Curr Genet (1986) 10:619–624

Current Genetics

Cytochrome *b*-559 genes from *Oenothera hookeri* and *Nicotiana tabacum* show a remarkably high degree of conservation as compared to spinach

The enigma of cytochrome *b*-559: highly conserved genes and proteins but no known function

Nestor Carrillo, Patrick Seyer*, Akhilesh Tyagi, and Reinhold G. Herrmann**

Botanisches Institut der Universität, Universitätsstrasse 1, D-4000 Düsseldorf 1, Federal Republic of Germany

Summary. Recent data suggest that cytochrome *b*-559, an intrinsic membrane protein of the oxygen-evolving photosystem II in chloroplasts, is a heme cross-linked heteromeric polypeptide unit (Herrmann et al. 1984, FEBS Lett 176:239–244). The genes for this cytochrome, designated psbE and psbF, have been located on the chloroplast chromosome of *Oenothera hookeri* and *Nicotiana tabacum* by hybridization with fragments of the corresponding spinach genes, and characterized. In both cases, the nucleotide sequence discloses 2 uninterrupted reading frames of 83 and 39 codons separated by a few nucleotides, as in spinach. The amber translation stop codon of psbE overlaps the putative ribosome-binding site for psbF in all cases. The predicted molecular weights of the proteins are 9.2 and 4.3 kd, respectively. In each of the three plant species, the clustered genes are transcribed into a single RNA species, and the direction of transcription is opposite to that of the gene for cytochrome *f* which is located distal to the *b*-559 genes. Comparison of the deduced amino acid sequences with those from the corresponding spinach genes shows 97% homology. The ubiquitous presence and remarkably high degree of structural and functional conservation of this gene cluster supports the hypothesis of a heteromeric assembly and an important functional role for cytochrome *b*-559.

Key words: Cytochrome *b*-559 genes – Amino acid sequence – Topological model – Thylakoid membrane – *Oenothera* – *Nicotiana*

Present addresses:

* Université de Grenoble, Physiologie Cellulaire, Végétale Domaine Universitaire St. Martin D'Heres, F-38041 Grenoble Cédex, France)

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

Offprint requests to: R. G. Herrmann

Introduction

The chloroplast cytochrome *b*-559 is a membrane protein intrinsic to the core complex of photosystem II (Butler 1978; Cramer and Crofts 1982; Matsuda and Butler 1983). Although biochemical and biophysical data link this cytochrome with the function of this membrane complex, elucidation of its physiological role has remained elusive and some challenging proposals have been made without resolution among the different possibilities (Cramer et al. 1981; De La Rosa et al. 1981; Cramer and Crofts 1982). A 9–10 kd polypeptide associated with this cytochrome has been purified from maize (Metz et al. 1983) and spinach (Widger et al. 1984). Its gene, designated psbE, has been located on the plastid chromosome from spinach (Westhoff et al. 1985) and sequenced (Herrmann et al. 1984). The sequence data suggest that the protein is composed of 83 amino acids and has a molecular mass of 9.2 kd. However, the predicted primary structure as well as amino acid composition data (Widger et al. 1984) show the presence of only a single histidine residue, whereas optical, resonance Raman and electron paramagnetic resonance spectroscopic data clearly indicate bis(histidine) axial ligation for the heme group (Babcock et al. 1985). Therefore, coordination of a single heme demands that this cytochrome is at least a polypeptide dimer (Herrmann et al. 1984; Cramer et al. 1985). The sequence analysis of the spinach gene suggested the striking possibility that a second gene, designated psbF, located distal to the 83 residue reading frame and encoding a 39-residue polypeptide (4.3 kd), might be part of

the cytochrome *b*-559 holoprotein and fulfil this role (Herrmann et al. 1984). This protein again contains only a single histidine residue at an appropriate position within the polypeptide chain. A small component of similar size detected in gels of purified photosystem II particles (Metz et al. 1983; Widger et al. 1984) could represent this component and favour this heteromer model.

The discrepancy between amino acid sequence and biophysical data and the intriguing topological model for holocytochrome *b*-559, that is, that a structure with heme cross-linking of separate polypeptides has not been previously described for any other cytochrome, prompted us to extend our experiments to two other plant species, *Oenothera hookeri* and *Nicotiana tabacum*. This paper, presents the results of this study.

Materials and methods

Oenothera hookeri subsp. hookeri and *Nicotiana tabacum* var. Wisconsin were grown in a greenhouse. Details of the isolation of DNA and RNA from these sources, restriction analysis, agarose and polyacrylamide gel electrophoresis as well as Northern and Southern blot analysis have been described previously (Herrmann et al. 1980, 1983; Westhoff et al. 1983, 1985).

Recombinant DNA techniques. The spinach plastid DNA fragments used to localize the cytochrome *b*-559 genes, psbE and psbF, were (a) the 130 bp TaqI-EcoRI fragment, encompassing nucleotides 52-185 (corresponding to amino acids 18–61) of psbE, (b) a 190 bp DdeI-EcoRI fragment starting 5 bp upstream the ATG codon, and (c) a 550 bp EcoRI fragment containing the C-terminus (amino acid residues 62–83) of psbE, 8 bp intercistronic sequence, the entire gene psbF, and 3′ flanking regions (for details see Herrmann et al. 1984). The 135 bp TaqI-AluI fragment, specific for psbF (nucleotides 272-402 in Fig. 2 of Herrmann et al. 1984 corresponding to amino acid residues 5-39 and 30 nucleotides 3′ untranslated region), was used as an additional probe in Northern blot experiments.

The fragments SalI-5 (15.6 kbp) from *Oe. hookeri* and SalI-9 (5.7 kbp) from *N. tabacum* plastid DNA, each containing both cytochrome *b*-559 genes, were ligated into pBR322 and the recombinant DNAs, designated pOhhp205 and pNtp209, cloned in *E. coli* C600. Plasmid DNA was isolated in large or small scale according to Birnboim and Doly (1979).

Subclones were prepared in pUC18 (Vieira and Messing 1982) after digestion of the *Oenothera* fragment SalI-5 with EcoRI. The chimeric plasmids obtained were designated pOhhp205/E1 to E14 in order of descending insert sizes (Tyagi and Herrmann 1986). Fragments EcoRI-5 (1.2 kbp) and EcoRI-8 (0.9 kbp) were used for sequence analysis. In addition, the second largest, 4.8 kbp subfragment, obtained after digestion of SalI-5 with HindIII (Fig. 1), was incubated with BamHI, yielding a 2.1 kbp HindIII-BamHI fragment that contains the apocytochrome genes. This fragment which overlaps the EcoRI cleavage site was cloned into pEMBL9$^+$ and the recombinant DNA designated pOhhp205/BH1.

The fragment SalI-9 of *Nicotiana* plastid DNA was digested with EcoRI and the first (1.6 kbp) and third (0.65 kbp) largest subfragments were found to carry information for the apocytochrome *b*-559 genes. These fragments were, therefore, selected for sequence studies. Analysis of the region overlapping the EcoRI cleavage site was performed by digestion of the fragment SalI-9 with DdeI. A 0.6 kbp DdeI-DdeI subfragment, encompassing both genes, was directly used.

DNA sequence analysis. Excised DNA fragments prepared as above were digested with appropriate restriction endonucleases to obtain small secondary fragments for sequencing. These fragments were labelled at their 3′-termini by fill-in synthesis using α-^{32}P-labelled deoxyribonucleotide triphosphates and the large fragment of DNA polymerase I (Morris and Herrmann 1984). After digestion with a second restriction endonuclease radiolabelled fragments were recovered from polyacrylamide gels and sequenced according to Maxam and Gilbert's (1980) procedure, essentially as described previously (Herrmann et al. 1984).

Nuclease S1 mapping. Labelling of dephosphorylated 5′ ends with γ-^{32}P-ATP and T_4 polynucleotide kinase was performed according to Maxam and Gilbert (1980). S1 nuclease protection experiments (Berk and Sharp 1977) on the 5′ end of the cytochrome *b*-559 gene were carried out essentially as described by Weaver and Weissmann (1979) using a temperature gradient (48 °C to 42 °C) for hybridization.

Results and discussion

In order to locate the gene for cytochrome *b*-559 on the previously published physical maps of the plastid chromosomes from *Oe. hookeri* (Gordon et al. 1982) and *N. tabacum* (Seyer et al. 1981), a 130 bp TaqI-EcoRI fragment of the spinach gene psbE, encoding most of the 83 amino acid residue chain (Herrmann et al. 1984), was chosen as a hybridization probe to SalI restricted plastid DNA from both organisms. The probe hybridized to the fifth largest (15.6 kbp) fragment SalI-5 of *Oe. hookeri* and to the 5.7 kbp fragment SalI-9 of *N. tabacum* (Herrmann et al. 1985; Bisanz-Seyer 1985). These fragments were selected for further study.

To determine the presence of both apocytochrome gene, psbE and psbF, and to locate them more precisely, refined maps of the gene-containing regions were constructed by comparing cross-cleavage patterns and DNA-DNA hybridization to derivative fragments obtained after digestion of SalI-5 (*Oenothera;* Fig. 1) or SalI-9 (*Nicotiana*) with BamHI, HindIII and EcoRI. Signals were obtained in all cases with the three chosen probes that are representative for different regions of the spinach genes (see above; Fig. 2). These findings suggested the existence of both genes in the *Oenothera* and tobacco chromosomes. The N-terminal region of psbE in *Oenothera* was found on the 1.2 kbp (EcoRI-5), the C-terminus and psbF on the 0.9 kbp (EcoRI-8) EcoRI subfragments, respectively (Fig. 2). A similar strategy was employed to determine the fine structure of the tobacco gene on fragment SalI-9 (data not shown). Again, psbE was found to be dissected by an EcoRI cleavage site. A 1.6 kbp subfragment (EcoRI-1) carries the N-terminus, a 0.65 kbp

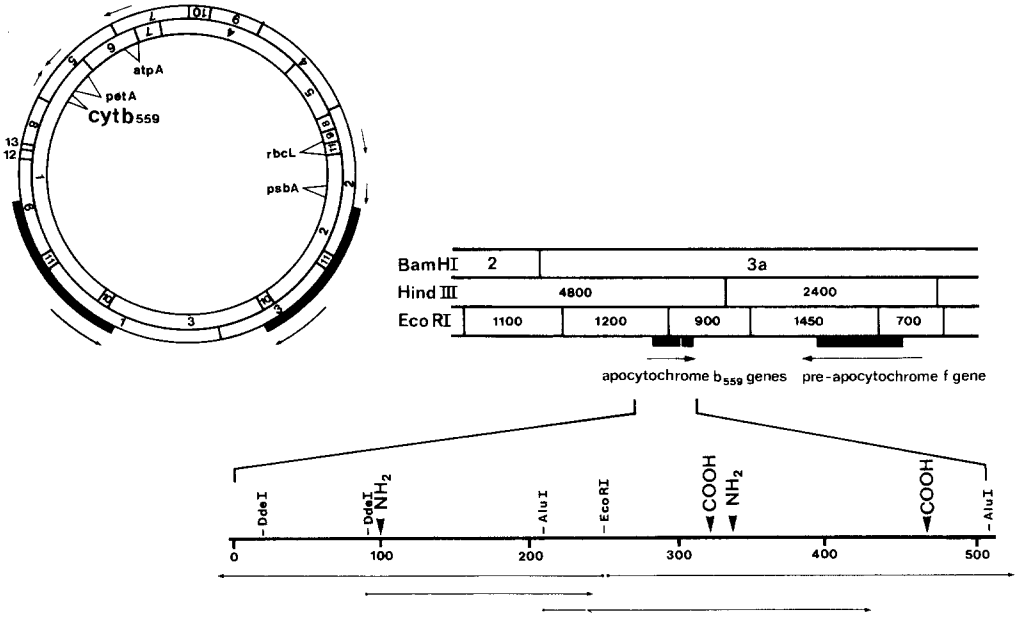

Fig. 1. Detail restriction map and sequencing strategy diagram of the region of the *Oe. hookeri* plastid chromosome carrying the genes for apocytochrome *b*-559. The position of the genes for cytochrome *b*-559 and *f* (*bold areas*) and their direction of transcription (*arrows*) are indicated in the above map. DNA fragments are designated in descending size (BamHI) or given in bp (HindIII, EcoRI). Restriction sites used for DNA sequence analysis and the direction and extent of individual sequence reactions (*arrows*) are indicated in the lower part of the figure. The circular map of the *Oenothera* plastid chromosome shows the restriction sites for SalI (*outer circle*), KpnI (*inner circle;* Gordon et al. 1982) and the location of the genes for cytochrome *b*-559 (cyt b559), cytochrome *f* (petA), the herbicide-binding 32 kd protein of photosystem II (psbA), ATP synthase subunit alpha (atpA) and the large subunit of ribulose bisphosphate carboxylase (rbcL); the transcription polarities are indicated by *arrows*. The fragment numbers are in decreasing order. The inverted repeat is depicted by *bold lines* and the *outer parallel arrows* indicate the position and transcription polarity of the two rDNA operons

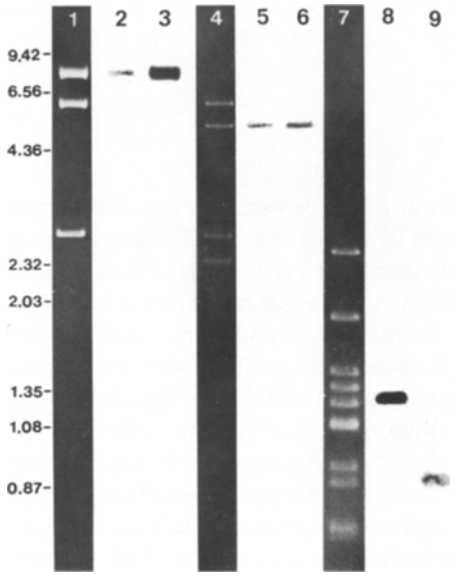

Fig. 2. Hybridization of nick-translated fragments of the spinach apocytochrome *b*-559 to restriction digests of *Oenothera* SalI-5. *Lanes 1, 4* and *7:* digests obtained with BamHI, HindIII and EcoRI, respectively (1.5% agarose gel). Hybridization with the N- (*lanes 2, 5, 8*) and C-terminal (*lanes 3, 6, 9*) spinach cytochrome *b*-559 probes (fluorography, cf. also text). DNA fragments sizes are given in kbp (*left*)

(EcoRI-3) fragment the carboxyl end of the gene and psbF (data not shown). The conserved EcoRI site within all psbE genes helped to assign the relative position of both genes and their transcription polarity. In all three instances, the genes are located at equivalent positions on the circular plastid chromosomes, within the large single-copy region distal to and on the opposite strand of the gene for cytochrome *f* (Herrmann et al. 1984; Seyer et al. 1981; Tyagi and Herrmann 1986; Fig. 1).

The respective EcoRI fragments and the overlapping HindIII-BamHI (2.1 kbp; *Oenothera*) and DdeI-DdeI (0.6 kbp; *Nicotiana*) fragments were subjected to sequence analysis. The nucleotide sequence of the protein-coding and flanking regions, determined according to the strategy illustrated by Fig. 1, confirmed the existence of both genes (Fig. 3).

Alignment of these sequences and of the predicted amino acid sequence with those from spinach (Herrmann et al. 1984; Fig. 3) reveals that the genes of the three organisms are highly conserved. They share more than 90% of the nucleotides, 97% homology at the amino acid level, and in all cases encode proteins of the same size, of 83 and 39 amino acid residues. Only two changes are observed in the first polypeptide. The glycine residue (GGA codon) at position seven in the predicted *Oenothera* pro-

Fig. 3. The nucleotide sequences of the cytochrome *b*-559 genes, psbE and psbF, and flanking region from *Oenothera hookeri,* tobacco and spinach. The sequences of the non-transcribed strand of the genes are arranged in codons and the corresponding amino acids of the *Oenothera* proteins are indicated. Numbering starts at the ATG triplet at which translation is probably initiated. Nucleotide differences between the *Oenothera* gene and the spinach (Herrmann et al. 1984) and tobacco genes and any resulting amino acid changes are shown below the *Oenothera* sequence. Possible ribosome binding sites are *boxed,* inverted and direct repeats are indicated by *arrows.* Hydrophobic regions which might span the membrane are indicated by *horizontal brackets;* the heme-complexing histidines are *boxed. Thick arrows* mark the RNA 5′ termini as determined by S1 protection analysis (see text)

tein is replaced by glutamic acid (GAA) in spinach and tobacco, and the serine (TCT) at position 72 in *Oenothera* and spinach by proline (CCT) in the tobacco protein. Similarly, only two amino acid changes were observed in the 39 residue polypeptide, namely, an isoleucine (ATT) instead of valine (GTT) at the 17th position in spinach and tobacco, and a phenylalanine residue (TTT) replaced serine (TCT) at position 26 in tobacco. Most of these changes are conservative replacements. The intercistronic region in the case of *Oe. hookeri* and *N. tabacum* is short as in spinach but contains an additional nucleotide (A and T, respectively) which shifts the second reading frame "in-phase" with the first one. The TAG termination codon of psbE overlaps the ribosome binding sites of psbF, GGAGG, in all organisms. The untranslated 5′ and 3′ flanking regions diverge markedly, with the exception of a few sequence motifs including the putative ribosome binding sites (Shine and Dalgarno 1974) that precede both reading frames (Fig. 3).

In spinach, both *b*-559 proteins are decoded from a single RNA species (Westhoff et al. 1985). The similarity in the arrangements of the genes in *Oenothera* and tobacco suggested that their functional organization might be comparable. Northern blot and S1 protection analysis confirms this and extends the uniformity of the protein-coding regions to the functional organization of this DNA segment. When gene-specific sequences from the 1.2 and 0.9 kbp EcoRI fragments of *Oe. hookeri* (Fig. 1) were hybridized to size-fractionated *Oenothera,* spinach or tobacco chloroplast RNA, immobilized on nitrocellulose strips, a prominent 1.1 kb long RNA and a minor species of 1.7 kb were detected in each instance, regardless of the chosen probe (Fig. 4, lanes 1–6, cf. also Westhoff et al. 1985). The 135 bp TaqI-AluI fragment, specific to the second reading frame from spinach, selects the same RNA species (cf. Fig. 4, lane 7).

The 5′ ends of the cytochrome *b*-559 RNA species in spinach and *Oenothera* were mapped by S1 protection

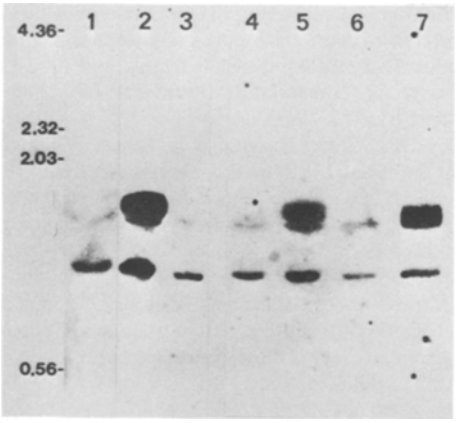

Fig. 4. Autoradiographs of Northern blots obtained after hybridization of nick-translated fragments of spinach plastid DNA to size-fractionated (1.2% agarose slab gel) chloroplast RNA from spinach (*lanes 1, 4*), *Oenothera* (*lanes 2, 5, 7*) and tobacco (*lanes 3, 6*). Hybridization with (*lanes 1–3*) the 1.2 kbp fragment EcoRI-5 from *Oenothera*, (*lanes 4–6*) the 0.9 kbp fragment EcoRI-8 from *Oenothera* and (*lane 7*) the 135 bp TaqI-AluI fragment of the spinach gene psbF. For description of the fragments see text. The strong signal in the *Oenothera* lanes represents an unspecifically hybridizing contaminant, probably mucilage. *Left:* molecular weight standards in kb

experiments. Fragments corresponding to the 5′ terminal region of the gene were prepared by digestion of the EcoRI subfragments with DdeI, labelled at the DdeI 5′ ends at −3 (spinach) and −75 (*Oenothera*) and annealed with homologous RNA (Berk and Sharp 1977). After incubation of the DNA/RNA hybrids with S1 nuclease, the protected DNA fragments were sized on polyacrylamide/urea gels. Fragments of 85 and 30 nucleotides were observed in spinach and *Oenothera*, respectively. When these fragments were co-electrophoresed with the end-labelled DNA strands that had been chemically modified and cleaved for sequence analysis, the two 5′ ends were found to be coincident with positions −90 to −92 (spinach) and −105 (*Oenothera*) nucleotides upstream the translational initiation codon of the gene (data not shown; Fig. 3). Although these RNA ends are preceded by potential promotor-like "−10" and "−35" consensus sequences (Fig. 3; Rosenberg and Court 1979), it is possible that the common RNA species are the result of processing and part of a longer transcript.

The exact pathway and mechanism of function of cytochrome *b*-559 is currently a subject of investigation in many laboratories but until recently surprisingly little information was available about this electron carrier and its association with the photosystem II complex. The key to this cytochrome were the findings of the downstream 39 residue gene in spinach, that the 9.2 and 4.3 kd polypeptides contained only one histidine residue each (Herrmann et al. 1984), and that heme coordination was found to be bis(histidine) from spectroscopic studies (Babcock et al. 1985). The detection of a small protein in purified *b*-559 preparations with the expected amino acid composition and in an approximate 1:1 stoichiometry with the 9.2 kd component (Widger et al. 1985), and the recently determined partial N-terminal sequence from the spinach protein that was found to match the predicted amino acid sequence from psbF (Widger et al. 1985) demonstrate that the downstream 39 residue gene is expressed. The high degree of structural and functional conservation among the genes from the three plant species including the positions of the key amino acids Arg-18, His-23 and Asp-45 in psbE, and Arg-13, His-18 and Arg-39 in psbF proposed to be involved in the topographical model of this electron carrier (Herrmann et al. 1984), implies that the resulting structures and functions of *b*-559 from different organisms are superimposable. A similar gene organization has been found in *Synechocystis* (H. Pakrasi, personal communication). We conclude from the ubiquitous presence of both genes that the 39 residue protein is an integral part of the *b*-559 heme-binding unit. The proximity of the genes, their co-transcription, the existence of putative ribosome binding sites preceding each of these genes, and the structural similarity of both proteins in having a remarkably related N-terminal sequence (see below), a single potential membrane spanning segment, and a single histidine residue as a possible heme ligand at the same relative position within the hydrophobic domain of each polypeptide, all support strongly that the cytochrome contains both the 9.2 and 4.3 kd polypeptides and that the simplest cytochrome *b*-559 structural unit is a heme cross-linked dimer containing one copy of each polypeptide.

The predicted membrane folding patterns suggest that each of the polypeptides spans the thylakoid membrane once with the position of the single heme-linked histidine at an equivalent position within the hydrophobic domain, five residues from charged arginine residues that border these transmembrane segments on the N-terminal side (Fig. 3) and probably arrest the spans in the polar headgroups of the lipid phase (for schemes see Herrmann et al. 1984; Widger et al. 1985). This model explains most features of this cytochrome. It is consistent with the spectroscopic data and with the known lability in heme-binding and redox potential of this cytochrome. Although details such as number and arrangement of polypeptide chains remain to be settled (for discussion of more complicated tetramer structures see Widger et al. 1985) the crucial feature of such a unit, the heme cross-linking of discrete polypeptide chains, has not been previously reported for any other cytochrome. The conclusion that the unit heme-binding *b*-559 structure in the photosystem II complex contains two discrete polypeptide species is, therefore, an important change in viewpoint.

The topological models and the gene organization for cytochrome *b*-559 outlined here provide a paradigma of the potential of primary structural data deduced from DNA sequence analysis. In addition to serving as a basis for the solution of some as yet more pressing problems concerning architecture and the enigmatic function of this – highly conserved – electron carrier, e.g., by site-specific mutagenesis, rearrangement of protein domains and reconstitution or transformation, the sequence analysis can help in elucidating aspects of phylogeny and biogenesis of this cytochrome. It will be of interest to find out in this context whether cytochrome *b*-559 represents a residual split gene or has arisen by gene duplication (see above), and why this component is detectable in etiolated material in contrast to all other photosystem II core components (P. Westhoff and R. G. Herrmann, manuscript in preparation).

Acknowledgements. N. C. received support from the German Research Foundation (DFG), P.S. from the DFG/CNRS exchange program, A.T. is a Visiting Academic Fellow under the special gene-technology programme of the German Academic Exchange Service (DAAD). This work was supported by the DFG, the Fonds der Chemischen Industrie and the Stiftung Volkswagenwerk.

References

Babcock GT, Widger WR, Cramer WA, Oertling WA, Metz JG (1985) Biochemistry 24:3638–3645

Berk AJ, Sharp PA (1977) Cell 12:721–732

Birnboim HC, Doly J (1979) Nucleic Acids Res 7:1513–1523

Bisanz-Seyer C (1985) Dissertation, University of Düsseldorf

Butler WL (1978) FEBS Lett 95:19–25

Cramer WA, Crofts AR (1982) In: Govindjee (ed) Photosynthesis: energy conversion by plants and bacteria, vol 1. Academic Press, New York London, pp 387–467

Cramer WA, Whitmarsh J, Widger WR (1981) In: Akoyunoglou G (ed) Photosynthesis: electron transport and phosphorylation, vol 2. International Science Service, Philadelphia, pp 509–522

Cramer WA, Widger WR, Herrmann RG, Trebst A (1985) Trends Biochem Sci 10:125–129

De La Rosa FF, Galvan F, Losada M (1981) In: Akoyunoglou G (ed) Photosynthesis: electron transport and phosphorylation, vol 2. International Science Service, Philadelphia, pp 531–543

Gordon KHJ, Crouse EJ, Bohnert HJ, Herrmann RG (1982) Theor Appl Genet 61:373–384

Herrmann RG, Seyer P, Schedel R, Gordon K, Bisanz C, Winter P, Hildebrandt JW, Wlaschek M, Alt J, Driesel AJ, Sears BB (1980) In: Bücher T, Sebald W, Weiss H (eds) Biological chemistry of organelle formation. Springer, Berlin Heidelberg New York, pp 97–112

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: Ciferri O, Dure III L (eds) Structure and function of plant genomes. Plenum Press, New York, pp 143–154

Herrmann RG, Alt J, Schiller B, Widger WR, Cramer WA (1984) FEBS Lett 176:239–244

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 233–255

Matsuda H, Butler WL (1983) Biochim Biophys Acta 725:320–324

Maxam AM, Gilbert W (1980) Methods Enzymol 65:499–560

Metz JG, Ulmer G, Bricker TM, Miles D (1983) Biochim Biophys Acta 725:203–209

Morris J, Herrmann RG (1984) Nucleic Acids Res 12:2837–2850

Rosenberg M, Court D (1979) Annu Rev Genet 13:319–353

Seyer P, Kowallik KV, Herrmann RG (1981) Curr Genet 3: 189–204

Shine J, Dalgarno L (1974) Proc Natl Acad Sci USA 71:1342–1346

Tyagi AK, Herrmann RG (1986) Curr Genet 10:481–486

Vieira J, Messing J (1982) Gene 19:259–268

Weaver RF, Weissmann C (1979) Nucleic Acids Res 7:1175–1193

Westhoff P, Alt J, Nelson N, Bottomley W, Bünemann H, Herrmann RG (1983) Plant Mol Biol 2:95–107

Westhoff P, Alt J, Widger WR, Cramer WA, Herrmann RG (1985) Plant Mol Biol 4:103–110

Widger WR, Cramer WA, Hermodson M, Meyer D, Gullifor M (1984) J Biol Chem 259:3870–3876

Widger WR, Cramer WA, Hermodson M, Herrmann RG (1985) FEBS Lett 191:186–190

Communicated by C. P. Hollenberg

Received December 19, 1985