Properties of a Primer RNA-DNA Hybrid at the Mouse Mitochondrial DNA Leading-strand Origin of Replication*

Daniel Y. Lee‡ and David A. Clayton§

From the Department of Developmental Biology, Beckman Center for Molecular and Genetic Medicine, Stanford University School of Medicine, Stanford, California 94305-5427

Primers for vertebrate mitochondrial leading-strand DNA replication are products of transcription synthesized by mitochondrial RNA polymerase. The precursor primer RNA exists as a persistent RNA-DNA hybrid, known as an R-loop, formed during transcription through the replication origin (Xu, B., and Clayton, D. A. (1996) EMBO J. 15, 3135–3143). In an attempt to examine the precise structure of this primer RNA intermediate, we have used two methods to reconstitute model R-loops containing the mouse mitochondrial DNA origin sequence. First, we demonstrate that bacteriophage SP6 RNA polymerase can efficiently catalyze the formation of an R-loop at the mouse mtDNA origin sequence. Second, the R-loop can be assembled by annealing presynthesized RNA and supercoiled DNA template in the presence of formamide. R-loop formation by either method is dependent on specific template sequences. The reconstituted R-loop is exceptionally stable and exhibits an unexpected structure. Structural studies indicate that the RNA strand is organized within the RNA-DNA base-paired region, suggesting that the heteroduplex interaction occurs through a specific conformation. We propose that the organized structure of the R-loop is critical for primer RNA function in vitro with important implications for the RNA processing and DNA replication machinery.

The mammalian mitochondrial genome (mtDNA) is a small, closed circular molecule of ~16,000 base pairs (1, 2). Each strand of this compact supercoiled genome is transcribed from a single major promoter, the heavy-strand promoter and the light-strand promoter (LSP).1 Analyses of newly replicated DNA strands from the mtDNA control region (called the “D-loop region”), which houses both promoters and the leading-strand replication origin (O_h), have revealed an intimate relationship between transcription and the initiation of DNA synthesis. All of the major nascent DNA strands were found to initiate within ~200 nucleotides downstream of the LSP in human and mouse mitochondria (3, 4). For both systems, two species of LSP transcripts were observed: transcripts that extended beyond the replication start sites and transcripts with 3’-termini mapping precisely at or closely to the start sites of nascent DNA strands. In the mouse system, a third type of LSP transcript was demonstrated: those that were covalently attached at the 5’-termini of nascent DNA strands (4), confirming earlier observations of RNA-linked DNAs from the O_h region (5). These findings supported a model for transcriptional activation of leading-strand DNA replication. Recent results from in vitro transcription experiments using purified mitochondrial RNA polymerase (RNAP) activities have suggested that origin activation occurs through the formation of a persistent RNA-DNA hybrid at a specific region encompassing the sites of DNA replication (6). This persistent RNA-DNA hybrid, also termed an “R-loop,” has been postulated to represent the precursor form of the primer RNA that, upon endonucleolytic processing, leads to mature primers that are properly positioned on the template for DNA synthesis.

To understand the physical basis for the stable RNA-DNA interaction and to facilitate in vitro studies of primer RNA metabolism, we have reconstituted a model R-loop with the mouse mtDNA O_h sequence. In this report, we describe two methods of generating a stable triple-stranded complex of RNA and plasmid DNA. We first show that bacteriophage SP6 RNAP can substitute for mitochondrial RNAP in forming an R-loop at a discrete region that encompasses all of the known DNA synthesis start sites at the mouse mtDNA O_h. We then demonstrate that an R-loop can be assembled using a defined length RNA synthesized in vitro and annealing it onto a supercoiled plasmid template in the presence of a mild denaturant. The reconstituted R-loop is resistant to branch migration and can tolerate temperatures of up to 95 °C. Due to this stability, the R-loop can be purified to homogeneity by conventional liquid chromatography. Nuclease probing assays indicate that the RNA strand in the R-loop is precisely positioned on the DNA template and is folded in a configuration that is similar to its free form in solution. We propose that the RNA conformation facilitates the RNA-DNA interaction, which stabilizes the precursor primer RNA on the DNA template.

EXPERIMENTAL PROCEDURES

Nucleic Acids—pMR718B contains an RsaI fragment of mouse mtDNA as described previously (7). pMR718X.2 was constructed by subcloning the mtDNA fragment of pMR718B into the BamHI-PvuII sites of pSP65 such that SP6 transcription generates RNA of the reverse orientation relative to that of pMR718B. Construction of pMR718B derivatives with mutations or deletions of CSB II or CSB III was described previously (8). Plasmids were grown in DH5α and purified by two consecutive CsCl/ethidium bromide gradients to obtain the superhelical form. Buoyant density gradient-purified DNA was passed through Sepharose CL-4B (Pharmacia Biotech Inc.) and quantitated by UV absorption. A 338-nucleotide run-off RNA, corresponding to the mtDNA light-strand transcript, was synthesized from pMR718B that had been digested with NcoI endonuclease (New England Biolabs Inc.) using 40–80 units of SP6 RNA polymerase in a 50–100-μl reaction volume containing 0.5 mM each of four ribonucleoside triphosphates, 40 mM Tris (pH 7.9), 10 mM MgCl₂, 1 mM spermidine, and 1 mM dithio-
that the reaction was stopped by phenol extraction and ethanol-purification. The DNA pellet was redis- solved in 50 μl of pure water and passed through Sepahex G-50 (Pharmacia Biotech Inc.) to remove unincorporated ribonucleoside triphosphates. The DNA was precipitated again to concentrate the samples. Radiolabeling at the 5′-terminus was achieved by treatment with calf thymus alkaline phosphatase (New England Biolabs Inc.) followed by direct phosphorylation with T4 polynucleotide kinase (Pharmacia Biotech Inc.) and γ-[32P]ATP (Du Pont NEN). Radiolabeling at the 3′-terminus was achieved with [5′-32P]PPCP (Du Pont NEN) and T4 RNA ligase (New England Biolabs Inc.). Radiolabeled DNA was purified electrophoretically on denaturing polyacrylamide gels and eluted as described (7). Isolated RNA was then purified by gel filtration through Sepahex G-50, stored in diethyl pyrocarbonate-treated water, and quantified by UV absorption.

R-loop Formation by Transcription in Vitro—Transcription reactions were carried out in 10–20-μl volumes containing ~0.10 pmol of superhelical pMR718B DNA (~0.25 μg), 0.10 mM sodium triphosphates, 50 mM KCl, 20 mM Tris (pH 8.0), 1.0 mM MgCl2, and 1 mM dithiothreitol. Approximately 40–80 units of E. coli SP6 RNA polymerase were added to initiate the reactions. After 30–60 min at 37°C, the reactions were quenched by the addition of 0.1% Sarkosyl and 40 μg/ml proteinase K (Boehringer Mannheim). Proteolysis was carried out at 37°C for an additional 15 min. Samples were adjusted to ~5% glycerol and loaded onto ~0.8% agarose gels in TBE buffer (45 mM Tris (pH 8.3), 45 mM borate, and 5 mM EDTA).

Nonenzymatic R-loop Assembly—An earlier protocol used by Thomas et al. (9) to assemble R-loops in vitro was modified to optimize efficiency. DNA and RNA (10 pmol each) were mixed in formamide buffer (62% distilled deionized formamide, 25 mM HEPES (pH 7.5), 0.4 mM NaCl, and 1.25 mM EDTA) in a total volume of 100 μl in a siliconized microen- trifuge tube. The mixture was topped with an equal volume of light weight mineral oil and heat-denatured at 56°C for 2 h in a water bath. The mixture was gradually cooled to 42°C over a 2-h period, followed by an additional 12–16 h at 42°C. Thereafter, the temperature was slowly adjusted to 37°C over a 4-h period, followed by a 1-h equilibration at room temperature (22–25°C). Separation of the unincorporated RNA in the annealing mixture was achieved by passage through Sephacryl CL-4B. The free DNA and the complex were fractionated by ion exchange chromatography using BND-cellulose (10). Medium mesh BND-cellulose (Sigma) was rehydrated in 0.1 M NaCl in TE buffer (20 mM Tris (pH 8.0) and 1 mM EDTA), and resin fines were removed prior to column packing (1.5 × 12 cm polypropylene). 1 g of wet resin was used to purify ~0.01–0.1 mg of the input DNA + RNA in the annealing mixture. The sample was added to 1 ml NaCl/TE buffer and loaded onto the column (~0.5 ml/min). The column was washed with 5 ml of 0.3 M NaCl/TE buffer at the same flow rate, and the free DNA was eluted with 5 ml of 1.0 M NaCl/TE buffer. The RNA-DNA complex was eluted with 5 ml of 1.0 M NaCl + 1.0% caffeine/TE buffer. All steps were carried out at room temperature. Eluted samples were pooled and concentrated with ultrafiltration units (Centricon 50) or precipitated with ethanol for 12–16 h at ~20°C. When ethanol precipitation was carried out, the pellet was washed with ice-cold 70% ethanol and redissolved in water or low ionic strength buffers (10–50 mM KCl) and stored at 4°C; the R-loop was rehydrated at 37°C for 2 h, followed by an additional 10 min with 20–30 units of RNase-free DNase I (Boehringer Mannheim). Thereafter, the reaction was stopped by phenol extraction and ethanol-purification. The DNA pellet was redis- solved in 50 μl of pure water and passed through Sepahex G-50 (Pharmacia Biotech Inc.) to remove unincorporated ribonucleoside triphosphates. The RNA was precipitated again to concentrate the samples. Radiolabeling at the 5′-terminus was achieved by treatment with calf thymus alkaline phosphatase (New England Biolabs Inc.) followed by direct phosphorylation with T4 polynucleotide kinase (Pharmacia Biotech Inc.) and γ-[32P]ATP (Du Pont NEN). Radiolabeling at the 3′-terminus was achieved with [5′-32P]PPCP (Du Pont NEN) and T4 RNA ligase (New England Biolabs Inc.). Radiolabeled RNA was purified electrophoretically on denaturing polyacrylamide gels and eluted as described (7). Isolated RNA was then purified by gel filtration through Sepahex G-50, stored in diethyl pyrocarbonate-treated water, and quantified by UV absorption.

Nonenzymatic R-loop Assembly—An earlier protocol used by Thomas et al. (9) to assemble R-loops in vitro was modified to optimize efficiency. DNA and RNA (10 pmol each) were mixed in formamide buffer (62% distilled deionized formamide, 25 mM HEPES (pH 7.5), 0.4 mM NaCl, and 1.25 mM EDTA) in a total volume of 100 μl in a siliconized microen- trifuge tube. The mixture was topped with an equal volume of light weight mineral oil and heat-denatured at 56°C for 2 h in a water bath. The mixture was gradually cooled to 42°C over a 2-h period, followed by an additional 12–16 h at 42°C. Thereafter, the temperature was slowly adjusted to 37°C over a 4-h period, followed by a 1-h equilibration at room temperature (22–25°C). Separation of the unincorporated RNA in the annealing mixture was achieved by passage through Sephacryl CL-4B. The free DNA and the complex were fractionated by ion exchange chromatography using BND-cellulose (10). Medium mesh BND-cellulose (Sigma) was rehydrated in 0.1 M NaCl in TE buffer (20 mM Tris (pH 8.0) and 1 mM EDTA), and resin fines were removed prior to column packing (1.5 × 12 cm polypropylene). 1 g of wet resin was used to purify ~0.01–0.1 mg of the input DNA + RNA in the annealing mixture. The sample was added to 1 ml NaCl/TE buffer and loaded onto the column (~0.5 ml/min). The column was washed with 5 ml of 0.3 M NaCl/TE buffer at the same flow rate, and the free DNA was eluted with 5 ml of 1.0 M NaCl/TE buffer. The RNA-DNA complex was eluted with 5 ml of 1.0 M NaCl + 1.0% caffeine/TE buffer. All steps were carried out at room temperature. Eluted samples were pooled and concentrated with ultrafiltration units (Centricon 50) or precipitated with ethanol for 12–16 h at ~20°C. When ethanol precipitation was carried out, the pellet was washed with ice-cold 70% ethanol and redissolved in water or low ionic strength buffers (10–50 mM KCl) and stored at 4°C; the R-loop was stable for at least 6 months when stored in this manner. Samples subjected to ethanol precipitation behaved identically to those treated under all-aqueous conditions. Stabilities of R-loops assembled with nonmitochondrial origin sequences were highly variable in low ionic strength buffers or under enzyme reaction conditions, with half-lives usually ~1 h at 37°C, as determined by native gel electrophoresis and Escherichia coli RNase H digestions.

Topoisomerase Analysis—The BND-cellulose-purified R-loop substrate was enzymatically relaxed in a standard reaction buffer (50 mM KCl, 10 mM MgCl2, 20 mM Tris (pH 8), and 1 mM dithiothreitol). Approximately 3–5 μg of DNA or purified R-loop was relaxed with 4–8 units of calf thymus topoisomerase I in a 25-μl volume for 30 min at 37°C. As a control, the R-loop was pretreated with E. coli RNase H (1–3 units) at 37°C for 30 min under the same conditions, followed by an additional 30 min with topoisomerase I. Supercoiled pMR718B was partially relaxed with 4 units of topoisomerase I for 2 min at 37°C. The gel used in the reaction buffer introduced a net positive unwinding of the DNA. In the absence of MgCl2, ~5-fold more enzyme was required to achieve complete relaxation per unit time. Reactions were quenched by the addition of protease K (40 μg/ml) and SDS (0.1%) and incubated at 65°C for 10 min, followed by phenol extraction and ethanol precipitation. The DNA pellet was redisolved in 25 μl of standard reaction

![Fig. 1. Recombinant plasmids containing the mouse mtDNA control region under the control of an SP6 promoter.](http://www.jbc.org/)

**RESULTS**

Transcription by SP6 RNA Polymerase Generates a Stable R-loop at the Mouse mtDNA Replication Origin—The recombinant plasmid pMR718B contains a restriction fragment containing the D-loop control region of mouse mtDNA positioned downstream of an SP6 promoter as described previously (Fig. 1A) (7). The elements characteristic of this mtDNA control region include the three evolutionarily conserved sequences called CSBs (CSB III, CSB II, and CSB I) positioned downstream of the LSP (Fig. 1A) (11). Low yields in obtaining partially purified mouse mitochondrial RNP devoid of nucleic contaminants prompted us to use bacteriophage RNA polymerase to examine transcription-dependent R-loop formation at the mouse mtDNA sequence. This was justified based on the previously observed capacity of the bacteriophage RNAPs to promote R-loop formation, although at reduced efficiency (6). We have found that SP6 RNAP is capable of efficient R-loop formation on a superhelical template containing the mouse mtDNA control region. The addition of SP6 RNAP and nucleo-
oxide triphosphates to supercoiled pMR718B resulted in an RNA-DNA complex with altered electrophoretic migration (Fig. 1B, lane 3). Digestion with E. coli RNase H, an endonuclease specific for the RNA strand in RNA-DNA hybrids, reversed the shifted mobility to that of the original supercoiled plasmid (Fig. 1B, lane 4). No such complex was formed with transcription in the reverse orientation through the mouse mtDNA sequence in pMR718X.2 (Fig. 1B, lanes 5 and 6) or with the pSP65 vector DNA (data not shown). We conclude that the persistent RNA-DNA complex is generated in a template sequence-dependent manner by transcription through the mouse mtDNA control region in the direction of leading-strand DNA replication. These results also indicate that R-loop formation is efficiently catalyzed by bacteriophage SP6 RNAP; thus, the enzymatic requirement for this phenomenon is not restricted to mitochondrial RNAP. In further support of this point, we have observed template sequence-dependent R-loop formation at the human mtDNA O_H region with SP6 RNAP that is comparable to that generated by the partially purified human mitochondrial enzyme.\(^2\) Templates with mutations or deletions of either CBS II or CSB III were ineffective in R-loop formation with SP6 RNAP. In the course of these experiments, we have observed certain transcription reaction parameters that strongly affected R-loop formation, implicating the velocity of the ternary elongation complex in the efficiency of this process.\(^3\)

**Polymerase-independent Assembly of an R-loop at the Mouse O_H Region**—The ability to form an R-loop utilizing a heterologous RNAP suggested that the formation of the RNA-DNA complex was governed largely by the primary nucleotide sequence. To test this hypothesis, we examined whether the free RNA and a DNA template could stably associate as an R-loop in the absence of RNAP or other protein factors. An SP6 RNAP run-off transcript of this region (338 nucleotides) was synthesized and annealed onto superhelical pMR718B by partial denaturation in formamide to promote strand exchange. The regional melting temperature that allowed annealing of the RNA was determined empirically to be 42 ± 1°C in a neutral pH buffer containing 62% formamide and 0.5 M NaCl (see “Experimental Procedures”). At an equal stoichiometry of RNA and DNA, the efficiency of hybridization as assayed by native gel electrophoresis was ~50%. Quantitative separation of hybrid molecules from unincorporated nucleic acids was achieved in two column steps. The free RNA was removed from the annealing mixture by gel filtration in which the free DNA and the RNA-DNA complex were excluded (Fig. 2A). Partitioning of the hybrid complex from free DNA was achieved by ion exchange on BND-cellulose. Pure DNA was eluted from this column with 1 M NaCl, while the RNA-containing complex required the addition of 1% caffeine (Fig. 2B). Purification of the RNA-DNA complex was followed by electrophoretic analysis of pooled column fractions (Fig. 3A). The hybrid complex migrated as a relaxed circular plasmid (Fig. 3A, lane 5), consistent with the altered mobility of the complex generated by transcription (lane 7).

The R-loop reconstituted in this manner was highly stable with no detectable dissociation rate even after 6 months at 4°C or at room temperature in low ionic strength buffers (data not shown). The R-loop was tolerant to thermal melting as exemplified by the >50% retention of the hybrid molecule even after heating at 95°C for up to 5 min, a condition that fully denatures a simple RNA-DNA heteroduplex containing the same nucleotide sequence (data not shown). This complex can be cleaved with restriction endonucleases for 1–24 h of reaction time at 37°C and maintain at least 50% of the RNA-DNA complex as a linear fragment (Fig. 3A, lane 10). Although the stability of an assembled R-loop is not strictly dependent on negative superhelical tension, torsional stress on the DNA template appears to be required for R-loop assembly in vitro. Under various hybridization conditions, we have failed to observe stable R-loop assembly on nicked circular pMR718B (data not shown). Furthermore, mutated templates with deletions of CSB II or CSB III were ineffective for R-loop reconstitutions by the formamide annealing method, indicating a template sequence requirement as observed for transcription-generated hybrid formation.\(^3\)

The specificity of the RNA-DNA interaction of the reconstituted R-loop was confirmed by restriction analysis. Digestions with AseI endonuclease of free pMR718B resulted in five fragments ranging in size from 59 to 1477 bp (Fig. 3A, lane 9). AseI digestion of the R-loop resulted in the same cleavage products with an additional species migrating slower than the 610-bp fragment (Fig. 3A, lane 10). This species was determined to contain the radiolabeled RNA (Fig. 3B, lane 4) and the DNA sequence of the 610-bp AseI fragment containing the mtDNA insert (lane 6). These results confirmed that the R-loop reconstituted in formamide is assembled with the RNA strand specifically base-paired to its corresponding DNA sequence. Since R-loops can be assembled with defined RNA strands radiolabeled internally or at either terminus, this reconstitution method was employed for subsequent R-loop analyses.

**Analysis of the RNA-DNA Base-paired Region**—To obtain quantitative information on the extent of RNA-DNA base pairing within the R-loop, we examined the amount of DNA duplex unwinding associated with the annealing of the RNA strand by exploiting the relationship between linking number, duplex

---

\(^2\) D. Y. Lee and D. A. Clayton, unpublished observations.

\(^3\) D. Y. Lee and D. A. Clayton, manuscript in preparation.
helicity, and superhelicity of covalently closed circular DNAs (12). We determined the relative change in supertwisting (\(\tau\)) between the original plasmid template and the reconstituted R-loop by comparing the topoisomers generated by enzymatic relaxation with calf thymus topoisomerase I. After relaxation, the RNA in the complex was removed, and the products were analyzed by native gel electrophoresis to determine the absolute \(\tau\) (Fig. 4A). The handedness of the superhelical turns was determined by two-dimensional native/chloroquine gel electrophoresis (Fig. 4, B and C). The free DNA relaxed to an average of 2.5 positive supertwists (Fig. 4B), while the relaxed R-loop revealed an average of 15 negative supertwists (Fig. 4C), thus indicating a \(\Delta\tau\) of \(-17.5\). If a turn of the DNA helix is estimated to contain 10.5 bp, then the supercoiled plasmid was unwound by \(-184\) bp upon the annealing of the 338-nucleotide RNA. Hence, at most, approximately half of the RNA can be directly base-paired to the DNA template strand via canonical base pair interactions.

This partial RNA-DNA interaction was further examined by \(E.\ coli\) RNase H digestions of the R-loop assembled with internally radiolabeled RNA (Fig. 4D). Interestingly, RNase H digestion did not result in a random ladder of products as observed for simple RNA-DNA heteroduplexes, but instead generated products of discrete lengths ranging in size from \(-136\) to \(-195\) nucleotides. The most abundant RNA fragment of \(-185\) nucleotides is consistent with the amount of DNA unwinding as assayed by linking number analysis. These results suggest that the RNA strand is not in contact with the DNA template strand in a contiguous manner. To map the RNA-DNA contacts precisely, we reconstituted the R-loop with the RNA strand radiolabeled at either terminus and assayed for cleavage by \(E.\ coli\) RNase H.

Assembly of the R-loop with 5’-end-labeled RNA also provided an opportunity for direct comparison with the R-loop formed by transcription with SP6 RNAP. Since the first nucleotide of the SP6 transcript was a guanosine residue, transcription in the presence of \(\gamma\)-[\(\text{32}P\)]GTP permitted specific labeling of the 5’-terminus. Comparison of the \(E.\ coli\) RNase H cleavage patterns produced on the transcription-generated (Fig. 5A, lanes 2 and 3) and the formamide-reconstituted (lane 4) R-loops revealed a cluster of cleavage sites positioned within CSB II extending into the 3’-region of CSB III. This RNase H cleavage pattern was consistently observed at various levels of transcriptional activity on the plasmid template and also correlated with the extent of transcriptional pausing that occurred downstream of CSB II (Fig. 5A, lane 1). An important conclusion from this comparative RNase H cleavage assay is that the 5’-boundary of the DNA-RNA interaction in the reconstituted R-loop is similar, if not identical, to that generated by transcription.

When the RNase H assays were repeated with the R-loop assembled with 3’-end-labeled RNA, a group of prominent cleavage sites was observed within the CSB II-CSB III region (Fig. 5B). Multiple, weaker cleavage sites downstream of this region were also evident, suggesting that the RNA-DNA base pairing downstream of CSB II was either less efficient or sterically occluded from RNase H. Nonetheless, these results indicate that the RNA-DNA interactions extend from the CSB II-CSB III region to as far downstream as \(-190\) nucleotides, again consistent with the estimated amount of the DNA duplex
unwinding and the discrete RNase H-generated RNA fragments. The apparent RNase H hypersensitivity observed at CSB II-CSB III suggests that this RNA-DNA base-paired region is accessible and required for R-loop stability. Even under limit digestion conditions, the relative abundance of cleavage products did not change significantly (Fig. 5B, lane 2). RNase H nicking within this region results in dissociation of the RNA-DNA complex, thereby generating the biased cleavage pattern. This explanation is consistent with the observed requirement for CSB II-CSB III in R-loop formation via transcription.

Single-stranded Regions of the RNA in the R-loop—To determine the unpaired regions of the RNA in the R-loop, we assayed for residues susceptible to cleavage by single strand-specific ribonucleases. The RNA strand was radiolabeled at either terminus and assembled into an R-loop. 3'-End-labeled free RNA and R-loops were digested briefly with RNase T1, which cleaves at unpaired G residues, and the partial cleavage products were resolved on denaturing gels. Surprisingly, the RNase T1 cleavage pattern was nearly identical for most of the RNA sequence flanking the three CSBs (Fig. 6A, lanes 2 and 3). The regions encompassing CSB III and CSB II were clearly protected from cleavage, consistent with the deduced position of the RNA-DNA contacts as determined by RNase H digestions (Fig. 6A, lane 8). A similar nuclease protection pattern relative to the free RNA control was also observed with RNase T2, a nonspecific ribonuclease active at all unpaired RNA bases. Notable differences are the distortion of the cleavage pattern around the 3'-boundary of CSB I and the increased protection in the region immediately 3' of that site (Fig. 6A, lanes 4 and 5). Even with mung bean nuclease, an enzyme active on both single-stranded RNA and DNA, the relative cleavage protection pattern was similar at dilute enzyme concentrations and short digestion times (Fig. 6B, lanes 6 and 7). In all cases, the 5'-region upstream of CSB III was not protected from cleavage, suggesting that this region is not in contact with the DNA template. The sites protected from single-strand nicking on the R-loop correlate with sites of RNase H cleavage. Probing of the R-loop where the RNA strand was radiolabeled at the 5'-terminus confirmed the RNase T1 cleavage patterns generated on the 3'-end-labeled substrate and permitted assessment of the 5'-boundary of the strong RNA-DNA interaction at CSB II-CSB III (Fig. 6B). A summary of the results obtained from RNase T1 and RNase H digestions (Fig. 7) reveals an organized RNA strand that is base-paired at
discrete regions along the DNA template.

This structured configuration of the R-loop is accentuated when compared with an RNA-DNA heteroduplex formed with the same single-stranded RNA and its cDNA strand. This type of heteroduplex is cleaved by *E. coli* RNase H at almost every residue with nearly equal efficiency (Fig. 8, lane 4). The relatively random cleavage pattern indicates that the RNA strand is base-paired with its cDNA throughout its entire length. In contrast with what is observed for the R-loop, limit RNase H digestions result in small RNA oligomers with no distinct cleavage product (data not shown). We conclude from these results that the R-loop exhibits a unique structure that requires the strands of the DNA template to be in a certain topological state.

**DISCUSSION**

We have demonstrated the ability of an RNA strand containing the mouse mtDNA OH sequence to adopt a stable configuration with the corresponding double-stranded DNA. On supercoiled templates, RNAP drives the synthesis and spontaneous formation of an R-loop at the OH region only in the direction of leading-strand DNA replication. The enzymatic requirement for this effect at the mammalian mtDNA sequence is not stringent since bacteriophage SP6 RNAPs can efficiently catalyze this process, a feature not surprising considering the structural homology between mitochondrial RNAP of *Saccharomyces cerevisiae* and bacteriophage RNAPs (13). We have also shown that in the absence of RNAP or other protein factors, free light-strand RNA can stably associate with the DNA under denaturing conditions using formamide. This method of assembly is also template-topology-dependent and results in a stable complex that is susceptible to cleavage by *E. coli* RNase H.

A notable feature of the cleavage patterns generated on the free RNA is the discrete regions that are inaccessible to all three single-stranded nucleases we have tested. These protected regions likely represent RNA-RNA interactions as a result of intrastrand folding and thus provide a comparative measure for the extent of RNA folding in the R-loop conformation. The apparent structural similarity of the free form in solution and that in the R-loop suggests that the RNA is contorted in the path of the RNA-DNA paired region. The RNA strand appears to be partially or fully dissociated from the DNA template at discrete regions, and paradoxically, this conformation is associated with the stability of the R-loop. Since templates without CSB II and CSB III do not support R-loop formation by either method, these short elements probably facilitate and/or stabilize the RNA-DNA complex.

**Structural Implications of RNA-DNA Interactions Mediated by Poly(G) Tracts**—The strong heteroduplex interaction observed at the CSB II-CSB III region implicates a role for poly(G) tracts in the exceptional stability of the R-loop. At the yeast mtDNA replication origin (called *ori* or *rep*), the G-rich element homologous to CSB II (called GC cluster C) was necessary and sufficient for R-loop formation (6). For human mtDNA, efficient R-loop formation required CSB II as well as CSB III, an element absent in yeast mtDNA (31). Deletions of CSB II or CSB III rendered templates ineffective for R-loop assembly. Isolated poly(G) tracts (G₆ and G₇) from two other regions of mouse mtDNA were also ineffective (data not shown), suggesting a requirement for a certain nucleotide se-
sequence. The R-loop stability cannot be fully explained simply by the number of hydrogen bonds in Watson-Crick G:C base pair interactions since transcription in the reverse orientation should not change this amount in the predicted hybrid; anti- sense transcription across the mtDNA control region clearly does not support R-loop formation. Another perplexing feature of the R-loop is that the RNA strand does not base pair with the DNA template throughout its entire length. Again, a model relying only on Watson-Crick base pairing does not adequately explain the stable RNA-DNA complex since there are fewer base interactions than are potentially available in this partially melted RNA-DNA hybrid. The stability of the complex is not a function of ionic strength since the R-loop is fully stable in pure water or in salt solutions containing up to 1 M NaCl. Since the R-loop is efficiently assembled in the complete absence of proteins, stabilization by proteins such as RNAP is an unlikely explanation. Based on these observations, we suggest that the poly(G) tracts within CSB II-CSB III contribute more than the canonical base interactions, perhaps by participating in structures that are stabilized by additional hydrogen bonds.

Guanosine-rich nucleic acids have long been known to exhibit unusual properties (14–18). The presence of contiguous G and T/U tracts within CSB II and CSB III (Table I) opens the possibility of potential intrastrand and/or multiple-stranded hydrogen-bonding structures, not ruled out by the data presented here. Oligonucleotides containing the characteristic poly(G)-poly(T) sequences of telomeric DNA have been shown to associate spontaneously into intramolecular structures involving G-G base pairs stabilized by Hoogsteen hydrogen bonds (15); the monovalent cation-dependent base pairing of these telomeric sequences led to the “G-quartet” tetraplex model (16). This phenomenon is not restricted to DNA as demonstrated by the stable RNA tetraplex formed with the short RNA oligonucleotide, rUGGGGU, also in a monovalent cation-dependent manner (19). These higher order structures have been suggested to be involved in telomere function and in certain recombination reactions (15–18). It is tempting to speculate that the extraordinary stability of the mtDNA R-loop is conferred by structures involving multiple-base interactions.

Another type of potential structure formed by RNA and double-stranded DNA is the triple helix, which has been observed to form during transcription at long polypurine tracts (20–23). Triplex formation has also been proposed to modulate transcription in a direction-specific manner (22). In these examples, triplex formation was detected by the formation of persistent RNA-DNA hybrids that were resistant to single strand-specific ribonuclease, but sensitive to RNase H. Unfortunately, these enzymatic tests alone do not distinguish an RNA-double-stranded DNA triplex, in which all strands are hydrogen-bonded, from an R-loop, in which the RNA strand pairs only with the complementary DNA strand, leaving the homologous DNA strand as a non-interacting entity. Whether any of these structural motifs occur in the mtDNA R-loop awaits further physicochemical studies.

Biological Significance—Demonstration of R-loop formation at the mtDNA OH region has provided evidence that supports a model for transcription initiation of leading-strand mtDNA replication. The LSP transcripts produced by mitochondrial RNAP exist as primer precursors in a stable RNA-DNA hybrid, which is then subsequently processed by the action of a primer maturation endonuclease. The mature 3’-termini of processed RNAs are then extended by mtDNA polymerase. Thus, transcription generates primer RNAs that are properly positioned on the DNA template for replication. While this general type of origin activation mechanism has been described for prokaryotic systems such as bacteriophage T7 (24, 25) and the bacterial ColEI replicon (26, 27), the mitochondrial example is unique among eukaryotic origins.

Bacteriophage T7 replication occurs by the sequential actions of phage-encoded RNAP and DNA polymerase without the formation of an extensive RNA-DNA hybrid (25). ColEI primer synthesis involves an extensively folded 550-nucleotide RNA (called RNA II) synthesized by the host RNAP that remains stably base-paired at the origin region; cleavage with host RNase H provides the mature primer (27). Only one true chromosomal origin is known to involve transcriptional activation: oriC of E. coli. However, transcription is required only under certain conditions, and the resulting R-loop functions by melting duplex DNA for loading of DnaA protein; the transcript RNA does not prime DNA synthesis (28).

At the mouse mtDNA OH, transcripts that originate at the LSP have been shown to be covalently attached to nascent heavy strands (4), thus providing strong evidence for transcription products in directly priming DNA synthesis. If mtDNA replication requires the transcript to be stably annealed to the DNA template, then only RNAs stably base-paired at the CSB II-CSB III region would serve as functional primers. Supporting this notion, the vast majority of nascent heavy strands was previously shown to initiate from positions downstream of CSB II in both human and mouse mitochondria (3, 4). A cluster of minor but detectable DNA start sites have been mapped to within CSB II (3, 4), which may be explained by the strong transcriptional pausing observed in vitro with human mitochondrial RNAP (31) and with SP6 RNAP (Fig. 6A).

The exceptionally stable structure of the R-loop also poses a possible impediment to the completion of heavy-strand DNA replication since the RNA must be removed before covalent joining of the new daughter strands. Mitochondrial DNA polymerase appears to lack an associated RNase H activity (29); therefore, we postulate that a separate RNase H-like activity must exist to degrade the RNA primers after DNA replication has initiated. Such RNase H-like activities exist in the mitochondria since they have been observed as contaminants in partially purified mitochondrial RNAP preparations (31).4

Although transcription elements have been recognized as components of eukaryotic replication origins (reviewed in Ref. 30), the specific roles of transcription in DNA replication remain unclear. In mitochondria, the leading-strand origin requires, in principle, only the transcription machinery for activation since the transcript RNA stably associates with the DNA template without the action of accessory proteins. Perhaps nucleic acid dynamics at this region of mtDNA have evolved to function with minimal demands for trans-acting factors to be imported into the organelle. We are currently refining the structure of the mouse mitochondrial R-loop by sequence mutation analysis and structural probing assays using the reconstitution methods we have described in this report. These results also provide the opportunity to study the

---

4 D. D. Chang and D. A. Clayton, unpublished observations.

---

**TABLE I**

Naturally occurring poly(G) and poly(T/U) tracts

| Name       | Sequence              | Ref. |
|------------|-----------------------|------|
| CSB II (mouse) | GGGGGGTTTTTTGGTTTGG | 15   |
| CSB II (human) | GGGGGGTTTTTTGGTTTGG | 15   |
| GC cluster C (S. cerevisiae) | GGGGGGTTTTTTGGTTTGG | 6    |
| CSB III (mouse) | TGTTTTGGGGTTTGGCA | 11   |
| CSB III (human) | TGTTTTGGGGTTTGGCA | 11   |
| Telomeric repeat (Tetrahymena) | (TTTTGGG) | 15   |
| Telomeric repeat (Arabidopsis) | (TTTTGGG) | 15   |
| Telomeric repeat (Oxytricha) | (TTTTGGG) | 15   |
| 5 S rRNA (E. coli) | 5'-terminus | 19   |

**Biological Significance—Demonstration of R-loop formation at the mtDNA OH region has provided evidence that supports a model for transcription initiation of leading-strand mtDNA replication. The LSP transcripts produced by mitochondrial RNAP exist as primer precursors in a stable RNA-DNA hybrid, which is then subsequently processed by the action of a primer maturation endonuclease. The mature 3’-termini of processed RNAs are then extended by mtDNA polymerase. Thus, transcription generates primer RNAs that are properly positioned on the DNA template for replication. While this general type of origin activation mechanism has been described for prokaryotic systems such as bacteriophage T7 (24, 25) and the bacterial ColEI replicon (26, 27), the mitochondrial example is unique among eukaryotic origins.**
maturation of replication primers by examining the R-loop as a potential substrate for a previously identified mitochondrial RNA-processing endonuclease (7).

Acknowledgments—We thank Alison Davis, Tim Brown, Jan Paluh, and Jin Shang for helpful comments and critical review of the manuscript.

REFERENCES
1. Clayton, D. A. (1982) Cell 28, 693–705
2. Bogenhagen, D., and Clayton, D. A. (1978) J. Mol. Biol. 119, 69–81
3. Chang, D. D., and Clayton, D. A. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 351–355
4. Chang, D. D., Hauswirth, W. W., and Clayton, D. A. (1985) EMBO J. 4, 1559–1567
5. Gillum, A. M., and Clayton, D. A. (1979) J. Mol. Biol. 135, 353–368
6. Xu, B., and Clayton, D. A. (1995) Mol. Cell. Biol. 15, 580–589
7. Chang, D. D., and Clayton, D. A. (1987) EMBO J. 6, 409–417
8. Bennett, J. L., and Clayton, D. A. (1990) Mol. Cell. Biol. 10, 2191–2201
9. Thomas, M., White, R. L., and Davis, R. W. (1976) Proc. Natl. Acad. Sci. U. S. A. 73, 2294–2298
10. Gamper, H. B., and Hearst, J. E. (1982) Cell 29, 81–90
11. Walberg, M. W., and Clayton, D. A. (1981) Nucleic Acids Res. 9, 5411–5421
12. Vinograd, J., Lebowitz, J., and Watson, R. (1986) J. Mol. Biol. 193, 173–197
13. Masters, B. S., Stohl, L. L., and Clayton, D. A. (1987) Cell 51, 89–99
14. Zimmerman, S. B., Cohen, G. H., and Davies, D. R. (1975) J. Mol. Biol. 92, 181–192
15. Henderson, E., Hardin, C. C., Walk, S. K., Tinoea, I., and Blackburn, E. H. (1987) Cell 51, 899–908
16. Williamson, J. R., Raghuaraman, M. K., and Cech, T. R. (1988) Cell 59, 871–880
17. Sen, D., and Gilbert, W. (1989) Nature 344, 364–366
18. Sen, D., and Gilbert, W. (1990) Nature 344, 410–414
19. Cheong, C., and Moore, P. B. (1992) Biochemistry 31, 8406–8414
20. Reaban, M. E., and Griffith, J. A. (1990) Nature 348, 342–344
21. Reaban, M. E., Lebowitz, J., and Griffith, J. A. (1994) J. Biol. Chem. 269, 21850–21857
22. Grabczyk, E., and Fishman, M. C. (1995) J. Biol. Chem. 270, 1791–1797
23. Daniels, G. A., and Lierer, M. R. (1995) Nucleic Acids Res. 23, 5006–5011
24. Romano, L. J., Tannmori, F., and Richardson, C. C. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 4107–4111
25. Fuller, C. W., and Richardson, C. C. (1985) J. Biol. Chem. 260, 3185–3196
26. Itoh, T., and Tomizawa, J. (1990) Proc. Natl. Acad. Sci. U. S. A. 77, 2450–2454
27. Masukata, H., and Tomizawa, J. (1990) Cell 62, 331–338
28. Baker, T. A., and Kornberg, A. (1988) Cell 55, 113–123
29. Kornberg, A., and Baker, T. A. (1992) DNA Replication, 2nd Ed., pp. 209–210
30. DePamphilis, M. L. (1988) Cell 52, 635–638
31. Xu, B., and Clayton, D. A. (1996) EMBO J., 15, 3133–3143
Properties of a Primer RNA-DNA Hybrid at the Mouse Mitochondrial DNA Leading-strand Origin of Replication
Daniel Y. Lee and David A. Clayton

J. Biol. Chem. 1996, 271:24262-24269.
doi: 10.1074/jbc.271.39.24262

Access the most updated version of this article at http://www.jbc.org/content/271/39/24262

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 30 references, 9 of which can be accessed free at http://www.jbc.org/content/271/39/24262.full.html#ref-list-1