PRIMARY STRUCTURE OF BARLEY GENES ENCODING QUINONE AND CHLOROPHYLL *a* BINDING PROTEINS OF PHOTOSYSTEM II

by

EVA MARIA NEUMANN

Department of Physiology, Carlsberg Laboratory,
Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby

Present address: Department of Biology, Texas A & M University,
College Station, Texas 77843-3258, USA

Keywords: Chloroplast DNA, D-1 polypeptide, D-2 polypeptide, Chl_{*a*}-P3, psbD, psbC, psbA

The psbA, psbD and psbC genes encoding the quinone binding D-1 and D-2 apoproteins and the 44 kD chlorophyll *a*-apoprotein 3 have been located in the chloroplast genome of barley. They are found on a 23 kbp SalI restriction endonuclease fragment in the large single copy region of the chloroplast DNA adjacent to the inverted repeat. As in other species the psbD and psbC genes have reading frames which overlap by 53 bp. They are transcribed in the same direction but translated with a frameshift of one nucleotide. Ten amino acid substitutions are found among the 18 N-terminal residues of the D-2 polypeptide if barley, spinach, tobacco, pea and the liverwort *Marchantia* are compared. Only 8 substitutions are present among the 335 other residues of the D-2 polypeptide. The amino acid residues located in the putative binding site for the special pair reaction center chlorophyll and the residues probably serving as ligands to non-heme iron in the D-1 and D-2 proteins of barley are strictly conserved when compared to those of purple bacteria and of other higher plants. Identity is also observed for the residues of importance in the binding of quinones.

1. INTRODUCTION

Photosystem II (PSII) catalyses the reduction of plastoquinone (PQ) by H₂O. The functional components of PSII are located in the thylakoid membrane and constitute an integral membrane protein complex. Four integral membrane proteins, known as D-1, D-2, Chl_{*a*}-P2 (47 kD protein) and Chl_{*a*}-P3 (44 kD protein) have recently elicited particular interest as to their functional role. The 47 and the 44 kD proteins are recognized as chlorophyll *a* binding proteins since chlorophyll remains attached to the

protein after SDS-polyacrylamide gel electrophoresis. D-1 was identified as target for inhibitors of photosynthetic electron transport (35) suggesting that D-1 binds also the secondary electron acceptor plastoquinone Q_B. Little is known about the function of D-2, but it has been noted that D-2 exhibits structural similarities with D-1, suggesting that D-2 also has a quinone binding site (39).

The precise localization of the primary charge separation in PSII of higher plants remains to be determined. Because of correlations of PSII

Abbreviations: bp = basepair; Chl_{*a*}-P = Chlorophyll *a*-protein; cpDNA = Chloroplast DNA; Da = Dalton; kbp = kilobasepair; kD = kilodalton; PQ = plastoquinone; PSII = photosystem II; SDS = sodium dodecyl sulphate; SSC = 0.15 M-sodium chloride + 0.015 M-sodium citrate.

activity with amount of the 47 kD polypeptide in isolated fractions (5) and because of the fluorescence emission at 695 nm by Chl_a-P2 (31) it was considered until recently that the 47 kD polypeptide binds the P680 reaction center chlorophyll. Analysis of the crystallized photosynthetic reaction center of purple bacteria (9, 28, 29) led to new hypotheses about the structural and functional role of PSII components in higher plants. The pronounced similarities of the location of certain amino acid residues and of the distribution of hydrophobic domains in the reaction center polypeptides M and L from purple bacteria and in the proteins D-1 and D-2 from PSII gave rise to the suggestion that D-1 and D-2 bind the reaction center chlorophyll, the electron acceptors Q_A and Q_B as well as non-heme iron (45, 46). This notion is supported by the isolation of a PSII reaction center complex consisting of D-1, D-2 and cytochrome *b*-559 polypeptides with five chlorophyll *a*, two pheophytin *a*, one β -carotene and two cytochrome *b*-559 molecules (25, 32). The complex exhibits reversible absorbance changes indicative of accumulation of reduced pheophytin.

Chl_a-P2 and Chl_a-P3 serve as photosystem II antenna complex. Additionally, proper insertion of the apoproteins of 47 and 44 kD into the thylakoid membrane together with D-1 and D-2 is essential for the physiological function of PSII (2, 11, 48, 49) and suggests transcriptional and/or translational regulation of the coordinated expression of their genes. All four polypeptides are coded for by the chloroplast DNA and they have been mapped and sequenced in several higher plants as well as in green algae and cyanobacteria. Information on the D-1 coding gene *psbA* is reviewed in (23), that for the D-2 coding gene *psbD* is presented in (1, 18, 38, 42, 47), while the structure of the *psbB* gene, coding for the 47 kD apoprotein, and the *psbC* gene, coding for the 44 kD apoprotein, is reported in (1, 18, 21, 30, 42, 47). Transcriptional features of the genes for these proteins are complex in that hybridization of plastid RNA to the neighbouring *psbD* and *psbC* genes revealed a heterogeneous population of RNA transcripts (1, 3, 33). For example in plastids of dark-grown barley plants six different RNAs, ranging in size

from 5.7 kb to 1.7 kb, hybridize to the *psbD*/*psbC* gene region (3, 33). Mapping of the transcripts reveals that the larger ones hybridize to both genes and thus belong to a transcriptional unit as it is also shown for *psbD* and *psbC* in spinach (1). The shorter transcripts, however, hybridize only to the *psbC* coding region. The physiological significance of this RNA heterogeneity and the molecular basis of the regulation are being investigated.

In this paper the nucleic acid sequences for the barley *psbD* and the *psbC* genes as well as for the adjacent region upstream from the *psbD* gene are determined. Structural features of the *psbD* and the *psbC* genes which elucidate the RNA heterogeneity and the complex transcriptional regulation of these genes are discussed. Whereas the *psbD* and the *psbC* genes have been sequenced from several dicotyledonous plants and from the green alga *Chlamydomonas*, herein the nucleic acid sequence for *psbD* from a monocotyledonous plant is presented together with a partial nucleic acid sequence from the barley *psbA* gene. In a subsequent paper the in vitro expression of the D-2 polypeptide and the possibility of binding quinones to it is explored.

2. MATERIAL AND METHODS

2.1. Isolation of chloroplast DNA (cpDNA)

cpDNA was isolated from seedlings of *Hordeum vulgare* L. var Bonus. Seeds were germinated and grown in moist vermiculite for six days in the dark at 23 °C. Before cpDNA isolation seedlings were illuminated for eight hours in white light (2400 lux). Isolation of cpDNA was performed according to (36).

2.2. Hybridization probes

Two probes were used for mapping the *psbA* gene to the barley cpDNA: 1) An oligonucleotide was synthesized on an Applied Biosystems Oligonucleotide Synthesizer. It had the sequence 5'-GCTGGTGTATTCGGCGGC-3' which is homologous to a highly conserved sequence in the spinach (52), tobacco (52) and soybean (43) *psbA* genes. The oligonucleotide was 5' endlabelled by T4 DNA polynucleotide kinase with (α -³²P)-dATP (> 7000 Ci/mmol) (24). 2) As

second probe served a 790 bp PstI/XbaI fragment of the spinach *psbA* gene, obtained by digestion and insert preparation of the plasmid pSocS15, kindly provided by J. BOHNERT and described in (52).

For mapping the *psbD* gene a 1.2 kbp EcoRI/EcoRI fragment was used. It was isolated from the plasmid 531-4 which contains on a 3.4 kbp BamHI/BamHI fragment the barley *psbD* and part of the *psbC* genes. This plasmid was a gift from J. MULLET (3).

For hybridization experiments DNA probes were nick-translated using *E. coli* DNA polymerase I and (α - 32 P)-dATP (> 3000 Ci/mmol). Hybridization was carried out in 6 $\times$ SSC, 0.25% blotto (20), 0.1% SDS at 55 °C. Filters were washed at 60 °C (or 70 °C with use of the oligonucleotide as probe) with 0.2 $\times$ and 0.1 $\times$ SSC.

Southern blots to nitrocellulose filters were performed according to (24).

2.3. Preparation of plasmid DNA and inserts, subcloning, transformation

Plasmid DNA minipreparations were made according to (15), omitting the polyethylene glycol precipitation step. For sequencing reactions using the dideoxy chain termination method plasmid DNA was purified from amplified bacteria cultures by caesium chloride gradient centrifugation (7, 16).

Insert preparations from agarose or acrylamide gels were either performed by electroelution into dialysis bags (41) or by use of the "GeneClean"-kit, obtained from BIO 101, La Jolla, CA, USA, according to the manufacturers' instructions.

Subcloning was performed into the multiple cloning site of pUC 13, pUC 18 (50) or pGEM 3 (Promega Biotec, Madison, WI, USA). If no appropriate cloning site was available, the DNA insert was cloned into the SmaI site after filling in of protruding ends either by use of *E. coli* DNA polymerase I (Klenow fragment) or T4 DNA polymerase (24) in the presence of deoxynucleotidetriphosphates.

Transformation: *E. coli* K12 strain HB 101 (27) was used as host for pGEM 3 recombinant and strain DH5 α (BRL, Cambridge, United

Kingdom) for pUC recombinant plasmids. Transformation was performed using the CaCl₂ procedure according to (24).

2.4. DNA sequence determination

Three different methods were used to determine the DNA sequences. 1) After endlabelling with (γ - 32 P)-ATP and T4 polynucleotide kinase, endonucleolytic cleavage, and purification of the DNA fragments, the sequencing procedure of MAXAM and GILBERT (26) was employed. 2) The dideoxy chain termination sequencing method using alkali denatured plasmid templates followed the modified procedure of CHEN and SEEBURG (6): 2 μ g denatured and ethanol precipitated supercoiled DNA plasmid template was solubilized in 9 μ l sterile H₂O. 50 ng primer (1 μ l), 1.5 μ l 10 $\times$ sequencing buffer (75 mM-Tris HCl pH 7.6; 75 mM-DTT; 50 mM-MgCl₂) and 2 μ l 32 P-dATP (> 600 Ci/mmol) were added. The annealing reaction was carried out at 70 °C for 15 min. After slowly cooling down to room temperature 1.5 μ l (5U/ μ l) *E. coli* DNA polymerase I (Klenow fragment) was added and 3.5 μ l samples dispensed to each of the four reaction tubes containing the appropriate dideoxy- and deoxynucleotide triphosphates (40). The chain elongation and the "chase" (40) reaction temperature was 37 °C. The samples were then dried down and dissolved in 4 μ l 10 mM-EDTA pH 8.0; 3 mg/ml bromophenolblue; 3 mg/ml xylene cyanol in deionised formamide. 3) When it was not possible to obtain a reasonable sequence determination by either of these methods, denatured plasmid templates were then sequenced by use of a modified T7 DNA polymerase, included in the "Sequenase"-kit (United States Biochemical Corporation, Cleveland, Ohio, USA) following the manufacturers' instructions. Primers were synthesized on an Applied Biosystems Oligonucleotide Synthesizer and purified according to the manufacturers' instructions. Sequences were chosen which allow annealing to the right or left hand borders of the multiple cloning site of pUC or pGEM plasmids. Other sequences were chosen which were homologous to sequences from the inserted gene. This selection facilitated further sequencing from the same or opposite strand.

Figure 1. Hybridization of psbA and psbD probes to barley cpDNA. Four μ g of barley cpDNA were digested with SalI, EcoRI, BamHI and HindIII, electrophoresed on a 0.8% agarose gel and transferred to nitrocellulose. Filter A was hybridized with a synthetic psbA oligonucleotide probe (SalI, EcoRI, BamHI, HindIIIa) and a spinach psbA probe (HindIIIb). Filter B was hybridized with the barley psbD probe. Standard sizes are shown in bp on the left.

2.5. Enzymes and chemicals

Radioactive nucleotides were purchased from New England Nuclear. Most of the enzymes as well as deoxy- and dideoxynucleotides were from Boehringer Mannheim, except restriction endonuclease HgiAI, which was obtained from New England Biolabs, Beverly, MA, USA.

3. RESULTS

3.1. Mapping of the psbA and the psbD genes in the barley cpDNA genome

Figure 1 shows the radioactive hybridization patterns obtained upon incubation of filter

bound cpDNA from barley, digested with different restriction endonucleases, with gene specific psbA and psbD probes. As probe for the psbA gene served a synthetic oligonucleotide of 18 bases corresponding to the codons for amino acids 203 to 208 in the tobacco D-1 polypeptide (52) which are conserved in the psbA genes of spinach (52) and soybean (43). This oligonucleotide hybridized to a 23 kbp SalI, a 2.26 kbp EcoRI, a 6.7 kbp BamHI and a 1.6 kbp HindIII fragment (Figure 1A). A second probe for the psbA gene was a 760 bp PstI/XbaI heterologous DNA fragment obtained from the spinach psbA gene. This probe hybridized to the same 1.6 kbp

Figure 2. A: Physical map of barley cpDNA showing the localization and orientation of the 20.3 kb *SalI* fragment, B: Map of a 3.4 kb *BamHI* subclone and C: Map of a 1.6 kb *HindIII* subclone, located on a *HindIII*/*EcoRI* fragment. B and C display also the sequencing strategy of the subclones. The arrows indicate the origin and extent of DNA sequences. Arrows marked with "M/G" indicate a sequence determination following the procedure of MAXAM and GILBERT. Arrows marked with "T7" indicate the use of T7 DNA polymerase in sequence determination.

HindIII fragment as the synthetic oligonucleotide.

In order to map the *psbD* gene on the cpDNA genome from barley a homologous, nick-translated 1.2 kbp *EcoRI* fragment (2.2) was hybridized to endonuclease digested filter bound

cpDNA. As is shown in Figure 1B hybridization occurred to a 23 kbp *SalI*, a 3.4 kbp *BamHI* and a 3.34 kbp *HindIII* fragment. But when *EcoRI* digested cpDNA was used the probe did not only hybridize to the predicted 1.2 kb *EcoRI* fragment but in addition to a 2.26 kb *EcoRI* frag-

E. M. NEUMANN: Genes encoding photosystem II proteins

```

                                30                                60
[ATG] ACT ATA GCC CTT GGT AGA GTT CCT AAA GAA GAA AAT GAT CTA TTT GAT ACT [ATG] GAT
Met Thr Ile Ala Leu Gly Arg Val Pro Lys Glu Glu Asn Asp Leu Phe Asp Thr Met Asp

                                90                                120
GAC TGG TTA CGA AGG GAC CGT TTC GTT TTT GTA GGA TGG TCT GGC CTA TTG CTC TTT CCT
Asp Trp Leu Arg Arg Asp Arg Phe Val Phe Val Gly Trp Ser Gly Leu Leu Leu Phe Pro

                                150                                180
TGT GCT TAT TTC GCT TTA GGG GGT TGG TTT ACA GGG ACA ACT TTT GTA ACT TCT TGG TAT
Cys Ala Tyr Phe Ala Leu Gly Gly Trp Phe Thr Gly Thr Thr Phe Val Thr Ser Trp Tyr

                                210                                240
ACC CAT GGA TTG GCA AGT TCC TAT TTG GAA GGT TGT AAT TTC TTA ACC GCA GCA GTT TCT
Thr His Gly Leu Ala Ser Ser Tyr Leu Glu Gly Cys Asn Phe Leu Thr Ala Ala Val Ser

                                270                                300
ACC CCT GCC AAT AGT TTA GCA CAC TCT TTG TTG CTA CTA TGG GGG CCA GAA GCA CAA GGA
Thr Pro Ala Asn Ser Leu Ala His Ser Leu Leu Leu Leu Trp Gly Pro Glu Ala Gln Gly

                                330                                360
GAT TTT ACT CGT TGG TGT CAA TTA GGC GGT CTA TGG ACT TTT GTA GCT CTC CAC GGG GCT
Asp Phe Thr Arg Trp Cys Gln Leu Gly Gly Leu Trp Thr Phe Val Ala Leu His Gly Ala

                                390                                420
TTT GCA CTA ATA GGT TTC ATG TTA CGC CAA TTT GAA CTT GCT CGG TCT GTT CAA TTG CGG
Phe Ala Leu Ile Gly Phe Met Leu Arg Gln Phe Glu Leu Ala Arg Ser Val Gln Leu Arg

                                450                                480
CCT TAT AAT GCA ATC TCA TTC TCT GGT CCA ATT GCT GTT TTT GTT TCG GTA TTC CTT ATT
Pro Tyr Asn Ala Ile Ser Phe Ser Gly Pro Ile Ala Val Phe Val Ser Val Phe Leu Ile

                                510                                540
TAT CCA CTG GGG CAA TCT GGT TGG TTC TTT GCG CCG AGT TTT GGC GTA GCA GCG ATA TTT
Tyr Pro Leu Gly Gln Ser Gly Trp Phe Phe Ala Pro Ser Phe Gly Val Ala Ala Ile Phe

                                570                                600
CGA TTC ATC CTT TTC TTC CAA GGA TTT CAT AAT TGG ACG TTG AAC CCA TTT CAT ATG ATG
Arg Phe Ile Leu Phe Phe Gln Gly Phe His Asn Trp Thr Leu Asn Pro Phe His Met Met

                                630                                660
GGA GTT GCC GGA GTA TTA GGC GCG GCT CTG CTA TGC GCT ATT CAT GGA GCA ACC GTA GAA
Gly Val Ala Gly Val Leu Gly Ala Ala Leu Leu Cys Ala Ile His Gly Ala Thr Val Glu

                                690                                720
AAC ACT CTA TTT GAG GAC GGT GAT GGT GCA AAT ACC TTC CGT GCT TTT AAC CCA ACT CAA
Asn Thr Leu Phe Glu Asp Gly Asp Gly Ala Asn Thr Phe Arg Ala Phe Asn Pro Thr Gln

                                750                                780
GCT GAA GAA ACT TAT TCA ATG GTC ACT GCT AAC CGC TTT TGG TCC CAA ATC TTT GGT GTT
Ala Glu Glu Thr Tyr Ser Met Val Thr Ala Asn Arg Phe Trp Ser Gln Ile Phe Gly Val

                                810                                840
GCT TTT TCC AAT AAA CGT TGG TTA CAT TTC TTT ATG CTA TTT GTA CCC GTC ACC GGT TTA
Ala Phe Ser Asn Lys Arg Trp Leu His Phe Phe Met Leu Phe Val Pro Val Thr Gly Leu

                                870                                900
TGG ATG AGT GCT ATT GGC GTA GTT GGC TTG GCT CTG AAC TTA CGT GCC TAT GAC TTT GTT
Trp Met Ser Ala Ile Gly Val Val Gly Leu Ala Leu Asn Leu Arg Ala Tyr Asp Phe Val

                                930                                960
TCC CAG GAA ATC CGT GCA GCG GAA GAT CCT GAA TTC GAG ACT TTC TAC ACC AAA AAT ATT
Ser Gln Glu Ile Arg Ala Ala Glu Asp Pro Glu Phe Glu Thr Phe Tyr Thr Lys Asn Ile

                                990                                1020
CTT TTA AAC GAG GGT ATT CGT GCG TGG ATG GCA GCT CAG GAT CAG CCT CAT GAA AAT CTT
Leu Leu Asn Glu Gly Ile Arg Ala Trp Met Ala Ala Gln Asp Gln Pro His Glu Asn Leu
Met Lys Ile Leu

                                1050
ATA TTC CCT GAG GAG GTT CTA CCA CGT GGA AAC GCT CTT [TAA]T GGA ACT TTC GTT TTA GCT
Ile Phe Pro Glu Glu Val Leu Pro Arg Gly Asn Ala Leu End
Tyr Ser Leu Arg Arg Phe Tyr His Val Glu Thr Leu Phe Asn Gly Thr Phe Val Leu Ala

                                1100                                1130
GGT CGT GAC CAA GAA ACC ACC GGC TTT GCT TGG TGG GCT GGG AAT GCC AGA CTT ATC AAT
Gly Arg Asp Gln Glu Thr Thr Gly Phe Ala Trp Trp Ala Gly Asn Ala Arg Leu Ile Asn

                                1160                                1190
TTG TCC GGT AAA CTA CTT GGA GCT CAC GTA GCC CAT GCC GGA TTA ATC GTA TTC TGG GCC
Leu Ser Gly Lys Leu Leu Gly Ala His Val Ala His Ala Gly Leu Ile Val Phe Trp Ala

```

220
 GGA GCA ATG AAC CTA TTT GAA GTG GCC CAT 1TC GTA CCA GAA AAG CCC ATG TAT GAA CAA
 Gly Ala Met Asn Leu Phe Glu Val Ala His Phe Val Pro Glu Lys Pro Met Tyr Glu Gln

280
 GGG TTG ATT TTA CTT CCA CAC TTA GCT ACT CTA GGT TGG GGA GTA GGG CCA GGG GGA GAA
 Gly Leu Ile Leu Leu Pro His Leu Ala Thr Leu Gly Trp Gly Val Gly Pro Gly Gly Glu

340
 GTT CTA GAT ACT TTT CCA TAC TTT GTA TCT GGC GTA CTT CAC CTA ATT TCC TCC GCA GTC
 Val Leu Asp Thr Phe Pro Tyr Phe Val Ser Gly Val Leu His Leu Ile Ser Ser Ala Val

400
 TTA GGC TTC GGT GGC ATT TAT CAC GCG CTT CTG GGA CCC GAG ACT CTT GAA GAA TCT TTT
 Leu Gly Phe Gly Gly Ile Tyr His Ala Leu Leu Gly Pro Glu Thr Leu Glu Glu Ser Phe

460
 CCA TTC TTT GGT TAT GTC TGG AAA GAT AGA AAT AAA ATG ACT ACA ATT TTG GGT ATT CAC
 Pro Phe Phe Gly Tyr Val Trp Lys Asp Arg Asn Lys Met Thr Thr Ile Leu Gly Ile His

520
 TTA ATT TTG TTA GGT CTA GGT GCT TTT CTT CTA GTA CTC AAG GCT CTT TAT TTT GGC GGT
 Leu Ile Leu Leu Gly Leu Gly Ala Phe Leu Leu Val Leu Lys Ala Leu Tyr Phe Gly Gly

580
 GTA TAT GAT ACC TGG GCC CCT GGG GGG GGA GAT GTA AGA AAA ATT ACC AAT TTG ACC CTT
 Val Tyr Asp Thr Trp Ala Pro Gly Gly Gly Asp Val Arg Lys Ile Thr Asn Leu Thr Leu

640
 AGT CCC AGT GTT ATA TTT GGT TAT TTA CTA AAA TCT CCT TTT GGG GGA GAA GGG TGG ATT
 Ser Pro Ser Val Ile Phe Gly Tyr Leu Leu Lys Ser Pro Phe Gly Gly Glu Gly Trp Ile

700
 GTT AGT GTA GAT GAT TTA GAA GAT ATA ATT GGT GGA CAT GTA TGG TTG GGT TTT ATT TGT
 Val Ser Val Asp Asp Leu Glu Asp Ile Ile Gly Gly His Val Trp Leu Gly Phe Ile Cys

760
 GTA TTT GGG GGA ATT TGG CAT ATT TTA ACC AAA CCC TTC GCA TGG GCT CGC CGT GCA TTT
 Val Phe Gly Gly Ile Trp His Ile Leu Thr Lys Pro Phe Ala Trp Ala Arg Arg Ala Phe

820
 GTA TGG TCT GGA GAA GCT TAC TTG TCT TAT AGT TTA GCT GCT TTA TCT GTC TTT GGT TTT
 Val Trp Ser Gly Glu Ala Tyr Leu Ser Tyr Ser Leu Ala Ala Leu Ser Val Phe Gly Phe

880
 ATC GCT TGT TGT TTT GTA TGG TTC AAT AAT ACA GCT TAT CCG AGT GAG TTT TAT GGA CCC
 Ile Ala Cys Cys Phe Val Trp Phe Asn Asn Thr Ala Tyr Pro Ser Glu Phe Tyr Gly Pro

940
 ACC GGG CCA GAA GCT TCT CAA GCT CAA GCA TTT ACT TTT CTA GTT AGA GAC CAG CGT CTT
 Thr Gly Pro Glu Ala Ser Gln Ala Gln Ala Phe Thr Phe Leu Val Arg Asp Gln Arg Leu

1000
 GGA GCT AAT GTG GGA TCC GCT CAA GGA CCC ACA GGT TTA GGG AAA TAT CTA ATC CGT TCC
 Gly Ala Asn Val Gly Ser Ala Gln Gly Pro Thr Gly Leu Gly Lys Tyr Leu Ile Arg Ser

1060
 CCA ACT GGG GAG GTT ATC TTT GGA GGG GAA ACT ATG CGT TTT TGG GAC CTT CGT GCT CCA
 Pro Thr Gly Glu Val Ile Phe Gly Gly Glu Thr Met Arg Phe Trp Asp Leu Arg Ala Pro

1120
 TGG TTA GAA CCT CTA AGG GGC CCC AAC GGT TTG GAC TTG AGT AGG TTG AAA AAA GAC ATA
 Trp Leu Glu Pro Leu Arg Gly Pro Asn Gly Leu Asp Leu Ser Arg Leu Lys Lys Asp Ile

1180
 CAA CCT TGG CAA GAA CGG CGC TCA GCA GAA TAT ATG ACC CAC GCT CCT TTA GGC TCT TTA
 Gln Pro Trp Gln Glu Arg Arg Ser Ala Glu Tyr Met Thr His Ala Pro Leu Gly Ser Leu

1240
 AAT TCC GTG GGT GGC GTA GCG ACC GAG ATC AAT GCA GTT AAT TAT GTC TCT CCT AGA AGT
 Asn Ser Val Gly Gly Val Ala Thr Glu Ile Asn Ala Val Asn Tyr Val Ser Pro Arg Ser

1300
 TGG TTA TCA ACT TCT CAT TTT GTT CTA GGA TTC TTC TTT TTT GTG GGC CAT TTA TGG CAT
 Trp Leu Ser Thr Ser His Phe Val Leu Gly Phe Phe Phe Phe Val Gly His Leu Trp His

1360
 GCA GGA AGA GCC CGA GCT GCT GCA GCA GGT TTT GAA AAG GGA ATC GAT CGT GAT TTG GAA
 Ala Gly Arg Ala Arg Ala Ala Ala Ala Gly Phe Glu Lys Gly Ile Asp Arg Asp Leu Glu

1420
 CCT GTT CTT TAC ATG AAC CCT CTT AAC TAA
 Pro Val Leu Tyr Met Asn Pro Leu Asn End

- ◀ Figure 3. DNA sequence and deduced amino acid sequence of the barley *psbD* and *psbC* genes. Only the mRNA like strands are given. Those ATG codons, which are discussed in the text as putative translation start sites as well as translational stop codons are framed. A putative “-10” promoter sequence “TATAAT” is underlined as is the Shine-Dalgarno sequence GGAG. Elements which are capable to form stable secondary stem/loop structures are marked by arrows. In the overlapping region of *psbD* and *psbC* only the nucleic acid sequence of *psbD* is shown in the right frame. The upper line of the deduced amino acid residues refers to the *psbD* encoded protein, the lower line to the *psbC* encoded protein. The numbering of the *psbD* gene starts with the first ATG in the reading frame and ends with the stop codon TAA (position 1062). The numbering of *psbC* starts with the first ATG in the *psbC* related reading frame in the overlap region and is continued to the first TAA codon (position 1422).
-

ment. Since the *psbD* and the *psbA* genes exhibit homologies of up to 53% at the amino acid sequence level (3.3) it was assumed that hybridization of the *psbD* probe to the larger *EcoRI* fragment was a cross-reaction with the *psbA* gene. This assumption is supported by the fact that the hybridization signal to the 2.24 kbp fragment was weaker than to the 1.2 kbp *EcoRI* fragment.

The fragments which hybridized to the *psbA* and *psbD* gene probes could be placed on the barley cpDNA map by using the following data: 1) 80% of the *HindIII* map of barley cpDNA genome has been established by POULSEN (36), leaving two gaps of 11 kbp and 3 kbp in the map. 2) As evident from Figure 1, the *psbA* and the *psbD* genes are both located on the 23 kbp *Sall* fragment.

3) Determination of the partial nucleic acid sequence of the barley *psbA* gene (Figure 5) and its alignment with the spinach *psbA* nucleic acid sequence (52) revealed that the 1.6 kbp *HindIII* fragment contained the 3' half of the *psbA* gene with an *EcoRI* site downstream of the stop codon.

4) POULSEN (36) described a 6 kbp *HindIII* fragment which has not yet been assigned coordinates on the cpDNA map of barley.

5) BERENDS et al. (3) report that the *psbD* gene in barley is followed by the *psbC* gene, which extends into a 4.3 kb *HindIII* fragment previously cloned, mapped (36) and sequenced by OLIVER and POULSEN (33).

It is thus concluded that the 1.6 kb *HindIII* fragment, containing the 3' end of the *psbA* gene, is framed on one side by the previously cloned 5.8 kb *HindIII* inverted repeat fragment with the designation pHvC18 (36), and on the other side by the so far not cloned *HindIII*

fragment of 6 kbp. This places *psbA* in the same region of the chloroplast genome as has been reported for other species. As the direction of transcription is towards the inverted repeat (Figure 5) the 5' end of the *psbA* gene is to be sought in the 6 kbp *HindIII* fragment. The 3.34 kbp *HindIII* fragment, hybridizing to the *psbD* probe and thus containing *psbD* and part of the *psbC* gene, joins the other end of the 6 kbp *HindIII* fragment. The organization of the *HindIII* fragments and the localization of the *psbA*, *psbD* and *psbC* genes on the barley cpDNA genome is shown in Figure 2A, B.

3.2. Nucleic acid sequence determination of barley *psbD* and *psbC*; Comparison of the deduced amino acid sequences with those from other plants

The restriction site map and the sequencing strategy for 2282 bp of the 3.4 kbp *BamHI* fragment is given in Figure 2C. One open reading frame for 353 amino acids was identified and is terminated by a TAA codon (Figure 3). Comparison with published sequences showed that this open reading frame contains the entire *psbD* gene (1062 bp). A second open reading frame started 50 bp upstream from the *psbD* TAA stop codon. Comparison with the literature revealed this to be the 5' end of *psbC* which is transcribed in the same direction as *psbD* but read in a different frame. The 985 bp part of the *psbC* gene, which is located on the 3.4 kb *BamHI* fragment was sequenced and joined in Figure 3 with the sequence of the 3' part of the gene determined in (33). The sequencing strategy is shown in Figure 2C. The sequence of both strands of DNA has been obtained in all but one case, in which sequencing was performed by the

CAAAGGTTT ATTTC TA TAAGT AT A AGATA AG AAAAACA⁻²⁰⁰AAA TCGAATCAAATTCATGGATTTCACACG⁻¹⁷⁰ACC
 GAGCTCTTTCAAAGAITA ATAGT TGAGTCTAATA AGTTAAATAAAGGAATAATTGAATTTATAAATTTACTTTAACC
 TT AGT CAG A AGGAA TAA TCCAATGGAGTTCATGGATTTAC CTAGG
 ATTAGTAAA TTTTGAAGTTAATAGATTTTCTTTTATACAATATAAAAA AACTATTAGA TAATGAATTTAAAATAAAA

 TCGGTGTGACTCCATA GATAAA AATGGAAAAATTTCT⁻¹³⁰CTC TTCGAGACC ATTGAAAAAGGGCATTGA⁻¹⁰⁰ACGAGAAAGAA
 TAAGTT AATTAT GA ACC ACTAAAGAATTTAGTTT TTCGAAACCTATAGTTAAGGGTCAGTATACGAGAAA TC
 TCAGT TTATGGGCTA AT C AATAAGCATTITTTATC TTCGAAACCC ATTGGAAAGGGCAGTG CAAG AGAA
 AAAGACATAAAAAATTATTACAAAGAAAAGAAAAGTT TACC TACTTTTA ATTTTAATCG ATTTTA GAAATTATA
 AAA AATAAT TTTTCTATTTCGAAACCCAGTCGAAAAC AAGGGAGAGTCTACGAGAAATCATATCAT

 ATCGTCCACAGAT ATAA⁻⁶⁰AACTATCATATGCCTT GGAAAGTGATATGAGGTGCTCGGAAAT⁻³⁰TGG TTGAAGTAATTGAATA
 ATCCATAAATGAA ATGATCGAATTCTCGGACTC CAAAAAATGCTATGGGGTGCTCGGAAATGG TTGAAGTAGTTGAATA
 ATCATACAAAAATGATCGAATCTTCGGACGCC CCGAAAAAGATATGAGGTGCTCGGAAATGG TCGAAGTAGTTGAATA
 TTAATAAAAAAAA TTAAATTTGAAGGGATCAT AAAAAAATTTTTTTAGGTACTTAAAAATGG TTGAAGTAACCTAATA
 AT ATAAAATGAA AAAAGCCTCGAAGGTTCCATCCCTGAAAATGCTAGGAGGTGTTTCGAAATGGTTTGAAGTAGTTGAATA

GGAGGATCACT ⁻¹	Barley	psbD 5' end
GGAGGATCACT	Spinach	psbD 5' end
GGAGGATCACT	Nicotiana	psbD 5' end
GGAGAATCATT	Liverwort	psbD 5' end
GGAGGATTACT	Pea	psbD 5' end

Figure 4. Comparison of DNA sequences of untranslated psbD 5' upstream regions of the psbD region from barley, spinach, tobacco (Nicotiana), liverwort and pea. Only the 5'-3' direction in regard to the psbD coding region is shown. Presumptive "-35" and "-10" promoter consensus sequences are underlined, a Shine-Dalgarno Sequence and the translational stop codon are framed. The numbering is related to the first ATG (position 1) in the reading frame of the psbD gene.

dideoxy chain termination procedure as well as by sequencing according to MAXAM and GILBERT.

In addition to the two protein coding regions, 238 bp of the untranslated 5' upstream region of *psbD* were sequenced. Figure 4 shows an alignment of this sequence from barley with the 5' leader sequences of *psbD* from spinach (1), tobacco (42), liverwort (47) and pea (38). Starting from about position -36 bp the nucleic acid sequences exhibit a striking similarity towards position -1. This includes a putative ribosome-binding site (GGAGGA), which consistently is separated by 5 bp from the first ATG codon in the reading frame. Also conserved is a translational stop codon TAA or TAG in position -21. The barley sequence contains elements which resemble the prokaryotic "-10" and "-35" promoter consensus elements. The former is found in position -69 and repeated in position -77, while the latter is seen at position -104 and repeated in position -116. The significance of these elements is, however, weakened by the fact that they are not conserved in position or detailed sequence in the other species analyzed.

The deduced amino acid sequence for *psbD* comprises 353 amino acids (Figure 3), which calculates to a molecular weight of 39,562 Da. The *psbC* gene codes for 473 amino acids giving a molecular weight of 52,125 Da. In the nucleotide sequences of both genes secondary stem-and-loop structures can be predicted. The most interesting one was located in the 3' end of the *psbD* gene. It exhibits a ΔG of folding of -30 kcal, if calculated according to TINOCO et al.

(44). The possible relevance of this structure for regulation of transcription will be discussed later.

No sequences resembling the prokaryotic promoter consensus sequence are found preceding the first ATG of the reading frame in the *psbC* gene. However, at position 11 and 188 bp upstream from the second ATG was the sequence TATATT located, which resembles the prokaryotic TATA box. In addition, a putative ribosome binding site (GGAG) could be identified 6 bp upstream from the second ATG in the reading frame.

On the level of the nucleic acid sequence the barley *psbD* gene exhibits 90% identity to the *psbD* genes of spinach, tobacco and pea and 83% to the one of liverwort. On the amino acid level the similarity is even more pronounced reaching an identity of 94-95%. The amino acid substitutions in the D-2 protein (coded for by *psbD*) among these plants are set out in Table I. It is remarkable that between 55% (tobacco) and 70% (spinach) of the differences in the amino acid sequence occur in only 5% of the protein that is in the 18 NH₂-terminal amino acids located between the first and the second methionine.

Comparison of the barley *psbC* gene and its product, the Chl_a-P3 apoprotein, with those in other plant species exhibits 85% identity on the nucleic acid level and 94 to 96% on the amino acid level. Amino acid substitutions among the Chl_a-P3 apoproteins of barley, liverwort, spinach, tobacco and maize are shown in Table II.

Table I. Amino acid substitutions among *psbD* encoded D-2 proteins from barley, spinach, pea, liverwort and tobacco (*Nicotiana*). Underlined amino acid residues mark a difference in comparison to the barley sequence. A substitution located in a putative membrane spanning helix is indicated by the number of the helix in roman letters.

POSITION																		
	5	7	8	9	11	12	13	14	17	18	46	102	113	122	146	241	285	341
Barley	Leu	Arg	Val	Pro	Glu	Glu	Asn	Asp	Asp	Thr	Leu	Phe	Thr	Ala	Ser	Ala	Ile	Ile
Spinach	<u>Val</u>	<u>Lys</u>	<u>Phe</u>	<u>Thr</u>	<u>Asp</u>	Glu	<u>Lys</u>	Asp	Asp	<u>Ser</u>	Leu	Phe	<u>Ala</u>	Ala	<u>Ala</u>	Ala	<u>Leu</u>	Ile
Pea	Leu	<u>Lys</u>	<u>Phe</u>	<u>Thr</u>	<u>Asp</u>	<u>Gln</u>	Asn	Asp	<u>Tyr</u>	<u>Ser</u>	<u>Val</u>	<u>Leu</u>	Thr	<u>Gly</u>	<u>Ala</u>	Ala	<u>Leu</u>	Ile
Liverwort	<u>Ile</u>	<u>Lys</u>	<u>Ser</u>	<u>Ser</u>	Glu	<u>Pro</u>	<u>Lys</u>	<u>Gly</u>	Asp	<u>Ser</u>	Leu	Phe	Thr	<u>Gly</u>	<u>Ala</u>	<u>Ser</u>	Ile	<u>Val</u>
Nicotiana	Leu	<u>Lys</u>	<u>Phe</u>	<u>Thr</u>	<u>Asp</u>	Glu	Asn	Asp	Asp	<u>Ile</u>	<u>Val</u>	Phe	Thr	<u>Gly</u>	<u>Ala</u>	Ala	<u>Leu</u>	Ile
	between 1. and 2. ATG										I	II			III		V	

Table II. Amino acid substitutions among the psbC encoded 44 kD proteins (Chl_a-P3) from barley, spinach, tobacco (Nicotiana), liverwort and maize. Underlined amino acid residues mark a difference in comparison to the barley sequence.

	POSITION																	
	3	7	12	21	22	24	105	106	135	154	170	177	185	202	207	208	212	
Barley	Ile	Leu	His	Phe	Val	Ala	Val	Leu	Leu	Arg	Leu	Leu	Val	Leu	Ser	Val	Tyr	
Spinach	<u>Thr</u>	Leu	<u>Pro</u>	<u>Leu</u>	<u>Thr</u>	Ala	Val	<u>Ile</u>	Leu	Arg	<u>Ile</u>	<u>Phe</u>	Val	<u>Val</u>	Ser	<u>Ile</u>	<u>Cys</u>	
Nicotiana	<u>Thr</u>	Leu	His	<u>Leu</u>	<u>Ala</u>	Ala	Val	<u>Ile</u>	Leu	Arg	Leu	<u>Phe</u>	Val	Leu	Ser	<u>Ile</u>	Tyr	
Liverwort	Ile	<u>Gln</u>	<u>Pro</u>	<u>Leu</u>	<u>Ala</u>	<u>Gly</u>	<u>Ile</u>	<u>Val</u>	<u>Ile</u>	<u>Lys</u>	<u>Ala</u>	<u>Phe</u>	<u>Ile</u>	Leu	<u>Gly</u>	Val	Tyr	
Maize	Ile	Leu	His	Phe	Val	Ala	Val	Leu	Leu	Arg	Leu	Leu	Val	Leu	<u>Gly</u>	Val	Tyr	

	POSITION																	
	229	240	242	245	246	264	277	279	280	341	342	349	379	427	437	463	468	470
Barley	Asp	Leu	Val	Val	Phe	Phe	Ala	Leu	Ser	Leu	Ile	Val	Arg	Ser	Phe	Leu	Tyr	Asn
Spinach	Asp	<u>Ile</u>	Val	Ile	<u>Leu</u>	<u>Leu</u>	Ala	Leu	Ser	Leu	<u>Met</u>	Val	Arg	Ser	<u>Leu</u>	<u>Phe</u>	Ser	<u>Thr</u>
Nicotiana	Asp	Leu	Val	Ile	<u>Leu</u>	<u>Leu</u>	<u>Gly</u>	Leu	Ser	Leu	<u>Met</u>	Val	Arg	<u>Ala</u>	Phe	<u>Phe</u>	<u>Ser</u>	<u>Thr</u>
Liverwort	<u>Asn</u>	Leu	Val	<u>Ile</u>	Phe	<u>Leu</u>	<u>Gly</u>	<u>Ile</u>	<u>Ala</u>	<u>Ile</u>	<u>Met</u>	<u>Ile</u>	<u>Lys</u>	<u>Ala</u>	Phe	<u>Phe</u>	<u>Ser</u>	<u>Thr</u>
Maize	Asp	Leu	<u>Ser</u>	Val	<u>Leu</u>	Phe	<u>Gly</u>	<u>Ser</u>	Ser	Leu	Ile	Val	Arg	Ser	Phe	Leu	Tyr	Asn

3.3. Partial determination of the nucleic acid sequence of the barley psbA gene

When barley HindIII digested cpDNA was separated on an agarose gel, a 1.6 kbp HindIII fragment hybridized to the psbA probe. This fragment was recovered from the gel, cloned into the vector pUC 18 and partially sequenced. The sequence strategy is shown in Figure 2D, the localization of the psbA gene in Figure 2A, B. In Figure 5 is set out the partial nucleotide sequence and the amino acid sequence deduced from it. The amino acid sequence aligns with amino acid 170 to 278 of the spinach D-1 protein which has a sequence of 353 amino acids (52). Of particular interest is the presence of a lysine residue in position 238, since this amino acid is not present in D-1 of dicotyledonous plants. Specially marked are in Figure 5 those amino acid residues which have been suggested to be involved in binding of redox components (His₂₁₅ and His₂₇₂) (45, 46) and inhibitors of photosynthetic electron transport (Val₂₁₉, Ala₂₅₁, Phe₂₅₅, Ser₂₆₄, Arg₂₆₉, Leu₂₇₃) (10, 12, 17, 19).

In Figure 5 the deduced barley D-1 amino acid sequence has also been aligned with the barley D-2 protein residues 170 to 275. The overall identity between the two sequences is 38%. But

a comparison of the region between residues 192 and 219 of D-1 reveals 53% identity. This part of the protein should form a membrane spanning α -helix as suggested by drop-hobicity plots of the very similar D-1 protein of spinach (45). The importance of this helix in D-1 and D-2 will be discussed below.

4. DISCUSSION

Chloroplast DNAs of higher plants and green algae are circular molecules with a size of 120 – 160 kbp. They have been isolated and studied in a number of species. The barley cpDNA genome consists of about 130 kbp with two copies of a large inverted repeat of about 20.9 kbp, separated by a large and a small single copy region (33). This organization is typical for most plastid genomes. Hybridization experiments with heterologous and homologous DNA, presented in this paper, localize the psbD, psbC and psbA genes on the same 23 kbp SalI fragment. Thus, the psbD gene is separated by only 8 kbp from the psbA gene. This location agrees with one determined by BERENDS et al. (3), who mapped the psbD and the psbC genes together with several tRNAs downstream from psbC.

170																		
GAT	GGT	ATG	CCT	TTA	GGA	ATC	TCT	GGT	ACT	TTC	AAC	TTT	ATG	ATT	GTA	TTC		
Asp	Gly	Met	Pro	Leu	Gly	Ile	Ser	Gly	Thr	Phe	Asn	Phe	Met	Ile	Val	Phe		
Phe	Ala	Pro	Ser	Phe	Gly	Val	Ala	Ala	Ile	Phe	Arg	Phe	Ile	Leu	Phe	Phe		
CAG	GCA	GAG	CAC	AAC	ATC	CTT	ATG	CAT	CCA	TTC	CAC	ATG	TTA	GGT	GTA	GCT		
Gln	Ala	Glu	His	Asn	Ile	Leu	Met	His	Pro	Phe	His	Met	Leu	Gly	Val	Ala		
Gln	Gly	Phe	His	Asn	Trp	Thr	Leu	Asn	Pro	Phe	His	Met	Met	Gly	Val	Ala		

Figure 5. Partial DNA sequence of the barley psbA gene, the deduced amino acid sequence and alignment with the D-2 amino acid sequence. Only the mRNA like strand is given. Residues, that are important for binding of redox components or inhibitor molecules, (as discussed in the text), are boxed. Predicted α -helices are framed.

The same gene arrangement has also been described for wheat (8, 37), while for dicotyledonous plants like spinach (34) and tobacco (42) a 20 kbp inversion in the large single copy region is characteristic and places the psbD gene about 30 kbp apart from the psbA gene (1). It has been suggested by QUIGLEY and WEIL (37) and COURTICE et al. (8) that at least three inversions must have occurred during the evolution of the wheat chloroplast genome from a spinach like ancestor to explain the observed rearrangement of the genomes.

In spite of these rearrangements the organization of the psbD/psbC gene cluster has remained conserved in all plants investigated. The barley

psbD gene comprises 1062 bp and the psbC gene 1422 bp with an overlap of 53 bp and a reading frameshift of 1 nucleotide. Examination of the untranslated 5' nucleotide sequence in preceding psbD showed in barley prokaryotic "-35" and "-10" consensus sequences. The "-35" element consisted of the hexanucleotides TTGAAA and TTGAAC with a spacer of 7 nucleotides. This type of duplication is also characteristic for the "-10" element located 22 bp downstream and composed of the sequences TATAAA and TATCAT spaced 2 nucleotides apart. Suggested promoters in the form of the "-35" box with consensus TTG^{CC}AA^T (4, 14, 22) and in the form of the "-10" with consensus TAT^CAA^T

(4, 14, 22) are found in the 5' flanking sequences of many chloroplast genes and are separated by 14 to 25 nucleotides. The significance of these sequences for *psbA*, *rbcl* and *atpB* was tested by GRUISSEM and ZURAWSKI (13) in a chloroplast in-vitro transcription system on mutant templates with alterations in these 5' regions. They found that the spacing between the "-35" and the "-10" boxes is essential (13). This would imply for barley that only the second "-35" and the first "-10" like sequence could be a promoter structure. However, none of the six *psbD*/*psbC* transcripts found in dark grown barley plants map with their initiation points near the mentioned "-35" and "-10" sequence element (3). More recently performed studies on the transcription of these genes reveal that two transcripts of 4.0 and 3.2 kbp accumulate preferentially upon illumination. Although they are initiated closest to the *psbD* start codon their 5' ends still map about 600 nucleotides upstream of the *psbD* gene (P.E. GAMBLE, T. BERENDS SEXTON and J. MULLET, pers. commun.).

A presumptive ribosome binding site preceded the *psbD* coding region and is also conserved in all five compared plant species. It exhibits the sequence GGAGGA and is separated by 5 bp from the first ATG in the reading frame. This sequence element is the ribosome binding site in bacterial messenger RNA and it has been found that a complementary sequence is located at the 3' end of prokaryotic 16S rRNA. Such "Shine-Dalgarno" sequences have also been found for the *rbcl* genes from maize, spinach and pea (51). Furthermore, the 3' ends of the 16S rRNA of maize and tobacco chloroplasts are very similar to that of *E. coli* (51) suggesting that the requirements for efficient initiation of translation in chloroplasts are similar to those in bacteria.

Surprising is the localization and strong conservation of a translational termination codon TAA or TAG, respectively, 7 bp upstream from the ribosome binding site and in frame with the first ATG codon of the *psbD* gene. As will be shown in a subsequent paper, it was not possible to translate *psbD* derived transcripts in an *E. coli* in vitro translation system. This may imply that in the bacterial system binding of ribosomes to the initiation signal of *psbD* mRNA is accompa-

nied by the recognition of the signal for termination, thus inhibiting the translation of the D-2 transcript. This in turn may explain why it was no problem to clone the *psbD* gene in *E. coli* whereas cloning of the intact highly homologous *psbA* gene is almost always lethal to *E. coli* (R.B. HALLIK and H.-Y. BOHNERT, pers. commun.). Extensive expression of the *psbA* gene due to a strong promoter may lead to insertion of this very hydrophobic protein D-1 into the bacterial membranes with lethality of the host as a consequence. The absence of difficulties in cloning *psbD* then is due to the lack of translation of the similar D-2 protein in *E. coli*. The translation machinery of chloroplasts apparently has a feature which makes it different from the bacterial system by allowing *psbD* expression in the presence of the termination signal.

The 5' flanking region of the *psbC* gene exhibits an interesting nucleic acid sequence which might be involved in the regulation of gene expression. The first ATG in the coding region of the *psbC* gene is preceded by a sequence, which can be folded into a stable stem and loop structure ($\Delta G = -29.8$ kcal). This structure (Figure 6) includes the EcoRI restriction site in the 3' end of the *psbD* gene. It is this region to which the 5' ends of two transcripts of 2.6 kbp and 1.7 kbp length map (3). These two transcripts are possibly derived by processing from the 5.7 kbp primary transcript which maps with its 5' end about 1.2 kbp upstream from the *psbD* gene (3). The stem/loop structure could serve as a signal for correct processing in the formation of the shorter mRNA molecules as suggested for other chloroplast genes by GRUISSEM et al. (14).

Both the *psbD* and the *psbC* genes contain two ATG codons in the 5' part of the coding region. In *psbD* they are separated by 51 bp (coding for 17 amino acids) and in *psbC* by 195 bp (coding for 65 amino acids). In *psbC* a presumptive "-10" and a putative Shine-Dalgarno sequence were only found preceding the second ATG. The question thus arose which of the two methionine codons are initiation sites for translation. A comparison of *psbD* deduced amino acid sequences of different plant species (Table I) revealed that 55 to 70% of the observed amino

Figure 6. Predicted secondary structure in the 3' region of barley *psbD*.

acid substitutions occur between the first and the second methionine residue, which led to the speculation that translation is initiated at the second ATG codon. It was recently shown, however, that the first amino acid of the mature D-2 polypeptide (the *psbD* gene product) is an N-acetyl-O-phosphothreonine (H. MICHEL and J. BENNETT, pers. commun.) proving that the translation initiates at the first methionine prior to the posttranslational modifications of the NH_2 -terminus. Therefore, the highly polymor-

phic structure of the amino-terminal part of the protein indicates it to be of minor importance for the functional role of the D-2 protein. This is in agreement with the predicted folding of D-2 in the thylakoid membrane, which locates the hydrophilic NH_2 end on the stroma side whereas the functional components are located in hydrophobic regions of the polypeptide (45, 46). A similar tertiary structure was predicted for the D-1 polypeptide (45, 46).

The PSII polypeptides D-1 and D-2 have structural features in common with the reaction center polypeptides of purple bacteria and it has been proposed that D-1 and D-2 might also bind the reaction center chlorophyll molecule (28, 45, 46). As described in this paper, attempts were undertaken to clone the *psbA* gene encoding the D-1 polypeptide from barley. So far, hybridization with *psbA* probes identified a 1.6 kb HindIII fragment which could be cloned from the barley cpDNA. It contains the 3' half of the *psbA* gene (52) coding for the important transmembrane helices IV and V (amino acid residues 169 to 278). In this region (Figure 5) the identity of the nucleotide sequence with spinach is 95%, and only a single amino acid substitution from arginine to lysine is observed in position 238. In this position the wheat D-1 protein also contains a lysine (J. GRAY, pers. commun.) while in all other higher plants investigated an arginine has been found. Notably, the positions of the histidine ligands for the non-heme-iron in positions 215 and 272 and the histidine ligand for the special pair reaction center chlorophyll in position 198 are conserved. Identity extends also to amino acid neighbours of these histidine ligands in the reaction center subunits L and M from purple bacteria (29). Thus the residues Pro_{195} , Phe_{197} and Met_{199} bracketing the ligand to the special pair are conserved as are Ala_{213} , Met_{214} and Gly_{216} , while no identity is seen for the neighbours to His_{272} the other ligand to the non-heme-iron atom.

The corresponding D-2 protein domain extends from residues 170 to 275 (Figure 5). As in other plants and in the green alga *Chlamydomonas* (1, 39) about 50% identical residues can be observed in the transmembrane helices IV and V. Thus, the histidine 198 ligand for the special pair chlorophylls is conserved in D-1 and

D-2 together with the neighbouring amino acids. Ligands for the non-heme-iron are found at the same places in D-2 as in D-1 (His₂₁₅, His₂₇₂). The plastoquinone Q_A binds in *Rhodospseudomonas* to two helices in the subunit M with Trp₂₅₀ forming an important binding site (29). This tryptophan is conserved in position 255 of the D-2 polypeptide (Figure 5). In the purple bacteria the homologous residue in subunit L is Phe₂₁₆ which forms part of the binding site of plastoquinone Q_B (bound to two helices in the L subunit). Interestingly enough in D-1 this very position is occupied by Phe₂₅₅ (Figure 5). Besides Phe₂₁₆ the residues Ser₂₂₃ and Tyr₂₂₂ participate in forming the binding site of Q_B in the L subunit. In D-1, Ser₂₂₃ is found as Ser₂₆₄ while this residue is missing in D-2.

Analysis of herbicide tolerant mutants (10, 12, 17, 19) and binding studies with the photoaffinity label azido-atrazin revealed the amino acids Val₂₁₉, Ala₂₅₁, Phe₂₅₅, Ser₂₆₄, Arg₂₆₉ and Leu₂₇₅ (Figure 5) to be important for the action of inhibitors of photosystem II activity. These amino acid residues are found to be conserved in the barley D-1 protein and three of them in the D-2 protein.

ACKNOWLEDGEMENTS

I would like to thank Professor DITER VON WETTSTEIN for encouragement in this project and help with the manuscript, Dr. SANDY BERRY-LOWE for instructing me in molecular techniques, Dr. PER STRØMAN for the synthesis of oligonucleotides, Dr. GUNILLA HØYER-HANSEN for discussions and the Commission of the European Communities for a Fellowship. I am indebted to HANNE THEM NIELSEN, NINA RASMUSSEN, ANN-SOFI STEINHOLTZ and LISE TRILLOT for technical help in the preparation of the manuscript.

REFERENCES

- ALT, J., J. MORRIS, P. WESTHOFF & R.G. HERRMANN: Nucleotide sequence of the clustered genes for the 44 kd chlorophyll *a* apoprotein and the "32 kd"-like protein of the photosystem II reaction center in the spinach plastid chromosome. *Curr. Genet.* 8, 597-606 (1984)
- BENNOUN, P., M. SPIERER-HERZ, J. ERICKSON, J. GIRARD-BASCOU, Y. PIERRE, M. DELOSME & J.-D. ROCHAIX: Characterization of photosystem II mutants of *Chlamydomonas reinhardtii* lacking the *psbA* gene. *Plant. Molec. Biol.* 6, 151-160 (1986)
- BERENDS, T., P.E. GAMBLE & J.E. MULLET: Characterization of the barley chloroplast transcription units containing *psbA-psbB* and *psbD-psbC*. *Nucleic Acids Res.* 15, 5217-5240 (1987)
- BRIAT, J.F., A.M. LESCURE & R. MACHE: Transcription of the chloroplast DNA: a review. *Biochimie* 68, 981-990 (1986)
- CAMM, E.L. & B.R. GREEN: Relationship between the two minor chlorophyll *a* protein complexes and the photosystem II reaction centre. *Biochim. Biophys. Acta* 724, 291-293 (1983)
- CHEN, E.Y. & P.H. SEEBURG: Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* 4, 165-170 (1985)
- CLEWELL, D.B. & D.R. HELINSKI: Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. *Proc. Nat. Acad. Sci. USA* 62, 1159-1166 (1969)
- COURTICE, G.R.M., C.M. BOWMAN, T.A. DYER & J.C. GRAY: Localization of genes for components of photosystem II in chloroplast DNA from pea and wheat. *Curr. Genet.* 10, 329-333 (1985)
- DEISENHOFER, J., O. EPP, K. MIKI, R. HUBER & H. MICHEL: Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Å resolution. *Nature* 318, 618-623 (1985)
- ERICKSON, J.M., M. RAHIRE, J.-D. ROCHAIX & L. METS: Herbicide resistance and cross-resistance: changes at three distinct sites in the herbicide-binding protein. *Science* 228, 204-207 (1985)
- ERICKSON, J.M., M. RAHIRE, P. MALNOE, J. GIRARD-BASCOU, Y. PIERRE, P. BENNOUN & J.D. ROCHAIX: Lack of the D2 protein in a *Chlamydomonas reinhardtii* *psbD* mutant affects photosystem II stability and D1 expression. *EMBO J.* 5, 1745-1754 (1986)
- GOLDEN, S.S. & R. HASELKORN: Mutation to herbicide resistance maps within the *psbA* gene of *Anacystis nidulans* R2. *Science* 229, 1104-1107 (1985)
- GRUISSEM, W. & G. ZURAWSKI: Analysis of promoter regions for the spinach chloroplast *rbcL*, *atpB* and *psbA* genes. *EMBO J.* 4, 3375-3383 (1985)
- GRUISSEM, W., X.-W. DENG, H. JONES, D. STERN, J. TONKYN & G. ZURAWSKI: Transcriptional and post-transcriptional regulation of chloroplast gene expression. In: *Plant Molecular Biology*. D. von

- Wettstein & N.-H. Chua eds. NATO-ASI, Plenum Press, New York, Series A, Vol. 140, 135-148 (1987)
15. HATTORI, M. & Y. SAKAKI: Dideoxy sequencing method using denatured plasmid templates. *Analytical Biochem.* 152, 232-238 (1986)
16. HERSHFELD, V., H.W. BOYER, C. YANOFSKY, M.A. LOVETT & D.R. HELINSKI: Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA. *Proc. Nat. Acad. Sci. USA* 71, 3455-3459 (1974)
17. HIRSCHBERG, J. & L. MCINTOSH: Molecular basis of herbicide resistance in *Amaranthus hybridus*. *Science* 222, 1346-1349 (1983)
18. HOLSCHUH, K., W. BOTTOMLEY & P.R. WHITFIELD: Structure of the spinach chloroplast genes for the D2 and 44 kd reaction-centre proteins of photosystem II and for tRNA^{Ser} (UGA). *Nucleic Acids Res.* 12, 8819-8834 (1984)
19. JOHANNINGMEIER, U., U. BODNER & G.F. WILDNER: A new mutation in the gene coding for the herbicide-binding protein in *Chlamydomonas*. *FEBS Lett.* 211, 221-224 (1987)
20. JOHNSON, D.A., J.W. GAUTSCH, J.R. SPORTSMAN & J.H. ELDER: Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. *Gene Anal. Techn.* 1, 3-8 (1984)
21. KREBBERS, E.T.: Ph.D. Thesis. Harvard University (1983)
22. KUNG, S.D. & C.M. LIN: Chloroplast promoters from higher plants. *Nucleic Acids Res.* 13, 7543-7549 (1985)
23. KYLE, D.J.: The 32 000 dalton Q_B protein of photosystem II. *Photochem. Photobiol.* 41, 107-116 (1985)
24. MANIATIS, T., E.F. FRITSCH & J. SAMBROOK: Molecular cloning. Cold Spring Harbor Laboratory Press, New York (1982)
25. MARDER, J.B., D.J. CHAPMAN, A. TELFER, P.J. NIXON & J. BARBER: Identification of psbA and psbD gene products, D1 and D2, as reaction centre proteins of photosystem 2. *Plant Molec. Biol.* 9, 325-333 (1987)
26. MAXAM, A.M. & W. GILBERT: Sequencing end-labeled DNA with base-specific chemical cleavages. In: *Methods in Enzymology* 65, L. Grossman & K. Moldave, eds., Academic Press New York, pp. 499-560 (1980)
27. MESSING, J.: A multi-purpose cloning system based on the single-stranded DNA bacteriophage M13. *Recombinant DNA technical bulletin, NIH publications no. 79-99*, 2, No. 2, pp. 43-48 (1979)
28. MICHEL, H.: Three-dimensional crystals of a membrane protein complex. *J. Mol. Biol.* 158, 567-572 (1982)
29. MICHEL, H. & J. DEISENHOFER: Photosynthetic reaction center from the purple bacterium *Rhodospirillum rubrum* and its relevance to photosystem II. In: *Plant Molecular Biology*. D. von Wettstein & N.-H. Chua eds. NATO-ASI, Plenum Press, New York, Series A, Vol. 140, 597-607 (1987)
30. MORRIS, J. & R.G. HERRMANN: Nucleotide sequence of the gene for the P₆₈₀ chlorophyll *a* apoprotein of the photosystem II reaction center from spinach. *Nucleic Acids Res.* 12, 2837-2850 (1984)
31. NAKATANI, H.Y., B. KE, E. DOLAN & C.J. ARNTZEN: Identity of the photosystem II reaction center polypeptide. *Biochim. Biophys. Acta* 765, 347-352 (1984)
32. NANBA, O. & K. SATOH: Isolation of a photosystem II reaction center consisting of D-1 and D-2 polypeptides and cytochrome b-559. *Proc. Natl. Acad. Sci. USA* 84, 109-112 (1987)
33. OLIVER, R.P. & C. POULSEN: Structure of a heavily transcribed region of barley chloroplast DNA. Transfer RNA genes for serine (UGA), glycine (GCC, UCC), formyl-methionine and threonine (GGU). *Carlsberg Res. Commun.* 49, 647-673 (1984)
34. PALMER, J.D. & W.F. THOMPSON: Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 29, 537-550 (1982)
35. PFISTER, K., K.E. STEINBACK, G. GARDNER & C.J. ARNTZEN: Photoaffinity labeling of an herbicide receptor protein in chloroplast membranes. *Proc. Natl. Acad. Sci. USA* 78, 981-985 (1981)
36. POULSEN, C.: The barley chloroplast genome: physical structure and transcriptional activity in vivo. *Carlsberg Res. Commun.* 48, 57-80 (1983)
37. QUIGLEY, F. & J.H. WEIL: Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. *Curr. Genet.* 9, 495-503 (1985)
38. RASMUSSEN, O.F., G. BOOKJANS, B.M. STUMMANN & K.W. HENNINGSEN: Localization and nucleotide sequence of the gene for the membrane polypeptide D2 from pea chloroplast DNA. *Plant Molec. Biol.* 3, 191-199 (1984)
39. ROCHAIX, J.-D., M. DRON, M. RAHIRE & P. MALNOE: Sequence homology between the 32 K dalton and the D2 chloroplast membrane polypeptides of *Chlamydomonas reinhardtii*. *Plant Molec. Biol.* 3, 363-370 (1984)
40. SANGER, F., S. NICKLEN & A.R. COULSON: DNA sequencing with chain-terminating inhibitors.

- Proc. Natl. Acad. Sci. USA 74, 5463-5467 (1977)
41. SHAH, D.M., R.C. HIGHTOWER & R.B. MEAGHER: Complete nucleotide sequence of a soybean actin gene. Proc. Natl. Acad. Sci. USA 79, 1022-1026 (1982)
 42. SHINOZAKI, K., M. OHME, M. TANAKA, T. WAKASUGI, N. HAYASHIDA, T. MATSUBAYASHI, N. ZAITA, J. CHUNWONGSE, J. OBOKATA, K. YAMAGUCHI-SHINOZAKI, C. OHTO, K. TORAZAWA, B.Y. MENG, M. SUGITA, H. DENO, T. KAMOGASHIRA, K. YAMADA, J. KUSUDA, F. TAKAIWA, A. KATO, N. TOHDOH, H. SHIMADA & M. SUGIURA: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. 5, 2043-2049 (1986)
 43. SPIELMANN, A. & E. STUTZ: Nucleotide sequence of soybean chloroplast DNA regions which contain the *psbA* and *trnH* genes and cover the ends of the large single copy region and one end of the inverted repeats. Nucleic Acids Res. 11, 7157-7167 (1983)
 44. TINOKO, I., P.N. BORER, B. DENGLER, M.D. LEVINE, O.C. UHLENBECK, D.M. CROTHERS & J. GRALLA: Improved estimation of secondary structure in ribonucleic acids. Nature New Biology 246, 40-41 (1973)
 45. TREBST, A.: The topology of the plastoquinone and herbicide binding peptides of photosystem II in the thylakoid membrane. Z. Naturforsch. 41c, 240-245 (1986)
 46. TREBST, A.: The three-dimensional structure of the herbicide binding niche on the reaction center polypeptides of photosystem II. Z. Naturforsch. 42c, 742-750 (1987)
 47. UMESONO, K., H. INOKUCHI, K. OHYAMA & H. OZEKI: Nucleotide sequence of *Marchantia polymorpha* chloroplast DNA: a region possibly encoding three tRNAs and three proteins including a homologue of *E. coli* ribosomal protein S14. Nucleic Acids Res. 12, 9551-9565 (1984)
 48. VERMAAS, W.F.J., J.G.K. WILLIAMS & C.J. ARNTZEN: Sequencing and modification of *psbB*, the gene encoding the CP-47 protein of photosystem II, in the cyanobacterium *Synechocystis* 6803. Plant Molec. Biol. 8, 317-326 (1987)
 49. VERMAAS, W.F.J., M. IKEUCHI & Y. INQUE: Protein composition of the photosystem II core complex in genetically engineered mutants of the cyanobacterium *Synechocystis* 6803. Photosynthesis Res. (in press)
 50. VIEIRA, J. & J. MESSING: The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259-268 (1982)
 51. WHITFIELD, P.R. & W. BOTTOMLEY: Organization and structure of chloroplast genes. Ann. Rev. Plant. Physiol. 34, 279-310 (1983)
 52. ZURAWSKI, G., H.J. BOHNERT, P.R. WHITFIELD & W. BOTTOMLEY: Nucleotide sequence of the gene for the M_r 32,000 thylakoid membrane protein from *Spinacia oleracea* and *Nicotiana debneyi* predicts a totally conserved primary translation product of M_r 38,950. Proc. Natl. Acad. Sci. USA 79, 7699-7703 (1982)

Accepted by E. LUND