The chloroplast ribosomal intron of *Chlamydomonas reinhardii* codes for a polypeptide related to mitochondrial maturases

J.D.Rochaix, M.Rahire and F.Michel

Departments of Molecular and Plant Biology, University of Geneva, 1211 Geneva, Switzerland, and
+Centre de Génétique Moléculaire du CNRS, 91190 Gif-sur-Yvette, France

Received 4 December 1984; Accepted 3 January 1985

**ABSTRACT**

The sequences of the 888bp chloroplast ribosomal intron and of the flanking 23S rRNA gene regions of *Chlamydomonas reinhardii* have been established. The intron can be folded with a secondary structure which is typical of group I introns of fungal mitochondrial genes. It contains a 489bp open reading frame encoding a potential polypeptide that is related to mitochondrial maturases.

**INTRODUCTION**

It is well documented that several mitochondrial genes from lower eukaryotes and chloroplast genes from algae and higher plants contain introns. In yeast mitochondria split genes include those of the large rRNA (1), of apocytochrome b (2,3) and of subunit I of cytochrome oxidase (4). Chloroplast introns of higher plants occur mostly in tRNA genes (5) and more rarely in protein genes (6). In green algae chloroplast introns have been found in numerous protein genes (7,8,9,10) and in the 23S rRNA gene (11). Recently Michel and Dujon have performed an extensive sequence comparison of introns from fungal mitochondria and from rRNA genes of lower eukaryotes (12, 13). These introns, which do not follow the GT...AG rule of introns from higher eukaryotes, can be divided into two groups, I and II, based on their secondary structure and on conserved sequence elements. Among the chloroplast introns that have been sequenced at least two, the introns of the tRNA^ile^ and tRNA^ala^ genes of *Zea mays* (14), appear to fit into group II (13). In contrast, the introns of the tRNA^leu^ genes from *Vicia faba* (15) and maize (16) belong to group I.

The occurrence in yeast mitochondria of splicing deficient mutations within introns that can be complemented in trans and
the finding of large intron open reading frames that are in phase with the preceding exons has led to the concept of maturase, a chimeric exon-intron encoded polypeptide that participates in the splicing reactions (3). Comparison of the sequences of these intron open reading frames has revealed two conserved elements P1 and P2 each coding for 12 amino acids (12,31,34). While similar open reading frames have been found in mitochondrial introns from other fungi they have not yet been detected in other organisms. Most chloroplast intron open reading frames are rather small (14-19). However a large open reading frame of 894 bp has been discovered recently in an intron of a chloroplast gene of C. reinhardii (20).

Here we report the sequence of the intron of the chloroplast 23S rRNA gene of Chlamydomonas reinhardii. This 888bp intron, which can be folded with a secondary structure typical of group I introns, contains an open reading frame coding for a putative polypeptide of 163 amino acids. A striking feature of the amino acid sequence is the presence of a dodecapeptide that is highly related to the P1 element of group I introns (12). This is the first evidence for a protein coding sequence of this sort outside of fungal mitochondria.

MATERIALS AND METHODS
DNA and RNA

The plasmid containing the chloroplast ribosomal BamHI-HindIII fragment which includes the ribosomal intron was described previously (21). Plasmid DNA was prepared as described by Katz et al. (22). DNA sequencing was performed by the chemical cleavage method of Maxam and Gilbert (23). The DNA sequence analysis was performed on a Hewlett Packard computer, model 9845. The folding model of the intron (fig. 3) was obtained as described by Michel et al. (12). C. reinhardii RNA was isolated as described (24).

Mapping of the 3' end of the 23S rRNA gene

The 166bp AluI fragment (fig. 1) was 3'end labelled with DNA polymerase (Klenow fragment), denatured and annealed with C. reinhardii RNA. The hybrids were digested with S1 nuclease (25) and the S1 resistant DNA products were sized on a 5% sequencing
RESULTS AND DISCUSSION

Previous studies have revealed the presence of an intron in the chloroplast 23S rRNA gene of *C. reinhardii* near its 3' end (11) and they have allowed us to establish the sequence of the two junctions between the intron and the flanking rRNA coding sequences (21). Fig. 1 displays a restriction map of the intron and of the flanking 23S rRNA coding regions which are contained in a 1623 bp BamHI-HindIII fragment. The sequencing strategy is also indicated. The BamHI site maps within the 23S rRNA gene sequence and the HindIII site is located 50 bp downstream of the 3' end of the 23S rRNA gene (figs. 1, 2, cf. Materials and Methods). The intron is located in a region of the 23S rRNA gene that has been highly conserved in different organisms (26,27). Comparison with the corresponding *E. coli* sequence (26) reveals a sequence homology of 78% for the 389bp upstream region and 72.5% for the 190bp downstream region (relative to the intron). The homology between the last 100 bases of the 23S rRNA genes of *C. reinhardii* and *E. coli* is only 59%. It is noted that the intron is located...
The open reading frame extends between residues 2204 and 2904. One base insertion and deletions in the E. coli sequence are indicated by △ and ▼ respectively. The number of nucleotides is indicated for larger insertions. The two sites in the 23S rRNA gene whose homologues are changed in the yeast mitochondrial 21S rRNA gene of chloramphenicol resistant mutants (28) are marked with 0. S1 refers to the 3'end of the 23S rRNA gene. The intron ends are indicated by large dark wedges (21). The open reading frame starts at position 571 and ends at position 1059. The conserved dodecapeptide is underlined (12). Regions corresponding to box 9 and box 2 are indicated (30). Parts of the intron flanking regions determined previously (21) have been corrected.
near the homologous region of the yeast mitochondrial 21S rRNA
gene where mutations have been sequenced that confer resistance
to chloramphenicol (28, fig. 2). A uniparental chloramphenicol
resistant mutant has been isolated in C. reinhardii (29). It is
therefore possible that this mutation is located in the region of
the chloroplast 23S rRNA gene indicated in fig. 2. If true, this
region would provide a new correlation site between the physical
and genetic maps of the chloroplast genome of C. reinhardii.

The ribosomal intron is significantly richer in AT (63.4%)
than the surrounding rRNA gene region (49.6%). This 888bp intron
contains an open reading frame which could encode a polypeptide
of 163 amino acids (fig. 2). An unusual property of this basic
protein is its high lysine content (11.6%). Fig. 3 shows that the
secondary structure of the intron resembles closely the structure
of group I introns with several characteristic helical regions (a
to e) and loops (fig. 3). Typical features include the U residue
preceding the 5' splice junction that can basepair with a G
within the intron in helix a and the presence of a G at the 3'end
of the intron. The position of the open reading frame is unusual:
It starts in the loop of helix d and continues through helix d
(fig. 3). In most of the group I introns the open reading frame
is in the loop of helix e (12). Another distinctive feature of
group I introns are two elements related to the box 9 and box 2
sequences of the fourth intron of the yeast mitochondrial
cytochrome b gene (30). Mutations in these elements are
cis-dominant and are thought to destroy recognition sequences
involved in splicing. The C. reinhardii box 2 homologue is highly
related to the fungal consensus sequence (figs. 2,3). In contrast
the C. reinhardii box 9 homologue (figs. 2,3) is a rare variant
which has also been found in aI5α and aI5β, two introns in the
yeast mitochondrial gene of subunit I of cytochrome oxidase (31).
There are only four complementary bases between the box 2 and box
9 regions (f and f' in fig. 3) whereas in most cases the two
elements have 5 complementary bases (13). Two additional helices
are present on the 3' side of f' as has also been observed in
fungal ribosomal introns (32). Davies et al. (33) have proposed
that the two exon-intron boundaries may come into close contact
with each other through the pairing of short complementary guide
Fig. 3. Secondary structure of the chloroplast ribosomal intron of *C. reinhardii*. Thick arrows indicate the 5'and 3'ends of the intron. a, U1, b, c, d, and D1 refer to the helices of group I introns (12). The two complementary regions of the box 9 and box 2 elements are indicated by f and f', respectively. Nucleotides that are conserved in the ribosomal intron of *Kluyveromyces thermotolerans* (32) are shaded and those that are maintained in the a15a intron of *S. cerevisiae* (31) are marked with dots. The last six codons of the open reading frame are indicated with the corresponding amino acids. Codons used rarely in chloroplast protein genes are marked with asterisks. X designates the stop codon. The arrows with perpendicular tails near the ends of the intron (marked by short arrows) indicate two possible complementary guide sequences (33).

sequences located in the first loop of the intron and in the region downstream of the other intron-exon junction. Sequences of this type are also found in *C. reinhardii* (marked by arrows with perpendicular tails in fig. 3). However an alternative pairing which brings the two intron ends into close proximity is also possible (fig. 4a, b).

An interesting feature of the open reading frame of the ribosomal intron of *C. reinhardii* is the presence near its N terminal end of a dodecapeptide YLAGFVDGDGSI which is highly related to one of the group I consensus sequences YLAGFVDGDGSI.
also referred to as P1 (12, 31, 34, fig. 2). The second conserved element P2 is missing in this open reading frame, the shortest found among intron open reading frames of this sort. The homology with mitochondrial intron open reading frames suggests that the chloroplast protein has a similar function and that it may be involved in the splicing reactions. There is however no proof that this protein is synthesized. The fact that there is no apparent ribosome binding site near the ATG initiation codon of the open reading frame does not necessarily imply that it is not expressed since the absence of a ribosome binding site has also been reported for the chloroplast gene of the β subunit of ATP synthase in spinach (35). In yeast it has recently been possible to prepare antibodies against maturases and to demonstrate their presence in splicing deficient mutants, but not in wild type cells presumably because these proteins are highly unstable (38, 39). In contrast to typical maturases whose coding regions include both exon and intron sequences, the chloroplast ribosomal intron open reading frame is contained entirely within the intron as is the case for its homologues in the introns of the mitochondrial 21S rRNA genes from the yeasts Saccharomyces cerevisiae (28) and Kluyveromyces thermotolerans (32). Recently intron open reading frames that prolong the upstream exons have been found in a chloroplast gene of C. reinhardii (10).

The codon usage in the chloroplast ribosomal open reading frame is rather unusual and differs considerably from the restricted codon usage found in chloroplast protein coding
sequences of \textit{C. reinhardii}. Only 51 different codons were found among 1505 codons examined from several chloroplast genes (37). It is interesting to note that of the 10 missing codons eight are present in the ribosomal open reading frame and they all end with C or G (Table I). It can be seen in fig. 3 that several of the codons which participate in the helical structure d are rarely used in chloroplast gene sequences of \textit{C. reinhardii} (Table I) suggesting that in this case codon usage is governed by secondary structure requirements. Similar observations have been made for open reading frames of mitochondrial introns (36).

Little is known on the mechanisms of splicing of the chloroplast rRNA precursor in \textit{C. reinhardii} except that hybridization of RNA with an intron specific probe reveals the presence of a transcript of equal size to the intron (40). Since group I introns also include the ribosomal intron of \textit{Tetrahymena pyriformis} which is capable of autocatalytic splicing (41), it will be of interest to determine whether the same holds for the chloroplast intron. In this case, there may be no strict requirement for the intron encoded polypeptide, which could merely act as a cofactor for improving the efficiency of splicing.

\textbf{ACKNOWLEDGEMENTS}

We thank O. Jenni for drawings and photography, J. Erickson and M. Goldschmidt-Clermont for helpful comments. This work was supported by grants 3.258.082 and 3.587.084 from the Swiss National Science Foundation.

\textbf{Abbreviation used}

\begin{tabular}{ll}
 rRNA & ribosomal RNA \\
\end{tabular}

\textbf{REFERENCES}

1. Bos, J.L., Heyting, C. and Borst, P. (1978) Nature \textbf{275}, 336–338.
2. Nobrega, F.G. and Tzagaloff, A. (1980) J. Biol. Chem. \textbf{255}, 9828–9837.
3. Lazowska, J., Jacq, C. and Slonimski, P.P. (1980) Cell \textbf{22}, 333–348.
4. Bonitz, S., Coruzzi, G., Thalenfeld, B., Tzagaloff, A. and Macino, G. (1980) J. Biol. Chem. \textbf{255}, 11927–11941.
5. Whitfeld, P.R. and Bottomley, W. (1983) Ann. Rev. Plant. Physiol. \textbf{34}, 279–310.
6. Koller, B. and Delius, H. (1984) Cell \textbf{6}, 613–622.
7. Stiegler, G.L., Matthews, H.M., Bingham, S.E. and Hallick, R.B. (1982) Nucl. Acid. Res. 10, 3427-3444.
8. Montandon, P.E. and Stutz, E. (1983) Nucl. Acids Res. 11, 5977-5982.
9. Karabin, G.D., Farley, M. and Hallick, R.B. (1984) Nucl. Acids Res 12, 5881-5812.
10. Erickson, J.M., Rahire, M. and Rochaix, J.D. (1984) EMBO J. in press.
11. Rochaix, J.D. and Malnoe, P.M. (1978) Cell 15, 661-670.
12. Michel, F., Jacquier, A. and Dujon, B. (1982) Biochimie 10, 867-881.
13. Michel, F. and Dujon, B. (1983) EMBO J. 2, 33-38.
14. Koch, W., Edwards, K. and Kössel, H. (1981) Cell 25, 203-213.
15. Bonnard, G., Michel, F., Weil, J.H. and Steinmetz, A. (1984) Molec. Gen. Genet. 194, 330-336.
16. Steinmetz, A.A., Gubbins, E.J. and Bogorad, L. (1982) Nucl. Acids Res 10, 3027-3037.
17. Takaiwa, F. and Sugiura, M. (1982) Nucl. Acids Res. 10, 2665-2676.
18. Takaiwa, F. and Sugiura, M. (1982) Eur. J. Biochem. 124, 13-19.
19. Deno, H., Kato, A., Shinozaki, K. and Sugiura, M. (1982) Nucl. Acids Res. 10, 7511-7520.
20. Erickson, J., Delépelaire, P. and Rochaix, J.D. (1984) in Molecular Biology of the Photosynthetic Apparatus, Cold Spring Harbor Laboratory, Eds. Arntzen, C.J., Bogorad, L., Bonitz, S. and Steinback, K.E., in press.
21. Allet, B. and Rochaix, J.D. (1979) Cell 18, 55-60.
22. Katz, L., Kingsbury, D.T. and Helinski, D.R. (1973) J. Bacteriol. 114, 577-591.
23. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 449-560.
24. Rochaix, J.D. and Malnoe, P.M. (1982) in Methods in Chloroplast Molecular Biology. M. Edelman, R.B. Hallick and N.H. Chua, Eds., pp. 477-490 Elsevier Biomedical Press.
25. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
26. Brosius, J., Dull, T.J. and Noller, H.F. (1980). Proc. Natl. Acad. Sci. USA 77, 201-204.
27. Edwards, K. and Kössel, H. (1981) Nucl. Acids Res 9, 2853-2869.
28. Dujon, B. (1980) Cell 20, 185-197.
29. Bennoun, P., Delépelaire, P. and Delosme, M. (1981) Current Genet. 3, 251-253.
30. De La Salle, H., Jacq, C. and Slonimski, P.P. (1982) Cell 28, 721-732.
31. Hengsens, L.A.M., Bonen, L., de Haan, M., van der Horst, G. and Grivell, L.A. (1983) Cell 32, 379-389.
32. Jacquier, A. and Dujon, B. (1983) Mol. Gen. Genet. 192, 487-489.
33. Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A. and Scanzocchio, C. (1982) Nature 300, 719-724.
34. Waring, R.B., Davies, R.W., Scanzocchio, C. and Brown, T.A. (1982) Proc. Natl. Acad. Sci. USA 79, 6332-6336.
35. Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1982) Proc. Nat. Acad. Sci. USA 79, 6260-6264.
36. Dujon, B. (1981) In The Molecular Biology of the Yeast Saccharomyces Strathern, S.N., Jones, E.W. and Broach, J.R. eds pp. 503-635 Cold Spring Harbor Laboratory, New York.
37. Rochaix, J.D., Dron, M., Rahire, M. and Malnoe, P.M. (1984) Plant Molec. Biol. 3, 363-370.
38. Jacq, C., Banroques, J., Becan, A.M., Slonimski, P.P., Guiso, N. and Danchin, A. (1984) EMBO J. 3, 1567-1572.
39. Guiso, N., Dreyfus, M., Siffert, O., Danchin, A., Spyridakis, A., Gargouri, A., Claisse, M. and Slonimski, P.P. (1984) EMBO J. 3, 1769-1772.
40. Rochaix, J.D. (1981) Experientia 37, 323-332.
41. Cech, T.R., Tanner, N.K., Tinoco, I. Jr., Weir, B.B., Zucker, M. and Perlman, P.S. (1983) Proc. Natl. Acad. Sci USA 80, 3903-3907.