Studies on the gene structure of photosystem I and the
reconstitution of the photosystem I function *in vitro*
with thermophilic cyanobacterium *Synechococcus vulcanus*

光化学系1遺伝子の機能および光化学系1
機能の試験管内再構成に関する分子遺伝学的研究

1992年3月

埼玉大学大学院理工学研究科（博士後期課程）
生産情報科学専攻（主指導教官　檜山哲夫教授）
清 水 徳 朗

CONTENTS

-Immunological identification of proteins
-Flash spectrophotometry of P_{700}
-Spinach photosystem I and CPI preparation
-*psa*C gene expression and holo PSI-C fraction
 isolation
-Reconstitution
-Electron microscopy
-Low temperature EPR analysis
-Abbreviations

CHAPTER 3
_____

Isolation and characterization of the genes
for *mtn*A–*psa*C of *Synechococcus vulcanus*
Introduction
Results and discussions
  – *psa*C gene
  – *mtn*A gene

CHAPTER 4
_____

Nucleotide sequences of the *psa*A and *psa*B
genes encoding the photosystem I core
proteins from thermophilic cyanobacterium
Synechococcus vulcanus
Abstract
Results and discussion
  – Structure of the *psa*A–*psa*B gene
  – Promoters
  – Terminators
  – Interrupted ORF
  – Homologies of the deduced proteins
  – Homologies of the *psa*A–*psa*B genes

CHAPTER ONE

GENERAL INTRODUCTION

Light has various effects on life. Light is utilized by living organisms for both an energy source and signals. All living organisms can be divided into two categories, photoautotrophs and chemoheterotrophs. Most of the radiation energy from the sun is converted to chemical energy *via* photosynthetic process by photoautotrophs, and the products are also utilized by chemoheterotrophs. Since we, human beings, as a chemoheterotroph owe all energy to photoautotrophic plants, studying the photosynthetic light energy conversion is one of the most important theme of biology.

Photosynthetic organisms have developed specialized apparata to capture and convert the light energy efficient-

1-1

ly. The conversion of light energy in plants are carried out in the process of 'photosynthetic electron transport'. In 1960, Hill and Bendall first postulated a 'zig-zag scheme' of photosynthetic electron transport which formulated the cooperative work of two photosystems. The scheme could satisfactorily explain the 'red-drop' of quantum yield of photosynthesis and the 'enhancement' of the quantum yield of photosynthesis by a weak blue light illumination which had been observed by Emerson.

Oxygen-evolving photosynthesis in higher plants, algae and cyanobacteria involves two primary photochemical reaction sites, namely, photosystem I and II. Photosystem II (PSII), an oxidative side of photosynthetic electron transport chain, catalyzes the light-dependent oxidation of water and reduction of plastoquinone (PQ), yielding an oxygen as a by-product. Photosystem I (PSI), a reductive side of photosynthetic electron transport chain, transfers electrons from plastocyanin to $NADP^+$ through ferredoxin and ferredoxin-NADP reductase(FNR). Both PSI and PSII are located on the thylakoid membranes in chloroplasts and cyanobacteria. During transfer of electrons from PSII to PSI through quinone-pool and cytochrome-b_6/f complex, protons were translocated across the thylakoid membrane,

which enables plants to accumulate and convert the photo-synthetic energy in the form of ATP.

Photosystem I as a supramolecular complex consists of the primary electron donor P_{700} (Kok, 1963), primary electron acceptor A_0, intermediate electron carriers A_1, A_2 (also called center F_X or 'P430') and terminal electron acceptors center A and center B (F_A, F_B), and functional pigment such as chlorophyll at 100 molecules per P_{700} (Golbeck, 1987). Four out of these six components, P_{700}, A_0, A_1 and A_2 are located on the two large subunits of PSI (PSI-A and PSI-B) which are encoded by psaA and psaB genes of chloroplast DNA in higher plants, respectively (Golbeck, 1987). Two other terminal electron acceptors center A and center B are both retained in a 9kDa subunit, namely PSI-C, which is also encoded by the psaC gene of chloroplast DNA in higher plants (Hayashida et al.1987).

PSI complex also contains several other lower molecular mass subunits associated with the complex (Bengis and Nelson, 1977) whose functions are not well characterized. The subunit II protein (PSI-D), one of these subunits encoded by psaD gene, is thought to be required for the binding of the ferredoxin (Zanetti and Merati, 1987; Zilber and Malkin, 1988) and essential for the efficient function of

PSI (Chitnis *et al.* 1989b). Although another subunit which is encoded by *psa*E (PSI-E) gene had been demonstrated to be crosslinked to the 9kDa subunit (PSI-C) with PSI-D subunit as well (Oh-oka *et al.*1989), destruction of the *psa*E gene in cyanobacteria scarcely affect the growth of the mutant (Chitnis *et al.* 1989a). On the other hands, a subunit encoded by a nuclear gene *psa*F was identified as the binding site of plastocyanin by a crosslink experiment (Hippler *et al.* 1989). The functions of these subunits and other subunits encoded by *psa*I,*psa*J,*psa*K (Koike *et al.* 1989; Ikeuchi *et al.* 1990) and 9- and 14kDa subunits (Ikeuchi and Inoue, 1991), are still not well understood.

Many kinds of proteins in chloroplast were synthesized by cytosolic ribosomes as larger precursors with a transit peptide and posttranslationally transported into chloroplast (Keegstra and Olsen, 1989). That was same for the several of the subunits of photosystem I in higher plants. While it has been easy to introduce a gene to a nucleus with different ways (Draper *et al.* 1988; de la Peña *et al.* 1987), gene specific destruction, substitution and target mutagenesis for the already existed nuclear-encoded gene is not possible except for a special case (Lee *et al.* 1990). As demonstrated for the *rbc*S and *cab* genes (Dean *et al.* 1989;

1-4

Keegstra and Olsen, 1989), the nuclear-encoded chloroplast genes exist in multiple copies and this makes it difficult to destruct or substitute them completely and study the function of a gene precisely. Furthermore, expression of these genes are also regulated by several factors such as light (Gilmartin *et al.* 1990), localization of organ (Fluhr *et al.* 1986) and developmental stage of the chloroplast (Mullet, 1988). The inhibition of the gene expression by an anti-sense RNA technique (Stockhaus *et al.* 1990) or ribozyme may improve the problems.

Chloroplasts (Boynton *et al.*1988) and mitochondria (Johnston *et al.* 1988) could be transformed by a micropro-jectiles cell bombardment technique, also called 'DNA-gun'. However, there are numbers of difficulties in studying the chloroplast encoded genes in higher plants for the follow-ing reasons: (1) a significant number of chloroplasts are usually found in a plant cell (2) each chloroplast contains multiple copies of a plastid genome (3) the numbers of chloroplasts will increase during the plant cell development (Mullet, 1988; Taylor, 1989) (4) expression of some chloro-plast-encoded genes are light-regulated (Klein and Mullet, 1987), and (5) chlorophyll is also required for the accumu-lation of the large-subunits of photosystem I (Klein *et al.*

1988). These many different factors should be taken into account when studying the expression and function of the chloroplast-encoded genes.

As a matter of fact, most recent investigations of transforming the chloroplast have been limited to an unicellular algae *Chlamydomonas*, which contains a single chloroplast (Blowers *et al*.1989; Przibilla *et al*. 1991; Takahashi *et al*. 1991).

Cyanobacteria, on the other hand, provide several advantages over plants for studying the photosystems. First, cyanobacteria are prokaryote and as easy to handle as bacteria. Second, since cyanobacterial photosystems resemble to those of higher plants (Nechushtai *et al*.1983; Lundell *et al*. 1985), many investigations on cyanobacteria are applicable to those of the higher plants. Third, several cyanobacteria have naturally occurring transformation systems for the uptake of exogenous DNA and introducing it into the genomic DNA by a homologous recombination manner. Fourth, some peculiar cyanobacteria like *Synechocystis* sp. PCC6803 (Williams, 1988; Chitnis *et al*. 1989a and 1989b) and *Anabaena variabilis* ATCC 29413 (Mannan *et al*. in press; Toelge *et al*. 1991) are capable to grow on hexoses like glucose or fructose, allowing the isolation and

growth of the mutants defective in photosynthesis . Fifth, the composition of membrane proteins could change in response to the quality of light (Aizawa *et al.* in press), implicating the existence of photoregulation.

However, recent works on genetical modification of cyanobacterial photosystem II indicated that any kinds of deletion or interruption of only one of those genes could cause a disappearance of entire photosynthetic apparatus (Yu and Vermaas, 1990), probably because the deficient of key-subunit might destabilize the whole assembly of the photosystem II.

The organization and the structure of photosystem I has currently been the focus of intense studies, but the diffi-culties which mentioned above prevent the further investi-gations. Since we do not have an effective way of subunit specific exclusion, the exact function of each subunits should be examined by an *in vitro* reconstitution analysis. Furthermore, isolation and production of these genes will make it possible to study the function by a reverse-genet-ics method using site specific mutagenesis technique. From these points of view, I have intended to isolate the genes of photosystem I and establish the *in vitro* reconstitution system with *in vivo* produced photosystem I subunits in

E.coli. For the experiments, I have used the thermophilic cyanobacterium *Synechococcus vulcanus* for the following reasons: (1) the components of photosystem I are similar to those of higher plants and (2) their amino acid sequences at N-terminal are already known (Koike *et al.*1989) (3) the photosystem I genes have not been isolated from the cyanobacterium (4) because *Synechococcus vulcanus* is a thermophilic cyanobacterium which grows at $50^{O}C$, produced proteins by the isolated genes are expected to be thermostable.

REFERENCES

Aizawa, K., Shimizu, T., Hiyama, T., Satoh, K., Nakamura, Y. and Fujita, Y. (1991) Changes in composition of membrane peptides in *Synechocystis* PCC6803 in response to light quality for photosynthetic growth. *Photosynthesis Res. in press.*

Bengis, C. and Nelson, N. (1977) Subunit structure of chloroplast photosystem I reaction center. *J.Biol.Chem.* 252, 4564-4569.

Blowers, A.D., Bogorad, L., Shark, K.B. and Sanford, J.C. (1989) Studies on *Chlamydomonas* chloroplast transformation: foreign DNA can be stably maintained in the chromosome. *Plant Cell*, 1, 123-132.

Boynton, J.E., Gillham, N.W., Harris, E.H., Hosler, J.P., Johnson, A.M., Jones, A.R., Randolph-Anderson, B.L., Robertson, D., Klein, T.M., Shark, K.B. and Sanford, J.C. (1988) Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science*, 240, 1534-1538.

Chitnis, P.R., Reilly, P.A., Miedel, M.C. and Nelson, N. (1989a) Structure and targeted mutagenesis of the gene encoding 8-kDa subunit of photosystem I from the cyano-bacterium *Synechocystis* sp.PCC6803. *J.Biol.Chem.* **264,** 18374–18380.

Chitnis, P.R., Reilly, P.A. and Nelson, N. (1989b) Insertional inactivation of the gene encoding subunit II of photosystem I from the cyanobacterium *Synechocystis* sp. PCC6803. *J.Biol.Chem.* **264,** 18381–18385.

Dean, C., Pichersky, E. and Dunsmuir, P. (1989) Structure, evolution, and regulation of *RbcS* genes in higher plants. *Ann.Rev.Plant Physiol. Plant Mol.Biol.* **40,** 415–439.

de la Peña, A. Lörz, H. and Schell. J. (1987) Transgenic rye plants obtained by injecting DNA into young floral tillers. *Nature,* **325,** 274–276 .

Draper, J., Scott, R., Armitage, P. and Walden, R. (edt.) Plant genetic transformation and gene expression (a labo-ratory manual), Blackwell, London, 1988.

Fluhr, R., Huhlemeier, C., Nagy, F. and Chua, N-H. (1986) Organ-specific and light-induced expression of plant genes. *Science* **232**, 1106-1112.

Gilmartin, P.M., Sarokin, L., Memelink, J. and Chua, N-H. (1990) Molecular light switches for plant genes. *Plant Cell* **2**, 369-378.

Golbeck, J.H. (1987) Structure, function and organization of the Photosystem I reaction center complex. *Biochim. Biophys.Acta*, **895**, 167-204 (review).

Hayashida, N., Matsubayashi, T., Shinozaki, K., Sugiura, M., Inoue, K. and Hiyama, T. (1987) The gene for the 9kd polypeptide, a possible apoprotein for the iron-sulfur centers A and B of the photosystem I complex, in tobacco chloroplast DNA. *Curr.Genet.*, **12**, 247-250.

Hippler, M., Ratajczak, R. and Haehnel, W. (1989) Identification of the plastocyanin binding subunit of photosystem I. *FEBS Lett.* **250**, 280-284.

Ikeuchi, M., Hirano, A., Hiyama, T. and Inoue, Y. (1990) Polypeptide composition of higher plant photosystem I complex: Identification of *psaI*, *psaJ* and *psaK* gene products. *FEBS Lett.* 1990, 274-278.

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

Johnston, S.A., Anziano, P.Q., Shark, K., Sanford, J.C. and Butow, R.A. (1988) Mitochondrial transformation in yeast by bombardment with microprojectiles. *Science*, 240, 1538-1541.

Keegstra, K. and Olsen, L.J. (1989) Chloroplastic precursors and their transport across the envelope membranes. *Annu.Rev. Plant Physiol. Plant Mol. Biol.*, 40, 471-501.

Klein, R.R. and Mullet, J.E. (1987) Control of gene expression during higher plant chloroplast biogenesis: protein synthesis and transcript levels of psbA, psaA-psaB, and rbcL in dark-grown and illuminated barley seedlings. *J. Biol.Chem.* 256, 4341-4348.

Klein, R.R., Gamble, P.E. and Mullet, J.E. (1988) Light-dependent accumulation of radiolabeled plastid-encoded chlorophyll *a*-apoproteins requires chlorophyll *a*: I. analysis of chlorophyll-deficient mutants and phytochrome involvement. *Plant Physiol.* 88, 1246-1256.

Koike, H., Ikeuchi, M., Hiyama, T. and Inoue, Y. (1989) Identification of photosystem I components from the cyanobacterium, *Synechococcus vulcanus* by N-terminal sequencing. *FEBS Lett.* 253, 257-263.

Kok, B. (1963) Significance of P_{700} as an intermediate in photosynthesis. *Proc.Int.Cong. Biochem.*, 6, 73-81.

Lee, K.Y., Lund, P., Lowe, K. and Dunsmuir, P. (1990) Homologous recombination in plant cells after *Agrobacterium*-mediated transformation. *Plant Cell*, 2, 415-425.

Lundell, D.J., Glazer, A.N., Melis, A. and Malkin, R. (1985) Characterization of a cyanobacterial photosystem I complex. *J.Biol.Chem.* 260, 646-654.

Mannan, R.M., Whitmash, J., Nyman, P. and Pakrasi, H.B.

Directed mutagenesis of an iron sulfur protein of the Photosystem I complex in the filamentous cyanobacterium *Anabaena variabilis* ATCC 29413. *Proc.Natl.Acad.Sci.USA in press.*

Mullet, J.E. (1988) Chloroplast development and gene expression. in *Annu. Rev. Plant Physiol. Plant Mol.Biol.*, 39, 475–502.

Nechushtai, R., Muster, P., Binder, A., Liveanu, V. and Nelson, N. (1983) Photosystem I reaction center from the thermophilic cyanobacterium *Mastigocladus laminosus. Proc. Natl.Acad.Sci.USA* 80, 1179–1183.

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

Przibilla, E. Heiss, S., Johanningmeier, U. and Trebst, A. (1991) Site-specific mutagenesis of the D1 subunit of photosystem II in wild-type *Chlamydomonas. Plant Cell*, 3,

169-174.

Smart, L.B., Anderson, S.L. and McIntosh, L. (1991) Targeted genetic inactivation of the photosystem I reaction center in the cyanobacterium *Synechocystis* sp. PCC6803. **10**, 3289-3296.

Stockhaus, J., Hoefer, M., Renger, G., Westhoff, P., Wydrzynski, T. and Willmitzer, L. (1990) Anti-sense RNA efficiently inhibits formation of the 10kd polypeptide of photosystem II in transgenic potato plants: analysis of the role of the 10kd protein. *EMBO J.*, **9**, 3013-3021.

Taylor, W.C. (1989) Regulatory interactions between nuclear and plastid genomes. *Ann.Rev.Plant Physiol. Plant Mol.Biol.* **40**, 211-233.

Takahashi, Y., Goldschmidt-Clermont, M., Soen, S.-Y., Franzen and Rochaix, J.-D. (1991) Directed chloroplast transformation in *Chlamydomonas reinhardtii*: insertional inactivation of the *psaC* gene encoding the iron sulfur protein destabilizes photosystem I. *EMBO J.*, **10**, 2033-2040.

Toelge, M., Ziegler, K., Maldener, I. and Lockau, W. (1991) Directed mutagenesis of the gene *psaB* of Photosystem I of the cyanobacterium *Anabaena variabilis* ATCC 29413. *Biochim.Biophys.Acta* **1060**, 233-236.

Williams, J.G.K. (1988) Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods in Enzymology* **167**[85], 766-778.

Yu, J. and Vermaas, W.F.J. Transcript levels and synthesis of photosystem II components in cyanobacterial mutants with inactivated photosystem II genes. (1990) *Plant Cell*, **2**, 315-322.

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.L. and Malkin, R. (1988) Ferredoxin cross-links to a 22kD subunit of Photosystem I. *Plant Physiol.* **88**, 810-814.

CHAPTER TWO

MATERIALS AND METHODS

Materials

Restriction endonucleases and DNA modification enzymes were obtained from TAKARA-SHUZO, TOYOBO, Nippon Gene and New England Biolabs, and used according to the instructions of suppliers. A strain JM109(DE3), and pGEMEX1 cloning vector were purchased from Promega(U.S.A). Unless otherwise stated, all reagents employed in this study were of the highest quality available.

Reagents

Reagents which were used for molecular biological experiments were prepared and stored according to Sambrook *et al.*(1989).

Bacterial strains

A strain *E.coli* JM109 (*recA*1 *endA*1 *gyrA*96 *thi hsdR*17 *supE*44 *relA*1 λ⁻ *delta*(*lac-proAB*)/F' *proAB lacI*^{q}*Z* ΔM15 *traD*36) (Yanisch-Perron *et al.* 1985) was usually used as a host to maintain plasmid vector and M13 phage vector. Other bacterial strains which were used in this experiments were listed below:

JM109(DE3)/F'(DE3 lysogen of JM109/F');

LE392 (F⁻ *hsdR*514(rk⁻,mk⁺) *supE*44 *supF*58 *lacY*1 or Δ(*lacIZY*)6 *galK2*);

HB101 (F⁻ *hsdS*20(r$_b$⁻,m$_b$⁻) *recA*1 *ara-14 proA2 lacY*1 *galK2* rpsL*20(Sm^r)* xyl-5 *mt1*-1 *supE*44 λ⁻ (*mcrA*⁺, *mcrB*⁻).

P2392 (P2 lysogen of LE392).

Maintenance of strains

Long-term storage of desired strain was accomplished by mixing 1ml of a stationary-phase culture with 1ml of stock medium (0.8% Bacto nutrient broth, 20%(w/v) glycerol) and freezing

at -80°C. Bacteria were revived by inoculating
aliquots on appropriate selective media and incu-
bating at 37°C(Miller, 1972).

Bacterial strain which exhibiting the resistan-
cy of antibiotics was cultured in the media
containing appropriate antibiotics.
Concentration of the each antibiotics were listed
below:

Ampicillin	100μg/ml
Streptomycin	25μg/ml
Tetracycline	12.5μg/ml
Kanamycin	50μg/ml
Chloramphenicol	25μg/ml

Cyanobacterium

Thermophilic cyanobacterium *Synechococcus vulcanus* was kindly provided from the solar energy research group of RIKEN. The cyanobacterium was cultured at 50°C as described by Shin et al. (1984) in a medium described by Katoh (1988) for *Synechococcus* sp.. For long-term storage, thick culture (100ml) was centrifuged and the precipitated cells were resuspended with few ml of resultant supernatant. One ml of the suspension was mixed with 0.24ml of 80%(w/v) glycerol in a sterilized vial and quickly frozen at -80°C. The frozen culture had never thawed. The cyanobacterium was revived by inoculating scraped frozen stock into prewarmed 10ml medium then cultivated immediately under weak illumination until early-log phase.

Cloning vectors

A plasmid vector pTZ19R(TOYOBO) was usually used as a vector to maintain DNA fragment. Other vectors pGEMEX1(Promega), pACYC177(Chang *et al.* 1978), pBR322 (laboratory stock) and phage vector λ-Dash (Stratagene) were also used.

The bacterial strains HB101 which harboring each plasmid of tobacco chloroplast DNA clone pTBa2, pTB20, pTB30 and pTX9 (Sugiura *et al.* 1986) were kindly provided from Prof. M.Sugiura at Nagoya university.

A plasmid pLysS (Studier *et al.* 1990) which containing T7 lysozyme gene in pACYC184 was obtained by courtesy of Mr.T.Konishi at Nagoya university.

The plasmid vectors and the helper phage M13KO7(Vieira and Messing, 1987) were maintained with *E.coli* strain JM109, HB101 and JM109(DE3). λ phage was maintained with *E.coli* strain LE392 and P2392.

DNA manipulation

Manipulation of DNA was achieved with the methods of Sambrook, Fritsch and Maniatis (1989), and Berger and Kimmel(1987). Concentration of DNA was estimated by UV-absorbance or ethidium bromide fluorescence spot assay(Sambrook *et al.* 1989) with already-known concentration λ-phage DNA as a standard. Plasmid DNA was prepared according to the alkaline-SDS method(Sambrook *et al.* 1989). Prepared DNA was dissolved in auto-claved Milli-Q water and stored at $4^{O}C$ (for short term storage) or $-20^{O}C$ (for long-term storage). Dissolving a DNA with 1xTE(pH8.0) occasionally prevent a digestion of the DNA with a restriction endonuclease.

Precipitation of DNA with ethanol, terminal repair of DNA with T4 DNA polymerase and ligation of DNA with T4 DNA ligase were accomplished with individual method as below:

2-6

Ethanol precipitation of DNA

DNA solution was mixed with 0.1 volume of 3M sodium acetate (pH8.0) and 2.5 volumes of ethanol, then standed at -80°C for 10minutes. The precipitated DNA was recovered by centrifugation at 16,000rpm for 10minutes at 4°C in a microfuge. The precipitated DNA was washed once with adding 1ml of 70%(v/v) ethanol and centrifugation at 16,000rpm for 10minutes at 4°C. The resultant DNA was dried *in vacuo* and dissolved in autoclaved Milli-Q water.

Terminal repair by T4 DNA polymerase

A solution of DNA was mixed with 10xT4 buffer (0.33M Tris-acetate, pH7.9, 0.66M sodium acetate, 0.1M magnesium acetate and 5mM DTT), 10xBSA and 10xdNTP (a mixture of dATP,dCTP, dGTP and dTTP for 10mM) and the volume was made to be 50 to 100ul. T4 DNA polymerase(TAKARA) was mixed (amount of the polymerase required for the reac-

tion was estimated by the concentration to the DNA would be below $1U/\mu g$ DNA) and incubated at $30^{O}C$ for 10minutes(prolonged reaction time occasionally caused a unexpected disappearing of the DNA). The reaction was terminated with extraction once with phenol:chloroform and once with chloroform:isoamylalcohol. The repaired DNA was recovered by ethanol precipitation.

Ligation of DNA by T4 DNA ligase

Reaction was performed in a mixture of 66mM Tris-HCl, pH7.6, 6.6mM $MgCl_2$, 10mM DTT, 0.15M NaCl, 10%(w/v)PEG-6000 and 0.1mM ATP with 100-200U of T4 DNA ligase(TAKARA), and incubated at $16^{O}C$ for few hours (usually overnight).

Preparation of λ phage and DNA

λ phage were amplified as described below:

Preparation of host cell

1. A host strain (LE392 or P2392) was inoculat-
ed in NZCYM or LB medium supplemented with
0.2%(w/v) maltose and shaked overnight(<12hours)
at $37^{O}C$.

2. 5ml of the overnight culture was taken into
a Falcon 2059 tube and centrifuged at 4,000xg for
10minutes at room temperature, and supernatant
was removed.

3. The cell pellet was suspended in ice-cold
sterilized 10mM $MgSO_4$ solution and standed in ice
until use.

4. A_{600} were adjusted to 2.0 with 10mM $MgSO_4$
and used within a day.

Liquid amplification of λ *phage*

1. A lysate which containing 0.5×10^6 λ phage particles was take into a Falcon 2063 tube and standed at a clean bench for 15minutes to evaporate the residual chloroform in the lysate.

2. 0.05-0.1ml of the host cell suspension was added and incubated at 37^OC or 15minutes to adsorb the phage. The mixture was transferred into 4-5ml of NZCYM medium and shaked at 37^OC for 6 to 12hours until the cell almost lysed.

3. Few drops of chloroform was added to the lysed culture and shaked further 15minutes.

4. The culture was transferred to a Falcon 2059 tube and centrifuged at 4,000xg for 10minutes at 4^OC. The supernatant was transferred to a new Falcon 2063 tube.

5. The amplified λ-phage stock was stored with few drops of chloroform at 4^OC. The titer after

the amplification should be over 1×10^{10} pfu/ml.

λ _phage DNA preparation_

1. 0.4ml of the amplified λ-phage lysate was taken into an eppendorf tube and mixed with 16μl of 0.5M EDTA(pH8) and 4μl of 10%(w/V) SDS, and incubated at 55^OC for 5minutes to denature the coat proteins.

2. The lysate was then mixed with equal volume of phenol:chloroform, shaken gently for 10minutes at room temperature and aqueous phase was separated by centrifugation at 16,000rpm for 4minutes. The upper aqueous phase was transferred to a new eppendorf tube, and repeated this step several times until a fluffy interlayer (protein) almost disappeared.

3. The aqueous phase was extracted once with equal volume of chloroform:isoamylalcohol for 10minutes and separated by centrifugation as above. The phage DNA was recovered by ethanol

precipitation.

4. The resultant λ phage DNA was dissolved in 100-200μl of autoclaved Milli-Q water and stored at 4°C (short term) or -20°C (long term). λ phage DNA often required considerable time(1-2 days) for dissolving, but should not use a vortex mixer.

Screening of the genomic λ phage library

The genomic λ phage library was screened with probes for the *psa*A, *psa*B and *psa*C genes of tobacco chloroplast DNA. The plaques were transferred onto Colony/PlaqueScreen (NEN, DuPont) disks ($\emptyset$82mm) by capillary action. The blotted disks were then hybridized with ^{32}P-labeled appropriate probe as described at 'Hybridization'. I could finally isolate the 6 and 12 clones from ~30,000 clones with the probes of the *psa*A-*psa*B and the *psa*C genes, respectively.

DNA subcloning and transformation

A DNA fragment was separated by agarose or polyacrylamide gel electrophoresis (Sambrook *et al.* 1989) then recovered into a small sealed dialysis tube by electroelution with a Mupid-2 electrophoresis tank (15-20minutes at 100V was sufficient for complete elution from the gel). The electroeluted DNA fragment solution was extracted once with phenol:chloroform and then with chloroform:isoamylalcohol, and recovered by ethanol precipitation. A digested cloning vector DNA, previously treated with bacterial alkaline phosphatase, was simultaneously separated and recovered from the gel. The isolated DNA frag-ment and the vector DNA was ligated by T4 DNA-ligase with vector to insert ratio of 1:1.2. If necessary, DNA was recovered by ethanol precipi-tation to remove PEG because residual PEG occa-sionally causes trouble in the following steps.

The ligated DNA fragment-vector was introduced into an appropriate *E.coli* strain by using a highly efficient transformation protocol (Hana-

han, 1985). When JM109/F' was used as a recipi-
ent, cultured cells in SOB medium were harvested
at A_{600}=1.2 to obtain maximum efficiency of
transformation. The transformed cells were
screened with appropriate antibiotics and, if
possible, by suitable indicator agar plates.

Single-stranded DNA preparation using M13KO7 helper phage

Since cloning vectors pTZ19R, pGEMEX1 and derivatives, also called phagemid, contain the replication origin of phage-f1, these plasmids could produce a single-stranded DNA(ssDNA) which would be used as a template for sequence analysis with an aid of appropriate helper phage like M13KO7 (Vieira *et al.* 1987).

The production of ssDNA and maintenance of a helper phage M13KO7 was performed according to the method of Vieira and Messing (1987). Produced ssDNA was isolated and purified as below:

1. 1.2ml of a culture was taken into an eppendorf tube, centrifuged for 5minutes and 1ml of the supernatant was transferred to a new eppendorf tube.

2. The viral particles were precipitated by mixing 0.25ml of 20%(w/v) PEG-6000-2.5M NaCl solution for 15minutes at room temperature and recovered by centrifugation at 16,000rpm for

5minutes. The supernatant was completely removed
by second centrifugation since remaining PEG
would cause pseudoladders when used as a template
for the sequence analysis.

3. Contaminated DNA and RNA could be eliminated
from the particle by a treatment with 0.1mg/ml
DNaseI and 0.01mg/ml RNaseA in a 100μl buffer
(50mM Tris-HCl, pH7.5, 10mM $MgCl_2$) at 37°C for
60 minutes. The particle was recovered again by
mixing with PEG-NaCl as described above.

4. The recovered viral particles were resus-
pended in 100μl of 1xTE (pH8.0) and extracted
once with phenol:chloroform and then with chloro-
form:isoamylalcohol. The ssDNA in aqueous phase
was precipitated with ethanol, washed once with
70% (v/v) ethanol and dissolved in 30μl of 1xTE
(pH8.0).

5. The resultant ssDNA solution was stored at
-20°C until use.

Unidirectional DNA deletion for sequence analysis

Unidirectional deletion was introduced into the target site of a plasmid DNA according to the original methods of Henikoff (1984), but some points were improved. The Reaction was performed as described below:

1. An intact plasmid DNA was digested with a set of two restriction endonucleases to create a 3'-proximal site to protect from a digestion by exonuclease III and a 5'-proximal site to be deleted with exonuclease III. The digested plasmid DNA was purified by gel-electrophoresis and dissolved in water.

2. A portion of digested and purified DNA (ca. 1.5pmole) was digested unidirectionally in 100μl solution (reaction buffer: 50mM Tris-HCl,pH8.0, 100mM NaCl, 5mM $MgCl_2$ and 10mM 2-ME) with 180-200U of exonuclease III at 37^OC, and an aliquot was removed at 30 to 90-second intervals (the time required to delete desired size should be

determined by a preliminary experiment for each
DNA fragment) to produce a size-varied deletion
set. The digested DNA was mixed and pooled in a
mung bean nuclease buffer (reaction buffer: 40mM
sodium-acetate, pH4.5, 100mM NaCl, 2mM $ZnCl_2$, 10%
glycerol) in another eppendorf tube in ice, then
incubated at 65^OC to inactivate the exonuclease
III. The DNA could be deleted at the rate of
300bp/minute.

3. The deleted DNA was treated with 50U of mung
bean nuclease at 37^OC for 30-60minutes (time was
dependent on the size to be deleted) to digest
the single-stranded region . The sample was ex-
tracted once with phenol:chloroform and precipi-
tated by ethanol (the pH of the solution at the
phenol:chloroform extraction should be maintained
at alkaline pH by adding a small amount of 1M
Tris-HCl, pH8.0 for the good recovery of the
DNA).

4. The recovered DNA was dissolved in water and
the termini were repaired with T4-DNA polymerase

and recovered by ethanol precipitation. The repaired DNA was self-ligated with T4 DNA ligase and ATP (see the 'DNA manipulation' section for the detail).

5. The resultant DNA was provided for the transformation of a strain JM109 with a high efficiency protocol of Hanahan(1985). Transformed cells were screened by ampicillin (100μg/ml).

Dideoxy chain termination method for sequence analysis

Nucleotide sequence of a DNA was determined by Sanger's dideoxy chain termination method with a 7-deaza M13 sequence kit (TAKARA) and a 7-deaza Sequenase version 2.0 kit (TOYOBO). A single-stranded template DNA was annealed with a primer (0.5pmole) by a series of incubation at 60°C for 5minutes, 37°C for 60 minutes and room temperature for 15minutes. Reaction protocol was according to the supplier's instruction but several points were altered. In the case of the 7-deaza Sequenase version 2.0 (TOYOBO), the labeling step was at 37°C for 5 to 10minutes with 10-fold diluted labeling mixture in the kit, and the elongation step was with 2.0μl of 7-deaza mixture plus 0.5μl of the elongation mixture (a solution of 180μM each of cold dNTP), or 1.5μl of the 7-deaza mixture plus 1.0μl of the elongation mixture at 37°C for 10minutes. In both methods, sequencing reaction were terminated by mixing with 6μl (for the kit of TAKARA) or 5μl (for the kit of TOYOBO) of formamide dye

2-20

(98%(v/v) formamide(BRL), 0.1%(w/v) xylenecya-nole FF, 0.1%(w/v) BPB and 1mM EDTA, pH8.0), boiled for 5minutes and quickly cooled in an ice bath, and 1.3-1.5μl was applied onto sequencing gel.

Gel electrophoresis was performed on buffer-gradient (Biggin *et al*.1983) and acrylamide-gradient gel (6%T-5.0xTBE:bottom to 4%T-1.5xTBE :top) with 1.5xTBE for an electrode buffer. The gel was fixed with 10%(v/v) methanol and 10%(v/v) acetic acid for 5-10minutes after electrophore-sis, transferred onto a Whatmann 3MM filter paper and dried at 65-75°C for 40-60 minutes. Autoradi-ography was done at -80°C for 8-18hours without intensify screen. I could usually read 350-400bases per template.

Nucleotide sequence analysis

Nucleotide sequences were analyzed by a PC9801RX computer (NEC,Japan) with a Genetyx software, version 6 (SDC,Japan).

Chemical synthesis of oligodeoxyribonucleotide

The oligodeoxyribonucleotides (oligo DNA) which were used for the oligonucleotide-directed site specific mutagenesis were synthesized in the 0.2μmole scale with the phosphoramidite method using a Cyclone DNA synthesizer (Milligen/Biosearch). Synthesized oligo DNA were provided for ammonia cleavage/de-protection at 55°C for 10hours according to the supplier's protocol. The synthesized 5'-DMTr-oligo DNA was injected into a C18 HPLC column (μbondasphere C18 (300A), 5μm, 0.39x15cm) then eluted with 5 to 40% linear gradient of acetonitrile in 0.1M TEAA (pH7.0) within 20minutes at 1ml/minutes for flow rate, and the elution was monitored by the absorbance at 260nm. The slowest eluted major peak was collected and treated with 80% acetic acid, extracted with diethylether and lyophilized to remove cleaved DMTr. The recovered oligo DNA was injected again into the same HPLC column and major peak was recovered. The recovered peak was lyophilized, dissolved in water, the amount was

analyzed by the absorbance at 260nm, and stored at -20°C until use.

Because the synthesized oligo DNA had -OH at 5'-terminus, oligo DNA should be phosphorylated with ATP and T4 polynucleotide kinase before use. If necessary, the efficiency of 5'-terminal kination reaction was analyzed with gamma-^{32}P-ATP as described below:

5'-phosphorylation analysis of oligo DNA

1. The 5'-OH oligo DNA was mixed with the following solutions and incubated at 37°C for 45-60minutes.

10xPNK oligo buffer	3μl
oligo DNA	0.5μl (20pmole/μl)
gamma-^{32}P-ATP	1μl
(ca.4,500Ci/mmole,10mCi/ml)	
water	8.5μl
T4 polynucleotide kinase	1μl (4U/μl)

---

2. A 2μl aliquot was taken to a eppendorf tube, added 8μl sequencing dye, and incubated at 65°C for 5minutes.

3. Aliqouts of the mixture were applied onto a 20% polyacrylamide slab gel containing 8M urea and 1.5xTBE (1mm-thick), and electrophoresed at 150V until a BPB front marker migrated to the bottom of the gel.

4. After electrophoresis, the gel was washed with 10%(v/v) acetic acid, 10%(v/v)methanol for 5 minutes, covered with Kre-wrap, and autoradiographed at room temperature for few hours.

5. The efficiency of the phosphorylation was determined by the already known concentration and length of oligo-DNA which was simultaneously phosphorylated and electrophoresed as a standard. The efficiency of the reactions should usually be over 90%.

Oligo nucleotide-directed site-specific mutagenesis

The oligo nucleotide-directed site-specific mutagenesis was performed by a modified method of Inouye and Inouye(1989) for gapped duplex plasmid DNA. Although the following protocol was originally adjusted to introduce a mutation for the pSVCX2 plasmid, it could be easily applicable to other plasmid vectors (see original report of Inouye and Inouye, 1989, for detail). Since mutations were designed on purpose to cause the introduction of a restriction endonuclease digestion site, the resulted mutant could easily be isolated with restriction endonuclease digestion analysis without other isotopic methods.

Step 1. Restriction endonuclease digestion of pSVCX2

The pSVCX2 plasmid DNA was digested with each of *Xba*I/*Eco*RI and *Sca*I. The *Sca*I digested DNA was treated with bacterial alkaline phosphatase(BAP). Both digested DNA were purified by gel electrophoresis. Digested and purified DNA fragments

were tentatively designated as fragment-I (for *Xba*I/*Eco*RI fragment) and fragment-II (for *Sca*I/BAP fragment), respectively. The concentration of both DNA fragments were adjusted to 0.1pmole/μl with water.

Step 2. Oligo-DNA for mutation

The nucleotide sequence of the each oligo DNAs used for the experiment were:

C21S: 5'-GTGCGGGCTAGCCCCACCG-3'   (*Nhe*I)

C17L: 5'-CCCAGCTGGTGCGGGC-3'   (*Pvu*II)

C34T: 5'-TGGGATGGTACCAAAGCTGGCCAA-3'   (*Kpn*I)

C48I: 5'-ACGGAAGATATCGTTGGCTGCAAA-3'   (*Eco*RV)

ϕ15G: 5'-CATTGGCTGCGGTACCCAGTGCACC-3'   (*Kpn*I)

ϕ53S: 5'-CTGCAAACGCAGCTGTGAAACGGCC-3'   (*Pvu*II)

When a desired mutation was introduced into the plasmid, each mutated DNA would be able to be digested by restriction endonucleases where were indicated in the parentheses.

2-26

Step 3.Phosphorylation of oligo DNA

The HPLC-purified oligo DNA was dissolved to be 20pmole/μl in autoclaved Milli-Q water. Following solutions were mixed and incubated at 37°C for 60minutes.

10xPNK oligo buffer*	5μl
oligoDNA(20pmole/μl)	5μl
1mM ATP	5μl
T4 polynucleotide kinase	0.5μl (5U)
water	34.5μl
total	50μl

Reaction was terminated by incubating the mixture at 65°C for 10minutes to denature the PNK. The resultant phosphorylated oligo DNA was stored at -20°C until use.

*10xPNK oligo buffer
500mM Tris-HCl,pH8.0, 100mM MgCl$_2$, 100mM 2-ME.

Step 4. Formation and confirmation of heteroduplex DNA

The following mixture was prepared and kept on ice.

Fragment- I(0.1pmole/μl)   2μl

Fragment-II(0.1pmole/μl)   2μl

Phosphorylated oligoDNA(2pmole/μl)   2μl

5xPL buffer**   2μl

Autoclaved Milli-Q water   2μl

_____

Total   10μl

A 2μl aliquot was removed to another eppendorf tube and stored in ice until use (Mix-A).

3. To construct the heteroduplex DNA, the mixture was incubated first at 100°C for 6minutes and immediately transferred to a 30°C bath, then kept there for 30minutes, and then in refrigerator (ca.4°C) for 30minutes and on ice for 30minutes. The mixture was collected at the bottom of the tube by centrifugation and 2μl was

2-28

transferred to another tube (Mix-B). The remainder was stored on ice.

4. Each mix-A and mix-B were mixed with $6\mu l$ of 2xdye and applied on 0.8% agarose gel. Ethidium bromide staining should reveal two bands for the mix-A and three bands for the mix-B. The slowest band appeared on the mix-B means newly formed heteroduplex DNA.

**5xPL buffer: 500mM NaCl, 32.5mM Tris-HCl,pH7.5, 40mM $MgCl_2$, 5mM 2-ME.

Step 5. Filling-in and ligation

To the remainder of the preceding step, solutions listed below was added:

Annealed mixture(step4)	6μl
10xT	2μl
10xBSA	2μl
10mM dNTP	1.5μl
10mM ATP	1μl
T4 DNA polymerase(1U/μl)	2μl
T4 DNA ligase(5U/μl)	1μl
Milli-Q water	4.5μl
total	20μl

The reaction was initiated by adding the enzymes to the mixture. The mixture was incubated first at 0°C for 5minutes and at room temperature for 5minutes to initiate correct priming reaction from the primer. Then the mixture was incubated at 37°C for 10-30minutes and then for another 20minutes at 15°C to complete ligation reaction.

To the reaction mixture, $2\mu l$ of 0.1M EDTA, pH8.0 and the enzyme was added. The enzymes were heat-denatured at 65^OC for 10minutes. The resultant DNA was provided for the first transformation.

Step 6. First transformation and screening

A competent cell for the transformation was prepared according to a 'simple-method' of Hana-han(1985). Whole amount of the DNA at step 5 was used for the transformation and all transformed cells were spread onto a 100μg/ml ampicillin containing agar plate. Over one-hundred colonies per plate should appear after incubation at 37^OC overnight. Strain JM109/F' should be chosen as a host for the first transformation because the strain did not have a gene for T7 RNA polymerase so the *psa*C gene in pSVCX2 could not express in the strain. This would suppress the appearance of undesired spontaneous mutant. All of the ap-peared colony was dispersed into 50ml of LB medium containing 200μg/ml ampicillin and shaken at 37^OC for 1hour. Plasmid DNA was prepared from

the culture by the alkaline-SDS method and dissolved in water.

The recovered plasmid DNA was digested with a appropriate restriction endonuclease and applied onto preparative scale 0.8% agarose gel. The mutant which had been introduced a new restriction endonuclease digestion site for the plasmid was recovered from the gel and self-ligated. (To ascertain the relative mobility of the digested DNA, digested plasmid DNA at a single site by a restriction endonuclease was simultaneously electrophoresed). In the case of a mutant which lost restriction endonuclease digestion site, undigested cccDNA band which migrated fastest was recovered.

The recovered DNA was provided again for transformation. The plasmid DNA of transformed colonies were individually analyzed by restriction endonuclease digestion and desired mutant was stored. (If desired mutant could not be found this stage, repeated step4.)

2-32

Step 7. Strain construction

The mutated plasmid DNA was introduced into a strain JM109(DE3). If required, plasmid pLysS (Studier *et al.*1990) was also introduced. Nucleotide sequence of the mutant was analyzed by the dideoxy chain termination method with the T7 RNA polymerase promoter-specific primer(Promega).

Preparation of total DNA of Synechococcus vulcanus

The total DNA of *Synechococcus vulcanus* was prepared by a method as described below. Although the method was originally described for *Synechocystis* sp. PCC6803 (Williams, 1988), some steps were altered to improve the yield of the DNA.

1. 1-2liter of *S.vulcanus* dense culture (A_{730} ~5) was centrifuged at 7,000xg for 5minutes. The harvested cells were weighed (1.5-2.5g) and stored at -80°C until use.

2. The frozen cells were thawed and treated with saturated sodium iodide solution (dissolve 4g of NaI with 2ml water) at 2ml/gram-wet weight. After the incubation of the cells at 37°C for 20minutes, add water to be 40ml/gram-wet weight and centrifuged at 11,000xg for 10minutes at room temperature. The supernatant was completely removed to exclude the residual NaI.

3. The NaI treated cells were suspended with

Lys-buffer (50mM Tris-HCl, pH8.0, 50mM NaCl, 5mM Na-EDTA) at 10ml/gram-wet weight, and egg white lysozyme powder was dissolved to be 75mg/gram-wet weight. The suspension was incubated at 37^OC for 45minutes then mixed with 10% N-lauroylsarcosine to be 0.5% and incubated further 20minutes. The lysozyme-treated spheroplasts would be lysed and the lysate should be viscous. Phase-constrast microscopy observation would reveal the residual intact cells. If the cells did not lysed completely, ca. 1g of baked grass beads (0.4mm∅) were added and burst few times by moderate vortexing. But this treatment should not be used if very high molecular weight DNA (>30kb) was required. Following phenol-chloroform treatment could also help the lysis of the cells.

4. Equal amount of phenol-chloroform was mixed and extracted for 30-60minutes by inverting the capped tube. The aqueous phase was separated and recovered by centrifugation at 12,000xg for 10minutes. If the aqueous phase was still remained green color, white interphase (DNA and

protein) was also recovered, mixed and extracted again. The aqueous phase was then extracted once with equal volume of chloroform-isoamylalcohol for 30 minutes and the aqueous phase was recovered again.

5. The cyanobacterial DNA was recovered by ethanol precipitation and dissolved into 10ml of 0.2x TE. Considerable period often required for the complete dissolving. $10\mu l$ RNase A (10mg/ml) was mixed and incubated for 15minutes at room temperature, then centrifuged at 14,000xg for 10minutes to remove insoluble matter. For the recovered supernatant 2.5ml of 20%(w/v) PEG-6000/2.5M NaCl was added. The mixture was mixed by inverting the capped tube then stand at room temperature for 30minutes (or at 4^OC for overnight).

6. The partly purified cyanobacterial DNA was recovered again by centrifugation at 14,000xg for 10minutes. The residual PEG-containing supernatant should be removed with a brief centrifugation. The precipitated DNA was then dissolved

2-36

into 8ml of 1xTE at 4°C for overnight.

7. To 8ml of the DNA solution, 8.8g of CsCl and 0.8ml of 10mg/ml EtBr solution were added and mixed by inverting. The insoluble red matter was excluded by centrifugation at 12,000xg for 20-30minutes. The supernatant was transferred into a Hitachi 12PA tube then overlaid with mineral oil. After ultracentrifugation at 49,000rpm for 21 hours in Hitachi RP-55T rotor at 4°C, single DNA band was carefully transferred into a new clean tube through an 18G needle under the illumination of black light.

8. The CsCl purified DNA was extracted once with equal amount of water-saturated n-buthanol. The separated n-buthanol phase was back-extracted with 0.5ml water 2 times. The recovered aqueous phases were combined and extracted once with equal amount of chloroform/isoamylalcohol, then extracted once with equal amount of diethylether. The DNA was finally precipitated with ethanol and dissolved into 200-400μl water. Over 100μg DNA

usually obtained from culture of one liter.

Preparation of cyanobacterial total RNA

The total RNA of *Synechococcus vulcanus* was extracted by guanidium thiocyanate(GTC) method (Shimizu, unpublished) as follows (all solutions, tips and tubes used in this experiment were prepared to be RNase-free. See Berger and Kimmel(1987); Golden *et al.*(1987) for detail):

1. 5-10ml cyanobacterial culture of early to mid-log phase was harvested by centrifugation at 6,000rpm for 5minutes at room temperature. Supernatant was removed and cell-precipitate was immediately frozen by liquid-nitrogen, and stored at -80°C until use.

2. The frozen cells were suspended in 200μl of a GTC-mix (4M guanidium thiocyanate, 8mM EDTA, 50mM Tris-HCl, pH7.6) and transferred to an eppendorf tube after thawing.

3. The following solutions were added into the GTC-mix suspension:

200μl TE-phenol (pH8)

    with 0.2%(w/v) 8-hydroxyquinoline

200μl Chloroform

100mg Acid washed, baked glass beads(0.4mm$\emptyset$)

 10μl 10%(w/v) SDS

---

4. The mixture was vortexed at top speed in three burst of 10seconds each, chilling the tube for a few minutes after each burst, then centrifuged at 15,000rpm for 1minute to separate the organic and aqueous phases.

5. The upper aqueous phase was transferred to a new eppendorf tube, extracted 2 to 3 times more with equal volume of phenol/chloroform(1:1) and centrifuged at 15,000rpm for 3minutes.

6. The upper aqueous phase was extracted 3 to 4 times with equal volume of diethylether. The final lower aqueous phase should be clear and may be pale-pink.

7. The RNA in the aqueous phase was precipitated by mixing with 0.4 volume of 10M LiCl, stand at $-20^{O}C$, overnight and recovered by centrifugation at 16,000 rpm for 20minutes in a microfuge.

8. The isolated total RNA was suspended in 100μl of RNase-free water, and dissolved RNA was recovered again by precipitation with 0.1volume sodium-acetate (pH5.0) and 2.5volumes ethanol to eliminate LiCl, washed once with 70%(v/v) ethanol, dried *in vacuo*, and suspended in 20μl of water. The resultant total RNA was mature and almost free of DNA when analyzed by gel electrophoresis and northern hybridization.

Southern blotting

The gel electrophoresed DNA was blotted onto nylon membrane (Immobilon-N, Millipore) by capillary action, alkaline blotting (Sambrook *et al*.1989) or by electrophoresis (Shimizu, unpublished).

The electroblotting method has the advantage of obtain several blotted membranes from a gel within few hours.

Electroblotting method of DNA

1. After electrophoresis, gel (agarose or polyacrylamide) was immersed into 0.5xTBE solution and shaken several times for 30minutes each at room temperature for lowering the ionic strength of the gel.

2. Electroblotting was carried out in a blotting tank filled with 0.5xTBE at 4°C. DNA was transferred onto hydrated Immobilon-N membrane at constant current (150mA) for 30minutes.

3. If sufficient amounts of DNA (5~10μg) were
electrophoresed, it was possible to make several
sheets of membrane of blot under the following
conditions that:

Order of membrane	Current	time for blot(minutes)
First	150mA	20
Second- third	180mA	20-25
Fourth- sixth	200mA	30

4. The blotted sheets were treated immediately
with alkaline solution (0.5N NaOH, 1.5M NaCl)
for 10minutes at room temperature with blotted-
side up.

5. Denatured DNA on the membrane was neutralized
with a solution (0.2M Tris-HCl, pH7.5) for

5~10minutes and the membrane was washed once in a 5xSSC solution, dried with filter papers and stored at -20^OC until use. I could store the blotted membranes for at least several months under this storage condition.

RNA gel electrophoresis and blotting

Glyoxylation, electrophoresis and staining with acridine orange was performed according to Carmichael and McMaster(1980). Relative molecular size was estimated by using ribosomal RNAs (23S: 2,904nt; 16S: 1,541nt; 5S:550nt ; values obtained from the E.coli) and an RNA ladder(1.77, 1.52, 1.28, 0.78, 0.53, 0.40, 0.28, 0.16-kilobases) purchased from BRL as standards.

For northern hybridization, electrophoresed RNA were directly alkaline transferred without staining onto a Hybond-N$^+$ membrane (Amersham Japan), overnight by capillary action (Sambrook et al. 1989). Blotted membrane was washed once with 5xSSC for 5minutes and immediately provided for hybridization.

^{32}P-labeled DNA probe preparation

Labeling of a DNA fragment with ^{32}P was achieved by a random-prime labeling method (Feinberg and Vogelstein, 1983) or a replacement synthesis method with T4 DNA polymerase. Both labeling reaction used high specific activity (1,000-4,000Ci/mmole) alpha-^{32}P-dCTP aqueous solution (Amersham Japan and ICN). The long DNA fragment (>500bp) was ^{32}P-labeled by random-prime labeling method and the short DNA fragment (<500bp) was ^{32}P-labeled with T4 DNA polymerase. DNA labeling by replacement synthesis with T4 DNA polymerase was performed as described below:

DNA labeling by replacement synthesis

1. A DNA fragment should be purified by gel electrophoresis and dissolved in water for 100 ng/μl prior to use. The purified DNA was mixed with following solution in an eppendorf tube in this order:

DNA	$2 \mu l$
10xT4 buffer*	$1 \mu l$
10xBSA	$1 \mu l$
water	$4 \mu l$
total	$8 \mu l$

2. $2 \mu l$ of a T4 DNA polymerase ($0.4U/\mu l$)** was added and incubated few minutes to digest some part of the DNA (a time required for the digestion was determined by preliminary experiment). Although T4 DNA polymerase could digest DNA at 25-40bp/minute under this condition, a precise speed should be determined for each DNA probe.

3. The tube was quickly transferred to an ice bath and following premixed solutions were added immediately:

```
10xT4 buffer          1.5 μl

10xBSA                1.5 μl

1mM dATP              2.5 μl

1mM dGTP              2.5 μl

1mM dTTP              2.5 μl

alpha-32P-dCTP          4 μl

( 40μCi,3,000Ci/mmol )

water                 0.5 μl
_____

total                  15 μl
```

4. The mixture was incubated at 30-37°C for 5-10minutes to synthesize a new strand and incorporate the isotope into the DNA (prolonged reaction period should not be recommended because 3'->5' exonuclease activity would decompose the DNA even in the presence of dNTP).

5. The reaction was stopped by adding 1μl of 0.5M EDTA (pH8.0) and extracted once with equal volume of phenol-chloroform(1:1).

7. Unincorporated isotope was removed by pass-

ing through a Sephadex G-50 (Pharmacia, DNA grade) mini column (column size: ⌀0.9 x 4.4cm).

8. Elution was monitored by a GM survey meter or Celenkov counting by liquid scintillation counter, and fastest peak was collected. The recovered probe DNA was mixed to the hybridization mixture after heat denaturing.

*10xT4 buffer: 0.33M Tris-acetate, pH7.9, 0.66M sodium-acetate(pH8), 0.1M $(CH_3COO)_2Mg$, 5mM DTT.

**T4 DNA polymerase was diluted with the following solution: 50mM Tris-HCl, pH7.5, 10mM 2-ME, 0.1M $(NH_4)_2SO_4$, 0.1% BSA. T4 DNA polymerase was diluted just prior to use, and the remainder should be discarded.

Hybridization

Hybridization of immobilized DNA/RNA was performed usually under the following conditions: 40-50% formamide(BRL), 4 to 6x SSPE or SSC, 0.1mg/ml salmon sperm DNA (Wako), 0.5%(w/v) SDS and 10x Denhartd't solution (100xDenhardt's solution: 2% BSA, 2% Ficoll-400 (Pharmacia) and 2% polyvinyl pyrolidone (Sigma P-5288)).

Hybridization temperature were determined on each experiments according to the empirical calculating methods of Meinkoth and Wahl(1984).

Hybridized membrane was washed with appropriate solution by increasing the temperature. The efficiency of washing was monitored by GM-survey meter. Existence of SDS at washing decreased the background efficiently, but since SDS confuse the autoradiograph pattern so membrane was treated with SDS-free solution at the end.

Protein assay

Concentration of protein was determined with a CBB G-250 dye method(Bradford, 1976). Bovine plasma protein (Bio-Rad) was used as a standard for assay.

SDS-PAGE

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to the method of Ikeuchi and Inoue(1988), but urea was omitted from the recipe. A relative size of protein was estimated from a standard protein marker kit (phosphorylase b, 94kDa; bovine serum albumin, 67kDa; ovalbumin, 43kDa; carbonic anhydrase, 30kDa; soybean trypsin inhibitor, 20.1kDa; alpha-lactoalbumin, 14.4kDa) and a peptide marker kit (myoglobin and its BrCN fragments of 14.6, 8.2, 6.4, 2.6 and 1.7kDa) of Pharmacia. Gels were fixed first in 7%(v/v) acetic acid for 30 minutes then stained with 0.25% (w/v) CBB R-250 (Kodak) (Diezel *et al*.1972), 10%(v/v) metha-

nol-7%(v/v) acetic acid for 1hour, and destained
in 7% (v/v) acetic acid with several sheets of
Kimwipe paper (Kimberly-Clark) (Ikeuchi and
Inoue, 1988) until background dye almost disap-
peared.

Immunological identification of proteins

Electrophoretically separated proteins in SDS-
PAGE gel were blotted onto a polyvinylidene
difluoride membrane (PVDF;Millipore) in 25mM
Tris (pH was NOT adjusted) and 20%(v/v) metha-
nol at 150mA current constant for 2 to 3hours at
4^OC in a blotting tank. A portion of the mem-
brane where the standard protein was blotted was
cut off and stained with India ink (Hughes *et
al.* 1988). The remainder was blocked with
2%(w/v) skim milk in PBS-Tween (1xPBS with
0.05%(w/v) Tween-20) at room temperature for
2hours with gentle shaking or at 4^OC, overnight
(Tsang *et al.*1983). The blot was then treated
with rabbit first antibody in PBS-Tween-skimmilk
at room temperature for 2hours with gentle shak-

ing, and washed more than three times with PBS-Tween for 15minutes each. After that, the blot was treated with 300-500 folds diluted horse-radish peroxidase conjugated goat anti-rabbit IgG polyclonal antibody (E-Y labs.) in PBS-Tween-skimmilk at room temperature for 1hour, and washed same as above. The antibody absorbed protein band was visualized by staining with Immuno stain kit (Konica).

If necessary, cross reactive components in first antibody against *E.coli* originated proteins were adsorbed by the following method:

1. A strain of *E.coli* appropriate for experiment was cultured in 50ml LB. Ampicillin and IPTG were added to the medium if required. The cultured cell was harvested by centrifugation, suspended in 2ml of 1xPBS and autoclaved at 105°C for 5minutes.

2. To the heat treated cell suspension, 1.5g of skim milk powder was added and filled up for 50ml with 1xPBS.

3. Appropriate volume of a rabbit polyclonal antisera (laboratory stock) was mixed with the heat-treated *E.coli* suspension and kept at 4°C, overnight.

4. The residual precipitate was removed by centrifugation at 18,000rpm for 30minutes and supernatant was transferred into a clean bottle, mixed with Tween-20 to be 0.05% and 0.01% NaN$_3$ or streptomycin, and stored at 4°C until use.

Flash spectrophotometry of P$_{700}$

Absorbance change of P$_{700}$ was measured at near UV region (430nm) by a laboratory made single beam spectrophotometer connected with a signal averager (Hiyama and Ke, 1971). P$_{700}$ was coupled with TMPD and sodium ascorbate (Hiyama and Ke, 1972) containing buffer-C mixture (100mM Tris-HCl, pH8.8, 0.2M Sucrose, 0.5mM 2-ME and, if required, 0.05% Triton X-100). Excitation of P700 was achieved by a xenon flash through red-filters. The transient absorbance changes of

P_{700} were averaged 4 to 16 flashes at 30-seconds intervals each.

Spinach photosystem I and CPI preparations

PSI and CPI preparations of spinach (Hiyama *et al.*in preparation) were kindly provided from Mr.S.Kobayashi. CPI was prepared from spinach PSI by treating the spinach PSI with 2M NaI at $4^{O}C$, and immediately ultlafiltrated by a UK-200 membrane (Advantec TOYO, Japan) with nitrogen pressure. Ultrafiltration was performed to near-dryness on the filter. Both preparations were stored at -20 or $-80^{O}C$ until use.

psaC *gene expression and holo PSI-C fraction isolation*

An *E.coli* strain pSVCX1/JM109(DE3) were inocu-
lated into 5ml of EX-broth (1.6%(w/v) Polypep-
tone, 1.0%(w/v) Bacto Yeast Extract, 1.0% NaCl)
supplemented with 100μg/ml ampicilin and
0.2%(w/v) glucose, and shaken at 37°C for 12-14
hours. 500μl of the pre-culture was inoculated
to 50ml of new EX-broth containing 50μM FeSO$_4$
but glucose-free and shaken for 3hours at 37°C.
An IPTG solution was added to the media to be
0.5mM and shaken further 2hours. The cells were
harvested by a centrifugation at 4,000xg for
5minutes at 4°C, wet-weight was estimated, and
suspended into 5ml of solution-A (50mM Tris-HCl,
pH8.8, 1mM EDTA, 50mM NaCl, 5mM DTT and 1mM
PMSF) per g-wet weight of the cells (the solu-
tion and centrifuge tube used bellow were filled
with nitrogen gas before use). Suspended cells
were passed two times through French press at
8,000p.s.i., transferred into a centrifuge tube
and 5 to 10 times volume of nitrogen-replaced
solution-A was mixed. After centrifugation at

4,200xg for 5minutes at 4°C, the supernatant was discarded and the resultant pellet was resuspended in 1.5ml of solution-A, and centrifuged again at 1,000xg for 1minutes to get rid of the unbroken cells. The prepared PSI-C rich inclusion body fraction was stored at 4°C under nitrogen atmosphere and used immediately or frozen at -80°C until use.

Amino acid sequence analysis

Amino acid sequence of PSI-C protein was determined with Shimadzu PSQ-1 protein sequencer. A crude extract containing the PSI-C protein whose expression was promoted with IPTG in *E.coli* was subjected for SDS-PAGE. After the electrophoresis, separated protein band was directly electroblotted from the gel onto Immobilon PVDF membrane and stained with amido-black. The stained protein band which corresponded to the PSI-C protein was excised from the membrane and carefully washed with Milli-Q water. The excised membranes were provided for sequence analysis.

Reconstitution

The isolated inclusion body fraction was solubilized just prior to reconstitution by 0.5%(w/v) Triton X-100 for 1hour at 4^OC in nitrogen atmosphere. The mixture was centrifuged at 7,600xg for 5minutes and the supernatant was recovered. A 3 volumes of the spinach CPI preparation (3 to 6μM as P_{700}) was mixed with the solubilized inclusion body fraction and kept for 1hour at room temperature, then ultlafiltrated with Molcut II(Millipore, exclusion limit 30,000Da) at 4^OC. The concentrated, tar-like mixture was resuspended with aliquot of 0.1M Tris-HCl(pH8.8), concentrated again and resuspended with the same buffer. Reconstituted PSI(rPSI) was provided for the flash spectrophotometry analysis and for low temperature EPR spectroscopy analysis.

Electron microscopy

E.coli cells were fixed directly in culture medium with 3% glutaraldehyde for 45minutes. They were collected by low speed centrifugation (3,000xg) for 5 minutes, and the resuspended in 0.2M cacodylate buffer (pH7.2) containing 3% glutaraldehyde and left for 2hours. Following two times of washing (30minutes each) with the same buffer, they were post-fixed with 1% osmium tetraoxide in 0.1M cacodylate buffer for 2hours, briefly washed and centrifuged into a pellet, and taken up into agar blocks by adding 4% agar in distilled water to the pellet.

The agar blocks were then excised into small pieces, dehydrated through a graded series of acetone and embedded in Spurr resin. Thin sections were stained with aqueous uranyl acetate and then lead citrate, and observed with a Hitachi H700H electron microscope at 100kV.

Low temperature EPR analysis

Electron paramagnetic resonance (EPR) signals of PSI, CPI and rPSI at low temperature were measured according to the method of Hiyama *et al.*(1985).

ABBREVIATIONS

cccDNA, covalently closed circular DNA;

CPI, Core complex of photosystem I;

Dithionite, sodium hydrosulfite;

DMTr, Dimethoxytrithyl;

F_A, Iron-sulfur cluster center A;

F_B, Iron-sulfur cluster center B;

GTC, Guanidium thiocyanate;

IPTG, Isopropyl-β-$_D$-thiogalactopyranoside;

moi, multiplicity of infection;

PEG, Polyethyleneglycol;

pfu, plaque formation unit;

PMS, *N*-methylphenazinium methylsulphate

   (phenazine methosulphate);

PMSF, phenylmethylsulfonylfluoride;

PNK, T4 polynucleotide kinase;

PSI, Photosystem 1;

rPSI, reconstituted photosystem 1

TEMED, *N,N,N',N'*-tetramethylethylenedinamie;

TMPD, *N,N,N',N'*-tetramethyl-*p*-phenylenediamine;

REFERENCES

Berger, S.L. and Kimmel, A.R. (1987) *Methods in Enzymology* **152** (Guide to molecular cloning techniques), Academic press, Orlando.

Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl.Acad.Sci.USA.* **80**, 3963–3965.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal.Biochem.* **72**, 248–254.

Carmichael, G.G. and McMaster, G.K. (1980) The analysis of nucleic acids in gels using glyoxal and acridine orange. *Methods in Enzymology* **65**, 380–391.

Chang, A.C.Y. and Cohen, S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J.Bacteriol.*

2-61

134, 1141-1156.

Diezel, W., Kopperschläger, G. and Hofmann, E. (1972) An improved procedure for protein staining in polyacrylamide gels with a new type of coomassie brilliant blue. *Anal. Biochem.* 48, 617-620.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal.Biochem.* 132, 6-13.

Golden, S.S., Brusslan, J. and Haselkorn, R. (1987) Genetic engineering of the cyanobacterial chromosome. *Methods in Enzymology* 153, 215-231.

Hanahan, D(1985) Techniques for transformation of *E.coli* in **DNA Cloning** vol.I, 109-135,edited by Glover, D.M., IRL press, Washington DC.

Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene*, 28 351-359.

Hiyama, T. and Ke, B. (1971) A further study of P430: a possible primary electron acceptor of photosystem I. *Arch.Biochem.Biophys.* **147**. 99-108.

Hiyama, T. and Ke, B. (1972) Difference spectra and extinction coefficients of P_{700}. *Biochim.Biophys.Acta* **267**, 160-171.

Hiyama, T., Murakami, A., Itoh, S., Fujita, Y. and Sakurai, H. (1985) A low potential iron-sulfur cluster bound to membranes in blue-green algae. *J.Biochem.* **97**, 89-95.

Hobot, J.A. and Rogers, H.J. (1991) Intracellular location of the autolytic *N*-acetylmuramyl-L-alanine amidase in *Bacillus subtilis* 168 and in an autolysis-deficient mutant by immunoelectron microscopy. *J.Bacteriol.* **173**, 961-967.

Hughes, J.H., Mack, K. and Hamparian, V.V. (1988) India ink staining of proteins on nylon and hydrophobic membranes. *Anal.Biochem.* **173**, 18-25.

Ikeuchi, M. and Inoue, Y. (1988) A new 4.8-kDa polypeptide intrinsic to the PSII reaction center, as revealed by modified SDS-PAGE with improved resolution of low-molecular-

weight proteins. *Plant Cell Physiol.* **29**, 1233-1239.

Inouye, S. and Inouye, M. (1989) Oligonucleotide-directed site-specific mutagenesis using double-stranded plasmid DNA. in *DNA and RNA synthesis* (edited by Narang, S.) Academic press.

Katoh, S. (1988) Photosystem I and photosystem II preparations from thermophilic *Synechococcus*. *Methods in Enzymology* **167**, 263-269.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Meinkoth, J. and Wahl, G. (1984) Hybridization of nucleic acids immobilized on solid supports(review). *Anal.Biochem.* **138**, 267-284.

Miller, J.H. (1972) *Experiments in Molecular Genetics* Cold Spring Harbor Laboratory Press,New York.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular*

Cloning: a laboratory manual (second edition) Cold Spring Harbor Laboratory Press, New York.

Shin, M., Sakihama, N., Koike, H. and Inoue, Y. (1984) Purification of C-type cytochromes from a thermophilic blue-green alga, *Synechococcus vulcanus*, by hydrophobic column chromatography using Toyopearls. *Plant Cell Physiol.* **25**, 1575-1578.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, 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) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO J.* **5**, 2043-2049.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* **185**, 60-89.

Sugiura, M., Shinozaki, K., Zaita, N., Kusuda M. and Kumano

M. (1986) Clone bank of the tobacco(*Nicotiana tabacum*) chloroplast genome as a set of overlapping restriction endonuclease fragments:mapping of eleven ribosomal protein genes. *Plant Science* **44**, 211-216.

Tsang, V.C.W., Peralta, J.M. and Simons, A.R. (1983) Enzyme-linked immunoelectrotransfer blot techniques (EITB) for studying the specificities of antigens and antibodies separated by gel electrophoresis. *Methods in Enzymology* **92**, 377-391.

Vieira, J. and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods in Enzymology* **153**, 3-11.

Williams, John, G.K. (1988) Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods in Enzymology* **167**, 766-778.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Isolation and characterization of the genes for *mtnA-psaC* of *Synechococcus vulcanus*

Introduction

The *psaC* gene is found in chloroplast genomes of higher plants: tobacco (Hayashida *et al.* 1987), liverwort (Oh-oka *et al.*1987), barley (Hoj *et al.*1987), maize (Schantz and Bogorad, 1988), pea (Dunn and Gray, 1988) and wheat (Dunn and Gray, 1988). It encodes a photosystem-I associated polypeptide, variously designated as Subunit VII, 9kDa subunit, PsaC, PSI-C, etc., presumably the apoprotein for iron-sulfur clusters, Centers A and B. Koike *et al.*(1989) previously reported a partial N-terminal amino acid sequence of Subunit VII of a thermophilic

cyanobacterium, *Synechococcus vulcanus* and it was
a first and only report of the sequence of cyano-
bacterial 9kDa subunit. Because of the potential
importance of the cyanobacterial photosystem I
gene isolation for the genetical manipulation and
protein production, we attempted to isolate the
corresponding gene from the *S.vulcanus* genomic
library cloned into λ-Dash with tobacco *psaC* DNA
probe.

psaC gene

A 2,702bp DNA fragment which had been well hybridized to a tobacco *psa*C probe was subcloned into pTZ19R and nucleotide sequence was determined. The *psa*C gene was located at 2,059-2,261 of the fragment (Figure 3-1) and consists of 81 codons (Figure 3-2), the same length as the tobacco gene that Hayashida *et al.*(1987) previously reported as well as other higher plants (Oh-oka *et al.*1987; Høj *et al.* 1987; Schantz and Bogorad, 1988; Dunn and Gray, 1988). The G-C content of this thermophilic prokaryote gene is 54.1% which is considerably higher than those of mesophilic plants (less than 42%).

The deduced amino acid sequence is matched the previously reported amino-terminal sequence (Koike *et al.*1989) except for cysteine and tryptophan residues, and indicated comparable homologies of 87.5% for the deduced amino acid se-

quence(Figure 3-3), and 73.2% for the nucleo-
tide sequence when compared with the tobacco
sequence. All nine cysteine residues were en-
tirely conserved(Figure 3-4). These are now
known to be also conserved in other cyanobacteri-
al PSI-C proteins (Figure 3-3) determined later
(Bryant *et al*.1990; Anderson and McIntosh, 1991;
Takahashi *et al*.1991). Figure 3-5 indicates the
codon usage of the *psa*C gene.

Five nucleotides upstream of the start codon of
*psa*C is the AGGAG(Figure 3-2), which has simi-
larity to the Shine-Dalgarno sequence and may
constitute a ribosome binding site. A sequence
(TAGCTT) similar to the *E.coli* consensus -10
sequence (TATAAT) lies 67 nucleotides upstream of
the start codon of the *psa*C gene, while a se-
quence (TCTCCA) similar to the *E.coli* -35 se-
quence (TTGACA) lies 17 nucleotides upstream of
the potential -10 sequence (Figure 3-2). We
could not find meaningful similarities for 5'- or
3'-proximal regions of the *psa*C gene with those
of the chloroplast DNA in higher plants or cyano-
bacterium *Synechocystis* sp. PCC6803 (Anderson

3-4

and McIntosh, 1991). The Southern hybridization analysis of *S.vulcanus* genomic DNA revealed that the *psa*C gene would exist as an single copy gene in the cyanobacterium(Figure 3-10A).

There were several nucleotide sequences which were expected to form a hairpin structure at the downstream of the *psa*C gene. A nucleotide sequence which expected to form a hairpin structure and may work as the terminator of the *psa*C gene positioned between 2,342 to 2,376 (Figure 3-2). It consisted of an 11-nucleotide length stem and an 11-nucleotide length loop accompanied with T-rich sequences which were reminiscent of the structure of the bacterial *rho*-independent transcription terminator (Lewin, 1987).

The nucleotide sequence of the *psa*C gene was registered to the DDBJ/EMBL/GenBank databases under accession No.D00590.

References

Anderson, S.L. and McIntosh, L. (1991) Partial conservation of the 5' *ndhE-psaC-ndhD* 3' gene arrangement of chloroplasts in the cyanobacterium *Synechocystis* sp. PCC6803: implications for NDH-D function in cyanobacteria and chloroplasts. *Plant Mol.Biol.* 16, 487-499

Bryant, D.A., Rhiel, E., Lorimier, R.D., Zhou, J., Stirewalt, V.L., Gasparich, G.E., Dubbs, J.M. and Snyder, W. (1990) Analysis of phycobilisome and photosystem I complexes of cyanobacteria. *in M.Baltscheffsky(ed.), Current Research in Photosynthesis* II, 1-9.

Dunn, P.P.J. and Gray, J.C. (1988) Localization and nucleotide sequence of the gene for the 8kDa subunit of photosystem I in pea and wheat chloroplast DNA. *Plant Mol.Biol.* 11, 311-319.

Hayashida, N. Matsubayashi, T., Shinozaki, K., Sugiura, M., Inoue K. and Hiyama, T. (1987) The gene for the 9kd polypeptide, a possible apoprotein for the iron-sulfur centers A and B of the photosystem I complex, in tobacco

chloroplast DNA. *Curr.Gen.* **12**, 247-250.

Høj, P.B., Svendsen, I. Scheller, H.V. and Møller, B.L.
(1987) Identification of a chloroplast-encoded 9-kDa
polypeptide as a 2[4Fe-4S] protein carrying centers A and
B of photosystem I. *J.Biol.Chem.* **262**, 12676-12684.

Koike, H., Ikeuchi, M., Hiyama, T. and Inoue, Y. (1989)
Identification of photosystem I components from the cyano-
bacterium, *Synechococcus vulcanus* by N-terminal sequenc-
ing. *FEBS Lett.* **253**, 257-263.

Lewin, B. (1987) *Genes*(third edition) John Wiley & Sons,
New York.

Oh-oka, K., Takahashi, Y., Wada, K., Matsubara, H., Ohya-
ma, K. and Ozeki, H. (1987) The 8kDa polypeptide in photo-
system I is a probable candidate of an iron-sulfur center
protein coded by the chloroplast gene *frxA*. *FEBS Lett.*
218, 52-54.

Oh-oka, H., Takahashi, Y. and Matsubara, H. (1989) Topo-
rigical considerations of the 9-kDa polypeptide which

contains centers A and B, associated with the 14- and 19-
kDa polypeptides in the photosystem I complex of spinach.
Plant Cell Physiol. **30**, 869-875.

Schantz, R. and Bogorad, L. (1988) Maize chloroplast genes
*ndh*D,*ndh*E, and *psa*C. Sequences, transcripts and transcript
pools. *Plant Mol.Biol.* **11**, 239-247.

Shimizu, T., Hiyama, T., Ikeuchi, M., Koike, H. and Inoue,
Y. (1990) Nucleotide sequence of the *psa*C gene of the
cyanobacterium *Synechococcus vulcanus*. *Nuc.Acids Res.* **18**,
3644.

Takahashi, Y., Goldschmidt-Clermont, M., Soen, S.-Y,
Franzén, L.G. and Rochaix, J.-D. (1991) Directed chloro-
plast transformation in *Chlamydomonas reinhardtii*: inser-
tional inactivation of the *psaC* gene encoding the iron
sulfur protein destabilizes photosystem I. *EMBO J.* **10**,
2033-2040.

mtnA gene

Metallothioneins (MTs), cysteine-rich heavy-metal binding proteins, are ubiquitously present in eukaryotes (Hamer, 1986) and in some proka-ryotes, *Synechococcus* sp.(Olafson, 1984) and *Pseudomonas putida* (Higham et al. 1984). The amino acid sequence of the protein of *Synechococ-cus* sp. was previously determined where signifi-cant differences from eukaryotic counterparts were noted (Olafson et al. 1988). Genes for eukaryotic MTs have been reported in yeast(Butt et al. 1984) and sea urchin (Nemer et al. 1985). Recently, Robinson et al.(1990) reported the identification of a gene for metallothionein (*smt*A) from cyanobacterium *Synechococcus* PCC6301 and demonstrated that a *smt*A homologue was also expressed in *Synechococcus* PCC7942 (*Anacystis nidulans* R2).

We previously reported the nucleotide se-quence of the *psa*C gene for a photosystem-1 subunit of a thermophilic cyanobacterium, *Synech-ococcus vulcanus* (Shimizu et al. 1990). Subse-quently, we determined the nucleotide sequence of

3-9

the upstream region of this gene and found an ORF which is expected to direct a 57 amino acid long protein (Figure 3-1) and rich in cysteine (Figure 3-6). The nucleotide codon usage was indicated at Figure 3-7.

This ORF57 is located at the 1,757 bp upstream region of the *psa*C gene with the same direction of transcription (Figure 3-2). The six nucleotides sequence GGAGGA is located in upstream from the start codon of ORF57 and has similarity to the Shine-Dalgarno sequence. The nucleotide sequences TTGACA and TATACT which are also founded at -75 and -51 nucleotides upstream of the start codon of ORF57 are quite similar to the *E.coli* consensus -35 sequence (TTGACA) and -10 sequence (TATACT), respectively.

The expected amino acid sequence has homology to those reported for an MT from *Synechococcus* sp. (Olafson et al. 1984), especially for the arrangement of the cysteine residues (Figure 3-8). The ORF57 also has high homology to that of the gene for MT from *S*. PCC6301(Robinson et al. 1990) in nucleotide sequence (63.2%) as well as amino acid sequence (52.1%), and nine out of ten cysteine residues were also conserved (Figure

3-8). While it has little homology with those of eukaryote proteins, a predicted secondary structure by the Chou-Fasman method indicates that the ORF57 protein has some tendency to form a beta-structure near the amino terminal side and a beta-turn near the carboxyl terminal which is reminiscent of those of eukaryote MTs (Hamer, 1986).

A Northern hybridization analysis of the total RNA revealed that the ORF57 was mainly tran- scribed as a 320-base RNA together with other longer transcripts which remained even after a highly stringent washing (Figure 3-9).   This suggests that transcription of ORF57 is monocis- tronic.   There are several nucleotide sequences which would form the stem-loop structure at the downstream of the *mtn*A gene (data not shown). Some of these sequences should work for the terminator of the *mtn*A gene. The Southern hybri- dization analysis of *S.vulcanus* genomic DNA revealed that the *mtn*A gene would exist as an single copy gene in the cyanobacterium (Figure 3-10B).

The ORF57 is tentatively designated as the *mtn*A gene (DDBJ accession no. X53839).

3-11

The ORF57 is tentatively designated as the *mtn*A
gene (DDBJ accession no. X53839).

References

Butt, T.R., Sternberg, E.J., Gorman, J.A., Clark, P., Hamer, D., Rosenberg, M. and Crooke, S.T.(1984) Copper metallothionein of yeast, structure of the gene, and regulation of expression. *Proc. Nat. Acad. Sci. USA* 81: 3332-3336.

Hamer, D.H. (1986) Metallothionein *Ann. Rev. Biochem.* 55: 913-951.

Higham, D.P., Sadler, P.J. and Scawen, M.D. (1984) Cadmium-resistant *Pseudomonas putida* synthesizes novel cadmium proteins. *Science* 225: 1043-1046.

Nemer, M., Wilkinson, D.F., Travaglini, E.C., Sternber, E.J. and Butt, T.R. (1985) Sea urchin metallothionein sequence: Key to an evolutionary diversity. *Proc. Nat. Acad. Sci. USA* 82: 4992-4994.

Olafson, R.W. (1984) Prokaryotic metallothionein. *Int. J. Pept. Protein Res.* 24: 303-308.

Olafson, R.W., McCubbin, W.D. and Kay, C.M.
(1988) Primary- and secondary-structural analysis of a
unique prokaryotic metallothionein from a *Synechococcus*
sp. cyanobacterium. *Biochem. J.* 251: 691-699.

Robinson, N.J., Gupta, A., Fordham-Skelton, A.P., Croy,
R.R.D., Whitton, B.A. and Huckle, J.W. (1990) Prokaryotic
metallothionein gene characterization and
expression:chromosome crawling by ligation-mediated PCR.
Proc. R. Soc. London B 242: 241-247.

Shimizu, T., Hiyama T., Ikeuchi M., Koike H. and Inoue
Y.(1990) Nucleotide sequence of the *psa*C gene of the
cyanobacterium *Synechococcus vulcanus. Nucl. Acids Res.*
18: 3644.
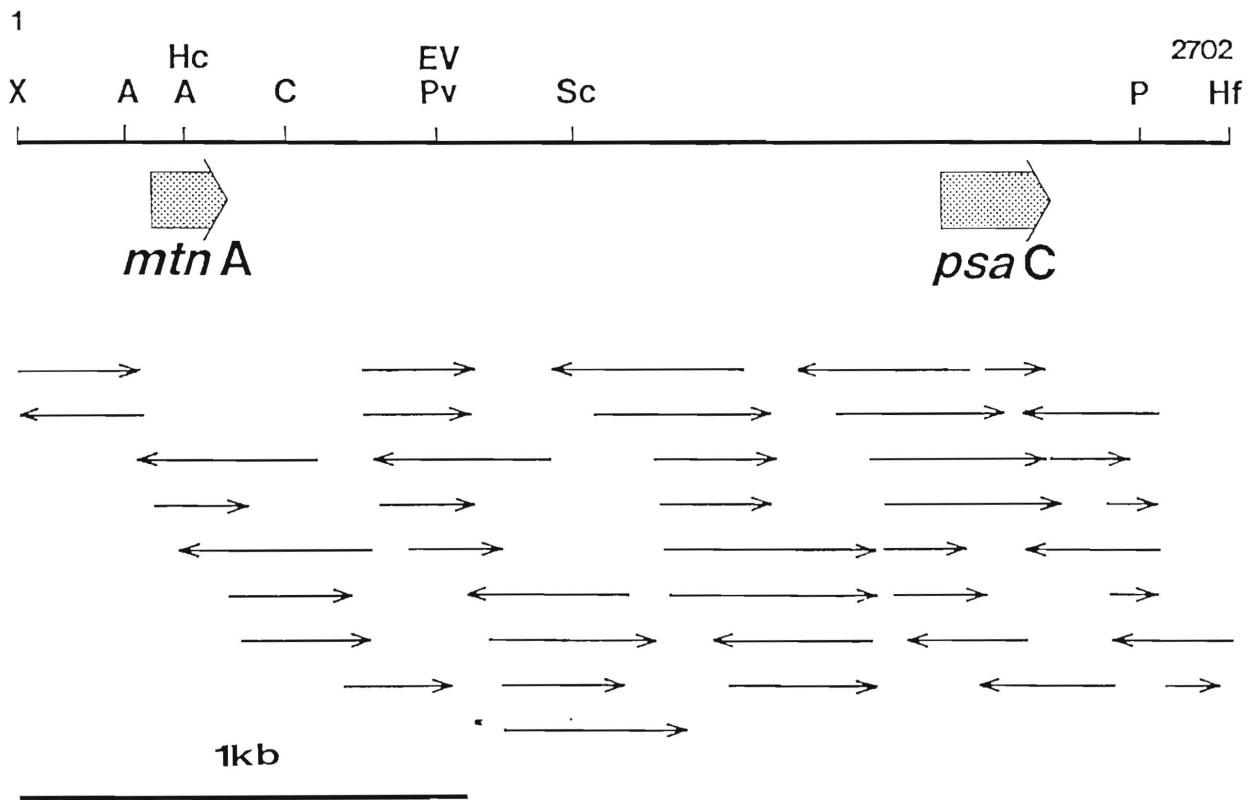
Restriction endonuclease map, sequencing strategy and gene location of mtnA-psaC gene. Dotted boxes represent the protein coding regions of *mtn*A(ORF57) and *psa*C genes. Restriction map is shown on the solid line: X, *Xba*I; A, *Acc*I; Hc, *Hinc*II; C, *Cla*I; EV, *Eco*RV; Pv, *Pvu*II; Sc, *Sca*I; P, *Pst*I; Hf, *Hinf*I. Arrows represent sequencing strategies.

5' - TCTAGAGCCACGGCAATATCCCTGACACAAAACTCCCCTTGGGCCAGGAGGGCCACAATGCGCACTCGATTCGTGTCAGCCAGTAGGCCA　90

AAAAACTGTGCCATCCGCTGGGCTTTTTCTGTAGAGAGCAGGCGATCGCTCACCTGCGCTAAAGCCTCTGGATGATGAGGATCAAGAGTATTCATATCTTCATCATAGGCGCTGGTTAGG　210

CAATAACCAATTGACATATGAACAGTTGTTCAAGTATACTGAACTAGAGGCGGAAAACAACCTCACTGATTCGTCATTGGAGGAAAAGCC ATG ACA ACC GTG ACC CAA ATG　322
　　　 *mtn*A　M　T　T　V　T　Q　M

AAA TGC GCC TGT CCC CAC TGC CTG TGC ATC GTC TCC CTC AAC GAT GCC ATT ATG GTC GAC GGC AAA CCC TAC TGC TCC GAA GTC TGT GCC　412
K　 C　 A　 C　 P　 H　 C　 L　 C　 I　 V　 S　 L　 N　 D　 A　 I　 M　 V　 D　 G　 K　 P　 Y　 C　 S　 E　 V　 C　 A

AAT GGC ACC TGC AAA GAG AAT AGC GGC TGT GGC CAC GCA GGT TGC GGC TGT GGC TCT GCC TAGAAGTTGCAATCTGTAACAGTGGATAATGACGAGGCGA　512
N　 G　 T　 C　 K　 E　 N　 S　 G　 C　 G　 H　 A　 G　 C　 G　 C　 G　 S　 A

TCGCCCCATCGACTCTATACAATTAAGGAGCTAGAGATTGTCCTCGTCATTTCCTATTAGTTGCGATCCAGCGTGGCGGACCTCCCATCGATTTCCTTGATTACAGCCACCTATAACGCT　632

GCCCAGCACTTGCCGGGGCTGATTGAGTCGGTTCGGCAACAGAGCGATCGCGCCTTCGAATGGATTGTGATTGATGGCGCCTCTGAGGATGGCACCCTTGATCTCATCAGAGCGGCCCAA　752

GATGTGCTCACAGACTATGTGAGTGAGCCAGATTTTGGCATCTTTCATGCATTGAACAAAGGAATCCAACGGGCAACCGGAGACTATTACCTTGTGGTTGGCGCAGATGATCGCCTCGAC　872

CCCCTAGCGATCGCCAACTACAAACAAGCCGTGAGAATGAGTCATGCCGACATTATTGCAGCTGATATCTATTCAGAAAATCAACGACATATACAACCCAAAAAAACACCCGTTTGGTTA　992

TCAGGTGCCAGATCCCTGATAAGCGCCCACGCCGTAGGAACGTTAATTAAAAAAACTCTCCACGATACTTACGGATATTATTGCCCTCTATACCCAACGATCGCCGATAGTGTGTTCCTT　1112

GTCGATGTTTTTCAAGCAGGTGTCAGCATTTATCATGCTCCCTTCATTGCTGGCCAACACTGCACGCAGGGCTTTTCTTCCCGCTACAAAGCCCTCAGCTTTGCTGAAAACTTCACGAAT　1232

CAGAAAAAGTACTCAATAAAATCTATCCAGTTCCTGCTTCTCATTGTGAGGCTACTTTGGTTTTATTTCCACTTGTGACCAATATTAACGCCCTTGCAATCGCAACATCTCGGCTTGGAT　1352

ACGTCGGCTCAAGGTGCTATCTTGCTGTGCTCGCAGTGTGATTGCATTGAATTGCTCTAGGGACATATTATATTTTTTCAAGACCTGTATTGCCTCCCACCCATAGTTAGCACAGATCGC　1472

CTCCAATCCACTGGGGATATGTTGCGGATTCATGCCAAAGCAAGCGTTGGACGGCACCTGACCTTGGAAAGCTGCCTGAGCTTGGCGATAGTATTTTTGGCGAATGGGCTCAATTTCGAG　1592

AACCCACACGGGCATAGGTCGTGATCTCCCACGGAGCAATCACATCCGTGTGCGGCGAAGTCTGAGCCAAGGAATAAGGTGAAAGCGCAGGGACCAGCAGCAGACCCACACTACACATAGC　1712

AAGGGGCAACCACTTCCAGGACATTGACCACCACATACAACGCGCCCGACTCCACTCTACTTGGAATCCCTAGACCCCTTTCAAGTTCCTCAACGCAGGCAAGCACCAGCAATTAAAAAC　1832

TGTTTTGGCCAAGATTAAGGGCACAATAGGCAAATTTACCAGACTTTACAAAAAAGTTACATTCTTTGAAGTTTGCCCATCCAGCCCGCCTAGCACTGAGTCACTGCGGTATCCTAAATC　1952

TCGGTATTTCTCTCCAGACTGCCCTTATTGTCGTAGCTTAGGGGCTTCCTGAGTTGGCGGCAGGGCATCCACCTGTCATCCTGATCGTTTTAGGAGCGGTTACTAC　ATG GCT CAC　2067
　　　 *psa*C　M　A　H

ACT GTC AAA ATT TAC GAT ACT TGC ATT GGC TGC ACC CAG TGT GTG CGG GCC TGC CCC ACC GAT GTG CTG GAA ATG GTG CCT TGG GAT GGT　2157
T　 V　 K　 I　 Y　 D　 T　 C　 I　 G　 C　 T　 Q　 C　 V　 R　 A　 C　 P　 T　 D　 V　 L　 E　 M　 V　 P　 W　 D　 G

TGC AAA GCT GGC CAA ATT GCC TCC TCC CCT CGC ACG GAA GAT TGC GTT GGC TGC AAA CGC TGT GAA ACG GCC TGC CCA ACT GAC TTT TTG　2247
C　 K　 A　 G　 Q　 I　 A　 S　 S　 P　 R　 T　 E　 D　 C　 V　 G　 C　 K　 R　 C　 E　 T　 A　 C　 P　 T　 D　 F　 L

AGT ATT CGT GTC TAT TTG GGT GCA GAA ACC ACC CGC AGC ATG GGT CTA GCC TAC TAGTCCCCTATTTTCGTGAAACTTTCGGCAAATTCTTGACTAGGCTTG　2349
S　 I　 R　 V　 Y　 L　 G　 A　 E　 T　 T　 R　 S　 M　 G　 L　 A　 Y　　　　　　　　　　　　　　　　　　　　　　　 ────────

AGGCAACATCCCTACGTCTCAAGTCTTTTATTTGTTTATTGGTTGTCCTTAGCACCAACCAACCCCTACTCACCTACAACTCAGGGCTGGTCTTCGTTACTAAGGCGGCTTTGTAGGTCA　2469
　→　　　　　　　　　　　　　　←

TCCACTGCTTCATTGATTTGGGCAATTTTAGCCTCTAGCTCTAGGCGCATCGCCTCCATCACAGGGTCATCATTCCCCTGCAGGGGAGTGGATCGCTGATCTTTAATCCGTGGCCTGTTG　2589

TCCGTGGCGGCGGCGGCAACTCTCTCCTTGAGTAGGAGCGTGACCAGAGAGCCAGTGATGCCGCCAAAGAAGGCGCCCACCAGAAACCCCAGCCAAACCCATTGCCACGATT - 3'　2702

Figure 3-2.

Nucleotide and deduced amino acid sequences of the cloned
*Synechococcus vulcanus mtn*A(ORF57)-*psa*C region.

```
Cyanobacteria  MxHxVKIYDTCIGCTQCVRACPxDVLEMVPWDFCKAxQIASSPRTEDCVGCKRCETACPTDFLSIRVYLGAETTRSMGLAY

               1         10        20        30        40        50        60        70        80     ( % )
S. vulcanus    MAHTVKIYDTCIGCTQCVRACPTDVLEMVPWDGCKAGQIASSPRTEDCVGCKRCETACPTDFLSIRVYLGAETTRSMGLAY   100.0
S. PCC7002     ·S·S·····················L·········································                      96.3
S. PCC8009     ·S·························A·······································                       97.5

S. PCC6803     ·S·S······················I·······K····A·············S·······V····WH··········        88.9
C. paradoxa    ······················R·N····A·············S··························G·               93.8
Chlamydomonas  ···I················L··········S·M··A·················V·····S·S······S·               88.9

Spinach        S·S························I·······K····A·············S·······V····WH·····N··G·         88.8
Tobacco        S·S························I·······K····A·············S·······V····WH·····N·····       87.5
Maize          S·S·········H············I·······K····A·············S·······V·····P·····NA·S·         85.0
Pea            S·S·····················I··G····K····A·············S·······V····WH·····N····          86.3
Wheat          S·S························I·······K····A·············S·······V·····P·····NA·S·         86.3
Barley         S·S························I·······K····A·············S·······V·····P······A·S·         87.5
Liverwort      ··A······················I·······N····A··········SR·······V·····N·····NA·S·           86.3
Rice           ·S·S······················I·······K····A·············S·······V·····P······A·S·        87.7

Higher plants  MxHxVKIYDTCIGCTxCVRACPTDVLEMIPWxGCKAxQIASAPRTEDCVGCKRCESxCPTDFLSVRVYLxxETTRSxxLxY
```

Figure 3-3.

Comparison of amino acid sequence between other organisms.

Figure 3-4.

Amino acid composition of the deduced *psa*C
protein.

Amino acid	residues
Gly	6
Ala	7
Val	6
Leu	4
Ile	4
Met	3
Phe	1
Trp	1
Pro	4
Ser	4
Thr	9
Asn	0
Gln	2
Cys	9
Asp	5
Glu	4
Lys	3
His	1
Arg	5
Tyr	3
Mw(Da)	8,800

Figure 3-5.

Amino acid codon usage of the *psa*C gene.

1	2	T	2	C	2	A	2	G	3
T	Phe	1	Ser	0	Tyr	1	Cys	2	T
	Phe	0	Ser	2	Tyr	2	Cys	7	C
	Leu	0	Ser	0	STOP	0	STOP	0	A
	Leu	2	Ser	0	STOP	0	Trp	1	G
C	Leu	0	Pro	2	His	0	Arg	1	T
	Leu	0	Pro	1	His	1	Arg	3	C
	Leu	1	Pro	1	Gln	1	Arg	0	A
	Leu	1	Pro	0	Gln	1	Arg	1	G
A	Ile	4	Thr	3	Asn	0	Ser	1	T
	Ile	0	Thr	4	Asn	0	Ser	1	C
	Ile	0	Thr	0	Lys	3	Arg	0	A
	Met	3	Thr	2	Lys	0	Arg	0	G
G	Val	1	Ala	2	Asp	4	Gly	3	T
	Val	2	Ala	4	Asp	1	Gly	3	C
	Val	0	Ala	1	Glu	4	Gly	0	A
	Val	3	Ala	0	Glu	0	Gly	0	G

Figure 3-6

Amino acid composition of the deduced *mtn*A
protein.

Amino acid	residues
Gly	7
Ala	5
Val	4
Leu	2
Ile	2
Met	3
Phe	0
Trp	0
Pro	2
Ser	4
Thr	4
Asn	3
Gln	1
Cys	10
Asp	2
Glu	2
Lys	3
His	2
Arg	0
Tyr	1
Mw(Da)	5,775

Figure 3-7.

Amino acid codon usage of the *mtn*A gene.

1	2	T	2	C	2	A	2	G	3
T	Phe	0	Ser	1	Tyr	0	Cys	4	T
	Phe	0	Ser	2	Tyr	1	Cys	6	C
	Leu	0	Ser	0	STOP	0	STOP	0	A
	Leu	0	Ser	0	STOP	0	Trp	0	G
C	Leu	0	Pro	0	His	0	Arg	0	T
	Leu	1	Pro	2	His	2	Arg	0	C
	Leu	0	Pro	0	Gln	1	Arg	0	A
	Leu	1	Pro	0	Gln	0	Arg	0	G
A	Ile	1	Thr	0	Asn	2	Ser	0	T
	Ile	1	Thr	3	Asn	1	Ser	1	C
	Ile	0	Thr	1	Lys	3	Arg	0	A
	Met	3	Thr	0	Lys	0	Arg	0	G
G	Val	0	Ala	0	Asp	1	Gly	1	T
	Val	3	Ala	4	Asp	1	Gly	6	C
	Val	0	Ala	1	Glu	1	Gly	0	A
	Val	1	Ala	0	Glu	1	Gly	0	G

```
                    1        10       20        30       40        50

S. sp. MT       TSTTLVKCACEPCLCNVDPSKAIDRNGLYYCCEACADGHTGGSKGCGHTGCNC

                *   *   **** *** *     **   *  ** * ** *     **** ** *

S. vulcanus MT  MTTVTQMKCACPHCLCIVSLNDAIMVDGKPYCSEVCANGTCKENSGCGHAGCGC

                *   *   **** *** *     **   *  **** ** *     **** ** *

S. PCC6301 MT   TSTTLVKCACEPCLCNVDPSKAIDRNGLYYCSEACADGHTGGSKGCGHTGCNC
```

Figure 3-8.

Comparison of the amino acid sequences of *mtn*A protein
between other organisms.

1   2   3   4

←320bases
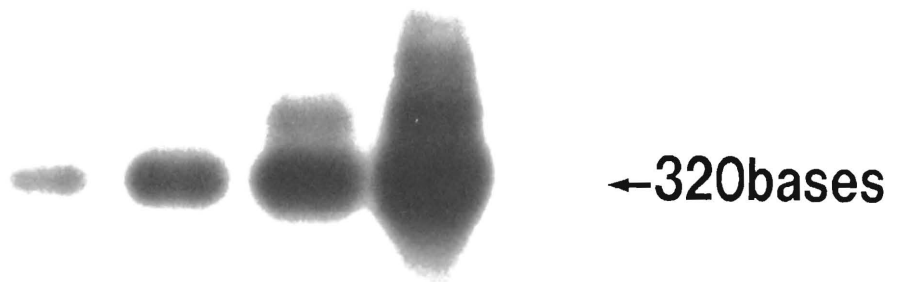
Northern hybridization analysis of the *mtn*A gene expres-
sion. The total RNA of *S.vulcanus* on Hybond-N$^+$ (Amersham)
were hybridized with a ^{32}P-labeled DNA fragment of *mtn*A
gene.   The amount of total RNA applied onto the gel were:
lane 1, 46.7ng; lane 2, 117ng; lane3, 230ng; lane 4, 700ng.
Hybridization was performed overnight at 42°C in 50%
formamide, 6xSSPE, 0.1mg/ml salmon sperm DNA, 10xDen-
hardt's and 0.5% SDS, then washed with 0.1xSSPE, 0.1% SDS
at 59°C for 5minutes.   The washed membrane was autoradi-
ographed for 52 hours at −80°C.
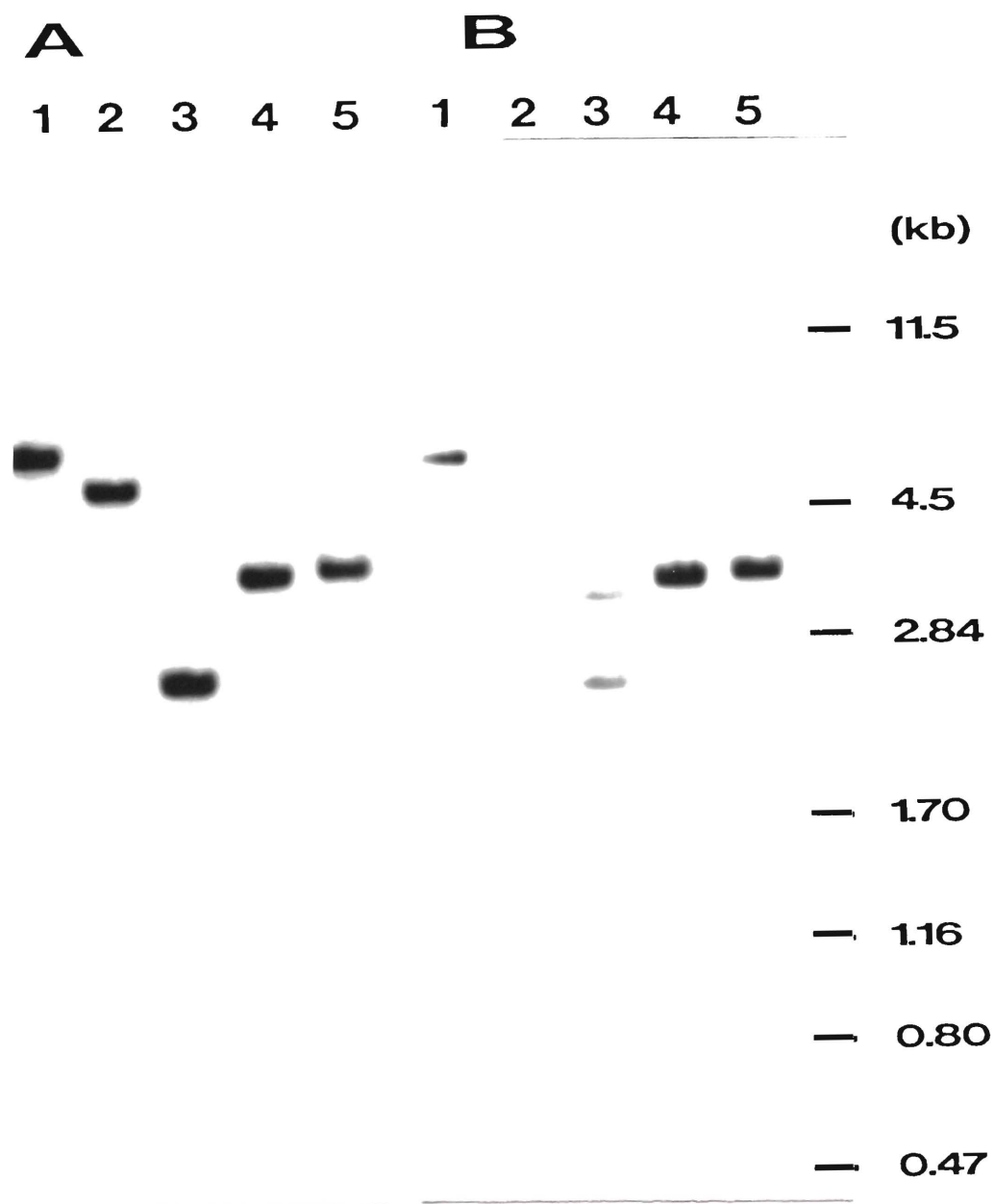
Figure 3-10.

The Southern hybridization analysis of the *Synechococcus vulcanus psa*C(A) and *mtn*A(B) gene. The genomic DNA of *S.vulcanus* (ca.8μg) was digested with 80U of *Xba*I (lane1), *Xba*I-*Cla*I (lane2), *Hinc*II(lane3), *Pst*I (lane4) and *Eco*RI (lane5) for 12hours, and electrophoresed in 0.8% agarose gel. After alkaline blotting onto the Immobilon-N (Milli-pore) by capillary action, each membranes was hybridized with ^{32}P-labeled DNA probe. The DNA probe for the *psa*C gene was a smaller DNA fragment of *Pst*I digested pSVC6 plasmid DNA. The DNA probe for the *mtn*A gene was iden-tical to those of the northern hybridization analysis (see Figure 3-9).

CHAPTER FOUR

Nucleotide sequences of the *psa*A and *psa*B genes encoding the photosystem I core proteins from thermophilic cyanobacterium *Synechococcus vulcanus*

Abstract

The cyanobacterial *psa*A and *psa*B genes for the photosystem I core proteins (PSI-A and PSI-B) were isolated with a tobacco chloroplast DNA probe and their nucleotide sequences were determined. Nucleotide and deduced amino acid sequences were quite similar to those of higher plants, rather than previously determined sequence of *Synechocystis* PCC7002. We could find interrupted ORF at the downstream of the *psa*B gene with same direction. One of the most characteristic feature of this proteins is five and

three cysteine residues in the *psa*A and *psa*B proteins, respectively.

RESULTS AND DISCUSSION

Structure of the psaA-psaB gene

A genomic DNA library of Synechococcus vulcanus which had been cloned into λ-Dash cloning vector (Shimizu et al.1990) were screened with the three kinds of ^{32}P-labeled DNA fragments of the tobacco psaA and psaB region (Shinozaki et al.1986) (Figure 4-1). The 5,685bp XbaI fragment and the 412bp SmaI-ClaI fragment which partly overlapped each other were both subcloned into the pTZ19R. The nucleotide sequences were determined with dideoxynucleotide chain termination method. The 5,945bp genomic DNA sequence (Figure 4-2) contained three open reading frames which were expected to encode 755, 741 and 223 amino acid residues (Figure 4-3). The 755 and 741 amino acid residues open reading frames were highly homologous with known psaA and psaB genes of higher plants and cyanobacterium (Cantrell and Bryant, 1987; Shinozaki et al. 1986; Kirsch et al. 1986; Fish et al. 1985; Kück et

4-3

al. 1987; Hiratsuka et al. 1989; Ohyama et al. 1986; Lehmbeck et al. 1986) respectively. These psaA and psaB genes predicted 83,198 and 83,043Da proteins, which were highly hydrophobic (62% and 60% of amino acids were hydrophobic each; Table 4-2) and contained eleven expected hydrophobic membrane spanning regions (Kirsch et al.1986) (Figure 4-4). The psaA and psaB genes have been found to express strongly and co-transcribed without the other ORF of downstream in the cyano-bacterial cell by northern hybridization analysis(data not shown). Furthermore, the Southern hybridization analysis of the S.vulcanus genomic DNA revealed that the psaA-psaB genes would exist as an single copy genes in the cyanobacterium, respectively (data not shown).

Promoters

The three possible sequences which should work for the initiation of transcription (promoter) were detected at positions, P1:ATGAGG (57-62)-TAACAT(80-85); P2: TTCTCT(149-154)-AACACT (172-

177); P3: TCGACT(211-216)-GAGAAT(237-242), that corresponded to -35 and -10 regions, respectively. The P1 was most likely to function as the promoter of the *psa*A-*psa*B operon. The nucleotide sequence homologies at the region between other genes of *S. vulcanus* will be discussed in the Chapter-5.

Terminators

We could find several nucleotide sequences which would form a hairpin-structure and work as the terminators at the immediately downstream of the *psa*B gene. One of these sequence (T1) positioned at 4,861-4,884 (Figure 4-6A) consisted of 9-nucleotide length stem and 4-nucleotide length loop at the top of the stem, and was accompanied by 8-T(U) residues. Other sequence (T2) positioned at 4,939-4,980 (Figure 4-6B) consisted of 6-nucleotide length basal stem and 2- and 4-nucleotide length other halfway stems with 4- and 9-nucleotide length loops, but was not accompanied by such an oligo-(dT) sequence

like the T1. The T2 is located at 55-nucleotides
downstream of the T1 and the T1 is located at
28-nucleotide downstream of the stop codon of
psaB. The T1 resembles the structure of bacteri-
al rho-independent terminator and the T2 looks
like the structure of bacterial rho-dependent
terminator (Lewin, 1987). Since psaA-psaB
operon may be a most vigorous genes of expression
in cyanobacterial cells, potent termination
signal must be required as well. These discovered
potential terminators, T1 and T2, seems suffi-
cient to terminate the transcription efficiently.

Interrupted ORF

The 223 amino acid residues open reading frame
which is located at the downstream of the psaB
gene is interrupted at a XbaI site of 5945bp and
the direction is same with the psaA and psaB
genes. The amino acid composition of this open
reading frame protein (partial) is highly hydro-
phobic (65% of known amino acid is hydrophobic)
and expected to contain some membrane spanning

4-6

regions like the *psa*A and *psa*B proteins. North-
ern hybridization analysis revealed that this
ORF is weakly expressed as a ca.1,900base-length
transcript (data not shown). So far, we have
not found neither nucleotide nor protein sequence
with any homologies to this open reading frame in
the current EMBL and SWISS-PROT databases, the
function of this open reading frame is not appre-
ciated at the moment.

Homologies of the deduced proteins

Comparison of the *psa*A and *psa*B protein se-
quences with those of other higher plants and
cyanobacteria (table 4-1) revealed strong simi-
larity (approximately 80% each). It should be
noted that the *psa*A protein of *S.vulcanus* is the
most homologous to the Liverwort (Ohyama *et
al.*1986), but not to the *S.* sp. PCC 7002 (Can-
trell and Bryant, 1987). This is probably due
to a significant difference at the amino-terminal
of the *psa*A protein. In fact the *S.vulcanus psa*B
protein is the most homologous to that of the

4-7

S. sp.PCC7002 *psa*B protein where there is no such difference at the amino-terminal region. There are some extra amino acid sequences in both *psa*A and *psa*B (264-267 for *psa*A and 313-316 for *psa*B) which have not been found in other organisms. These sequences might have something to do with the thermostability. It should also be noted that the unique amino acid sequence motif FPCDGPGRGGTC(604-615 in *psa*A, 568-579 in *psa*B in Figure 4-3, respectively), which has been speculated to act as the binding site for the 4Fe-4S iron sulfur cluster center-X (Scheller *et al.* 1989) is completely conserved among any *psa*A and *psa*B proteins of all the organisms reported so far (Cantrell and Bryant, 1987; Shinozaki *et al.* 1986; Kirsch *et al.*1986; Fish *et al.*1985; Kück *et al.*1987; Hiratsuka *et al.*1989; Ohyama *et al.*1986; Lehmbeck *et al.* 1986; Manzara *et al.* 1987). Leucine-zipper motifs (Weber and Malkin, 1990; Kössel *et al.*1990) are seen in both *psa*A and *psa*B protein. One characteristic difference is that the *psa*A protein contains five cysteine residues and *psa*B, three, compared with

4-8

four or three and two or three in other organisms (Cantrell and Bryant, 1987; Shinozaki *et al.* 1986; Kirsch *et al.*1986; Fish *et al.*1985; Kück *et al.*1987; Hiratsuka *et al.*1989; Ohyama *et al.*1986; Lehmbeck *et al.* 1986; Manzara *et al.* 1987). The function of these cysteine residues are not understood.

Homologies of the psaA-psaB *genes*

Cantrell and Bryant(1987) previously reported that, in *Synechococcus* sp. PCC 7002, the amino-terminal region of the *psa*A protein was shorter than that of higher plants. They also reported that the distance of the *psa*A and *psa*B intergenic region was 173bp where they found a 39 amino acid length open reading frame (Cantrell and Bryant, 1987). They proposed these differences as a unique feature of the cyanobacterial *psa*A and *psa*B genes and their proteins. In thermo-phillic cyanobacterium *Synechococcus vulcanus*, however, we could not find these differences. The intergenic region of the *psa*A to *psa*B genes is

4-9

23bp-length, quite similar to the length (22 to 23bp) of the intergenic region of the chloroplast DNA encoded ones of higher plants (Shinozaki *et al.* 1986; Kirsch *et al.*1986; Fish *et al.*1985; Kück *et al.*1987; Hiratsuka *et al.*1989; Ohyama *et al.*1986; Lehmbeck *et al.* 1986). Together with the sequence at amino-terminal region of the *psa*A described above, it is concluded that the overall features of the *psa*A and *psa*B genes of *S.vulcanus* is quite similar to the genes of higher plants.

The nucleotide sequence data reported in this paper will appear in the DDBJ,EMBL and GenBank nucleotide sequence databases under the accession number D01126.

REFERENCES

Cantrell, A. and Bryant, D.A. (1987) Molecular cloning and nucleotide sequence of the *psa*A and *psa*B genes of the cyanobacterium *Synechococcus* sp. PCC 7002. *Plant Mol.Biol.* **9**, 453-468.

Fish, L.E., Kück, U. and Bogorad, L. (1985) Two partially homologous adjacent light-inducible maize chloroplast genes encoding polypeptides of the P700 chlorophyll *a*-protein complex of photosystem I. *J.Biol.Chem.* **260**(3), 1413-1421

Hiratsuka. J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y.Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989) The complete sequence of the rice(*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol.Gen.Genet.* **217**, 185-194.

Kössel, H., Dory, I., Igloi, G. and Maier R (1990) A

leucine-zipper motif in photosystem I. *Plant Mol.Biol.* **15**, 497-499.

Kück, U., Choquet, Y., Schneider, M., Dron, M. and Bennoun, P. (1987) Structure and transcription of two homologous genes for the P700 chlorophyll *a*-apoprotein in *Chlamydomonas reinhardii*:evidence for *in vivo trans*-splicing. *EMBO J.* **6**, 2185-2195.

Kirsch, W., Seyer, P. and Herrmann, R.G. (1986) Nucleotide sequence of the clustered genes for two P700 chlorophyll *a* apoproteins of the photosystem I reaction center and the ribosomal protein S14 of the spinach plastid chromosome. *Curr.Genet.* **10**, 843-855.

Lehmbeck. J., Rasmussen, O.F., Bookjans, G.B., Jepsen, B.R., Stummann, B.M. and Henningsen, K.W. (1986) Sequence of two genes in pea chloroplast DNA coding for 84 and 82kD polypeptides of the photosystem I complex. *Plant Mol.Biol.* **7**, 3-10

Lewin, B. (1987) *Genes*(third edition) John Wiley & Sons, New York.

Manzara, T., Hu, J-X., Price, C.A. and Hallick, R.B.
(1987) Characterization of the *Trn*D, *Trn*K, *Psa*A locus of
Euglena gracilis chloroplast DNA. *Plant Mol.Biol.* **8**, 327-
336.

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) Chloroplast
gene organization deduced from complete sequence of liver-
wort *Marchantia polymorpha* chloroplast DNA. *Nature* **322**(7),
572-574.

Scheller, H.V., Svendsen, I. and Møller, B.L. (1989)
Subunit composition of photosystem I and identification of
center X as a [4Fe-4S]iron-sulfer cluster. *J.Biol.Chem.*
264, 6929-6934.

Shimizu, T., Hiyama, T., Ikeuchi, M., Koike, H. and Inoue,
Y. (1990) Nucleotide sequence of the *psa*C gene of the
cyanobacterium *Synechococcus vulcanus*. *Nucl.Acids Res.* **18**,
3644.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Haya-
shida, N., Matsubayashi, 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) The complete nucleotide
sequence of the tobacco chloroplast genome:its gene organ-
ization and expression. *EMBO J.* **5**, 2043-2049.

Weber, A.N. and Malkin, R. (1990) Photosystem I reaction-
centre proteins contan leucine zipper motifs(A proposed
role in dimer formation). *FEBS Lett.* **264**, 1-4.
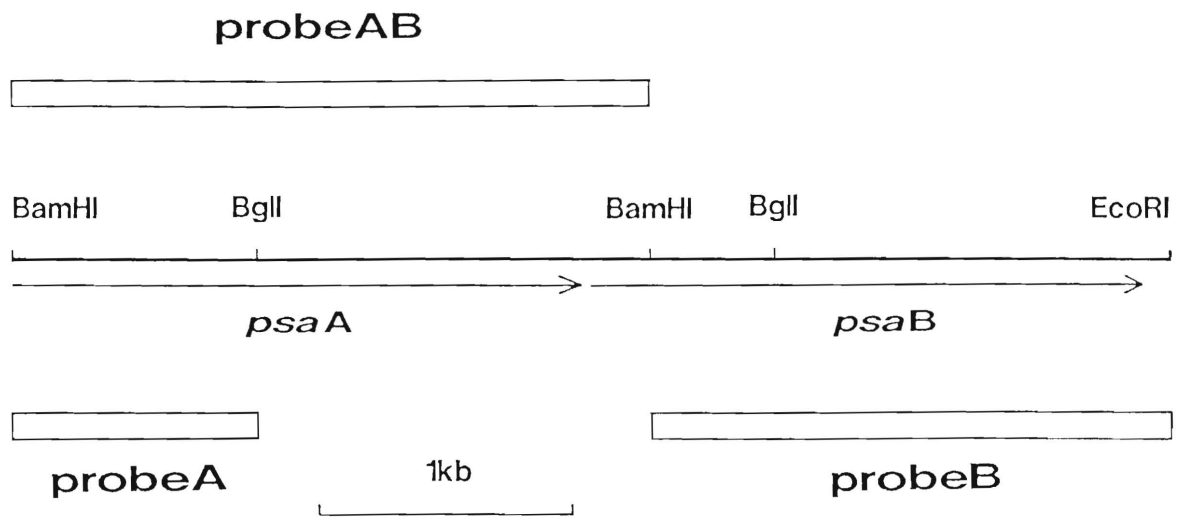
Figure 4-1.

A map of the tobacco *psa*A and *psa*B gene probes used to screen the *S.vulcanus* genomic DNA library. The three probes, probe-A, probe-B and probe-AB, include tobacco *psa*A, *psa*B and *psa*A-*psa*B regions, respectively.

4-15
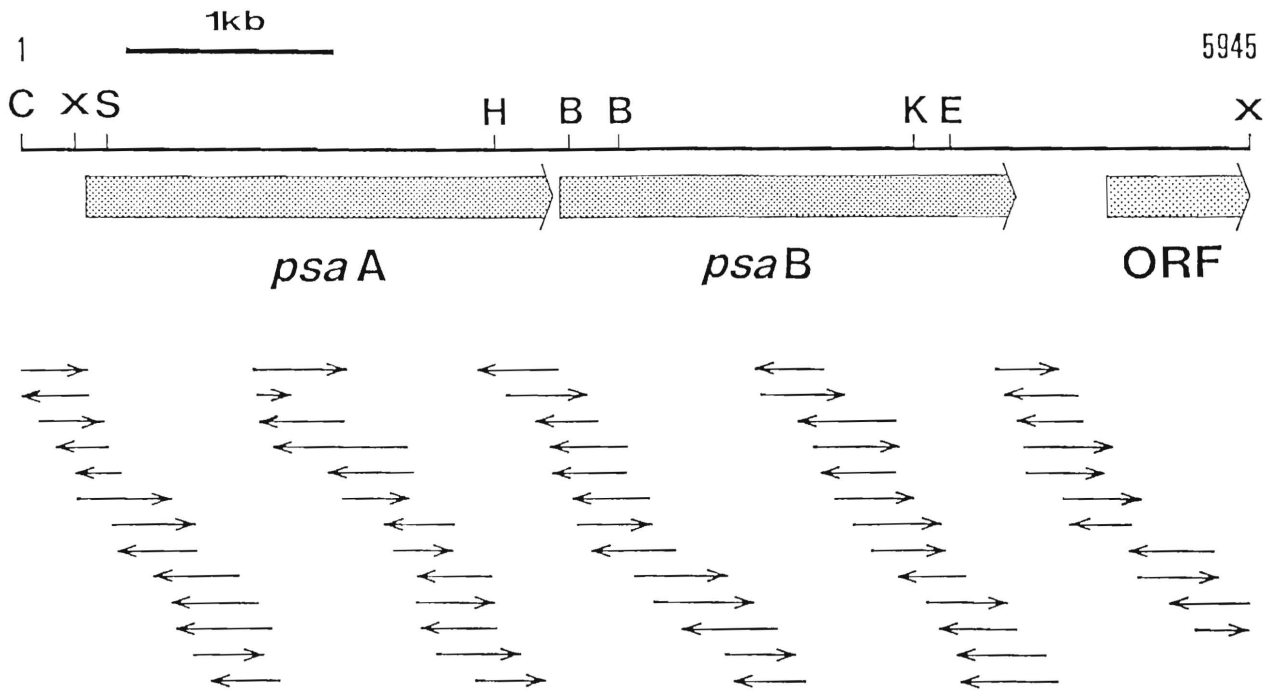
Figure 4-2.

A restriction endonuclease map, sequencing strategy and gene location of *psa*A-*psa*B operon and an unknown ORF. Dotted boxes represent the protein coding regions of *psa*A, *psa*B genes and unknown ORF. Restriction map is shown on the solid line: X, *Xba*I; C, *Cla*I; E, *Eco*RI; S, *Sma*I; H, *Hind*III; K, *Kpn*I; B, *Bam*HI. Arrows represent sequencing strategies.

CGATGCACATTAATTGCTAGCGAAAATGCAAGAGCAAAGACGAAAACAT   49

GCCACACATGAGGAATACCGATTCTCTCATTAACATATTCAGGCCAGTTATCTGGGCTTAAAAGCAGAAGTCCAACCCAGATAACGATCA   139

TATACATGGTTCTCTCCAGAGGTTCATTACTGAACACTCGTCCGAGAATAACGAGTGGATCTCTAGAGCTGTCGACTAATACGACTCACT   229

ATAGGGCGAGAATTCGGATCTAAAGATCTTTCTAGATTCTCGAGGTGAGTGCCCTCGTTCCTTCCTTGCCTAAGAGAGAGGAGAGACTCG   319
***psa*A**

ATGACCATCAGTCCACCGGAGCGAGAGCCAAAGGTCAGAGTCGTGGTTGATAATGACCCGGTGCCCACATCTTTTGAAAAATGGGCAAAA   409
M T I S P P E R E P K V R V V V D N D P V P T S F E K W A K

CCCGGGCATTTTGACCGCACGTTGGCCAGAGGACCCCAAACCACCACATGGATTTGGAACCTCCACGCTCTTGCCCACGATTTTGATACA   499
P G H F D R T L A R G P Q T T T W I W N T H A T A H D F D T

CACACGAGCGACCTTGAAGATATCTCCCGCAAAATCTTCAGTGCACACTTCGGCCATCTGGCTGTGGTGTTCATCTGGCTGAGTGGGATG   589
H T S D T E D I S R K I F S A H F G H T A V V F I W T S G M

TACTTCCACGGTGCAAAATCTCAAACTATGAGGCTTGGCTGGCCGATCCCACCGGTATCAAGCCCAGTGCTCAAGTGGTCTGGCCCATT   679
Y F H G A K F S N Y E A W T A D P T G I K P S A Q V V W P I

GTGGGTCAGGGCATTCTCAATGGTGATGTCGGCGGTGGTTTCCACGGCATCCAAATCACCTCGGGGCTATTCCAACTGTGGCGTGCCTCT   769
V G Q G I T N G D V G G G F H G I Q I T S G T F Q T W R A S

GGGATCACCAATGAGTTCCAGCTTTACTGCACCGCAATCGGTGGCTTGGTCATGGCTGGCTTAATGCTCTTTGCAGGCTGGTTCCACTAT   859
G I T N E F Q T Y C T A I G G L V M A G L M T F A G W F H Y

CACAAGCGCGCTCCTAAGCTGGAATGGTTCCAAAACGTTGGAATCCATGCTCAACCACCACCTTGCCGGTTTACTTGGCTTGGGATCTTTG   949
H K R A P K T E W F Q N V E S M T N H H T A G L T G L G S L

TCTTGGGCAGGTCACCAGATCCACGTCTCACTACCCATCAACAAACTTTTGGATGCAGGGGTCGCTGCTAAGGATATTCCCTTGCCCCAC   1039
S W A G H Q I H V S T P I N K T L D A G V A A K D I P L P H

GAGTTTATCCTTAACCCCAGCTTGATGGCCGAGCTATATCCCAAAGTGGATTGGGGTTTCTTTAGTGGTGTCATTCCCTTTTTTCACCTTT   1129
E F I T N P S L M A E T Y P K V D W G F F S G V I P F F T F

AATTGGGCTGCCTACTCGGATTTCCTCACCTTTAACGGTGGCTTGAATCCCGTTACCGGTGGCCTGTGGCTGTCGGATACGGCTCACCAC   1219
N W A A Y S D F T T F N G G L N P V T G G T W T S D T A H H

CATCTGGCGATCGCCGTCCTCTTCATCATTGCTGGTCATATGTACCGCACCAATTGGGGGATCGGCCACAGCCTAAAAGAAATCCTTGAA   1309
H T A I A V T F I I A G H M Y R T N W G I G H S T K E I T E

GCCCACAAAGGCCCCTTCACAGGTGCTGGCCATAAAGGTCTCTATGAAGTGCTGACCACCTCTTGGCATGCCCAACTGGCGATCAACCTT   1399
A H K G P F T G A G H K G T Y E V T T T S W H A Q T A I N T

GCCATGATGGGCTCCCTGAGCATTATTGTGGCACAGCACATGTATGCAATGCCCCCCTATCCCTACTTGGCCACCGACTATCCAACTCAA   1489
A M M G S T S I I V A Q H M Y A M P P Y P Y L A T D Y P T Q

CTGTCGCTGTTTACCCACCACATGTGGATTGGTGGCTTCCTGGTCGTGGGGGGTGCTGCCCATGGTGCCATCTTCATGGTGCGTGACTAC   1579
T S T F T H H M W I G G F T V V G G A A H G A I F M V R D Y

GATCCGGCCATGAATCAGAACAACGTTTTGGATCGGGTTCTGCGCCATCGTGATGCCATCATTTCTCACCTGAACTGGGTGTGCATCTTC   1669
D P A M N Q N N V L D R V T R H R D A I I S H T N W V C I F

TTGGGCTTCCACAGCTTCGGTCTGTACGTCCACAACGACACGATGCGGGCCTTTGGTCGTCCCCAAGATATGTTCTCCGATACGGGGATT   1759
L G F H S F G I Y V H N D T M R A F G R P Q D M F S D T G I

CAGCTTCAGCCCGTGTTTGCCCAATGGGTGCAAAATTTACATACCCTAGCCCCCGGTGGTACTGCTCCCAATGCAGCAGCGACGGCTAGT   1849
Q T Q P V F A Q W V Q N L H T T A P G G T A P N A A A T A S

GTCGCCTTTGGGGGTGATGTGGTTGCTGTCGGTGGCAAAGTGGCCATGATGCCCATTGTCTTGGGAACTGCCGACTTCATGGTGCATCAT   1939
V A F G G D V V A V G G K V A M M P I V L G T A D F M V H H

ATTCACGCTTTCACCATTCATGTGACAGTGCTGATTCTGCTGAAGGGCGTACTCTTTGCCCGCAGCTCTCGCCTGATTCCCGATAAAGCC   2029
I H A F T I H V T V T I T T K G V T F A R S S R T I P D K A

AACTTAGGCTTCCGCTTCCCCTGCGATGGTCCCGGTCGTGGCGGGCACCTGCCAAGTCTCCGGTTGGGATCACGTCTTCTTGGGTCTGTTC   2119
N L G F R F P C D G P G R G G T C Q V S G W D H V F L G I F

TGGATGTACAACTGCATCTCCGTTGTGATTTTCCACTTTAGTTGGAAGATGCAGTCGGATGTCTGGGGTACTGTTGCCCCCGATGGTACG   2209
W M Y N C I S V V I F H F S W K M Q S D V W G T V A P D G T

GTATCTCACATCACGGGCGGTAACTTTGCCCAAAGTGCCATCACCATCAATGGCTGGCTACGGGATTTCCTGTGGGCACAAGCTTCTCAG   2299
V S H I T G G N F A Q S A I T I N G W T R D F T W A Q A S Q

```
GTGATTGGCTCCTATGGTTCAGCCCTATCCGCCTATGGTTTGTTGTTCTTGGGTGCTCACTTCATTTGGGCCTTCAGCCTCATGTTCCTC    2389
 V  I  G  S  Y  G  S  A  T  S  A  Y  G  L  L  F  L  G  A  H  F  I  W  A  F  S  T  M  F  T

TTCAGTGGCCGTGGCTACTGGCAAGAGCTAATTGAGTCCATTGTTTGGGCCCACAACAAACTGAAAGTTGCACCGGCTATTCAGCCCCGT    2479
 F  S  G  R  G  Y  W  Q  E  T  I  E  S  I  V  W  A  H  N  K  T  K  V  A  P  A  I  Q  P  R

GCCTTGAGCATCATTCAAGGCCGTGCCGTCGGTGTAGCCCATTACCTCCTAGGGGGGATTGCCACTACCTGGGCATTCTTCCTAGCTCGG    2569
 A  L  S  I  I  Q  G  R  A  V  G  V  A  H  Y  T  T  G  G  I  A  T  T  W  A  F  F  T  A  R

                                                   *psaB*
ATTATTTCTGTAGGATAGGGGCGAGGAGGAGATAACCAACTATGGCAACTAAATTTCCGAAGTTTAGCCAAGACCTCGCACAGGATCCGA    2659
 I  I  S  V  G                         M  A  T  K  F  P  K  F  S  Q  D  T  A  Q  D  P  T

CCACACGCCGGATTTGGTACGGCCATCGCCATGGCCCATGACTTTGAAAGCCACGATGGCATGACTGAGGAGAATCTTTACCAAAAGATTT    2749
 T  R  R  I  W  Y  A  I  A  M  A  H  D  F  E  S  H  D  G  M  T  E  E  N  T  Y  Q  K  I  F

TTGCCTCCCACTTTGGGCATCTGGCCATCATCTTCCTGTGGGTGTCTGGTAGCCTATTCCACGTTGCATGGCAGGGGAACTTTGAGCAAT    2839
 A  S  H  F  G  H  T  A  I  I  F  T  W  V  S  G  S  T  F  H  V  A  W  Q  G  N  F  E  Q  W

GGGTTCAAGACCCTGTCAACACCCGTCCCATCGCCCATGCGATCTGGGATCCCCAATTTGGTAAAGCGGCCGTGGACGCCTTCACCCAAG    2929
 V  Q  D  P  V  N  T  R  P  I  A  H  A  I  W  D  P  Q  F  G  K  A  A  V  D  A  F  T  Q  A

CGGGGGGCTTCTAACCCTGTGGACATTGCCTACTCTGGTGTCTATCACTGGTGGTACACCATCGGTATGCGCACCAACGGCGACCTGTATC    3019
 G  A  S  N  P  V  D  I  A  Y  S  G  V  Y  H  W  W  Y  T  I  G  M  R  T  N  G  D  T  Y  Q

AAGGTGCCATCTTCCTGCTGATTCTGGCGTCGCTGGCTCTCTTTGCTGGCTGGCTGCACTTGCAACCCAAATTCCGTCCTAGCCTCTCTT    3109
 G  A  I  F  T  T  I  T  A  S  T  A  T  F  A  G  W  T  H  L  Q  P  K  F  R  P  S  T  S  W

GGTTTAAAAATGCTGAATCGCGGTTGAACCACCACTTGGCAGGTCTATTTGGGGTTAGCTCCTTAGCTTGGGCAGGCCACCTGATTCACG    3199
 F  K  N  A  E  S  R  L  N  H  H  L  A  G  T  F  G  V  S  S  L  A  W  A  G  H  T  I  H  V

TAGCCATTCCTGAGTCCCGTGGTCAGCACGTGGGCTGGGATAATTTCCTGAGCACCATGCCCCACCCTGCCGGTCTGGCACCTTTCTTTA    3289
 A  I  P  E  S  R  G  Q  H  V  G  W  D  N  F  T  S  T  M  P  H  P  A  G  T  A  P  F  F  T

CGGGGAACTGGGGCGTCTATGCCCAAAACCCTGACACGGCTAGCCACGTCTTTGGCACGGCACAGGGTGCTGGCACTGCGATTCTCACCT    3379
 G  N  W  G  V  Y  A  Q  N  P  D  T  A  S  H  V  F  G  T  A  Q  G  A  G  T  A  I  T  T  F

TCCTGGGTGGATTCCATCCCCAAACAGAGTCCCTGTGGCTCACGGATATGGCTCACCACCACCTTGCTATTGCTGTGCTCTTTATCGTGG    3469
 T  G  G  F  H  P  Q  T  E  S  T  W  T  T  D  M  A  H  H  H  T  A  I  A  V  T  F  I  V  A

CAGGTCACATGTACCGCACCCAGTTTGGGATTGGCCACAGTATCAAAGAGATGATGGATGCCAAGGATTTCTTTGGCACCAAGGTGGAAG    3559
 G  H  M  Y  R  T  Q  F  G  I  G  H  S  I  K  E  M  M  D  A  K  D  F  F  G  T  K  V  E  G

GTCCCTTCAACATGCCTCACCAAGGCATCTATGAAACCTACAACAACTCACTGCACTTCCAACTGGGCTGGCACTTGGCCTGCTTGGGCG    3649
 P  F  N  M  P  H  Q  G  I  Y  E  T  Y  N  N  S  T  H  F  Q  T  G  W  H  L  A  C  L  G  V

TGATCACTTCCTTGGTGGCACAACACATGTACTCGCTGCCACCCTATGCCTTCATTGCCCAAGACCATACCACGATGGCTGCCCTTTACA    3739
 I  T  S  L  V  A  Q  H  M  Y  S  T  P  P  Y  A  F  I  A  Q  D  H  T  T  M  A  A  T  Y  T

CCCATCACCAGTACATTGCTGGCTTCTTGATGGTGGGTGCCTTTGCCCATGGTGCTATCTTCTTGGTGCGTGACTACGATCCAGCCCAAA    3829
 H  H  Q  Y  I  A  G  F  L  M  V  G  A  F  A  H  G  A  I  F  L  V  R  D  Y  D  P  A  Q  N

ATAAAGGTAATGTGTTGGATCGGGTGCTGCAACACAAAGAGGCGATCATTTCCCACTTGAGCTGGGTGTCGCTCTTCTTGGGCTTCCACA    3919
 K  G  N  V  L  D  R  V  T  Q  H  K  E  A  I  I  S  H  L  S  W  V  S  T  F  L  G  F  H  T

CGTTGGGTCTCTATGTCCACAACGATGTGGTGGTGGCCTTTGGTACTCCTGAGAAGCAAATCCTGATTGAGCCGGTGTTTGCCCAGTTCA    4009
 L  G  T  Y  V  H  N  D  V  V  V  A  F  G  T  P  E  K  Q  I  T  I  E  P  V  F  A  Q  F  I

TTCAAGCAGCCCACGGAAAACTACTCTATGGGTTTGATACATTGCTGTCGAATCCCGATAGCATTGCCAGCACTGCTTGGCCGAACTATG    4099
 Q  A  A  H  G  K  T  T  Y  G  F  D  T  L  T  S  N  P  D  S  I  A  S  T  A  W  P  N  Y  G

GCAACGTCTGGCTACCCGGTTGGCTCGATGCCATCAACAGTGGCACGAACTCTCTGTTTTTGACAATTGGTCCTGGGGACTTCTTGGTGC    4189
 N  V  W  T  P  G  W  T  D  A  I  N  S  G  T  N  S  T  F  L  T  I  G  P  G  D  F  L  V  H

ACCACGCCATTGCCCTAGGTCTGCACACCACCACCCTGATTTTGGTCAAGGGTGCATTGGATGCCCGTGGCTCCAAACTGATGCCAGATA    4279
 H  A  I  A  T  G  T  H  T  T  T  T  I  L  V  K  G  A  L  D  A  R  G  S  K  T  M  P  D  K

AGAAGGACTTCGGCTATGCTTTCCCCTGCGATGGCCCTGGCCGCGGCGGTACCTGCGATATTTCCGCATGGGATGCCTTCTATCTGGCTA    4369
 K  D  F  G  Y  A  F  P  C  D  G  P  G  R  G  G  T  C  D  I  S  A  W  D  A  F  Y  T  A  M

TGTTCTGGATGCTGAACACCATTGGCTGGGTGACTTTCTACTGGCACTGGAAACACCTCGGTGTCTGGGAAGGCAACGTGGCGCAGTTCA    4459
 F  W  M  T  N  T  I  G  W  V  T  F  Y  W  H  W  K  H  T  G  V  W  E  G  N  V  A  Q  F  N

ATGAAAGCTCCACCTACCTCATGGGTTGGCTACGGGATTACCTGTGGTTGAATTCATCCCAGCTCATTAATGGCTACAACCCCCTTCGGCA    4549
 E  S  S  T  Y  T  M  G  W  T  R  D  Y  T  W  L  N  S  S  Q  T  I  N  G  Y  N  P  F  G  T
```

4-18

```
CCAATAACTTGTCAGTGTGGGCATGGATGTTCCTGTTTGGTCACCTTGTGTGGGCCACTGGCTTCATGTTCCTGATTAGCTGGCGGGGCT    4639
 N  N  L  S  V  W  A  W  M  F  T  F  G  H  T  V  W  A  T  G  F  M  F  T  I  S  W  R  G  Y

ACTGGCAAGAGCTAATTGAAACCCTGGTGTGGGCACACGAGCGGACTCCCTTGGCCAACTTGGTGCGCTGGAAAGACAAACCTGTGGCGC    4729
 W  Q  E  T  I  E  T  T  V  W  A  H  E  R  T  P  L  A  N  L  V  R  W  K  D  K  P  V  A  T

TGTCGATTGTTCAGGCCCGTTTGGTCGGTTTGGCACACTTCAGCGTTGGCTATATCTTGACCTATGCTGCCTTCCTAATTGCTTCAACGG    4819
 S  I  V  Q  A  R  L  V  G  L  A  H  F  S  V  G  Y  I  L  T  Y  A  A  F  T  I  A  S  T  A

CAGCCAAGTTCGGTTGATCGACTTGCGTTAAGTTCTCTTGAGTCCCCCGCCCCCTGGTGGGGGATTTTTTTTCCGATGGCAAAACTGCCT    4909
 A  K  F  G                            ------------>  <------------

GACTAATTAGGTATGCAAATCCCTTGAGAGACGATATAAGTGGTTGATTCACAAACAACTCTCCTATCGTGACCTTTAGGGTGTGGCAAC    4999
                              ------------>  <------------

AGGCGATAAATCTGTTGATCGGATGCGCAACGAAACTGCTTGGGAGACGATTAAATAGGCTGTCACTTTCGCAAGATCAATACAGTTGAC    5089

TGTAATTTTGTATCCCTTTGTATCTCTTTGTTTAGAAGGGGTTTAGAAGGTTGCTGTGTTGCCCTAAAGGTTATGAAC⌐TTGCAA⌐CCCTC    5179

TCAATACTTCT⌐AAGAAC⌐TGATAACCTATTTTTAACACGCTCACTTTATTGCTGAGCTATGATAGGACTGAATTGTTCTGTCTGGAGAAAG    5269
           ORF

TAGTAATGCAGGATAGATTCATATTCTCTATTTCTACATTTCCGTACAAAAACTATATTGAACAGTTATTAGCACAAGCAGATGTTGGCG    5359
    M  Q  D  R  F  M  F  S  I  S  T  F  P  Y  K  N  Y  I  E  Q  L  L  A  Q  A  D  V  G  E

AATTTCCCCAGGTTGTGGGCGAAAATTCAGCCAGTTTCACCTTGGGTCTCAATAATTTGGGTGCACTTTTAGTGCAGGTGGTGATTGCAG    5449
 F  P  Q  V  V  G  E  N  S  A  S  F  T  L  G  T  N  N  L  G  A  T  L  V  Q  V  V  I  A  V

TTGCGTTGTTGTTCGTTGGCTGGATTGTTGCAACTGTTTTAGCAAGGGTCACACAATCAATTTTGAAACGGGTGCGCCTCGATGAATGGT    5539
 A  L  L  F  V  G  W  I  V  A  T  V  L  A  R  V  T  Q  S  I  L  K  R  V  R  T  D  E  W  L

TGACCCAGTTTTTGGGGGGGGAATGAATCACTACAAGGCCTATCACCAACAGCAATTTTGTCAGGAGCTGTTTTTTGGATTATCTTTCTGC    5629
 T  Q  F  L  G  G  N  E  S  T  Q  G  T  S  P  T  A  I  L  S  G  A  V  F  W  I  I  F  T  T

TGGGAGTGGTGGCCTTCTTGGATGCCCTGCGACTCACAACCGTTTCCCAGCCCCTCAATGCTTTTCTGAATCAAATTTTTAGTTTCCTGC    5719
 G  V  V  A  F  L  D  A  T  R  T  T  T  V  S  Q  P  T  N  A  F  T  N  Q  I  F  S  F  T  P

CAAAACTAGGGGTAGCTATTCTGCTTTTGGATCTGGCTTGGGTCGTAGCAACGATATCAAAAATGCTCGTGACTCAGTCTGCCCGCTCTC    5809
 K  T  G  V  A  I  T  T  L  D  T  A  W  V  V  A  T  M  S  K  M  T  V  T  Q  S  A  R  S  T

TCAATTTGGATCGCTCACTACCGCTGGAATCTCGCGAAGAGGGCGCTCCCACAATGTCTGTGGCTGAGATGTTGGGAAATACCCTCTACT    5899
 N  L  D  R  S  T  P  T  E  S  R  E  E  G  A  P  T  M  S  V  A  E  M  L  G  N  T  T  Y  W

GGTTTGTGTTCCTCTTCTTCCTGCCCTTGATTTTAGGGGTGCTGAA    5945
 F  V  F  T  F  F  T  P  L  I  L  G  V  T
```

Figure 4-3.

Nucleotide and deduced amino acid sequences of the *Synechococcus vulcanus* *psa*A–*psa*B gene and unknown ORF. The positions of the each genes are: 320–2584 (*psa*A), 2611–4833 (*psa*B) and 5275–5945 (unknown ORF). The boxed sequences at the preceding region of the *psa*A gene indicate the possible promoter sequences (P1, P2 and P3 : see the text for detail). Another boxed region at 5,168–5,173 and 5,191–5,196 indicates possible –35 and –10 sequences for the interrupted ORF. The underlined regions represent expected ribosome binding site (Shine-Dalgarno sequence). The arrows represent potential terminators of *psa*A–*psa*B operon.

4-19

psa A

```
a: MTI-S-PPER-----EPKVRVVVDNDPVPTSFEKWAKPGHFDRTLARG-PQTTTWIWNI HAI ANDFDTHTSDLEDISRKIFSAHFGHLAVVFIWLSGHYFHGAKFSNYEAWLADPTGIK
b:             MA·K---·K·······G···········K·········SQ···············V····················NN··V··
c: ·I·R········-E·KIL·R··K··E·R····S·I·K···D·········D·····S······E···V·····Q·SII·L·········R········S···II·G
d: ···R········-E·KI·EK··K·······S···K··S·······D·····S·N···E···V·····Q··II··········R········S···II··
e: ·M·R········-E·KI··R··K··E·R····S··K···D·········D··S·G···E···V·····Q·SII·L·········R········S···II·G
f: ·I·R·S·······-E·KIA·R··IK··E·R····S·I·K·N·D·······D·····S·G···E···V·····Q·SII·L·········R········S···II·G
g: ·I·R········-E·KIL·R··K··A·······S·I·K··E·······D·····S······E·········Q·SII·L·········R········S···II·G
h: ·····T······-EAK··KIA·RN··E·······S··SK··N·······D·····S······E···V·····Q·GII·········R········S···II·
i: ·I·R····-PKVQILAD·E·KIL·R··IK····Q·······S·I·K···D·····D·····S······E···V·····Q·SII·L·········R········N···II·R
j: ·····T····QQV····-KR··AFVSN··E······SR····S·L·SK·PNT··········D···N··T·····V·····Q··IIE········F···R········L··IIN·
```

```
                                                                                                                    240
a: PSAQVVWPIVGQGILNGDVGGGFHGIQITSGLFQLWRASGITNEFQLYCTAIGGLVMAGLMLFAGWFHYIIKRAPKLEWFQNVESMLNHHLAGLLGLGSLSWAGHQIHVSLPTNKLLDAGV
b: ·········V··········S········Y··A·F··SY··V········A··V········A·····A·A·M···SV··C··G·T·L·····V·····S··
c: ············E·········R·····F·I····S·L·····A·F·A········A··A··D···········V····QF·N···
d: ············E·········Q·········F·········S·L··S······F·A········A··A··D········V····Q····
e: ············E·········R··S·F·I····S·L·····A·IF·S········A··A·SHD······QF····
f: ············E·········R·····F·I····S·L·····A·IF·S········A··A·D········QF····
g: ············E·········R·····F·I····S·L·····A·F·A········A··A··D···········QF·N···
h: ············E·········Q·········R·······S·L··T······AA·F········G·······A······V····
i: ············E·········R·····F·I····S·L·····A·F·A········A··A··D········V····QF·N···
j: ····I········E·········N·Q·········SC···S····I·LT····IFSAVLF········A·········S······C·········S·
```

```
                                                                                                                    360
a: AAKDIPLPHEFILNPSLMAELYPKVDWGFFSGVIPFFTFNWAAYSDFLTFNGGLNPVTGGLWLSDTAHHHLATAVLFIIAGHMYRTNWGIGHSLKETLEAHKGPFTGAGHKG-------
b: ·············-FDA·V·········-----S·AQ·LK···LD·········K·····T·S····V················GF·····M·········E·······----
c: DP·E··········RD·L·Q·----S·AE·AT····L·SK·A·····R···D········T·I········I·L·········G··D·········Q········----
d: DP·E··········RD·L·-----S·AK·LT····L·SE·········I············LV·········F·········Q········----
e: DP·E··········RD·L·Q·----S·AERAT····L·SK·AE··S·R··D·I········I········I·L·········G··D·········QR·······----
f: DP·E····L···RD·L·Q·----S·AE·AT····L·SK·AE··S·R··D·I········I········I·L·········G··D·········Q·········----
g: DP·E··········RD·L·Q·----S·AE·AT····L·SK·A·····R···D········V········LV·········M·········R···E·V······----
h: DP·E······DLL··RAI·D·------S·AK·IA····L·SE·········K·········V········LV·········M·········R···E·V······----
i: D-------------L·Q·----S·AE·AT····L·SK·A·····R··D·L········T·I········I·L·········GI·D·········Q·····PFTGQGHK
j: NPAEL·····D···DK··ISQ·----S·SK·LA····II·SE·······R····N········T·V·····L·····L·····K···K···DI·GL··S·T·····Q········
```

```
                                                                                                                    480
a: ---------LYEVLTTSWHAQLATNLAHMGSLSIIVAQHMYAMPPYPYLATDYPTQLSLFTHHMWIGGFLVVGGAAHGAIFMVRDYDPAHNQNNVLDRVLRHRDAITSHLNWVCIFLGF
b: ---------I········V·LL···T·VI·H··S·····M··G·········T·····AG········K·V··L···I·······
c: ---------I········SL··L···T·V·H··S·······G·················I·A···A·······TTRY·DL·········A·····
d: ---------I·········L···L···I··H··········G·················I·A···A·······TIQY·L·········
e: ---------I········SL··L··TT·V·H··S·······G·················I·A···A·······TTRY·DL·········
f: ---------I········SL··L··TT·V··H··S·······G·················I·A···A·······TTRY·DL·········
g: ---------I·········L···L··T·V·H··········G·················I·A···A·······TTRY·DL·········A·····
h: ---------I··········LF·······II··········G·········T·····CI··AG··A·······TN·Y··L···I········
i: GPFTGQGHKG···I·········S····L···T·A·H··I···I········G·················I·A···A·······TTRY·DL·········
j: -----------IF·N····SL·············S····I·I··G·E·······Y···*
```

```
                                                                                                                    600
a: HSFGLYVHNDTMRAFGRPQDMFSDTGIQLQPVFAQWVQNIIITLAPGGTAPNAAATASVA-FGGDVVAVGGKVAHMPIVLGTADFMVHHHHHAFTHHVTVLILLKGVLFARSSRLIPDKANL
b: ·····I·····L·········SA····I····I··I·········I·S··QV·····I·N···L··T··············L·YS·····V··GQ·
c: ·····I····S·L·········A·········I··T·A···A··G·T·ST··LTWG··L·······LL··P····L·············A·········I·····
d: ·····I····S·L·········A·········I··T·A···NF····L·SL·LTWG····I···S···LL··P·····L··········
e: ·····I····S·L·········A····I·······AG··RL···G·TTST··LTWG··EL·········LL··P····L··········
f: ·····I····S·L·········AA····I····I··I·AG···V···G·TTST··LTWG··EL··I·····LL··P····L··········
g: ·····I····S·L·········A·········I··T·A···SA··G·T·ST··LTWG·S·L·········LL··P····L··········
h: ·····I····S·L·········A·········I··T·F···QL····L·AT··LTWG··ELG·II·········S··S·········
i: ·····I····S·L·········A·········I··T·A····T··G·TTST··LIWG··L·S·······LL··P····L··········
```

```
                                                                                                                    720
a: GFRFPCDGPGRGGTCQVSGWDHVFLGLFWMYNCISVVIFHFSWKMQSDVWGTVAPDGTV-SHTTGGNFAQSATTINGWLRDFLWAQASQVIGSYGSALSAYGILFLGAHFIWAFSLMFLF
b: ················SL·I········IL··S···V····T·········S···N··········IM··AG··VF·······
c: ·········A··········A··············S··S·QG·VT··········Q···S·······F·····V······
d: ··········S···············I·SEQG·VT··········Q···S·······F·····V······
e: ·········A··········A··········I·S·QG·VT··········Q···S·······F·····V······
f: ·········A··········S·········I·S·QGIVI··········S·······Q···S·······F·····V······
g: ·········A··········S··········SI·S·QG·VT··········S·······Q···S·······F·····V······
h: ················SL·I········TASG········N··········S···Q·········I·····V······
i: ·········A·V········A············SIN··QG·VT·····R····S·········Q···S·······F·····V······
```

		784		Residues	Homology (%)
a:	SGRGYWQELIESIVWAHNKLKVAPAIQPRALSIIQGRAVGVAHYLLGGIATTWAFFLARIISVG		*S. vulcanus*	755	100
b:	·························L·········V·················V·········SL·I·		*S.*sp PCC7002	739	84
c:	···············T················T·················A··		tobacco	750	82
d:	·······················T·························A··		liverwort	750	85
e:	··············T················T·················A··		rice	751	82
f:	·············T················T···········A··		maize	752	82
g:	··············T·······V········T·················A··		spinach	750	83
h:	···············T················T·········S···········		Chlamydomonas reinhardii	751	84
i:	··············T·······V········T·················A··		pea	761	84
			j) Euglena gacilis		

4-20

psa B

```
                                                                                                        . 120
a: MATK-FPKFSQDLAQDPTTRRIWYAIAMAIIDFESHDGMTEENLYQKIFASIIFGHLAIIFLWVSGSLFIIVAWQGNFEQWVQDPVNTRPIAIIATWDPQFGKAAVDAFTQAGASNPVDIAYSG
b: ·················G·T····T·····················T·T·········IK··L·V·········II··QG·I···········N·····
c: ··LR··R···G·········FG·I·····DI··R··N·········Q·····T·N·········S····LIIV·········II··QP·E···RG··LG··N·····
d: ··SR·······G·S·······FG·T·········D···R········Q·····T·N·········A·G···LIIV·········II··QP·E···RG··LG··N·····
e: ·ELR··R···G·········FG·I·····DI··R··N·········Q·····T·N·········S·I····LIIV·········II··QP·E···RG··AG··N·····
f: ·ELR··R···G·········FG·T·····DI··R··N·········Q·····T·N·········A·I····LIIV·········II··QP·E···RG··AG··N·····
g: ··LR··R···G·········FG·T·····DI··R··N·········Q·····T·N·········S····LIIV·········II··QP·E···RG··LG··N·····
h: ····L·····G·R·A·I···GL·······VW·········IS··Q·S·····T·N·········T··HI··C·········II··QP·E···RG··G·N·ST·
i: ··LRI··R···GI········FG·T·········DI··GR··N·········Q·····T·N·········A·····FIIV·········H··QP·E···RG··LG··NN···
```

```
                                                                                                        . 240
a: VYIMWYTIGMRTNGDLYQGAIFLLILASLALFAGWLIILQPKFRPSLSWFKNAESRLNIIILAGLFGVSSLAWAGHILIIVAIPE-SRGQHVGWDNFLSTMPHPAGLAPFFTGNWGVYAQNPD
b: ····F·····T·A·····V·····S·F·········A··D·········T··V·········V········VK·····M······
c: ··Q·····L···E··T··L···F·SAIS·I·········WK··V·········S·········T··V·····A·····EY·R·N··DVL··Q··G·L···Q·NL··
d: ··Q·····L···Q··N·L·V··S·IS·I·········WK·KV·········S·········T··V·····E··R·····TKL··E··G···A·Q·NI····V·
e: ··Q·····L···E··T··L···F·ST·S·IG·······WK·········S·········T··V·····A·····EY·R·N··DVL·Y·Q··G·LL···Q·NL··
f: ··Q·····L···E··T··L···F·ST·S·IG·······WK·········S·········T··V·····GS·····EY·R·N··DVL·Y·Q··G·LL···Q·NL··
g: ··Q·····L···E··T··L···F·SVIS·LG·······WK··V·········S·········T·········G·····EY·R·N··DVL····Q··G·L···Q·NL··
h: ··Q·······Q··V·SV··ALVSA1F·········N·Q·····D·········S·········T··V·····LQ·········VL··Q·T·······AA···S··
i: ··Q·····L···E··T······F·SFIS·L·······WK··V·········S·········V·········G·····EY·R·N··DVL·Y·Q··G·LL···Q·NL···
```

```
                                                                                                        . 360
a: TAS-HVFGTAQGAGTAILTFLGGFHPQTESLWLTDMAIIIIILAIAVLFIVAGIMYRTQFGIGHSIKEMMDAKDFFGTKVEGPFNMPHQGIYETYNNSLHFQLGWIILACLGVITSLVAQIMY
b: ····G·····S·······················I·········I······NVR·······ILN·HQ····PPS·KLGAG·K·L·D·V·········LA··S·········
c: SS···L·········L·······Q·····I······FI·L·········N····M·DLL··HI····PPG·RLGRG·K·L·D·I·········LA··S·········
d: -S·N·A···S·········I·······Q·····I······V·I·······N············ILETHIT····PPG·RLGRG·K·L·D·I·········LA··S·········
e: -S·N·L···T···········L·······Q·····I······FI·LI········N············DLLE·HIT····PPG·RLGRG·K·L·D·I····I·····LA··S·········
f: -S·N·L···T···········L·······Q·····I······FI·LI········N············DLLE·HIT····PPG·RLGRG·K·L·D·I····I·····LA··S·········
g: SS···S·········L·······Q·····I······FV·L········N····M·DLLE·HI····PPG·RLGRG·K·L·D·I·········LA··S·········
h: ··········S·Q········Q·······I·······N··N··RMQAILE·HII····PPS·SLGAG·K·LFD·V········I·LA··SV·T·········
i: SS·N·L···T···········I·······Q·····V······F··LIG·L····N··········YILE·III····PPG·RLGRG·K·L·D·I····I·····LA··S·········
```

```
                                                                                                        . 480
a: SLPPYAFIAQDIITTMAALYTHIIQYIAGFLMVGAFAIIGAIFLVRDYDPAQNKGNVLDRVLQHKEAIISIILSWVSLFLGFHTLGLYVIINDVVAFGTPEKQILIEPVFAQFIQAAHGKLLYG
b: ···S······K·Y··Q·············F·····EA·D···Y·M·E···L·····················V··S·A··
c: ···A·········F··Q·········I·T······FI··N·E··ED··A·M·E·········A·········ML·········I··W··S····TS··
d: ·······L···F··Q·········I·T······FI··N·E··D···A·M·E·········A·········ML·········I··W··S····A··
e: ···S·········F··Q·········I·T······FI··N·E··ED··A·M·D·········A·········ML·········I··W··S····TT··
f: ···A·········F··Q·········I·T······FI··N·E··ED··A·M·D·········A·········P·········I··W··S····TT··
g: ···A·········F··Q·········I·T······FI··N·E··ED··A·M·D·········A·········ML·········I··W··S····TS··
h: ·······Q·I·F··Q·········I·C······FI··E········A·M·D···L·········MQ·········W········A··
i: ···A·········F··Q·········I·T······P·FI··N·E··AD···A·M·E·········A·········ML·········I··W··S····TT··
```

```
                                                                                                        . 600
a: FDTLLSNPDSI-ASTAWPNYGNVWLPGWLDAINSGTNSLFLTIGPGDFLVIIHAIALGLIITTTLILVKGALDARGSKLMPDKKDFGYAFPCDGPGRGGTCDISAWDAFYLAMFWMLNTIGW
b: ··V····A·····N·SAA·······S·····N···························S············L··
c: ··V··S·T·GP·FN····GRSI·····N·V·ENS··············S··········V·····
e: ··V··S·TNNP·FN····GQSI······NNS··············E··S··········V·····
f: ··I··S·T·GP·FN····GRTL·····N·V·ENS··············S··········V·····
g: ··I··S·TNGPTFN····GR·I······C·········G···············S··········V·····
h: ··V··S·T·GP·FN····GRSI·····N·V·ENS··············S··········V·····
i: ··F···SKT·AAFAN····GQSL······NNQ·····L···················S·········Y···V·····
j: ··IP··S·TNGP·LN····GR·I······N··ENS··············S············D··V·····
```

```
                                                                                                        . 720
a: VTFYWIIWKHILGVWEGNVAQFNESSTYLMGWLRDYLWLNSSQLINGYNPFGTNNLSVWAWMFLFGIILVWATGFMFLISWRGYWQELIETLVWAHERTPLANLVRWKDKPVALSIVQARLVG
b: L········I·Q·······K······F····A·A·······Y·V···········I···········I·······
c: ·········ITL·Q···S······················M·S················A·········I·R······
d: ·········ITL·Q···A······················M·S················A·············
e: ·········ITL·Q···S······················M·S················A·········I·R······
f: ·········ITL·Q···S······················M·S··············?····A·········I·R·····V····
g: ·········ITL·Q···S······················M·S················A·········I·R······
h: ·········IL·Q···D······················M·S···T······IY·············K····Y······
i: ·········ITL·R···S·····················IT·LVC·S··········A·········I·R······
```

```
         .     745              Residues    Homology(%)
a: LAHFSVGYILTYAAFLIASTAAKFG    S. vulcanus      743       100
b: ····T··V··········SG···     S.sp PCC7002      733        86
c: ·········F··········SG···    tobacco          734        80
d: ·········F··········SG···    liverwort        734        81
e: ·········F··········SG···    rice             735        79
f: ·········F··········SG···    maize            735        78
g: ·········F··········SG···    spinach          734        80
h: ·········F··········SGR··    Chlamydomonas reinhardii  736  81
i: ·V·······F··········SG···    pea              734        78
```

4-21

Figure 4-4.

Alignment of the *Synechococcus vulcanus psa*A and *psa*B protein with other organisms derived protein sequences. The alphabets at the top of the lines indicates: (a)*Synechococcus vulcanus* (this study) (b)*Synechococcus* sp. PCC 7002 (Cantrell and Bryant,1987) (c)tobacco (Shinozaki *et al.*1986) (d)Liverwort (Ohyama *et al.* 1986) (e)Rice (Hiratsuka *et al.*1989) (f)Maize (Fish *et al.* 1985) (g)Spinach (Kirsch *et al.*1986) (h) *Chlamydomonas reinhardii* (Kück *et al.*1987) (i)Pea (Lehmbeck *et al.*1986) (j) *Euglena glacilis* (Manzara *et al.*1987). The conserved amino acid is indicated(.) an the skipped amino acids,(-). The termination position of *Euglena glacilis* (Manzara *et al.* 1987) *psa*A protein is indicated with asterisk. The underlined regions indicates the deduced membrane spanning sections.

4-22

Figure 4-5.

Nucleotide codon usage of the *psa*A and the *psa*B genes.

1	2	T	2	C	2	A	2	G	3
T	Phe	16/21	Ser	10/5	Tyr	9/12	Cys	0/0	T
	Phe	36/32	Ser	9/10	Tyr	10/15	Cys	5/3	C
	Leu	4/1	Ser	3/4	STOP	0/0	STOP	0/1	A
	Leu	17/24	Ser	5/6	STOP	1/0	Trp	28/33	G
C	Leu	10/4	Pro	1/12	His	13/7	Arg	8/6	T
	Leu	11/13	Pro	27/12	His	29/35	Arg	8/5	C
	Leu	10/8	Pro	3/3	Gln	15/19	Arg	1/0	A
	Leu	24/30	Pro	4/4	Gln	10/10	Arg	4/6	G
A	Ile	26/25	Thr	5/9	Asn	10/10	Ser	9/2	T
	Ile	25/16	Thr	20/21	Asn	16/20	Ser	8/13	C
	Ile	0/0	Thr	5/4	Lys	13/12	Arg	2/0	A
	Met	22/19	Thr	8/8	Lys	7/9	Arg	0/0	G
G	Val	9/5	Ala	21/18	Asp	23/19	Gly	36/27	T
	Val	18/9	Ala	35/35	Asp	7/12	Gly	27/29	C
	Val	4/1	Ala	14/16	Glu	7/7	Gly	4/2	A
	Val	22/26	Ala	3/8	Glu	8/11	Gly	11/8	G

*psa*A/*psa*B

(A)

```
                C-C
               C    T
                C=G
                C=G
               G    T
                C=G
                C=G
                C=G
                C=G
                C=G
   5'          T=A                        3'
  -TCTCTTGAG┘ └TTTTTTTTTCCGATGGCAAAACTGCCTG-
        4861    4884
```

(B)

```
                   A
              C       C
             T         A
             T         A
             A         A
              G=C
              T=A
              T=A
              G=C┐
             G    ┤T
             T    ┘
              G=C┘
              A=T
             A    C
             T    C
              A=T
              T=A
              A=T
              G=C
              C=G
   5'         A=T                                          3'
  -TCCCTTGAGAG┘ └GACCTTTAGGGTGTGGCAACAGGCGATAAATCTGTTGATCGGATGC-
          4939    4980
```
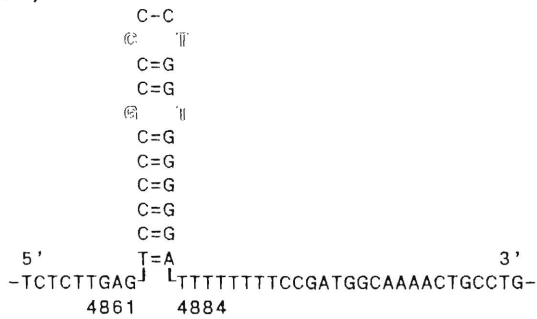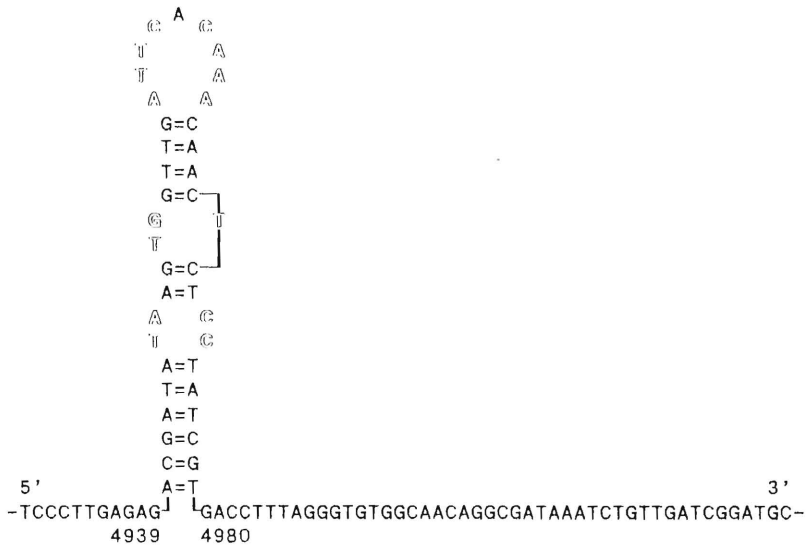
Figure 4-6.

Structure of the terminators (T1 and T2) for the *psa*A–*psa*B operon. (A) A structure of the supposed *rho*-independent terminator T1 (from positions 4,861 to 4,884).(B) A structure of the supposed *rho*-dependent terminater T2 (from positions 4,939 to 4,980). An equal mark (=) means a complementary base pair (see the text for the detail).

Table 4-1.

The amino acid sequence homology of the *psa*A and *psa*B proteins and the number of amino acid residues.

organisms	*psa*A (residues)	*psa*B (residues)	ref.
S.vulcanus	100.0 (755)	100.0 (743)	this study
S. sp. PCC7002	84.0 (739)	86.0 (733)	Cantrell *et al.*(1987)
Tobacco	82.4 (750)	79.8 (734)	Shinozaki *et al.*(1986)
Liverwort	85.4 (750)	80.8 (734)	Ohyama *et al.*(1986)
Rice	81.6 (751)	78.9 (735)	Hiratsuka *et al.*(1989)
Maize	81.8 (752)	78.0 (735)	Fish *et al.*(1985)
Spinach	82.9 (750)	79.5 (734)	Kirsch *et al.*(1986)
Chlamydomonas	84.3 (751)	80.5 (736)	Kück *et al.*(1987)
Pea	78.6 (761)	77.5 (734)	Lehmbeck *et al.*(1986)

Table 4-2.

Deduced amino acid composition, expected molecular
weight and G-C content of the *psa*A and *psa*B genes.

Amino acid	*psa*A	*psa*B
Gly	78	66
Ala	73	77
Val	53	41
Leu	76	80
Ile	51	41
Met	22	19
Phe	52	53
Trp	28	33
Pro	35	31
Ser	44	40
Thr	38	42
Asn	26	30
Gln	25	29
Cys	5	3
Asp	30	31
Glu	15	18
Lys	20	21
His	42	42
Arg	23	17
Tyr	19	27
Mw(Da)	83,198	83,043
G-C(%)	53.62	53.86

CHAPTER FIVE

Comparison of the nucleotide sequences
at 5'-proximal region of the genes of
Synechococcus vulcanus

Abstract

The nucleotide sequences of the genes of ther-
mophilic cyanobacterium *Synechococcus vulcanus*
were compared each other for the 5'-proximal
regions including the promoter. In addition to
the probable -35 and -10 sequences, several
conserved characteristic nucleotide sequences
were recognized each other. One of these se-
quences (TTCTCTCCAGA : box-E) located within
promoters of *psa*A-*psa*B and *psa*C might control the
expression of the photosystem I genes.

Comparison of the promoter sequences

The *cis*-acting elements which located upstream
of the start codon of particular gene(s) would
affect the efficiency of the gene expression.
These include the Shine-Dalgarno sequence (Shine
and Dalgarno, 1974) for ribosome binding and the
-35 and -10 sites for the initiation of the tran-
scription (promoter). In *E.coli*, consensus
sequence for the promoter are well character-
ized: TTGACA for the -35 site and TATAAT for the
-10 site (also called the **Pribnow box**), respec-
tively (Lewin, 1987). Any kinds of mutation
within a promoter should affect the level of
expression of the genes. It should be remembered,
however, a very few promoters lack a recognizable
sequences. They are often recognized by auxiliary
protein(s), *trans*-acting elements, in addition to
the RNA polymerase for the expression of the
gene. Some other conserved sequences could

5-2

occasionally be recognized around the promoter site by many investigators. It is presumed that the similar consensus sequence might be recognized around the promoter site for cyanobacteria. Since these putative sequences are often different from strain to strain, discreet determination is required for the analysis.

In thermophilic cyanobacterium *Synechococcus vulcanus*, we recognized possible promoter sequences of genes of *S. vulcanus* based on the similarities to the consensus sequence of the *E.coli* promoter. There were three candidates for the promoter to the *psa*A-*psa*B operon, namely the P1, P2 and P3 (Figure 5-1). The nucleotide sequence which exhibited significant similarities to the consensus sequence of the *E,coli* promoter were also found for *psa*C, *mtn*A and interrupted ORF at the downstream of the *psa*B (Figure 5-1).

The gap between the -35 regions and the -10 regions of possible promoters P1, P2 and P3 of *psa*A-*psa*B were 17-, 17- and 20-bp long, respectively. Although the P1 is most probable promoter to the *psa*A-*psa*B operon, the nucleotide se-

quence at the gap between the -35 site and the -10 site of the P1 and the P3 resemble each other (Figure 5-1). Since the *psa*A-*psa*B operon is expressed very strongly in the cells, this observation suggest that tandem promoter sequences would work for the efficient expression of the genes. These deduced promoter sequences are not highly homologous to the consensus sequence of *E.coli* in contrast to the previously reported genes for the photosynthesis of other cyanobacteria (Borthakur and Haselkorn, 1989).

The nucleotide sequence at the gap between the -35 region and the -10 region of the P2 looks like the promoter sequence of the *psa*C gene. We could not find significant similarities to the promoter sequences of the *psa*A-*psa*B genes of *Synechococcus* sp. PCC7002 (Cantrell and Bryant, 1987), and the *psb*A and *wox*A genes of the *Anabaena* sp. PCC7120. There are numerous reports of the isolation and determination of the photosynthetic genes from cyanobacteria, but the description of promoter sequences are found in only a few report, which makes it difficult to compare

5-4

these sequences among cyanobacteria. The precise judgment of the promoter structure by the S1-mapping are required.

The nucleotide sequences found at the upstream region of the genes

The cautious observation of the nucleotide sequences of *psa*A-*psa*B, *psa*C and *mtn*A genes of *S.vulcanus* have enabled us to find out the several conserved sequences for the upstream region of the genes. The three nucleotide sequences which are denoted AAAANNTG (box-A), GGCCA (box-B) and GGGC (box-C), are recognized in common at the upstream region of these three genes (Figure 5-2). A box-A and a box-C are arrayed after three box-B sequences before the promoter of the *mtn*A gene. The box-B, box-C and promoter of the *psa*C and *mtn*A genes are arranged alike with the distances of 7, 107 and 52bp (*psa*C) and 19, 107 and 52bp (mtn*A*), respectively. We could also recognize another sequence of GAATAA (box-D) at

the 5'-upstream region between *psa*A-*psa*B operon and *psa*C gene.

Moreover, the unique sequence TTCTCTCCAGA (box-E) are located within the promoter of *psa*C gene and P2 of *psa*A-*psa*B operon(Figure 5-2). A part of the box-E sequence TTCTCTC (box-E') is also found at the P1 site of the *psa*A-*psa*B operon. These box-D, box-E and box-E' structures could not be found at the 5'-upstream region of *mtn*A gene. These nucleotide sequences (from box-A to box-E') which could not be found at the upstream region of the interrupted ORF after the *psa*B gene possibly regulate the expression of these genes.

The box-E and box-E' structures found at only *psa*A-*psa*B operon and *psa*C probably are the *cis*-acting element specific for the photosystem I. Much more research is needed to determine the roles of these structures through a comparison of the nucleotide sequences to other genes of photosystem I, gel retardation assay and *in vitro* transcription analysis.

REFERENCES

Borthakur, D. and Haselkorn, R. (1989) Nucleotide sequence of the gene encoding the 33kDa water oxidizing polypeptide in *Anabaena* sp. strain PCC7120 and its expression in *Escherichia coli*. *Plant Mol.Biol.* 13, 427-439.

Cantrell, A. and Bryant, D.A. (1987) Molecular cloning and nucleotide sequence of the *psa*A and *psa*B genes of the cyanobacterium *Synechococcus* sp. PCC7002. *Plant Mol.Biol.* 9, 453-468.

Lewin, B. (1987) *Genes*(third edition) John Wiley & Sons, New York.

Shimizu, T., Hiyama, T., Ikeuchi, M., Koike, H. and Inoue, Y. (1990) Nucleotide sequence of the *psa*C gene of the cyanobacterium *Synechococcus vulcanus*. *Nuc.Acids Res.* 18, 3644.

Shine, J. and Dalgarno, L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc.Natl. Acad.Sci.USA* 71, 1342-1346.

promoter	-35		-10	references
psaA-psaB				
P1	ATGAGG	AATACCGATTCTCTCAT	TAACAT	this study
P2	TTCTCT	CCAGAGGTTCATTACTG	AACACT	this study
P3	TCGACT	AATACGACTCACTATAGGGC	GAGAAT	this study
psaC	TCTCCA	GACTGCCCTTATTGTCG	TAGCTT	this study
mtnA	TTGACA	TATGAACAGTTGTTCAAG	TATACT	this study
partial				
-ORF	TTGCAA	CCCCTCTCAATACTTCT	AAGAAC	this study
psaA-psaB				
(PCC7002)	ATGATT	TCAGGGAGTTTTAAA	TATTGT	(1)
psbA				
(PCC7120)	TAGTAA	ATTTGCGTGAATTCATG	TAAATT	(2)
woxA				
(PCC7120)	TTGATT	TTGAGTTAAAAAAAATCG	TATGAT	(2)
E.coli				
consensus	TTGACA	(16-19bases)	TATAAT	(3)

Figure 5-1.

Comparison of the promoter sequences of *S.vulcanus* psaA-
psaB, psaC, mtnA and partial ORF at the downstream of the
psaB with other bacterial promoters. *psbA*, D1 protein of
PSII reaction center; *woxA*; 33kDa extrinsic polypeptide of
photosystem II (manganese-stabilizing polypeptide: MSP).
Numbers cited in the parentheses corresponded to: (1)
Cantrell and Bryant (1987); (2) Borthakur and Haselkorn
(1989); (3) Lewin (1987).

psaA–psaB operon

```
           [6]   [5]      [7]    [42]  [7]      [19]  [77]
   box-A                box-B  box-C         box-D
5'-AAAACATG— P1 —GGCCA——GGGC— P2 —GAATAA— P3———psaA
            |  |                    |  |
         TTCTCTCATTAA          TTCTCTCCAGA
         box-E'                box-E
```

psaC

```
           [158]       [4]      [7]    [107] [52]
   box-D     box-A      box-B  box-C
5'-GAATAA——AAAAACTG—GGCCA——GGGC—— P —— psaC
                                  |  |
                               TTCTCTCCAGA
                               box-E
```

mtnA
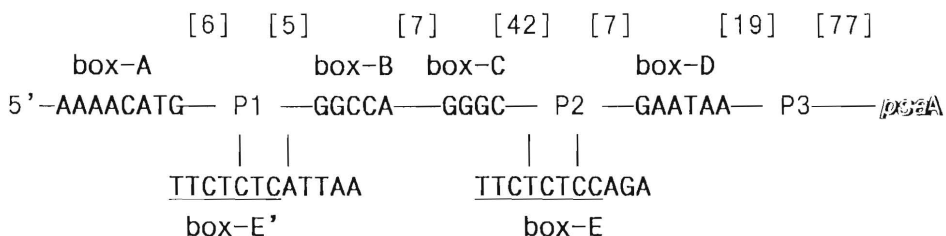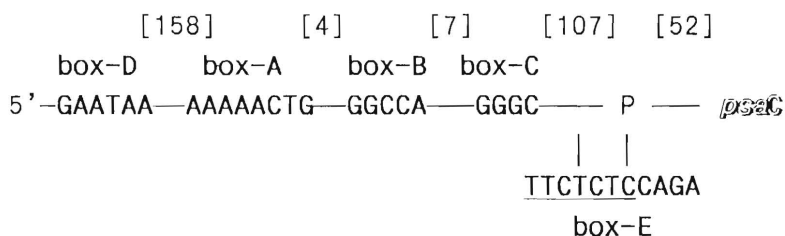
```
           [4]    [30]   [0]      [11]  [107] [51]
   box-B  box-B  box-B  box-A    box-C
5'-GGCCA—GGCCA—GGCCA|AAAAACTG—GGGC— P —— mtnA
```
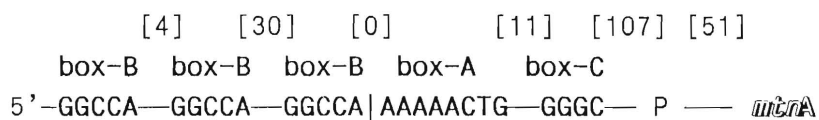
Figure 5-2.

A map of the conserved nucleotide sequence which were
found around the 5'-upstream region of the particular
genes. Each boxes indicates the conserved nucleotide
sequences. The P, also includes P1, P2 and P3 indicate the
possible promoter sites. Numbers in square brackets
indicate the length(bp) of the distance to the next struc-
ture. The underlined regions of both the box-E and box-
E' are the conserved sequence among them.

5-9

Gene specific production of a photo-
system-I subunit (PSI-C) with functional
iron-sulfur clusters *E.coli in vivo* and
reconstitution of photochemical reaction
in vitro

Abstract

A *psa*C gene from a thermophillic cyanobacteri-
um, *Synechococcus vulcanus* was introduced into an
E.coli expression vector under the regulation of
T7 RNA polymerase specific promoter. By adding
IPTG at 0.5mM to the JM109(DE3) strain harboring
the plasmid provoked the production of the PSI-C
protein in a high amount. Electron microscopy
revealed distinct cytoplasmic inclusion bodies
present in the gene-expressed *E.coli* cells.

These inclusion bodies were enriched with iron-sulfur clusters as evidenced by low temperature EPR spectra, without any *in vitro* supplementation of either iron or sulfur. By adding this fraction to a photosystem-1 reaction center preparation that consisted only of the large subunits (PSI-A and PSI-B), typical EPR signals due to Center A and B were restored without adding any other photosystem I subunit, contrary to a previous results by Golbeck *et al.*(1990).

Introduction

The PSI-C protein, a 9kDa subunit of photosystem I, contains two [4Fe-4S] iron-sulfur clusters, centers A and B. The *psa*C genes encoding PSI-C protein were first found in the chloroplast DNA of tobacco (Hayashida *et al*. 1987), liverwort (Oh-oka *et al*. 1987) and barley (Høj *et al*. 1987) and later in other plants and cyanobacteria including the gene from thermophilic cyanobacterium *Synechococcus vulcanus* (Shimizu, *et al*.1990). In this report, we describe the production and isolation of the PSI-C protein with intact iron-sulfur clusters, centers A and B, in *E.coli* to which the *S.vulcanus psa*C gene has been introduced.

pSVCX1 and pSVCX2 construction

A *S.vulcanus* 524bp DNA fragment containing the *psa*C gene was isolated from pSVC6 by using the *Xba*I/*Eco*RI digestion and cloned into the *Xba*I/*Eco*RI site of pGEMEX10 to make the pSVCX1 (3,637bp) plasmid by replacing the original T7-gene10 region of pGEMEX10. The pSVC6 is a plasmid which contains the *psa*C gene at the multi-cloning site of pTZ19R and able to produce the PSI-C protein under the specific regulation of the *lac*-promoter, although a part of *lac*Z protein was fused at the N-terminal (Shimizu, unpublished). The cloned DNA fragment contains an entire sequence of the *S.vulcanus psa*C gene with a putative ribosome binding sequence (S-D sequence) preceding the start codon (Figure 6-1). As pSVCX1 was not suitable for the following experiments due to too many restriction

endonuclease digestion sites, these extra sites were dislodged from the plasmid DNA by treating with *Bam*HI followed by self-ligation of the large fragment, yielding pSVCX2 (Figure 6-2). As the *psa*C gene of both pSVCX1 and pSVCX2 are oriented in the same direction with the T7-promoter, it is expected to express the *psa*C gene in *E.coli* to the same extent using the *in vivo* produced T7 RNA polymerase (Studier, *et al.* 1990) by the IPTG induction (data not shown).

psa*C gene expression analysis*

An *E.coli* strain JM109(DE3) harboring the pSVCX2 was treated with 0.5mM IPTG in a $50 \mu M$ $FeSO_4$ supplemented media to express the *psa*C gene *in vivo*. Since the expression of the *psa*C gene at an early-logarithmic period had strongly sup- pressed the following bacterial growth (data not shown), IPTG was added at a late-logarithmic period (A_{600} ~1.0) to the medium. The ex- pressed cells were collected by centrifugation

6-5

and disrupted by passing through a French pressure cell (8,000 p.s.i). After centrifugation at 4,000xg for 5 minutes, the precipitate and the supernatant were analyzed by using SDS-PAGE and immunostaining with an antisera raised against the spinach PSI-C protein. A band due to the expressed PSI-C protein was discernible on both CBB-stained gel and immunostained PVDF membrane (Figure 6-3). The amino acid sequence of the expressed PSI-C protein was confirmed with a protein sequencer. The obtained amino acid sequence was AHTVKIYDT and these result demonstrated that the methionine residue at the N-terminus of the expressed PSI-C protein was eliminated. Therefore, the PSI-C protein whose structure was completely identical to those of higher plants and cyanobacteria could be produced with E.coli.

Although the produced PSI-C protein is present mainly in the precipitate fraction, it is more likely that the product was precipitated as cytoplasmic inclusion bodies than integrated into the cytoplasmic membrane. It has been known

that the expression of an exogenous gene in *E.coli* would frequently result in an accumulation of the expressed protein as cytoplasmic inclusion bodies (Williams *et al.* 1982). Electron micrographs revealed the inclusion bodies only in IPTG treated cells, as shown in Figure 6-4.

Böhme and Haselkorn (1989) had previously reported that the holoprotein of a plant type [2Fe-2S] ferredoxin of *Anabaena* could be produced with functional iron-sulfur cluster in *E.coli*. Since PSI-C protein contains two characteristic amino acid sequence motifs (C-x-x-C-x-x-C-x-x-x-C) that resembled to those of bacterial ferredoxin (Adman *et al.*1973), it is reasonable to assume that the PSI-C proteins produced in *E.coli* might hold functional iron-sulfur clusters within a molecule. It can also be expected that, although the iron-sulfur clusters in the PSI-C protein are quite oxygen-labile (Oh-oka *et al.* 1989), the PSI-C protein originated from thermophilic cyano-bacterium would be more stable than those of mesophilic organisms. Therefore, we investigated a possibility of the holo PSI-C protein produced

in *E.coli*.

*Isolation   and   characterization   of   holo
PSI-C protein*

The isolated cytoplasmic inclusion bodies were
solubilized by 0.5%(w/v) Triton X-100 in an ice
bath under nitrogen atmosphere. Then, the solu-
bilized fraction was mixed with a PSI reaction
center core complex preparation (CPI) which was
free of lower molecular mass subunits including
the PSI-C and showed no EPR signals due to either
center A or center B.

The low temperature EPR spectrum of this mix-
ture showed typical ordinary signals due to
centers A and B just like those in chloroplast
(Figure 6-5). The solubilized cytoplasmic
inclusion bodies showed a characteristic EPR
spectrum which did not resemble those of centers
A and B but that reported by Oh-oka *et al.*(1988)
for the isolated holo PSI-C protein from spinach
(Figure 6-5). The differences in EPR spectrum

would be explained by a possible conformational change of the PSI-C protein when bound to the PSI complex thus affecting the magnetic properties of iron-sulfur clusters.

We could not determine the reconstitution by flash kinetic analysis due to technical difficulties; contamination of some unidentified electron carriers originated in *E.coli*, possibly cytochromes, complicated the decay kinetics of flash induced P_{700} photooxidation (Shimizu, unpublished). Therefore, we examined EPR signals to estimate for the reconstitution .

The present work has shown that the PSI-C protein is produced with functional iron-sulfur clusters in *E.coli*. The reconstitution required only the produced holo PSI-C protein and PSI core complex. Exogenous supplementation of either iron and sulfur or other lower molecular mass polypeptides was not essential. There were no detectable amount of low molecular-mass subunit which include PSI-C, PSI-D, PSI-E and PSI-F in our PSI-core complex preparation (Hiyama *et al. in preparation*), this enabled us to exclude the possibilities that the residual low molecular-mass subunit might facilitate the interaction between the PSI-C protein and the PSI core complex. The PSI-D was first recognized as a ferredoxin binding subunit of photosystem I (Zanetti and Merati,1987; Zilber and Malkin,1988). Cross-linking study has demonstrated that the 9-kDa subunit (PSI-C) closely interacts with the 14- and 19kDa subunits of spinach photosystem I (Oh-

oka *et al*.1989). These subunit interaction between PSI-C subunit and PSI-D and PSI-E subunits was also suggested in recent study of cyanobacteria through flash kinetic analysis and low temperature EPR analysis (Li *et al*. 1991). We could observe a typical low-temperature EPR spectrum in a mixture of the PSI-C protein and PSI core complex. In particular, the EPR spectra demonstrated that the binding of the PSI-C subunit to the photosystem I core complex affects the magnetic properties of centers A and B as reported by Oh-oka *et al*. (1988) and the recent report of Mehari *et al*. (1991).

Zhao *et al*.(1990) reported the *in vitro* reconstitution with *in vivo* produced PSI-C1 (Psa-C1) apo protein and they described an absolute requirement of the PSI-D protein to the reconstitution. They postulated that the requirement of the PSI-D protein is due to the existence of a five amino acid (MEHSM...) extension at the amino terminal of the PSI-C1 protein. Since the PSI-C protein we produced is the authentic one without an extra amino acid residue, it is likely that

the PSI-C protein has much affinity and easy to adhere to the PSI core complex without other low molecular mass subunit in contrast to their PSI-C1 protein. These results had prompted us to study the roles of the PSI-C protein with the site-directed mutagenized *psa*C gene and the *in vitro* reconstitution technique, in progress.

The recent progress of the *psa*C gene destruction in photosynthetic organisms were accomplished both unicellular green algae *Chlamydomonas reinhardtii* (Takahashi *et al.* 1991) and filamentous cyanobacterium *Anabaena variabilis* ATCC 29413 (Mannan *et al.* in press). The *psa*C gene destruction of *Chlamydomonas reinhardtii* by DNA-gun affected the accumulation of the large subunits (Takahashi *et al.*1991). This observation suggests the function of the PSI-C subunit to stabilize the photosystem I complex. However, they did not account the possibility about the deficient of the large subunits of the photosystem I was not caused by an undesired recombination during the isolation. Since they isolated *psa*C gene deficient mutants in expectation of

mutants lacking the photosystem I activity, it is reasonable that the mutants lack the photosystem I activity or large subunits of photosystem I as well. In cyanobacterium, although the *psa*C gene destruction caused the complete loss of the PSI-C protein and the PSI activity, the large subunits (PSI-A/PSI-B) still assembled in thylakoid membrane (Mannan *et al.* in press).

The function of the PSI-C subunit in photosystem I should be discussed after more careful investigations.

REFERENCES

Adman, E.T., Sieker, L.C. and Jensen, L.H. (1973) The structure of a bacterial ferredoxin. *J.Biol.Chem.* **248**, 3987-3996.

Böhme, H. and Haselkorn, R. (1989) Expression of *Anabaena* ferredoxin genes in *Escherichia coli. Plant Mol.Biol.* **12**, 667-672.

Golbeck, J.H. (1987) Structure, function and organization of the Photosystem I reaction center complex. *Biochim. Biophys.Acta*, **895**, 167-204 (review).

Hayashida, N. Matsubayashi, T., Shinozaki, K., Sugiura, M., Inoue K. and Hiyama, T. (1987) The gene for the 9kd polypeptide, a possible apoprotein for the iron-sulfur centers A and B of the photosystem I complex, in tobacco chloroplast DNA. *Curr.Gen.* **12**, 247-250.

Høj, P.B., Svendsen, I. Scheller, H.V. and Møller, B.L. (1987) Identification of a chloroplast-encoded 9-kDa

polypeptide as a 2[4Fe-4S] protein carrying centers A and B of photosystem I. *J.Biol.Chem.* **262**, 12676-12684.

Li, N., Warren, P.V., Golbeck, J.H., Frank, G., Zuber, H. and Bryant, D.A. (1991) Polypeptide composition of the Photosystem I complex and the Photosystem I core protein from *Synechococcus* sp. PCC6301. *Biochim.Biophys.Acta* **1059**, 215-225.

Mannan, R.M., Whitmash, J., Nyman, P. and Pakrasi, H.B. Directed mutagenesis of an iron sulfur protein of the Photosystem I complex in the filamentous cyanobacterium *Anabaena variabilis* ATCC 29413. *Proc.Natl.Acad.Sci.USA in press*.

Mehari, T., Parrett, K.G., Warren, P.V. and Golbeck, J.H. (1991) Reconstitution of the iron-sulfur clusters in the isolated F_A/F_B protein: EPR spectral characterization of same-species and cross-species Photosystem I complexes. *Biochim.Biophys.Acta* **1056**, 139-148.

Oh-oka, K., Takahashi, Y., Wada, K., Matsubara, H., Ohyama, K. and Ozeki, H. (1987) The 8kDa polypeptide in photo-

system I is a probable candidate of an iron-sulfur center
protein coded by the chloroplast gene *frxA*. *FEBS Lett.*
218, 52-54.

Oh-oka, H., Takahashi, Y., Kuriyama, K., Saeki, K. and
Matsubara, H. (1988) The protein responsible for center
A/B in spinach photosystem I: isolation with iron-sulfur
cluster(s) and complete sequence analysis. *J.Biochem.* **103**,
962-968.

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

Parrett, K.G., Mehari, T. and Golbeck, J.H. (1990) Resolu-
tion and reconstitution of the cyanobacterial Photosystem
I complex. *Biochim.Biophys.Acta* **1015**, 341-352.

Shimizu, T., Hiyama T., Ikeuchi M., Koike H. and Inoue
Y.(1990) Nucleotide sequence of the *psaC* gene of the
cyanobacterium *Synechococcus vulcanus*. *Nucl. Acids Res.* 18

3644.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* 185, 60-89.

Takahashi, Y., Goldschmidt-Clermont, M., Soen, S.-Y., Franzèn and Rochaix, J.-D. (1991) Directed chloroplast transformation in *Chlamydomonas reinhardtii*: insertional inactivation of the *psaC* gene encoding the iron sulfur protein destabilizes photosystem I. *EMBO J.*, 10, 2033-2040

Williams, D.C., Van Frank, R.M., Muth, W.L. and Burnett, J.P. (1982) Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. *Science* 215, 687-689.

Zhao, J., Warren, P.V., Li, N., Bryant, D.A. and Golbeck, J.H. (1990) Reconstitution of electron transport in photosystem I with PsaC and PsaD proteins expressed in *Escherichia coli*. *FEBS Lett.* 276, 175-180.

Zanetti, G. and Merati, G. (1987) Interaction between photo-

system I and ferredoxin: Identification by chemical cross-linking of the polypeptide which binds ferredoxin. *Eur.J. Biochem.* **169**, 143-146.

Zilber, A.L. and Malkin, R. (1988) Ferredoxin cross-links to a 22kD subunit of Photosystem I. *Plant Physiol.* **88**, 810-814.

```
                                   GATCGTTTTAGGAGCGGTTACTAC   -1

ATGGCTCACACTGTCAAAATTTACGATACTTGCATTGGCTGCACCCAGTGTGTGCGGGCC   60
MetAlaHisThrValLysIleTyrAspThrCysIleGlyCysThrGlnCysValArgAla

TGCCCCACCGATGTGCTGGAAATGGTGCCTTGGGATGGTTGCAAAGCTGGCCAAATTGCC  120
CysProThrAspValLeuGluMetValProTrpAspGlyCysLysAlaGlyGlnIleAla

TCCTCCCCTCGCACGGAAGATTGCGTTGGCTGCAAACGCTGTGAAACGGCCTGCCCAACT  180
SerSerProArgThrGluAspCysValGlyCysLysArgCysGluThrAlaCysProThr

GACTTTTTGAGTATTCGTGTCTATTTGGGTGCAGAAACCACCCGCAGCATGGGTCTAGCC  240
AspPheLeuSerIleArgValTyrLeuGlyAlaGluThrThrArgSerMetGlyLeuAla

TACTAGTCCCCTATTTTCGTGAAACTTTCGGCAAATTCTTGACTAGGCTTGAGGCAACAT  300
Tyr

CCCTACGTCTCAAGTCTTTTATTTGTTTATTGGTTGTCCTTAGCACCAACCAACCCCTAC  360

TCACCTACAACTCAGGGCTGGTCTTCGTTACTAAGGCGGCTTTGTAGGTCATCCACTGCT  420

TCATTGATTTGGGCAATTTTAGCCTCTAGCTCTAGGCGCATCGCCTCCATCACAGGGTCA  480

TCATTCCCCTGCAGGGGAGTG                                         501
```

Figure 6-1。

The nucleotide and expected amino acid sequences of the
S.vulcanus psaC gene in pSVCX1 and pSVCX2. An under-
lined sequence indicates an exprected ribosome binding site
(Shine-Dalgarno sequence).
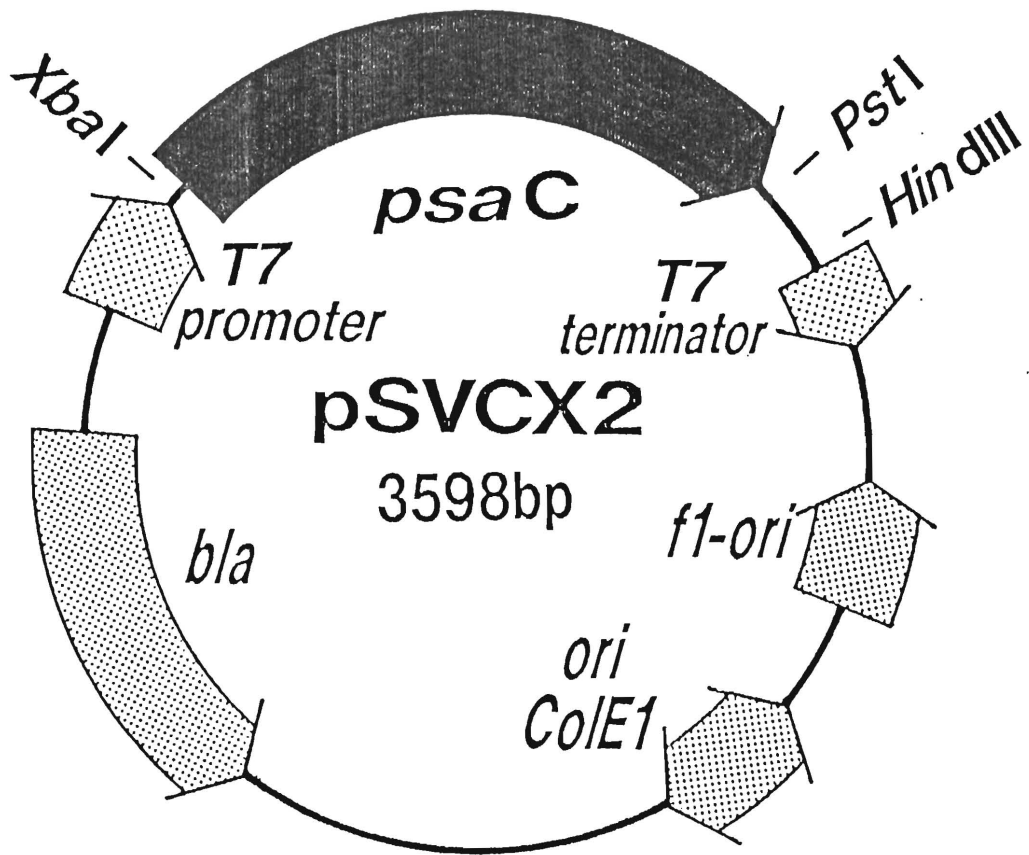
Figure 6-2.

A map of the plasmid pSVCX2. The pSVCX2 has unique
sites of *Xba*I, *Pst*I and *Hin*d III, but can not be digested
with *Eco*RI, *Sma*I, *Kpn*I, *Sac*I, *Acc*I, *Sa*lI, *Hin*cII, *Xho*I, *Nhe*I,
*Pvu*II, *Sna*BI, *Nco*I, *Eco*RV and *Cla*I.

6-20

A(CBB)   B(EIA)

M 1 2 3 4   1 2 3 4 5 M   kDa

— 94
— 67

— 43

— 30

— 20.1

— 14.4
8.24
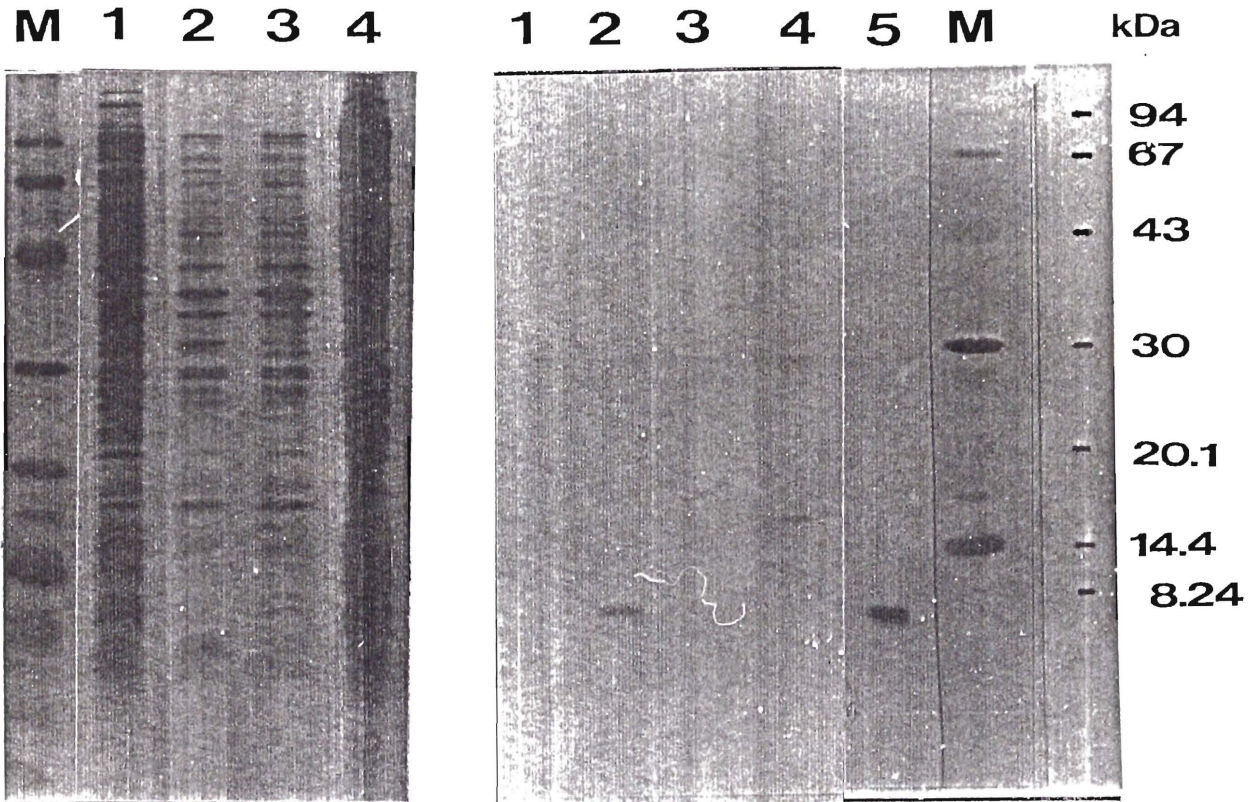
Figure 6-3.

Gel electrophoresis analysis of the PSI-C protein produced
in *E.coli*. A.The CBB-stained gel. B.The immunostained gel.
Each lanes correspond to: (1)total cell proteins of gene ex-
pressed cell (2)cytoplasmic inclusion body fraction
(3)membrane fraction (4)cytosol fraction (5) *Synechococcus
vulcanus* PSI preparation (B only).
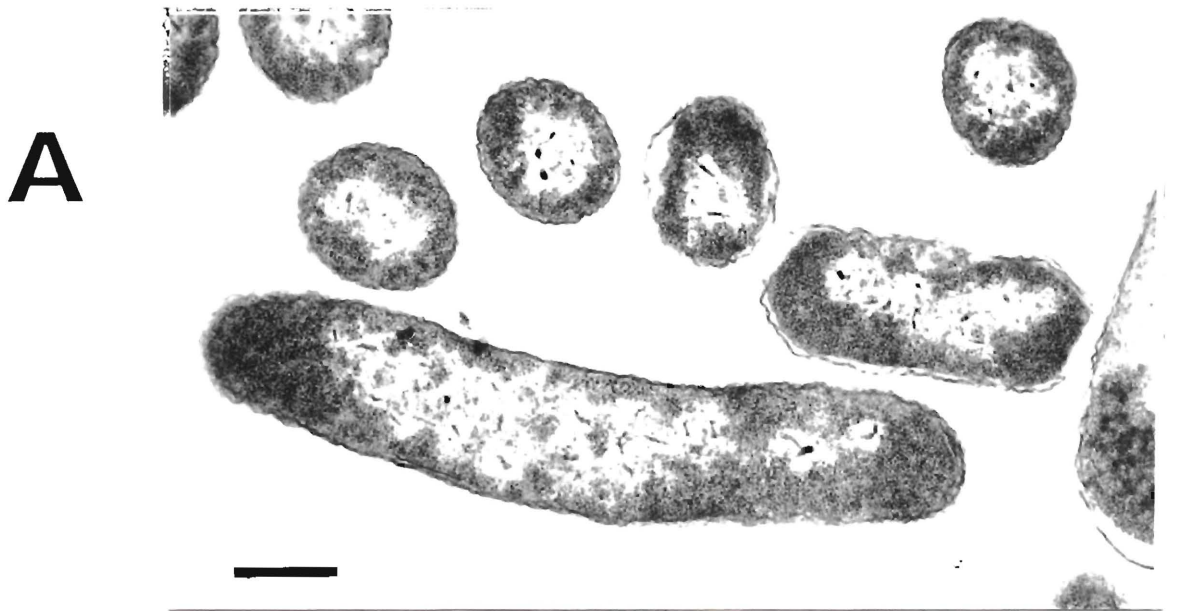
Figure 6-4.

Electron microscopy of the *psa*C gene expressed, and not expressed cells. A. Control pSVCX2/JM109(DE3) strain (*psa*C gene did not express) B.IPTG treated, gene expressed pSVCX2/JM109 (DE3) strain. Bars indicate 0.5μm.

6-22

ESR analysis of F_A/F_B restoration

PSI-C

g=1.926

PSI core +PSI-C

PSI

1.92 1.89

g=2.05   2.00   1.94 1.92   1.89   1.86

3.1   3.2   3.3   3.4   3.5

Magnetic field, kG
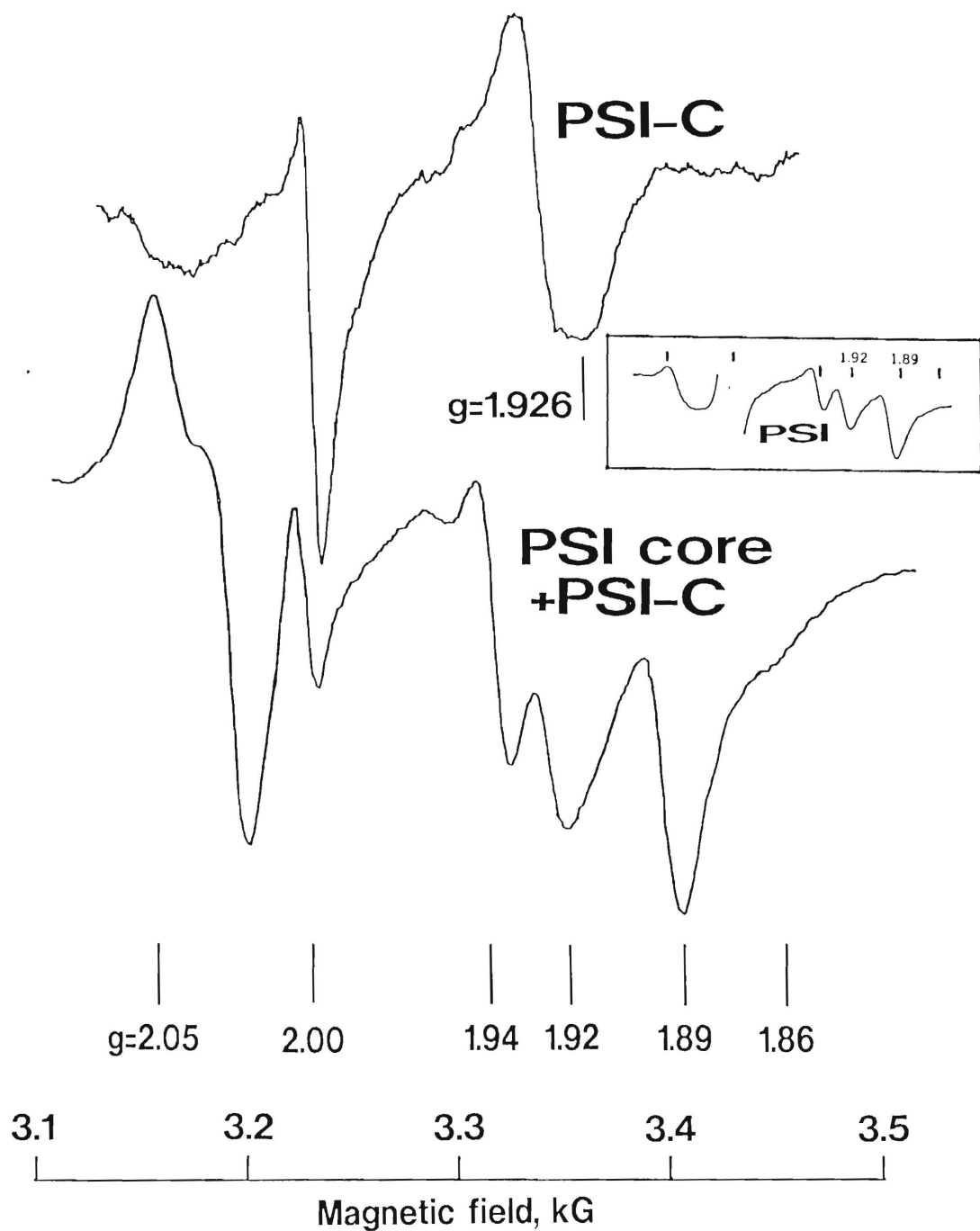
Figure 6-5.

Low-temperature EPR analysis.

CHAPTER SEVEN

Site-directed mutants of the *psaC* gene of
Synechococcus vulcanus

The terminal electron acceptors of photosystem I are [4Fe-4S] iron-sulfur cluster center A and center B which are bound to the PSI-C subunit like bacterial ferredoxin (Adman *et al.* 1973) (Figure 7-1). Many investigator have presumed the order of electron-flow mechanism of the photosystem I. Nevertheless, there is no decisive evidence about the order of the electron flow of centers A and B; whether they transfer electrons in parallel manner or sequential manner from center F_x to ferredoxin. Furthermore, we do not know about which of the two cysteine motifs

ligates the iron-sulfur cluster centers A and B, and the function of another cysteine residue which is located in the middle of the PSI-C sequence and does not ligate the iron-sulfur clusters.

I have established the *in vitro* reconstitution analysis method with an *in vivo* produced PSI-C protein of *Synechococcus vulcanus* (Shimizu *et al*. 1990) containing functional iron-sulfur clusters and photosystem I core complex (Shimizu *et al*. in preparation). The reconstitution analysis with the site-directed mutants of the PSI-C subunit must be able to answer the question.

An amino acid residue of the PSI-C protein was replaced by another amino acid residue by using a site specific mutagenesis. I produced six *psa*C mutants, each of which has replaced amino acid residues around the ligand of iron-sulfur cluster binding site. I have focused to mutate the amino acid residues just around the cysteinyl residues in two ways: (1) replace the cysteine residue to another residue (2) insert an amino acid residue

into the cysteine clusters (Figure 6-2). I also tried to produce a mutant C14S in which the cysteine-14 residue was replaced by a serine. However, I could not isolate a mutant after several attempts for unknown reasons.

Although the [4Fe-4S] iron-sulfur clusters were retained in the PSI-C protein by cysteinyl ligands, the non-cysteinyl ligation of a [4Fe-4S] iron-sulfur cluster was found in a ferredoxin of hyperthermophilic archaebacterium *Pyrocuccus furiosus* (Aono *et al*.1989 ;Conover *et al*. 1990). It has atypical arrangement of cysteinyl residues: X_{10}-Cys-X_2-Asp-X_2-Cys-X_3-Cys-Pro-X_{25}-Cys-X_7-Cys-Pro-X_9. While the cysteine residues were arranged analogous to that of the other [4Fe-4S] ferredoxin, a second cysteine residue in the ligand of iron-sulfur cluster was replaced with an aspartate residue and an iron atom in the [4Fe-4S] cluster could ligate to the aspartate residue as well (Conover *et al*.1990). Therefore, since replacement of the second cysteinyl residue to another amino acid residue such as aspartate may effect little to the iron-sulfur

clusters, replacement of the second cysteinyl
residue was excluded from the original.

All six mutants could produce the PSI-C protein
E.coli in vivo as efficiently as wild type PSI-C
protein (Figure 7-2). However, the growth of a
mutant ø15G is suppressed after the gene expres-
sion was initiated by an IPTG (data not shown).

Because I could not examine the flash induced
kinetic analysis of P_{700} as described in the
chapter 6, it is indispensable to isolate the
PSI-C holo protein from a *E.coli* inclusion body
fraction using a method similar to that of Oh-oka
et al.(1987), or reconstitute the photosystem I
in vitro with purified PSI-C protein by *in vitro*
reconstitution manner, as demonstrated by Zhao *et
al.* (1990).

Further work is needed to identify the iron-
sulfur clusters binding site of PSI-C protein and
the order of electron transfer by the reconstitu-
tion analysis.

REFERENCES

Adman, E.T., Sieker, L.C. and Jensen, L.H. (1973) The structure of a bacterial ferredoxin. *J.Biol.Chem.* **248**, 3987-3996.

Aono, S., Bryant, F.O. and Adams, M.W.W. (1989) A novel and remarkably thermostable ferredoxin from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *J.Bacteriol.* **171**, 3433-3439.

Conover, R.C., Kowal, A.T., Fu, W., Park, J-B., Aono, S., Adams, M.W.W. and Johnson, M. (1990) Spectroscopic characterization of the novel iron-sulfur clusters in *Pyrococcus furiosus* ferredoxin. *J.Biol.Chem.* **265**, 8533-8541.

Oh-oka, H., Takahashi, Y., Kuriyama, K., Saeki, K. and Matsubara, H. (1988) The protein responsible for center A/B in spinach photosystem I: isolation with iron-sulfur cluster(s) and complete sequence analysis. *J.Biochem.* **103**, 962-968.

Shimizu, T., Hiyama T., Ikeuchi M., Koike H. and Inoue Y.(1990) Nucleotide sequence of the *psa*C gene of the cyanobacterium *Synechococcus vulcanus*. *Nucl. Acids Res.* 18 3644.

Zhao, J., Warren, P.V., Li, N., Bryant, D.A. and Golbeck, J.H. (1990) Reconstitution of electron transport in photo-system I with PsaC and PsaD proteins expressed in *Escheri-chia coli*. *FEBS Lett.* **276**, 175-180.

A
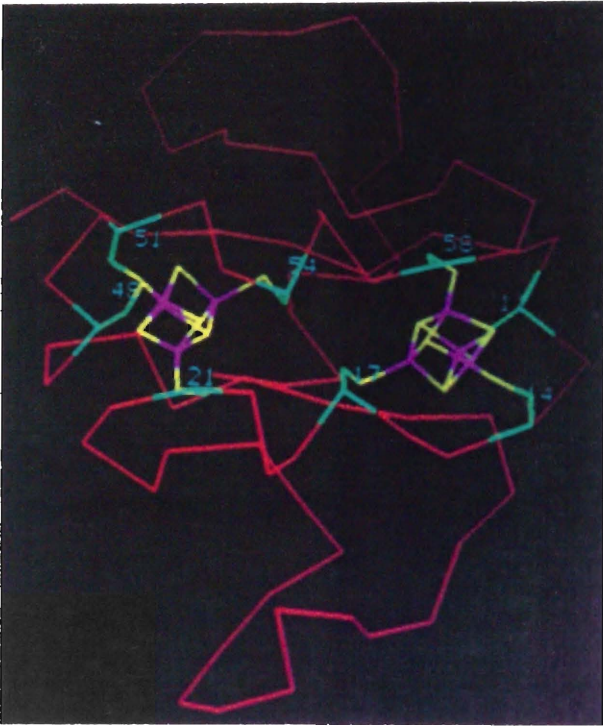
B
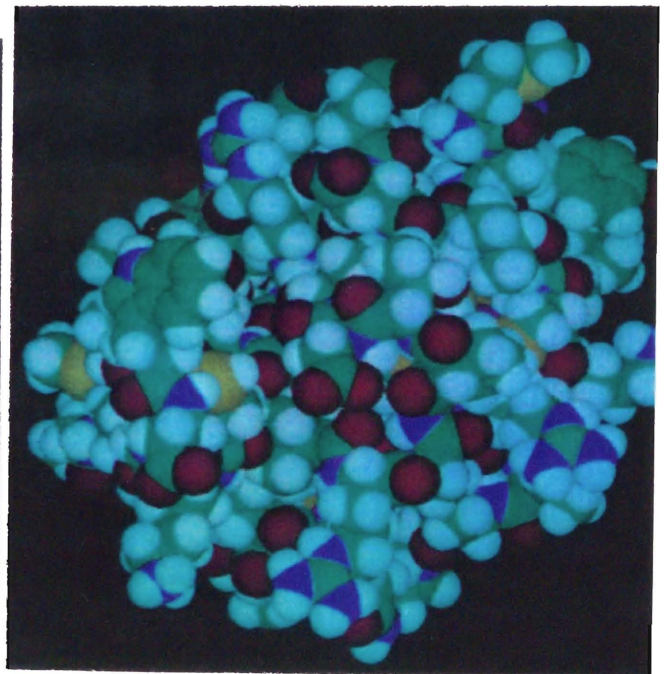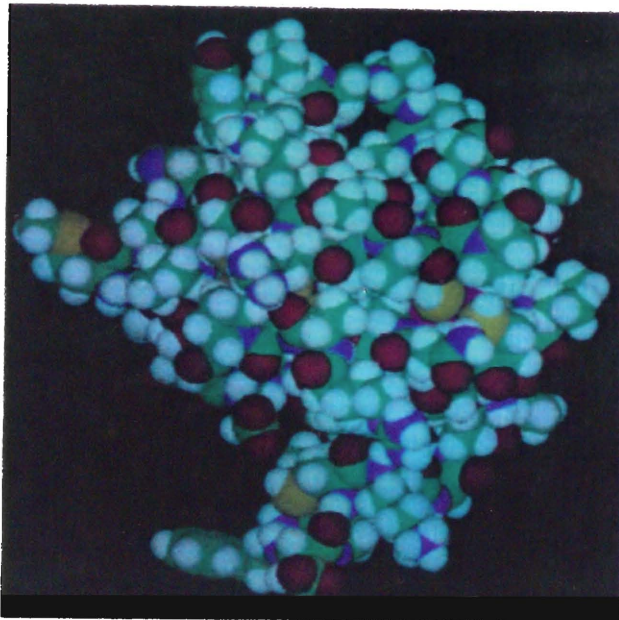
Figure 7-1.

   The predicted three dimensional structure of the *S.vulcanus* PSI-C protein, drawn by a computer. Each pictures demonstrate (A) wire models and (B) space filling models of deduced structure with different angles. Each colors correspond to: (A) red – α carbons, yellow – sulfur, purple – iron, green – α carbons of cysteine (B) red – oxygen, purple – iron, white – hydrogen, yellow – sulfur, green – carbon, blue – nitrogen. The photos were kindly offered from Prof.Rachel Nechushtai, the institute of life science, the Hebrew university of Jerusalem.

```
        1       10       20       30       40       50       60       70       80
MAHTVKIYDTCIGC-TQCVRACPTDVLEMVPWDGCKAGQIASSPRTEDCVGCKR-CETACPTDFLSIRVYLGAETTRSMGLAY


C17L    ..............-..L.................................-............................
C21S    ..............-......S............................-............................
C34T    ..............-..................T................-............................
C48I    ..............-.......................................I.....-..................
ø15G    ..............G...................................-............................
ø53S    ..............-.................................S..............................
```

Figure 7-2.

The amino acid sequences of the wild type, and the mutants of the PSI-C proteins. Bars(-) indicate the inserted positions for the mutants ø15G and ø53S. All conserved amino acid residues were represented by dots(.).

A(CBB)

wild C34T C48L C17L C21S ¢15G ¢53S -control

S P S P S P S P S P S P S P S P
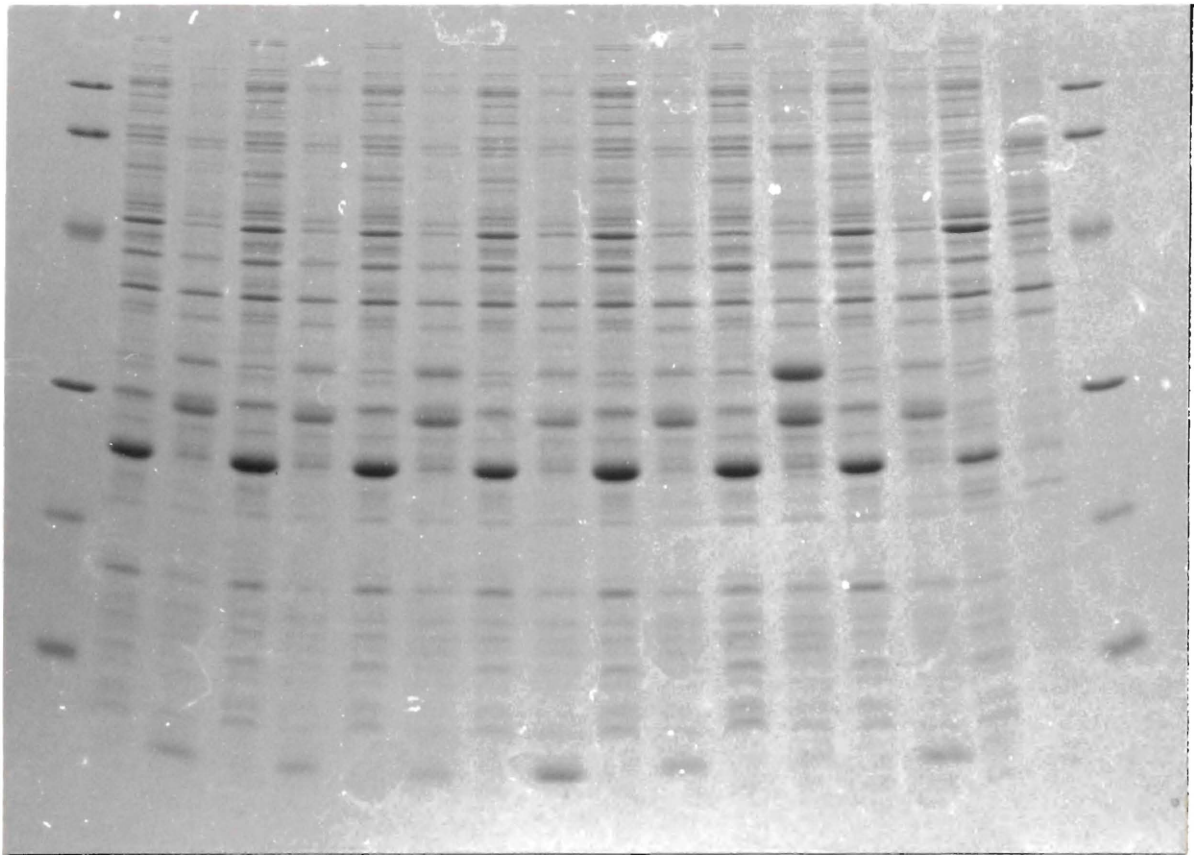
B(EIA)

S P S P S P S P S P S P S P S P

Figure 7-3.

The SDS-PAGE analysis of the mutated PSI-C protein production by (A) CBB-staining and (B) immunostaining. Each P and S represent the precipitated and soluble fractions of centrifugation after cell disruption. Arrows indicate the PSI-C and mutant proteins.

CHAPTER EIGHT

GENERAL CONCLUSION

In this thesis, I have revealed the structure, function and reconstitution of cyanobacterial photosystem I as described below:

(1) The *psa*C gene of photosystem I which encodes the 9kDa subunit containing the iron-sulfur centers A/B was isolated from thermophilic cyanobacterium *Synechococcus vulcanus*. The deduced amino acid sequence was well agreed to the formerly determined amino acid sequence except for cysteine and tryptophan residues.

(2) The *psa*A-*psa*B genes of photosystem I which encode the large subunit of the photosystem I core complex (PSI-A and PSI-B) were also isolated from the cyanobacterium.

(3) The nucleotide sequences of these isolated genes (5,945bp for the *psa*A-*psa*B genes and 2,702bp for the *psa*C gene) were both determined.

(4) The structure of the *psa*A-*psa*B operon was quite similar to those of the higher plants, in contrast to those of cyanobacteria. The deduced amino acid sequences were also homologous to those of the higher plants.

(5) Although the nucleotide sequence of the *psa*C gene resembled those of higher plants, the nucleotide sequence around the *psa*C gene was not homologous to those of the higher plants.

(6) I have found an ORF which exhibits a characteristic distribution of cysteine residues. Since the deduced amino acid sequence of the found ORF was homologous to those of the other cyanobacterial metallothineins, the ORF was presumed the to be gene encoding a metallothionein (*mtn*A).

(7) All these cloned genes, *psa*C, *mtn*A and *psa*A-*psa*B, were found to exist as a single copy gene by genomic southern hybridization analysis, respectively.

(8) The cloned *psa*C gene was inserted into an expression

vector of *E.coli* and expressed *in vivo* under the regulation by IPTG. The produced PSI-C protein has an identical mass on the electrophoresis gel to the native PSI-C protein, and cross-reacted to the PSI-C protein specific antibody. The sequence analysis of the amino acid residues of the produced PSI-C protein revealed that the methionine residue at the N-terminus was removed and the PSI-C protein was completely identical to those of the higher plants.

(9) Cytoplasmic inclusion bodies were present in the *psa*C gene expressed cells when observed by an electron microscopy. It was likely that produced PSI-C proteins were accumulated in cells as cytoplasmic inclusion bodies. No membranes or layers which surround the inclusion bodies could not be found.

(10) The cytoplasmic inclusion bodies were isolated after the cell disruption through a French pressure cell and solubilized by Triton X-100 under nitrogen atmosphere. After low speed centrifugation, the solubilized inclusion bodies were mixed with a photosystem I core complex of spinach. The low temperature EPR spectrum of this mixture showed typical ordinary signals due to centers A and B just like those in chloroplasts. The solubilized cytoplasmic inclusion bodies showed a characteristic EPR spectrum which did not

resemble those of centers A and B but that for the isolated holo PSI-C protein from spinach.

(11) These observations strongly indicated the production of the PSI-C protein with intact iron-sulfur clusters, centers A and B, in *E.coli* to which the *S.vulcanus psa*C gene had been introduced.

(12) I have also constructed six *psa*C mutants, each of which was replaced amino acid residues around the ligand of iron-sulfur cluster binding sites. These mutants could be used for the further study of the iron-sulfur centers A and B.

This is a first observation of the production of the PSI-C protein with functional iron-sulfur clusters in *E.coli*, and the restoration of the photosynthetic EPR signals of iron-sulfur clusters center A and center B without other lower molecular mass subunits of photosystem I, such as PSI-D.