Volume 228, number 1, 157–161 FEB 05546 February 1988

N-terminal amino acid sequence analysis of the subunits of pea photosystem I

Paul P.J. Dunn[+]*, Leonard C. Packman*, Darryl Pappin° and John C. Gray[+]

*Departments of [+] Botany and * Biochemistry, University of Cambridge, Cambridge* and *° AFRC Sequencing Laboratory, Department of Biochemistry, University of Leeds, Leeds, England*

Received 30 October 1987; revised version received 14 December 1987

Six 'core' subunits of pea photosystem I have been isolated and their N-terminal amino acid sequences determined by gas-phase or solid-phase sequencing. On average more than thirty residues were determined from the N-terminus of each polypeptide. This sequence analysis has revealed three polypeptides with charged N-terminal regions (21, 17 and 11 kDa subunits), one polypeptide with a predominantly hydrophobic N-terminal region (9 kDa subunit), one polypeptide which is cysteine-rich (8 kDa subunit) and one which is alanine-rich (13 kDa subunit).

Photosystem I subunit; N-terminal sequence; Amino acid sequence; (Pea)

1. INTRODUCTION

Electrons carried by plastocyanin, a soluble protein of the thylakoid lumen, are ultimately utilised in the reduction of ferredoxin, a soluble protein found in the chloroplast stroma. This transfer of electrons from one region of the chloroplast to another is mediated by photosystem I. Photosystem I has been isolated from a variety of higher plants, including pea [1–3], spinach [4], barley [5] and wheat [6]. The subunit composition of this complex has not been well defined due to the variation in the number and size of the subunits reported for different photosystem I preparations. For example, pea photosystem I has been reported to have 6 [2], 10 [1], or 16 [3] subunits and barley photosystem I has been reported to contain only 3 [7] or 5 [5] subunits. However, photosystem I preparations from both spinach [4] and wheat [6] are reported to have 7 subunits.

Photosystem I is known to contain two large polypeptides which bind P700 and the primary acceptors A_0, A_1 and X [5,8]. The genes for these polypeptides have been located and sequenced in pea [9], maize [10], and spinach [11] chloroplast DNA. In addition to these large subunits there are a number of smaller 'core' subunits which are presumed to be involved in the functional and structural roles of this complex. One of these 'core' subunits (8 kDa) has recently been identified as an iron-sulphur protein and its gene has been located and sequenced in the chloroplast DNA of tobacco [12], pea [13] and wheat [13]. Little is known about the function of the other 'core' subunits.

As an important step in the characterisation of photosystem I we have electrophoretically purified each of the six smaller 'core' subunits from a well resolved pea complex and subjected these to gas-phase or solid-phase sequence analysis.

2. MATERIALS AND METHODS

Photosystem I was purified from washed thylakoid mem-

Correspondence address: P.P.J. Dunn, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, England

Abbreviations: DITC, *p*-phenylenediisothiocyanate; PTH, phenylthiohydantoin

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

branes isolated from 200 g of 10 day old pea leaves. Thylakoid membranes were depleted of ATP synthase and cytochrome *b–f* complex by treatment with 1% (w/v) octyl β-D-glucopyranoside and 0.5% (w/v) sodium cholate at 0°C [14]. Photosystem I was solubilised in 25 mM Tricine·KOH, pH 7.8, 2% (v/v) Triton X-100 at a chlorophyll concentration of 0.5–1.0 mg/ml and then applied to a 3 × 12 cm column of DEAE-Sephacel (in 25 mM Tricine·KOH, pH 7.8, 0.2% (v/v) Triton X-100). The column was washed with one column volume of equilibration buffer. A fraction enriched in photosystem I was eluted using equilibration buffer containing 0.2 M NaCl at a flow rate of 1 ml/min. The photosystem I-enriched sample was then gel filtered using a 3.5 × 90 cm column of Ultrogel AcA34 equilibrated in 25 mM Tricine·KOH, pH 7.8, 50 mM NaCl, 0.1% (v/v) Triton X-100. The complex was isolated at a flow rate of 25 ml/h.

Protein was analysed by SDS-polyacrylamide gel electrophoresis by the method of Fling and Gregerson [15] except that the resolving gel contained acrylamide and urea gradients of 12.5–25% and 2–4 M, respectively. The stacking gel contained 2 M urea. Høj et al. [16] have already shown that the method of Fling and Gregerson [15] resolves the minor subunits of photosystem I.

For electroelution, polyacrylamide gels were stained in 0.5% (w/v) Coomassie brilliant blue R in double-distilled water for 15 min and then briefly washed in water. Bands of interest were excised with a razor blade and electroeluted overnight using 25 mM Tris, 92 mM glycine and 0.1% SDS as electrophoresis buffer. Electroeluted proteins were then dialysed against 5 l of 0.02% SDS with 4–5 changes. Proteins were lyophilised prior to preparation for N-terminal sequence analysis.

Electroeluted proteins were sequenced either by automated solid-phase or gas-phase Edman degradation. For solid-phase sequencing lyophilised peptide was dissolved in 0.05 ml of 0.2 M $NaHCO_3$ containing 0.25% (w/v) SDS and added to 15 mg DITC glass (170 Å pore-size, 200–400 mesh) prepared by a modification of the method of Wachter et al. [17]. The glass was incubated at 56°C for 60 min, then washed with water and methanol containing 0.5% (v/v) *n*-propylamine to remove non-covalently bound material. Approx. 5 mg of the glass-coupled peptide was then sequenced by automated solid-phase Edman degradation [18] using the microsequencing facility of the AFRC Sequencing Laboratory, Department of Biochemistry, Leeds.

Samples were prepared for gas sequence analysis as follows. The protein sample, containing residual SDS and Coomassie brilliant blue was dissolved in 40 μl H_2O. The protein was precipiated with 10 vols of 96% ethanol at −70°C for 20 min and recovered by centrifugation. The procedure was then repeated on the resuspended protein. This process removed SDS, unbound Coomassie brilliant blue and UV-absorbing contaminants which may interfere with PTH-amino acid analysis. The protein sample was dried under vacuum, redissolved in trifluoroacetic acid (40 μl) and a portion applied to a preconditioned glass fibre disc containing 3 mg of Biobrene plus. Each sample was treated with iodoacetic acid to modify any cysteine residues present (Packman, L., unpublished). Sequence analysis was performed, using RUN 470-1 software, on an Applied Biosystems 470 A gas-phase sequencer coupled to a 120A PTH-amino acid analyser under the control of a 900A data controller at the Protein Sequencing Facility, University of Cambridge.

3. RESULTS AND DISCUSSION

Photosystem I isolated from pea chloroplasts as described in section 2 and analysed by the electrophoretic system of Fling and Gregerson [15] shows seven well resolved protein bands (fig.1). Fig.1 also shows minor contamination from ATP synthase subunits (≈55 kDa) and from subunits of the light harvesting complex of photosystem I (≈29–26 kDa). In addition to the larger polypeptides of 62 and 60 kDa, which are not resolved in this gel system, there are six 'core' subunits of 21, 17, 13, 11, 9 and 8 kDa. These values are apparent molecular masses determined from a single percentage (15%) polyacrylamide gel. Fig.1 shows that the smaller subunits (<12 kDa) are sufficiently separated to enable electroelution as described in section 2.

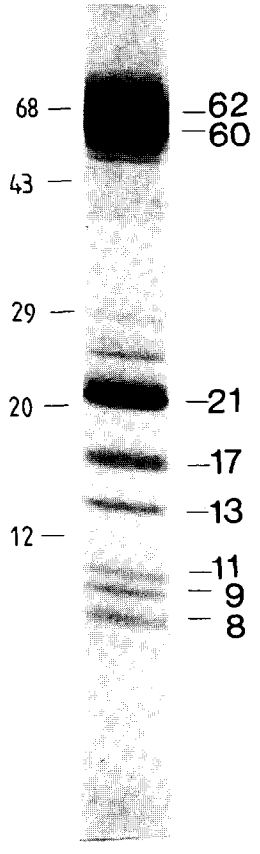

Fig.1. Pea photosystem I complex isolated as described in section 2 and analysed on a polyacrylamide gel containing SDS and urea. Values on the left refer to molecular mass markers in kDa and values on the right refer to subunit sizes in kDa.

The N-terminal amino acid sequence determined for each 'core' subunit of pea photosystem I is shown in fig.2. Databases of protein sequences have been scanned for these sequences but no significant homologies were detected. Therefore, with the exception of the 8 kDa subunit (the gene of which has recently been sequenced [12,13]), the amino acid sequences shown in fig.2 are unique.

The 21 kDa subunit was sequenced using a solid-phase sequencer. The first amino acid was not identified. Thirty-one amino acids were successfully identified from thirty-six cycles. This polypeptide appears to be rich in charged residues (four aspartic acids, three glutamic acids and one lysine) and polar residues (six threonines and one serine). The pea 21 kDa subunit may correspond to the 22 kDa subunit of spinach which is supposed to mediate the transfer of electrons from plastocyanin [19]. It may also be equivalent to the 20 kDa nuclear-encoded subunit of pea which is synthesized in the cytoplasm as a 26 kDa precursor [3].

Twenty amino acids were determined from the N-terminus of the 17 kDa subunit using a gas-phase sequencer. Both alanine and aspartic acid gave high yields at the first cycle and both yields

```
                                                    10
21 kDa      X  Thr Glu Asp Lys Thr Asp Ala Ala Thr Asp Val Ala Thr  X  Glu Ala
                  20                                          30
           Pro Val Gly Phe Thr Pro  X  Glu Leu Asp Pro Asn Thr  X  Ser  X  Ile

           Phe Gly

                                                    10
17 kDa Asp/Ala Ile Ser Gly Leu Thr Pro Cys Lys Glu Ser Lys Gln Phe Ala Lys Arg
                  20
           Glu Lys Gln

                                                    10
13 kDa      X  Ser Glu Asp Thr Ala Glu Ala Ala  X  Pro Ser Ala
                                                    10
13 kDa     Ala Ser Glu Asp Thr Ala Glu Ala Ala Ala Pro

                                                    10
11 kDa     Lys Tyr Gly Asp Lys Ser Val Tyr Phe Asp Leu Glu Asp Ile Gly Asn Thr
                  20                                          30
           Thr Gly Gln Trp Asp Leu Tyr Gly Ser Asp Ala Pro Ser Pro Tyr Ser  X

           Leu Gln

                                                    10
 9 kDa     Glu Leu Asn Pro Ser Leu Val Ile Ser Leu Ser Thr Gly Leu Ser Leu Phe
                  20                                          30
           Leu Gly Arg Phe Val Phe Phe Asn Phe Gln Arg Glu Asn Val Ala Lys Gln

           Gly Leu Pro Glu Gln

                                                    10
 8 kDa     Ser His Ser Val Lys Ile Tyr Asp Thr Cys Ile Gly Cys Thr Gln Cys Val
                  20                                          30
           Arg Ala Cys Pro Thr Asp Val Leu Glu Met Ile Pro Trp Gly Gly Cys Lys
                             40                                           50
           Ala Lys Gln Ile Ala Ser Ala Pro Arg  X  Glu Asp  X  Val  X   X   K
```

Fig.2. N-terminal amino acid sequences of pea photosystem I. The sequences of the 21 kDa subunit and the first 13 kDa subunit sequence were obtained by solid-phase sequencing. All other subunits were sequenced using gas-phase chemistry. Initial yields and repetitive yields for each polypeptide were: 21 kDa, 80 pmol, 92.5%; 17 kDa, 125 pmol, 93.8%; 13 kDa (solid phase), 106 pmol, 94.1%, (gas phase) 1 nmol, 93.8%; 11 kDa, 475 pmol, 92.4%; 9 kDa, 525 pmol, 93.7%; 8 kDa, 457 pmol, 94.6%. X indicates positions at which no clear identification of a single phenylthiohydantoin could be made. Only firm assignments are shown.

decreased at the second cycle. The lag data favour alanine but yields are such that aspartic acid (which can arise artefactually from contaminating ammonia) cannot be ruled out. Similarly, microheterogeneity cannot be excluded as an explanation for two N-terminal amino acids. Thus, the first residue of the 17 kDa subunit has been given as alanine or aspartic acid in fig.2. Nine of the first 20 N-terminal amino acids are charged residues (if aspartic acid at residue number 1), five are positively charged (four lysine and one arginine) and three are negative (two glutamic acids and one aspartic acid). The N-terminus of this polypeptide is highly charged indicating that it is likely to be exposed to an aqueous environment.

The 13 kDa subunit of pea photosystem I is alanine-rich at the N-terminus. Six of the 13 amino acids determined are alanine. This polypeptide was sequenced using a gas-phase sequencer and a solid-phase sequencer and both sequences obtained are shown (fig.2). Both methods revealed the presence of a second sequence (approx. 50% abundant) representing the same protein lacking the N-terminal amino acid. This suggests that the 13 kDa subunit may undergo a degree of N-terminal processing which may be important, for example, for the correct insertion of the mature polypeptide in the membrane in vivo.

Gas-phase sequence analysis of the 11 kDa subunit revealed that this polypeptide, like the 21 and 17 kDa subunits, contains a significant number of charged residues. There are 6 acidic and two basic residues in the 35 amino acids sequenced which suggests that the N-terminus of this subunit also protrudes from the thylakoid membrane and is not involved in forming a membrane span.

Thirty-nine amino acids from the N-terminus of the 9 kDa subunit were sequenced. There are 16 hydrophobic amino acids (i.e. one isoleucine, three valines, five phenylalanines and seven leucines) in the first 36 residues. Between residues 21 and 26 there is an unusual phenylalanine-rich sequence: Phe-Val-Phe-Phe-Asn-Phe. The presence of such a hydrophobic N-terminus suggests that this region of the polypeptide may be membrane-located.

N-terminal amino acid sequence analysis of the 8 kDa subunit revealed a polypeptide rich in cysteine residues (fig.2). Five cysteine residues were identified in 51 cycles. The protein sequence shown in fig.2 is identical to the 29 N-terminal residues determined for an 8 kDa spinach photosystem I subunit [20] and identical to the 29 N-terminal amino acids of a 9 kDa barley photosystem I subunit [16]. This indicates a functional role for this protein and we have recently suggested that the 8 kDa subunit of pea photosystem I carries the two iron sulphur centres A and B involved in electron transfer to ferredoxin [13]. The gene for the 8 kDa subunit has been located and sequenced in pea and wheat chloroplast DNA [13] and the predicted amino acid sequence from this gene (which is identical with the determined sequence in fig.2) reveals a total of 9 cysteine residues, 8 of which are arranged like bacterial 2(4Fe-4S) ferredoxins [21].

These N-terminal sequences have been compared to proteins predicted from open reading frames located in tobacco chloroplast DNA [22], but, with the exception of the 8 kDa polypeptide, no matches were obtained. This suggests that the genes for these other subunits are located in nuclear DNA. The sequence data presented in this paper will be useful for identifying genes for photosystem I subunits and for indicating the extent of any presequences necessary to direct the proteins to the chloroplast.

Acknowledgements: This work was supported by grants from the AFRC and the Wellcome Trust.

REFERENCES

[1] Mullet, J.E., Burke, J.J. and Arntzen, C.J. (1980) Plant Physiol. 65, 814–822.
[2] Smith, A.G. and Gray, J.C. (1984) Mol. Gen. Genet. 194, 471–476.
[3] Mullet, J.E., Grossman, A.R. and Chua, N.-H. (1982) Cold Spring Harbor Symp. Quant. Biol. 46, 979–984.
[4] Westhoff, P., Alt, J., Nelson, N., Bottomley, W., Bunneman, H. and Hermann, R.G. (1983) Plant Mol. Biol. 2, 95–107.
[5] Høj, P.B. and Møller, B.L. (1986) J. Biol. Chem. 261, 14292–14300.
[6] Obokata, J. (1986) Plant Physiol. 81, 705–707.
[7] Møller, B.L., Høyer-Hansen, G. and Hiller, R.G. (1981) Photosynth., Proc. 5th Int. Congr. vol.iii, 245–256.
[8] Malkin, R. (1982) Annu. Rev. Plant Physiol. 33, 455–479.
[9] Lehmbeck, J., Rasmussen, O.F., Bookjans, G.B., Jepsen, B.R., Stummann, B.M. and Henningsen, K.W. (1986) Plant Mol. Biol. 7, 3–10.
[10] Fish, L.E., Kuck, U. and Bogorad, L. (1985) J. Biol. Chem. 260, 1413–1421.
[11] Kirsch, W., Seyer, P. and Hermann, R.G. (1986) Curr. Genet. 10, 843–855.

[12] Hayashida, N., Matsubayashi, T., Shinozaki, K., Suguira, M., Inoue, K. and Hiyama, T. (1987) Curr. Genet. 12, 247–250.
[13] Dunn, P.P.J. and Gray, J.C. (1987) Plant Mol. Biol., submitted.
[14] Nechustai, R. and Nelson, N. (1981) J. Biol. Chem. 256, 11624–11628.
[15] Fling, S.P. and Gregerson, D.S. (1986) Anal. Biochem. 155, 83–88.
[16] Høj, P.B., Svendsen, I., Scheller, H.V. and Møller, B.L. (1987) J. Biol. Chem. 262, 12676–12684.
[17] Wachter, E., Machleidt, W., Hofner, H. and Otto, J. (1973) FEBS Lett. 35, 97–102.
[18] Laursen, R.A. (1971) Eur. J. Biochem. 20, 89–102.
[19] Bengis, C. and Nelson, N. (1977) J. Biol. Chem. 252, 4564–4569.
[20] Oh-oka, H., Takahashi, Y., Wada, K., Matsubara, H., Ohyama, K. and Ozeki, H. (1987) FEBS Lett. 218, 52–54.
[21] Adman, E.T., Sieker, L.C. and Jensen, L.H. (1973) J. Biol. Chem. 248, 3987–3996.
[22] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinosazi, 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 Suguira, M. (1986) EMBO J. 5, 2043–2049.