Volume 242, number 2, 263-269 FEB 06622 January 1989

# N-terminal sequencing of photosystem II low-molecular-mass proteins

## 5 and 4.1 kDa components of the $O_2$-evolving core complex from higher plants

Masahiko Ikeuchi, Koji Takio* and Yorinao Inoue

*Solar Energy Research Group* and **Laboratory for Aging Process, Frontier Research Program, The Institute of Physical and Chemical Research (RIKEN), Wako, Saitama 351-01, Japan*

Received 10 October 1988

High resolution gel electrophoresis in the low-molecular-mass region combined with electroblotting using polyvinylidene difluoride membranes enabled us to sequence the low-molecular-mass proteins of photosystem II membrane fragments from spinach and wheat. The determined N-terminal sequences, all showing considerable homology between the two plants, involved two newly determined sequences for the 4.1 kDa protein and one for the 5 kDa proteins. The sequence of the 4.1 kDa protein did not match any part of the chloroplast DNA sequence from tobacco or liverwort, suggesting that it is encoded by the nuclear genome. In contrast, the sequence of the 5 kDa protein matched ORF38, which is located just downstream of *psb*E and *psb*F in the chloroplast DNA and is assumed to be co-transcribed with them. These two components were associated with the $O_2$-evolving core complex. Sequences of other low-molecular-mass proteins confirmed the previous identification as photosystem II components.

Chloroplast gene; $O_2$ evolution; Photosystem II core complex

## 1. INTRODUCTION

Recently many low-molecular-mass proteins have been reported in various PS II complexes, although their physiological functions have not yet been elucidated [1–4]. Recent technical development in protein sequencing using PVDF membranes [5] has enabled us to sequence PS II low-molecular-mass proteins [6–8]. However, there are still more proteins in PS II which have not yet been sequenced, and it is still not easy to assign each sequenced protein to a Coomassie-stained band on SDS-PAGE profiles among different reports.

We developed a new SDS-PAGE technique to attain high resolution of low-molecular-mass proteins and classified them into three groups: reaction center complex-associated, core complex-associated and PS II membrane fragment-associated [4]. Of these, the 4.8 kDa protein, which was registered as a new component of the reaction center complex, was sequenced and its corresponding gene was identified [8]. Here, we systematically sequenced all the other low-molecular-mass proteins of PS II, and report that two of them provide unreported sequences.

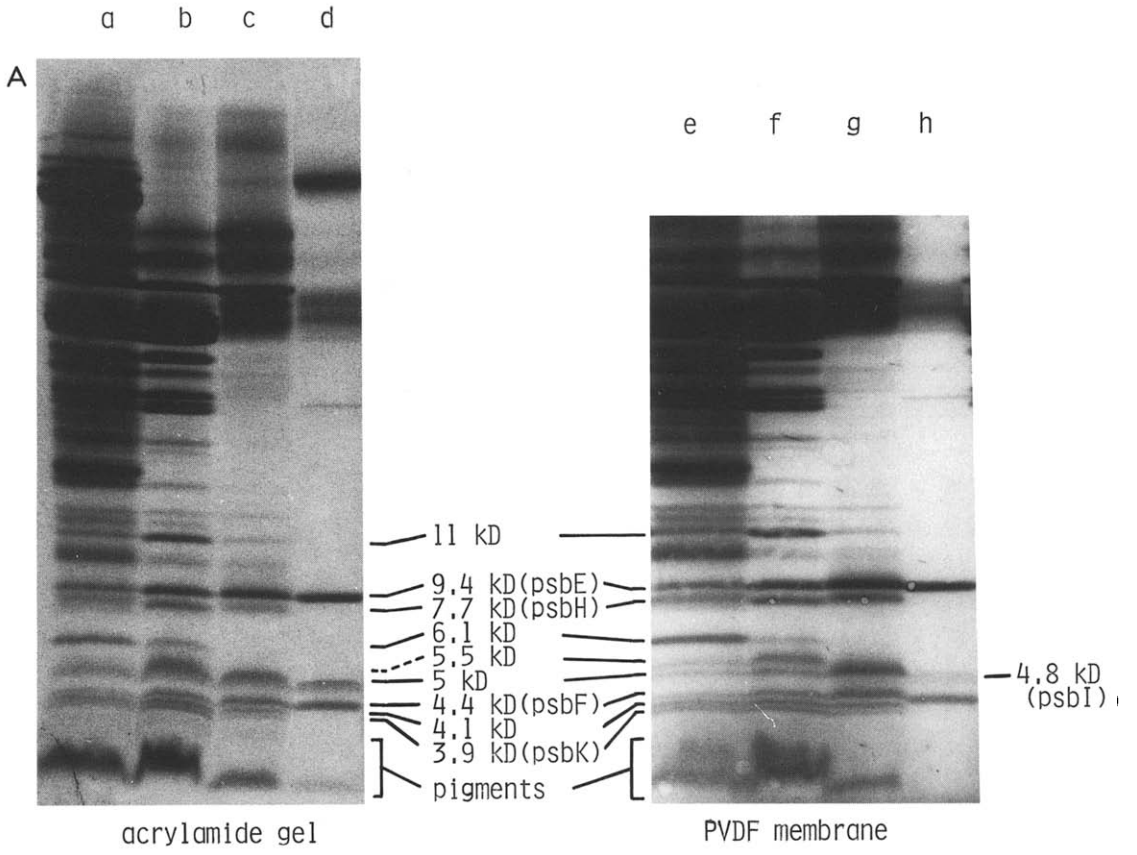
*Correspondence address:* M. Ikeuchi, Solar Energy Research Group, The Institute of Physical and Chemical Research (RIKEN), Wako, Saitama 351-01, Japan

*Abbreviations:* ORF, open reading frame; PS II, photosystem II; PVDF, polyvinylidene difluoride; SDS-PAGE, SDS-polyacrylamide gel electrophoresis

## 2. MATERIALS AND METHODS

PS II membrane fragments, $O_2$-evolving core complexes and reaction center complexes were prepared from spinach and

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/89/$3.50 




wheat according to [4]. SDS-PAGE with 7.5 M urea and a 16–22% (w/v) polyacrylamide gradient was performed following [4]. Proteins were stained with 0.04% (w/v) Coomassie brilliant blue R-250 (Bio-Rad) in 50% (v/v) methanol and 10% (v/v) acetic acid for 2 h. To obtain reproducible resolution in the low-molecular-mass region, 5% (w/v) lithium dodecylsulfate instead of 2% (w/v) was used to solubilize thylakoid membranes. For protein blotting, an unstained gel was equilibrated with a blotting buffer composed of 30 mM Tris, 17 mM boric acid, 0.055% (w/v) SDS and 20% (v/v) methanol for 5 min at 25°C, then placed on a PVDF membrane (Millipore) and sandwiched between two sets of nine sheets of filter paper (no.2, Advantech, Japan) saturated with the blotting buffer. Electroblotting was carried out at a constant voltage of 10 V/cm for 2 h at 25°C using a semidry-type electroblotter (Atto, Japan). Transferred proteins were stained with 0.1% (w/v) Amido black 10B (Bio-Rad) in 50% (v/v) methanol and 10% (v/v) acetic acid for 1 min and destained with distilled water. The stained bands were cut out and subjected to a protein sequencer (model 477A, Applied Biosystems). To release N-terminal blockage, proteins on the PVDF membrane were treated with 0.6 N HCl for 24 h at 25°C. Determined amino acid sequences were compared by computer matching with all the possible protein-coding sequences longer than 10 frames [8] deduced from chloroplast DNA sequences of tobacco (*Nicotiana tabacum*) [9] and liverwort (*Marchantia polymorpha*) [10].

## 3. RESULTS AND DISCUSSION

As already shown in our previous report [4], at least nine low-molecular-mass proteins (below 11 kDa) have been confirmed in PS II membrane fragments, and they give very similar separation profiles among various PS II complexes from spinach and wheat (fig.1, lanes a–d and i–l). When electroblotted by using semidry-type equipment, all these low-molecular-mass proteins except for the 4.8 kDa reaction center protein [4,8] were almost quantitatively transferred to a PVDF membrane and easily detected by Amido black staining (fig.1, lanes e–h and m–p). These stained bands obtained from PS II membrane fragments were cut out and subjected to N-terminal sequencing (table 1).

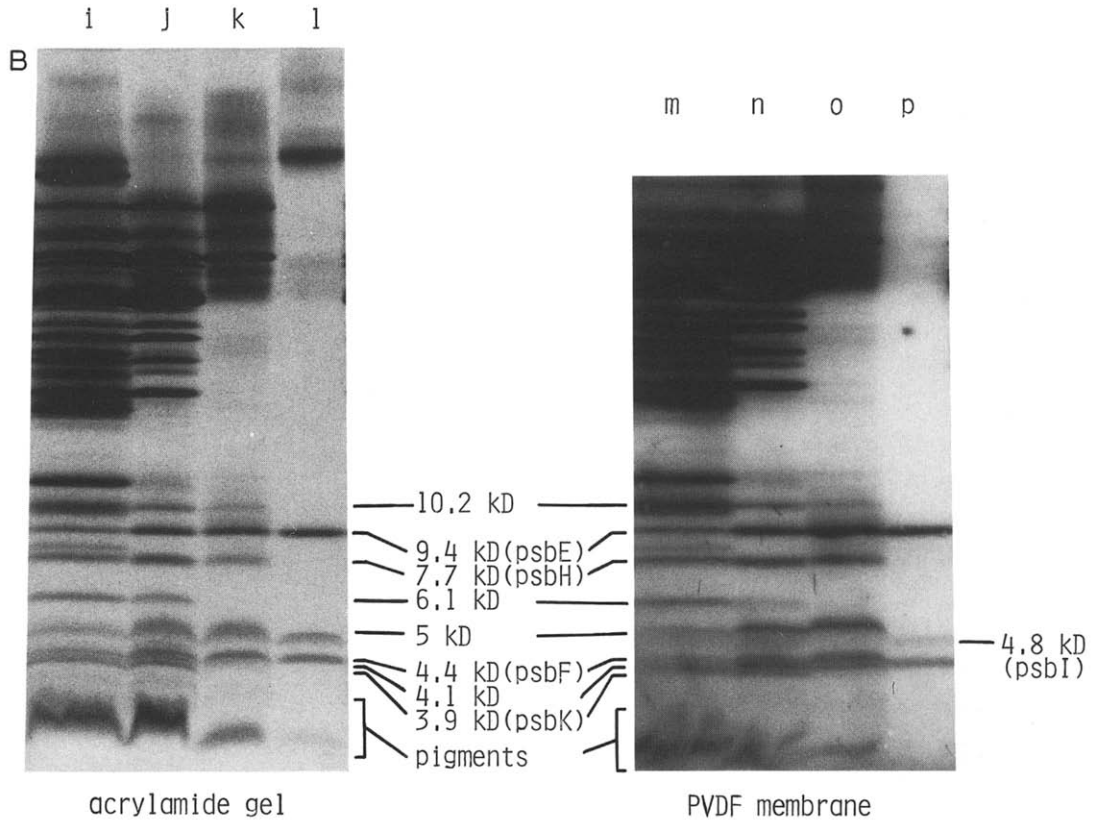



Fig.1. Low-molecular-mass proteins in various PS II preparations resolved by SDS-PAGE followed by electroblotting on PVDF membranes. Thylakoid membranes (lanes a,e,i,m), PS II membrane fragments (b,f,j,n), the $O_2$-evolving core complex (c,g,k,o) and the reaction center complex (d,h,l,p) from spinach (a–h) and wheat (i–p).

Sequencing of the 4.1 kDa protein revealed an unreported sequence having homology between spinach and wheat (table 1). This sequence, however, did not correspond to any possible reading frame in the chloroplast DNA of tobacco or liverwort, suggesting that the protein is encoded by the nuclear genome. This protein was always associated with the $O_2$-evolving core complex but not retained in the reaction center complex.

Sequencing of the 5 kDa band (a composite of more than two proteins) from PS II membrane fragments of both plants gave a pair of amino acid signals, in which one major signal was reproducibly accompanied by a minor signal in every cycle of sequencing (table 1). The amplitude of the minor signals was usually about 30% of the major ones. These paired sequencing signals were observed for the 5 kDa band obtained from the $O_2$-evolving core complex, but not for the similar band from the reaction center complex, so that we may conclude that both proteins giving the major and minor signals are real components of the $O_2$-evolving core complex. The difference in signal amplitude can be interpreted by different stoichiometry or different efficiency of protein transfer to PVDF membranes, but we could not prove that this was due to comigration of both components as a single diffused band.

The sequence determined from the minor signals (denoted as 5 kDa-I protein) was identical in spinach and wheat, and corresponded to a hypothetical product of ORF38 which is located just downstream of the *psb*F gene in both the tobacco and liverwort chloroplast DNA (fig.2). The first Met residue of the ORF38 was missing in the protein of both wheat and spinach. This ORF38 was first reported in *Cyanophora* based on the presence of a preceding ribosome-binding site

Table 1

N-terminal partial sequence of the low-molecular-mass proteins from PS II membrane fragments of spinach (SP) and wheat (WH)

| Proteins | N-terminal sequence | Genome (gene) | Minimum unit associated |
|---|---|---|---|
| SP 11 kDa | S G V K K I K V D K P L ? | N[a] | PS II membrane |
| WH 10.2 kDa | . . G . . . . . G | | |
| SP 9.4 kDa | S G S T G E R S F A D I I T | C(*psb*E) | reaction center |
| WH 9.4 kDa | . . . . . | | |
| SP 7.7 kDa | A T Q T V E S S S R | C(*psb*H) | $O_2$-evolving core |
| WH 7.7 kDa | . . . . . D . . K | | |
| SP 6.1 kDa | L V D E R M S T E G T G L P F G L S N N | N | PS II membrane |
| WH 6.1 kDa | . . . . . S L . . . . . | | |
| SP 5 kDa-I[b] | T Q S N P N E Q N V E L N R T | C(*psb*L) | $O_2$-evolving core |
| WH 5 kDa-I | . . . . . | | |
| SP 5 kDa-II | E - E P K R G T P E A K K K Y A P V V V T | N | $O_2$-evolving core |
| WH 5 kDa-II | . S D V . K . S . . . . . ? ? ? | | |
| SP 4.8 kDa[c] | M L ? L K L F V Y T V V I F F V S L F I | C(*psb*I) | reaction center |
| SP 4.4 kDa | T I D R T Y P I F T | C(*psb*F) | reaction center |
| WH 4.4 kDa | . . . . . | | |
| SP 4.1 kDa | A - - E L A P S L K N F L | N | $O_2$-evolving core |
| WH 4.1 kDa | . S P G . S . . . . . | | |
| SP 3.9 kDa | K L P E A Y A F L S P I V D | C(*psb*K) | PS II membrane |
| WH 3.9 kDa | . . . . . I F N . ? ? ? | | |

[a] N, nuclear genome; C, chloroplast genome
[b] 5 kDa-I and -II, deduced from minor and major sequencing signals, respectively
[c] [8]

A hyphen denotes a deletion and a dot means the same residue as in spinach

[11,12], and a homologous ORF (or parts) has also been found in pea, spinach, wheat, tobacco, liverwort, *Euglena* and *Synechocystis* [9,10,12–14]. In tobacco and liverwort, possible ribosome-binding sites (GAG and GAA, respectively) can be assumed 5 bp upstream of the proposed initiation codons (fig.2). However, as the initiation codon in tobacco and spinach is ACG, which generally codes not Met but Thr, final assignment of the beginning of the gene may be reserved, even though the protein sequencing has confirmed its expression. Since the N-terminus of the 5 kDa-I protein corresponds to the second amino acid residue of ORF38, the homologous proteins of tobacco and liverwort are assumed to consist of 37 amino acid residues with molecular masses of 4366 and 4348 Da, respectively. These values are slightly lower than the value (5 kDa) estimated for the spinach and wheat proteins by mobility in SDS-PAGE. The protein deduced from the gene sequence contains a single membrane-spanning segment at the C-terminal side. Comparison of the tobacco sequence with the liverwort reveals 92% homology at amino acid level and 85% at nucleotide level. This homology is slightly lower than that (94% at amino acid level and 86% at nucleotide level) of the 4.8 kDa reaction center component (the product of *psb*I) [8].

When the ORF38 was reported in *Cyanophora* and *Euglena*, it was nominated to be *psb*I without identification of its product [12,13]. However, since the ORF36 located between the genes for $tRNA^{Ser}$ and $tRNA^{Gln}$ in chloroplast DNA has been appointed to *psb*I [16] based on its product identified as the 4.8 kDa reaction center component [8], we propose to designate the gene encoding the 5 kDa-I protein as *psb*L. In *Euglena*, the ORF38 and another ORF (denoted as ORF40/42 in [13]) were reported to be co-

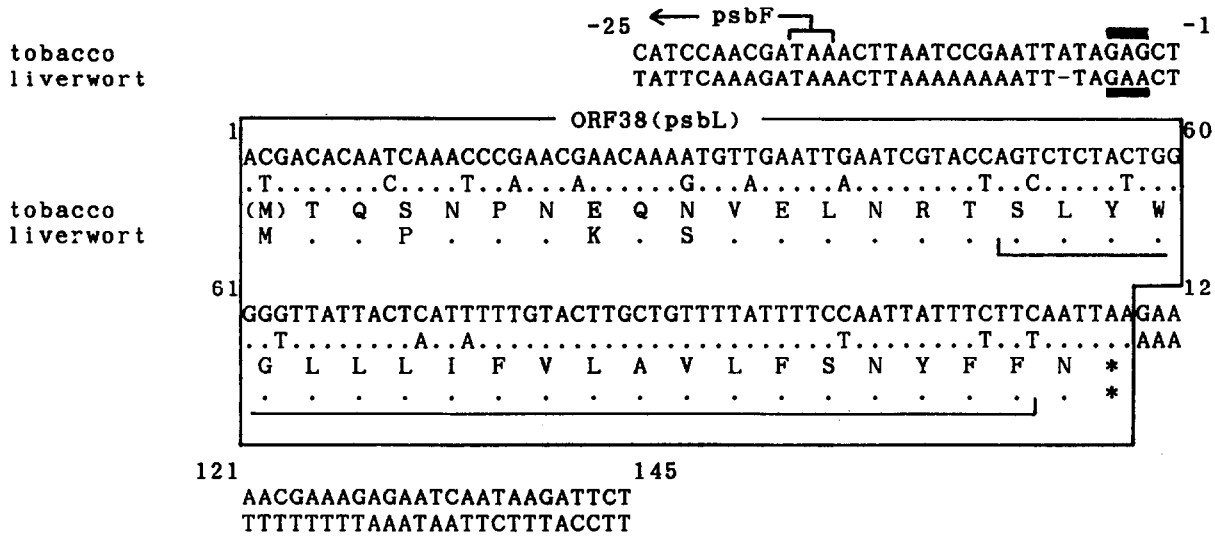



Fig.2. The nucleotide sequences of ORF38 (*psb*L) encoding the 5 kDa-I protein and its flanking region in the chloroplast DNA from tobacco [9] and liverwort [10]. Numbering starts at the ACG codon in tobacco at which translation is probably initiated. Corresponding amino acids deduced from the tobacco sequence are indicated below the codons. Nucleotide differences in the coding region between tobacco and liverwort and any resulting amino acid differences are shown below the tobacco sequence. A dot in the liverwort sequence indicates the same residue as in tobacco. An asterisk denotes a stop codon and a hyphen means a deletion. The identified gene is boxed. A possible ribosome-binding site is shown by bold lines above and below. Hydrophobic region, which may span the membrane, is indicated by a horizontal bracket. The termination codon for *psb*F is also shown.

transcribed with *psb*E and *psb*F [13]. In this study we could not identify an ORF40/42 product in spinach or wheat PS II membrane fragments. However, if we take into account our experience that the 4.8 kDa reaction center component was hardly identified in PS II membrane fragments or core complexes due to its extremely poor transfer efficiency to PVDF and its N-terminal block, it seems still possible to expect the presence of an ORF40/42 product in PS II.

The sequence determined from the major signals (denoted as 5 kDa-II protein) showed homology between spinach and wheat: many substitutions and insertion/deletion were found in the N-terminal region but most of the substitutions were conservative, keeping the number of charged residues unchanged. The same spinach sequence has already been reported as a hydrophilic 5 kDa protein which was extracted from PS II membrane fragments [1]. Our sequence data clearly demonstrate that the hydrophilic 5 kDa protein (identified as 5 kDa-II protein) is a component of the $O_2$-evolving core complex.

The sequences of the 6.1 kDa protein showed high homology between spinach and wheat (table 1). Namely, only 2 out of 20 amino acid residues were different between the two plants. The sequence did not correspond at all to any part of the tobacco or liverwort chloroplast genome, suggesting that it is encoded by the nuclear genome. This protein was associated with the PS II membrane but not retained in the $O_2$-evolving core complex. Although the staining intensity of the 6.1 kDa band was much higher in thylakoids than in PS II membrane fragments (fig.1), both bands gave the same amino acid signals with similar intensity, suggesting an almost quantitative recovery in the PS II membrane fragments. This discrepancy in staining intensity turned out to be due to comigration of another protein superimposed on the 6.1 kDa band of thylakoids: after acid pretreatment, the 6.1 kDa band from thylakoids, but not from PS II membrane fragments, produced another series of major amino acid signals, Met-Asn-Pro-Leu-Ile-Ala-Ala-Ala-..., which corresponded to the ATPase proteolipid encoded by a chloroplast gene, *atp*H [17].

A part of the 6.1 kDa sequence corresponded to

the recently reported sequences for two proteins with molecular masses of 7 and 6.5 kDa from the spinach preparation [7] analogous to our PS II membrane fragments. However, in our gel system the 6.1 kDa band of spinach, wheat and pea PS II has never split into such a doublet band as reported in [7]. Moreover, a degraded fragment of the 16 kDa extrinsic protein, which was not detected in our preparation, was obtained from their preparation [7]. These facts suggest that the reported heterogeneity of the 6.1 kDa protein might be an artefact due to proteolysis during membrane preparation.

The sequence of the 3.9 kDa protein corresponded to that recently reported for the 'K polypeptide' which is encoded by a chloroplast gene, *psb*K [6]. It is of note that the N-terminal seven residues of this protein are completely conserved in spinach and wheat (table 1). This conservation might be related with the proteolytic processing proposed to occur between the identified residues and the preceding segment consisting of 61 amino acid residues only found in the gene [6].

Although the 10.2 kDa wheat protein has an electrophoretic mobility different from the 11 kDa spinach protein, its N-terminal sequence indicates a clear homology between the two plants (table 1). The spinach sequence is identical with that already reported for a 10 kDa protein [1]. The gene sequence of this protein encoded by the nuclear genome was recently determined [18]. This protein is not retained in the $O_2$-evolving core complex (fig.1) and, therefore, may not be involved in $O_2$ evolution.

The sequences of 9.4 kDa and 7.7 kDa proteins from spinach and wheat were identical with those deduced from *psb*E [15] and *psb*H genes [19], respectively, indicative of the large subunit of cytochrome *b*-559 and the '10 kDa' phosphoprotein. Sequencing of the 4.4 kDa protein was successful only after acid pretreatment like the 4.8 kDa reaction center protein, indicative of N-terminal blocking [8]. This sequence revealed that it is the small subunit of cytochrome *b*-559 encoded by the *psb*F gene, as reported before for spinach [20]. It is of note that in wheat the initiation Met residue is absent in these three proteins (9.4, 7.7 and 4.4 kDa) as already reported for spinach [19–21]. The 5.5 kDa band from spinach revealed an unreported sequence (not shown). However, no corresponding band or amino acid sequence was found in wheat PS II membrane fragments. N-terminal sequencing is a simple but a powerful technique for identifying not only unknown proteins but also known proteins in different samples under different separation conditions.

*Acknowledgements:* We thank Dr T. Yasunaga (RIKEN) for computer-assisted homology search and Dr K. Narahashi (RIKEN) for helpful instructions for protein sequencing. This work was supported by a Grant on Solar Energy Conversion by Means of Photosynthesis awarded to The Institute of Physical and Chemical Research (RIKEN) by the Science and Technology Agency of Japan, and partly by a Grant on Frontier Research Program at RIKEN awarded by the Science and Technology Agency of Japan.

## REFERENCES

[1] Murata, N., Kajiura, H., Fujimura, Y., Miyao, M., Murata, T., Watanabe, A. and Shinozaki, K. (1987) in: Progress in Photosynthesis Research (Biggens, J. ed.) vol.I, pp.701–704, Martinus Nijhoff, Dordrecht.
[2] Ljungberg, U., Henrysson, T., Rochester, C.P., Åkerlund, H.-E. and Andersson, B. (1986) Biochim. Biophys. Acta 849, 112–120.
[3] Henrysson, T., Ljungberg, U., Franzen, L.-G., Andersson, B. and Åkerlund, H.-E. (1987) in: Progress in Photosynthesis Research (Biggens, J. ed.) vol.II, pp.125–128, Martinus Nijhoff, Dordrecht.
[4] Ikeuchi, M. and Inoue, Y. (1988) Plant Cell Physiol. 29, 1233–1239.
[5] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[6] Murata, N., Miyao, M., Hayashida, N., Hidaka, T. and Sugiura, M. (1988) FEBS Lett. 235, 283–288.
[7] Schröder, W.P., Henrysson, T. and Åkerlund, H.-E. (1988) FEBS Lett. 235, 289–292.
[8] Ikeuchi, M. and Inoue, Y. (1989) FEBS Lett., in press.
[9] 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.
[10] Ohyama, K., Fukuzawa, H., Kohchi, T., Shira, 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.
[11] Cantrel, A. and Bryant, D.A. (1987) in: Progress in Photosynthesis Research (Biggens, J. ed.) vol.IV, pp.659–662, Martinus Nijhoff, Dordrecht.
[12] Cantrel, A. and Bryant, D.A. (1988) Photosynth. Res. 16, 65–81.

[13] Cushman, J.C., Christopher, D.A., Little, M.C., Hallick, R.B. and Price, C.A. (1988) Curr. Genet. 13, 173–180.
[14] Hird, S.M., Willey, D.L., Dyer, T.A. and Gray, J.C. (1986) Mol. Gen. Genet. 203, 95–100.
[15] Herrmann, R.G., Alt, J., Schiller, B., Widger, W.R. and Cramer, W.A. (1987) FEBS Lett. 176, 239–244.
[16] Sugiura, M., Shinozaki, K., Tanaka, M., Hayashida, N., Wakasugi, T., Matsubayashi, T., Ohto, C., Torazawa, K., Meng, B.Y., Hidaka, T. and Zaita, N. (1987) in: Plant Molecular Biology (Von Wettstein, D. and Chua, N.-H. eds) pp.65–76, Plenum, New York.
[17] Sebald, W. and Hoppe, J. (1981) Curr. Top. Bioenerg. 12, 1–64.
[18] Lautner, A., Klein, R., Ljungberg, U., Reiländer, H., Bartling, D., Andersson, B., Reinke, H., Beyreuther, K. and Herrmann, R.G. (1988) J. Biol. Chem. 263, 10077–10081.
[19] Farchaus, J. and Dilley, R.A. (1986) Arch. Biochem. Biophys. 244, 94–101.
[20] Widger, W.R., Cramer, W.A., Hermodson, M. and Herrmann, R.G. (1985) FEBS Lett. 191, 186–190.
[21] Widger, W.R., Cramer, W.A., Hermodson, M., Meyer, D. and Gullifor, M. (1984) J. Biol. Chem. 259, 3870–3876.