Volume 237, number 1,2, 218-224 FEB 06307 September 1988

Nucleotide sequence of cDNA clones encoding the entire precursor polypeptides for subunits IV and V of the photosystem I reaction center from spinach

J. Steppuhn, J. Hermans, R. Nechushtai*, Ulf Ljungberg, F. Thümmler, F. Lottspeich+ and R.G. Herrmann

*Botanisches Institut der Ludwig-Maximilians-Universität, Menzinger Straße 67, D-8000 München 19, FRG, *Botany Department, The Hebrew University, Jerusalem 91904, Israel* and +*Max-Planck-Institut für Biochemie, Am Klopferspitz, 8033 Martinsried, FRG*

Received 22 July 1988

Using λgt11 expression cloning and immunoscreening, cDNA-containing recombinant phages for subunits IV and V of the photosystem I reaction center were isolated, sequenced and used to probe Northern blots of polyadenylated RNA prepared from spinach seedlings. The mRNA sizes for both components are ~1000 and 850 nucleotides, respectively. The 968 nucleotide cDNA sequence and derived amino acid sequence for subunit IV predict a single open reading frame of 231 amino acid residues (25.4 kDa). Comparison with a 13-residue N-terminal amino acid sequence determined for subunit IV suggests a mature protein of 17.3 kDa (154 residues) and a transit sequence of 77 amino acids (8.1 kDa). The corresponding data for subunit V are 677 bp (cDNA), 167 residues for the precursor protein (18.2 kDa), 98 residues for the mature polypeptide (10.8 kDa) and 69 residues for the transit peptide (7.4 kDa). Secondary structure predictions indicate that both proteins possess greatly different transit sequences and that none is membrane-spanning.

Photosynthesis; Photosystem I; Subunit; cDNA; Nucleotide sequence; (Spinach)

1. INTRODUCTION

The primary light reactions in oxygenic photosynthesis are catalysed by two multisubunit polypeptide complexes of the thylakoid membrane, photosystem I and the oxygen-evolving photosystem II, which act in series and are associated with light-harvesting antenna. Photosysem I (PS I) refers to a chlorophyll *a*-protein complex which is functionally defined by its ability to catalyse the light-mediated transfer of electrons from reduced plastocyanin to ferredoxin. The complete electron path in the reaction center has not been fully settled but includes a series of acceptors, namely A_0 (possibly a special chlorophyll *a* pigment), A_1 (probably a phylloquinone [vitamin K_1] molecule) and the iron-sulfur centers X, A and B (e.g. [1,2]). PS I particles, with and without attached antenna (LHCI), have been prepared from a variety of lower and higher organisms (e.g. [3–6]). Their protein complement has been resolved into up to 13 components by SDS slab gel electrophoresis, including four LHCI polypeptides in the molecular mass range of 21.0–24.5 kDa [4,6], two chlorophyll *a*-binding polypeptides (subunits Ia, b) of 82–84 kDa that complex the P-700 pigment involved in the primary light reaction, and possibly the primary and secondary acceptors A_0, A_1 and X [1, 3,7] as well as low molecular mass subunits. The exact number of the latter subunits varies considerably from laboratory to laboratory and probably depends upon preparation, plant source, and possibly cell type. In the PS I reaction center core of spinach, up to six small polypeptides have con-

Correspondence address: R. Herrmann, Botanisches Institut der Ludwig-Maximilians-Universität, Menzinger Straße 67, D-8000 München 19, FRG

The nucleotide sequences presented here have been submitted to the EMBL/GenBank database under the accession no. Y00967

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

sistently been observed with molecular masses of 22, 18.5, 18, 16, 12 and 10 kDa, designated subunits II–VII [5,6]. Little is known about their function and spatial arrangement. Their selective depletion abolishes NADP photoreduction [1,8]. The [4Fe-4S] centers A and B that serve in photosynthetic electron transport from the reaction center to ferredoxin have been assigned to subunit VII [9]. Subunit II appears to be involved in docking both ferredoxin and ferredoxin oxidoreductase [10,11]; subunit III may be implicated in the interaction between plastocyanin and the P-700 pigment [8]. The roles of subunits IV–VI are unknown.

Information about the primary structures of subunits, deduced from the investigation of their genes, will ultimately need to be known for understanding this photosystem. As other thylakoid membrane assemblies, the PS I core has dual genetic origin implying that its biogenesis involves both nuclear and plastid control. While genes for three of the probably four [12,13] plastome-encoded components, namely subunits Ia, b, and subunit VII, have been mapped on plastid chromosomes of various plants and sequenced [5,9,14–17], the analysis of nuclear components lags behind (cf. [18]).

3 years ago, we initiated a systematic study of the nuclear-encoded complement for thylakoid membrane polypeptides including those for PS I ([19] and unpublished) in order to contribute to further understanding and refinement of models for this reaction center, its biogenesis, and evolution. We have cloned and characterized cDNAs for five nuclear-encoded core components (subunits II–VI) and for two LHCI polypeptides of this membrane complex from spinach ([19] and unpublished). Here, we report the analysis of cDNA clones that encode for the complete precursor proteins for subunits IV and V. Equivalent data from recombinant DNAs that carry information for other PS I subunits is presented elsewhere.

2. MATERIALS AND METHODS

2.1. *Library screening*

The amplified λgt11 library made from polyadenylated RNA of spinach seedlings was screened immunologically [19] and subsequently with ^{32}P-labelled speficic cDNA probes as described [20]. Positive plaques were purified by successive plating and phage DNA was prepared by routine procedures as collected [21].

2.2. *DNA sequencing*

All restriction fragments were recloned into M13mp19 and Bluescript M13$^+$ vectors for sequencing. Subclones for further sequencing analysis were generated by (i) restriction of the p6SocPI4-7 insert with *Hae*III and cloning of the resulting subfragments into the *Eco*RI/*Sma*I or the *Sma*I sites of M13mp18 and M13mp19; (ii) restriction of the p6SocPI5-7 insert with *Dde*I, S_1 nuclease treatment and subsequent cloning of the subfragments into the *Sma*I site of M13mp18 (fig.1). Additional sequencing was performed on 'sequential deletions', i.e. on independent cDNA clones with inserts varying in size that were selected during library screening. Sequence determination was carried out according to [22], and the data were analysed on an IBM-AT computer with PCGENE programs compiled by Amos Broich (Genofit, Geneva).

2.3. *Purification of PS I reaction center; protein sequencing*

The PS I reaction center was isolated from spinach thylakoid membranes as in [5]. The purified membrane complex was separated on 10–15% SDS-polyacrylamide gels, blotted onto modified glass fibers (Biometra, Göttingen), stained with Coomassie brilliant blue and the zone corresponding to subunit IV was excised for sequence analysis. Protein sequencing was performed on a gas-phase protein sequencer (Applied Biosystems model 470A). The released derivative amino acids were analysed as described [23].

3. RESULTS AND DISCUSSION

3.1. *Clone selection*

Six immunopositive phages for both subunit IV and V of the PS I reaction center were isolated from a λgt11 library [19] after screening approx. 1.5×10^5 plaque-forming units. Their DNA was digested with *Eco*RI. The resulting fragments were sized on 1.2% agarose slab gels and found to contain inserts in the range of 450–700 nucleotides. These inserts were subcloned into M13mp18 and Bluescript M13$^+$ vectors, designated p6SocPI4-1 to -6 and p6SocPI5-1 to -6 (Bluescript vectors), respectively, and sequenced. RNA copies of the largest clones (p6SocPI4-1, p6SocPI5-1) obtained with T_3 RNA polymerase [24,25] after linearization of the respective plasmid DNAs with *Bam*HI were used to probe polyadenylated RNA from spinach seedlings. In each instance, the probe hybridized to a single mRNA species of 1000 (subunit IV) and 850 (subunit V) nucleotides when polyadenylated RNA was prepared from material illuminated for 12–16 h after etiolation (not shown). The same RNA copies were used as probes to screen the library for recombinant phage that bear the entire protein-coding regions for subunits IV or V. In each instance, positive signals for 14 plaques

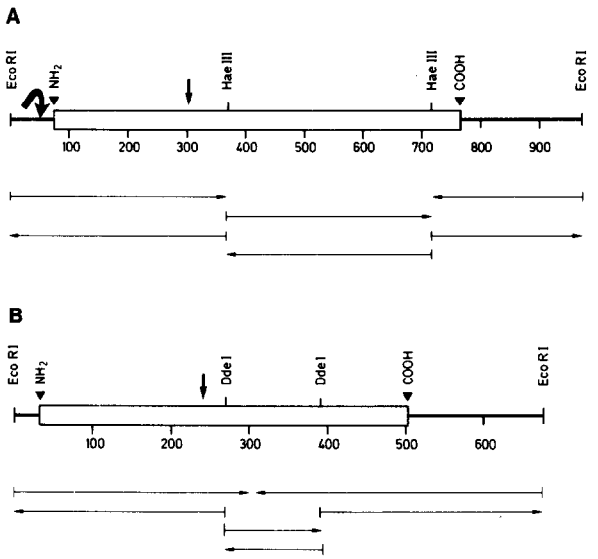

Fig.1. Partial restriction map and sequencing strategy for p6SocPI4-7 cDNA (A) and p6SocPI5-7 (B) from spinach. The coding sites for PS I subunit IV (A) and V (B) are boxed. Cleavage sites for relevant restriction endonucleases, N- and C-termini (NH_2, COOH) and the putative terminal processing sites of the respective transit sequences (arrow) are marked. Horizontal arrows indicate the direction and extent of individual DNA sequencing reactions. The scales are in bp. Nucleotides −30 to −71 of p6SocPI4-7 (curved arrow) are disregarded in fig.2 for they are complementary to the sequence interval 461-502 (cf. fig.2A) and probably a cloning artefact that is quite frequently observed with 'full-length' cDNAs (cf. [25,26]).

resulted from a screening of 60000 clones. Five (subunit IV) and three (subunit V) of the 14 recombinant phage that contained inserts approx. 200–250 bp larger than p6SocPI4-1 or p6SocPI5-1 were recloned to yield plasmids p6SocPI4-7 to p6SocPI4-11 and p6SocPI5-7 to p6SocPI5-9, and chosen for further analysis (fig.1).

3.2. *Sequence analysis*

The sequences of the largest inserts are detailed in fig.2. Identification of translational start and stop codons by computer analysis of the p6SocPI4-7 sequence predicts a single open reading frame from base position 77 to position 769, beginning with the first possible in-phase methionine residue and terminating at the phenylalanine residue before the predicted opal stop codon. This reading frame encodes a protein of 25.4 kDa (231 amino acid residues) and is followed by a 3′-untranslated region of at most 210 bp which contains no typical consensus polyadenylation element (AATAAA [27]) but a poly(A) tract varying in size (10–30 nucleotides) between individual clones. Two clones contained these tracts at different locations, with p6SocPI4-1 and -7 representing longer RNA species (fig.2A). The remaining sequences were identical. It is possible, therefore, that either side is used and that the cDNA inserts represent mRNA copies of one gene (*psa*F).

The equivalent data for p6SocPI5-7 DNA that bears information for subunit V are presented in fig.2B. It consists of 677 bp followed by a poly(A) tail of 16 residues. The longest open reading frame (501 bp) codes for a protein of 167 residues (18.2 kDa). It is preceded by a 33 bp 5′- and 127 bp 3′-untranslated region. The latter includes a potential polyadenylation signal. No nucleotide difference was found between individual clones that were sequenced. We designate this gene *psa*G.

To assess the nature of the cDNAs and to determine the junction between the presumed transit sequence and mature protein, PS I reaction center (free of LHCI) was purified from spinach thylakoids [5,6]. Its subunits were separated on 10–15% SDS polyacrylamide gels, blotted onto glass fiber [23], and subunit IV was partially sequenced. The N-terminal sequence of 13 residues fits precisely within the open reading frame shown in fig.2, substantiating the identity of the cDNA. The deduced amino acid sequences for both proteins were also compared with the published N-terminal sequences for PS I polypeptides from pea [28]. The derived subunit V sequence for spinach corresponding to the nucleotide segment 208–321 matches the 39 N-terminal residues determined for the pea protein. The comparable partial sequences for both proteins are conserved to a remarkably high degree. Only a conservative substitution, Ser 3 to Ala 3 is found in spinach subunit IV, and four changes are noted in subunit V, the replacements of Ser 3 by Asn 3, of Val 31 by Met 31, of Leu 36 by Val 35 in spinach, and the addition of Gly 35 in pea. The amino terminus of mature subunit IV corresponds to the 78th codon, GAC (Asp), at nucleotide position 262–264. The remaining downstream open reading frame encodes a polypeptide of 154 amino acids (17.3 kDa). The corresponding data for subunit V are GAG (Glu, 70th codon, nucleotide position 208–210) and 98 residues (12.2 kDa) for the mature protein. Estimates of the

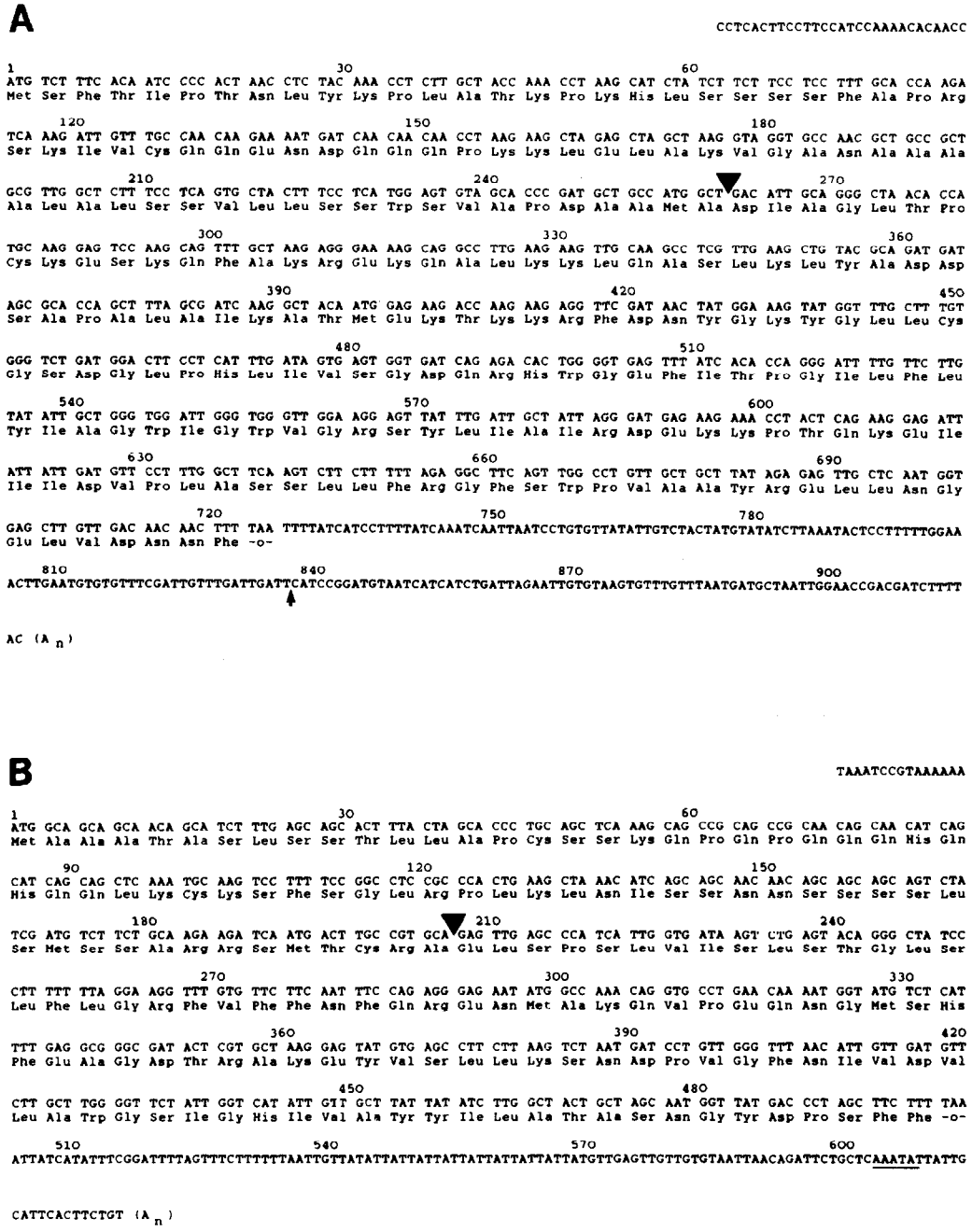

Fig.2. Nucleotide sequence and derived amino acid sequence of P6SocPI4-7 DNA (A) and p6SocPI5-7 DNA (B) from spinach. The putative junction between transit peptide and mature protein is marked by an arrowhead. A potential polyadenylation signal is underlined (B). The N-terminal amino acid sequence, determined from purified subunit IV protein spans amino acid residues 78–90. The additional polyadenylation site is indicated by an arrow (A).

molecular mass of the mature proteins based on SDS-polyacrylamide gel electrophoresis, 18 and 12 kDa [5], agree well with these values.

3.3. *Transit sequences*

The 77 and 69 amino acid residues preceding the N-terminal sequence of subunits IV and V respectively, may constitute the transitory transit sequences that are cleaved during or after import into the organelle and possibly thylakoid membrane. These presumed presequences are relatively long, similar in size to transit sequences that direct polypeptides into the thylakoid lumen. Although there is almost no homology in primary sequence to any other transit sequence for nucleus-encoded chloroplast proteins [29], the presequences do conform to the general characteristics of sequences directing polypeptides into chloroplasts (e.g. [30]). They are remarkably polar and rich in hydroxylated as well as positively charged residues (8 positive, 3 negative charges in subunit IV; 8 positive charges in subunit V). This reinforces previous suggestions [25,31] that secondary rather than primary structure accounts for chloroplast recognition and intraorganelle routing. One of the striking findings of this study is the marked difference of the predicted secondary structure for both transit sequences (fig.3). While the subunit IV transit sequence shares characteristics with transit sequences of extrinsic luminal proteins such as plastocyanin [32] and components of the water oxidation complex [25,31], notably a C-terminal hydrophobic domain and a preceding amphipathic β-sheet (Lys-Val-Gly-Ala-Asn-Ala, residues 49–54) that for luminal precursor polypeptides can include the intermediary processing site (unpublished), that of subunit V is hydrophilic and resembles transit sequences of the 10 kDa polypeptide of the PS II reaction center or of stromal enzymes such as small subunit of ribulose-bisphosphate carboxylase/oxygenase, ferredoxin, ferredoxin oxidoreductase or thioredoxin f (cf. [29]). This could reflect intraorganelle routing differences (see below). It will therefore be interesting to determine whether the endoproteolytic maturation of both subunits differs, i.e. that of subunit IV involves at least two steps. PS I and PS II assemblies are located in different, appressed and non-appressed regions of the thylakoid membrane, respectively [33]. An important and not yet addressed facet of protein import results from this lateral heterogeneity. It should be interesting therefore to study the routes of these proteins, and to delineate intraorganelle sorting signals, if any, between these two polypeptide classes. Apart from a highly hydrophilic domain shared by all transit sequences of PS I subunits (fig.3) there are no other notable differences between target sequences of PS I and PS II polypeptides.

3.4. *Protein structure predictions*

Various algorithms [34–36] were applied to estimate the secondary structures for both subunits. An unexpected result of this analysis was that subunit V and, possibly, subunit IV are probably not integral membrane proteins (fig.3). Although each of the polypeptides displays two segments up to 19 residues that are rich in amino acids with apolar side chains, mainly helical, the preferred structure of transmembrane segments, and some of them are bounded by charged residues (e.g. subunit IV, residues 82 and 102; subunit V, residues 1 and 20), their hydrophobic moments are rather low so that they might not serve to hold the polypeptide chains by transmembrane anchoring (a possible exception could be domain I in subunit IV). Both proteins, as subunit II [18], III (unpublished) and VII [9], may therefore reside atop of the membrane, exposed to the aqueous phase. However, they should be buried within the PS I assembly, since they are quite well protected from protease digestion and relatively insensitive to extraction with chaotropic salts or low concentrations of detergent [3–5]. This arrangement provides also a reasonable explanation for the different subunit compositions reported for PS I complexes.

The intraorganelle position of subunits IV and V remains ambiguous. Differential extraction of thylakoid membranes has not yet led to conclusive results. In addition, small subunits may change their relative migration velocities in different gel systems [5,6] which precludes an unequivocal correlation of the polypeptides defined by sequence and by electrophoretic mobility. Nearest-neighbour analysis has suggested that at least two subunits, namely subunit III and one of the smaller components, are located at the oxidizing (luminal) side, in close proximity to the subunit I pair [8]. If the anatomy of transit sequences is indicative of suborganelle targeting (see above), one would ex-

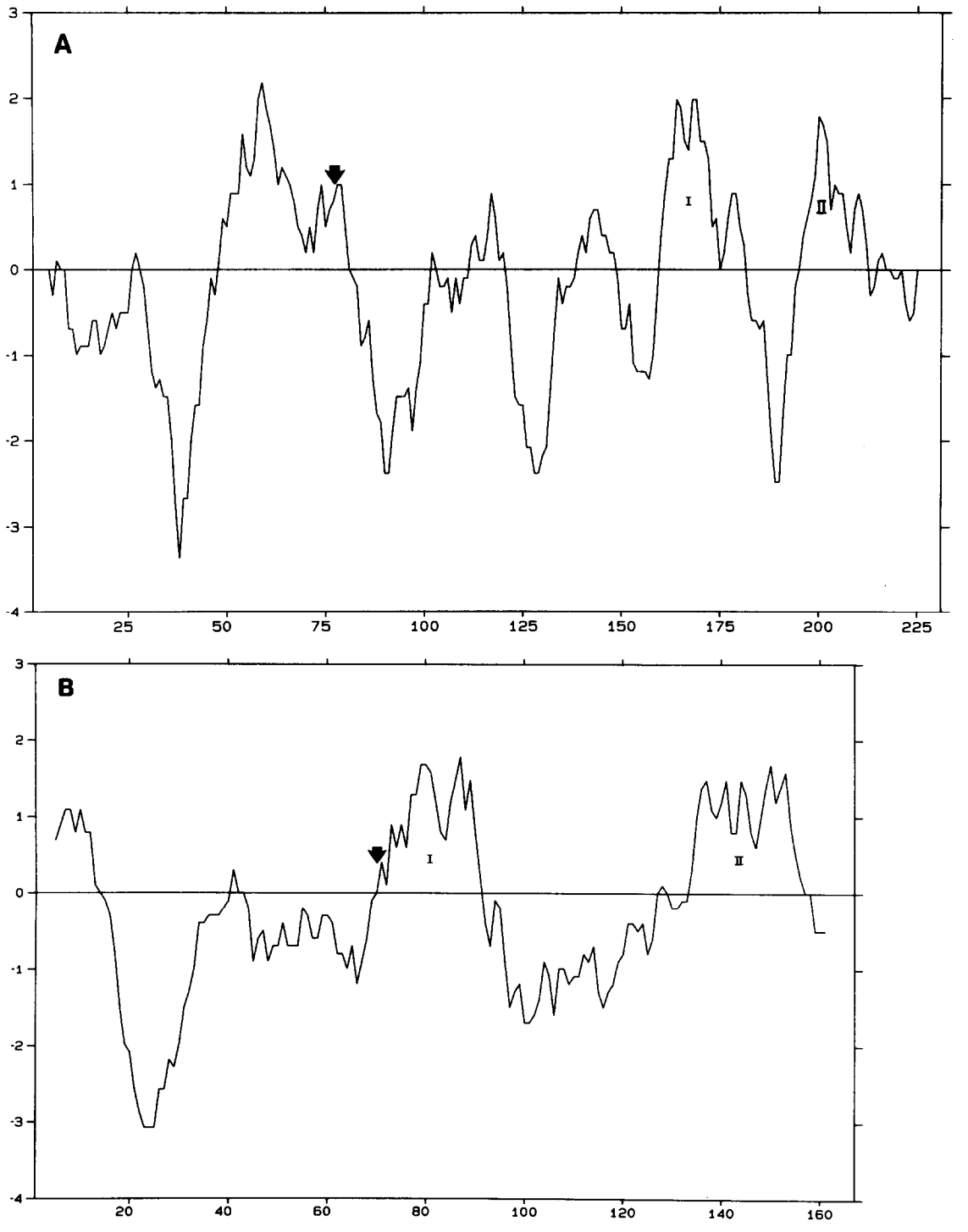

Fig.3. Hydropathy blots of the precursor polypeptides of subunits IV (A) and V (B) of the PS I reaction center from spinach. The profiles were produced from the algorithms of [36] with a window average across 11 residues. The putative cleavage site between transit peptide and mature protein is indicated by an arrow; predicted hydrophobic domains are marked by Roman numbers (see text).

pect subunit IV to be located at the luminal, and subunit V at the stromal side of the reaction center. Obviously, further studies are required to settle this point, and the role of each subunit in PS I activity. Both proteins possess a relatively balanced ratio of basic and acidic residues. Their abundance (>20% in subunit IV) and partial clustering as well as the absence (subunit V) or presence of only two (widely spaced) Cys residues (subunit IV) suggest that both proteins fulfil primarily structural roles and are not directly involved in binding Fe or in electron transport. We are confident that the recombinant DNAs now available will be of use to study structure and functions of these polypeptides as well as their suborganelle locations by import assays, fractionation and protease-sensitivity techniques (e.g. [37]).

Acknowledgements: The technical assistance of Ms Christa Jansen is gratefully acknowledged. The work was supported by the German Research Foundation (grant He 623/12-2; SFB 184) and the Fonds der Chemischen Industrie.

REFERENCES

[1] Malkin, R. (1982) Annu. Rev. Plant Physiol. 33, 455–479.
[2] Rutherford, A.W. and Heathcote, P. (1985) Photosynth. Res. 6, 295–316.
[3] Bengis, C. and Nelson, N. (1975) J. Biol. Chem. 250, 2783–2788.
[4] Mullet, J.E., Burke, J.J. and Arntzen, C.J. (1980) Plant Physiol. 65, 814–822.
[5] Westhoff, P., Alt, J., Nelson, N., Bottomley, W., Bünemann, H. and Herrmann, R.G. (1983) Plant Mol. Biol. 2, 95–107.
[6] Herrmann, R.G., Westhoff, P., Alt, J., Winter, P., Tittgen, J., Bisanz, C., Sears, B.B., Nelson, N., Hurt, E., Hauska, G., Viebrock, A. and Sebald, W. (1983) in: Structure and Function of Plant Genome (Ciferri, O. and Dure, L. iii eds) pp. 143–153, Plenum, New York.
[7] Høj, P.B. and Møller, B.L. (1986) J. Biol. Chem. 261, 14292–14300.
[8] Bengis, C. and Nelson, N. (1977) J. Biol. Chem. 252, 4564–4569.
[9] Høj, P.B., Svendsen, I., Scheller, H.V. and Møller, B.L. (1987) J. Biol. Chem. 262, 12676–12684.
[10] Zanetti, G. and Merati, G. (1987) Eur. J. Biochem. 169, 143–146.
[11] Zanetti, G., Morelli, D., Ronchi, S., Negri, A., Aliverti, A. and Curti, B. (1988) Biochemistry 27, 3753–3759.
[12] Nechustai, R., Nelson, N., Mattoo, A.K. and Edelman, M. (1981) FEBS Lett. 125, 115–119.
[13] Mullet, J.E., Grossman, A.R. and Chua, N.H. (1981) Cold Spring Harbor Symp. Quant. Biol. 46, 979–984.
[14] Fish, L.E., Kuck, U. and Bogorad, L. (1985) J. Biol. Chem. 260, 1413–1421.
[15] Kirsch, W., Seyer, P. and Herrmann, R.G. (1986) Curr. Genet. 10, 843–855.
[16] Hayashida, N., Matsubayashi, T., Shinozaki, K., Sugiura, M., Inoue, K. and Hiyama, T. (1987) Curr. Genet. 12, 247–250.
[17] Oh-oka, H., Takahashi, Y., Wada, K., Matsubara, H., Ohyama, K. and Ozeki, H. (1987) FEBS Lett. 218, 52–54.
[18] Hofmann, N.E., Pichersky, E., Malik, V.S., Ko, K. and Cashmore, A.R. (1988) Plant Mol. Biol. 10, 435–445.
[19] Tittgen, J., Hermans, J., Steppuhn, J., Jansen, T., Jansson, C., Andersson, B., Nechushtai, R., Nelson, N. and Herrmann, R.G. (1986) Mol. Gen. Genet. 204, 258–265.
[20] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[21] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY.
[22] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[23] Eckerskorn, C., Mewes, W., Goretzki, H. and Lottspeich, F. (1988) Eur. J. Biochem., in press.
[24] Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucleic Acids Res. 12, 7035–7056.
[25] Tyagi, A., Hermans, J., Steppuhn, J., Jansson, C., Vater, F. and Herrmann, R.G. (1987) Mol. Gen. Genet. 207, 288–293.
[26] Smeekens, S., Van Binsbergen, J. and Weisbeek, P. (1985) Nucleic Acids Res. 13, 3179–3194.
[27] Proudfoot, N. (1982) Nature 298, 516–517.
[28] Dunn, P.P.J., Packman, L.C., Pappin, D. and Gray, J.C. (1988) FEBS Lett. 228, 157–161.
[29] Von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1988) EMBO J., submitted.
[30] Schmidt, G.W. and Mishkind, M.L. (1986) Annu. Rev. Biochem. 55, 879–912.
[31] Jansen, T., Rother, C., Steppuhn, J., Reinke, H., Beyreuther, K., Jansson, C., Andersson, B. and Herrmann, R.G. (1987) FEBS Lett. 216, 234–240.
[32] Rother, C., Jansen, T., Tyagi, A., Tittgen, J. and Herrmann, R.G. (1986) Curr. Genet. 11, 171–176.
[33] Anderson, J.M. and Andersson, B. (1982) Trends Biochem. Sci. 7, 288–292.
[34] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[35] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[36] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[37] Bartlett, S.G., Grossmann, A.R. and Chua, N.H. (1982) in: Methods in Chloroplast Molecular Biology (Edelman, M. et al. eds) pp. 1081–1091, Elsevier, Amsterdam, New York.