Volume 216, number 2, 234-240 FEB 04727 June 1987

Nucleotide sequence of cDNA clones encoding the complete '23 kDa' and '16 kDa' precursor proteins associated with the photosynthetic oxygen-evolving complex from spinach

T. Jansen, C. Rother, J. Steppuhn, H. Reinke+, K. Beyreuther+, C. Jansson*, B. Andersson* and R.G. Herrmann

Botanisches Institut der Ludwig-Maximilians-Universität, Menzinger Str. 67, D-8000 München 19, +Institut für Genetik der Universität Köln, Im Weyertal 121, D-5000 Köln 41, FRG and **Dept of Biochemistry, Arrhenius Laboratory, University of Stockholm, S-106 91 Stockholm, Sweden*

Received 20 March 1987

We present the nucleotide sequences and derived amino acid sequences of cDNAs that encode the complete precursors of the extrinsic '23 kDa' and '16 kDa' polypeptides associated with the photosynthetic oxygen-evolving complex from spinach. The luminal proteins consist of 267/186 (precursor/mature 23 kDa protein) and 232/149 (16 kDa polypeptide) amino acid residues corresponding to molecular masses of 28.5/20.2 and 24.9/16.5 kDa, respectively. Secondary structure predictions disclose epitopes that are potential candidates for two-step processing of the precursors during import and intraorganelle routing as well as for calcium sequestering, chloride binding and subunit/subunit interaction.

Photosynthesis; Oxygen evolution; 23 kDa protein; 16 kDa protein; cDNA nucleotide sequence; Transit peptide; (Spinach)

1. INTRODUCTION

Light-induced oxidation of water is a fundamental process of oxygenic photosynthetic biomembranes. This process is catalyzed by photosystem II, a multisubunit polypeptide complex that contains both integral and peripheral components (review [1,2]). Constituents of the reaction center core that include the subunits involved in the primary charge separation belong to the first category (review [3]). The latter group is dominated by three hydrophilic polypeptides with apparent molecular masses of 33, 23 and 16 kDa which are located in the interior of stacked thylakoid membranes in close proximity to the catalytic Mn-cluster involved in water oxidation. The 33 kDa component appears to stabilize the manganese [4], while the 23 and 16 kDa proteins are required for binding or sequestering calcium and/or chloride necessary for optimal oxygen evolution [1,2].

Photosynthetic membranes of eukaryotes are of dual genetic origin (cf. [3,5]). The seven polypeptide species constituting the photosystem II reaction center core appear all to be plastome-encoded and their genes are well studied (cf. [3,6]). On the other hand the three extrinsic subunits originate from nuclear genes [7]. They are synthesized on cytosolic ribosomes as precursors that are imported into the organelle and, on their route,

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

The nucleotide sequences presented here have been submitted to the EMBL/GenBank database under the accession number Y00681

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

traverse three membranes, the two envelopes and the thylakoid membrane. Little is known about their genes.

We have recently succeeded in the isolation of cDNA clones for more than 20 nuclear-derived thylakoid proteins from spinach ([5], and unpublished) including the three extrinsic photosystem II polypeptides. The fine structure analysis of cDNA for the 33 kDa protein has been detailed [8]. In this communication we provide corresponding data for the extrinsic 23 kDa and 16 kDa proteins of the oxygen-evolving system.

2. MATERIALS AND METHODS

The rationale for sequencing cDNA clones complementary to the 23 kDa and 16 kDa subunits associated with the oxygen-evolving complex of spinach was the same as that described for the 33 kDa protein [8] and plastocyanin [9], except that an exclusively λgt11 cDNA expression library was used [5] to select an appropriate recombinant phage with polyclonal monospecific antisera elicited against each of the two spinach proteins [10].

The two proteins were isolated according to Jansson [11]. Their N-terminal amino acid sequences were determined by automated gas-phase Edman degradation with on-line identification of the PTH amino acids [12].

3. RESULTS

3.1. *Clone selection*

We have previously described the construction of spinach cDNA expression libraries in the λ phage gt11 [5]. In a screen of 1×10^5 recombinants with a polyclonal antiserum directed against the 23 kDa and 16 kDa polypeptides, respectively, of the water splitting complex, in separate experiments the absorbed antibodies recognized a total of 16 and 15 plaques. Since cDNAs selected by immunoscreening are inherently incomplete, excised 5′-part of one of the inserts for each of the two proteins was recloned into Bluescribe vector $M13^+$, transcribed to high specific activity with T_7 and/or T_3 polymerase [13], and used to isolate additional chimeric phage by plaque hybridization [14]. A total of 2 (23 kDa protein) and 25 (16 kDa protein) phage were obtained.

A complication arising from restriction analysis with *Eco*RI of these recombinant DNAs was that inserts for the 23 kDa component exceeding ~700 nucleotides consistently gave two fragments including one fragment of constant size. In all instances, this fragment was inserted in the same orientation relative to lacZ, at the 3′-end of the gene. We have subcloned these fragments separately into the *Sma*I insertion site of the Bluescribe vector $M13^+$ [13] after filling in 3′-recessed ends, and checked by hybrid release translation with poly(A^+) RNA [15] that both of them selected mRNA for the same protein, i.e., that the gene contains an internal *Eco*RI restriction site (not shown). Sequence analysis confirmed that this assignment was correct, since the *Eco*RI site was found within the stretch of amino acid residues that was determined by N-terminal sequencing of the authentic protein.

3.2. *Sequence analysis*

The inserts from several clones for each of the polypeptides were sequenced (fig.1). The two largest inserts were found to contain the entire amino acid-coding regions of the proteins including their transit peptide as well as 5′- and 3′-untranslated segments. The nucleotide sequences and derived amino acid sequences of these clones, 123Soc23-9 and 123Soc16-11 totalling 1013 and 875 bp, respectively, are shown in fig.2.

Inspection of these sequences raises 6 major points of interest:

(i) The amino acid sequences deduced from the nucleotide sequences were identified as those of the 23 and 16 kDa proteins by comparison with partial amino acid sequences of the corresponding mature spinach proteins. The determined first 25 (23 kDa polypeptide) and 19 (16 kDa polypeptide) N-terminal amino acid residues of the respective proteins are found within the predicted amino acid sequence from residue 82–106 (23 kDa protein) and 84–101 (16 kDa protein). The data are also consistent with the recently published, incompletely determined N-terminal sequence for both proteins [16]. Only the methionine residue at position 100 in the sequence for the 23 kDa precursor protein is replaced by threonine. Furthermore, the amino acid compositions calculated from the sequence data of the derived polypeptides agree with those

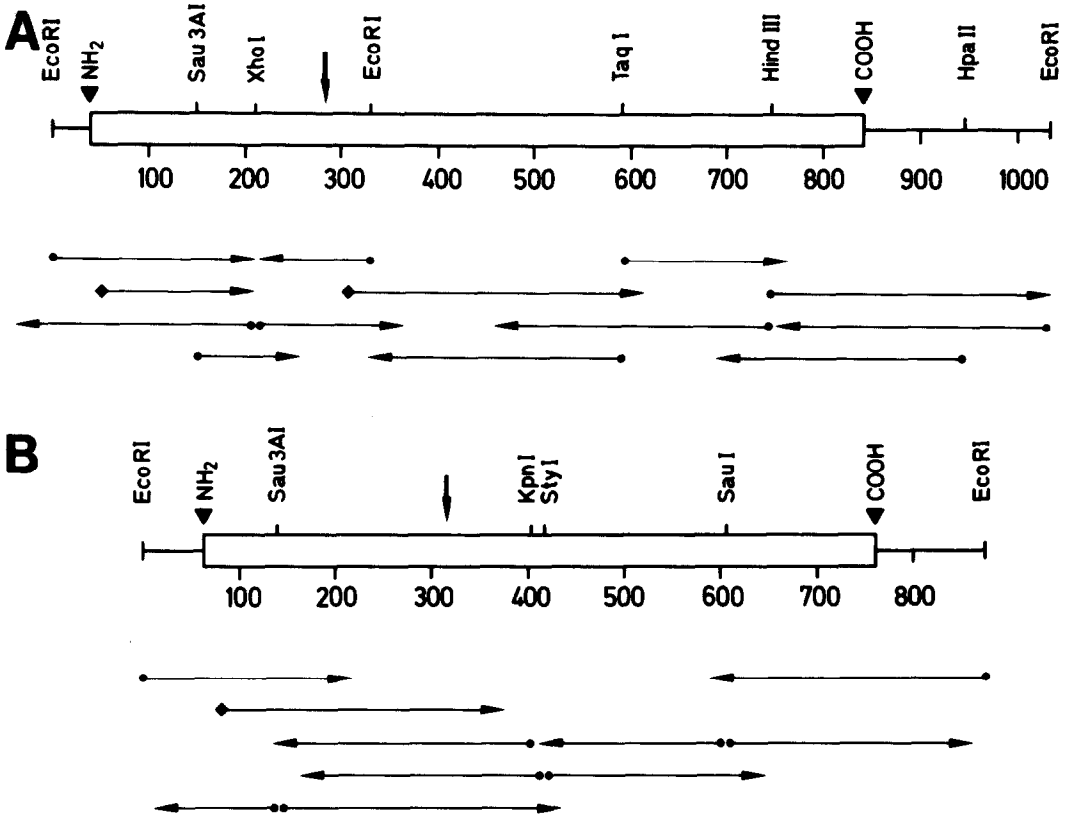

Fig.1. Restriction maps and sequencing strategy of p6Soc23-9 (A) and p6Soc16-11 inserts (B). The regions coding for the 23 kDa (A) and 16 kDa polypeptides (B) associated with the photosynthetic water-oxidation complex are boxed. The positions of the NH_2- and COOH-termini as well as the terminal processing sites of the precursor proteins are given by arrowheads and arrows, respectively. Horizontal arrows denote the direction and extent of individual sequence reactions; the diamonds mark sequence reactions starting from 5′-termini of various recombinant cDNAs.

determined for the authentic proteins [11] within the limits of error, except for Tyr (6 residues predicted, 3 determined), Cys (0/1) and Trp (1/0) in the 16 kDa polypeptide (cf. [17]).

(ii) The presence of poly(A) tails in the range of 8–23 nucleotides indicates that the cDNA inserts include the entire 3′-ends of the respective mRNAs. The 3′-untranslated segments of p6Soc23-9 DNA and p6Soc16-11 DNA between the TAA stop codon and the putative polyadenylation site are 165 and 106 bp long. The regions include putative polyadenylation signals [18], TATAAT (positions 931–936; 23 kDa polypeptide) and AATGAA (positions 770–775; 16 kDa polypeptide).

(iii) The 3′-untranslated sequence is identical in all the clones studied for the two proteins indicating that they are derivatives of the same or from identical gene(s); 3′-untranslated regions generally diverge quite rapidly in gene families (e.g. [19]). All size variations of the cDNA inserts are caused by variation in the 5′-terminal segment (fig.1).

(iv) The mature 23 and 16 kDa proteins are predicted to contain 186 and 149 amino acid residues corresponding to molecular masses of 20209 and 16522 Da, respectively. In contrast to the results obained with the 33 kDa protein [8,20], these calculated molecular masses are in reasonable agreement with the experimentally determined values (23 kDa; 16–18 kDa) [1,2].

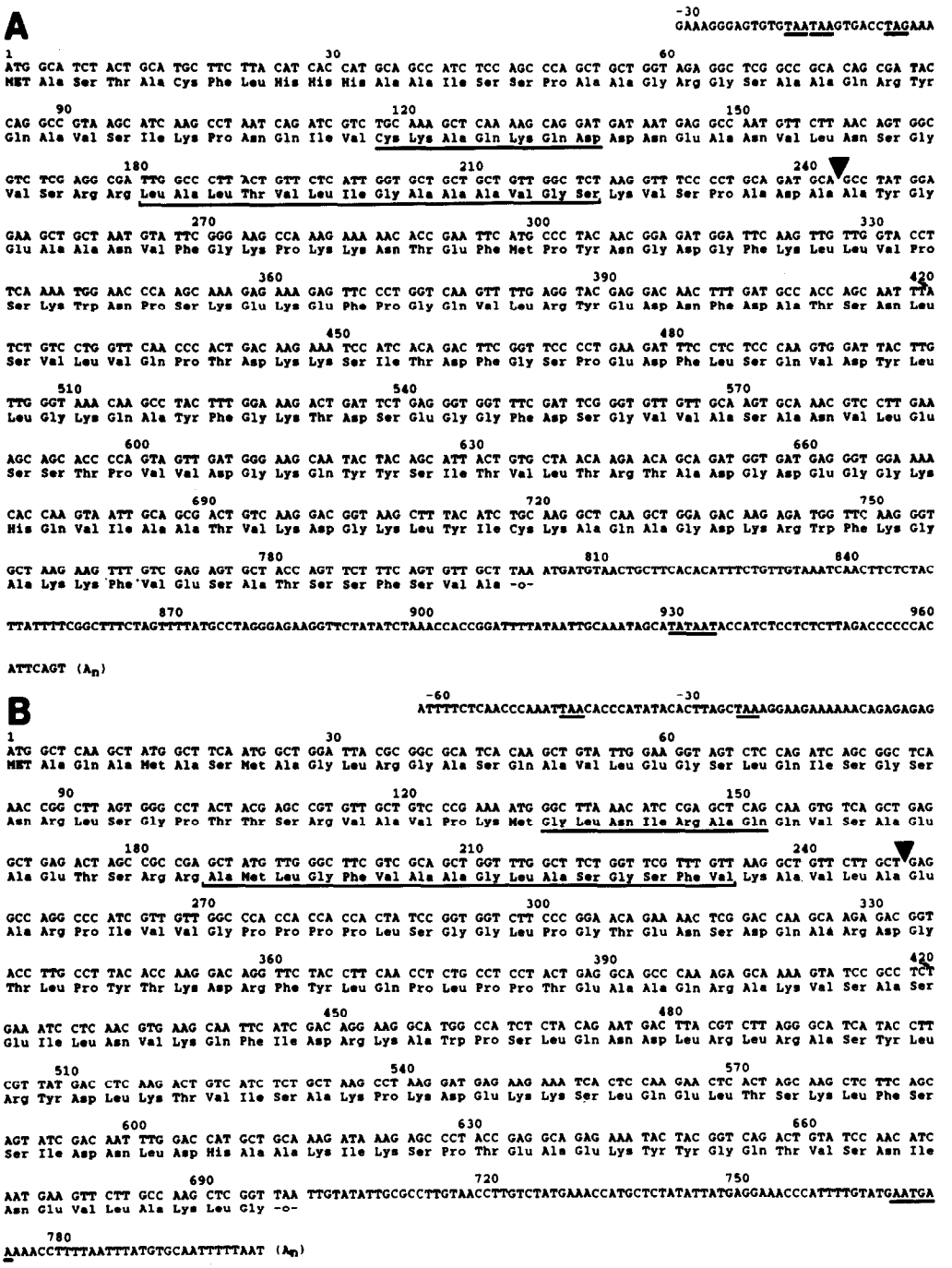

Fig.2. Nucleotide sequence and deduced amino acid sequences of recombinant cDNAs encoding the entire 23 kDa (A) and 16 kDa polypeptides (B) of the photosynthetic water-oxidation complex from spinach. The sequences of the mRNA-sense strands are given. The arrowheads indicate the putative terminal processing sites of the respective precursor proteins. Putative polyadenylation signals, in-frame stop codons in the untranslated 5′-regions and β-sheets possibly involved in intermediate processing are underlined; proposed hydrophobic domains in the putative transit signals are bracketed (see text).

(v) The sequence data are in accord with the synthesis of higher molecular mass forms for these proteins [7]. Translation of the nucleotide sequence gave one coherent reading frame in each instance, of 267 and 232 codons. Unfortunately, uncertainties in electrophoretic molecular mass estimates preclude an unambiguous identification of the translational initiation for the 16 kDa component. The nucleotide sequence indicates that there are three potential in-frame AUG codons which could act as initiation start signals. Comparison of the transit sequences for the four luminal proteins (see below) as well as the presence of G at -3 and $+4$ and in-phase stop codons at -24 to -22 and -45 to -43 that are in line with the translational scanning model of eukaryotic 80 S ribosomes [21] suggest that the indicated methionine residue (fig.2) may act as a translation start.

(vi) Transit sequence: the precursors of the 33, 23 and 16 kDa subunits from spinach are transported into and processed inside the chloroplasts. During import, sequences corresponding to molecular masses of 6 (33 kDa polypeptide) and 10 kDa (23 kDa and 16 kDa polypeptides) as determined electrophoretically in denaturing polyacrylamide gels are cleaved off the amino-terminal end of the precursors to produce the mature proteins [7]. The presumptive sequence of the transit peptides derived from the DNA sequences contain 81 (8.3 kDa; 23 kDa protein) and 83 residues (8.4 kDa; 16 kDa protein) which correlates with this finding.

Components of the oxygen-evolving complex, such as plastocyanin [9,22], traverse three membranes, viz. the thylakoid membrane in addition to the two envelopes of the organelle. It has been proposed that the transitory sequences for plastocyanin [9,22] and the three extrinsic polypeptides of the water-oxidizing complex [8] are a composite of two substructures, i.e. are functionally divided, and are dissected in at least two steps during import. These reactions are catalysed by different endopeptidases [23]. The transit sequences of these four luminal proteins lack significant homology in primary sequence but contain hydroxylated and charged (basic and acidic) residues that are suggested to be involved in the import process [24,25]. However, they share striking characteristic secondary structure including a β-sheet within or in vicinity of the presumed intermediate cleavage site which is amphipathic in the case of the 16 kDa protein (cf. [8]) and a large hydrophobic domain preceding the terminal processing site (figs 2 and 3). We are presently testing the possibility that the presence of these elements and the charge patterns surrounding them reflects a functional requirement for intracellular sorting, intraorganelle routing and the proteolytic processing system.

3.3. *Protein structure predictions*

Our sequence data are consistent with the biochemical findings (review [1,2]) that all three proteins are hydrophilic (fig.3) and protrude into the intrathylakoid space, associated with the membrane via subunits of the photosystem II core [26] and/or the 22 kDa polypeptide [27] but with little contact with the lipid bilayer. Assuming a globular shape for the proteins, their protrusion into the thylakoid lumen would be 4 nm. Such particles have been deduced from electron spin measurements [28] and found by freeze-etch electron microscopy ([29] Staehelin, L.A., personal communication).

The 23 kDa component appears to serve four specific functions. It binds to the 33 kDa component, binds the 16 kDa subunit, seems to be involved in sequestering Ca^{2+} and, in a cooperative effect with the 16 kDa compound, to reduce the requirement for chloride that are both necessary factors for optimal oxygen evolution [30]. Consequently, the 23 kDa and 16 kDa polypeptides contain a relatively large number of acidic and basic amino acid residues (25–27% of the total) throughout their polypeptide chains, although some clustering is obvious (fig.2). These are potential candidates for Ca^{2+} or Cl^--binding or could play a role in subunit/subunit interaction if they are brought together in the tertiary structure to form basic patches that might interact with acidic patches of surrounding proteins and vice versa. In this context, clusters of charged residues, e.g., found at nucleotide positions 478–501 or 502–519 in the chain of the 16 kDa protein that are capable of forming sided β-sheets, are of especial interest. Similarly, limited proteolysis of the 16 kDa subunit to a lower molecular mass component removes a short N-terminal sequence to give a product that has lost the ability to associate with the

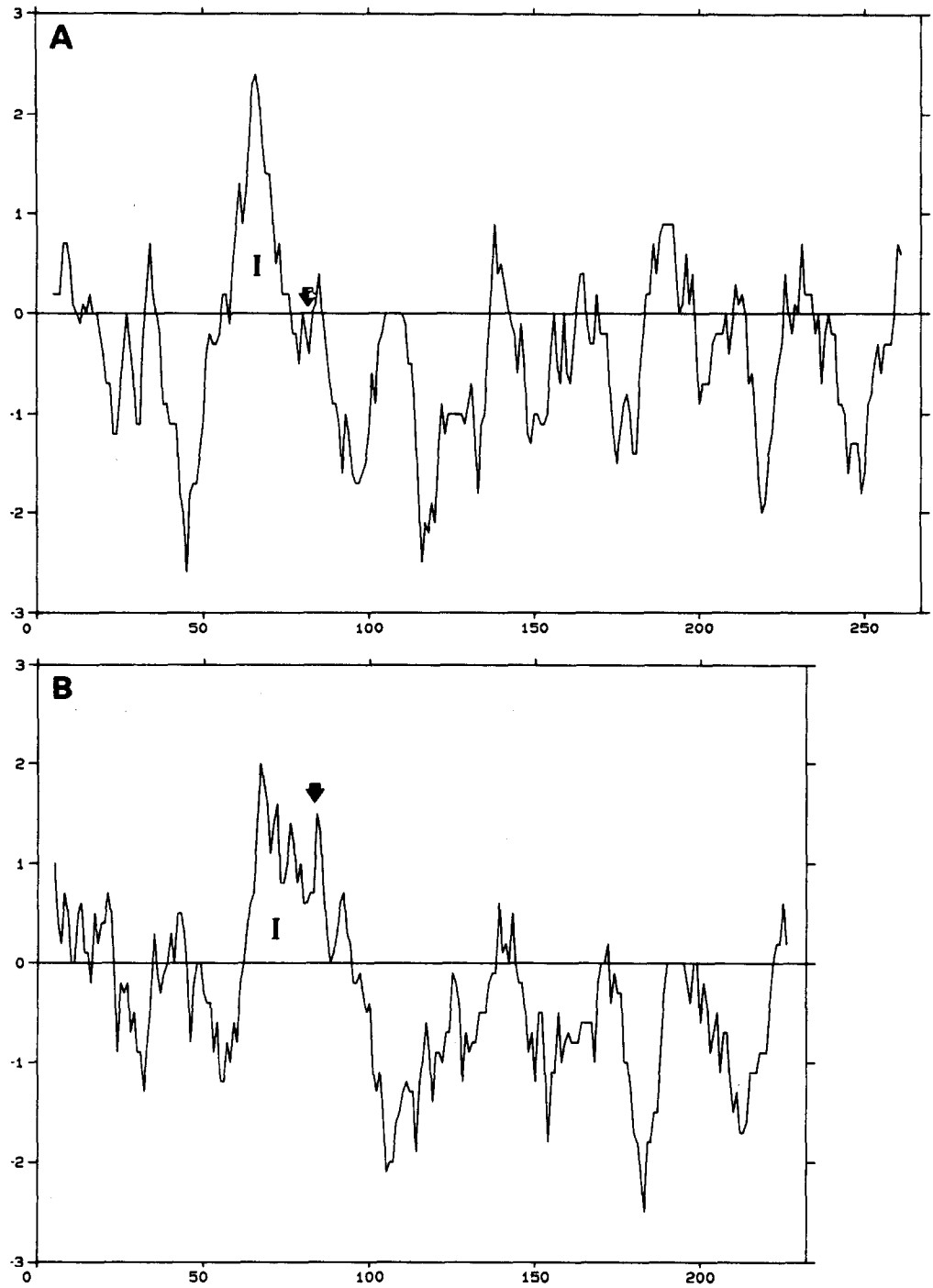

Fig.3. Hydropathy profiles of the 23 kDa (A) and 16 kDa (precursor) proteins (B) associated with the photosynthetic water-oxidation complex from spinach. The profiles were calculated according to Kyte and Doolittle [32] using an 11-point moving interval. The arrows indicate the terminal processing site of the precursors in the chloroplast. A possible h-domain of the transit sequence [25] is marked by a roman number (I; cf. fig.2).

23 kDa component [31]. It is tempting to assume that the discrete cluster of 4 proline residues in the N-terminal region that presumably forms kinks in the polypeptide chain is involved in the loss of this short appendix.

Although the functional characterization of constituents of the oxygen-evolving complex has progressed substantially [1,2], ultimate information on genes and, derived from their investigation, protein structure are indispensible for a detailed understanding of photosynthetic processes as well as biogenesis and physiological adaptation of this specialized biomembrane. The outlined study has provided regions in proteins allowing correlations between epitopes and functions of subunits of the water oxidation complex. Although these correlations are only tentative at present, the methodology, specifically the combination of site-specific mutagenesis, expression cloning and reconstitution, is now available to answer the type of questions outlined. In this way the regions where functional or structural sequences in proteins are located will soon be known.

ACKNOWLEDGEMENTS

This work was supported by the SFB 184 and a grant from the Fonds der Chemischen Industrie.

REFERENCES

[1] Andersson, B. (1986) in: Encyclopedia of Plant Physiology (Staehelin, L.A. and Arntzen, C.B. eds) vol. 19, pp. 447–456, Springer, Berlin.
[2] Miyao, M. and Murata, N. (1987) Top. Photosynthesis 9, in press.
[3] Herrmann, R.G., Westhoff, P., Alt, J., Tittgen, J. and Nelson, N. (1985) in: Molecular Form and Function of the Plant Genome (Van Vloten-Doting, L., et al. eds) pp. 233–256, Plenum, New York.
[4] Miyao, M. and Murata, N. (1985) FEBS Lett. 180, 303–308.
[5] 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.
[6] Westhoff, P., Farchaus, T. and Herrmann, R.G. (1986) Curr. Genet. 11, 165–169.
[7] Westhoff, P., Jansson, C., Klein-Hitpaß, L., Berzborn, R., Larsson, C. and Bartlett, S.G. (1985) Plant Mol. Biol. 4, 137–146.
[8] Tyagi, A., Hermans, J., Steppuhn, J., Jansson, C., Vater, F. and Herrmann, R.G. (1987) Mol. Gen. Genet., in press.
[9] Rother, C., Jansen, T., Tyagi, A., Tittgen, J. and Herrmann, R.G. (1986) Curr. Genet. 11, 171–176.
[10] Andersson, B., Larsson, C., Jansson, C., Ljungberg, U. and Åkerlund, H.-E. (1985) Biochim. Biophys. Acta 766, 21–28.
[11] Jansson, C. (1983) in: Advances in Photosynthesis Research (Sybesma, C. ed.) vol. 1, pp. 375–378, Martinus Nijhoff/Junk, Dordrecht, The Netherlands.
[12] Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.
[13] Stratagene, San Diego, USA.
[14] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[15] Bünemann, H., Westhoff, P. and Herrmann, R.G. (1982) Nucleic Acids Res. 10, 7163–7180.
[16] Vater, J., Salnikow, J. and Jansson, C. (1986) FEBS Lett. 203, 230–234.
[17] Murata, N., Miyao, M. and Kuwabara, T. (1983) in: The Oxygen Evolving System of Photosynthesis (Inoue, Y., et al. eds) pp. 213–222, Academic Press, Tokyo.
[18] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[19] Dunsmuir, P., Smith, S.M. and Bedbrook, J. (1983) J. Mol. Appl. Genet. 2, 285–300.
[20] Oh-oka, H., Tanaka, S., Wada, K., Kuwabara, T. and Murata, N. (1986) FEBS Lett. 197, 63–66.
[21] Kozak, M. (1986) Cell 44, 283–292.
[22] Smeekens, S., De Groot, M., Van Binsbergen, J. and Weisbeek, P. (1985) Nature 317, 456–458.
[23] Robinson, C. and Ellis, R.J. (1984) Eur. J. Biochem. 142, 337–342.
[24] Schmidt, G.W., Devillers-Thiery, A., Desruisseaux, H., Blobel, G. and Chua, N-H. (1979) J. Cell Biol. 83, 615–622.
[25] Von Heijne, G. (1986) J. Mol. Biol. 189, 239–242.
[26] Sayre, R.T., Andersson, B. and Bogorad, L. (1986) Cell 47, 601–608.
[27] Ljungberg, U., Åkerlund, H.-E. and Anderson, B. (1986) Eur. J. Biochem. 158, 477–482.
[28] Itoh, S., Isogai, Y., Tang, X.-S. and Satoh, K. (1987) in: Proceedings of the VII Congress in Photosynthesis, in press.
[29] Simpson, D.J. and Andersson, B. (1986) Carlsberg Res. Commun. 51, 467–474.
[30] Akabori, K., Imaoka, A. and Toyoshima, Y. (1984) FEBS Lett. 173, 36–40.
[31] Kuwabara, T., Murata, T., Miyao, M. and Murata, N. (1986) Biochim. Biophys. Acta 850, 146–155.
[32] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.