11. DNA manipulations and gel electrophoresis were performed by standard methods (*23*). The *Bacillus subtilis* tRNA^{Asp} gene was excised from pT7-1 pre-tRNA^{Asp} (*25*) and cloned into pUC19 to yield pUC153. Subsequently, pT7-1 pre-tRNA^{Asp} was digested with either Bst NI or Bam HI, its ends were filled in by treatment with the Klenow fragment of DNA polymerase, and it was digested with Eco RI and gel-purified. These fragments were ligated to pUC153 DNA that had been digested with Kpn I, treated with Klenow as above, digested with Eco RI, and gel-purified. The DNA fragments from pT7-1 pre-tRNA^{Asp} were ligated to pUC153 and used to transform a *rec*$^-$ *E. coli* host (SURE, Stratagene, La Jolla, CA); the resulting tandem duplications were named p153Bamtan and p153Bsttan depending on the enzyme used in the construction. The constructions were confirmed by DNA sequencing. The p153Bamtan duplication was used as a template for amplification of the genes for cptRNA$_{53}$ and cptRNA$_{64}$; the p153Bsttan duplication was used for amplification of cptRNA$_{53Bst}$. Amplifications were performed in 50 to 200 μl of standard buffer (*12*) that contained 6 mM $MgCl_2$ and DNA template (approximately 10 ng/ml); the amplification program was 30 cycles at 92°C for 1.5 min, 50°C for 2 min, and 72°C for 1 min. Primer pairs used for cptRNA$_{53}$ were forward (53F) (5′-TAATACGACTCACTATAGTTCGAGTCCCGTCC-3′) and reverse (52R) (5′-CCGCGACCTCCTGCGTG-3′); for cptRNA$_{64}$ the primers were forward (64F) (5′-TAATACGACTCACTATAGTCCGGACCGCCAGGAT-3′) and reverse (63R) (5′-GGGACTCGAACCCGCGA-3′). The PCR products were purified by extraction with phenol and ethanol precipitation. Product ends were made blunt with Klenow, phenol extracted, and purified by Sephadex G-50 (Pharmacia, Piscataway, NJ) spin column chromatography. The eluate was precipitated and resuspended in TE [10 mM tris-HCl (pH 8.0) and 1 mM EDTA] at one-tenth of the original PCR reaction volume.
12. R. K. Saiki *et al.*, *Science* **239**, 487 (1988).
13. J. F. Milligan and O. C. Uhlenbeck, *Methods Enzymol.* **164**, 51 (1989).
14. C. Guerrier-Takada, N. Lumelsky, S. Altman, *Science* **246**, 1578 (1989).
15. *Escherichia coli* RNase P RNA was transcribed from a derivative of pDW27 (*25*) and purified by gel electrophoresis as described (*5, 13*). Native tRNA was transcribed from pT7-1 pre-tRNA^{Asp} that had been cleaved with either Bst NI or Bam HI. Transcriptions were performed as above but with 4 mM guanosine monophosphate (GMP), 0.2 mM guanosine triphosphate (GTP), [α-^{32}P]GTP (0.3 mCi/ml), plus the other nucleoside triphosphates (1 mM each). PCR products were transcribed as for tRNA, except that polyethylene glycol 8000 (Sigma) was added to a final concentration of 8%. One microliter (~0.1 μg) of PCR template was added per 20 μl of reaction buffer. Transcripts were purified as above. The CCA half of the bipartite tRNAs was transcribed from cptDNA$_{53Bam}$ that had been digested with either Bam HI or Bst NI. The template for the precursor half of the bipartite tRNAs was amplified from a derivative of pT7-1 pre-tRNA^{Asp} that has a pre-tRNA^{Asp} gene under control of a T7 promoter but lacks most of the polylinker sequences present in pT7-1 pre-tRNA^{Asp}. The precursor half of this tRNA gene was amplified with the use of 52R (above) as the reverse primer and an M13 universal primer (5′-CCCAGTCACGACGTTGT-3′) as the forward primer. Amplifications were performed as described for cptRNA genes (*11*) with a 45°C annealing temperature. Both halves of the bipartite tRNAs were transcribed and purified as for cptRNAs. For analytical cross-linking, GMPS-primed transcripts were prepared as above, with GMPS substituted for GMP and modified with azidophenacyl bromide (Fluka, Ronkonkoma, NY) as described (*5*). The RNase P RNA used for preparative cross-links was pEcKS RNA, which is *E. coli* RNase P with an additional 26 nucleotides at the 3′ end, complementary to the pBluescript KS primer (Stratagene). The pEcKS plasmid is an Nco I–Sna BI fragment containing the T7 promoter and *E. coli* RNase P gene from pDW27 (*25*) that was treated with Klenow and ligated to pBluescript KS+ that was digested with Cla I and Nae I and treated with Klenow. Then pEcKS was digested with Xho I and transcribed with T7 RNA polymerase to produce pEcKS RNA. Similar results were seen with normal-length RNase P RNA transcripts and a primer complementary to the 3′ terminal 14 nucleotides of RNase P RNA (*16*). Unlabeled GMPS transcripts were prepared and modified with azidophenacyl bromide as described (*5*). Kinetic parameters were determined by linear regression analysis with the Enzyme Kinetics computer program (dogStar Software, Bloomington, IN).
16. J. M. Nolan, unpublished data.
17. T. Inoue and T. R. Cech, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 648 (1985).
18. D. Smith, A. B. Burgin, E. S. Haas, N. R. Pace, *J. Biol. Chem.* **267**, 2429 (1992).
19. C. Reich *et al.*, *Science* **239**, 178 (1988).
20. J. W. Brown and N. R. Pace, *Nucleic Acids Res.* **20**, 1451 (1992).
21. J. J. Dunn and F. W. Studier, *J. Mol. Biol.* **166**, 477 (1983).
22. A. Buchwalder, H. Szadkowski, K. Kirschner, *Biochemistry* **31**, 1621 (1992).
23. F. M. Ausubel *et al.*, *Current Protocols in Molecular Biology* (Wiley, New York, 1989), vols. 1 and 2.
24. C. I. Reich, thesis, Universidad de Buenos Aires (1988).
25. D. S. Waugh, C. J. Green, N. R. Pace, *Science* **244**, 1569 (1989).
26. We thank members of the Pace lab for critical review of this manuscript and for many helpful discussions. Dedicated to the memory of T. F. Nolan. Supported by Public Health Service grant GM34527 from the National Institute of General Medical Sciences and Department of Energy grant FG02-92ER2088.

17 February 1993; accepted 27 May 1993

Primers for Mitochondrial DNA Replication Generated by Endonuclease G

Jacques Côté and Adolf Ruiz-Carrillo*

Endonuclease G (Endo G) is widely distributed among animals and cleaves DNA at double-stranded $(dG)_n \cdot (dC)_n$ and at single-stranded $(dC)_n$ tracts. Endo G is synthesized as a propeptide with an amino-terminal presequence that targets the nuclease to mitochondria. Endo G can also be detected in extranucleolar chromatin. In addition to deoxyribonuclease activities, Endo G also has ribonuclease (RNase) and RNase H activities and specifically cleaves mouse mitochondrial RNA and DNA-RNA substrates containing the origin of heavy-strand DNA replication (O_H). The cleavage sites match those found in vivo, indicating that Endo G is capable of generating the RNA primers required by DNA polymerase γ to initiate replication of mitochondrial DNA.

Endo G is a low-abundance endonuclease of vertebrates (*1*) that has been purified from calf thymus as a dimer of two identical subunits of 26 to 28 kD (*2*). The term Endo G was originally coined (*1*) to reflect the preference of the partially purified nuclease for the G strand of $(dG)_n \cdot (dC)_n$ DNA tracts. However, the highly purified nuclease nicks either strand of the homopolymer with similar kinetics (*2*), producing 5′-phosphoryl and 3′-hydroxyl ends (*1*). Although the occurrence of $(dG)_n \cdot (dC)_n$ tracts next to regions of deletions, translocations, and DNA hypervariability suggested a causal relation (*1*), the function of Endo G remained undetermined.

To clone the Endo G cDNA, the NH_2-terminal sequence of the purified nuclease from calf thymus and liver (*3*) was determined (Fig. 1) (*4, 5*). A short Endo G cDNA probe was obtained by polymerase chain reaction (PCR) of total calf liver cDNA with degenerate oligonucleotide primers derived from the NH_2-terminal (AGLPAVP) and COOH-terminal (ELAKYGL) Endo G sequence (*6*) (Fig. 1). This DNA probe hybridized with RNA of ~1000 nucleotides in length in blots of polyadenylated RNA from bovine and other mammalian species, including human (HeLa cells).

An end-labeled 53-nucleotide fragment of the Endo G cDNA hybridized with 24 clones in a partial calf liver cDNA expression library (*7*). All 24 clones contained essentially the same sequence, and none was in the proper reading frame to produce a β-galactosidase fusion protein, suggesting that expression of Endo G was deleterious in *Escherichia coli*. From the initial cDNA enrichment and the frequency of positive clones, we estimate that Endo G mRNA represents ≈1/60,000 of all bovine liver mRNA. This low concentration is in accord with the low abundance of the protein (*2*).

The sequence (*8*) of the longest cDNA clone (pbEG4) contained an open reading frame (ORF) predicting a protein of molecular weight 32,261 (Fig. 1), larger than the Endo G monomer (*2*). The NH_2-terminal sequence of the purified Endo G protein did not correspond to that predicted by the cDNA but rather coincided with amino acids 49 to 75 (Fig. 1). Hybridization of

Cancer Research Center and Department of Biochemistry, Medical School of Laval University, L'Hôtel-Dieu de Québec, Québec, Canada G1R 2J6.

*To whom correspondence should be addressed.

genomic DNA with pbEG4 indicated that Endo G is encoded by a single-copy gene. Because there is no translation initiation site corresponding to the mature Endo G protein within the cDNA sequence, we concluded that Endo G was synthesized as a precursor protein and subsequently processed between amino acids 48 and 49 to yield the mature protein of 27,340 daltons.

Comparison with known protein sequences indicated that Endo G is homologous to Nuc1, a 329–amino acid endonuclease from *Saccharomyces cerevisiae* mitochondria (9). Amino acids 68 to 294 of Endo G are 42% identical (67% similar, adding conservative substitutions) to amino acids 61 to 295 of the yeast protein (Fig. 1). In contrast, the processed presequences have not been conserved, although both share the peptide GLGA in the middle of the presequence. Amino acids 140 to 202 of both nucleases also show homology to the central region (amino acids 106 to 171) of a 266 amino acid–long secreted endonuclease from *Serratia marcesens* (*10*) (42% identity and 51% similarity with Endo G, and 45% identity and 54% similarity with Nuc1, Fig. 1), which also has the GLGA sequence near its NH_2-terminus.

Despite the homology between Endo G and Nuc1, their enzymatic activities are different. Digestion of linear p554-3′ DNA with Nuc1 did not produce the 0.4- and 2.9-kb fragments diagnostic of Endo G cleavage at the $(dG)_{17}\cdot(dC)_{17}$ tract (*1*, *2*) (Fig. 2). The cleavage pattern produced by Nuc1 is similar to that produced by *Neurospora crassa* endo-exonuclease (*11*) (Fig. 2), a cellular nuclease that shares immunological epitopes with Nuc1 but not with Endo G (*12*). Therefore, the ability to preferentially recognize $(dG)_n\cdot(dC)_n$ tracts is a distinct property of the vertebrate nuclease.

Antibodies raised against the synthetic peptide AGLPAVPGAPAG (NH_2-terminus of mature Endo G) (*13*) recognized a single protein band corresponding in size to the mature Endo G monomer (*2*) in immunoblots of crude extracts from calf liver and of purified Endo G preparations. Under the same conditions, the in vitro translation products of the complete and truncated (lacking the 48 amino acid–long presequence) reading frames (*14*) were distinguishable in size (*12*). Endo G immune sera but not preimmune sera precipitated Endo G activity, subsequently assayed by the endonucleolytic cleavage of linear p554-3′ DNA (Fig. 2). Therefore, we concluded that the presequence was removed shortly after synthesis, resulting in active Endo G.

To determine whether the Endo G cDNA directed expression of the nuclease in vivo and to examine the role of the presequence, we constructed expression vectors containing the complete Endo G cDNA (pSVT7EG) or a truncated cDNA in which the presequence was deleted (pSVT7tEG) (*14*). These plasmids, and their antisense constructs, were then cotransfected with pSV_2CAT into monkey COS-7 cells (*15*), and cell extracts were assayed for Endo G activity. Only cells transfected with the plasmid containing the complete cDNA in the sense orientation showed increased Endo G activity (Fig. 2). Immunoblot analysis indicated the presence of mature Endo G in extracts of cells transfected with pSVT7EG, but no precursor protein could be detected (*12*). In contrast, cells transfected with pSVT7tEG showed no increase in Endo G activity (Fig. 2) even though the protein was detected by immunoblot analysis and by immunocytochemistry (Fig. 3). Endo G translated in vitro from the truncated cDNA was also inactive. We

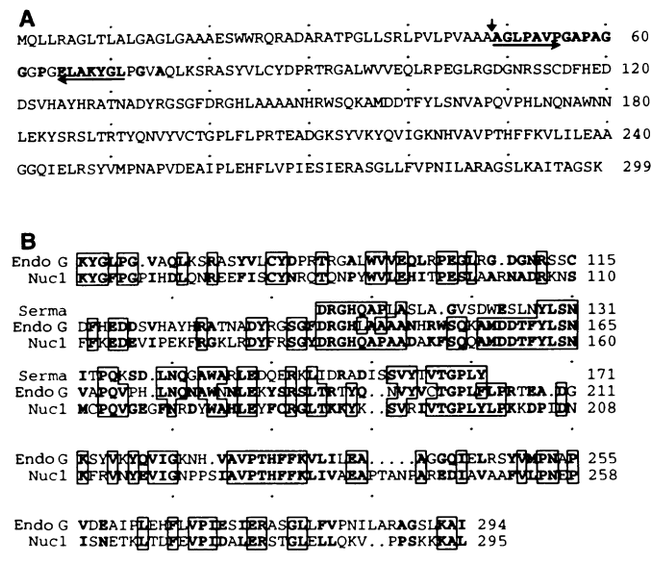

Fig. 1. (**A**) Sequence of the Endo G protein deduced from the cDNA clone pbEG4 (*5*, *8*). Amino acids in bold correspond to the NH_2-terminal sequence of purified Endo G. Horizontal arrows indicate the regions covered by the primers used in step PCR. The vertical arrow shows the boundary between the precursor and mature Endo G. (**B**) Comparison of the amino acid sequences of Endo G, yeast Nuc1, and *S. marcesens* secreted nuclease. Only the most conserved regions are shown. Dots in the sequences represent interruptions for maximal alignment. Conservative substitutions are in bold and identities are boxed.

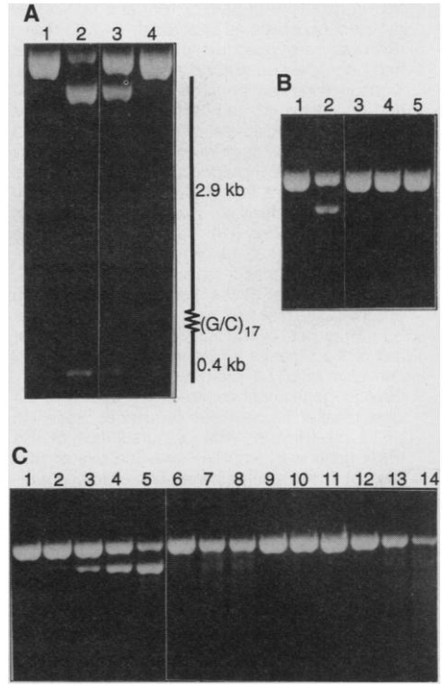

Fig. 2. Enzymatic activities of Endo G. (**A**) Endo G activity precipitated by peptide antibodies. Step (x) Endo G was reacted with IgG from preimmune and immune sera, and the nuclease activity was determined from equal portions of the immunosupernatant or immunoprecipitate by digestion of Hind III–linearized p554-3′. Lanes 1 and 2, precipitate from preimmune and immune sera; lanes 3 and 4, supernatant from preimmune and immune sera. (**B**) The cDNA of pbEG4 directs expression of Endo G in vivo. Expression vectors pSVT7EG or pSVT7tEG were cotransfected with pSV_2CAT into COS-7 cells. Endo G activity was determined in fractions of whole extracts normalized to have equal CAT activity. Lane 1, no transfection; lanes 2 and 3, transfection with pSVT7EG in the sense and antisense orientations; lanes 4 and 5, transfection with pSVT7tEG in the sense and antisense orientations. Reactions were carried out for 60 min at 25°C. (**C**) Specificity of Endo G, Nuc1, and *N. crassa* endonucleases. Hind III–linearized p554-3′ DNA was digested at 25°C with 2.5 U of step (xi) Endo G for 0, 5, 15, 30, and 60 min (lanes 1 through 5); 0.3 μg of a mitochondrial extract from wild-type *S. cerevisiae* for 5, 15, and 30 min (lanes 6 through 8); 0.38 μg of a mitochondrial extract from *S. cerevisiae Nuc1* null mutant (*37*) for 5, 15, and 30 min (lanes 9 through 11; no endonuclease activity could be detected); and 0.75×10^{-3} U of *N. crassa* endo-exonuclease for 1, 5, and 10 min (lanes 12 through 14). In (B) and (C) the 0.4-kb fragment is not shown for illustration purposes.

suggest, therefore, that the presequence is necessary for the correct folding of Endo G (*16*).

We have found Endo G activity in nuclei (*1*) and mitochondria (*2, 12*) from several vertebrates, and a nuclease activity related to Endo G has independently been found in bovine heart and rat liver mitochondria (*17, 18*). We used immunocytochemistry to analyze the cellular location of Endo G. COS-7 cells transfected with pSVT7EG were reacted with affinity-purified antibodies to bovine Endo G and a monoclonal antibody (MAB1273) that recognizes a mitochondrial antigen (*19*). The granular cytoplasmic fluorescence of the two antibodies coincided (Fig. 3) in all cases. When cells were transfected with pSVT7tEG, the fluorescence was located in the nucleus and not in mitochondria. Cells transfected with the antisense constructs showed no Endo G staining. Thus, in vivo, most Endo G is located in mitochondria, and the presequence is required for its targeting to the organelle. The mitochondrial localization of Endo G and its sequence homology with the *Serratia* endonuclease supports the hypothesis that eukaryotic mitochondria originated from prokaryotic endosymbionts (*20*).

Endo G in cells transfected with pSVT7tEG was confined to the extranucleolar chromatin of the nucleus (Fig. 3). A portion of Endo G seems normally to be localized in nuclei. Hepatocytes of calf liver sections (*19*) (Fig. 3) displayed intense granular cytoplasmic fluorescence, as expected from their high content of mitochondria, and also fluorescence in the nuclei. Nuclear fluorescence did not appear to be spurious because the nucleoli were always visible as dark spots. No fluorescence was seen when the antibody was preincubated with an excess of the immunizing peptide. As is the case with Endo G, removal of the presequence from another protein shared by nuclei and mitochondria, the N^2,N^2-dimethylguanosine–specific tRNA methyltransferase (m_2^2Gtase) of yeast, results in the inactivity of the nuclear form and absence from the mitochondria (*21*).

The mitochondrial targeting signal of Endo G has no obvious homology to that of other nuclear-encoded mitochondrial proteins, as is often found to be the case with such proteins, and it does not have an overt amphiphilic character (*16*). No nuclear targeting signal was found, although that is not unusual (*22*).

A likely target for Endo G activity in animal mitochondrial DNA (mtDNA) would be the region upstream of the $tRNA^{Phe}$, and the conserved sequence box–II (CSB-II) in the displacement-loop (D-loop) region, both of which contain $(dG)_n \cdot (dC)_n$ tracts (of length 6 to 19 in bovine mtDNA) (*23, 24*). Endo G efficiently cleaves these lengths of tracts (*1, 2*), and the G-specific nuclease from bovine heart mitochondria preferentially recognizes those sequences in vitro (*17*). Endo G may thus be involved in maintenance of mtDNA copy number or initiation of recombination. However, Endo G could also have other functions if it had activities other than deoxyribonuclease.

We examined whether Endo G had RNase activity by digesting RNA synthesized from pBT29 (*25*) that contains $(dG)_{29} \cdot (dC)_{29}$. Endo G cleaved the C transcript virtually at every C with little attack at other positions (Fig. 4) or the G transcript. The lack of recognition of the G runs in vitro may not reflect an intrinsic property of the nuclease but rather the secondary structure of RNA, because the G tract can form hairpin fold-back structures (*26*). This possibility was supported by the resistance of the central G residues of the G strand to ribonuclease (RNase) T1 digestion at 50°C in the presence of 8 M urea (Fig. 4). Pretreatment of Endo G with micrococcal nuclease had no effect on its RNase activity or specificity. The RNase activity of Endo G could be involved in processing of polycistronic transcripts by cleaving the C tract immediately upstream of the $tRNA^{Phe}$. However, as is the case of its single-stranded DNase activity (*2*), Endo G may specifically recognize other RNA sequences.

The D-loop region of vertebrate mtDNA contains the promoters for transcription of the light strand (LSP) and heavy strand (HSP) (*27*). Downstream from the LSP are three conserved sequence blocks (CSB-III, CSB-II, and CSB-I) and the origin of leading-strand replication (O_H). Leading-strand replication appears to be primed by RNA transcripts from LSP, and the transition from RNA to DNA encompasses sequences CSB-I, CSB-II, and CSB-III. To examine whether Endo G could be implicated in this process, we assayed its RNase activity on in vitro L-strand transcripts of the O_H region of the mouse mtDNA (pMR718B) (*28*). We used the mouse substrate, which has an organization of the D-loop region similar to that of the bovine mtDNA, because it is the only system in which RNA has been found covalently linked to the 5′ ends of nascent heavy-strand DNA (H-DNA), and the in vivo RNA 3′ ends have been aligned with DNA 5′ termini (*29*). The use of the heterologous enzyme is justified because rodent and bovine Endo G have the same size and specificity and are homologous at the nucleic acid level (*1, 12*). Endo G cleaved the mouse substrate at three main regions, two of which (B and C) correspond to RNA 3′ ends produced in vivo (Fig. 4) (*29*).

The presence of ribonucleotides at the 5′ ends of nascent H-DNA argues that a putative RNase H activity could be responsible for generation of the 3′-hydroxyl RNA ends required by DNA polymerase γ to

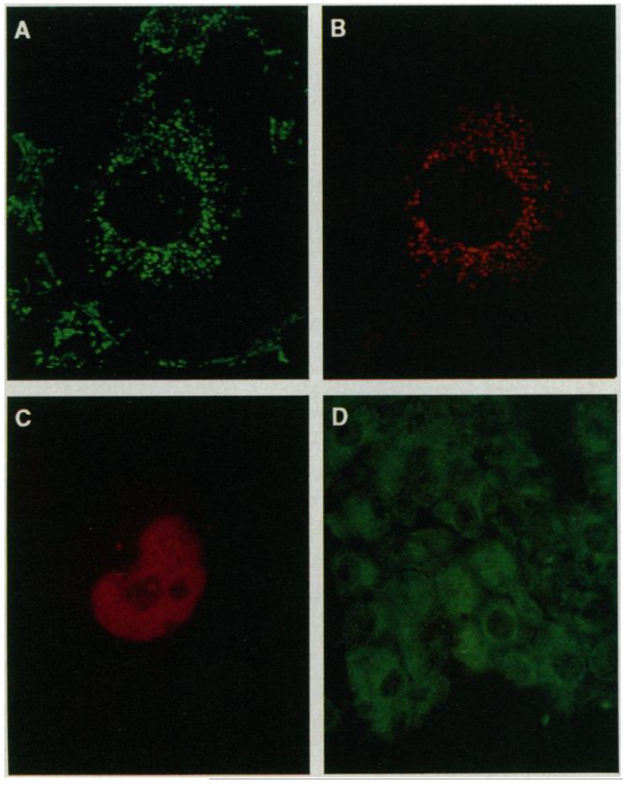

Fig. 3. Subcellular localization of Endo G. COS-7 cells were transfected with pSVT7EG (A and B) or pSVT7tEG (C) and reacted with rabbit antibodies to Endo G and mouse MAB1273 antibodies to mitochondria and then stained with the suitable secondary antibodies. (**A**) Staining for antibody to mitochondria; (**B**) staining for antibody to Endo G [same field as in (A)]; (**C**) staining for Endo G antibodies. Because the bovine Endo G antibody does not cross-react with simian Endo G, only fluorescence of the transfected cells can be detected. (**D**) Staining of a section of calf liver reacted with Endo G antibody and FITC-conjugated secondary antibody.

initiate replication. Endo G has also an RNase H activity that produced a series of specific cleavages on the RNA moiety of DNA-RNA hybrids of the O_H region (*30*). Site 4 mapped at the middle of CSB-II (Fig. 4), precisely where the transition from ribonucleotide to deoxyribonucleotide of the 5′ ends of nascent H-DNA was found to take place, downstream of the RNase site B (Fig. 4) where a perfect alignment of in vivo RNA 3′ ends and DNA 5′ ends was observed (*29*). In addition, sites 3 and 5 mapped at CSB-III and between CSB-II and CSB-I where 3′ ends of RNA have been mapped in human (*31*) and mouse mtDNA (*29*), respectively. The transition site at CSB-I (*29*) was also cleaved by Endo G, although at lower frequency than the other sites, and could only be seen at a higher extent of digestion (*12*).

The relative sensitivity of each site does not strictly parallel the relative abundance of the in vivo ends. This could be due to differential stability or rate of processing of the products, which may depend on changing physiological stimuli. For example, the type of 5′ termini of 7*S* D-loop heavy-strand DNA (DH-DNA) from bovine mitochondria differs as a function of cell growth (*32*). Although the absence of ribonucleotides at the 5′ termini of bovine H-DNA (*32*) suggests that the molecules were processed, the main 5′ termini in proliferating cells mapped at CSB-II, and between CSB-II and CSB-I, in the same regions recognized by Endo G in the mouse mtDNA. Representation of primers in replicons could also be modulated by several mechanisms, such as an intrinsically different elongation rate of RNA polymerase through the O_H region. Thus, different regions of the RNA-DNA hybrid and the nascent RNA transcript would be exposed to the action of Endo G for different times.

Although the function of Endo G in vivo remains to be shown, we suggest that it is responsible for the generation of RNA primers required by DNA polymerase γ for initiation of mtDNA replication. A similar mechanism appears to be used by plasmid ColE1 (*33*), where the primers for DNA polymerase are produced near the replication origin by RNase H cleavage of hybrids formed by primary transcripts with template DNA. The role that we propose for Endo G was previously suggested for RNase MRP (an RNA-containing RNase) (*28*). However, recent evidence has cast serious doubts about the presence of such an activity in mitochondria (*34*).

The D-loop region of animal mtDNA is the most divergent portion of the mitochondrial genome. However, the CSBs have been conserved in vertebrates (*35*). Hence, it appears that CSB-II, which is G·C–rich, and the preference of Endo G for G·C tracts have been mutually conserved during vertebrate evolution.

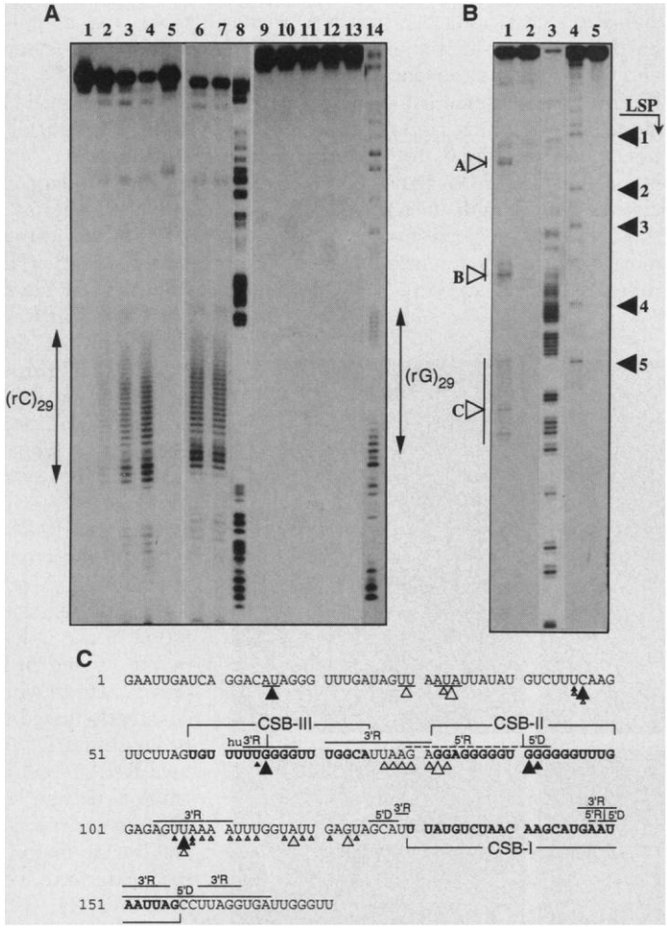

Fig. 4. Endo G has RNase and RNase H activities. (**A**) RNA containing an $(rC)_{29}$ tract (lanes 1 through 8) and an $(rG)_{29}$ tract (lanes 9 through 14) was labeled at the 3′ end and digested at 25°C with 1 U of step (xi) Endo G for 0 min (lanes 1 and 9), 5 min (lanes 2 and 10), 15 min (lanes 3 and 11), and 30 min (lanes 4 and 12). Lanes 5 and 13, incubation with Endo G buffer alone; lanes 6 and 7, preincubation of Endo G for 15 min at 37°C with, respectively, 2.5 mM $CaCl_2$ and 2.5 mM $CaCl_2$ + 1 U of micrococcal nuclease, before digestion in the presence of 5 mM EGTA; lanes 8 and 14, digestion with T1 RNase. (**B**) Recognition of the L-strand transcript and DNA-RNA hybrid of the mouse mtDNA O_H region by Endo G. A 268 nucleotide–long transcript from pMR718B containing the L-strand transcript from +1 to +205 labeled at the 3′ end was incubated with 3.5 U of Endo G (lane 1), with the nuclease buffer (lane 2) for 10 min, or with RNase T1 (lane 3). A DNA-RNA hybrid of the same region, labeled at the RNA 3′ end, was incubated for 20 min at 25°C with 3.5 U of Endo G (lane 4) or with the nuclease buffer (lane 5). The site of transcription initiation of the L-strand promoter (LSP) is indicated. (**C**) Mapping of the Endo G cleavage sites of the L-strand transcript (open arrowheads) and RNA·DNA hybrids (solid arrowheads). The positions of CSB-I, CSB-II, and CSB-III and the distribution of in vivo RNA 3′ ends (3′R) and nascent H-DNA 5′ ends (5′R and 5′D) are indicated. The initiation site of the L-strand transcript (+1) corresponds to nucleotide 16,183 of the mouse mtDNA sequence (*38*).

REFERENCES AND NOTES

1. A. Ruiz-Carrillo and J. Renaud, *EMBO J.* **6**, 401 (1987).
2. J. Côté, J. Renaud, A. Ruiz-Carrillo, *J. Biol. Chem.* **264**, 3301 (1989).
3. Endo G was purified essentially as described (*2*) by 11 (i to xi) fractionation steps.
4. The NH_2-terminal sequence of Endo G from calf liver and thymus was AGLPAVPGAPAGGXP(V/I/G)ELAKYGLX(G/P)XA [M. Moos, Jr., N. Y. Nguyen, T.-Y. Liu, *J. Biol. Chem.* **263**, 6005 (1988)].
5. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
6. The sense TCTGAATTCGCTGGSCTNCCNGCNGTNCC and antisense CTYAANCGNTTYATRCCSGATTCGAACGACGTCT primers (underscored nucleotides correspond to Eco RI and Hind III sites) were used to amplify single-stranded cDNA synthesized by Moloney murine leukemia virus reverse transcriptase from oligo(dT)-primed calf liver poly(A)$^+$ RNA. PCR conditions were as follows: 94°C for 1 min, 70°C for 0.5 min (2 cycles); 94°C for 1 min, 68°C for 0.5 min (2 cycles); 94°C for 1 min, 66°C for 0.5 min (2 cycles); 94°C for 1 min, 60°C for 0.5 min (30 cycles). DNA (average size, 90 bp) was purified and re-amplified by 30 cycles, each being 94°C for 1 min, 60°C for 0.5 min. The products of the second amplification were digested with Eco RI and Hind III and cloned in Bluescript SK$^+$ (Stratagene). DNA was sequenced with a T7 DNA polymerase kit (Pharmacia).
7. A library was constructed (*36*) with size-fractionated calf liver cDNA of 0.7 to 1.3 kb, with the use of λZAP II DNA (Stratagene). DNA was packaged with Gigapack II Gold extracts (Stratagene) as recommended by the manufacturer. Hybond-N phage lifts were hybridized at 55°C in 6× standard saline citrate (SSCE), 5× Denhardt's solution, 40% formamide, 0.2% SDS, 20 mM sodium phosphate (pH 7.4), and yeast tRNA (100 μg/ml) with 5′ end–labeled oligonucleotide AGGCCGTATTTGGCTAACTCGCCCGGGCCGCCACCCGCGGGAGCCCCTGGCAC (6 ng/ml; 7 × 10^8 cpm/μg) for 18 hours. The filters were washed in 6× SSCE, 0.1% SDS at 67°C.

8. Each strand of pbEG4 was sequenced twice with both normal and a mixture of deaza–deoxyadenosine triphosphate and deaza–deoxyguanosine triphosphate, and specific oligonucleotide primers. The cDNA of pbEG4 is 1008 nucleotides long, and its sequence has been submitted to EMBL, GenBank, and DDBJ nucleotide sequence databases under the accession number X72802.
9. R. D. Vincent, T. J. Hofmann, H. P. Zassenhaus, *Nucleic Acids Res.* **16**, 3297 (1988).
10. T. K. Ball, P. N. Saurugger, M. J. Benedik, *Gene* **57**, 183 (1987).
11. T. Y. K. Chow and M. J. Fraser, *J. Biol. Chem.* **258**, 12010 (1983).
12. J. Côté and A. Ruiz-Carrillo, unpublished results.
13. The immunizing synthetic peptide had the additional KC sequence at the COOH-terminus for solubility and cross-linking purposes. Preimmune and immune rabbit immunoglobulin G (IgG) were purified by ammonium sulfate precipitation and CM Affi-gel Blue (Bio-Rad) chromatography. For purification of the antibody by affinity chromatography, the peptide was cross-linked to bovine serum albumin and coupled to cyanogen bromide–activated Sepharose (Pharmacia). Antibodies bound to the resin were eluted with 3.5 M $MgCl_2$, 10 mM sodium phosphate (pH 7.3).
14. The expression vector pSVT7EG was constructed by ligation of the Eco RI fragment of pbEG4 at the Eco RI site of pSVT7 [P. Bird, M.-J. Gething, J. Sambrook, *J. Cell Biol.* **105**, 2905 (1987)]. The two orientations of the cDNA were obtained. For pSVT7tEG, the sequences of the cDNA coding for amino acids 1 through 48 were deleted by site-directed mutagenesis using PCR. The mutagen sense primer was TCTGAATTCGCCATGGCGG-GGCTTCCCGCGGTGCCA, where the underlined nucleotides surrounding the initiator codon conformed to Kozak's rule [M. Kozak, *Cell* **44**, 283 (1986)]. The antisense primer was ATCGAATTC-GTCCGCCCCAAGGACTTTTCAG. Vent polymerase (NEB) was used for amplification according to the manufacturer's instructions. Thirty cycles of amplification were used, each cycle being 94°C for 1 min, 72°C for 2 min. Amplified DNA was digested with Eco RI and ligated to the Eco RI site of pSVT7. The constructs were verified by DNA sequencing.
15. Exponentially growing COS-7 cells (5×10^5 per 78 cm^2 dish) were transfected for 20 hours with 25 μg of the Endo G expression vectors and 4.5 μg of pSV_2CAT by the calcium phosphate method. Cells were harvested 48 hours after transfection and extracted (*1*). Chloramphenicol acetyltransferase (CAT) activity was used to normalize the efficiency of transfection.
16. F. U. Hartl, N. Pfanner, D. W. Nicholson, W. Neupert, *Biochim. Biophys. Acta* **988**, 1 (1989); G. Attardi and G. Schatz, *Annu. Rev. Cell Biol.* **4**, 289 (1988).
17. R. L.,Low, O. W. Cummings, T. C. King, *J. Biol. Chem.* **262**, 16164 (1987).
18. R. L. Low, J. M. Buzan, C. L. Couper, *Nucleic Acids Res.* **16**, 6427 (1988).
19. Cells on cover slides were fixed with 3.7% formaldehyde in phosphate-buffered saline and post-fixed with methanol and 90% acetone. Cells were reacted with affinity-purified rabbit antibodies to Endo G and MAB1273 (Chemicon, Inc.) Fluorescein isothiocyanate (FITC)–labeled goat antibody and Texas red–labeled donkey antibody were used as secondary antibodies. Calf liver samples were obtained from a local slaughterhouse, frozen within 10 min of the animal's death, and sliced into sections of about 10 μm in thickness. Slides were observed with a Bio-Rad MRC-600 confocal imaging system mounted on a Nikon Diaphot-TMD (×60 objective lens with a 1.4 numerical aperture). A 630-nm filter was used for dual labeling to cut FITC emission in the Texas red range.
20. D. Yang, Y. Oyaizu, H. Oyaizu, G. J. Olsen, C. R. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4443 (1985).
21. A. M. Rose, P. B. M. Joyce, A. K. Hooper, N. C. Martin, *Mol. Cell. Biol.* **12**, 5652 (1992).
22. C. Dingwall and R. A. Laskey, *Trends Biochem. Sci.* **16**, 478 (1991).
23. S. Anderson *et al.*, *J. Mol. Biol.* **156**, 683 (1982).
24. W. W. Hauswirth, M. J. van de Valle, P. J. Laipis, P. D. Olivo, *Cell* **37**, 1001 (1984).
25. Plasmid pBT29 was constructed by cloning of the Bam HI insertion of pUC12T29 (*2*) at the Bam HI site of pBluescript SK^+ (Stratagene). In vitro transcription of Not I–linearized pBT29 with T7 RNA polymerase, and Xba I–linearized pMR718B (*28*) with SP6 polymerase was carried out according to the manufacturer's instructions (Gibco-BRL). RNA was labeled at the 3′ end with [^{32}P]pCp (specific activity 3000 Ci/mmol, DuPont-NEN) and RNA ligase (BRL) (*28*) and purified by electrophoresis in a 6% polyacrylamide gel containing 8 M urea.
26. I. G. Panyutin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 867 (1990).
27. D. A. Clayton, *Trends Biochem. Sci.* **16**, 107 (1991).
28. D. D. Chang and D. A. Clayton, *EMBO J.* **6**, 409 (1987).
29. A. M. Gillum and D. A. Clayton, *J. Mol. Biol.* **135**, 353 (1979); D. D. Chang, W. W. Hauswirth, D. A. Clayton, *EMBO J.* **4**, 1559 (1985).
30. DNA-RNA hybrids were obtained by annealing of 3′ end–labeled RNA to the oligonucleotide CTAC-GGTGAAGAATC (corresponding to nucleotides 15984 through 15998 of the mtDNA) (*38*) and extension at 45°C with Superscript RNase H^- reverse transcriptase (Gibco-BRL) according to the manufacturer's instructions. Hybrids were also obtained by annealing of the labeled RNA to denatured DNA. Hybrids were purified by electrophoresis in a 5% polyacrylamide gel in tris-borate EDTA.
31. D. D. Chang and D. A. Clayton, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 351 (1985).
32. T. C. King and R. L. Low, *J. Biol. Chem.* **262**, 6204, 6214 (1987).
33. T. Itoh and J. Tomizawa, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2450 (1980).
34. T. Kiss and W. Filipowicz, *Cell* **70**, 11 (1992); J. N. Tooper, J. L. Bennett, D. A. Clayton, *ibid.*, p. 16.
35. M. W. Walberg and D. A. Clayton, *Nucleic Acids Res.* **9**, 5411 (1981).
36. E. Prats, L. Cornudella, A. Ruiz-Carrillo, *ibid.* **17**, 10097 (1989).
37. H. P. Zassenhaus *et al.*, *ibid.* **16**, 3283 (1988).
38. M. J. Bibb *et al.*, *Cell* **26**, 167 (1981).
39. We thank C. Lazure and M. Blum for NH_2-terminal sequence determination, H. P. Zassenhaus for *S. cerevisiae* mitochondria from wild-type and *Nuc1* null mutants, M. Fraser for *N. crassa* endo-exonuclease and for rabbit antisera against it, D. A. Clayton for plasmid pMR718B, S. St.-Pierre for oligopeptide synthesis, T. Moss for revision of the manuscript, S. Rioux for pBT29, S. Descôteaux for DNA sequencing, E. Belzile for cell fractionation and marker determinations, C. Chamberland for confocal microscopy, M. Lambert for oligonucleotide synthesis, and G. Langlois and P. Paquin for photographic reproduction. Supported by the Medical Research Council of Canada (doctoral fellowship to J.C. and scholarship to A.R.-C.) and by a grant from the National Cancer Institute of Canada.

25 February 1993; accepted 25 May 1993

Selective and ATP-Dependent Translocation of Peptides by the MHC-Encoded Transporter

Jacques J. Neefjes, Frank Momburg, Günter J. Hämmerling*

Major histocompatibility complex (MHC) class I molecules present peptides derived from nuclear and cytosolic proteins to $CD8^+$ T cells. These peptides are translocated into the lumen of the endoplasmic reticulum (ER) to associate with class I molecules. Two MHC-encoded putative transporter proteins, TAP1 and TAP2, are required for efficient assembly of class I molecules and presentation of endogenous peptides. Expression of TAP1 and TAP2 in a mutant cell line resulted in the delivery of an 11–amino acid oligomer model peptide to the ER. Peptide translocation depended on the sequence of the peptide, was adenosine triphosphate (ATP)–dependent, required ATP hydrolysis, and was inhibited in a concentration-dependent manner.

Cytotoxic T lymphocytes recognize peptides presented by MHC class I molecules. It is likely that two steps precede the binding by class I molecules of peptides originating from cytosolic or nuclear antigens. First, proteins undergo a limited degradation in which the multicatalytic or proteasome complex may be involved (*1*). Second, the resulting peptides are translocated from the cytosol into the lumen of the ER. Analysis of mutant cell lines with defective class I assembly and antigen presentation led to the identification of two related genes, located in the MHC class II region of humans and rodents, that encode the transporters associated with antigen processing, TAP1 and TAP2. These proteins form heterodimers (*2*, *3*) and are expressed in the ER membrane (*4*). By structural homology they belong to a family of transporter proteins that have multiple membrane-spanning sequences and contain an ATP-binding consensus sequence (*5*). Other members are the mammalian P-glycoproteins (multidrug resistance pumps), the cystic fibrosis transmembrane conductance regulator (CFTR), bacterial hemolysin transporters (HLyB), and the yeast protein Ste6.

Two lines of evidence suggested that the TAP1-TAP2 heterodimer translocates peptides to the ER: The reconstitution of class I assembly (*2*, *3*, *6*–*9*) in mutant cells transfected with TAP genes and the observation that expression of different alleles of

J. J. Neefjes, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. F. Momburg and G. J. Hämmerling, Tumor Immunology Program, German Cancer Research Center, Im Neuenheimer Feld 280, 6900 Heidelberg, Germany.

*To whom correspondence should be addressed.