Structure of the spinach chloroplast genes for the D2 and 44 kd reaction-centre proteins of photosystem II and for tRNA^{Ser} (UGA)

Karl Holschuh*, Warwick Bottomley and Paul R. Whitfeld⁺

Division of Plant Industry, CSIRO, GPO Box 1600, Canberra ACT 2601, Australia

Received 2 October 1984; Accepted 24 October 1984

ABSTRACT

We have determined the sequence of the spinach (*Spinacia oleracea*) chloroplast genes for the photosystem II proteins, D2 and the 44 kd reaction-centre, chlorophyll *a*-binding protein, and for tRNA^{Ser} (UGA). The 3' end of the D2 gene overlaps the first 50 bp of the 5' end of the gene for the 44 kd protein. Northern RNA hybridization analysis indicates the two genes are cotranscribed into a single 3.5 kb RNA. The predicted molecular weight of the 353-residue D2 protein is 39536 and that of the 473-residue 44 kd protein is 51816. Both proteins are hydrophobic containing at least five possible membrane-spanning domains. D2 shows significant homology to the 32 kd herbicide-binding protein (Zurawski *et al.*, (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7699-7703), and parts of the 44 kd protein show obvious similarities to parts of the 51 kd reaction-centre, chlorophyll *a*-binding protein of photosystem II (Morris and Herrmann (1984) *Nucleic Acids Res.* **12**, 2837-2850). The gene for tRNA^{Ser} (UGA) which is on the opposite strand to and transcribed towards the photosystem II genes is 72% homologous with the corresponding *Escherichia coli* tRNA^{Ser}.

INTRODUCTION

In higher plants, photosystem II is one of the supramolecular complexes of the chloroplast thylakoid membranes that participate in the trapping and conversion of light energy into chemical energy (ATP) and reducing power. It is composed of a number of proteins some of which are nuclear encoded and some chloroplast encoded. The genes for the chlorophyll *a/b*-binding proteins of the light-harvesting complex are located in the nucleus (1), whereas those for the 51 kd and 44 kd chlorophyll *a*-binding proteins of the reaction centre (2), and the 32 kd herbicide-binding protein (3,4) are located in the chloroplast. A protein referred to as D2 (designated D because of the diffuse nature of the band which the protein forms on acrylamide gel electrophoresis (5)), now generally considered to be a component of photosystem II (6), is also chloroplast encoded (7-9).

Our understanding of the photosystem II complex has been considerably advanced by the determination of the primary structure of a

number of the component proteins. The sequence of the chlorophyll a/b-binding proteins of the light-harvesting complex (10-12) and of the 32 kd herbicide-binding protein (13-16) from a number of plant species is known. We also know the sequence of the 51 kd chlorophyll a reaction-centre apoprotein from spinach (17) and the D2 protein from *Chlamydomonas reinhardtii* (9) and pea (8). In all these cases the amino acid sequence has been deduced from the nucleotide sequence of the corresponding cDNA or gene. In this paper we present the nucleotide and deduced amino acid sequence of another photosystem II gene, that for the 44 kd chlorophyll a reaction-centre apoprotein from spinach (*Spinacia oleracea*) chloroplasts. The gene for this protein overlaps by 50 bp the 3' end of the gene for the D2 protein, the sequence of which is also presented. The two genes are located on the spinach chloroplast genome between the previously sequenced gene for tRNA^{Ser} (18) and the gene for tRNA^{UGA}, the sequence of which is included in this paper. Comparison of the D2 protein with the 32 kd herbicide-binding protein (13) and the 44 kd protein with the 51 kd chlorophyll a apoprotein (17) reveals features of the amino acid sequences which may be important with respect to the binding of quinone and chlorophyll a respectively.

MATERIALS AND METHODS

Recombinant plasmids containing BamHI fragments 8b and 16 from spinach chloroplast DNA were constructed as described previously (19). EcoRI and EcoRI/BamHI subfragments from BamHI fragment 8b were subcloned into pBR325 and pBR322. Plasmid DNA was purified from bacterial lysates by ethidium bromide-CsCl density gradient centrifugation (20). All enzymatic reactions were carried out using commercial enzymes in accordance with the supplier's instructions. DNA fragments were end-labeled with [α -³²P]dNTPs by fill-in reaction using DNA polymerase (Klenow fragment) or with [γ -³²P]ATP by kinase reaction after dephosphorylation with alkaline phosphatase (21). They were then cut asymmetrically with an appropriate restriction enzyme, fractionated on 6.5% acrylamide-urea gels and electroeluted. The base-specific chemical cleavage reactions were performed according to Maxam and Gilbert (21).

Total spinach chloroplast RNA (22) was reacted with glyoxal (23), fractionated by electrophoresis on a 1.2% agarose gel, blotted onto nitrocellulose membranes and hybridized with nick-translated DNA probes (24).

RESULTS

The location of BamHI fragments 8b and 16 on the SalI restriction map of spinach chloroplast DNA is shown in Figure 1A. We have previously published (18,25) the sequence of that part of Bam 8b which contains the genes for 4 tRNAs (BamHI - HinfI fragment, right-hand end of Figure 1B). The sequence of the remainder of Bam 8b and part of the adjacent Bam 16 fragment has now been determined according to the strategy shown in Figure 1C. Our analysis was prompted initially by our observation that DNA fragments derived from the region hybridized to a single high molecular weight component of spinach chloroplast RNA (see later).

The sequence of 4381 bp extending from the HinfI site in Bam 8b to the double EcoRI site in Bam 16 (Figure 1B) is given in Figure 2. It contains two major open reading frames (353 and 473 codons) on one strand and the gene for a tRNA on the opposite strand. The translated sequence

Figure 1. Location of the genes for the D2 (psbD) and 44 kd (psbC) proteins of photosystem II and the gene for tRNA^{Ser} (trnS) on spinach chloroplast DNA. **A.** SalI/PstI restriction endonuclease map of spinach chloroplast DNA (26) showing the position of Bam 8b and Bam 16. The shaded areas denote the inverted repeat regions. Location of rbcl (20) and psbA (13) are shown for reference purposes. **B.** Complete BamHI, EcoRI, PstI, PvuII, ClaI, BglII, HindIII restriction map of Bam 8b and Bam 16. The only HinfI site marked is that at which the sequence shown in Figure 2 commences. The sequence of Bam 8b to the right of the HinfI site has been published (18,25). Black bars indicate the position of psbD and psbC which overlap, and trnS as well as the genes for the four other tRNAs in Bam 8b. The arrows show the direction of transcription of the genes. **C.** Strategy for determining the sequence given in Figure 2. Fragments were 3' or 5' end-labelled with [α -³²P]dNTPs or [γ -³²P]ATP by fill-in (→) or kinase (•→) reactions and sequenced (21) from the sites indicated in the direction of the arrows. The sequence was determined on both strands through all restriction sites. The sequence through the HinfI site at the extreme right-hand end originated at a HpaII site about 350 bp further to the right in Bam 8b.

[illegible]

of the open reading frames, both of which are read in the same direction as the gene for the large subunit of ribulose biphosphate carboxylase (rbcL, Figure 1A), is also shown in Figure 2.

Sequence of the gene for the D2 protein

Comparison of the spinach sequence with the sequence of the gene for the D2 protein of pea chloroplasts (8) clearly establishes the identity of the smaller open reading frame as the gene for the D2 protein (Figure 2). In the protein coding region (coordinates 1-1059, Figure 2) the two nucleotide sequences are 93% homologous. In accordance with the proposed nomenclature for chloroplast genes (28) we will refer to this gene as psbD. There is significant homology (76%) between the spinach and pea sequences upstream from the ATG codon for approximately 160 bp but beyond that point the two sequences diverge. Downstream from the psbD stop codon for as far as the published pea sequence extends (360 bp to position 1379, Figure 2 (8)), the level of homology between the two species is as high as it is within the D2 coding region (Figure 2). As will be seen below, this is because this region also is part of a protein coding sequence.

Of the 75 nucleotide differences between the spinach and pea psbD genes, 65 are third-position codon changes, and the proteins differ in only 9 out of the 353 amino acids (97% homology) (Figure 3). Figure 3 also includes a comparison with the D2 sequence of C. reinhardtii chloroplasts (9). Overall the homology between the spinach and C. reinhardtii proteins is 83% but, if 30 N-terminal and 28 C-terminal residues are excluded, the homology of the remaining 294 amino acids is 96%. More than half (51%) of the amino acid residues of spinach D2 are hydrophobic and the protein has a derived molecular weight of 39536.

Figure 2. Nucleotide sequence of a 4381 bp fragment of spinach chloroplast DNA and the deduced amino acid sequence (single letter code) of the D2 and 44 kd proteins. The sequence starts at the HinfI site in Bam 8b and ends at the double EcoRI site in Bam 16 (Figure 1) and only the mRNA-like strand is shown. Part of the sequence is compared with the published sequence of psbD from pea chloroplast DNA (8). Only those nucleotides which are different are shown below the spinach sequence. In the region upstream from the psbD initiation codon gaps have been introduced into the spinach sequence in order to improve the alignment. No attempt is made to compare the spinach and pea sequences 5' to -145 because the homology is so poor. The protein coding regions are boxed, the small box in the centre defining the region of overlap. A potential ribosome binding site (-7 to -11) and three sets of possible 'Pribnow-box'-like and '-35 region'-like sequences (27) in the 5' leader sequence are underlined. A sequence of dyad symmetry downstream from psbD is marked by inverted arrows. The sequence complementary to trnS is underlined (2682-2774). A 90-codon open reading frame on the same strand as trnS is defined by the underlined triplets complementary to the ATG start and TAG stop codons.

```

              19                      49                      80
S      MTIAGKFTKDEKDLFSDMDLRRDRFVFGWSGLLLFPCAYFALGGWFTGTTFTVTSWYTHGLASSYLEGCNFLTAAVS
P      ++++L+++++QN+++I+++++V+++++
C      MSCFFILEIHA+++I+TYQEKRTW ++DA+++++Q+++++L+++++T+++++

              110                     140                     171
S      TPANSLAHSLLLLWGPEAQGDFTRWCQLGGLWAFVALHGAFALIGFMLRQFELARSQQLRPYNAIAFSGPIAVFVSFLIYPLGQSGWFFA
P      ++++++L+++++T+++++G+++++N+++++A+++++
C      ++++M+++++FV+++++G+++++N+++++A+++++

              201                     231                     262
S      PSFGVAAIFRFLFFQGFHNWTLNPFHMMGVAGVLGAALLCAIHGATVENTLFEDGDGANTFRAFNPQTAEETYSMTANRFWSQIFGVAF
P      ++++++
C      ++++++

              292                     322                     353
S      SNKRWLHFFMLFVPYTLGWSALGVVGLALNLRAIDFVSQEIRAAEDPEFETFYTKNILLNEGIRAWMAAQDQPHENLIFPEEVLPRGNAL
P      ++++++
C      ++++++L+++++I+++++FFSIFIIPNHIINGSYFFNKSQKQIVYI

```

Figure 3. Comparison of the amino acid sequences of the D2 protein from spinach (S), pea (P) (8) and *C. reinhardtii* (C) (9). Only differences from the spinach sequence are shown. Numbering refers to the spinach sequence. The gap in the *C. reinhardtii* sequence was inserted in order to maximize homology.

The open reading frame overlapping psbD encodes the 44 kd photosystem II protein

The second open reading frame (473 codons, coordinates 1009-2428) shown in Figure 2 commences 50 bp prior to the TAA stop codon of psbD. It is on the same strand as psbD but in a different reading frame and encodes a protein of molecular weight 51816. The location and polarity of this coding region which spans the junction of Bam 8b and Bam 16 matches with the location and polarity of the gene for the 44 kd reaction-centre protein of photosystem II as recently mapped on spinach chloroplast DNA by Westhoff *et al.*, (2). We have sequenced all of Bam 16 and Bam 8b and there is no other open reading frame large enough to code for a 44 kd protein. We conclude this second open reading frame is the gene for the 44 kd protein and designate it psbC (28). The percentage of hydrophobic amino acids in the 44 kd protein (48%) is less than in D2 but in the range expected for a membrane located protein.

Rasmussen *et al.* (8) noted that in pea chloroplast DNA an open reading frame, extending for at least 123 codons, overlapped the 3' end of psbD by 50 bp. As the deduced amino acid sequences from spinach and pea are 96% homologous over these 123 codons (coordinates 1010-1380, Figure 2), it is probable that in pea, psbC is also located next to and overlaps psbD.

psbD and psbC are cotranscribed

DNA fragments specific for psbD and psbC and for the flanking regions were hybridized to Northern blots of total spinach chloroplast RNA

Figure 4. Hybridization of subfragments of Bam 8b and Bam 16 to a Northern blot of spinach chloroplast RNA. The origin of the probes used is indicated in the restriction map above the autoradiograph and is as follows: Lane A, 344 bp EcoRI fragment from Bam 16; Lane B, 371 bp BamHI-PstI fragment from Bam 16 containing psbC sequences; Lane C, 989 bp EcoRI-PvuII fragment from Bam 8b carrying psbD sequences; Lane D, 205 bp HindIII fragment from Bam 8b; Lane E, 266 bp HinfI fragment from Bam 8b, immediately upstream from the sequence shown in Figure 2 and containing trnT. The molecular weight markers are *E. coli* ribosomal RNAs. The size of the hybridizing band marked by the arrow is estimated as 3.5 kb. Apart from the tRNA in Lane E the other bands are ribosomal RNA shadow bands and are seen in most gel patterns.

(Figure 4). Sequences derived from the coding regions and from the region immediately upstream from psbD all hybridized to a single RNA species estimated to be approximately 3.5 kb in length. As the 344 bp EcoRI fragment downstream from psbC (Figure 4, lane A) and the 266 bp HinfI fragment immediately upstream from the beginning of the sequence given in Figure 2 (Figure 4, lane E) do not hybridize to this large RNA, the sequence transcribed to give the RNA must lie between the HinfI site shown in Figure 1B and the first EcoRI site downstream from psbC (Figure 1B), a total distance of 4037 bp. The 3.5 kb transcript must therefore be derived from both psbD and psbC. Cotranscription of these two genes is not unexpected in view of their overlapping reading frames.

The gene for tRNA^{Ser} (UGA) lies downstream from psbC

250 bp downstream from the psbC stop codon is the 92 bp gene for tRNA^{Ser} (trnS-UGA) (coordinates 2682-2774, Figure 2). It is on the DNA strand opposite to the one carrying psbC and its identification is based on the fact that the transcribed RNA sequence can be written in a conventional clover-leaf form, that the anticodon triplet is UGA and the fact that the sequence is 72% homologous with the sequence of the corresponding trnS gene from *E. coli* (Figure 5) (29). The sequence is also 70% homologous with trnS-GCU from *Euglena gracilis* chloroplasts (30) and 68% homologous

Figure 5. Clover-leaf representation of spinach chloroplast tRNA^{Ser}. The 3' terminal CCA is not coded by the gene. Arrows indicate nucleotide differences to the corresponding *E. coli* isoaccepting tRNA^{Ser} (29) except in the region of the extra arm where the *E. coli* sequence is shown in the inset.

with trnS-GGA from maize chloroplasts (31). It has been shown previously that tRNA^{Ser} hybridizes to this region of spinach chloroplast DNA (32).

Between psbC and trnS there is an 11 bp inverted repeat sequence (coordinates 2573-2583, 2591-2601, Figure 2) which might function as a transcription terminator for either of the genes. The other feature of note in the sequence is a 90-codon open reading frame (coordinates 3050-3319, Figure 2) which is read from the same strand as trnS. However, it probably is not functional because the 344 bp EcoRI fragment containing the sequence did not hybridize to chloroplast RNA (lane A, Figure 4).

DISCUSSION

The smaller open reading frame in the sequence shown in Figure 2 was identified as the gene for the D2 protein (psbD) on the basis of its homology to pea psbD (8), and the larger open reading frame as the gene for the 44 kd photosystem II protein (psbC) on the basis of the coincidence of its location with the previously determined position of this gene on spinach chloroplast DNA (2). In both cases the open reading frames predict proteins much larger than those expected from acrylamide gel electrophoresis of the purified proteins. There is no evidence that synthesis of the 44 kd protein involves a precursor protein (2) and although post-translational phosphorylation of the D2 protein of *C. reinhardtii* has been reported, this increases rather than decreases the apparent molecular weight (6). Synthesis of proteins smaller than those predicted would result if initiation of translation occurred at the second methionine (ATG) codon in

the open reading frames. This is very unlikely to be the case for psbD because the first ATG is preceded by the sequence GGAGG (-7 to -11, Figure 2) which is complementary to the 3' end of chloroplast 16S rRNA and is considered a strong ribosome binding site (33). There is no equivalent sequence immediately prior to the first ATG codon in psbC but the sequence GGAG occurs three nucleotides prior to the second ATG (position 1208, Figure 2) tending to support the notion that translation may be initiated at this site. However, the presence of a G-rich sequence preceding an ATG codon is not mandatory for translation initiation of chloroplast proteins, as evidenced by the gene for the β subunit of chloroplast ATPase (34).

The simplest explanation for the discrepancy between the estimated and derived molecular weights is that molecular weight estimations of highly hydrophobic proteins based on electrophoretic mobility in sodium dodecyl sulphate-polyacrylamide gels are frequently unreliable (35,36) and it is interesting to note that the difference between the estimated molecular weights (44 kd and 51 kd) of the two photosystem II reaction-centre proteins is almost the same as that between their derived molecular weights (51816 and 56246) (this paper and ref. 17). Resolution of the problem may be possible by directly sequencing the N-terminal region of the 44 kd protein.

One of the striking features of our analysis is the 50 bp overlap between the psbD and psbC genes (assuming translation commences at the first ATG in psbC). The only other report of overlapping chloroplast genes is that of the 1 bp overlap between the β and ϵ genes of ATPase (34,37). The reality of the overlap is substantiated by the fact that it probably also occurs in pea chloroplast DNA (8). Although psbC has not yet been identified in pea, the homology between the spinach gene and the pea sequence downstream from psbD (see Figure 2) leaves little room to doubt that this sequence in pea codes for the 44 kd protein and that the psbD/psbC arrangement is the same in both species. In C. reinhardtii chloroplast DNA, however, psbD does not appear to overlap with any gene (9) and the location of psbC has not yet been reported.

Cotranscription of psbD and psbC would seem to be unavoidable with overlapping genes and is supported experimentally by our identification of a 3.5 kb transcript from this region. Westhoff *et al.*, (38) reported transcripts of size 2.4, 4.2, 5.5 and over 6 kb from Bam 16 and of size 2.4 and 4.2 kb from Bam 8b. The 4.2 kb RNA might be equivalent to our 3.5

kb species as exact sizes are difficult to determine in this molecular weight range. Our Northern analysis argues against the existence of larger transcripts from the psbD/psbC region. This would also seem to follow from the fact that there is an inverted repeat sequence 140 bp downstream from psbC which could act as a transcription terminator and from the fact there is a tRNA gene 250 bp downstream from psbC which is transcribed in the reverse direction. The smaller, 2.4 kb RNA (38) could be a product of processing which we failed to detect.

Upstream from the psbD initiation codon there are numerous putative RNA polymerase binding sites (Pribnow boxes and '-35' regions (27)) and three sets of these have been marked in Figure 2. Since the synthesis of photosystem II proteins is stimulated by light it is possible that there is a transcription start site for low level constitutive mRNA synthesis and a transcription start site which is brought into action following exposure of plants to light. What is interesting is that none of the putative promoter sites occur in the region which shows strong homology with the pea sequence (-1 to -150, Figure 2) and presumably the sites are different in the two species. There is no obvious homology between the 5' leader sequences of spinach and C. reinhardtii (9).

Comparison of the D2 and 32 kd herbicide-binding proteins

The arguments for initially regarding the D2 protein as a constituent of photosystem II have been documented by Rasmussen et al. (8). Delepelaire (6), using three different approaches - C. reinhardtii mutants deficient in photosystem II activity, purified active photosystem II particles and protein synthesis inhibitors differentially specific for 70S and 80S ribosomes - has now unequivocally shown that D2 is a component of photosystem II and is synthesised in the chloroplasts. The gene for the D2 protein was first mapped onto C. reinhardtii chloroplast DNA (7) and subsequently by heterologous hybridization onto pea chloroplast DNA (8).

Comparison of the D2 protein from spinach with that from pea shows the sequence to be highly conserved (> 97%). The degree of conservation is considerably greater than that of say the β subunits of ATPase (39) from pea and spinach (89%) and is more reminiscent of the absolute conservation of the 32 kd herbicide binding protein of spinach and Nicotiana debneyi (13) which, like D2, is a component of photosystem II. Although the D2 protein of C. reinhardtii is only three amino acids shorter than the spinach and pea proteins, optimal alignment of the sequences is obtained by displacing them by 11 residues (Figure 3). Thus the second methionine in

Figure 6. Hydropathy plots of the D2 (A) and 32 kd herbicide-binding proteins (B) of spinach chloroplasts. Each point on the profile represents the sum of the hydropathy values of 9 adjacent amino acids calculated according to Kyte and Doolittle (40). The window of 9 amino acids is moved one residue at a time and the values are plotted against the amino acid sequence (Figure 7). Hydrophobic regions large enough to span the thylakoid membrane (20 or more residues) are denoted by roman numerals. The locations of histidine residues are indicated by arrow heads. The short bars mark those regions of the D2 and 32 kd protein sequences which show approximately 50% homology (Figure 7). The long bar in (B) indicates the region which has some homology with the M subunit of *R. sphaeroides* (41).

the *C. reinhardtii* protein is equivalent to the first methionine in the spinach and pea proteins, lending support to the proposal (9) that translation of *psbD* in the algal chloroplasts may initiate at the second methionine codon. The considerable divergence which is apparent between the higher plant and *C. reinhardtii* D2 sequences at both the N-terminal and C-terminal regions (Figure 3) indicates that these parts of the sequence are not critical with respect to the protein's function.

As 51% of the D2 amino acids are hydrophobic the protein is likely to be an intrinsic membrane protein. A hydropathy plot (40) of the protein shows the hydrophobic residues cluster into seven major regions (Figure 6A), five (possibly six) of which (marked I-V) are clearly large enough to be membrane spanning. The similarity of this pattern to the hydropathy plot of the 32 kd herbicide binding protein ((13) and Figure 6B) is striking not only in terms of the distribution of the hydrophobic regions but also in

```

                                30                                60                                90
D2  MTIAVGKFTKDEKDLFDSMDWLRDRFVFGVSGLLLFPCAYFALGGWFTGTTFTVTSWYTHGLASSYLEGCNFLTAAVSTPANSLAHS
   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
32kd MT AILERRESESLWGRFCNWTSTENRLIGWGVGLMIPTLLTATSVFIIAFAAPPYDIDGIREPVSGSLLYGNNIISGAIPTSAAI

                                119                                149                                179
D2  LLLWGPEAQGD FTRNCQLGGLWAFVALHGAFALIGFMLRQFELARSVQLRPYNAIAFSGPIAVFVSFVLIYPLGQSGWFFAPSGVAAI
   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
32kd GLHFYPIWEAASVDEWLYNGGPYELIVLHFLLGACVYMGREWELSFRLGMRPWIAVAYSAPVAAATAVFLIYPIGQSGFSDGMPLGISGT

                                209                                238                                266
D2  FRFILFFQGFHNWTLNPFHMMGVAGVLGAALLCAIHGATVENTLFEDGDGANTFRA FNPTQAEETYSMYTANR FWSQIFGVA FSNKR
   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
32kd FNFMIVFQAEHNILMHPFHLGVAGVGGSLFSAMHGLSVTSSLIRETTEENESANEGYRFGQEEETYNIVAAGHYFGRLIFQYASFNSR

                                296                                326
D2  WLHFFLMFVPPYTLGWSALGVVGLALNLRAVDFVSQEIARAADPEFETFYTKNILLNEGIRAWMAAQDPHENLIFPEEVLPRGNL
   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
32kd SLHFFLAAPVVGWIFLTALGISTMAFMNLGNFNFNQSVVDSQGRVINTWADIINRANLDMVEVMHERNAHNFPLDLAAIEAPSTNG

```

Figure 7. Comparison of the amino acid sequences of the D2 and 32 kd (13) proteins from spinach chloroplasts. Asterisks denote those amino acids which are identical in the two sequences. Five single gaps have been inserted in order to maximize the homology. Numbering refers to the D2 sequence.

terms of the correlation of the size of the interspersing hydrophilic regions.

Comparison of the amino acid sequences of the D2 and 32 kd proteins (Figure 7) shows that, although overall the two sequences are only 26% homologous, there are three regions (140-165; 180-219; 240-286, Figure 7) where the homology is 50%. These highly conserved sequences derive from equivalent positions in the hydropathy plots of the two proteins (Figure 6A,B) and relate mainly to the putative transmembrane segments III, IV and V. One of these regions (180-219) partly overlaps with the region (160-216) of the 32 kd protein which has been reported to show significant homology with part of the sequence of the M subunit of the reaction centre of *Rhodospseudomonas sphaeroides* (Figure 6 and ref. 41). Furthermore, residues 196-221 include the sequence noted by Hearst and Sauer (42) to be common to the L and M subunits of *R. capsulata* reaction centres and the 32 kd protein. This sequence (Figure 8) is characterized by the conserved peptides, Pro-Phe-His-Met(or Gly) and Ala-Met(or Ile)-His-Gly spaced in each case so that the histidines are exactly 16 residues apart (42). It has

```

D2      P F H M M G V A G V L G A A L L C A I H G A T V
32 kd   P F H M L G V A G V F G G S L F S A M H G S L V
R.c.L   P F H M L G I S L F F T T A W A L A M H G A L V
R.c.M   P F H G L S I A A L Y G S A L L F A M H G A T I
R.s.M   P F H G L S I A F L Y G S A L L F A M H G A T I

```

Figure 8. Comparison of the amino acid sequence of residues 196 to 219 of the D2 and 32 kd proteins of spinach (Figure 7) with presumptive quinone-binding regions of the L and M subunits of *R. capsulata* (42) and the M subunit of *R. sphaeroides* (41). Identical amino acids are boxed.

Figure 9. Hydropathy plot of the 44 kD reaction-centre photosystem II protein of spinach chloroplasts. See Legend to Figure 6 for an explanation of the plot. Histidine residues are indicated by arrow heads.

been suggested that the His-Met dipeptide might be involved in quinone binding (42). If this is in fact the case then the role of the D2 protein, whatever it might be, will presumably involve quinone. The D2 protein, like the M subunit, contains only one of the two His-Met dipeptides whereas the 32 kD protein, like the L subunit, contains both.

There are at least two regions in the 32 kD protein that play a role in herbicide binding. One is the tryptic peptide from position 141 to 225 to which azidoatrazine binds covalently (43); the other is the serine residue at position 264 which, in atrazine-resistant mutants of *Amaranthus hybridus* (14) and *Solanum nigrum* (44), is changed to a glycine residue. As the region 141-225 coincides with that part of the sequence which is well conserved between D2 and the 32 kD protein, and as D2 contains a serine residue at position 263, and as the relative location of these two key regions in the hydrophobicity profile is the same for both proteins, it might be expected that the D2 protein would also bind atrazine.

Comparison of the 44 kD and 51 kD reaction-centre proteins of photosystem II

The 44 kD protein, like the D2 protein, is highly hydrophobic. The hydrophobic amino acids cluster into seven well-defined regions but of these only five are clearly large enough to be membrane spanning domains (Figure 9). This is in contrast to the 51 kD protein which appears to contain seven transmembrane segments (17). Hydropathy plots of the two proteins have an overall similarity including a long hydrophilic segment in the C-terminal half which is interrupted by one hydrophobic region 30-40 residues from the C-terminus (Figure 9 and Ref. 17).

Although the 44 kD and 51 kD proteins do not cross-react

Figure 10. Comparison of the amino acid sequences of the 44 kd and 51 kd reaction-centre photosystem II proteins from spinach chloroplasts. The 51 kd protein sequence (lower line) is from Morris and Herrmann (17). Asterisks denote amino acids that are identical in the two proteins. Arrow heads indicate histidine residues. Gaps have been inserted in the sequence in order to align regions which are similar.

immunologically, both bind chlorophyll *a* and for that reason we compared their amino acid sequences (Figure 10). It can be seen that there are discrete regions which have significant homology and these have been aligned by introducing gaps into the two sequences. The sequences which match in the two proteins derive from equivalent regions of the respective hydropathy profiles (Figure 9 and Ref. 17). The coincidence of these specific residues implies some relatedness in terms of their function.

Comparison of mitochondrial and chloroplast cytochrome *b* sequences from a number of species has shown that the position of four histidine residues in two membrane spanning segments of the proteins are invariant (36,45). In each of the spans the histidines are separated by 13 residues and this has led to the formulation of a model in which the two protohaems are sandwiched between the two transmembrane segments (36,45). In the 44 kd protein there are two histidine residues in transmembrane segment II, two in segment IV and three in the C-terminal hydrophobic segment (Figure 9). In each segment two histidines are separated by 13 residues. In the 51 kd protein there are five sets of similarly spaced histidine residues in

membrane spanning segments (17) including the particular arrangement of three histidines in the C-terminal hydrophobic domain. The conservation of the histidine spacings raises the possibility that these residues may be involved in the non-covalent binding of chlorophyll *a* to the two proteins.

The structure of four of the chloroplast DNA coded components of spinach photosystem II is now known : the D2 and 44 kd proteins, the 51 kd protein (17) and the 32 kd herbicide binding protein (13). The gene for cytochrome *b*₅₅₉ has been identified (see footnote in Ref. 46) and its structure will presumably soon be forthcoming. The characterization of the nuclear coded components of the complex is now of immediate importance.

ACKNOWLEDGEMENTS

K.H. acknowledges the support of a Deutsche Forschungsgemeinschaft stipend. We are grateful to Dr. A.J. Gibbs, Research School of Biological Sciences, Australian National University, for providing access to computing facilities. We thank Dr. J.D. Rochaix for kindly letting us see a copy of his manuscript prior to publication.

+To whom correspondence should be addressed

*Present address: E.Merck, 6100 Darmstadt, FRG

REFERENCES

1. Broglie, R., Bellemare, G., Bartlett, S.G., Chua, N.H. and Cashmore, A.R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 7304-7308.
2. Westhoff, P., Alt, J. and Herrmann, R.G. (1983) *EMBO J.* 2, 2229-2237.
3. Bedbrook, J.R., Link, G., Coen, D.M., Bogorad, L. and Rich, A. (1978) *Proc. Natl. Acad. Sci. USA* 75, 1572-1576.
4. Driesel, A.J., Speirs, J. and Bohnert, H.J. (1980) *Biochim. Biophys. Acta* 610, 297-310.
5. Chua, N.H., and Gillham, N.W. (1977) *J. Cell Biol.* 77, 441-452.
6. Delepelaire, P. (1984) *EMBO J.* 3, 701-706.
7. Rochaix, J.D. (1981) *Experientia* 37, 323-340.
8. Rasmussen, O.F., Bookjans, G., Stumann, B.M. and Henningsen, K.W. (1984) *Plant Mol. Biol.* in press.
9. Rochaix, J.D., Dron, M., Rahire, M. and Malnoe, P. (1984) *Plant Mol. Biol.* in press.
10. Coruzzi, G., Broglie, R., Cashmore, A. and Chua, N.H. (1983) *J. Biol. Chem.* 258, 1399-1402.
11. Dunsmuir, P., Smith, S.M. and Bedbrook, J. (1983) *J. Mol. Appl. Gen.* 2, 285-300.
12. Broglie, R., Coruzzi, G., Lampka, G., Keith, B., and Chua, N.H. (1983) *Biotechnology* 1, 55-61.
13. Zurawski, G., Bohnert, H.J., Whitfeld, P.R. and Bottomley, W. (1982) *Proc. Natl. Acad. Sci. USA* 79, 7699-7703.
14. Hirschberg, J., and McIntosh, L. (1983) *Science* 222, 1346-1349.
15. Spielmann, A. and Stutz, E. (1983) *Nucleic Acids Res.* 11, 7157-7167.

16. Link, G. and Langridge, U. (1984) *Nucleic Acids Res.* 12, 945-958.
17. Morris, J. and Herrmann, R.G. (1984) *Nucleic Acids Res.* 12, 2837-2850.
18. Holschuh, K., Bottomley, W. and Whitfeld, P.R. (1984) *Plant Mol. Biol.* in press.
19. Whitfeld, P.R. and Bottomley, W. (1980) *Biochem. Internat.* 1, 172-178.
20. Zurawski, G., Perot, B., Bottomley, W. and Whitfeld, P.R. (1981) *Nucleic Acids Res.* 9, 3251-3270.
21. Maxam, A.M. and Gilbert, W. (1980) *Methods in Enzymology* 65, 499-560.
22. Whitfeld, P.R., Herrmann, R.G. and Bottomley, W. (1978) *Nucleic Acids Res.* 5, 1741-1751.
23. Thomas, P.S. (1983) *Methods in Enzymology* 100, 255-266.
24. Maniatis, T., Kee, S.G., Efstratiadis, A. and Kafatos, F.C. (1976) *Cell* 8, 163-182.
25. Holschuh, K., Bottomley, W. and Whitfeld, P.R. (1983) *Nucleic Acids Res.* 11, 8547-8554.
26. Herrmann, R.G., Whitfeld, P.R. and Bottomley, W. (1980) *Gene* 8, 179-191.
27. Rosenberg, M. and Court, D. (1979) *Ann. Rev. Genetics* 13, 319-353.
28. Hallick, R.B. and Bottomley, W. (1983) *Plant Mol. Biol. Reporter* 1, 38-43.
29. Sprinzl, M. and Gauss, D.H. (1982) *Nucleic Acids Res.* 10, r1-r55.
30. Karabin, G.D. and Hallick, R.B. (1983) *J. Biol. Chem.* 258, 5512-5518.
31. Steinmetz, A.A., Krebbers, E.T., Schwarz, Z., Gubbins, E. and Bogorad, L. (1983) *J. Biol. Chem.* 258, 5503-5511.
32. Driesel, A.J., Crouse, E.J., Gordon, K., Bohnert, H.J., Herrmann, R.G., Steinmetz, A., Mubumbila, M., Keller, M., Burkard, G. and Weil, J.H. (1979) *Gene* 6, 285-306.
33. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1342-1346.
34. Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6260-6264.
35. Darley-Usmar, M.V. and Fuller, S.D. (1981) *FEBS Lett.* 135, 164-166.
36. Saraste, M. (1984) *FEBS Lett.* 166, 367-372.
37. Krebbers, E.T., Larrinua, I.M., McIntosh, L. and Bogorad, L. (1982) *Nucleic Acids Res.* 10, 4985-5002.
38. Westhoff, P., Alt, J., Nelson, N., Bottomley, W., Bünemann, H. and Herrmann, R.G. (1983) *Plant Mol. Biol.* 2, 95-107.
39. Whitfeld, P.R., Zurawski, G. and Bottomley, W. (1982) In: *Structure and function of plant genomes*. Eds. O. Ciferri and L. Dure III. pp. 193-198. Plenum Press, New York.
40. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.
41. Williams, J.C., Steiner, L.A., Ogden, R.C., Simon, M.I. and Feher, G. (1983) *Proc. Natl. Acad. Sci. USA* 80, 6505-6509.
42. Hearst, J.E. and Sauer, K. (1983) *Z. Naturforsch.* 39C, 421-424.
43. Wolber, P.K. and Steinback, K.E. (1983) *Z. Naturforsch.* 39C, 425-429.
44. Hirschberg, J., Bleecker, A., Kyle, D.J., McIntosh, L. and Arntzen, C.J. (1983) *Z. Naturforsch.* 39C, 412-420.
45. Widger, W.R., Cramer, W.A., Herrmann, R.G. and Trebst, A. (1984) *Proc. Natl. Acad. Sci. USA* 81, 674-678.
46. Widger, W.R., Cramer, W.A., Hermodson, M., Meyer, D. and Gullifor, M. (1984) *J. Biol. Chem.* 259, 3870-3876.