
Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA

Mark W. Walberg and David A. Clayton

Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA

Received 3 August 1981

ABSTRACT

The sequences of the displacement-loop (D-loop) regions of mitochondrial DNA (mtDNA) from mouse L cells and human KB cells have been determined and provide physical maps to aid in the identification of sequences involved in the regulation of replication and expression of mammalian mtDNA. Both D-loop regions are bounded by the genes for tRNA^{Phe} and tRNA^{Pro}. This region contains the most highly divergent sequences in mtDNA with the exceptions of three small conserved sequence blocks near the 5' ends of D-loop strands, a 225 nucleotide conserved sequence block in the center of the D-loop strand template region, and a short sequence associated with the 3' ends of D-loop strands. A sequence similar to that associated with the 3' termini of D-loop strands overlaps one of the conserved sequence blocks near the 5' ends of D-loop strands. The large, central conserved sequence probably does not code for a protein since no open reading frames are discretely conserved. Numerous symmetric sequences and potential secondary structures exist in these sequences, but none appear to be clearly conserved between species.

INTRODUCTION

The mammalian mitochondrial genome has a unique asynchronous mode of replication in which the two DNA strands have physically distinct origins of replication (1). Replication begins by initiation of heavy strand (H-strand) synthesis at the H-strand origin (O_H). Light strand (L-strand) synthesis does not begin until the nascent H-strand has been elongated to the origin of L-strand synthesis (O_L), two-thirds of the way around the circular DNA genome. O_L has been well characterized in mouse and human mtDNA and provides a clear example of correlation of structure and function. 5' ends of nascent L-strands isolated from replicative intermediates map within the stem of a large dyad symmetry which is highly conserved between the two species (2).

In contrast to the high degree of conservation of sequence, structure, and function of O_L , the region around O_H contains the most highly divergent sequences in mtDNA, as first noted by Upholt and Dawid (3). The O_H region contains the D-loop which consists of three DNA strands (4). The short D-loop strand, varying between approximately 520-700 nucleotides in length, is complementary to the L-strand, thus

displacing the homologous region of H-strand. Multiple, discrete classes of D-loop strands exist which vary in length. The nature of the heterogeneity of classes of D-loop strands is different in different species (5,6).

In addition to a site of initiation of H-strand DNA synthesis, the D-loop region is thought to contain a site of initiation of H-strand transcription required to promote expression of the tightly packed rRNA and associated tRNA genes (7,8) and, possibly, all of the H-strand genes. A promoter of L-strand transcription (9) and a membrane attachment site (10) may exist in the D-loop region. In order to provide a physical map to aid in the identification of functionally important sequences and structures in the D-loop region of mtDNA, the nucleotide sequence of this region has been determined for mtDNA from mouse LA9 cells and human KB cells. These sequences are compared with each other as well as with D-loop sequences of mtDNA from human placenta (11) and rat (12).

MATERIALS AND METHODS

Isolation of mtDNA

mtDNA from KB and LA9 cells was isolated as described by Bogenhagen and Clayton (13). The region of the mouse sequence from 437 to 832 is that of mtDNA from LA9 cells. The remainder of the mouse sequence is that of pHRB, which is the largest HindIII/EcoRI fragment inserted into pBR322. Cloned mtDNA was propagated and isolated under P1-EK1 conditions. The KB sequence is that of mtDNA isolated from KB cells.

DNA sequence determination and analysis

Nucleotide sequence determinations were performed by the method of Maxam and Gilbert (14). DNA sequences were compared with each other and were analyzed for repeated sequences, direct symmetries, dyad symmetries and open reading frames using the Molgen Program at the SUMEX-AIM computer facility at Stanford.

RESULTS AND DISCUSSION

The complete nucleotide sequences of the D-loop regions of mtDNA from mouse LA9 cells and human KB cells are presented and provide physical maps on which the DNA sequence components of the molecular events important in the regulation of mtDNA replication and gene expression can be identified and localized (Figures 1,2). Both D-loop regions, which are the largest segments of mtDNA which are not transcribed into RNA, are bounded by the gene for tRNA^{Phe}, encoded on the H-strand adjacent to the 12S rRNA gene, and by the gene for tRNA^{Pro}, encoded on the L-strand adjacent to the tRNA^{Thr} gene. This arrangement is conserved in human placental mtDNA (11) and in the available portion of the rat mtDNA sequence (12). A comparison of the

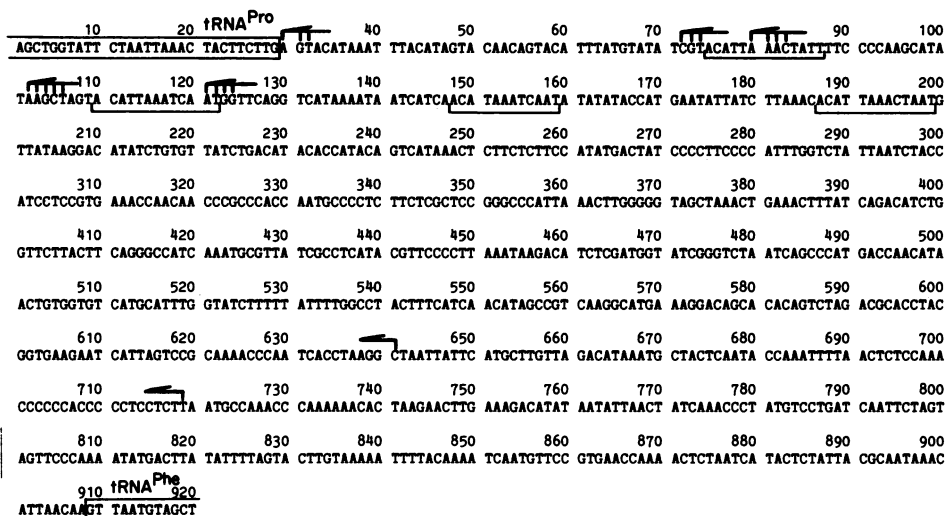


Figure 1. Nucleotide Sequence of the D-loop Region of Mouse L cell mtDNA.

The sequence shown is L-strand, 5'→3'. tRNA genes are boxed. The predominant 5' end of D-loop strands is indicated by the arrow at position 641. The 5' end of the longest D-loop strand is indicated by the arrow at position 719. 3' ends of D-loop strands are indicated by the five arrows between nucleotides 30-125. The lines extending below these arrows indicate the degree of heterogeneity which has been detected for each class of 3' ends. The D-loop strand termination associated sequences (16) are underlined.

D-loop region sequences from mouse, human and rat mtDNAs has allowed identification of a number of blocks of conserved primary sequence of various lengths (Figure 3). These will be discussed below. The regions of greatest divergence of primary sequence occur at both ends of the D-loop region, as has been noted in electron microscopic studies of heteroduplexes of sheep and goat mtDNAs (3).

The heterogeneity of classes of D-loop strands is portrayed in Figure 3. Six classes of D-loop strands have been identified in mouse mtDNA which differ primarily at their 3' ends (5) but also at the 5' end in the less abundant, longest D-loop strand (15). The four classes of D-loop strands in human mtDNA share a common 3' end, but differ in length at their 5' ends (6). The termini of D-loop strands have been mapped at the nucleotide level and are shown in Figures 1 and 2 (2,15,16). The lengths of D-loop strands in LA9 and KB cells are listed in Table I.

The region between the tRNA^{Phe} gene and O_H

The region between the tRNA^{Phe} gene and the 5' ends of major D-loop strands varies in length (267 nucleotides (nt) in mouse, 240 nt in rat, and 385 nt in human

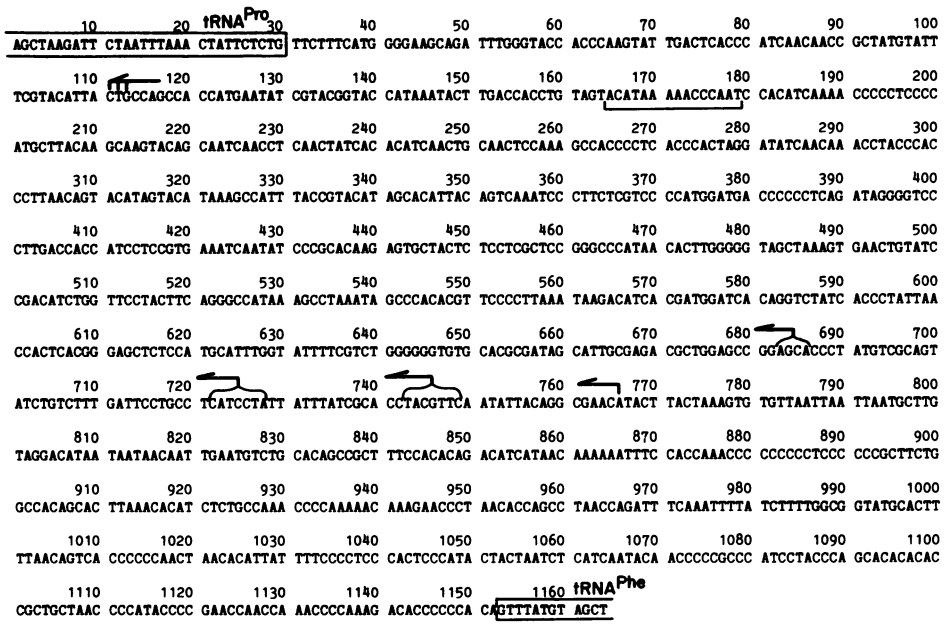


Figure 2. Nucleotide Sequence of the D-loop Region of Human KB Cell mtDNA. The sequence shown is L-strand 5' → 3'. tRNA genes are boxed. 5' ends of D-loop strands are indicated by the arrow between positions 766-683. The brackets indicate the uncertainty, as well as the heterogeneity of map positions of 5' ends (2). 3' ends of D-loop strands are indicated at nucleotides 111-113. The D-loop strand termination associated sequence is underlined. There are 11 differences between the KB sequence and the human placenta sequence (11). They are at positions 131 and 525, C → T; position 284, T → C; positions 648 and 838, G → A (KB → placenta). KB nucleotides 885, 886 and 888 are deleted in the placenta sequence. Placenta nucleotides 525, 523 and 16221 (11) are deleted in the KB sequence. KB nucleotide 828 is deleted in the limited HeLa mtDNA sequence (17).

mtDNA) and in sequence. The mouse and rat sequences are more similar to each other, being about 69% homologous, than either is to the human sequence in this region. The level of conservation of sequences in the three mtDNAs falls greatly, to below 35% for mouse versus human, immediately outside of the tRNA^{Phe} gene. There are three blocks of conserved sequence near the 5' ends of D-loop strands (Figure 4). Conserved sequence block 1 (CSB-1) is just upstream from the map positions of the major D-loop strands in each sequence. The positions of these D-loop strand 5' ends relative to CSB-1 varies by 20 nucleotides between species. A large dyad symmetry has been noted in HeLa cell mtDNA (17) in a region which overlaps CSB-1 and which contains the template sequence of the 5' end of the longest human D-loop strand shown

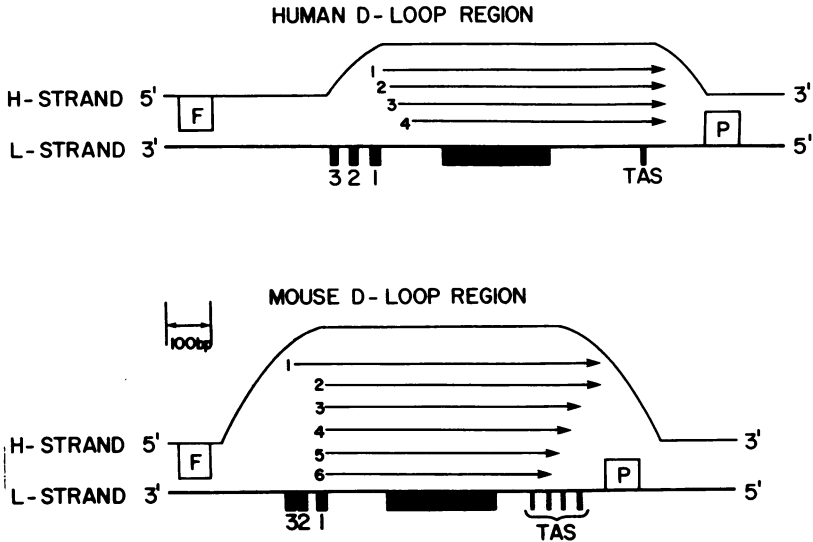


Figure 3. The D-loop Regions of Mouse and Human mtDNA.

Both the L-strands and H-strands are shown. tRNA genes are indicated by boxes on the coding strand. F = Phe, P = Pro. D-loop strands are represented by arrows which indicate both their relative map positions and their 5' → 3' polarity. The numbering of D-loop strands corresponds to that in Table I. The solid blocks below the L-strands represent conserved sequence blocks (CSBs). CSBs 1-3 are those shown in Figure 4. Those labeled TAS represent the repeated sequence associated with termination of D-loop strand synthesis (16). The largest CSB, in the middle of the D-loop region, corresponds to nucleotides 298-524 in the mouse sequence and 408-632 in the human sequence.

Table I. Map Positions of Human and Mouse D-loop Strands.

	D-Loop Strand	Length	Map positions	
			5' ends	3' ends
Mouse	1	688-690	719	30-32
	2	610-612	641	30-32
	3	568-570	641	72-74
	4	559-562	641	80-83
	5	540-536	641	102-106
	6	521-518	641	121-124
Human	1	654-656	766	111-113
	2	634-636+3	746+3	111-113
	3	613-615+3	725+3	111-113
	4	571-576	683 or 686	111-113

D-loop strands are numbered as in Figure 3. Map positions are numbered as in Figures 1 and 2 for mouse and human, respectively. Heterogeneity of length or terminal map position is indicated as a range. Uncertainty is indicated as ±.

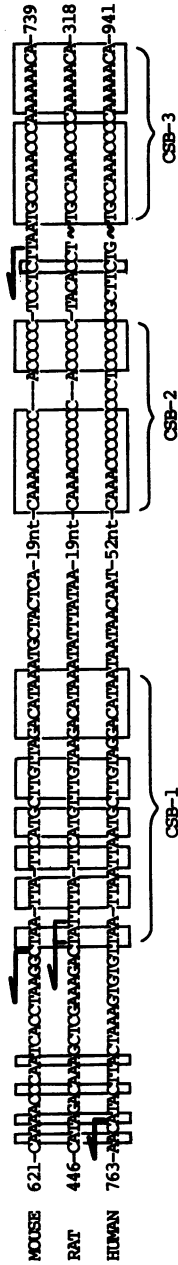


Figure 4. Conserved Sequence Blocks Near the 5' ends of D-loop Strands.

The sequences are L-strand 5' → 3'. Numbering corresponds to that in Figures 1 and 2 and Sekiya et al. (12) for mouse, KB cell and rat mtDNA, respectively. Template map positions of 5' ends of D-loop strands are indicated by arrows. Nucleotides conserved in all three sequences are boxed. The extent of conserved sequence blocks are indicated by brackets. ~ represents 9 nucleotides in the rat sequence and 23 nucleotides in the KB cell sequence. A dash (-) indicates the absence of a nucleotide at that position.

in Figure 3. A dyad symmetry is present in mouse mtDNA in a similar location, also overlapping CSB-1. It is, however, different in both sequence and structure from that present in human mtDNA. Although 5' ends of D-loop strands have been mapped within both the mouse and human hairpin structures, they do not map in similar locations within the symmetries. No clearly analogous structure can be drawn for rat mtDNA. No potential secondary structures in this region are as clearly conserved as is the dyad symmetry present at O_L in mouse and human mtDNA (2).

Overlapping CSB-1 in each species is a sequence remarkably similar to a DNA sequence which has been implicated in the termination of D-loop strand synthesis in mouse and human mtDNA (16) (Figure 5). The position of these sequences just upstream from the position of 5' ends of nascent H-strands (Figure 4) suggests that they may be involved in the termination of nearly completed H-strand displacement synthesis during replication of mtDNA. It is also possible that a specific stop point for H-strand synthesis is required by the as yet unknown mechanism of separation of daughter molecules near the end of the replication cycle.

CSB-2, a 16-19 nucleotide sequence which is $\geq 75\%C$, is located just downstream from the map position of the 5' end of the longest D-loop strand in mouse mtDNA. The 5' end of this D-loop strand contains exclusively ribonucleotides, suggesting that it may be a primary product of initiation of H-strand DNA replication (15). The location of CSB-2 relative to the 5' end of this D-loop strand suggests that CSB-2 may play some role in the initiation of H-strand mtDNA replication. D-loop strands mapping in analogous positions near CSB-2 in human and rat mtDNA have not been reported.

CSB-3, an 18 nucleotide sequence which is conserved at a level of $>94\%$, is located just upstream from the map position of the 5' end of the longest D-loop strand in mouse mtDNA. As for CSB-2, the location of CSB-3 relative to the 5' end of this D-loop strand suggests that it may play some role in the initiation of H-strand mtDNA



Figure 5. Homology of Termination Associated Sequences to Sequences Overlapping CSB-1.

Sequences are L-strand, 5' → 3'. Numbering is as in Figure 4. Sequences associated with the termination of D-loop strand synthesis are shown at the top of each column. Similar sequences overlapping CSB-1 are shown below. Conserved nucleotides are boxed.

replication.

A very large hairpin structure can be drawn which includes CSB-1, involving nucleotides 719-820 in mouse, 927-995 in KB, and 248-338 (using the numbering in Sekiya *et al.* (12)) in rat mtDNA. The length, sequence and potential secondary structure differ significantly between species. The significance of such a large hairpin structure of questionable stability is unclear. Since the structures involve many G-T pairs, they are not true dyad symmetries (which would be likely candidates for recognition sites for the initiation of a bidirectional process, such as transcription).

Attardi *et al.* (18) have presented data from HeLa cells suggesting that a highly abundant 200-220 nt RNA is transcribed from the L-strand in the region between the 5' end of D-loop strands and the gene for tRNA^{Phe}. An exhaustive search by both S1 nuclease protection and RNA blotting experiments has failed to demonstrate the presence of any RNA encoded in this region of length greater than 50 nucleotides in mitochondria of LA9 or KB cells (in log or stationary growth phase), or in mitochondria of leukocytes from a human with chronic myelogenous leukemia (19). A comparison of the sequence thought to code for this small RNA in HeLa cells (18) with the sequences presented here and with the rat sequence (12) indicates that much of this sequence has been deleted in the rodent mtDNA, but not in KB cell mtDNA. No RNAs have been mapped in the remainder of the D-loop region in LA9 or KB cells. Thus, there appear to be no stable RNAs encoded in the D-loop region between the genes for tRNA^{Phe} and tRNA^{Pro} in these cells.

The region encoding D-loop strands

The sequences encoding 5' ends of the major D-loop strands are quite divergent between species. In human mtDNA, where there are multiple D-loop strand 5' ends, no common 5' sequence has been identified. The 5' ends of nascent H-strands isolated from replicative intermediates are identical to the 5' ends of D-loop strands in KB cells (2) and in mouse cells (D.P. Tapper and D.A. Clayton, unpublished results). It has not been demonstrated whether the various 5' ends of D-loop strands result from different points of initiation of H-strand synthesis or whether they are processed from longer D-loop strands. Predominant sites of ribosubstitution in mouse mtDNA have been identified (20); however, there is no discrete correlation of sites of ribosubstitution with the positions of D-loop strand 5' ends, with the positions of ribonucleotides found at the 5' ends of D-loop strands (15), or with any of the CSBs.

There is an ~225 nt block of sequence in the center of the D-loop strand coding region which is conserved at a level of >65% (Figure 3). This degree of conservation of sequence is as high as that for the rRNA and structural genes of mouse and human mtDNA, thus suggesting that this large block of sequence plays an important role in expression or maintenance of the mammalian mitochondrial genome. It is unlikely that

this conserved region is a protein coding sequence. Although short open reading frames exist in this region, few are conserved between species due to numerous insertions and deletions of one or two nucleotides. Also, no RNA has been shown to be encoded by this region of mtDNA. 225 nt is far more than would be statistically required to identify a unique protein binding site in mtDNA. No protein-DNA interaction is known to involve such a large number of nucleotides in a sequence specific manner. Thus, the role of this large conserved sequence remains obscure. This region contains no remarkable symmetries which are clearly conserved between species.

The region encoding 3' ends of D-loop strands

The region between the highly conserved block of sequence in the middle of the D-loop region and the tRNA^{Pro} gene is highly divergent between mouse and human mtDNA both in length and in sequence. A striking exception is the presence of four repeats of a 13 nucleotide sequence in mouse mtDNA and a single copy of a very similar sequence in KB cell mtDNA (Figures 1-3). Because of a correspondence of the locations of these sequences with the map positions of 3' ends of the multiple mouse D-loop strands and the common 3' end of KB cell D-loop strands, they have been implicated in the termination of D-loop strand synthesis (16). Exact copies of this repeated termination associated sequence do not occur at any site on either strand of human placental mtDNA (11) or mouse mtDNA (21) except for the copies which are located near the 3' ends of D-loop strands. The mechanism of action of this signal is unknown. Symmetric sequences exist in this region, including a large hairpin predicted to be relatively stable ($\Delta G \cong -18\text{Kcal}$) in the KB sequence, but none are well conserved between species nor is there a clear relationship of symmetries to 3' ends of D-loop strands.

The role of the D-loop

The disparity between species of both the sequence and number of nucleotides in the D-loop regions of mtDNAs is, with the exceptions noted in Figure 3, markedly higher than that in the remainder of the genome. This is somewhat surprising in light of the highly conserved nature of the remainder of the genome. This disparity raises the possibility that, in addition to containing recognition elements for sequence directed events, the D-loop region of mtDNA may be important in a process which involves higher order structure, presumably that of the D-loop itself. The functional roles of the D-loop have not been clearly identified. Although it has often been assumed that D-loop strands are intimately involved in the replication of mtDNA, there is no compelling evidence that they are used directly as primers to be extended into nascent H-strands. Both the high frequency of turnover of D-loop strands relative to frequency of DNA replication (22) and the maintenance of multiple classes of D-loop strands, even though the nature of the heterogeneity is vastly different in different species,

suggest that the function of the D-loop is more complex than would be expected if D-loop strands acted solely as replication primers. Although it is possible that D-loop strands do act as primers, it is likely that the D-loop plays a role in other important functions related to replication or expression of mtDNA, such as regulation of initiation of DNA replication or transcription, membrane attachment, or some other process yet to be identified.

The presence of blocks of conserved sequence within and around the D-loop (Figure 3) suggest that sequence directed events occur in the D-loop region. Genetic studies and in vitro systems will be required to identify the functionally important components of each of these sequences.

ACKNOWLEDGEMENTS

We thank J.N. Doda and C.T. Wright for technical assistance in DNA sequencing and D.P. Tapper and R.A. Van Etten for helpful discussions. We acknowledge the Stanford Molgen Project at the NIH SUMEX-AIM Facility for use of facilities used in DNA sequence analysis. This investigation was supported by grants CA-12312-II from the National Cancer Institute and NP-9G from the American Cancer Society, Inc. M.W.W. is a Medical Scientist Training Program trainee (GM-07365) and D.A.C. is a Faculty Research Awardee (FRA-136) of the American Cancer Society, Inc.

REFERENCES

1. Berk, A.J. and Clayton, D.A. (1974) *J. Mol. Biol.* **86**, 801-824
2. Tapper, D.P. and Clayton, D.A. (1981) *J. Biol. Chem.* **256**, 5109-5115
3. Upholt, W.B. and Dawid, I.B. (1977) *Cell* **11**, 571-583
4. Kasamatsu, H., Robberson, D.L. and Vinograd, J. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2252-2257
5. Gillum, A.M. and Clayton, D.A. (1978) *Proc. Nat. Acad. Sci. USA* **75**, 677-681
6. Brown, W.M., Shine, J. and Goodman, H.M. (1978) *Proc. Nat. Acad. Sci. USA* **75**, 735-739
7. Van Etten, R.A., Walberg, M.W. and Clayton, D.A. (1980) *Cell* **22**, 157-170
8. Eperon, I.C., Anderson, S. and Nierlich, D.P. (1980) *Nature* **286**, 460-467
9. Cantatore, P. and Attardi, G. (1980) *Nucleic Acids Res.* **8**, 2605-2624
10. Albring, M., Griffith, J. and Attardi, G. (1977) *Proc. Nat. Acad. Sci. USA* **74**, 1348-1352
11. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) *Nature* **290**, 457-465
12. Sekiya, J., Kobayashi, M., Seki, T. and Koike, K. (1980) *Gene* **11**, 53-62
13. Bogenhagen, D. and Clayton, D.A. (1974) *J. Biol. Chem.* **249**, 7991-7995
14. Maxam, A.M. and Gilbert, W. (1980) in *Methods in Enzymology*, Grossman, L. and Moldave, K. Eds., Vol. **65**, pp. 449-560, Academic Press, New York.
15. Gillum, A.M. and Clayton, D.A. (1979) *J. Mol. Biol.* **135**, 353-368
16. Doda, J.N., Wright, C.T. and Clayton, D.A. (1981) *Proc. Nat. Acad. Sci. USA* **78**, in press
17. Crews, S., Ojala, D., Posakony, J., Nishiguichi, J. and Attardi, G. (1979) *Nature* **277**, 192-198

18. Attardi, G., Cantatore, P., Ching, E., Crews, S., Gelfand, R., Merkel, C., Montoya, J. and Ojala, D. (1980) in *The Organization and Expression of the Mitochondrial Genome*, Kroon, A.M. and Saccone, C. Eds., pp. 103-119, North-Holland, Amsterdam.
19. Van Etten, R.A., Michael, N.L., Bibb, M.J., Brennicke, A. and Clayton, D.A. (1981) in *Mitochondrial Genes*, Attardi, G., Borst, P. and Slonimski, P. Eds., Cold Spring Harbor, New York, in press
20. Brennicke, A. and Clayton D.A. (1981) *J. Biol. Chem.* **256**, in press
21. Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) *Cell* **26**, in press
22. Bogenhagen, D. and Clayton, D.A. (1978) *J. Mol. Biol.* **119**, 49-68