Replication of linear mitochondrial DNA from *Paramecium*: Sequence and structure of the initiation-end crosslink

(DNA synthesis/palindromic sequences/DNA replication origin/replicative intermediates)

ARTHUR E. PRITCHARD AND DONALD J. CUMMINGS

Department of Microbiology and Immunology, University of Colorado Health Sciences Center, Denver, Colorado 80262

Communicated by David Marshall Prescott, August 17, 1981

ABSTRACT Replication of the 14-μm linear *Paramecium* mitochondrial DNA is initiated by a covalent crosslink at a unique molecular (1, 2). Unidirectional synthesis of the dimer form produces a head-to-head dimer molecule that contains a long-range palindromic sequence centered about the middle of the dimer. The dimer is then processed to two monomers. It was determined that the crosslink region does not contain protein or RNA and is cleaved by S1 nuclease. To study the replication mechanism in more detail, restriction enzyme fragments of the dimer form of the mtDNA from four species of *Paramecium* were cloned and characterized (3). The characteristics of the cloned sequences support the Goddard and Cummings model for all four species. Specifically, the cloned sequences were studied by Southern hybridization to noncloned mtDNA fragments containing the initiation region, by determination of the size of snapback products, and by determination of the symmetry of restriction endonuclease sites within the clones. Nucleotide sequences for species 1 and 4 are reported herein. The results provide details of the crosslinking event. Replication of linear DNA is currently being studied in a number of systems. One aim of these studies is to determine the mechanism of completion at the 5′ ends of the progeny strands. It appears that there is no scheme that is generally applicable. Replication of *Paramecium* mtDNA also seems to be unique.

MATERIALS AND METHODS

*Paramecium* Growth and Mitochondrial DNA Purification. *Paramecium* species 1 (primaurevul, stock 513) and species 4 (tetrauarevul, stock 51) were grown and the mtDNA was isolated as described (1).

Preparation of Plasmid DNA. Clones containing the inserted mtDNA sequences were grown, and the DNA was purified as described (3).

Sequence Determination. DNA fragments used for sequence determination were isolated from preparative poly-

acrylamide gels by electrophoresion in a dialysis bag containing electrophoresis buffer at 1:20 dilution. Fragments were dephosphorylated and then labeled at their 5′ ends by using polynucleotide kinase (Bethesda Research Laboratories, Rockville, MD) and [γ-32P]ATP (ICN) as described (4). The labeling reaction was followed by digestion with an appropriate restriction enzyme and the desired fragment was isolated by preparative polyacrylamide gel electrophoresis. Sequence determination of the isolated fragments was accomplished by using the C, C + T, A > C, and G cleavages described by Maxam and Gilbert (4).

Enzymes. Restriction enzymes were purchased from New England BioLabs, Bethesda Research Laboratories, or Boehringer Mannheim. Digestions were performed as specified by the supplier.

RESULTS

Sequence Determination Strategy. Initial experiments with the species 1 and 4 dimer initiation-region clones (3) revealed that there were few restriction endonuclease sites close to the center of the cloned insert and that, with one exception, the restriction sites were symmetric about the center as expected for a palindromic sequence. The exception was a single Sau 3A site in the species 4 clone which, as we shall show, lies within a central nonpalindromic region.

Because the known restriction sites for the species 1 clone were symmetric, a special strategy was needed to obtain small fragments that were end labeled at a unique 5′ terminus for Maxam–Gilbert sequence determination. The plasmid pBR322, containing the mtDNA inserted into the HindIII site, was cut at the EcoRI and BamHI sites in the vector, and the fragment containing the insert was isolated (Fig. 1). This procedure created a fragment with an asymmetric arrangement of *Msp* I sites. This fragment was subjected in turn to partial digestion with *Msp* I, which has two symmetric sites in the insert and two sites in the adjoining vector fragment. The partial digestion products were end labeled, and the desired fragments, containing a label at only one of the insert *Msp* I sites and a label at either the EcoRI or BamHI site (see Fig. 1), were isolated by preparative polyacrylamide gel electrophoresis and then excised with *Msp* I. The central insert fragments, possessing a label at only one of the insert *Msp* I sites, were isolated, and the sequence was determined. This fragment was 526 base pairs (bp) long, and it was possible to assay the sequence inwards approximately 330 bp from each end; thereby, sequences for both strands were obtained for the central region of this fragment. A typical sequence gel for species 1 is shown in Fig. 2. It is clear that the distal regions are palindromic; the sequences are identical from both 5′ termini, but in the central region they diverge. The point at which they diverge shall be referred to as the nonpal-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.
indrome point, the central area defined by these points is the nonpalindromic region, and the distal areas are the palindromic regions.

The strategy for species 4 was somewhat different because there is a Sau 3A site within the nonpalindromic region (Fig. 1). We used the symmetric Pst I and Alu I sites in the palindromic regions, together with the asymmetric Sau 3A site to obtain fragments labeled at a unique 5' end by the conventional technique of a second restriction digest after end labeling (4) (Fig. 1). The small species 4 fragment that was sequenced has 217 bp; therefore, it was possible to determine the nucleotides of both strands for almost the entire molecule. The larger of the two fragments that were sequenced is 688 bp long and is bounded by Alu I and Sau 3A restriction sites. Sequence determinations were made from both ends of this fragment, but it was impossible to read the exact nucleotide sequence in the central region of this fragment. However, this region contains a unit of 34 bp which apparently is repeated in tandem 11 times, and then the first 7 bp of the sequence are repeated again. It was possible to read the exactly identical sequence of several of the repeats at either end of the molecule, and the number of times the sequence was repeated could be determined from each end by the obvious repeat pattern on the sequence gels (Fig. 3). It is possible, that there may be very minor alterations of the repeat sequence, which we were unable to detect, in the very middle region of this fragment.

Determination of Nonpalindromic Sequences in Noncloned mtDNA. It was necessary to determine if the nonpalindromic sequences are also present in the noncloned mtDNA. The replication model predicts a single long-range palindromic sequence centered at the middle of the dimer, and it is possible that the nonpalindromic sequence found represents an artifact introduced by the cloning procedure. There are several lines of evidence that do not support this possibility. (i) The cloned fragments comigrated with the noncloned mtDNA on agarose and polyacrylamide gels (data not shown), and they are homologous according to Southern hybridization experiments (3). (ii) The nonpalindromic region of the species 4 mtDNA contains a 34-bp sequence that is repeated 11 times. To convert a palindromic sequence to the one that was found, the same insert, deletion, or mutation would have to occur in all 11 of the repeats. (iii) The sequences of limited regions of noncloned species 4 mtDNA were determined, and the cloned and noncloned sequences were identical in an area (containing the asymmetric Sau 3A site) that includes the nonpalindromic point (Fig. 4). Similar experiments were performed with species 1 DNA, and the same result was obtained. We conclude that the nonpalindromic region exists in the noncloned mtDNA.

Comparison of Species 1 and 4 Sequences. The sequences that were determined as described by Fig. 1 are shown in Figs. 5 and 6. The arrangement of the nonpalindromic region sequences for the two species are compared in Fig. 7. Although the two sequences are different, they have several similar features.

First, the internal area of the nonpalindromic region contains direct tandem repeats for both species. Species 1 has a 70-bp unit (positions 134–203) that is repeated again (positions 204–275) with an additional A and T included (around position 240); the exact position cannot be determined because this re-
The sequence studies presented here provide specific details concerning the replication scheme for *Paramecium* mtDNA proposed by Goddard and Cummings (2). Because of a cross-linking event at the initiation end of the linear monomer, their model predicts a dimer replicative intermediate whose sequence, when considered in its entirety, is palindromic. We determined that the sequence in the center of the dimer molecule for species 1 and 4 is nonpalindromic and provided evidence that this nonpalindromic region is not due to a cloning artifact. The sequence in the more distal regions is palindromic as predicted. This result can be reconciled with the replication model by postulating that the nonpalindromic sequence constitutes a single-stranded region that links the 5' and 3' termini at the initiation end of the linear duplex monomer (Fig. 8). The essential unique feature of this scheme is that the crosslink sequence is not palindromic. Evidence for this structure is provided by electron micrographs of the crosslinked monomer. A bubble is seen at the initiation end of crosslinked monomers under partially denaturing conditions (2).

This model is relevant to mechanisms for completing replication of the 5' ends of progeny strands in linear DNA.
FIG. 5. Nucleotide sequence of the species 1 dimer initiation region: the sequence of the central fragment bordered by symmetric Msp I sites (Fig. 1a). The nonpalindrome points (NP) are indicated by the arrow; the first nucleotide at the 5'-end of the nonpalindrome region is arbitrarily called position 1. Parentheses enclose the three repeat units described in the text.

All known DNA polymerases require a 3'-OH terminus on which to initiate strand elongation (5). Various models have been proposed, including ones involving a protein primer that has been found attached to the 5' termini of some linear DNAs (6, 7) or a mechanism requiring special terminal structures such as terminal repeats (8), palindromic termini (9), and covalently crosslinked termini (10). Experimental evidence has been obtained for these models in a number of systems such as adenovirus (11, 12), the related autonomous parvovirus (13), vaccinia virus (14, 15), and yeast nuclear DNA (16). A similar mechanism of herpes simplex virus replication has been proposed based on available experimental evidence (17). In all of these systems, the DNA display terminal symmetry elements, but detailed analysis has not provided a generally applicable replication mechanism.

Most of these models are patterned after that of Cavalier-Smith (9), which utilizes terminal palindromes that are able to form hairpin structures with a 3'-OH terminus to prime DNA synthesis. A logical extension of this type of model to mtDNA replication would be nicking of opposite strands of the duplex dimer molecule near each nonpalindrome point. The long complementary nonpalindromic region sequences would then be separated (facilitated by the A + T content), leaving

FIG. 6. Nucleotide sequence of the species 4 dimer initiation region: sequences of the two fragments enclosed by the two symmetric Alu I sites (Fig. 1b). The repeat unit is bracketed with a subset 11 to indicate the number of times it is repeated. The parenthesis after position 551 marks off the first seven nucleotides of the unit, which are repeated again. Arrows indicate nonpalindrome points (NP).
within virus repeats of yeast petite mutants, (29). mtDNA + and replication regions are replication removal of both to DNA sequences we this dict a for the termini of the parvovirus genome. The nucleotide sequences of the 3' termini of the DNA from four autonomous rodent parvoviruses have been determined (13). These results demonstrate that the terminus of each genome has a Y-shaped hairpin structure, involving 115 or 116 nucleotides. Almost all of the nucleotides in this structure are base-paired. In contrast, our sequences predict a large single-stranded terminal loop. Some base pairing in this loop is possible but examination and comparison of the sequences we have obtained does not reveal an obvious structure. The function of this region is therefore not clear. Perhaps the repeated sequences serve as a recognition site for priming enzymes. Sequences resembling the T-A-T-A box, thought to be a eukaryotic promoter, are found throughout this region.

Replication intermediates of vaccinia virus DNA also have similarities with *Paramaecium* mtDNA. The complementary strands of DNA are crosslinked at both termini (18). The sequences at the two ends of the genome are inverted repeats, about 10,000 bp in length (19). Within each inverted terminal repeat are tandem repeats, approximately 70 bp repeated 30 times (15). Although no satisfactory replication model for this genome exists, removal of both crosslinks from the parental genome appears to be an early step (20), which distinguishes it from the mtDNA replication scheme.

Although the functional significance of the obtained sequences are not known, a comparison with other sequenced replication origin regions could be useful. One obvious feature of the mtDNA initiation region is the high A + T content. Regions of varying sizes rich in A + T have been found near other replication origins, such as those of lambdoid phages (21, 22); simian virus 40, BK virus (23), and polyoma DNA (24); αX174 and G4 viral strand synthesis (25); phage T7 DNA (26); and mtDNA from *Drosophila* (27), human (28), and yeast petite mutants (29). For a number of these systems, such as mtDNA from yeast petite mutants, phage G4 DNA, and phage αX174 DNA, the region rich in A + T is flanked by clusters rich in G + C, just as we have noted here. Another significant feature of the sequences reported here is the direct tandem repeats in the nonpalindrome region, a unit of 34 bp repeated 11 times in species 4, and a 70-bp unit repeated twice in species 1. Direct repeats have been found near other replication origins. Simian virus 40 DNA has two tandem 72-bp repeats on the "late" side of the origin; deletion of one repeat plus part of the second leads to a replication-defective mutant (30). The plasmid RK2 contains eight closely spaced 17-bp direct repeats with five located within the region required for a functional origin and three within the region involved in incompatibility (31). Other examples include repeats in vaccinia virus DNA, four or five repeats of 13–19 bp near lambdoid phage origins (21, 22) and two repeats of 124 bp in the origin region of the 2-μm yeast plasmid (32).

This work was supported by a grant from the U.S. Public Health Service (GM 21949).