Volume 13 Number 8 1985 Nucleic Acids Research

A mitochondrial reading frame which may code for a maturase-like protein in *Saccharomyces cerevisiae*

Bertrand Séraphin, Michel Simon and Gérard Faye

Institut Curie, Section de Biologie, Bâtiment 110, Centre Universitaire, 91405 - Orsay, France

Received 26 January 1985; Revised and Accepted 18 March 1985

ABSTRACT

In S. cerevisiae, the large oxi3/oli2 mitochondrial transcript contains the products of the oxi3, aap1 and oli2 genes and an unassigned reading frame, RF3. In the work presented here, we have completed the nucleotide sequence of RF3. We have shown that RF3 is composed of four fairly large ORFs which overlap within GC rich sequences. Furthermore, a shift of +1 base was found between each pair of consecutive reading frames. We discuss how these frameshifts could be removed to produce a 500 aminoacid long protein containing the two well conserved P1 and P2 oligopeptide sequences featuring several mitochondrial intron reading frames, suggesting, thereby, a RNA-maturase-like activity for the putative RF3 protein. In addition, we suggest that the insertion of GC clusters in a gene could provide a novel way of regulating its expression.

INTRODUCTION

In S. cerevisiae, the whole span of mitochondrial DNA coding for oxi3 (cytochrome oxidase subunit 1 gene) (1), aap1 (ATPase subunit 8) (2) and oli2 (ATPase subunit 6) (3) is transcribed as a large multigenic primary transcript that extends beyond an unassigned reading frame called RF3 (formerly named URF2) (4). This large oxi3/oli2 message undergoes several processing reactions (4,5,6). Whereas the introns of oxi3 are excised and its exons are spliced together, cuts downstream of oxi3 liberate a polycistronic RNA carrying aap1, oli2 and RF3, which is further processed into shorter pieces (4). Both mitochondrial products (RNA-maturases) (7) and nuclear-coded proteins are involved at these different steps (8,9).

In order to study the RNA processing downstream of oli2, we completed the nucleotide sequence surrounding RF3, which had been partially established by Macino and Tzagoloff (3). We then discovered, in the 1450 nucleotide long DNA stretch starting about 80 bases downstream of the oli2 stop codon, four fairly large overlapping reading frames. Moreover, each overlapping occurs in a region containing a G+C rich sequence. A shift of +1 base was found between each pair of consecutive reading frames. Mechanism(s) (either at the transcriptional, post-transcriptional or translational level) able to annihilate the frameshifts would give rise to a protein some 500 aminoacids long. This structure is reminiscent of that of the RF2 region recently

described by Michel (10). Furthermore, the putative RF3 protein would contain the P1 and P2 oligopeptide sequences found in several mitochondrial intron reading frames as well as in the RF1 and RF2 sequences (10), which suggests that it could be endowed with a RNA-maturase-like activity.

MATERIALS AND METHODS

Yeast and bacterial strains :

The yeast strains employed were D273-10B/A α met (11), JM6 a ade2-1, lys-1-1, his4-580, met8-1, SUP4-3 (12) and X4005-11A a met5 leu2. This latter is from the Yeast Genetic Stock Center. Escherichia coli strain RR1 was used for transformation (13).

Preparation of mitochondrial DNA :

Mitochondrial DNA of strains D273-10B/A, JM6 and X4005-11A was purified by bisbenzimide/CsCl buoyant density centrifugation (14). The EcoRI restriction fragment RR8 (Figure 1) of the D273-10B/A mitochondrial genome was cloned in pBR328 (15). Restriction analysis and cloning in E.coli were carried out as described in Maniatis et al. (16).

DNA sequencing:

DNA sequencing was performed as described by Maxam and Gilbert (17). The products of sequencing reactions of fragments carrying GC cluster II were run on acrylamide gels containing 90 % formamide (18).

RESULTS

The RR8 restriction fragment (Figure 1) of strain D273-10B/A was cloned in the plasmid pBR328. It was sequenced according to the strategy shown in Figure 2. The corresponding sequence is presented in Figure 3. The 1784 bp long RR8 fragment contains the ultimate 81 bp of the oli2 reading frame (3). Three G+C rich clusters punctuate this RR8 sequence. About 100 GC clusters (19) are scattered throughout the

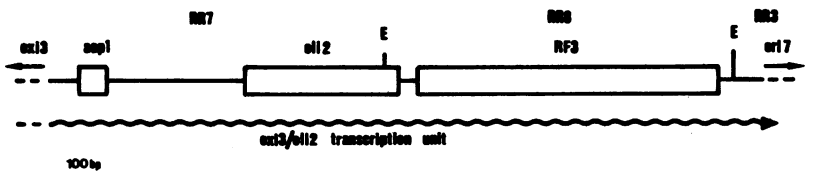

Figure 1 : Organization of the oxi3/oli2 transcription unit in the RF3 region. oxi3 : cytochrome oxidase subunit 1, aap1 : ATPase subunit 8, oli2 : ATPase subunit 6, RF3 : unassigned reading frame, ori7 : origin of replication. E : EcoRI restriction sites. RR7, RR8, RR3 : EcoRI restriction fragments. The long arrow indicates the direction of transcription.

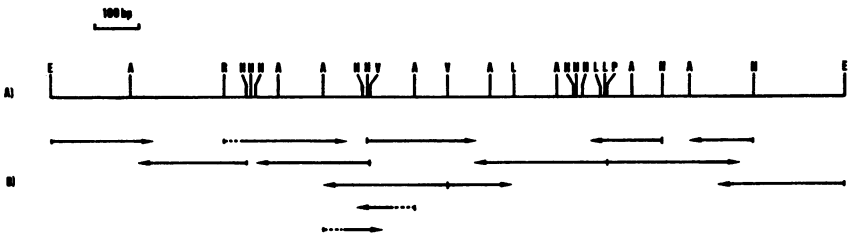

Figure 2 : EcoRI restriction fragment RR8. Part A of the diagram shows some restriction recognition sites in fragment RR8. A : AhaIII, E : EcoRI, H : HpaII, L : AluI, N : HinfI, P : PstI, R : RsaI, V : AvaII. Part B depicts the sequencing strategy of DNA fragment RR8.

mitochondrial genome ; they may be arranged into several classes according to their sequence. The three GC rich runs of the RR8 fragment are delimited by the sequences 5'-TAGT and AAGGAG-3' and fall into two of these classes, several members of which are compared in Figure 4. Class A clusters I and III are 43 bp long, classe B cluster II is 33 bp long ; each of them has the same orientation. These clusters belong to the very unusual "gene" RF3. Its first reading frame starts with a ATG codon (nucleotide 161) and stops just downstream of cluster I. Starting from the 5'-end of cluster I, a second reading frame stops downstream of cluster II and it is followed by a third frame which starts at the 5'-end of cluster II and ends shortly after cluster III. Finally, a fourth reading frame begins at the 5'-end of the last cluster and stops at position 1700 (Figures 3 and 5). These four reading frames are not in phase with each other, a shift of +1 base is present between each consecutive pair of reading frames. Any mechanism capable of suppressing the +1 frameshift would allow (see discussion) the synthesis of a rather large protein. It is interesting to note that this organization of RF3 is reminiscent of that of RF2 which is located downstream from the oxi2 gene (10). RF2 is also interrupted by three GC clusters : one belonging to class A and two to class B. Two other similar structures may in fact exist in the mitochondrial genome. The RF1 reading frame located downstream from the oxi1 gene (20) and the aI5β open reading frame present in an optional intron of the cytochrome oxidase subunit 1 gene (21) both contain a class A GC cluster. As published, RF1 and aI5β are not associated with a frameshift ; however, as stated by the authors themselves (20), the former published sequence could be off by at least one base. The putative RF3 sequence was compared with that of the RF1 and RF2 reading frames. No sequence homology was found between these proteins except for two short conserved oligopeptide sequences featuring the ORFs of group IB introns and called P1 and P2 by Michel et al. (22) and decapeptides in Waring et al. (23) (Figures 3 and 6). The homology between RF3 and these intron sequences is restricted to only these two oligopeptides, however.

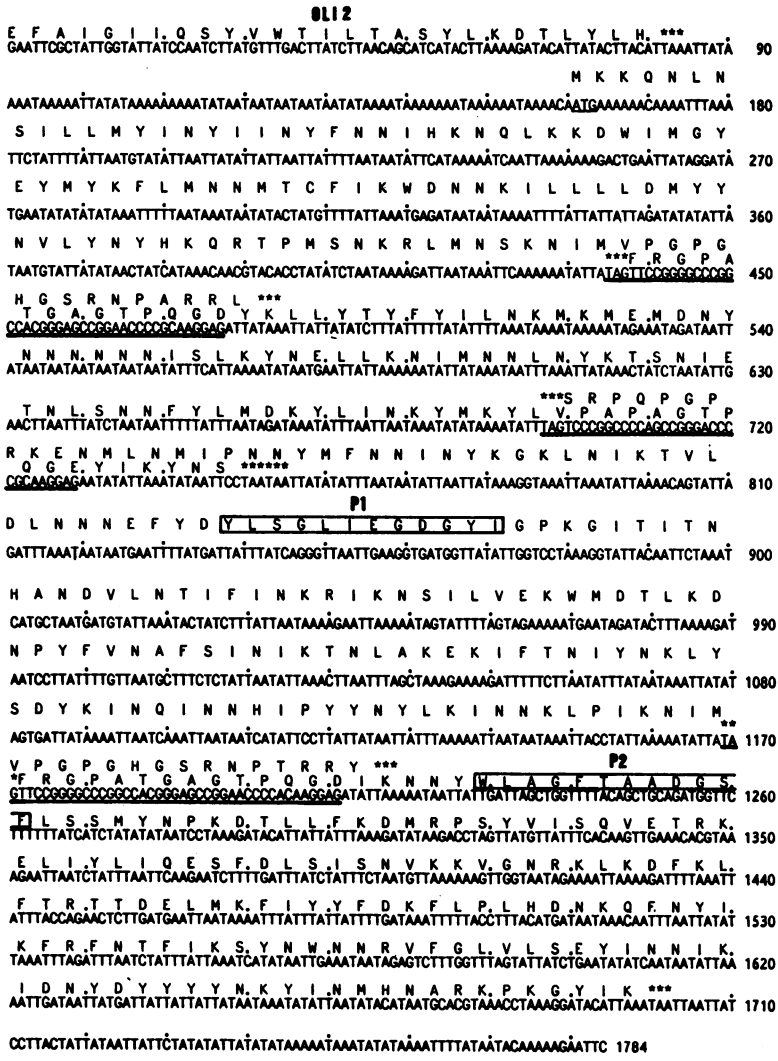

Figure 3 : Nucleotide sequence of the RR8 restriction fragment of strain D273-10B/A. The last 27 amino-acids of oli2 and the sequence of the RF3 putative protein are indicated by the single-letter amino-acid code (IUPAC-IUB Commission, 1969) using the codon recognition rule of yeast mitochondria were TGA is used for Trp, CTN for Thr and ATA for Met. The GC clusters are underlined. The oligopeptide sequences P1 and P2 are boxed. +1 frameshifts are observed at the level of RF3 GC clusters I, II and III. RF2 GC cluster I which has the same orientation as the RF3 clusters also introduces a +1 frameshift, whereas RF2 GC clusters II and III which are oriented in the opposite direction give -1 frameshifts (10). If, actually, a base is missing in the RF1 GC cluster sequence, as stated by Coruzzi et al. (20), RF1 would be composed of two reading frames with a +1 shift. However, the orientation of the RF1 cluster is the inverse of that found in RF3 clusters.

Figure 4 : Homologies between GC clusters. GC rich sequences related to RF3 clusters I or III and II are aligned to illustrate their sequence similarity. Bases diverging from the consensus are in lower case letters. Extra bases are indicated below the sequences. The sources of the sequences are as follows : a : 20 (1904-2359) ; b : 10 (89, 354, 1046, 1663) ; c : 21 (2043) ; d : 36 (508, 480, 95, 84, 944, 1603) ; e : 37 (269) ; f : 38 (1894) ; g : 39 (853) ; h : 40 (1385, 3452) ; i : 41 (3639, 4951). Bracketed numbers point out the position of the first nucleotide of the GC cluster according to the nucleotide enumeration used in the references. + or - indicate the orientation of GC cluster on the mitochondrial genome. The arrows above or below the consensus sequences show dyad symetries as described by Goursot et al. (36) or Michel (10) respectively.

The codon usage of RF3 appears to be similar to that of the intron reading frames (24) and to that of RF1 and RF2 (10). When it is compared with that of the major mitochondrial-encoded protein genes, we notice an excess of the AUA codon (Met) as well as a paucity of the UUC codon (Phe). Furthermore, RF3, like RF1 and RF2, is rather rich in asparagine and lysine.

As previously shown by Nagley et al. (12), a class (class II) of S.cerevisiae strains is lacking a 1,8 kb mtDNA fragment that is normally present in long (class I)

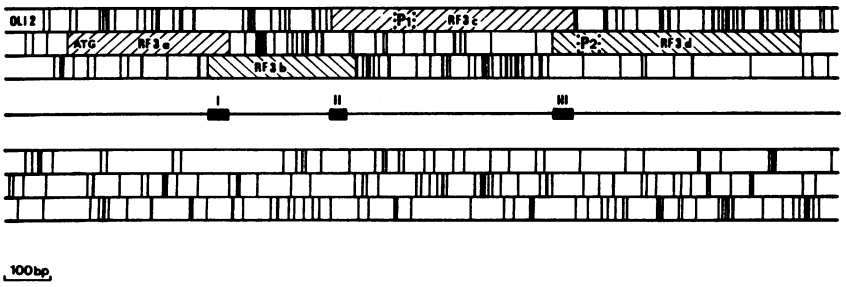

Figure 5 : Termination codons in the restriction fragment RR8. Codons TAA and TAG, in the three frames of both DNA strands are represented by vertical bars. The four reading frames RF3a, b, c and d which overlap at the level of GC clusters are 321, 312, 516 and 528 bases long, respectively. Only the RF3a reading frame starts with a ATG codon. The black boxes on the horizontal median line indicate the position of GC clusters. The two oligopeptide motifs P1 and P2 are depicted. oli2 : ATPase subunit 6 gene.

strains downstream of the oli2 gene. We have sequenced the mtDNA of two class II strains, namely JM6 and X4005-11A, over a region overlapping the deleted fragment. The comparison of these sequences to that of (class I) strain D273-10B/A is shown in Figure 7. Both class II strains are lacking a DNA fragment, 1667 bp in length, that include the whole RF3 sequence. Instead, an additional 40 bp long fragment was found in their DNA.

DISCUSSION

The odd structure of the RF3 gene, which consists of four fairly large consecutive overlapping ORF's (107, 104, 172, 176 codons long, respectively), suggests that it may be an inactive gene. It is possible that a formerly functional RF3 gene was made silent by the insertion of GC clusters which may be considered to be mobile elements (25,26). The same hypothesis could apply as well to other mitochondrial genes such as RF2 and very probably RF1 (10). However, the probability that three different genes could be inactivated by the same process and during the same recent evolutionary time period (since no nonsense mutations occured in the ORF's) is very low. We find it much more likely and stimulating that RF3 is active. Indeed, we have recently shown that the RF3 region is transcribed as part of the oxi3/oli2 transcription unit which contains the highly expressed oxi3 and oli2 genes (4).

If RF3 is actually a functional gene, how is the RF3 protein produced since the four ORF's are not in phase with each other and their overlapping region are centered around three GC clusters ? Any mechanism able to introduce changes in the reading frames at the level of the GC clusters could give rise to a large continuous

		P1		P2	REF
Sc	RF3	-YLSGLIEGDGYI-	104aa	-WLAGFTAADGSF-	
Sc	AI3	-YLAGLIEGDGSI-	120aa	-WLAILTDADGNF-	1
Sc	AI4	-WLAGLIDGDGYF-	94aa	-WFVGFFDADGTI-	1
Sc	AI5β	-YLSGLFEGDGNI-	97aa	-WLTGFNDADGYF-	21
Sc	BI4	-WLAGLIDGDGYF-	94aa	-WFMGFFDADGTI-	41
Sc	RF1	-WLVGFTDGDGSF-	124aa	-WILGFIEAEGTF-	20,10
Sc	RF2	-WLSGFVVGDGYF-	121aa	-FILGFIEAEGTF-	10
An	AI2	-YLAGLIEGDGTI-	107aa	-WLSGFIEADGSF-	42
An	AI3	-YLAGLIDGDGHF-	99aa	-WLAGFSDADASG-	42
An	BI1	-YLVGLFEGDGYF-	120aa	-WLVGFIEAEGCF-	23
Nc	BI2	-FLAGLVDGDGYI-	119aa	-WIVGFIYSKGSF-	43
Nc	BI2	-FLAGLVDGDGYI-	144aa	-WIVGFTCSEGSF-	44
Nc	OI2	-YITGFVDGEG -	158aa	-WIRGFIEGEG -	43
Sp	AI2	-YLAGLIDGDGHF-	99aa	-WLAGFSDADASF-	45
CONSENSUS		-YLAGL*DGDG+F-		-WL*GF/DADGSF-	
		W S E I		I E E T	

Figure 6 : Comparison of the dodecapeptide sequences P1 and P2 which feature the intron ORF's of group IB and are present in RF1 and RF2. This comparison extends the selection of published sequences presented by Michel (10). Distances between P1 and P2 are in amino acids. The meaning of the symbols drawn in the consensus sequence are the following : * : majority of apolar AA ; + : majority of non-ionic polar AA ; / : any AA. Sc : Saccharomyces cerevisiae ; An : Aspergillus nidulans ; Nc : Neurospora crassa; Sp : Schizosaccharomyces pombe.

```
                                                                    oli2
JM6   GAATTCGCTATTGGTATCATTCAGGGATATGTCTGGGCTATTTTAACAGCATCATATTTAAAAGATGCAG
D273  GAATTCGCTATTGGTATTATCCAATCTTATGTTTGACTTATCTTAACAGCATCATACTTAAAAGATACAT

JM6   TATACTTACATTAAATTATAAAATAAAA-TTATA-AAATAAAA-TAATTTACATATGGAGTATTAAACTA
D273  TATACTTACATTAAATTATAAAATAAAAATTATATAAAAAAAAATA     1667 NUCLEOTIDES

JM6   TAATAAATACAATATACCCC-ATCCCCCCCTTTTAATAATATTCTTTTATCTAATAAAATATTTATTTAT
D273  INCLUDING RF3   CTCATATTCTTTTTTTTTTAATAATATTCTTTCATAAATATATCTATTTATTTA-
                                                          2480

JM6   TAATATTATTATTATCTTCTTCAAGGACTTATTTAATATATTTAATAAC-TTATTATACTTATTTATATT
D273  ------TATTATTATCATCTTAAAAGACTTATTTAATATATTTAATAATTTTATTATACTTATTTATATT

JM6   T-ATAATTAATACAAATATATTATTAA
D273  TTATAATTAATATAAATATATTATTAA
```

Figure 7 : Comparison of the mitochondrial DNA sequences of strains D273-10B/A, JM6 and X4005-11A, downstream of oli2. Mitochondrial DNA of JM6 and X4005-11A was digested with EcoRI. The restriction fragments RR3 were purified, 3'-end labelled with Klenow DNA polymerase (17), then digested with HpaII. The EcoRI-HpaII fragments were isolated on polyacrylamide gel slabs and their sequencing was performed as described by Maxam and Gilbert (17). JM6 and X4005-11A sequences are identical. The D273-10B DNA sequence downstream of RF3 is from Macino and Tzagoloff (3). The processing consensus sequence AATAATATTCTT (4) is underlined ; the number 2480 indicates the position of the last T and refers to the nucleotide numeration used in Macino and Tzagoloff (3).

reading frame with the potential of coding for a 500 amino-acid long protein, setting in the same frame the two well conserved P1 and P2 oligopeptide sequences. Such a mechanism could be accomplished by the translational machinery. Because of their secondary structure and sequence (which contains rarely used codons) the GC clusters might reduce the speed of mitoribosome movement and promote frameshifts (27). If GC cluster actually induce frameshifts, they most probably occur at a low frequency since the expression of the var1 protein is comparable to those of other mitochondrial components of the respiratory chain in spite of the presence of either one or two GC rich sequences in phase with its coding sequence (28). In fact it has been shown that frameshifting is necessary for the expression of several genes (29,30,31,32) although no clear secondary structure that could promote frameshift is to be found in their sequence. In addition, it is interesting to note that the mitochondrial translational machinery was shown to suppress frameshift mutations (33).

As an alternative explanation, we could envisage that the excision of the GC cluster from the RF3 transcript by a cut and splice mechanism would allow the expression of RF3.

Whatever the mechanism by which the putative RF3 protein may be expressed, the insertion of GC clusters in an open reading frame offers a potential novel way to regulate gene expression assuming that the more an open reading frame is interrupted by GC clusters, the less it may be expressed. It has been shown that low levels of RNA-maturases are sufficient to excise mitochondrial introns (34) The presence of the oligopeptide sequences P1 and P2 suggests that the RF3 putative protein may be endowed with a RNA-maturase-like activity and therefore should be found in low amount in the mitochondria. The peculiar organization of RF3 (and of RF2) seems to indicate that GC clusters could act as regulatory elements reducing protein levels although this may be a subsidiary function.

To explain why RF3 (as well as RF2) is lacking in some yeast mitochondrial genomes (cf. Figure 7) we may suppose that the RF3 open reading frame is absent in some strains either because the RF3 putative protein is necessary to catalyse the RNA processing reaction in the RF3 region itself or because its function is compensated by another coding sequence located elsewhere in the mitochondrial genome or in the nucleus. Clearly, the finding of mutations within the RF3 sequence would allow one to answer these questions. We are currently searching for such mutants.

Although RF3 is related to the IB family of intron open reading frames (22), the RF3 optional region is probably not an intron sequence for it does not retain the two highly conserved oligonucleotide motifs GACUA and AUAGUC (35). In this respect RF3 resembles RF1 and RF2, which also belong to multigenic transcription

units (5) and are located immediately downstream of non-mosaic genes coding for the respiratory proteins.

ACKNOWLEDGEMENTS

We thank H. Fukuhara for encouragements and supports, F. Michel for communicating his manuscript prior to publication and E. Sena for critical reading of the manuscript. This investigation was supported by the CNRS (ATP 3141 and ATP 2486) and by the INSERM (CRE 831011). B.S. is supported by the Ecole Normale Supérieure.

REFERENCES

1. Bonitz, S., Coruzzi, G., Thalenfeld, B., Tzagoloff, A. and Macino, G. (1980) J. Biol. Chem. 255, 11927-11941.
2. Novitski, C., Macreadie, I., Maxwell, R., Lukins, H., Linnane, A. and Nagley, F. (1984) Current Genetics 8, 135-146.
3. Macino, G. and Tzagoloff, A. (1980) Cell 20, 507-517.
4. Simon, M. and Faye, G. (1984a) Mol. Gen. Genet. 196, 266-274.
5. Osinga, K., De Vries, E., Van der Horst, G. and Tabak, H. (1984) EMBO J. 3, 829-834.
6. Cobon, G., Beilharz, M., Linnane, A. and Nagley, P. (1982) Current Genet. 5, 97-107.
7. Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E. and Slonimski, P. (1983) Cell 35, 733-742.
8. Faye, G. and Simon, M. (1983) Cell 32, 77-87.
9. Pillar, T., Lang, B., Steinberger, I., Vogt, B. and Kaudewitz, F. (1983) J. Biol. Chem. 258, 7954-7959.
10. Michel, F. (1984) Current Genetics 8, 307-317.
11. Foury, F. and Tzagoloff, A. (1976) Mol. Gen. Genet. 149, 43-50.
12. Nagley, P., Cobon, G., Linnane, A. and Beilharz, M. (1981) Biochemistry International 3, 473-481.
13. Bolivar, F., Rodriguez, P., Greene, P., Betlach, M., Heyeker, H., Boyer, H., Crosa, J. and Falkow, S. (1977) Gene 2, 95-113.
14. Simon, M. and Faye, G. (1984b) Proc. Natl. Acad. Sci. USA 81, 8-12.
15. Soberon, X., Covarrubias, L. and Bolivar, F. (1980) Gene 9, 287-305.
16. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory, New York.
17. Maxam, A. and Gilbert, W. (1980) Methods in Enzymology 65, 499-560.
18. Frank, R., Müller, D. and Wolff, C. (1981) Nucl. Acids Res. 9, 4967-4979.
19. Prunell, A. and Bernardi, G. (1977) J. Mol. Biol. 110, 53-74.
20. Coruzzi, G., Bonitz, S., Thalenfeld, B. and Tzagoloff, A. (1981) J. Biol. Chem. 256, 12780-12787.
21. Hensgens, L., Bonen, L., de Haan, M., Van der Horst, G. and Grivell, L. (1983) Cell 32, 379-389.
22. Michel, F., Jacquier, A. and Dujon, B. (1982) Biochimie 64, 867-881.
23. Waring, R., Davies, R., Scazzochio, C. and Brown, T. (1982) Proc. Natl. Acad. Sci. USA 79, 6332-6336.
24. Dujon, B. (1981) in "The Molecular Biology of the Yeast Saccharomyces. Life Cycle and Inheritance", Strathern, Jones, Broach (Eds), Cold Spring Habor Laboratory, New York, pp. 503-635.
25. Dujon, B. (1980) Cell 20, 185-197.
26. Sor, F. and Fukuhara, H. (1982) Nucl. Acids Res. 10, 1625-1633.
27. Roth, J. (1981) Cell 24, 601-602.

28. Hudspeth, M., Vincent, R., Perlman, P., Shumard, D., Treisman, L. and Grossman, L. (1984) Proc. Natl. Acad. Sci. USA 81, 3148-3152.
29. Kastelein, R.A., Remant, E., Fiers, W. and Van Duin, J. (1982) Nature 295, 35-41.
30. Hensgens, L., Brakenhoff, J., De Vries, B., Sloof, P., Tromp, M., Van Boom, J. and Benne, R. (1984) Nucleic Acids Res. 12, 7327-7344.
31. Mellor, J., Fulton, S., Dobson, M., Wilson, W., Kingsman, S. and Kingsman, A. (1985) Nature 313, 243-246.
32. Schwartz, D., Tizard, R. and Gilbert, W. (1983) Cell 30, 853-869.
33. Fox, T. and Weiss-Brummer, B. (1980) Nature 288, 60-63.
34. Jacq, C., Banroques, J., Becam, A.M., Slonimski, P., Guiso, N. and Danchin, A. (1984) EMBO J. 3, 1567-1572.
35. Michel, F. and Dujon, B. (1983) EMBO J. 2, 33-38.
36. Goursot, R., Mangin, M. and Bernardi, G. (1982) EMBO J. 1, 705-711.
37. Bonitz, S. and Tzagoloff, A. (1980) J. Biol. Chem. 255, 9075-9081.
38. Miller, D., Underbrink-Lyon, K., Najarian, D., Krupp, J. and Martin, N. (1983) in "Mitochondria 1983. Nucleo-mitochondrial Interactions". Schweyen, R., Wolf, K., Kaudewitz, F. (Eds), De Gruyter, Berlin, pp. 151-164.
39. Sor, F. and Fukuhara, H. (1980) C.R. Acad. Sci. Paris 291, 933-936.
40. Tzagoloff, A., Nobrega, M., Akai, A. and Macino, G. (1980) Current Genetics 2, 149-157.
41. Nobrega, F. and Tzagoloff, A. (1980) J. Biol. Chem. 255, 9828-9837.
42. Waring, R., Brown, T., Ray, J., Scazzochio, C. and Davies, R. (1984) EMBO J. 3, 2121-2128.
43. Citterich, M., Morelli, G. and Macino, G. (1983) EMBO J. 2, 1235-1242.
44. Burke, J., Breitenberger, C., Heckman, J., Dujon, B. and RajBhandary (1984) J. Biol. Chem. 259, 504-511.
45. Lang, F. (1984) EMBO J. 3, 2129-2136.