Reprinted from

GENETICS AND BIOGENESIS OF CHLOROPLASTS AND MITOCHONDRIA

Interdisciplinary Conference on The Genetics and Biogenesis of Chloroplasts and Mitochondria, Munich, Germany, August 2–7, 1976, held under the auspices of the Bayerische Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse.

Editors
Th. Bücher
W. Neupert
W. Sebald
S. Werner

1976

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

Genetics and Biogenesis of Chloroplasts and Mitochondria, Th. Bücher et al. eds.

THE MITOCHONDRIAL GENOME OF YEAST : ORGANIZATION AND RECOMBINATION

Giorgio BERNARDI

Laboratoire de Génétique Moléculaire,
Institut de Biologie Moléculaire,
Paris, France.

Introduction. Previous investigations (1-6) led us to the recognition of two sequence elements in the mitochondrial genome of wild-type Saccharomyces cerevisiae cells : 1) AT-rich stretches (GC < 5 %), essentially formed by short dAdT:dAdT and dA:dT sequences ; these stretches are characterized by a remarkable compositional homogeneity, and are very largely responsible for the "anomalous" properties of yeast mitochondrial DNA ; 2) GC-rich stretches (average GC = 32 %), characterized by an extreme compositional heterogeneity ; the GC level in these stretches ranges from 22 % in the ribosomal cistrons (7) to 65 % in double-stranded fragments which can be prepared by micrococcal nuclease digestion and which represent 10 % of mitochondrial DNA (6).

These two sequence elements each represent 50 % of the mitochondrial genome of yeast, are interspersed with each other, and their average size is larger than $1.6.10^5$. We called the AT-rich stretches "spacers", and proposed that the GC-rich stretches correspond to genes with their regulatory elements (6). Both sequence elements were first postulated on the basis of the physical and chemical properties of yeast mitochondrial DNA (1,2), and then characterized by studying the mitochondrial DNA hydrolysates obtained by degradation with spleen acid DNase (3,4) and micrococcal nuclease (6) or by depurination (5).

An analysis of mitochondrial DNA hydrolysates obtained with two restriction enzymes, Hae III and Hpa II (referred to hencefrom as Hae and Hpa), led us (8,9) to the identification of two new sequence elements, which we call restriction site clusters and GC-rich clusters. Both sequence elements are present in the micrococcal nuclease fragments richest in GC, which belong to the GC-rich stretches. These new sequence elements appear to play a regulatory role.

Restriction site clusters. A clustering of Hae and Hpa sites among themselves and with each other can be demonstrated with two simple experiments : 1) Chromatography of either Hae or Hpa digests on hydroxyapatite revealed that close to 10^6 of DNA per genome unit of 50.10^6 were present in the digests as single-stranded oligonucleotides, which could only arise from the melting of DNA segments where Hae or Hpa sites were clustered. Double hydrolysates released

$1.3\text{-}1.5.10^6$ of DNA per genome unit ; the GC content of this material was 45-62 %. 2) The fragment pattern of Hae + Hpa double hydrolysates (8, 9) showed that almost all Hae sites are clustered with Hpa sites ; the number of such (Hae,Hpa) site clusters is therefore slightly lower than the number of Hae fragments, 71 in the case of S.carlsbergensis DNA. Since $1.3\text{-}1.5.10^6$ of DNA is present in 60-70 site clusters, each site cluster is formed, on the average, by 30 base pairs. It is unlikely that these melt unless they become shorter than about 10 base pairs; this implies that each site cluster contains, on the average, no less than 3-4 sites, namely 12-16 base pairs. Since Hae sites and Hpa sites (GGCC and CCGG, respectively) have a GC level of 100 %, and (Hae, Hpa) clusters a GC level of 45-62 %, then most base pairs alternating with the restriction sites in the clusters must be A:T base pairs.

GC-rich clusters. Degradation of the micrococcal nuclease fragments richest in GC with Hpa showed that, though these restriction sites were abundant in them, they were mostly formed by sequences resistant to Hpa (as well as to Hae and HhaI, an enzyme splitting the sequence GCGC) and having a GC content of 60 %. These GC-rich clusters have two important points in common with the (Hae, Hpa) clusters : 1) they are highly non-statistical in sequence since they do not contain any of the quadruplets GGCC, CCGG and GCGC, though their GC content is 60 % (close to that of the restriction site clusters) ; 2) they are not localized within the spacers (6), nor within the genes, which would otherwise contain such peculiar sequences ; the only possible localization remains that at the border of subsequent gene-spacer units, where GC-rich clusters and (Hae,Hpa) restriction site clusters may be contiguous to each other. The amount of DNA in (Hae,Hpa) clusters is about 3 %, that in GC-rich clusters about 5 %.

Significance of (Hae,Hpa) site clusters and GC-rich clusters. The restriction site clusters are characterized by 1) a high concentration in symmetrical nucleotide sequences, the Hae and Hpa restriction sites ; and 2) the alternation of short A:T and G:C sequences. The first feature suggests by itself a binding role for proteins having a dyad axis of symmetry ; such a symmetry relationship (10,11) has been found in all specific protein-DNA interactions (12). Both features found in the (Hae, Hpa) site clusters are also present in all prokaryotic promotor and operator sites sequenced so far (13). This suggests that the (Hae, Hpa) clusterscorrespond to regulatory sequences. Two additional points in favor of this view are 1) the localization and 2) the number of such sequences. In fact,this number is about twice as large as the number of yeast mitochondrial genes known so far (24 t-RNA genes (7,14), 2 r-RNA genes (15), and, very probably, 7 genes coding for polypeptide chains (16)). These genes, however, only account for half of the potentially coding DNA ; furthermore, 60-70 mitochondrial genes would mean an average gene size of about 350,000, which is a reasonable figure for a genome in which the small t-RNA genes represent over 30 % of the

total.

Another possible role of the restriction site clusters is that they contain sites for the initiation of mitochondrial DNA replication. It is well known that the defective genomes of cytoplasmic "petite" mutants can arise by a deletion mechanism from any region of the wild-type mitochondrial genome, suggesting that multiple initiation sites for replication exist on the latter. Such sites are not likely to be located inside genes, nor inside spacers, and could well be in the clusters.

It is very probable that the GC-rich clusters also have the same biological role of the (Hae, Hpa) restriction site clusters; this is strongly suggested by both the sequence properties of the latter and by the contiguity of both types of clusters.

The genetic units of the mitochondrial genome. Several lines of evidence (9) suggest that genes and spacers are stoichiometric with each other and so are restriction site clusters and GC-rich clusters. If the stoichiometry and the topology of these four sequence elements are those just discussed, then the mitochondrial genome of yeast is organized in a number of genetic units, each one of which contains the four sequences (Figure). The relative order of (Hae,Hpa) site clusters and GC-rich cluster is unknown.

	GC rich cluster	site cluster	gene	spacer
MW	$\sim 3 \cdot 10^4$	$\sim 2 \cdot 10^4$	$\sim 3.5 \cdot 10^5$ (av.)	$\sim 4 \cdot 10^5$ (av.)
G + C	60%	45-62%	26%	< 5%

Evolution of the mitochondrial genome of yeast. The comparison of Hae and Hpa fragment patterns of mitochondrial DNAs from one S.carlsbergensis and three genetically unrelated S.cerevisiae wild-type strains revealed large differences (17-20). In particular, the fragment homology between S.carlsbergensis and S.cerevisiae DNAs is extremely small, almost all fragments having different electrophoretic mobilities. In apparent full contradiction with these results, a number of findings indicate a high level of sequence homology. For instance, (21) DNA-DNA hybridization experiments indicate 100 % homology, with no mismatch in the hybrids, for the DNAs from S.carlsbergensis and S.cerevisiae. Two, not mutually exclusive, possibilities can explain this apparent paradox : 1) Point mutations and/or base modifications at the restriction sites ; 2) Deletions and/or additions between restriction sites. The first possibility has been ruled

out on several grounds (17-20) as a mechanism accounting for more than a minor part of the changes. Therefore, the differences in Hae or Hpa restriction patterns must be due to deletions and additions between restriction sites.

Recalling that practically all Hae fragments and most Hpa fragments correspond to the genetic units of the mitochondrial genome (8,9), the important question now is which one of the sequence elements forming the units is affected by deletions and additions. Obviously such deletions and additions cannot affect the genes themselves ; the two remaining possibilities are the AT-rich spacers or the site and GC-rich clusters. The first one is the most likely since 1) the AT-rich spacers represent 50 % of the genome, the clusters less than 10 % ; if the number of these sequences is the same, as it seems to be the case, the clusters are just too short to allow for the changes in length observed ; 2) if the clusters correspond to regulatory sequences, obviously they cannot be altered without functional losses ; in contrast the AT-rich spacers are not likely to play the same role nor to be under the same evolutionary constraints.

Several mechanisms may conceivably be responsible for the deletion-addition events occurring in the AT-rich spacers of the mitochondrial genome of yeast. The basic mechanism we favor is one involving unequal crossing-over events in the AT-rich spacers. Such a mechanism is likely to involve segment exchanges among genome units within the same cell. This process can potentially lead to extensive gene translocations, a phenomenon which is compatible with the Hae and Hpa patterns, since these enzymes unlink genetic units. Gene order is, however, very largely preserved, as shown by the great similarity of the band patterns produced from all DNAs by two restriction enzymes, Eco RI and Hind II + III, which only release ten fragments or so (8).

The cytoplasmic "petite" mutation. Several years ago, it was demonstrated (22,23) that the mitochondrial DNAs from two independent, acriflavine-induced petite mutants had a greatly altered base composition (GC = 4 %) compared to the DNAs from their parent wild-type cells (GC = 18 %). Subsequent investigations (1,2) showed that the DNAs from three spontaneous petite mutants 1) had GC contents in the 16 % - 13 % range ; 2) had multimodal melting profiles exhibiting the main low-melting component corresponding to dAT:dAT and dA:dT sequences, but lacking, to different degrees, high-melting components ; 3) renatured very rapidly.

On the basis of these data and of what was known on the structure of mitochondrial DNA from wild-type cells, it was proposed (24,4,6) that petite mutants had defective genomes in which mitochondrial genes had been deleted as a result of internal recombinations in the AT-rich stretches. The high frequency of the spontaneous petite mutation was considered to be the result of high recombination rates, the latter being due to the high level of sequence homology in the AT-rich stretches.

More recently, the organization of the mitochondrial genomes from several spontaneous petite mutants has been investigated again using the same approaches applied to the study of the mitochondrial genome from wild-type cells. Physical and chemical analyses and the use of enzymatic degradations by spleen acid DNase and by micrococcal nuclease has demonstrated that the basic organization of the petite genome is characterized, like that of the wild-type genome, by an interspersion of AT-rich stretches and GC-rich stretches. The most interesting results came, however, from the use of restriction enzyme degradations (25).

In one typical case, that of the spontaneous "petite" strain a_1 (18), the Hae and Hpa fragment patterns showed the following features : 1) practically all the bands had mobilities identical to those of the DNA from the parent strain A ; only 2 Hpa and 4 Hae bands had different mobilities ; 2) about one third of the bands (34 Hpa and 32 Hae bands) of the DNA from strain A were missing ; 3) a certain number of bands in the petite DNA had higher intensities than the corresponding bands in the wild-type DNA (8,25). In sharp contrast, 8 "new bands" out of 17, are seen in the Hind II + III pattern of DNA from "petite" a_1. Since the sum of the molecular weights of all Hind II + III bands (regardless of their multiplicity) is 70 millions, more than twice that of all Hae or Hpa bands, 30 millions, this result suggests that all or most of Hind II + III "new bands" contain tandem repeats of sequences present in the wild-type DNA ; such tandem repeats obviously lack Hind II + III sites, yet contain Hae and Hpa sites and lead, upon digestion with the latter enzymes, to the production of reinforced bands. These findings provide a direct evidence for the deletion and amplification phenomena characterizing the "petite" mutation.

As far as the mechanism of the "petite" mutation is concerned, the "petite" genomes may conceivably occur in either of the two repetitive sequences of the wild-type genome, the AT-rich spacers or the GC-rich sequences, (restriction and GC-rich clusters). In the first case, the process would be essentially similar to that just described for the divergence of wild-type genomes, except that sequences essential for the respiratory function would have been deleted. The two processes, which are not mutually exclusive, have different consequences on the Hae and Hpa restriction patterns. In the case of crossing-overs in the AT-rich spacers, each event will cause the appearance in the "petite" pattern of a "new band", namely a band not present in the pattern of the parental wild-type strain; in the case of crossing-overs in the GC-rich sequences, no "new band" will be formed. It would be very interesting to assess the relative importance of the two processes by looking at the "new bands" present in the "petite" patterns. This is, however, made difficult by the heterogeneity of the mitochondrial genome of "petite" strains". The "petite" strain a_1 falls in this category, since clones issued from it show a different suppressiveness and a different Hae and Hpa restriction pattern.

Recombination in crosses. This was investigated (18,26) by studying the Hpa restriction patterns of cloned diploids issued from 1) zygotes obtained in mass mating experiments ; 2) individual buds from two cell lineages ; 3) individual zygotes. In all cases, crosses concerned the two S.cerevisiae wild-type strains A and B, previously investigated. The three major results of this work concern 1) the demonstration of physical recombination in crosses of wild-type yeast cells ; most bands from diploid clones are identical in mobility with those of either parent strain ; 2) the demonstration of "new bands" namely bands absent in either parent DNA ; these "new bands" appear to be due to a recombinational event in the spacers of two "homologous Hpa segments" of different length in the two parent DNAs; 3) the rate of the process ; recombination in zygotes, and possibly in vegetative cells as well, is an extremely fast process compared to point mutation. Results concerning the segregation of the mitochondrial genome into the buds will not be mentioned here.

Recombination in mitochondrial DNA fragments cloned in E.coli. Experiments (27) in which Eco RI fragments of mitochondrial DNA from S.carlsbergensis were cloned in E.coli revealed, in addition to plasmids which harbor some of the smallest Eco RI fragments, plasmids carrying mitochondrial DNA fragments showing an electrophoretic mobility different from those of the original Eco RI fragments. These cloned fragments are likely to represent deletion products originating from the larger Eco RI fragments which were not found in any of the plasmids. Since such deletion processes do not take place in the same rec A^- host-vector system for any of the Eco RI fragments obtained from λ DNA, it is likely that repetitive sequences once more play a role in the recombination process.

Conclusions. Two main conclusions can be drawn from the results briefly summarized here. The first one is that the mitochondrial genome of yeast has an eukaryotic and not a prokaryotic type of organization. In fact, this genome, like the nuclear genome of eukaryotes (28), is an interspersed system of unique and repetitive nucleotide sequences. In round figures, 60 to 70 gene sequences, averaging 530 base pairs in length, alternate with repetitive sequences, averaging 380 base pairs in length. These repetitive sequences are formed by the long AT-rich spacers ($\sim$ 600 base pairs) and by the short (Hae,Hpa) site clusters ($\sim$ 30 base pairs) and GC-rich clusters ($\sim$ 50 base pairs). The AT-rich spacers appear to play an important role in recombination phenomena ; the other two elements may correspond to promotor and operator sequences and may also contain initiation sites for DNA replication. Again as in the case of the nuclear genome of eukaryotes, the transcription of the mitochondrial genome appears to be monocistronic : it is known (29), that the two ribosomal RNA cistrons are very far apart from each other ; on the other hand no evidence has been found for transcription products originating from AT-rich spacers. The coordination of the expression of the mitochondrial genome does not, therefore, take place through polycistronic

transcription, but rather through the simultaneous transcription at possibly identical promotor sites all over the genome ; similarly, the repression of this genome might occur through the interaction of repressors, possibly of nuclear origin, with a number of operator sites.

The second main conclusion is that in an interspersed system of unique and repetitive sequences, like the mitochondrial genome of yeast, recombination phenomena taking advantage of the sequence homology in the repetitive sequences, play a very important role. Genome divergence and evolution proceed essentially by such a mechanism, which is orders of magnitude faster than the point mutation mechanism. The production of the defective "petite" genomes seems to be just a very frequent accident in the general phenomenon of evolution ; these mutants, "lethal" in terms of mitochondrial function, can be observed simply because yeast can dispense with functional mitochondria.

An important implication of these results is that the mitochondrial genome of yeast really is a simple model of the nuclear genome of eukaryotes. A knowledge of its organization, expression and evolution may therefore be of great value in providing interesting clues for probing and understanding the much more complex nuclear genome. In this connection, the features of the evolution of the mitochondrial genome of yeast suggest that recombination processes taking place at a high rate at the interspersed repetitive sequences of the nuclear genome of eukaryotes may have played a very important role in evolution. A key feature of evolution by recombination at interspersed repetitive sequences is that it can generally affect the organization and the expression of the genome by causing translocations, inversions, duplications and deletions of genes, whereas it does not affect the nucleotide sequences of the latter ; only the much slower process of point mutation is operative on these. A major difference between prokaryotes and eukaryotes might precisely be that the latter developed an interspersed system of repetitive sequences allowing them to evolve much more rapidly than prokaryotes.

REFERENCES

1. Bernardi G., Faurès M., Piperno G., Slonimski P.P. (1970), J. Mol. Biol., 48. 23-42
2. Bernardi G., Timasheff S.N. (1970) J. Mol. Biol., 48. 43-52
3. Bernardi G., Piperno G., Fonty G. (1972) J. Mol. Biol., 65. 173-189
4. Piperno G., Fonty G., Bernardi G. (1972) J. Mol. Biol., 65. 191-205
5. Ehrlich S.D., Thiery J.P., Bernardi G. (1972) J. Mol. Biol., 65. 207-212
6. Prunell A., Bernardi G. (1974) J. Mol. Biol., 86. 825-841
7. Reijnders L., Borst P. (1972) Biochem. Biophys. Res. Comm., 47. 126-133
8. Bernardi G., Prunell A., Fonty G., Kopecka H., Strauss F. (1976) in The Genetic Function of Mitochondrial DNA (Saccone C., Kroon A.M.,eds.), North-Holland, Amsterdam

9. Prunell A., Bernardi G. (1976) J. Mol. Biol., submitted

10. Bernardi G. (1965) J. Mol. Biol., 13. 603-605

11. Bernardi G. (1968) in Advances in Enzymology (F.F. Nord,ed.), 31. 1-49, Wiley, New York

12. Sobell H.M. (1973) in Advances in Genetics (E. Caspari,ed.), 17. 411-490, Academic Press, New York

13. Dykes G., Bambara R., Marians K., Wu R. (1975) Nucleic Acids Res., 2. 327-345

14. Casey J., Cohen M., Rabinowitz M., Fukuhara H., Getz G.S. (1972) J. Mol. Biol., 63. 431-

15. Reijnders L., Kleisen C.M., Grivell L.A., Borst P. (1972) Biochim. Biophys. Acta, 272. 396-407

16. Schatz G., Mason T.L. (1974) Ann. Rev. Biochem., 43. 51-87

17. Bernardi G., Prunell A., Kopecka H. (1975) in Molecular Biology of Nucleocytoplasmic Relationships (S.Puiseux-Dao,ed.), Elsevier, Amsterdam, 85-90

18. Bernardi G. (1975) in Proceedings of the 10th FEBS Meeting, vol. 38 (G. Bernardi and F. Gros,eds.), North-Holland, 41-56

19. Bernardi G. (1976) in Euratom Symposium on "Mitochondrial Genetics Biogenesis and Bioenergetics" De Gruyter, Berlin (in press)

20. Prunell A., Kopecka H., Strauss F., Bernardi G. (1976) J. Mol. Biol., submitted

21. Groot S.P., Flavell R.A., Sanders J.P.M. (1975) Biochim. Biophys. Acta, 378. 186-194

22. Bernardi G., Carnevali F., Nicolaieff A., Piperno G., Tecce G. (1968) J. Mol. Biol., 37. 493-505

23. Mehrotra B.D., Mahler H.R. (1968) Arch. Biochem. Biophys., 128. 685-703

24. Bernardi G. (1969) Unpublished

25. Prunell A., Fonty G., Bernardi G. (1976) J. Mol. Biol., to be submitted

26. Fonty G., Kopecka H., Wilkie D., Bernardi G. (1976) J. Mol. Biol., to be submitted

27. Bernardi A., Bernardi G. (1976) In preparation

28. Davidson E.H., Galau G.A., Angerer R.C., Britten R.J. (1975) Chromosoma, 51. 253-259

29. Sanders J.P.M., Heyting G., Borst P. (1975) Biochem. Biophys. Res. Comm., 65. 699-707.