Volume 220, number 1, 61–66 FEB 04971 August 1987

Splicing of group II introns in mRNAs coding for cytochrome b_6 and subunit IV in the liverwort *Marchantia polymorpha* chloroplast genome

Exon specifying a region coding for two genes with the spacer region

Hideya Fukuzawa[+], Tadashi Yoshida, Takayuki Kohchi, Takeshi Okumura*, Yasuji Sawano[°] and Kanji Ohyama

Laboratory of Plant Molecular Biology, Research Center for Cell and Tissue Culture, Faculty of Agriculture, Kyoto University, Kyoto 606 and **Kyoto Analytical Application Laboratory, Shimadzu Corp., Kyoto 604, Japan*

Received 23 April 1987; revised version received 16 June 1987

Two genes, *petB* and *petD*, encoding cytochrome b_6 and subunit IV of the cytochrome b_6/f complex, respectively, of the chloroplast genome of *Marchantia polymorpha*, a liverwort, have been mapped between the *psbH* and *rpoA* genes and each of the *petB* and *petD* genes has been predicted to contain a group II intron in its coding sequence [(1986) Nature 322, 572–574]. Primer extension analysis of the liverwort chloroplast RNA using two synthetic oligodeoxyribonucleotides, complementary to parts of the coding sequences of the *petB* and *petD* transcripts, as primers has provided evidence that the two genes are co-transcribed to a precursor mRNA which is then precisely spliced at the predicted sites.

Chloroplast; Cytochrome b_6/f complex; RNA splicing; Primer extension; (*Marchantia polymorpha*)

1. INTRODUCTION

The cytochrome b_6/f complex of chloroplasts catalyzes the photosynthetic electron transport from plastoquinol to plastocyanin and thus forms a link between photosystems I and II. The complex has four major polypeptides, cytochrome *f* (34 kDa), cytochrome b_6 (23.5 kDa), Rieske FeS center (20 kDa), and subunit IV (17.5 kDa) [2]. Genes coding for cytochrome *f* (*petA*), cytochrome b_6 (*petB*) and subunit IV (*petD*) are located in the chloroplast genome and have been sequenced in spinach and some other plants [3–8]. Rieske FeS protein, on the other hand, is encoded by a nuclear gene and synthesized in the cytoplasm [9].

The complete nucleotide sequence of the chloroplast genome of *Marchantia polymorpha*, a liverwort, has recently been determined and the organization of gene in this genome has been deduced [1]. In this genome, the leucine tRNA(UAA) gene is thought to be split by a group I intron, whereas 17 genes including *petB* and *petD* have been predicted to contain group II introns. This prediction was based on their possession of 5′- and 3′-consensus sequences specific for group II introns: 5′-GTGYG, and RAGCCGNATGAA-

Correspondence address: K. Ohyama, Laboratory of Plant Molecular Biology, Research Center for Cell and Tissue Culture, Faculty of Agriculture, Kyoto University, Kyoto 606, Japan

[+] Present address: Institute of Applied Microbiology, University of Tokyo, Tokyo 113, Japan

[°] Present address: General Research Institute, Ozeki Sake Brewing Co. Ltd, Nishinomiya 663, Japan

The nucleotide sequence presented has been submitted to the EMBL/GenBank database under the accession number Y00686

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

NNGAAANNTTCATGTNCGGTTY--CTAYYN-YNAY-3′, where R and Y stand for purine and pyrimidine nucleotides, respectively [10]. This paper reports experimental evidence to indicate that the predicted introns in the *petB* and *petD* genes, which are tandemly arranged in the chloroplast genome [1], are in fact spliced out in the predicted sites.

2. MATERIALS AND METHODS

Chloroplast RNA was prepared from the chloroplasts of 10-day-old cultured cells of *Marchantia polymorpha*, a liverwort, by phenol extraction as described [11]. RNA was purified three times by LiCl precipitation and twice by ethanol precipitation to remove a trace amount of DNA contamination.

Two oligodeoxyribonucleotides were synthesized by a DNA synthesizer (Shimadzu NS-1, Shimadzu Corp., Japan). One was a 17-mer primer (P1; 5′-GCAATCGCTTGAATCTC-3′) that is complementary to part of the coding region (exon 2) for the *petB* gene. The other was a 17-mer primer (P2; 5′-GATCGTTTGGCCAAGCA-3′) complementary to part of the coding region (exon 3) for the *petD* gene. The 5′-ends of primers were labeled with [γ-^{32}P]ATP (5000 Ci/mmol, Amersham, Japan) and with T_4 polynucleotide kinase (Takara Shuzo Co. Ltd, Japan) and purified by 10% polyacrylamide gel electrophoresis.

Purified chloroplast RNA (175 μg and 5′-end labeled primer (6 ng) were first incubated in a 22-μl reaction mixture containing 50 mM Tris-HCl (pH 8.3), 6 mM $MgCl_2$, and 40 mM KCl at 42°C for 5 min. Reverse transcriptase from Avian myeloblastosis virus (AMV; 25 units, Life Sciences Inc., USA) was added to the reaction mixture. Primer extension experiments were done by the addition of 1 μl of a 1 mM dNTP solution to 5 μl of an RNA-primer reaction mixture at 42°C for 20 min. For RNA sequencing, 1 μl of 1 mM dNTP and 1 μl of 1 mM dideoxy NTP were added to 4 μl of the RNA-primer reaction mixture at 42°C for 20 min. The radioactive label was then chased with 1 μl of 1 mM dNTP and 2.5 units of AMV reverse transcriptase for another 20 min of incubation. The reaction was stopped by the addition of 7 μl of formamide solution containing dyes into the reaction mixture, which was kept at 95°C for 2 min. The reaction mixtures were put on either a 6% (P2 primer) or 10% (P1 primer) polyacrylamide gel containing 8.3 M urea for RNA sequencing electrophoresis (constant voltage of 30 V/cm, 4 h).

3. RESULTS AND DISCUSSION

The *petB* and *petD* genes in the liverwort chloroplast genome have previously been mapped between *psbH* (encoding a 10 kDa phosphoprotein of photosystem II) and the *rpoA* gene (coding for the α-subunit of RNA polymerase) [1]. The predicted organization of these genes in the genome is illustrated in fig.1, and the nucleotide sequence of the corresponding region is shown in fig.2. The locations of introns have been predicted with 5′- and 3′-consensus intron sequence as markers. In the predicted organization, the coding sequence for cytochrome b_6 (*petB* product) starts 108 bp downstream from the termination codon of the *psbH* gene. The first exon (exon 1) covers a 5′-flanking region plus first two codons (initiator ATG and GGT for glycine) for cytochrome b_6. This exon is followed by a 495-bp intron (intron 1) starting with GTGCG, a 5′-terminal consensus intron sequence and ending with 3′-terminal consensus sequences for group II introns. The next exon (exon 2) is unique in that it covers the coding

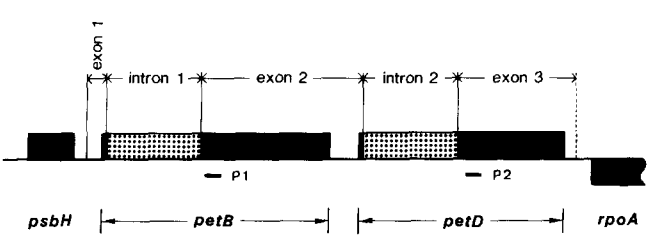

Fig.1. Organization of the liverwort genes *psbH*, *petB*, *petD*, and a part of *rpoA*, α-subunit of RNA polymerase. Synthetic oligodeoxyribonucleotides used as primers in the primer extension experiments are shown as P1 and P2.

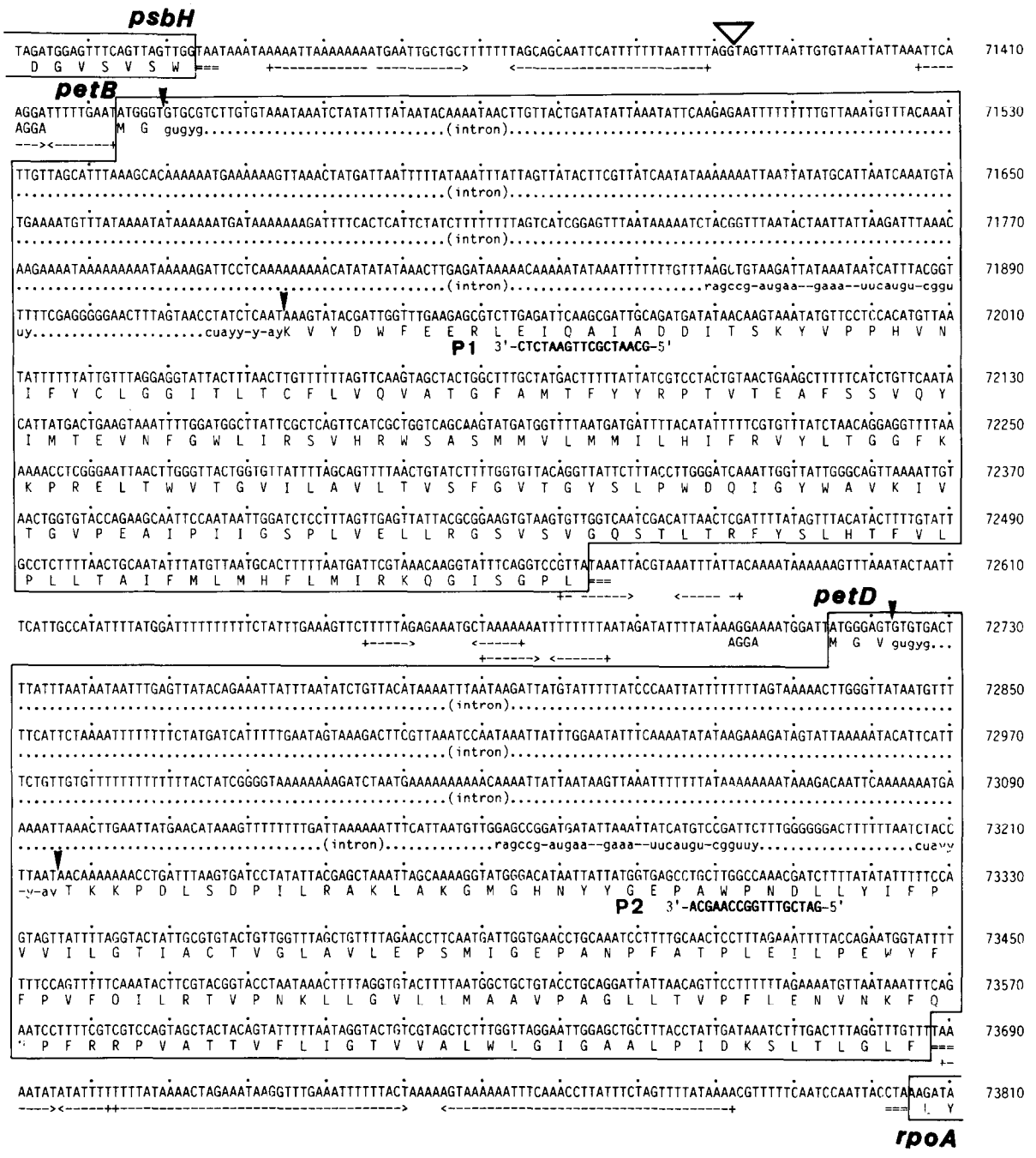

Fig.2. Nucleotide sequence and information deduced about *petB* and *petD* genes. Amino acid sequences deduced from the nucleotide sequence are shown by one-letter symbols. The coding sequences are boxed. Intron sequences have dots under the DNA sequence. The 5'- and 3'-terminal consensus sequences of group II introns are shown by lower-case letters under the DNA sequence. Oligodeoxyribonucleotides (P1 and P2) used in the primer extension experiments are shown as complementary sequences. Dotted lines with arrows indicate stem-and-loop structures. Double underlining shows termination codons. Vertical arrowheads are for splicing sites of exon-introns. Sequences (AGGA) given under the DNA sequence are possible ribosome binding sites. The open triangle indicates a processing site of mRNA.

region for the remaining 213 residues of cytochrome b_6 and the sequence ATGGGAGT coding for the N-terminal portion of subunit IV (*petD* product) in addition to a 151-bp spacer region connecting the two genes. This organization of exon 2 indicates that the mature mRNA to be formed should be polycistronic encoding both cytochrome b_6 and subunit IV. This exon is followed by a second 493-bp intron (intron 2) that starts with GTGTG, a 5′-terminal consensus intron sequence, and lasts until a 3′-terminal consensus sequence for group II introns. Finally, exon 3 covers the coding sequence for the remaining C-terminal portion of subunit IV and a 3′-flanking region of the *petD* gene.

The purpose of this study was to examine experimentally the validity of the predicted exon-intron organization outlined above. Liverwort chloroplasts, prepared from 10-day-old cells in suspension culture, accumulate considerable amounts of precursor mRNA compared to the mature mRNA. On the other hand, a chloroplast RNA preparation from rapidly growing cells (1–2-day-old) had a small amount of precursor mRNA. Therefore, this system helps us to investigate RNA processing and splicing events in the liverwort chloroplasts by means of primer extension analysis. To this end, primer extension analysis of the liverwort chloroplast RNA was carried out using synthetic oligodeoxyribonucleotides, P1 and P2, as primers. P1 and P2 were designed to be complementary to parts of exons 2 and 3, respectively (figs 1 and 2). As shown in fig.3, when primer P1 was used, the autoradiograms gave the sequence, 5′-AAAGUAUACGAUUGGUUU-3′, corresponding to the predicted 5′-terminal portion of exon 2. Upstream from this point, however, doublet sequence bands could be detected, suggesting the presence in the liverwort chloroplasts of two species of RNA that could hybridize with primer P1. The first six nucleotides (5′-AUGGGU-3′, boxed in fig.3) immediately upstream from the junction site, obtained from the major bands, were identical with the sequence predicted for the 3′-terminal portion of exon 1. Moreover, the ribosomal binding site (AGGA) [12] was seen 10 bases upstream from the initiation codon of the *petB* gene. It was thus certain that this RNA represented the mature mRNA that had been precisely spliced as predicted. The upstream

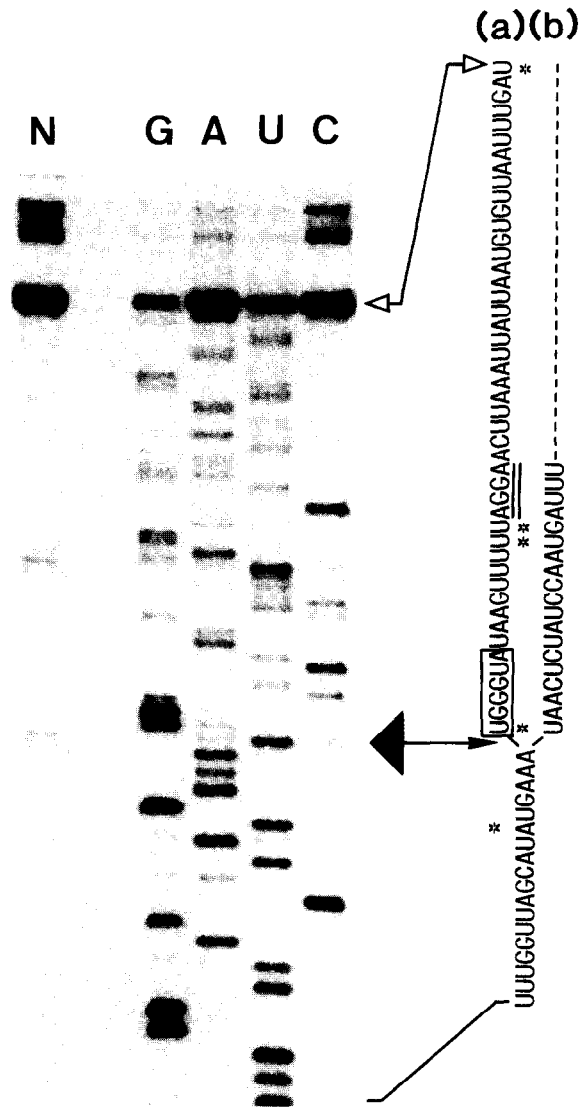

Fig.3. Sequence analysis of the 5′-end of *petB* mRNA. Lanes G, A, U, and C indicate sequences of cDNA synthesized with the use of the P1 primer in the presence of dideoxy NTP (C, T, A, and G, respectively) and lane N is without dideoxy NTP. Mature and premature mRNA sequences obtained from the autoradiogram are shown in (a) and (b), respectively. The closed arrow indicates the splicing site. The open arrow indicates the location of the 5′-ends of the *petB* mRNA. Nucleotide sequence coding for an initiation codon and N-terminal cytochrome b_6 protein in exon 1 is boxed. Asterisks indicate non-specific stop signals of primer extension. The ribosome binding site is shown by double underlining.

nucleotide sequence from the junction site obtained from the minor bands, on the other hand, was identical with the 3′-terminal consensus sequence of intron 1, indicating that this RNA corresponded to the unspliced precursor mRNA. These results can only be explained by assuming the presence of an intron (intron 1) in the coding sequence of the *petB* gene as shown in figs 1 and 2. The occurrence of intron 2 in the *petD* gene and splicing of its transcript at the predicted sites could be similarly confirmed by primer extension analysis using primer P2. As shown in fig.4, doublet sequence bands were again detected in the upstream region from the sequence 5′-AACAAAA-AAACCUGAUUUAAG-3′, which corresponds to the 5′-terminal portion of predicted exon 3. The first eight nucleotides (5′-AUGGGAGU-3′) in the sequence obtained from the major bands were identical with those predicted for the 3′-terminal portion of exon 2, whereas the sequence obtained from the minor bands matched perfectly with 3′-terminal portion of intron 2. The ribosomal binding site (AGGA) could again be detected 10 bases upstream from the initiation codon of the *petD* gene.

Primer extension analysis of *petB* gene transcript also indicated that the major 5′-terminal end for the mRNA of the *petB* gene was 41 bases upstream from the initiation codon of this gene (open arrow in fig.3). This signal corresponds to a major processing site of precursor mRNA between *psbH* and *petB* genes. Typical prokaryotic promoter sequences were not seen in the 107-bp spacer region between the *psbH* and *petB* genes, but a 25-bp-long stem-and-loop structure could be formed in the region. This structure may be functional in the processing of precursor mRNA between the two genes *psbH* and *petB* as described above. On the other hand, primer extension analysis of the *petD* gene transcript did not show any processing signal in the spacer region between *petB* and *petD* genes (fig.4). It is, therefore, certain that the mature mRNA to be formed is polycistronic to the two genes. A large stem-and-loop structure could be formed in the 3′-flanking region of the *petD* gene (fig.2). This may be important in the binary termination of transcripts for both *petD* and *rpoA* genes.

Primer extension analysis demonstrated the presence and the precise splicing of introns of both

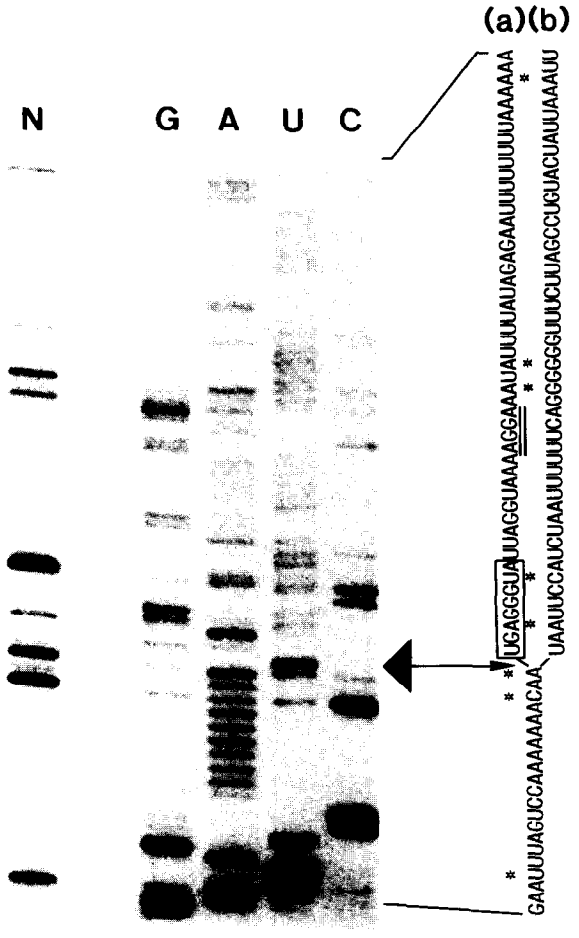

Fig.4. Sequence analysis of the 5′-end of *petD* mRNA. Lanes G, A, U, and C indicate the sequence of cDNA synthesized with the use of the P2 primer in the presence of dideoxy NTP (C, T, A, and G, respectively) and lane N is without dideoxy NTP. Mature and precursor mRNA sequences obtained from the autoradiogram are shown in (a) and (b), respectively. The closed arrow indicates the splicing site. Nucleotide sequence coding for an initiation codon and N-terminal subunit IV protein in exon 2 is boxed. Asterisks indicate non-specific stop signals of primer extension. The ribosome binding site is shown by double underlining.

petB and *petD* genes in the liverwort chloroplasts. Introns in the coding sequences for cytochrome b_6 and subunit IV have not been described for spinach [7] or pea [8]. The tobacco chloroplast genome has been sequenced, and the presence of an intron has been predicted in the *petB* gene, but not in the *petD* gene [13]. To our understanding,

however, an intron can be predicted in a tobacco *petD* coding sequence by application of the 5′- and 3′-consensus group II intron sequence. These facts indicate that introns are characteristic of *petB* and *petD* genes in chloroplast genomes from land plants. Chloroplast cytochrome b_6 and subunit IV in the cytochrome b_6/f complex have sequence homology and structural similarity to the corresponding N-terminal portion and C-terminal portion of cytochrome *b* of mitochondrial complex III [14]. Perhaps split chloroplast genes (*petB* and *petD*) evolved into a single polypeptide as seen in mitochondria, being related to introns in the coding sequences.

ACKNOWLEDGEMENTS

This research was supported in part by a Grant-in-Aid for Special Research Projects from the Ministry of Education, Science, and Culture of Japan.

REFERENCES

[1] Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.

[2] Hurt, E. and Hauska, G. (1981) Eur. J. Biochem. 117, 591–599.

[3] Alt, J. and Herrmann, R.G. (1984) Curr. Genet. 8, 551–557.

[4] Willey, D.L., Howe, C.J., Auffret, A.D., Bowman, C.M., Dyer, T.A. and Gray, J.C. (1984) Mol. Gen. Genet. 194, 416–422.

[5] Willey, D.L., Auffret, A.D. and Gray, J.C. (1984) Cell 36, 555–562.

[6] Tyagi, A.K. and Herrmann, R.G. (1986) Curr. Genet. 10, 481–486.

[7] Heinemeyer, W., Alt, J. and Herrmann, R.G. (1984) Curr. Genet. 8, 543–549.

[8] Phillips, A.L. and Gray, J.C. (1984) Mol. Gen. Genet. 194, 477–484.

[9] Alt, J., Westhoff, P., Sears, B.B., Nelson, N., Hurt, E., Hauska, G. and Herrmann, R.G. (1983) EMBO J. 2, 979–986.

[10] Michel, F. and Dujon, B. (1983) EMBO J. 2, 33–38.

[11] Yamano, Y., Ohyama, K. and Komano, T. (1984) Nucleic Acids Res. 12, 4621–4624.

[12] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.

[13] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.

[14] Widger, W.R., Cramer, W.A., Herrmann, R.G. and Trebst, A. (1984) Proc. Natl. Acad. Sci. USA 81, 674–678.