The *Neurospora* mitochondrial genome: the region coding for the polycistronic cytochrome oxidase subunit I transcript is preceded by a transfer RNA gene

Hans de Vries, Peter Haima, Marja Brinker and Jenny C. de Jonge

Laboratory of Physiological Chemistry, State University at Groningen, Bloemsingel 10, 9712 KZ Groningen, The Netherlands

Received 13 November 1984

We have sequenced a 682 bp fragment of *Neurospora crassa* mitochondrial DNA to complete the sequence between the gene for cytochrome *b* and the unassigned reading frame, URF U. The sequence contains a gene for a cysteine tRNA. The 5′ end of the 6 kb polycistronic transcript of cytochrome *c* oxidase subunit 1 is immediately downstream from this tRNA. This shows that also in fungal mitochondria tRNAs can be used as processing signals, whereas palindromic sequences containing double *Pst* I sites, also present in this region, are not used for processing.

Mitochondrial DNA *Fungus* Neurospora crassa *DNA sequence* *Cysteine tRNA* *Transcription*

1. INTRODUCTION

In *Neurospora* the gene for C.O.I is located on mitochondrial DNA [1–3] near the cob gene. Transcript mapping and sequencing of the C.O.I region of the mitochondrial DNA revealed that the only transcript detected contains the information of the C.O.I gene as well as of two additional protein genes, URF U [2,3] and URF N [4]. This led us [5] and Burger and Werner [4] to the preliminary conclusion that this transcript of about 5700 nucleotides is a polycistronic messenger RNA. The function of the proteins encoded by the URFs remains unknown.

We have sequenced the region upstream from URF U up to the place where the known sequence following the cob gene [6,7] ends and determined the position of the 5′-end of the presumptive C.O.I messenger by S1 protection analysis. On the basis of these results we discuss the typical aspects of transcription and RNA processing in *Neurospora* mitochondria.

Abbreviations: C.O.I, subunit I of cytochrome *c* oxidase; cob, cytochrome *b*; URF, unassigned reading frame; bp, basepair

2. MATERIALS AND METHODS

2.1. *Fragments and plasmids*

For the isolation of the fragments used for sequencing and S1 nuclease analysis we used the plasmid pBE3 [2], containing *Eco*RI-3 in pBR322. The appropriate subfragments were separated by agarose gel electrophoresis.

2.2. *DNA sequencing*

Fragments were labeled at their 5′-ends using [γ-^{32}P]ATP (Amersham) and polynucleotide kinase (Boehringer). Double- and single-strand sequencing was performed by the chemical method [8].

S1 nuclease mapping was performed as described by Osinga and Tabak [9].

2.3. *Computer analysis*

The DNA sequence was analyzed for tRNA structures by using a matrix-plot program designed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

by Dr Peter Terpstra (Biochemical Laboratory, State University Groningen).

3. RESULTS AND DISCUSSION

The distance between the coding sequence of the cytochrome *b* apoprotein gene [6,7] and the start of URF U is about 1170 bp, as indicated in fig.1. The part downstream from the cytochrome *b* gene has been sequenced up to the *Pst*I palindrome by other groups [6,7], whereas we reported the sequence of 550 bp upstream from URF U [5]. We have now sequenced the *Xho*I-*Hin*dIII fragment containing the remaining 515 nucleotides. The sequence strategy is given in fig.2. The sequence between the *Xho*I site and URF U is presented in fig.3. The most prominent features of this region are: a $tRNA^{cys}$ gene at nucleotide 122 to 189 (fig.4), the absence of further open reading frames and the presence of several strongly palindromic structures. In particular, the GC-rich double *Pst*I and *Hin*dIII palindromes, both types already encountered abundantly in *Neurospora* mitochondrial DNA [10], are conspicuous.

We have shown [5] that the long transcript containing the C.O.I message has its 5′-terminus in this region. Fig.5 shows the result of S1 mapping experiments, performed with 5′-end-labeled fragments containing the tRNA gene and surrounding sequences and total mitochondrial RNA. The length of the protected fragment in fig.5a is estimated at about 500 nucleotides. The few faint shorter bands may be artefacts of the procedure. Hence, the 5′-end of the C.O.I transcript roughly coincides with the 3′-end of the tRNA gene. Fig.5b shows a more accurate determination of the 5′-end

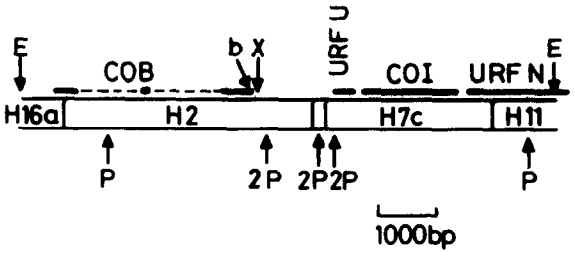

Fig.1. Map of the region of the *Neurospora* mitochondrial genome containing the cob and C.O.I genes. E, *Eco*RI; P, *Pst*I (2P, double *Pst*I site); H, *Hin*dIII; X, *Xho*I; b, *Bgl*II.

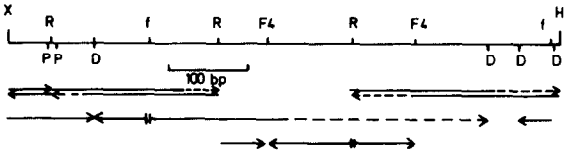

Fig.2. Sequencing strategy for the *Xho*I-*Hin*dIII fragment. The labelled sites, directions and extent to which sequences were read are indicated by the arrows. X, *Xho*I; H, *Hin*dIII; P, *Pst*I; R, *Rsa*I; f, *Hin*fI; F4, *Fnu*4H I; D, *Dde*I; h, *Hae*III; P_1, *Hin*P_1 I.

of the long transcript. Two hybrids are now visible. The least intense starts at nucleotide 194, which is 4 nucleotides behind the $tRNA^{cys}$ gene. The major band is localized at position 200. Hence, it appears that two initiation or processing sites are present immediately downstream from the $tRNA^{cys}$ gene. A similarity between these sites is that both contain the sequence UUAG, whereas the transcripts start at AG. These data show that the C.O.I transcript is delimited by tRNAs: this $tRNA^{cys}$ preceding the 5′-end and a $tRNA^{arg}$ immediately following its 3′-end [4]. This finding yields several clues to the processes of transcription and transcript processing in *Neurospora* mitochondria:

The polycistronic transcript URF U-C.O.I-URF N, presumably the mRNA for these 3 genes, starts immediately downstream from the $tRNA^{cys}$ gene. At present, no promoter sequences are known for *Neurospora* mitochondria, so we have to rely here on extrapolation from yeast. Since no sequence like the yeast mitochondrial ATATAAGTA box [9,11–13] is present at this site, we assume that the S1 hybrids indicate processing rather than initiation. Moreover, the mere fact that the long transcript starts immediately downstream from the $tRNA^{cys}$ strongly indicates that processing of a larger transcript occurs at this site.

Since it appears to be not a primary transcript, transcription has to start more upstream, possibly even before the cytochrome *b* gene.

The processing of long transcripts evidently does not always occur at the double *Pst*I palindromes which occur so ubiquitously throughout the *Neurospora* mitochondrial genome. This conclusion has been drawn by us [5] and by Burger and Werner [4]. Therefore, unlike the situation in *Saccharomyces cerevisiae* mitochondria but similar to

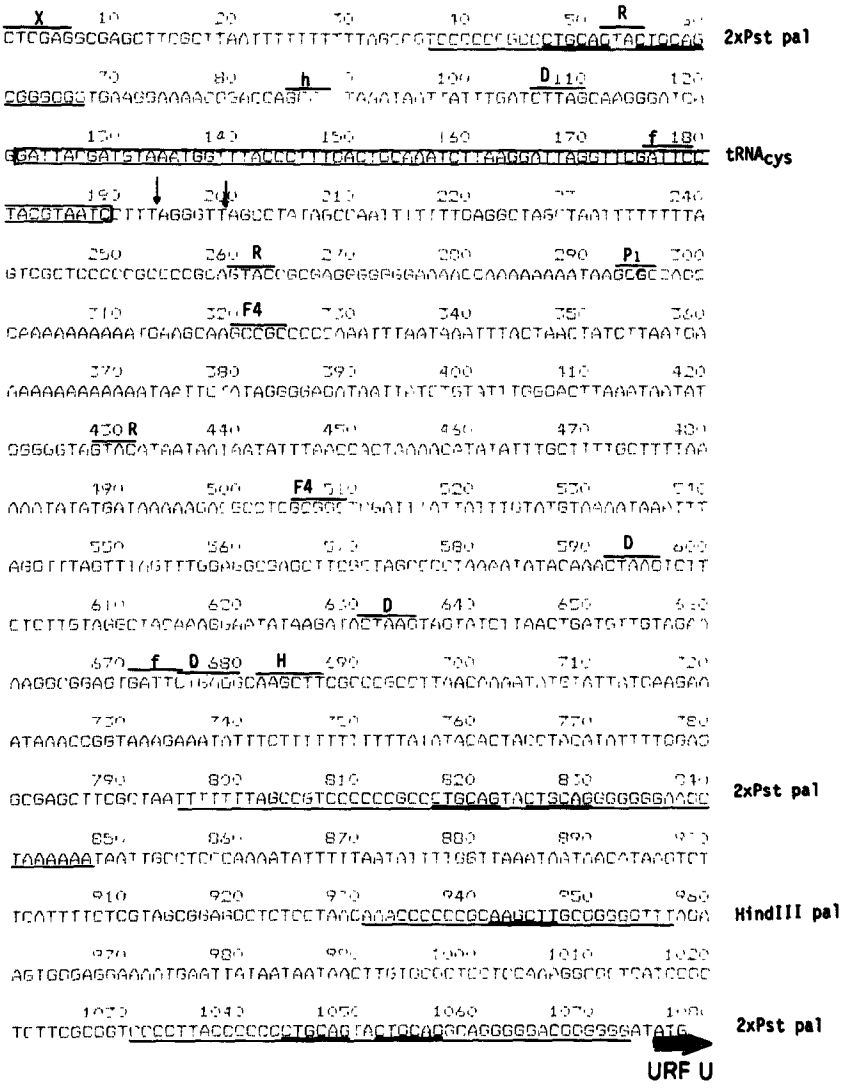

Fig.3. Nucleotide sequence of the region between the *Xho*I site and URF U. The nucleotide numbered 1 is at position −1525 when the T of the (TCA) codon for the first residue (Ser) of the mature C.O.I is taken as 1. Restriction sites (see fig.2) are given above the sequence. The $tRNA^{cys}$ sequence is boxed; palindromic regions are underlined. The sequence to the left of the *Hin*dIII site at 682 was taken from [5]. The arrows indicate the processing sites between tRNA and the long C.O.I transcript.

that in HeLa [14] and probably also in *Schizosaccharomyces pombe* [15] mitochondria, it is rather the tRNAs that are used as processing signals in *Neurospora* mitochondria. Nevertheless, it cannot be that tRNAs are essential to generate all separate transcripts. For example, the region of the mitochondrial genome in and around fragment *Eco*RI-4, contains at least 3 different genes, viz., for ATPase subunit 6 [16], for the MAL (ATPase proteolipid-like) gene product [17] and for C.O.II [18]. These genes are not surrounded by tRNA genes, nevertheless long transcripts encompassing all 3 genes as well as transcripts unique to these genes were found [5]. Hence, it appears that several transcript processing mechanisms are operative in *Neurospora* mitochondria, of which processing at tRNAs is one.

Other information drawn from the sequence in

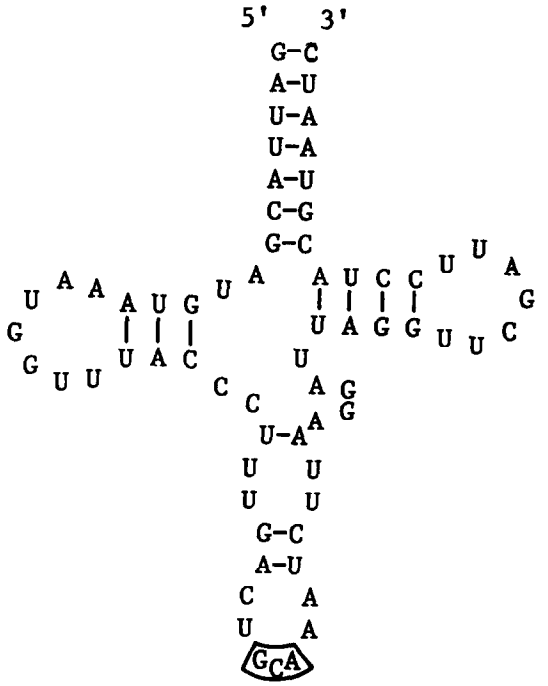

Fig.4. Secondary structure of the $tRNA^{cys}$ as deduced from the DNA sequence of its gene (fig.3).

fig.3 concerns the structure of the $tRNA^{cys}$ gene and the *Pst*I palindrome immediately downstream from the *Xho*I site:

The tRNA structure presented in fig.4 is rather orthodox for a mitochondrial tRNA, except for the anticodon stem: only 3 out of 5 base couples are in fact Watson-Crick paired. The anticodon loop is classical and permits the translation of the codons UGC and UGU into cysteine. Until now, no other $tRNA^{cys}$ gene has been uncovered in *Neurospora* mitochondria.

Helmer Citterich et al. [6] have described the sequence of the entire cytochrome *b* gene down to nucleotide 96 in our fig.3. Their conclusion was that the GC-rich palindrome following the *Xho*I site contains only a single *Pst*I site, which would have indicated the first aberrant *Pst*I palindrome to be found. On reading their paper, we have carefully rechecked our autoradiograms, and found a total of 4 base differences, all As or Gs. One of these differences concerns this palindrome: we find the now classical sequence CTGCAGTAC-TGCAG instead of CTACAGTACTGCAG. With the sequence reported by Burke et al. [7], extending to nucleotide 52 in our fig.3, we found only one difference, again a G-A discrepancy, at position 10.

4. CONCLUSIONS

Our results indicate that, contrary to our original expectation [5], the long polycistronic transcript for C.O.I, is not a primary transcript but is rather a processing product from an even larger RNA molecule. This larger molecule may even include the cytochrome *b* transcript, although we have no data pertinent to this point. It should be stressed that in Northern hybridization experiments we could not find any signal larger than the 6 kb C.O.I transcript. Hence, it may well be that processing of transcripts at tRNAs takes place before transcription termination, i.e., while the nascent transcript is still on the RNA polymerase.

The anatomy of the entire C.O.I region, as deduced from our present data on sequence and transcription (fig.6) and from [2–5], is: $tRNA^{cys}$ – set of GC-rich palindromes – URF U – C.O.I gene – URF N – *Pst*I palindrome – $tRNA^{arg}$. After transcription and cutting at the tRNAs a polycistronic messenger RNA remains.

At present we are investigating mitochondrial transcription in *Neurospora* to determine the number and structure of the transcription initiation sites.

ACKNOWLEDGEMENTS

This research was supported in part by a grant to A.M.K. from the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO). We wish to thank Dr Peter Terpstra for kindly making his matrix plot program available to us, Dr Etienne Agsteribbe for his cooperation with the mitochondrial RNA and his advice, Dr Ab Kroon for his continuing interest, Rinske Kuperus for typing the manuscript, Harm Schokkenbroek for drawing the figures and Bert Tebbes for photography.

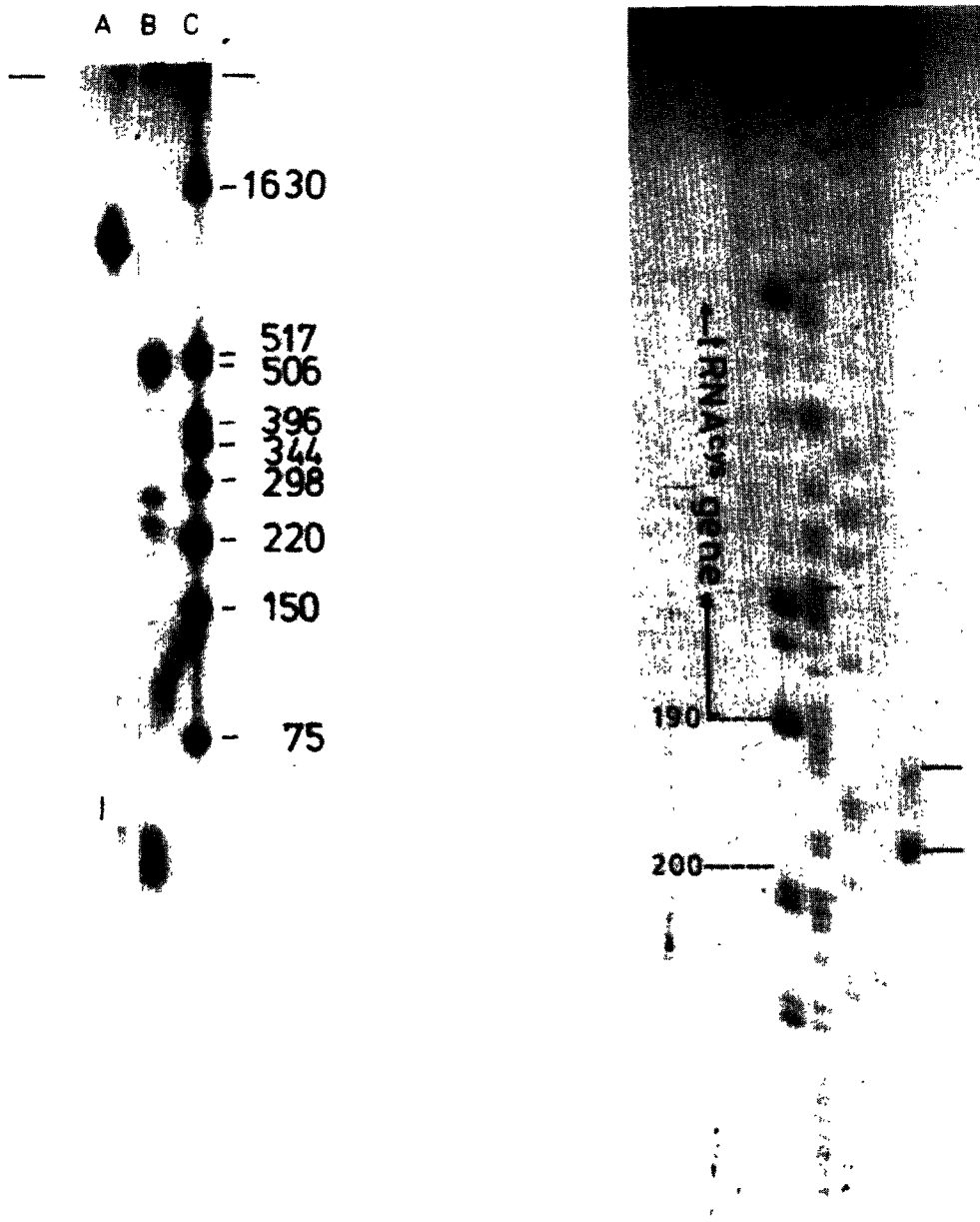

Fig.5. S1 nuclease mapping of the 5′-end of the C.O.I transcript. (a) A 5′-end-labeled fragment stretching from the *Bgl*II site indicated in fig.1 to the *Hind*III site at position 682 (lane A) was denatured, hybridized to 30 μg total mitochondrial RNA at 50.5°C and treated with 3000 U S1 nuclease (lane B). Lane C: *Hin*fI digest of pBR322. (b) A *Hae*III-*Hin*P_1 I fragment (positions 85–293 in fig.3) which was 5′-labeled at the *Hin*P_1 I site was used for S1 analysis and run along a sequence ladder of the same fragment. The length of the hybrids was calculated by subtracting 1.5 nucleotides [9]. Numbering is as in fig.3.

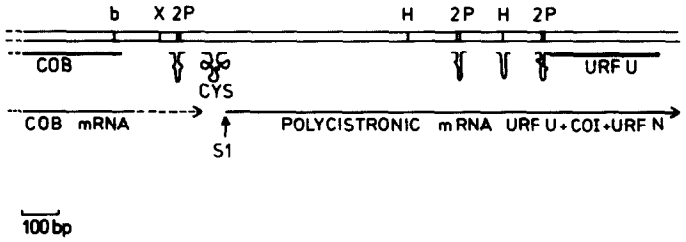

Fig.6. Schematic representation of the region between the cob and C.O.I gene.

REFERENCES

[1] Agsteribbe, E., Samallo, J., De Vries, H., Hensgens, L.A.M. and Grivell, L.A. (1980) in: The Organization and Expression of the Mitochondrial Genome (Kroon, A.M. and Saccone, C. eds) pp.51–59, Elsevier/North-Holland, Amsterdam.

[2] De Jonge, J.C. and De Vries, H. (1983) Curr. Genet. 7, 21–28.

[3] Burger, G., Scriven, C., Machleidt, W. and Werner, S. (1982) EMBO J. 1, 1385–1391.

[4] Burger, G. and Werner, S. (1983) in: Mitochondria 1983, Nucleo-mitochondrial Interactions (Schweyen, R.J. et al. eds) pp.331–342.

[5] De Vries, H., De Jonge, J.C., Arnberg, A., Peijnenburg, A.A.C.M. and Agsteribbe, E. (1983) in: Mitochondria 1983, Nucleo-mitochondrial Interactions (Schweyen, R.J. et al. eds) pp.343–356.

[6] Helmer-Citterich, M., Morelli, G. and Macino, G. (1983) EMBO J. 2, 1235–1242.

[7] Burke, J.M., Breitenberger, C., Heckman, J.E., Dujon, B. and RajBhandary, U.L. (1984) J. Biol. Chem. 259, 504–511.

[8] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.

[9] Osinga, K. and Tabak, H.F. (1982) Nucleic Acids Res. 10, 3617–3627.

[10] Burke, J.M. and RajBhandary, U.L. (1982) Cell 31, 509–520.

[11] Baldacci, G. and Bernardi, G. (1982) EMBO J. 1, 987–994.

[12] Christianson, T. and Rabinowitz, M. (1983) J. Biol. Chem. 22, 14025–14033.

[13] Osinga, K.A., De Vries, E., Van der Horst, G.T.J. and Tabak, H.F. (1984) Nucleic Acids Res. 12, 1889–1900.

[14] Anderson, S.A., Bankier, B.T., Barrell, B.G., De Bruyn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457–465.

[15] Lang, B.F., Ahne, F., Distler, S., Trinkl, H., Kaudewitz, F. and Wolf, K. (1983) in: Mitochondria 1983, Nucleo-mitochondrial Interactions (Schweyen, R.J. et al. eds) pp.313–330.

[16] Macino, G. (1980) J. Biol. Chem. 255, 10563–10565.

[17] Van den Boogaart, P., Samallo, J. and Agsteribbe, E. (1982) Nature 298, 187–189.

[18] Van den Boogaart, P., Van Dijk, S. and Agsteribbe, E. (1982) FEBS Lett. 147, 97–100.