REVIEW ARTICLE

The molecular biology of the mitochondrion

John ROSAMOND*

Microbiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

The 15 years since the identification of a genetic system in mitochondria have seen the publication of a wealth of information on mitochondrial genome structure, organization and expression. This has been due in part to the relative simplicity of mtDNA, which ranges in size from 1×10^7 – 7×10^7 daltons in different organisms (Borst, 1972). The larger genome size is more common in fungi and higher plants (Morimoto *et al.*, 1975; Quetier & Vedel, 1977), while mammalian mitochondria tend to have smaller genomes, and are thus the smallest DNAs to code for all three species of RNA. Nonetheless, mtDNA from these different organisms appears to code for similar numbers of proteins (Borst & Grivell, 1978; Tzagoloff *et al.*, 1979).

In addition, the application of modern techniques of DNA analysis and manipulation has been essential to the mapping of the mitochondrial genome (Sanders *et al.*, 1977; Morimoto *et al.*, 1977; Heyting *et al.*, 1979; Martin *et al.*, 1979; Grivell *et al.*, 1980), especially in view of the limitations of the more classical genetic mapping techniques (Dujon *et al.*, 1974). Undoubtedly though, the single most important development has been the application of DNA sequencing techniques to elucidate the nucleotide sequence of parts of the yeast mt genome (Dujon, 1980; Lazowska *et al.*, 1980; Tzagoloff *et al.*, 1980; Bonitz & Tzagoloff, 1980; Nobrega & Tzagoloff, 1980; Bonitz *et al.*, 1980b; Thalenfeld & Tzagoloff, 1980; Macino & Tzagoloff, 1980), and the whole of the human mt genome (Anderson *et al.*, 1981). Genetic maps of the genomes of yeast and mammalian mitochondria are shown in Fig. 1.

Quite apart from their contribution to our understanding of mitochondrial biogenesis, the nucleotide sequence data have served to highlight several unusual and unique features in the organization and expression of both yeast and mammalian mtDNA. Thus, for example, mitochondria exhibit variations on what was previously thought to be a universal

genetic code, process their primary RNA transcripts in novel fashion, and use less than the predicted minimum number of tRNAs for translation. This review will concentrate on some of these exceptional features of mitochondrial molecular biology, especially those that have become apparent from DNA sequencing or the use of related techniques. More comprehensive accounts can be found in other recent reviews and symposia proceedings (Saccone & Kroon, 1976; Linnane & Nagley, 1978; Borst & Grivell, 1978; Tzagoloff *et al.*, 1979; Locker & Rabinowitz, 1979; Bernardi, 1979; Kroon & Saccone, 1980; Attardi, 1981a,b; Dujon, 1981).

Mitochondrial DNA replication

Replication of the mammalian mt genome begins at a unique origin (Crews *et al.*, 1979; Gillum & Clayton, 1979), with the synthesis of about 680 nucleotides of H-strand DNA (7S DNA). This forms a displacement loop (D-loop) by base-pairing to the parental L-strand (Kasamatsu *et al.*, 1971). Replication continues by extension of some of the 7S DNA molecules, thus extending the D-loop. By determining the sequence of the 5'-proximal portion of 7S DNA, Crews *et al.* (1979) have been able to identify precisely the site at which HeLa mtDNA synthesis is initiated, although it is uncertain whether 7S DNA is preceded by a 5'-ribonucleotide primer (Crews *et al.*, 1979; Gillum & Clayton, 1978). Synthesis of the complementary L-strand also begins at a unique point some two-thirds of the genome length from the H-strand origin, and probably requires that H-strand synthesis has exposed the origin before L-strand synthesis can occur (Berk & Clayton, 1974; Martens & Clayton, 1979). Synthesis continues asymmetrically until two complete DNA molecules are formed.

The mechanism of yeast mtDNA replication is less clear, although structures have been observed that are consistent with a D-loop mechanism (Locker *et al.*, 1974). Mapping and hybridization studies have shown that the yeast mt genome contains at least seven putative origins of replication (*ori* sequences), which shows a remarkable degree of homology (Banc & Dujon, 1980; de Zamaroczy

Abbreviation used; mt, mitochondria(l).

* Present address: Department of Biochemistry, University of Manchester, Oxford Road, Manchester M13 9PL, U.K.

Fig. 1. Organization of human (a) and yeast (b) mitochondrial genomes

(a) Human mt genome, based on Anderson *et al.* (1981). Sequences coding for known gene products are shown in black, tRNA genes are dotted, while unassigned reading frames (URF) are cross-hatched. Transcription of the 16 569-base-pair genome is clockwise from a promoter near the replication origin, except for those few genes which are transcribed from the opposite strand, and which are identified by a small arrow (Attardi *et al.*, 1980b). (b) Yeast mtDNA. Assignments are as above, while the double-hatched area represents the *var-1* determinant, and open boxes represent either non-coding regions (within the COI and cytochrome *b* genes), or regions for which the DNA sequence is unknown (elsewhere).

et al., 1981). Moreover, both yeast and mammalian mt *ori* sequences can be folded in a similar manner, to form the (theoretically) stable hairpin structures (Crews *et al.*, 1979; Gillum & Clayton, 1979; de Zamaroczy *et al.*, 1981), which are frequently found at replication origins, but whose involvement in the initiation of DNA synthesis *in vivo* is still a matter for conjecture.

As with many other aspects of mitochondrial molecular biology, the reason for the novel mechanism, and the purpose of 7S DNA, is not obvious. Crews *et al.* (1979) have suggested that 7S DNA might be involved in the interaction between the mitochondrial membrane and the mt genome at the replication origin (Albring *et al.*, 1977). Another possibility is that it may be involved in the regulation of mtDNA synthesis, acting in a manner analogous to the mechanism of attenuation of the *trp* operon in *Escherichia coli* (Yanofsky, 1981). Alternatively, since 7S DNA is unstable and only a small number of these molecules serve as precursors for further DNA synthesis (Bogenhagen & Clayton, 1978), 7S DNA may have a function other than in DNA replication. An interesting possibility is that since transcription of human mtDNA occurs from a promoter near to the origin of replication (Aloni & Attardi, 1971a; Cantatore & Attardi, 1980), 7S DNA might be necessary to facilitate transcription by exposing the parent H-strand.

Organization of the yeast mitochondrial genome

Yeast mtDNA shows an unusual bias in base composition, being 82% AT base pairs, although four domains of distinctive nucleotide composition can be distinguished (Prunell & Bernardi, 1974, 1977; Bernardi *et al.*, 1976; Prunell *et al.*, 1977). About 50% of the total genome consists of AT-rich segments, with a GC content of less than 5%. Moreover, these GC base pairs are clustered within the AT-rich segments, and can themselves be assigned to one of two categories, depending on whether or not they are susceptible to the restriction enzymes *HpaII* (C-C-G-G) and *HaeIII* (G-G-C-C). The remaining 50% of the genome has a variable, moderate GC content (26% average).

The maintenance of these distinct regions of unusual base sequence within the yeast mt genome requires some sort of explanation for both their function and their origin. One possible function, suggested by Prunell & Bernardi (1977), is that the repeated sequences within the GC clusters and AT spacers provide regions of homology for illegitimate, site-specific recombination, giving rise to the observed polymorphism of the mitochondrial genome as a result of duplication, inversion, translocation and deletion. In a more extreme case, DNA sequence analysis has shown that the rearrangements of the

genome that occur during the formation of the petite mutation follow the excision of a small fragment of the genome, and that this excision occurs preferentially at a repeated site within the CG clusters or AT spacers (Faugeron-Fonty *et al.*, 1979; Baldacci *et al.*, 1980; Gaillard *et al.*, 1980). Whether this is indeed a function of the intergenic sequences is a matter for speculation. An alternative would be that the relative instability of yeast mtDNA which leads to the high rate of formation of petite mutants is not a function but a consequence of maintaining the AT-rich regions. This would in turn argue for strong selective pressures for their maintenance, and fundamentally important functions for these regions. One possibility then would be the regulation of mtDNA expression, either by interaction with sequence-specific DNA-binding proteins acting at the level of replication or transcription (Prunell & Bernardi, 1977), or in the maturation of precursor mRNA (Borst *et al.*, 1977). An intriguing possibility is suggested by observations on the control of gene expression by regions distant from the coding sequences, particularly the possibility that supercoiling and superhelix density may be important in the regulation of transcription (reviewed in Smith, 1981). The AT spacers may then influence mitochondrial gene expression by an effect on DNA structure, rather than by a direct effect on the processes themselves. Domains of high AT content are not found in the mitochondrial genomes of other organisms though (see below), so that the function of these regions and the selective pressures that encourage their retention may be a specific consequence of an aspect of the metabolism of *Saccharomyces cerevisiae*, and the relationship between the mitochondrion and the nucleus in this organism, rather than being a general property of mitochondria.

Mosaic genes in yeast mitochondria

Recent work on the analysis of gene organization in eukaryotes has established that, typically, the DNA sequences that specify the amino-acid sequence of the protein product (exons) are interrupted by other base sequences (introns). The origin and function of introns remains uncertain, but the observation that they occur in the yeast mt genome has led to an extensive genetic and physical analysis of the yeast mt genes that contain introns, which in turn has suggested one possible function for the 'non-coding' regions.

The most extensive analysis has been done on the *COB* gene, which codes for apocytochrome *b*, a hydrophobic protein which is a component of the ubiquinone-cytochrome *c* reductase complex. The original suggestion that the *COB* gene might contain introns came from genetic studies which showed the

following. (a) Mutations which affected the synthesis of cytochrome *b* spanned a region of the genome that was five times larger than ought to have been necessary to code for a protein of this size. Moreover, the mutations were highly clustered, and some were pleiotropic, affecting the synthesis of the *OXI3* gene product (Slonimski *et al.*, 1978; Alexander *et al.*, 1979). (b) The smallest transcript of this region hybridizes to non-contiguous segments of the DNA (Church *et al.*, 1979; Grivell *et al.*, 1979). (c) Some mutants produced a smaller polypeptide, presumably by premature chain termination. Other mutants, interspersed with those of the first class, produced polypeptides up to twice the size of cytochrome *b* but which nonetheless cross-reacted with antibody to cytochrome *b*, suggesting that translation was now occurring in regions that were not normally translated (Haid *et al.*, 1979; Kreike *et al.*, 1979; Solioz & Schatz, 1979). More direct evidence for the presence of introns in *COB* has been obtained by electron microscopy of DNA-RNA hybrids (Grivell *et al.*, 1979).

Complementation testing of the non-conditional mutants established two classes: those which failed to complement, exemplified by mutations *box-1*, -2, -4, -5, -6, -8, -9, and which are assumed to contain mutations in exons (although this may not be the case for *box-2* and *box-9*); and mutants in *box-3*, -7 and -10 which complement both each other and mutants of the first class (Slonimski *et al.*, 1978), and which are interspersed amongst the other mutants (Fig. 2a). Moreover, *box-3*, -7 and -10 mutants are pleiotropic, and block expression of the *OXI3* gene. Thus, these mutations are recessive but affect some function that normally acts *in trans* on both the *COB* and *OXI3* genes, one possibility of course being that the introns in the cytochrome *b* gene were needed for RNA splicing (Kreike *et al.*, 1979; Solioz & Schatz, 1979; Church *et al.*, 1979).

The physical structure of the *COB* gene is shown in Fig. 2; surprisingly, the gene structure is strain-dependent and exists in at least two forms (van Ommen *et al.*, 1979; Nobrega & Tzagoloff, 1980; Dhawale *et al.*, 1981). In its long form the gene contains six exons (B_1 - B_6) and five introns (I_1 - I_5), while in its short form it contains three exons and two introns (Figs. 2b and 2c). The short form of the gene is related to the long form, and can be derived by removal of I_1 - I_3 and fusion of B_1 - B_4 (van Ommen *et al.*, 1979; Nobrega & Tzagoloff, 1980; Borst & Grivell, 1981). In addition, the long form of the gene contains at least two long open reading frames which are continuous with the preceding exon. One of these extends from the terminal region of I_1 , through the 14-base-pair B_2 (and in register with it), and for about 840 base-pairs into I_2 ; this reading frame is characterized by the *box-3* mutation (Van Ommen *et al.*, 1979). A second open

Fig. 2. Organization and expression of the *COB* gene in yeast mitochondria

(a) Location of the *box* mutation clusters in relation to the exon (shaded boxes) and intron (thin line) components of the gene, based on Church *et al.* (1979), Jacq *et al.* (1981) and Mahler *et al.* (1981). (b) Short form of the gene with three exons (shaded) and two introns (unshaded). (c) The long form of the gene with six exons (shaded) and five introns (unshaded). Open reading frames within the introns are cross-hatched. (d)–(f) Proposed scheme for the synthesis of mature apocytochrome *b* mRNA from primary transcript with removal of intron *I*₁ by a nuclear-coded gene product (c)–(d), removal of *I*₂ by *box-3* RNA maturase (d)–(e) and subsequent removal of *I*₃, *I*₄ and *I*₅ by an undefined mechanism (e)–(f). Redrawn from Borst & Grivell (1981), with data from Church *et al.* (1979) and Lazowska *et al.* (1980).

reading frame extends from *B*₄ for 1158 nucleotides into *I*₄ and contains the *box-7* mutation (Nobrega & Tzagoloff, 1980; Jacq *et al.*, 1980).

A mechanism for the maturation of the primary transcript of the *COB* gene has been proposed by Lazowska *et al.* (1980), and is depicted schematically in Figs. 2c–2f. The first splicing reaction fuses *B*₂ (and *box-3*) to *B*₁, with the removal of *I*₁ as a circular 10S RNA molecule (Church *et al.*, 1979; van Ommen *et al.*, 1980). This step is dependent on nuclear coded enzyme(s), and produces an RNA molecule which serves as the mRNA for synthesis of a protein (called *box-3* RNA maturase) by translation of the open reading frame from *B*₁*B*₂ into the *box-3* region of *I*₂ (Fig. 2; Lazowska *et al.*, 1980). This protein contains the first 143 amino acids of cytochrome *b* at its *N*-terminus coupled to 280 amino acids from *I*₂, and is necessary for the second processing step, specifically, the removal of *I*₂ and fusion of *B*₃ to *B*₁*B*₂ (Figs. 2d and 2e; Lazowska *et al.*, 1980; Borst & Grivell, 1981). Subsequent, as yet undefined, processing events lead ultimately to mature cytochrome *b* mRNA by removal of *I*₃, *I*₄ and *I*₅ (Figs.

2e and 2f). Whether the removal of any of these last three introns requires intron-coded protein is uncertain, but a good candidate can be found in *I*₄, which has a long, open reading frame and which contains the *box-7* mutation (Nobrega & Tzagoloff, 1980; Jacq *et al.*, 1980). However, if a *box-7* RNA maturase does exist, it must differ from the *box-3* maturase since the two mutants complement each other (Lamouroux *et al.*, 1980).

These results raise several points. Firstly, it would seem that intron-coded proteins may act in maturation steps of more than one gene. For example, the *OXI3* gene, which codes for cytochrome oxidase subunit 1, is also a mosaic gene, comprising seven or eight exons, with four open reading frames amongst its introns (Bonitz *et al.*, 1980b). However, expression of *OXI3* is abolished by *box-7* and *box-3* mutations. The mechanism underlying this pleiotropic effect is not understood, but it seems likely that only *box-7* has a direct effect on *OXI3* expression, while the effect of *box-3* is a consequence of its effect on *box-7* (Dujon, 1981; Dhawale *et al.*, 1981). Furthermore, a mutation which phenotypically suppresses the *box-7* mutation

(*MIM2*) has been mapped in *OXI3* (Dujardin *et al.*, 1980). Lazowska *et al.* (1980) have speculated that *MIM2* affects a maturase coded by an intron in *OXI3*, such that the mutant protein is able to perform both its own normal functions as well as those of the *box-7* maturase.

The second important point concerns the function of the *box-3* RNA maturase, for there is no evidence as to whether it has splicing activity, or whether it serves as a 'co-maturase' by influencing the activity of other splicing enzymes. More important thought is the observation that *box-3* maturase controls its own synthesis by negative feedback, since it acts to degrade the RNA responsible for its synthesis (Fig. 2; Lazowska *et al.*, 1980). This property could also be applicable to a *box-7* maturase, and may be the key function of these mitochondrial introns in allowing yet another level of regulation in the expression of some mitochondrial genes. However, the function of *box-3* maturase is not essential to the synthesis of cytochrome *b*, since the short form of the *COB* gene, in which B_1 - B_4 are fused, makes cytochrome *b* without involving *box-3* maturase.

Organization of the mammalian mitochondrial genome

In contrast with the profligate nature of the yeast mt genome, which maintains large non-coding regions at the expense of genome instability, the mammalian mt genome is a model of economy and efficiency. By determining the base sequence of the entire 16 569-base-pair genome of HeLa cell mtDNA, Anderson *et al.* (1981) have shown how minimal the requirements for gene expression can be, and how markedly the mechanisms of gene expression differ between yeast and mammalian mitochondria.

HeLa mtDNA codes for 22 tRNA molecules, two rRNA species and at least 16 polyadenylated RNA molecules which are presumed to be mRNA and precursors (Attardi, 1981a). In contrast with the symmetry of transcription, though, which sees both strands of the DNA duplex almost completely transcribed from a single promoter near the origin of replication (Aloni & Attardi, 1971b; Murphy *et al.*, 1975; Cantatore & Attardi, 1980), the utilization of these primary transcripts shows a remarkable asymmetry. Pulse-labelling experiments have shown that L-strand transcripts are synthesized faster than are H-strand transcripts, but that the former have a much shorter half-life and so do not accumulate to any significant extent (Aloni & Attardi, 1971; Cantatore & Attardi, 1980). Consequently, it is not surprising to find that most of the functional RNA molecules are transcribed from the H-strand.

Analysis of the products of H-strand transcription by hybridization and S_1 -nuclease protection has

shown several general features. In contrast with yeast mtDNA, there is no evidence for intervening sequences within the regions of HeLa cell mtDNA coding for poly(A)-containing RNAs, nor is there any evidence for overlapping of the H-strand transcripts (Attardi *et al.*, 1980a, b). However, the most important point to emerge from these studies concerned the organization of the coding sequences in mammalian mtDNA. It had previously been shown by electron microscopy that the two rRNA coding sequences were flanked by 4S RNA genes, which were probably tRNA genes (Wu *et al.*, 1972; Angerer *et al.*, 1976). However, alignment of the positions of the tRNA genes derived from the mtDNA sequence has shown that this is a more general phenomenon, and that the sequences coding for the rRNA and poly(A)-containing RNA species on the H-strand are often immediately adjacent to tRNA molecules on both sides (Ojala *et al.*, 1980; Anderson *et al.*, 1981). This arrangement of RNA transcripts has led several groups to suggest a model for the transcription of the H-strand, the tRNA punctuation model, in which a single, large RNA transcript is processed by precise endonucleolytic cleavages before and after the tRNA molecules, to produce the mature RNA molecules or intermediates for further processing (Ojala *et al.*, 1980; Anderson *et al.*, 1981). Even though H-strand transcription is thought to start at a single promoter, and to cover almost all of the genome (Cantatore & Attardi, 1980), no large H-strand transcript has been found, suggesting that if the tRNA punctuation model is correct, processing occurs while the RNA is still associated with the transcription complex.

The mechanism by which the primary transcripts are processed is unknown. It has been suggested that the secondary structure of the tRNA molecules may be important, rather than any specific nucleotide sequence, in which case an enzyme or enzymes similar to *E. coli* RNAase P might be involved (Schmidt *et al.*, 1976; Anderson *et al.*, 1981; Ojala *et al.*, 1981).

It might be expected that, as a consequence of this mechanism, all RNA molecules should be present in equimolar amounts. However, estimates of the amounts of the various RNA species suggest that rRNA is present at concentrations at least 50 times higher than mRNA, and that this difference is accounted for by a differential rate of synthesis (Attardi *et al.*, 1980b; Gelfand & Attardi, 1981). Since transcription is unlikely to be regulated at initiation by any of the conventional means, the most likely means of obtaining differential transcription lies in regulating the processing of the primary transcript or in some form of premature transcription termination (transcription attenuation). This latter possibility is suggested by the high concentration of nascent transcripts from the rRNA

region of the genome (Brown *et al.*, 1978; Ojala *et al.*, 1981).

There are two major consequences of this extreme economy of the mammalian mt genome. The first became apparent after the sequence of the DNA and putative mRNA molecules had been determined (Anderson *et al.*, 1981; Ojala *et al.*, 1981). It was then found that many reading frames lack an intact termination codon and instead, the RNAs have a U or UA at the 3'-terminus, immediately before a tRNA gene or another reading frame. Anderson *et al.* (1981) have proposed that the termination codon is created as a result of polyadenylation which converts U and UA to UAA in phase with the reading frame. RNA sequencing suggests that this does indeed occur (Ojala *et al.*, 1981).

The second consequence is that the putative mRNA molecules lack any significant 5' leader regions (Attardi *et al.*, 1980a; Montoya *et al.*, 1981), while many have an initiation codon at the actual 5' terminus. This raises several points regarding the initiation of translation in mitochondria, notably the way in which mitochondrial ribosomes bind to the message, since mRNA is conventionally thought to contain a ribosome-binding site within the 5' leader (Shine & Dalgarno, 1974; Steitz & Jakes, 1975). Thus, accurate ribosome binding in the mitochondria is probably achieved by a combination of mRNA secondary structure, thereby masking any internal methionine codons, and some special features of the mitochondrial ribosomes. This latter may be related to the unusual nature of the mammalian mt rRNAs which, with 1559 and 954 nucleotides, are significantly smaller than most other rRNAs, while a 5S RNA as such is totally lacking (Anderson *et al.*, 1981).

The rRNAs though are not the only component of the mitochondrial machinery to differ from their cytoplasmic counterparts. The mitochondrial tRNA molecules have highly unusual structures, vary widely in size, lack many of the features common to cytoplasmic tRNAs from most organisms and probably have a more flexible tertiary structure, allowing novel interactions with the mitochondrial ribosome (Anderson *et al.*, 1981).

Perhaps the most startling observation though, has been that the mammalian mt genome codes for only 22 tRNA species and that translation can proceed quite satisfactorily with this number, even though it is less than the minimum predicted number. This is achieved by using a single tRNA to read those families of four codons which code for a single amino-acid. In mammalian mitochondria, the codons GCN which code for alanine are read by a single tRNA in which U is at the 5' position of the anticodon (Barrell *et al.*, 1980). This is attained either by an extension of the 'wobble' mechanism, allowing U-N base pairing, or else by a 'two-

out-of-three' interaction between the tRNA and the message (Crick, 1966; Lagerkvist, 1978). Furthermore, the genetic code used by mammalian mitochondria is different from that used in other systems, and previously regarded as universal. Thus, UGA codes for tryptophan and not 'stop', AUA for methionine not isoleucine, AGA and AGG code for termination and not arginine, while AUA codes for methionine and may act as an initiation codon (Barrell *et al.*, 1980; Anderson *et al.*, 1981). Mitochondria from other species seem to have their own variations on this theme, though, since yeast mitochondria also read UGA as tryptophan, but then read CUN as leucine rather than threonine, and do not share the unusual assignments of AUA, AGA and AGG (Bonitz *et al.*, 1980a).

Summary

It is clear that attempts to classify mitochondria as prokaryote- or eukaryote-like are inappropriate. For example, we cannot tell whether the decoding mechanisms in mitochondria represent a highly evolved form that has emerged in response to selective pressures on the mitochondrion, or whether it represents the remnant of a primitive system. Nonetheless, it may be significant that the basic principles on which the decoding mechanisms for protein synthesis are based appear to be the same in mitochondria of different organisms (Barrell *et al.*, 1980; Bonitz *et al.*, 1980a; Heckman *et al.*, 1980). However, in contrast to this, there is the extreme diversity in the organization of mitochondrial genes from different organisms, exemplified by the amount of non-coding, intergenic material found in the yeast mt genome, and the total absence of any non-coding material in HeLa mtDNA (Bernardi, 1979; Anderson *et al.*, 1981). These two extremes may represent different responses to the selective pressures that the mitochondria experience, as well as organism-specific effects resulting from differences in the relationship between the nucleus, cytoplasm and mitochondrion.

Recombinant DNA technology, DNA sequencing and RNA sequencing have been combined to particular effect to answer questions about the structure, organization and expression of mitochondrial genomes, but as yet have left unanswered the questions of why such extremes of organization and the particular processes of gene expression have been selected. It seems likely that the future will see answers to these, and other, outstanding questions, which will serve not only to increase our understanding of a fundamental genetic system, but also to suggest avenues that might have been followed in the development of the more sophisticated mechanisms of nuclear gene expression.

I am extremely grateful to the many people who generously supplied preprints, and also wish to thank Joel Mandelstam, Keith Dyke and Roger Scaise for discussions and assistance with the manuscript.

References

- Albring, M., Griffith, J. & Attardi, G. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1348–1352
- Alexander, N. J., Vincent, R. D., Perlman, P. S., Miller, D. H., Hanson, D. K. & Mahler, H. R. (1979) *J. Biol. Chem.* **254**, 2471–2479
- Aloni, Y. & Attardi, G. (1971a) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1757–1761
- Aloni, Y. & Attardi, G. (1971b) *J. Mol. Biol.* **55**, 251–270
- Anderson, A., 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. & Young, I. G. (1981) *Nature (London)* **290**, 457–465
- Angerer, L., Davidson, N., Murphy, W., Lynch, D. & Attardi, G. (1976) *Cell* **9**, 81–90
- Attardi, G. (1981a) *Trends Biochem. Sci.* **6**, 86–89
- Attardi, G. (1981b) *Trends Biochem. Sci.* **6**, 100–103
- Attardi, G., Cantatore, P., Ching, E., Cres, S., Gelfand, R., Merkel, C., Montoya, J. & Ojala, D. (1980a) in *Organisation and Expression of the Mitochondrial Genome* (Kroon, A. M. & Saccone, C., eds.), pp. 103–119, North-Holland, Amsterdam
- Attardi, G., Cantatore, P., Ching, E., Cres, S., Gelfand, R., Merkel, C., Montoya, J. & Ojala, D. (1980b) *Int. Congr. Cell Biol.* **2nd**, 225–238
- Baldacci, G., de Zamaroczy, M. & Bernardi, G. (1980) *FEBS Lett.* **114**, 234–236
- Barrell, B. G., Anderson, S., Bankier, A. T., de Bruijn, M. H. L., Chen, E., 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. & Young, I. G. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3164–3166
- Berk, A. J. & Clayton, D. A. (1974) *J. Mol. Biol.* **86**, 801–824
- Bernardi, G. (1979) *Trends Biochem. Sci.* **4**, 197–201
- Bernardi, G., Prunell, A., Fonty, G., Kopecka, H. & Strauss, F. (1976) in *The Genetic Function of Mitochondrial DNA* (Saccone, C. & Kroon, A. M., eds.), pp. 185–198, North-Holland, Amsterdam
- Blanc, H. & Dujon, B. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3942–3946
- Bogenhagen, D. & Clayton, D. A. (1978) *J. Mol. Biol.* **119**, 49–68
- Bonitz, S. G. & Tzagoloff, A. (1980) *J. Biol. Chem.* **255**, 9075–9081
- Bonitz, S. G., Berlan, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F. G., Nobrega, M. P., Thalenfeld, B. E. & Tzagoloff, A. (1980a) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3167–3170
- Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagoloff, A. & Macino, G. (1980b) *J. Biol. Chem.* **255**, 11927–11941
- Borst, P. (1972) *Annu. Rev. Biochem.* **41**, 333–376
- Borst, P. & Grivell, L. A. (1978) *Cell* **15**, 705–723
- Borst, P. & Grivell, L. A. (1981) *Nature (London)* **289**, 439–440
- Borst, P., Bos, J. L., Grivell, L. A., Groot, G. S. P., Heyting, C., Moorman, A. F. M., Sanders, J. P. M., Talen, J. L., van Kreijl, C. F. & van Ommen, G. J. B. (1977) in *Mitochondria 77: The Genetics and Biogenesis of Mitochondria* (Bandlow, W., Schweyen, R. J., Wolfe, K. & Kaudewitz, F., eds.), pp. 213–254, de Gruyter, Berlin
- Brown, W. M., Shine, J. & Goodman, H. M. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 735–739
- Cantatore, P. & Attardi, G. (1980) *Nucleic Acids Res.* **8**, 2605–2625
- Church, G. M., Slonimski, P. P. & Gilbert, W. (1979) *Cell* **18**, 1209–1215
- Crews, S., Ojala, D., Posakony, J., Nishiguchi, J. & Attardi, G. (1979) *Nature (London)* **277**, 192–198
- Crick, F. H. C. (1966) *J. Mol. Biol.* **19**, 548–555
- de Zamaroczy, M., Marotta, R., Faugeron-Fonty, G., Goursot, R., Mangin, M., Baldacci, G. & Bernardi, G. (1981) *Nature (London)* **292**, 75–78
- Dhawale, S., Hanson, D. K., Alexander, N. J., Perlman, P. S. & Mahler, H. R. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1778–1782
- Dujardin, G., Pajot, P., Groudinsky, O. & Slonimski, P. P. (1980) *Mol. Gen. Genet.* **179**, 469–482
- Dujon, B. (1980) *Cell* **20**, 185–197
- Dujon, B. (1981) in *Molecular Biology of the Yeast Saccharomyces* (Strathern, J. H., Jones, E. W. & Broach, J. R., eds.), Cold Spring Harbor Press, in the press
- Dujon, B., Slonimski, P. P. & Weill, L. (1974) *Genetics* **78**, 415–437
- Faugeron-Fonty, G., Culard, F., Baldacci, G., Goursot, R., Prunell, A. & Bernardi, G. (1979) *J. Mol. Biol.* **134**, 493–537
- Gaillard, C., Stauss, F. & Bernardi, G. (1980) *Nature (London)* **283**, 218–220
- Gelfand, R. & Attardi, G. (1981) *Nucleic Acids Res.* in press
- Gillum, A. M. & Clayton, D. A. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 677–681
- Gillum, A. M. & Clayton, D. A. (1979) *J. Mol. Biol.* **135**, 353–368
- Grivell, L. A., Arnberg, A. C., Boer, P. H., Borst, P., Bos, J. L., van Bruggen, E. F. J., Groot, G. S. P., Hecht, N. B., Hensgens, L. A. M., van Ommen, G. J. B. & Tabak, H. F. (1979) *Extrachromosomal DNA, ICN-UCLA Symp. Mol. Cell. Biol. Proc. 15th*, 305–324
- Grivell, L. A., Arnberg, A. C., Hensgens, L. A. M., Roosendaal, E., van Ommen, G. J. B. & van Bruggen, E. F. J. (1980) *Dev. Genet.* **2**, 37–49
- Haid, A., Schweyen, R. J., Bechmann, H., Kaudewitz, F., Solioz, M. & Schatz, G. (1979) *Eur. J. Biochem.* **94**, 451–464
- Heckman, J. E., Sarnoff, J., Alzner-DeWeerd, B., Yen, S. & Rajbhandary, U. L. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3159–3163
- Heyting, C., Meijlink, F. C. P. W., Verbeet, M. P., Sanders, J. P. M., Bos, J. L. & Borst, P. (1979) *Mol. Gen. Genet.* **168**, 231–250
- Jacq, C., Lazowska, J. & Slonimski, P. P. (1980) in *Organisation and Expression of the Mitochondrial Genome* (Kroon, A. M. & Saccone, C., eds.), pp. 138–152, North-Holland, Amsterdam

- Jacq, C., Pajot, P., Lazowska, J., Dujardin, G., Claisse, M., Groudinsky, O., de la Salle, H., Grandchamp, C., Labouesse, M., Gargouri, A., Guiard, B., Spyridakis, A., Dreyfus, M. & Slonimski, P. P. (1981) in *Mitochondrial Genes* (Slonimski, P., Attardi, G. & Borst, P., eds.), Cold Spring Harbor Press, in the press
- Kasamatsu, H., Robberson, D. L. & Vinograd, J. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2252–2257
- Kreike, J., Bechmann, H., van Hemert, F. J., Schweyen, R. J., Boer, P. H., Kaudewitz, F. & Groot, G. S. P. (1979) *Eur. J. Biochem.* **101**, 607–617
- Kroon, A. M. & Saccone, C. (1980) (eds.) *The Organisation and Expression of the Mitochondrial Genome*, North-Holland, Amsterdam
- Lagerkvist, U. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1759–1762
- Lamouroux, A., Pajot, P., Kochko, A., Halbreich, A. & Slonimski, P. P. (1980) in *Organisation and Expression of the Mitochondrial Genome* (Kroon, A. M. & Saccone, C., eds.), pp. 153–156, North-Holland, Amsterdam
- Lazowska, J., Jacq, C. & Slonimski, P. P. (1980) *Cell* **22**, 333–348
- Linnane, A. W. & Nagley, P. (1978) *Plasmid* **1**, 324–345
- Locker, J. & Rabinowitz, M. (1979) *Methods Enzymol.* **56**, 3–16
- Locker, J., Rabinowitz, M. & Getz, G. S. (1974) *J. Mol. Biol.* **88**, 489–507
- Macino, G. & Tzagoloff, A. (1980) *Cell* **20**, 507–517
- Mahler, H. R., Perlman, P. S., Hanson, D. K., Lamb, M. R., Anziano, P. G., Glaus, K. R. & Haldi, M. L. (1981) in *Mitochondrial Genes* (Slonimski, P., Attardi, G. & Borst, P., eds.), Cold Spring Harbor Press, in the press
- Martens, P. A. & Clayton, D. A. (1979) *J. Mol. Biol.* **135**, 327–351
- Martin, N. C., Miller, D. L. & Donnelson, J. E. (1979) *J. Biol. Chem.* **254**, 11729–11734
- Montoya, J., Ojala, D. & Attardi, G. (1981) *Nature (London)* **290**, 465–470
- Morimoto, R., Lewin, A., Hsu, H.-J., Rabinowitz, M. & Fukuhara, H. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3868–3872
- Morimoto, R., Lewin, A. & Rabinowitz, M. (1977) *Nucleic Acids Res.* **4**, 2331–2351
- Murphy, W. I., Attardi, B., Tu, C. & Attardi, G. (1975) *J. Mol. Biol.* **99**, 809–814
- Ojala, D., Montoya, J. & Attardi, G. (1981) *Nature (London)* **290**, 470–474
- Nobrega, F. G. & Tzagoloff, A. (1980) *J. Biol. Chem.* **255**, 9828–9837
- Ojala, D., Merkel, C., Gelfand, R. & Attardi, G. (1980) *Cell* **22**, 393–403
- Prunell, A. & Bernardi, G. (1974) *J. Mol. Biol.* **86**, 825–841
- Prunell, A. & Bernardi, G. (1977) *J. Mol. Biol.* **110**, 53–74
- Prunell, A., Kopecka, H., Strauss, F. & Bernardi, G. (1977) *J. Mol. Biol.* **110**, 17–52
- Quetier, F. & Vedel, F. (1977) *Nature (London)* **268**, 365–368
- Saccone, C. & Kroon, A. M. (1976) (eds.) *The Genetic Function of Mitochondrial DNA*, North-Holland, Amsterdam
- Sanders, J. P. M., Heyting, C., Verbeet, M. Ph., Meijlink, F. C. P. W. & Borst, P. (1977) *Mol. Gen. Genet.* **157**, 239–261
- Schmidt, F. J., Seidman, J. G. & Bock, R. M. (1976) *J. Biol. Chem.* **251**, 2440–2445
- Shine, J. & Dalgarno, L. (1974) *Biochem. J.* **141**, 609–615
- Slonimski, P. P., Pajot, P., Jacq, C., Foucher, M., Perrodin, G., Kochko, A. & Lamouroux, A. (1978) in *Biochemistry and Genetics of Yeast* (Bacila, M., Horecker, B. L. & Stoppani, A. O. N., eds.), pp. 339–369, 391–401, Academic Press, New York
- Smith, G. R. (1981) *Cell* **24**, 599–600
- Soloz, M. & Schatz, G. (1979) *J. Biol. Chem.* **254**, 9331–9334
- Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4734–4738
- Thalenfeld, B. E. & Tzagoloff, A. (1980) *J. Biol. Chem.* **255**, 6173–6180
- Tzagoloff, A., Macino, G. & Sebald, W. (1979) *Annu. Rev. Biochem.* **48**, 419–441
- Tzagoloff, A., Nobrega, M., Akai, A. & Macino, G. (1980) *Curr. Genet.* **2**, 149–157
- van Ommen, G. J. B., Groot, G. S. P. & Grivell, L. A. (1979) *Cell* **18**, 511–523
- van Ommen, G. J. B., Boer, P. H., Groot, G. S. P., de Haan, M., Roosendaal, E., Grivell, L. A., Haid, A. & Schweyen, R. J. (1980) *Cell* **20**, 173–183
- Wu, M., Davidson, N., Attardi, G. & Aloni, Y. (1972) *J. Mol. Biol.* **71**, 81–93
- Yanofsky, C. (1981) *Nature (London)* **289**, 751–758