Volume 237, number 1,2, 108-112 FEB 06284 September 1988

A cDNA clone encoding a 10.8 kDa photosystem I polypeptide of barley

Jens Sigurd Okkels, Lars Bæk Jepsen, Lisbeth Skou Hønberg*, Jan Lehmbeck+, Henrik Vibe Scheller, Peter Brandt*, Gunilla Høyer-Hansen*, Bjarne Stummann+, Knud W. Henningsen+, Diter von Wettstein* and Birger Lindberg Møller

Department of Plant Physiology and +Department of Genetics, Royal Veterinary and Agricultural University, Frederiksberg C and **Department of Physiology, Carlsberg Laboratory, Valby, Denmark*

Received 19 July 1988

A cDNA clone encoding the barley photosystem I polypeptide which migrates with an apparent molecular mass of 16 kDa on SDS-polyacrylamide gels has been isolated. The 634 bp sequence of this clone has been determined and contains one large open reading frame coding for a 15 457 Da precursor polypeptide. The molecular mass of the mature polypeptide is 10 821 Da. The amino acid sequence of the transit peptide indicates that the polypeptide is routed towards the stroma side of the thylakoid membrane. The hydropathy plot of the polypeptide shows no membrane-spanning regions.

Photosystem I; 16 kDa polypeptide; cDNA sequence; Transit peptide; (Barley)

1. INTRODUCTION

Photosystem I (PS I) catalyzes a coupled photoreduction of ferredoxin and photooxidation of plastocyanin. To elucidate the mechanism of these reactions at the molecular level, one approach is to determine the primary structure of the proteins involved. PS I preparations from barley contain the 82/83 kDa reaction center protein P700-chlorophyll *a*-protein 1 and smaller polypeptides with apparent molecular masses of 18, 16, 14, 9.5 and 9 kDa [1,2]. The P700-chlorophyll *a*-protein 1 and the 9 kDa polypeptide have been identified as iron-sulfur proteins carrying the electron acceptors X, A and B [1,2]. The role of the remaining polypeptides in PS I is unclear. Whereas the genes for P700-chlorophyll *a*-protein 1 and the 9 kDa polypeptide are located in the chloroplast genome [2-4], studies of protein synthesis in greening chloroplasts [5,6], combined with the use of protein synthesis inhibitors [7], and studies on the in vitro translation of poly(A^+) RNA [8,9] have indicated that the 18 and 16 kDa polypeptides are synthesized on cytoplasmic ribosomes. Partial amino acid sequencing of the 18 and 16 kDa polypeptides has been carried out [10]. These sequences reveal no homology to sequences deduced from the DNA sequence of the chloroplast genome [11,12]. A cDNA clone encoding the 20 kDa PS I polypeptide from spinach has recently been sequenced [13]. A comparison of the deduced amino acid sequence of the 20 kDa spinach polypeptide and of the partial amino acid sequences of the 18 kDa polypeptide from barley demonstrates that these two polypeptides are homologous.

We present here the sequence of a cDNA clone encoding the full-length precursor for the 16 kDa polypeptide in barley.

Correspondence address: J.S. Okkels, Department of Plant Physiology, Royal Veterinary and Agricultural University, 40 Thorvaldsensvej, DK-1871 Frederiksberg C, Denmark

Abbreviations: PS I, photosystem I; bp, basepairs

The nucleotide sequence presented here has been submitted to the EMBL/GenBank data base under the accession no. Y00966

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

2. MATERIALS AND METHODS

From the partial amino acid sequence data of the 16 kDa polypeptide [8], two different mixtures of oligonucleotide probes were synthesized, probes N and C, specifying the N- and C-terminal regions of the polypeptide, respectively:

$$\overset{N}{H_2N\text{-}A\overline{EEPTAA}APA.....NYAL\overline{DEIKEV}AA\text{-}COOH}\overset{C}{}$$

N: 5′-GA$^{G}_{A}$GA$^{G}_{A}$CCIACIGCIGC-3′ (4 different oligonucleotides; I, inosine)

C: 5′-AC$^{T}_{C}$TC$^{T}_{C}$TT$^{A}_{T}$GAT$^{T}_{C}$TC$^{G}_{A}$TC-3′ (48 different oligonucleotides)

The oligonucleotide probes were 5′-end labeled by incubation for 1 h at 37°C of 20 pmol oligonucleotide with 5 U T_4 polynucleotide kinase (BRL) and 20 pmol [γ-^{32}P]ATP (5000 Ci/mmol, Amersham) in a reaction mixture containing 0.1 M Tris-HCl, pH 7.5, 10 mM $MgCl_2$ and 5 mM 1,4-dithiothreitol.

Total RNA was isolated from 5-day-old etiolated barley seedlings after prior exposure to light for 5 h. Poly(A)$^+$ mRNAs were purified by two cycles of oligo(dT)-cellulose column chromatography and used for construction of a λgt11 cDNA library essentially as described by Glover [14].

The library was screened with the oligonucleotide probes using nitrocellulose filters treated as described by Berger and Kimmel [15]. Southern blotting and hybridisation were performed using Zetaprobe membranes according to manufacturer (BioRad).

The immunological screening was conducted as in [16]. The fusion proteins were immobilized on nitrocellulose filters (Schleicher & Schuell). The primary antibody used was a monoclonal antibody towards the barley 16 kDa polypeptide [8]. Antigen-antibody complexes were detected enzymatically with alkaline phosphatase-conjugated antibodies (Dakopatts).

DNA from λ phages was prepared according to Grossberger [17]. Inserts from λ phages were subcloned in either the pTZ18/19 plasmids [18] or M13 vectors [19]. Single-stranded DNAs were obtained from the pTZ18/19 plasmids using helper phage M13K07 [20]. DNA sequencing was carried out by the dideoxy chain-termination method of Sanger et al. [21], using [α-^{32}P]dATP (Amersham) and a sequenase kit (US Biochemicals).

3. RESULTS AND DISCUSSION

Five positive clones were isolated from approx. 6×10^3 recombinants upon screening the barley cDNA library with probe C. One of these clones was also positive with a monoclonal antibody to the 16 kDa polypeptide [8] and as well as with probe N on Southern blotting analysis. This clone was sequenced in both directions with overlaps for all restriction sites (fig.1). The nucleotide sequence is shown in fig.2. Analyses of the sequence reveal one large open reading frame of 441 bp. The deduced amino acid sequence corresponding to bp 119–204 and 304–441 is identical to the partial amino acid sequences obtained by protein sequencing of the 16 kDa polypeptide [10]. The open reading frame codes for a precursor protein with a calculated molecular mass of 15457 Da. Since the N- and C-terminal amino acid sequences of the mature polypeptide have been determined [10], the molecular mass of the mature polypeptide and the N-terminal transit peptide can be calculated to be 10821 and 4636 Da, respectively. No processing takes place at the C-terminal end of the precursor. The precursor and the mature polypeptides migrate with apparent molecular masses of 23 and 16 kDa, respectively, on SDS-polyacrylamide gels [8,2]. The discrepancy from the actual molecular mass, as deduced from the nucleotide sequence data, reflects an unusual migration of the polypeptide in SDS-polyacrylamide gels. The total charge of the precursor and the mature polypeptide at pH 7 is estimated to be +2 and +7, respectively. The charge of the mature polypeptide is not sufficiently high to explain the anomalous migration of the polypeptide on SDS-polyacrylamide gels. The polypeptide contains a possible glycosylation site (Asn-X-Thr), but chloroplast membrane and PS I preparations have been shown to contain little or no carbohydrate ([22] and unpublished). In addition, plasma desorption mass spectrometry indicates that, apart from the removal of the transit peptide, the polypeptide is not posttranslationally modified (Scheller et al., unpublished). The presence at −33 bp of a stop codon in-frame with the open reading frame (fig.2) eliminates the possibility of a larger translation product from the same gene. Preliminary Southern and Northern analyses have further indicated that only one gene for this polypeptide is present in the barley genome and only one transcript is formed (not shown).

Translation could start at the three in-frame AUGs located at positions 1, 16 and 40 of fig.2. We assume that the first AUG is the initiation codon, because translation is initiated at the first AUG in 95% of all eucaryotic mRNAs [23,24]. According to the modified scanning model of translation in eucaryotes [25], a consensus sequence (GCC)GCCACCAUGG is found in mammalian cells around the initiation codon [26] with a purine nucleotide at the −3 position as the most important: [27].

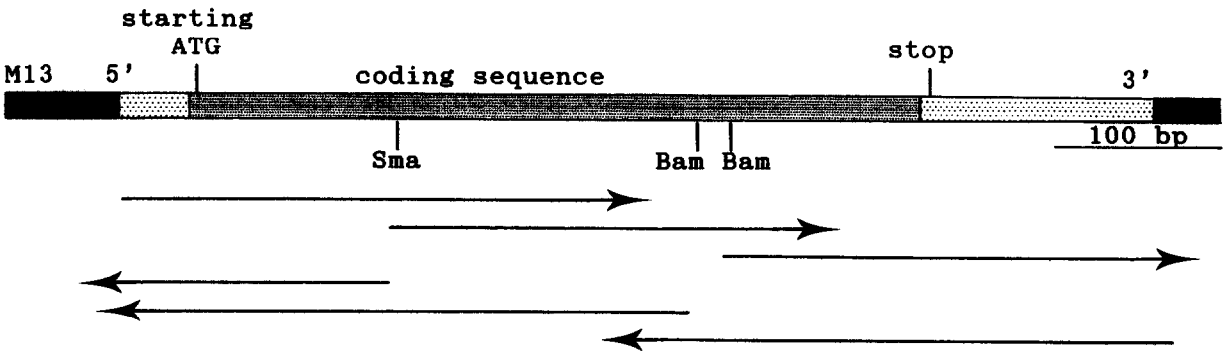

Fig.1. Sequencing strategy and the different sets of overlapping subclones.

However, Lütcke et al. [28] have reported that the in vitro translation systems of plants and animals exhibit different initiation codon selections with the nucleotide at the −3 position having no influence on the translation efficiency in the wheat germ extract. They also found that the initiation codon regions of plants have another consensus sequence: AACAAUGGC with a lower preference for a purine at the −3 position. The segment of the +40 AUG has no similarity to the plant consensus sequence, while both the +1 AUG and +16 AUG regions differ by three nucleotides. This means that the +16 AUG also could serve as an initiation codon, resulting in a smaller precursor polypeptide.

The hydropathy profile for the precursor polypeptide, presented in fig.3, shows no hydrophobic area sufficiently long to represent a membrane-

```
                                                CTAGCTAGCACCCACAAACAGCACCTGCTGCCGTCCCG
                                          -38  -33                                  -1
met ala ser thr asn met ala ser ala thr ser arg phe met leu ala ala gly ile pro
ATG GCG AGC ACC AAC ATG GCG TCG GCC ACC TCC AGA TTC ATG CTG GCG GCG GGC ATC CCC
1                                                                               60
ser gly ala asn gly gly val ser ser arg val ser phe leu pro ser asn arg leu gly
AGC GGC GCC AAC GGC GGC GTG AGC AGC CGT GTC AGC TTC CTC CCG TCC AAC CGG CTC GGC
                                                                               120
leu lys leu val ala arg ala glu glu pro thr ala ala ala pro ala glu pro ala pro
TTG AAG CTC GTG GCC CGG GCC GAG GAG CCG ACT GCC GCC GCG CCG GCG GAA CCA GCA CCG
                                                                               180
ala ala asp glu lys pro glu ala ala val ala thr lys glu pro ala lys ala lys pro
GCG GCG GAC GAG AAA CCG GAA GCC GCC GTG GCC ACC AAA GAG CCC GCC AAA GCC AAG CCG
                                                                               240
pro pro arg gly pro lys arg gly thr lys val lys ile leu arg arg glu ser tyr trp
CCG CCG AGG GGA CCC AAG AGG GGC ACC AAG GTG AAG ATC CTG AGG AGG GAG TCC TAC TGG
                                                                               300
tyr asn gly thr gly ser val val thr val asp gln asp pro asn thr arg tyr pro val
TAC AAC GGG ACT GGA TCC GTC GTC ACG GTT GAT CAG GAT CCC AAC ACC CGT TAC CCG GTG
                                                                               360
val val arg phe ala lys val asn tyr ala gly val ser thr asn asn tyr ala leu asp
GTG GTG CGT TTC GCC AAG GTG AAC TAC GCC GGC GTG TCG ACC AAC AAC TAC GCC CTG GAC
                                                                               420
glu ile lys glu val ala ala STOP
GAG ATC AAG GAG GTT GCT GCT TGA ACGATCGAGGCTGCCGCGTGCTCAATCCAATGTTTGTATCAGTAGCT
                                                                               491
CGTCAAGTGGCGATGTGAATGTTAGCCTCACAAATCTTATGTGTAATACCTCTGCGATTATATGTATTTGCCTGCTTCC
                                                                               570
TC(A)24
```

Fig.2. DNA sequence of the cDNA clone. The nucleotide sequence is written in the mRNA sense. The deduced amino acid residues are denoted in the three-letter code. The stop codon at the 5'-end, in-frame with the open reading frame, is underlined. Above the amino acid sequence are delineated different interesting domains: wavy line indicates predicted helices; maturation site is shown by an arrowhead and the positively charged area is overlined.

spanning region. A positively charged region is found just after the potential α-helix (fig.3). A similar region is found in the spinach 20 kDa PS I polypeptide [13].

A homology search using the National Biomedical Research Foundation protein bank revealed a 60% homology of the amino acids corresponding to bp 154–229 with myosin L1 and L4 light chains from skeletal muscle. The significance of this finding is unclear. The homology is in the region which also reveals a similarity to the 18 kDa polypeptide of barley PS I [10]. This similarity is, however, not conserved in the homologous 20 kDa polypeptide of spinach PS I [13].

The first 46 N-terminal amino acids represent the transit peptide. Transit peptides possess information for targeting of the polypeptide into the chloroplast and for its routing to a specific location. Three homology blocks on the transit peptides of the chloroplast have been suggested to be responsible for these functions [30]. The N-terminal sequence of the 15.5 kDa precursor polypeptide, MASTNMASA, exhibits similarity with the homology box I, MA.S.M.SS, which is suggested to be essential for uptake of the precursor into chloroplasts [30]. A weak similarity to box II (P.F.G.K.), possibly important for an intermediate processing event [30], can be found at PSNRLGLK. No obvious similarity to box III (G.GRV) is detectable except for a sequence with similar properties, GLKLVAR, immediately preceding the final maturation site. This sequence is also a disrupter of ordered secondary structure (due to glycine), and contains positively charged as well as hydrophobic amino acids; all characteristics that could make a region susceptible to proteolytic cleavage.

The length of the transit peptides of polypeptides located in the lumen of the thylakoid, such as plastocyanin [31] and oxygen-evolving enhancer proteins 1-3 [32–34], exceeds 56 amino acid residues. The hydrophobic thylakoid transfer domain shared by these four lumen polypeptides is lacking in the 10.8 kDa polypeptide. Shorter transit peptides comparable to that of the 10.8 kDa polypeptide have been observed for typical intrinsic polypeptides, such as the chlorophyll *a/b*-binding polypeptides [35,36] and for some stroma localized polypeptides, such as ferredoxin [37]. The properties of the transit peptide of the 10.8 kDa polypeptide show similarities with those of ferredoxin [38], suggesting a routing of the 10.8 kDa mature polypeptide to the stromal side of the thylakoid membrane. The function of the polypeptide could be to facilitate electron transport between P700 and ferredoxin.

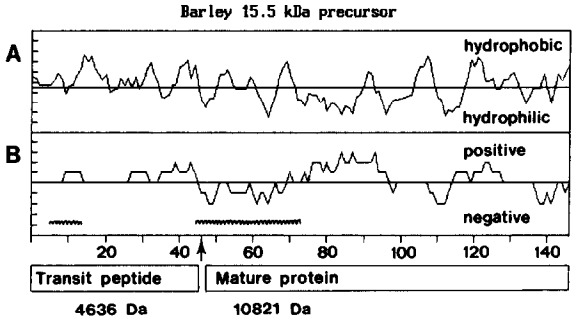

Fig.3. (A) Hydrophobicity plot of the 15.5 kDa precursor polypeptide according to Kyte and Doolittle [29] with an averaging window on 5 amino acids. (B) Charge distribution in the 15.5 kDa precursor polypeptide. The predicted helices are delineated by a wavy line and the maturation site by an arrow.

Acknowledgements: The authors are grateful to Peter Stougård and Finn Okkels, DDS, for synthesising the oligonucleotides, Dr Søren K. Rasmussen for advice with respect to nucleotide sequencing, and Inga Olsen and Hanne Linde Nielsen for skillful technical assistance. This work was supported in part by grants from the Danish Governmental Program for Biotechnology Research.

REFERENCES

[1] Høj, P.B. and Møller, B.L. (1986) J. Biol. Chem. 261, 14292–14300.
[2] Høj, P.B., Svendsen, I., Scheller, H.V. and Møller, B.L. (1987) J. Biol. Chem. 262, 12676–12684.
[3] Fish, L.E., Kück, U. and Bogorad, L. (1985) J. Biol. Chem. 260, 1413–1421.
[4] Hayashida, N., Matsubayashi, T., Shinozaki, K., Sugiura, M., Inoue, K. and Ozeki, H. (1987) Curr. Genet. 12, 247–250.
[5] Obokata, J. (1984) Plant Cell Physiol. 25, 821–830.
[6] Obokata, J. (1986) Plant Physiol. 81, 705–707.
[7] Nechustai, R. and Nelson, N. (1981) J. Biol. Chem. 256, 11624–11628.
[8] Høyer-Hansen, G., Hønberg, L.S. and Høj, P.B. (1985) Carlsberg Res. Commun. 50, 211–221.
[9] Westhoff, P., Alt, J., Nelson, N., Hermann, R.G., Bottomley, W., Bünemann, H. and Hermann, R.G. (1983) Plant Mol. Biol. 2, 95–107.
[10] Scheller, H.V., Høj, P.B., Svendsen, I. and Møller, B.L. (1988) Biochim. Biophys. Acta 933, 501–505.

[11] Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-I., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.
[12] Shinozaki, K., 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, T., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.
[13] Lagoutte, B. (1988) FEBS Lett. 232, 275–280.
[14] Glover, D.M. (1985) DNA Cloning, vol. 1, A Practical Approach, IRL, Oxford.
[15] Berger, S.L. and Kimmel, A.R. (1987) Methods Enzymol. 152.
[16] Young, R.A., Bloom, B.R., Grosskinsky, C.M., Ivanyi, J., Thomas, D. and Davis, R.W. (1985) Proc. Natl. Acad. Sci. USA 82, 2583–2587.
[17] Grossberger, D. (1987) Nucleic Acids Res. 15, 6737.
[18] Mead, D.A., Szczesna-Skorupa, E. and Kemper, B. (1986) Protein Eng. 1, 67–74.
[19] Messing, J. and Vieira, J. (1982) Gene 19, 269–276.
[20] Vieira, J. and Messing, J. (1988) Methods Enzymol. 153, 3–11.
[21] Sanger, F., Niklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[22] Keegstra, K. and Cline, K. (1982) Plant Physiol. 70, 232–237.
[23] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[24] Cigan, A.M. and Donahue, T.F. (1987) Gene 59, 1–18.
[25] Kozak, M. (1981) Nucleic Acids Res. 9, 5233–5252.
[26] Kozak, M. (1987) Nucleic Acids Res. 15, 8125–8148.
[27] Kozak, M. (1987) J. Mol. Biol. 196, 947–950.
[28] Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) EMBO J. 6, 43–48.
[29] Kyte, J. and Doolittle, W.F. (1982) J. Mol. Biol. 157, 105–132.
[30] Karlin-Neumann, G.A. and Tobin, E.M. (1986) EMBO J. 5, 9–13.
[31] Smeekens, S., De Groot, M., Van Binsbergen, J. and Weisbeek, P. (1985) Nature 317, 456–458.
[32] Jansen, T., Rother, C., Steppuhn, J., Reinke, H., Beyruther, K., Jansson, C., Andersson, B. and Hermann, R.G. (1987) FEBS Lett. 216, 234–240.
[33] Mayfield, S.P., Rahire, M., Frank, G., Zuber, H. and Rochaix, J.-D. (1987) Proc. Natl. Acad. Sci. USA 84, 749–753.
[34] Tyagi, A., Hermans, J., Steppuhn, J., Jansson, C., Vater, F. and Hermann, R.G. (1987) Mol. Gen. Genet. 207, 288–293.
[35] Cashmore, A.R. (1984) Proc. Natl. Acad. Sci. USA 81, 2960–2964.
[36] Hoffman, N.E., Pichersky, E., Malik, V.S., Castresana, C., Ko, K., Darr, S.C. and Cashmore, A.R. (1987) Proc. Natl. Acad. Sci. USA 84, 8844–8848.
[37] Smeekens, S., Van Binsbergen, J. and Weisbeek, P. (1985) Nucleic Acids Res. 13, 3179–3194.
[38] Smeekens, S., Bauerle, C., Hageman, J., Keegstra, K. and Weisbeek, P. (1986) Cell 46, 365–375.