# Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA

Shuji Mita[1], Rosario Rizzuto[1], Carlos T.Moraes[2], Sara Shanske[1], Enrica Arnaudo[1], Gian Maria Fabrizi[1], Yasutoshi Koga[1], Salvatore DiMauro[1] and Eric A.Schon[1,2]*
Departments of [1]Neurology and [2]Genetics and Development, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, NY 10032, USA

Received October 4, 1989; Revised and Accepted December 18, 1989

## ABSTRACT

**Large-scale deletions of mitochondrial DNA (mtDNA) have been described in patients with progressive external ophthalmoplegia (PEO) and ragged red fibers. We have determined the exact deletion breakpoint in 28 cases with PEO, including 12 patients already shown to harbor an identical deletion; the other patients had 16 different deletions. The deletions fell into two classes. In Class I (9 deletions; 71% of the patients), the deletion was flanked by perfect direct repeats, located (in normal mtDNA) at the edges of the deletion. In Class II (8 deletions; 29% of patients), the deletions were not flanked by any obviously unique repeat element, or they were flanked by repeat elements which were located imprecisely relative to the breakpoints. Computer analysis showed a correlation between the location of the deletion breakpoints and sequences in human mtDNA similar to the target sequence for *Drosophila* topoisomerase II. It is not known how these deletions originate, but both slipped mispairing and legitimate recombination could be mechanisms playing a major role in the generation of the large mtDNA deletions found in PEO.**

## INTRODUCTION

Mitochondrial diseases include myopathies and multisystem disorders that are defined by biochemical abnormalities of mitochondria, and are often accompanied by ragged red fibers (RRF), a hallmark of mitochondrial proliferation in muscle biopsies. A common symptom of mitochondrial diseases is paralysis of extraocular muscles, called progressive external ophthalmoplegia (PEO). Some patients have PEO alone or associated with limb weakness, a clinically mild condition called ocular myopathy (OM). Other patients have a severe multisystem disorder, called Kearns-Sayre Syndrome (KSS), characterized by the invariant triad of PEO, pigmentary retinal degeneration, and onset before age 20, plus one of the following: ataxia, heart block or cerebrospinal fluid protein above 100 mg/dl. Most patients with OM or KSS have been 'sporadic', with no other cases in the family.

We have recently found that essentially all patients with KSS, and many patients with OM and RRF, have large deletions of muscle mtDNA, ranging in size from 1.3 to 7.6 kb (1–3). The deletions in 30 of the 33 patients described by us were mapped by analyzing each deleted genome for the absence of known restriction sites on the mtDNA map. The 30 patients were found to harbor at least 17 types of deletion: 12 patients apparently had an identical deletion (the 'common' deletion), while at least 16 different deletions were identified among the remaining 18 patients. Although deletions of mtDNA had been reported in lower eukaryotes and plants, large deletions of mtDNA had not been observed in humans until 1988, when Holt *et al.* (4) described them in muscle biopsies from patients with mitochondrial myopathy.

To understand the mechanism of deletion, we had previously analyzed at the nucleotide level the deletion breakpoint of the twelve PEO patients with the 'common' deletion (3). In normal mtDNA, the region of the 'common' deletion was flanked by 13-bp direct repeats, suggesting that legitimate recombination or slipped mispairing were possible causes of mtDNA deletions (3). Subsequently, the same deletion was also found in a patient with Pearson's marrow/pancreas syndrome (5).

We have now sequenced the deletion breakpoints in 16 other patients with PEO, in order to elucidate further the mechanism(s) of deletion of large regions of human mtDNA, and to determine whether recombination via flanking direct repeats is a generalized phenomenon in PEO, or is restricted only to the hotspot represented by the 'common' deletion. Each patient has been found to harbor a different, unique, deletion. We report here the features of these breakpoints and discuss possible mechanisms leading to the formation of large deletions of mtDNA in these diseases.

## MATERIALS AND METHODS

### Mapping of deleted mtDNAs

Preparation of DNAs and mapping of the deletions in each patient have been described previously (1,2).

### PCR amplification and DNA sequencing

Based on the mapping data, a segment of each patient's mtDNA encompassing the deletion was amplified by the polymerase chain

* To whom correspondence should be addressed at Istituto di Patologia Generale, Via Loredan 16, 35100 Padova, Italy.

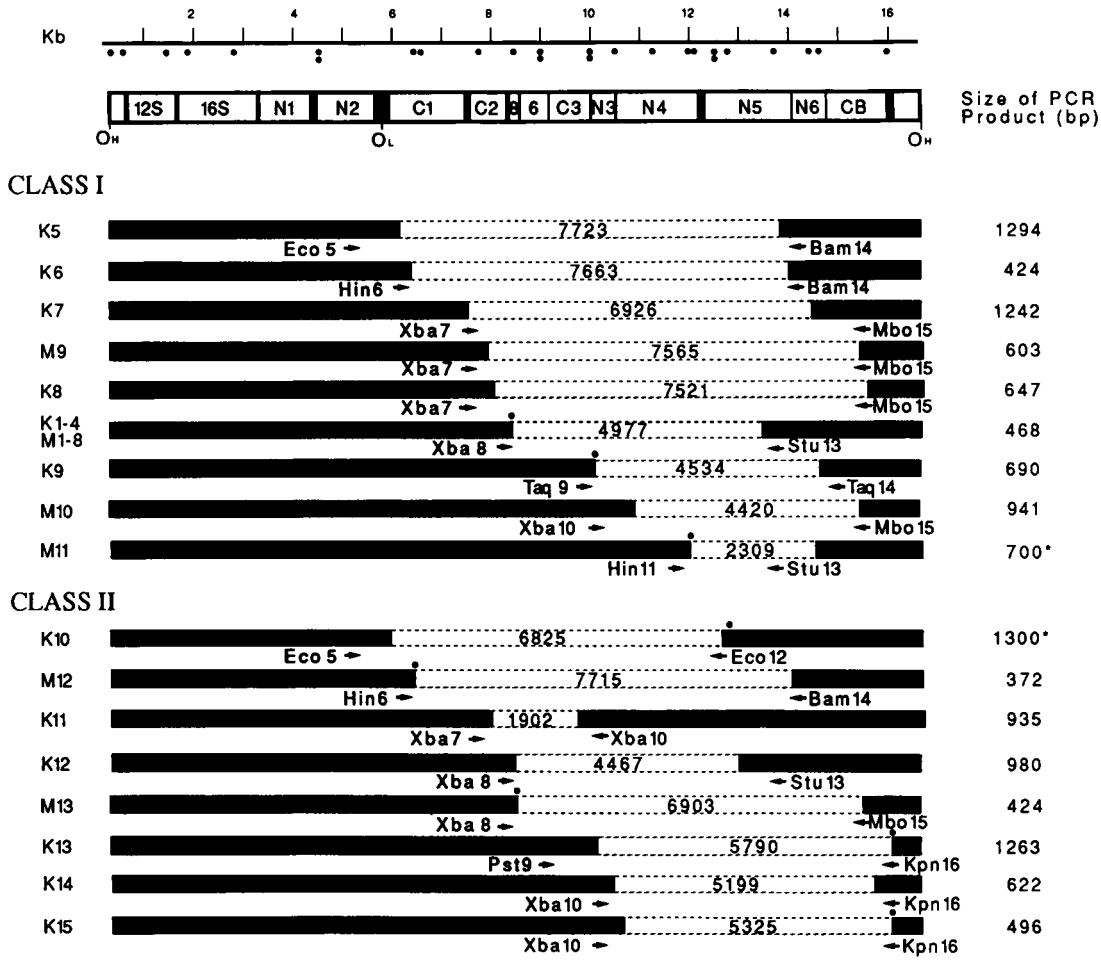


**Figure 1.** Localization of mtDNA deletions in PEO patients. A linearized mtDNA map is at top. $O_H$ and $O_L$ denote the origins of heavy- and light-strand replication. The 12S and 16S rRNA genes, and the subunits of cytochrome c oxidase (C), NADH dehydrogenase (N), cytochrome b (CB), and ATPase (6,8) are shown. Transfer RNA genes are in filled boxes. Dots below the map scale denote the locations of *Drosophila* topoisomerase-II-like sequences (map positions 205, 481, 1537, 1968, 2742, 4695, 4738, 6405, 6630, 7733, 8486, 9010, 9060, 9966, 10038, 10482, 11245, 11980, 12138, 12486, 12489, 12670, 13654, 14314, 14666, and 15897). Muscle mtDNA from each KSS (K) or OM (M) patient is represented by a bar map, and the deletion is identified as being in either Class I or Class II (see text). Intact regions are denoted by the solid areas; deleted regions by the open areas. The size of each deletion is shown in the deleted area, in bp. The 'common' deletion was present in 4 KSS patients (K1–K4) and 8 OM patients (M1–M8). Those topo II-like sites located within 50 bp of a deletion breakpoint are shown above individual maps as dots. The PCR primer pairs (→, forward; ←, backward) and the size of the PCR product spanning each deletion (in bp) is shown. The 'forward' PCR primers (5′→3′ on the L strand) were: Eco5F (mtDNA positions 5260–5288); Hin6F (6190–6220); Xba7F (7433–7468); Xba8F (8274–8305); Pst9F (9008–9036); Taq9F (9744–9765); Xba10F (10240–10269); and Hin11F (11668–11696). The 'backward' primers (5′→3′ on the H strand) were: Xba10B (10240–10269; Eco12B (12629–12653); Stu13B (13692–13720); Bam14B (14247–14276); Taq14B (14944–14967); Mbo15B (15574–15600); and Kpn16B (16033–16060). The asterisks (*) denote two PCR products that theoretically should not have been amplified, because the sequence homologous to the B primer, which yielded a PCR product, was found, after sequencing, to be located within the deleted region. We believe this occurred due to sequence similarity between the primer and a region located fortuitously just outside the deletion breakpoint. However, we have not yet sequenced the extreme 3′ end of either PCR product to confirm this supposition.

reaction, or PCR (6), using a pair of synthesized single-stranded oligonucleotide primers (Genetic Designs) corresponding to mtDNA sequences located immediately upstream and downstream of each breakpoint (Fig. 1). Several candidate primers were used to amplify a DNA fragment as small as possible, and the resulting smallest fragment was sequenced directly (3). In some cases, a third primer inside the amplified DNA fragment was used for sequencing. Twenty-five PCR cycles were performed with 1 μg of patient DNA, 100 pmole of each primer, 20 nmole of each deoxynucleotide triphosphate (dNTP), and 2.5 units of Taq polymerase (Perkin-Elmer Cetus), in 50 mM KCl, 10 mM Tris pH 8.3, 1.5 mM $MgCl_2$, 0.01% Gelatin, 0.01% Tween-20, and 0.01% NP40, in a total volume of 100 ul. Denaturation was at 94°C for 1 min, annealing was at 55°C for 2 min, and extension was at 72°C for 3 min. In the last cycle the extension proceeded for 10 min without any denaturation step. After electrophoresis of PCR products through 1.4% low melting temperature agarose (IBI), the DNA fragment was excised and extracted using GeneClean (BIO 101), and then sequenced directly using the Sequenase Kit (USB). About 200–400 ng of the amplified DNA was used for each sequencing reaction: after boiling for 10 min in 1× Sequenase buffer (40 mM Tris, pH 7.5, 20 mM $MgCl_2$, 50 mM NaCl) containing 10% DMSO (7) and 10 pmole of synthetic primer, annealing was allowed to proceed by quick-cooling of the template-primer mixture on ice. Dideoxy sequencing was performed as previously described (3).

## Computer analysis

Computer analysis of the sequences surrounding the deletions was performed using the MATCH algorithm of the GenBank data base, and the BESTFIT, GAP, FOLD, and SQUIGGLES

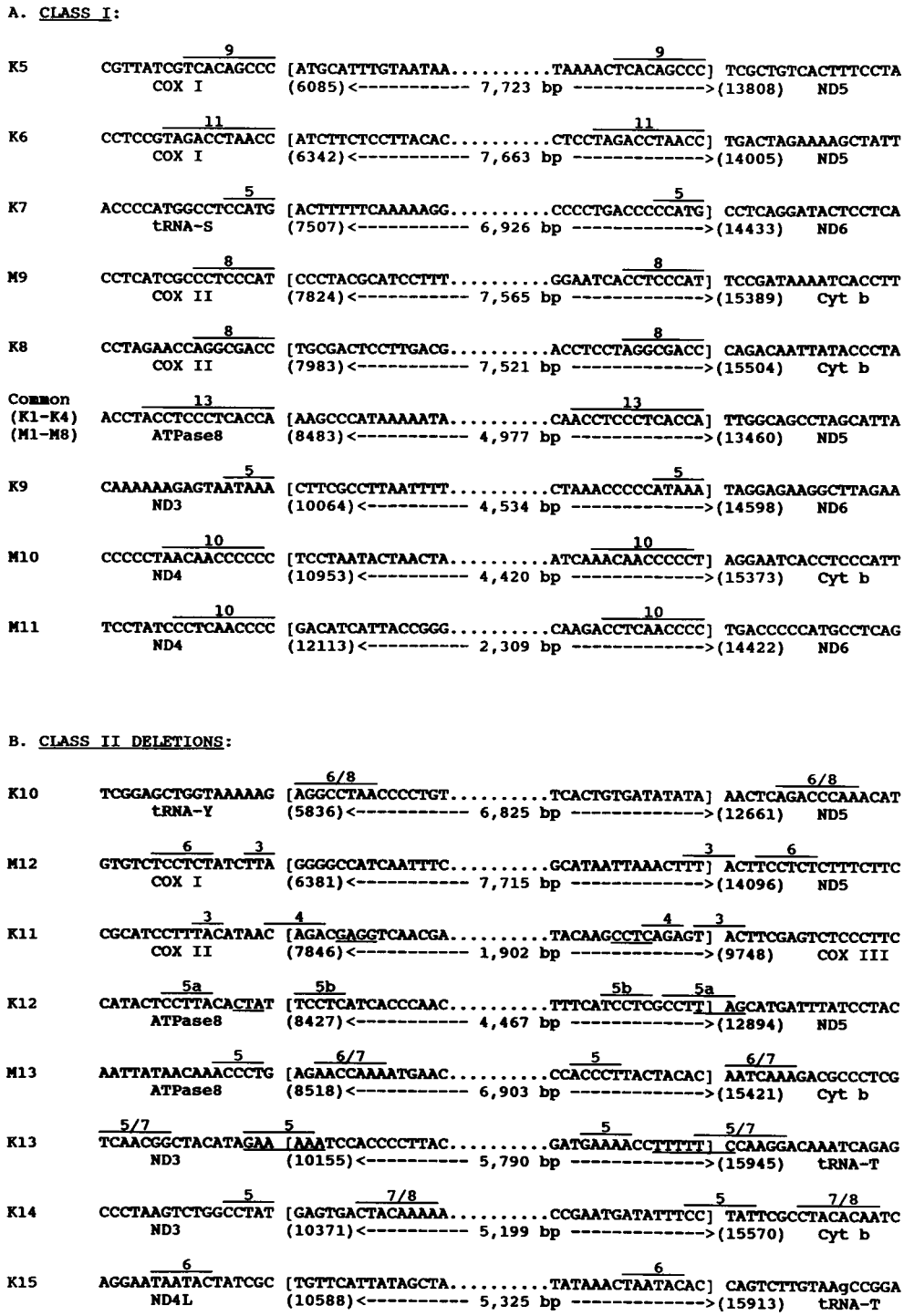


**Figure 2.** DNA sequences surrounding the deletion breakpoints. The sequences are of the L-strand, reading 5′ to 3′. Brackets denote the deletion breakpoints; mtDNA numbering (9) is of the first nt at the 5′ end and of the first nt immediately following the 3′ end of each deletion. The patient code is shown to the left of each deleted region. The interrupted genes are shown below the undeleted areas flanking the deletion. The length of the deletion is shown below the sequence. Repeated elements are overlined, showing the number of nucleotides in the repeat shown; inverted repeats are underlined. A. Class I deletions. B. Class II deletions. The lower-case G in K15 is a polymorphism relative to the published mtDNA sequence. 5a and 5b in K12 are two sets of 5-bp repeats.

programs provided by the Genetics Computer Group of the University of Wisconsin (8).

## RESULTS

### Sequences of the deletion breakpoints

Using PCR primers specific for the mtDNA deletion mapped in each patient, we amplified and sequenced the mtDNA in the regions flanking the breakpoints in 16 patients with KSS or OM (Fig. 1). The deletion breakpoints were located precisely by comparing the sequence of the region of the deleted mtDNA to that of the wild type mtDNA (9). The nucleotide sequences around the deletion breakpoints of the patients are shown in Fig. 2. The twelve patients with the 'common' deletion (patients K1-K4 with KSS and patients M1-M8 with OM; Fig. 1) had an

identical 4,977-bp deletion, which was flanked by a 13-bp direct repeat, as reported previously (3). The other 16 patients harbored uniquely different deletions. The sizes of the deletions ranged from 1,902 bp in patient K11 to 7,723 bp in patient K5. In all but one case, the exact size of the deletion as determined by sequencing agreed with the approximate size obtained from the mapping data. Patient M11 had a 2,309-bp deletion, instead of the 1.3 kb previously reported (2).

None of the 5′ and 3′ breakpoints were the same in any analyzed patient, except for those with the 'common' deletion. In four patients (K7, K10, K13, and K15), one of the breakpoints interrupted a tRNA gene. In the other 13 cases, both deletion breakpoints interrupted structural genes. In nine of these (the 'common' deletion and deletions K6, K11, K12, K14, M9, M10, M11, and M12), the resulting 'fusion gene' spanning the deletion breakpoint encoded a truncated 'fusion' polypeptide, due to alteration of the translational reading frame following the deletion breakpoint. In four other patients (K5, K8, K9, and M13), there was no alteration in the reading frame across the deletion breakpoint; thus a full-length fusion protein could theoretically be translated in these patients.

### Distribution of deletion breakpoints on the mtDNA map

In 16 of 17 deletions the 5′ breakpoints were concentrated in a 6.3-kb region between nt-5835 and nt-12112, and the 3′ breakpoints were located in an even smaller 3.3-kb region between nt-12661 and nt-15945. K11, which was the smallest deletion of all those examined (1,902 bp), extended from nt-7846 to nt-9748.

In some regions these breakpoints were even more tightly clustered. For example, four 5′ breakpoints were in a 388-bp region (between nucleotides 7506 and 7894), 4 were in a 545-bp region (5835−6380), and 4 were in an 800-bp region (10152−10952). Three 5′ breakpoints (including the 'common' deletion) were in a 93-bp region (8425−8518). Six 3′ breakpoints (including the 'common' deletion) were in a 973-bp region (13459−14432), and seven other 3′ breakpoints were in a 571-bp region (15373−15944). There was no obvious correlation between the location or clustering of the deletions and the class into which the deletions fell (see below).

### Classification of the deletions

The deletions fell into two classes. In Class I (9 different deletions, including the 'common' deletion; Fig. 2A), the deletion breakpoints were flanked precisely by direct repeats. The definition of members of Class I was specific: in all but one case, one repeat was found at the 5′ end immediately prior to the 'left' deletion breakpoint, while the other repeat was found (in normal, undeleted, mtDNA) at the extreme 3′ end of the deleted region, just prior to the 'right' deletion breakpoint. In one case, M10, the direct repeats were located 1 bp to the 'left' of the each deletion breakpoint.

The direct repeats ranged in size from 5 bp to 13 bp. Allowing for single mismatches between repeats, longer imperfect repeats could also be identified in three patients: 16/17 nucleotides in patient K6, 16/19 in patient K7, and 10/11 in patient K8.

In Class II (8 different deletions; Fig. 2B), the deletions were not flanked exactly by direct repeats. Thus, the definition of members of Class II was more ambiguous: the deletion breakpoints were either flanked imprecisely by perfect direct repeats, or there were no obvious repeat elements immediately adjacent to the deletion breakpoints. Small direct repeats (either perfect or imperfect) were often observed in the general area of the breakpoints, but in no case was a repeat pair found exactly at the 5′ and 3′ ends of the deletion, as was true for Class I. In three patients (K11, K12, and K13), small inverted repeats could be identified near the breakpoints, but this pattern was not found in the other patients in Class II.

### Computer analysis

We analyzed the sequences immediately surrounding the deletion breakpoints for the presence of direct and inverted repeats which might be associated with the formation of deletions. In each patient, a region of 100 bp flanking either side of the 5′ deletion breakpoint (i.e. 200 bp total) was compared to a similar 200-bp region surrounding the 3′ breakpoint, and examined for the presence of both direct and inverted (i.e. palindrome) repeat structures. We found homologous sequences in the 200-bp regions flanking the 5′ and 3′ breakpoints in all the patients (i.e. >50% identity), whether they had breakpoints in Class I or Class II. Similar results were also obtained when the regions were compared for the presence of inverted repeats. However, in a control computer search performed on randomly-picked 200-bp regions from areas of the mitochondrial genome that either contained deletion breakpoints or were conspicuously devoid of deletion breakpoints, we obtained similar degrees of sequence identity. Thus, we do not believe that any particular degree of sequence identity in these flanking regions, whether direct or inverted, can be correlated with confidence to the presence or location of the deletions, whether in Class I or II.

An analysis of DNA secondary structures of minimum free energy in regions in which many of the 5′ and 3′ breakpoints were clustered (see above) showed no definite pattern, but there was one interesting exception: the secondary structures around both the 5′ and 3′ deletion breakpoints in the 'common' deletion were remarkably similar, with the 13-bp repeats each located in a large single-stranded loop region (not shown).

## DISCUSSION

We have extended our analysis of the breakpoints of the 'common' deletion (3) to 16 other deletions in patients with PEO. We undertook this study in the belief that detailed analysis of the spectrum of mtDNA deletions in these patients might enable us to define a mechanism for the formation of large-scale deletions of human mtDNA, and to shed light on the pathogenesis of these diseases.

Analysis of 17 different deletions in 28 patients suggests that at least two classes of deletions can be distinguished: those that are precisely flanked by direct repeats (Class I) and those that are not (Class II). The mechanism(s) by which any of these deletions arose is currently unknown.

The Class I deletions were the most frequent (20 of 28 patients, or 71%, including the 12 patients with the 'common' deletion). While the length of the direct repeats ranged from 5 bp to 13 bp, the only deletion found to occur more than once among all 28 patients studied—the 'common' deletion—was flanked by the longest direct repeat, 13 bp. This is the largest perfect repeat sequence in the human mitochondrial genome outside the zone between the two origins of replication. The 'common' deletion thus appears to be flanked by the longest direct repeat capable of producing a viable deletion (3).

These results imply that large-scale deletions of human mtDNA can occur via a recombination mechanism mediated by direct

repeats. Also, the high prevalence of Class I deletions in OM and KSS implies that this type of recombination is a major source of the mtDNA deletions found in these disorders.

Since the 'common' deletion found by us is identical to the deletion reported by other groups (5,10−13), the 13-bp repeat must be a hotspot for deletion (3), but the reason for this is still speculative. Albertini *et al.* (14) pointed out the importance of the length of sequence homology in the generation of large deletions in *E. coli*: altering a single base pair within this homology reduced the frequency of deletion by an order of magnitude. Our findings on the 'common' deletion are in agreement with this observation, indicating that the length of the direct repeat may play a crucial role in the formation of deletions. In the next-longest direct repeat, in patient K6 (Class I), the 11-bp repeat is actually part of a larger region of sequence identity of 16/17 nucleotides. In fact, the K6 deletion was recently reported to exist in at least one other patient (11). The presence of 32 reported examples of the 'common' deletion and only 2 examples of the K6 deletion suggests that a complete direct repeat may be more important than a longer incomplete direct repeat. Furthermore, computer calculations of minimum free-energy secondary structures of single-stranded DNA showed that both sites of the 13-bp direct repeat flanking the 'common' deletion may actually reside in large single-stranded loops, and that these two loops might be physically close to each other in three dimensions (data not shown). Therefore, both the length of the direct repeat and the secondary structure of the DNA region may be two of the factors that render the 13-bp repeat a hotspot for mtDNA deletion (see also Ref. 3).

In non-mitochondrial genomes, deletion termini characterized by short direct repeats are relatively common in phage (15,16), in prokaryotes (14,17,18), and to a lesser extent, in higher eukaryotes (19−21). In mtDNA, large-scale deletions have been observed in lower eukaryotes and plants, but other than one report of mtDNA deletions in mice (22), they had not been found in mammals prior to 1988. Mitochondrial recombination involving repeated elements has been observed in yeast (23−25), in stopper mutants of *Neurospora crassa* (26,27), in the senescence phenotype of *Podospora anserina* (28) and in plants (29−31). One likely mechanism for the occurence of Class I mtDNA deletions is that they arise by means of an intragenomic recombination event via slipped mispairing during DNA synthesis (3,10,12,14,17,18,20, 21,25).

We have placed deletion M10 in Class I, even though both the 5′ and 3′ repeats are displaced by one nucleotide to the 5′ side (on the L-strand) of the deletion breakpoint. We felt this was appropriate, because it appears as if the 10-bp repeats in M10 were perfectly located with respect to the deletion breakpoint at the time of alignment between the pairing strands (i.e. a Class I alignment), but that prior to resolution of the recombination intermediates, a one-bp slip occurred, reminiscent of branch migration.

The sequence around the M10 breakpoint is also informative as to which of the two repeats was actually deleted during the recombination event. The sequence across the deletion breakpoint in M10 (5′-AACAACCCCCC|AGGAA-3′, where the vertical line denotes the deleted region) contains a C (underlined) that is present immediately after the 5′ 10-bp repeat; the 3′ 10-bp repeat is followed by a T, not a C (see Fig. 2A). Thus, if M10 is representative of Class I, we can infer that it is the 5′ repeat (on the L-strand) that was retained, and the 3′ repeat that was deleted. Deletion of the 3′ repeat was also inferred in a different

```
                                    16/19 (84%)
Alignment 1:  5' side:  TTCAAG---CCAACCCCATGGCCTCCATGacttttttcaaaaaggtatt
                          ****    ******* ** ** ***** **       *      * *
              3' side:  accaagacctcaaccccc-tgacc-ccatgCCTCAGGATACTCCTCAAT

                                    15/18 (83%)
Alignment 2:  5' side:  GGTTTCAAGCCAACCCCATGGCCTCCATGacttttt-tcaaaaaggtat
                          *      * ********* **** * **    *  **** *
              3' side:  cctcaaccccctgaccccatg-CCTC-AGGATACTCCTCAATAGCCATC
```

**Figure 3.** Two possible alignments around the deletion breakpoint in patient K7. Upper case letters denote undeleted mtDNA; lower case letters denote deleted mtDNA. 5′ and 3′ denote sequences at the 5′ and 3′ ends of the deletion, respectively (see Fig. 2A). Sequences have been aligned to maximize homology. Asterisks denote sequence identities. The highly-similar regions are overlined.

deletion (12). If this conclusion can be generalized to all Class I deletions, it may limit specific models of recombination to explain these deletions.

The Class II deletions were found in the minority of patients (8 of 28 patients, or 29%). One deletion suggests that the mechanism of deletion in the two classes may be related. K7 belongs to Class I, and is flanked precisely by a 5-bp perfect repeat, CCATG. However, closer inspection of the sequence surrounding the breakpoints indicates that there is a much larger area of sequence homology between the 5′ and 3′ breakpoints. In fact, two different alignments can be made, in part because CCATG is found at two positions around the 5′ breakpoint, but at only one position around the 3′ breakpoint (Fig. 3). In alignment 1, which probably did occur in the deletion, there is a region of 84% identity (16/19 nt, including single-base insertions/deletions in the alignment), with the 5′ and 3′ CCATG repeats aligned so that a crossover at the 5-bp repeat results in removal of the intervening 6,926 bp, as observed in the patient. This crossover occurred even though an even longer direct repeat of 7 bp (CAACCCC) can be found in alignment 1, but was not chosen as the crossover point. A similar degree of sequence identity, 83% (15/18 nt), can be observed in a second alignment. In alignment 2, the 'other' 5′ CCATG repeat is aligned with the single 3′ CCATG repeat, such that an even longer perfect repeat of 8 bp (ACCCCATG) can be generated. Based on degree of sequence identity and the observations of Albertini *et al.* (14), one would have expected alignment 2 (83%, with an 8-bp repeat) to have been slightly favored over alignment 1 (84%, with a 7-bp repeat). Thus, it is conceivable that this was in fact the initial alignment of the DNA strands, but, due to 'slippage', a new alignment was created prior to the crossover event.

This type of 'slippage' may be occurring in the Class II deletions. There are numerous repeats in the human mitochondrial genome, due to the asymmetry in composition between the heavy and light strands; in a search for long repeats in human mtDNA, we found almost 60 unique pairs of 11-bp repeats and more than 250 pairs of 10-bp repeats (3). Thus, the likelihood of finding a direct repeat of 5 or 6 bp between any two randomly-compared mtDNA regions is quite high. For this reason, it is not clear whether the short repeats highlighted in Fig. 2B have any functional significance in the generation of Class II deletions. Nevertheless, if these particular repeats are associated with the deletions in Class II, they cannot 'slip' to the eventual crossover point via classical branch migration, because in all the deletions the distance traversed from the 5′ repeat to the 5′ breakpoint is not equal to the distance traversed from the 3′ repeat to the 3′ breakpoint (although in two cases, K11 and K14, the movement of the two repeats differs by only 1 bp). On the other hand, other regions of high homology flanking the Class II deletion breakpoints, similar to the two alternative K7 alignments noted

above, might allow for 'slippage' and result in the breakpoints observed in the Class II deletions.

Besides direct repeats, palindromic sequences are frequently associated with deletion termini, and deletions appear to eliminate these potential hairpin structures (14,17). Vanin *et al.* (32) suggested that four large deletions in the human beta-globin gene cluster were generated by a common mechanism involving palindromic elements, in which the 5′ and 3′ breakpoints were brought physically close to each other, even though they were far apart in the primary DNA sequence. Therefore, palindromic structures as well as incomplete direct repeats in mtDNA may also play a role in deletion formation, by shortening the distance between greatly separated sequences.

Transfer RNA sequences may also participate in the deletion process, because they can form secondary structures which may bring the breakpoint regions into close proximity (28). Thus, it may be more than coincidental that four of our deletions had breakpoints which fell within tRNA genes, and eight deletions had breakpoints that fell within 200 bp of tRNA genes.

Besides global structure, specific sequences may also render certain regions of the mtDNA prone to deletion. The distribution of the deletion breakpoints, in fact, suggests that there may be preferred regions for deletions in the mtDNA. Inspection of the sequences shown in Fig. 2 shows no obvious similarity, except that many breakpoint regions are pyrimidine-rich on the L-strand (especially M11 and M12); Conversely, K9 is in a purine-rich region. The one specific motif that stands out is the pentanucleotide CCTCC, which is found at or near 7 of the 9 Class I deletion breakpoints ('common', K6, K7, K8, M9, M10, and M11), but in none of the Class II deletions. We had already noted that polypyrimidine- and AT-rich regions may favor the formation of bent DNA, and thus provide a region of single-stranded DNA as a target for slipped mispairing (3). Such regions may also be sequence-specific targets for components of the putative recombination machinery associated with deletions. For example, the GC clusters associated with yeast mtDNA deletions (24) are flanked, on their complementary strands, by a pyrimidine-rich element, 5′-TCTCCTTT-3′. The mouse mitochondrial RNA-processing endoribonuclease (RNAse MRP) RNA, which cleaves mouse mtRNA to produce a primer at the origin of heavy-strand replication, has the sequence 5′-CCCCUCC-3′ at the point of RNA cleavage (33).

Moreover, there may be sequences in the mtDNA associated with recombination, similar to the bacterial Chi site (5′-GCTGGTGG-3′) (34), the *var1* GC-rich element in yeast (35), TG-rich sequences associated with mtDNA rearrangements in *Podospora anserina*, *Plasmodium falciparum*, and *Saccharomyces cerevisiae* (summarized in 36), or the *Drosophila* (37) and vertebrate (38) topoisomerase II sites.

A computer search failed to show any obvious correlation between these sequences and sequences found at or near the mtDNA deletion breakpoints, with two exceptions. The 5′ breakpoint of the common deletion contains the sequence 5′-CACCA|A-3′, which is complementary to the TG-rich element in the *Podospora* and *Plasmodium* rearrangements (36). In addition, the computer search for sequences similar to the *Drosophila* topoisomerase II recognition sites suggested that the deletion breakpoints might be non-randomly associated with topo-II-like sites (Fig. 1). There are no perfect matches in human mtDNA to the *Drosophila* topo-II recognition site (5′-GTNWAYATTNATNNR-3′); however, there is one topo-II-like site with 1 mismatch, and 25 with 2 mismatches, of which 19 are in the region viable for mtDNA deletions (dots under map in Fig. 1). Of the 17 5′ deletion breakpoints, 14 were within 250 bp of a topo-II-like sequence, and 7 ('common', K9, K10, K13, K15, M11, M12, and M13) were within 50 bp. In fact, the 5′ breakpoint closest to a topo-II-like site was the 'common' deletion; the 13-bp repeat was only 3 bp away from a topo-II-like site. Three 3′ breakpoints ('common', K10, and K15) were within 25 bp of such a site, and K15 was only 1 bp away.

Toposiomerase II is a gyrase that introduces negative superhelical turns into DNA circles by a double-stranded breakage and rejoining mechanism. As such, it can catenate, decatenate, and supercoil DNA circles. We had previously noted that extrusion of bent DNA regions as putative targets for slipped mispairing requires torsion generated by superhelicity (3). Thus, the correlation of topo-II-like sites with deletion breakpoints renders this hypothesis more attractive, but only under the further assumption that the proximity of a topo II-like target to the breakpoint produces localized regions of supercoiling. A topo II-like enzyme probably exists in mitochondria, for two reasons: mtDNA circles are supercoiled (39), and catenated mtDNA circles have been observed during replication (39) and in gametogenesis (40).

Thus, another model for deletion formation may be based on recombination events mediated by enzymes that recognize some particular sequences or structure. King *et al.* (41) reported that in-vivo recombinants between bacteriophage lambda DNA and pBR322 in *E. coli* may have been generated by the RecA recombination system, in which only a short region of homology is required. Naito *et al.* (42) studied the nucleotide sequences of several recombination junctions of lambda-pBR322 recombinants, and found that as little as 3 bp of identity between the parent DNAs could promote DNA gyrase-mediated recombination between heterologous DNA in a cell-free system.

With respect to the molecular mechanism producing large-scale deletion of human mtDNA, Zeviani et al. (43) discussed the possibility of a nuclear-encoded protein associated with the formation of mtDNA deletions in a family with autosomal dominant mitochondrial myopathy. This putative protein appears to mediate an intragenomic homologous recombination event, since all 19 deletions analyzed started within a 12-nt stretch at the 5′ end of the D-loop region and were flanked by short direct repeats (3 bp to 12 bp), most of which were imperfect (i.e. Class II, in our nomenclature).

Finally, we note that in the 28 patients harboring 17 different deletions, not one had a deletion in mtDNA which did not eliminate at least one tRNA gene. We speculate that deletion of tRNA genes in mitochondria containing only deleted mtDNAs could render the organelles incompetent for translation. Conversely, if a deletion leaves all the tRNA genes intact, the genetic defect may not manifest itself phenotypically as PEO. This supposition is consistent with other findings by us. We have shown that (i) mitochondria containing deleted genomes seem to be competent for transcription but incompetent for translation (44), and (ii) COX deficiency in muscle sections from a patient with the 'common' deletion was not only segmental (i.e. normal in some regions of a fiber but deficient in others), but RNAs that were not encompassed by the deletion were transcribed (as measured by in-situ hybridization) but were not translated (as measured by immunofluorescence) (45). If this argument is correct, one would predict that large-scale deletions of mtDNA which remove structural genes but not any tRNA genes should result in a pathological phenotype different from PEO,

characterized mainly by deficiency of just those specific respiratory chain enzyme activities corresponding to those subunit genes deleted in the mtDNA. Mitochondria in such a pathology would be competent for translation of the undeleted subunit gene transcripts, and might not even proliferate and cause RRF.

We have shown that recombination via flanking direct repeats, perhaps via slipped mispairing, is a major cause of large-scale deletions of mtDNA in patients with PEO. The presence of at least two classes of deletions implies that more than one mechanism may exist for the generation of large deletions of mtDNA. Nevertheless, the demonstration of recombination of mtDNA in human pathological states should enable us to search for heretofore unidentified components of the recombination machinery in human mitochondria.

## ACKNOWLEDGEMENTS

We thank Mr. Dan Wasserburg for assistance with the computer analyses, Mr. James Sadlock for expert technical assistance, and Dr. Rodney Rothstein for helpful discussions. This work was supported by grants from the National Institutes of Health (NS11766), the Muscular Dystrophy Association (MDA), the Aaron Diamond Foundation, and by a gift from Dr. and Mrs. Libero Danesi. R.R. and Y.K. are MDA Postdoctoral Fellows. C.T.M. was supported by the Brazilian Research Council (CNPq).

## REFERENCES

1. Zeviani, M., Moraes, C.T., DiMauro, S., Nakase, H., Bonilla, E., Schon., E.A. and Rowland L.P. (1988) Neurology **38**, 1339–1346.
2. Moraes C.T., DiMauro, S., Zeviani, M., Lombes, A., Shanske, S., Miranda, A.F., Nakase, H., Bonilla, E., Werneck, L.C., Servidei, S., Nonaka, I., Koga, Y., Spiro, A.J., Brownell, K.W., Schmidt, B., Schotland, D.L., Zupanc, M., Devivo, D.C., Schon, E.A. and Rowland, L.P. (1988) N. Engl. J. Med. **320**, 1293–1299.
3. Schon, E.A., Rizzuto, R., Moraes, C.T., Nakase, H., Zeviani, M. and DiMauro, S. (1989) Science **244**, 346–349.
4. Holt, I.J., Harding, A.E. and Morgan-Hughes, J.A. (1988) Nature **331**, 717–719.
5. Rotig, A., Colonna, M., Bonnefont, J.P., Blanche, S., Fisher, A., Saudubray J.M. and Munnich, A. (1989) Lancet **i**, 902–903.
6. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T. and Erlich H.A. (1985) Science **230**, 1350–1354.
7. Winship, P.R. (1989) Nucl. Acids Res. **17**, 1266.
8. Devereaux, J., Haeberli, P. and Smithies, O. (1984) Nucl. Acids Res. **12**, 387–395.
9. Anderson, S., Bankier, A.T., Barrell, B.G., DeBruyn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nielich, D.P., Rae, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature **290**, 457–465.
10. Holt, I.J., Harding, A.E. and Morgan-Hughes, J.A. (1989) Nucl. Acids Res. **17**, 4465–4469.
11. Johns, D.R., Rutledge, S.L., Stine, O.C., and Hurko, O. (1989). Proc. Natl. Acad. Sci. USA **86**, 8059–8062.
12. Shoffner, J.M., Lott, M.T., Voljavec, A.S., Soueidan, S.A., Costigan, D.A., and Wallace, D.C. (1989). Proc. Natl. Acad. Sci. USA **86**, 7952–7956.
13. Tanaka, M., Sato, W., Ohno, K., Yamamoto, T., and Ozawa, T. (1989). Biochem. Biophys. Res. Comm. **164**, 156–163.
14. Albertini, A.M., Hofer, M., Calos, M.P. and Miller, J.H. (1982) Cell **29**, 319–328.
15. Studier, F.W., Rosenberg, A.H., Simmon, M.N. and Dunn, J.J. (1979) J. Mol. Biol. **135**, 917–937.
16. Pribnow, D., Sigurdson, D.C., Gold, L., Singer, B.S., Napoli, C., Brosius, J., Dull, T.J. and Noller, H.F. (1981) J. Mol. Biol. **149**, 337–376.
17. Schaaper, R.M., Danforth, B.N. and Glickman B.W. (1986) J. Mol. Biol. **189**, 273–284.
18. Farabaugh, P.J, Schmeissner, U., Hofer, M. and Miller, J.H. (1978) J. Mol. Biol. **126**, 847–857.
19. Nalbantoglu, J., Hartley, D., Phear, G., Tear, G. and Meuth, M. (1986) EMBO J. **5**, 1199–1204.
20. Canning, S. and Dryja, T.P. (1989) Proc. Natl. Acad. Sci. USA **86**, 5044–5048.
21. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connel, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weisman, S.M., Slinghtom, J.L., Blechl, A.E., Smithies, O., Barralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell **21**, 653–668.
22. Boursot, P., Yonekawa, H. and Bonhomme, F. (1987) Mol. Biol. Evol. **4**, 46–55.
23. Foury, F. and Lahaye, E. (1987) EMBO J. **6**, 1441–1449.
24. Dieckmann, C.L. and Gandy, B. (1987) EMBO J. **6**, 4197–4203.
25. Ahne, A., Muller-Derlich, J., Merlos-Lange, A.M., Kanbay, F., Wolf, K. and Lang, B.F. (1988) J. Mol. Biol. **202**, 725–734.
26. Gross, S.R., Hsieh, T.-S. and Levine P.H. (1984) Cell **38**, 233–239.
27. Almasan, A. and Mishra, N.C. (1988) Genetics **120**, 935–945.
28. Turker, M.S., Domenico J.M. and Cummings, D.J. (1987) J. Mol. Biol. **198**, 171–185.
29. Moon, E., Kao, T.-H. and Wu, R. (1988) Mol. Gen. Genet. **213**, 247–253.
30. Ogihara, Y., Terachi, T. and Sasakuma, T. (1988) Proc. Natl. Acad. Sci. USA **85**, 8573–8577.
31. vom Stein, J. and Hachtel, W. (1988) Mol. Gen. Genet. **213**, 513–518.
32. Vanin E.F., Henthorn, P.S., Kioussis, D., Grosveld, F. and Smithies, O. (1983) Cell **35**, 701–709.
33. Chang, D.D. and Clayton, D.A. (1989) Cell **56**, 131–139.
34. Kobayashi, I., Stahl, M. and Stahl, F. (1984) Cold Spr. Harb. Symp. Quant. Biol. **49**, 497–506.
35. Zinn, A.R., Pohlman, J.K., Perlman, P.S. and Butow, R.A. (1988) Proc. Natl. Acad. Sci. USA **85**, 2686–2690.
36. Koll, F. and Belcour, L. (1989) Trends Genet. 5, 173.
37. Sander, M. and Hsieh, T.-S. (1985) Nucl. Acids Res. **13**, 1057–1072.
38. Spitzner, J.R. and Muller, M.T. (1988) Nucl. Acids Res. **16**, 5533–5556.
39. Clayton, D.A. (1982) Cell **28**, 693–705.
40. Piko, L. and Matsumoto, L. (1976). Dev. Biol. **49**, 1–10.
41. King, S.R., Krolewski M.A., Marvo, S.L., Lipson P.J., Pogue-Geile, K.L., Chung, J.H. and Jaskunas, S.R. (1982) Mol. Gen. Genet. **186**, 548–557.
42. Naito, A., Naito, S. and Ikeda, H. (1984) Mol. Gen. Genet. **193**, 238–243.
43. Zeviani, M., Servidei, S., Gellera, C., Bertini, E., DiMauro, S. and DiDonato, S. (1989) Nature **339**, 309–311.
44. Nakase H, Moraes CT, Lombes A, Miranda AF, DiMauro S, and Schon EA (1990). Am. J. Hum. Genet., in press.
45. Mita, S., Schmidt, B., Schon, E.A., DiMauro, S. and Bonilla, E. (1989) Proc. Natl. Acad. Sci. USA 86, 9509–9513.