The EMBO Journal vol.6 no.3 pp.713-721, 1987

Three suppressor mutations which cure a mitochondrial RNA maturase deficiency occur at the same codon in the open reading frame of the nuclear *NAM2* gene

M.Labouesse, C.J.Herbert, G.Dujardin and P.P.Slonimski

Centre de Génétique Moléculaire du CNRS, Laboratoire propre associé à l'Université Pierre et Marie Curie, 91190 Gif-sur-Yvette, France

Communicated by P.P.Slonimski

Dominant mutations of the nuclear *NAM2* gene are able to compensate for a deficiency of the maturase encoded by the fourth intron of the mitochondrial cytochrome b gene. We have determined the complete nucleotide sequence of the *NAM2-1* suppressor allele. The results of S1 nuclease protection experiments show that two overlapping poly(A)$^+$ RNAs are transcribed from the gene using different promoters. The longer transcript contains two open reading frames (ORFs), a long ORF which could encode a protein of 894 amino acids, mol. wt 102 000 daltons, and a short ORF of 51 codons which is omitted from the shorter transcript. The wild-type *nam2*$^+$ and two other suppressor alleles, *NAM2-6* and *NAM2-7*, have been cloned. A comparison of the sequence of the wild-type and the three suppressor alleles shows that on three separate occasions the same codon specifying glycine was mutated (once to serine and twice to cysteine). Finally sequence comparisons identified two regions in the long ORF, distinct from the position of the suppressor mutations, that could correspond to binding domains for a nucleotide and a nucleic acid.

Key words: intron splicing/mitochondria/nuclear suppressor/RNA maturase/yeast

Introduction

There is considerable evidence that the RNA splicing reactions of mitochondrial and nuclear mRNA introns and of the rRNA intron of *Tetrahymena* are mechanistically related, the intron excision and exon ligation reactions proceeding via a series of transesterifications (Cech and Bass, 1986). Schematically, three levels of complexity can be distinguished in the requirement for various types of macromolecules in the splicing process. A few *in vitro* splicing reactions occur spontaneously in the absence of extraneous energy sources, proteins or RNAs (Kruger *et al.*, 1982; cf. Cech and Bass, 1986 for review). However, many *in vivo* splicing reactions require specific proteins which can be encoded by the intron itself or by another gene (cf. Lazowska *et al.*, 1980; Pillar *et al.*, 1983; McGraw and Tzagoloff, 1983; Last *et al.*, 1984; Lee *et al.*, 1984). Finally most nuclear mRNA splicing reactions depend on the presence of extraneous small nuclear RNAs and a number of associated proteins which form a complex particle, the spliceosome (for review, see Padgett *et al.*, 1986).

Although our understanding of the RNA chemistry of *in vitro* splicing reactions is progressing rapidly, little is known about the factors which are essential for *in vivo* splicing. The observation that an intron may self-splice *in vitro* does not exclude the involvement of other factors in its *in vivo* splicing. The importance of these factors has recently been demonstrated by the isolation of a nuclear mutant of *Neurospora crassa* which prevents the *in vivo* splicing of a mitochondrial intron which is capable of self-splicing *in vitro* (Collins and Lambowitz, 1985). It is therefore of interest to study the genes encoding these factors. For such a study, lower eukaryotes are of great value as they are amenable to genetic analysis; the yeast mitochondrial system is particularly attractive as mitochondrial splicing deficient mutants are not lethal, being able to grow by fermentation.

In an attempt to identify nuclear genes whose products might be involved in mitochondrial RNA splicing, we have isolated specific nuclear suppressors of yeast mitochondrial splicing-deficient mutants; the suppressor *NAM2-1* was isolated in this way (Dujardin *et al.*, 1980). Three types of alleles of the *NAM2* gene have been characterized: (i) the wild-type allele; (ii) suppressor alleles, capable of alleviating a splicing defect of two mitochondrial introns; and (iii) null alleles, leading to the loss of the mitochondrial genome.

Dominant suppressor alleles of the *NAM2* gene (*NAM2-1* . . . *NAM2-7*) compensate for mutations in the maturase encoded by the fourth intron of the mitochondrial gene coding for cytochrome b (bI4) (Groudinsky *et al.*, 1981). Such maturase mutations lead to a respiratory negative phenotype as this maturase is required for the excision of the intron bI4 (De la Salle *et al.*, 1982) and also for the excision of the fourth intron of the mitochondrial gene encoding subunit I of cytochrome oxidase (aI4) (Labouesse *et al.*, 1984). Suppression is dependent on the presence of a product encoded by the intron aI4 (Dujardin *et al.*, 1983). After the *NAM2-1* allele was cloned, the null allele was constructed by making an *in vitro* deletion or disruption which was reintroduced into the chromosome. Inactivation of the *NAM2* gene

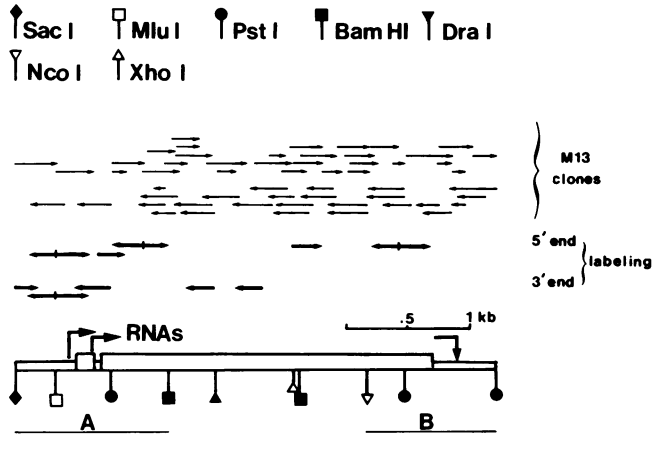

Fig. 1. Restriction map and strategy for sequencing the *NAM2-1* allele. The lower part of the figure shows the restriction map of the gene; the arrows in the upper part indicate the length and direction of the sequences determined by the methods of Sanger from M13 clones (thin arrows) or Maxam and Gilbert from end-labelled fragments (thick arrows). Above the restriction map are the positions of the major 5' and 3' ends of the two *NAM2* transcripts (arrows) and the positions of the two ORFs (large open bars). Fragments A and B were subcloned in M13 vectors and used in the S1 mapping experiments.

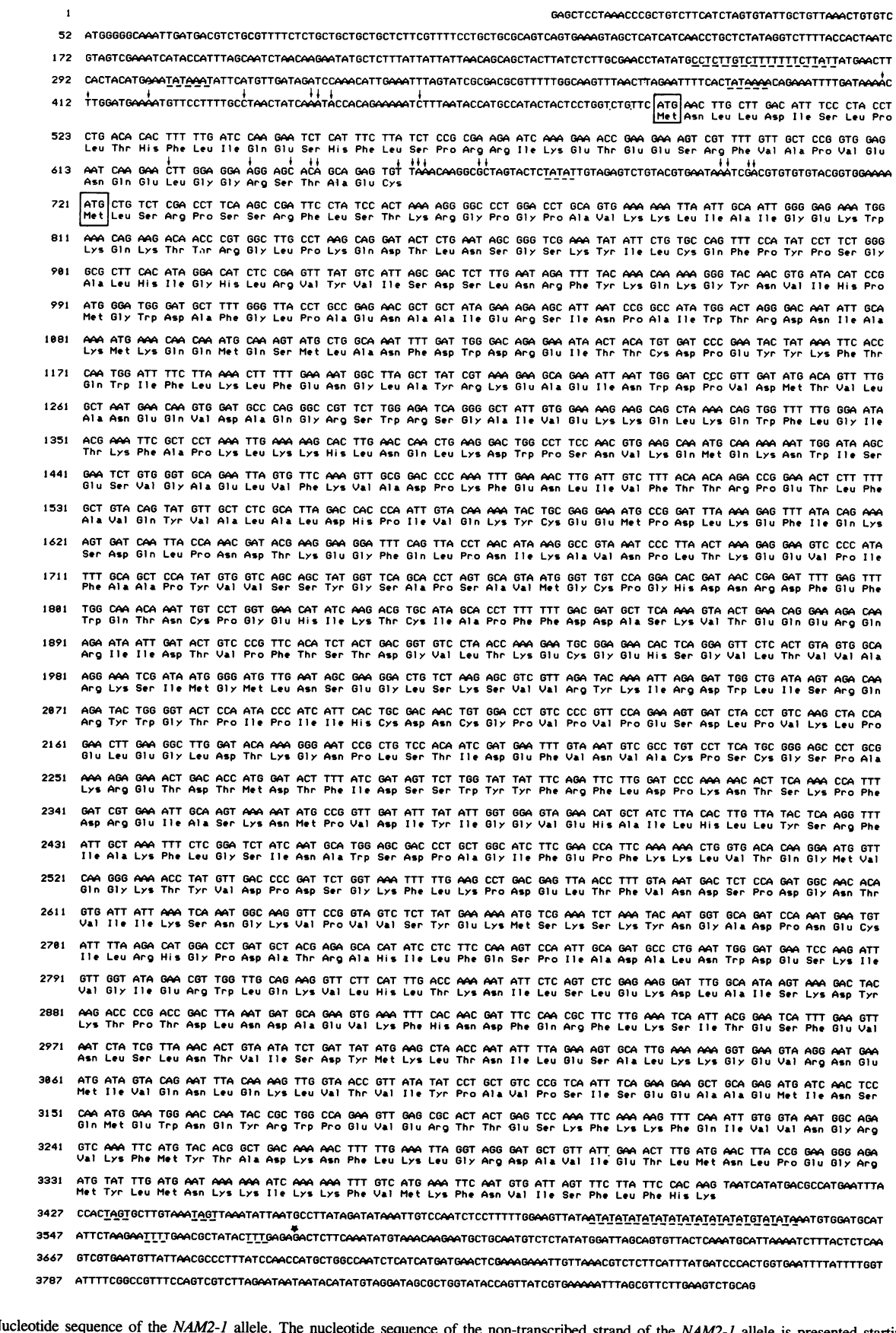

Fig. 2. Nucleotide sequence of the *NAM2-1* allele. The nucleotide sequence of the non-transcribed strand of the *NAM2-1* allele is presented starting at a *Sac*I site and ending at a *Pst*I site. The putative initiator codons are boxed, and the deduced amino acid sequence for the two ORFs are shown below the sequence. Arrowheads indicate the position of the major 5′ ends of the two RNAs transcribed from the gene, while the asterisk marks their common 3′ end as deduced from S1 mapping experiments (see Figures 3 and 4). Putative TATA boxes, a tract of pyrimidines which may form part of the promoter and a yeast 3′ end of messenger consensus sequence are underlined by dotted lines.

is not lethal to the cell but leads to the production of 100% cytoplasmic *petites* (Labouesse *et al.*, 1985); this shows that the expression of the gene is essential for the maintenance of an intact mitochondrial genome.

Here we present the complete nucleotide sequence of the *NAM2-1* suppressor allele and its flanking regions, the cloning and sequencing of the wild-type *nam2*$^+$ and two other suppressor alleles, and the precise mapping of the termini of the RNAs transcribed from the gene. From these data we conclude that two poly(A)$^+$ RNAs are transcribed from the gene and that they differ in their points of initiation. The longer RNA contains a short 51 codon open reading frame (ORF) followed by a long 894 codon ORF, while the shorter RNA only contains the long 894 codon ORF. The translation of the long ORF would give rise to a protein of 102 000 daltons, in which the same amino acid is changed in all three suppressor mutations sequenced. Finally, sequence comparisons revealed putative nucleic acid and nucleotide binding domains.

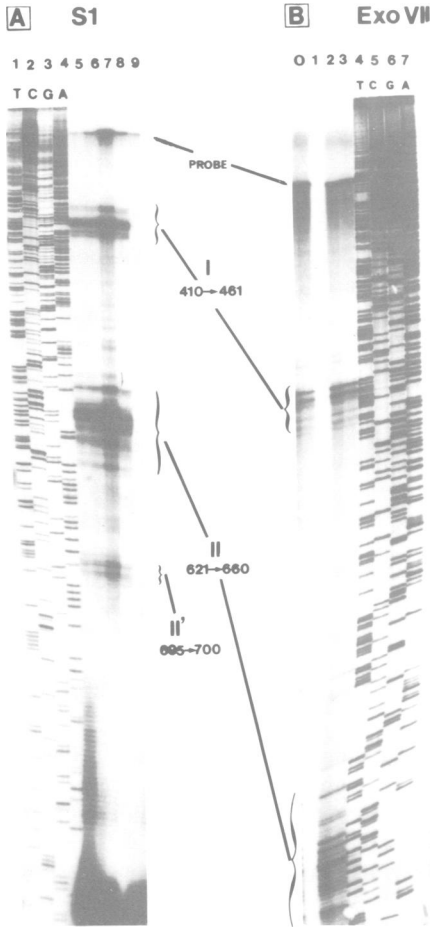

Fig. 3. Mapping of the 5′ ends of the two RNAs transcribed from the *NAM2* gene. A single-stranded radioactive probe was prepared from the recombinant M13 phage mp18−cr17 as described in Materials and methods and used to map the 5′ ends of the *NAM2* transcripts by S1 nuclease digestion **(panel A)** or *Exo*VII nuclease digestion **(panel B)**. **Panel A, lanes 1−4:** T, C, G, A dideoxy chain termination reactions carried out on the clone mp18−cr17 extended from the primer 01N2; **lanes 5−9:** S1 resistant products: **(5)** 1 μg poly(A)$^+$ RNA, enzyme at 20 U/ml final concentration; **(6)** 1 μg poly(A)$^+$ RNA, enzyme at 50 U/ml final concentration; **(7)** 2.5 μg poly(A)$^+$ RNA, enzyme at 50 U/ml final concentration; **(8)** 1 μg poly(A)$^+$ RNA, enzyme at 200 U/ml final concentration; **(9)** 10 μg *E. coli* tRNA, enzyme at 20 U/ml final concentration. **Panel B lanes 0−3:** *Exo*VII-resistant products: **(1)** 10 μg *E. coli* tRNA, 1 U enzyme; **(2)** 3 μg poly(A)$^+$ RNA, 1 U enzyme; **(0)** and **(3)** 1 μg poly(A)$^+$ RNA, 1 U enzyme; **lanes 4−7** are as **lanes 1−4** in **panel A**. The map coordinates indicated for the 5′ end of the two major RNAs (I, II and II′) refer to the sequence shown in Figure 2. II′ (bands around 695−700) is a minor RNA species non distinguishable by Northern blotting from the 2.85 kb RNA (bands around 621−660). Its significance is unclear but it would cover the entire large ORF.

Results and Discussion

Sequence of the NAM2-1 allele

We have determined the DNA sequence of both strands of the cloned *NAM2-1* allele. The restriction map and sequencing strategy are shown in Figure 1, while the DNA sequence of the 3888 nucleotide *Sac*I−*Pst*I fragment and the translation of the major ORFs are shown in Figure 2.

Analysis of the sequence shows a single long ORF with a methionine codon at nucleotide 721 and a TAA stop codon at nucleotide 3403. This ORF could code for a protein of 894 amino acids. The two remaining reading frames on the same strand and the three on the opposite strand contain between 43 and 96 stop

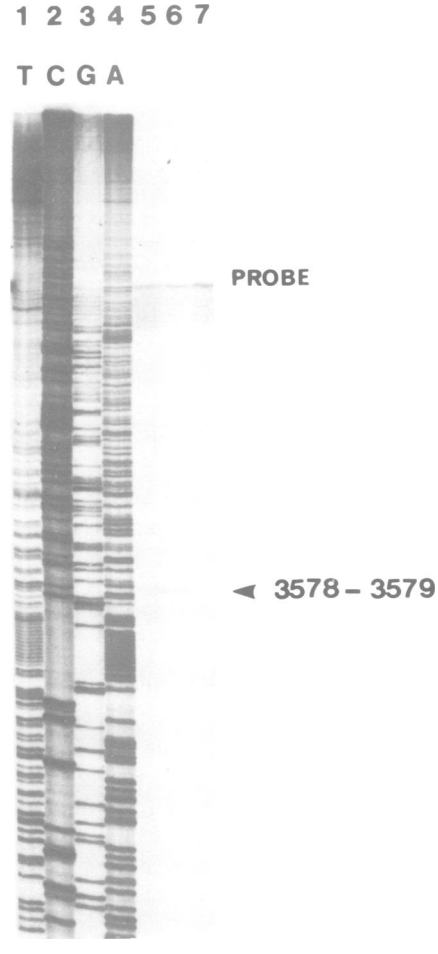

Fig. 4. Mapping of the 3′ end of the two RNAs transcribed from the *NAM2* gene. A single-stranded radioactive probe was prepared from the recombinant M13 phage mp10−cr6 as described in Materials and methods and used to map the 3′ ends of the *NAM2* transcripts by S1 nuclease digestion. In this experiment each band on the sequencing ladder contains 29 extra bases (derived from the universal primer and the polylinker); these should be added to the DNA fragment protected from digestion by S1 to determine the actual 3′ end of the RNAs. **Lanes 1−4:** T, C, G, A are dideoxy chain termination reactions carried out on the clone mp10−cr7 extended from the 17 mer 'universal primer'; **lanes 5, 6** and **7:** S1 nuclease resistant products: **(5)** 1 μg poly(A)$^+$ RNA, 50 U/ml; **(6)** 3 μg poly(A)$^+$ RNA, 50 U/ml; **(7)** 10 μg *E. coli* tRNA, 50 U/ml. The map coordinates refer to the sequence shown in Figure 2.

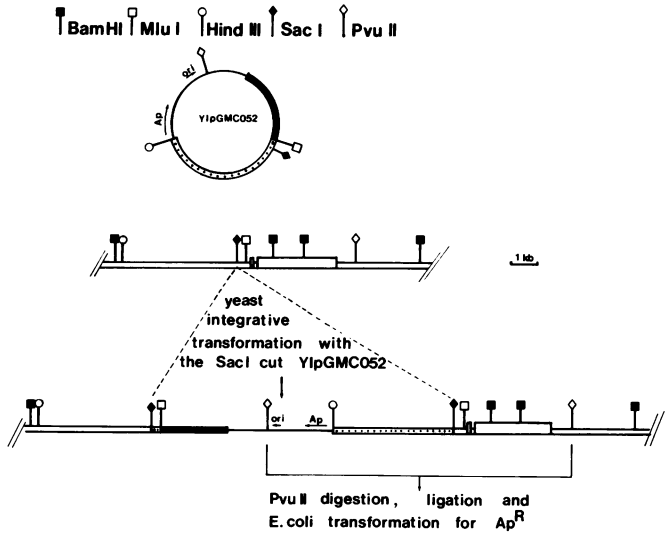

Fig. 5. Cloning the *NAM2-6* and *NAM2-7* alleles. The circle shows the structure of the plasmid YIpGMC052 which contains the 4.5 kb *Hin*dIII−*Mlu*I fragment (dotted open bar) derived from the region adjacent to the *NAM2* gene and is inserted between the *Hin*dIII and *Sal*I sites of the integrative plasmid YIpGMC019 which has a *LEU2* selection marker (black bar) (Labouesse *et al.*, 1985). Below is shown the restriction map of the *NAM2* chromosomal region. The lower line shows the expected structure of the *NAM2* chromosomal region after homologous integration of *Sac*I cut YIpGMC052. In the resulting leucine prototrophic transformants, the chromosomal *NAM2* gene (large open bar) is close to the origin of replication of the bacterial plasmid (ori) and the ampicillin resistance gene (Ap); after digestion of the DNA of such a transformant with *Pvu*II and ligation in dilute conditions, the mutated gene can be recovered upon transformation of *E. coli* and selection for ampicillin resistance.

codons and cannot code for a protein longer than 120 amino acids. The long ORF is preceded by a short ORF with a methionine codon at nucleotide 496 and a TAA stop codon at nucleotide 649; this ORF could code for a protein of 51 amino acids. The two ORFs are in the same reading frame but are separated by two additional stop codons. Finally, the gene was searched for the sequences GTATGT and TACTAAC, which have been characterized as the 5′ splice site and branch site respectively in yeast nuclear mRNA introns (for review, see Padgett *et al.*, 1986); these sequences were not found.

Mapping the termini of the NAM2 transcripts

Two RNAs of 3.05 kb and 2.85 kb which share a common 3′ end are transcribed from the *NAM2* gene (Labouesse *et al.*, 1985). In order to map the termini of these RNAs and to determine whether the smaller RNA is derived from the longer RNA by the excision of a small intron at the 5′ end, we performed S1 and *Exo*VII nuclease protection experiments. The results of 5′ end mapping using a probe homologous to the 810 first nucleotides of the gene are shown in Figure 3. Two clusters of DNA fragments are protected from digestion by S1 nuclease (Figure 3, panel A), each of them corresponding to one of the *NAM2* transcripts previously detected. The first cluster (I) is reproducibly composed of eight species which all map upstream of the first AUG codon of the short ORF (position 496), the two most abundant are at position 446 and 449. The second cluster (II and II′) is reproducibly composed of at least 12 species which all map upstream of the first AUG of the long ORF (position 721), the two most abundant are at positions 637 and 639. Essentially the same set of DNA fragments are protected from digestion by exonuclease VII (Figure 3, panel B).

The 3′ end of the *NAM2* RNAs was mapped using a single-stranded probe homologous to the non-transcribed strand of the small *Pst*I fragment (position 3882−3131). A single DNA fragment appears to be protected from S1 nuclease digestion; this maps the 3′ end of the two *NAM2* transcripts to nucleotide 3578 or 3579 (Figure 4). These results, together with data obtained previously using the method of Weaver and Weissman (Labouesse *et al.*, 1985) and the absence of the characteristic intron sequences, allow us to conclude that the *NAM2* gene does not contain an intron but codes for two mRNAs which are transcribed from separate promoters.

Signals for transcription and translation

In general the sequences which regulate the transcription of a gene are found upstream of the 5′ end of the messenger, and the site of initiation of transcription is thought to be determined by 'TATA boxes'. In yeast these are located 30 to 120 nucleotides upstream from the point of initiation. Because the 5′ ends of the yeast messengers are normally heterogeneous, each TATA box is believed to determine a set of 5′ termini (Hahn *et al.*, 1985). Two canonical TATAAA sequences are present upstream of the longer *NAM2* RNA at positions 304 and 387; the most abundant transcripts have their 5′ ends about 60 bases after the TATA box at nucleotide 387 and only a few minor transcripts start after the TATA box at nucleotide 304. No TATA box exists upstream of the shorter transcript, the TATA sequence found at position 669 being downstream of the 5′ end of the most abundant of the shorter *NAM2* transcripts. Only four minor transcripts (II′) start downstream from this TATA box, at positions 695−700. Similarly, in the gene encoding subunit VI of cytochrome oxidase, which also has a short ORF upstream of the main ORF, putative promoter elements are only found upstream of the short ORF (Wright *et al.*, 1984). A tract of pyrimidines (mainly Ts) which has been implicated in promoter functions (Struhl *et al.*, 1985) is present around position 275 just before the first TATA box.

The yeast transcription termination sequence, (T rich) . . .TAG . . . TAGT . . . (A−T rich) . . . TTT (Zaret and Sherman, 1982), is found between the stop codon (3403) of the long ORF and position 3578 (end of messenger).

In general, eukaryotic ribosomes initiate translation at the AUG closest to the cap site of the mRNA (Kozak, 1983). In yeast the preferred sequence surrounding the initiator codon is AXXAUGXXU for highly expressed genes (Dobson *et al.*, 1982). The sequence UUCAUGAAC which surrounds the first AUG on the longer transcript and opens the short ORF does not fit to this consensus, while the sequence surrounding the first AUG on the shorter transcript (AAAAUGCUG), which opens the long ORF, is a better fit. This is consistent with the hypothesis that one protein is translated from the smaller transcript containing the long ORF. Whether two proteins are translated from the longer transcript (one from the short and one from the long ORF) is unclear, but such an organization opens up the possibility of controlling the expression of the *NAM2* gene at the level of translation. It is interesting to note that in the above-mentioned case of the gene encoding subunit VI of cytochrome oxidase the AUG beginning the subunit VI coding sequence is also a better fit with the consensus than the AUG beginning the short upstream ORF (Wright *et al.*, 1984).

Cloning and sequencing of the nam2$^+$ allele and of the suppressor alleles NAM2-6 and NAM2-7

To determine the nature of the mutations enabling the *NAM2* gene to compensate for a bI4 maturase deficiency, the wild-type *nam2*$^+$ allele and two suppressor alleles, *NAM2-6* and *NAM2-7*, were cloned. The wild-type allele *nam2*$^+$ was cloned by screen-

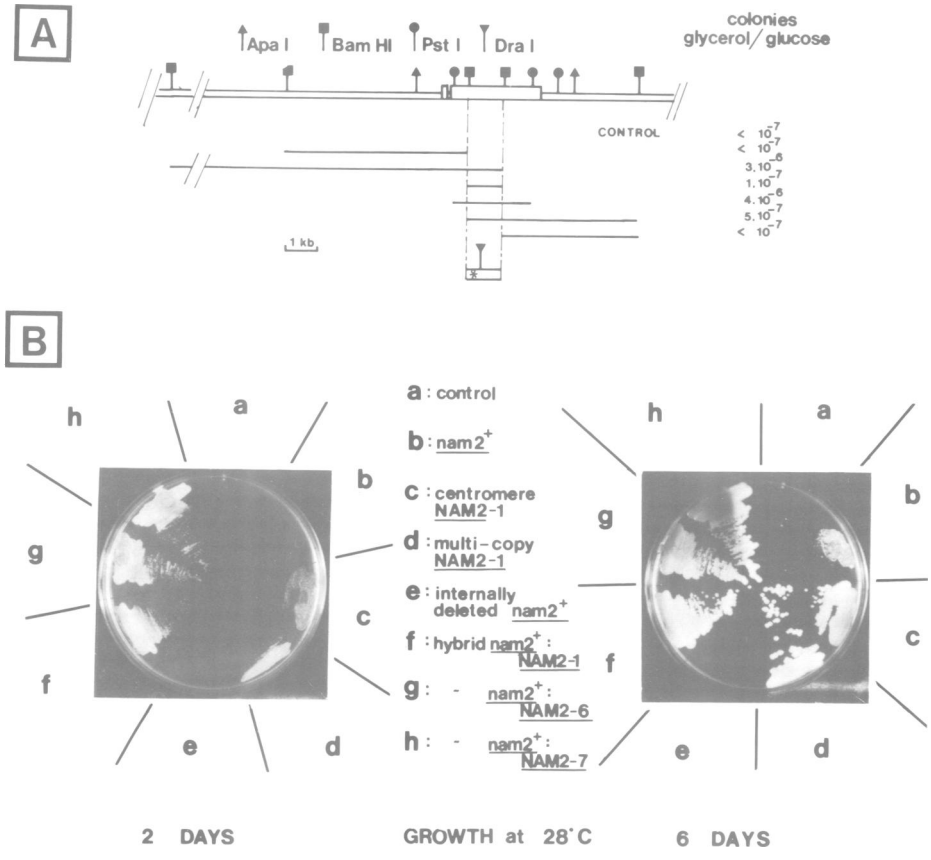

Fig. 6. Mapping of the *NAM2-1*, *NAM2-6* and *NAM2-7* mutations within the 1.05 kb *Bam*HI fragment. **Panel A:** the top line presents the restriction map of the *NAM2* chromosomal region. Below are shown the different restriction fragments derived from the plasmid YEpGMC001 carrying the *NAM2-1* allele (Labouesse *et al.*, 1985) which were subcloned in the episomal vector YEp13. These plasmids were used to transform the RNA maturase-deficient, respiratory negative, yeast strain CW01. Two independent transformants were grown in glucose minimal medium and, after being washed, cells were diluted and plated on glucose or glycerol. The figures on the right are the mean values of the number of glycerol colonies versus the number of glucose colonies obtained for each pair of transformants. **Panel B:** the figure shows growth on glycerol of the strain CW01 transformed with various *NAM2* derivatives. Negative controls **(a)** and **(e)** plasmids pEMBLYe30 and YEpGMC051 respectively show no growth; **(b)** multicopy plasmid YEpGMC050 carrying the wild-type *nam2*$^+$ allele shows no growth at 2 days and faint growth at 6 days; **(c)** centromeric plasmid YCpGMC103 carrying the *NAM2-1* allele shows faint growth after 2 days and strong growth after 6 days; **(d)** multicopy plasmid YEpGMC049 carrying the *NAM2-1* allele already displays strong growth after 2 days; **(f)**, **(g)** and **(h)** hybrid multicopy plasmids YEpGMCO53, YEpGMC058 and YEpGMC059 which are derived from the plasmid YEpGMC051 by the addition in the proper orientation of the central 1.05 kb *Bam*HI fragment of the alleles *NAM2-1*, *NAM2-6* and *NAM2-7* respectively; these display the same growth as with the plasmid YEpGMC049.

ing a genomic library constructed in the vector lambda EMBL4 (see Materials and methods). The suppressor alleles *NAM2-6* and *NAM2-7* were selected in a respiratory deficient strain lacking the bI4 intron and were cloned as depicted in Figure 5 using an integration eviction strategy (Orr-Weaver *et al.*, 1981; Roeder and Fink, 1980).

Before starting sequencing experiments we tried to map the position of the different suppressor mutations relative to the restriction map of the gene. To do this, we made use of the observation that autonomously replicating yeast plasmids can recombine with homologous chromosomal sequences (Falco *et al.*, 1983). Episomal plasmids carrying various portions of the *NAM2-1* gene were introduced into the yeast strain CW01 which has the wild-type *nam2*$^+$ allele and a bI4 maturase deficiency (Table III) and is unable to respire or grow on glycerol. If the *NAM2-1* mutation is introduced into the chromosomal *nam2*$^+$ gene by homologous recombination with one of the plasmids, respiratory competent glycerol-positive colonies should appear at rates significantly above the spontaneous reversion rate of the bI4 maturase mutation (Figure 6). The results of these mapping experiments indicate that the *NAM2-1* mutation is localized within the central 1.05 kb *Bam*HI fragment close to the 5′ *Bam*HI site.

Table I. Comparison of the DNA sequence of the wild-type *nam2*$^+$ and the three suppressor alleles *NAM2-1*, *NAM2-6* and *NAM2-7*

Allele	DNA sequence and predicted protein sequence		
nam2$^+$		1438	
	A T A	G G C	G A A
	I	**G**	**E**
		240	
NAM2-1	A T A	A G C	G A A
	I	**S**	**E**
NAM2-6	A T A	T G C	G A A
	I	**C**	**E**
NAM2-7	A T A	T G C	G A A
	I	**C**	**E**

The table shows the DNA sequence of the four alleles sequenced and the predicted amino acids around position 240 of the polypeptide chain. The sequence of the *NAM2-1* allele is taken from Figure 2. The sequence of the three other alleles was determined from M13 recombinant phages constructed by cloning the two *Bam*HI–*Dra*I fragments (positions 1236–1599 and positions 1599–2313) in both orientations (see Figure 6).

The location of the mutation was confirmed by making hybrid genes between the *nam2*$^+$ and the suppressor alleles. The *nam2*$^+$ allele (4.5 kb fragment *Apa*I, Figure 6) was cloned in pEMBLYe30 and a derivative was made by deleting the central 1.05 kb *Bam*HI fragment which was then replaced by the homologous fragment derived from either the *NAM2-1*, *NAM2-6* or *NAM2-7* allele. These plasmids were introduced into CW01 and growth on glycerol was monitored. The results are shown in Figure 6, panel B, and indicate that the *NAM2-1*, *NAM2-6* and *NAM2-7* mutations reside in the 1.05 kb *Bam*HI fragment, as all hybrid genes allow growth on glycerol at a level comparable to the *NAM2-1* allele.

We have determined the DNA sequence of the *nam2*$^+$, *NAM2-6* and *NAM2-7* alleles between the two *Bam*HI sites. In each case a single substitution was detected at nucleotide 1438 relative to the DNA sequence of the *NAM2-1* allele (Table I). As predicted (Figure 6) the *NAM2-1* mutation falls close to the *Bam*HI site at nucleotide 1236 and this probably explains why fragments starting at this position recombine poorly with the chromosome. The *NAM2-1* allele was isolated after ethyl methyl sulfonate treatment and is a transition (G to A), as are most mutations induced by this mutagen; the other two suppressor alleles occurred spontaneously and are transversions (G to T). The fact that the same codon has been mutated in three independent isolates demonstrates that the mutated residue (position 240 of the polypeptide chain) is of critical importance. In the wild-type gene it is a glycine, whereas in the mutant alleles it is a serine or a cysteine. Serine and cysteine have side chains of comparable size (CH_2OH versus CH_2SH) and both can form hydrogen bonds. Structure prediction programs place this residue in a region of the protein which is not structured, so it is not possible to predict the effect of the mutations on the conformation of the protein.

Figure 6 (sector B) also shows that the *nam2*$^+$ allele when placed on a multicopy vector confers slight but significant growth on glycerol to a maturase deficient strain. This indicates that the suppressor mutations do not create a new activity but enhance an activity which already exists in the wild-type protein. In general the suppressor action of *NAM2* seems to be sensitive to copy number since a centromeric *NAM2-1* allele (low copy number) is less effective than an episomal multicopy *NAM2-1* allele.

Table II. Codon usage in the long *NAM2* ORF

	T	C	A	G	
T	28	14	16	6	T
	18	6	11	6	C
	21	13	0	0	A
	21	4	0	19	G
C	5	20	7	5	T
	7	9	9	3	C
	7	14	22	4	A
	10	11	12	0	G
A	30	18	30	12	T
	10	11	17	11	C
	20	11	58	17	A
	24	6	22	6	G
G	18	19	34	12	T
	14	6	17	9	C
	15	21	45	14	A
	14	3	11	11	G

The bias in the codon usage was determined using the rules of Bennetzen and Hall (1982). The value of 0.044 is typical of proteins expressed at a low level.

Nature of the NAM2 encoded protein

The localization of the three suppressor mutations within the main ORF and the demonstration that all three create single amino acid replacements at residue 240 is a strong indication that the translation product of this ORF is the principal effector of the *NAM2* function(s). The translation of the main ORF would generate a protein of 102 000 daltons, containing mainly polar residues with a slight excess of basic over acidic residues (111 Lys plus Arg against 101 Asp plus Glu). The codon usage in the 894 codons of the main ORF (see Table II) is typical of genes expressed at a low level (Bennetzen and Hall, 1982), and this is consistent with the poor fit of the initiator AUG region to the consensus described for highly expressed genes. The translation product of the short ORF would be rather hydrophobic and we have not been able to detect any homology between this ORF and sequences in available databanks.

As the *NAM2* gene is required for processes occurring inside the organelle, the most likely subcellular location of the *NAM2* protein is intramitochondrial. Mitochondrial address sequences reside in the NH_2 terminal section of the protein and are enriched for Arg, Leu and Ser, have few Asp, Glu, Val and Ile and may form amphiphilic sequences (Roise *et al.*, 1986; Von Heijne, 1986). Inspection of the first 30 amino acids shows that the *NAM2* protein contains all the information typical of a protein imported into the mitochondrial matrix.

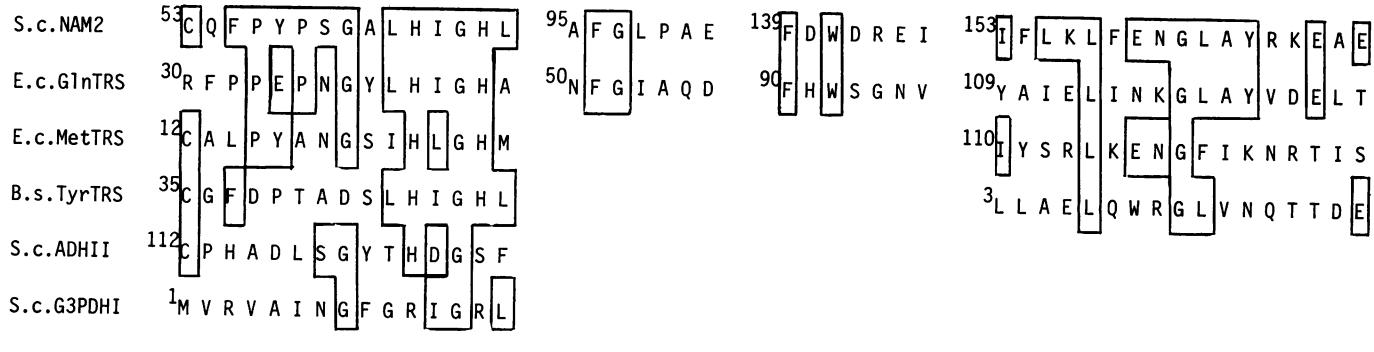

Fig. 7. Putative nucleotide binding domain of the *NAM2* protein. The figure shows a region homologous to a nucleotide binding domain previously identified in other proteins (Rossman *et al.*, 1974). The sequence of the glutaminyl−tRNA synthetase (*E. coli* GlnTRS) was taken from Hoben *et al.* (1982), of the methionyl−tRNA synthetase (*E. coli* MetTRS) from Barker *et al.* (1982), of the tyrosyl−tRNA (*B. stearothermophilus* TyrTRS) synthetase from Winter *et al.* (1983), of alcohol dehydrogenase II (*Saccharomyces cerevisiae* ADHII) from Russell *et al.* (1983) and of glyceraldehyde-3-phosphate dehydrogenase (*S. cerevisiae* G3PDHI) from Holland and Holland (1980). Identical residues are boxed.

Homologies between NAM2 and other proteins. The predicted amino acid sequence of the *NAM2* ORF was used to search for homologies with protein sequences stored in databanks. In this way, two regions of the *NAM2* protein were found to be homologous with other proteins. First, we have detected a strong homology with a number of different proteins which bind a nucleotide using a Rossmann fold (Rossmann *et al.*, 1974). Second, we have found a suggestive homology with proteins which bind nucleic acids through a presumed finger metal binding domain (Miller *et al.*, 1985; Berg, 1986).

Figure 7 shows the match with the nucleotide binding domain of proteins such as dehydrogenases and aminoacyl tRNA-synthetases. This domain consists of a parallel β sheet formed by three extended polypeptide strands connected in most cases by two α-helices (Rossman *et al.*, 1974). A number of residues are conserved among proteins folded in this way. Two blocks of homology with these proteins are found at the beginning of the *NAM2* protein around positions 60 and 160 of the polypeptide chain; the strongest match is with the *Escherichia coli* glutamine tRNA synthetase. Thus conservation of residue type and position suggests that the *NAM2* protein may bind a nucleotide.

Figure 8 shows the match with the nucleic acid binding domains of a few proteins such as retroviral low mol. wt proteins processed from the *gag* polyprotein, aminoacyl tRNA-synthetases and four intron encoded proteins from the mitochondria of fungi, one of which has a RNA−maturase function (Carignani *et al.*, 1983). The homology consists of a complete conservation of the motif Cys-X_2-Cys followed by a less stringent but significant conservation of several other residues present in *NAM2*. Many other proteins known to interact with nucleic acids possess related motifs (Berg, 1986) which are believed to chelate a Zn atom (Miller *et al.*, 1985). Between positions 462 and 506 the *NAM2* protein presents two repeated Cys-X_2-Cys units which are separated by 37 amino acids and which could possibly form a metal binding domain. Three other motifs Cys-X_2-His or Cys-X_3-His are also found at positions 350, 365 and 399. Interestingly the mutations are remote from the two domains revealed by computer aided comparisons, thus it is unlikely that the suppressor function is simply the result of a changed affinity for a nucleotide or nucleic acid.

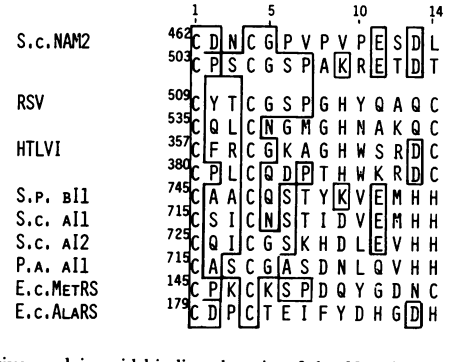

Fig. 8. Putative nucleic acid binding domain of the *NAM2* protein. The figure shows the homology with sequences from Rous sarcoma virus (RSV, Schwartz *et al.*, 1983), human T-cell leukemia virus (HTLV I, Copeland *et al.*, 1983) and four class II mRNA maturases: *Schizosaccharomyces pombe* bI1 from Lang *et al.* (1985); *S. cerevisiae* aI1 and aI2 from Bonitz *et al.* (1980); *P. anserina* aI1 from Osiewacz and Esser (1984); *E. coli* methionyl tRNA-synthetase (*E. coli* MetRS) from Barker *et al.* (1982) and *E. coli* alanyl tRNA-synthetase (*E. coli* AlaRS) from Putney *et al.* (1981). Identical residues are boxed.

Possible functions of the NAM2 protein. To understand the function of *NAM2*, one has to explain how the suppressor activity of RNA splicing deficiencies is acquired and why the gene is essential for mitochondrial DNA integrity. We show here that the suppressor activity of *NAM2-1* . . . *NAM2-7* results from the replacement of glycine at position 240 by serine or cysteine. We have previously shown (Dujardin *et al.*, 1983) that this activity requires the protein encoded by the intron aI4. Both findings can be interpreted in two ways: either the mutated *NAM2* protein modifies the synthesis and the primary amino acid sequence of the aI4 protein which has a latent RNA maturase activity (Dujardin *et al.*, 1982), or the mutated *NAM2* protein interacts post-translationally with the normal aI4 protein and induces the latent maturase activity. We shall illustrate the two alternatives by more specific models.

In the first model *NAM2* would code for a mitochondrial aminoacyl tRNA-synthetase. As a result of amino acid replacements at the position 240, this synthetase would produce a certain rate of misacylation. This degree of misacylation would not be sufficient to hamper the overall synthesis of normal mitochondrial proteins (cells carrying the mutated *NAM2-1* allele display normal growth and do not produce abnormal levels of *rho*$^-$ petites, data not shown), but would produce sufficient specific translation errors to render active the normally inactive aI4 protein. It should be recalled that the mitochondrial suppressor of bI4 maturase deficiencies, *mim2-1*, is a single base substitution in the middle

Table III. List of yeast strains

Name	Nuclear genotype	Mitochondrial genotype	Reference or origin
AB1-4A/8/55	*a his4 nam2*$^+$	*rho*^{0}	Groudinsky *et al.* (1981)
S912	*a his4 NAM2-1*	*rho*$^+$ *box7*-G1659	Groudinsky *et al.* (1981)
GRF18/50	α *his3-11,15 leu2-3,112 can*^{R} *nam2*$^+$	*rho*^{0}	Labouesse *et al.* (1985)
W303-1B/50	α *his3-11,15 leu2-3,112 ade2-1 ura3-1 trp1-1 can1-100 nam2*$^+$	*rho*^{0}	R.Rothstein (personal communication)
CK1000	*a his3 kar1-1 nam2*$^+$	*rho*$^+$ *box7*-V328	Labouesse *et al.* (1985)
CW01	α *his3-11,15 leu2-3,112 ade2-1 ura3-1 trp1-1 can1-100 nam2*$^+$	*rho*$^+$ *box7*-V328	This study
HM51/24-1A	*a ade1 NAM2-6*	*rho*$^+$ intron free *cob-box* gene	Labouesse *et al.* (1985)
HM54-12A	*a ade1 NAM2-7*	*rho*$^+$ *box7*-G1659	Labouesse *et al.* (1985)
HM182-15B	α *ade1 leu2 NAM2-6 can*^{R}	*rho*$^+$ intron free *cob-box* gene	This study
HM183-1A	α *ade1 leu2 ura3-1 can*^{R}	*rho*$^+$ *box7*-G1659	This study

Strain CW01 is a cytoductant with the nuclear genotype of W303-1B obtained by crossing W303-1B/50 with CK1000. Strains HM182-15B and HM183-1A are the meiotic products of crosses involving HM51/24-1A with GRF18/50 and HM54-12A with W303-1B/50 respectively.

of the aI4 ORF, which results in the replacement of a glutamic acid by lysine (Dujardin *et al.*, 1982). By analogy the *NAM2* gene could code for the mitochondrial glutamine tRNA-synthetase, and replacement of the glycine at position 240 could induce misacylation with lysine. Some misacylation would have to occur already with the wild-type gene since an increased gene dosage due to a multicopy plasmid results in a slight resumption of growth (Figure 6). Unfortunately nothing is known about the misacylation performed by mitochondrial tRNA-synthetases. In *E. coli* a mischarging phenotype conferred by a mutation in an aminoacyl tRNA synthetase gene has been described (Inokuchi *et al.*, 1984).

In the second model, activation by the mutated *NAM2* protein could result from post-translational alterations of the aI4 protein. Some maturases undergo proteolytic cleavages in relation to their RNA-splicing activities (De la Salle *et al.*, 1982; Carignani *et al.*, 1983) and one could hypothesize that the *NAM2* protein is involved in such a process. Alternatively, the *NAM2* protein could be a part of the mitochondrial RNA splicing machinery and act in conjunction, via protein–protein interactions, with the aI4 product. This interaction would be inefficient (detectable only with an increased gene dosage, see Figure 6) with glycine at position 240 and efficient when this residue is replaced by serine or cysteine.

At present we do not possess enough information to allow a clear rejection or acceptance of any of these hypotheses. The primary sequences of twelve tRNA-synthetases charging different amino acids have been established in various organisms. Breton *et al.* (1986) have recently described four regions of homology (regions A, B, C, D) characteristic of a number of tRNA-synthetases. Only one of these (region A) is found in the *NAM2* protein and it corresponds to the nucleotide binding domain discussed above, which is also present in proteins other than tRNA-synthetases. Furthermore, no homology between *NAM2* and any known tRNA-synthetase is observed outside this region, although it is well established that yeast mitochondrial tRNA-synthetases display strong homologies to their cytosolic or bacterial counterparts which charge the same amino acid (Myers and Tzagoloff, 1985; Pape *et al.*, 1985; Natsoulis *et al.*, 1986). Taken at face value, these comparisons argue against the hypothesis that *NAM2* could be one of the mitochondrial tRNA-synthetases charging alanine, aspartic acid, glutamine, glutamic acid, phenylalanine, glycine, histidine, isoleucine, methionine, threonine, tryptophan or tyrosine. However, the final proof can be obtained only by demonstrating that a strain carrying an inactivated *NAM2* gene is (or is not) able to aminoacylate a given mitochondrial tRNA. Such experiments are in progress.

The second aspect of the function of *NAM2* is that it is essential for the integrity of the mitochondrial genome: disruption or deletion of the gene leading to a quantitative formation of cytoplasmic petites (Labouesse *et al.*, 1985). It is well known that a large fraction of all nuclear *pet*$^-$ mutants are in reality double mutants that carry large mitochondrial deletions in addition to the nuclear mutation ('double petites': *pet*$^-$ *rho*$^-$ or *pet*$^-$ *rho*^{0}, Chen *et al.*, 1950; Sherman and Slonimski, 1964; Schweizer *et al.*, 1977). The reason for the disintegration of the mitochondrial DNA in such mutants is not known. Mitochondrial protein synthesis appears to be a prerequisite for the maintenance of the *rho*$^+$ genome, but the reason for this remains obscure (Williamson *et al.*, 1971; Myers *et al.*, 1985). Thus, the hypothesis that *NAM2* codes for a protein essential in mitochondrial protein synthesis is not at odds with all available data. However, it is also conceivable that a severe deficiency in mitochondrial RNA splicing could also lead to the formation of *petites*. It should be remembered, however, that in mitochondria all nucleic acid 'transactions' occur in the same cellular compartment, so that the involvement of an individual protein in more than one type of 'transaction' is quite possible. It is already known that the bI4 encoded protein is involved in RNA splicing and DNA recombination (Banroques *et al.*, 1986; Kotylak *et al.*, 1985). Thus the *NAM2* protein could also be involved in several processes such as translation and RNA splicing.

Materials and methods

Yeast strains and media

The yeast strains used are listed in Table III; media and genetic methods are as described in Dujardin *et al.* (1980).

E. coli strains, media, phages and plasmids

Strains JA221 (*recA*$^-$, *hsdR*$^-$, *M*$^+$, *leu*B6, *trp*E5, *lac*Y) and MC1061 (F$^-$, *ara*D138, (*ara*, *leu*)7696, *lac*Y74, *gal*U$^-$, *gal*K$^+$, *hsr*$^-$, *hsd*R$_k^-$, *hsd*M$_k^-$, *str*A) were used for standard cloning experiments. Strains JM101 (*lac pro AB*), *thi*, *sup*E, F' *tra*D36, *pro* AB, *lac* I^Q, *lac*Z M15) and Q358 (*hsd*R$_k^-$, *hsd*M$_k^-$, *sup* E, 80^r) were used for the propagation of M13 and lambda phages, respectively. M13 vectors mp10, mp11, mp18 and mp19 were obtained from Pharmacia, the lambda vector EMBL4 (Frischauf *et al.*, 1983) was obtained from Drs F.Caron and B.Guiard, and the plasmid pEMBLYe30 (Baldari and Cesareni, 1985) was obtained from Dr Cesareni. The recombinant plasmids containing *NAM2-1* were as previously described (Labouesse *et al.*, 1985), other plasmids were constructed by standard methods.

DNA sequencing

DNA sequencing was performed using the chemical techniques of Maxam and Gilbert and the chain termination techniques of Sanger, in conjunction with the M13 phage system of Messing. Recombinant M13 phages were obtained by random subcloning of the central 2.35 kb *Pst*I fragment (see Figure 1) after sonication or by cloning defined restriction fragments. Enzymes were purchased from Appligene and Boehringer Mannheim, [α-^{35}S]dATP 480 Ci/mmol and [γ-^{32}P]-ATP 3000 Ci/mmol were obtained from Amersham International.

Cloning of the nam$^+$ *allele*

Genomic DNA from the yeast strain AB1-4A/8/55 (which is isonuclear to S912 from which the allele *NAM2-1* had been cloned) was partially digested by *Sau*3A; fragments of 15–20 kb were purified and ligated into *Bam*HI-cut lambda EMBL4. After *in vitro* packaging the library was screened by plaque hybridization using a nick-translated probe made from the plasmid pGMC031, which is an *Apa*I fragment containing the entire *NAM2-1* allele cloned in pBR322 (Labouesse *et al.*, 1985) (*in vitro* packaging extracts were kindly donated by Drs F.Caron and B.Guiard). Out of 8000 lambda plaques screened, three appeared to hybridize with the probe; one of these contained a 16 kb insert with the complete *nam2*$^+$ allele.

Cloning of the NAM2-6 and NAM2-7 alleles

The *NAM2-6* and *NAM2-7* alleles were cloned from strains HM182-15B and HM183-1A respectively (see Table III), using an integration–eviction strategy (Roeder and Fink, 1980) with the plasmid YIpGMC052 as the integrative plasmid (Figure 5).

S1 nuclease and ExoVII nuclease mappping

Single-stranded radiolabelled DNA probes were prepared from M13 recombinant phage DNA (Miller, 1984). Two recombinant phages were used: mp18–cr17 contains the non-transcribed strand of the 1232 base long *Bam*HI–*Sac*I fragment (Fragment A, Figure 1) cloned in mp18; and mp10–cr6 contains the non-transcribed strand of the 751 base long *Pst*II fragment (Fragment B, Figure 1) cloned in mp10. The phage mp18–cr17 was elongated from the 18-mer primer 01N2 5'-CCATTTCTCCCCAATTGC-3' which is homologous to the non-transcribed strand of the *NAM2* gene from positions 793–810, and was kindly made by Dr C.Jacq. The phage mp10–cr6 was elongated from the 17-mer universal primer 5'-GTAAAACGACGGCCAGT-3' (Biolabs). The resulting double-stranded DNAs were cleaved with the restriction enzyme *Sac*I prior to gel purification of the radioactive strand. Hybridization, S1 and *Exo*VII nuclease digestions were performed essentially as described by Miller (1984). Nuclease resistant products were separated by electrophoresis through 8 M urea/6% polyacrylamide gels.

Miscellaneous

Yeast cells were transformed using the lithium chloride procedure (Ito *et al.*, 1983). DNA and RNA preparations, Southern and Northern blotting were performed as described previously (Labouesse *et al.*, 1985).

Acknowledgements

This paper is dedicated to the memory of Hugues Blanc. We are grateful to Dr P.Pajot for the gift of the yeast strain carrying the *NAM2-7* allele. Sequence data analysis was done using computer programs provided by Dr Ian Giles and with the CITI2 system with the help of the French Ministère de la Recherche et de la Technologie; Dr J.L.Risler, J.Renowicki and M.O.Mosse are thanked for their help in the computer analysis. We also thank Dr C.Jacq for the oligonucleotide synthesis and B.Poirier for help in preparing the manuscript. This work was supported by grants from the CNRS, ATP Biologie Moléculaire du Gène and Ligue Nationale contre le Cancer. C.J.H. is the recipient of a European Community Biomolecular Engineering Training Contract.

References

Baldari,C. and Cesareni,G. (1985) *Gene*, **35**, 27–32.
Banroques,J., Delahodde,A. and Jacq,C. (1986) *Cell*, **46**, 837–844.
Barker,D.G., Ebel,J.-P., Jakes,R. and Bruton,C.J. (1982) *Eur. J. Biochem.*, **127**, 449–457.
Bennetzen,J.F. and Hall,B.D. (1982) *J. Biol. Chem.*, **257**, 3026–3031.
Berg,J.M. (1986) *Science*, **232**, 485–487.
Bonitz,S., Coruzzi,G., Thalenfeld,B., Tzagoloff,A. and Macino,G. (1980) *J. Biol. Chem.*, **255**, 11927–11941.
Breton,R., Sanfaçon,H., Papayannopoulos,I., Biemann,K. and Lapointe,J. (1986) *J. Biol. Chem.*, **261**, 10610–10617.
Carignani,G., Groudinsky,O., Frezza,D., Schiavon,E., Bergantino,E. and Slonimski,P.P. (1983) *Cell*, **35**, 733–742.
Cech,T.R. and Bass,B.L. (1986) *Annu. Rev. Biochem.*, **55**, 599–630.
Chen,P.Y., Ephrussi,B. and Hottinger,H. (1950) *Heredity*, **4**, 337–351.
Copeland,T.D., Orozszlan,S., Kalyanaranam,V.S., Sarngadharan,M.G. and Gallo,R.C. (1983) *FEBS Lett.*, **162**, 390–394.
Collins,R.A. and Lambowitz,A.M. (1985) *J. Mol. Biol.*, **184**, 413–428.
De la Salle,H., Jacq,C. and Slonimski,P.P. (1982) *Cell*, **28**, 721–732.
Dobson,M.J., Tuite,M.F., Roberts,N.A., Kingsman,A.J. and Kingsman,S.M. (1982) *Nucleic Acids Res.*, **10**, 2625–2637.
Dujardin,G., Pajot,P., Groudinsky,O. and Slonimski,P.P. (1980) *Mol. Gen. Genet.*, **179**, 469–482.
Dujardin,G., Jacq,C. and Slonimski,P.P. (1982) *Nature*, **298**, 628–632.
Dujardin,G., Labouesse,M., Netter,P. and Slonimski,P.P. (1983) In *Mitochondria 1983*. Schweyen,R.J. *et al.* (eds), W. de Gruyter, Berlin, pp. 233–250.
Falco,S.C., Rose,M. and Botstein,D. (1983) *Genetics*, **105**, 843–856.
Frischhauf,A.M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827–842.
Groudinsky,O., Dujardin,G. and Slonimski,P.P. (1981) *Mol. Gen. Genet.*, **184**, 493–503.
Hahn,S., Hoar,E.T. and Guarente,L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8562–8566.
Hoben,P., Royal,N., Cheung,A., Yamao,F., Biemann,K. and Söll,D. (1982) *J. Biol. Chem.*, **257**, 11644–11650.
Holland,J.P. and Holland,M.J. (1980) *J. Biol. Chem.*, **255**, 2596–2605.
Inokuchi,H., Hoben,P., Yamao,F., Ozeki,H. and Söll,D. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5076–5080.
Ito,H., Fukuda,M., Murata,K. and Kimma,A. (1983) *J. Bacteriol.*, **153**, 163–168.
Kotylak,Z., Lazowska,J., Hawthorne,D. and Slonimski,P.P. (1985) In *Achievements and Perspectives in Mitochondrial Research*. Quagliariello,E. and Palmieri,P. (eds), Elsevier, Amsterdam, Vol. **II**, pp. 1–20.
Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1–45.
Kruger,K., Grabowski,P.J., Zaug,A.J., Sands,J., Gottschling,D.E. and Cech,T.R. (1982) *Cell*, **31**, 147–157.
Labouesse,M., Netter,P. and Schroeder,R. (1984) *Eur. J. Biochem.*, **144**, 85–93.
Labouesse,M., Dujardin,G. and Slonimski,P.P. (1985) *Cell*, **41**, 133–143.
Lang,B.F., Ahne,F. and Bonen,L. (1985) *J. Mol. Biol.*, **184**, 353–366.
Last,R.J., Stavenhagen,J.B. and Woolford,J.L., Jr. (1984) *Mol. Cell. Biol.*, **4**, 2396–2405.
Lazowska,J., Jacq,C. and Slonimski,P.P. (1980) *Cell*, **22**, 333–348.
Lee,M.G., Young,R.A. and Beggs,J.D. (1984) *EMBO J.*, **3**, 2825–2830.
McGraw,P. and Tzagoloff,A. (1983) *J. Biol. Chem.*, **258**, 9459–9468.
Miller,A.M. (1984) *EMBO J.*, **3**, 1061–1065.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1615.
Myers,A.M., Pape,L.K. and Tzagoloff,A. (1985) *EMBO J.*, **4**, 2087–2092.
Myers,A.M. and Tzagoloff,A. (1985) *J. Biol. Chem.*, **260**, 15371–15377.
Natsoulis,G., Hilger,F. and Fink,G.R. (1986) *Cell*, **46**, 235–243.
Orr-Weaver,T.L., Szostak,J.W. and Rothstein,R.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6354–6358.
Osiewacz,H.D. and Esser,K. (1984) *Curr. Genet.*, **8**, 299–305.
Padgett,R.A., Grabowski,P.J., Konarska,M.M., Seiler,S. and Sharp,P.A. (1986) *Annu. Rev. Biochem.*, **55**, 1119–1150.
Pape,L.K., Koerner,T.J. and Tzagoloff,A. (1985) *J. Biol. Chem.*, **260**, 15362–15370.
Pillar,T., Lang,B.F., Steinberger,I., Vogt,B. and Kaudewitz,F. (1983) *J. Biol. Chem.*, **258**, 7957–7959.
Putney,S.D., Royal,N.J., de Vegvar,H.N., Herlihy,W.C., Biemann,K. and Schimmel,P. (1981) *Science*, **213**, 1497–1501.
Roeder,G.S. and Fink,G.R. (1980) *Cell*, **21**, 239–249.
Roise,D., Horvarth,S.J., Tomich,J.M., Richards,J.H. and Schatz,G. (1986) *EMBO J.*, **5**, 1327–1334.
Rossmann,M.G., Moras,D. and Olsen,K.W. (1974) *Nature*, **250**, 194–199.
Russell,D.W., Smith,M., Williamson,V.M. and Young,E.T. (1983) *J. Biol. Chem.*, **258**, 2674–2682.
Schwartz,D.E., Tizard,R. and Gilbert,W. (1983) *Cell*, **32**, 853–869.
Schweizer,E., Demmer,W., Holzner,V. and Tahedl,H.W. (1977) In *Mitochondria 1977*. Bandlow,W. *et al.* (eds), W. de Gruyter, Berlin, pp. 91–106.
Sherman,F. and Slonimski,P.P. (1964) *Biochim. Biophys. Acta*, **90**, 1–15.
Struhl,K., Chen,W., Hill,D.E., Hope,I.A. and Oettinger,M.A. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 489–503.
Von Heijne,G. (1986) *EMBO J.*, **5**, 1335–1342.
Williamson,D.H., Maroudas,M.G. and Wilkie,D. (1977) *Mol. Gen. Genet.*, **111**, 209–223.
Winter,G., Koch,G.E.L., Hartley,B.S. and Barker,D.G. (1983) *Eur. J. Biochem.*, **132**, 383–387.
Wright,R.M., Ko,C., Cumsky,M.G. and Poyton,R.O. (1984) *J. Biol. Chem.*, **259**, 15401–15407.
Zaret,K.S. and Sherman,F. (1982) *Cell*, **28**, 563–573.

Received on December 3, 1986; revised on January 7, 1987