Volume 253, number 1,2, 178-182 FEB 07507 August 1989

N-terminal sequencing of low-molecular-mass components in cyanobacterial photosystem II core complex

Two components correspond to unidentified open reading frames of plant chloroplast DNA

Masahiko Ikeuchi, Hiroyuki Koike and Yorinao Inoue

Solar Energy Research Group, Institute of Physical and Chemical Research, Wako, Saitama 351-01, Japan

Received 6 June 1989

We recently reported the presence of several low-molecular-mass protein components in the PS II O_2-evolving core complex from the thermophilic cyanobacterium, *Synechococcus vulcanus* [(1989) FEBS Lett. 244, 391–396]. Here we have characterized the three components (4.1, 4.7, 5 kDa) of the same cyanobacterial core complex by N-terminal sequencing. There were two components in the 4.7 kDa region, both having a blocked N-terminus. One has a sequence highly homologous to open reading frame 34 of plant chloroplast DNA (tentatively designated *psb*M), while the other has a sequence partially homologous to open reading frame 43 of chloroplast DNA (designated *psb*N), although neither of the two gene products has yet been confirmed in chloroplasts. The cyanobacterial 4.1 kDa protein partially corresponds to the 4.1 kDa nuclear-encoded core component of higher plant PS II. The cyanobacterial 5 kDa component, however, shows a sequence that is unrelated to any other known proteins.

Chloroplast DNA; Core complex; Cyanobacterium; Photosystem II; (*Synechococcus vulcanus*)

1. INTRODUCTION

Several low-molecular-mass protein components have recently been resolved from various PS II complexes of higher plants by high-resolution SDS-PAGE [1,2]. Since the entire chloroplast DNA sequence of tobacco and liverwort has been determined [3,4], the partial amino acid sequencing of these proteins enables us to conclude that some are encoded by the chloroplast genome, the others probably by the nuclear genome [5-9]. However, even after determination of gene and/or protein sequences, the function or significance of these components in PS II remains uncertain. A promising approach to the elucidation of the roles of these components may be genetical elimination of the proteins. This approach, however, is not realistic at present for higher plants, and is possible only for cyanobacteria. It is thus necessary to study cyanobacterial PS II and to compare its protein composition with that of higher plant PS II. Such comparisons may also be of interest from an evolutionary point of view in relation to the origin of chloroplasts.

In the previous paper, we reported that the O_2-evolving PS II core complex of *Synechococcus vulcanus* contains many low-molecular-mass components like higher plants, and confirmed the presence of the *psb*H and *psb*K gene products by N-terminal sequencing [10]. We also detected the gene products of *psb*I and *psb*L as well as *psb*E and *psb*F in the cyanobacterial reaction center complex [11]. However, several other low-molecular-mass

Correspondence address: M. Ikeuchi, Solar Energy Research Group, Institute of Physical and Chemical Research, Wako, Saitama 351-01, Japan

Abbreviations: ORF, open reading frame; PS, photosystem; SDS-PAGE, SDS-polyacrylamide gel electrophoresis

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

proteins in PS II remain unidentified. Here we report the N-terminal sequences of the unidentified components of 4–5 kDa. Unexpectedly, the genes for the two components appear to correspond to the two open reading frames of plant chloroplast DNA, whose gene products, however, have not yet been observed.

2. MATERIALS AND METHODS

O_2-evolving PS II core complex was prepared from allophycocyanin-containing PS II particles from *S. vulcanus* Copeland by treatment with Triton X-100 followed by DEAE-Toyopearl column chromatography as in [10]. Low-molecular-mass proteins were resolved by SDS-PAGE with 7.5 M urea and a 16–22% (w/v) acrylamide gradient according to [2]. Separated proteins were stained with Coomassie brilliant blue R-250 (Bio-Rad) or electroblotted onto a polyvinylidene difluoride membrane (Immobilon, Millipore) using 0.055% (w/v) SDS, 20% (v/v) methanol and 17 mM boric acid as transfer buffer as described [11]. Transferred proteins were visualized using amido black 10B (Bio-Rad) and subjected to sequencing on a protein sequencer (model 477A, Applied Biosystems).

The amino acid sequences determined were compared by computer matching with those in the Protein Sequence Database or all possible protein-coding sequences longer than 10 frames [7] deduced from the chloroplast DNA sequences of tobacco (*Nicotiana tabacum*) [3] and liverwort (*Marchantia polymorpha*) [4].

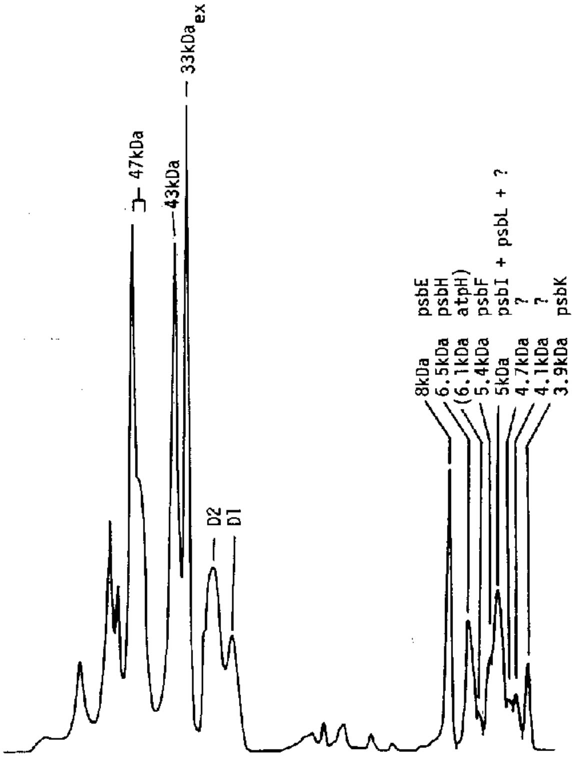
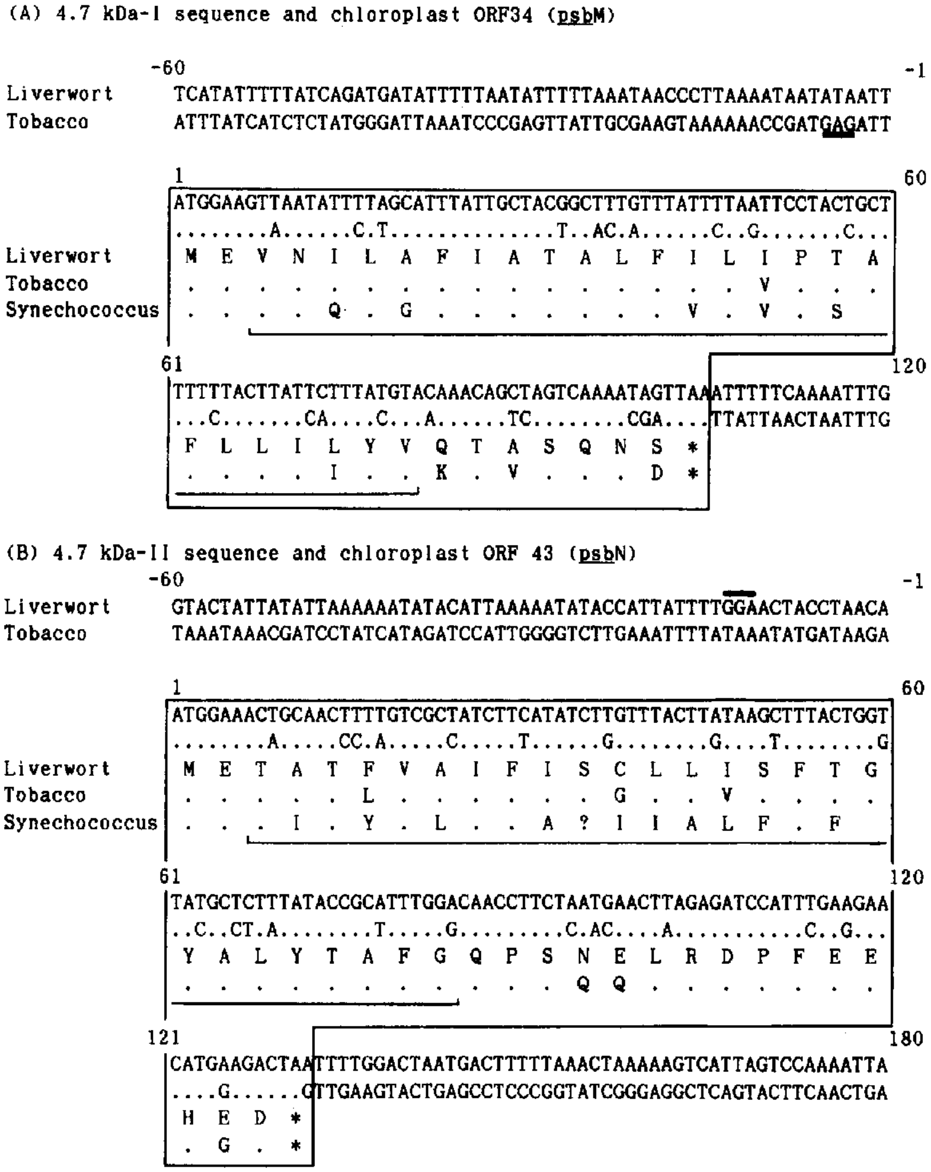


Fig.1. SDS-PAGE densitogram of O_2-evolving PS II core complex.

3. RESULTS

A typical SDS-PAGE separation profile of our cyanobacterial O_2-evolving core complex (fig.1) shows seven bands below 10 kDa in addition to a shoulder at 6.1 kDa due to contaminating *atp*H product (proteolipid component of ATPase: coupling factor). This pattern is almost identical with our previously reported profile [10], except for the absence of a minor band of 4.4 kDa. In many cases, the 4.4 kDa protein was not recovered in O_2-evolving core complex and may well be no real PS II component. As noted in fig.1, many of the bands have been identified by N-terminal sequencing as gene products of *psb*E, *psb*H, *psb*F, *psb*I, *psb*L and *psb*K [10,11] as well as *atp*H (unpublished). Hence, two bands, at 4.7 and 4.1 kDa, remain to be characterized. The dye-staining intensities of these bands are relatively low vs PS II bands. Since the content of these bands is appreciably higher in the initial PS II particles (not shown), their association with the PS II core complex may be weaker than that of the other components. On the other hand, the intensity of the 5 kDa band is about 3-times higher than that of the adjacent *psb*F or *psb*H band. This suggests the presence of an unknown protein(s) in this region in addition to the previously characterized *psb*I and *psb*L products [11].

When analyzed after acid treatment, the 4.7 kDa band from the O_2-evolving core complex provided two sets of amino acid signals at each cycle that had comparable signal intensities, indicative of the superposition of two unique components. Since neither signal was obtained without acid pretreatment, the N-termini of both proteins must be blocked. In order to distinguish between the two, we separately analyzed the 4.7 kDa band obtained from a partially purified reaction center preparation. As mentioned in [11], the reaction center complex isolated by use of Triton X-100 and urea sometimes retains part of the 4.7 kDa protein while the other core components are completely removed. This 4.7 kDa band gave only one amino acid sequence (fig.2A), which corresponded exactly to



```
(A) 4.7 kDa-I sequence and chloroplast ORF34 (psbM)

              -60                                                          -1
Liverwort      TCATATTTTTATCAGATGATATTTTTAATATTTTTAAATAACCCTTAAAATAATATAATT
Tobacco        ATTTATCATCTCTATGGGATTAAATCCCGAGTTATTGCGAAGTAAAAAACCGATGAGATT

               1                                                           60
               ATGGAAGTTAATATTTTAGCATTTATTGCTACGGCTTTGTTTATTTTAATTCCTACTGCT
               ........A......C.T..............T..AC.A......C..G.......C...
Liverwort       M  E  V  N  I  L  A  F  I  A  T  A  L  F  I  L  I  P  T  A
Tobacco         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  V  .  .  .
Synechococcus   .  .  .  .  Q  .  G  .  .  .  .  .  .  .  V  .  V  .  S

               61                                                          120
               TTTTTACTTATTCTTTATGTACAAACAGCTAGTCAAAATAGTTAAATTTTTCAAAATTTG
               ...C.......CA....C...A......TC........CGA....TTATTAACTAATTTG
                F  L  L  I  L  Y  V  Q  T  A  S  Q  N  S  *
                .  .  .  .  I  .  .  K  .  V  .  .  .  D  *

(B) 4.7 kDa-II sequence and chloroplast ORF 43 (psbN)
              -60                                                          -1
Liverwort      GTACTATTATATTAAAAAATATACATTAAAAATATACCATTATTTTGGAACTACCTAACA
Tobacco        TAAATAAACGATCCTATCATAGATCCATTGGGGTCTTGAAATTTTATAAATATGATAAGA

               1                                                           60
               ATGGAAACTGCAACTTTTGTCGCTATCTTCATATCTTGTTTACTTATAAGCTTTACTGGT
               ........A.....CC.A.....C.....T......G........G....T........G
Liverwort       M  E  T  A  T  F  V  A  I  F  I  S  C  L  L  I  S  F  T  G
Tobacco         .  .  .  .  .  L  .  .  .  .  .  .  G  .  .  V  .  .  .  .
Synechococcus   .  .  .  I  .  Y  .  L  .  .  A  ?  I  I  A  L  F  .  F

               61                                                          120
               TATGCTCTTTATACCGCATTTGGACAACCTTCTAATGAACTTAGAGATCCATTTGAAGAA
               ..C..CT.A........T.....G.........C.AC....A...........C..G...
                Y  A  L  Y  T  A  F  G  Q  P  S  N  E  L  R  D  P  F  E  E
                .  .  .  .  .  .  .  .  .  .  .  Q  Q  .  .  .  .  .  .  .

              121                                                          180
               CATGAAGACTAATTTTGGACTAATGACTTTTTAAACTAAAAAGTCATTAGTCCAAAATTA
               ....G......GTTGAAGTACTGAGCCTCCCGGTATCGGGAGGCTCAGTACTTCAACTGA
                H  E  D  *
                .  G  .  *
```

Fig.2. Partial N-terminal amino acid sequences of the cyanobacterial 4.7 kDa-I (A) and 4.7 kDa-II (B) components and their alignments with tobacco and liverwort ORFs. Numbering starts at the ATG codon at which translation is probably initiated. Corresponding amino acid residues deduced from the liverwort sequence are expressed in the single-letter code. Nucleotide differences in the coding region and amino acid differences are indicated below the liverwort sequence. A dot indicates the same residue as in liverwort and an asterisk denotes a stop codon. The identified gene is boxed. Possible ribosome-binding sites, GAG and GGA, are shown by bold lines. A hydrophobic region, which may span the membrane, is indicated by the horizontal bracket. ORF34 is located between *trn*L (CAA) and *rpo*B in liverwort chloroplast DNA [4] and between *trn*D (GUC) and *rpo*B in tobacco [3]. ORF43 is located between *psb*B and *psb*H but on the opposite strand in both liverwort and tobacco.

one of the two sets of signals found for the band from the core complex. This tightly associated component was designated 4.7 kDa-I, the component that readily dissociated from the reaction center being termed 4.7 kDa-II. The amino acid sequence of the 4.7 kDa-II component was obtained by subtracting the 4.7 kDa-I signals from the entire set (fig.2B). Since the 1st, 2nd and 9th signals were single and relatively intense, we postulate that both components share the same amino acid residues in these positions.

Computer-assisted homology search revealed that the sequence of the 4.7 kDa-I component is highly homologous to the predicted products of ORF34 from tobacco and liverwort chloroplast DNA (fig.2A). Most of the amino acid differences between *Synechococcus* and the two plants are due to substitutions of conservative residues. Both tobacco and liverwort ORF34 sequences predict the presence of a single possible membrane-spanning segment consisting of 25 amino acid residues. The hydrophobicity of this segment is conserved in the *Synechococcus* 4.7 kDa-I protein. The molecular masses of the ORF34 products were estimated to be 3.7 kDa (liverwort) and 3.8 kDa (tobacco). These values differ slightly from that of the *Synechococcus* counterpart (4.7 kDa). It may be possible that the C-terminal part of this protein, which was not sequenced here, differs between *Synechococcus* and higher plants, however the present results seem to suggest rather that a new PS II gene corresponding to the higher plant ORF34 encodes the 4.7 kDa-I protein in *Synechococcus*.

As regards the 4.7 kDa-II sequence, the computer homology search indicated ORF43 of the chloroplast DNA sequences from both liverwort and tobacco (fig.2B). The homology between the 4.7 kDa-II sequence and plant ORF43 is rather low compared with the 4.7 kDa-I sequence vs ORF34. However, many of the sequence differences can be interpreted as being the result of replacement of conservative residues. A single hypothetical membrane-spanning hydrophobic segment predicted in ORF43 is also maintained as hydrophobic in the 4.7 kDa-II sequence. The calculated molecular mass for the plant ORF43 product is 4.9 kDa for liverwort and 4.7 kDa for tobacco, both of which are almost identical to the apparent molecular mass of the *Synechococcus* protein. These indicate that the 4.7 kDa-II protein is encoded by a cyanobacterial gene homologous to the ORF43 of plant chloroplast DNA. Based on this agreement in the data, we tentatively designate ORF34 and ORF43 as *psb*M and *psb*N, respectively, although their gene products remain thus far undetected in higher plant thylakoids.

The 4.1 kDa band provided a unique N-terminal sequence (fig.3A). The sequence contains only one charged amino acid, Lys, at position 7. Subsequent to this Lys residue, there follows a hydrophobic amino acid stretch of at least 24 residues, which may span the membrane. This sequence shows only slight homology with the 4.1 kDa nuclear-encoded core component of spinach and wheat (fig.3A and [8]). Despite the relatively low homology, the Lys residue and the following hydrophobic stretch appear to be conserved in higher plants as far as the two sequenced species are concerned. Moreover, this component is associated with the PS II core complex in both higher plants and cyanobacteria and its apparent molecular mass is very similar between *Synechococcus* and higher plants. The above data strongly support the correspondence of the 4.1 kDa component between cyanobacteria and higher plants.

```
A.    4.1 kDa sequence

                          possible membrane spanning segment
Synechococcus        TITPSLKGFFIGLLSGAVVLGLTFAVLIAISQ
Spinach             AELA....N.LFSIAA.G...VQIVG
Wheat             ASPGLS....N.LLSIV..G..FAGIAG

B.    5 kDa sequence
                     ATKSAKPTYAFRTF?AVLLLAINFLVAAY
```

Fig.3. Partial N-terminal amino acid sequences of the cyanobacterial 4.1 kDa (A) and 5 kDa proteins. The seuqence of the 4.1 kDa protein is aligned with those of higher plant counterparts ([8] and unpublished). A dot indicates the same residue as in the *Synechococcus* sequence. (?) Residue not determined. Note that all the sequences start from the N-terminus and are partial.

The 5 kDa band showed a unique sequence superimposed on the sequence of *psb*L products without acid treatment (fig.3B). The N-terminal part of the sequence was enriched in positively charged residues. No sequence homologous to this 5 kDa protein was found in the Protein Sequence Database or in chloroplast DNA.

4. DISCUSSION

Here we have described four N-terminal sequences of cyanobacterial PS II core components which had remained to be characterized. The sequence homology suggests that counterparts to these cyanobacterial proteins, except in the case of the 5 kDa protein, are present in higher plants. Combining these results with our previous data [10,11], it follows that all the low-molecular-mass components of *Synechococcus* PS II core complex have been registered as follows: gene products of *psb*E, *psb*H, *psb*F, *psb*I, *psb*L, *psb*M (ORF34), *psb*N (ORF43) and *psb*K and 4.1 and 5 kDa proteins. The counterpart homologs of all these components except for the 5 kDa protein have been identified in plant PS II and/or chloroplast DNA. In contrast, based on the currently limited amount of sequence data available, the 5 kDa component of *Synechococcus* PS II and three nuclear-encoded higher plant components (10–11, 6.1 and 5 kDa [8]) do not appear to have any counterparts. Hence, we can infer that the common components play basic roles in supporting the function or structure of PS II, whereas species-specific components may be involved in differentiated aspects of PS II such as the light-harvesting system.

The gene products of ORF34 (*psb*M) and ORF43 (*psb*N) have not yet been detected in plants. However, it is known that ORF43 is actively transcribed as monocistronic mRNA in liverwort and pea [12]. Since ORF34 as well as ORF43 is highly conserved between tobacco and liverwort (fig.2), ORF34 may also be expressed. In liverwort, ORF34 is surrounded by inverted sequences [13], suggestive of monocistronic transcription. Possible ribosome-binding sequences are found upstream of tobacco ORF34 and liverwort ORF43 but not tobacco ORF43 or liverwort ORF34. The absence of the ribosome-binding sequence is not unusual for chloroplast genes, probably because the secondary structure of mRNA provides an equivalent sequence [14].

The sequence of the 5 kDa protein appears to be unique to *Synechococcus*. However, we have recently demonstrated the same sequence in a 5 kDa protein of the *Synechococcus* PS I core complex retaining the iron-sulfur protein and other core components (Koike, H. et al., unpublished). The content of 5 kDa protein is almost stoichiometric in both the PS I and PS II complex. Since this protein appears to be hydrophobic as suggested by extraction using *n*-butanol, it might connect PS I with the PS II complex in cyanobacterial thylakoids.

Acknowledgements: We thank Dr T. Yasunaga (RIKEN) for computer-assisted homology search and Dr Y. Narahashi (RIKEN) for helpful instructions on protein sequencing. We are also grateful to Mr K. Mamada for help in the preparation of PS II core complexes. This work was supported by a Grant on Solar Energy Conversion by Means of Photosynthesis awarded to The Institute of Physical and Chemical Research (RIKEN) by the Science and Technology Agency of Japan, and partly by a Grant on Frontier Research Program at RIKEN awarded by the Science and Technology Agency of Japan.

REFERENCES

[1] Ljungberg, U., Henrysson, T., Rochester, C.P., Akerlund, H.-E. and Andersson, B. (1986) Biochim. Biophys. Acta 849, 112–120.

[2] Ikeuchi, M. and Inoue, Y. (1988) Plant Cell Physiol. 29, 1233–1239.

[3] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubaybashi, 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.

[4] 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.

[5] Widger, W.R., Cramer, W.A., Hermodson, M. and Herrmann, R.G. (1985) FEBS Lett. 191, 186–190.

[6] Murata, N., Miyao, M., Hayashida, N., Hidaka, T. and Sugiura, M. (1988) FEBS Lett. 235, 283–288.

[7] Ikeuchi, M. and Inoue Y. (1988) FEBS Lett. 241, 99–104.

[8] Ikeuchi, M., Takio, K. and Inoue, Y. (1989) FEBS Lett. 242, 263–269.

[9] Webber, A.N., Packman, L., Chapman, D.J., Barber, J. and Gray, J.C. (1989) FEBS Lett. 242, 259–262.

[10] Koike, H., Mamada, K., Ikeuchi, M. and Inoue, Y. (1989) FEBS Lett. 244, 391–396.

[11] Ikeuchi, M., Koike, H. and Inoue, Y. (1989) FEBS Lett., submitted.

[12] Kohchi, T., Yoshida, T., Komano, T. and Ohyama, K. (1988) EMBO J. 7, 885–891.

[13] Umesono, K., Inokuchi, H., Shiki, Y., Takeuchi, M., Chang, Z., Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K. and Ozeki, H. (1988) J. Mol. Biol. 203, 299–331.

[14] Bonham-Smith, P.C. and Bourque, D.P. (1989) Nucleic Acids Res. 17, 2057–2080.