Volume 314, number 1, 61-66 FEBS 11830 December 1992
© 1992 Federation of European Biochemical Societies 00145793/92/$5.00

The single-copy gene *psbS* codes for a phylogenetically intriguing 22 kDa polypeptide of photosystem II

N. Wedel[a], R. Klein[a], U. Ljungberg[a], B. Andersson[b] and R.G. Herrmann[a]

[a]*Botanisches Institut der Ludwig-Maximilians-Universität, Menzinger Str. 67, D-8000 München 19, Germany* and [b]*Department of Biochemistry, Arrhenius Laboratories, Stockholm University, S-10691 Stockholm, Sweden*

Received 5 October 1992

Recombinant phages that encode the complete precursor polypeptide for the 22 kDa polypeptide associated with photosystem II have been serologically selected from two λgt11 expression libraries made from polyadenylated RNA of spinach seedlings. The cDNAs hybridize to a 1.3 kb RNA species. The precursor protein is comprised of 274 amino acid residues and carries an N-terminal transit peptide of probably 69 amino acid residues. The mature protein exhibits four predicted transmembrane segments and is shown to be an integral component of photosystem II originating in a single-copy gene. The unique characteristics of this protein are: (i) it is the result of a gene-internal duplication of an ancestor with two membrane spans, (ii) a striking resemblance to LHC I/II, CP24/CP29 apoproteins, and ELIPs, although it does not bind chlorophyll and is present in cyanobacteria, and, as these proteins, (iii) it integrates into the membrane with uncleaved routing signals that display remarkable resemblance to patterns found in bipartite transit peptides.

22 kDa polypeptide gene; Photosystem II; *cab* genes; Phylogeny; Protein import; Protein assembly

1. INTRODUCTION

Photosystem II (PSII), which functions as a water-plastoquinone oxidoreductase in photosynthetic electron transport, is the most complex multiprotein assembly of thylakoid membranes, located predominantly in appressed membrane regions [1]. To date, more than 20 different PSII polypeptides have been described, disregarding the elaborate set of light-harvesting chlorophyll *a/b* apoproteins associated with the complex. As all other thylakoid membrane complexes, PSII is genetically bipartite [2,3]. The genes for nine polypeptide species of the core complex, all membrane-spanning have been located in plastid chromosomes of several higher plants. These encode the D1/D2 heterodimer (genes: *psbA*, *psbD*) carrying all the redox ligands of the primary photochemistry and possibly also the manganese cluster, the two-chain cytochrome b_{559}, (*psbE*, *psbF*), the 51 and 43 kDa chlorophyll *a* apoproteins (*psbB*, *psbC*) serving as internal antenna, the 10 kDa phosphoprotein (*psbH*), and several low molecular mass polypeptides [1,2,4]. At least six PSII proteins originate in nuclear genes, in addition to the lumenal extrinsic 33, 23 and 16 kDa polypeptides of the oxygen-evolving system (*psbO*, *psbP*, *psbQ*), a non-phosphorylated 10 kDa component (*psbR*) as well as components of apparent molecular masses of 5 and 22 kDa [5,6]. The 22 kDa protein exhibits an extreme lateral location to the appressed thylakoid regions [5] which is not affected by disassembly events following photoinhibitory-induced D1 protein degradation [7] and has recently been shown to be present in cyanobacteria [8], but its structure and functional significance are unknown. It has been suggested a structural role to bind extrinsic lumenal proteins [5] or to confer DCMU sensitivity to PSII [6,9].

We describe here the isolation and characterization of recombinant λgt11 phage that bear cDNA inserts for the 22 kDa polypeptide. The sequence of the isolated cDNA turned out to be highly intriguing both with regard to functional and phylogenetic aspects. Preliminary accounts of this work have been presented at the NATO Advanced Study Institute on Plant Molecular Biology, Elmau [2].

Correspondence address: R.G. Herrmann, Botanisches Institut der Universität München, Menzinger Str. 67, D-8000 München 19, Germany.

2. MATERIALS AND METHODS

The spinach variety used was *Spinacia oleracea* var. Monatol. All strategies and procedures employed in this study have been detailed before as follows: growth of material, isolation of polyadenylated RNA [10], library construction and screening [11,12], radiolabeling of DNA fragments, Northern and Southern analysis as well as nucleotide sequence analysis [12], the preparation of the 22 kDa polypeptide by ion exchange chromatography from spinach PSII membranes solubilized with Triton X-100 and Zwittergent TM-314 [5], the generation of antisera and their use in Western and immunoprecipitation analysis [5,12,13], transcription of cDNA, translation of the resulting transcripts in vitro, and in organello assays [13], fractionation of thylakoid membranes into grana and stroma lamellae [14], the preparation of BBY particles [15], and electrophoresis in various gel systems [5,10,13]. The accession number of the gene *psbS* in the EMBL data bank is X68552.

3. RESULTS AND DISCUSSION

Eight phage, recombinant for the 22 kDa polypeptide cDNA, were isolated from 7 × 10⁴ pfu of two cDNA libraries [11,12] in two screening cycles including selection with a polyclonal, specific antiserum [5] and subsequent hybridization of phage plating lifts with the serologically selected cDNAs. These phage were designated p6SocPSII22-1 to -8. Sequence analysis of their inserts, ranging from 300 to 1024 bp, uncovered that all of them fell into a single class disregarding a difference in a single nucleotide at position 459 (A instead of G) in p6SocPSII22-3 DNA. This overall sequence identity suggested that the cDNAs originated in the same gene and that the exchange of the nucleotide in position 459 is due to an error during cDNA synthesis or cloning in *E. coli*. This inference is consistent with copy number determinations via Southern analysis which indicated only a single gene for this polypeptide per haploid spinach genome (not shown). We designate this gene *psbS* [2].

The sequence of the largest, 1024 bp cDNA insert, detailed in Fig. 1, possesses only one open reading frame comprising 822 nucleotides which code for a (precursor) protein of 274 amino acid residues. The deduced molecular mass of this protein is 29.2 kDa. The first possible ATG triplet of the sequence probably operates as the translational initiation codon. It is preceeded by a 5′ untranslated region containing an in-frame stop codon at position −6 to −4. Moreover, its flanking sequences fit the eukaryotic translational initiation consensus, PuXX/ATG/G [16], and the presumed initiator Met is followed by an Ala residue characteristic for chloroplast transit sequences [17 and below]. The open reading frame continues with a 3′ untranslated region of 126 bp which contains two consensus polyadenylation signal sequences (Fig. 1) [18]. Hydropathy analysis [19] predicts a protein with four transmembrane segments (cf. Fig. 1 and below).

As all organelle polypeptides of nuclear origin, the 22 kDa polypeptide is expected to be synthesized as a precursor molecule with a transit peptide. Attempts to pinpoint its junction to the transit peptide by determining the N-terminal sequence of the mature protein failed, probably because this protein is N-terminally blocked. Similarly, application of von Heijne's rule does not predict a discrete cleavage site [20]. However, inspection of the sequence shows that the stretch of the first 69 amino acid residues possesses all attributes of a typical stroma-targeting chloroplast transit sequence [17]. This sequence starts with the residues MAQAM which are also found in the transit sequence of the extrinsic, lumenal 16 kDa polypeptide associated with the oxygen-evolving system [21], is positively charged, contains hydroxylated amino acid residues, and includes a predicted β-sheet at position 65–70 that could serve as the recognition signal for the stroma-located endopeptidase [17].

To verify this prediction and to substantiate that p6SocPSII22-cDNAs indeed encode a constituent polypeptide of PSII we examined whether ^{35}S-labelled translation products made in vitro from linearized (*Eco*RV) and transcribed p6SocPSII22-3 cDNA (T7

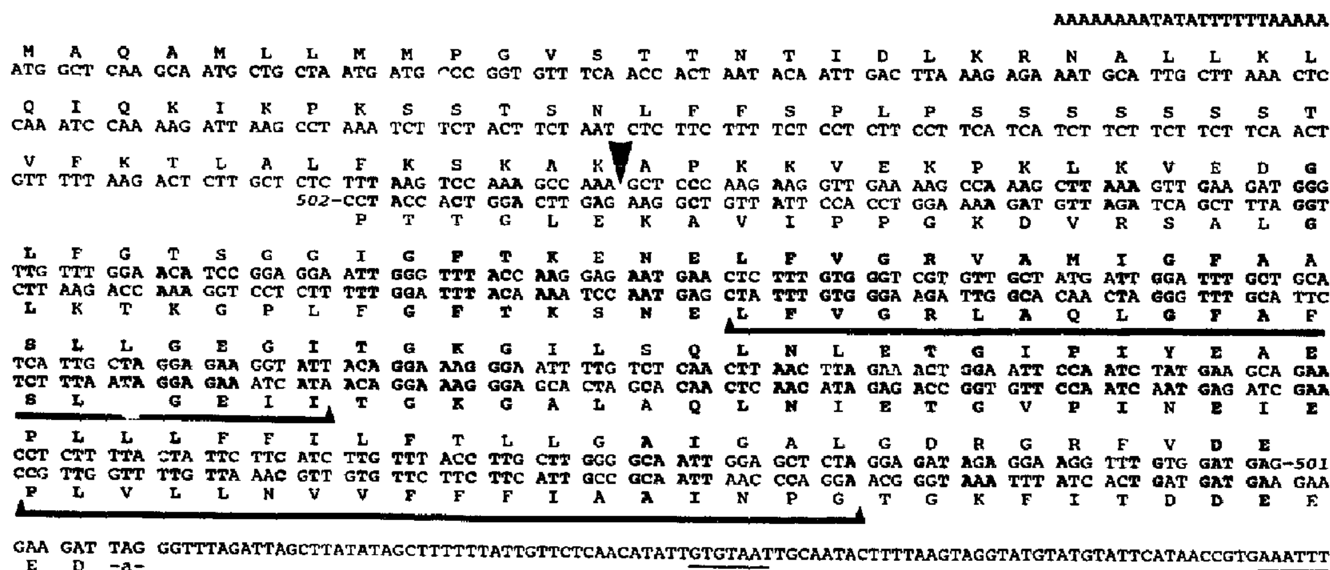

Fig. 1. Nucleotide sequence and derived amino acid sequence of p6SocPSII22-1 cDNA encoding the 22 kDa polypeptide associated with photosystem II from spinach. The cDNA sequence was dissected at nucleotide position 501 and aligned from nucleotides 213 and 502 to highlight the gene-internal sequence homology. The putative junction between transit peptide and mature protein is marked by an arrowhead; predicted transmembrane segments are indicated by brackets. Potential polyadenylation signals are underlined.

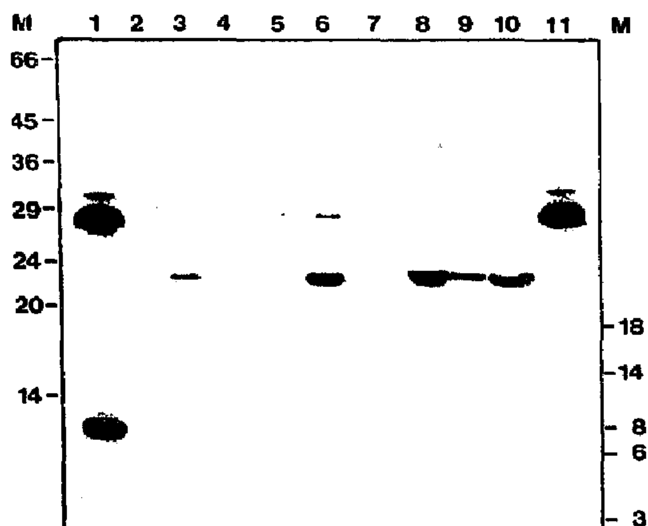

Fig. 2. Import, processing and assembly of the spinach 22 kDa precursor protein by isolated spinach chloroplasts (fluorography). The translation product obtained from p6SocPSII22-1 cDNA in vitro (lane 1) was incubated with isolated, intact spinach chloroplasts as quoted in section 2. Import, processing and intraorganelle location of the 22 kDa polypeptide was then monitored by reisolation and subfractionation of the organelle, by treatment of organelles and thylakoid membranes with trypsin, and separating the products on a 10-20% SDS polyacrylamide gel according to [22]. Lanes (2) stroma fraction, (3) membrane fraction, (4) membrane fraction treated with trypsin, (5) membrane fraction treated with 5% Triton X-100 and trypsin, (6) grana thylakoids, (7) stroma thylakoids. Lanes (8-10) Western analysis of (8) total chloroplast proteins, (9) membrane fraction, and (10) isolated photosystem II particles with anti-22 kDa serum. (11) Immunoprecipitate of the in vitro translation product shown in lane 1. Left and right: molecular weight standards in kDa.

polymerase) in a wheat germ extract could be imported, processed and finally assembled by isolated intact spinach chloroplasts (Fig. 2). The translated precursor could be selectively precipitated with the antiserum elicited against the 22 kDa polypeptide (Fig. 2, lane 11). Removal of its transit peptide, presumably upon entry into the organelle, shortens the protein to a component of 22 kDa apparent mass which had an identical disposition in the membrane and was indistinguishable in size from the native form, as judged by coelectrophoresis of the in organello import fractions with the authentic mature protein in several gel systems. After 30 min incubation in the in organello assay and subfractionation of thylakoids, the newly imported protein is exclusively found in grana lamellae or BBY-type of photosystem II-enriched membrane fractions (Fig. 2, lanes 6, 7, 10 and data not shown) in agreement with previous biochemical data [5,7]. This thylakoid fraction contains also the only immunologically detectable membrane polypeptide in Western blots. The deduced mass of 21.7 kDa of a 205 residue mature protein starting with Ala at position 70 corresponds well with that estimated on the basis of electrophoretic mobility. Furthermore, the amino acid composition calculated from the sequence data of the derived polypeptide fits with that determined for the authentic protein [5] within the limits of error, except for isoleucin (19 residues predicted, 12 determined), leucine (18/22) and phenylalanine (17/12), although their relative amounts fit the trend of analysis. Collectively these and subsequently outlined data substantiate the identity of the isolated cDNA clones and demonstrate that p6SocPI22-1 encodes the entire 22 kDa precursor polypeptide.

The available biochemical data do not give much information on the function of the 22 kDa protein. However, closer inspection of the sequence data has uncovered three striking aspects of considerable general interest with new perspectives for the understanding of the phylogeny and biogenesis of this and related proteins.

(1) Figs. 1 and 3 illustrate that the mature 22 kDa protein must be the result of a gene-internal duplication of an ancestor protein containing two transmembrane segments, since the first and third, and the second and fourth of these segments respectively, correspond to each other. This homology is also evident at the DNA level. Gene duplications and subsequent divergence have been crucial in the phylogeny of energy transducing membranes. For instance, various thylakoid structures have evolved by duplications of primordial DNA segments and subsequent divergence, at least in some instances from homodimeric to heterodimeric (asymmetric) forms [23]. However, in all these cases this has resulted in two individual proteins in contrast to the 22 kDa protein. The selective advantage of this difference is unknown.

(2) Sequence comparison (Figs. 1 and 3), secondary structure predictions (Fig. 4), and the conservation of discrete epitopes leave no doubt that the 22 kDa protein represents a distant member to the CP24/CP29 genes and the *cab* I/II and ELIP (early light-inducible proteins) gene families [2,24], although the isolated 22 kDa protein does not carry chlorophyll or any other pigment [5], consistent with the absence of histidine, asparagine and methionine residues presumed to act as chlorophyll ligands (Fig. 4). Moreover, the data are commensurate with the view that the 22 kDa protein represents an ancient member of these proteins. This is primarily borne out by serological crossreactions of the spinach antiserum which suggest that even cyanobacterial thylakoid membranes, although not equipped with a chlorophyll *a/b* antenna type, possess an equivalent of this protein [8], and that most subclasses of this putative superfamily have apparently lost the fourth predicted transmembrane segment (Fig. 4). If the existence of a prokaryotic predecessor for the 22 kDa protein is correct, this inherently implies that, at the eukaryotic level, the gene(s) must have been translocated to the nucleus, since all these proteins originate now in nuclear genes. They *subsequently* diverged into families of multiple biologically active peptides, including the development of

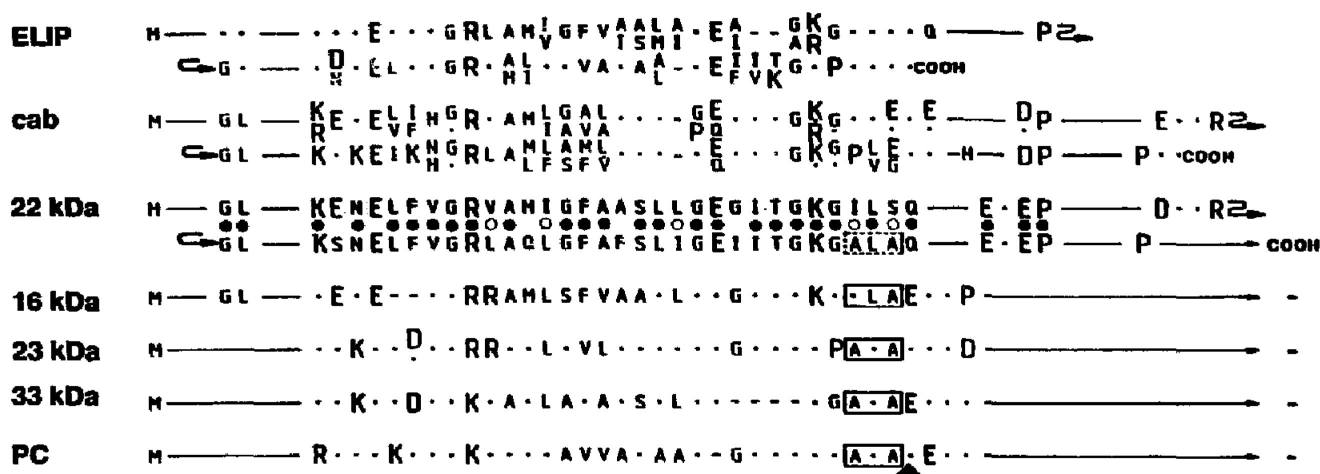

Fig. 3. Comparison of sequences for hydrophobic domains in CAB and lumenal proteins that operate as membrane anchors and/or targeting signals (cf. Fig. 4). The two halfs of the ELIP, CAB and 22 kDa proteins are aligned to illustrate the gene duplications. For the first two subfamilies, a consensus was derived from published data for the pea and tomato proteins (refs). Amino acid identities in the duplicated N- and C-terminal parts of the 22 kDa protein are marked by dots, conservative replacements by open circles. Variant residues in the ELIP and CAB proteins are indicated as alternatives. Small dots designate non-conserved residues, dashes refer to lacking residues, and lines mark sequences of various lengths. The AXA motifs preceding the respective cleavage sites (arrowhead) are boxed, a related motif in uncleaved position in the 22 kDa protein is boxed by a dotted line. Note the conserved charge pattern that is indicated by large letters and also found in various other components.

the chlorophycean antenna type, that perform coordinate function in stress protection, light harvesting and conversion.

The phylogenetic translocation of genes from plastids to the nucleus – undisputable from the comparison of equivalent prokaryotic and eukaryotic genes and of operons in prokaryotes and plastid chromosomes [25] – must have been accompanied by the acquisition of DNA segments ensuring a proper expression in the nucleo/cytosolic compartment and import of the resulting protein into the organelle. The observation that functionally equivalent plastid import sequences [17] and the promotors of the corresponding genes [2,26] are very different within one organism between different proteins and genes respectively, but relatively conserved for the same protein/gene in different organisms is indicative of separate phylogenetic translocation events. The dissimilarity of transit peptides (and promotors) between the individual groups of the superfamily, their similarity within each group even between different organisms as well as the different chromosomal location of the genes and gene subfamilies as judged from RFLP mapping studies with the spinach relative *Beta vulgaris* (sugar beet; Pillen, unpublished) suggest that the evolution of the obvious hierarchy of these functionally diverged branches of subfamilies included two qualitatively different events. Most members of a given protein subclass may have been doubled from a preexisting translocated gene (or a duplicated copy) and the resulting gene sequences drifted through time, in line with the fact that plant gene families often tend to be relatively conserved, clustered in the genome, and arranged in head-to-tail fashion, but the superfamily itself is very likely the result of multiple translocation events. Phylogenetic hierarchy may even be seen in wider perspective. The comparison of the amino acid sequences of a large number of antenna polypeptides from purple bacteria and cyanobacteria with those from chlorophyll *a* and *a/b* complexing proteins of plants (peripheral LHC complexes I and II, PSI and II core complexes) and of the PSI and -II reaction centres revealed surprising phylogenetic relationships which suggest both a common origin between these polypeptides and that the diversity of antenna systems seen today may have been derived from specific protein moduls (H. Zuber, personal communication).

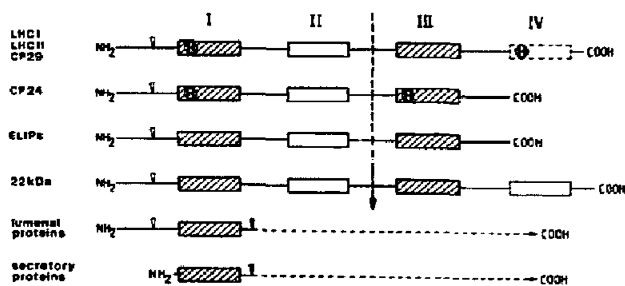

Fig. 4. Scheme illustrating the relationship of the so-called gene superfamily. The scheme includes domains of lumenal and secretory proteins functional in protein targeting (cf. Fig. 3). It is based on sequence comparison and hydropathy analysis. The vertical arrow marks the junction of the gene-internal duplication; arrowheads indicate processing sites. Homologous predicted transmembrane segments are numbered I to IV and the two related pairs presented as hatched and empty boxes, respectively. H designates histidine residues that may act as potential chlorophyll ligands. The arrowheads indicate processing sites.

The genetic setup of the eukaryotic cell is compatible

with this hypothesis, because the genes of the highly reiterated plastome per cell [27] would outnumber an intracellularly relocated gene-bearing DNA fragment. As promiscuous DNA [28] indicates, such gene transfers should be relatively frequent as are the chances for picking up a DNA segment with appropriate features for targeting [29], and probably expression. Usually however such a transfer will confer no selective advantage and an extra gene copy will usually be lost. Current ideas about gene duplication suggest the existence of stringent mechanisms for keeping to a minimum number of genes and duplications would be tolerated only where they are of functional use [30,31]. Thus, only if a transfer occurs into a favorable genetic context, may the nuclear gene copy be selected and its duplication in combination with advantageous mutations will preserve a new gene and the old. More detailed knowledge about chromosomal arrangements of these genes and their promotors is therefore of considerable interest.

(3) The distribution of topogenic sequences that contribute to protein import involving targeting, trafficking and assembly is quite complex. Considering only two elements essential for the integration into or translocation of a protein across the thylakoid membrane, charged residues that interact with the membrane surface and hydrophobic residues that contribute to partitioning into the thylakoid lipid phase [32], these may be found in both parts of a precursor molecule, transit peptide and mature protein [32,33]. For instance, the few known hydrophilic lumenal proteins plastocyanin and the three extrinsic polypeptides of the oxygen-evolving system, but also transmembrane proteins such as subunit 3 of PSI or subunit CFo-II of the thylakoid ATP synthase, that possess bipartite, stroma- *plus* thylakoid-targeting transit peptides (Fig. 3) [17] carry these epitopes in the transitory part of the molecule. On the other hand, the Rieske FeS protein [32] and, relevant in the context of this paper, the proteins of the *cab* superfamily, despite their dissimilar transit peptides, operate with (hydrophilic) transitory import and *internal* uncleaved thylakoid-targeting sequences [33,34]. Unfortunately, all attempts to determine precisely the exact nature of these signals failed (e.g. [35]) probably because the integration and assembly of these integral *polytopic* proteins depends on the cooperative effects of several such domains. The 22 kDa protein is a particularly interesting candidate of the *cab* gene superfamily in this context. If the helices 1 and 3, and 2 and 4 respectively, are functionally, and not only structurally equivalent helix 2 cannot be critically involved in signaling since its equivalent helix 4 is lacking in CP24, CP29, LHCPs and ELIPs (Fig. 4). Furthermore, sequence alignment discloses a remarkably conserved pattern (charged residues) between the mature moiety of the 22 kDa protein and transit peptides, specifically of the lumenal proteins (Fig. 3). Substitution of such conserved residues rather than following the generally used strategy of deleting protein segments [34,35] may bypass the risk caused by changing two parameters simultaneously and provide easier access to mechanisms and regulation of protein traffic or to processes of subcellular differentiation.

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft (SFB 184) and by a Postdoctoral Fellowship (U.L.) of the Swedish Natural Science Research Council.

REFERENCES

[1] Andersson, B. and Styring, S. (1991) in: Current Topics in Bioenergetics (Lee, C.P. ed.) vol. 16, pp. 2–81, Academic Press, San Diego.
[2] Herrmann, R.G., Oelmüller, R., Bichler, J., Schneiderbauer, A., Steppuhn, J., Wedel, N., Tyagi, A.K. and Westhoff, P. (1991) in: Plant Molecular Biology 2 (Herrmann, R.G. and Larkins, B.A. eds.) NATO ASI Ser. A, vol. 212, pp. 411–427, Plenum, New York/London.
[3] Andersson, B. and Herrmann, R.G. (1988) in: Plant Membranes, Structure, Assembly and Function (Harwood, J.L. and Walton, T.J. eds.) pp. 33–45, Biochem. Soc., London.
[4] Erickson, J.-M. and Rochaix, J.-D. (1992) in: Topics in Photosynthesis (Barber, J. ed.) vol. 11, pp. 101–178, Elsevier, Amsterdam.
[5] Ljungberg, U., Akerlund, H.-E. and Andersson, B. (1986) Eur. J. Biochem. 158, 477–482.
[6] Henrysson, T., Ljungberg, U., Franzén, L.-G., Andersson, B. and Akerlund, H.-E. (1987) in: Progress in Photosynthesis Research (Biggins, J. ed.) vol. II, pp. 125–128, Martinus Nijhoff, Dordrecht.
[7] Hundal, T., Virgin, I., Styring, S. and Andersson, B. (1990) Biochim. Biophys. Acta 1017, 235–241.
[8] Nilsson, F., Andersson, B. and Jansson, C. (1990) Plant Mol. Biol. 14, 1051–1054.
[9] Ghanotakis, D.F., Demetrion, D.M. and Yocum, C.F. (1987) Biochim. Biophys. Acta 891, 15–21.
[10] Westhoff, P., Nelson, N., Bünemann, H. and Herrmann, R.G. (1981) Current Genet. 4, 109–120.
[11] Tittgen, J., Hermans, J., Steppuhn, J., Jansen, Th., Jansson, C., Andersson, B., Nechushtai, R., Nelson, N. and Herrmann, R.G. (1986) Mol. Gen. Genet. 204, 258–265.
[12] Lautner, A., Klein, R., Ljungberg, U., Bartling, D., Andersson, B., Reinke, H., Beyreuther, K. and Herrmann, R.G. (198?) J. Biol. Chem. 263, 10077–10081.
[13] Westhoff, P., Jansson, Ch., Klein-Hitpaß, L., Berzborn, R., Larsson, Ch. and Bartlett, S.G. (1985) Plant Mol. Biol. 4, 137–146.
[14] Kyle, D.J., Ohad, I. and Arntzen, C.J. (1984) Proc. Natl. Acad. Sci. USA 81, 4070–4074.
[15] Berthold, D.A., Babcock, G.T. and Yocum, C.F. (1981) FEBS Lett. 134, 231–234.
[16] Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) EMBO J. 6, 43–48.
[17] von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535–545.
[18] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 203, 211–214.
[19] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[20] Gavel, Y. and von Heijne, G. (1990) FEBS Lett. 261, 455.
[21] Jansen, Th., Rother, Ch., Steppuhn, J., Reinke, H., Beyreuther, K., Jansson, C., Andersson, B. and Herrmann, R.G. (1987) FEBS Lett. 216, 234–240.
[22] Schägger, H. and von Jagow, G. (1987) Anal. Biochem. 166, 368–379.
[23] Büttner, M., Xie, D.-L., Nelson, H., Pinther, W., Hauska, G. and Nelson, N. (1992) Proc. Natl. Acad. Sci. USA 89, 8135–8139.
[24] Green, B.R., Pichersky, E. and Kloppstech, K. (1991) Trend Biochem. Sci. 16, 181–186.

[25] Palmer, J.D. (1992) in: Plant Gene Research (Herrmann, R.G. ed.) vol. 7 Cell Organelles, pp. 99–133, Springer, Vienna/New York.
[26] Flieger, K., Tyagi, A., Sopory, S., Csèplö, A., Herrmann, R.G. and Oelmüller, R. (1992) Plant J. (in press).
[27] Herrmann, R.G., Westhoff, P. and Link, G. (1992) in: Plant Gene Research (Herrmann, R.G. ed.) vol. 7, Cell Organelles, pp. 275–349, Springer, Vienna/New York.
[28] Ellis, J. (1982) Nature 299, 678–679.
[29] Baker, A. and Schatz, G. (1987) Proc. Natl. Acad. Sci. USA 84, 3117–3121.
[30] Ainsworth, C.C., Miller, T.E. and Gale, M.D. (1987) Gen. Res. 49, 93–103.
[31] Mardsen, J.E., Schawager, S.J. and May, B. (1987) Genetics 116, 299–311.
[32] Bartling, D., Clausmeyer, S., Oelmüller, R. and Herrmann, R.G. (1990) Bot. Mag. (Tokyo), Spec. Iss. 2, 119–144.
[33] Lamppa, G.K. (1988) J. Biol. Chem. 263, 14996–14999.
[34] Cai, D., Herrmann, R.G. and Klösgen, R.B. (1992) Plant J. (in press).
[35] Reinero, A. and Tobin, E.M. (1991) Photosynth. Res. 30, 25–33.