Plant Physiol. (1993) 102: 1259-1267

Molecular Heterogeneity of Photosystem I[1]

psaD, psaE, psaF, psaH, and *psaL* Are All Present in Isoforms in *Nicotiana* spp.

Junichi Obokata*, Kohki Mikami, Nobuaki Hayashida, Masayuki Nakamura, and Masahiro Sugiura
Department of Botany, Faculty of Science, Hokkaido University, Sapporo 060, Japan (J.O., K.M., M.N.); The Institute of Physical and Chemical Research, Tsukuba 305, Japan (N.H.); and Center for Gene Research, Nagoya University, Nagoya 464-01, Japan (M.S.).

The protein composition of photosystem I (PSI) was examined in *Nicotiana* spp. by high-resolution polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and immunoblot analysis. Five PSI proteins show polymorphism in an amphidiploid species, *Nicotiana tabacum*, but not in its ancestral diploid species, *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. These *Nicotiana* spp. appear to have at least 18 PSI proteins per genome that range in molecular mass from 3 to 20 kD. They include the products of nuclear genes *psaD, psaE, psaF, psaG, psaH, psaK,* and *psaL*, the product of chloroplast gene *psaC*, N-terminally blocked proteins of 4.5 and 3.0 kD, and an unidentified protein of 12.5 kD. The *psaD, psaF, psaH,* and *psaL* products have two isoforms each that are distinguished by different mobilities in polyacrylamide gel electrophoresis, and the *psaE* product has four isoforms. The two isoforms of the *psaD* product have distinct amino acid sequences, indicating that they are encoded by different genes within the genome. Four isoforms of the *psaE* products can be classified into two groups by N-terminal amino acid sequence, indicating that at least two *psaE* genes are present in the genome. To examine whether the polymorphic nature of PSI is peculiar to *Nicotiana* spp., we carried out immunoblot analysis of the *psaD* and *psaE* products in isogenic lines of tomato (*Lycopersicon esculentum*), *Arabidopsis thaliana*, red bean (*Vigna angularis*), and corn (*Zea mays*). Two electrophoretically distinct isoforms were found for the *psaD* products of tomato, *A. thaliana*, and corn, and two isoforms of *psaE* products were detected in tomato, *A. thaliana*, and red bean. These results suggest that the nuclear-encoded subunits of PSI, except for the *psaG* and *psaK* products, generally have two isoforms.

PSI is a multiprotein complex in the thylakoid membrane of chloroplasts and mediates light-driven electron transfer from plastocyanin to Fd. Because the biogenesis of PSI is dependent on the nuclear genome as well as the chloroplast genome (Obokata, 1987; Scheller and Møller, 1990; Herrmann et al., 1991; Bryant, 1992; Ikeuchi, 1992), it provides an interesting system for studying the nuclear-chloroplast relationship in plant cells.

PSI consists of two large subunits of 82 kD and several small subunits of less than 20 kD. Until recently, the subunit composition of PSI has been a subject of confusion, but now it is considered to contain at least 12 subunits, and the genes encoding these subunits are designated *psaA* through *psaL* (Bryant, 1992; Ikeuchi, 1992). In higher plants, *psaA, psaB, psaC, psaI,* and *psaJ* are chloroplast genes, and the remainder are nuclear encoded. The two large subunits, the products of *psaA* and *psaB*, bear the reaction center pigment P700, the primary electron acceptor A_0, and the intermediate electron acceptors A_1 and F_X (Golbeck, 1992). A small subunit of about 9 kD is encoded by *psaC* and has iron/sulfur centers F_A and F_B that accept electrons from F_X (Scheller and Møller, 1990; Golbeck, 1992; Ikeuchi, 1992). These subunits are central members of the reaction center and are highly conserved over a wide range of photosynthetic organisms (Shimada and Sugiura, 1991; Bryant, 1992). The *psaD* and *psaF* products are the docking sites of Fd (Zanetti and Merati, 1987; Zilber and Malkin, 1988) and plastocyanin (Wynn and Malkin, 1988), respectively. The functions of the other PSI subunits remain to be resolved.

Although 12 PSI gene products have been identified in cyanobacteria and in a few higher plants to date (Scheller and Møller, 1990; Iwasaki et al., 1991; Bryant, 1992; Ikeuchi, 1992), inconsistent reports of the number and molecular masses of the small subunits have created confusion concerning this topic (Golbeck, 1987). We previously suggested that alloploidy, which is found in many cultivated plant species, is sometimes responsible for the polymorphism of PSI subunits (Obokata et al., 1990). Multigene families are another source of protein polymorphism. Fd is encoded by a multigene family, at least in higher plants (Hase et al., 1991a), and, hence, consists of several isoforms in plant cells. Although little is directly understood about the function of the PSI protein isoforms, multigene families are known to provide a diversity of physiological capabilities and permit fine tuning of cellular processes in many eukaryotic systems over a range of physiological conditions (Singer and Berg, 1991).

In this study we examined the subunit composition of PSI complexes in *Nicotiana* spp., with special emphasis on subunit polymorphism. Because *Nicotiana* spp. have been extensively

[1] This work was supported in part by Grants-in-Aid from the Ministry of Education, Science, and Culture in Japan and from the Akiyama Foundation (J.O.).

* Corresponding author; fax 81-11-757-5994.

Abbreviations: LDS, lithium dodecyl sulfate; ; LHCI(II), light-harvesting chlorophyll-protein complex I(II); TLCK, tosyl-lysine chloromethyl ketone; TPCK, tosyl-phenylalanine chloromethyl ketone.

studied genetically (Smith, 1968; Gray et al., 1974; Kung et al., 1982; Shimada and Sugiura, 1991), this genus is suitable for studies of divergence of genes and their products during plant evolution. *Nicotiana tabacum* is an amphidiploid species and is thought to be the progeny of a cross between two autogamous diploid species, *Nicotiana sylvestris* and *Nicotiana tomentosiformis*, which derived from phylogenetic branches that diverged at the early stages of *Nicotiana* evolution (Smith, 1968; Kung et al., 1982). Comparative studies of the PSI subunits among these species have permitted us to discriminate subunit polymorphisms that can be attributed to either alloploidy or multigene families and, consequently, has revealed a set of PSI proteins commonly encoded by each nuclear (diploid) genome. One diploid set of PSI proteins includes all of the isoprotein variants that are products of multigene families and appear to be essential to the fine regulation of PSI biogenesis and function in plant cells. The variant PSI proteins that are present because of alloploidy are not included in a diploid set.

Here we present a detailed analysis of the protein composition of PSI in *Nicotiana* sp. and clearly elaborate the highly polymorphic nature of the PSI complex for the first time. In addition, we also provide evidence that the polymorphic nature of this complex is not peculiar to *Nicotiana* spp. but is ubiquitous in higher plants.

MATERIALS AND METHODS

Plant Material

Nicotiana tabacum cv Bright Yellow 4, *Nicotiana sylvestris, Nicotiana tomentosiformis, Nicotiana otophora, Lycopersicon esculentum*, VFNT LA line 1221, *Arabidopsis thaliana* ecotype Columbia, *Vigna angularis* cv Chagarawase, and an inbred *Zea mays* line, Ho9, were grown in a greenhouse.

Preparation of PSI

Chloroplasts were isolated from the leaves of *Nicotiana* spp. by silica sol gradient centrifugation (Obokata, 1987), and subsequently PSI was purified as previously described (Obokata, 1986) with slight modifications. All of the following procedures were carried out at 0°C. Chloroplast membranes were washed with 2 M NaBr and dissolved in 0.4 M Suc, 10 mM NaCl, 5 mM $MgCl_2$, and 10 mM Tricine-KOH (pH 8.3) to give a Chl concentration of 1.1 mg mL^{-1}. Digitonin was added to a concentration of 1% (w/v), and then the mixture was incubated for 1 h and centrifuged at 30,000*g* for 10 min. NaCl and digitonin were added to the supernatant to concentrations of 0.1 M and 1.5% (w/v), respectively, and the preparation was incubated overnight. The mixture was subsequently spun at 180,000*g* for 2 h, and the pellet was resuspended in 0.4 M Suc, 10 mM NaCl, 10 mM Hepes-NaOH (pH 8.0), and 5% (w/v) Triton X-100 to give a final Chl concentration of 1 to 3 mg mL^{-1}. Following an overnight incubation, the mixture was applied to a DEAE-cellulose column (Whatman DE-32) equilibrated with 50 mM Tris-HCl (pH 8.0) and 0.2% (w/v) Triton X-100. The column was washed once with the equilibration buffer and then eluted with the same buffer containing 0.13 M NaCl. Fractions rich in PSI proteins were collected and subsequently subjected to Suc density gradient centrifugation (5–25% [w/v] Suc, 0.2% [w/v] Triton X-100, 20 mM Tris-HCl [pH 8.0]) at 100,000*g* for 15 h. The lower green band was enriched for PSI. When PSI was isolated in the presence of proteinase inhibitors, 1 mM each of PMSF, TLCK, and TPCK were included in all steps from the lysis of intact chloroplasts to Suc gradient centrifugation.

PAGE

For LDS-PAGE, a combination of (a) long-resolving gels (40 × 20 × 0.1 cm), (b) cooling the gel plate at 4°C with an electric fan, (c) a high concentration of Tris in the resolving gel buffer, and (d) a gradient of acrylamide concentration (15–23% [w/v]) resulted in high-resolution electrophoretograms, by which we could discriminate proteins of 19 kD or less differing in length by one amino acid residue. A discontinuous buffer system was used, essentially according to the method of Fling and Gregerson (1986). The stacking and resolving gel buffers were 125 mM Tris-HCl (pH 6.8) and 750 mM Tris-HCl (pH 8.8), respectively. The electrode buffer contained 50 mM Tris and 0.384 M Gly with unadjusted pH. SDS was added only to the electrode buffer in the upper reservoir to give a concentration of 0.1% (w/v). The PSI samples were dissolved in sample buffer containing 50 mM Tris-HCl (pH 6.8), 2% (w/v) LDS, 2% (w/v) 2-mercaptoethanol, and 5% (w/v) Suc at room temperature. Electrophoresis was performed once at room temperature while the sample migrated in the stacking gel and again at 4°C after it entered the resolving gel. For two-dimensional PAGE analysis, the first-dimension PAGE was carried out as described above, and the second-dimension PAGE was performed with 12% (w/v) resolving gels (30 cm in length) in a resolving gel buffer of 375 mM Tris-HCl (pH 8.8) at 4°C. The protein bands in the gels were detected by silver staining (Wako Ltd., Japan) or Coomassie brilliant blue staining. Densitometric analysis of the gels stained with Coomassie brilliant blue was carried out with a Shimadzu scanner, model CS-930.

Protein Sequencing

The electrophoretically separated PSI subunits were blotted onto polyvinylidene difluoride membranes and subsequently applied to a gas-phase protein sequenator (Applied Biosystems model 470A) to determine the N-terminal amino acid sequences (Obokata et al., 1990).

Preparation of Chloroplast Membranes

The leaves of the indicated plant species were homogenized with a Polytron homogenizer in isolation medium (0.35 M sorbitol, 2 mM EDTA, 2 mM sodium isoascorbate, and 25 mM Hepes-NaOH [pH 7.6]), and the resulting material was filtered through eight layers of gauze and centrifuged at 12,000*g* for 10 min at 0°C. The pellets containing plastids were resuspended in isolation medium and then layered onto a 50% (w/v) Suc cushion containing 2 mM EDTA and 25 mM Hepes-NaOH (pH 7.6), followed by centrifugation at 12,000*g* for 30 min at 0°C. The green bands on the Suc cushions were recovered and washed three times with 5 mM Tris-HCl (pH 7.5) to obtain chloroplast membrane fractions.

Immunoblot Analysis

Proteins of the PSI complex and the chloroplast membranes were subjected to LDS-PAGE and then electrophoretically blotted onto polyvinylidene difluoride membranes. The western blots were probed with rabbit anti-PSI-D, anti-PSI-E, and anti-PSI-L antibodies and then reacted sequentially with biotinylated donkey antibody against rabbit immunoglobulin (Amersham) and streptavidin-alkaline phosphatase conjugate (Amersham). Specific antibodies against PSI-D and PSI-E subunits were purified from rabbit antiserum raised against spinach PSI complexes (Oh-oka et al., 1989) according to the protocol of Kelly et al. (1986). An anti-PSI-L rabbit antibody was raised against the PSI-L protein of barley (Okkels et al., 1991; Andersen et al., 1992).

Nomenclature for PSI Proteins

In this work, the PSI genes are designated according to the nomenclature proposed by Hallick (1989). The PSI subunits are identified by the corresponding gene name: for example, the products of the genes of *psaD* and *psaE* are denoted PSI-D and PSI-E, respectively. The isoforms of PSI subunits, which are derived from a single diploid genome, and not present because of alloploidy, are numbered in order of decreasing molecular mass: for example, the high and low molecular mass isomers of PSI-D are denoted PSI-D1 and PSI-D2, respectively. If, for example, multiple copies of a given PSI gene are present within a diploid genome, we denote them *psaDa* and *psaDb* in the order in which they were cloned.

RESULTS

Isolation of PSI Complex from *Nicotiana* spp.

PSI complexes were isolated from *Nicotiana* spp. and subjected to Suc gradient centrifugation. Figure 1 shows a high-resolution electrophoretogram of the PSI proteins from *N. tabacum*, with the left side of the panel corresponding to the top of the Suc gradient and the right side to the bottom. Twenty proteins appear to be associated with the large subunits (PSI-A and PSI-B) as they sedimented through the Suc gradient. The peak of PSI distribution on the gradient appears in lanes C in Figure 1. We refer to the enriched PSI complex in this fraction as the PSI core complex. A part of the PSI complex appears to be associated with 20- to 23-kD proteins and sedimented in heavier fractions than did the core complex (Fig. 1, lanes N). This 20-kD protein (Fig. 1B, arrowhead) had a blocked N terminus, but the amino acid sequence was partially determined after HCl hydrolysis and found to be SAPGXFGFXPLG, which matches the partial sequence of the product of the type I PSI *Cab* genes, *Cab-6A* and *Cab-6B* (Pichersky et al., 1987). We refer to the PSI complex enriched in lanes N as the native PSI complex, which refers to the PSI core complex bound to LHCI. When PSI was prepared in the presence of proteinase inhibitors (1 mM each of PMSF, TLCK, and TPCK), the yield of the native complex was substantially greater, with only slight recovery of the core complex (data not shown). Very low levels of PSII proteins and LHCII were present in the fractions above the PSI core fraction (Fig. 1B, lane L), but the electrophoretic images of these PSII proteins were very faint and did not reproduce well in this photograph.

Characterization of PSI Proteins by N-Terminal Sequence and Immunoblot Analysis

Figure 2A represents a part of the high-resolution electrophoretogram of the PSI core proteins from *N. tabacum*, corresponding to apparent molecular masses of 3 to 20 kD. At least 24 proteins can be distinguished in this photograph, and we tentatively refer to them as number 1 (highest molecular mass) through number 24 (lowest molecular mass, Fig. 2A). These proteins, except for numbers 19, 21, and 22, produced sedimentation profiles very similar to those of the PSI-A/B subunits in the Suc gradient (Fig. 1), indicating that they are components of the PSI core complex. The sedimentation profile of protein 21 is quite different from that of the PSI complex, and the amount of this protein in PSI fractions correlates with that of the LHCII contaminants (data not

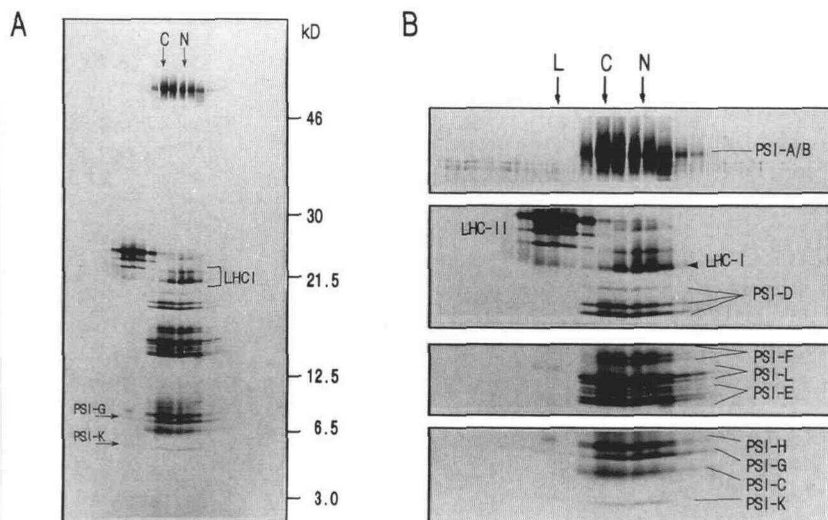

Figure 1. LDS-PAGE profiles of PSI complexes from *N. tabacum* fractionated after Suc gradient centrifugation. Parts of A are enlarged in B. Left side, Top of the Suc gradient; right side, bottom of the Suc gradient. PSI native complexes, PSI core complexes, and LHCII complexes are most abundant in the fractions labeled N, C, and L, respectively. An arrowhead indicates type I PSI *Cab* protein. Protein bands were visualized by silver staining.

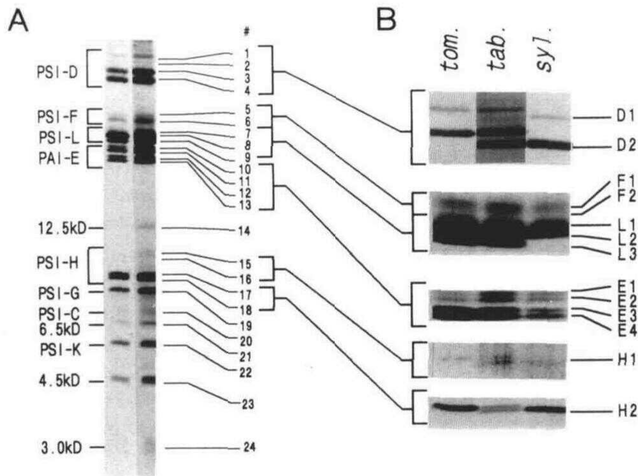

Figure 2. LDS-PAGE profiles of PSI proteins from *Nicotiana* spp. A, A part of the high-resolution electrophoretogram of PSI core proteins from *N. tabacum*, in the apparent molecular mass range of 3 to 20 kD. B, Comparison of PSI proteins between *N. tomentosiformis* (tom.), *N. tabacum* (tab.), and *N. sylvestris* (syl.). Protein bands were visualized by silver staining.

shown). Hence, we conclude that protein 21 is not a PSI component, but we suggest that it may be functionally and/or physically related to LHCII. Proteins 19 and 22 are described in more detail below.

The proteins shown in Figure 2A were subjected to N-terminal amino acid sequence analysis (Fig. 3), and PSI-D, PSI-F, PSI-E, PSI-H, PSI-G, PSI-C, and PSI-K subunits were identified among them (indicated at the left of Fig. 2A). It is clear from this analysis that PSI-D, PSI-F, PSI-E, and PSI-H are polymorphic in *N. tabacum*.

Proteins 5 to 13 in Figure 2A have apparent molecular masses of approximately 15 kD in PAGE profiles, and their electrophoretic mobilities are very similar to each other. We used two-dimensional gels to further analyze these proteins in *N. sylvestris*, because *N. sylvestris* is a diploid ancestor of *N. tabacum*, an amphidiploid, and its protein composition is simpler than that of *N. tabacum*. Figure 4A clearly demonstrates that *N. sylvestris* contains eight proteins that migrate with apparent molecular masses of approximately 15 kD in the first dimension. N-terminal amino acid sequence analysis revealed that they include four PSI-E (Figs. 3 and 4A) and two PSI-F isoforms (Figs. 3 and 4A) and two proteins with blocked N termini. The blocked proteins were identified as the PSI-L subunit, using an antibody directed against the barley PSI-L subunit (Fig. 4B).

Proteins 21, 23, and 24 in Figure 2A have blocked N termini; therefore, these N-terminal sequences could not be identified. The abundance of protein 14 was very low relative to the other subunits, and we could not obtain clear results with respect to its N-terminal amino acid sequence.

Subunit Polymorphism due to Alloploidy in *N. tabacum*

As summarized in Figure 2A, the PSI-D, PSI-F, PSI-L, PSI-E, and PSI-H subunits were found to be polymorphic in *N. tabacum*. Because *N. tabacum* is an amphidiploid with a genome composed essentially of two diploid genomes, we examined the extent to which alloploidy results in polymorphism of these subunits. *N. tabacum* has 24 pairs of chromosomes and is thought to be a hybrid between two autogamous diploid species, *N. sylvestris* and *N. tomentosiformis*, both of which have 12 pairs of chromosomes (Smith, 1968; Kung et al., 1982). Comparison of the PSI proteins among *N. tabacum*, *N. sylvestris*, and *N. tomentosiformis* reveals that alloploidy results in polymorphism in the PSI-D, PSI-L, and PSI-H subunits but not in PSI-F and PSI-E in *N. tabacum* (Fig. 2B). The electrophoretic profiles of the other PSI proteins are almost identical among these three plant species (data not shown).

N. sylvestris and *N. tomentosiformis* each has two types of PSI-D proteins (Fig. 2B). We refer to the high molecular mass-type PSI-D protein in each species as PSI-D1 and the low molecular mass type as PSI-D2 (indicated in Fig. 2B). *N. tabacum* has four PSI proteins including two PSI-D1 and two PSI-D2 proteins. One PSI-D1 protein is derived from *N.*

```
PSI-D

 (#1) D1 tab 19.3 kD  ADEKAQSATKEAEPAAPVGFTP
      D1 syl 19.0 kD  AVEKAQSATKEAEPA
                      | | |  ||||||
      D2 tom 18.5 kD  AEEAAAAATKEAEAP
 (#3) D2 tab 18.5 kD  AEEAAAAATKEAEAPVGFTPPQLDPNTPSXIFGG
 (#4) D2 tab 17.5 kD  AEEAA---TKEAEAPVGFTPPQLDPNTPSXIFG
      D2 syl 17.5 kD  AEEAA---TKEAEA

PSI-F

 (#5) F1 tab 16.1 kD  DISGLTPXKESKQFAK
      F1 syl 16.1 kD  DISGLTPXKESKQFAKXEKQXXKKL
                      ||||||| ||||||||
 (#6) F2 tab 15.9 kD  DISGLTPXKESKQFAK
      F2 syl 15.9 kD  DISGLTPXKESKQFXXXEXQ

PSI-E

(#10) E1 tab 14.4 kD  AAEEAAPPAAAATAEPAEAPVKAXKPXXIG
      E1 syl 14.4 kD  AAEEAAPPAAAATAEP
                       |||||||||||
(#11) E2 tab 14.3 kD   AEEAAPPAAAA
      E2 syl 14.3 kD   AEEAAPPAAAATAEPAEAPVKA
                       ||||||||| ||||    |
      E3 syl 14.1 kD  AAEEAAPPAATATAEGEAPP
(#12) E3 tab 14.1 kD  AAEEAAPPAA
                       |||||||||
(#13) E4 tab 14.0 kD   AEEAAPPAAT
      E4 syl 14.0 kD   AEEAAPPAATATAEGEAPPA

PSI-H

      H1 syl 10.4 kD  KYGDK
                      |||||
(#17) H2 tab  9.2 kD  KYGDKSVYFDLEDLGNTTGQXDXY
(#18) H2 tab  9.0 kD  KYGDKSVYFDLEDLANTTGQXDXY

PSI-G

(#19)    tab  8.0 kD  ELNTSLVISLSTGLSQFXG
         syl  8.0 kD  ELNTSLVISLSTGLSLFLGGFVF

PSI-C

(#20)    tab  7.5 kD  SHSVKIYDTXIGXTQXVXAXPTDVLEMI

PSI-K

(#22)    tab  5.6 kD  DFIGSPTNLIMVTSTXLM
         syl  5.6 kD  DFIGSPTNLIMVTSTSLMLFAGGFGLAP
```

Figure 3. N-terminal amino acid sequences determined for some PSI proteins in *Nicotiana* spp. Abbreviations and numbers (1–22) correspond to the bands in Figure 2A (highest to lowest molecular mass, respectively).

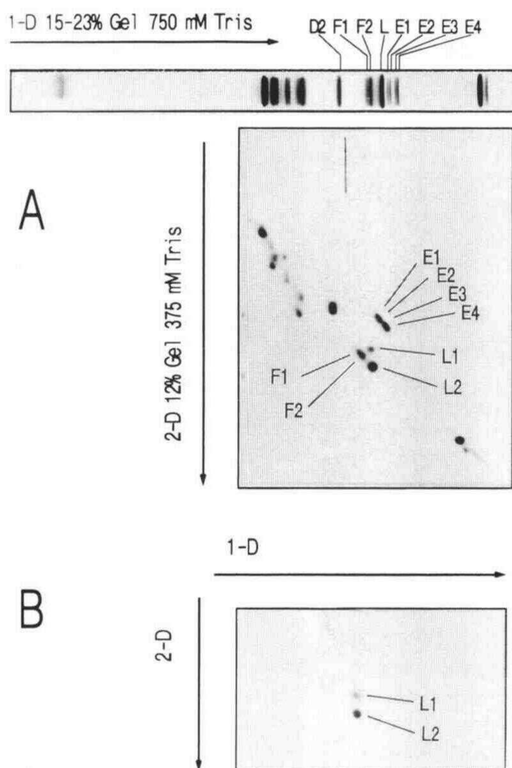

Figure 4. Two-dimensional electrophoretogram of PSI proteins from *N. sylvestris*. A, Proteins were stained by Coomassie brilliant blue. B, The PSI-L subunit was detected by antibodies raised against barley PSI-L.

sylvestris and one from *N. tomentosiformis*. The same is true for the two PSI-D2 proteins. The amount of the *N. tabacum* PSI-D1 protein derived from *N. sylvestris* was too small to reproduce well in photographs (Fig. 2B).

In *N. sylvestris*, PSI-L has two electrophoretically distinct isoforms, which we designate PSI-L1 and PSI-L2 in the order of decreasing molecular mass (Figs. 2B and 4). Both *N. tomentosiformis* and *N. tabacum* have a third isoform, PSI-L3, the apparent molecular mass of which is somewhat smaller than that of PSI-L2 (Fig. 2B). *Nicotiana otophora*, which is a close relative of *N. tomentosiformis*, has PSI-L1 and L2 but not L3 (data not shown), similar to *N. sylvestris*, which originated from a different phylogenetic branch of the *Nicotiana* genus than did *N. tomentosiformis* and *N. otophora*. These branches diverged at the initial stages of *Nicotiana* evolution (Smith, 1968; Kung et al., 1982). Therefore, it is likely that PSI-L1 and PSI-L2 are present in most *Nicotiana* spp. and that PSI-L3 of *N. tabacum* was generated by a gene duplication that occurred in *N. tomentosiformis* or in a closely related progenitor.

N. tabacum has four types of PSI-H, two of which are derived from the genome of *N. sylvestris*. The other two are from *N. tomentosiformis* (Fig. 2), which indicates that PSI-H is present in two isoforms per diploid genome in *Nicotiana*. We designate the high and low molecular mass isomers as PSI-H1 and PSI-H2, respectively.

Analysis of the PSI subunit polymorphism, as described above, suggests that PSI-D, PSI-L, and PSI-H are more variable than the rest of the PSI proteins during the evolution of *Nicotiana* spp.

The Complete Set of PSI Proteins per Diploid Genome in *Nicotiana*

Comparison of the PSI proteins among alloploid and diploid species of *Nicotiana* (Fig. 2) permitted us to determine the complete set of PSI proteins encoded by each diploid genome in members of this genus (Table I). This set is composed of 20 proteins, with every one present in all of the *Nicotiana* spp. we have examined, including *N. tabacum* (Fig. 2), *N. sylvestris*, *N. tomentosiformis*, and *N. otophora* (data not shown). A notable feature of this classification scheme is that at least two or more isoforms of PSI-D, PSI-E, PSI-F, PSI-L, and PSI-H are present per genome. Densitometric analysis of Coomassie brilliant blue-stained gels revealed that PSI-C, PSI-D, PSI-E, PSI-H, and PSI-L are present in nearly equimolar ratios in the PSI complex in this study (data not shown), implying that one molecule of each subunit is present in each PSI complex. Therefore, it appears likely that the isoforms of these subunits are alternatively integrated into PSI complexes as they are formed to give rise to isocomplexes of PSI. PSI-F is present in less than one copy each in this study (data not shown), which may be ascribed to the loss of this protein during the purification with DEAE-cellulose chromatography. Iwasaki et al. (1991) reported that anion-exchange column chromatography of PSI complexes resulted in a considerable loss of PSI-F from the complex.

PSI-D Isomers

Two isoforms of PSI-D, namely, PSI-D1 and PSI-D2 (Table I), are homologous but have distinct N-terminal sequences (Fig. 3), indicating that they are encoded by different genes within a genome. PSI-D1 is much less abundant relative to PSI-D2, and these proteins were obtained in a ratio of about 1:10 in purified PSI.

PSI-E Isomers

PSI-E has four isoforms per genome as shown in Figure 2, and they are designated PSI-E1, PSI-E2, PSI-E3, and PSI-E4 in order of descending molecular mass. The N-terminal sequences of E1 and E2 are identical except for the N-terminal Ala, which is present in E1 but absent in E2 (Fig. 3). Similarly, E3 and E4 have the same N-terminal sequence except for an N-terminal Ala, which is present in E3 and absent in E4 (Fig. 3). The sequences of the E1 and E3 proteins are clearly different (Fig. 3), and these results indicate that at least two PSI-E genes are present per diploid genome in *Nicotiana* spp.

The amounts of PSI-E1 and -E2 are approximately equimolar, and likewise, the E3 and E4 proteins are present in nearly equal amounts. The relative amount of the E3- and E4-type proteins is larger than the amount of E1 and E2 proteins, but the ratio between them varied according to the chloroplast membrane preparations from which PSI was purified.

Table I. *PSI proteins per genome in Nicotiana spp.*

Subunit	Isoform	kD[a]	Polymorphism in *N. tabacum*	Gene (genome)	Gene Copies
PSI-A		68		*psaA* (c)	
PSI-B		68		*psaB* (c)	
PSI-D	PSI-D1	19.0	Polymorphic	*psaD* (n)	*psaDb*[c]
	PSI-D2	17.5	Polymorphic		*psaDa*[c]
PSI-F	PSI-F1	16.1		*psaF* (n)	?
	PSI-F2	15.9			
PSI-L	PSI-L1	15.4	Polymorphic	*psaL* (n)	?
	PSI-L2	15.2			
PSI-E	PSI-E1	14.4		*psaE* (n)	*psaEa*[d]
	PSI-E2	14.3			
	PSI-E3	14.1			*psaEb*[d]
	PSI-E4	14.0			
12.5-kD protein[b]		12.5		?	
PSI-H	PSI-H1	10.4	Polymorphic	*psaH* (n)	*psaHa*[e] *psaHb*[e] *psaHc*[e]
	PSI-H2	9.0	Polymorphic		
PSI-G		8.0		*psaG* (n)	
PSI-C		7.5		*psaC* (c)	
PSI-K		5.6		*psaK* (n)	
Blocked protein		4.5		? *psaJ* (c)[f]	
Blocked protein		3.0		? *psaI* (c)[f]	

[a] Apparent molecular masses of *N. sylvestris* proteins on the LDS-PAGE profile. [b] See text. [c] PSI-D1 and PSI-D2 are encoded by *psaDb* and *psaDa*, respectively (Yamamoto et al., 1991, 1993). [d] PSI-E1 and E2 are encoded by *psaEa*, whereas E3 and E4 are by *psaEb* (J. Obokata, K. Mikami, Y. Yamamoto, and N. Hayashida, unpublished results). [e] Correspondence between isoforms and gene copies is not known (Hayashida et al., 1992; M. Nakamura and J. Obokata, unpublished results). [f] Putative correspondence judging from the apparent molecular mass of the proteins.

Isomers of PSI-F, PSI-L, and PSI-H

Two isoforms of PSI-F, PSI-L, and PSI-H can be distinguished from each other by their electrophoretic mobility in polyacrylamide gels (Figs. 2 and 4). Among these subunits, high molecular mass isoforms are less abundant than low molecular mass forms, as in the cases of PSI-D and PSI-E. The relative abundance of PSI-H1 compared to PSI-H2 was so low that the electrophoretic image of PSI-H1 did not reproduce well in photographs (Fig. 2).

PSI-G and PSI-K Subunits

Although PSI-G and PSI-K are found in the PSI core fraction (numbers 19 and 22 in Fig. 2A, respectively), their sedimentation profiles in Suc gradients are different from those of the other core components and are rather similar to those of LHCI proteins (Fig. 1B). PSI-G and PSI-K are distributed in very similar ways in Suc gradients, being more abundant in fractions containing native PSI complexes (Fig. 1, lanes N) than in those containing the core complex (Fig. 1, lanes C).

The Polymorphic Nature of PSI Subunits Is Ubiquitous in Higher Plants

Detailed analysis of the subunit structure of PSI complexes in *Nicotiana* spp. raises the question of whether or not isoforms of PSI subunits are also observed in other higher plants. We conducted immunoblot analysis of PSI-D and PSI-E subunits in several diploid plant species. Figure 5 shows that PSI-D has two isoforms with distinct electrophoretic mobilities not only in *Nicotiana* spp. but also in tomato, *A. thaliana*, and maize. Two forms of PSI-E are also present in tomato, *A. thaliana*, and red bean (Fig. 5, bottom). Thus, two isoforms of both PSI-D and PSI-E subunits are present in a broad range of higher plant species.

DISCUSSION

We have demonstrated that polymorphism in the nuclear-encoded subunits of PSI complexes in *Nicotiana* spp. can be attributed to at least two sources. Alloploidy is one source of polymorphism, as in the case of the PSI-D, PSI-L, and PSI-H subunits of PSI in *N. tabacum* (Fig. 2B). The genes present in divergent *Nicotiana* spp. were themselves also divergent, and subsequent formation of the interspecific hybrids resulted in a multiplicity of divergent forms of a similar protein. A second source of polymorphism is duplication and divergence of genes within a genome. Evidence for this type of event is observed in the PSI-D and PSI-E subunits (Table I). Species that are diploid and show no evidence of having been polyploid in the past contain isoforms of both PSI-D and PSI-E with distinct N-terminal sequences (Fig. 3). In accordance with this observation, we recently isolated cDNA and genomic clones encoding PSI-D1, PSI-D2, PSI-E1, and PSI-E3 proteins from *N. sylvestris*, and we designated the

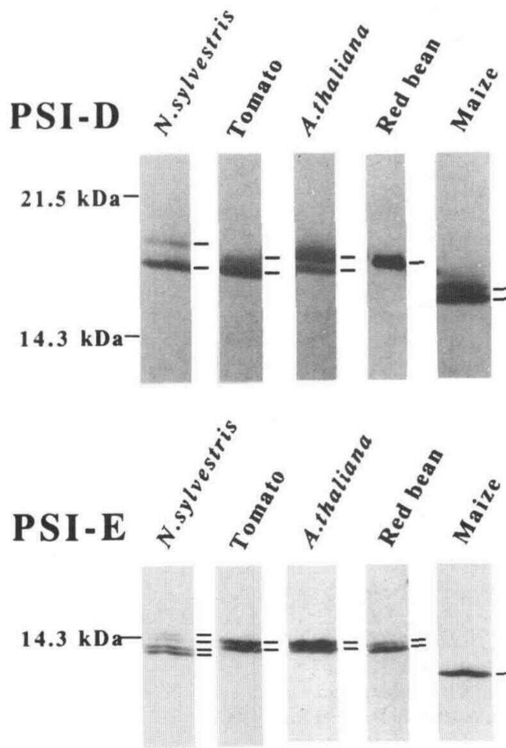

Figure 5. Immunoblot analysis of PSI-D and PSI-E subunits in isogenic plant lines. Chloroplast membranes were prepared from *N. sylvestris*, *L. esculentum*, *A. thaliana*, *V. angularis*, and *Z. mays*, and aliquots containing 25 μg of protein were subjected to high-resolution LDS-PAGE, followed by immunodetection with antibodies directed against spinach PSI-D (top) or PSI-E (bottom).

corresponding genes *psaDb* (Yamamoto et al., 1993), *psaDa* (Yamamoto et al., 1991), *psaEa*, and *psaEb* (J. Obokata, unpublished results), respectively. Along this same line, we found that three copies of *psaH* are present in the genome of *N. sylvestris* (M. Nakamura, unpublished results). All of these results suggest that molecular diversity in the nuclear genome encoding the photosynthetic apparatus has occurred by a number of mechanisms, including the formation of polyploids, interspecific hybrids, and multigene families and mutation during plant evolution. Thus, the potential of the nuclear genome to diverge is vast, which is in contrast with the highly conserved nature of the chloroplast genome.

The PSI-E subunits seem to exemplify formation of multigene families, which results in protein polymorphism. PSI-E occurs as four isoforms, PSI-E1, PSI-E2, PSI-E3, and PSI-E4, as summarized in Table I. The N-terminal sequences of these isoforms suggest that PSI-E2 and PSI-E4 are derived from PSI-E1 and PSI-E3, respectively, by the loss of N-terminal Ala (Fig. 3). PSI-E2 and PSI-E4 are unlikely to be proteolytic artifacts because (a) the presence or absence of protease inhibitors (1 mM each of PMSF, TPCK, and TLCK) during PSI preparation had no effect on the electrophoretic mobilities of these proteins and (b) both E1 and E2, and E3 and E4, were always present in 1:1 ratios in many batches of PSI preparations. The processing of the PSI-E precursors may be somewhat nonspecific, making the two kinds of products nearly but not always precisely equimolar. A detailed examination of the structures of both the PSI-E subunits and the *psaE* cDNAs is described elsewhere.

Detailed examination of protein polymorphism revealed a set of PSI proteins commonly shared by *Nicotiana* spp. It includes at least 20 proteins as summarized in Table I, 7 of which are isoforms of one of the other 13 protein species. The full set of proteins appears to consist of 13 distinct subunits, including PSI-A, PSI-B, PSI-C, PSI-D, PSI-E, PSI-F, PSI-G, PSI-H, PSI-K, PSI-L, two N-terminally blocked proteins of 4.5 and 3.0 kD, and an unidentified protein of 12.5 kD. The proteins of 4.5 and 3 kD were not analyzed in this study, but their electrophoretic mobilities suggest that they correspond to the products of chloroplast genes *psaJ* and *psaI* (Iwasaki et al., 1991; Bryant, 1992; Ikeuchi, 1992), respectively (Table I). This list of PSI subunits is in agreement with those reported for barley (Scheller and Møller, 1990), spinach (Ikeuchi and Inoue, 1991), and cucumber (Iwasaki et al., 1991), except for the presence of the 12.5-kD protein. The abundance of this 12.5-kD protein is very low relative to the other subunits, and we were unable to obtain clear results regarding its N-terminal sequence. At present, it is not clear whether the difficulty in determining its N-terminal sequence is due to a blocked N terminus or to low levels of this protein. Further analysis is necessary to ascertain the identity of the 12.5-kD protein.

The most notable finding of this study is that all of the nuclear-encoded subunits except for PSI-G and PSI-K are present in two isoforms in *Nicotiana* spp. (Fig. 2, Table I). Results of immunoblot analyses using anti-PSI-D and anti-PSI-E antibodies (Fig. 5) suggest that this polymorphism is not peculiar to *Nicotiana* spp. but is also present in many higher plants. We are also interested in ferns, mosses, and algae in this respect, but isogenic lines of these organisms were not available for this study. The immunoblot profiles did not reveal isoforms of PSI-D in red bean or PSI-E in maize (Fig. 5); however, other methods such as HPLC or two-dimensional PAGE may resolve them if they are present. We also carried out immunoblot analysis with an anti-PSI-L antibody, but this antibody produced dark smears when reacted with total thylakoid membrane proteins.

From the results described here, it appears likely that the nuclear genomes of higher plants generally have at least two genes encoding each subunit of the PSI core complex. This assumption is in agreement with the results of recent studies of the nuclear genes in *N. sylvestris* (Yamamoto et al., 1991; Obokata et al., 1992; Yamamoto et al., 1993) and spinach (Herrmann et al., 1991). In contrast to this, *psaD* (Reilly et al., 1988), *psaE* (Chitnis et al., 1989), and *psaF* (Chitnis et al., 1991) were reported to be single-copy genes in the cyanobacterium *Synecchocystis* spp. PCC 6803. Why are the PSI genes of higher plants diversified within the nuclear genome? One possible advantage to gene families is that different genes encoding the same subunit might respond to different signal transduction pathways. This seems to be the case for *psaD* gene family, because the genes for PSI-D1 and PSI-D2 are differentially expressed during leaf development (Yamamoto et al., 1993).

Relationships between the organization of the genes and the regulatory mechanisms of PSI biogenesis is an interesting

subject that deserves further attention. The second possibility is that isoforms of the PSI subunits may be adapted to slightly different functions in plant cells. Recently, Hase et al. (1991b) reported that, in maize, isoproteins of Fd, FdI and FdIII, have different electron transport activities. It is interesting to speculate, for example, that the granal and stromal PSI complexes may be composed of different isoproteins to differentiate their functions. Hypotheses relating to possible functional differentiation between isoforms of the PSI complex subunits are also rich subjects for future investigation. At present, we do not know how many different isocomplexes are present for PSI or how many combinations of the PSI isoproteins are possible in vivo.

In this study, we have provided a glimpse into the molecular heterogeneity of the PSI complex in plant cells. This heterogeneity may hinder us in resolving PSI isocomplexes or in obtaining high-grade crystals of this complex from higher plants. Recent advances in transgenic plant technology will afford us many opportunities to investigate the biological significance of this phenomenon in the near future.

ACKNOWLEDGMENTS

The authors express their thanks to Mr. Y. Yamamoto, Dr. K. Inoue, Dr. S. Tanifuji, and Dr. M. Sugita for useful discussions and encouragement and to Dr. K. Hong, Dr. A. Kimura, and Mr. M. Takata for their help in amino acid sequence determination. Antiserum raised against spinach PSI proteins was kindly supplied by Dr. H. Oh-oka, and antibody against PSI-L was the generous gift of Dr. H.V. Scheller. The seeds of *N. sylvestris*, *N. tomentosiformis*, and *N. otophora* were kindly provided by Japan Tobacco, Inc., the seeds of red bean were from Dr. T. Mikami, the seeds of *Z. mays* were from the Hokkaido National Agricultural Experimental Station, and the seeds of *Arabidopsis* were from Dr. S. Toki. This study was carried out at the Research Center for Molecular Genetics, Hokkaido University, and a part of the amino acid sequence analysis was performed at the Faculty of Agriculture, Hokkaido University, and at Ohtsuka Pharmaceutical Company.

Received March 5, 1993; accepted May 10, 1993.
Copyright Clearance Center: 0032-0889/93/102/1259/09.

LITERATURE CITED

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

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

Chitnis PR, Purvis D, Nelson N (1991) Molecular cloning and targeted mutagenesis of the gene *psaF* encoding subunit III of photosystem I from the cyanobacterium *Synechocystis* spp. PCC 6803. J Biol Chem **30:** 20146–20151

Chitnis PR, Reilly PA, Miedel MC, Nelson N (1989) Structure and targeted mutagenesis of the gene encoding 8-kDa subunit of photosystem I from the cyanobacterium *Synechocystis* spp. PCC 6803. J Biol Chem **264:** 18374–18380

Fling SP, Gregerson DS (1986) Peptide and protein molecular weight determination by electrophoresis using a high-molarity Tris buffer system without urea. Anal Biochem **155:** 83–88

Golbeck JH (1987) Structure, function and organization of the photosystem I reaction center complex. Biochim Biophys Acta **895:** 167–204

Golbeck JH (1992) Structure and function of photosystem I. Annu Rev Plant Physiol Plant Mol Biol **43:** 293–324

Gray JC, Kung SD, Wildman SG, Sheen SJ (1974) Origin of *Nicotiana tabacum* L. detected by polypeptide composition of fraction I protein. Nature **252:** 226–227

Hallick RB (1989) Proposals for the naming of chloroplast genes. II. Update to the nomenclature of genes for thylakoid membrane polypeptides. Plant Mol Biol Rep **4:** 266–275

Hase T, Kimata Y, Yonekura K, Matsumura T, Sakakibara H (1991a) Molecular cloning and differential expression of the maize ferredoxin gene family. Plant Physiol **96:** 77–83

Hase T, Mizutani S, Mukohata Y (1991b) Expression of maize ferredoxin cDNA in *Escherichia coli*. Plant Physiol **97:** 1395–1401

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

Herrmann RG, Oelmuller R, Bichler J, Schneiderbauer A, Steppuhun J, Wedel N, Tyagi AK, Westhoff P (1991) The thylakoid membranes of higher plants: genes, their expression and interaction. *In* RG Herrmann, B Larkins, eds, Plant Molecular Biology 2. Plenum Press, New York, pp 411–427

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

Ikeuchi M, Inoue Y (1991) Two new components of 9 and 14 kDa from spinach photosystem I complex. FEBS Lett **280:** 332–334

Iwasaki Y, Ishikawa H, Hibino T, Takabe T (1991) Characterization of genes that encode subunits of cucumber PSI complex by N-terminal sequencing. Biochim Biophys Acta **1059:** 141–148

Kelly JL, Greenleaf AL, Lehman LR (1986) Isolation of the nuclear gene encoding a subunit of the yeast mitochondrial RNA polymerase. J Biol Chem **261:** 10348–10351

Kung SD, Zhu YS, Shen GF (1982) *Nicotiana* chloroplast genome III, chloroplast DNA evolution. Theor Appl Genet **61:** 73–79

Obokata J (1986) Synthesis of a photosystem I polypeptide of 15 kilodaltons in isolated etiochloroplasts of wheat. Plant Physiol **81:** 705–707

Obokata J (1987) Synthesis and assembly of the polypeptides of photosystem I and II in isolated etiochloroplasts of wheat. Plant Physiol **84:** 535–540

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

Obokata J, Yamamoto Y, Kubota T, Nakamura M (1992) Structure of the nuclear genes coding for photosystem I subunits in *Nicotiana sylvestris*. *In* N Murata, ed, Research in Photosynthesis, Vol III. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 367–370

Oh-oka H, Takahashi Y, Matsubara H (1989) Topological consideration of the 9-kDa polypeptide which contains center A and B, associated with the 14- and 19-kDa polypeptides in the photosystem I complex of spinach. Plant Cell Physiol **30:** 869–875

Okkels JS, Scheller HV, Svendsen I, Møller BL (1991) Isolation and characterization of a cDNA clone encoding an 18-kDa hydrophobic photosystem I subunit (PSI-L) from barley (*Hordeum vulgare* L.). J Biol Chem **266:** 6767–6773

Pichersky E, Hoffman NE, Bernnatzky R, Pieculla B, Tanksley SD, Cashmore AR (1987) Molecular characterization and genetic mapping of DNA sequences encoding the type I chlorophyll *a/b* binding polypeptide of photosystem I in *Lycopersicon esculentum* (tomato). Plant Mol Biol **9:** 205–216

Reilly P, Hulmes JD, Pan YH, Nelson N (1988) Molecular cloning and sequencing of the *psaD* gene encoding subunit II of photosystem I from cyanobacterium, *Synechocystis* spp. PCC 6803. J Biol Chem **263:** 17658–17662

Scheller HB, Møller BL (1990) Photosystem I polypeptides. Physiol Plant **78:** 484–494

Shimada H, Sugiura M (1991) Fine structural features of the chlo-

roplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Res **19:** 983–995

Singer M, Berg P (1991) Genes and Genomes: A Changing Perspective. Blackwell Scientific Publications, Oxford, UK

Smith HH (1968) Recent cytogenetic studies in the genus *Nicotiana.* Adv Genet **14:** 1–54

Wynn RM, 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, Hayashida N, Inoue K, Obokata J (1991) Nucleotide sequence of cDNA clones encoding PSI-D2 protein of photosystem I in *Nicotiana sylvestris.* Plant Mol Biol **17:** 1251–1254

Yamamoto Y, Tsuji H, Obokata J (1993) Structure and expression of a nuclear gene for PSI-D1 subunit of photosystem I in *Nicotiana sylvestris.* Plant Mol Biol (in press)

Zanetti G, 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 AL, Malkin R (1988) Ferredoxin cross-links to a 22 kD subunit of photosystem I. Plant Physiol **88:** 810–814