Volume 314, number 1, 67–71 FEBS 11829 December 1992

Characterization of a spinach *psbS* cDNA encoding the 22 kDa protein of photosystem II

Sunyoung Kim[a], Peter Sandusky[a], Neil R. Bowlby[b], Ruedi Aebersold[c], Beverley R. Green[d], Sari Vlahakis[c], Charles F. Yocum[a] and Eran Pichersky[a]

[a]*Biology Department, University of Michigan, Ann Arbor, MI 48109, USA,* [b]*MSU/DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824, USA,* [c]*Biomedical Research Centre and Biochemistry Department and* [d]*Botany Department, University of British Columbia, Vancouver, BC, V6T 1WS Canada* and [e]*Protein and Carbohydrate Structure Facility, University of Michigan, Ann Arbor, MI 48109, USA*

Received 9 September 1992

An intrinsic 22 kDa polypeptide is found associated with the oxygen-evolving photosystem II (PSII) core complex in all green plants and cyanobacteria so far examined, although it does not appear to be required for oxygen evolution. Amino acid sequence information obtained from the purified 22 kDa protein was used to construct a probe that was employed to isolate a full-length cDNA clone encoding the 274-residue precursor of the 22 kDa protein. Hydropathy plot analysis predicts the existence of four membrane-spanning helices in the mature protein. The two halves of the approximately 200-residue mature protein show high sequence similarity to each other, suggesting that the *psbS* gene arose from an internal gene duplication. The 22 kDa protein has some sequence similarity to chlorophyll *a/b*-binding proteins.

Photosystem II; Polypeptide; Chlorophyll *a/b*-binding (CAB) protein; Gene evolution

1. INTRODUCTION

Photosystem II (PSII) is a multi-subunit structure that contains the catalytic components necessary for light-driven oxidation of water to O_2 coupled to reduction of plastoquinone. Biochemical resolution has identified three functional protein groups within PSII: a core complex (47, 43, 34 (D1), 32 (D2), and 9 and 4.5 kDa (cytochrome b_{559}) species), a light-harvesting chlorophyll–protein complex (LHCII), and an ensemble of extrinsic polypeptides of 33, 23, and 17 kDa. An examination of O_2-evolving PSII preparations from spinach has revealed the presence of additional intrinsic protein components (28, 22, 10, 5–3 kDa species) [1]. Removal of the 22 and 28 kDa polypeptides modifies the ability of PSII to reduce exogenous acceptors (substituted *p*-benzoquinones, $Fe(CN)_6^{3-}$), and the sensitivity of these reactions to the herbicide DCMU is also weakened. However, removal of the 28 kDa protein appears to produce the major change observed with respect to acceptor reduction in PSII [2].

Thus, the role of the 22 kDa PSII protein remains unclear. Ljungberg et al. [3] have examined the associations among PSII polypeptide components after solubilization of the complex in detergent. Using antibodies against the extrinsic 33 and 23 kDa PSII polypeptides, it could be shown that 24, 22 and 10 kDa proteins were co-precipitated by antisera directed against the extrinsic proteins. On the basis of these results, it was proposed that the 24, 22 and 10 kDa polypeptides are closely associated with the PSII complex, and that they might be required for binding of the extrinsic polypeptides. In this communication we report the sequence of a cDNA clone encoding the 22 kDa polypeptide associated with the PSII reaction center complex and analyze the sequence of the encoded polypeptide.

Correspondence address: E. Pichersky, Biology Department, University of Michigan, Ann Arbor, MI 48109, USA. Fax: (1) (313) 747-0884.

2. MATERIALS AND METHODS

2.1. *Isolation and protein sequencing of spinach 22 kDa polypeptide*

PSII reaction center complex was prepared from market spinach as previously described [4,5]. From this preparation, referred to as G&Y, the 22 kDa protein was prepared for amino acid sequencing in two ways.

(i) A 0.5 mg chlorophyll/ml suspension of the G&Y preparations was solubilized with 0.5% dodecyl maltoside (DM) on ice for 10 min in a medium containing 20 mM MES, pH 6, 2 mM $CaCl_2$, and 400 mM sucrose. The detergent was added slowly as a solid. After solubilization the RC complex–DM mixture was centrifuged (40,000 × *g*, 15 min) to remove undissociated material. The supernatant was then loaded onto a Pharmacia Fast Flow S cation exchange column attached to a Pharmacia FPLC system. The column was pre-equilibrated with solubilization buffer and packed in ice before the sample was loaded. The green material (containing O_2-evolving material that retains the 28 kDa protein) eluted in the void volume. When all traces of chlorophyll had been washed from the column, the 22 kDa protein was eluted with 2 M NaCl (Fig. 1A) The colorless solution formed a white precipitate when dialyzed against 20 mM MES, pH 6. The precipitated 22 kDa protein was collected by centrifugation, resolubilized in 1% DM/2 M NaCl and stored at −45°C until analyzed. Protein

sequence analysis was performed on fragments prepared by the WACO Chemicals lysyl-endopeptidase method (digestion in 4 M urea, 100 mM Tris (pH 8) overnight at 37°C). Proteolytic fragments were separated on a 0.2 mm Vydac C4 column attached to an Applied Biosystems 130 HPLC system by elution with an acetonitrile–water gradient. Sequences determination was performed on an Applied Biosystems 473 protein sequencer.

(ii) Alternatively, the G&Y preparation was washed with 1 M NaCl, 10 mM MES, pH 6.5, to remove the extrinsic 23 kDa polypeptide of the water-splitting complex [6]. The polypeptides were resolved on a 12–22% polyacrylamide gradient gel according to Pichersky et al. [7], and the 22 kDa polypeptide was identified using antibodies (Fig. 1B) raised to the purified protein [6]. After transferring the 22 kDa protein band to PVDF membranes, tryptic peptides were obtained and sequenced as in Pichersky et al. [7].

2.2. *Gene cloning and sequencing*

Oligonucleotides were prepared to correspond to peptide sequences obtained in this study and elsewhere [8]; for KAKAPKKV (residues 67–74), the sense 23-mer 5'-AA(A/G)GC(A/G/C/T)AA(A/G)GC(A/G/C/T)CC(A/G/C/T)AA(A/G)AA(A/G)GT; and for VEDGLF (residues 81–86), the antisense 20-mer 5'-AA(A/G/C/T)A(A/G)(A/G/C/T)CC(A/G)TC(C/T)TC(A/G/C/T)AC(C/T)TT. Plasmid DNA (0.1 μg) from a spinach leaf cDNA library was used as the source for DNA amplification with *Taq* DNA polymerase (35 cycles at 92, 37, and 72°C for 2, 3, and 2 min, respectively). After amplification, the 60-bp amplified DNA fragment was purified from a 2% agarose gel by electroelution and labelled with ^{32}P using the random-primer method. The labelled DNA was used as a probe to screen spinach leaf cDNA libraries in the lambda phage vector, Uni-Zap XR (a gift from Drs. W. Gruissem and J. Narita). Probe labeling, library screening, and the isolation and characterization of clones was performed as previously described [9,10]. DNA was sequenced on both strands by the enzymatic method [11].

3. RESULTS

N-Terminal sequencing of the purified spinach 22 kDa protein was unsuccessful, indicating that the N-terminus is probably blocked. Proteolysis as described in section 2 produced several peptides, and four unique fragments were successfully sequenced (Fig. 2). One peptide sequence, LKVEDGLFGTSGGIGPT, matched perfectly the last 12 amino acids of the N-terminal sequence for the precursor of a 22 kDa spinach PSII protein reported by von Heijne et al. [8]. The latter sequence was derived from the nucleotide sequence of a cDNA cloned, but neither the complete amino acid sequence nor any portion of the DNA sequence has been reported.

Two regions of this protein (see section 2) were used to prepare synthetic oligonucleotides for amplification of a 60-bp gene segment by the PCR method. The amplified DNA segment was used as a probe to screen a spinach leaf cDNA library. A single clone was thus identified. The nucleotide sequence of this clone was determined and is presented in Fig. 2. The clone contains 1,012 nucleotides and has an open reading frame encoding a polypeptide of 274 amino acids, starting with a methionine. The four peptide sequences previously determined are all in complete agreement with the predicted sequence determined from the DNA sequence (Fig. 2). In addition, the sequence of the first 90 amino acids is in perfect agreement with the partial sequence of the spinach '22 kDa' clone reported by von Heijne et al. [8]. We thus conclude that we have isolated a clone encoding the precursor of the spinach 22 kDa polypeptide. In accordance with the nomenclature for PSII genes proposed by Hallick [12], we designate the gene as *psbS*. Because the correct open reading frame in the isolated clone has no ATG codons upstream of the first ATG codon indicated in Fig. 2, and a stop codon exists in that region, we conclude that the isolated *psbS* cDNA clone contains the entire coding region.

Since we were unable to obtain the N-terminal sequence of the mature polypeptide, the precise size of the transit peptide could not be determined. However, the position of the tryptic peptide, LKVEDGLFGTSGGIGFT, in the 22 kDa protein (Fig. 2) indicates that the mature protein includes at least 198 residues. Secondary structure analyses (Figs. 3 and 4) predict that the mature 22 kDa protein possesses four hydrophobic regions of sufficient length to be membrane-spanning helices, and two regions of β-turns, preceding the first and third helices. Overall, the protein has a very hydrophobic character. There appear to be only three short hydrophilic regions; the mature N-terminus, the region which lies between the second and third helices, and the C-terminus.

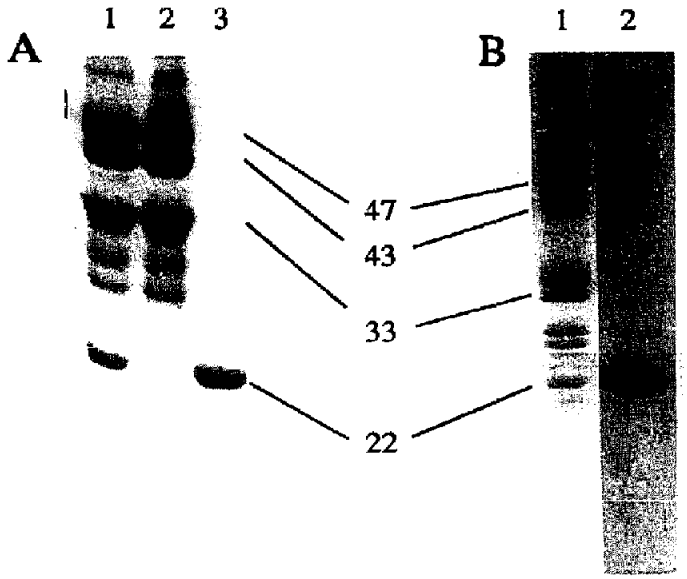

Fig. 1. Isolation and purification of the 22 kDa protein. (A) Polypeptides were resolved on a 12.5% SDS-polyacrylamide gel with 4 M urea as described in [4], and stained with Coomassie blue. Lane 1, G&Y preparation; lane 2, DM-solubilized G&Y components that eluted from the Fast Flow S column in the void volume; lane 3, purified 22 kDa protein after elution from Fast Flow S column with 2 M NaCl. (B) Polypeptides were resolved on an SDS-polyacrylamide gel (no urea) with a 12–22% polyacrylamide gradient as described in [7]. Lane 1, G&Y preparation stained with Coomassie blue; lane 2, G&Y preparation immunoblotted with anti-22 kDa antibodies. Faint bands in the approximately 28 kDa range are the CP29 proteins, which weakly cross-react with the anti-22 kDa antibodies.

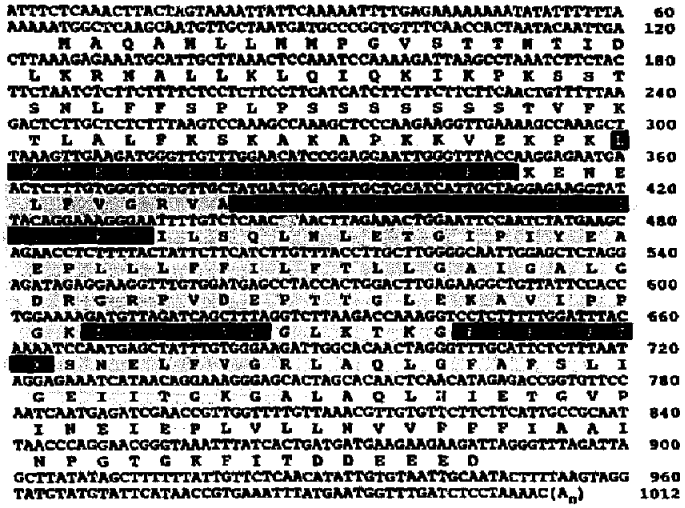

Fig. 2. The nucleotide sequence of the spinach *psbS* cDNA clone. The predicted amino acid sequence of the encoded polypeptide is shown below the DNA sequence. The proteolytic peptide sequences are highlighted.

Examination of the amino acid sequence of *psbS* itself revealed that the sequence from residue 93–144 is highly similar to another region in the protein, from residue 197–248 (Fig. 3B). These two regions have 38/52 identical matches (73% identity), with conservative substitutions in most of the other positions. The region from residue 93–144 includes the first putative transmembrane helix (TMH) and part of the second TMH, and the region from residue 197–248 corresponds to the third TMH and part of the fourth TMH.

Database searches revealed no sequences with substantial similarities (>50%) to the protein encoded by the *psbS* clone. However, significant similarity was observed with chlorophyll *a/b*-binding (CAB) proteins (Fig. 4). Although the overall similarity was <20%, a higher level of similarity was found in some regions of the protein, notably between the first and third TMHs of the 22 kDa polypeptide and the first and third TMHs of the CAB polypeptides, respectively (up to 50% sequence identity with some CAB sequences, with additional conservative substitutions). In addition, the region in the 22 kDa protein from residue 50–73 (which probably includes part of the transit peptide and the N-terminus of the mature polypeptide) was found to match in 18 positions (=75%) with the corresponding region in the CP29 Type I protein, but only negligibly with the corresponding region in other CAB proteins (comparison not shown).

4. DISCUSSION

The 22 kDa polypeptide is part of the middle layer of PSII polypeptides located between the reaction center sensu stricto (polypeptides obligately required for charge separation) and the light-harvesting antennae, LHCII and CP24 [13,14]. The polypeptides in this layer include those of the chlorophyll *a* antennae, CP47 and CP43, the chlorophyll *a/b* antenna, CP29, a 10 kDa polypeptide and several minor components. Removal of the 22 and 10 kDa polypeptides has little effect on the kinetics of water oxidation, but the fact that the 22 kDa polypeptide is found in PSII of all plants examined [13], and also in PSII of *Synechocystis* 6803 [15], suggests that it must have some role in PSII function or assembly. Its amino acid sequence indicates that the 22 kDa protein is a highly hydrophobic protein, with four potential membrane-spanning regions. However, additional experiments are required to determine the exact function of this protein and its orientation in the membrane.

Previous observations that both polyclonal and monoclonal antibodies specific for the CAB polypeptides of CP29 showed a small degree of cross-reaction with the 22 kDa polypeptide [5], and that, conversely, the anti-22

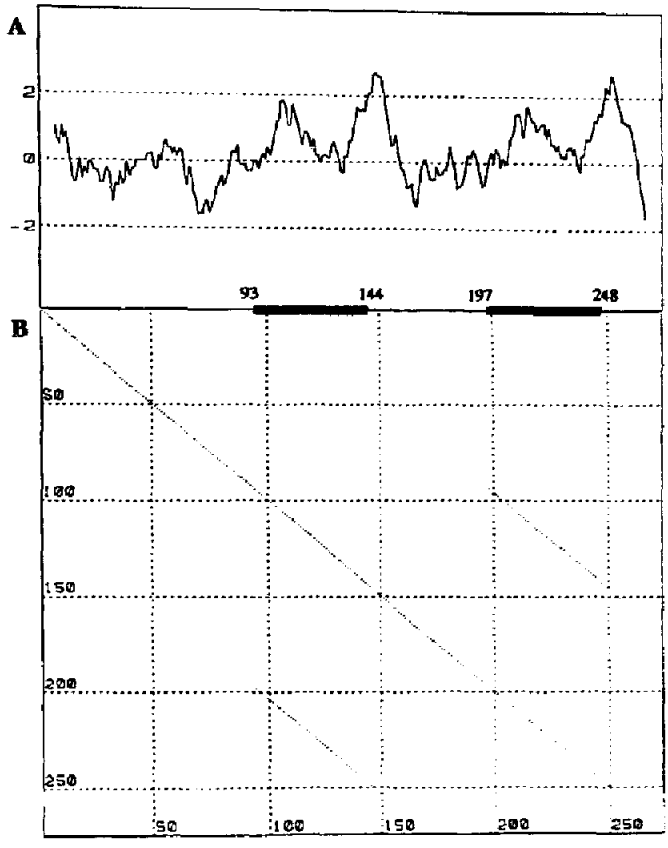

Fig. 3. Analyses of the sequence of the protein encoded by the spinach *psbS* cDNA clone. (A) Hydropathy plot based on the method of Kyte and Doolittle [21] with a window size of 15. (B) Comparison of the 22 kDa sequence with itself (using the Eugene software's Unit Odds Cost-Matrix program with a window size of 15). A dot represents residue identity in a larger context of high sequence identity. The internal duplication (regions 93–144 and 197–248) is clearly visible.

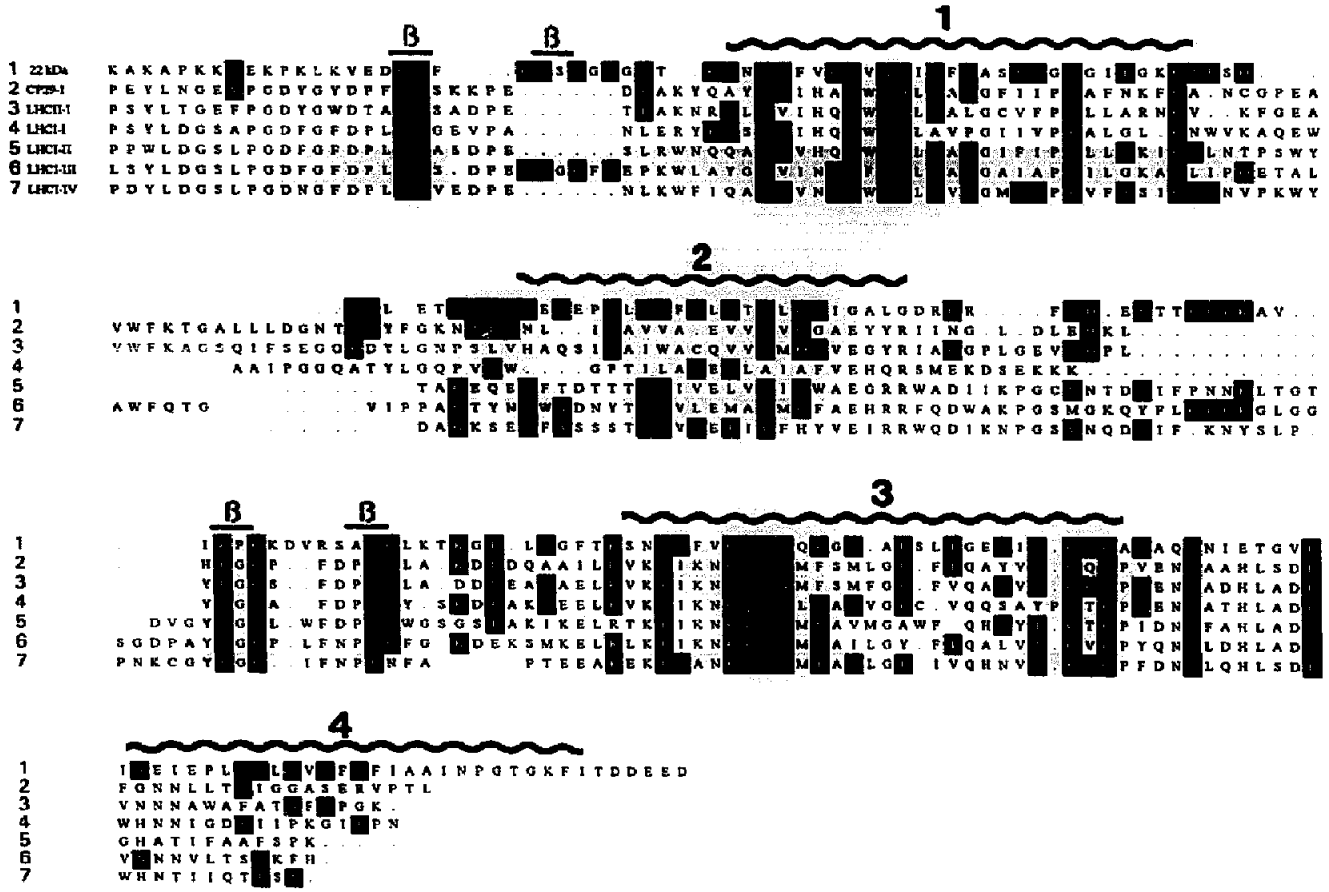

Fig. 4. Comparison of the 22 kDa sequence with that of several tomato CAB proteins of PSI and PSII (see [22] for complete references for the tomato sequences). Matches between a residue in the 22 kDa protein with a residue in any of the CAB polypeptides are highlighted. The structural features predicted by a combination of methods [23] (the four transmembrane helices, the β-turn regions) are also indicated. The transit peptides and mature N-termini of the proteins are not shown. A dot represents a gap introduced to maximize identity.

kDa antibodies cross-reacted weakly with polypeptides of CP29 ([12,14], see also Fig. 1B), suggested that the 22 kDa protein is structurally related to CAB proteins. Comparisons of the 22 kDa protein sequence with those of the tomato CAB proteins of both PSII and PSI (Fig. 4) confirm this hypothesis. It should be noted that no evidence has so far been reported that the 22 kDa polypeptide binds chlorophylls or any other pigments. The degree of similarity of the 22 kDa protein to CAB proteins is comparable to that between CAB proteins and ELIPs (early light-induced proteins), another group of proteins not known to bind pigments [16–18]. The highest level of sequence identity among all the proteins from these three groups occurs in the first and third TMHs (Fig. 4 and additional comparisons not shown).

Since the *psbS* polypeptide appears to be the result of the duplication of a two-helix protein to form a four-helix protein, and its first and third TMHs can be aligned with the first and third TMHs of the 3-TMHs CAB polypeptides (the similarity between the first and third TMHs of CAB proteins was previously noted [19]), we hypothesize that their most recent common ancestor may have had four helices, the last one of which was subsequently lost in the CAB lineage. Secondary structure analysis also indicates a high probability of β-turns in the analogous region in the CAB and 22 kDa proteins, although little primary sequence similarity exists in these regions (Fig. 4). These observations suggest that the tertiary structures of these proteins may still be similar, consistent with the hypothesis that protein folding patterns are conserved even when no detectable amino acid sequence similarity remains [20].

Acknowledgements: We thank Drs. J. Narita and W. Gruissem for their generous gift of spinach cDNA libraries, Dr. Bertil Andersson for the generous gift of anti-22 kDa antiserum, D. Shen for excellent technical assistance, and C. Bladen for help with the figures. This research was supported by grants from the United States Department of Agriculture (E.P.), National Science Foundation (C.F.Y.), Natural Sciences and Engineering Research Council of Canada (B.R.G.) and the Medical Research Council of Canada (R.A.).

REFERENCES

[1] Berthold, D.A., Babcock, G.T. and Yocum, C.F. (1981) FEBS Lett. 134, 231–234.
[2] Bowlby, N.R. (1990) Ph.D. thesis, University of Michigan.
[3] Ljungberg, U., Akerlund, H.E., Larsson, C. and Andersson, B. (1984) Biochim. Biophys. Acta 767, 145–152.
[4] Ghanotakis, D.F., Demetrious, D.M. and Yocum, C.F. (1987) Biochim. Biophys. Acta 891, 15–21.
[5] Camm, E.L. and Gree.., B.R. (1989) Biochim. Biophys. Acta 974, 180–184.
[6] Ljungberg, U., Akerlund, H.E. and Andersson, B. (1986) Eur. J. Biochem. 158, 477–482.
[7] Pichersky, E., Subramaniam, R., White, M.J., Reid, J., Aebersold, R. and Green, B.R. (1991) Mol. Gen. Genet. 227, 277–284.
[8] von Heijne, G., Harai, R., Klosgen, R.B., Steppuhn, J., Bruce, B., Keegstra, K. and Herrmann, R. (1991) Plant Mol. Biol. Rep. 9, 104–126.
[9] Pichersky, E., Bernatzky, R., Tanksley, S.D., Breidenbach, R.B., Kausch, A.P. and Cashmore, A.R. (1985) Gene 40, 247–258.
[10] Pichersky, E., Hoffman, N.E., Malik, V.S., Bernatzky, R., Tanksley, S.D., Szabo, L. and Cashmore, A.R. (1987) Plant Mol. Biol. 9, 109–120.
[11] Sanger, R., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Hallick, R.B. (1989) Plant Mol. Biol. Rep. 7, 266–275.
[13] Ghanotakis, D.F., Waggoner, C.M., Bowlby, N.R., Demtrious, D.M., Babcock, G.T. and Yocum, C.F. (1987) Photosyn. Res. 14, 191–199.
[14] Green, B.R. and Camm, E.L. (1990) in: Current Research in Photosynthesis (Baltscheffsky, M., ed.), pp. 659–662, Kluwer, Dordrecht.
[15] Nilsson, F., Andersson, B. and Jansson, C. (1990) Plant Mol. Biol. 14, 1051–1054.
[16] Green, B.R., Pichersky, E. and Kloppstech, K. (1991) Trends Biochem. Sci. 16, 181–186.
[17] Grimm, B., Kruse, E. and Kloppstech, K. (1989) Plant Mol. Biol. 13, 583–593.
[18] Lers, A., Levy, H. and Zamir, A. (1991) J. Biol. Chem. 266, 13698–13705.
[19] Hoffman, N.E., Pichersky, E., Malik, V.S., Castresana, C., Ko, K., Darr, S.C. and Cashmore, A.R. (1987) Proc. Natl. Acad. Sci. USA 84, 8844–8848.
[20] Branden, C. and Tooze, J. (1991) Introduction to Protein Structure, Garland Publishing Inc., London.
[21] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[22] Schwartz, E., Shen, D., Aebersold, R., McGrath, M.J., Pichersky, E. and Green, B.R. (1991) FEBS Lett. 280, 229–234.
[23] Green, B.R. (1990) in: Current Research in Protein Chemistry (Vilafranca, L. ed.) pp. 395–404, Academic Press.