.* **88:** 6432–6436.

Wong, Z., V. Wilson, I. Patel, S. Povey, and A.J. Jeffreys. 1987. Characterization of a panel of highly variable minisatellites cloned from human DNA. *Annu. Rev. Hum.* **51:** 269–288.

Wu, T.C. and M. Lichten. 1994. Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science* **263:** 515–518.

Yeramian, E., F. Schaeffer, B. Caudron, P. Claverie, and H. Buc. 1990. An optimal formulation of the matrix-method in statistical-mechanics of one-dimensional interacting units—efficient iterative algorithmic procedures. *Biopolymers* **30:** 481–497.

Received September 28, 1998; accepted in revised form December 10, 1998.