Volume 235, number 1,2, 283–288 FEB 06154 August 1988

Identification of a new gene in the chloroplast genome encoding a low-molecular-mass polypeptide of photosystem II complex

Norio Murata, Mitsue Miyao, Nobuaki Hayashida[+], Tadashi Hidaka[+] and Masahiro Sugiura[+]

National Institute for Basic Biology, Myodaiji, Okazaki 444 and [+]*Center for Gene Research, Nagoya University, Chikusa, Nagoya 464, Japan*

Received 3 June 1988

Low-molecular-mass polypeptides in spinach photosystem II membranes were separated by SDS-polyacrylamide gel electrophoresis. The partial amino acid sequence of one of the polypeptides was determined. Comparison of this sequence with the entire nucleotide sequence of the tobacco chloroplast genome revealed that this polypeptide is encoded in the chloroplast genome. The gene for the polypeptide, designated as *psb*K, is located between the genes for $tRNA^{Ser}$ and $tRNA^{Gln}$ in the large single-copy region and oriented in the direction opposite to the tRNA genes. The amino acid sequence deduced from the gene indicates that the translation product consists of 98 amino acid residues and its 62nd residue corresponds to the amino-terminus of the mature polypeptide. The putative polypeptide consists of 37 amino acid residues with a molecular mass of 4285 Da and has a single membrane-spanning segment. Northern blot hybridization analysis revealed that *psb*K was transcribed in the chloroplast.

Chloroplast DNA; Photosynthesis; Photosystem II; *psb*K Gene; (Spinach, Tobacco)

1. INTRODUCTION

The photosystem (PS) II complex (in this article, the PS II core complex represents the complex composed of the 47 kDa, 43 kDa, D1 and D2 proteins and apoproteins of cytochrome *b*-559, which catalyzes the photochemical reaction, and the PS II complex represents a supramolecular complex present in the thylakoid membrane which contains the PS II core complex, LHC II and three extrinsic proteins of 33, 23 and 18 kDa and carries out light trapping, the photochemical reaction and electron transport from water to plastoquinone), a supramolecular complex which spans the thylakoid membrane, catalyzes the light-driven electron transport reaction from water to a plastoquinone molecule. Polypeptide analysis by SDS-polyacrylamide gel electrophoresis indicates that more than 15 protein components exist in the PS II complex [1].

A recent biochemical study on the PS II complex has clarified the function and characteristics of part of the protein components. Intrinsic hydrophobic proteins of 47 kDa, 43 kDa, 34 kDa and 32 kDa and those of 9 kDa and 4 kDa (apoproteins of cytochrome *b*-559) form the PS II core complex [1,2]. These six proteins and a phosphoprotein of 10 kDa [3] are encoded in the chloroplast genome and the nucleotide sequences of their genes have been determined (as reviews, see [2] and [4]). Other protein components of the PS II complex characterized previously, such as apoproteins of LHC II [5], three extrinsic proteins of 33 kDa, 23 kDa and 18 kDa involved in the binding of inorganic components, Mn, Ca^{2+} and Cl^-, essential for oxygen evolution [1], hydrophobic proteins of 22 kDa, 20 kDa and 10 kDa [6], and a hydrophilic protein of 5 kDa [7], are all encoded in the nuclear genome and their

Correspondence address: N. Murata, National Institute for Basic Biology, Myodaiji, Okazaki 444, Japan

Abbreviations: Caps, 3-(cyclohexylamino)-1-propanesulphonic acid; LHC II, light-harvesting chlorophyll *a/b* protein of photosystem II; ORF, open reading frame; PVDF, polyvinylidene difluoride

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

partial amino acid sequences and the nucleotide sequences of their cDNAs have been determined ([8–11] and Herrmann, R.G., personal communication). The presence of several low-molecular-mass polypeptides of less than 10 kDa in the PS II complex has also been reported previously [7], but they have not yet been characterized.

Another approach to identify the protein components of the PS II complex is to analyze the gene products encoded in the chloroplast genome. Steinmetz et al. [12] identified the *psb*G gene in the maize chloroplast genome, using antibodies raised against a synthetic polypeptide corresponding to a part of the gene product. Using the same technique, two genes in the maize chloroplast genome, *psb*I and *psb*J, have also been found (Katoh, K., Sayer, R.T. and Bogorad, L., personal communication). On the other hand, Cushman et al. [13] reported another set of *psb*I and *psb*J in the *Euglena* chloroplast genome, which are located downstream from the *psb*E and *psb*F genes. Two genes corresponding to *psb*I and *psb*J of *Euglena* have also been found in the genome of cyanobacterium [14]. However, the presence of these gene products in the PS II complex has not yet been biochemically proved, except for the *psb*I gene of *Euglena*: the N-terminal amino acid sequence of a protein of 4 kDa in spinach PS II membranes matched that deduced from the nucleotide sequence of *psb*I of *Euglena* (Gray, J.C., unpublished).

Here we report a new gene encoding for one of the polypeptides which are present in the spinach PS II complex. The partial amino acid sequence at the N-terminal side of the polypeptide matched a part of an ORF of 98 codons in the tobacco chloroplast genome. The ORF, designated as *psb*K, is located between the genes for $tRNA^{Ser}$ and $tRNA^{Gln}$ and oriented in the direction opposite to the tRNA genes.

2. MATERIALS AND METHODS

PS II membranes were prepared from spinach thylakoids with Triton X-100 according to Kuwabara and Murata [15]. A crude preparation of the oxygen-evolving PS II core complex was prepared by removing LHC II from the PS II complex by treating the PS II membranes with 1-*O*-*n*-octyl-β-D-glucopyranoside (octylglucoside) in the presence of 0.4 M NaCl and subsequent differential centrifugation to pellet the released LHC II according to Ghanotakis et al. [16]. The oxygen-evolving PS II core complex was purified by treating the PS II membranes with octylglucoside followed by centrifugation on a sucrose density gradient according to Ikeuchi and Inoue [17]. The D1–D2–cytochrome *b*-559 complex was prepared from the PS II membranes with Triton X-100 according to Nanba and Satoh [18]. PS I particles were prepared from spinach thylakoids with digitonin according to Ikegami and Katoh [19].

Polypeptides in these preparations were separated by SDS-urea polyacrylamide gel electrophoresis with a gel plate 0.7 mm thick, using the buffer system of Laemmli [20] with slight modifications. The gel plate contained 6.0 M urea but no SDS, since the absence of SDS from the gel led to better separation of pigments from closely located polypeptide bands. The stacking gel contained 5% polyacrylamide and the separation gel consisted of a linear gradient of polyacrylamide concentration from 10 to 22%. The samples were solubilized with 2.0% SDS/4.0 M urea/50 mM dithiothreitol/10% glycerol/62.5 mM Tris-HCl (pH 6.8) at 50°C for 30 min. Electrophoresis was performed at 4°C. After electrophoresis, the gel plate was stained with 0.07% Coomassie brilliant blue R-250/50% (v/v) methanol/10% (v/v) acetic acid and destained with 30% (v/v) methanol/10% (v/v) acetic acid. Apparent molecular masses of polypeptides smaller than the 9 kDa subunit of cytochrome *b*-559 were estimated from their migration distances using molecular mass marker proteins prepared from horse-heart myoglobin by cyanogen bromide cleavage (BDH Chemicals) as standards.

For analysis of the amino acid sequences of the polypeptides in the gel, the polypeptides were transferred onto a PVDF membrane according to Matsudaira [21], except that 0.02% SDS was added to the transfer buffer, as follows. The gel plate was soaked in 0.02% SDS/10% (v/v) methanol/10 mM Caps-NaOH (pH 11.0) (transfer buffer) for 5 min, and sandwiched with PVDF membranes (Immobilon Transfer, 0.45-μm pore size, Millipore) which had been rinsed with 100% methanol and stored in the transfer buffer. Polypeptides in the gel were electroeluted onto the PVDF membrane facing the anode in the transfer buffer at 0.5 A for 30 min at room temperature. Next, the membrane was stained with 0.1% Coomassie brilliant blue R-250 in 50% (v/v) methanol for 5 min, destained with 50% (v/v) methanol/10% (v/v) acetic acid until the polypeptide bands emerged, and finally rinsed with deionized water, air-dried and stored at −20°C.

The stained polypeptide bands on the PVDF membrane were cut out with a razor, and the membrane pieces were placed in the cartridge block of a gas-phase protein sequence analyzer (470A, Applied Biosystems) and the amino acid sequence at the N-terminal side of the polypeptide was determined. The obtained amino acid sequences were compared with the amino acid sequence derived from the complete nucleotide sequence of the chloroplast genome of tobacco (*Nicotiana tabacum* var. Bright Yellow 4) [22] with a Micro VAXII computer using the IDEAS program and an NEC PC98XA computer using the GENETYX program (Software Development, Japan).

Northern blot hybridization was carried out as described in [23]. Total chloroplast RNA was prepared from young tobacco leaves and separated by electrophoresis with a 1% agarose gel. The RNA was transferred onto nylon membrane sheets and hybridized with the 5′-[^{32}P]-labelled synthetic oligonucleotide probe (see fig.3).

3. RESULTS

Among the proteins in the PS II complex, polypeptides smaller than the 9 kDa subunit of cytochrome *b*-559 have not well been characterized. Ljungberg et al. [7] previously reported five polypeptides of 7, 6.5, 5.5, 5 and 4 kDa in the oxygen-evolving PS II core complex. Under the conditions of SDS-polyacrylamide gel electrophoresis in the present study, eight polypeptide bands emerged below the band of the 9 kDa subunit in the PS II membranes and the crude preparation of the oxygen-evolving PS II core complex (fig.1, marked with arrowheads). Three of them correspond to already isolated and characterized proteins; the 10 kDa phosphoprotein [3], the 4 kDa subunit of cytochrome *b*-559 [24], and the hydrophilic 5 kDa protein [7]. The amino acid sequence of a polypeptide of the highest mobility in the gel, which had an apparent molecular mass of 2 kDa, was successfully determined up to the 13th amino acid residue as Lys-

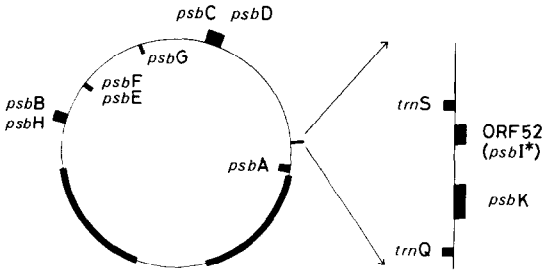

Fig.2. Location of the *psb*K gene on the tobacco chloroplast genome. The right-hand part shows an enlarged map of the region in which *psb*K is located. The inverted repeat is shown by bold lines. Genes indicated outside the circle are located on a strand transcribed counterclockwise and genes indicated inside are on another strand transcribed clockwise. *psb*A, gene for the D1 protein; *psb*B and *psb*C, genes for antenna chlorophyll proteins of 47 kDa and 43 kDa, respectively; *psb*D, gene for the D2 protein; *psb*E and *psb*F, genes for the apoproteins of cytochrome *b*-559; *psb*G, gene corresponding to the maize *psb*G [12]; *psb*H, gene for the 10 kDa phosphoprotein; *psb*I*, gene corresponding to the maize *psb*I (Katoh, K., Sayre, R.T. and Bogorad, L., personal communication); *trn*Q and *trn*S, genes for $tRNA^{Gln}$ and $tRNA^{Ser}$. *psb*J, which is present in the maize chloroplast genome, is missing from the tobacco chloroplast genome.

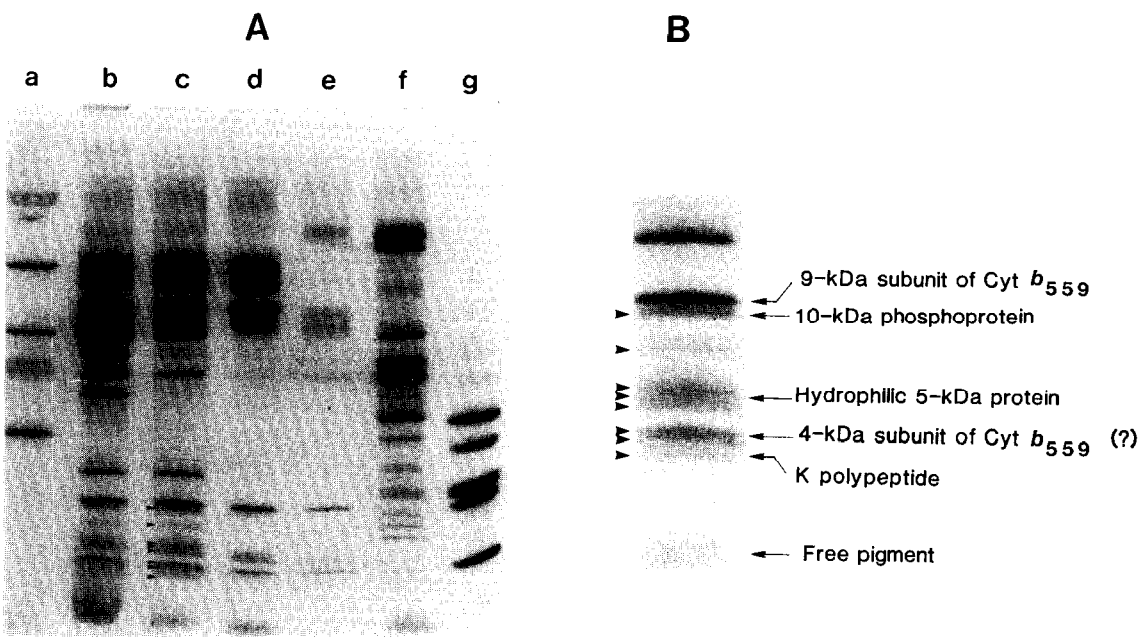

Fig.1. Separation of low-molecular-mass polypeptides of spinach PS II preparations by SDS-urea polyacrylamide gel electrophoresis. The gel was stained with Coomassie brilliant blue. (A) Lanes: a, molecular mass marker proteins (phosphorylase *b*, 94 kDa; bovine serum albumin, 67 kDa; ovalbumin, 43 kDa; carbonic anhydrase, 30 kDa; trypsin inhibitor, 20.1 kDa; α-lactalbumin, 14.4 kDa); b, PS II membranes; c, crude preparation of the oxygen-evolving PS II core complex; d, purified oxygen-evolving PS II core complex; e, D1–D2–cytochrome *b*-559 complex; f, PS I particles; g, molecular mass marker proteins prepared from horse-heart myoglobin by cyanogen bromide cleavage [16.9, 14.4, 8.16, 6.21 and 2.51 kDa, respectively (purchased from BDH Chemicals)]. (B) Enlarged photograph of the low-molecular-mass region below the 9 kDa subunit of cytochrome (Cyt) *b*-559 in the crude preparation of the oxygen-evolving PS II core complex (lane c).

Leu-Pro-Glu-Ala-Tyr-Ala-Phe-Leu-X-Pro-Ile-Val from the N-terminus.

Computer-assisted comparison of this partial amino acid sequence with those deduced from ORFs in the tobacco chloroplast genome [22] revealed that it was identical with a part of an ORF of 98 codons (ORF98). As previously shown [25], the ORF98 is located between the genes for $tRNA^{Ser}$ and $tRNA^{Gln}$ in the large single-copy region, and oriented in the direction opposite to the tRNA genes (fig.2). Another ORF of 52 codons which corresponds to *psb*I of the maize chloroplast genome is located downstream of the ORF98 [25].

The determined amino acid sequence of the polypeptide matched the deduced amino acid sequence from the 62nd to 74th residues of the ORF98 (fig.3). The putative polypeptide deduced from the ORF consists of 37 amino acid residues with a molecular mass of 4285 Da.

Transcripts of the ORF98 were analyzed using a synthetic oligodeoxyribonucleotide corresponding to positions from 237 to 274 in the ORF as a probe (see fig.3). This probe was hybridized to RNA bands of about 2.7–0.6 kb, indicating that the ORF98 is expressed in the chloroplast (fig.4). Therefore, we concluded that the ORF98 is the gene encoding for the polypeptide of the highest mobility of the PS II complex. Hereafter, the ORF98 is designated as *psb*K, in order to distinguish it from *psb*I and *psb*J which have been reported independently for different genes ([13,14] and Katoh, K., Sayer, R.T. and Bogorad, L., personal communication), and the encoded polypeptide as the K polypeptide.

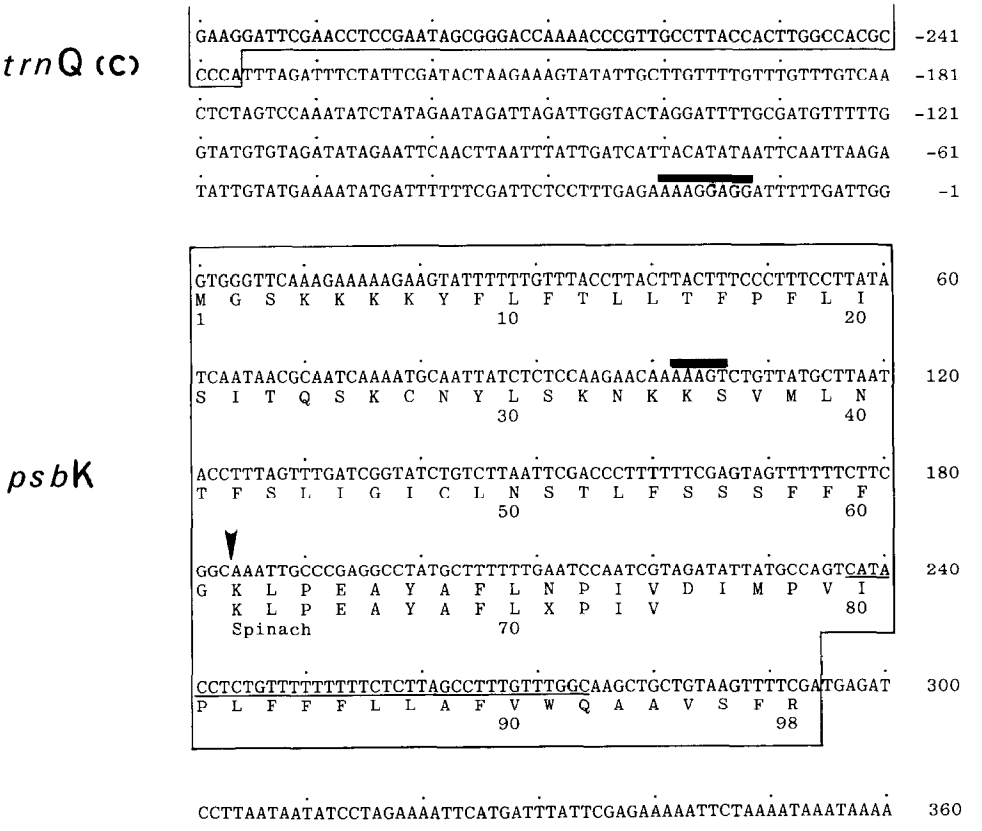

Fig.3. Nucleotide sequence of *psb*K and the derived amino acid sequence of the K polypeptide of the PS II complex of tobacco and the chemically determined amino acid sequence of the K polypeptide of spinach. The ORF is boxed. An arrowhead indicates the putative processing site. Bold lines above the nucleotide sequence indicate the SD sequences. The underlined sequence, position 237–274, corresponds to the probe used for the Northern blot hybridization. *trn*Q is located on the complementary strand.

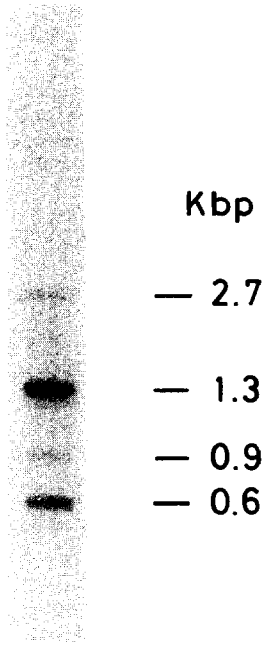

Fig.4. Detection of transcripts from the tobacco *psb*K gene. An autoradiogram of the Northern blot of tobacco chloroplast RNA hybridized with the *psb*K probe, 5′-[^{32}P]-labelled 38-mer oligonucleotide (see fig.3).

Fig.5 shows the hydropathy profile of the K polypeptide deduced from the nucleotide sequence of *psb*K. Obviously, the putative mature polypeptide consists of a single hydrophobic segment which may span the thylakoid membrane.

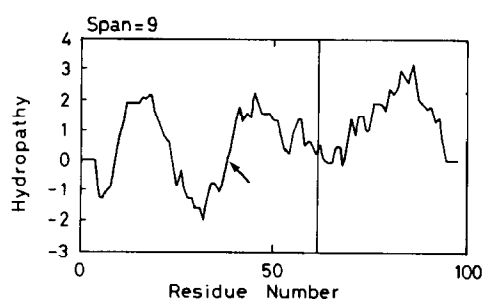

Fig.5. Hydropathy plot of the translation product of 98 amino acid residues deduced from the nucleotide sequence of the tobacco *psb*K gene. Hydropathy was calculated with a 9-point window according to Kyte and Doolittle [29]. The vertical line shows the putative processing site and the arrow indicates the second methionine at the 38th amino acid residue.

4. DISCUSSION

SDS-polyacrylamide gel electrophoresis of the present study indicates that eight polypeptides smaller than the 9 kDa subunit of cytochrome *b*-559 exist in spinach PS II membranes (fig.1). The present study shows that one of the polypeptides having an apparent molecular mass of 2 kDa, the K polypeptide, is encoded in the chloroplast genome.

The nucleotide sequence of the gene for the K polypeptide, *psb*K, indicates that the mature polypeptide consists of 37 amino acid residues with a molecular mass of 4285 Da. This is larger than the apparent molecular mass of the spinach K polypeptide estimated from the migration distance in the polyacrylamide gel. Since the C-terminus of the K polypeptide has not been analyzed yet, it cannot be ruled out that the mature polypeptide is smaller than the putative polypeptide in size. However, taking into account the fact that the putative polypeptide is highly hydrophobic (fig.5), it is more likely that the migration distance of the K polypeptide in the gel does not reflect the real molecular mass.

*psb*K has two possible initiation codons, GTG from position 1 and ATG from position 112 (fig.3). Both have sequences similar to the SD sequence in their upstream regions (fig.3), and therefore, the real initiation codon of *psb*K cannot be specified at present. If the first codon of the gene is the real initiation codon, the translation product consists of 98 amino acid residues and a polypeptide of 61 amino acid residues at the N-terminus would be removed post-translationally. Since the K polypeptide is likely to span the thylakoid membrane, its precursor may have the leading peptide to cross the membrane like other membrane-spanning proteins such as subunits I and IV of H^+-ATPase [26,27] and apoprotein of cytochrome *f* [28]. However, the polypeptide of 61 amino acid residues is large compared with previously reported pre-sequences of intrinsic proteins encoded in the chloroplast genome: subunits I and IV of H^+-ATPase have pre-sequences of 17 and 18 amino acid residues, respectively [26,27] and the apoprotein of cytochrome *f* has one of 35 residues [28]. There is no homology among these pre-sequences either in the amino acid sequences or hydropathy profiles.

The K polypeptide in PS II membranes remained bound to the PS II complex when LHC II was removed on treatment with octylglucoside (fig.1, lane c), but was absent in the purified oxygen-evolving PS II core complex or the D1–D2–cytochrome *b*-559 complex (fig.1, lanes d and e). This indicates that this polypeptide does not participate in the photochemical reaction nor in oxygen evolution. Further study is required to clarify the function of the K polypeptide in PS II, together with the mechanism of the translation and incorporation into the PS II complex of the polypeptide.

Acknowledgements: The authors are grateful to Ms Hiroko Kajiura, National Institute for Basic Biology, for the operation of the protein sequence analyzer, to Ms Yoko Fujimura of the same institute for preparing the PS II preparations and to Dr Shigeru Itoh of the same institute for kindly providing the PS I particles. This work was supported in part by Grants-in-Aid for Scientific Research on Priority Areas of the Molecular Mechanism of Photoreception (63621003) and for Cooperative Research (61304004) from the Japanese Ministry of Education, Science and Culture, and by a grant from Toray Science Foundation, Japan, to N.M.

REFERENCES

[1] Miyao, M. and Murata, N. (1987) in: Topics in Photosynthesis, vol.9, Photoinhibition (Kyle, D.J. et al. eds) pp.289–307, Elsevier, Amsterdam.
[2] Satoh, K. (1988) Physiol. Plant. 72, 209–212.
[3] Farchaus, J. and Dilley, R.A. (1986) Arch. Biochem. Biophys. 244, 94–101.
[4] Sugiura, M. (1987) Bot. Mag. Tokyo 100, 407–436.
[5] Zuber, H. (1985) Photochem. Photobiol. 42, 821–844.
[6] Ljungberg, U., Akerlund, H.-E. and Andersson, B. (1986) Eur. J. Biochem. 158, 477–482.
[7] Ljungberg, U., Henrysson, T., Rochester, C.P., Akerlund, H.-E. and Andersson, B. (1986) Biochim. Biophys. Acta 849, 112–120.
[8] Coruzzi, G., Broglie, R., Cushmore, A. and Chua, N.-H. (1983) J. Biol. Chem. 258, 1399–1402.
[9] Jansen, T., Rother, C., Steppuhn, J., Reinke, H., Beyreuther, K., Jansson, C., Andersson, B. and Herrmann, R.G. (1987) FEBS Lett. 216, 234–240.
[10] Tyagi, A., Hermans, J., Steppuhn, J., Jansson, C., Vater, F. and Herrmann, R.G. (1987) Mol. Gen. Genet. 207, 288–293.
[11] Murata, N., Kajiura, H., Fujimura, Y., Miyao, M., Murata, T., Watanabe, A. and Shinozaki, K. (1987) in: Progress in Photosynthesis Research, vol.I (Biggins, J. ed.) pp.701–704, Martinus Nijhoff, Dordrecht, The Netherlands.
[12] Steinmetz, A.A., Castroviejo, M., Sayre, R.T. and Bogorad, L. (1986) J. Biol. Chem. 261, 2485–2488.
[13] Cushman, J.C., Christopher, D.A., Little, M.C., Hallick, R.B. and Price, C.A. (1988) Curr. Genet. 13, 173–180.
[14] Cantrell, A. and Bryant, D.A. (1987) in: Progress in Photosynthesis Research, vol.IV (Biggins, J. ed.) pp.659–662, Martinus Nijhoff, Dordrecht, The Netherlands.
[15] Kuwabara, T. and Murata, N. (1982) Plant Cell Physiol. 23, 533–539.
[16] Ghanotakis, D.F., Demetriou, D.M. and Yocum, C.F. (1987) in: Progress in Photosynthesis Research, vol.I (Biggins, J. ed.) pp.681–684, Martinus Nijhoff, Dordrecht, The Netherlands.
[17] Ikeuchi, M. and Inoue, Y. (1986) Arch. Biochem. Biophys. 247, 97–107.
[18] Nanba, O. and Satoh, K. (1987) Proc. Natl. Acad. Sci. USA 84, 109–112.
[19] Ikegami, I. and Katoh, S. (1975) Biochim. Biophys. Acta 376, 588–592.
[20] Laemmli, U.K. (1970) Nature 227, 680–685.
[21] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[22] 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.
[23] Ohme, M., Kamogashira, T., Shinozaki, K. and Sugiura, M. (1985) Nucleic Acids Res. 13, 1045–1056.
[24] Widger, W.R., Cramer, W.A., Hermodson, M. and Herrmann, R.G. (1985) FEBS Lett. 191, 186–190.
[25] Deno, H. and Sugiura, M. (1983) Nucleic Acids Res. 11, 8407–8414.
[26] Bird, C.R., Koller, B., Auffret, A.D., Huttly, A.K., Howe, C.J., Dyer, T.A. and Gray, J.C. (1985) EMBO J. 4, 1381–1388.
[27] Fromme, P., Graber, P. and Salnikow, J. (1987) FEBS Lett. 218, 27–30.
[28] Willey, D.L., Howe, C.J., Auffret, A.D., Bowman, C.M., Dyer, T.A. and Gray, C.J. (1984) Mol. Gen. Genet. 194, 416–422.
[29] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.