Volume 232, number 2, 275–280 FEB 05860 May 1988

Cloning and sequencing of spinach cDNA clones encoding the 20 kDa PS I polypeptide

Bernard Lagoutte

Service de Biophysique, Département de Biologie, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

Received 14 March 1988

Apart from the 8 kDa subunit, which is of chloroplast origin, most of the small polypeptides of the PS I reaction center from higher plants are encoded in nuclear genes. We describe here the first nucleotide sequence of a nuclear component of this photosystem, the precursor of the 20 kDa protein. The deduced sequence of the large transit peptide (55–60 amino acids) is rich in serine/threonine residues and has a net positive charge, which are classical features of these precursors. The sequence itself is mainly hydrophilic, with no possibility of classical membrane-spanning α-helices; it exhibits an interesting stretch of five basic amino acids in close vicinity: Thr-**Arg**-Leu-**Arg**-Ser-**Lys**-Tyr-**Lys**-Ile-**Lys**-Tyr.

Photosystem I; 20 kDa subunit; cDNA sequence; Transit peptide

1. INTRODUCTION

A detailed description of both the structure and function of the PS I reaction center at the molecular level requires precise knowledge of the primary structure of the polypeptides involved. Currently, the most popular approach in this field is to select and sequence the genes coding for these polypeptides. Elucidation of the gene sequence of the two large subunits [1] and more recently of the small 8 kDa protein [2,3] has proved very useful and almost definitively ended the long-standing discussion concerning the localization of the three Fe-S clusters associated with PS I [4–7]. It now appears quite clear that F_X is linked to the large subunits [8,9], at a site involving a conserved, short amino acid sequence including two cysteines. The association of F_A and F_B with the small 8 kDa polypeptide is also virtually an established fact, in agreement with our first proposal based on the amino acid composition [5]. Different sequences for this component confirmed the predicted similarity with small 2 [4Fe-4S] proteins found in anaerobic organisms [2,3].

The roles of the other small polypeptides required for correct transfer of electrons from plastocyanin to ferredoxin remain to be clarified. The largest polypeptide (20 kDa) was at first claimed to be involved in electron donation from plastocyanin to PS I [10]. Other results have indicated a possible stabilization effect by a similar component on the EPR signals of centers F_A and F_B [7]. More recently, direct association of this subunit with ferredoxin has been reported [11]. Determination of the primary sequence of this polypeptide is thus of great interest to progress in this discussion. Here, we report the isolation and sequencing of cDNA clones which encode the full length of the 20 kDa precursor protein.

Correspondence address: B. Lagoutte, Service de Biophysique, Département de Biologie, CEN Saclay, 91191 Gif-sur-Yvette, France

Abbreviations: Cab protein, chlorophyll *a/b*-binding protein; mAb, monoclonal antibody; OEE2, oxygen-evolving enhancer protein 2; PC, plastocyanin; SSu, small subunit of ribulose-1,5-bisphosphate carboxylase

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no.Y00759

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

2. MATERIALS AND METHODS

Total RNA was isolated from young spinach leaves after grinding in liquid nitrogen [12]. Poly(A)$^+$ mRNAs were purified on poly(U)-Sepharose (Pharmacia) following the instructions of the manufacturer.

In vitro synthesis experiments were performed using a reticulocyte lysate system (BRL) with [^{35}S]methionine as labelling reagent (1100 Ci/mmol, New England Nuclear). The usual incubation period was 1 h at 30°C at an mRNA concentration of 0.1 μg/10 μl and the radioactivity added amounted to 30 μCi.

Proteins were analyzed by SDS-polyacrylamide gel electrophoresis according to Laemmli [13]; the resolving gel contained 15% acrylamide. Gels were treated with a solution of 20% 2,5-diphenyloxazole in DMSO for fluorography.

Using poly(A)$^+$ mRNA, a λgt11 cDNA library was constructed following the simplified protocol of Watson and Jackson [14]. AMV reverse transcriptase was from Life Sciences, the λ arms and packaging kit being from Promega Biotec.

Immunological screening of the library [15] was performed using nitrocellulose filters treated as described for Western blotting [16]. Antigen/antibody complexes were detected enzymatically with an alkaline/phosphatase conjugate from Promega Biotec.

Preparation of phages and DNA extraction were according to Maniatis et al. [17]. DNAs to be sequenced were cut at the appropriate sites and subcloned in M13 vector [18], either directly or after purification of the inserts on an agarose gel.

DNA sequencing was carried out by the dideoxy chain-termination method of Sanger et al. [19], with a New England Biolabs kit and ^{35}S-labelled dATP from Amersham (600 Ci/mmol).

Nucleotide and protein sequences were analyzed using CITI2 programs.

3. RESULTS AND DISCUSSION

A complete set of mAbs directed against the small polypeptides of a spinach PS I particle [5] has been recently characterized [16]. Antibodies specific for the 20 and 17 kDa polypeptides were the most abundant by far. We first tested different anti-20 kDa mAbs for the ability to recognize in vitro translation products directed by poly(A)$^+$ mRNA. Earlier studies with protein synthesis inhibitors indicated that all the polypeptides in the 20–25 kDa range were likely to be translated on cytoplasmic ribosomes [20]. This was confirmed by in vitro translation experiments [21,22]. However, it remains a difficult task to correlate unambiguously the various PS I polypeptides and their molecular masses as obtained from different organisms and as reported by various authors. We thus checked again the nuclear origin of the 20 kDa protein in our PS I particle.

Five anti-20 kDa mAbs were mixed in order to increase the potential number of recognition sites. During the course of this experiment, a classical immunoprecipitation protocol failed to work [23], due to the very low efficiency of binding of protein A to mouse γ_1 immunoglobulins. To overcome this problem, we used the solid-phase procedure previously designed for screening and titration of these mAbs [15]. Briefly, the specific mAbs were immobilized in the wells of a microtitration plate via anti-mouse immunoglobulins. After 10 min centrifugation at 100000 × g in a Beckman airfuge, the translation supernatant was diluted as in [23], placed in the wells of the plate and allowed to react for 1 h at 37°C. After extensive washing in the presence of detergents (1% Triton X-100/0.1% SDS), the Ag/Ab complex was dissociated by direct use of electrophoretic sample buffer and analyzed by SDS electrophoresis. The result is shown in fig.1C; apart from non-specific binding at 18 kDa (lane D), a very clear band is detected at 26 kDa, which corroborates previous results obtained with polyclonal Abs [21,22].

The same mixture of anti-20 kDa mAbs was then used to select positive clones from a λgt11 library. The starting library consisted of 1.7×10^6 inserts and was amplified 10^4 times before use. In the course of several screenings, 8 positive clones were retained after three successive rounds of plating. Three were used more extensively for sequencing experiments: clones 11, 20a and 20b. As described in fig.2A, clones 20a and 20b still had both *Eco*RI insertion sites, but clone 11 lost the left site, as determined by nuclease fragmentation with *Kpn*I, *Sac*I and *Eco*RI. The 20a and 20b *Eco*RI inserts were subcloned in both orientations in the vector M13 mp19. We also used forced cloning in M13 mp19 for the *Kpn-Eco* insert of clone 11. The sequencing strategy is described in fig.2B, and was entirely performed using the dideoxy chain-termination method of Sanger et al. [19], together with M13 deletion subcloning [24]. Clone 20b has been sequenced from the 3′-poly(A)$^+$ tail to the 5′-leader sequence. Both ends were confirmed, starting the reading of clone 20a in opposite directions. Finally, the full sequence was also read in the 5′- to 3′-direction using clone 11.

Fig.3 shows the different cDNA sequences thus

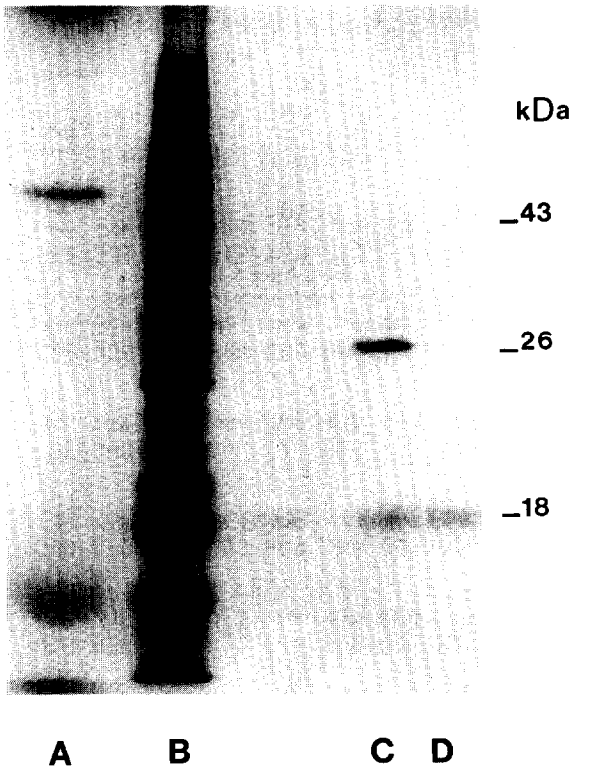

Fig.1. Fluorography of in vitro translation experiments directed by spinach poly(A)$^+$ mRNA. (A) Translation without added mRNA, (B) total translation products, (C) total translation after specific adsorption by anti-20 kDa mAb, (D) total translation after adsorption by non-specific mAb.

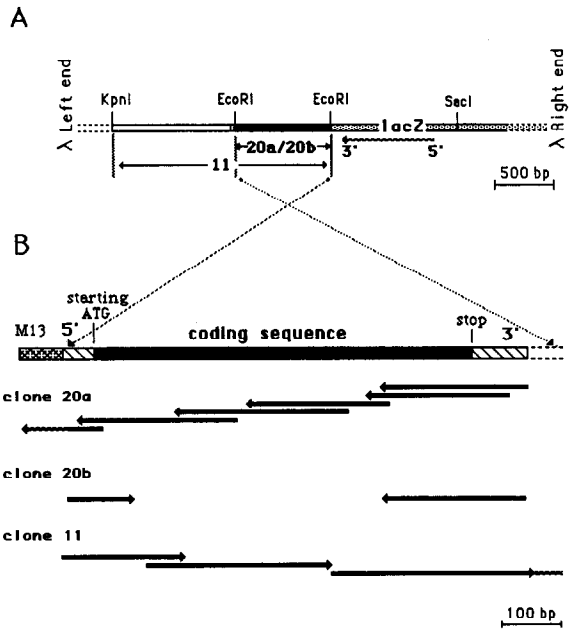

Fig.2. (A) Restriction map of λgt11 inserts of clones 11, 20a and 20b. (B) M13 sequencing strategy and the different sets of overlapping sub-clones analyzed.

obtained. There are some differences between the three clones, which do not affect the coding region. The number of A in the 3'-tail varies, and there are 12 more nucleotides at the 5'-end of clone 11. Surprisingly, a poly(T) tail is present at the 5'-end of clones 20a and 20b. Such a poly(T) tail has been described previously in the case of a cDNA coding ferredoxin [25], but at a distance of 500 bp from the starting ATG; this was attributed to cDNA exchange in the library. Reorganization might also have occurred in clones 20a and 20b as no poly(T) is observed in clone 11, which extends a little further. The minor differences observed at this end can be explained by such possible reorganization. As observed in clones 20a and 20b, the 3'-poly(A) tail is rather short and relatively close to the stop codon compared to other cDNAs coding for chloroplast proteins. Located 19 nucleotides upstream of the poly(A) is the characteristic sequence AATACAA already described for a pea Cab gene [26] and similar to the AATGAA reported for PC and SSu [27]. These sequences have been postulated to be equivalent [26] to the classical polyadenylation signal AATAAA commonly found in eukaryotic mRNA [28]. The 3'-end of clone 6 will not be described here; it has the same common sequence with about 90 extra nucleotides preceding a larger poly(A) tail.

The large open reading frame of 636 nucleotides commences with 2 ATG codons in phase, at the first and third positions, resulting in a double dipeptide Met-Ala at the beginning of the protein. This coding sequence is also terminated by a double TAA stop codon. Just after the stop codons is a T-rich sequence (12 out of 16); this sequence could form a large loop by associating with the poly(A) tail, thus enhancing the effect of the stop signals.

The reading frame codes for a precursor protein of 23161 Da, about 3 kDa less than the predicted apparent value deduced from in vitro translation studies (fig.1). This discrepancy can be due to the large excess of basic amino acids, giving an approximate residual positive charge of 12 at pH 7.

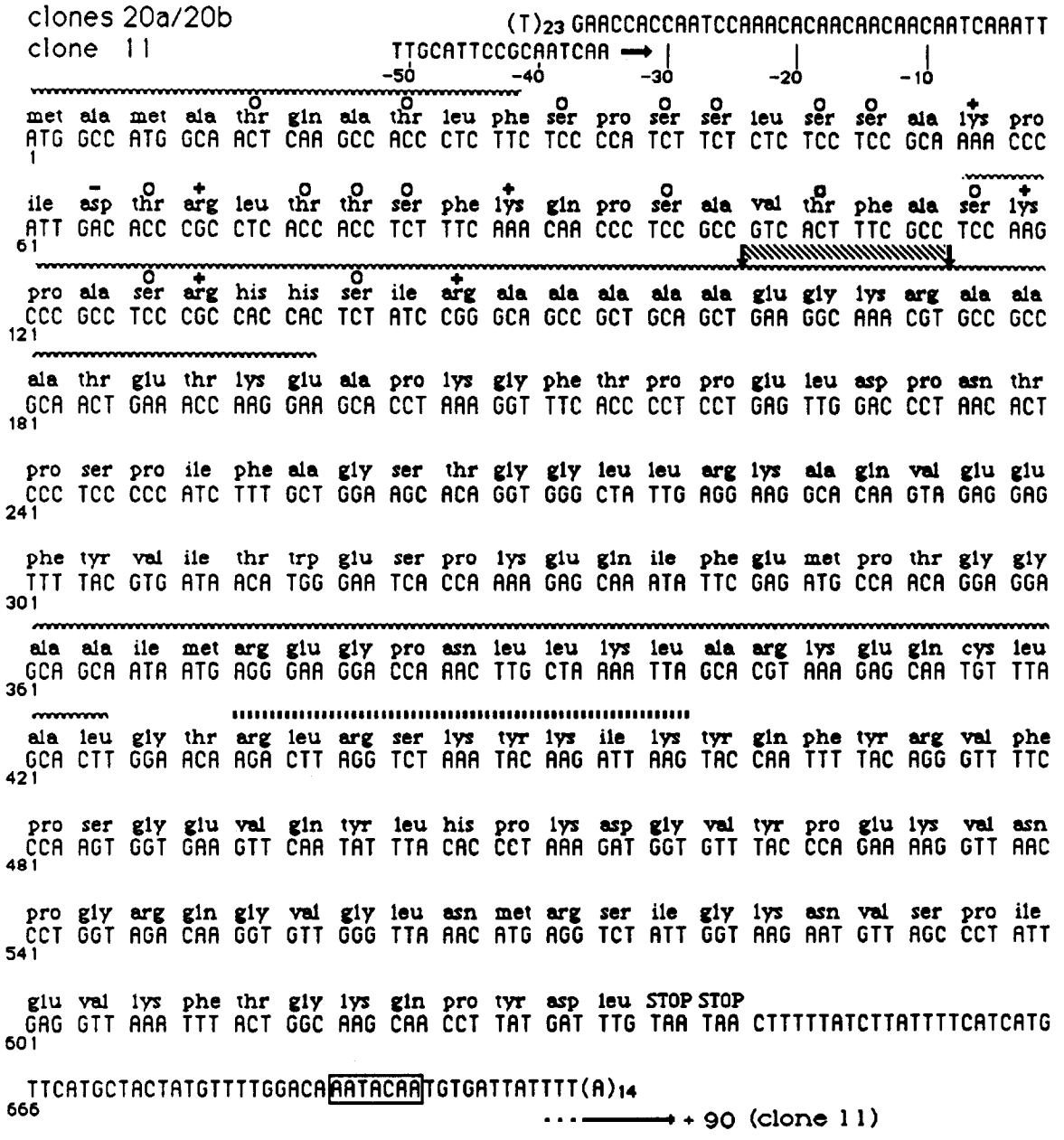

Fig.3. DNA sequences of cDNA clones (11,20a,20b) encoding the 20 kDa subunit of PS I. The nucleotide sequence is written in the mRNA sense. Deduced amino acid residues are designated the three-letter code. Above the protein sequence different interesting domains are delineated: predicted helices are indicated by wavy lines; the putative maturation site is within the hatched zone, between the two arrows; the highly basic sequence is denoted by the dotted line. Hydroxy amino acids of the transit peptide are shown by an open circle.

Such a positive charge can partly balance the global negative charge of SDS resulting in slower migration than expected. Tentative microsequencing of the protein extracted from preparative gels failed to give any clear result concerning the N-terminal end of the mature protein. Nevertheless, a precise determination of the amino acid composition provided interesting information concerning

the site of maturation. Of the three His predicted in the total sequence, only one is present in the mature protein, excluding the dipeptide His-His at position 46. Moreover, the number of Gly and Ala residues is about 15–16 and 10–11, respectively, arguing for a maturation site located between the two Ala sequences (at positions 49/54 and 59/61). The presence of basic amino acids within this short region is also in favor of this hypothesis. A very recent work by Gray and co-workers [29] strongly supports this prediction. The N-terminal sequence of the equivalent mature protein from pea is almost entirely the same all along 25 residues; the only differences occur near the maturation site, as has been observed for other thylakoid proteins [30].

47 first residues (35%) together with a large excess of positive charges, which seem now to be classical features of the chloroplast transit peptides [30].

The hydrophilicity profile [34] for the precursor protein is presented in fig.4. Hydrophilic domains are predominant and only short segments of the mature protein are susceptible to interactions with a hydrophobic environment.

Secondary structure predictions based on two methods [35,36] gave one short α-helix between Met 1 and Phe 9, and two others of a length compatible with a transmembrane location. The first extends in the domain 39/66, including the maturation site, and the second in the domain 120/142. A large number of charged amino acids are present in these two helices (8 and 6, respectively), virtually excluding membrane-spanning models. Both putative helices of the transit peptide could in any case play a role during interactions with the chloroplast envelope and thylakoid membrane. The absence of a classical transmembrane sequence argues for an extrinsic localization of this 20 kDa subunit. It remains to be decided whether this localization is outside the thylakoid, as supported by labelling, proteolysis and cross-linking experiments [10,37], or inside, as deduced from functional studies [9].

```
spinach AAEGKRA AATE--TKEAPVGFTPPELDPNTPSPIFAG
        |  |    ||||  |||||||||||||||||||||||||
pea     XTEDKTD AATDVATXEAPVGFTPXELDPNTXSXIF-G
```

The length of the transit peptide (55–59 residues) would then be close to that of proteins present in the thylakoid lumen, such as PC [31] or OEE2 [32]. Shorter lengths (35–40) have been observed for typical intrinsic proteins, such as the Cab protein of the LHCII [26] or LHCI polypeptides [33]. However, the length of the transit peptide is extremely variable for proteins of the stroma [30], and might be of no significance. A very high density of serine and threonine is found throughout the

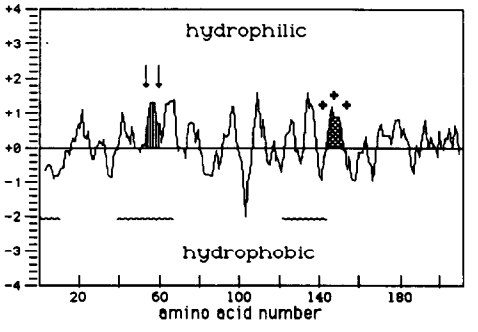

Fig.4. Hydrophilicity plot of the 20 kDa precursor protein according to Hopp and Woods [34]; averaging window was on 6 amino acids. The predicted helices detailed in fig.3 are delineated by a wavy line, the putative maturation site by the two arrows and the basic sequence by positive charges.

The most striking feature of this sequence is probably the long stretch of basic residues located just after the third potential α-helix: Thr-**Arg**-Leu-**Arg**-Ser-**Lys**-Tyr-**Lys**-Ile-**Lys**-Tyr. This regular alternation could result in a very high density of positive charges on the same side of the polypeptide backbone. Interactions with acidic domains of other polypeptides of the system are thus to be expected; this might be the case with the N-terminal fragments of PS I A1 or A2 proteins, or the C-terminal end of ferredoxin. It has been shown recently that the alternating copolymer poly(Leu-Lys) is able to undergo a β-sheet or α-helix transition in the presence of lysolecithin or lysolecithin plus cholate [38]. Upon interaction with polar lipids of the thylakoid, this particular sequence of the 20 kDa polypeptide could also induce some

modifications in the regular arrangement of the lipid bilayer.

When tested for cross-reactivity, most of our anti-20 kDa mAbs recognized various photosynthetic membranes from higher plants to cyanobacteria (unpublished), a result which favors a high degree of conservation. The extensive homology observed with the N-terminal sequence of the equivalent protein from pea [29] corroborates this assumption, and irrespective of its actual function, it is likely that the 20 kDa component plays a major role in PS I.

Acknowledgements: I particularly would like to thank Professor L. Bogorad for providing all essential facilities to start this work and Dr P. Mathis, through whom this project became a reality. I also acknowledge the collaboration with Dr P. Pradelles and his immunology group whose efforts helped to initiate this study. I express my sincere gratitude to Dr S. Merchant for helpful advice in constructing the cDNA library. I also wish to thank my colleagues in the Biology Department for fruitful discussions and Mike and Stan for technical assistance in reading DNA sequences.

REFERENCES

[1] Fish, L.E., Kück, U. and Bogorad, L. (1985) J. Biol. Chem. 260, 1413–1421.
[2] Oh-oka, H., Takahashi, Y., Wada, K., Matsubara, H., Ohyama, K. and Ozeki, H. (1987) FEBS Lett. 218, 52–54.
[3] Hayashida, N., Matsubayashi, T., Shinozaki, K., Sugiura, M., Inoue, K. and Hiyama, T. (1987) Curr. Genet. 12, 247–250.
[4] Nechustai, R. and Nelson, N. (1981) J. Biol. Chem. 256, 11624–11628.
[5] Lagoutte, B., Setif, P. and Duranton, J. (1984) FEBS Lett. 174, 24–29.
[6] Sakurai, H. and San Pietro, A. (1985) J. Biochem. 98, 69–76.
[7] Bonnerjea, J., Ortiz, W. and Malkin, R. (1985) Arch. Biochem. Biophys. 240, 15–20.
[8] Bordier Høj, P. and Lindberg Møller, B. (1986) J. Biol. Chem. 261, 14292–14300.
[9] Golbeck, J.H., Panett, K.G. and McDermott, A.E. (1987) Biochim. Biophys. Acta 893, 149–160.
[10] Bengis, C. and Nelson, N. (1977) J. Biol. Chem. 252, 4564–4569.
[11] Zanetti, G. and Merati, G. (1987) Eur. J. Biochem. 169, 143–146.
[12] Bogorad, L., Gubbins, E.J., Krebbers, E., Larrinua, J.M., Mulligan, J., Muskavitch, K.M.T., On, E.A., Rodermel, S.R., Schantz, R., Steinmetz, A.A., De Vos, G. and Ye, Y.K. (1983) Methods Enzymol. 97, 524–554.
[13] Laemmli, U.K. (1970) Nature 227, 680–685.
[14] Watson, C.J. and Jackson, J.F. (1985) in: DNA Cloning, a Practical Approach (Glover, D.M. ed.) pp.79–88, IRL, Oxford.
[15] Young, R.A., Bloom, B.R., Grosskinsky, C.M., Ivanyi, J., Thomas, D. and Davis, R.W. (1985) Proc. Natl. Acad. Sci. USA 82, 2583–2587.
[16] Grassi, J., Frobert, Y., Lamourette, P. and Lagoutte, B. (1988) Anal. Biochem. 168, 436–450.
[17] Maniatis, T., Fritsch, E. and Sambrook, J. (1982) in: Molecular Cloning, Cold Spring Harbor Laboratory, NY.
[18] Messing, J. and Vieira, J. (1982) Gene 19, 269–276.
[19] Sanger, F., Niklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[20] Gray, J.C. (1987) in: Photosynthesis (Amesz, J. ed.) pp.319–342, Elsevier, Amsterdam, New York.
[21] Westhoff, P., Alt, J., Nelson, N., Bottomley, W., Bünemann, H. and Hermann, R.G. (1983) Plant Mol. Biol. 2, 95–107.
[22] Tittgen, J., Hermans, J., Steppuhn, J., Jansen, T., Jansson, C., Andersson, B., Nechuschtai, R., Nelson, N. and Herrmann, R.G. (1986) Mol. Gen. Genet. 204, 258–265.
[23] Westhoff, P. and Zetsche, K. (1981) Eur. J. Biochem. 116, 261–267.
[24] Dale, R.M.K., McClure, B.A. and Houchins, J.P. (1985) Plasmid 13, 31–40.
[25] Smeekens, S., Binsbergen, J.V. and Weisbeek, P. (1985) Nucleic Acids Res. 13, 3179–3194.
[26] Cashmore, A.R. (1984) Proc. Natl. Acad. Sci. USA 81, 2960–2964.
[27] Coruzzi, G., Broglie, R., Cashmore, A. and Chua, N.H. (1983) J. Biol. Chem. 258, 1399–1402.
[28] Nevins, J.R. (1983) Annu. Rev. Biochem. 52, 441–466.
[29] Dunn, P.P.J., Packman, L.C., Pappin, D. and Gray, J.C. (1988) FEBS Lett. 228, 157–161.
[30] Schmidt, G.W. and Mishkind, M.C. (1986) Annu. Rev. Biochem. 55, 879–912.
[31] Smeekens, S., De Groot, M., Binsbergen, J.V. and Weisbeek, P. (1985) Nature 317, 456–458.
[32] Mayfield, S.P., Rahire, M., Frank, G., Zuber, H. and Rochaix, J.D. (1987) Proc. Natl. Acad. Sci. USA 84, 749–753.
[33] Stayton, M.M., Brosio, P. and Dunsmuir, P. (1987) Plant Mol. Biol. 10, 127–137.
[34] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3838.
[35] Chou, P.Y. and Fasman, G.D. (1980) Annu. Rev. Biochem. 47, 251–276.
[36] Garnier, J., Osguthorpe, D.J. and Robson (1978) J. Mol. Biol. 120, 97–120.
[37] Ortiz, W., Lam, E., Chollar, S., Munt, D. and Malkin, R. (1985) Plant Physiol. 77, 389–397.
[38] Bazzi, M.D., Woody, R.W. and Brack, A. (1987) Biopolymers 26, 1115–1124.