© 1996 Oxford University Press

Nucleic Acids Research, 1996, Vol. 24, No. 17 **3337–3340**

The germline gene encoding DNA polymerase α in the hypotrichous ciliate *Oxytricha nova* is extremely scrambled

David C. Hoffman and David M. Prescott*

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309-0347, USA

Received May 30, 1996; Revised and Accepted July 16, 1996

GenBank accession no. U58368

ABSTRACT

We report the structure of the micronuclear (germline) gene encoding the large catalytic subunit of DNA polymerase α (DNA pol α) in the ciliate *Oxytricha nova*. It contains 44 internal eliminated segments (IESs) that divide the gene into 45 macronuclear-destined segments (MDSs) that are in a non-randomly scrambled order with an inversion near the gene center. Odd numbered MDSs 29–43, containing 230 bp out of a total of 4938 bp of macronuclear sequence, are missing from the 14 kb cloned gene. The missing MDSs have not been located but are at least several kilobases from the main body of the gene. The remarkably scrambled DNA pol α gene must be extensively cut, re-ordered and spliced and an inversion must occur to produce an unscrambled, functional version of the gene during development of a new macronucleus. Unscrambling is hypothesized to occur by a homologous recombination mechanism guided by repeat sequences at MDS ends.

INTRODUCTION

Genes in the germline micronuclear genome in hypotrichous ciliates are interrupted by multiple non-coding DNA segments called internal eliminated segments (IESs) (1). The micronuclear gene segments separated by IESs are called macronuclear-destined segments (MDSs), which are spliced when IESs are excised during development of a somatic macronucleus from a micronucleus after cell mating. In some micronuclear genes MDSs are in a scrambled order and are spliced in the correct order during macronuclear development (2,3). The analysis of a long, highly scrambled gene, encoding the large, catalytic subunit of DNA polymerase α (DNA pol α) is reported here.

MATERIALS AND METHODS

Oxytricha nova was grown and harvested and DNA prepared as described elsewhere (3). A single purified clone was selected from a λ library (1) and amplified and purified according to established protocols (4). The insert was digested into three *Sac*I fragments, which were subcloned into the plasmid vector pGEM-7zf+ (Promega, Madison, WI). Sequencing was by the dideoxynucleotide chain termination method using Sequenase (Amersham Life Sciences) and [α-^{35}S]dATP (Dupont/New England Nuclear) incorporation.

RESULTS

A 14 kb micronuclear DNA fragment encoding DNA pol α was selected from a λ genomic library of *O.nova* micronuclear DNA using fragments of the macronuclear DNA pol α gene (5) as selection probes. The probes were a 483 bp fragment from the 5′-end of the macronuclear gene, a fragment of 852 bp from the open reading frame (ORF) and a *Sac*I fragment of 628 bp from the 3′-end of the macronuclear molecule. A purified λ clone that hybridized with all three fragments was digested with *Sac*I, yielding fragments of ~6.0, ~6.7 and ~1.3 kb. All three were subcloned into the plasmid pGEM-7zf+. The ~6.0 and ~6.7 kb inserts contained 5999 bp of the DNA pol α gene, including IESs, the 5′ leader and 3′ trailer and putative flanking sequences of 776 bp at one end and 5963 bp at the other. The ~1.3 kb fragment contained additional flanking sequence of 1265 bp extending from the 776 bp of flanking sequence.

The macronuclear DNA pol α gene sized molecule is 4938 bp long (5). Comparison of the macronuclear gene sized molecule with the micronuclear sequence revealed that the micronuclear DNA pol α gene is divided into at least 45 MDSs separated by at least 44 IESs. Numbers were assigned to MDSs according to the orthodox order in which they appear in the macronuclear gene. Thirty seven MDSs are present in the λ clone, but eight MDSs containing 230 bp are absent. The 37 MDSs are in the order shown in Figure 1. The 13 MDSs at one end of the gene (MDS 27 and even numbered MDSs 26–4) are inverted with respect to the remaining 24. The MDSs missing from the λ clone are 29, 31, 33, 35, 37, 39, 41 and 43. Repeats of 6–16 bp are present at the ends of those MDSs that must be re-ordered (Table 1) and repeats of 2–4 bp are present where MDSs are already in the correct order with respect to one another and need only to be spliced, i.e. MDSs 1, 2 and 3; 26 and 27; 44 and 45 (Fig. 1). None of the 230 bp sequence in the eight missing MDSs is present in the DNA that flanks the ends of the cloned gene (flanking DNA of 2041 bp at one end and 5963 bp at the other end). Library screens and PCR with micronuclear DNA have so far failed to detect the missing

* To whom correspondence should be addressed

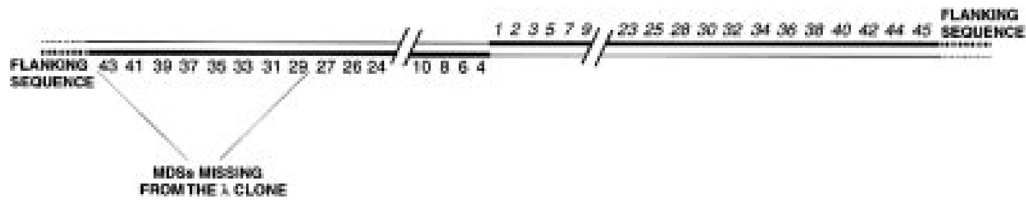

Figure 1. MDSs in the scrambled micronuclear DNA pol α gene. The odd numbered MDSs 29–43 are absent from the clone and their arrangement is hypothetical. The thin strand is transcribed. GenBank accession no. U58368.

230 bp elsewhere in the genome, suggesting that the missing sequence may be broken up into many more than eight MDSs.

Table 1. Pairs of repeats at the ends of MDSs

Left MDS/IES junction	MDS no.	Right MDS/IES junction
5′ Telomere addition site	1	TA
TA	2	GAGT
GAGT	3	GAATGGCAATA
GAATGGCAATA	4	AAGAATGC
AAGAATGC	5	TAAGAAT
TAAGAAT	6	ACTCTTTCTTT
ACTCTTTCTTT	7	TGAAGAAAATACCG
TGAAGAAAATAACG	8	TGCTCT
TGCTCT	9	AGAGCTATT
AGAGCTATT	10	AAGAATAGGCAA
AAGAATATGCAA	11	TTTGAGCTG
TTTGAGTTG	12	TTAAGATCA
TTAAGATCA	13	GAGCTAA
GAGCTAA	14	TGGTTCAA
TGGTGCAA	15	ATGATTTGAA
ATGATTTGAA	16	AAAACTT
AAAACTT	17	GAGATAGCAAT
GAAATAGCAAT	18	ATGGAAAG
ATGGAAAG	19	AGAGACA
AGAGACA	20	TGGCTCATAAT
TGGCTCATAAT	21	AACAAGAA
AACAAGAA	22	AATTATG
ATTATG	23	ATTCCAA
ATTCCAA	24	TAGCTAAG
TAGCTAAG	25	TTGACTTTG
TTGACTTTG	26	AGAC
AGAC	27	CTAACAGTATGTATGG
CTAACAGTATGTATGG	28	Not determined
Not determined	44	TCTT
TCTT	45	3′ Telomere addition site

MDSs 29, 31, 33, 35, 37, 39, 41 and 43 have not yet been located in the micronuclear genome and the relevant repeat sequences have not been determined, although MDSs 28, 30, 32, 34, 36, 38, 40 and 42 are present in the λ micronuclear clone and have been sequenced. Mismatches are underlined.

The micronuclear MDS sequence differs in ~2% of nucleotide positions from the previously sequenced macronuclear gene (5). This suggests the presence of at least two versions of the DNA pol α gene. Two versions of the macronuclear gene were demonstrated by sequencing of a 608 bp fragment generated from macronuclear DNA (data not shown). One version corresponds to the previously reported macronuclear gene sequence (5) and the other corresponds to the micronuclear gene sequence analyzed here.

DISCUSSION

Of the seven micronuclear genes of *O.nova* that have been sequenced, R1 (6), C2 (1), the gene encoding β telomere binding protein (7) and the gene encoding heat shock protein 70 (8) contain IESs but are non-scrambled. Three genes, encoding actin I (3), α telomere binding protein (αTP) (2) and DNA pol α are scrambled. For example, the micronuclear αTP gene consists of 14 MDSs in the order 1–3–5–7–9–11–2–4–6–8–10–12–13–14. The DNA pol α gene is much longer and its non-random scrambling pattern is more complex than the patterns in the other two scrambled genes. Like the αTP gene (14 MDSs), most of its MDSs fall into odd and even numbered series. However, ~1.9 kb of the DNA pol α gene are inverted relative to the rest of the gene. In this characteristic it is different from the αTP gene, which has no inversion, but is similar to the actin I gene, which has a single, inverted MDS, however, the scrambling pattern of the DNA pol α gene is unlike the actin I gene, whose nine MDSs are randomly scrambled in the order 3–4–6–5–7–9–2–1–8 (3). Thus, the three scrambled micronuclear genes characterized so far show three different combinations of inversions and random/non-random scrambled patterns of MDSs.

The DNA pol α gene contains 12 domains of amino acid sequence conserved among a wide variety of other organisms from bacteriophage to humans and yeast (9,10), but these domains do not correspond in any discernible way to individual MDSs or groups of MDSs in the *O.nova* gene. In some cases, conserved domains are actually split between two or three MDSs.

The 14 kb micronuclear DNA λ clone contains 4708 of the 4938 bp in the macronuclear DNA pol α gene. Missing from the micronuclear clone are 230 bp of the ORF. The simplest assumption is that these missing 230 bp are present in eight missing MDSs, numbered 29, 31, 33, 35, 37, 39, 41 and 43, as depicted in Figure 2. This MDS pattern would complement the MDS pattern 28–30–32–34–36–38–40–42–44–45 present in the micronuclear clone. The 2041 bp and 5963 bp of the insert that flank the cloned gene contain none of the missing MDS sequences. A logical possibility is that the missing MDSs are located >2041 bp into the DNA flanking MDS 27 and the 2041 bp

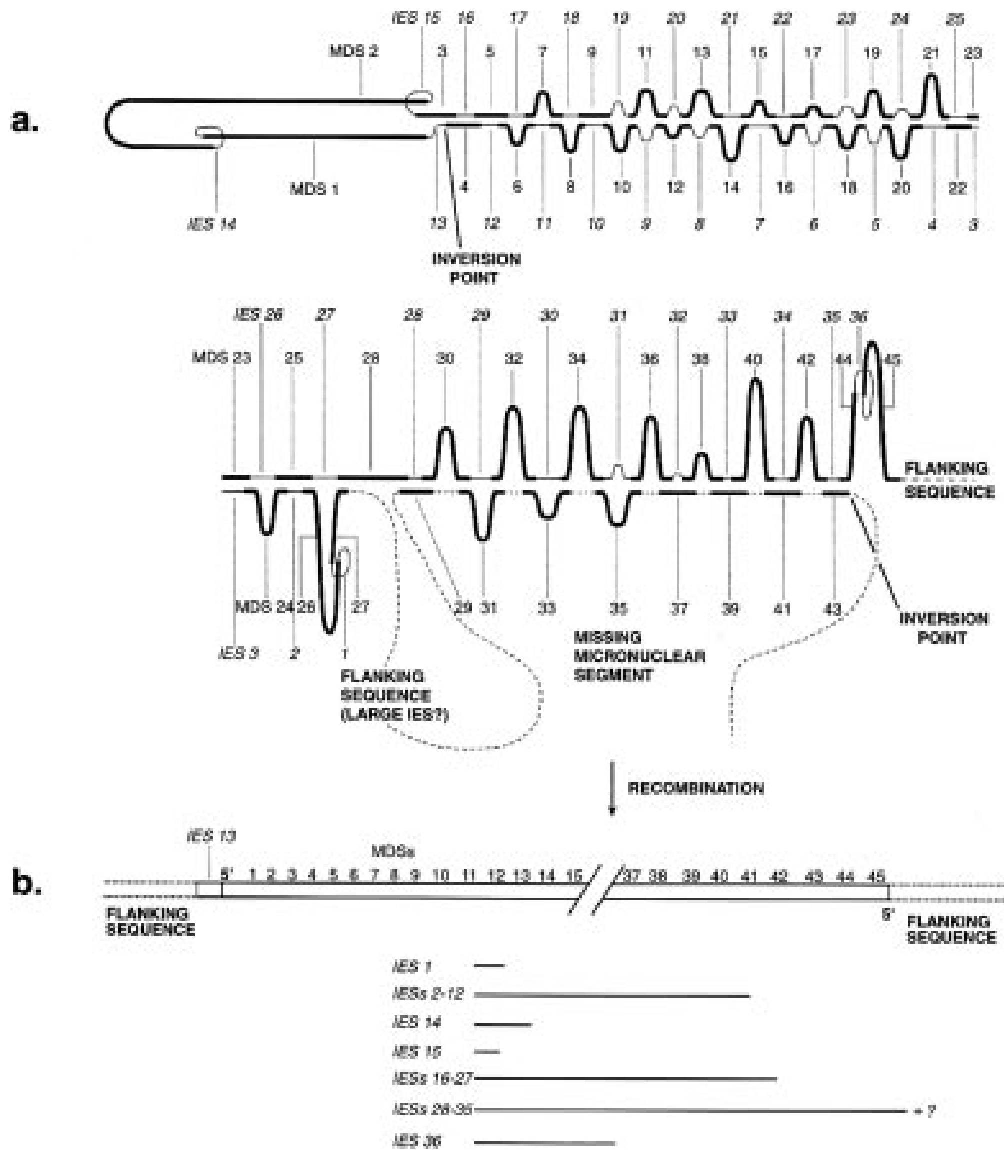

Figure 2. Recombination model for unscrambling of the micronuclear DNA pol α gene in *O.nova*. (**a**) Folding of the micronuclear DNA to bring the two members of each repeat pair into parallel alignment. Repeats occur where MDSs overlap at their ends. The MDSs and IESs are drawn to approximately proportional lengths. The missing, odd numbered MDSs 29–43 are hypothesized to be contained in a single segment of DNA. These could possibly be separated from MDS 27 in the main body of the gene by a very long (>2041 bp) IES. The lengths of the missing MDSs are calculated from the macronuclear gene sequence, knowing the sequences of even numbered MDSs 28–44 in the micronuclear clone. The unknown IESs between the missing MDSs are indicated as short dotted lines. An inversion point is present between IES 13 and MDS 4. The second inversion point is presumably present in the flanking DNA that extends from MDS 43. IESs 1, 14, 15 and 36 are excised as individual molecules. IESs 2–12 are spliced when their associated MDSs are spliced and then excised as a single, composite IES molecule. IES 13 becomes flanking sequence at the 5′-end of the unscrambled gene (see b). IESs 16–27 are spliced and excised together as a single molecule. IESs 28–35 are hypothesized to fuse with the IESs in the missing micronuclear segment and then undergo excision as a single molecule. (**b**) Splicing of MDSs in the orthodox order by recombination removes all IESs from the micronuclear gene without excision of the gene from the chromosome.

are part of a very long IES connecting MDS 27 with MDS 29, as shown in Figure 2.

The previously sequenced macronuclear DNA pol α gene (5) differs in ~2% of nucleotide positions (including leader and trailer) and in ~0.9% of amino acid positions. The question remains whether the particular micronuclear gene version in Figure 1 becomes unscrambled during macronuclear development. Sequencing of a PCR product of macronuclear DNA proved that macronuclear DNA contains two versions of the DNA pol α gene, one of which matches the previously published macronuclear sequence (5) and one of which matches the micronuclear sequence. Therefore, we conclude that the scrambled micronuclear version is processed into an unscrambled version during macronuclear development.

The MDSs of the DNA pol α gene that are numerically consecutive but disordered by scrambling have pairs of repeat sequences of 6–16 bp at their ends (Table 1). These repeat pairs are thought to guide recombination between MDSs, splicing them in the orthodox order, with accompanying elimination of one copy in each repeat pair and elimination of the intervening IESs

(2,3). For recombination to occur, it is postulated that the two repeat sequences in each pair align side-by-side in register to facilitate recombination by folding of the DNA into the pattern shown in Figure 2. Recombination of the end of MDS 3 with the end of MDS 4, MDS 4 with MDS 5, etc., would splice the MDSs into the orthodox order. Splicing of MDS 3 with MDS 4 would release the end of IES 13 at the inversion point, which could then splice to the end of the flanking sequence released by recombination of MDS 43 with MDS 44. This leaves the re-ordered DNA pol α gene still integrated in its chromosome. Excision of the gene from the chromosome is envisaged as a separate, subsequent event during or after fragmentation of the polytene chromosomes in the developing macronucleus.

In the recombination model, IES 1 (between MDSs 26 and 27; see Fig. 2 for numbering of IESs) is released as a separate molecule, IESs 2–12 are spliced into a single excised molecule, IES 13 is not excised but is displaced to a flanking position next to MDS 1 in the chromosome, IESs 14 and 15 are each excised as separate molecules and IESs 16–27 are spliced and excised as a single molecule. IES 36 would also be excised as a single molecule. The excision form of the remaining IESs is uncertain because the disposition of MDSs 29, 31, 33, 35, 37, 39, 41 and 43 is not known.

For scrambled MDS in the actin I, αTP and DNA pol α genes, repeat lengths range from 6 to 19 bp (average ~10 bp). These repeats may be long enough to signal precise recombination. A sequence of 6 bp will occur by random chance once every 4^6 bp (4096 bp). In the non-scrambled MDSs present in the three scrambled genes (actin I, αTP and DNA pol α) and in all the MDSs in the four non-scrambled genes (C2, R1, βTBP and hsp70), the repeat pairs range from 2 to 6 bp in length (average 4 bp). A repeat of 4 bp will occur by random chance once every 256 bp, which is clearly inadequate to guide correct recombination. Additional, unidentified information must help to guide removal of IESs from the extremely scrambled gene encoding DNA pol α in *O.nova*.

ACKNOWLEDGEMENT

Supported by NSF grant MCB9506287 to DMP.

REFERENCES

1. Klobutcher,L.A., Jahn,C.L. and Prescott,D.M. (1984) *Cell*, **36**, 1045–1055.
2. Mitcham,J.L., Lynn,A.J. and Prescott,D.M. (1992) *Genes Dev.*, **6**, 788–800.
3. Prescott,D.M. and Greslin,A.F. (1992) *Dev. Genet.*, **13**, 66–74.
4. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
5. Mansour,S.J., Hoffman,D.C. and Prescott,D.M. (1994) *Gene*, **144**, 155–161.
6. Ribas-Aparicio,R.M., Sparkowski,J.J., Proulx,A.E., Mitchell,J.D. and Klobutcher,L.A. (1987) *Genes Dev.*, **1**, 323–336.
7. Mitcham,J.L., Prescott,D.M. and Miller,M.K. (1994) *J. Eukaryotic Microbiol.*, **41**, 478–480.
8. Anderson,R.C., Lindauer,K.L. and Prescott,D.M. (1996) *Gene*, **168**, 103–107.
9. Wang,T.S.-F., Wong,S.W. and Korn,D. (1989) *FASEB J.*, **3**, 14–21.
10. Damagnez,V., Tillit,J., deRecondo,A.-M. and Baldacci,G. (1991) *Mol. Gen. Genet.*, **226**, 182–189.