Repeat instability at human minisatellites arising from meiotic recombination

Alec J. Jeffreys¹, David L. Neil and Rita Neumann

Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

¹Corresponding author
e-mail: ajj@le.ac.uk

Little is known about the role of meiotic recombination processes such as unequal crossover in driving instability at tandem repeat DNA. Methods have therefore been developed to detect meiotic crossovers within two different GC-rich minisatellite repeat arrays in humans, both in families and in sperm DNA. Both loci normally mutate in the germline by complex conversion-like transfer of repeats between alleles. Analysis shows that inter-allelic unequal crossovers also occur at both loci, although at low frequency, to yield simple recombinant repeat arrays with exchange of flanking markers. Equal crossovers between aligned alleles, resulting in recombinant alleles but without change in repeat copy number, also occur in sperm at a similar frequency to unequal crossovers. Both crossover and conversion show polarity in the repeat array and are co-suppressed in an allele showing unusual germline stability. This provides evidence that minisatellite conversion and crossover arise by a common mechanism, perhaps by alternative processing of a meiotic recombination initiation complex, and implies that minisatellite instability is a by-product of meiotic recombination in repeat DNA. While minisatellite recombination is infrequent, crossover rates indicate that the unstable end of a human minisatellite can act as a recombination warm-spot, even between sequence-heterologous alleles.

Keywords: conversion/meiosis/minisatellite/mutation/recombination

Introduction

Tandem repeat DNA is a common and frequently functionally significant component of higher eukaryotic genomes. However, little is known about the germline processes involved in the amplification, maintenance and diversification of repeat arrays. Possible mechanisms include recombination (Smith, 1976) and errors during replication (Levinson and Gutman, 1987). Recombination-based processes such as unequal crossover between alleles or sister chromatids have long been suspected to be major contributors to the evolution of tandem repeat DNA (Schalet, 1969; Smith, 1976; Petes, 1980; Jeffreys *et al.*, 1985). However, direct evidence is lacking in mammals due to the lack of appropriate systems for monitoring meiotic recombination.

GC-rich minisatellites provide useful loci for analysing recombination-based repeat turnover in humans. These loci, typically 1–20 kb long, frequently show substantial germline instability that can be monitored in families (Jeffreys *et al.*, 1988; Buard and Vergnaud, 1994) and by single sperm analysis (Jeffreys *et al.*, 1990, 1994). Clues concerning mutation processes can be obtained using repeat unit sequence variation in the array to characterize allele structure before and after mutation (Jeffreys *et al.*, 1991). These approaches have shown that germline instability is a complex process involving the copying and transfer of blocks of repeats from one allele to the other (Jeffreys *et al.*, 1991, 1994; Buard and Vergnaud, 1994; May *et al.*, 1996). These transfers are usually restricted to one end of the repeat array and appear to be influenced profoundly by upstream flanking DNA (Monckton *et al.*, 1994). This gene conversion-like process occurs only in the germline (Jeffreys and Neumann, 1997) and, given that it involves interactions between alleles, is likely to be related to meiotic recombination. However, there is no evidence to date that true meiotic crossovers occur within minisatellite repeat arrays (Wolff *et al.*, 1989; Jeffreys *et al.*, 1994). To determine the relationship between minisatellite mutation and crossing over, and more generally to begin to define at the molecular level the rules that govern meiotic recombination in humans, we have combined methods for studying repeat turnover with analysis of flanking marker exchange to detect meiotic crossovers within minisatellite repeat arrays.

Results

Germline recombination events at human minisatellite MS31

MS31 (D7S21) is a highly variable minisatellite with a mutation rate to new length alleles of 1.2 and 0.3% per gamete in the male and female germline, respectively (Jeffreys *et al.*, 1988; Henke *et al.*, 1993). Previous analysis of five *de novo* mutant alleles using minisatellite variant repeat mapping by polymerase chain reaction (MVR-PCR; Jeffreys *et al.*, 1991) showed that mutation involved the polar transfer of repeats between alleles but did not result in the exchange of upstream flanking DNA markers (Jeffreys *et al.*, 1994). This analysis was extended to additional pedigree mutants identified mainly by paternity testing laboratories (Henke *et al.*, 1993). Among 34 mutants (32 paternal, two maternal) informative for at least one flanking marker, there were two examples, both paternal in origin, that showed exchange of one or more upstream markers (Figure 1).

One MS31 mutant (Figure 1A) had arisen by simple unequal crossover between the two paternal alleles, yielding a recombinant repeat array with exchange of all informative 5' markers including a simple tandem repeat

allele no. repeats

Fig. 1. Germline recombination events at human minisatellite MS31. (A) Unequal crossover. (B) Flanking co-conversion. The four-state MVR code (Jeffreys *et al.*, 1994) and 5'-flanking haplotype of each paternal allele is shown, not to scale and with homozygous flanking sites omitted, together with the mutant paternal allele identified in the child. The -10 200 polymorphism is a (CCAT)_n STR with allele sizes indicated in bp. Allele 2 and the mutant in (A) share identical structures for at least 60 repeats further into the array (data not shown).

(STR) marker 10 kb upstream of the repeat array. Unequal crossover therefore occurs between MS31 alleles in the male germline at a frequency of $\sim 4 \times 10^{-4}$ per sperm (1.2% paternal mutation rate, one out of 32 paternal mutants showing unequal crossover). The second mutant (Figure 1B) was more complex and involved the in-phase transfer of four repeat units between alleles near the beginning of the repeat array, together with the proximal flanking marker only. This mutant is therefore the product, not of crossover, but of a patchwork gene conversion event encompassing the beginning of the repeat array and the immediate flanking DNA.

Screening mutant molecules of minisatellite MS32 in sperm DNA for unequal crossovers

Human minisatellite MS32 (D1S8) is another unstable locus with 29 bp GC-rich repeat units, a mean mutation rate of 0.8% per gamete and with most sperm mutations involving polar repeat transfers between alleles (Jeffreys *et al.*, 1988, 1994). MS32 is not used in paternity testing, and crossover analysis in pedigrees was, therefore, not feasible. Recombinant molecules were isolated instead directly from sperm DNA, by using allele-specific PCR (Newton *et al.*, 1989) to amplify selectively molecules that had exchanged DNA markers flanking each side of the repeat array. Sequence analysis revealed five single

nucleotide polymorphisms (SNPs) in Caucasians 5' to the repeat array, including three described previously (Monckton *et al.*, 1993), plus one SNP 3' to the array in an Alu element and a length-polymorphic STR based on degenerate (AAGGG)_n repeats derived from the poly(dA) tail of this element.

A Caucasian man was identified who was heterozygous at three of the 5'-flanking sites and at both 3' sites and who had a large (A) and small (B) MS32 allele (Figure 2A). Small pool PCR (SP-PCR) analysis (Jeffreys *et al.*, 1994) of sperm DNA from this individual showed a normal spectrum of length change mutants biased towards repeat gains (Figure 2Bi) and a typical sperm mutation rate of 1.4%. To simplify the detection of unequal crossovers, sperm DNA was size-fractionated (Jeffreys and Neumann, 1997) to recover gain mutant molecules that are the main products of sperm mutation. Batches of these mutants were then surveyed for any molecules showing exchanges of markers 5' and 3' to the repeat array. Fractionation also eliminated the bulk of progenitor allele molecules, minimizing the risk of inter-allelic jumping PCR creating artefactual 'recombinants' during DNA amplification.

Sperm DNA from this man was therefore digested with *ScaI*, which cleaves just outside the flanking polymorphic sites (Figure 2A), and fractionated by preparative gel

Fig. 2. Screening minisatellite MS32 mutant molecules in sperm DNA for unequal crossovers. (A) MS32 haplotypes in the semen donor, showing flanking base substitutional heterozygous positions and a length-heterozygous STR (alleles STR-S and STR-L). (Bi) Spectrum of new length mutants in sperm DNA from this man, established from the analysis of 10 800 progenitor molecules by small pool PCR (Jeffreys *et al.*, 1994). (Bii) Spectrum of mutants recovered following digestion of sperm DNA with *ScaI* and size fractionation. This spectrum was determined by dilution of 2% of the enriched fraction into 30 long PCRs followed by amplification using primers -0.9F and +1.4F. (C) Examples of long PCR products from mutant and progenitor molecules in these dilutions of the enriched fraction, detected by Southern blot hybridization following agarose gel electrophoresis. (D) A total of 128 batches of the enriched fraction each containing on average 29 mutant molecules plus 41 remaining allele B progenitor molecules were amplified using primers -0.9F and +1.4F; PCR products in a sample of these primary PCRs were analysed as in (C). (E) Scanning primary PCR products for recombinant alleles by reamplification using allele-specific primers. The examples shown used primers H2T plus HpC to amplify selectively PCR products derived from recombinant molecules with the 5' end of allele A fused to the 3' end of allele B. (F) 'Clonality' testing. Primary PCR products containing a H2T/HpC recombinant allele were retested by amplification with HpC plus either H2C (C) or H2T (T).

electrophoresis. Single molecule long PCR (Cheng *et al.*, 1994) analysis of the enriched fraction (Figure 2C) showed that it contained 5700 mutant molecules larger than allele B. There were also 8200 remaining molecules of allele B, but no molecules of the larger allele A. Comparison of the size distribution of these enriched mutant molecules

with the overall mutation spectrum in sperm (Figure 2B) showed that the mutant fraction contained a representative sample of mutants gaining 4–20 repeats derived from the equivalent of 2.3×10^6 sperm (1.15×10^6 amplifiable molecules of each allele).

Batches of these enriched mutant MS32 molecules were

pre-amplified by long PCR using universal (not allele-specific) primers, ensuring that all molecules, whether recombinant or not, were amplified with equal efficiency (Figure 2D). These primary PCRs were then surveyed systematically by allele-specific long PCR, using all combinations of allele-specific primers 5' and 3' to the array, to reamplify selectively any mutants carrying at least one flanking marker from the larger allele A (Figure 2E). A survey of 3700 enriched mutant molecules revealed 32 such mutants (Figure 2E). In each case, the hybridization signal from the mutant was similar to that of other mutants in the same primary PCR reamplified using non-recombinant combinations of flanking primers (Figure 2F). Furthermore, 26 mutants (81%) appeared to be 'clonal', in that they could only be reamplified using recombinant combinations of flanking primers (Figure 2F). The other six mutants showed relatively small length changes and were more likely to co-migrate with other mutants present in the primary PCR (Figure 2Bii). All recombinants lay within the known size distribution of mutants in the enriched fraction. Together, this provides strong evidence that these recombinant molecules are authentic and are not PCR artefacts arising either by inter-allelic jumping PCR or by base misincorporation during PCR leading to an apparent switch of flanking marker in a proportion of PCR products.

Properties of unequal crossover events at MS32

Structural analysis by MVR-PCR (Jeffreys *et al.*, 1991; Tamaki *et al.*, 1993) showed that three of these 32 mutants had arisen by simple deletion of repeats from the larger allele A with no exchange of flanking markers (data not shown); such large deletions have been seen previously in sperm DNA (Jeffreys *et al.*, 1990). All remaining mutants were recombinant (Figure 3) and most (27 out of 29) were different, as expected for products of meiotic recombination. Most recombinants consisted of the 5' end of allele A, including all 5'-flanking markers, fused to the 3' end of allele B to create a simple recombinant array without rearrangement of repeat units derived from allele A; repeat homogeneity in allele B prevented detection of possible rearrangements downstream of the breakpoint. These simple unequal crossovers contrasted with the frequently complex rearrangements seen in conversion mutants (Figure 3). Only three crossovers were more complex: mutant 15 showed an additional e-type repeat acquired upstream of the point of crossover, mutant 16 a Y- to e-type repeat switch in the segment derived from allele A, and mutant 17 an anomalous eEY motif at the crossover junction.

Knowing the number of sperm from which these recombinants were derived, it was possible to estimate the frequency of unequal crossover at MS32 resulting in gains of 4–20 repeats relative to allele B. This frequency (3×10^{-5} per sperm) is far lower (270-fold) than the frequency of conversion events which dominate MS32 instability in sperm. Remarkably, only one crossover was seen resulting in the simple fusion of the 5' end of allele B with the 3' end of allele A (mutant 22). Thus, unequal crossover at MS32 resulting in 4–20 repeat unit gains is strongly asymmetric, and implies that breakpoints are located preferentially towards the beginning of both interacting alleles. This asymmetry predicts that equivalent

Fig. 3. Structures of recombinant MS32 mutants. The flanking haplotypes of the progenitor alleles are shown, not to scale, together with allele structures established by MVR-PCR which assays two sites of base substitutional variation within MS32 repeat units (Tamaki *et al.*, 1993). Representative examples of conversion mutants were recovered by small pool PCR of unfractionated sperm DNA. Duplications are arrowed, and repeats of uncertain origin shown as smaller boxes. Recombinant and co-convertant molecules identified among size-enriched mutants are similarly shown, with the change in repeat number relative to allele B indicated. Positions of recombination breakpoints are shown by a filled circle where the location is unambiguous, and by a shaded circle if uncertain due to the unequal crossover occurring within a region of e-type repeats in both alleles; the range of possible positions is indicated by a horizontal line. For the 5' recombinants, six representative mutants (9–14) arising by simple unequal crossover are shown, together with all three mutants showing more complex rearrangements (15–17). The total number of recombinants of each class identified among 3700 mutant molecules is indicated on the right.

mutants similar in size to the large and unanalysable allele A would similarly be strongly biased towards crossovers fusing the beginning of the small allele B with the end of allele A.

While most recombinants had exchanged all upstream markers, there were four instances of incomplete exchange, three involving the proximal H2 site alone and one the more distal Sa site (Figure 3). Three mutants appeared to be simple conversion products resulting from the inter-allelic transfer of repeats plus proximal flanking DNA (mutants 18–20). The other mutant (number 21) involved a more complex patchwork transfer of DNA between alleles.

Fig. 4. Detection of unequal crossovers at minisatellite MS32 in unfractionated sperm DNA. Multiple aliquots of sperm DNA from the man used for size fractionation, each containing 29 ng of DNA digested with *ScaI* (4000 amplifiable molecules of each allele), were pre-amplified by long PCR using primers $-0.9F$ and $+1.4F$ (Figure 2A). These primary PCRs were reamplified using primers H2T and HpC (Figure 2A) to amplify selectively PCR products derived from recombinant molecules containing the 5' end of allele A plus the 3' end of allele B. PCR products in primary and secondary PCRs were detected by gel electrophoresis and Southern blot hybridization.

Detection of unequal crossovers at MS32 in unfractionated sperm DNA

Size fractionation greatly reduces the complexity of pools of molecules analysed for crossovers and largely eliminates the risk of inter-allelic jumping PCR generating false recombinants. However, unequal crossover mutants lying outside the size fraction will not be detectable. Analysis was therefore extended to unfractionated sperm DNA from the same man. Batches of DNA were pre-amplified as before with universal primers, then reamplified using allele-specific primers to amplify selectively the most common class of recombinant consisting of the 5' end of allele A fused to the 3' end of allele B (Figure 4). This procedure revealed putative unequal crossover products that increased in frequency linearly with the number (500–4000) of progenitor molecules per PCR. There was also a fainter signal from a complex heterodisperse set of 'recombinant' PCR artefacts arising from inter-allelic jumping PCR; above 4000 molecules, this signal became too great for the reliable detection of genuine recombinants.

A survey of 400 000 molecules from this man revealed 19 unequal crossover mutants, all of which were exchanged at both the M1 and H2 sites. As before, all mutants contained recombinant repeat arrays, in most cases (16 out of 19) without evidence of more complex rearrangements (data not shown). Twelve of these mutants resulted in gains of 4–20 repeats, giving an unequal crossover rate for this size class of 3×10^{-5} per sperm, identical to the estimate derived by size fractionation and providing strong evidence that these recombinants are authentic. The other recombinants ranged from a loss of nine repeats relative to the small allele B to a gain of 37 repeats. Larger gains were not detected, even though the jumping PCR artefacts

provided an internal control to show that recombinant molecules with gains of at least 100 repeats could be amplified efficiently.

The overall MS32 unequal crossover rate in this man was therefore 4.7×10^{-5} per sperm (19 recombinants per 400 000 amplifiable molecules of each progenitor allele). To determine whether this rate is typical, a second man heterozygous at the 5' M1 site and the 3' Hp site and with 600 and 33 repeat-long alleles C and D was analysed similarly for unequal crossovers resulting in the beginning of the larger allele becoming fused to the end of the smaller allele. A survey of 720 000 progenitor molecules yielded 31 unequal crossover products (data not shown), giving an unequal crossover rate of 4.3×10^{-5} per sperm for this individual, a rate very similar to that seen in the first man. Both alleles in this second man had informative internal structures (Figure 5C), and MVR mapping showed that most unequal crossovers (28 out of 31) were simple, with no additional rearrangements either 5' or 3' to the recombination breakpoint.

Unequal crossover is suppressed at a stable MS32 allele

A subset of MS32 alleles associated with a G→C transversion at the 'O1' site 48 bp upstream of the repeat array show unusual stability in sperm DNA (Monckton *et al.*, 1994). For one 38 repeat O1C allele, this suppression has been quantitated at 110-fold relative to O1G alleles (Jeffreys and Neumann, 1997). None of the residual mutants detected among 740 000 molecules of this allele showed simple recombinant arrays or flanking marker exchange (data not shown). This man also had a large (770 repeat) O1G-linked MS32 allele and further was heterozygous at the 5' M1 and 3' Hp sites. His sperm DNA was therefore screened for recombinants, exactly as for the second man. Again, no unequal crossover mutants were seen among 680 000 progenitor molecules (data not shown). Taken together, this suggests an unequal crossover frequency of $< 3 \times 10^{-6}$ ($P > 0.95$) at the O1C allele, and therefore that unequal crossover is suppressed at least 15-fold relative to the O1G alleles in the first two men analysed. The lack of unequal crossovers seen in unfractionated sperm DNA from the O1C/G individual also provides strong evidence that crossovers recovered from O1G homozygotes using this method are authentic and not PCR artefacts.

Distribution of recombination breakpoints in MS32

The location of unequal crossover breakpoints in MS32 mutants recovered from the first and second man are shown in Figure 5. In most cases, breakpoints could be mapped to a resolution of a single repeat unit (29 bp). To clarify further discussion, we define the allele contributing to the beginning of the recombinant as the 'donor' and to the end of the recombinant as the 'recipient'. Breakpoints in recipient alleles were located preferentially towards the beginning of the array, though allele D showed a second central hot-spot. Donor allele breakpoints also showed 5' polarity; while this polarity may be artefactual for size-constrained mutants recovered by size fractionation (Figure 5A), it is genuine for the more representative mutants isolated from unfractionated DNA (Figure 5B and C) in which all breakpoints are restricted to the first

Fig. 5. Location of unequal crossover breakpoints within the MS32 repeat array. Breakpoints are indicated by filled circles for unequal crossovers resulting in gains, empty circles for deletions, and shaded circles for 5' co-conversion events, with lines (dashed for co-conversions) connecting the breakpoints in each mutant. Positional uncertainties in the donor allele are indicated by horizontal lines; corresponding uncertainties for the recipient allele are omitted for clarity. (A) Recombinants recovered following size enrichment of sperm DNA (Figure 2 and 3). (B) Recombinants recovered from unfractionated sperm DNA from the same man. (C) Recombinants from a second man heterozygous for haplotypes C (M1T–600 repeats–HpC) and D (M1C–33 repeats–HpT) recovered from unfractionated sperm DNA using allele-specific primers M1T and HpT.

54 repeats. Donor allele A showed evidence of clustering of breakpoints at repeat 2 and more noticeably at repeats 10 and 18, suggesting a possible eight repeat periodicity to recombination. The additional apparent cluster at repeat 26 represents crossovers between sequence-identical regions in alleles A and B which cannot be localized with precision. Clustering is not evident in the second donor allele C. Finally, while the unequal crossover rates in these two men are very similar, the detailed patterns of allelic interaction are different (Figure 5B and C), with a significant proportion of events in the second but not first man yielding deletions relative to the recipient allele (12 out of 31 versus one out of 19, $P = 0.016$).

Equal crossover events in sperm DNA

The second potential class of recombination event between MS32 repeat arrays involves alleles of different lengths

pairing in-register over the beginning of the array prior to exchange, yielding an isometric recombinant array equal in length to the progenitor allele (Figure 6A). Such crossovers cannot be detected by size enrichment nor by analysis of unfractionated sperm DNA, given the potential of jumping PCR in the flanking DNA to yield isometric 'recombinants'. Allele B from the first man (Figure 2) was therefore separated from the other allele by size fractionation, pre-amplified using universal primers, then subjected to two rounds of allele-specific PCR (Figure 6B) to achieve the necessary level of allele specificity. Fifteen quantal isometric recombinants were detected among 370 000 progenitor molecules, each showing a similar signal intensity to the unequal crossover mutants still present in this fractionated sperm DNA. In contrast, no false recombinants were obtained from a 55 repeat allele similarly purified from homozygous sperm DNA and included as an internal control (Figure 6B), estab-

Fig. 6. Detection of equal crossovers at MS32 by analysis of sperm DNA. (A) Haplotypes of the man tested, also used for unequal crossover analysis (Figure 2). Equal crossovers will yield recombinants equal in length to allele B but flanked on the 5' side by markers from allele A. Reciprocal recombinants equal in length to allele A have not been analysed. (B) Allele B was separated from allele A following digestion of sperm DNA with *ScaI*. Sperm DNA from a second man homozygous for M1C, H2C and HpC was processed similarly to yield the control 55 repeat allele E. Alleles B and E were mixed and pre-amplified using primers $-0.9F$ and $+1.4F$ (2000 molecules of each allele per PCR). These primary PCRs were reamplified using allele-specific primers M1T and HpC to enrich for recombinants. The secondary PCRs were amplified further using the nested allele-specific primer H2T plus the universal primer 32E to yield tertiary PCR products. PCR products were analysed by gel electrophoresis and Southern blot hybridization. The examples shown reveal two unequal crossovers still present in the purified allele B fraction and detectable in both the secondary and tertiary PCRs, one recombinant equal in length to allele B but only detectable in the tertiary PCR, and no false recombinants derived from the control allele E. Prolonged autoradiography (not shown) revealed a faint and uniform signal from allele E in all tertiary PCRs arising from mispriming. (C) Recombination breakpoints in seven equal crossovers detected among 370 000 molecules. Variant repeats are encoded as in Figure 5.

lishing that most or all of these isometric recombinants were authentic. Seven recombinants had arisen by equal crossover within the repeat array, with recombination breakpoints again located towards the beginning of the array (Figure 6C). Equal crossovers in MS32 therefore do occur, at a frequency of $\sim 1.9 \times 10^{-5}$ (7 out of 370 000) per sperm and $\sim 40\%$ of the unequal crossover rate. The remaining eight recombinants had MS32 repeat arrays identical to the progenitor allele B, indicating that crossover had occurred in the 147 bp interval between H2 and the beginning of the MS32 array. This suggests that recombinational proficiency at MS32 may not necessarily be restricted to the repeat array but may extend into flanking DNA.

A polar recombination warm-spot in minisatellite MS32

Unequal and equal crossover data were combined to establish the profile of recombination events along MS32

alleles and to compare their frequency with that expected if crossovers are distributed at random in the human genome (Figure 7B). This analysis revealed a gradient of recombinational proficiency in both donor and recipient alleles, with the beginning of the repeat array showing an ~ 15 -fold enhancement of recombination relative to the genome average, and the end of the array being recombinationally inert. The much more frequent MS32 conversion events in sperm DNA also show polarity but more extreme than for recombination, with most conversion breakpoints in recipient alleles mapping to the first few repeats (Figure 3 and 7A).

Discussion

Recombination and repeat instability

This work establishes that fine-scale characterization of recombination events in human DNA is possible by single

Fig. 7. Conversion and crossover efficiency within the MS32 repeat array. (A) Distribution of breakpoints in the recipient allele in sperm mutants arising from inter-allelic conversion, determined from 126 conversion events characterized in five different alleles 22–63 repeats long (Figure 3; Jeffreys *et al.*, 1994; Jeffreys and Neumann, 1997; unpublished data). Breakpoints are binned into six-repeat intervals; 67% of breakpoints map within the first two repeats of the array. (B) Unequal crossover breakpoints (light shading) in recombinants recovered from alleles A+B and C+D by analysis of unfractionated sperm DNA (Figure 5), and equal crossover breakpoints (dark shading) from alleles A+B (Figure 6C) were combined and used to estimate the frequency of crossover in six-repeat (174 bp) intervals along the repeat array. These frequencies are shown as the ratio of the observed number of recombinants per interval to that expected from the genome average male recombination frequency of 8.9×10^{-3} per Mb (1.55×10^{-6} per six-repeat interval, dotted line) (Weissenbach *et al.*, 1992; Gyapay *et al.*, 1994). Donor and recipient alleles are defined as the alleles contributing to the 5' and 3' half of the recombinant, respectively.

sperm analysis and provides the first evidence that inter-allelic crossover plays a role in minisatellite repeat turnover in the human germline, albeit at a frequency far lower than the complex conversion-like repeat transfers between alleles that dominate sperm mutation at MS31 and MS32. While crossovers are generally much simpler in structure than convertants, with simple recombinant repeat arrays containing unarranged segments from the two recombining alleles, they nevertheless appear to arise by a common pathway. Thus, both crossover and conversion share the property of polarity within the repeat array. Furthermore, both processes are suppressed at a stable OIC-linked MS32 allele. Previous analysis of this allele has shown that residual sperm mutation is still dominated by inter-allelic conversion, suggesting that the OIC variant blocks an early stage in the establishment of a conversion complex but does not affect downstream processing (Jeffreys and

Neumann, 1997). If correct, then co-suppression of meiotic crossover and conversion suggests that both processes originate from the same initiation complex. This in turn implies that minisatellite instability occurs specifically at meiosis and is a by-product of meiotic recombination.

Alternative processing

We have proposed previously, by analogy with yeast recombination models (Szostak *et al.*, 1983), that minisatellite conversion is initiated by a controlling flanking DNA element, blocked by the OIC variant at MS32, that serves to introduce a double strand break or staggered nicks into the beginning of the repeat array (Jeffreys *et al.*, 1994; Buard and Jeffreys, 1997). The resulting gap is then bridged by strand invasion from the other allele to provide a template for gap repair, leading to conversion. Strand invasion would also create a recombination junction which, if isomerized, could be resolved to yield a meiotic crossover of the type identified in this study. This implies that recombination events are initiated at very high frequency (>1% of meioses) near the beginning of minisatellite MS32, but that almost all (>99.6%) fail to be resolved to yield recombinants but instead are processed to generate convertants (Buard and Vergnaud, 1994; Jeffreys *et al.*, 1994). There is evidence in lower eukaryotes that conversion rather than recombination can dominate turnover of repeat DNA (Fogel *et al.*, 1984), though minisatellite MS32 integrated near a meiotic recombination hot-spot in yeast shows meiosis-specific repeat instability with inter-allelic recombination events more frequent than conversion (Appelgren *et al.*, 1997). It remains unclear why so few recombination initiation events in minisatellites in humans yield crossovers; one possibility is that mismatch repair systems might identify and abort recombination between heterologous repeats (Rayssiguier *et al.*, 1989; Borts *et al.*, 1990) (but see below).

Rules of engagement

Fine-scale mapping of minisatellite breakpoints makes it possible to explore the rules that govern inter-allelic recombination. Breakpoint locations are influenced by polarity, though different allele pairs can interact differently; thus allele A is usually displaced to the left of allele B before unequal crossover, while alleles C and D show leftward and rightward displacements with similar frequency (Figure 5B and C). Breakpoints do not appear to be influenced by sequence differences between recombining alleles. Analysis of MVR codes around breakpoints in the informative MS32 alleles C and D (Figure 5C) shows that repeat segments in the donor and recipient allele adjacent to the breakpoint are typically 3.5% diverged in DNA sequence. This divergence is typical for any segment of allele C and D compared at random and does not decline in repeats immediately abutting the breakpoint (data not shown). Similarly, MS32 alleles A and B share regions of sequence (MVR) identity as well as regions of substantial divergence; again, there is no tendency for crossovers to be targeted disproportionately towards regions of complete homology (Figure 5A and B). Thus, recombination appears to occur with equal efficiency irrespective of the degree of sequence mismatch between alleles at the point of crossover. This is consistent with data on a hot-spot within the mouse major histocompatibility

complex, where the frequency of meiotic recombination appears not to be influenced substantially by base substitutional divergence between haplotypes (Yoshino *et al.*, 1995), but contrasts with meiotic recombination in yeast which is reduced by sequence divergence between recombining alleles (Borts and Haber, 1987; Borts *et al.*, 1990).

Given the number of degrees of freedom for unequal crossover, it is unlikely that two alleles would recombine by chance in-register to yield an equal crossover product. It is, therefore, highly significant that equal crossovers do occur at MS32, at ~40% of the rate of unequal crossover. A similar in-register phenomenon occurs in minisatellite conversion, where the repeat segment frequently is transferred from donor to recipient allele in-phase with respect to the beginning of the repeat array (Figure 3) (Jeffreys *et al.*, 1994). Registration of both conversion and crossover events again points to the potential importance of upstream flanking DNA in controlling allele pairing.

Crossover and co-conversion

The unequal crossovers detected at minisatellites MS31 and MS32 appear to be true meiotic recombination events with exchange of all flanking markers. Analysis of reciprocal products of recombination at MS32 has not yet been carried out, due to the lack of semen donors informative for flanking variants and with both alleles sufficiently short for recombination analysis. For the MS31 recombinant, exchange of flanking markers extended at least 10 kb upstream of the unstable end of the array. For MS32, flanking markers were limited to 832 bp upstream of the repeat array, and the possibility of conversions transferring repeats plus all proximal 5' markers cannot be excluded. However, 86% of exchanges involved all flanking markers, inconsistent with a monotonic gradient of co-conversion extending upstream of the repeat array and suggesting that most or all of these events represent true meiotic crossovers.

Flanking co-conversion without exchange of distal markers does occur at both minisatellites examined, though the rate for MS32 appears to be very low, at ~15% of the crossover rate. Some co-conversions appear to be the simple product of inter-allelic transfer of a contiguous segment of proximal flanking DNA plus adjacent repeats, but others are more complex. The MS31 mutant (Figure 1B) shows an anomalous repeat unit in the transferred segment apparently retained from the recipient allele. The simplest interpretation is that conversion was initiated in this region, followed by branch migration of a recombination junction to generate heteroduplex DNA which was repaired segmentally on either the donor or recipient strand to give the patchwork convertant observed (Borts and Haber, 1987). A similar phenomenon could underlie MS32 mutant 21 (Figure 3) that shows co-conversion of the middle but not distal or proximal 5'-flanking markers, as well as some of the crossovers restricted to the repeat array but with additional repeat switches located several repeat units away from the point of crossover (e.g. mutant 16, Figure 3).

Recombination warm-spot

Assuming that crossovers at minisatellite MS32 represent full meiotic recombination events, it is possible to estimate the recombination rate across MS32 repeat arrays at

$\sim 6.3 \times 10^{-5}$, with 32% of events arising by equal crossover. For MS31, only one crossover has been seen, leaving the unequal crossover rate uncertain but in excess of 2×10^{-5} ($P > 0.95$) and possibly similar to MS32. While these rates are low, the concentration of crossovers into such a short segment of the repeat array results in the beginning of MS32 being a recombination warm-spot. This is the first evidence that minisatellites might serve as meiotic recombination hot-spots, and rekindles earlier speculation (Jeffreys *et al.*, 1985) concerning the involvement of these loci, preferentially (e.g. MS31) but not exclusively (e.g. MS32) located in subterminal recombination-proficient regions of the human genome (Royle *et al.*, 1988), in meiotic phenomena such as homologue recognition and the initiation of recombination.

Materials and methods

PCR primers

MS31 primers were as follows. Universal primers: 31GGR, 5'-GCC TCA CAC AGG ACT TCA GC-3'; 31R, 5'-AAG TCC CGG GTG AGA GCG TG-3'; 31FF, 5'-CTG AGG TTC AGC CAC TTG CC-3'; 31MM, 5'-GCA TGG AAC CTT GGA CCT GC-3'; 31NNR, 5'-TGT TTG TTC CCT GCC GCG TC-3'. Forced mismatch primer: 31FspI, 5'-CTT TTC CAG ATC CTT CTC AAT GCG C-3'. Allele-specific primers: 31-759C, 5'-GGC GCT CAT GCC TGG AAC-3'; 31-759A, 5'-GGC GCT CAT GCC TGG AAA-3'; 31-457C, 5'-CAT CTA TAT TTC TTT CTG AGT ATA AAA C-3'; 31-457T, 5'-CAT CTA TAT TTC TTT CTG AGT ATA AAA T-3'; 31-457RT, 5'-CAG ATC CTT CTC AAT GGG CA-3'; 31-457RC, 5'-CAG ATC CTT CTC AAT GGG CG-3'.

MS32 primers were as follows. Universal primers: -0.9F, 5'-CAG AGG TCA CCA AGC TAC GG-3'; +0.6R, 5'-GTC AAG GCA GAT TTG AAG GGA G-3'; +1.4F, 5'-GGT AGC CAC ATC TCT GCA TGA TCT C-3'; 32E, 5'-CTT CCT CGT TCT CTT CAG CCC TAG-3'; 32SR, 5'-GCT CTG GCC ATT GCT CCA TC-3'. Forced mismatch primer: 32SDF, 5'-GAT ATT TTT GTG TAT GAA CAT CYA CTT GAT-3'. Allele-specific primers: M1C, 5'-GGT TTT ACT GGA TTA TGG CCA TAC C-3'; M1T, 5'-TAC TGG ATT ATG GCC ATA CT-3'; SaC, 5'-GAT ATT TTT GTG TAT GAA CAT CCA CTT CAT C-3'; SaT, 5'-GAT ATT TTT GTG TAT GAA CAT CCA CTT CAT T-3'; H2C, 5'-TGA TGC GTC GTT CCC GTA TC-3'; H2T, 5'-TGA TGC GTT GTT CCC GTA TT-3'; HpC, 5'-GCA GCC TCC GCC TCC CG-3'; HpT, 5'-TGC AGC CTC CGC CTC CCA-3'.

Polymorphism typing

MS31 variants at -109 and -220 were assayed as described elsewhere (Neil and Jeffreys, 1993; Huang *et al.*, 1996). The -457 and -759 polymorphisms were assayed by PCR amplification of 5'-flanking DNA from genomic DNA using primers 31GGR and 31R, followed by reamplification using the nested primers 31FF plus 31FspI to generate a 526 bp product containing a mismatch forced by 31FspI to create an *FspI* polymorphism at the -457 site. The -457 and -759 variants were assayed by digestion with *FspI* and *NlaIV*, respectively. Linkage phase between these sites and the MS31 repeat array was determined by allele-specific MVR-PCR (Neil and Jeffreys, 1993; Huang *et al.*, 1996) using primers 31-759C/A and 31-457C/T. Length polymorphism at the -10 kb STR was assayed by PCR amplification using primers 31MM and 31NNR. Haplotypes between this STR and the -457 variant were established by allele-specific long PCR amplification of 5'-flanking DNA using 31MM and the reverse orientation allele-specific primers 31-457RT or 31-457RC, followed by reamplification of the STR using 31MM and 31NNR.

MS32 5' variants M1C/T and SaC/T were typed by PCR amplification of 693 bp 5'-flanking DNA from genomic DNA using primers -0.9F and 32SR. M1C/T creates an *MboII* restriction fragment length polymorphism (RFLP) in this amplicon. Reamplification with 32SR plus the nested primer 32SDF produced a 397 bp PCR product containing a forced mismatch generating an *MboI* RFLP at the SaC/T site. Other 5' variants were assayed as described elsewhere (Monckton *et al.*, 1993). The 3' STR was assayed by amplification of genomic DNA using primers +0.6R and +1.4F to generate 730 and 840 bp amplicons from STR-S and STR-L alleles, respectively. HpC/T creates an *HpaII* RFLP

within this PCR product. Full 5'- and 3'-flanking haplotypes were determined using allele-specific primers directed to each flanking heterozygous position plus a universal primer on the other side of the MS32 repeat array to amplify selectively one or other MS32 allele and thus link flanking variants to MS32 allele length.

Size fractionation of sperm DNA

The preparation of sperm DNA and all subsequent manipulations were carried out under conditions designed to minimize the risk of contamination (Jeffreys *et al.*, 1990, 1994). To enrich for length gain mutants, three batches of 13 µg of sperm DNA were digested with *ScaI*, electrophoresed through a 40 cm long 0.9% SeaKem HGT (FMC BioProducts) agarose gel, and multiple fractions ranging from 3.3 to 4.3 kb and including the smaller 3.4 kb progenitor allele were collected from each batch (Jeffreys and Neumann, 1997). MS32 molecules from 0.5% of each fraction were amplified and analysed by gel electrophoresis (Jeffreys and Neumann, 1997) to identify fractions either enriched in mutant molecules or containing the progenitor allele. Appropriate mutant-enriched fractions were pooled and the frequency and spectrum of mutants analysed by Poisson analysis of multiple limiting dilutions amplified by single molecule long PCR (see below). Similarly, the number of amplifiable progenitor molecules in the progenitor allele pool was estimated by Poisson analysis of 80 limiting dilution long PCRs, each containing on average 0.5–1.0 amplifiable molecules (Jeffreys *et al.*, 1994). All dilutions of DNA were into 5 mM Tris–HCl, pH 7.5 plus 5 µg/ml carrier herring sperm DNA.

Long PCR and MVR-PCR mapping of recombinants

Long PCRs were performed in 9 µl reactions on a GeneAmp PCR system 9600 thermal cycler (Perkin Elmer Cetus) using the buffer described previously (Jeffreys *et al.*, 1990) supplemented with 12.5 mM Tris base, plus 0.2 µM of each primer, 1 µg/ml carrier herring sperm DNA (Jeffreys and Neumann, 1997), 0.07 U/µl *Taq* polymerase (Advanced Biotechnologies) and 0.007 U/µl *Pfu* polymerase (Stratagene). Primary PCRs using primers –0.9F and +1.4F were performed at 96°C for 20 s, 60°C for 30 s, 70°C for 4 min for 23 cycles (size-enriched mutant analysis) or 18 cycles (unfractionated sperm DNA, equal crossover analyses), followed by a chase at 56°C for 1 min and 70°C for 10 min for one cycle. Allele-specific secondary PCRs were seeded with 1 µl of a 1:40 dilution of primary PCR products in 5 mM Tris–HCl pH 7.5, 5 µg/ml carrier herring sperm DNA and amplified by touchdown PCR (Don *et al.*, 1991) optimized for each primer combination. Cycling was at 96°C for 20 s, A°C for 30 s, 70°C for 4 min, with the annealing temperature A being varied as follows (annealing temperature/number of cycles at that temperature): M1C plus HpC 64°C/5, 63°C/10, 62°C/8; M1T plus HpC 61°C/6, 59°C/12, 58°C/7; SaC/T plus HpC 66°C/8, 65°C/11, 64°C/4; H2C/T plus HpC 68°C/3, 67°C/18; H2C/T plus HpT 68°C/6, 67°C/6, 65°C/10. Tertiary PCRs in the equal crossover analysis were seeded similarly with 1 µl of a 1:40 diluted secondary PCR products and amplified by touchdown PCR using primers H2T plus E (69°C/2, 68°C/4, 66°C/8). Aliquots (3 µl) of each PCR were analysed by electrophoresis in 20 cm long 0.8% SeaKem LE (FMC BioProducts) agarose gels and PCR products detected by Southern blot hybridization with ³²P-labelled MS32 repeat probe (Wong *et al.*, 1987). Recombinant mutants in secondary or tertiary PCRs were if necessary purified away from progenitor allele PCR products as described elsewhere (Jeffreys *et al.*, 1994; Jeffreys and Neumann, 1997), and subjected to four-state MVR-PCR mapping (Tamaki *et al.*, 1993).

Acknowledgements

We thank Mike Webb (Cellmark Diagnostics, UK), Lisa Forman (Cellmark Diagnostics, USA), Margurethe Stenersen (Rettsmedisinsk Institutt, UiO Rikshospitalet, Oslo, Norway), Gunilla Holmlund (State Institute of Forensic Medicine, Linköping, Sweden) and Lotte Henke (Institut für Blutgruppenforschung, Düsseldorf, Germany) for providing DNA samples from families with MS31 mutants. We are grateful to Jane Blower for supplying semen samples from anonymous donors, to Kathryn Lilley and Stuart Bayliss for oligonucleotide synthesis and assistance with automated sequencing, to Yuri Dubrova for statistical analyses 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.

References

- Appelgren, H., Cederberg, H. and Rannug, U. (1997) Mutations at the human minisatellite MS32 integrated in yeast occur with high frequency in meiosis and involve complex recombination events. *Mol. Gen. Genet.*, **256**, 7–17.
- Borts, R.H. and Haber, J.E. (1987) Meiotic recombination in yeast—alteration by multiple heterozygosities. *Science*, **237**, 1459–1465.
- Borts, R.H., Leung, W.Y., Kramer, W., Kramer, B., Williamson, M., Fogel, S. and Haber, J.E. (1990) Mismatch repair-induced meiotic recombination requires the *pms1* gene-product. *Genetics*, **124**, 573–584.
- Buard, J. and Jeffreys, A.J. (1997) Big, bad minisatellites. *Nature Genet.*, **15**, 327–328.
- Buard, J. and Vergnaud, G. (1994) Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *EMBO J.*, **13**, 3203–3210.
- Cheng, S., Fockler, C., Barnes, W.M. and Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl Acad. Sci. USA*, **91**, 5695–5699.
- Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. (1991) Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.*, **19**, 4008.
- Fogel, S., Welch, J.W. and Louis, E.J. (1984) Meiotic gene conversion mediates gene amplification in yeast. *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 55–65.
- Gyapay, G. *et al.* (1994) The 1993–94 Génethon human genetic linkage map. *Nature Genet.*, **7**, 246–339.
- Henke, J., Fimmers, R., Baur, M.P. and Henke, L. (1993) DNA-minisatellite mutations: recent investigations concerning distribution and impact on parentage testing. *Int. J. Leg. Med.*, **105**, 217–222.
- Huang, X.-L., Tamaki, K., Yamamoto, T., Suzuki, K., Nozawa, H., Uchihi, R., Katsumata, Y. and Neil, D.L. (1996) Analysis of allelic structures at the D7S21 (MS31A) locus in the Japanese, using minisatellite variant repeat mapping by PCR (MVR-PCR). *Ann. Hum. Genet.*, **60**, 271–279.
- Jeffreys, A.J. and Neumann, R. (1997) Somatic mutation processes at a human minisatellite. *Hum. Mol. Genet.*, **6**, 129–136.
- Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Hypervariable ‘minisatellite’ regions in human DNA. *Nature*, **314**, 67–74.
- Jeffreys, A.J., Royle, N.J., Wilson, V. and Wong, Z. (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature*, **332**, 278–281.
- Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying allelic variation and mutation by single molecule analysis. *Cell*, **60**, 473–485.
- Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L. and Monckton, D.G. (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature*, **354**, 204–209.
- Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L. and Armour, J.A.L. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.*, **6**, 136–145.
- Levinson, G. and Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.*, **4**, 203–221.
- May, C.A., Jeffreys, A.J. and Armour, J.A.L. (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., Tamaki, K., MacLeod, A., Neil, D.L. and Jeffreys, A.J. (1993) Allele-specific MVR-PCR analysis at minisatellite D1S8. *Hum. Mol. Genet.*, **2**, 513–519.
- Monckton, D.G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., MacLeod, A. and Jeffreys, A.J. (1994) Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nature Genet.*, **8**, 162–170.
- Neil, D.L. and Jeffreys, A.J. (1993) Digital DNA typing at a second hypervariable locus by minisatellite variant repeat mapping. *Hum. Mol. Genet.*, **2**, 1129–1135.
- Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C. and Markham, A.F. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.*, **17**, 2503–2516.
- Petes, T.D. (1980) Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA. *Cell*, **19**, 765–774.
- Rayssiguier, C., Thaler, D.S. and Radman, M. (1989) The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature*, **342**, 396–401.

- Royle,N.J., Clarkson,R.E., Wong,Z. and Jeffreys,A.J. (1988) Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics*, **3**, 352–360.
- Schalet,A. (1969) Exchanges at the bobbed locus of *Drosophila melanogaster*. *Genetics*, **63**, 133–153.
- Smith,G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science*, **191**, 528–535.
- Szostak,J.W., Orr-Weaver,T.L., Rothstein,R.J. and Stahl,F.W. (1983) The double-strand-break repair model for recombination. *Cell*, **33**, 25–35.
- Tamaki,K., Monckton,D.G., MacLeod,A., Allen,M. and Jeffreys,A.J. (1993) Four-state MVR-PCR: increased discrimination of digital DNA typing by simultaneous analysis of two polymorphic sites within minisatellite variant repeats at DIS8. *Hum. Mol. Genet.*, **2**, 1629–1632.
- Weissenbach,J., Gyapay,G., Dib,C., Vignal,A., Morissette,J. and Millasseau,P. (1992) A second-generation linkage map of the human genome. *Nature*, **359**, 794–801.
- Wolff,R.K., Plaetke,R., Jeffreys,A.J. and White,R. (1989) Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. *Genomics*, **5**, 382–384.
- Wong,Z., Wilson,V., Patel,I., Povey,S. and Jeffreys,A.J. (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.*, **51**, 269–288.
- Yoshino,M., Sagai,T., Lindahl,K.F., Toyoda,Y., Moriwaki,K. and Shiroishi,T. (1995) Allele-dependent recombination frequency: homology requirement in meiotic recombination at the hot spot in the mouse major histocompatibility complex. *Genomics*, **27**, 298–305.

*Received March 31, 1998; revised May 18, 1998;
accepted May 20, 1998*