The gene for the 10 kDa phosphoprotein of photosystem II is located in chloroplast DNA

Sean M. Hird*+, Tristan A. Dyer+ and John C. Gray*

*Botany School, University of Cambridge, Downing St., Cambridge CB2 3EA and Plant Breeding Institute, Trumpington, Cambridge CB2 2LQ, England

Received 6 October 1986

Nucleotide sequencing of a region of wheat chloroplast DNA between the genes for the 47 kDa chlorophyll *a*-binding protein of photosystem II (*psb*B) and cytochrome *b*-563 (*pet*B) has revealed an open reading frame of 73 codons. This open reading frame has been identified as the gene (*psb*H) for the 10 kDa phosphoprotein of photosystem II by comparison with the published N-terminal amino acid sequence and amino acid composition of the purified spinach protein. The predicted sequence of the protein shows some homology with the N-terminal region of the light-harvesting chlorophyll *a*/*b*-binding protein of photosystem II (LHCII).

Phosphoprotein *Photosystem II* *Chloroplast* *DNA* *(Wheat)*

1. INTRODUCTION

The chloroplast genome contains the genetic information for approximately half of the identified polypeptides of the thylakoid membrane system, including components of photosystems I and II, the cytochrome *b*-*f* complex and ATP synthase [1]. However, the location of the genetic information for a number of smaller polypeptides, particularly in photosystem II preparations [2], has not been determined. Although seven chloroplast genes for polypeptides of photosystem II have been identified [3–7] the possibility that other photosystem II polypeptides are encoded in chloroplast DNA must be considered. Nucleotide sequencing has revealed a number of small open reading frames which have been shown to be cotranscribed with known photosystem II genes. These include an open reading frame of 62 codons which is co-transcribed with the genes for the D2 and 44 kDa chlorophyll *a*-binding polypeptides [8,9] and two open reading frames of 38 and 40 codons which are cotranscribed with the genes for cytochrome *b*-559 polypeptides [10]. However, the putative products of these open reading frames have not yet been identified as photosystem II polypeptides.

We now report that nucleotide sequencing downstream of the gene (*psb*B) for the 47 kDa chlorophyll *a*-binding protein in wheat chloroplast DNA has revealed an open reading frame of 73 codons which may be identified as the gene (*psb*H) for the 10 kDa phosphoprotein of photosystem II by reference to the published N-terminal amino acid sequence and amino acid composition of the spinach protein [11].

2. MATERIALS AND METHODS

Plasmid pTac 2.33, containing the 3.3 kbp (B15), 4.6 kbp (B11) and 1.7 kbp (B20) *Bam*HI fragments of wheat chloroplast DNA inserted in the *Bam*HI site of pBR322, was a gift from C.M. Bowman. The location of these fragments on the restriction map of wheat chloroplast DNA is shown in fig.1. Nucleotide sequencing of 5′-end-labelled restriction fragments of wheat chloroplast DNA was carried out by the method of Maxam and Gilbert [12] with the modifications described

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

by Barker et al. [13]. The sequencing strategy is outlined in fig.1.

Comparisons of amino acid compositions were made using the parameter $S\Delta n$ [$= \frac{1}{2}\Sigma(n_{iA} - n_{iB})^2$ where n_i is the number of residues of an individual amino acid in proteins A or B] of Cornish-Bowden [14]. Comparisons of amino acid sequences were made using the DIAGON programme described by Staden [15].

3. RESULTS

Nucleotide sequencing of a region of wheat chloroplast DNA between the gene (*psb*B) for the 47 kDa chlorophyll *a*-binding polypeptide of photosystem II and the gene (*pet*B) for the apoprotein of cytochrome *b*-563 revealed an open reading frame of 73 codons. A detailed restriction map of a region of wheat chloroplast DNA delimited by *Hin*dIII and *Xba*I sites and its location in the wheat chloroplast genome are shown in fig.1. The nucleotide sequence of the open reading frame and its flanking regions is shown in fig.2. The open reading frame starts with an ATG codon and ends with a TAG codon. The N-terminal sequence of the protein predicted from the nucleotide sequence matches in 8 out of 9 residues with the determined N-terminal sequence (Ala-Thr-Gln-Thr-Val-Glu-Ser-Ser-Ser) of the 10 kDa phosphoprotein of spinach thylakoid membranes [11]. The probability of the matching sequence of 9 amino acid residues occurring by chance in a protein with the amino acid composition deduced from the wheat open reading frame is 1.6×10^{-11}, suggesting that the open reading frame is the gene for the 10 kDa phosphoprotein. Comparison of the amino acid composition of the protein deduced from the nucleotide sequence of the open reading frame with the amino acid composition determined for the spinach phosphoprotein (table 1), by the method of Cornish-Bowden [14], suggests considerable similarities between the proteins. The parameter $S\Delta n$ is an unbiased estimator of the number of differences between two sequences, and a value of 7.2 (or $0.10N$, where $N = 72$ amino acid residues, allowing for the removal of the N-terminal methionine residue) suggests that the sequences are 90% identical. The value of $S\Delta n$ of 7.2 is considerably less than $0.42N$ indicating a greater than 95% probability that the similarities between

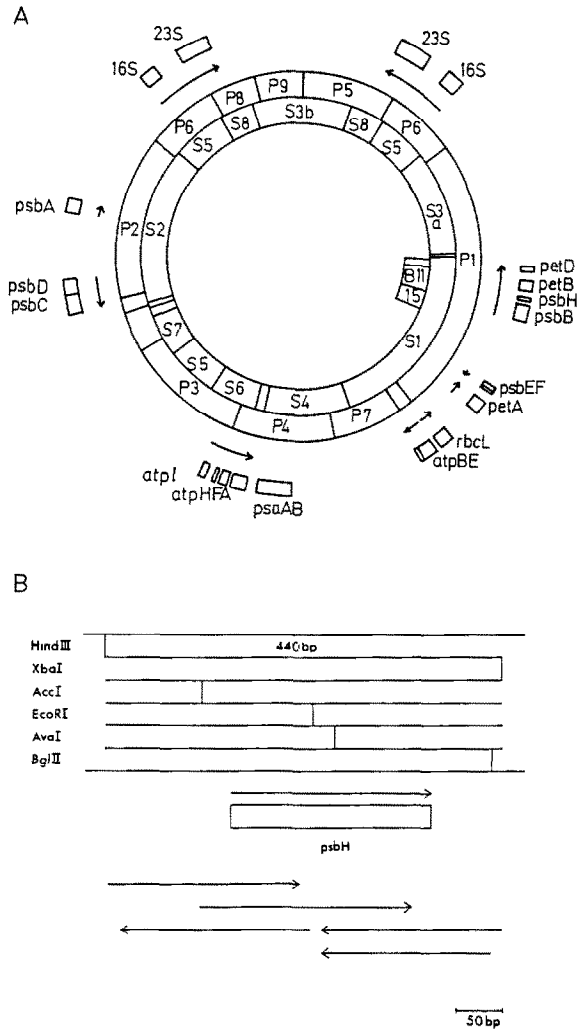

Fig.1. (A) Restriction map of wheat chloroplast DNA showing the cleavage sites for *Pst*I and *Sal*I [16] and the *Bam*HI sites in pTac 2.33. The position of the gene (*psb*H) for the 10 kDa phosphoprotein is shown as a filled box. The locations of other genes are given in [17,18]. The directions of transcription of genes are indicated by the arrows. (B) Detailed restriction map of the 440 bp *Hin*dIII-*Xba*I fragment and sequencing strategy. The direction and extent of sequencing are shown by the arrows. The dotted area represents plasmid DNA sequences not investigated in this study.

the proteins did not arise by chance [14]. The very close match of both the N-terminal sequence and the amino acid composition of the spinach protein with the deduced wheat sequence indicates beyond any reasonable doubt that the open reading frame

```
                                                                  50
ATGGATCTAC GAAAAGATCG TGTATTTACA ACTACAACGG AATAGTATAC AAAGTCAACA CCAATGATTA AATAGAATTT

                                 100
ATG GCT ACA CAA ACC GTT GAA GAT AGT TCT AAA CCT AGG CCA AAA CGA ACT GGT GCA GGT AGT TTA

Met Ala Thr Gln Thr Val Glu Asp Ser Ser Lys Pro Arg Pro Lys Arg Thr Gly Ala Gly Ser Leu
 1                                   10                                  20

    150                                                             200
CTG AAA CCC TTG AAT TCG GAA TAT GGG AAA GTC GCC CCG GGT TGG GGG ACT ACT CCT TTT ATG GGG

Leu Lys Pro Leu Asn Ser Glu Tyr Gly Lys Val Ala Pro Gly Trp Gly Thr Thr Pro Phe Met Gly
                                30                                  40

                                                        250
GTC GCA ATG GCT TTA TTC GCT ATA TTC CTA TCT ATC ATT TTA GAA ATT TAT AAT TCT TCC GTT TTA

Val Ala Met Ala Leu Phe Ala Ile Phe Leu Ser Ile Ile Leu Glu Ile Tyr Asn Ser Ser Val Leu
                    50                                  60

                                    300
CTG GAC GGA ATT TTA ACC AAT TAG GTTTCTACTA ACTAAAACTA

Leu Asp Gly Ile Leu Thr Asn
              70
```

Fig.2. Nucleotide sequence of the *psb*H gene and predicted amino acid sequence of the 10 kDa phosphoprotein.

Table 1

Amino acid compositions of the predicted wheat protein and the isolated spinach phosphoprotein

Amino acid	mol/100 mol[a]		mol/mol protein[b]	
	Spinach	Wheat	Spinach	Wheat
Asx	7.5	6.9	5.4	5
Thr	6.4	8.3	4.6	6
Ser	11.4	9.7	8.2	7
Glx	6.7	5.5	4.8	4
Pro	8.3	6.9	6.0	5
Gly	10.0	9.7	7.2	7
Ala	9.2	8.3	6.6	6
Val	7.7	5.5	5.5	4
Met	3.4	2.8	2.4	2
Ile	4.3	6.9	3.1	5
Leu	11.7	12.5	8.4	9
Tyr	2.8	2.8	2.0	2
Phe	2.9	4.2	2.1	3
His	0.2	0	0.1	0
Lys	4.6	5.5	3.3	4
Arg	3.1	2.8	2.2	2
Cys	0	0	0	0
Trp	0	1.4	0	1

[a] Values calculated from [11]
[b] Values calculated for polypeptide of 72 residues, after removal of N-terminal methionine residue

of 73 codons is the gene (*psb*H) for the phosphoprotein. This is the eighth gene for a photosystem II polypeptide to be located in chloroplast DNA. A similar open reading frame of 73 codons has been recognised in liverwort [19] and tobacco [20] chloroplast DNAs.

4. DISCUSSION

The identification of the open reading frame of 73 codons in wheat chloroplast DNA as the gene for the 10 kDa phosphoprotein provides information on the chemical composition of the protein. The protein deduced from the nucleotide sequence has a relative molecular mass of 7647 Da discounting the contribution of the initiating methionine residue. The protein is highly charged near the N-terminus with 6 basic and 3 acidic residues in the region of residues 7–32; two additional acidic residues are located near the C-terminus at residues 59 and 68. Overall the protein is predicted to be slightly basic, although on phosphorylation the protein would become negatively charged. A hydrophobic region near the C-terminus of the protein (residues 42–58) may fold to produce a membrane-spanning α-helical sequence. The presence of a single membrane-spanning region

would suggest that the N- and C-termini of the protein are located on opposite sides of the thylakoid membrane. Bennett [21] has shown that the protein is phosphorylated on a threonine residue(s) and the phosphorylated residues are removed by treatment of unstacked thylakoid membranes with trypsin [22]. Although the deduced wheat sequence contains 6 threonine residues, the one that is the most likely candidate for the phosphorylation site is Thr 17. The two threonines (Thr 3 and Thr 5) in the N-terminal sequence of the spinach phosphoprotein were reported not to be phosphorylated [11]. Thr 39, 40 and 72 would not be released by trypsin cleavage if residues 42–58 form a membrane-spanning sequence. Thr 17 would be released in a small peptide by trypsin cleavage at Lys 24 or 32. This threonine residue is also flanked by basic residues, a feature commonly observed at phosphorylation sites [23]. Removal of the phosphorylation site by trypsin treatment of unstacked thylakoid indicates that the phosphorylation site must be on the stromal side of the membrane. This suggests that the N-terminal sequences of the protein are located on the stromal side of the membrane, with the C-terminal residues in the intrathylakoid space. The mechanism by which the protein attains this configuration in the membrane is of considerable interest, but is currently unknown. One other feature of the synthesis of the mature protein is the removal of the initiating *N*-formylmethionine residue. In common with animal and bacterial systems, chloroplasts remove the N-terminal methionine residues of some proteins, for example the polypeptides of cytochrome *b*-559 [24,25] and the ϵ-subunit of ATP synthase [26], but not others [27].

Although the nucleotide sequence of the gene for the phosphoprotein can give valuable information on characteristics of the protein, it is unable to indicate the function of the protein. There have been several suggestions in the past concerning the identity and function of the 10 kDa phosphoprotein. These have included the proteolipid subunit of the ATP synthase [28], cytochrome *b*-559 [29] and cytochrome *b*-563 [30]. Although all of these suggestions can be discounted on the basis of the N-terminal sequence of the spinach phosphoprotein alone [11], comparisons of amino acid compositions using the parameter $S\Delta n$ are presented in table 2. Values for $S\Delta n$ of 86–136 for comparison of the spinach phosphoprotein and these other proteins clearly indicate the ability of the parameter $S\Delta n$ to discriminate between non-identical proteins. Also included in table 2 is a comparison with the N-terminal 90 residues of the wheat light-harvesting chlorophyll *a*/*b*-binding

Table 2

Comparison of the amino acid composition of the spinach 10 kDa phosphoprotein [11] and amino acid compositions of various other chloroplast proteins, deduced from nucleotide sequences of gene

Protein	$S\Delta n$[a]	N[b]	Ref.
Wheat ORF 73 product	7.2	72	this paper
Wheat proteolipid subunit of ATP synthase	96.0	81	[25]
Wheat 9 kDa polypeptide of cytochrome *b*-559	88.0	83	[18]
Spinach cytochrome *b*-563 residues 1–100	135.7	100	[29]
Spinach cytochrome *b*-563 residues 101–211	86.0	111	[29]
Wheat LHCII polypeptide residues 1–90	25.0	90	[30]

[a] $S\Delta n$ is defined in section 2

[b] N = no. of residues used for calculation of n_i for spinach phosphoprotein

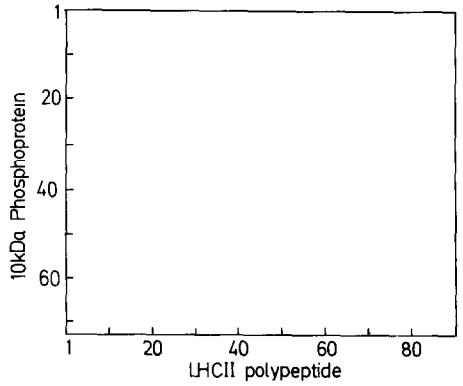

Fig.3. DIAGON plot of the amino acid sequences of the 10 kDa phosphoprotein (vertical axis) and the N-terminal 90 amino acid residues of wheat LHCII [32], constructed according to Staden [15]. Spans of 41 amino acid residues were compared and each dot represents a match at a probability of 0.5%.

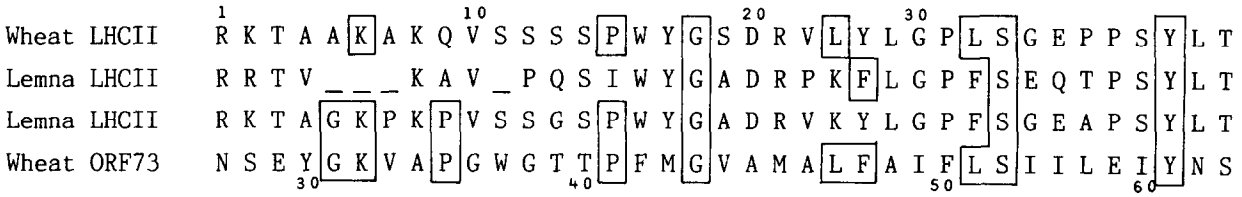

Fig.4. Alignment of the amino acid sequences of the 10 kDa phosphoprotein and the N-terminal residues of LHCII from wheat [32] and *Lemna* [34,35]. Identical residues are boxed.

protein (LHCII), which is the other main phosphoprotein of chloroplast membranes [21]. This protein is also phosphorylated on threonine residues [21] located near the N-terminus of the protein [33]. The value of $S\Delta n$ of 25 suggests some similarity between the sequences of the 10 kDa phosphoprotein and the LHCII polypeptide. Weak sequence homology between residues 27–63 of the wheat phosphoprotein and residues 1–37 of the LHCII polypeptide is indicated by a DIAGON plot (fig.3). A direct sequence comparison is shown in fig.4, indicating identical residues in 30% of the positions in this region of the wheat phosphoprotein and the LHCII polypeptides of monocotyledonous plants. Allowing for conservative replacements, these regions of the proteins show approx. 50% similar residues. The significance of this sequence homology is not clear, but presumably reflects a common evolutionary origin of the genes for these proteins. Allen and Holmes [36] have suggested that the phosphoprotein may bind chlorophyll and be involved in excitation energy distribution. However, the deduced protein sequence shows that the protein contains no histidine or cysteine residues which could be involved in chlorophyll binding. A role in excitation energy distribution may be the function of the protein. Phosphorylation of the 10 kDa polypeptide would result in a change in the charge on the N-terminal region of the polypeptide. As with LHCII, this may result in changes in the appression of the stacked granal membranes, allowing greater lateral mobility of the LHCII complexes and transfer of excitation energy to photosystem I. The lateral mobility of the 10 kDa polypeptide has not been investigated, and it is not presently known if the 10 kDa phosphoprotein shows the same lateral mobility as LHCII complexes. Differences in the lateral mobility of these two proteins may explain the differences in the rates of dephosphorylation of the phosphorylated proteins [22] if the protein phosphatase responsible for the dephosphorylation is located primarily in the stromal lamellae. Although much remains to be learned of the function of the 10 kDa phosphoprotein, the characterisation of the gene for this polypeptide provides some of the basic information needed for studies on its function.

ACKNOWLEDGEMENTS

We are grateful to Dr C.M. Bowman for the gift of plasmid DNA, and to E.R. Blyden and R.F. Barker for help and advice. S.M.H. was supported by a CASE award from SERC.

REFERENCES

[1] Gray, J.C., Bird, C.R., Courtice, G.R.M., Hird, S.M., Howe, C.J., Huttly, A.K., Phillips, A.L., Smith, A.G., Willey, D.L., Bowman, C.M. and Dyer, T.A. (1986) Biochem. Soc. Trans. 14, 25–27.
[2] Ljungberg, U., Henrysson, T., Rochester, C.P., Akerlund, H.-E. and Andersson, B. (1986) Biochim. Biophys. Acta 849, 112–120.
[3] Bedbrook, J.R., Link, G., Coen, D.M., Bogorad, L. and Rich, A. (1978) Proc. Natl. Acad. Sci. USA 75, 3060–3064.
[4] Westhoff, P., Alt, J. and Herrmann, R.G. (1983) EMBO J. 2, 2229–2237.
[5] Rochaix, J.-D., Dron, M., Rahire, M. and Malnoe, P. (1984) Plant. Mol. Biol. 3, 363–370.
[6] Herrmann, R.G., Alt, J., Schiller, B., Widger, W.R. and Cramer, W.A. (1984) FEBS Lett. 176, 239–244.
[7] Steinmetz, A.A., Castroviejo, M., Sayre, R.T. and Bogorad, L. (1986) J. Biol. Chem. 261, 2485–2488.
[8] Holschuh, C., Bottomley, W. and Herrmann, R.G. (1984) Nucleic Acids Res. 12, 8819–8834.
[9] Oliver, R.P. and Poulsen, C. (1984) Carlsberg. Res. Commun. 49, 647–673.

[10] Willey, D.L. and Gray, J.C. (1986) Mol. Gen. Genet., submitted.
[11] Farchaus, J. and Dilley, R.A. (1986) Arch. Biochem. Biophys. 244, 94–101.
[12] Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[13] Barker, R.F., Idler, D.B., Thompson, D.V. and Kemp, J.D. (1983) Plant. Mol. Biol. 2, 335–350.
[14] Cornish-Bowden, A. (1977) J. Theor. Biol. 65, 735–742.
[15] Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961.
[16] Bowman, C.M., Koller, B., Delius, H. and Dyer, T.A. (1981) Mol. Gen. Genet. 183, 93–101.
[17] Courtice, G.R.M., Bowman, C.M., Dyer, T.A. and Gray, J.C. (1985) Curr. Genet. 10, 329–333.
[18] Hird, S.M., Willey, D.L., Dyer, T.A. and Gray, J.C. (1986) Mol. Gen. Genet. 202, 95–100.
[19] Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.
[20] Shinozaki, J., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.
[21] Bennett, J. (1977) Nature 269, 344–346.
[22] Bennett, J. (1980) Eur. J. Biochem. 104, 85–89.
[23] Williams, R.E. (1976) Science 192, 473–474.
[24] Widger, W.R., Cramer, W.A., Hermodson, M., Meyer, D. and Gullifor, M. (1984) J. Biol. Chem. 259, 3870–3876.
[25] Widger, W.R., Cramer, W.A., Hermodson, M. and Herrmann, R.G. (1985) FEBS Lett. 191, 186–190.
[26] Howe, C.J., Fearnley, I.M., Walker, J.E., Dyer, T.A. and Gray, J.C. (1985) Plant. Mol. Biol. 4, 333–345.
[27] Howe, C.J., Auffret, A.D., Doherty, A., Bowman, C.M., Dyer, T.A. and Gray, J.C. (1982) Proc. Natl. Acad. Sci. USA 79, 6903–6907.
[28] Alfonzo, R., Nelson, N. and Racker, E. (1980) Plant Physiol. 65, 730–734.
[29] Widger, W.R., Farchaus, J., Cramer, W.A. and Dilley, R.A. (1984) Arch. Biochem. Biophys. 233, 72–79.
[30] Suss, K.-H. (1981) Biochem. Biophys. Res. Commun. 102, 724–729.
[31] Heinemeyer, W., Alt, J. and Herrmann, R.G. (1984) Curr. Genet. 8, 543–549.
[32] Lamppa, G.K., Morelli, G. and Chua, N.-H. (1985) Mol. Cell. Biol. 5, 1370–1378.
[33] Mullet, J.E. (1983) J. Biol. Chem. 258, 9941–9948.
[34] Kohorn, B.D., Harel, E., Chitnis, P.R., Thornber, J.P. and Tobin, E.M. (1986) J. Cell Biol. 102, 972–981.
[35] Karlin-Neumann, G.A., Kohorn, B.D., Thornber, J.P. and Tobin, E.M. (1985) J. Mol. Appl. Genet. 3, 45–61.
[36] Allen, J.F. and Holmes, N.G. (1986) FEBS Lett. 202, 175–181.