# Plastocyanin is encoded by an uninterrupted nuclear gene in spinach\*

## C. Rother, T. Jansen, A. Tyagi, J. Tittgen, and R. G. Herrmann

Botanisches Institut, Ludwig-Maximilians-Universität, Menzinger Strasse 67, D-8000 München 19, Federal Republic of Germany

Summary. Plastocyanin is a member of photosynthetic electron transport chains that transfers electrons from cytochrome f to the oxidized  $P_{700}$  chlorophyll a pigment of the photosystem I reaction center. We have isolated and characterized cDNA- and genomic clones from spinach (Spinacia oleracea) encoding the complete plastocyanin-precursor polypeptide. The amino acid sequence derived from the nucleotide sequence shows that the precursor consists of 168 amino acid residues including a transit sequence of 69 residues. The precursor polypeptide has a predicted  $M_r$  of 16,917, the mature protein of 10,413. The available data indicate that plastocyanin derives probably from a single-copy gene. The coding region contains no intron. The size of the mRNA as determined by S1 nuclease protection experiments is approximately 660 nucleotides, although analysis of different cDNA clones suggests that longer RNA species do exist, approaching the size of the mRNA (850 bases) estimated by Northern blot techniques.

Key words: Plastocyanin – Photosynthesis – Genomic and cDNA clones – Sequence analysis – Transcript – Spinach

# Introduction

Plastocyanin (PC) is a member of photosynthetic electron transport chains. In higher plants, it is a soluble

Offprint requests to: R. G. Herrmann

type I copper protein with an approximate molecular mass of 11 kd that serves in the transfer of electrons from cytochrome f to the P<sub>700</sub> chlorophyll a molecule in the photosystem I reaction center (reviewed in Haehnel 1986). The protein is located in the thylakoid lumen, originates in the genome and, like other chloroplast proteins encoded by genes in nuclear DNA, is decoded in the cytosol as precursor with an N-terminal transit peptide that is processed during its post-translational import into the organelle (Grossman et al. 1982; Herrmann et al. 1983; Smeekens et al. 1985, 1986; Tittgen et al. 1986). The complete amino acid sequence of the mature protein has been determined for several plants, including spinach, and a topological model has been derived from X-ray crystallography at a resolution of 1.6 Å (reviewed in Haehnel 1986). Recently, cDNA clones for spinach and Silene plastocyanin have been isolated (Herrmann et al. 1983; Smeekens et al. 1985) including a clone encoding the complete precursor protein from the latter organism (Smeekens et al. 1985).

We are interested in understanding structure/function relationships in photosynthetic energy transduction as well as mechanisms operating during biogenesis of thylakoid membranes and their physiological adaptation. The synthesis of this specialized biomembrane depends upon the adequacy of the expression of nuclear and plastid genes that is controlled by internal and external factors, especially light (reviewed in Herrmann et al. 1985). We have recently isolated recombinant cDNA plasmids or phage for 20 components of the nuclear encoded complement of thylakoid membrane proteins, including those for plastocyanin (Herrmann et al. 1983; Tittgen et al. 1986, and unpublished results). In this paper we describe the analysis of cDNA and genomic clones that contain the complete coding region for plastocyanin from spinach. Our data indicate the presence of only a single, uninterrupted gene for the protein in this organism.

<sup>\*</sup> Dedicated to Prof. Dr. Fritz Kaudewitz on the occasion of his 65th birthday

## Materials and methods

Restriction enzymes, Klenow fragment of DNA polymerase I, SP6 RNA polymerase, calf intestinal alkaline phosphatase, T4 DNA ligase, T4 polynucleotide kinase and Proteinase K were purchased from Boehringer (Mannheim);  $\alpha^{-32}$ P-dNTPs and  $\gamma^{-32}$ P-ATP from Amersham (Braunschweig). All chemicals used were of analytical reagent grade.

Library construction and DNA subcloning. Genomic DNA from purified spinach nuclei was isolated as described in Tittgen et al. (1986). Briefly, the final pellet of nuclei from 100 g young leaves prepared according to Willmitzer and Wagner (1981) with the modifications detailed in Tittgen (1985), was suspended in 10 ml of a lysis buffer consisting of 50 mM Tris/HCl, 0.5% SDS, 100 mM EDTA, 50 mM NaCl, pH 8.3, and incubated with gentle shaking at 37 °C for a minimum of 2 h. This step was followed by phenol/chloroform (1:1, v/v) extraction. The DNA was ethanol-precipitated and dissolved in 10 mM Tris/HCl, 1 mM EDTA, pH 7.5. Solid CsCl (0.83 g/ml) was dissolved in the DNA solution and, after addition of 100  $\mu$ l of ethidium bromide (5 mg/ ml), the mixture was centrifuged at 45,000 rpm in a Beckman VTi 50 rotor for 24 h at 18 °C. The DNA was collected using a longwave UV light source to visualize the band. Ethidium bromide was removed by extraction with isoamyl alcohol, and CsCl by exhaustive dialysis against 10 mM Tris/HCl, 1 mM EDTA, pH 8.0. The DNA was ethanol-precipitated, dissolved and stored at 4 °C.

Approximately 100  $\mu$ g of the nuclear DNA was partially digested for 1 h with EcoRI (Sau3A or BamHI) under the conditions specified by the supplier. The DNA was then ethanol precipitated and sized on a 10-40% sucrose gradient containing 20 mM Tris/HCl, 1 M NaCl, 5 mM EDTA, pH 8.0 (Beckman SW 40 rotor, 38,000 rpm for 20 h at 10 °C). The size class of DNA molecules in each fraction was determined by agarose gel electrophoresis. Fractions containing fragments between 15 and 23 kbp were concentrated by ethanol precipitation and ligated into the EcoRI (BamHI) site of the lambda phage EMBL4 (Frischauf et al. 1983). The vector was previously prepared by double-digestion with the corresponding enzyme and Sall, and fractionated by differential precipitation to remove the stuffer fragment and recover the purified phage arms (Tittgen 1985). Ligated DNA was packaged in vitro into lambda particles in 0.1 µg aliquots (Hohn and Murray 1977). After transfection, the phage libraries were amplified (Maniatis et al. 1982) and recombinants were selected by in situ plaque hybridization (Benton and Davis 1977) using SP6 RNA polymerase derived RNA (Melton et al. 1984) from PC-cDNA clones. Positive phase were purified by replating.

The construction of the pBR322 and lambda gt11 cDNA libraries of spinach seedling poly  $A^+$ -RNA, selection and characterization of cDNA clones for 14 intrinsic and extrinsic thylakoid membrane proteins, including plastocyanin as well as subclonings in pUC18, pUC118, Bluescribe M13 and pSP64 expression vectors have been described (Tittgen et al. 1986).

SI protection analysis. Approximately  $1-2 \mu g$  of 3'- or 5'-endlabelled restriction fragment was hybridized with either 25  $\mu g$  of *E. coli* tRNA or 5  $\mu g$  poly A<sup>+</sup>-RNA from spinach seedlings, and S1 protection mapping (Berk and Sharp 1977) was performed essentially as described by Weaver and Weissmann (1979) including a temperature gradient of 48 to 42 °C for hybridization.

Nucleotide sequence analysis. Sequence analysis of the cDNA and genomic clones was carried out by the chemical modification and controlled chain cleavage procedure (Maxam and Gilbert 1980). Sequence data were analysed on a IBM-AT computer with programs compiled by Kröger and Kröger-Block (1982).

*Miscellaneous.* Small- and large scale phage and DNA purification, preparation of poly  $A^+$ -RNA, nick-translation of DNA, transcription mediated by SP6 RNA polymerase, southern transfer analysis, fillin synthesis of 3' recessed ends with Klenow fragment of DNA polymerase I, 5' end-labelling with T4 polynucleotide kinase, agarose and polyacrylamide gel electrophoresis were performed as previously reported (Morris and Herrmann 1984; Westhoff 1985; Tittgen et al. 1986).

## **Results and discussion**

#### cDNA clone selection and identification

The initial approach used to clone DNA sequences complementary to plastocyanin mRNA was to screen a library of cDNA recombinant plasmids (pBR322) by hybrid select translation, immunology and electrophoresis of the resulting product. Three positive clones were isolated from about 1,500 transformants (Herrmann et al. 1983) and used to screen a lambda gt11 expression library (Tittgen et al. 1986). Twenty clones were selected from approximately 2 x 10<sup>5</sup> transformants. Translation of hybrid-selected mRNA (Bünemann et al. 1982) yielded a single, prominent 21 kd band on a sodium dodecyl sulfate polyacrylamide gel comigrating with the plastocyanin precursor obtained by immunoprecipitation from poly A+-RNA translation products. Furthermore, the antiserum precipitated an 11 kd processed and trypsin-resistant translation product after import by isolated unbroken chloroplasts which, in turn, comigrated with authentic spinach plastocyanin (cf. Tittgen et al. 1986; data not shown). The cDNA inserts in these clones cross-hybridide, confirming the identification of the recombinant plasmids and phage.

For sequence analysis, the cDNA clones were rescreened for insert sizes. The largest inserts characterized, 0.7 kb, approached the average size of the single hybridizing RNA band in Northern blots (0.85 kb; Tittgen et al. 1986). The inserts from the two largest spinach PC-cDNA clones isolated designated l23SocPC-1 and -2 (Tittgen et al. 1986) were sequenced and one of them, l23SocPC-1, was found to begin within the proposed initiation codon and to contain the entire amino acid coding region including the transit peptide (Fig. 1). The inserts of these hybrid phage were subcloned in pSP-64 and pUC118 and used to probe a genomic library of spinach nuclear DNA.

#### Isolation of the spinach PC gene

Screening plaque blots of the genomic EcoRI library allowed the selection of one positive recombinant phage,



Fig. 1A, B. Restriction endonuclease cleavage map and sequencing strategy of genomic (A) and cDNA (B) clones for spinach plastocyanin. The maps are drawn 5' to 3' and are aligned relative to each other on the basis of restriction sites. A The position and transcription polarity of the plastocyanin gene within the genomic clone is indicated in the map above. The expanded part marks the transcription initiation and termination signals; *horizontal arrows* denote the direction and extent of individual sequence reactions. The positions of the amino and carboxyl termini as well as the terminal processing site are given by *arrowheads* and *arrow*, respectively. B The different 3' ends of two otherwise identical cDNA clones are distinguished by the *horizontal arrow* (3') in the expanded scheme and by the cDNA clones depicted in Panel B (*open boxed* and *rightmost arrow*, respectively). Scales and fragment sizes are given in bp

123SonPC-1, which had an insert of approximately 20 kbp. This fragment was mapped with BamHI, HindIII, EcoRI, SalI, PstI and AvaI, and the EcoRI as well as SalI fragments were subcloned into pUC18. A physical map of the DNA fragment is shown in Fig. 1. Hybridization with the radiolabelled PC-cDNA insert gave signals with a single 8.2 kbp EcoRI fragment and two adjacent SalI fragments of 4.4 and 2.1 kbp (Fig. 1) suggesting that only one copy of the plastocyanin gene is present on the clone.

## Sequence analysis

Based on the DNA hybridization data, nucleotide sequencing was performed on part of the genomic 4.4 and 2.1 kbp SalI fragments and the 8.2 kbp EcoRI fragment. The regions sequenced as well as the strategy used, are included in Fig. 1. The complete nucleotide sequence and derived amino acid sequence for the plastocyanin precursor are depicted in Fig. 2.

The sequence of the *cDNA clone* 123SocPC-1 shows that the insert contained (i) a 30 nucleotide long 3' terminal poly A-tract, (ii) 164 bp 3' untranslated region in-

cluding a putative polyadenylation signal (CATAAA, position 642-647) and (iii) 502 bp coding region for the plastocyanin precursor. Sequence analysis of several independent plastocyanin cDNA clones indicates that their 3' untranslated regions are not diverged, substantiating our previous assumption (Tittgen et al. 1986) that the corresponding gene is located in single-copy DNA. The finding of 3' untranslated regions, polymorphic with respect to length (Figs. 1 and 2), does not contradict this assumption (see below and Dean et al. 1986; Hernández-Lucas et al. 1986; Tyagi et al., submitted).

Analysis of the *genomic DNA* and the determined amino acid sequences reinforce the above assumption. Only one open reading frame was detected in the genomic clone. The sequence in the coding and noncoding regions is identical with that deduced from the cDNA sequence. The gene is uninterrupted, 504 bp long, encoding a polypeptide of 168 amino acid residues (including a 69 residue presequence) with a predicted molecular mass of 16.9 kd (precursor) and 10.4 kd (mature protein). The amino acid sequence deduced from the nucleotide sequence matches 100% with the known amino acid sequence of the processed mature protein (Scawen et al. 1975) and indicates that the signal peptide has a molecular weight of 6.5 kd.

The polypeptide chain contains a relatively large number of charged amino acid residues (17% of the total). It includes the invariant copper-binding site His(37) and Cys(84)-X-X-His(87)-X-X-X-Met(92) and the acidic patch Asp(18), Asp-Glu-Asp-Glu(42-45) and Glu-Glu-Asp(59-61) that are located in spans 2, 4 and 5 in close spatial proximity. This patch might interact with basic regions of cytochrome f and/or the P<sub>700</sub> chlorophyll a apoprotein (reviewed in Haehnel 1986).

It was not possible to identify the N-terminal residue of the transit sequence directly, because the open reading frame continuous 5' for another 112 triplets, including two in-frame ATG codons (positions -162 to -160and -183 to -181). In addition, the experimentally determined molecular weight for plastocyanin precursors differs considerably (8–15 kd; Grossman et al. 1982; Herrmann et al. 1983; Robinson and Ellis 1984). This difference is difficult to explain, although various thylakoid proteins are known to show abnormal behaviour in polyacrylamide gels depending on sample preparation and on the gel system used. Several lines of evidence suggest, however, that the assumed translational initiation codon (Fig. 2) is correct. First, S1 protection analysis (see below) places the mRNA 5' end at about nucleotide -65 implying that the upstream Met residues lie outside the transcribed part of the gene. Second, only this latter residue is flanked by a highly favourable translation initiation element, ACA ATG G, according to the translation initiation scanning model of 80 S ribosomes (Kozak 1986). Third, the amino acid sequence

-240 GAGCACAA<mark>CCAAT</mark>GAAAAACTAGATAAT

-210 ACCCTTTATTGGCCCACACATCACAAA								-180 FCCTTTAAT <mark>CCAAT</mark> GCCACTAAAAATCCCA							–150 АЛАТ <mark>СААААССАССАСАЛААССАТСТ</mark> АААССАСАС <i>А</i>								-120 TCACCTCATCTATCCTCCA					
-90 -30																												
TTTTATCACCTCTTATTAAATCCCATCCTCGCTCTCAGCACTCTTTGGCAATTGTCATTTCTTAATTGCATTCCACTTAAACCCCCTCCAAAATCCTCCTCCATT															raca													
1	1 30													60														
ATG	GCC	ACC	GTC	GCT	тсс	тсс	GCT	GCC	GTA	GCC	GTC	ccc	тсс	TTT	ACC	GGC	CTT	AAG	GCC	TCA	GGA	тсс	ATC	AAG	сст	ACC	ACC	
MET	Ala	Thr	Val	Ala	Ser	Ser	Ala	Ala	Val	Ala	Val	Pro	Ser	Phe	Thr	Gly	Leu	Lys	Ala	Ser	Gly	Ser	Ile	Lys	Pro	Thr	Thr	
	9	0								120											150							
GCA	AAA	ATC	ATC	CCA	ACC	ACC	ACC	GCC	GTC	CCA	AGA	TTG	тст	GTC	AAG	GCT	TCC	TTG	AAG	AAT	GTC	GGA	GCC	GCC	GTG	GTA	GCA	
Ala	Lys	Ile	Ile	Pro	Thr	Thr	Thr	Ala	Val	Pro	Arg	Leu	Ser	Val	Lys	Ala	Ser	Leu	<b>Lys</b>	Asn	Val	Gly	Ala	Ala	Val	Val	Ala	
180 👿 2												10									24	10						
ACC	GCA	GCC	GCC	GGA	CTT	СТА	GCC	GGA	AAC	GCC	ATG	GCC,	GTT	GAG	GTG	TTG	CTC	GGA	GGG	GGT	GAC	GGA	TCA	TTG	GCA	TTC	CTT	
Thr	Ala	Ala	Ala	Gly	Leu	Leu	Ala	GIY	Asn	Ala	Met	Ala	Val	Glu	Val	Leu	Leu	GIY	Gly	GIY	Asp	Gly	Ser	Leu	Ala	Phe	Leu	
					2	70						3				00										330		
CCA	GGA	GAC	TTC	AGC	GTA	GCC	TCG	GGC	GAG	GAG	ATC	GTA	TTC	AAG	AAC	AAT	GCC	GGA	TTC	000	CAC	AAC	GTA	GTG	TTT	GAC	GAG	
Pro	GIY	Asp	ьvé	ser	vai	Ala	Ser	GTÄ	GIU	eru	ile	vai	Phe	Lys	ASN	ASN	ATS	GIY	Pne	Pro	His	ASN	vai	vai	рле	Asp	Giu	
	360														390											420		
GAC	GAG	ATT	CCC	TCC	GGT	GTC	GAC	GCC	GCG	AAG	ATT	TCG	ATG	TCC	GAG	GAG	GAT	TTG	CTG	AAT	GCA	CCA	GGG	GAA	ACT	TAC	AAA	
Asp	GIU	ITe	Pro	Ser	GIÀ	vai	Asp	AIA	AIa	ràz	TTe	Ser	met	Ser	GIU	GIU	Asp	rea	Leu	ASN	AIA	Pro	GIY	GIU	Thr	Tyr	Lys	
							450												4	480								
GTT	ACC	CTT	ACT	GAG	AAA	GGA	ACT	TAC	AAG	TTC	TAC	TGC	TCA	ccc	CAC	CAG	GGT	GCT	GGT	ATG	GTG	GGA	AAA	GTA	ACT	GTC	AAC	
vai	Thr	Leu	Inr	GIU	гля	GIY	TUL	TÄL	цуs	Pne	TYT	Cys	Ser	PTO	HIS	GIN	GIY	Ala	GIΫ	met	vai	619	Lys	vai	Thr	vai	Asn	
510 <b>•••</b> 540							540	40						570						<u>▼</u> 600								
TAA	FATT	TTAA	TGAA	AAAT	TGAN	FTTT	TATT	TATA	TGAC	TTTG	TTGT.	AGAG	CTAT	TTAT	ATAA.	ACGY.	FCTC.	ATCA.	ATTG	STAC	GTTT.	AGCA	CATT	rg <u>rg</u>	FAGA	GCTAT	TTA	
630												560 🗸				690						720						
<u>ттсдалатттдотттадатттдатстсаталя</u> тсаталатссалалттдоаттатасс <u>одоттадатттда</u> аттдалстасадтасалстасадтасалалатдосал														ACCA														
					75	b																						
TTC.	AATC	TTTA'	TCTG.	AAAG.	AATT	c																						

Fig. 2. The nucleotide sequence (coding strand) and predicted amino acid sequence of that part of the l23SonPC-1 genomic clone carrying a plastocyanin gene from spinach. The terminal processing site of the precursor is marked by a *triangle*, a putative h-domain (v. Heijne 1986a) indicated by a *horizontal bracket*. Potential "TATA"-, "CAAT"- and polyadenylation motifs are *boxed*. Three *overlined* sequences plus corresponding *arrowheads* mark mRNA 5' and 3' termini as determined by S1 protection analysis (Berk and Sharp 1977). The *arrowheads* at position 594/595 and 669/670 indicate the 3' ends of two different classes of cDNA clones as discussed in the text

deduced was more than 70% identical with that of *Silene* plastocyanin (Smeekens et al. 1985) throughout the entire 168 amino acid residues. It is somewhat less in the presequence (64%), but diverges markedly further upstream. It seems unlikely that such conservation would exist outside the protein-coding region again suggesting that translation begins at the indicated methionine.

The N-terminal transit sequence alone or in conjunction with sequence elements of the structural protein should possess the topogenic signals for routing the protein into the chloroplast and, within the organelle, into the thylakoid lumen. Targeting sequences including those for plastids are quite variable in their primary structure but they share features reflecting probably functionally discrete regions that form the basis for the current domain models for transiently attached transit peptides. These include positively charged N-terminal regions, which are also rich in hydroxylated amino acid residues, a hydrophobic core which is generally somewhat less hydrophobic and shorter than transmembrane anchor segments, and a region surrounding the cleavage site, designated n-, h- and c-domains (von Heijne 1986a, b). On the basis of a comparison predominantly of small subunits of ribulose bisphosphate carboxylases/oxygenases and chlorophyll a/b apoproteins of the light-harvesting complexes II from different sources, Karlin-Neumann and Tobin (1986) have deduced a sequence consensus pattern (homology blocks I-III, interblock regions I and II) for transit sequences of nuclear-encoded chloroplast proteins. The application of this model to plastocyanin has not yet led to an unambiguous picture, probably because the import of this protein into the organelle is complex. It has to traverse the thylakoid membrane in addition to the envelopes. Recent evidence suggests that the plastocyanin precursor is, in fact, processed in at least two steps, by two different protease activities (Robinson and Ellis 1984; Robinson, personal communication), the intermediate processing site being in the amino acid range 42 to 48 in spinach (cf. Smeekens et al. 1986). [The spinach transit sequence is 3 amino acids longer than that of Silene (Smeekens et al. 1985)]. Both transit sequences share the homology blocks I (MAT-VASS, position 1-7) and II (PSFTGLK, positions 13-19) as well as the interblock regions I (5 residues, all hydrophobic) and II (which is unusually long with 41 residues) with ribulose bisphosphate carboxylase/oxygenase

small subunit protein and chlorophyll a/b apoprotein presequences. However, in view of the proposed twostep processing the significance of element III at position 61-67 (GLLAGNA) within the h-region of the Silene transit sequence (cf. Karlin-Neumann and Tobin 1986) is difficult to envisage. Apart from the fact that it is not well conserved in spinach, a block III equivalent should be located *before* the intermediate cleavage site, since ribulose bisphosphate carboxylase/oxygenase small subunits and chlorophyll a/b subunits are not imported into the lumen. It is remarkable that the putative intermediate presequence in itself exhibits properties for signal peptides, in that the hydrophobic segment is preceded by charged amino acid residues. These could aid directing the intermediate through the thylakoid membrane, provided they are retained after the first processing event (cf. Smeekens et al. 1986). It is worth noting that the single negatively charged residue in this segment of the Silene transit sequence is absent in spinach precluding that such a residue would generally facilitate transport of plastocyanin across the thylakoid membrane (von Heijne 1986a; Karlin-Neumann and Tobin 1986). The anatomy of the spinach plastocyanin transit peptide clearly deserves further investigation. A comparative discussion for more than half a dozen pre-sequences will be subject of a separate paper.

## S1 protection analysis

Poly A<sup>+</sup>-RNA of greening spinach seedlings contain substantial amounts of plastocyanin transcripts that hybridize to the excised, nick-translated insert or a SP6 RNA polymerase transcribed RNA probe (Tittgen et al. 1986). The 5' end of the PC-mRNA was mapped using the S1 nuclease protection technique (Berk and Sharp 1977). The 6.6 kbp BamHI fragment (Fig. 1) was labelled at the 5' end and digested with ClaI to yield a 840 bp fragment marked at the BamHI cleavage site at position 64 within the structural gene. This ClaI/BamHI fragment was hybridized to poly A<sup>+</sup>-RNA. In parallel, the same DNA fragment was subjected to sequence analysis (Maxam and Gilbert 1980) in order to provide a standard for comparison with S1resistant hybrids. A 130 bp fragment was protected from S1 nuclease. Taking into account a single base pair correction for the extra residue on the S1-protected fragment (Sollner-Webb and Reeder 1979) and excluding the possibility of mRNA 5' processing, the principal 5' end of plastocyanin mRNA, which is probably the transcriptional start of the plastocyanin gene, is about 65 nucleotides upstream from the AUG start codon. We note that this start is neither at an A residue nor within a PyAPy motif. The sequence TATTAAA, located -27 to -33 nucleotides upstream from the transcriptional initiation site, resembles the "TATA"-box, three options exist for the "CAAT"-motif (positions -96 to -100, -117 to -121 and -173 to -177). The positions and spacings of these elements are consistent with promoter models of nuclear genes (Breathnach and Chambon 1981;Heidecker and Messing 1986). Moreover, no internal AUG inition codons are contained between these signals and the actual start of translation (Kozak 1986).

Similarly, the 4.4 kbp Sall fragment was labelled at the 3' recessed ends, digested with EcoRI and the purified 390 bp SalI-EcoRI fragment (starting at position 355 within the structural gene) used to determine the S1 3' protected end. Two strong signals were obtained corresponding to fragment sizes of ca. 172 and 236 bp. The former band reflects probably an artifact. The simplest explanation of this hybrid is that it results from an  $(A)_5$ (T)<sub>5</sub> inverted repeat at positions 518-522 and 528-532. The latter component would indicate a mRNA of approximately 660 nucleotides, a finding that is consistent with the isolation of a corresponding cDNA clone (Figs. 1 and 2), but is less than the size of the major RNA band detected by Northern analysis (850 bp; Tittgen et al. 1986). The isolation of a cDNA clone extending 73 bp beyond the latter protection site with a sequence identical to that of the genomic DNA strongly suggests that longer plastocyanin mRNA species do exist. Interestingly, both 3' termini are preceded by potential polyadenylation sites TATAAA (position 563-568) and CATAAA (642-647) starting 26 and 21 nucleotides, respectively, before the poly-A stretch of the respective cDNA clones. We interpret these results as indicating heterogeneity within the plastocyanin mRNA population. This finding as well as the significance of two direct repeats (549-563/602-616, and 626-638/674-686, Fig. 2) deserve further study.

Acknowledgements. This work was supported by the Deutsche Forschungsgemeinschaft (grant He 693) AT is a recipient of a fellowship from the German Academic Exchange Service (DAAD) under a Special Gene Technology Programme.

## References

- Benton WD, Davis RW (1977) Science 196:180-182
- Berk AJ, Sharp PA (1977) Cell 12:721-732
- Breathnach R, Chambon P (1981) Ann Rev Biochem 50:349-383
- Bünemann H, Westhoff P, Herrmann RG (1982) Nucleic Acids Res 10:7163-7180
- Dean C, Tamaki S, Dunsmuir P, Favreau M, Katayama C, Dooner H, Bedbrook J (1986) Nucleic Acids Res 14:2229-2240
- Frischauf A-M, Lehrach H, Poustka A, Murray N (1983) J Mol Biol 170:827-842
- Grossman AR, Bartlett SG, Schmidt GW, Mullet JE, Chua N-H (1982) J Biol Chem 257:1558-1563
- Haehnel W (1986) Encycl Plant Physiol 19:547-559
- Heidecker G, Messing J (1986) Ann Rev Plant Physiol 37:439-466

- Heijne G von (1986a) J Mol Biol 189:239-242
- Heijne G von (1986b) EMBO J 5:1335-1342
- Hernández-Lucas C, Royo J, Paz-Ares J, Ponz F, Garcia-Olmedo F, Carbonero P (1986) FEBS Lett 200:103-106
- Herrmann RG, Westhoff P, Alt J, Winter P, Tittgen J, Bisanz C, Sears BB, Nelson N, Hurt E, Hauska G, Viebrock A, Sebald W (1983) In: Cifferri O, Dure L III (eds) Structure and function of plant genomes. Plenum Press, New York, pp 143-153
- Herrmann RG, Westhoff P, Alt J, Tittgen J, Nelson N (1985) In: van Vloten-Doting L, Groot GSP, Hall TC (eds) Molecular form and function of the plant genome. Plenum Press, New York, pp 233-256
- Hohn B, Murray K (1977) Proc Natl Acad Sci USA 74:3259-3263
- Karlin-Neumann GA, Tobin EM (1986) EMBO J 5:9-13
- Kozak M (1986) Cell 44:283–292
- Kröger M, Kröger-Block A (1982) Nucleic Acids Res 10:229-236
- Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: Laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Maxam AM, Gilbert W (1980) Methods Enzymol 65:499-560
- Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR (1984) Nucleic Acids Res 12:7035-7056

- Morris J, Herrmann RG (1984) Nucleic Acids Res 12:2837-2850
- Robinson C, Ellis RJ (1984) Eur J Biochem 142:337-342
- Scawen MD, Ramshaw JAM, Boulter D (1975) Biochem J 147: 343-349
- Smeekens S, de Groot M, van Binsbergen J, Weisbeek P (1985) Nature 317:456-458
- Smeekens S, Bauerle C, Hageman J, Keegstra K, Weisbeek P (1986) Cell 46:365-375
- Sollner-Webb B, Reeder RH (1979) Cell 18:484-499
- Tittgen J (1985) Dissertation, Universität Düsseldorf
- Tittgen J, Herrmans J, Steppuhn J, Jansen T, Jansson C, Andersson B, Nechushtai R, Nelson N, Herrmann RG (1986) Mol Gen Genet 204:258-265
- Tyagi AK, Steppuhn J, Hermans J, Jansson Ch, Vater F, Herrmann RG (Mol Gen Genet, submitted)
- Weaver RF, Weissmann C (1979) Nucleic Acids Res 7:1175-1193
- Westhoff P (1985) Mol Gen Genet 201:115-123
- Willmitzer L, Wagner KG (1981) Exp Cell Res 135:69-77

Communicated by K. Wolf

Received October 6, 1986