Plant Cell Physiol. 35(2): 203–209 (1994)

Microheterogeneity of PSI-E Subunit of Photosystem I in *Nicotiana sylvestris*

Junichi Obokata[1], Kohki Mikami[1], Yoshiharu Yamamoto[1,2] and Nobuaki Hayashida[3]

[1] *Graduate school of Environmental Earth Science, Hokkaido University, Sapporo, 060 Japan*
[2] *Department of Botany, Faculty of Science, Kyoto University, Kyoto, 606 Japan*
[3] *The Institute of Physical and Chemical Research, Tsukuba, 305 Japan*

Microheterogeneity of a photosystem I (PSI) subunit encoded by a nuclear gene *psaE* was examined in *Nicotiana sylvestris*, with the aid of cDNA cloning, peptide mapping analysis and protein sequencing. The *psaE* product of this plant has four isoforms whose mobilities in PAGE are slightly different from each other. We isolated two types of *psaE* cDNAs from a *N. sylvestris* cDNA library, and designated the corresponding genes as *psaEa* and *psaEb*, respectively. The *psaEa* and *psaEb* genes are 77% homologous at DNA level, and their translation products share 80.4% homology for the precursor proteins and 89.1% for the mature forms. Comparative analysis of the four isoproteins and the putative products of the two *psaE* genes revealed that two isoproteins out of four are derived from *psaEa* gene, and the difference between these two isoproteins lies in the respective presence or absence of N-terminal alanine. Likewise, the other two proteins are derived from *psaEb* with similar N-terminal heterogeneity. These results indicate that multi-gene organization and heterogeneous N-terminal formation at post-translational level are two possible causes for PSI subunit polymorphism in isogenic plant lines.

Key words: Chloroplast — Isoprotein — *Nicotiana sylvestris* — Photosystem I subunit — Processing — *psaE*.

PSI is a multi-protein complex spanning the thylakoid membranes, and mediates light-driven electron transfer from plastocyanin to ferredoxin. In higher plants, PSI complex consists of at least 12 subunits, including the products of chloroplast genes *psaA, psaB, psaC, psaI*, and *psaJ*, and also those of nuclear genes *psaD, psaE, psaF, psaG, psaH, psaK* and *psaL* (Scheller and Møller 1990, Bryant 1992, Ikeuchi 1992, Obokata et al. 1993). The gene products of *psaA* through *psaL* are designated as PSI-A through PSI-L (Scheller and Møller 1990, Obokata et al. 1993), or as PsaA through PsaL (Bryant 1992), respectively. The chloroplast-encoded subunits are highly conserved among a wide range of photosynthetic organisms in their primary structures, and constitute the central part of PSI (Bryant 1992). The nuclear-encoded subunits, on the other hand, are more variable during plant evolutions and constitute the peripheral part. PSI-D (Zanetti and Merati 1987, Zilber and Malkin 1988) and PSI-F (Wynn and Malkin 1988) are docking sites of ferredoxin and plastocyanin, respectively. PSI-E locates very near to PSI-D on the stromal surface of PSI (Andersen et al. 1992), and its involvement in cyclic electron flow is suggested (Yu et al. 1992).

Although chloroplast genes have single-copy organization in the organelle genome, many nuclear genes constitute multi-gene families and hence bring birth to sets of isoproteins. This situation is also true to the genes for PSI components in higher plants (Obokata et al. 1993, Yamamoto et al. 1993). For example, *Nicotiana* species have two *psaD* genes encoding slightly different proteins, and their expressions are subjected to different regulations during leaf development (Yamamoto et al. 1993). Alternative integration of the isoproteins results in molecular heterogeneity of PSI complex within a plant body, but its biological significance is not evident as yet (Obokata et al. 1993). It is an interesting subject whether this molecular diversity is involved in fine tunings of PSI function or biogenesis.

PSI-E is another example for the polymorphic subunit of PSI, and a set of isoforms with distinct electrophoretic mobilities are present in a wide range of plant species

Abbreviations: LDS, lithium dodecyl sulfate; PMSF, phenylmethanesulfonyl fluoride; TLCK, tosyl-lysine chloromethyl ketone; TPCK, tosyl-phenylalanine chloromethyl ketone.

(Obokata et al. 1993). In order to investigate the biological significance of this phenomenon, here we carried out the comparative analysis of the individual PSI-E isoforms derived from an isogenic plant line.

Materials and Methods

Preparation of PSI and electrophoresis—PSI complex was prepared from the leaves of *Nicotiana sylvestris* (Obokata et al. 1990), and subjected to LDS PAGE as previously described (Obokata et al. 1993).

cDNA cloning—A mixed oligonucleotide probe of 41 mer (5'-GCIGCIGA$^{A}/_{G}$GA$^{A}/_{G}$GCI-GCICCICCIGCIGCI-GCIGCIACIGC-3') was synthesized according to the N-terminal amino acid sequence (NH_2-AAEEAAPPAAAATA) of the PSI-E1 subunit from *N. sylvestris* (Obokata et al. 1993), and used for screening of a *Nicotiana sylvestris* cDNA library constructed in λgt10 as previously described (Hayashida et al. 1992). The inserts of positive clones were subjected to DNA sequencing analysis.

Genomic Southern hybridization—Genomic Southern analysis was carried out as previously described (Yamamoto et al. 1993).

Peptide mappings and protein sequencing—PSI subunits were resolved by LDS PAGE, and blotted to PVDF membranes (Obokata et al. 1993). The PSI-E1, E2, E3, and E4 subunits each were eluted from the membrane by incubating the membrane strips in 50 mM Tris-HCl (pH 9.0) containing 2% SDS and 1% Triton X-100 (Szewczyk and Summers 1988), and subsequently recovered as precipitates in 80% acetone. The isolated subunits were dissolved into 20 mM Tris-HCl (pH 9.0) containing 2 M urea, and then incubated with 2 pmole of lysil endopeptidase (Wako, Japan) over night at 30°C. The proteolytic digests were resolved by reverse-phase HPLC on a 5 μm Brownlee C18 column (2.1 × 30 mm) at 28°C. The elution system was a linear gradient between buffer A (0.1% trifluoroacetic acid) and buffer B (0.088% trifluoroacetic acid in 70% acetonitrile), and the gradient was from 0 to 100% buffer B in 50 min. The flow rate was 0.2 ml min^{-1}, and the elution of peptides was detected by monitoring the absorbance at 215 nm. The proteolytic fragments purified by HPLC were applied to a pulse-liquid protein sequenator, Applied Biosystems model 477 A.

Results

Electrophoretic dissection of PSI-E isoforms—Fig. 1 (left lane) represents a LDS PAGE profile of PSI proteins from *Nicotiana sylvestris*. PSI-D, PSI-F, PSI-L, PSI-E, PSI-H, PSI-G, and PSI-C subunits were identified with the aid of their N-terminal sequences (Obokata et al. 1993). PSI-E subunit was further resolved into four components by high resolution PAGE (Fig. 1, adjoining half lane), and these components were all shown to be PSI-E subunit by protein sequencing as described in the following paragraphs. We designated these proteins as PSI-E1, PSI-E2, PSI-E3 and PSI-E4, in the order of increasing mobility in LDS PAGE, and their apparent molecular masses are 14.4 kDa, 14.3 kDa, 14.1 kDa and 14.0 kDa, respectively. As *N. sylvestris* is an autogamous diploid species, occurrence of the four electrophoretically distinct isoforms may suggest that PSI-E is encoded by four distinct *psaE* genes in the nuclear genome. In order to examine this possibility, we actually analyzed *psaE* cDNA clones from *N. sylvestris*.

Isolation of psaE cDNA clones—psaE cDNA clones were isolated by use of mixed oligonucleotide probe whose sequences were derived from the N-terminal 14 amino acids of PSI-E1. From 8×10^4 phage clones of a library constructed from poly(A)$^+$ RNA of *N. sylvestris*, we isolated 6 clones which gave positive hybridization signals with the probe. The nucleotide sequences were determined for these clones, revealing that they all encode PSI-E subunit. Five of them were overlapping clones derived from the identical gene, and we designated this gene as *psaEa*. The remaining one clone was evidently derived from a different gene, and hence we denoted the second *psaE* gene as *psaEb*. The nucleotide sequences of full length *psaEa* and *psaEb* cDNA clones, ob5A3 and ya5A51, respectively, are presented in Fig. 2. These clones are 77% homologous at nucleotide sequence level (Fig. 2), and encoded proteins share 80.4% homology for the precursor proteins and 89.1% for the mature forms (Fig. 3).

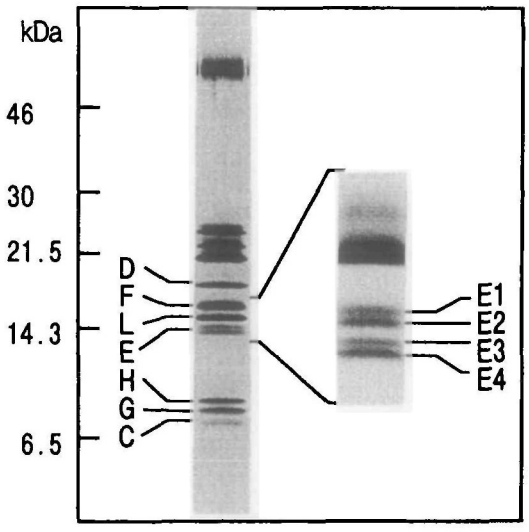

Fig. 1 LDS PAGE profiles of PSI-E isoforms from *Nicotiana sylvestris*. Left lane, PSI proteins of *N. sylvestris* were electrophoresed in a 15–23% polyacrylamide gel; right half lane, proteins corresponding to the part of the left lane were further resolved in long separating gel of 40 cm under high resolution condition (Obokata et al. 1993).

Using these cDNA clones as probe, we attempted to isolate the third and the fourth type of *psaE* cDNAs according to the following strategy. After the first screening of the cDNA library, we picked up the mixture of the recombinant phages around each positive signal, and subsequently the cDNA inserts of the phage clones in every mixtures were amplified by PCR using λgt10 primers. The amplified DNA fragments were digested with several kinds of restriction enzymes including *Eco*R I, *Hind* III, *Bam*H I, *Hinf* I, *Hha* I and *Taq* I, and then cutting patterns of the positive cDNA clones in every mixture were analyzed by agarose gel electrophoresis and visualized by Southern hybridization

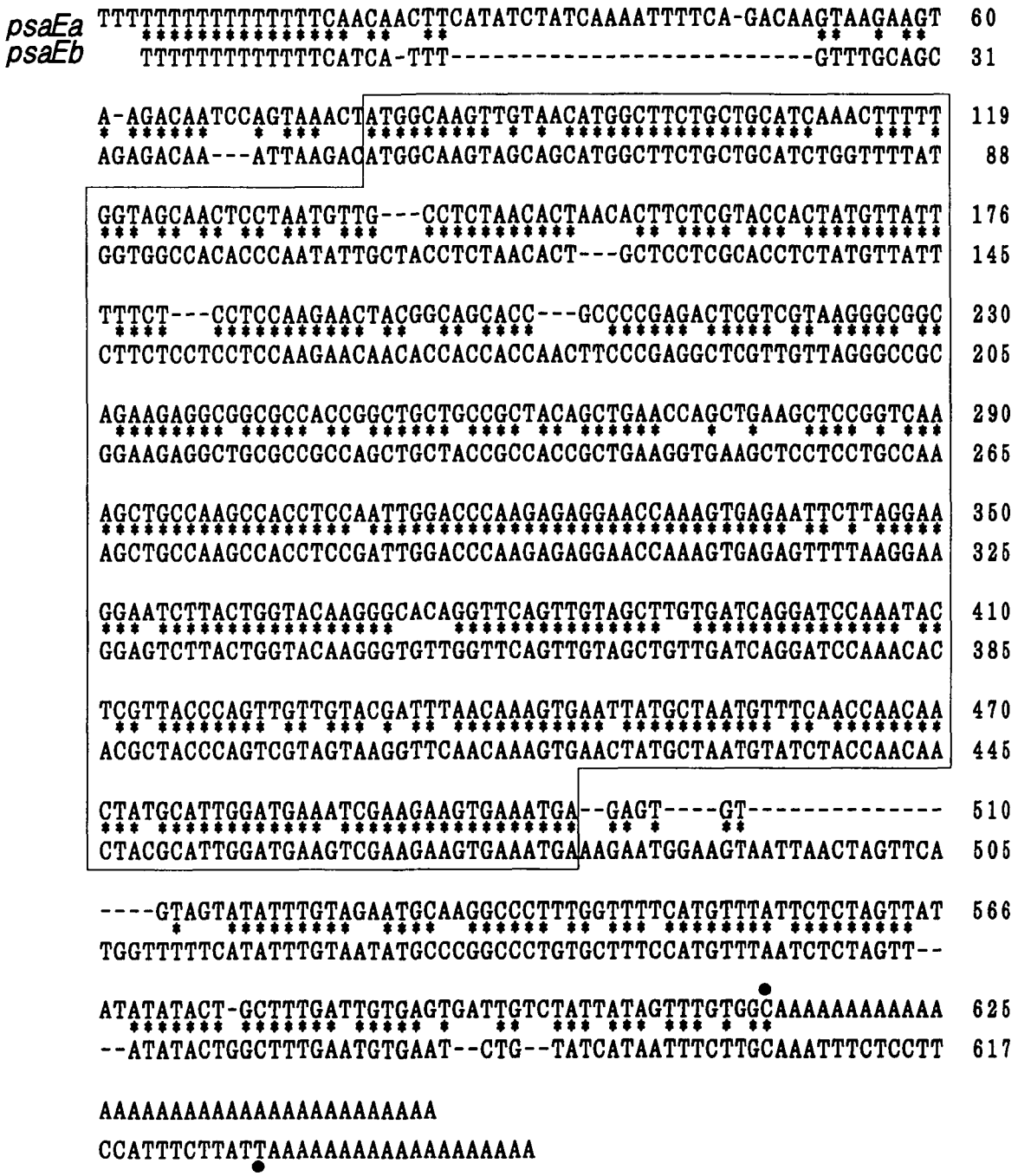

Fig. 2 Nucleotide sequences of cDNAs derived from distinct *psaE* genes. *psaEa* represents the sequence of ob5A1 clone; *psaEb*, ya5A51 clone. Protein coding regions are surrounded by a box.

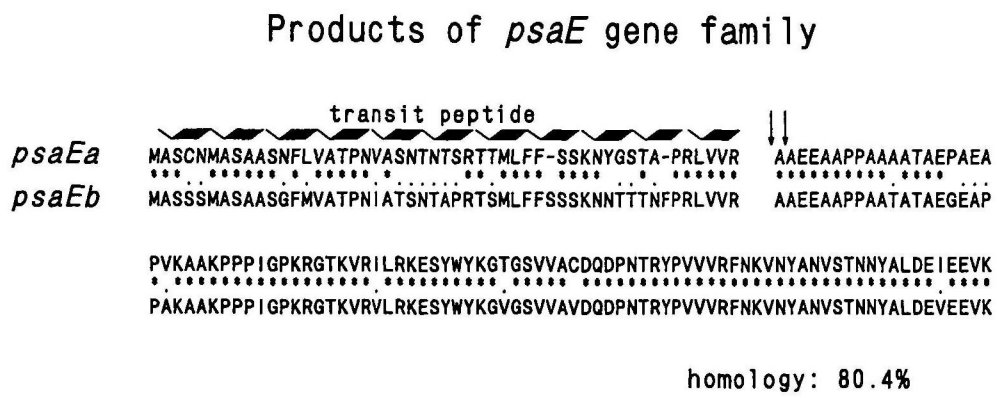

Fig. 3 Complete amino acid sequences of the PSI-E precursor proteins encoded by *psaEa* and *psaEb* genes. Arrows indicate the N-terminal residues determined by protein sequencing.

with the probe. From 1.8×10^5 recombinant phages we detected 44 positive clones, and they were classified into two types by this RFLP analysis method (data not shown). However, the cutting patterns of these two types were the same as those of *psaEa* and *psaEb* cDNAs, and we could not find the third or the fourth types of *psaE* cDNAs. Several of the positive clones obtained thus far were actually sequenced, and all of these were demonstrated to be either *psaEa* or *psaEb* cDNA as was expected from their RFLP profiles. These results imply that *psaE* has no more than two gene copies in the genome of *N. sylvestris*. This view of *psaE* copy number was examined by Genomic Southern analysis as is described in the next paragraph.

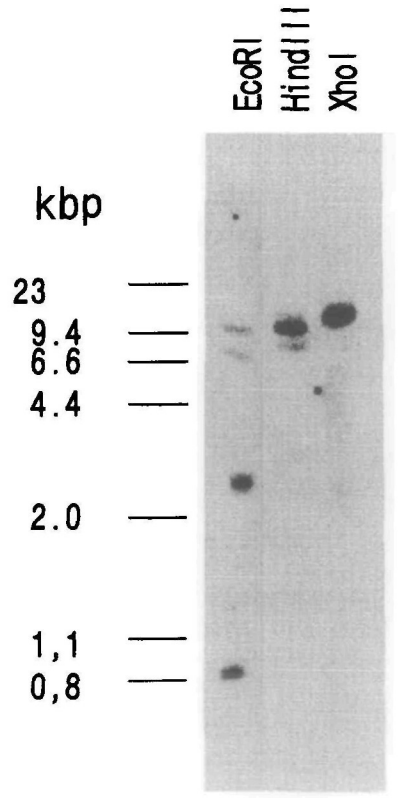

Fig. 4 Genomic Southern analysis of *psaE*. Restriction enzymes used for the digestion of *N. sylvestris* DNA are shown above the lanes. The blot was probed with a mixture of *psaEa* and *psaEb* cDNAs.

Genomic Southern analysis of psaE—Nuclear DNA of *N. sylvestris* was digested with *Eco*R I, *Hin*d III or *Xho* I, and subjected to Southern blot hybridization probed with a mixture of *psaEa* and *psaEb* cDNAs (Fig. 4). Four *Eco*R I fragments, two *Hin*d III fragments, and one *Xho* I fragment were found to be hybridized with *psaE* cDNAs. Since *Eco*R I gives one cutting site in the middle of *psaEa* cDNA, these hybridization signals suggest that *psaE* has two to three gene copies in the genome of *N. sylvestris*. This estimation of *psaE* copy number is consistent with that from cDNA analysis described above.

Structural comparison between PSI-E isoforms and putative gene products—PSI-E1, E2, E3 and E4 proteins each were isolated by LDS PAGE, digested with lysil endopeptidase to completion, and subsequently fractionated by a reverse phase HPLC column (Fig. 5). Elution profiles of the proteolytic fragments of PSI-E1 and PSI-E2 are very similar to each other, and likewise, PSI-E3 and PSI-E4 gave very similar elution profiles. These results indicate that the PSI-E1 through E4 proteins can be classified into two types, E1-E2 type and E3-E4 type, by their primary structures.

In order to compare the primary structures of these isoforms in more detail, their proteolytic fragments recovered from HPLC were subjected to amino acid sequence determination. Fig. 6 represents amino acid sequences determined for the recovered fragments, aligned with those deduced from the *psaEa* and *psaEb* cDNAs. Putative cleavage sites of lysil endopeptidase are shown by vertical arrows, and predicted fragments are denoted as P1

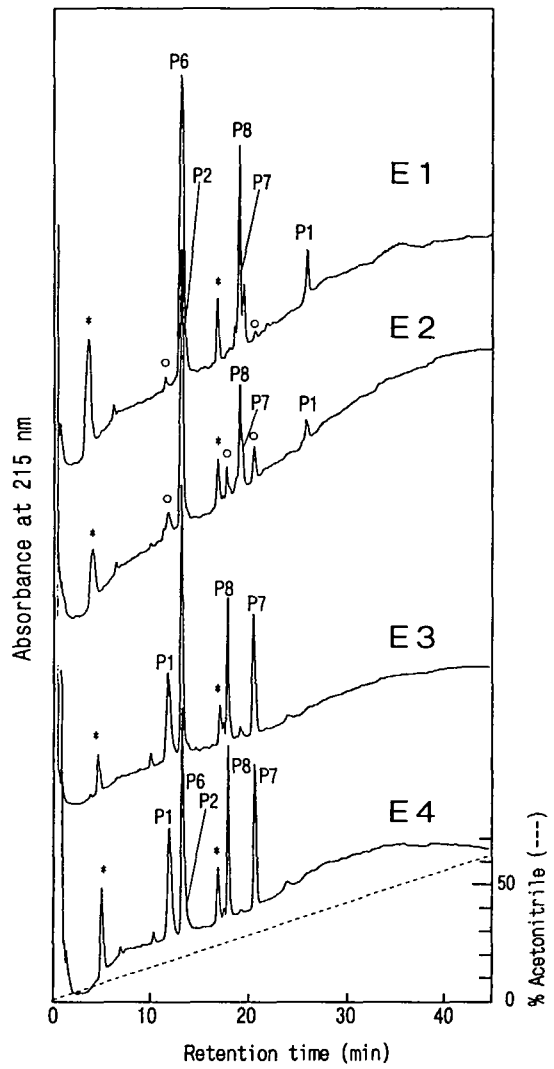

Fig. 5 HPLC elution profiles of the proteolytic digests of PSI-E isoforms. PSI-E1, E2, E3 and E4 each were isolated by LDS PAGE, digested with lysil endopeptidase, and subjected to reverse phase HPLC column. Asterisks indicate the peaks derived from urea. Open circles are referred to in Discussion.

through P8 (Fig. 6). We actually identified P1, P2, P6, P7 and P8 fragments in the PSI-E1 digest (Fig. 5 E1, Fig. 6), P1, P7 and P8 fragments in the PSI-E2 digest (Fig. 5 E2, Fig. 6), P1, P7 and P8 fragments in the PSI-E3 digest (Fig. 5 E3, Fig. 6), and P1, P2, P6, P7 and P8 fragments for PSI-E4 (Fig. 5 E4, Fig. 6). As far as these analyzed fragments are concerned, PSI-E1, PSI-E2, and a putative *psaEa* product have the same sequences except for the N-terminal alanine, which is present in both PSI-E1 and the *psaEb* product but absent from PSI-E2. Similarly, the sequences determined for the PSI-E3 and PSI-E4 fragments match with that of the *psaEb* product, and the only difference between PSI-E3 and PSI-E4 lies in respective presence or absence of N-terminal alanine.

Discussion

In this study, four PSI-E isoforms of *N. sylvestris* were classified into two types, E1–E2 type and E3–E4 type, by their peptide mappings with lysil endopeptidase and HPLC (Fig. 5). However, PSI-E2 gave some minor peaks whose retention time coincides with those of the P1, P7 and P8 of the E3–E4 type proteins (Fig. 5, open circle), and corresponding peaks were scarcely detected in PSI-E1. These minor peaks were probably due to the contamination of PSI-E2 with PSI-E3. The PSI-E3 is more abundant than PSI-E2, and, during preparative PAGE, the tailing of PSI-E3 tended to overlap the PSI-E2 fraction. Hence, PSI-E2 seems to provide almost the same peptide mapping profile with PSI-E1.

Taking these peptide mapping profiles (Fig. 5) and partial amino acid sequences (Fig. 6) into consideration, it seems most likely that PSI-E1 and PSI-E2 share the same amino acid sequences except for the N-terminal alanine which is present in PSI-E1 but absent from PSI-E2. Likewise, PSI-E3 and PSI-E4 appear to have the same sequences except for N-terminal alanine. These views were buttressed by the cDNA analysis in which we could not find more than two kinds of *psaE* cDNAs even after extensive survey on 1.8×10^5 cDNA clones (Fig. 2). In addition, these two *psaE* genes, *psaEa* and *psaEb*, appear to be sufficient for encoding all the PSI-E isoforms, PSI-E1 through PSI-E4 (Fig. 6). All these results strongly suggest that PSI-E1 and PSI-E2 are produced from the same precursor proteins directed by *psaEa* gene, and that PSI-E3 and PSI-E4 are from *psaEb*.

It is unlikely that PSI-E2 and PSI-E4 are respective artifacts of PSI-E1 and PSI-E3 by losing their N-terminal alanines, because both PSI-E1 and E2, and PSI-E3 and E4, always occurred in 1 : 1 ratio in many batches of thylakoid membrane and PSI preparations. In addition, presence or absence of proteinase inhibitors (0.25 mM each of PMSF, TLCK and TPCK) during purification steps of chloroplasts and PSI complex did not cause any change in the abundance ratio among these four isoforms at all (data not shown). Similar phenomenon on the N-terminal heterogeneity was reported in spinach thioredoxin *m* (Wedel et al. 1992). Thioredoxin *m* is encoded by the nuclear genome and post-translationally imported into the stroma of chloroplasts. Spinach thioredoxin *m* has N-terminally heterogeneous isoforms in vivo (Maeda et al. 1986) as in the case of *N. sylvestris* PSI-E protein, and Wedel et al. demonstrated that two isoforms of equimolar abundances arose from the same precursor protein imported into isolated spinach chloroplasts (Wedel et al. 1992). It is likely that similar mechanisms operate, in chloroplasts, for the precur-

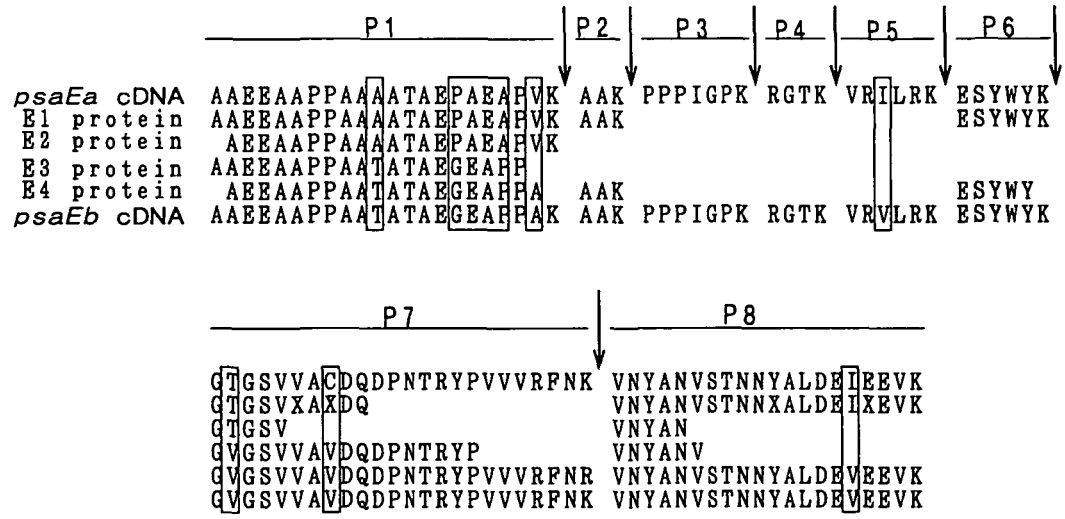

Fig. 6 Comparison of amino acid sequences among PSI-E isoforms and the putative products of *psaEa* and *psaEb* cDNAs. Vertical arrows indicate putative cutting sites of the *psaEa* and *psaEb* products by lysil endopeptidase, and P1 through P8 represent predicted proteolytic fragments. Residues indicated by boxes are different between *psaEa* and *psaEb* products.

sors of thioredoxin *m* and PSI-E to produce N-terminally heterogeneous isoforms. These phenomena may afford a clue to the molecular recognition mechanism of chloroplast processing machineries.

Nucleotide sequences of *psaE* cDNAs, and hence their deduced amino acid sequences for the PSI-E precursor proteins, have been reported in spinach (Münch et al. 1988), barley (Okkels et al. 1988), and *Chlamydomonas* (Franzén et al. 1989). Their amino acid sequences differ in the transit peptide and N-terminal regions except for the vicinity of the processing sites; RLVVRA(A/E)E is conserved among these organisms and *N. sylvestris*. Within this consensus motif, E1–E3 type N-termini, the alanine next to arginine, were found in barley (Okkels et al. 1988) and spinach (Münch et al. 1988, Zilber and Malkin 1992), while E2–E4 type which is next to the above alanine was reported in *Chlamydomonas* (Franzén et al. 1989) and spinach (Zilber and Malkin 1992). It requires further examination whether the N-terminal heterogeneity of PSI-E occurs only in limited number of organisms or ubiquitous in higher plants and algae. Both of the N-terminal sites, E1–E2 type and E3–E4 type, fit a cleavage site rule of chloroplast transit peptides (Gavel and von Heijne 1990).

This study demonstrated that the polymorphism of PSI-E in *N. sylvestris* is caused by two different sources, namely a couple of gene copies and post-translational modifications of each gene product. The next problem about PSI-E polymorphism is whether individual isoforms or corresponding gene copies have differentiated functions or not. In fact, several nuclear-encoded PSI subunits have two isoforms in a wide range of higher plants (Obokata et al. 1993). We examined the organ-specific expression of *psaE* genes with the aid of immunoblot and Northern hybridization analyses. However, we could not find significant difference in tissue specific nor developmental stage specific manners between the *psaEa* and *psaEb* products at both mRNA and protein levels (data not shown). High homology between these gene products imply that their tertially structures are scarcely different (Figs. 3, 4 and 5). Hence, at present there is no evidence supporting the functional difference among PSI-E isoforms.

The authors express their thanks to Drs. M. Sugiura, H. Tsuji, and A. Kimura for their useful discussions and encouragement, and to Ms. Y. Kondo for her help in some immunoblot experiments. This study was carried out in part at the Research Center for Molecular Genetics, Hokkaido University, and amino acid sequence analysis and HPLC analysis were performed at Biopolymer Analysis Laboratory, Faculty of Agriculture, Hokkaido University. Antisera raised against spinach PSI proteins was kindly supplied from Dr. H. Oh-oka. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science and Culture in Japan.

References

Andersen, B., Koch, B. and Scheller, V. (1992) Structural and functional analysis of the reducing side of photosystem I. *Physiol. Plant.* 84: 154–161.

Bryant, D.A. (1992) Molecular biology of photosystem I. *In* The Photosystems: Structure, Function and Molecular Biology. Edited by Barber, J. pp. 501–549. Elsevier Science Publishers, Amsterdam.

Franzén, L-G., Frank, G., Zuber, H. and Rochaix, J-D. (1989) Isolation and characterization of cDNA clones encoding the 17.9 and 8.1 kDa subunits of photosystem I from *Chlamydomonas reinhardtii*. *Plant Mol. Biol.* 12: 463–474.

Gavel, Y. and von Heijne, G. (1990) A conserved cleavage-site motif in chloroplast transit peptides. *FEBS Lett.* 261: 455–458.

Hayashida, N., Izuchi, S., Sugiura, M. and Obokata, J. (1992) Nucleotide sequence of cDNA clones encoding PSI-H subunit of photosystem I in tobacco. *Plant Cell Physiol.* 33: 1031–1034.

Ikeuchi, M. (1992) Subunit proteins of photosystem I. *Plant Cell Physiol.* 33: 669–676.

Maeda, K., Tsugita, A., Dalzoppo, D., Vilbois, F. and Shürmann, P. (1986) Further characterization and amino acid sequence of m-type thioredoxins from spinach chloroplasts. *Eur. J. Biochem.* 154: 197–203.

Münch, S., Ljungberg, U., Steppuhn, J., Schneiderbauer, A., Nechshtai, R., Beyreuther, K. and Herrmann, R.G. (1988) Nucleotide sequences of cDNAs encoding the entire precursor polypeptides for subunits II and III of the photosystem I reaction center from spinach. *Curr. Genet.* 14: 511–518.

Obokata, J., Mikami, K., Hayashida, N. and Sugiura, M. (1990) Polymorphism of a photosystem I subunit caused by alloploidy in *Nicotiana*. *Plant Physiol.* 92: 273–275.

Obokata, J., Mikami, K., Hayashida, N., Nakamura, M. and Sugiura, M. (1993) Molecular heterogeneity of photosystem I: *psaD, psaE, psaF, psaH*, and *psaL* are all present in isoforms in *Nicotiana* spp. *Plant Physiol.* 102: 1259–1267.

Okkels, J.S., Jepsen, L.B., Hønberg, L.S., Lehmbeck, J., Scheller, H.V., Brandt, P., Høyer-Hansen, G., Stummann, B., Henningsen, K.W., von Wettstein, D. and Møller, B. (1988) A cDNA clone encoding a 10.8 kDa photosystem I polypeptide of barley. *FEBS Lett.* 237: 108–112.

Scheller, H.V. and Møller, B.L. (1990) Photosystem I polypeptides. *Physiol. Plant.* 78: 484–494.

Szewczyk, B. and Summers, D.F. (1988) Preparative elution of proteins blotted to immobilon membranes. *Anal. Biochem.* 168: 48–53.

Wedel, N., Clausmeyer, S., Herrmann, R.G., Gardet-Salvi, L. and Shürmann, P. (1992) Nucleotide sequence of cDNAs encoding the entire precursor polypeptide for thioredoxin m from spinach chloroplasts. *Plant Mol. Biol.* 18: 527–533.

Wynn, R.M. and Malkin, R. (1988) Interaction of plastocyanin with photosystem I: a chemical cross-linking study of the polypeptide that binds plastocyanin. *Biochemistry* 27: 5863–5869.

Yamamoto, Y., Tsuji, H. and Obokata, J. (1993) Structure and expression of a nuclear gene for PSI-D subunit of photosystem I in *Nicotiana sylvestris*. *Plant Mol. Biol.* 22: 985–994.

Yu, L., Golbeck, J.H., Zhao, J., Schluchter, W.M., Mühlenhoff, U. and Bryant, D. (1992) The PsaE protein is required for cyclic electron flow around photosystem I in the cyanobacterium *Synechococcus* PCC 7002. *In* Research in Photosynthesis. Edited by Murata, N. Vol. 1, pp. 565–568. Kluwer Academic Publishers, Dordrecht.

Zanetti, G. and Merati, G. (1987) Interaction between photosystem I and ferredoxin: identification by chemical cross-linking of the polypeptide which binds ferredoxin. *Eur. J. Biochem.* 169: 143–146.

Zilber, A. and Malkin, R. (1992) Organization and topology of photosystem I subunits. *Plant Physiol.* 99: 901–911.

(Received October 8, 1993; Accepted November 30, 1993)