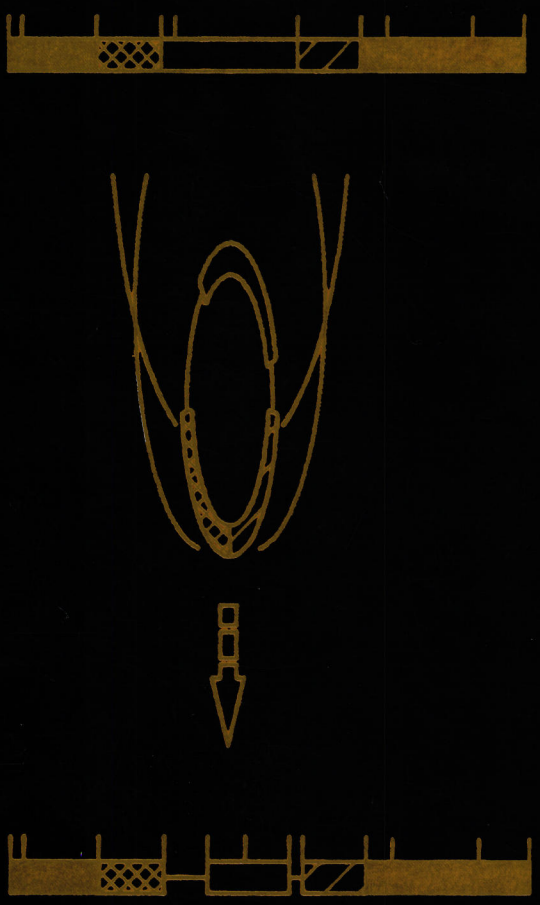
Gene analysis in the cyanobacterium *Synechococcus* sp. PCC 7942



Jan van der Plas

576.8

GENE ANALYSIS IN THE CYANOBACTERIUM *SYNECHOCOCCUS* SP. PCC 7942

Gen analyse in de cyanobacterie *Synechococcus* sp. PCC 7942

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Rijksuniversiteit Utrecht, op gezag van de rector magnificus, Prof. Dr. J.A. van Ginkel, volgens besluit van het College van Dekanen in het openbaar te verdedigen op maandag 20 november 1989 des namiddags om 4.15 uur

door

Jan van der Plas

geboren op 18 augustus 1955 te Landsmeer

Promotores: Prof. Dr. G.A. van Arkel

Prof. Dr. P.J. Weisbeek

(verbonden aan de Faculteit der Biologie
van de Rijksuniversiteit Utrecht)

The studies described in this thesis were performed in the Department of Molecular Cell Biology and Institute for Molecular Biology and Medical Biotechnology, University of Utrecht, Transitorium III, Padualaan 8, 3584 CH Utrecht, The Netherlands.

The publication of this thesis was partly supported by the TNO Division for Nutrition and Food Research, Zeist, The Netherlands.

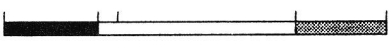
"Its the journey that matters, not the arrival"
(R. Y. Stanier)

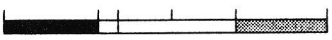
Aan Korry

Contents

Chapter I

Introduction





Introduction

1. General introduction

Cyanobacteria constitute the largest and most widely distributed group of photosynthetic prokaryotes (Stanier and Cohen-Bazire, 1977). They form a heterogeneous group of gram negative bacteria, encompassing wide differences in morphology, physiology and reproduction. Their dominant unifying feature is the presence of a photosynthetic apparatus, comprised of two photoreaction centres and an electron transport chain, similar to that found in higher plant chloroplasts (Crofts and Wood, 1978; Ho and Krogmann, 1982; Hladík and Sofrová, 1983). The two photosystems (PS II and PS I) operating in series are capable of oxidizing water to O_2 and generating ATP and reductant for use in biosynthetic reactions. All cyanobacteria are capable of CO_2 fixation and, while some are facultative heterotrophs, many appear to be obligate photoautotrophs (Stanier and Cohen-Bazire, 1977; Rippka *et al.*, 1979).

In response to environmental fluctuations cyanobacteria display a wide variety of regulatory processes. Well-known examples are: 1) adjustment of the light-harvesting apparatus to the quantity (antennae acclimation; Lönneborg *et al.*, 1985) and the quality (complementary chromatic adaptation; Tandeau de Marsac, 1983) of the available light, 2) shifting between equivalent electron transport chain components depending on the availability of their metal-cofactors (e.g. ferredoxin/flavodoxin, plastocyanin/cytochrome c_{553}; Bryant, 1987; Riethman *et al.*, 1988), 3) induction of specific systems for the mobilization of reserve material from a range of storage facilities (cyanophycin, glycogen and polyphosphate granules, phycobilisomes; Yamanaka and Glazer, 1980; Allen, 1984; Simon, 1987), 4) the formation of heterocysts, cells specialized in fixing atmospheric N_2 under aerobic conditions (Wolk, 1982; Van Baalen, 1987) in which specific

genomic rearrangements are found (JW Golden *et al.*, 1987), and 5) buoyancy regulation by formation of gas vesicles (Walsby, 1987).

The unicellular cyanobacterium *Synechococcus* sp. PCC 7942, formerly called *Anacystis nidulans* R2 and close relative of *Synechococcus* sp. PCC 6301 (Golden *et al.*, 1989a), is often used as a model for the analysis of cyanobacterial regulatory phenomena. *Synechococcus* sp. PCC 7942 is one of the best studied cyanobacterial strains from a physiological, structural and genetical point of view, it is highly transformable with both chromosomal and plasmid DNA, and a range of cloning vehicles for this strain has been developed (Kuhlemeier and Van Arkel, 1987; Tandeau de Marsac and Houmard, 1987). In the next paragraphs some aspects of cyanobacterial research relevant to gene analysis in the cyanobacterium *Synechococcus* sp. PCC 7942 are presented.

1. Dynamic aspects of the cyanobacterial photosynthetic apparatus

The cyanobacterial photosynthetic apparatus is localized in the thylakoids, intracytoplasmatic membranes, which are generally arranged peripherally in three to six concentric layers (Fig. 1; Golecki and Drews, 1982). This apparatus consists of five multiprotein complexes: the phycobilisome, the PS II and oxygen evolving complex, the cyt b_6/f complex, the PS I complex, and the ATPase complex (Fig.2; Sherman *et al.*, 1987). Each of these complexes, with the exception of the phycobilisome, is the structural and functional equivalent of its higher plant analog (Ho and Krogmann, 1982; Hladík and

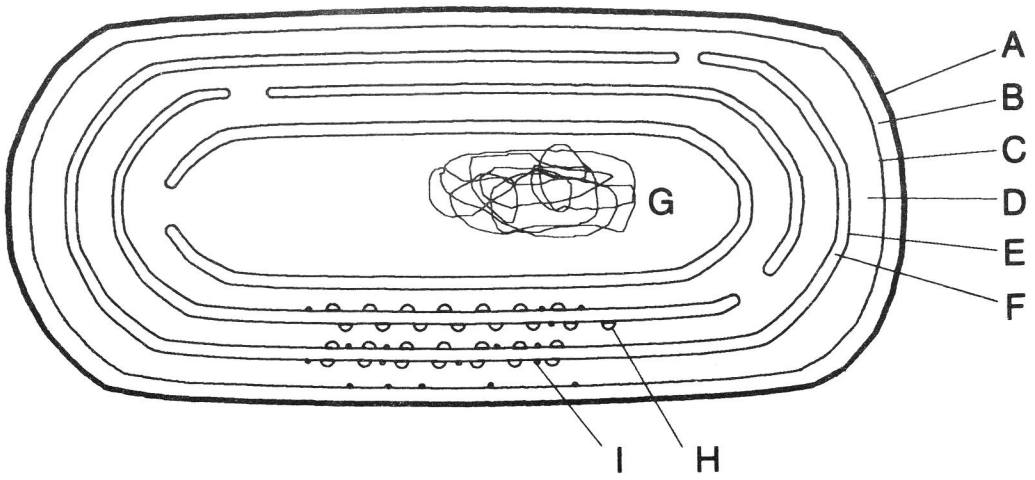


Fig. 1. Schematic drawing of a unicellular cyanobacterium. A, outer membrane and cell wall; B, periplasmic space; C, cytoplasmic membrane; D, cytoplasm; E, thylakoid membrane; F, thylakoid lumen; G, nucleoplasmic region; H, ATPase; I, phycobilisome.

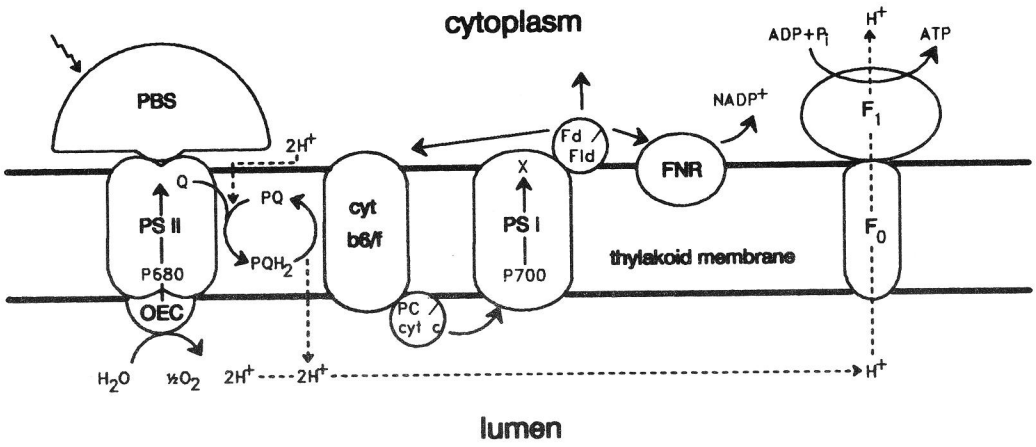


Fig. 2. Schematic representation of the cyanobacterial photosynthetic electron transport chain. $cyt_{b6/f}$, cytochroom b_6/f complex; cyt_c, cytochrome c_{553}; F_0 and F_1, parts of the ATPase complex; Fd, ferredoxin; Fld, flavodoxin; FNR, ferredoxin NADP reductase; PBS, phycobilisome; PC, plastocyanin; PQ, plastoquinon; PS I, photosystem I; PS II, photosystem II; OEC, oxygen evolving complex.

Sofrová, 1983). The interactions between these multiprotein complexes are accomplished via several small, diffusible proteins and molecules: plastoquinone, cytochrome c_{553} (or plastocyanin), ferredoxin (or flavodoxin), and protons. In cyanobacteria phycobiliproteins compose a large fraction (up to 45%) of the total soluble protein (Tandeau de Marsac, 1977), and constitute the major light-harvesting antenna. These phycobiliproteins are chromoproteins, organized in multimolecular structures called phycobilisomes, which are attached to the surface of the photosynthetic membranes (Glazer, 1984, 1987).

One of the most remarkable properties of cyanobacteria is their ability to alter the composition of their photosynthetic apparatus in response to certain environmental factors. In photosynthetic prokaryotes the amount of photopigment is generally inversely proportional to the light intensity received by the cells during growth. Similarly, cyanobacteria respond to different light intensities by increasing (under low light intensity) or decreasing (under high light intensity) their chlorophyll *a* and phycobiliprotein contents (Bryant, 1987). During periods of shortage in nitrogen, carbon, sulphur or phosphate, phycobiliproteins serve as reserve material in addition to the special cell inclusions like polyglucose, cyanophycin and polyphosphate granules (Allen, 1984; Simon, 1987). Specific enzyme systems are believed to exist for the degradation of the phycobiliproteins under nutritional stress (Wood and Haselkorn, 1980; Yamanaka and Glazer, 1980).

Cytochrome c_{553} is generally found as the primary donor of electrons to PS I in cyanobacteria (Ho and Krogmann, 1982; Sandmann, 1986). Both acidic and basic forms of this soluble cytochrome have been isolated from different cyanobacterial species (Ho and Krogmann, 1984; Cohn *et al.*, 1989). Plastocyanin, which is the primary electron donor to PS I in higher plants and most algae, has been found less frequently in cyanobacteria. Also of this electron carrier acidic forms (Stewart and Kaethner, 1983; Tan and Ho, 1989), as well as basic forms have been purified (Aitken, 1975; Ho and Krogmann, 1984). Some cyanobacteria apparently possess the ability to synthesize both plastocyanin and cytochrome c_{553}, and they regulate the amount of these components according to the nutritional availability of copper (Ho *et al.*, 1979; Sandmann and Böger, 1980; Ho and Krogmann, 1982; Sandmann, 1986).

The principal acceptor of electrons from PS I in cyanobacteria is a soluble [2Fe-2S] ferredoxin (Rogers, 1987). Reduced ferredoxin serves as a common reductant for many processes. It is involved in the reduction of NADP, nitrate, nitrite, sulphate, etc. (Rogers, 1987; Andriesse *et al.*, 1989). Many cyanobacteria contain two types of ferredoxin with similar molecular weights, but differing in amino acid sequence (Matsubara and Hase, 1983; Rogers, 1987). The functional basis of this heterogeneity, which is also observed in higher plants, has not yet been established. When grown under iron-limiting conditions cyanobacteria can replace (most of) their ferredoxin complement by the FMN-containing electron carrier flavodoxin (Hutber *et al.*, 1977; Sandmann and Malkin, 1983; Rogers, 1987), as an initial step in a complex adaptation process (Boyer *et al.*, 1987; Sherman *et al.*, 1987).

Although the ferredoxin/flavodoxin exchange, the plastocyanin/cytochrome c_{553} exchange and the effects of many nutritional stress factors on the metabolism of phycobiliproteins are known in some detail, little is known about the molecular mechanisms behind the responses observed. These aspects of cyanobacterial physiology raise interest in the control mechanisms of cyanobacterial gene expression.

2. Topological aspects of the cyanobacterial photosynthetic apparatus

The existence of structural and functional relationships (if any) between the cytoplasmic membranes and the thylakoid membranes still is a controversial issue. In general, the two membrane systems appear as topologically distinct (Stanier and Cohen-Bazire, 1977; Kunkel, 1982; Staehelin, 1986), but indications for contacts between the thylakoid membranes and the cytoplasmic membrane have been reported (Nierzwicki-Bauer *et al.*, 1983; Bryant, 1987; Stone *et al.*, 1988). These putative contact zones are suggested to be sites for thylakoid synthesis, by invagination of the cytoplasmic membrane, but they

might as well function to maintain the three-dimensional arrangement of the thylakoid system (Nierzwicki-Bauer *et al.*, 1983). Since phycobilisomes are never observed on the inner surface of the cytoplasmic membrane, this strongly implies that the two membrane systems are at least functionally distinct (Stanier and Cohen-Bazire, 1977).

The cytoplasmic and thylakoid membranes divide the cyanobacterial cell into three different compartments: the cytoplasm, the periplasm and the thylakoid lumen (Fig. 1). Some constituents of the photosynthetic apparatus, like plastocyanin and the oxygen evolving complex, function in the thylakoid lumen (Fig. 2), so these molecules or their precursors have to be transported to this location during or after their synthesis in the cytoplasm. So in contrast to most other prokaryotes, cyanobacteria require an intracellular sorting mechanism to distinguish polypeptides with trans thylakoid membrane and trans cytoplasmic membrane destination. How this is accomplished is as yet unknown. The isolation and analysis of the cyanobacterial genes for thylakoid lumen proteins and the subsequent identification and comparison of their protein leader sequences can offer insight into the sorting mechanism. The first such sequence reported was for the Mn-stabilizing protein involved in photosystem II water oxidation (*WoxA*) from *Synechococcus* sp. PCC 7942 (Kuwabara et al, 1987). The derived amino-terminal extension of the precursor of this thylakoid lumen protein strongly resembles the typical bacterial signal peptide for targeting over bacterial cytoplasmic membranes (Watson, 1984). This observation implies that topologically the thylakoid lumen can be considered as extracellular space and it suggests that the thylakoids originate (evolutionary) from the cytoplasmic membrane. At the same time it makes the still unanswered question of how the cell distinguishes between outward and inward destinations even more intriguing. The isolation of the plastocyanin gene of *Anabaena* sp. PCC 7937 (Chapter III) and its transfer to the genetically accessible strain *Synechococcus* sp. PCC 7942 (Chapter IV), may provide the tools to solve this problem.

3. Gene regulation in cyanobacteria

Regulation in cyanobacteria has been encountered in different stadia of gene expression (Tandeau de Marsac and Houmard, 1987). Basically, the ways to modulate gene expression can be split in control exerted on the mRNA level and on the protein level. The amount of functional mRNA can successively be regulated by the rate of transcription initiation, elongation (pausing), termination (attenuation), and degradation or processing of the transcript. The amount of functional protein finally synthesized from the mRNA template depends on the rate of translation initiation, peptide chain elongation, protein maturation and degradation (Higgins, 1986).

3.1 regulation at the mRNA level

Although the cyanobacterial RNA polymerase (Schneider *et al.*, 1987; Schneider and Haselkorn, 1988; Xie *et al.*, 1989) has a subunit composition substantially different from that of *E. coli* and other eubacteria, their promoter specificities seem to overlap. The cloning of cyanobacterial genes with their putative regulatory sequences in *E. coli* showed that some promoter sequences present in the cloned fragments are capable of directing transcription in the heterologous host (e.g. Bryant *et al.*, 1985). Likewise, the introduction of foreign genes into *Synechococcus* sp. PCC 7002 showed that some foreign bacterial (e.g. Buzby *et al.*, 1985) and even chloroplast regulatory sequences (Dzelzkalns *et al.*, 1984) can be recognized in vivo by the cyanobacterial polymerase. The fact that for any cyanobacterial species only a small number of promoters has been mapped for their exact location, and given the possibility that some of them may require regulatory effectors in transcription initiation, prevents the description of typical cyanobacterial promoters beyond a resemblance to certain aspects of *E. coli*–like promoters (Hawley and McClure, 1983; Tandeau de Marsac and Houmard, 1987). Many putative cyanobacterial promoter sequences contain *E. coli*–like motifs in the –10 region, but they often match poorly the –35 consensus sequence. Explanations for this observation may be the requirement for an activator protein (Raibaud and Schwartz, 1984), or the different preferences of cyanobacterial and *E. coli* RNA polymerases. The promoter regions of the iron–stress inducible *isiAB* operon and *irpA* gene of *Synechococcus* sp. PCC 7942 are found to contain sequences that are similar to the consensus operator sequences involved in iron–stress regulation in *E. coli* (Laudenbach and Straus, 1988; Reddy *et al.*, 1988; Riethman *et al.*, 1988).

The regulatory sequences immediately upstream of the transcriptional start of two messengers induced by nitrogen starvation in *Anabaena* sp. PCC 7120 are very poor promoters for *E. coli* RNA polymerase; they do however point to a nitrogen fixation (*nif*) gene promoter for this strain that shows some similarity with bacterial *nif* gene promoters (Tumer *et al.*, 1983; Schneider *et al.*, 1987; Tandeau de Marsac and Houmard, 1987). The selective transcription from these *nif* promoters during heterocyst differentiation in *Anabaena* is thought to be at least in part effected by modification of the transcription apparatus, e.g. the RNA polymerase requires the presence of a special σ factor to recognise the *nif* promoter sequences (Schneider *et al.*, 1987). Sigma factors confer upon a core RNA polymerase the ability to recognize a specific group of promoters (Higgins, 1986). Indications whether in cyanobacteria specialized σ factors play a role in other situations, might come from immunochemical analysis using antisera against σ factors of RNA polymerase from *E. coli* and *Salmonella typhimurium* (Klimpel *et al.*, 1989).

The most elaborated examples of regulation in cyanobacteria concern the effects of alterations in light intensity and quality on gene expression (Bryant, 1987). This research focused in particular on the regulation of phycobiliprotein synthesis (e.g. Belknap and Haselkorn, 1987; Kalla *et al.*, 1988; Tandeau de Marsac *et al.*, 1988). Despite all the efforts little knowledge has been obtained on the cyanobacterial mechanisms of light regulation on the molecular level. The segmental regulation of the relative levels of expression of phycobilisome components from the phycocyanin operons in *Anacystis nidulans* UTEX625 appears to involve attenuation of transcriptional activity or post-transcriptional processing, whereas the absolute levels of phycobiliproteins in this organism seem to be regulated at the translational level (Kalla *et al.*, 1988). For *Anabaena* sp. PCC 7120 and Fremyella sp. PCC 7601 similar results were obtained (Belknap and Haselkorn, 1987; Tandeau de Marsac *et al.*, 1988). An example of transcriptional regulation by light is represented by the differential expression of members of the *psbA* gene family in *Synechococcus* sp. PCC 7942 (Schaefer and Golden, 1989).

In analogy to the situation in *E. coli*, also in cyanobacteria sequences capable of forming stable hairpin structures in the primary transcript have been implicated in transcription attenuation and termination (Csiszàr *et al.*, 1987). This mechanism provides a possible explanation for finding transcripts of different lengths starting from the same promoter in front of an operon. Segmental differences in mRNA stability could also explain the data. The secondary structures would then protect the mRNA from 3' exonuclease attack (Brawerman, 1987, 1989). The factors determining the stability of bacterial mRNA are rather poorly understood, because in prokaryotic cells transcription and translation are coupled. The attachement of ribosomes on an mRNA molecule not only influences its stability, but also makes it difficult to separate experimentally the processes of RNA synthesis, translation, and degradation. In addition, the half-life of most bacterial mRNA molecules is very short, in the range of 1 to 2 minutes (Newbury *et al.*, 1987).

3.2 regulation at the protein level

The start codon AUG (or GUG) is a necessary but not sufficient requirement for translation. The start codon needs also to be embedded in an appropriate sequence of about 50 bp (Scherer *et al.*, 1980). Initiation codons of cyanobacterial genes are preceded by sequences resembling the ribosome binding site of *E. coli* (Shine and Dalgarno, 1974; Tandeau de Marsac and Houmard, 1987). These sequences are complementary to the 3' end of 16S rRNA from *Synechococcus* sp. PCC 6301 (Tomioka and Sugiura, 1983). There are indications that this complementarity supports translation initiation, but that much of the information necessary to determine the translation efficiency is located in the mRNA sequence outside de SD region and the startcodon (De Boer *et al.*, 1983; Hui *et al.*, 1984;

Looman *et al.*, 1987). Alterations in the ribosome binding site sequence have been reported that enhance drastically the amount of protein synthesized (Olins *et al.*, 1988; Olsen *et al.*, 1989).

Next to AUG, GUG is also functional as a startcodon in *Synechococcus* sp. PCC 7942 and related strains (Golden and Stearns, 1988; Meng *et al.*, 1989; Reddy *et al.*, 1988; Yasui *et al.*, 1988). Whether the use of a GUG start codon results in a reduction of the amount of the gene product as compared to the level that would have been obtained by an AUG start remains to be investigated. For *E. coli* the reduction factor is 50–70% (Yasui *et al.*, 1988).

One of the four different transcripts from the region of the *Calothrix* sp. PCC 7601 gas vesicle protein operon (*gvp*) corresponds to an antisense RNA. This antisense RNA most likely impairs translation and/or modifies mRNA stability by forming duplexes with the three other transcripts (Tandeau de Marsac *et al.*, 1988). Moreover, secondary structures upstream of the coding sequences are suspect to impair translation by masking the ribosomal binding site.

The efficiency of translation can possibly be influenced by codon usage (Higgins, 1986). Other possible ways to effect the steady state level of a functional protein are the rate of synthesis (and maturation if necessary) versus the rate of specific or aspecific degradation (turnover). In the green alga *Chlamydomonas* for example, the plastocyanin content is controlled by rapid degradation of constitutively synthesized apoprotein in the absence of Cu^{2+} (Merchant and Bogorad, 1986a, b).

4.0 Cyanobacterial genetics

In 1970 Shestakov and Khuyen demonstrated that the cyanobacterial strain *Synechococcus* sp. PCC 7943 is naturally competent to take up DNA from the medium (Shestakov and Khuyen, 1970). Since then only a few other strains have been found that can be transformed by exogenously added DNA (Porter, 1985; Shestakov and Reaston, 1987) or can receive mobilizable plasmid DNA from *E. coli* by conjugation relying on the action of helper plasmids (Wolk *et al.*, 1984; Thiel and Wolk, 1987). However, this still limited number of cyanobacterial strains amenable to genetic manipulation might soon be enlarged significantly, when more recipient strains are found for conjugation (Tandeau de Marsac and Houmard, 1987) or when the recent revolution in DNA transformation called electroporation becomes a common tool in cyanobacterial research (Chassy *et al.*, 1988). A serious difficulty often encountered in transfer of heterologous DNA to cyanobacteria is presence of restriction enzymes in most of the strains examined (for a recent survey see: Houmard and Tandeau de Marsac, 1988).

4.1 mutagenesis and genomic mapping

The use of transformation for genomic mapping in cyanobacteria has been minimal. A "classical" chromosomal mapping system like that existing for *E. coli* or *B. subtilis* could not be developed because of the difficulties encountered in the isolation of stable mutants and the absence of a generalized transduction or conjugation system (Porter, 1985). Some cyanobacteria have been shown to possess about 10 genomic copies per cell under standard growth conditions, which complicates the segregation (and detection) of (recessive) mutations (Mann and Carr, 1974; Labarre *et al.*, 1989). Until now only one *E. coli* transposon, Tn*901*, has been shown to be usable in *Synechococcus* sp. PCC 7942 (Van den Hondel *et al.*, 1980; Tandeau de Marsac *et al.*, 1982; Gendel, 1987; Madueño *et al.*, 1988), although it confers low levels (1–2 µg/ml) of Ap^R resistance to *Synechococcus* sp. PCC 7942. However, at the moment there are promising developments, both in genomic mutagenesis and in the field of mapping technology.

Both specific (Williams and Szalay, 1983; SS Golden *et al.*, 1987) and random cyanobacterial mutants (Buzby *et al.*, 1985; Chauvat *et al.*, 1989; Labarre *et al.*, 1989) are successfully made by genomic integration exploiting homologous recombination. Selectable DNA fragments (e.g. antibiotic resistance cassettes) bordered by homologous genomic sequences are used for the recombinational inactivation by gene–replacement or disruption (the principles of these techniques will be discussed in more detail in section 4.3). Another important development might be the identification and isolation of an endogenous cyanobacterial mobile element: IS*701*, a 1.4 kb insertion element from *Fremyella* sp. PCC 7601 (Tandeau de Marsac *et al.*, 1988). Such IS elements may provide a suitable basis for the construction of new tools for *in vivo* mutagenesis in cyanobacteria.

Genomic mapping, i.e. the generation of an ordered clone library that fully represents a genome in order to allow the accurate construction of a genetic map aligning the physical map, is a rapidly evolving field (Coulson and Sulston, 1988). A map of this kind has already been completed for *E. coli* (Kohara *et al.*, 1987). The principal purpose of such mapping is to facilitate both the cloning of known genes and the genetic placing of known fragments (Lee *et al.*, 1988). The construction of a physical map for *Anabaena* sp. PCC 7937 from overlapping cosmid clones is well on its way to completion (Herrero and Wolk, 1986). Cloned cyanobacterial genes, homologous bacterial and higher plant genes are now localized on this map by hybridization studies.

The advent of pulsed–field gel electrophoresis (PFGE; Schwartz and Cantor, 1984) and derived techniques (FIGE, OFAGE etc.) made the separation of DNA fragments in the Megabase range practicable (Bancroft and Wolk, 1988; Smith *et al.*, 1988). This gives the

opportunity to directly derive simple restriction endonuclease maps on the genome level, without depending on prior cloning (e.g. the complete restriction map for *E. coli* by Smith *et al.*, 1987). Genes can then be placed on such maps by hybridization analysis (Lee *et al.*, 1988). PFGE allows the comparison and distinction of even very closely related strains as *Synechococcus* sp. PCC 7942 and *Synechococcus* sp. PCC 6301 at the level of genomic organization (Golden *et al.*, 1989a).

4.2 plasmid vectors

One or several plasmids are found in about 60% of the cyanobacterial strains examined (for a recent survey see: Houmard and Tandeau de Marsac, 1988). They vary in size from 1.3 kb to approximately 1 Mb (Rebière *et al.*, 1986) and until now they all remain cryptic. No cyanobacterial plasmid is known to replicate in *E. coli* (Kuhlemeier *et al.*, 1983). Also *E. coli* replicons, even the so-called broad-host range plasmids, are not functional in the cyanobacterium (Kuhlemeier and Van Arkel, 1987; see, however, Lightfoot *et al.*, 1988, versus Daniell *et al.*, 1986).

Thus, the host-vector systems based on endogenous cryptic plasmids that have been developed, are mostly in the form of hybrid shuttle vectors, plasmids capable of replication in alternate hosts (Porter, 1985; Kuhlemeier and Van Arkel, 1987; Tandeau de Marsac and Houmard, 1987; Shestakov and Reaston, 1987). In these constructions little or no attention has been paid to plasmid replication and (in)stability, except for the selection of a minimal fragment containing the functions essential for cyanobacterial replication (Laudenbach *et al.*, 1983, 1985; Schmetterer and Wolk, 1988). Until now the mechanisms which control plasmid maintenance in cyanobacteria are totally unknown. An understanding of these processes will be important in future exploitation of cyanobacterial cloning vectors.

Based on pUH24, the small cryptic plasmid from *Synechococcus* sp. PCC 7942, a broad spectrum of (shuttle) plasmid vectors has been developed (for a review see: Tandeau de Marsac and Houmard, 1987). The complete nucleotide sequence of this plasmid, as well as the localization of replication and stability functions of are now known (Van der Plas *et al.*, manuscript in preparation; Chapter V). The few other vectors made for *Synechococcus* sp. PCC 7942 are based on pUH25, the large plasmid of this strain (Laudenbach *et al.*, 1985).

Four *Synechococcus* sp. PCC 7942 genes have been isolated from genomic libraries constructed in cyanobacterial (shuttle-) vectors: a *met1* gene (Tandeau de Marsac *et al.*, 1982) and three genes involved in nitrate assimilation, *narA*, *narB* and *narC* (Kuhlemeier *et al.*, 1984a, b). Only two of these genes, *narA* and *narC*, have been found by direct

phenotypic complementation of corresponding mutants in *Synechococcus* sp. PCC 7942. In addition, pCH1 (pUH24::Tn*901*) has been successfully used in transposon mutagenesis (Tandeau de Marsac *et al.*, 1982; Kuhlemeier *et al.*, 1984a, b; Madueño *et al.*, 1988).

The common use of autonomously replicating vectors in *Synechococcus* sp. PCC 7942 as well as in other strains has been limited, because recombination between a cloned gene on a cyanobacterial vector and its chromosomal counterpart occurs with high frequency (Tandeau de Marsac *et al.*, 1982; Kuhlemeier *et al.*, 1985). The development of recombination-deficient host strains is therefore required in order to avoid recombination of plasmid-borne DNA fragments with homologous sequences on the cyanobacterial genome. Several cyanobacterial *recA*-like genes have already been cloned (Geoghegan and Houghton, 1987; Murphy *et al.*, 1987; Owttrim and Coleman, 1987), but no successful inactivation of these genes has been reported. Preliminary evidence indicates that the *recA* function might be essential to the cyanobacterium (Bryant *et al.*, 1988; Borrias, pers. comm.).

4.3 genomic integration

After introduction into the cell by transformation or conjugation, plasmids or other chimeric DNA constructs carrying chromosomal sequences can integrate into the chromosome by homologous recombination (Fig. 3). This phenomenon, described first for *S. cerevisiae*, *B. subtilis* and *S. pneumoniae* (Pozzi and Guild, 1985), occurs also in *Synechococcus* sp. PCC 7942 and other cyanobacterial species (Williams and Szalay, 1983; Buzby *et al.*, 1985; Vermaas *et al.*, 1986). It can be exploited to introduce heterologous DNA into the chromosome, using flanking homologous sequences to direct it to a particular locus. The DNA constructs for transformation by homologous recombination can consist of either linear fragments or covalently closed (*E. coli*) plasmids. The latter form of donor DNA mainly integrates into the chromosome in either of two possible arrangements: (i) "plasmid addition" or "insert duplication" (Pozzi and Guild, 1985), leading to duplication of the region of cyanobacterial DNA which was present in the plasmid and integration of the complete plasmid including its antibiotic resistance marker (Fig. 3A); (ii) "gene replacement", in which the chromosomal target is replaced by the cyanobacterial insert from the donor molecule, and as a consequence the plasmid or other surrounding sequences are lost (Fig. 3B). It should be noted that additional transformants, carrying the plasmid marker and two identical "wildtype" chromosomal duplications, can arise by replacement of the plasmid insert for its chromosomal counterpart, followed by integration of the "repaired" vector molecule (Williams and Szalay, 1983; Kolowsky and Szalay, 1986).

Gene replacement is favoured over plasmid addition in *Synechococcus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803, and occurs at a 10-100 fold higher frequency (Williams and

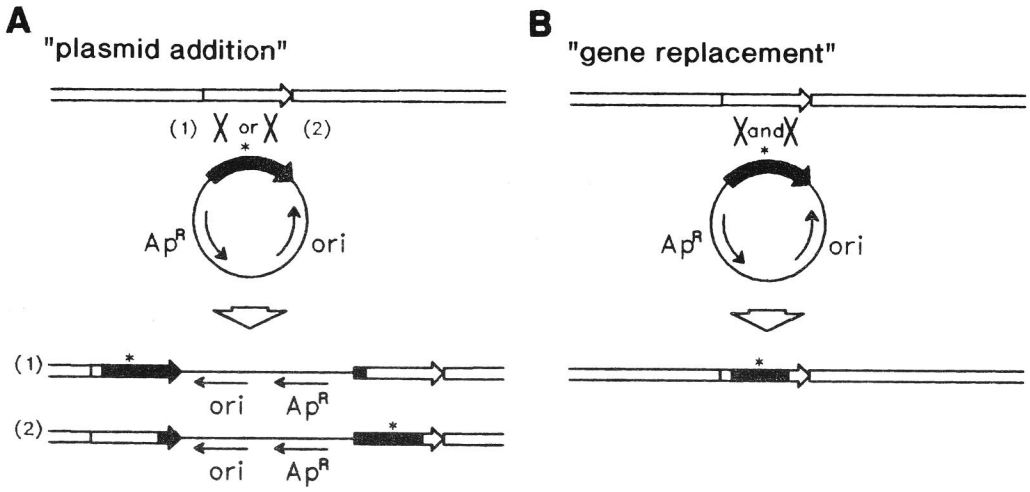


Fig. 3. Model depicting the recombination events between cloned cyanobacterial DNA fragments (filled arrow) and the chromosome (open arrow and bar). The open and filled arrows are merely used to indicate the origin of the homologous sequences and do not necessarily indicate a complete gene. The asterisk may represent any kind of mutation (e.g. point mutation, transposon insertion). The plasmid-borne antibiotic resistance gives the opportunity to directly select transformants of the "plasmid addition" type or to identify such transformants when selection depended on sequences contained in or interrupting the cyanobacterial insert (*). The duplication can be maintained by continued selection for the antibiotic resistance conferred by the integrated vector. The "gene replacement" type of transformant can be favoured by linearization of the donor molecule in the vector sequences prior to transformation.

Szalay, 1983; Kolowsky *et al.*, 1984). Integrational recombination events have been analysed in detail by Szalay and co-workers (Williams and Szalay, 1983; Kolowsky *et al.*, 1984; Kolowsky and Szalay, 1986). Plasmid addition is the expected result of a single "Campbell-like" crossing-over (SS Golden *et al.*, 1987). Gene replacement can be the result of either double crossing-over or gene conversion (via a heteroduplex DNA intermediate). We cannot distinguish between these two mechanisms which lead to gene replacement. However, the latter possibility seems more plausible given the fact that single crossovers in transformation by circular DNA are relatively rare (Porter, 1985; Labarre *et al.*, 1989).

In principle, both plasmid addition and gene replacement can be used for random mutagenesis. For effective gene inactivation by single cross-overs it is required that the homologous fragments present in the mutagenic plasmids are relatively small and consist of incomplete genes (either internal fragments or N-terminal parts without promoter), such that resulting duplications will not form functional genes. Following this mutation by vector integration, the mutated gene can easily be cloned by endonucleolytic excision.

Although successfully applied for *B. subtilis* and *S. pneumoniae* (Pozzi and Guild, 1985), no application in cyanobacteria has been reported yet. However, the replacement mechanism has been exploited by the random ligation of genomic restriction fragments of *Synechocystis* sp. PCC 6803 to selectable DNA fragments, e.g. antibiotic resistance cassettes (Labarre *et al.*, 1989). This resulted with a high frequency in stable and random insertional mutations of the gene replacement type, accompanied by precise deletions of the genomic segment normally located between the two border fragments. Interestingly, these varying deletions can be used to derive restriction maps of large (50–100 kb) regions of the chromosome (Chauvat *et al.*, 1989).

A special application of genomic integration is described in Chapter IV. Specific recombination targets, derived from pBR, are incorporated into a non-essential gene (*met1*) in the cyanobacterial chromosome. Recombination between a pBR-derived donor plasmid and such a chromosomal "integration platform" allows gene replacement based on pBR sequences alone. Furthermore the presence of the pBR origin of replication enables rescue of the plasmid with the integrated sequences. The main advantage of this system is that the integration is independent of the inserted DNA. In addition it allows complementation and dominance studies and utilization of the very many sophisticated constructs derived from pBR for gene analysis in the cyanobacterium.

4.4 cloning of genes

Evidently, any facility for analysing gene structure and function will only be fruitful in combination with cloned cyanobacterial genes of interest (for a recent survey see: Houmard and Tandeau de Marsac, 1988). Roughly, four methods for the isolation of cyanobacterial genes can be distinguished (Bryant and Tandeau de Marsac, 1988; Houmard and Tandeau de Marsac, 1988):

(i) The first and probably most widely used method involves heterologous hybridization. The availability of characterized bacterial or higher plant genes, provides a ready source of probes for the isolation of their cyanobacterial homologues, if existing. Well known examples are the identification of nitrogenase genes of *Anabaena* sp. PCC 7120 with probes derived from *Klebsiella pneumoniae nif* genes (Mazur *et al.*, 1980) and of the ribulose bisphosphate carboxylase genes with higher plant probes (Curtis and Haselkorn, 1983; Reichelt and Delaney, 1983; Shinozaki and Sugiura, 1983). In Chapter II (Van der Plas *et al.*, 1986a, b, 1988) is described how the ferredoxin I genes from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942 were cloned by low stringency hybridization with a *Silene pratensis* cDNA probe (Smeekens *et al.*, 1985).

(ii) Antibodies directed against either purified cyanobacterial proteins or homologous proteins from other bacteria or higher plants can be used as a selection criterion for the isolation of cyanobacterial genes. In particular Sherman and co-workers are exploiting their knowledge of the protein composition of cyanobacterial membranes (Bricker *et al.*, 1986; Sherman *et al.*, 1987) for obtaining specific antisera and subsequent cloning of genes for membrane proteins by immunological screening of expression libraries (Kuwabara *et al.*, 1987; Reddy *et al.*, 1988, 1989).

(iii) A third method for the identification of cyanobacterial genes rests on the use of synthetic oligonucleotide probes whose sequences are derived from determined amino acid sequences or from regions that can be predicted by homology. Several approaches exist for the design of suitable oligo sequences (Bryant and Tandeau de Marsac, 1988), ranging from full mixes representing all possible coding combinations (usually 14-17 mers) to long unique oligo's (20 mers and more). The complexity of the probe can be reduced by careful selection of less degenerate stretches, by allowing the less-destabilizing G-T basepairing (e.g. Lind *et al.*, 1985), by taking into account the characteristics of codon usage (e.g. Mazel *et al.*, 1988), and by the incorporation of deoxyinosine at ambiguous codon positions (e.g. Yasui *et al.*, 1988). The possibility of false positives can be minimized by combination of the results for two or more probes directed against different stretches of the demanded coding sequence (e.g. Alam *et al.*, 1986, Reith *et al.*, 1986, Laudenbach *et al.*, 1988). Mixed oligonucleotide probes designed after two conserved regions in the amino acid sequence of *Anabaena* sp. PCC 7118 plastocyanin, were used to clone the plastocyanin gene from *Anabaena* sp. PCC 7937 (Chapter III, Van der Plas *et al.*, 1989).

(iv) Finally, there is a possibility to isolate genes by complementation of characterized mutations. Both cyanobacterial (Kuhlemeier *et al.*, 1984a, 1984b; Dzelzkalns and Bogorad, 1988) and *E. coli* mutants (Williams and Szalay, 1983; Kodaki *et al.*, 1985; Porter *et al.*, 1986; Parsot and Mazel, 1987) have been employed for cloning by *in vivo* complementation. Besides, the inactivated target from transposon induced cyanobacterial mutants or mutants obtained by random cloning of antibiotic resistance genes (Labarre *et al.*, 1989) can be selected by their resistance marker, when cloned into an *E. coli* vector. Subsequently, the corresponding wildtype DNA can be isolated with the mutated fragment as a probe (Tandeau de Marsac *et al.*, 1982; Kuhlemeier, 1984a).

4.5 gene analysis

On the one hand the active recombination process in cyanobacteria is a powerful tool for their genetic engineering, on the other hand it has severely inhibited the use of autonomously replicating vectors in cloning and functional analysis of cyanobacterial DNA. However, shuttle vectors as well as non-replicating vectors designed for specific

recombination with the host genome, can in principle be used in gene analysis for the introduction of (i) genes for complementation or functional analysis, (ii) gene-fusions with an easily assayed product ("reporter") for regulation analysis, and (iii) altered genes for site-specific mutagenesis. Some examples of these applications are given in the following paragraphs.

ad i. Homologous or heterologous genes can be introduced for complementation analysis, e.g. as has been done in establishing the complementation groups of *nar* mutants with the 3 genes involved in nitrate reduction (*narA*, *narB* and *narC*) cloned from *Synechococcus* sp. PCC 7942 and present in cyanobacterial cloning vehicles (Kuhlemeier *et al.*, 1984a, b). Homologous or heterologous genes are also introduced for (controlled) expression in studies on gene function and application. The larvicidal gene of *Bacillus sphaericus* 1593M was expressed from its own promoter in *Synechococcus* sp. PCC 7942 (cloned in pUC303) for use as a potential biocontrol agent with amplified persistence in mosquito habitats (Tandeau de Marsac *et al.*, 1987). The *Cyanophora paradoxa* allophycocyanin genes carried by a shuttle vector were found to be functionally expressed in *Synechocystis* sp. PCC 7002 from the plastid promoter as well as from the promoter for aminoglycoside phosphotransferase (De Lorimier *et al.*, 1987). The plastocyanin gene from *Anabaena* sp. PCC 7937 was transferred to the genome of *Synechococcus* sp. PCC 7942 by the platform mediated integration system (Chapter V) in order to investigate the expression of this protein and its routing to the thylakoid lumen.

In many studies of this kind the use of strong and regulated promoters will be preferred to enhance expression and reduce potential deleterious effects on the viability of the host. ß-Galactosidase was expressed from the *lacZ* promoter of the *lac* operon present in a shuttle-vector for *Synechococcus* sp. PCC 7002, to an approximately equal level in the cyanobacterium as in *E. coli* (Buzby *et al.*, 1985) but inducibility of the *lac* operon in *Synechococcus* sp. PCC 7002 has not been reported yet. The first functional example of a vector constructed for controlled gene expression is a pUC303 derivative in which the O_LP_L operator-promoter region and the temperature-sensitive repressor gene *cI*857 of phage lambda are incorporated (Friedberg and Seijffers, 1986). Undoubtedly cyanobacterial promoters, like the iron-stress regulated *isiAB* or *irpA* promoters (Laudenbach and Straus, 1988; Reddy *et al.*, 1988) will be used in future constructs.

ad ii. Gene-fusions with an easily assayed product, such as fusions with the "reporter" genes encoding chloramphenicol acetyl transferase (CAT), ß-galactosidase, ß-glucuronidase or luciferase, permit the study of gene control mechanisms like transcription regulation and initiation of protein translation (Shapira *et al.*, 1983; Jefferson *et al.*, 1986; Klotsky and Schwartz, 1987; Boylan *et al.*, 1989). Dzelzkalns *et al.* (1984) found chloroplast *psbA* promoter driven CAT expression in *Synechococcus* sp. PCC

7942 for a plasmid construct based on pUC105 (Kuhlemeier *et al.*, 1981). Schmetterer *et al.* (1986) obtained expression of bacterial luciferase genes after conjugational transfer to filamentous cyanobacteria. After Buzby *et al.* (1985) had shown the potential of *lacZ* as a reporter gene in cyanobacteria, Schaefer and Golden (1989) elegantly applied the technique of plasmid addition for the chromosomal integration of translational *lacZ* fusions to measure the differential expression of the members of the *psbA* gene family from *Synechococcus* sp. PCC 7942 in response to light. This genomic modification can be accomplished without interfering with the functioning of the gene concerned, provided that at least the upstream regions of that gene, promoter and eventual regulatory elements included, are present in the plasmid with the gene fusion. Another feasible approach would be the application of pBR-based gene-fusion vectors for transfer to the recombination target of the genomic integration system described in Chapter IV. In principle, the latter genomic modification by gene replacement has the advantage of being stably maintained even in the absence of antibiotic selection pressure, which is in contrast to the tandem duplication created by plasmid addition (Kolowsky *et al.*, 1984).

ad iii. Introduction of modified genes for site-specific mutagenesis by genomic recombination already proved a very powerful tool in the genetic analysis of the photosynthetic apparatus in *Synechococcus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803 (SS Golden *et al.*, 1987). Two important options are: a) transfer of genes altered by insertion of a selectable marker (e.g. an antibiotic resistance) for analysis of the overall function of a particular gene under certain well-defined conditions and b) introduction of a particular gene with an altered nucleotide sequence in order to study the relationship between structure and function.

A good example of the first type of application, mentioned under a), is the selective inactivation in *Synechococcus* sp. PCC 7942 of two members at a time of the *psbA* multigene family (*psbAI, psbAII, psbAIII*) encoding D1, the herbicide binding component of photosystem II (Golden and Haselkorn, 1985). It showed that each of the genes alone is capable to support normal photoautotrophic growth (Golden *et al.*, 1986). Similar results were obtained with the two *psbD* genes (Golden and Stearns, 1988; Golden *et al.*, 1989b). A Tn*5* insertion in the coding region of the iron-regulated *irpA* gene was transferred to the *Synechococcus* sp. PCC 7942 genome and rendered the resulting mutant unable to grow under iron-deficient conditions (Reddy *et al.*, 1988). Attempts to replace the genes coding for ferredoxin I and thioredoxin *m* by a Km^R marker were unsuccessful, indicating an essential role for these proteins in the photoautotrophic *Synechococcus* sp. PCC 7942 (Van der Plas *et al.*, 1988; Muller and Buchanan; 1989). The particular advantage of *Synechocystis* sp. PCC 6803 is its facultative photoheterotrophy which enables growth on glucose in the absence of functional PS II. Thus in *Synechocystis* sp. PCC 6803 gene

replacement experiments could directly show that the PS II genes *psbB*, *psbE* and *psbF*, encoding CP-47 and the subunits of cytochrome b_{559} respectively, are essential for photoautotrophic growth (Vermaas *et al.*, 1986, 1987; Pakrasi *et al.*, 1988).

In addition to the relatively harsh alterations (interruptions and deletions) described above, it is also possible to introduce genes with subtle modifications, achieved by site-directed mutagenesis, in order to study the relationship between structure and function. A prominent example of this application is found in the identification of the enigmatic `Z' and `D' components of PS II. After indications that Z and D are tyrosine residues and not quinols as previously presumed (Barry and Babcock, 1987). Debus *et al.* (1988a, b) and Vermaas *et al.* (1988) independently proved by site-directed mutagenesis that Z and D indeed were tyrosine residues in the reaction-center proteins D1 (encoded by *psbA*) and D2 (encoded by *psbD*). Tyr-161 residues in D1 and D2 were changed in phenylalanine and the mutations were incorporated into the genome of *Synechococcus* sp. PCC 6803 by gene replacement. The resulting mutants showed defects in accordance with the elimination of Z and D, respectively. Another example of the transfer of a point mutation is the ingenous engineering of herbicide resistance into each of the three *psbA* genes of *Synechococcus* sp. PCC 7942 separately (Brusslan and Haselkorn, 1989). In this way it could be shown that resistance to the PS II herbicide diuron is dominant to sensitivity.

The transformation of cells with alleles of unknown phenotype can be achieved by selection for a linked marker, e.g. an antibiotic resistance gene located downstream of the allele in question (SS Golden *et al.*, 1987; Debus *et al.*, 1988a, b; Brusslan and Haselkorn, 1989). The construction of a restriction site distal from the antibiotic marker by the simultaneous introduction of a silent mutation and the desired point mutation, gives an opportunity to distinguish transformants with the mutated gene from those with the wild-type gene (Debus *et al.*, 1988b).

It is to be expected that the use of both site-directed mutations and gene-fusions will increase considerably in the near future, as a result of the definite facilitation of these approaches by the development of Polymerase Chain Reaction (PCR; Saiki *et al.*, 1988) derived techniques like site-directed mutagenesis by overlap extension (Ho *et al.*, 1989) and gene splicing by overlap extension (Horton *et al.*, 1989). Also the inspection of transformants after introduction of modified genes should be greatly facilitated by PCR, as the presence of specific sequences can easily be determined in very small amounts of DNA and within a few hours.

5. Outline of the thesis

The aim of the investigations presented in this thesis was to clone and analyse cyanobacterial genes for ferredoxin (chapter II) and plastocyanin (chapter III) and to study their regulation. The small metalloproteins ferredoxin and plastocyanin are essential electron carriers in oxygenic photosynthesis, and their presence is regulated by the availability of Fe^{3+} and Cu^{2+}, respectively. In addition efforts were made to forward gene analysis in *Synechococcus* sp. PCC 7942. A genomic integration system based on pBR vector sequences was developed for *Synechococcus* sp. PCC 7942 and applied in the transfer of genes for ferredoxin and plastocyanin (chapter IV). The nucleotide sequence of pUH24, the small cryptic plasmid of *Synechococcus* sp. PCC 7942 was analysed for its genetic content (chapter V), with prospect to improvements in stability and performance of pUH24 derived-plasmid vectors.

References

Aitken A (1975) Prokaryote-eukaryote relationships and the amino-acid sequence of plastocyanin from *Anabaena variabilis*. Biochem J 149:675-683

Alam J, Whitaker RA, Krogmann DW and Curtis SE (1986) Isolation and sequence of the gene for ferredoxin I from the cyanobacterium *Anabaena* sp. Strain PCC7120. J Bacteriol 168:1265-1271

Allen MM (1984) Cyanobacterial cell inclusions. Ann Rev Microbiol 38:1-25

Andriesse AJ, Weisbeek PJ and Van Arkel GA (1989) Biochemistry, regulatory aspects, and genetics of nitrate assimilation in cyanobacteria. In: Wray J and Kinghorn J (eds) Molecular and genetic aspects of nitrate assimilation. pp 40-50. Oxford University Press

Bancroft I and Wolk CP (1988) Pulsed homogeneous orthogonal field gel electrophoresis. Nucl Acids Res 16:7405-7418

Barry BA and Babcock GT (1987) Tyrosine radicals are involved in the photosynthetic oxygen-evolving system. Proc Natl Acad Sci USA 84:7099-7103

Belknap WR and Haselkorn R (1987) Cloning and light regulation of expression of the phycocyanin operon of the cyanobacterium *Anabaena*. EMBO J 6:871-884

Boyer GL, Gillam AH and Trick C (1987) Iron chelation and uptake. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 415-436, Amsterdam: Elsevier

Boylan M, Pelletier J and Meighen EA (1989) Fused bacterial luciferase subunits catalyze light emission in eukaryotes and prokaryotes. J Biol Chem 264:1915-1918

Brawerman, G. (1987) Determinants of Messenger RNA stability. Cell 48:5-6

Brawerman, G. (1989) mRNA decay: finding the right targets. Cell 57:9-10

Bricker TM, Guikema JA, Pakrasi HB and Sherman LA (1986) Proteins of cyanobacterial thylakoids. Encycl Plant Phys 19:640-652

Brusslan J and Haselkorn R (1989) Resistance to the photosystem II herbicide diuron is dominant to sensitivity in the cyanobacteria *Synechococcus* sp. PCC7942. EMBO J 8:1237-1245

Bryant DA, Dubbs JM, Fields PI, Porter RD and de Lorimier R (1985) Expression of phycobiliprotein genes in *Escherichia coli*. FEMS Microbiol Lett 29:343-349

Bryant DA (1987) The cyanobacterial photosynthetic apparatus: comparisons to those of higher plants and photosynthetic bacteria. Can Bull Fish Aquat Sci 214:423-500

Bryant DA, Dubbs JM, Gasparich GE, Gingrich JC, Murphy RC, Rhiel E, Stirewalt VL and Zhou J (1988) In: VI International Symposium on Photosynthetic Prokaryotes, Noordwijkerhout, The Netherlands. Abstract p.93

Bryant DA and Tandeau de Marsac N (1988) Isolation of genes encoding components of photosynthetic apparatus. Methods Enzymol 167:755–765.

Buzby JS, Porter RD and Stevens Jr SE (1985) Expression of the *Escherichia coli lacZ* gene on a plasmid vector in a cyanobacterium. Science 230–805–807

Chassy BM, Mercenier A and Flickinger J (1988) Transformation of bacteria by electroporation. Trends Biotechn 6:303–309

Chauvat F, Rouet P, Bottin H and Boussac A (1989) Mutagenesis by random cloning of an *Escherichia coli* kanamycin resistance gene into the genome of the cyanobacterium *Synechocystis* PCC 6803: selection of mutants defective in photosynthesis. Mol Gen Genet 216:51–5.

Cohn CL, Hermodson MA and Krogmann DW (1989) The amino acid sequence of cytochrome c_{553} from *Microcystis aeruginosa*. Arch Biochem Biophys 270:219–226.

Coulson A and Sulston J (1988) Genome mapping by restriction fingerprinting. In: Davies KE (ed) Genome analysis: a practical approach, pp 19–39. Oxford: IRL Press

Crofts AR and Wood PM (1978) Photosynthetic electron–transport chains of plants and bacteria and their role as proton pumps. Curr Top Bioenerg 7:175–244

Csiszàr K, Houmard J, Damerval T and Tandeau de Marsac N (1987) Transcriptional analysis of the cyanobacterial *gvpABC* operon in differentiated cells: occurrence of an antisense RNA complementary to three overlapping transcripts. Gene 60:29–37

Curtis SE and Haselkorn R (1983) Isolation and sequence of the gene for the large subunit of ribulose–1,5–bisphosphate carboxylase from the cyanobacterium *Anabaena* 7120. Proc Natl Acad Sci USA 80:1835–1839

Daniell H, Sarojini G and McFadden (1986) Transformation of the cyanobacterium *Anacystis nidulans* 6301 with the *Escherichia coli* plasmid pBR322. Proc Natl Acad Sci USA 83:2546–2550

De Boer HA, Hui A, Comstock LJ, Wong E and Vasser M (1983) Portable Shine–Dalgarno regions: a system for a systematic study of defined alterations of nucleotide sequences within the *E. coli* ribosome binding sites. DNA 2:231–235

Debus RJ, Barry BA, Babcock GT and McIntosh L (1988a) Site–directed mutagenesis identifies a tyrosine radical involved in the photosynthetic oxygen evolving system. Proc Natl Acad Sci USA 85:427–430

Debus RJ, Barry BA, Sithole I, Babcock GT and McIntosh L (1988b) Directed mutagenesis indicates that the donor to P_{680}^{+} in photosystem II is tyrosine–161 of the D1 polypeptide. Biochemistry 27:9071–9074.

De Lorimier R, Guglielmi G, Bryant DA and Stevens Jr SE (1987) Functional expression of plastid allophycocyanin genes in a cyanobacterium. J Bacteriol 169:1830–1835

Dzelzkalns VA, Owens GC and Bogorad L (1984) Chloroplast promoter driven expression of the chloramphenicol acetyl transferase gene in a cyanobacterium. Nucl Acids Res 12:8917–8925

Dzelzkalns VA and Bogorad L (1988) Molecular analysis of a mutant defective in photosynthetic oxygen evolution and isolation of a complementing clone by a novel screening procedure. EMBO J 7:333–338

Friedberg D and Seijffers J (1986) Controlled gene expression utilising Lambda phage regulatory signals in a cyanobacterium host. Mol Gen Genet 203:505–510

Gendel SM (1987) Instability of Tn*5* inserts in cyanobacterial cloning vectors. J Bacteriol 169:4426–4430

Geoghegan CM and Houghton JA (1987) Molecular cloning and isolation of a cyanobacterial gene which increases the UV and methylsulphonate survival of *recA* strains of *Escherichia coli* K12. J Gen Microbiol 133:119–126

Glazer AN (1984) Phycobilisome; a macromolecular complex optimized for light energy transfer. Biochem Biophys Acta 768:29–51

Glazer AN (1987) Phycobilisomes: assembly and attachment. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 69–94, Amsterdam: Elsevier

Golden JW, Mulligan ME and Haselkorn R (1987) Different recombination site specifity of two developmentally regulated genome rearrangements. Nature (London) 327:526–529

Golden SS and Haselkorn R (1985) Mutation to herbicide resistance maps within the *psbA* gene of *Anacystis nidulans* R2. Science 229:1104–1107
Golden SS, Brusslan J and Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. EMBO J 5:2789–2798
Golden SS, Brusslan J and Haselkorn R (1987) Genetic engineering of the cyanobacterial chromosome. Methods Enzymol 153:215–231
Golden SS and Stearns GW (1988) Nucleotide sequence and transcript analysis of three photosystem II genes from the cyanobacterium *Synechococcus* sp. PCC 7942. Gene 67:85–96
Golden SS, Nalty MS and Cho DC (1989a) Genetic relationship of two highly studied *Synechococcus* strains designated *Anacystis nidulans*. J Bacteriol 171:24–29
Golden SS, Cho DC and Nalty MS (1989b) Two functional *psbD* genes in the cyanobacterium *Synechococcus* sp. strain PCC 7942. J Bacteriol 171:4707–4713
Golecki JR and Drews G (1982) Supramolecular organization and composition of membranes. In: Carr NG and Whitton BA (eds) The Biology of Cyanobacteria, pp 125–141, Oxford: Blackwell
Hawley DK and McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucl Acids Res 11:2237–2255
Herrero A and Wolk CP (1986) Genetic mapping of the chromosome of the cyanobacterium, *Anabaena variabilis*. J Biol Chem 261:7748–7754
Higgins CF (1986) The regulation of gene expression in *Escherichia coli* and its bacteriophage. In: Rigby PWJ (ed) Genetic Engineering vol 5, pp 2–59, London: Academic Press
Hladík J and Sofrová D (1983) The structure of cyanobacterial thylakoid membranes. Photosynthetica 17:267–288
Ho KK, Ulrich EL, Krogmann DW and Gomez-Lojero C (1979) Isolation of photosynthetic catalysts from cyanobacteria. Biochim Biophys Acta 545:236–248
Ho KK and Krogmann DW (1982) Photosynthesis. In: Carr NG and Whitton BA (eds) The Biology of Cyanobacteria, pp 191–214, Oxford: Blackwell
Ho KK and Krogmann DW (1984) Electron donors to P-700 in cyanobacteria and algae an instance of unusual genetic variability. Biochim Biophys Acta 766:310–316
Ho SN, Hunt HD, Horton RM, Pullen JK and Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59
Horton RM, Hunt HD, Ho SN, Pullen JK and Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77:61–68
Houmard J and Tandeau de Marsac N (1988) Cyanobacterial genetic tools: current status. Meth Enzym 167:808–847
Hui A, Hayflick J, Dinkelspiel and de Boer HA (1984) Mutagenesis of the three bases preceding the start codon of the ß-galactosidase mRNA and its effect on translation in *Escherichia coli*. EMBO J 3:623–629
Hutber GN, Hutson KG and Rogers LJ (1977) Effect of iron deficiency on levels of two ferredoxins and flavodoxin in a cyanobacterium. FEMS Microbiol Lett 1:193–196
Jefferson RA, Burgess SM and Hirsh D (1986) ß-Glucuronidase from *Escherichia coli* as a gene-fusion marker. Proc Natl Acad Sci USA 83:8447–8451
Kalla SR, Lind LK, Lidholm J and Gustafsson P (1988) Transcriptional organization of the phycocyanin subunit gene clusters of the cyanobacterium *Anacystis nidulans* UTEX625. J Bacteriol 170:2961–2970
Klimpel KW, Lesley SA and Clark VL (1989) Identification of subunits of gonococcal RNA polymerase by immunoblot analysis: evidence for multiple sigma factors. J Bacteriol 171:3713–3718
Klotsky RA and Schwartz I (1987) Measurement of *cat* expression from growth-rate-regulated promoters employing ß-lactamase activity as an indicator of plasmid copy number. Gene 55:141–146
Kodaki T, Katagiri F, Asano M, Izui K and Katsuki H (1985) Cloning of Phosphoenolpyruvate carboxylase gene from a cyanobacterium, *Anacystis nidulans*, in *Escherichia coli*. J Biochem 97:533–539

Kohara Y, Akiyama K and Isono K (1987) The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library

Kolowsky KS, Williams JGK and Szalay AA (1984) Length of foreign DNA in chimeric plasmids determines the efficiency of its integration into the chromosome of the cyanobacterium *Synechococcus* R2. Gene 27:289–299

Kolowsky KS and Szalay AA (1986) Double-stranded gap repair in the photosynthetic prokaryote *Synechococcus* R2. Proc Natl Acad Sci USA 83:5578–5582

Kuhlemeier CJ, Borrias WE, Van den Hondel CAMJJ and Van Arkel GA (1981) Construction and characterization of two hybrid plasmids capable of transformation to *Anacystis nidulans* R2 and *Escherichia coli* K12. Mol Gen Genet 184:249–254

Kuhlemeier CJ, Thomas AAM, Van der Ende A, Van Leen RW, Borrias WE, Van den Hondel CAMJJ and Van Arkel GA (1983) A host-vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. Plasmid 10:156–163

Kuhlemeier CJ, Logtenberg T, Stoorvogel W, Van Heugten HAA, Borrias WE and Van Arkel GA (1984a) Cloning of nitrate reductase genes from the cyanobacterium *Anacystis nidulans*. J Bacteriol 159:36–41

Kuhlemeier CJ, Teeuwsen VJP, Janssen MJT and Van Arkel GA (1984b) Cloning of a third nitrate reductase gene from the cyanobacterium *Anacystis nidulans* R-2 using a shuttle cosmid library. Gene 31:109–116

Kuhlemeier CJ, Hardon EM, Van Arkel GA and Van de Vate C (1985) Self-cloning in the cyanobacterium *Anacystis nidulans* R2: fate of a cloned gene after reintroduction. Plasmid 14:200–208

Kuhlemeier CJ and Van Arkel GA (1987) Host-vector systems for gene cloning in cyanobacteria. Methods Enzymol 153:199–215

Kunkel DD (1982) Thylakoid centers: structures associated with the cyanobacterial photosynthetic membrane system. Arch Microbiol 133:97–99

Kuwabara T, Reddy KJ and Sherman LA (1987) Nucleotide sequence of the gene from the cyanobacterium *Anacystis nidulans* R2 encoding the Mn stabilizing protein involved in photosystem II water oxidation. Proc Natl Acad Sci USA 84:8230–8234

Labarre J, Chauvat F and Thuriaux P (1989) Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC 6803. J Bacteriol 171:3449–3457

Laudenbach DE, Straus NA, Gendel S and Williams JP (1983) The large endogenous plasmid of *Anacystis nidulans*: mapping, cloning and localization of the origin of replication. Mol Gen Genet 192:402–407

Laudenbach DE, Straus NA and Williams JP (1985) Evidence for two distinct origins of replication in the large endogenous plasmid of *Anacystis nidulans* R2. Mol Gen Genet 199:300–305

Laudenbach DE, Reith ME and Straus NA (1988) Isolation, sequence analysis, and transcriptional studies of the flavodoxin gene from *Anacystis nidulans* R2. J Bacteriol 170:258–265

Laudenbach DE and Straus NA (1988) Characterization of a cyanobacterial iron stress-induced gene similar to *psbC*. J Bacteriol 170:5019–5026

Lee CC, Kohara Y, Akiyama K, Smith CL, Graigen WJ and Caskey CT (1988) Rapid and precise mapping of the *E. coli* release factor genes by two physical approaches. J Bacteriol 170:4537–4541

Lightfoot DA, Walters DE and Wooton JC (1988) Transformation of the cyanobacterium *Synechococcus* PCC 6301 using cloned DNA. J Gen Microbiol 134:1509–1514

Lind LK, Kalla SR, Lönneborg A, Öquist G and Gustafsson P (1985) Cloning of the ß-phycocyanin gene from *Anacystis nidulans*. FEBS Lett 188:27–32

Lönneborg A, Lind LK, Kalla SR, Gustaffson P and Öquist G (1985) Acclimation processes in the light-harvesting system of the cyanobacterium *Anacystis nidulans* following a light shift from white to red light. Plant Physiol 78:110–114

Looman AC, Bodlaender J, Comstock LJ, Eaton D, Jhurani P, De Boer HA and van Knippenberg PH. Influence of the codon following the AUG initiation codon on the expression of a modified *lacZ* gene in *Escherichia coli*. EMBO J 6:2489–2492

Madueño F, Borrias WE, Van Arkel GA and Guerrero MG (1988) Isolation and characterization of *Anacystis nidulans* R2 mutants affected in nitrate assimilation: establishment of two new mutant types. Mol Gen Genet 213:223–228

Mann N and Carr NG (1974) Control of macromolecular composition and cell division in the blue-green alga *Anacystis nidulans*. J Gen Microbiol 83:399–405

Matsubara H and Hase T (1983) Phylogenetic consideration of ferredoxin sequences in plants, particularly algae. In: Jensen U and Fairbrothers DE (eds) Proteins and nucleic acids in plant systematics, pp 168–181, Berlin: Springer

Mazel D, Guglielmi G, Houmard J, Sidler W, Bryant DA and Tandeau de Marsac N (1988) Green light induces transcription of the phycoerythrin operon in the cyanobacterium *Calothrix* 7601. Nucleic Acids Res 14:8279–8290

Mazur BJ, Rice D and Haselkorn R (1980) Identification of blue-green algal nitrogen fixation genes by using heterologous DNA hybridization probes. Proc Nat Acad Sci USA 77:186–190

Meng BY, Shinozaki K and Sugiura M (1989) Genes for the ribosomal proteins S12 and S7 and elongation factors EF-G and EF-Tu of the cyanobacterium, *Anacystis nidulans*: structural homology between 16S rRNA and S7 mRNA. Mol Gen Genet 216:25–30

Merchant S and Bogorad L (1986a) Regulation by copper of the expression of plastocyanin and cytochrome c_{552} in *Chlamydomonas reinhardi*. Mol Cell Biol 6:462–469

Merchant S and Bogorad L (1986b) Rapid degradation of apoplastocyanin in Cu(II)-deficient *Chlamydomonas reinhardtii*. J Biol Chem 261:15850–15853

Muller EGD and Buchanan BB (1989) Thioredoxin is essential for photosynthetic growth. J Biol Chem 264:4008–4014

Murphy RC, Bryant DA, Porter RD and Tandeau de Marsac N (1987) Molecular cloning and characterization of the *recA* gene from the cyanobacterium *Synechococcus* sp. strain PCC 7002. J Bacteriol 169:2739–2747

Newbury SF, Smith NH, Robinson EC, Hiles ID and Higgins CF (1987) Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48: 297–310

Nierzwicki-Bauer SA, Balkwill DL and Stevens SE jr (1983) Three dimensional ultrastructure of a unicellular cyanobacterium. J Cell Biol 97:713–722

Olins PO, Devine CS, Rangwala SH and Kavka KS (1988) The T7 phage gene *10* leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. Gene 73:227–235

Olsen MK, Rockenbach SK, Curry KA and Tomich CSC (1989) Enhancement of heterologous polypeptide expression by alterations in the ribosome-binding-site sequence. J Biotechn 9:179–190

Owttrim GW and Coleman JR (1987) Molecular cloning of a *recA*-like gene from the cyanobacterium *Anabaena variabilis*. J Bacteriol 169:1824–1829

Pakrasi HB, Williams JGK and Arntzen CJ (1988) Targeted mutagenesis of the *psbE* and *psbF* genes blocks photosynthetic electron transport: evidence for a functional role of cytochrome b_{559} in photosystem II. EMBO J 7:325–332

Parsot C and Mazel D (1987) Cloning and nucleotide sequence of the *thrB* gene from the cyanobacterium Calothrix PCC7601. Mol Microbiol 1:45–52

Porter RD (1985) Transformation in cyanobacteria. Crit Rev Microbiol 13:111–132

Porter RD, Buzby JS, Pilon A, Fields PI, Dubbs JM and Stevens Jr SE (1986) Genes from the cyanobacterium *Agmenellum quadruplicatum* isolated by complementation: characterization and production of merodiploids. Gene 41:249–260

Pozzi G and Guild WR (1985) Modes of integration of heterologous plasmid DNA into the chromosome of *Streptococcus pneumoniae*. J Bacteriol 161:909–912

Raibaud O and Schwartz M (1984) Positive control of transcription initiation in bacteria. Ann Rev Genet 18:173–206

Rebière MC, Castets AM, Houmard J and Tandeau de Marsac N (1986) Plasmid distribution among unicellular and filamentous cyanobacteria: occurence of large and mega-plasmids. FEMS Microbiol Lett 37:269–275

Reddy KJ, Bullerjahn GS, Sherman DM and Sherman LA (1988) Cloning, nucleotide sequence, and mutagenesis of a gene (*irp*A) involved in iron-deficient growth of the cyanobacterium *Synechococcus* sp. Strain PCC7942. J Bacteriol 170:4466–4476

Reddy KJ, Masamoto K, Sherman DM and Sherman LA (1989) DNA sequence and regulation of the gene (*cbpA*) encoding the 42-kilodalton cytoplasmic membrane carotenoprotein of the cyanobacterium *Synechococcus* sp. strain PCC 7942. J Bacteriol 171:3486-3493

Reichelt BY and Delaney SF (1983) The nucleotide sequence for the large subunit of ribulose 1,5-bisphosphate carboxylase from a unicellular cyanobacterium, *Synechococcus* PCC6301. DNA 2:121-129

Reith ME, Laudenbach DE and Straus NA (1986) Isolation and nucleotide sequence analysis of the ferredoxin I gene from the cyanobacterium *Anacystis nidulans* R2. J Bacteriol 168:1319-1324

Riethman H, Bullerjahn G, Reddy KJ and Sherman LA (1988) Regulation of cyanobacterial pigment-protein composition and organization by environmental factors. Photosynth Res 18:133-161

Rippka R, Deruelles J, Waterbury JB, Herdman M and Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol 111:1-61

Rogers LJ (1987) Ferredoxins, flavodoxins and related proteins: structure, function and evolution. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 35-67, Amsterdam: Elsevier

Saiki RK, Gelfland DH, Stoffel S, Scharf Sf, Higuchi R, Horn RT, Mullis KB and Ehrlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491

Sandmann G and Böger P (1980) Copper-induced exchange of plastocyanin and cytochrome c_{553} in cultures of *Anabaena variabilis* and *Plectonema boryanum*. Plant Sci Lett 17:417-424

Sandmann G and Malkin R (1983) Iron-sulfur centers and activities of the photosynthetic electron transport chain in iron-deficient cultures of the blue-green alga *Aphanocapsa*. Plant Physiol 73:724-728

Sandmann G (1986) Formation of plastocyanin and cytochrome c_{553} in different species of blue-green algae. Arch Microbiol 145:76-79

Schaefer MR and Golden SS (1989) Differential expression of members of a cyanobacterial *psbA* gene family in response to light. J Bacteriol 171:3973-3981

Scherer GFE, Walkinshaw MD, Arnott S and Morré DJ (1980) The ribosome binding sites recognized by *E. coli* ribosomes have regions with signal character in both the leader and protein coding segments. Nucl Acids Res 8:3895-3907

Schmetterer G, Wolk CP and Elhai J (1986) Expression of luciferases from *Vibrio harveyi* and *Vibrio fisheri* in filamentous cyanobacteria. J Bacteriol 167:411-414

Schmetterer G and Wolk CP (1988) Identification of the region of cyanobacterial plasmid pDU1 necessary for replication in *Anabaena* sp. strain M-131. Gene 62:101-109

Schneider GJ, Tumer NE, Richaud C, Borbely G and Haselkorn R (1987) Purification and characterization of RNA polymerase from the Cyanobacterium *Anabaena* 7120. J Biol Chem 262:14633-14639

Schneider GJ and Haselkorn R (1988) RNA polymerase subunit homology among cyanobacteria, other eubacteria, and archaebacteria. J Bacteriol 170:4136-4140

Schwartz DC and Cantor CR (1984) Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis. Cell 37:67-75

Shapira SK, Chou JC, Richaud FV and Casabadan MJ (1983) New versatile plasmid vectors for expression of hybrid proteins coded by a cloned gene fused to *lacZ* gene sequences encoding an enzymatically active carboxy-terminal portion of ß-galactosidase. Gene 25:71-82

Sherman L, Bricker T, Guikema J and Pakrasi H (1987) The protein composition of the photosynthetic complexes from the cyanobacterial thylakoid membranes. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 1-33, Amsterdam: Elsevier

Shestakov SV, Khuyen NT (1970) Evidence for genetic transformation in the blue-green alga *Anacystis nidulans*. Mol Gen Genet 107:372-375

Shestakov SV, Reaston J (1987) Gene-transfer and host-vector systems of cyanobacteria. Oxford Surveys of Plant Molec Cell Biol 4:137-166

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

Shinozaki K and Sugiura M (1983) The gene for the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase is located close to the gene for the large subunit in the cyanobacterium *Anacystis nidulans* 6301. Nucl Acids Res 11:6957-6964

Simon RD (1987) Inclusion bodies in the cyanobacteria: cyanophycin, polyphosphate, polyhedral bodies. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 199-226, Amsterdam: Elsevier

Smeekens S, Van Binsbergen J and Weisbeek P (1985) The plant ferredoxin precursor: nucleotide sequence of a full length cDNA clone. Nucl Acids Res 113:3179-3194

Smith CL, Econome JG, Schutt A, Klco SR and Cantor CR (1987) A physical map of the *E. coli* K12 genome. Science 236:1448-1453

Smith CL, Klco SR and Cantor CR (1988) Pulsed-field gel electrophoresis and the technology of large DNA molecules. In: Davies KE (ed) Genome analysis: a practical approach, pp 41-72. IRL Press, Oxford

Staehelin LA (1986) Chloroplast structure and supramolecular organization of photosynthetic membranes. Ency Plant Phys 19:1-84

Stanier RY and Cohen-Bazire G (1977) Phototrophic procaryotes: the cyanobacteria. Ann Rev Microbiol 31:225-274

Stewart AC and Kaethner TM (1983) Extraction and partial purification of an acidic plastocyanin from a blue-green alga. Photobiochem Photobiophys 6:67-73

Stone BB, Boussiba S and Gibson J (1988) Penicillin-binding proteins in the unicellular cyanobacterium *Synechococcus* sp. strain R2 (PCC 7942): implications for membrane identification. J Gen Microbiol 134:2951-2958

Tan S and Ho KK (1989) Purification of an acidic plastocyanin from *Microcystis aeruginosa*. Biochim Biophys Acta 973:111-117

Tandeau de Marsac N (1977) Occurrence and nature of chromatic adaptation in cyanobacteria. J Bacteriol 130:82-91

Tandeau de Marsac N, Borrias WE, Kuhlemeier CJ, Castets AM, Van Arkel GA and Van den Hondel CAMJJ (1982) A new approach for molecular cloning in Cyanobacteria: cloning of an *Anacystis nidulans met* gene using a Tn*901*-induced mutant. Gene 20:111-119

Tandeau de Marsac N (1983) Phycobilisomes and complementary chromatic adaptation in cyanobacteria. Bull Inst Pasteur 81:201-254

Tandeau de Marsac N, de la Torre F and Szulmajster J (1987) Expression of the larvicidal gene of *Bacillus sphaericus* 1593M in the cyanobacterium *Anacystis nidulans* R2. Mol Gen Genet 209:396-398

Tandeau de Marsac N and Houmard J. (1987) Advances in cyanobacterial molecular genetics. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 251-302, Amsterdam: Elsevier

Tandeau de Marsac N, Mazel D, Damerval T, Guglielmi G, Capuano V and Houmard J (1988) Photoregulation of gene expression in the filamentous cyanobacterium *Calothrix* sp. PCC 7601: light-harvesting complexes and cell differentiation. Photosynth Res 18:99-132

Thiel T and Wolk CP (1987) Conjugal transfer of plasmids to cyanobacteria. Methods Enzym 153:232-243

Tomioka N and Sugiura M (1983) The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue-green alga, *Anacystis nidulans*. Mol Gen Genet 191:46-50

Tumer NE, Robinson SJ and Haselkorn R (1983) Different promoters for the *Anabaena* glutamine synthetase gene during growth using molecular or fixed nitrogen. Nature 306, 337-342

Van Baalen C (1987) Nitrogen fixation. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 187-198, Amsterdam: Elsevier

Van den Hondel CAMJJ, Verbeek S, Van der Ende A, Weisbeek P, Borrias WE and Van Arkel GA (1980) Introduction of transposon Tn901 into a plasmid of *Anacystis nidulans*: preparation for cloning in cyanobacteria. Proc Natl Acad Sci USA 77:1570-1574

Van der Plas J, de Groot RP, Weisbeek PJ and van Arkel GA (1986a) Coding sequence of a ferredoxin gene from *Anabaena variabilis* ATCC29413. Nucl Acids Res 14:7803

Van der Plas J, de Groot RP, Woortman MR, Weisbeek PJ and van Arkel GA (1986b) Coding sequence of a ferredoxin gene from *Anacystis nidulans* R2 (*Synechococcus* PCC7942). Nucl Acids Res 14:7804

Van der Plas J, De Groot R, Woortman M, Cremers F, Borrias M, Van Arkel G and Weisbeek P (1988) Genes encoding ferredoxins from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942: structure and regulation. Photosynth Res 18:179–204

Van der Plas J, Bovy A, Kruyt F, De Vrieze G, Dassen E, Klein B and Weisbeek P (1989) The gene for the precursor of plastocyanine from the cyanobacterium *Anabaena* sp. PCC 7937: isolation, sequence and regulation. Mol Microbiol 3:275–284

Vermaas WFJ, Williams JGK, Rutherford AW, Mathis P and Arntzen CJ (1986) Genetically engineered mutant of the cyanobacterium *Synechocystis* 6803 lacks the photosystem II chlorophyll-binding protein CP-47. Proc Natl Acad Sci USA 83:9474–9477

Vermaas WFJ, Williams JGK and Arntzen CJ (1987) Sequencing and modification of *psbB*, the gene encoding the CP-47 protein of photosystem II, in the cyanobacterium *Synechocystis* 6803. Plant Mol Biol 8:317–326

Vermaas WFJ, Rutherford AW, Hansson Ö (1988) Site-directed mutagenesis in photosystem II of cyanobacterium *Synechocystis* sp. PCC 6803: Donor D is a tyrosine residue in the D2 protein. Proc Natl Acad Sci USA 85:8477–8481.

Walsby AE (1987) Mechanisms of buoyancy regulation by planktonic cyanobacteria with gas vesicles. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 377–392, Amsterdam: Elsevier

Watson MEE (1984) Compilation of published signal sequences. Nucl Acids Res 12:5145–5164

Williams JGK and Szalay AA (1983) Stable integration of foreign DNA into the chromosome of the cyanobacterium *Synechococcus* R2. Gene 24:37–51

Wolk CP (1982) Heterocysts. In: Carr NG and Whitton BA (eds) The Biology of Cyanobacteria, pp 191–214, Oxford: Blackwell

Wolk CP, Vonshak A, Kehoe P and Elhai J (1984): Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. Proc Natl Acad Sci USA 81 1561–1565

Wood NB and Haselkorn (1980) Control of phycobiliprotein proteolysis and heterocyst differentiation in *Anabaena*. J Bacteriol 141:1375–1385

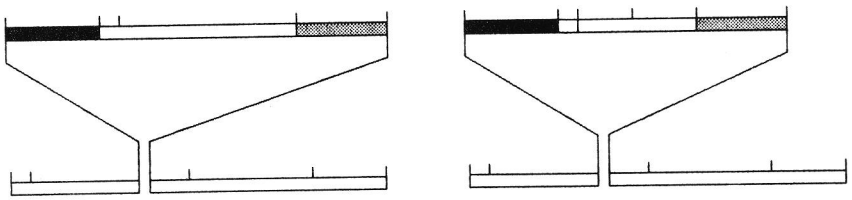
Xie WQ, Jäger K and Potts M (1989) Cyanobacterial RNA Polymerase genes *rpoC1* and *rpoC2* correspond to rpoC of *Escherichia coli*. J Bacteriol 171:1967–1973

Yamanaka G and Glazer AN (1980) Dynamic aspects of phycobilisome structure: phicobilisome turnover during nitrogen starvation in *Synechococcus* sp. Arch Microbiol 124:39–47

Yasui A, Takao, M, Oikawa A, Kiener A, Walsh CT and Eker APM (1988) Cloning and characterization of a photolyase gene from the cyanobacterium *Anacystis nidulans*. Nucl Acids Res 16:4447–4463

Chapter II

Genes encoding ferredoxins from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942: structure and regulation



J. van der Plas, R. de Groot, M. Woortman,
F. Cremers, M. Borrias, G. van Arkel and P. Weisbeek
(1988) Photosynthesis Research 18:179–204

Genes encoding ferredoxins from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942: structure and regulation

Key words: *Anabaena* sp. PCC 7937, cyanobacteria, ferredoxin, gene-expression, iron-regulation, recombination, *Synechococcus* sp. PCC 7942

Abstract. The gene encoding ferredoxin I (*petF1*) from the filamentous cyanobacterium *Anabaena* sp. PCC 7937 (*Anabaena variabilis* ATCC 29413) was cloned by low stringency hybridization with the ferredoxin cDNA from the higher plant *Silene pratensis*. The *petF1* gene from the unicellular cyanobacterium *Synechococcus* sp. PCC 7942 (*Anacystis nidulans* R2) was cloned by low stringency hybridization with the *petF1* gene from *Anabaena* sp. PCC 7937. One copy of the *petF* genes was detected in both organisms, and a single transcript of about 630 b was found for *Synechococcus* sp. PCC 7942. Both the *Synechococcus* sp. PCC 7942 and the *Anabaena* sp. PCC 7937 *petF1* genes contain a 297 bp open reading frame coding for a small acidic protein, consisting of 98 amino-acid residues, with a molecular mass of about 10.5 kDa.

The ferredoxin content of *Synechococcus* sp. PCC 7942 is strongly reduced under iron-limited growth conditions. The slight decrease in the amount of ferredoxin transcript found under iron limitation does not account for the more severe reduction in ferredoxin protein observed. The main regulation of the ferredoxin content probably is effected at the level of translation and/or degradation. Although ferredoxin expression can be strongly reduced by iron stress, the ferredoxin function seems to be indispensable, as *Synechococcus* sp. PCC 7942 appeared refractory to yield mutants lacking the *petF1* gene.

Abbreviations: b – bases, bp – basepair(s), kb – 10^3 basepairs, SSC – standard saline citrate

Introduction

Reduced ferredoxin has a central function in many light-dependent processes in both cyanobacteria and plants. Ferredoxin is the last component of their photosynthetic electron transport chains. It is not only involved in $NADP^+$ photoreduction and cyclic photophosphorylation, but also in such

diverse processes as nitrogen-fixation, nitrate- and nitrite-reduction, glutamate synthesis, and fatty acid metabolism (Hall and Rao 1977, Ho and Krogmann 1982, Smeekens et al. 1985). In addition, ferredoxin is involved in the regulation of many other metabolic activities via the thioredoxin system (Buchanan 1984).

Plant-type ferredoxins are small non-haem iron-sulfur proteins of about 11 kDa, containing one [2Fe–2S] cluster bound by four cysteinyl-sulfur bonds (Hall and Rao 1977). The amino-acid sequences of ferredoxins from many cyanobacteria and plants have been determined and used for the construction of phylogenetic trees (Matsubara and Hase 1983). Eighteen amino-acids were found to be invariant and centered mainly around the four cysteine residues that participate in the [2Fe–2S] cluster binding. Substitutions near these sites are usually equivalent amino-acids, while most non-conservative changes are encountered in other parts of the protein.

Two types of ferredoxin have been found in several plants and in some cyanobacteria (Matsubara and Hase 1983). The differences in primary structure require that there are two distinct structural genes involved. It is most plausible that the individual ferredoxins contribute to different functions within the organism. The biological significance of these differences, however, has not been established yet, with one possible exception. A second ferredoxin has been isolated from heterocysts of *Anabaena* sp. PCC 7937, which seems better suited to act as electron donor to nitrogenase when compared to the ferredoxin isolated from vegetative cells of the same organism (Schrautemeier and Böhme 1985, Böhme and Schrautemeier 1987).

Under conditions of iron limitation, algae and cyanobacteria have been reported to replace ferredoxin by flavodoxin, a flavoprotein (Bothe 1977, Hutber et al. 1977). Cyanobacterial cells from natural blooms frequently do not contain detectable quantities of ferredoxin (Ho et al. 1979). The regulatory processes which cause the shift from ferredoxin synthesis to flavodoxin production are unknown.

We pursued the isolation of the genes for ferredoxin in order to start the investigation both of the dynamic process of ferredoxin/flavodoxin exchange and of the relation between structure and function in ferredoxin. In this paper we report on the isolation and the DNA sequence of a *petF1* gene from the filamentous cyanobacterium *Anabaena* sp. PCC 7937 and of the correspondig gene from *Synechococcus* sp. PCC 7942. The transcription and the regulation of the *Synechococcus* sp. PCC 7942 *petF1* gene were analysed and mutagenesis experiments with recombinational deletion vectors were performed. A preliminary report of the coding sequences of the *petF1* genes has appeared (Van der Plas et al. 1986a,b).

Materials and methods

Materials

Nitrocellulose filter (PH79) and DEAE membrane (NA45) were manufactured by Schleicher & Schuell (Dassel, FRG). Oligonucleotides for probe labeling (random hexamers) were from Pharmacia (Uppsala, Sweden). α^{32}P-dCTP (specific activity 1.1×10^{11} Bq/mol) and α^{35}S-dATP (specific activity 1.85×10^{10} Bq/mol) were purchased from Amersham (Little Chalfont, UK).

Organisms and growth conditions

The cyanobacterial strains *Anabaena* sp. PCC 7937 (*Anabaena variabilis* ATCC 29413) (Duyvesteyn et al. 1983), *Synechococcus* sp. PCC 7942 (in this study the small-plasmid-cured derivative *Anacystis nidulans* R2-SPc has been used; Kuhlemeier et al. 1983), *Synechococcus* sp. PCC 6301, *Synechocystis* sp. PCC 6803, and *Calothrix* sp. PCC 7601 were grown in BG11 medium (i.e. with 25 μmol/l Fe^{3+}) (Rippka et al. 1979). Iron-limited growth conditions were obtained by replacement of ferric ammonium citrate in BG11 medium with equal molar amounts of ammonium citrate. The concentration of iron originating from impurities in the other components of the BG11 medium was less than 0.1 μmol/l. Iron-starved cells were grown for at least 100 generations in BG11 medium from which the Fe^{3+} component was omitted.

Escherichia coli PC 2495, a *recA*$^-$, *hsdS*$^-$ derivative of JM 101 *supE*, *thi*$^-$, Δ(*lac-proAB*), [F', *traD*36, *proAB*, *lacI*^{q}*Z* M15] (Vieira and Messing 1982) constructed by E. Kampert of this department, and *E. coli* HB 101, F$^-$, *hsdS20*, *recA13*, *ara-14*, *proA2*, *lacY1*, *galK2*, *rpsL20*, *xyl-5*, *mtl-1*, *supE44*, λ^- (Maniatis et al. 1982) were grown as described (Maniatis et al 1982).

Western blotting

Collected cyanobacterial cells were suspended to a density of 0.16 g wet weight per cm^3 in 25 mmol/l Tris-HCl, 10 mmol/l EDTA, pH 8.0, and disrupted by sonication for *Synechococcus* sp. PCC 7942, by a combined lysozyme and osmotic shock treatment for *Anabaena* sp. PCC 7937, *Synechococcus* sp. PCC 6301 and *Calothrix* sp. PCC 7601, or by French Press treatment for *Synechocystis* sp. PCC 6803. After removal of the celldebris by centrifugation (10 min, 10.000 × g, 4 °C) an ammonium sulphate fractionation was performed on the supernatant. The 60–100% saturation precipitate was dissolved in 10 mmol/l potassium phosphate buffer pH 7.6 and dialyzed against the same buffer. Protein concentration of the extracts was determined as described (Bradford 1976).

These partially purified protein extracts were used for SDS-PAGE (Laemmli 1970) and electro-blotted onto nitrocellulose filters (Towbin et al. 1979). Alternatively, *Synechococcus* sp. PCC 7942 cells were directly boiled in sample buffer and subjected to SDS-PAGE. The blots were probed with polyclonal rabbit antibodies raised against spinach ferredoxin. Immunoreactions were visualized by incubation with anti-rabbit IgG antibody conjugated to horse radish peroxidase, followed by enzymatic color development through incubation with peroxide and 3,3′,5,5′-tetramethylbenzidine. Lanes containing molecular mass markers were cut from the filters beforehand and stained with amido-black.

Electron microscopic investigation

Synechococcus sp. PCC 7942 mid-exponential phase cells were fixed with 2% paraformaldehyde and 0.5% glutaraldehyde in phosphate buffered saline (PBS). Preparation of ultra-thin cryo sections, immunolabeling and staining of the cryo sections was done as described previously (Van Bergen en Henegouwen and Leunissen 1986). The sections were examined on a Philips EM301 electron miscroscope.

Southern analysis and colony hybridization

Large scale preparations of total DNA from *Synechococcus* sp. PCC 7942 and *Anabaena* sp. PCC 7937 were made according to Mazur et al. (1980) and Curtis and Haselkorn (1983). Small scale isolations were performed as described by Dzelzkalns et al. (1984) with the addition of a lithium chloride precipitation step (Cathala et al. 1983) after the RNase treatment. Restricted DNAs were separated in agarose gels with Tris-borate-EDTA as electrophoresis buffer (Maniatis et al. 1982). Blotting of the separated DNA from the agarose gels to nitrocellulose filters was performed bidirectionally (Meinkoth and Wahl 1984). After baking in vacuo at 80 °C for 30 min the blots were washed at room temperature in 4 × SSC (1 × SSC is 150 mmol/l sodium chloride plus 15 mmol/l trisodium citrate) for 30 min and prehybridized at 50 °C for at least 30 min in a mix of 6.6 × SSC, 10 × Denhardt's solution, 0.1% SDS, 0.05% PPi and 0.1 mg/cm^3 denatured herring sperm DNA (0.25 cm^3 hybridization mix/cm^2 nitrocellulose) (Maniatis et al. 1982, Meinkoth and Wahl 1984). Hybridization was performed by adding denatured probe DNA, labeled to high specific activity ($\geqslant 1.8 \times 10^{16}$ Bq/g DNA) with ^{32}P by nick translation (Maniatis et al. 1982) or by random priming with oligodeoxyribonucleotides (Feinberg and Vogelstein 1983), to the prehybridization medium and continuing the incubation at 50 °C for 16 h. The hybridized blots were washed for 10–30 min at 50 °C, initially three times in 5 × SSC + 0.1% SDS and once in 5 × SSC. Next, the washed

filters were blotted dry, wrapped in plastic foil while still moist and autoradiographed at −20 °C, using X-ray film and intensifying screens (Eastman Kodak, Rochester, NY, USA). If necessary the stringency of hybridization was increased by repeating the washing steps at a higher temperature.

Ordered replicas of a library of 504 cosmid clones of *Synechococcus* sp. PCC 7942 DNA in *E. coli* HB 101, stored in microtiterplates (−70 °C, 15% glycerol), were made on nitrocellulose filters with a stainless steel replicating device and incubated overnight at 37 °C. After colony-filter processing (Maniatis et al. 1982), hybridization was performed exactly as described above for the Southern blots.

DNA cloning

Plasmid DNAs were prepared with the alkaline lysis method (Maniatis et al. 1982) for large scale isolations. 'Minipreps' of plasmid DNA were done by a modified "boiling" procedure (Maniatis et al. 1982): after boiling and spinning the lysates, the supernatant was first extracted with phenol/chloroform, before the DNA was precipitated with 0.6 volumes isopropanol in the presence of 2.5 mol/l ammonium acetate. For cloning or for probe preparation DNA fragments were isolated from agarose gels with DEAE-membrane (Lizardi et al. 1984). Dephosphorylation of linearized vector DNAs with calf intestinal phosphatase (Boehringer, Mannheim, FRG) was carried out 30 min at 37 °C. The phosphatase was inactivated by adding 20 mmol/l EGTA and heating the sample 15 min at 68 °C, followed by phenol /chloroform extraction. DNA ligations were done according to Dugaiczyk et al. (1975). A cosmid library consisting of 504 cosmid clones, was constructed by ligating the 35–55 kb fraction of partially *Sau*3A-cleaved *Synechococcus* sp. PCC 7942 DNA, fractionated on a 10–30% sucrose gradient, to *Bam*HI-linearized and phosphatase-treated DNA of cosmid vector pJB8, followed by in vitro packaging into λ phage particles and transduction to *E. coli* HB 101 (Ish-Horowicz and Burke 1981, Maniatis et al. 1982). Transformation of *E. coli* was performed by the calcium chloride method as described (Maniatis et al. 1982). Transformation of *Synechococcus* sp. PCC 7942 was carried out as described (Van den Hondel et al. 1980), except that after plating the cells, the plates were first incubated in the dark for 16 h.

DNA sequence analysis

DNA sequences were determined by the dideoxy chain termination method of Sanger et al. (1980) with ^{35}S-dATP as radiolabel and with standard primers. Most regions were sequenced several times and in both orientations. Templates for sequencing were single-strand DNAs of in vivo pack-

aged pEMBL8, -9, -18 and -19 clones (Dente et al. 1983). Analysis of the sequences produced was performed with aid of the computer program MICROGENIE (Beckman Instruments, Palo Alto, CA, USA) and of the sequence analysis package of Stephens (1985).

Transcription analysis
RNA was isolated from mid-exponential phase cultures of *Synechococcus* sp. PCC 7942 according to Dzelzkalns (1984) and was analysed on agarose gels run in 25 mmol/l potassium phosphate buffer, after denaturation with glyoxal-DMSO according to Thomas (1983). The RNA was blotted to nitrocellulose (Meinkoth and Wahl 1984) and hybridized with a homologous probe at 65 °C in aqueous medium as described for the Southern blots.

Optical density measurements
The integrated optical density of bands identified on autoradiographs of Northern blots and on prints from Western blots was determined with the IBAS automatic image analysis system (Kontron/Zeiss, FRG). The integrated optical densities were used as parameter for specific mRNA and protein content, respectively.

Results

Homology at the protein level
In Western blots, partially purified ferredoxin protein extracts from *Anabaena* sp. PCC 7937, *Calothrix* sp. PCC 7601, *Synechococcus* sp. PCC 6301, *Synechococcus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803 all gave a single, strongly cross-reacting band with approximately the same electrophoretic mobility as that for spinach ferredoxin when probed with antibodies raised against spinach ferredoxin (results not shown). This result agrees well with the functional exchangeability of ferredoxins in in vitro tests (Hall and Rao 1977, Ho and Krogmann 1982) and with the similarity in ferredoxin amino-acid sequences (Matsubara and Hase 1983). The observed conservation of the plant-type ferredoxin proteins was exploited for the isolation of cyanobacterial genes encoding ferredoxins.

Homology at the DNA level
Total *Anabaena* sp. PCC 7937 DNA was digested with the restriction enzymes *Cla*I, *Eco*RI, *Hin*dIII and *Xba*I, and *Synechococcus* sp. PCC 7942 DNA was digested with *Bam*HI, *Eco*RI, *Hind*III and *Pst*I. The restricted

DNAs were separated on agarose gels and transferred to nitrocellulose filters by Southern blotting. The blots were hybridized with a DNA probe derived from the *petF1* gene of *Silene pratensis* (Smeekens et al. 1985). This probe consisted of a 236 bp *Hin*fI fragment of the coding region for the mature protein and was labeled with ^{32}P by nick translation. The digests of *Anabaena* sp. PCC 7937 DNA gave one major hybridization signal at 55 °C, namely a 5.3 kb *Cla*I, a 20.0 kb *Eco*RI, a 2.7 kb *Hin*dIII, and a 4.9 kb *Xba*I fragment (Fig. 1A).

The blots with the *Synechococcus* sp. PCC 7942 chromosomal digests did not give unique signals. Therefore it was decided to isolate the *petF*-specific DNA from *Anabaena* sp. PCC 7937 first and to use this for the detection of the corresponding gene from *Synechococcus* sp. PCC 7942.

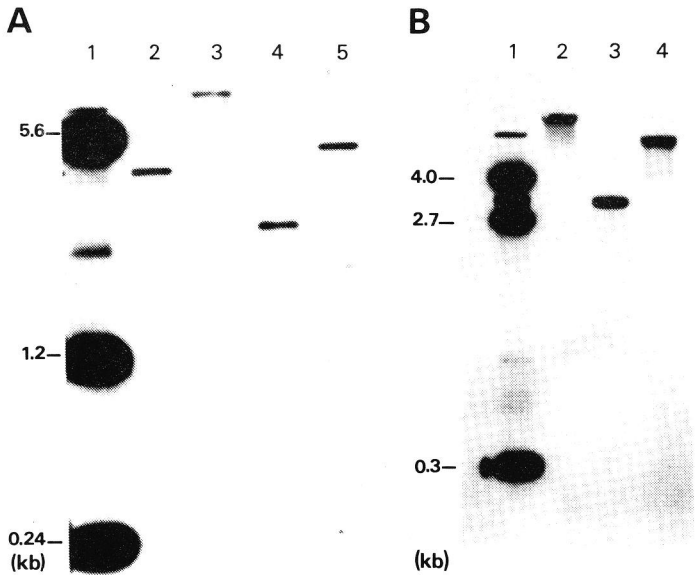


Fig. 1. Chromosomal DNA of *Anabaena* sp. PCC 7937 probed with *S. pratensis petF1* DNA and chromosomal DNA of *A. nidulans* probed with the *petF1* DNA from *Anabaena* sp. PCC 7937. A. Total *Anabaena* sp. PCC 7937 DNA (10 μg) was digested with *Cla*I (lane 2), *Eco*RI (land 3), *Hin*dIII (lane 4) and *Xba*I (lane 5) and run on a 0.8% agarose gel. Lane 1 contained a positive control mixture of three fragments (5.6, 1.2 and 0.24 kb respectively) from the ferredoxin cDNA clone pFD1 of *S. pratensis* (Smeekens et al. 1985). In the amount applied of each of these fragments ca 1 ng *petF1* probe DNA was present. The DNA was probed with the nick-translated 240 bp *Hin*fI fragment from pFD1. B. Total *Synechococcus* sp. PCC 7942 DNA (10 μg) was digested with *Bam*HI (lane 2), *Eco*RI (lane 3) and *Hin*dIII (lane 4), and run on an 0.8% agarose gel. Lane 1 contained a positive control mixture of three fragments (4.0, 2.7 and 0.3 kb respectively) from pVA1 (see Fig. 2). In the amount applied of each of these fragments ca. 1 ng of the *petF1* probe DNA was present. The DNA was probed with the 300 bp *Dra* I fragment from the *Anabaena* sp. PCC 7937 *petF1* gene, labeled by random oligo-priming.

The petF1 *gene from* Anabaena *sp. PCC 7937*

DNA fragments in the range of 4.5–5.5 kb were isolated from an *Xba*I digest of chromosomal *Anabaena* sp. PCC 7937 DNA by preparative agarose gel electrophoresis. These fragments were ligated to *Xba*I-linearized and phosphatase-treated pUC18 (Vieira and Messing 1982) and transformed to *E. coli* PC 2495. The colony library obtained was grown in 64 mixes of 12 clones. Plasmid DNAs of each of these mixed cultures were isolated, digested with *Xba*I, separated on agarose gels and transferred to nitrocellulose filters by Southern blotting. The resulting blots were hybridized with the *S. pratensis petF1* probe at 55 °C. Out of the 64 mixed plasmid preparations 10 were found to contain a 4.9 kb *Xba*I fragment which hybridized to the probe. Of the corresponding mixed cultures one was plated for single colonies. The separate plasmid DNAs subsequently obtained were screened by Southern hybridization analysis. This resulted in the detection of the plasmid pVA1, which contained the 4.9 kb *Xba*I fragment.

This 4.9 kb *Xba*I DNA fragment was analyzed for restriction enzyme cleavage sites. The few enzymes that did cut were *Acc*I, *Bal*I, *Cla*I, *Dra*I, *Eco*RV, *Hinc*II, and *Hind*III. Physical maps of the fragment constructed with these enzymes are shown in Fig. 2. A 1.7 kb *Eco*RV—*Hin*dIII subfragment, hybridizing with the plant *petF1* gene probe, was used to construct a detailed map for the enzymes *Dra*I and *Hinc*II (Fig. 2B). The hybridizing sequence within this fragment was found to be a 300 bp *Dra*I fragment.

The 1.7 kb *Eco*RV—*Hin*dIII fragment and some of its subfragments were cloned into pEMBL plasmid vectors and sequenced. Translation of the

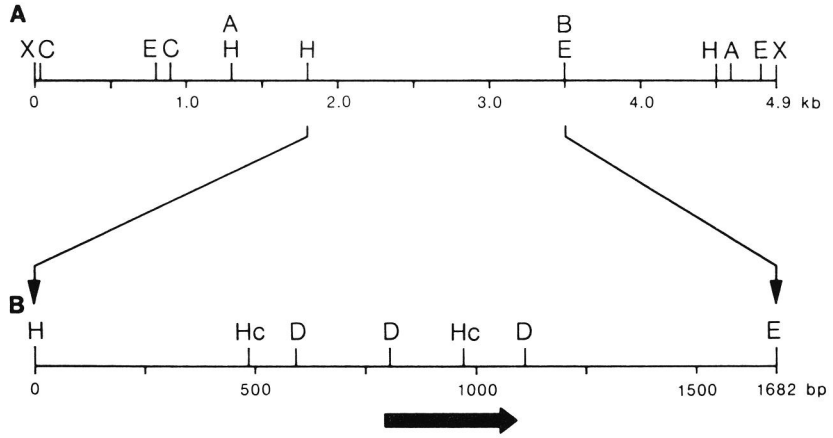


Fig. 2. Restriction maps of the 4.9 kb *Xba*I insert of pVA1 (A) and the sequenced *Eco*RV-*Hin*dIII region (B). The arrow in fig. 2B indicates the *petF1* coding region from *Anabaena* sp. PCC 7937. A = *Acc*I, B = *Bal*I, C = *Cla*I, D = *Dra*I, E = *Eco*RV, H = *Hin*dIII, Hc = *Hinc*II, X = *Xba*I.

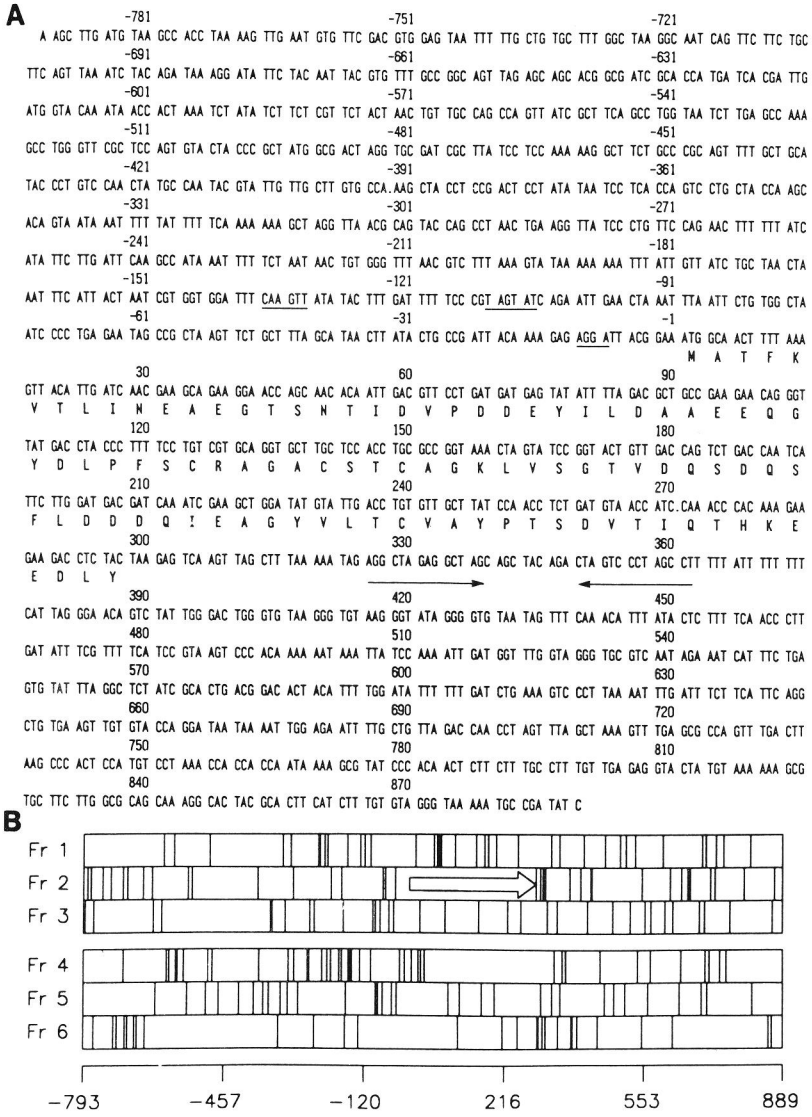


Fig. 3. Nucleotide sequence of the *Anabaena* sp. PCC 7937 *petF1* gene. A. The sequence corresponds to the 1.7 kb *Hin*dIII-*Eco*RV fragment from pVA1. The putative ribosome binding site (AGGA) and putative promoter sequences are underlined (see also Fig. 12). The sequence underlined by arrows, downstream of the coding region and centered at nucleotide 344, indicates a potential stable hairpin structure. B. Scan for stopcodons in the six possible reading frames of the 1682 bp sequence. Numbering of the bp is according to that in A. The arrow in frame 2 indicates the ferredoxin coding sequence.

1.7 kb sequence (Fig. 3A) in all six possible reading frames resulted in the identification of the coding sequence for a plant-type ferredoxin in the second reading frame (Fig. 3B). This open reading frame of 297 bp is preceded by a putative ribosome binding site AGGA (Tomioka and Sugiura 1983) 8 nucleotides upstream of the initiator codon ATG. Downstream of the gene is a palindromic sequence with the potential to form a structure in the mRNA that is similar to a ϱ-independent transcription termination signal of *E. coli* (Rosenberg and Court 1979), or that is involved in mRNA stabilization (Csiszàr et al. 1987). The calculated free energy of formation of this stem-loop structure is -6.9×10^4 J/mol (Tinoco et al. 1973).

The petF1 *gene from* Synechococcus *sp. PCC 7942*

The *petF1* gene from *Anabaena* sp. PCC 7937 was used to isolate the corresponding gene of *Synechococcus* sp. PCC 7942. Southern blots of total *Synechococcus* sp. PCC 7942 DNA digested with *Bam*HI, *Eco*RI and *Hin*dIII were hybridized with a ^{32}P-labeled 300 bp *Dra*I fragment containing almost the entire coding sequence of the *Anabaena* sp. PCC 7937 *petF1* gene (Fig. 2B), in order to investigate the DNA sequence similarity of the *petF1* genes. At 60 °C one major hybridizing fragment was found with the *Bam*HI, *Eco*RI and *Hin*dIII digests, with sizes of 11, 3.2 and 6.5 kb respectively (Fig. 1B). Subsequently, nitrocellulose replica filters of a *Synechococcus* sp. PCC 7942 cosmid library were screened with the same probe. One colony gave a strong and reproducible hybridization reaction. It contained a cosmid, designated pA612, of about 50 kb in size. Restriction and hybridization

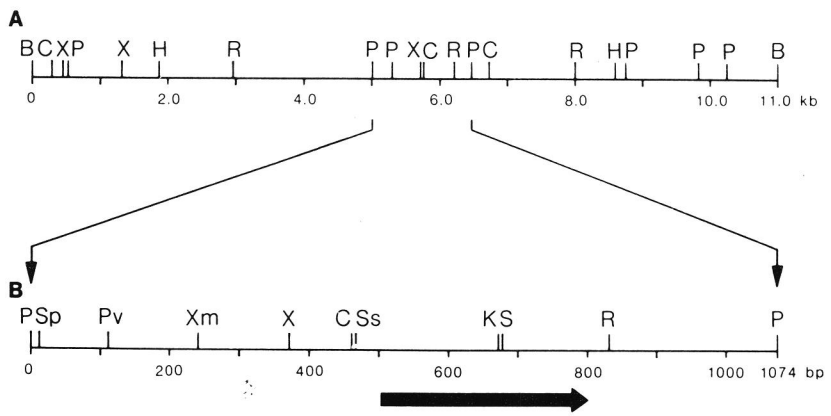


Fig. 4. Restriction maps of the 11 kb *Bam*HI fragment (A) and of the 1.1 *Pst*I fragment from cosmid pA612 (B). The arrow in Fig. 4b indicates the ferredoxin coding region from *Synechococcus* sp. PCC 7942. B = *Bam*HI, C = *Cla*I, H = *Hin*dIII, K = *Kpn*I, P = *Pst*I, Pv = *Pvu*II, R = *Eco*RI, S = *Sal*I, Sp = *Sph*I, Ss = *Sst*I, X = *Xba*I, Xm = *Xma*III.

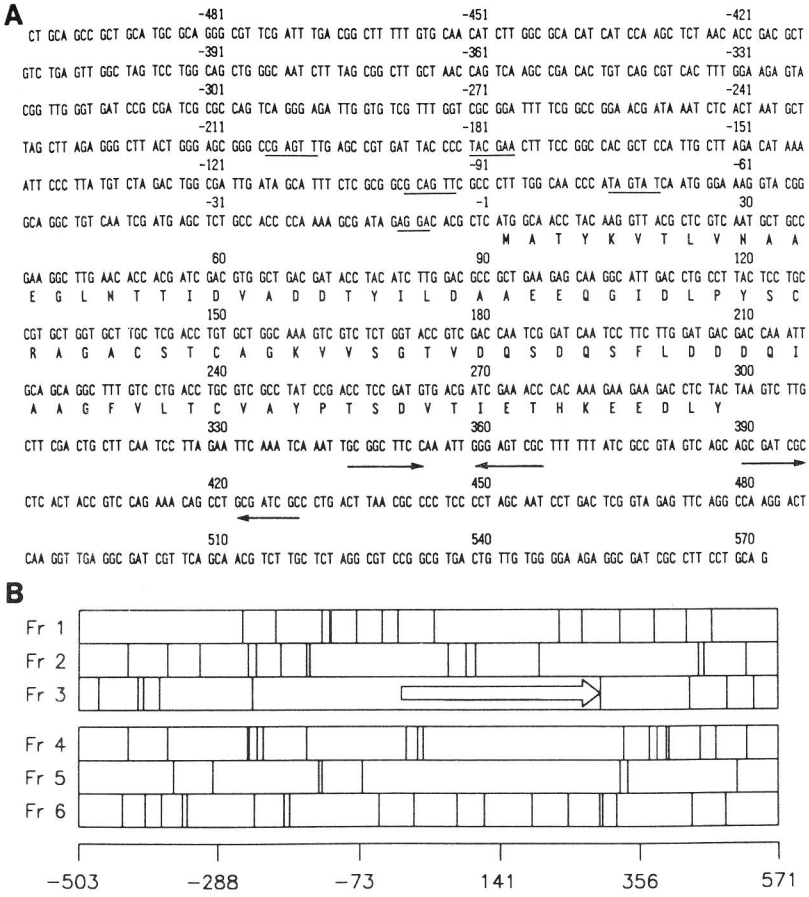


Fig. 5. Nucleotide sequence of the *Synechococcus* sp. PCC 7942 *petF1*gene. A. The sequence of the 1.1 kb *Pst*I fragment from pA612 is given. The ribosome binding site (AGGA) and putative promoter sequences for 'P1' and 'P2' (see also Fig. 12) are underlined. The sequences underlined by arrows, downstream of the coding region and centered at nucleotides 355 and 409, respectively, indicate potential stable hairpin structures. B. Scan for stopcodons in the six possible reading frames of the 1074 bp sequence. Numbering of the bp is according to that in A. The *petF1* coding sequence is indicated by the arrow in reading frame 3.

analysis of pA612 showed hybridizing *Bam*HI, *Eco*RI and *Hin*dIII fragments (data not shown) of the same size as those reacting in Southern blots of chromosomal DNA of *Synechococcus* sp. PCC 7942 (Fig. 1B). The 11 kb *Bam*HI fragment was subcloned in pUC18 and subsequently mapped (Fig. 4A). The 1.1 kb *Pst*I fragment that contains the hybridizing sequence was further subcloned in pEMBL vectors for mapping (Fig. 4B) and sequencing (Fig. 5A).

Translation of the 1.1 kb sequence (Fig. 5A) in all six possible reading

frames resulted in the identification of the coding sequence for a plant-type ferredoxin in the third reading frame (Fig. 5B). This open reading frame of 297 bp is preceded by a putative ribosome binding site AGGA (Tomioka and Sugiura 1983) 7 nucleotides upstream of the initiator codon ATG. The non-coding region following the gene contains two sequences (see Fig. 5A) able to form base-paired structures in mRNA with calculated free energies

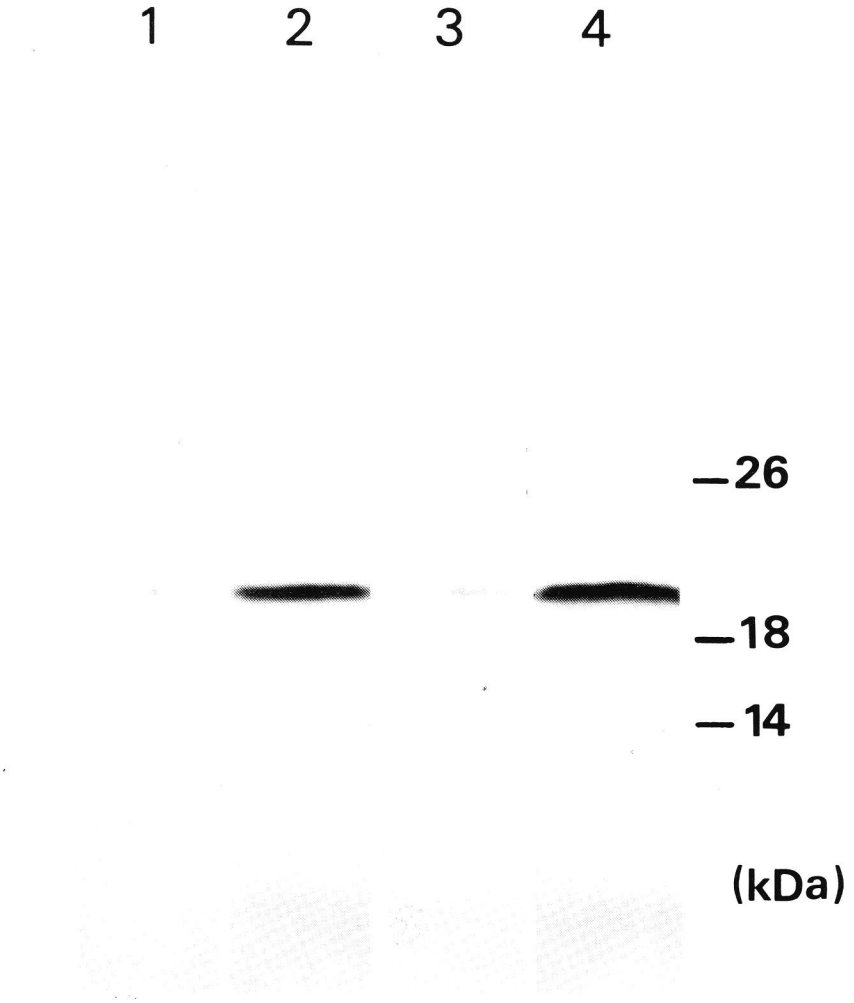


Fig. 6. Protein analysis of *Synechococcus* sp. PCC 7942 grown with or without Fe^{3+}. Cells were harvested by centrifugation, boiled in sample buffer and used for SDS-PAGE. After electrophoretic transfer to nitrocellulose, the proteins were incubated with anti-ferredoxin antibodies. Antibody reaction was visualized via horse radish peroxidase coupled second antibodies. Lanes 1 and 3: cells grown in BG11 minus Fe^{3+}. Lanes 2 and 4: cells grown under standard conditions with complete BG11 medium. In lanes 1 and 2: 5.10^7 cells and in lanes 3 and 4 10^8 cells were used, respectively.

of formation (Tinoco et al. 1973) of -4.4×10^4 and $-6.9 \times {}^4$ J/mol, respectively.

Regulation of gene expression

Iron limitation is known to induce in cyanobacteria a shift from the synthesis of ferredoxin to the production of flavodoxin (Bothe 1977, Hutber et al. 1977). To determine the level at which this regulation occurs, both the ferredoxin protein content and the level of mRNA encoding ferredoxin were studied.

The relative amounts of ferredoxin protein present in *Synechococcus* sp. PCC 7942 cells grown either under iron limitation or under standard conditions were measured immunologically. In Western blot analysis (Fig. 6) the samples from cells grown in the presence of iron reproducibly gave signals approximately 30 times stronger than the samples from iron-limited cultures, as determined by optical density measurement. As an alternative approach, electron microscopic analysis of cells from the two types of cultures described above was performed. Cryo sections were incubated with ferredoxin antibodies, after which antibody-antigen complexes were made visible under the electron microscope with protein A-gold (Van Bergen en Henegouwen and Leunissen 1986). Gold-labeling is found almost exclusively in the cytoplasm (Fig. 7). From this observation in can be concluded that ferredoxin is localized throughout the cytoplasm, rather than associated with thylakoid or cytoplasmic membranes. The gold-labeling of cells from iron-limited cultures when compared to that in cells from complete medium, is reduced to the same degree as the corresponding signal in the Western blots (Fig. 6).

The specific *petF1* mRNA levels were determined by Northern analysis. Total RNA was extracted from *Synechococcus* sp. PCC 7942 cells grown with and without Fe^{3+} added to the culture medium. Northern blots of total RNA hybridized with the 360 bp *Sst*I-*Eco*RI *petF1* gene fragment from *Synechococcus* sp. PCC 7942 as probe, showed a single signal, corresponding to a transcript of about 630 b (Fig. 8). The hybridization signal found with total RNA from iron-limited cells was 2–3 times weaker than obtained with the same amount of total RNA extracted from cells grown in standard medium.

In conclusion, the slight decrease in the amount of *petF1* transcript found under iron limitation does not account for the significantly stronger reduction in ferredoxin protein content observed.

Attempted deletion of the petF1 *gene*

The *petF1* gene from *Synechococcus* sp. PCC 7942 was used in mutagenesis

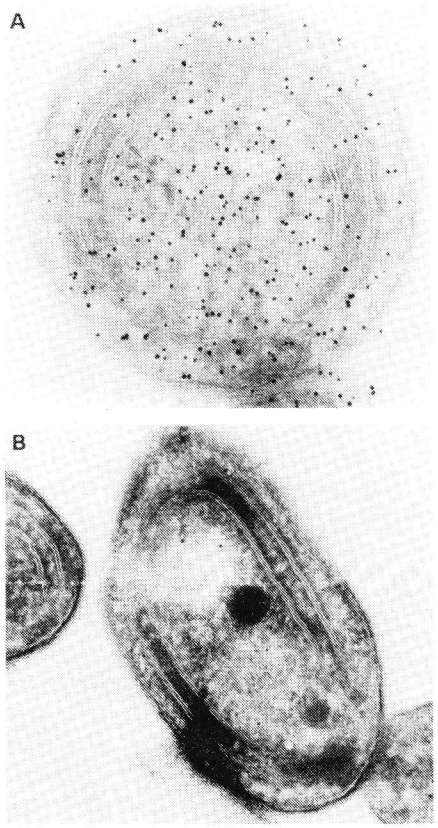


Fig. 7. Electron microscopic analysis of *Synechococcus* sp. PCC 7942 cells grown with or without Fe^{3+}. Cryo sections of cells, grown in A: BG11 medium and B: BG11 medium without iron, were incubated with anti-ferredoxin antibodies, followed by incubation with protein A-gold for electron microscopic investigation.

experiments, in order to investigate whether this organism can grow in the absence of ferredoxin and, if so, to create concomitantly a host-strain for studies on ferredoxin function. For the construction of a *petF1* deletion mutant by exploiting homologous recombination, a plasmid was made (Fig. 9A) in which the *petF1* coding sequence was replaced by a fragment from pUC4-K bearing the *neo* gene (Vieira and Messing 1982). The vector does not replicate in *Synechococcus* sp. PCC 7942 and stable kanamycin-resistant cells can only be formed when the vector or part of it integrates into the chromosome.

Before transformation, the plasmid pFDdel II (Fig. 9A) was linearized in the vector sequences with *Sca*I, in order to reduce the frequency of chro-

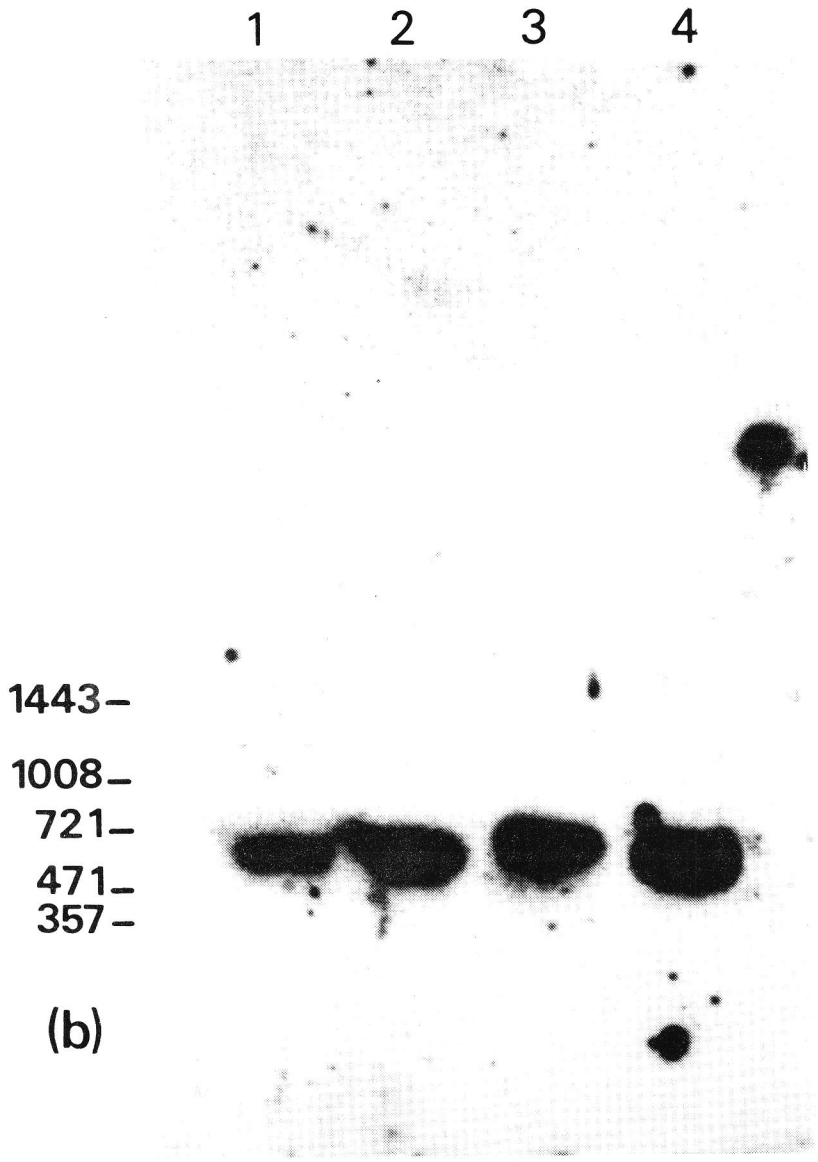


Fig. 8. Transcription analysis of *Synechococcus* sp. PCC 7942 cells grown with or without Fe^{3+}. Total RNA was separated on an 1.1% agarose gel, blotted and hybridized with an *Synechococcus* sp. PCC 7942 *petF1* DNA probe, consisting of the radioactively labeled 360 bp *Sst*I-*Eco*RI fragment. pEMBL8-*Taq*I DNA fragments were denatured and used as size markers. Lanes 1 and 3: RNA from cells grown inBG11 medium minus FE^{3+}. Lanes 2 and 4: RNA from cells grown in complete BG11 medium. In lanes 1 and 2 5 μg and in lanes 3 and 4 10 μg total RNA was loaded onto the gel

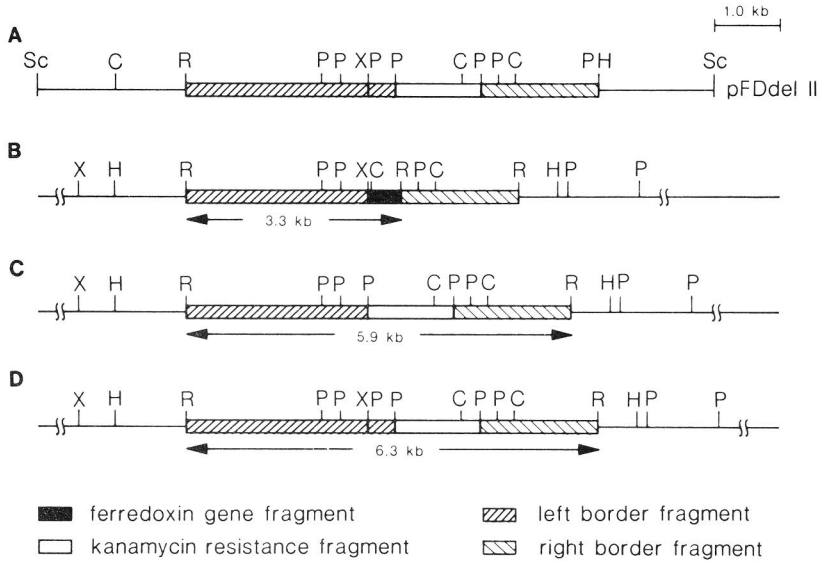


Fig. 9. Plasmid vector pFDdel II for mutagenesis of the *Synechococcus* sp. PCC 7942 *petF1* gene. Fragments flanking the *petF1* gene were ligated to the *neo* gene fragment from pUC4-K and inserted into the polylinker of pEMBL19. Details of the final construct pFDdel II (10.3 kb) are shown in Fig. 9A as linearized by *Sca*I in the *bla* gene of the pEMBL19 part. Note that the left border has a small duplication (it contains two copies of the 0.37 kb *Pst*I-*Xba*I fragment in direct repeat) and that the *Eco*RI site delimiting the *petF1* gene fragment to be deleted in vivo, has been removed. For comparison the original situation for the ferredoxin gene in the chromosome of *Synechococcus* sp. PCC 7942 is drawn in fig. 9B. The results expected from recombination between pFDdel II and the *Synechococcus* sp. PCC 7942 genome are drawn in Figs. 9C and 9D. C = *Cla*I, H = *Hin*dIII, Sc = *Sca*I, R = *Eco*RI, P = *Pst*I, X = *Xba*I.

mosomal integration of the complete plasmid by single-crossover recombination events. Both *Synechococcus* sp. PCC 7942 cells grown under iron-limitation and grown under standard conditions were transformed. Selection for kanamycin resistance (7.5 $\mu g/cm^3$) was applied by adding the antibiotic underneath the agar after overnight preincubation. With both types of cells rather large numbers of very tiny, slowly growing colonies (transformation frequency: 10^{-3}–$10^{-4}/\mu g$ DNA), as well as a few larger colonies (transformation frequency: 10^{-6}–$10^{-7}/\mu g$ DNA) were detected. The cells with small-colony phenotype could not be subcultured further under kanamycin pressure together with iron limitation.

Chromosomal DNA was extracted from wildtype *Synechococcus* sp. PCC 7942 and from a limited number of the larger type transformants. The chromosomal DNAs were digested with *Eco*RI, separated on an agarose gel and blotted to two nitrocellulose filters (Fig. 10). One filter was probed with

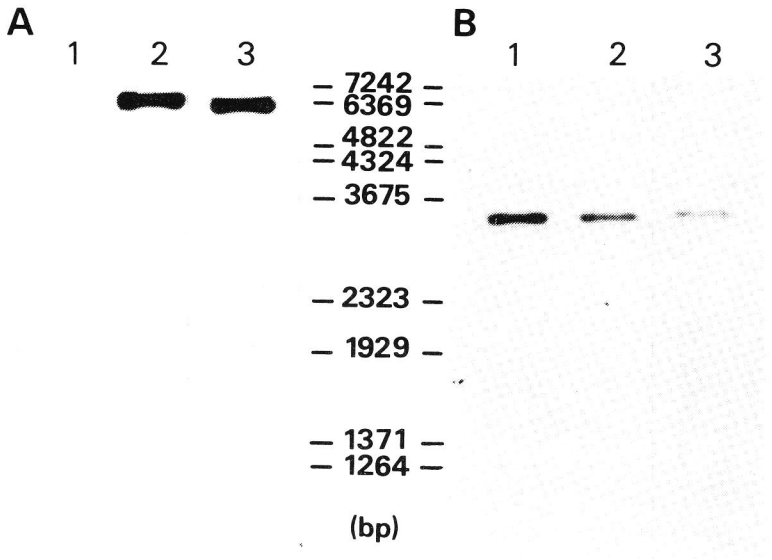


Fig. 10. Southern hybridization analysis of kanamycin-resistant *Synechococcus* sp. PCC 7942 recombinants produced by transformation with pFDdelII. Southern blots of chromosomal DNA (10 μg) from wildtype *Synechococcus* sp. PCC 7942 (lane 1) and from two different types of kanamycin-resistant recombinants (lanes 2 and 3) are shown. The blot in panel A was probed with a radioactively labeled *neo* gene fragment from PUC4-K and the blot in panel B was probed with the radioactively labeled 360 bp *Sst*I-*Eco*RI fragment from the *Synechococcus* sp. PCC 7942 *petF1* gene. the size markers indicated are from a *Bst*EII digest of phage lambda DNA.

a radioactively labeled *neo* gene fragment. The other filter was probed with the radioactively labeled 360 bp *Sst*I-*Eco*RI fragment, which is completely contained within the 460 bp *Xba*I-*Eco*-RI fragment to be deleted from the chromosome. The DNAs from all the transformants hybridized with the *neo* gene probe on one of the two expected fragments of 6.3 kb and 5.9 kb. Since the left border of pFDdel II has a small duplication of th 0.37 kb *Pst*I-*Xba*I fragment, recombination can take place in or adjacent to this duplication (Figs. 9C and 9D). The *neo* gene therefore must have been integrated correctly. Nevertheless, all strains still contained the 3.3 kb wildtype *petF1* gene fragment as in the wildtype control (Fig. 10B). So the transformants, which were colony-purified before further analysis, were heterozygous with respect to the *petF1* region. It is likely that segregation of the *petF1* gene causes loss of viability.

Discussion

The petF1 *gene from* Anabaena *sp. PCC 7937*

The *petF1* gene from the filamentous cyanobacterium *Anabaena* sp. PCC 7937 was cloned by low-stringency hybridization with the *petF1* cDNA from the higher plant *S. pratensis* (Smeekens et al. 1985). The DNA similarity between the coding region for the *Anabaena* sp. PCC 7937 ferredoxin and the coding region for the mature *S. pratensis* ferredoxin is rather homogeneously spread (the longest stretch of identical nucleotides amounts 11). Alignment of the deduced *Anabaena* sp. PCC 7937 ferredoxin amino-acid sequence with that of the *S. pratensis* mature ferredoxin (Fig. 11), gives an amino-acid similarity of 71%. This is slightly higher than the DNA similarity of 65%, because 50 out of the 100 base changes are silent.

The *Anabaena* sp. PCC 7937 *petF1* gene codes for a protein with a molecular mass of 10.6 kDa, consisting of 98 amino-acid residues when the first methionine is excluded. This methionine will most likely be processed, for all cyanobacterial ferredoxin amino-acid sequences determined begin with alanine (Matsubara and Hase 1983). The observed molecular mass of ferredoxin (Fig. 6) is higher than the calculated one because of the anoma-

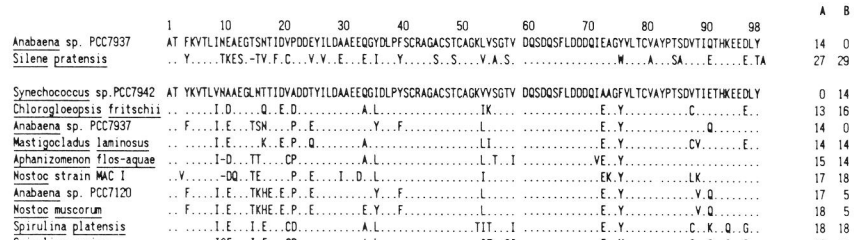


Fig. 11. Comparison of ferredoxin amino-acid sequences. The ferredoxin amino-acid sequences for *Synechococcus* sp. PCC 7942 and *Anabaena* sp. PCC 7937 are shown. The sequence of *Anabaena* sp. PCC 7937 is aligned to that of *S. pratensis* (Smeekens et al. 1985) and the differences are indicated. the sequence of *Synechococcus* sp. PCC 7942 is compared with the sequences known from other cyanobacteria (Alam et al. 1986, Chan et al. 1983, Cozens and Walker 1987, Matsubara and Hase 1983). For each ferredoxin the number of amino-acid residues that differ from *Synechococcus* sp. PCC 7942 ferredoxin (column A) and from *Anabaena* sp. PCC 7937 ferredoxin (column B) are given. At the bottom the 24 residues shared by all cyanobacterial ferredoxins are indicated. Cysteine residues involved in [2Fe-2S]-cluster formation are nos 41, 46, 49 and 79.

lous electrophoretic mobility of ferredoxin in this gel system. This effect might be produced by dimerization of ferredoxin (Böhme and Schrautemeier 1987), or by the highly negative charge of the protein, which prevents proper binding of SDS (Smeekens et al. 1985), or by both. The protein deduced from the DNA sequence contains 4 cysteinyl residues and their positions correspond to those in the other ferredoxin sequences (Fig. 11). Besides the 4 cysteinyl residues, which are required for the chelation of the two iron atoms of the [2Fe–2S] cluster, 20 amino-acid residues are conserved in all the known ferredoxins of cyanobacterial origin. The *Anabaena* sp. PCC 7937 ferredoxin, like all other known plant-type ferredoxins, has an acidic character, as predicted by the high content of aspartic acid and glutamic acid residues (20x) relative to that of arginine and lysine residues (4x). The amino-acid sequence of the ferredoxin from *Anabaena* sp. PCC 7937 is most similar to the three almost identical amino-acids sequences of the ferredoxins from *Nostoc muscorum* (Matsubara and Hase 1983), from an unspecified *Anabaena variabilis* strain (Chan et al. 1983) and of the ferredoxin encoded by the *petF1* gene of *Anabaena* sp. PCC 7120 (Alam et al. 1986) (Fig. 11).

In many plants and some cyanobacteria two types of ferredoxin are found, differing in amino-acid sequence and redox potential (Matsubara and Hase 1983). By convention the type I designation is given to the major ferredoxin component in the cell. When the amino-acid sequence of the *Anabaena* sp. PCC 7937 ferredoxin was compared with those of the type I and II ferredoxins from *Aphanothece sacrum* and *Nostoc* sp. strain MAC (Matsubara and Hase 1983), a higher degree of similarity to the type I sequences was found. Therefore the *Anabaena* sp. PCC 7937 ferredoxin can be assigned to the type I ferredoxins.

The *Anabaena* sp. PCC 7937 *petF1* gene probably encodes the ferredoxin found in vegetative cells of this strain, because the deduced amino-acid composition agrees almost completely with that determined for the purified protein (Böhme and Schrautemeier 1987). This protein is also found in heterocysts of this strain as the minor ferredoxin species. The amino-acid composition of the major ferredoxin from heterocysts, thought to be specialized in electron transfer to nitrogenase (Schrautemeier and Böhme 1985), appears to be quite different (Böhme and Schrautemeier 1987).

In the *Anabaena* sp. PCC 7937 genome *petF1* is probably present as a single-copy gene, as only one major hybridization signal was found, both with the plant probe and with the *petF1* gene itself as probe (results not shown). The genome of *Anabaena* sp. PCC 7937, however, is expected to possess at least two different *petF* genes, as suggested by the presence of two clearly different ferredoxin proteins in the heterocysts (Schrautemeier and

Böhme 1985, Böhme and Schrautemeier 1987). The second ferredoxin most likely is so divergent that a corresponding second hybridization signal can not readily be detected under the conditions applied. The *petF1* gene probably is not part of an operon, as directly adjacent to the *petF1* reading frame no other open reading frames were found (Fig. 3).

Alignment of the DNA sequence of the *petF1* gene and some flanking sequences (790 bp in total) from *Anabaena* sp. PCC 7937 with that from *Anabaena* sp. PCC 7120 (Alam et al. 1986) resulted in an overall similarity of 92% (93% for the coding region alone). Hence even the non-coding flanking sequences are strongly conserved. Further sequence comparisons might reveal whether the high similarity found is incidental, or whether it is indicative of close strain relatedness despite distinct differences in restriction endonuclease content (Duyvesteyn et al. 1983) and in heterotrophic growth capacity (Wolk and Schafer 1976, Rippka et al. 1979).

The petF1 *gene from* Synechococcus *sp. PCC 7942*

The *petF1* gene from the unicellular cyanobacterium *Synechococcus* sp. PCC 7942 was cloned by low-stringency hybridization with the *petF1* DNA from *Anabaena* sp. PCC 7937. The DNA similarity between the coding regions of the *petF1* genes of the two organisms is 76%. The 3′ part is somewhat more conserved, with a continuous stretch of 29 identical nucleotides (including the TAA stopcodon). Outside the coding regions sequence similarity is insignificant. Alignment of the deduced *Synechococcus* sp. PCC 7942 ferredoxin amino-acid sequence with that of *Anabaena* sp. PCC 7937 ferredoxin (Fig. 11) reveals 14 differences, 5 of which are exchanges for equivalent amino-acids. The amino-acid similarity of 86% is much higher than the DNA homology (76%), because 49 out of 72 base changes are silent.

The *Synechococcus* sp. PCC 7942 *petF1* gene codes for a small acidic protein with a molecular mass of 10.4 kDa, consisting of 98 amino-acid residues, the first methionine being excluded. It also contains 4 cysteinyl residues, in positions corresponding to those in the other ferredoxin sequences shown (Fig. 11). Its acidic character is inferred from the high content of aspartic acid and glutamic acid residues (18x) relative to arginine and lysine residues (4x). The codon usage resembles that seen in genes already sequenced from this organism (Tandeau de Marsac and Houmard 1987).

The *Synechococcus* sp. PCC 7942 ferredoxin must be assigned to the type I ferredoxins for the same reason as discussed for the *Anabaena* sp. PCC 7937 ferredoxin. Remarkably the *Synechococcus* sp. PCC 7942 ferredoxin shows much greater similarity with the type I ferredoxins of the filamentous cyanobacteria than with the type I ferredoxins of the unicellular strains *Aphanothece sacrum*, *Aphanothece halophitica*, *Synechocystis* sp. PCC 6714

and *Synechococcus* sp. (Fig. 11). This reflects the observed heterogeneity in the group of unicellular cyanobacteria (Rippka et al. 1979).

For two reasons *Synechococcus* sp. PCC 7942 is now expected to possess a second *petF* gene. The detection of a second, minor, ferredoxin component for *Anacystis nidulans* has been reported (Sakihama and Shin 1987), and a coding sequence for another [2Fe–2S] ferredoxin has been identified in *Synechococcus* sp. PCC 6301 (Cozens and Walker 1987), a strain very closely related to *Synechococcus* sp. PCC 7942. The latter gene, of which the expression is unknown, is located near the *atpC* gene and is clearly different from *petF1* described by Reith et al. (1986) and by us (Van der Plas et al. 1986b; this paper) (Fig. 11). The similarity in amino-acid sequence between the gene products of this second *petF* gene and the *petF1* gene amounts to only 42% (the DNA similarity is 57% at the most). The deduced amino-acid sequence also differs considerably from that of the type II ferredoxins (Fig. 11). In spite of these indications for a second *petF* gene, only one major hybridization signal was found with the *Anabaena* sp. PCC 7937 *petF1* probe. Using the *Synechococcus* sp. PCC 7942 *petF1* gene itself as probe for hybridization at low stringency (45 °C) with total *Synechococcus* sp. PCC 7942 DNA digested with *Eco*RI and *Hin*dIII, additional signals of 4.9 and 2.5 kb, respectively, were observed (results not shown).

Directly adjacent to the *petF1* reading frame no other open reading frames are found (Fig. 5). Together with the detection of a single messenger for ferredoxin with a length 630 b, this leads to the conclusion that the *petF1* gene is not part of an operon. This length of the *petF1* specific messenger is found reproducibly for *Synechococcus* sp. PCC 7942 under our conditions. In Fig. 8 pEMBL8-*Taq*I fragments were used as size markers, but the same transcript length was determined in Northern blots with the size markers phage ØX174 RF-DNA fragments, rRNAs, or RNA transcripts made in vitro with SP6 polymerase (data not given). This mRNA size differs significantly from the 430 b determined for *Synechococcus* sp. PCC 7942 by Reith et al. (1986). As the DNA sequence they reported is almost identical to ours (we find one extra C at position 396 and one extra G at position 553), the discrepancy in size can not be explained just by strain differences. Little is known about the transcription start and termintion processes in cyanobacteria. A mRNA of 430 b would fit a transcription start at position −64, as determined by S1-mapping (Reith et al. 1986), and a transcription stop directly after the first putative terminator structure downstream from the stopcodon. Directly in front of the start site the sequence TAGTAT is found, resembling the *E. coli* consensus −10 region (Rosenberg and Court 1979), but the −35 region shows no familiar consensus sequence (fig. 12. 'P2'). This presumable *Synechococcus* sp. PCC 7942 promoter has consider-

able homology with the putative promoter region of *Anabaena* sp. PCC 7120 *petF1* (Alam et al. 1986) (Fig. 12). The corresponding sequence is also conserved in *Anabaena* sp. PCC 7937 (Figs 3 and 12). However, at position −178 of the *Synechococcus* sp. PCC 7942 *petF1* sequence another possible promoter region was found (Fig. 5A and Fig. 12. 'P1'), which again resembles the *Anabaena* sequences. The −35 region of P1 matches better than that of P2, while the −10 region of P2 fits better than that of P1. Assuming that transcription starts from P1 at about −170, transcription termination after the second, more stable, palindromic structure would result in a messenger of 600 b, in agreement with our transcript length of about 630 b. It can not be completely ruled out, however, that the difference in size observed for the *petF1* transcripts corresponds only to different transcription stops.

The main regulation of the cell's ferredoxin content by the availability of Fe^{3+} most probably occurs at the level of translation and/or degradation. The reduction by a factor of 2–3 in the amount of *petF1* transcript found during growth under iron limitation (Fig. 8) is not sufficient to explain the 30-fold decrease in protein content (Figs 6 and 7). The small decrease in transcript level, as related to the amount of total RNA (mainly rRNA), might be caused by a small general reduction in transcript levels. As the cells remain viable and keep growing at an almost unchanged rate under conditions of iron limitation, the cells must be able to compensate for the drastically lowered ferredoxin levels. Flavodoxin is reported to take over the function of ferredoxin under such circumstances (Bothe 1977, Hutber et al. 1977, Ho and Krogmann 1982). However, it has not been established whether flavodoxin can fully replace ferredoxin in vivo. The presence of ferredoxin, even though at a very reduced level, might be essential for certain specific functions.

Indispensability of the petF1 *gene*

The data from the deletion mutagenesis experiments suggest that a cell

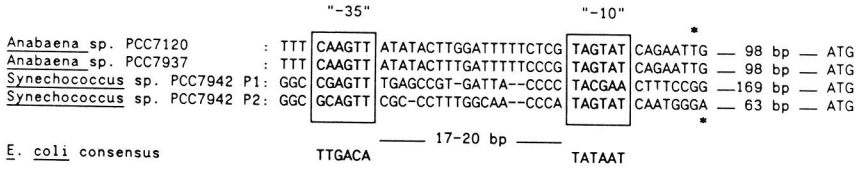


Fig. 12. Comparison of putative *petF1* promoter sequences. Putative *petF1* promoter regions for *Anabaena* sp. PCC 7120 (Alam et al. 1986), *Anabaena* sp. PCC 7937 (this study, Fig. 3A), and *Synechococcus* sp. PCC 7942 (P1: this study, Fig. 5A; P2: Reith et al. 1986) are aligned for optimal homology. the presumed '−10' and '−35' regions are boxed. The 5′ ends determined for the *petF1* transcripts of *Anabaena* sp. PCC 7120 (Alam et al. 1986) and *Synechococcus* sp. PCC 7942 (Reith et al. 1986) are marked with an asterisk.

lacking an intact *petF1* gene is not viable. Although the *neo* gene was inserted properly in the chromosomal DNA, thereby replacing the *petF1* gene, the kanamycin-resistant transformants still contained an intact *petF1* gene. These results are interpreted as an indication that the *petF1* ferredoxin is essential to the cell, even under conditions of iron-limitation, when the synthesis of flavodoxin is induced. The precise genetic make-up of these transformants is not yet understood. The intact and the impaired gene copy probably reside on different copies of the chromosome, because *Synechococcus* sp. PCC 7942 is thought to contain up to 16 genome equivalents per growing cell (Mann and Carr 1974). In experiments to eliminate the *petF1* gene, those cells seem to be selected in which a partial heterozyzgous situation ensures the cell of both the necessary *petF1* and *neo* gene products.

The mutagenesis technique by homologous recombination has proven to be effective for inactivation of *Synechococcus* sp. PCC 7942 genes: for the *met1* gene (Kuhlemeier et al. 1985; Van der Plas et al., to be published) and for up to two of the three expressed copies of the *psbA* genes (Golden et al. 1986). It seems that the desired mutants can only be produced if the mutagenized function is dispensable, i.e. as long as it can be compensated e.g. by the uptake of methionine added to the growth medium or by the expression of a remaining intact copy in the case of *psbA*. Attempts to inactivate the third *psbA* gene copy (Golden et al. 1986) or a single copy '*recA*'-like gene (Borrias et al., to be published) failed and gave results similar to those for the *petF1* gene. Further evidence regarding the dispensability of ferredoxin may come from conditional ferredoxin mutants, in which expression of the *petF1* gene is directed by a controlable promoter.

Acknowledgements

The authors thank Ans van Pelt, Gigi Jonk and Renske Oosterhoff-Teertstra for their help in the sequence determination, Sjef Smeekens for providing the *S. pratensis petF1* cDNA clone, Jenneke Huizingh and Geert de Vrieze for their technical assistance, Evelien Kampert for the construction of *E. coli* PC 2495, Maarten Terlou from the Department for Image Processing and Design for the optical density measurements, Dick Smit and Ronald Leito for help in preparing the figures, Mark Tuyl, René Pieterse and Bart Klein for their contribution to the blotting and mapping experiments, and Johan Hageman for his advice and support in handling computers and software.

References

Alam J, Whitaker RA, Krogmann DW and Curtis SE (1986) Isolation and sequence of the gene for ferredoxin I from the cyanobacterium *Anabaena* sp. Strain PCC 7120. J Bacteriol 168: 1265–1271

Böhme H and Schrautemeier B (1987) Comparative characterization of ferredoxins from heterocysts and vegetative cells of *Anabaena variabilis*. Biochim Biophys Acta 891: 1–7

Bothe H (1977) Flavodoxin. In: Trebst A and Avron M (eds) Photosynthesis (Encyclopedia of Plant Physiology 5), pp 216–221. Berlin: Springer

Bradford MM (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

Buchanan BB (1984) The ferredoxin/thioredoxin system: a key element in the regulatory function of light in photosynthesis. BioScience 34: 378–383

Cathala G, Savouret J, Mendez B, West BL, Karin M, Martial JA and Baxter JD (1983) A method for the isolation of intact, translationally active ribonucleic acid. DNA 2: 329–335

Chan TM, Hermodson MA, Ulrich EL and Markley JL (1983) Nuclear magnetic resonance studies of two-iron-two-sulfur ferredoxins. 2. Determination of the sequence of the *Anabaena variabilis* ferredoxin II, assignment of aromatic resonances in proton spectra and effects of chemical modifications. Biochemistry 22: 5988–5995

Cozens AL and Walker JE (1987) The organization and sequence of the genes for ATP synthase subunits in the Cyanobacterium Synechococcus 6301; support for an endosymbiotic origin of chloroplasts. J Mol Biol 194: 359–383

Csiszàr K, Houmard J, Damerval T and Tandeau de Marsac N (1987) Transcriptional analysis of the cyanobacterial *gvpABC* operon in differentiated cells: occurrence of an antisense RNA complementary to three overlapping transripts. Gene 60: 29–37

Curtis SE and Haselkorn R (1983) Isolation and sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from the cyanobacterium *Anabaena* 7120. Proc Natl Acad Sci USA 80: 1835–1839

Dente L, Cesareni G and Cortese R (1983) pEMBL: a new family of single stranded plasmids. Nucleic Acids Res 11: 1645–1655

Dugaiczyk A, Boyer HW and Goodman HM (1975) Ligation of *Eco*RI endonuclease-generated DNA fragments into linear and circular structures. J Mol Biol 96: 171–184

Duyvesteyn MGC, Korsuize J, De Waard A, Vonshak A and Wolk CP (1983) Sequence-specific endonucleases in strains of *Anabaena* and *Nostoc*. Arch Microbiol 134: 276–281

Dzelzkalns VA, Owens GC and Bogorad L (1984) Chloroplast promoter driven expression of the chloramphenicol acetyl transferase gene in a cyanobacterium. Nucleic Acids Res 12: 8917–8925

Feinberg AP and Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13

Golden SS, Brusslan J and Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. EMBO J 5: 2789–2798

Hall DO and Rao KK (1977) Ferredoxin. In: Trebst A and Avron M (eds) Photosynthesis (Encyclopedia of Plant Physiology 5), pp 206–216. Berlin: Springer

Ho KK, Ulrich EL, Krogmann DW and Gomez-Lojero C (1979) Isolation of photosynthetic catalysts from cyanobacteria. Biochim Biophys Acta 545: 236–248

Ho KK and Krogmann DW (1982) Photosynthesis. In: Carr NG and Whitton BA (eds) The Biology of Cyanobacteria, pp 191–214. Oxford: Blackwell

Hutber GN, Hutson KG and Rogers LJ (1977) Effect of iron deficiency on levels of two ferredoxins and flavodoxin in a cyanobacterium. FEMS Microbiol Lett 1: 193–196

Ish-Horowicz D and Burke JF (1981) Rapid and efficient cosmid cloning. Nucleic Acids Res 9: 2989–2998

Kuhlemeier CJ, Thomas AAM, Van der Ende A, Van Leen RW, Borrias WE, Van den Hondel CAMJJ and Van Arkel GA (1983) A host-vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. Plasmid 10: 156–163

Kuhlemeier CJ, Hardon EM, Van Arkel GA and Van de Vate C (1985) Self-cloning in the cyanobacterium *Anacystis nidulans* R2: fate of a cloned gene after reintroduction. Plasmid 14: 200–208

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 277: 680–685

Lizardi PM, Binder R and Short SA (1984) Preparative isolation of DNA and biologically active mRNA from diethylaminoethyl membrane. Gene Anal Techn 1: 33–39

Maniatis T, Fritsch EF and Sambrook J (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory

Mann N and Carr NG (1974) Control of macromolecular composition and cell division in the blue-green alga *Anacystis nidulans*. J Gen Microbiol 83: 399–405

Matsubara H and Hase T (1983) Phylogenetic consideration of ferredoxin sequences in plants, particularly algae. In: Jensen U and Fairbrothers DE (eds) Proteins and nucleic acids in plant systematics, pp 168–181. Berlin: Springer

Mazur BJ, Rice D and Haselkorn R (1980) Identification of blue-green algal nitrogen fixation genes by using heterologous DNA hybridization probes. Proc Natl Acad Sci USA 77: 186–190

Meinkoth J and Wahl G (1984) Hybridisation of nucleic acids immobilized on solid supports. Anal Biochem 138: 267–284

Reith ME, Laudenbach DE and Straus NA (1986) Isolation and nucleotide sequence analysis of the ferredoxin I gene from the cyanobacterium *Anacystis nidulans* R2. J Bacteriol 168: 1319–1324

Rippka R, Deruelles J, Waterbury JB, Herdman M and Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol 111: 1–61

Rosenberg M and Court D (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Ann Rev Genet 13: 319–353

Sakihama N and Shin M (1987) Evidence from high-pressure liquid chromatography for the existence of two ferredoxins in plants. Arch Biochem Biophys 256: 430–434

Sanger F, Coulson AR, Barrell BG, Smith AJH and Roe BA (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143: 161–178

Schrautemeier B and Böhme H (1985) A distinct ferredoxin for nitrogen fixation isolated from heterocysts of the cyanobacterium *Anabaena variabilis*. FEBS Lett 184: 304–308

Smeekens S, Van Binsbergen J and Weisbeek P (1985) The plant ferredoxin precursor: nucleotide sequence of a full length cDNA clone. Nucleic Acids Res 113: 3179–3194

Stephens RM (1985) A sequencers sequence analysis package for the IBM PC. Gene Anal Techn 2: 67–75

Tandeau de Marsac N and Houmard J (1987) Advances in cyanobacterial molecular genetics. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 251–302. Amsterdam: Elsevier

Thomas PS (1983) Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymol 100: 255–266

Tinoco J, Borer PN, Dengler B, Levine MD, Uhlenbeck OC, Crothers DM and Cralla J (1973) Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol 246: 40–41

Tomioka N and Sugiura M (1983) The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue-green alga, *Anacystis nidulans*. Mol Gen Genet 191: 46–50

Towbin H, Staehelin T and Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci USA 76: 4350–4354

Van Bergen en Henegouwen PMP and Leunissen JLM (1986) Controlled growth of colloidal gold particles and implications for labelling efficiency. Histochem 85: 81–87

Van den Hondel CAMJJ, Verbeek S, Van der Ende A, Weisbeek P, Borrias WE and Van Arkel GA (1980) Introduction of transposon Tn901 into a plasmid of *Anacystis nidulans*: preparation for cloning in cyanobacteria. Proc Natl Acad Sci USA 77: 1570–1574

Van der Plas J, De Groot RP, Weisbeek PJ and Van Arkel GA (1986a) Coding sequence of a ferredoxin gene from *Anabaena variabilis* ATCC 29413. Nucleic Acids Res 14: 7803

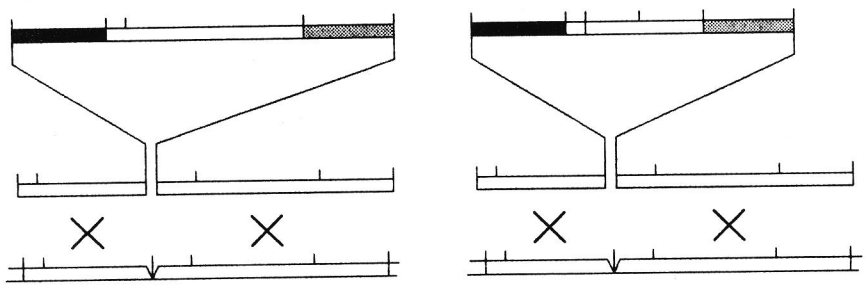
Van der Plas J, De Groot RP, Woortman MR, Weisbeek PJ and Van Arkel GA (1986b) Coding sequence of a ferredoxin gene from *Anacystis nidulans* R2 (*Synechococcus* PCC 7942). Nucleic Acids Res 14: 7804

Vieira J and Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259–268

Wolk CP and Schafer PW (1976) Heterotrophic micro- and macrocultures of a nitrogen-fixing cyanobacterium. Arch Microbiol 110: 144–147

Chapter III

The gene for the precursor of plastocyanin from the cyanobacterium *Anabaena* sp. PCC 7937: isolation, sequence and regulation



J. van der Plas, A. Bovy, F. Kruyt, G. de Vrieze,
E. Dassen, B. Klein and P. Weisbeek
(1989) Molecular Microbiology 3:275–284

The gene for the precursor of plastocyanin from the cyanobacterium *Anabaena* sp. PCC 7937: isolation, sequence and regulation

Summary

The gene encoding plastocyanin (*petE1*) from *Anabaena* sp. PCC 7937 was isolated using two sets of mixed oligonucleotide hybridization probes derived from conserved regions in the protein. Plastocyanin is encoded as a preprotein of 139 amino acids. The amino-terminal extension of 34 residues has all the characteristics of a signal peptide and is probably involved in translocation of preplastocyanin over the thylakoid membrane. The level of the *petE1* mRNA, a single transcript of about 740 bases, was found to be severely reduced under conditions of Cu^{2+} deficiency. The *petE1* gene was transferred to the genome of *Synechococcus* sp. PCC 7942, which did not appear to contain a structural gene for plastocyanin itself. The integrated gene becomes expressed at the transcriptional level, regardless of the amount of Cu^{2+} available.

Introduction

Plastocyanin is a soluble, low molecular-weight copper protein (MW 10.5–12.0kD) that functions as a redox carrier in the photosynthetic electron transport chain of organisms performing oxygenic photosynthesis (Boulter *et al.*, 1977). Plastocyanin catalyses electron transfer from the membrane-bound cytochrome b_6/f complex to P-700, the reaction centre of Photosystem I, by reversible oxidation (Cu^{II}) and reduction (Cu^{I}) of its active copper centre (Crofts and Wood, 1978; Ho and Krogmann, 1982).

The copper-containing plastocyanin can be replaced by the iron-haem protein, cytochrome c_{553} (Wood, 1978; Ho and Krogmann, 1984; Sandmann, 1986). Many cyanobacteria (e.g. *Fischerella muscicola*) have cytochrome c_{553} instead of plastocyanin. Other cyanobacterial species (e.g. *Anabaena variabilis*) and some green algae (e.g. *Chlamydomonas reinhardtii*, *Scenedesmus acutus*) express either plastocyanin, cytochrome c_{553} or both. Under growth conditions with insufficient Cu^{2+}, cytochrome c_{553} is produced as an alternative and additional electron carrier for plastocyanin. Likewise, depletion of endogenous copper seems to be the reason for the absence of plastocyanin in dense natural blooms of cyanobacteria and in *Anabaena* sp. stationary phase cells from laboratory cultures (Ho *et al.*, 1979). This Cu^{2+}-regulated exchange is comparable to the iron-controlled ferredoxin/flavodoxin exchange in cyanobacteria and algae (Wood, 1978). In the green alga *C. reinhardtii*, the plastocyanin content is controlled by rapid degradation of constitutively synthesized apoprotein in the absence of Cu^{2+} (Merchant and Bogorad, 1986b), whereas the cytochrome c_{553} content appears to be regulated by transcriptional control, probably via sensing of the endogenous Cu^{2+} concentration and not by the plastocyanin content of the cell (Merchant and Bogorad, 1987).

The complete amino-acid sequence of plastocyanin from *Anabaena* sp. PCC 7118 has been published (Aitken, 1975). When aligned with sequences of green algal and plant plastocyanin, their relationship is evident; the highest conservation resides in the regions around the amino acid residues involved in formation of the copper cluster (Boulter *et al.*, 1977; Guss and Freeman, 1983). Recently, the nucleotide sequence of a full-length cDNA clone encoding the complete precursor protein for *Silene pratensis* plastocyanin has been determined (Smeekens *et al.*, 1985), followed by the genomic sequences from spinach (Rother *et al.*, 1986), barley (Nielsen and Gausing, 1987) and *Arabidopsis thaliana* (Vorst *et al.*, 1988).

Plastocyanin is functional in the thylakoid lumen of the cyanobacterium and of algal and higher plant chloroplasts. In order to arrive at this specific location, it has to pass the thylakoid membrane during or after synthesis in the cytoplasm. Proteins targeted to transmembrane destinations, both in chloroplasts and in bacteria, are often made as a precursor protein with an amino-terminal extension that contains information for routing of the mature protein towards its specific location (Pugsley and Schwartz, 1985; Smeekens and Weisbeek, 1988). Concomitantly with the transport process, the precursor protein is proteolytically cleaved to its mature size. Similarly, cyanobacterial plastocyanin is expected to be encoded

and synthesized as a precursor polypeptide. At present, however, nothing is known about the mechanism of protein targeting to the functionally distinct cytoplasmic membrane and intracytoplasmic thylakoid membrane of cyanobacteria. Isolation of the gene for plastocyanin (*petE1*) should provide a tool for genetically and physiologically investigating the question of protein transport across the thylakoid membrane.

Analysis of the biogenesis and Cu^{2+} regulation of plastocyanin in cyanobacteria was initiated by isolation of the gene encoding the plastocyanin precursor from the cyanobacterium *Anabaena* sp. PCC 7937 and by its transfer to the genome of *Synechococcus* sp. PCC 7942. A structural gene for plastocyanin is normally absent in this strain, as was concluded from the absence of hybridization of its DNA with the *Anabaena* sp. PCC 7937 *petE1* gene as a probe.

Results

Cloning of the Anabaena *sp. PCC 7937* petE1 *gene*

Comparison of the amino acid sequence of plastocyanin from *Anabaena* (*Anabaena* sp. PCC 7118; Aitken, personal communication) with the sequences from a few algae and several plants shows the highest conservation in two stretches surrounding the residues involved in formation of the copper centre (His 39, Cys 89, His 92 and Met 97). The regions covering amino acid 39 to 44 and 87 to 92 were selected for the synthesis of the mixed oligonucleotide probes 1 and 2, respectively. All possible codons for each amino acid were taken. Both mixed oligonucleotide probes are 17 bases in length and are synthesized in the non-coding orientation (Fig. 1).

The two oligonucleotide probes were used to probe Southern blots of *Anabaena* sp. PCC 7937 total DNA,

```
                      39                                          44
AMINO ACID SEQUENCE   :   His - Asn - Val - Val - Phe - Asp

DNA SEQUENCE          : 5' CAT  AAT  GTN  GTN  TTT  GAT  3'
                           C    C              C    C

PROBE 1 (128x17-MERS): 3' GTA  TTA  CAN  CAN  AAA  CT   5'
                           G    G              G

                      87                                          92
AMINO ACID SEQUENCE   :   Phe - Tyr - Cys - Glu - Pro - His

DNA SEQUENCE          : 5' TTT  TAT  TGT  GAA  CCN  CAT  3'
                           C    C    C    G              C

PROBE 2 (64x17-MERS)  : 3' AAA  ATA  ACA  CTT  GGN  GT   5'
                           G    G    G    C
```

Fig. 1. Oligodeoxyribonucleotide probes used for the identification of the *Anabaena petE1* gene. The amino acid stretches used are derived from *Anabaena* sp. PCC 7118 plastocyanin (Aitken, 1975). In the DNA sequences, N represents any of the four nucleotides.

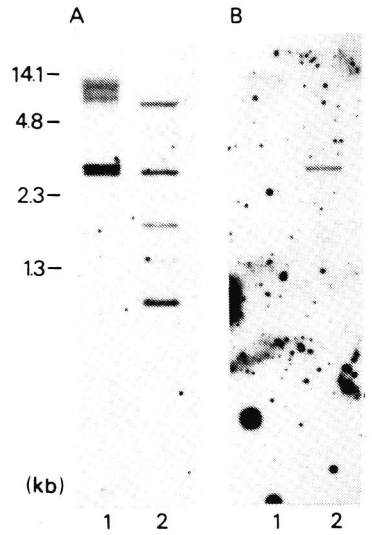


Fig. 2. Hybridization of oligonucleotide probes to *Anabaena* DNA. Total *Anabaena* DNA (10 µg) digested with *Eco*RI (lane 1) and *Hin*dIII (lane 2) was run on a 0.8% agarose gel and blotted to nitrocellulose bidirectionally. One blot was hybridized with probe 1 (panel A), the other with probe 2 (panel B). Probe 1 and 2 were end-labelled with ^{32}P. Hybridization was performed at 37°C and washing at 45°C. Molecular size markers indicated are fragments from *Bst*EII-digested phage lambda DNA.

digested with *Hin*dIII and *Eco*RI (Fig. 2). With probe 1 this resulted in four major hybridization signals in both digests, whereas single signals were obtained with probe 2. The *Hin*dIII fragment of approximately 3.1 kb hybridizing with probe 2 (Fig. 2: panel B), coincided with one out of the four *Hin*dIII fragments reacting with probe 1 (Fig. 2: panel A).

In order to isolate this 3.1 kb *Hin*dIII DNA fragment, a plasmid library was constructed by ligating DNA fragments in the range of 2.5–3.5 kb of *Hin*dIII-digested *Anabaena* sp. PCC 7937 DNA to *Hin*dIII-linearized and phosphatase-treated pEMBL8 DNA. The ligation mixture was transformed to *Escherichia coli* PC 2495 and 480 of the resulting colonies were grown in mixes of twelve. Plasmid DNAs from these mixed cultures were digested with *Hin*dIII and analysed by Southern blotting and hybridization. One of the 40 mixed plasmid preparations gave hybridization with both probes. The corresponding mixed culture was plated for single colonies and the selection procedure was subsequently repeated with the separate plasmid DNAs extracted from 36 single cultures. This resulted in the isolation of plasmid pPCV1, which contained a 3.1 kb *Hin*dIII fragment from *Anabaena* sp. PCC 7937 DNA hybridizing specifically both to probe 1 and 2 (results not shown). The 3.1 kb *Hin*dIII insert of pPCV1 was mapped with the restriction enzymes *Eco*RV,

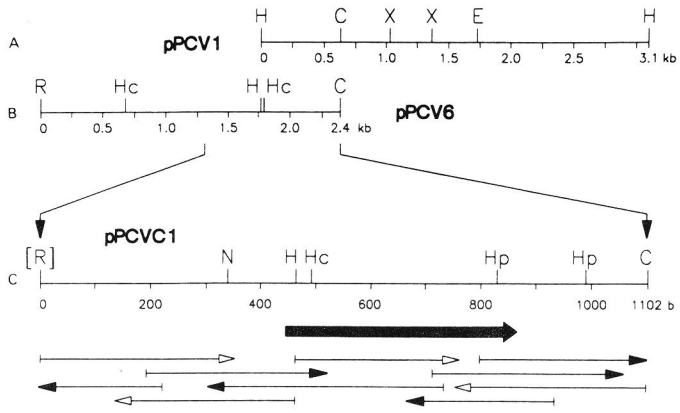


Fig. 3. Physical maps of the *petE1* gene region.
A. Restriction map of the 3.1 kb *Hin*dIII fragment of pPCV1.
B. Restriction map of the 2.4 kb *Eco*RI-*Cla*I fragment cloned in pPCV6. This fragment did not contain sites for *Xba*I and *Eco*RV.
C. Restriction map of the 1.1 kb insert of pPCVC1 that was sequenced. The arrows below the restriction map give the direction and extent of the DNA regions analysed. The closed arrowheads indicate the use of specific oligonucleotide primers, and the open arrowheads represent the use of standard M13 primer in the sequencing. Letters indicate restriction sites as follows: C, *Cla*I; E, *Eco*RV; H, *Hin*dIII; Hc, *Hin*cII; Hp, *Hin*PI; N, *Nru*I; R, *Eco*RI; X, *Xba*I. The solid arrow represents the plastocyanin coding region. The bracketed *Eco*RI site in Fig. 3C is not present in the genomic situation, but was positioned there by fragment shortening with nuclease Bal31.

*Xba*I and *Cla*I (Fig. 3A). The hybridizing sequence was localized on the 0.7 kb *Hin*dIII-*Cla*I terminal fragment of pPCV1.

Preliminary sequencing showed that the cloned fragment did not contain the complete *petE1* gene, because the coding sequence was truncated by a *Hin*dIII site. Thus an overlapping restriction fragment was identified by Southern hybridization analysis of single and double digests of *Anabaena* sp. PCC 7937 total DNA with the 0.7 kb *Hin*dIII-*Cla*I fragment as a probe (results not shown). A 2.4 kb *Eco*RI-*Cla*I fragment should contain the complete gene, including the 5′-non-coding region. This fragment was isolated by directional cloning of the 1.9–2.8 kb fraction of chromosomal *Eco*RI-*Cla*I fragments in pEMBL18, digested with *Eco*RI and *Acc*I, followed by transformation to *E. coli* PC2495 and subsequent colony hybridization. Several colonies hybridized with the 0.7 kb *Cla*I-*Hin*dIII probe and one of them was analysed further. A physical map of the 2.4 kb insert of the plasmid concerned (pPCV6) is given in Fig. 3B.

Sequence analysis of the petE1 *gene*

The sequencing strategy employed is shown in Fig. 3C together with plasmid pPCVC1. This plasmid was made by nuclease Bal31 digestion from the left *Hin*cII site of the insert from pPCV6 (Fig. 3B). Subclones for sequencing with the standard M13 primer were made, starting at the *Eco*RI, the *Hin*dIII and *Cla*I site. Further sequence analysis was performed with the aid of specific oligonucleotide primers in order to obtain sequence information on both strands and in overlapping stretches.

Framescan analysis of the 1.1 kb sequence obtained (Fig. 4A) in all six possible reading frames led to the identification of the coding sequence for plastocyanin in the first reading frame (Fig. 4B). The nucleotide stretches found at positions 217–233 and 361–377 in this coding sequence both have a full complement in the mixed oligonucleotide probes used (Figs 1 and 4). There was no indication for (parts of) other potential coding sequences. The ATG codon at position 1 has to be the start codon for *petE1*, because there is no other start codon present in the same open reading frame. The identified coding sequence of 417 bp is preceded by a putative ribosome binding site AGGA (Tomioka and Sugiura, 1983) at a distance of seven bases from the ATG (Fig. 4A). Downstream of the coding region, two sets of palindromic sequences are found (Fig. 4A), which have the potential to form stable 'hairpin' structures in the mRNA with calculated free energies of formation of −13 and −18 kCal mol^{-1}, respectively (Tinoco *et al.*, 1973). These structures could represent a transcription-termination signal or a barrier against 3′ degradation (Brawerman, 1987). The 450 bp upstream of the coding sequence were screened for the presence of potential promoter sequences, resembling the *E. coli* consensus promoter (Hawley and McClure, 1983), the cyanobacterial *petF1* promoters (Van der Plas *et al.*, 1988), or other cyanobacterial promoters. No obvious similarities were discovered. A striking feature, however, of the 250 bp upstream of the coding sequence is its high AT content (69%) relative to the average value (59%) for the complete 1102 bp sequence, which is close to the 58% reported for *Anabaena* sp. PCC 7937 (Herrero *et al.*, 1984).

Genomic situation

Southern blots of restricted *Anabaena* sp. PCC 7937 DNA probed with the 700 bp *Cla*I-*Hin*dIII DNA fragment (Fig. 3C) at various stringencies showed only a single, major hybridization signal for each digest tested (the result at

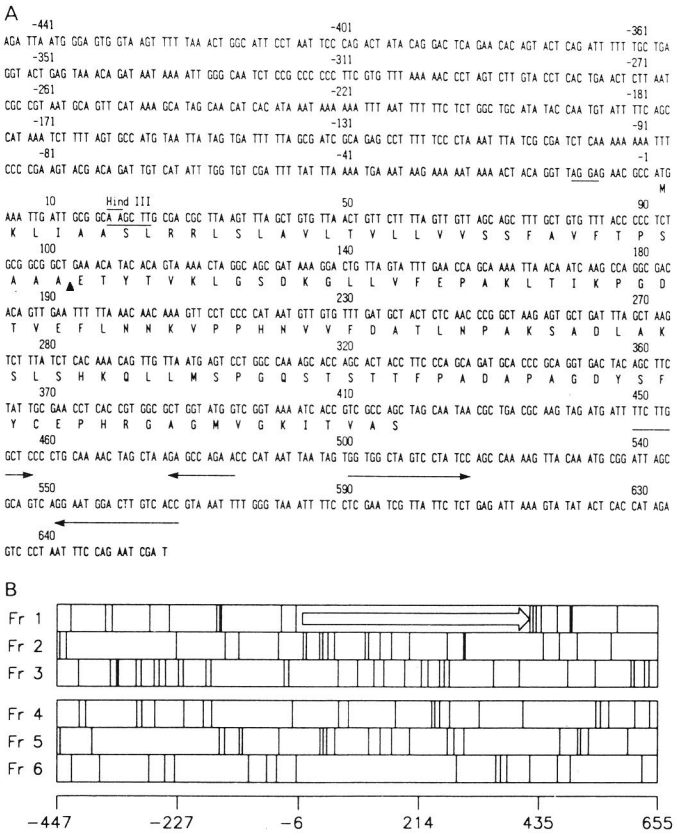


Fig. 4. Nucleotide sequence of the *petE1* gene.
A. The sequence presented begins 465 bp upstream from the *Hin*dIII site and ends at the *Cla*I site. The amino acid sequence deduced for the plastocyanin pre-apoprotein is displayed below the DNA coding region in the standard single-letter code. The presumed processing site for the removal of the signal peptide is indicated (▲). The putative ribosome binding site (AGGA) is underlined. Potential secondary structures, downstream of the coding region and centred at nucleotides 466 and 533, respectively, are indicated by arrows.
B. Scan for stop codons in the six possible reading frames of the 1102 bp sequence. Numbering is according to that shown in A. The *petE1* coding sequence is indicated by the arrow in frame 1. These sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Database Libraries under the accession number X14256.

60°C and 5X SSC for washing is shown in Fig. 5). This is strong evidence that the *petE1* gene of *Anabaena* sp. PCC 7937 is a single-copy gene.

In order to investigate whether the unicellular cyanobacterium *Synechococcus* sp. PCC 7942 also encodes a plastocyanin, Southern blots of chromosomal DNA digests of this strain were analysed with the *Anabaena* sp. PCC 7937 *petE1* probe at very low stringency (50°C and 5X SSC). This did not result in any hybridization signal at all. The cautious conclusion from this absence of specific hybridization is that *Synechococcus* sp. PCC 7942 does not contain a *petE1* gene (see *Discussion*).

Transcription of the petE1 *gene*

The plastocyanin content of many cyanobacteria and algae is dependent on the availability of copper to the cell (Ho and Krogmann, 1982; Sandmann, 1986). To investigate the level at which regulation occurs, *petE1*-specific mRNA levels were determined by Northern analysis. Total RNA was extracted from mid-exponential phase cells grown with and without Cu^{2+}. After denaturation, separation on agarose and blotting onto nylon membrane, the RNA was probed with the 360 bp *Hin*dIII-*Hin*PI *petE1* fragment. A single specific transcript of about 740 bp was detected (Fig. 6). The size of this mRNA species shows that the *Anabaena* sp. PCC 7937 *petE1* gene is transcribed as a monocistronic message. Total RNA extracted from Cu^{2+}-limited cells gave a much weaker hybridization signal. It differed by a factor of approximately 30 from that obtained with total RNA from cells grown in Cu^{2+}-rich medium. The level of the 520 bp transcript of the *petF1* gene encoding ferredoxin (Van der Plas *et al.*, 1988), however, was not influenced by changes in the Cu^{2+}

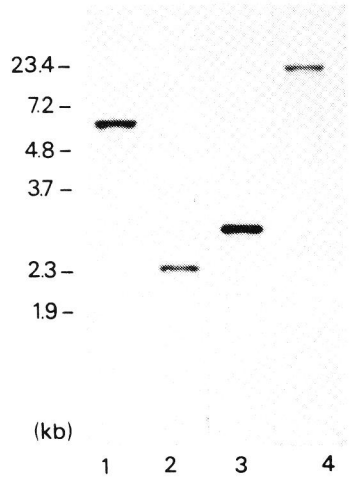


Fig. 5. Hybridization of the 700 bp *Hin*dIII-*Cla*I *petE1* fragment to total *Anabaena* DNA. The DNA (10 μg) was digested with *Cla*I (lane 1), *Cla*I-*Eco*RI (lane 2), *Hin*dIII (lane 3) and *Xba*I (lane 4), run on a 0.8% agarose gel blotted to nitrocellulose and probed with radiolabelled *petE1* DNA. Molecular size markers indicated are fragments from phage lambda DNA digested with *Bst*EII or with *Apa*I-*Xho*I.

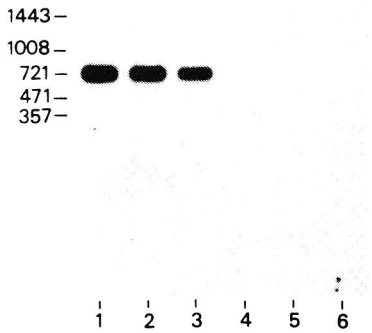


Fig. 6. Northern analysis of *Anabaena* cells grown with or without Cu^{2+}. Total RNA was denatured, electrophoresed on a 1.1% agarose gel, blotted and hybridized with an 0.36 kb *Hin*dIII-*Hin*PI *petE1* probe fragment. pEMBL8-*Taq*I DNA fragments were denatured and used as size markers. Lanes 1–3: RNA from cells grown in BG11 with 1 μM Cu^{2+} (40, 30, and 20 μg, respectively). Lanes 4–6: RNA from cells grown in BG11 without Cu^{2+} (40, 30, and 20 μg, respectively).

concentration and this transcript was subsequently used as an internal standard to check the amount of RNA analysed (data not shown).

Transfer of the petE1 *gene to* Synechococcus *sp. PCC 7942*

Both the absence of *petE1* and the availability of genetic tools for *Synechococcus* sp. PCC 7942 (Shestakov and Khuyen, 1970; Tandeau de Marsac and Houmard, 1987) make this strain an interesting host for the study of *Anabaena* sp. PCC 7937 plastocyanin. The *petE1* gene was integrated into the genome of *Synechococcus* sp. PCC 7942 by transformation of strain R2-PIM9, a universal acceptor developed for genomic integration of DNA fragments cloned in pBR322-derived vectors (Van der Plas *et al.*, unpublished), with the plasmid pPCVC1 (Fig. 3C).

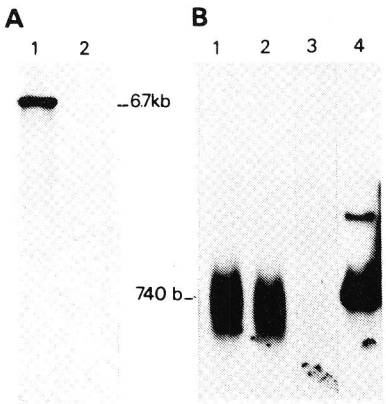


Fig. 7. Hybridization analysis of *Synechococcus* sp. PCC 7942 strain R2-PIM9-PC transformed with the *Anabaena petE1* gene.
A. Southern blot of *Eco*RV-digested chromosomal DNA from R2-PIM9-PC (lane 1) and from the corresponding host strain R2-PIM9 (lane 2) probed with the 0.36 kb *Hin*dIII-*Hin*PI *petE1* fragment.
B. Northern blot with total RNA (*c.* 40 μg) extracted from R2-PIM9-PC (lanes 1 and 2), R2-PIM9 (lane 3) and *Anabaena* sp. PCC 7937 (lane 4) probed with the 0.36 kb *petE1* fragment. The RNA in lane 2 is extracted from R2-PIM9-PC grown without Cu^{2+}. All the other RNAs were extracted from cells grown in the presence of Cu^{2+}.

The resulting transformants were analysed for the presence of the plastocyanin gene by Southern hybridization of chromosomal digests made with *Eco*RV. *Eco*RV does not cleave pPCVC1 carrying the *petE1* gene. A 6.6 kb hybridizing fragment was detected in the DNA of the transformed strain R2-PIM9-PC but not in the host strain R2-PIM9 (Fig. 7A). The size of this fragment exactly matches the size deduced from the *Eco*RV map of the integration site (Van der Plas *et al.*, unpublished) and from the size of the *petE1* clone used for transformation.

In a Northern blot made with RNA extracted from the transformed strain R2-PIM9-PC grown with and without Cu^{2+}, *petE1*-specific transcripts of *c.* 740bp were detected but are absent in RNA extracted from the corresponding wild-type strain, R2-PIM9 (Fig. 7B). Part of the RNA hybridizing to the probe is detected as a 'smear', probably produced by specific mRNA degradation, as the transcript of the ferredoxin gene (Van der Plas *et al.*, 1988) is detected on the same blot as a discrete band (results not shown).

From these results, it must be concluded that the *petE1* of *Anabaena* sp. PCC 7937 has been integrated successfully into the genome of *Synechococcus* sp. PCC 7942 and that in the transformant analysed the *petE1* gene is transcribed in a manner independent of the availability of Cu^{2+}.

Discussion

The gene encoding plastocyanin (*petE1*) from the filamentous cyanobacterium *Anabaena* sp. PCC 7937 was cloned by hybridization with two sets of synthetic oligonucleotides. The protein is encoded as a precursor of 139 amino acids. The mature protein probably begins with Glu at nucleotide position 103, as derived from comparison with the *Anabaena* sp. PCC 7118 plastocyanin mature sequence (Aitken, 1975). The deduced amino acid sequence of the mature plastocyanin from *Anabaena* sp. PCC 7937 differs from the amino acid sequence of the *Anabaena* sp. PCC 7118 protein in four positions. Three of these substitutions are for similar amino acids: position 46 Thr for Ala, position 84 Asp for Glu, and position 86 Ser for Thr, whereas Ser substitutes Gly at the carboxy terminus in the *Anabaena* PCC 7937 plastocyanin. The pre-apoplastocyanin has a calculated molecular mass of 14.6kD. The deduced mature protein consists of 105 amino acid residues and should have a molecular mass of 11.1kD. It is a basic protein, in contrast to plastocyanins of eukaryotic origin, which uniformly have an acidic character (Ho and Krogmann, 1984). The pI of *c.* 8 estimated for the mature plastocyanin is in accordance with the reported values of 8.4 (Davis *et al.*, 1980) and 7.74 (Ho and Krogmann, 1984).

Comparison of the nucleotide sequence coding for the mature cyanobacterial protein with the corresponding sequences for plastocyanin from the higher plants *Silene pratensis* (Smeekens *et al.*, 1985), *A. thaliana* (Vorst *et al.*, submitted), spinach (Rother *et al.*, 1986) and barley (Nielsen and Gausing, 1987) gives a percentage similarity of *c.* 51. The matching bases are scattered along the sequence, the longest stretch being eleven. At the amino acid level, the similarities for the deduced mature proteins were only *c.* 40%. The extent of the similarity between the *Anabaena* amino acid sequence and the green algal sequences from *Enteromorpha prolifera* (Simpson *et al.*, 1986) and *Chlorella fusca* (Kelly and Ambler, 1974) is of the same order (Fig. 8). When the higher plant sequences were compared to each other, similarities of 67–73% at nucleotide level, and 74–81% at amino acid level were found. The large difference and evolutionary distance between cyanobacterial plastocyanin and plant plastocyanin, concluded from sequence comparisons (Fig. 8) and differences in isoelectric points, is also seen in the absence of an immunological cross-reaction between anti-spinach plastocyanin antibodies and *Anabaena* sp. PCC 7937 plastocyanin (Van der Plas, unpublished results; see Sandmann and Böger, 1983). Also, the failure of initial

```
                    10        20        30        40        50        60        70        80        90       100
Anabaena sp. PCC 7937 ETYTVKLGSDKGLLVFEPAKLTIKPGDTVEFLNNKVPPHNVVFDATLNPAKSADLAKSLSHKQLLMSPGQSTSTTFPADAPAGDYSFYCEPHRGAGMVGKITVAS
Anabaena sp. PCC 7118 ..........................................A.....................................E.T.....................G 101

Chlorella fusca       -DV.....A.S.A.....SSV...A.E..TWV..AGF...I...EDEV.SGANAE.L.--.EDY.NA..E.Y.AK.--.T-..T.GYF....Q....K.T...Q-  52
Enteromorpha prolifera -AAI....G.D.S.A.V.NNI.VGA.ESI..I..AGF...I...EDAV..GVDAD.I.--AEDY.N.K..TVVRKL--TT.-.T.GV..D..S....KMT...Q-  43

Hordeum vulgare       --QD.L..ANG.V.....NDFSV.A.E.IT.K..AGY.......EDAV.SGVDVSKI.--QEEY.TA..ETF.V.L--TV.-.T.G......A.......V..N-  47
Silene pratensis      --AE.L...SD.G.A.V.SD.S.AS.EKIT.K..AGF.......EDEV..GVDVTKI.MPEED..NA..EEY.V.L--TE-K.T.K...A..A.......V..N-  46
Arabidopsis thaliana  --ME.L....D.S.A.V.SEF.VAK.EKIV.K..AGF.......EDEI.SGVDASKI.MDETA..NGA.ETYEV.L--TE.-.S.G...A..Q.......L..K-  44
Populus nigra         --ID.L..A.D.S.A.V.SEFS.S..EKIV.K..AGF...I...EDSI.SGVDASKI.M.EED..NAK.ETFEVAL--SN-K.E.....S..Q.......V..N-  43
Spinacia oleracea     --VE.L..GGD.S.A.L.GDFSVAS.EEIV.K..AGF.......EDEI.SGVDAAKI.M.EED..NA..ETYKV.L--TE-K.T.K...S..Q.......V..N-  42
                                                  *                                                  * *   *

                      | ||  | | | |       |    ||  ||| |||  |       |   |  |           | |  | ||  ||||   ||  (31)
```

Fig. 8. Comparison of mature plastocyanin amino acid sequences. The complete amino acid sequence for *Anabaena* sp. PCC 7937 plastocyanin is shown in single-letter code. Only the amino acids which differ are given for the other cyanobacterial (Aitken, 1975), green algal (Kelly and Ambler, 1974; Simpson *et al.*, 1986) and higher plant (Nielsen and Gausing, 1987; Rother *et al.*, 1986; Smeekens *et al.*, 1985; Vorst *et al.*, 1988) plastocyanin sequences. Numbering of the amino acid residues refers to the *Anabaena* sp. PCC 7937 sequence. Dashes denote gaps introduced to maximize similarity. Numbers at the end of each sequence indicate the number of amino acids matching to *Anabaena* sp. PCC 7937 plastocyanin. At the bottom of the figure, the 31 residues shared by the plastocyanins are indicated. The stretches used for probe synthesis are underlined in the *Anabaena* sp. PCC 7937 sequence. The four residues involved in the formation of the copper centre are marked with an asterisk.

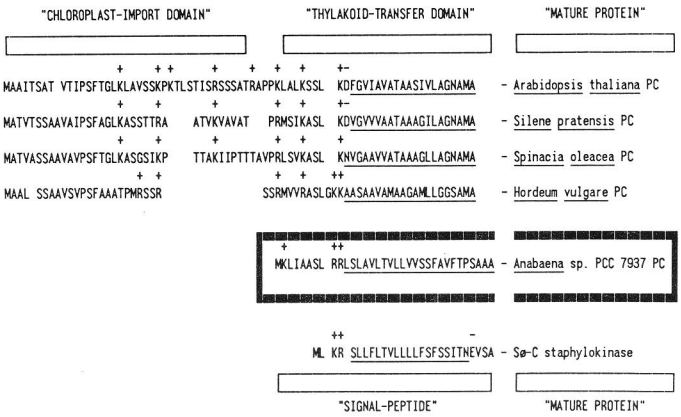


Fig. 9. Comparison of pre-plastocyanin amino-terminal extensions. The presumed signal peptide from *Anabaena* sp. PCC 7937 plastocyanin is aligned to the known plant plastocyanin transit peptides and to the bacterial signal peptide most similar to it, i.e. that of staphylokinase (Sako and Tsuchida, 1983; Sako, 1985). Strongly basic residues are indicated with a '+' sign. Strongly acidic residues are indicated with a '−' sign. The hydrophobic core regions are underlined. The transit sequences shown are from Smeekens *et al.* (1985), Rother *et al.* (1986), Nielsen and Gausing (1987) and Vorst *et al.* (1988).

attempts to identify and clone the *Anabaena* sp. PCC 7937 *petE1* gene by low-stringency hybridization with a higher plant *petE1* cDNA probe (Van der Plas, unpublished results), can be explained by such remote homology.

Regulation of the plastocyanin content of *Anabaena* sp. PCC 7937 occurs by means of influences on the steady state levels of plastocyanin transcripts, as the amount of the specific mRNA in cells grown under Cu^{2+}-limitation is much lower than that in cells grown in complete medium (Fig. 6). This can mean regulation of transcription initiation and/or differential stability of the mRNA under different conditions for growth. This regulation of the plastocyanin level is fundamentally different from that documented for *C. reinhardtii* (Merchant and Bogorad, 1986a; Merchant and Bogorad, 1986b). In this green alga, the regulation of plastocyanin content is determined by the very low stability of the copperless protein. The protein is rapidly degraded when no Cu^{2+} has been incorporated. The cytochrome c_{553} content of *C. reinhardtii*, on the other hand, appeared to be regulated at the transcriptional level by the internal Cu^{2+} concentration of the cell (Merchant and Bogorad, 1986a; Merchant and Bogorad, 1987). It is interesting to see how the cytochrome c_{553} of *Anabaena* sp. is regulated.

The amino-terminal extension of the plastocyanin precursor protein is 34 amino acids long (MW 3.5 kD) and consists mostly of hydrophobic and neutral amino acids, especially Ala, Leu and Val (19/34), and Ser and Thr (7/34), respectively (Fig. 4). The amino terminus contains three strongly basic amino acids (Arg, Arg and Lys). The carboxy-terminal end has the sequence Ala-Ala-Ala, preceded by Pro at position −5, which strongly resembles a signal-protease processing site (Ala-X-Ala). Though the hydrophobic core is somewhat longer than in most bacterial signal peptides (Watson, 1984), this means that all the features of a signal sequence are present (Pugsley and Schwartz, 1985). This conclusion is accentuated by the observation that a computer search finds that a different signal peptide, i.e. that of staphylokinase from the *Staphylococcus aureus* bacteriophage SØ-C (Sako and Tsuchida, 1983; Sako, 1985), is the most similar to the *Anabaena* signal peptide (Fig. 9).

As plastocyanin is synthesized in the cytoplasm and functions in the thylakoid lumen, it has to be translocated over the thylakoid membrane. Therefore the plastocyanin signal sequence probably has a function similar to that of the second domain of the transit peptide from higher plant plastocyanin and thus is functional during translocation across the thylakoid membrane (Smeekens and Weisbeek, 1988). In plants, plastocyanin is nuclear-encoded and synthesized in the cytoplasm as a precursor protein with an amino-terminal extension (transit peptide). This transit peptide is removed in two separate steps: one part during import in the chloroplast and the second part during transfer over the thylakoid membrane (Smeekens and Weisbeek, 1988). Comparison of the primary structure of the known thylakoid-transfer domains (Fig. 9) show that these second domains look very much like a signal-peptide and this supports a similar mechanism postulated for transport of proteins across cyanobacterial and chloroplast thylakoid-membranes.

This fits in well with the evolutionary model of the endosymbiont hypothesis (Gray and Doolittle, 1982). This model states that the chloroplast is descended from an ancestor in common with the cyanobacteria. With the transfer of genetic material, among which is the *petE1* gene, from this prokaryotic progenitor of the chloroplast to the nucleus, no new device had to be developed for thylakoid transfer. Only a chloroplast import domain had to be added to the ancestral signal peptide to give a

complete precursor which, when synthesized in the cytoplasm, is targeted for the chloroplast and ends up in the thylakoid lumen.

A second interesting aspect of the plastocyanin signal peptide is the matter of routing inside the cyanobacterial cell. This cell contains two different membrane systems: the cytoplasmic membrane and the thylakoid membrane. This situation requires a mechanism for distinguishing between both membranes in protein targeting. The fact that the plastocyanin signal is very similar to the staphylokinase signal, which directs this protein across the cytoplasmic membrane, suggests that only subtle differences in the signal sequence are required for distinguishing between the different membrane systems.

Southern hybridization analysis of *Synechococcus* sp. PCC 7942 chromosomal DNA, with the *Anabaena* sp. PCC 7937 *petE1* gene as probe, did not produce evidence for the presence of *petE1*-specific sequences. Moreover, attempts in the past to purify plastocyanin or to detect plastocyanin immunologically in *Anacystis nidulans* failed (Aitken, 1976). We conclude from the combined results that *Synechococcus* sp. PCC 7942 has no genetic information for the synthesis of plastocyanin. This, together with the genetic tools available for *Synechococcus* sp. PCC 7942, makes it an interesting host for the analysis of targeting, biogenesis and functioning of *Anabaena* sp. PCC 7937 plastocyanin. As a first step in this direction we successfully transferred the *petE1* gene to the genome of *Synechococcus* sp. PCC 7942 (Van der Plas *et al.*, submitted) and demonstrated its Cu^{2+}-independent transcription into *petE1*-specific mRNA. Antibodies raised against a synthetic amino-terminal fragment of *Anabaena* plastocyanin are presently used to investigate plastocyanin expression at the protein level and the destination of plastocyanin in its new host.

Experimental procedures

Materials

Nitrocellulose filters (PH79) and DEAE membranes (NA45) were manufactured by Schleicher & Schuell (Dassel, FRG). Oligonucleotides for probe labelling (random hexamers) were from Pharmacia (Uppsala, Sweden). Oligonucleotide probes were generously supplied by Professor Dr Van Boom (University of Leiden). Oligonucleotide primers were obtained from Dr Verheij (University of Utrecht). Radiolabelled nucleotides and Hybond-N membrane were purchased from Amersham (Little Chalfont, UK).

Organisms and culture conditions

The cyanobacterial strains *Anabaena* sp. PCC 7937 (*Anabaena variabilis* ATCC 29413) (Duyvesteyn *et al.*, 1983) and *Synechococcus* PCC 7942 (in this study the small-plasmid-cured derivative *A. nidulans* R2-SPc has been used; Kuhlemeier *et al.*, 1983) were grown in BG11 medium (Rippka *et al.*, 1979). Copper-limited growth conditions were obtained by omitting $CuSO_4$ from the medium. For copper-rich conditions, the Cu^{2+} concentration was brought to 1 μM for *Anabaena* and to 0.3 μM for *Synechococcus*.

E. coli PC2495, a *recA*, *hsdS* derivative of JM101 *supE*, *thi*, Δ(*lac-proAB*), (F′, *traD*36, *proAB*, *lacI*^{q}ZΔM15) (Vieira and Messing, 1982) constructed by E. Kampert of this department, was grown in LB medium as described (Maniatis *et al.*, 1982).

Southern analysis and colony hybridization

Large-scale preparations of total DNA from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942 were made according to Mazur *et al.* (1980) and Curtis and Haselkorn (1983). Restricted DNA was separated in agarose gels with Tris-borate-EDTA as electrophoresis buffer (Maniatis *et al.*, 1982). Blotting of the separated DNA from the agarose gels to nitrocellulose filters was performed bidirectionally (Meinkoth and Wahl, 1984). After being baked *in vacuo* at 80°C for 30 min, the blots were washed at room temperature in 4X SSC for 30 min.

Hybridization under stringent conditions with homogeneously labelled DNA probes was performed at 65°C. Prehybridization was for at least 30 min at 65°C in a mix of 6.6X SSC, 10X Denhardt's solution, 0.1% SDS, 0.05% PPi and 0.1 mg ml^{-1} denatured herring sperm DNA (0.25 ml hybridization mix cm^{-2} nitrocellulose) (Maniatis *et al.*, 1982; Meinkoth and Wahl, 1984). Hybridization was performed by adding denatured probe DNA, labelled to high specific activity (≥3.10^8 dpm μg^{-1} DNA) with ^{32}P by random priming with oligonucleotides (Feinberg and Vogelstein, 1983), to the prehybridization medium and continuing the incubation at 65°C for 16 h. The hybridized blots were washed for 10–30 min at 65°C, initially three times in 5X SSC + 0.1% SS and once in 5X SSC. Next, the washed filters were blotted dry, wrapped in plastic foil while still moist and then autoradiographed at −20°C, using X-ray film and intensifying screens (Eastman Kodak, Rochester, NY, USA). For hybridization analysis under low-stringency conditions, the above procedure was carried out at 50°C.

For oligonucleotide hybridization to Southern blots, the same hybridization procedure was applied at an initial temperature of 37°C, but the stringency had to be enhanced by washing at 45°C to obtain specific signals. The only further modification was the omission of herring sperm DNA from the hybridization medium, in order to avoid loss of probe by non-specific hybridization, thereby risking a somewhat higher background noise. The (mixed) oligonucleotide probes were end-labelled with γ-[^{32}P] and T4 polynucleotide kinase (Maniatis *et al.*, 1982).

One hundred colonies originating from a transformation experiment were transferred to agar plates in ordered arrays and after incubation at 37°C overnight, the colonies were transferred to nitrocellulose filters (HATF; Millipore, Molsheim, France) *in duplo*. After processing of the colony filters (Maniatis *et al.*, 1982), hybridization was performed exactly as described above for the Southern blots.

Cloning and sequence analysis

Plasmid DNAs were prepared using the alkaline lysis method (Maniatis *et al.*, 1982) for large-scale isolations. 'Minipreps' of plasmid DNA were prepared by a modified 'Boiling' procedure

(Van der Plas *et al.*, 1988). For cloning or for probe preparation DNA fragments were isolated from agarose gels with DEAE-membrane (Lizardi *et al.*, 1984). Dephosphorylation of linearized vector DNAs with calf-intestinal phosphatase (molecular biology grade; Boehringer, Mannheim, FRG) was carried out for 30 min at 37°C. The phosphatase was inactivated by the addition of 20 mM EGTA and heating of the sample for 15 min at 68°C, followed by phenol/chloroform extraction. DNA ligations were performed according to Dugaiczyk *et al.* (1975). Transformation to *E. coli* PC2495 was performed using standard procedures (Maniatis *et al.*, 1982).

DNA sequences were determined by the dideoxy chain termination method of Sanger *et al.* (1980) with [^{35}S]-dATP as radiolabel, with both specific and standard primers. Templates for sequencing were single-stranded DNAs of *in vivo* packaged pEMBL8, −9, −18 and −19 (Dente *et al.*, 1983) clones. The region of interest was sequenced several times and in both directions. Analysis of the sequences produced was performed with the aid of the computer program MICROGENIE (Queen and Korn, 1984) and the sequence analysis package of Stephens (1985).

Northern hybridization analysis

Total RNA was isolated from mid-exponential cultures of *Anabaena* sp. PCC 9737 and *Synechococcus* sp. PCC 7942. The cells were harvested by centrifugation and washed with a buffer containing 20 mM sodium acetate (pH 5.2) and 0.15 M sucrose. The collected cells were subsequently resuspended in the same buffer, frozen in liquid nitrogen and ground in a mortar, while frozen. The solution obtained after thawing the powder was brought to 1% SDS and extracted with phenol (saturated with 20 mM sodium acetate, pH 5.2, 1 mM EDTA and 0.5% SDS) at 65°C. After back-extraction of the phenol phase with chloroform and further phenol-, phenol/chloroform- and chloroform extractions, the RNA was precipitated with ethanol. The RNA was dissolved in a buffer with 20 mM sodium acetate, pH 5.2, 1 mM EDTA and 0.5% SDS and finally stored as an ethanol precipitate at −20°C. After denaturation with formamide-formaldehyde, RNA was analysed on formaldehyde-agarose gels run in 10 mM sodium phosphate buffer (Meinkoth and Wahl, 1984). RNA was blotted to nylon filters and hybridized with a homologous probe at 65°C in aqueous medium, as described for the Southern blots.

References

Aitken, A. (1975) Prokaryote-eukaryote relationships and the amino-acid sequence of plastocyanin from *Anabaena variabilis*. *Biochem J* **149:** 675–683.

Aitken, A. (1976) Protein evolution in cyanobacteria. *Nature* **623:** 793–796.

Boulter, D., Haslett, B.G., Peacock, D., Ramshaw, J.A.M., and Scawen, M.D. (1977) Chemistry, function, and evolution of plastocyanin. In *International Rev of Biochem, Plant Biochemistry II. Vol. 13.* Northcote, D.H. (ed.). Baltimore: University Park Press, pp. 1–40.

Brawerman, G. (1987) Determinants of messenger RNA stability. *Cell* **48:** 5–6.

Crofts, A.R., and Wood, P.M. (1978) Photosynthetic electron-transport chains of plants and bacteria and their role as proton pumps. *Curr Top Bioenerg* **7:** 175–244.

Curtis, S.E., and Haselkorn, R. (1983) Isolation and sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from the cyanobacterium *Anabaena* 7120. *Proc Natl Acad Sci USA* **80:** 1835–1839.

Davis, D.J., Krogmann, D.W., and San Pietro, A. (1980) Electron donation to Photosystem I. *Plant Physiol* **65:** 697–702.

Dente, L., Cesareni, G., and Cortese, R. (1983) pEMBL: a new family of single-stranded plasmids. *Nucl Acids Res* **11:** 1645–1655.

Dugaiczyk, A., Boyer, H.W., and Goodman, H.M. (1975) Ligation of *Eco*RI endonuclease-generated DNA fragments into linear and circular structures. *J Mol Biol* **96:** 171–184.

Duyvesteyn, M.G.C., Korsuize, J., De Waard, A., Vonshak, A., and Wolk, C.P. (1983) Sequence-specific endonucleases in strains of *Anabaena* and *Nostoc*. *Arch Microbiol* **134:** 276–281.

Feinberg, A.P., and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analyt Biochem* **132:** 6–13.

Gray, M.W., and Doolittle, W.F. (1982) Has the endosymbiont hypothesis been proven? *Microbiol Rev* **46:** 1–42.

Guss, J.M., and Freeman, H.C. (1983) Structure of oxidized poplar plastocyanin at 1.6 Å resolution. *J Mol Biol* **169:** 521–563.

Hawley, D.K., and McClure, W.R. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl Acids Res* **11:** 2237–2255.

Herrero, A., Elhai, J., Hohn, B., and Wolk, C.P. (1984) Infrequent cleavage of cloned *Anabaena variabilis* DNA by restriction endonucleases from *A. variabilis*. *J Bacteriol* **160:** 781–784.

Ho, K.K., Ulrich, E.L., Krogmann, D.W., and Gomez-Lojero, C. (1979) Isolation of photosynthetic catalysts from cyanobacteria. *Biochim Biophys Acta* **545:** 236–248.

Ho, K.K., and Krogmann, D.W. (1982) Photosynthesis. In *The Biology of Cyanobacteria.* Carr, N.G., and Whitton, B.A. (eds). Oxford: Blackwell Scientific Publications, pp. 204–207.

Ho, K.K., and Krogmann, D.W. (1984) Electron donors to P-700 in cyanobacteria and algae: an instance of unusual genetic variability. *Biochim Biophys Acta* **766:** 310–316.

Kelly, J., and Ambler, R.P. (1974) The amino acid sequence of plastocyanin from *Chlorella fusca*. *Biochem J* **143:** 681–690.

Kuhlemeier, C.J., Thomas, A.A.M., Van der Ende, A., Van Leen, R.W., Borrias, W.E., Van den Hondel, C.A.M.J.J., and Van Arkel, G.A. (1983) A host-vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. *Plasmid* **10:** 156–163.

Lizardi, P.M., Binder, R., and Short, S.A. (1984) Preparative isolation of DNA and biologically active mRNA from diethyl-aminoethyl membrane. *Gene Analyt Tech* **1:** 33–39.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Mazur, B.J., Rice, D., and Haselkorn, R. (1980) Identification of blue-green algal nitrogen fixation gene by using heterologous DNA hybridization probes. *Proc Natl Acad Sci USA* **77:** 186–190.

Meinkoth, J., and Wahl, G. (1984) Hybridisation of nucleic acids immobilized on solid supports. *Analyt Biochem* **138:** 267–284.

Merchant, S., and Bogorad, L. (1986a) Regulation by copper of the expression of plastocyanin and cytochrome c_{552} in *Chlamydomonas reinhardtii*. *Mol Cell Biol* **6:** 462–469.

Merchant, S., and Bogorad, L. (1986b) Rapid degradation of apoplastocyanin in Cu(II)-deficient *Chlamydomonas reinhardtii*. *J Biol Chem* **261:** 15850–15853.

Merchant, S., and Bogorad, L. (1987) Metal ion regulated gene expression: use of a plastocyanin-less mutant of *Chlamydomonas reinhardtii* to study the Cu(II)-dependent expression of cytochrome c_{552}. *EMBO J* **6:** 2531–2535.

Nielsen, P.S., and Gausing, K. (1987) The precursor of barley plastocyanin: sequence of cDNA clones and gene expression in different tissues. *FEBS Lett* **225:** 159–162.

Pugsley, A.P., and Schwartz, M. (1985) Export and secretion of proteins by bacteria. *FEMS Microbiol Rev* **32:** 3–38.

Queen, C., and Korn, L.J. (1984) A comprehensive sequence analysis program for the IBM personal computer. *Nucl Acids Res* **12:** 581–599.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111:** 1–61.

Rother, C., Jansen, T., Tyagi, A., Tittgen, J., and Herrmann, R.G. (1986) Plastocyanin is encoded by an uninterrupted nuclear gene in spinach. *Curr Genet* **11:** 171–176.

Sako, T. (1985) Overproduction of staphylokinase in *Escherichia coli* and its characterization. *Eur J Biochem* **149:** 557–563.

Sako, T., and Tsuchida, N. (1983) Nucleotide sequence of the staphylokinase gene from *Staphylococcus aureus*. *Nucl Acids Res* **11:** 7679–7693.

Sandmann, G. (1986) Formation of plastocyanin and cytochrome c_{553} in different species of blue-green algae. *Arch Microbiol* **145:** 76–79.

Sandmann, G., and Böger, P. (1980) Copper-induced exchange of plastocyanin and cytochrome c_{553} in cultures of *Anabaena variabilis* and *Plectonema boryanum*. *Plant Sci Lett* **17:** 417–424.

Sandmann, G., and Böger, P. (1983) The enzymological function of heavy metals and their role in electron transfer processes of plants. In *Encyclopedia of Plant Physiology 15*. Lauchli, A., and Bieleski, R.K. (eds). Berlin: Springer Verlag, pp. 563–596.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J Mol Biol* **143:** 161–178.

Shestakov, S.V., and Khuyen, N.T. (1970) Evidence for genetic transformation in the blue-green alga *Anacystis nidulans*. *Mol Gen Genet* **107:** 372–375.

Simpson, R.J., Moritz, R.L., Nice, E.C., Grego, B., Yoshizaki, F., Sugimura, Y., Freeman, H.C., and Murata, M. (1986) Complete amino-acid sequence of plastocyanin from a green alga, *Enteromorpha prolifera*. *Eur J Biochem* **137:** 497–506.

Smeekens, S., De Groot, M., Van Binsbergen, J., and Weisbeek, P. (1985) Sequence of the precursor of the chloroplast thylakoid lumen protein plastocyanin. *Nature* **317:** 456–458.

Smeekens, S., and Weisbeek, P. (1988) Protein transport towards the thylakoid lumen: post-translational translocation in tandem. *Photosynth Res* **16:** 177–186.

Stephens, R.M. (1985) A sequencers sequence analysis package for the IBM PC. *Gene Analyt Techn* **2:** 67–75.

Tandeau de Marsac, N., and Houmard, J. (1987) Advances in cyanobacterial molecular genetics. In *The Cyanobacteria*. Fay, P., and Van Baalen, C. (eds). Amsterdam: Elsevier Biomedical Press, pp. 251–302.

Tinoco, J., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., and Cralla, J. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nature New Biol* **246:** 40–41.

Tomioka, N., and Sugiura, M. (1983) The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue-green alga, *Anacystis nidulans*. *Mol Gen Genet* **191:** 46–50.

Van der Plas, J., De Groot, R., Woortman, M., Cremers, F., Borrias, M., Van Arkel, G., and Weisbeek, P. (1988) Genes encoding ferredoxins from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942: structure and regulation. *Photosynth Res* **18:** 179–204.

Vieira, J., and Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19:** 259–268.

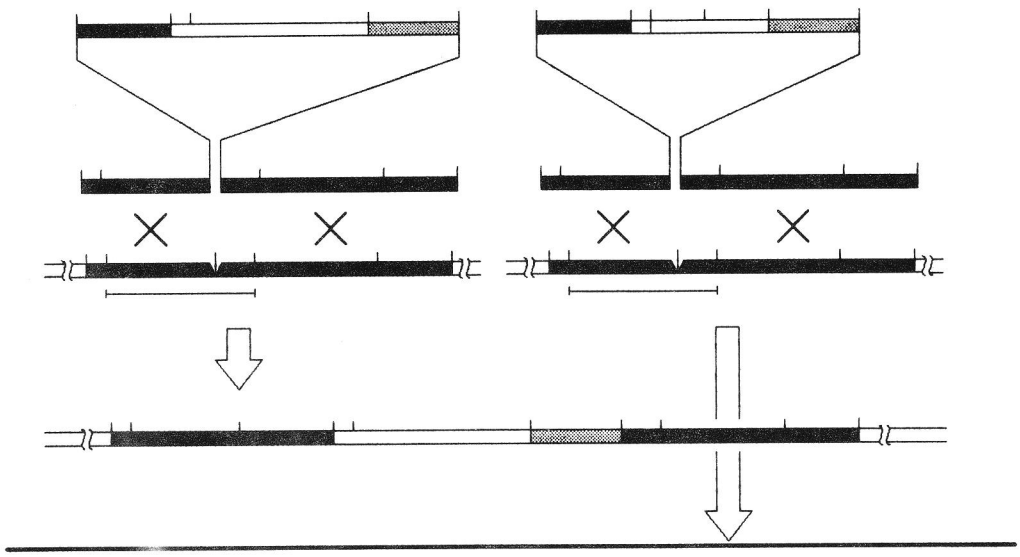
Vorst, O., Oosterhoff-Teertstra, R., Vankan, P., Smeekens, S., and Weisbeek, P. (1988) Plastocyanin of *Arabidopsis thaliana*; isolation and characterization of the gene and chloroplast import of the precursor protein. *Gene* **65:** 59–69.

Watson, M.E.E. (1984) Compilation of published signal sequences. *Nucl Acids Res* **12:** 5145–5164.

Wood, P.M. (1978) Interchangeable copper and iron proteins in algal photosynthesis: studies on plastocyanin and cytochrome c_{552} in *Chlamydomonas*. *Eur J Biochem* **87:** 9–19.

Chapter IV

Genomic integration system based on pBR sequences for the cyanobacterium *Synechococcus* sp. PCC 7942: transfer of genes encoding plastocyanin and ferredoxin



J. van der Plas, H. Hegeman, G. de Vrieze,
M. Tuyl, M. Borrias and P. Weisbeek,
accepted for publication in GENE

Genomic integration system based on pBR sequences for the cyanobacterium *Synechococcus* sp. PCC 7942: transfer of genes encoding plastocyanin and ferredoxin

Summary

Synechococcus sp. PCC 7942 recipient strains were constructed for the chromosomal integration of DNA fragments cloned in any pBR322 derived vector, which carries the Ap^R marker. The construction was based on the incorporation of specific recombination targets, the so-called "integration platforms", into the chromosomal *met1* gene. These platforms consist of an incomplete *bla* gene (Ap^S) and the pBR322 origin of replication separated from each other by a gene encoding an antibiotic resistance (Sm^R or Km^R). Recombination between a pBR-derived donor plasmid and such a chromosomal platform results with high frequency in restoration of the *bla* gene and replacement of the chromosomal marker (Sm^R or Km^R) by the insert of the donor plasmid. The integration into the platform depends on recombination between pBR *ori* and *bla* sequences only and is therefore independent of the DNA insert to be transferred. The desired recombinants are found by selection for a functional *bla* gene (Ap^R) and subsequent screening for absence of the chromosomal antibiotic marker. Gene transfer with this integration system was found to occur efficiently and reliably. Furthermore, the presence of the pBR322 origin of replication in the platform allowed for "plasmid rescue" of integrated sequences.

The system was applied successfully for the transfer of the gene encoding plastocyanin (*petE1*) from *Anabaena* sp. PCC 7937 and for the integration of an extra copy of the gene encoding ferredoxin I (*petF1*) from *Synechococcus* sp. PCC 7942 itself.

Introduction

Cyanobacteria offer a prokaryotic model system for the study of plant-type oxygenic photosynthesis, as they possess a photosynthetic apparatus very similar to that of higher plant and algal chloroplasts. Particularly attractive in this regard are those cyanobacterial strains that are amenable to genetic manipulation by transformation e.g. *Synechococcus* sp. PCC 7942, *Synechococcus* sp. PCC 7002 and *Synechocystis* sp. PCC 6803 (Porter, 1985; Kuhlemeier and van Arkel, 1987) or by conjugation, e.g. *Anabaena* sp. PCC 7120 (Wolk *et al.*, 1984; Thiel and Wolk, 1987).

For the unicellular obligate photo-autotrophic *Synechococcus* sp. PCC 7942 in particular, a broad spectrum of (shuttle) plasmid vectors has been developed (Van den Hondel *et al.*,1980; Kuhlemeier *et al.*, 1983; for a review see: Tandeau de Marsac and Houmard, 1988). These autonomously replicating vectors are based on the smaller of the two cryptic plasmids present in this strain, pUH24 (Van den Hondel *et al.*,1980). Use in the functional analysis of cloned *Synechococcus* DNA of these vectors has been limited by the lack of a recombination-deficient strain of *Synechococcus* sp. PCC 7942, as it has been shown that recombination between a cloned gene on a vector and its chromosomal counterpart occurs with high frequency (Kuhlemeier *et al.*, 1985).

A different approach for genetic manipulation of *Synechococcus* sp. PCC 7942 is to exploit homologous recombination for the stable integration of specific DNA fragments, of exogenous or endogenous origin, at a defined site of the chromosome (Golden et *al.*, 1987). Integration events of that kind have been analysed in detail by Szalay and coworkers (Williams and Szalay, 1983; Kolowsky *et al.*, 1984; Kolowsky and Szalay, 1986) Transformation of *Synechococcus* sp. PCC 7942 with not autonomously replicating plasmids, consisting of a fragment of *Synechococcus* sp. PCC 7942 chromosomal DNA cloned in an *E. coli* vector (pBR322) and interrupted by a piece of foreign DNA marked by an antibiotic resistance gene, generated three types of transformants. These differed with respect to the foreign element (pBR322 and/or the interrupting DNA) they have acquired. In the major type transformant (more than 90%), the sequence in the recipient chromosome is replaced by its interrupted counterpart from the donor molecule ("allele replacement"). This structure can be the result of either double crossing-over or gene conversion. The same procedure has proven to be useful for insertional inactivation of several genes in *Synechococcus* sp. PCC 7942, at least if their function was dispensable (for a discussion see: Van der Plas *et al.*, 1988).

We have exploited the possibilities of homologous recombination in *Synechococcus* sp. PCC 7942 by the construction of two strains containing an "integration platform" in their chromosome. At such a chromosomal platform, any fragment cloned in an Ap^R pBR322-derived vector (Balbás *et al.*, 1986) can integrate. The location of the platforms is the *metI* gene of *Synechococcus* sp. PCC 7942, which can be inactivated without affecting the viability of the host strain when grown on methionine supplemented media (Tandeau de Marsac *et al.*, 1982; Kuhlemeier *et al.*, 1985).

Materials and methods

(a) Strains, plasmids and growth conditions

A list of the strains and plasmids used in this study is given in Table I. The cyanobacterial strains were grown in the light in BG11 medium as described (Van den Hondel *et al.*, 1980). Strain R2-PIM8 was cultured in BG11 medium supplemented with 30 µg/ml DL-methionine and 5 µg/ml Sm. Strain R2-PIM9 was cultured in BG11 medium supplemented with 30 µg/ml DL-methionine and 7.5 µg/ml Km.

E. coli PC2495 (Van der Plas *et al.*, 1988) was the host for all plasmids. The following antibiotic concentrations were added to standard growth media for the selection of plasmids in *E. coli*: 50 µg/ml Ap, 50 µg/ml Km and 25 µg/ml Sm.

(b) DNA and RNA manipulations

Basic techniques (e.g., plasmid and chromosomal DNA isolation, gel electrophoresis, restriction fragment purification, ligation, transformation, oligo-primed labelling, Southern hybridisation, RNA extraction and Northern analysis) were performed by standard methods (Maniatis *et al.*, 1982), or modified as described by Van der Plas *et al.* (1988; 1989).

(c) Transformation of Synechococcus *sp. PCC 7942*

Transformation of *Synechococcus* sp. PCC 7942 platform strains R2-PIM8 and R2-PIM9 (Table I) with pBR-derived plasmids was carried out by a modification of the procedure described by Van den Hondel *et al.* (1980). Cells were grown into log phase ($5x10^7$-$1x10^8$ cells/ml), washed with fresh medium, concentrated to 1-2 $x10^9$ cells/ml in BG11 supplemented with methionine and used immediately. To 100 µl of cells 10 µl of plasmid DNA solution was added (approximately 100-200 ng/ml final concentration). The cell suspension was incubated in the light for 1 hr and plated in different amounts: 10^8, 10^7 and 10^6 cells on fresh, BG11 solid medium (60 ml with 1% Difco agar per 11cm plate) supplemented with methionine. After incubation of the plates first in the dark for 16 hr and then in the light for 8 hr, ampicillin was applied underneath the agar. By adjustment of the Ap concentration to the amount of cells used for plating (see Table II), a number of separate transformant colonies sufficient for further handling was readily obtained (approximately 10^3cfu/µg DNA).

The transformant colonies obtained after 5-7 days were rechecked for Ap resistance by transferring them to BG11 plates with methionine and Ap (0.3 µg/ml for R2-PIM8 and 0.2

Table I. Strains and Plasmids

Strain/plasmid	Relevant properties	Reference
Escherichia coli PC2495	*recA*, *hsdS*, *supE*, *thi*, Δ(*lac-proAB*), [F', *traD*36, *proAB*, *lacI*^{q} *Z* M15]	Van der Plas *et al.* (1988)
Anabaena sp. sp. PCC 7937	(*Anabaena variabilis* ATCC 29413)	Duyvesteyn *et al.* (1983)
Synechococcus sp. PCC 7942	Wild type (*Anacystis nidulans* R2), in this study the small-plasmid-cured derivative R2-SPc has been used throughout, Ap^S Km^S Sm^S Met$^+$	Kuhlemeier *et al.* (1983)
R2-PIM8	Derivative of R2-SPc with pPIM8 integrated into the *met1* gene, Ap^S Sm^R Met$^-$	This study
R2-PIM9	Derivative of R2-SPc with pPIM9 integrated into the *met1* gene, Ap^S Km^R Met$^-$	This study
pUC7-K[a]	3.9 kb, Ap^R Km^R	This study
pUC7-S[b]	4.7 kb, Ap^R Sm^R	This study
pEMBL8	4 kb, Ap^R	Dente *et al.* (1987)
pIM1[c]	6.6 kb, Ap^R, 3.7 kb *Cla*I-*Eco*RI *met1* fragment	This study
pPIM8	7.6 kb, Ap^S Sm^R, intermediate in the platform construction	This study
pPIM9	7,0 kb, Ap^S Km^R, intermediate in the platform construction	This study
pPCVC1	Ap^R, *Anabaena petE1* encoding the precursor of plastocyanin on a 1.1 kb fragment	Van der Plas *et al.* (1989)
pRA81[d]	Ap^R, *petF1* encoding ferredoxin I orientation opposite to *bla* gene	This study
pRA82[e]	Ap^R, *petF1* encoding ferredoxin I same orientation as *bla* gene	This study

[a] Plasmid pUC7-K was made by transferring the *Pst*I fragment carrying the *neo* gene from pUC4-K (Vieira and Messing, 1982) to the *Pst*I site in the polylinker of pUC7 (Vieira and Messing, 1982). The direction of the *neo* gene is opposite to that of the gene for β-lactamase (*bla*), as determined by restriction mapping.

[b] Plasmid pUC7-S was made by ligation of the *Bam*HI fragment from pHP45-Ω (Prentki and Krisch, 1984), made blunt ended with Mung bean nuclease, to the *Eco*RI vector fragment of pUC7 made blunt by filling in with Klenow. In this way the *Bam*HI sites were removed, but the *Eco*RI sites were restored. The Omega fragment encodes an aminoglycoside adenylyltransferase (*aad*) conferring resistance to Sm, bordered by transcriptional and translational stop signals in both orientations and by polylinker regions. The direction of the *aad* gene is the same as that of the *bla* gene, as determined by restriction mapping.

[c] Plasmid pIM1 consists of the 3.7 kb *Cla*I-*Eco*RI fragment from the *met1* gene region (Kuhlemeier *et al.*, 1985) spliced to the 2.9 *Cla*I-*Eco*RI vector fragment from pEMBL8 (Dente *et al.*, 1987).

[d][e] Plasmids pRA81 and pRA82 contain the complete *petF1* gene of *Synechococcus* sp. PCC 7942 in opposite orientations. They were constructed by insertion of the 1.1 kb *Pst*I genomic fragment for which the nucleotide sequence has been determined (Van der Plas *et al.*,1988) into the *Pst*I site of pEMBL8.

µg/ml for R2-PIM9). The Ap^R transformants were subsequently checked for loss of either Sm resistance (R2-PIM8) or Km resistance (R2-PIM9) on BG11 with methionine, Ap, and Sm or Km. Some of the colonies with the required phenotype, Ap^RSm^S in the case of R2-PIM8 and Ap^RKm^S in the case of R2-PIM9, were colony purified, i.e. transformants were grown to log phase in methionine- supplemented BG11 with Ap (0.3 µg/ml for R2-PIM8 and 0.2 µg/ml for R2-PIM9 transformants, respectively) and plated for single colonies on BG11 with methionine and Ap (0.3 µg/ml for R2-PIM8 and 0.1 µg/ml for R2-PIM9 transformants, respectively).

Table II. Ampicillin concentrations applied for the selection of transformant colonies in the integration system

Strain	Plates with 10^8 cells	Plates with 10^7 cells	Plates with 10^6 cells
R2-PIM8	0.7 µg Ap/ml[a]	0.3 µg Ap/ml	0.2 µg Ap/ml
R2-PIM9	0.4 µg Ap/ml	0.2 µg Ap/ml	0.1 µg Ap/ml

[a] Only freshly made Ap solutions should be used.

Results and discussion

(a) Development of a system for genomic integration

Many of the vectors that are used in recombinant DNA research today are members of the pBR family (Balbás *et al.*, 1986). It was our aim to exploit the fact that most of these vectors have the origin of replication (*ori*) and the ß-lactamase gene (*bla*) in common, to facilitate the genetic manipulation of the cyanobacterium *Synechococcus* sp. PCC7942. The strategy employed was to create a target in the chromosomal DNA where the common elements of the cloning vectors will integrate via homologous recombination. The target was named "integration platform" and consisted of a promoterless *bla* gene fragment and a fragment containing the complete pBR322 origin of replication, separated from each other by a functional antibiotic resistance gene. We intended the integration of a pBR-derived donor molecule into this platform to take place by recombination both in the promoterless *bla* gene and in the origin sequence, such that this will result in the formation of a functional *bla* gene (Ap^R) and concomitantly in the replacement of the platform-sited resistance marker by the cloned insert of the donor plasmid ("insert replacement").

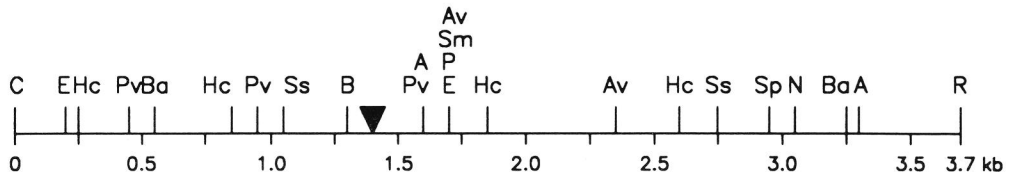


Fig. 1. Restriction map of a 3.7 kb *Cla*I–*Eco*RI fragment containing (part of) the *Synechococcus* sp. PCC 7942 *met1* gene. Restriction sites are: A, *Acc*I; Av, *Ava*I; B, *Bam*HI; Ba, *Bal*I; C, *Cla*I; E, *Eco*RV; Hc, *Hin*cII; N, *Nco*I; P, *Pst*I ; Pv, *Pvu*II; R, *Eco*RI; Sm, *Sma*I; Sp, *Sph*I; Ss, *Ssp*I. No sites were found for: *Apa*I, *Bgl*I, *Bgl*II, *Bst*EII, *Hin*dIII, *Hpa*I, *Kpn*I, *Nae*I, *Nsp*V, *Sal*I, *Tth*111I, *Xba*I, *Xho*I. The site of the Tn901 insertion resulting in methionine auxotrophy is indicated (▼).

(b) The met1 gene as chromosomal site for integration

The chromosomal *met1* gene involved in the pathway of methionine biosynthesis, can be inactivated without affecting the viability of the host strain (Tandeau de Marsac *et al.*, 1982; Kuhlemeier *et al.*, 1985). A restriction map of the relevant part of the *met1* gene region, a 3.7 kb *Cla*I–*Eco*RI fragment, is shown in Fig. 1. Insertion of transposon *Tn*901 in the *met1* gene close to the *Bam*HI site resulted in a methionine auxotroph called Met–1 (Tandeau de Marsac *et al.*, 1982). Also insertions in the *Bam*HI site give Met$^-$ mutants (Van der Plas and Tuyl, unpublished results). These auxotrophs grow with wildtype efficiency in methionine–supplemented media. Therefore the *Bam*HI site in the *met*–region seems suitable for the integration of foreign DNA.

(c) Construction of the integration platform plasmids

Two platforms designed for genomic integration were initially constructed in the form of the plasmids pPIM8 (Ap^SSm^R) and pPIM9 (Ap^SKm^R) (Table I). Each plasmid consists of an "integration platform" and flanking *met1*–sequences necessary for integration of the platform into the chromosome via homologous recombination (Fig. 2).

The integration platform of pPIM8 consists of the 1.9 kb *Ssp*I–*Pvu*II fragment of pUC19 containing the *ori* region plus the *bla* structural gene but not the *bla* promoter (Yanish–Perron *et al.*, 1985), linked to a *Bam*HI fragment (made blunt with Mung bean nuclease) carrying the *aad* gene. This 2.0 kb *Bam*HI fragment is derived from the "Omega" element (Prentki and Krisch, 1984), which consists of the Sm^R/Sp^R segment (*aad*) from R100.1 flanked by inverted repeats carrying the T4 transcription–termination signals, translational stop signals and a short polylinker.

In the integration platform of pPIM9 a 1.9 kb *Sph*I–"*Bam*HI/*Pvu*II" fragment from the pPIM8 platform with most of the Omega element was replaced by the 1.3 kb *Sal*I fragment from pUC4–K (Vieira and Messing, 1982) conferring resistance to kanamycin (*neo*). In this way the transcription terminator separating the *bla* structural gene from the *aad* gene was left in its place, in order to prevent transcription of the *bla* gene from the *neo* promoter.

The precursors described above were provided with flanking *met1* fragments by linearizing them between the *bla* gene and *ori* by partial *Xho*II digestion and linking them to the *Cla*I–*Bam*HI and *Bam*HI–*Eco*RI *met1* fragments as shown in Fig. 2A. In the ligation of the *Bam*HI sticky ends to the compatible *Xho*II sticky ends of the platform only the hybrid site located between *ori* and *met1* DNA became recleavable with *Bam*HI. The *Cla*I end was ligated to the *Eco*RI end, having both sites made blunt by filling in with Klenow, thus regenerating the *Eco*RI cleavage site. All constructs were checked by extensive restriction mapping in order to make certain that they were assembled correctly.

(d) Transfer of the platforms to the chromosomal met1 gene

Both platform vectors pPIM8 and pPIM9 were cleaved at their unique *Eco*RI site thus separating the flanking *met*1 fragments and this linear DNA was used to transform

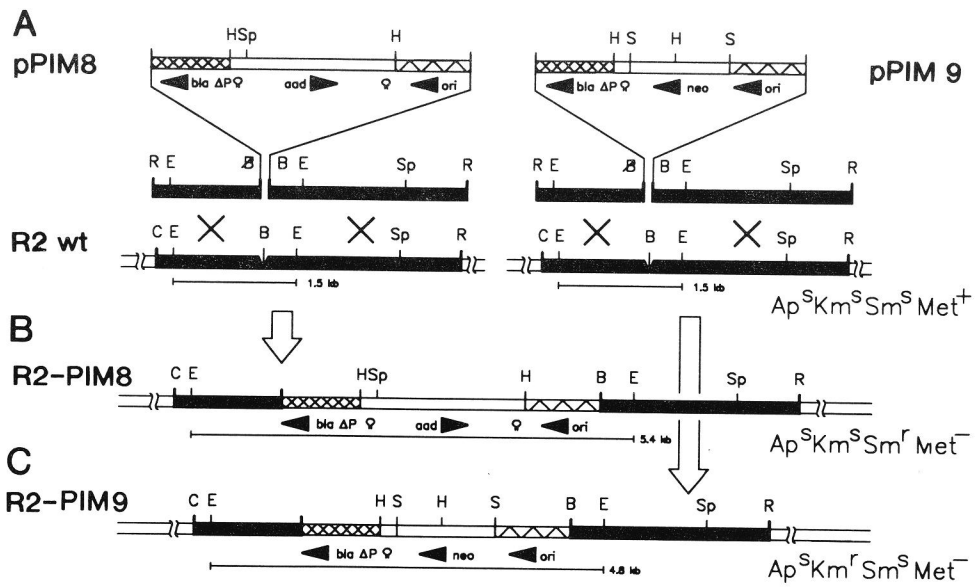


Fig. 2. Construction of the integration platforms in the *met1* gene of *Synechococcus* sp. PCC 7942. **(A)** Composition of the platform vectors pPIM8 and pPIM9 and their transfer by homologous recombination to the *met1* gene. Restriction sites are indicated by the abbreviations as stated for Fig. 1, with the addition of H, *Hin*dIII and S, *Sal*I. The ♀ symbols represent *E. coli* transcription termination signals. **(B)** Map of the platform region in strain R2–PIM8. **(C)** Map of the platform region in strain R2–PIM9. PIM is an acronym for "Platform for Integration in the Methionine gene".

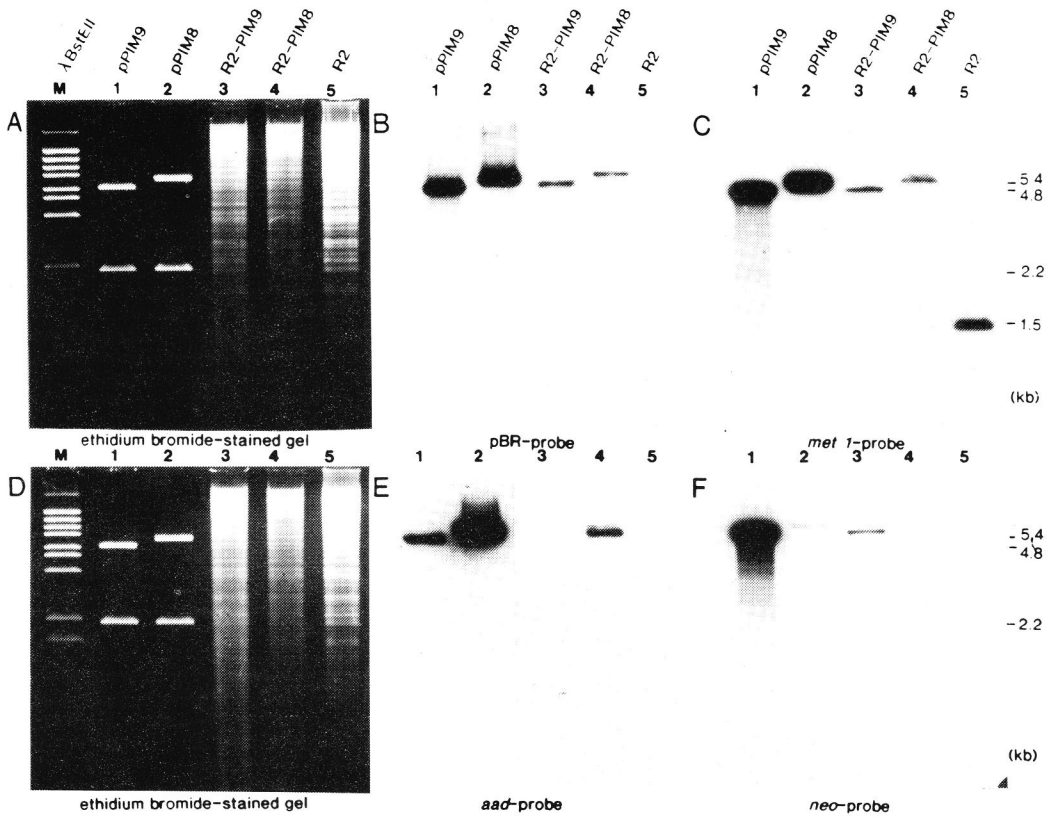


Fig. 3. Southern analysis of the platform strains R2-PIM8 and R2-PIM9. Two identical agarose gels (panel A and D) were run with the following samples: *Eco*RV digests of chromosomal DNAs from wildtype (lane 5), R2-PIM8 (lane 4) and R2-PIM9 (lane 3), together with *Eco*RV digests of the plasmid DNAs pPIM8 (lane 2), pPIM9 (lane 1) and a *Bst*EII digest of lambda DNA (lane M). The DNAs were blotted bidirectionally to nitrocellulose filters, (panels B, C, E, and F). The blots were hybridized with one of the following ^{32}P-labelled probes: total pUC18 DNA (pBR-probe; panel B), the central 1.5kb *Eco*RV fragment of the *met1* region, containing the *Bam*HI site (*met1*-probe; panel C), the 0.9kb *Bst*EII-*Sph*I fragment from the Omega element (*aad*-probe; panel E) and the 1.3kb *Pst*I *neo* fragment from pUC4-K (*neo*-probe; panel F).

wildtype *Synechococcus* sp. PCC7942 (Fig. 2). The transformed cells were plated on methionine-supplemented medium and selected for Sm^R (in the case of pPIM8), or for Km^R (in the case of pPIM9). Two transformants, designated R2-PIM8 ($Ap^S Sm^R Met^-$) and R2-PIM9 ($Ap^S Km^R Met^-$), were chosen for further analysis (Fig. 2). Chromosomal DNAs of the wildtype strain and platform strains R2-PIM8 and R2-PIM9 were analysed by Southern blotting to determine the exact nature of the integration. DNA digested with *Eco*RV was hybridized with different radiolabelled DNA-probes (Fig. 3). *Eco*RV cuts only in the flanking *met1* sequences and not in the integration platform itself. With the 1.5 kb *Eco*RV *met1*-probe

(Fig. 3C) the corresponding 1.5 kb *Eco*RV wildtype fragment (Fig. 3C, lane 5) is not detected in the DNAs from the transformed strains (Fig. 3C, lanes 3 and 4). Instead a 5.4 kb fragment (R2-PIM8, lane 4) and a 4.8 kb fragment (R2-PIM9, lane 3) reacted with the *met1*-probe. Fragments of the same size did hybridize in the digests of the corresponding platform vectors pPIM8 and pPIM9 (Fig. 3C, lanes 1 and 2). The labelled pUC18 DNA used as probe specific for the pBR-derived parts of the platform did not hybridize to wildtype DNA (Fig. 3B, lane 5), but did hybridize to the same fragments of the platform strains (Fig. 3B, lanes 3 and 4) and the platform vectors (Fig. 3B, lanes 1 and 2) as with the *met1*-probe (Fig. 3C). Hybridization with an *aad* gene probe (Fig. 3E) showed the 5.4 kb signal both in the chromosomal DNA of R2-PIM8 and in plasmid pPIM8, whereas a *neo* gene probe (Fig. 3F) produced the 4.8 kb signal in the chromosomal DNA of R2-PIM9 and in plasmid pPIM9. The minor crossreactions visible in the lanes of the plasmid DNAs are due to minute amounts of vector DNA contaminating the *aad* and *neo* gene probes.

The results of these Southern hybridizations show that the plasmid DNA of pPIM8 and pPIM9 had integrated in the chromosome of R2 as expected (see Fig. 2). Strains R2-PIM8 and R2-PIM9 were subsequently tested for their utility in serving as acceptor for genes cloned in pBR-derived vectors.

Table III. Transformation of pBR-derived plasmids to the platform strains R2-PIM8

	Host	Plasmid[a]	Transformant phenotype			
			Ap^R Sm^S [b]		Ap^R Sm^S Km^R [c]	
A)	R2-PIM8	pUC7-K	126	(85%)	126	(100%)
			Ap^R Km^S		Ap^R Km^S Sm^R	
	R2-PIM9	pUC7-S	58	(100%)	58	(100%)
B)	R2-PIM9	pPCVC1	53	(73%)		
	R2-PIM9	pRA81	35	(45%)		
	R2-PIM9	pRA82	25	(33%)		

[a] Plasmid donor DNA was not linearised prior to transformation.

[b] Efficiency as percentage of Ap^R transformants. The data given are totals from 2-3 independent experiments.

[c] Efficiency as percentage of the fraction Ap^R transformants that lost their platform resistance upon transformation.

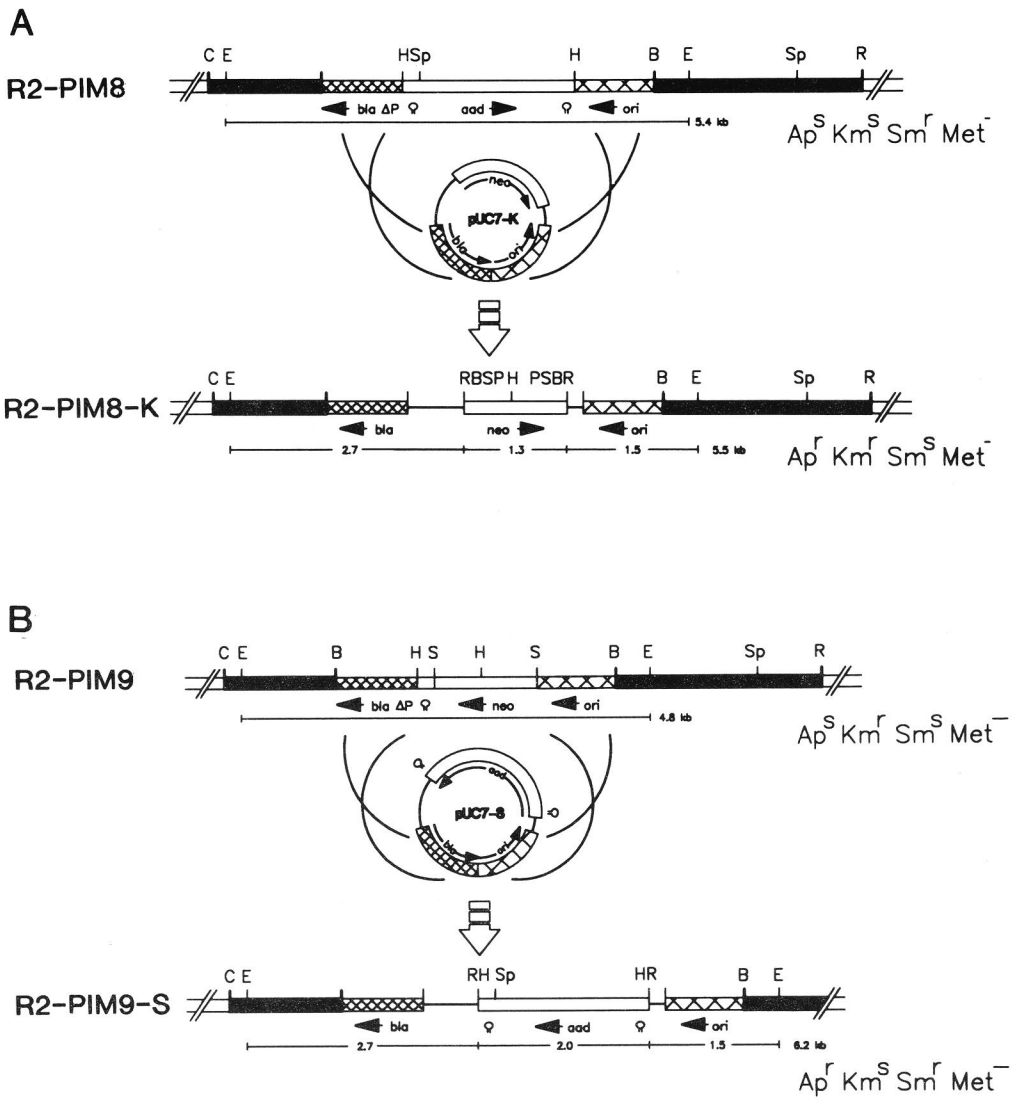


Fig. 4. Mechanism of the platform-mediated transfer of plasmid-borne antibiotic resistance genes to the genome of *Synechococcus* sp. PCC7942. **(A)** When strain R2-PIM8 ($Ap^{S}Sm^{R}$) is transformed with plasmid pUC7-K ($Ap^{R}Km^{R}$), recombination of the homologous regions of the *bla* gene and the *ori* sequence, will result in R2-PIM8-K. In this strain the *neo* gene of pUC7-K will become integrated into the genome, concomitant with restoration of the *bla* gene and deletion of the *aad* gene from the chromosome. **(B)** When strain R2-PIM9 ($Ap^{S}Km^{R}$) is transformed with plasmid pUC7-S ($Ap^{R}Sm^{R}$), recombination of the homologous regions of the *bla* gene and the *ori* sequence, will result in R2-PIM9-K. In such a strain the *aad* gene of pUC7-S will become integrated into the genome, concomittant with restoration of the *bla* gene and deletion of the *neo* gene from the chromosome. Arrows indicate the direction of transcription. The position of transcription termination signals is marked by an ♀ and the absence of the *bla* promoter by ΔP. The phenotype of the cells is indicated at the right.

(e) Cloning into the chromosomal platforms

The utility of platform strain R2-PIM8 (Sm^R) was tested with plasmid pUC7-K (Ap^R Km^R), that of strain R2-PIM9 (Km^R) with plasmid pUC7-S (Ap^R Sm^R) (Table I). Homologous recombination between the donor plasmid molecules and the integration platforms both in the *bla* and the *ori* sequences will result in restoration of a functional *bla* gene (Ap^R) and in concomitant replacement of the *aad* gene of the PIM8 platform by the *neo* gene from pUC7-K (Fig. 4A), or of the *neo* gene of the PIM9 platform by the *aad* gene from pUC7-S (Fig. 4B).

Synechococcus R2-PIM8 and R2-PIM9 were transformed with pUC7-K and pUC7-S, respectively, and the recombinants were selected on agar plates containing various concentrations of Ap (see Materials and Methods section c). The Ap^R transformant colonies were subsequently screened for the loss of their platform antibiotic resistance, i.e. for a Sm^S or Km^S phenotype, by streaking on methionine and Ap containing plates, with and without Sm or Km (Table IIIA). Finally the Ap^R Sm^S and Ap^R Km^S clones were investigated for the presence of the plasmid marker gene by applying the relevant antibiotic pressure:

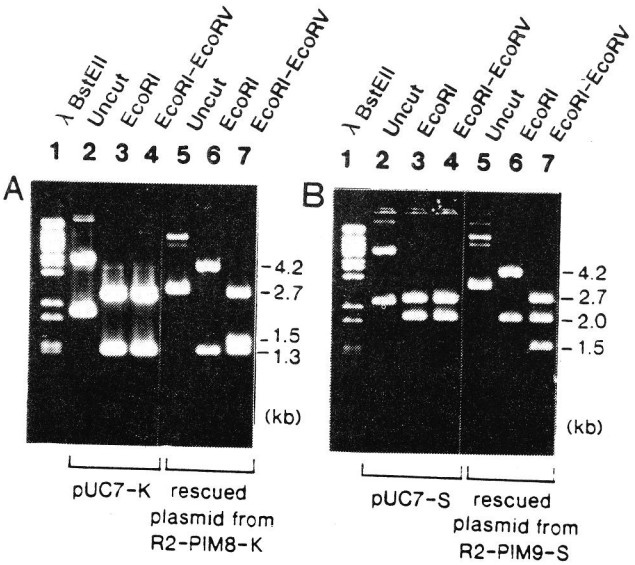
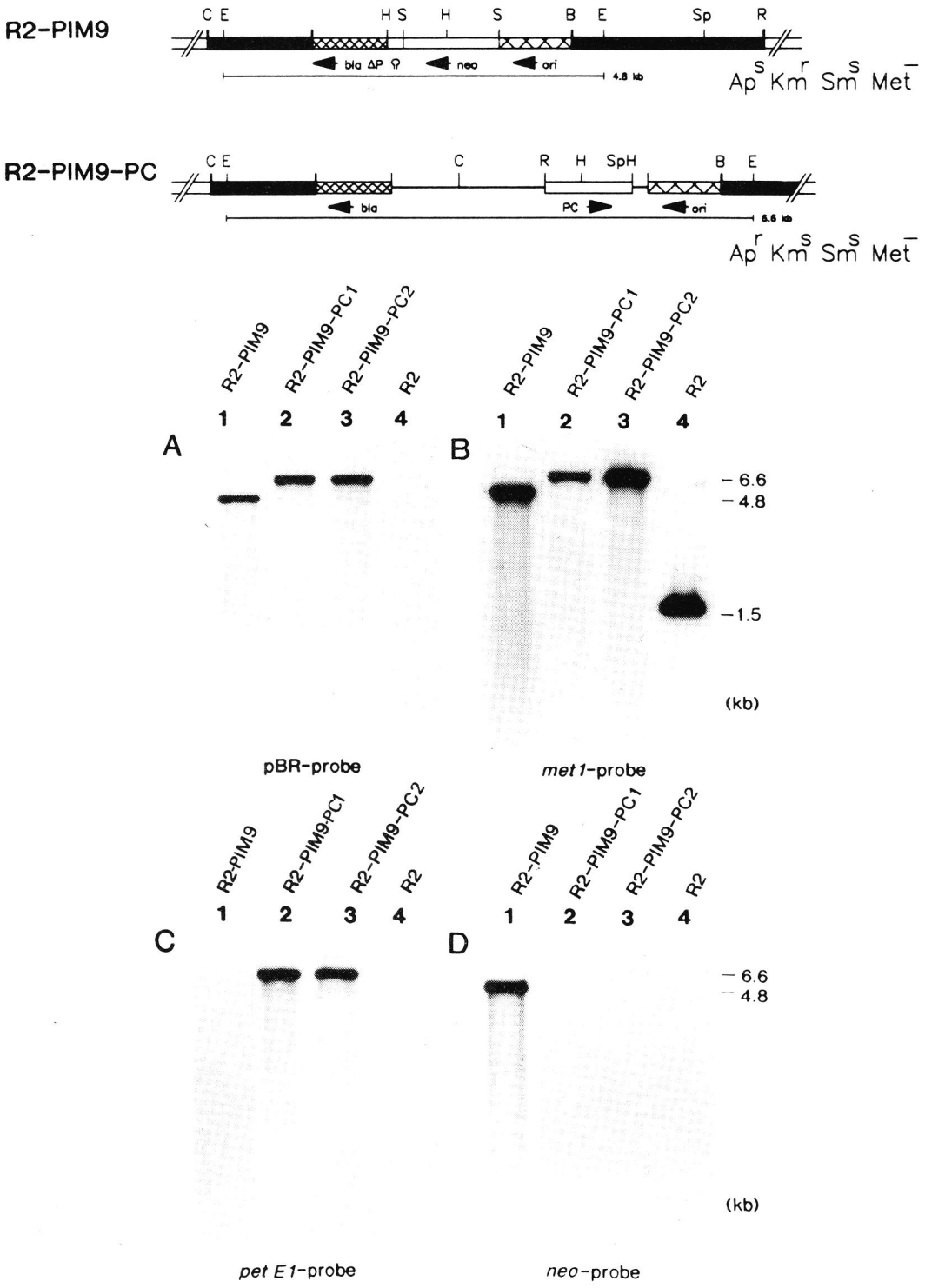


Fig. 5. Restriction analysis of rescued plasmids containing the recombined structures from R2-PIM8-K and R2-PIM9-S. (**A**) Undigested DNA (lane 5), *Eco*RI (lane 6) and *Eco*RI-*Eco*RV (lane 7) digests of a plasmid rescued from R2-PIM8-K (see Fig. 4) were compared with DNA of donor plasmid pUC7-K treated in the same way (lane 2, 3 and 4, resp.) on an ethidium-bromide stained agarose gel. (**B**) Undigested DNA (lane 5), *Eco*RI (lane 6) and *Eco*RI-*Eco*RV (lane 7) digests of a plasmid rescued from R2-PIM9-S (see Fig. 4) were compared with DNA of donor plasmid pUC7-S (lane 2, 3 and 4, resp.) on an ethidium-bromide stained agarose gel. The fragments of *Bst*EII digested lambda DNA were used as molecular weight markers (Panels A and B: lane 1).


R2-PIM9
C E H S H S B E Sp R
bla ΔP ♀ neo ori
4.8 kb
Ap^S Km^r Sm^S Met^-
R2-PIM9-PC
C E C R H SpH B E
bla PC ori
6.6 kb
Ap^r Km^S Sm^S Met^-
R2-PIM9
R2-PIM9-PC1
R2-PIM9-PC2
R2
1 2 3 4
A
B
6.6
4.8
1.5
(kb)
pBR-probe
met 1-probe
C
D
pet E1-probe
neo-probe

Km and Sm, respectively. The Ap^R colonies showing loss of the original platform-sited antibiotic resistance had in 100% of the cases acquired the new antibiotic resistance from the donor plasmids (Table IIIA). These data prove that the integration system functions efficiently.

(f) Plasmid rescue

The structure of the integration platforms allows the reisolation of the sequences cloned into the platform; this technique is called "plasmid rescue" (Perucho *et al.*, 1980). The transformed cells have a functional *bla* gene and a complete pBR origin of replication flanking the cloned DNA. Plasmid rescue begins with digestion of the chromosomal DNA with a restriction enzyme that does not cut in the platform-integrated structure, such as *Eco*RV, and circularization of the resulting fragments with ligase. Subsequent transformation of *E. coli* and selection for Ap^R, should result in the isolation of the original plasmid with some additional *met* DNA. Two pieces of the *met*1 gene, namely the 1.1 kb *Eco*RV-*Bam*HI and the 0.4 kb *Bam*HI-*Eco*RV fragment, will become connected by their *Eco*RV ends and then be located between the *bla* gene and *ori*.

Plasmid rescue was performed in strain R2-PIM8-K, a Km^R derivative of R2-PIM8 transformed with plasmid pUC7-K (Fig. 4A), and in strain R2-PIM9-S, a Sm^R derivative of R2-PIM9 transformed with plasmid pUC7-S (Fig. 4B). Chromosomal DNAs were digested to completion with *Eco*RV. The restriction fragments were ligated at low DNA concentrations ($\leq$ 0.01 µg/µl) to promote circularization over coupling of fragments and subsequently transformed to *E. coli* PC2495. This resulted in 2-5 x 10^3 Ap^R cfu/µg DNA. From each transformation 10 colonies were tested for the presence of the platform antibiotic marker and the antibiotic marker of the pUC7 derived initial donor plasmid. All contained next to the Ap resistance exclusively the donor resistance marker. Restriction analysis with *Eco*RI and *Eco*RI-*Eco*RV (Fig. 5) showed that the expected *Eco*RI fragments (see Fig. 4) containing the antibiotic resistance genes were present (1.3 kb for the *neo*-fragment and 2.0 kb for the *aad*-fragment) and that the rescued "pBR" fragment has

Fig. 6. Integration of the *Anabaena* sp. PCC 7937 plastocyanin gene (*petE1*) in R2-PIM9. **(Top)** Schematic representation of the platform region of R2-PIM9 before and after transfer of the *Anabaena* plastocyanin gene from the pPCVC1 donor plasmid to the genomic integration platform. **(Bottom)** Southern analysis of the integrated *Anabaena* plastocyanin gene in R2-PIM9. Four Southern blots were made containing *Eco*RV digests of chromosomal DNA from the acceptor strain R2-PIM9 (lane 1), from two Ap^RKm^S transformants of R2-PIM9 with pPCVC1 called R2-PIM9-PC1 and R2-PIM9-PC2 (lane 2 and 3, resp.) and from the wildtype strain R2 SPc (lane 4). These blots were hybridized with one of four different ^{32}P-labelled probes: total pUC18 DNA (pBR-probe; panel A), the 1.5kb *Eco*RV *met1* fragment containing the *Bam*HI site (*met1*-probe; panel B), the 360 bp *Hin*dIII-*Hin*PI fragment from the plastocyanin gene (*petE1*-probe; panel C) and the 1.3kb *Pst*I *neo* fragment from pUC4-K (*neo*-probe; panel D).

the additional 1.5 kb DNA from the *met1* gene (4.2 kb instead of 2.7 kb). As expected this *met1* fragment contained the *Eco*RV site of circularisation, whereas *Eco*RV sites are absent in the "pBR" sequences. The inserted sequence thus can be rescued.

(g) Integration of the Anabaena *plastocyanin gene*

The platform integration system was used to bring into *Synechococcus* sp. PCC 7942 a gene encoding plastocyanin (*petE1*) from *Anabaena* sp. PCC 7937 (Van der Plas *et al.*, 1989) (Fig. 6, Top of the figure). *Synechococcus* sp. PCC 7942 itself does not contain a gene for plastocyanin, as in this organism cytochrome c_{553} instead of plastocyanin functions as electron carrier between the cytochrome b_6/f complex and photosystem I. This makes *Synechococcus* sp. PCC 7942 an interesting host to analyse various aspects of the biogenesis and functioning of *Anabaena* sp. PCC 7937 plastocyanin.

The plasmid pPCVC1 (Table I) consisting of the 4 kb vector pEMBL19 (Dente and Cortese, 1983) with the *petE1* gene of *Anabaena* sp. PCC 7937 on a 1.1 kb insert, was transformed to strain R2-PIM9. Of the 73 resulting Ap^R transformants 53 were found to be Km^S (73%; Table IIIB). After colony purification, chromosomal DNA was extracted from two of these Ap^R Km^S transformants, digested with restriction endonuclease *Eco*RV and used for Southern hybridization analysis, with wildtype DNA and DNA from R2-PIM9 as controls.

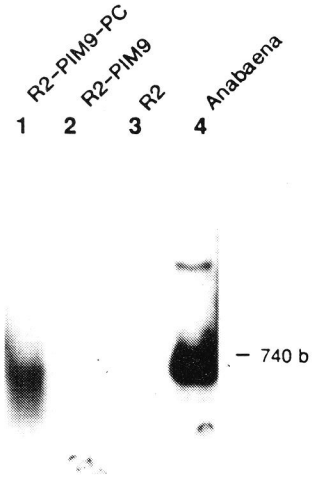


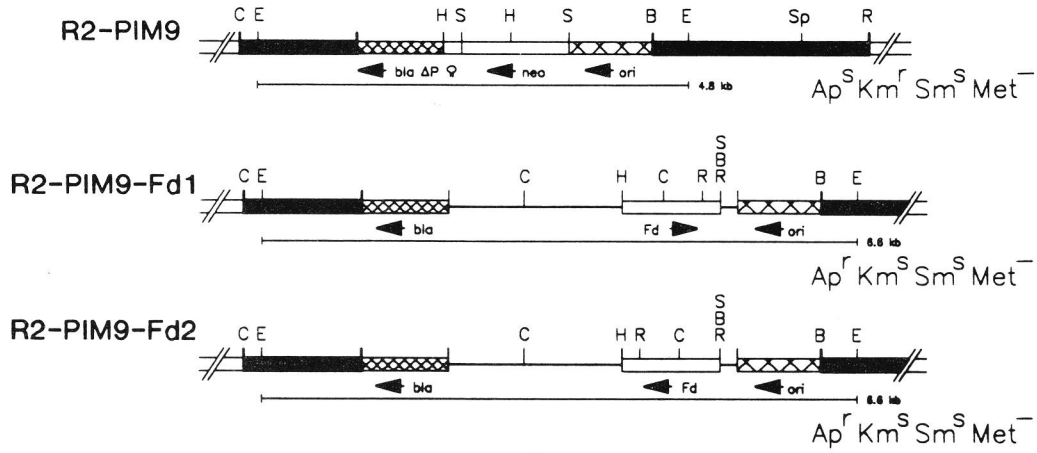
Fig. 7. Transcription analysis of the *Anabaena* plastocyanin gene in R2-PIM9-PC. A Northern blot with total RNA (approximately 40 µg) extracted from R2-PIM9-PC (lane 1), R2-PIM9 (lane 2), wildtype strain R2 (lane 3) and *Anabaena* sp. PCC 7937 (lane 4) was probed with the 0.36 kb *Hin*dIII-*Hin*PI *petE1* fragment. All RNAs were extracted from cells grown in the presence of Cu^{2+}. Denatured fragments from pEMBL8 DNA digested with *Taq*I were used as molecular size markers.

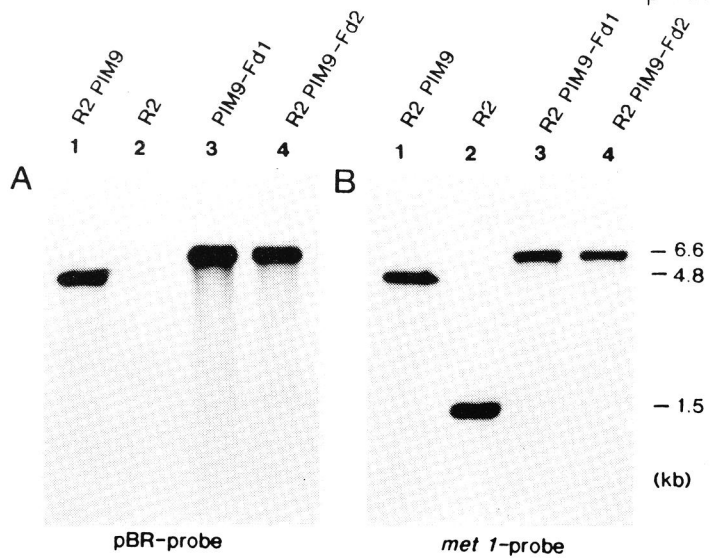
*Eco*RV cleaves only the flanking *met1* sequences and not the DNA interrupting the *met1* gene. Hybridization with an internal fragment of the plastocyanin gene (Van der Plas *et al.*, 1989), resulted in the detection of a 6.6 kb fragment in the two DNAs from the transformed Ap^R Km^S strains and not from the wildtype and the platform strain (Fig. 6C). The size of 6.6 kb for the fragment hybridizing to the *petE* probe matches the size calculated from the *Eco*RV map of the methionine DNA, the size of the platform and that of the *petE1* fragment (including flanking pEMBL sequences) in the clone used for transformation, indicating integration took place in the proper way (Fig. 6, Top of the figure). The results of hybridization with the pBR, *met1* and *neo* probe (Fig. 6A, B and D respectively) confirm this conclusion.

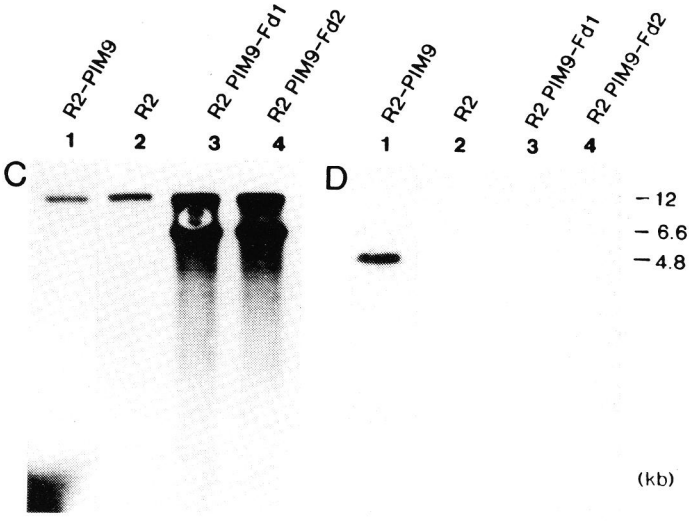
In a Northern blot made with RNA extracted from the transformed strain, a *petE1* specific transcript was detected of similar size as in *Anabaena* (Van der Plas *et al.*, 1989). This transcript of approximately 740 b is absent in RNA extracted from the corresponding wildtype and platform strains (Fig. 7). From these results it is concluded that the *petE1* of *Anabaena* sp. PCC 7937 has been integrated successfully and in the expected way into the genome of *Synechococcus* sp. PCC 7942 by means of the platform system and that in the transformant analysed the *petE1* gene is specifically transcribed into mRNA.

(h) Integration of an additional copy of the ferredoxin I gene

The system was used for the transfer of sequences for which an endogenous genomic copy is already present, in other words for the construction of partial diploids. An additional copy of the gene encoding ferredoxin I (*petF1*) (Van der Plas *et al.*, 1986; Van der Plas *et al.*, 1988) was transferred to *Synechococcus* sp. PCC 7942 (Fig. 8, Top of the figure). This *petF1* copy was present as a 1.1 kb *Pst*I fragment in pRA81 and pRA82 (Table I), representing both orientations cloned in pEMBL8. After transformation of these plasmids to strain R2-PIM9 about 40% of the Ap^R transformants also were Km^S (Table IIIB). After colony purification, the DNA was extracted from two of these $Ap^R Km^S$ transformants, digested with restriction endonuclease *Eco*RV and used for Southern hybridization analysis. Hybridization with a ferredoxin probe (Van der Plas *et al.*, 1988), showed two *petF1* genes in the transformed $Ap^R Km^S$ strains, the resident copy present on a 12 kb fragment as in the wildtype and the R2-PIM9 strain, and an additional copy present on a 6.6 kb fragment (Fig. 8, Panel C). The size of 6.6 kb matches the size calculated from the *Eco*RV map of the methionine DNA, the size of the platform and that of the *petF1* fragment in the clone used for transformation, indicating integration took place in the proper way (Fig. 8). The results of hybridization with the pBR, *met1* and *neo* probe (Fig. 8A, B and D repectively) confirm this conclusion. So, even in the presence of a homologous


R2-PIM9
C E
H S
H
S
B E
Sp
R
bla ΔP
neo
ori
4.8 kb
Ap^S Km^r Sm^S Met^-
R2-PIM9-Fd1
C E
C
H C R
S B R
B E
bla
Fd
ori
6.6 kb
Ap^r Km^S Sm^S Met^-
R2-PIM9-Fd2
C E
C
H R C
S B R
B E
bla
Fd
ori
6.6 kb
Ap^r Km^S Sm^S Met^-


R2 PIM9
R2
PIM9-Fd1
R2 PIM9-Fd2
1
2
3
4
A
B
6.6
4.8
1.5
(kb)
pBR-probe
met 1-probe


R2-PIM9
R2
R2 PIM9-Fd1
R2 PIM9-Fd2
1
2
3
4
C
D
12
6.6
4.8
(kb)
pet F1-probe
neo-probe

resident gene the desired integration of an additional copy at the site of the platform can easily be obtained.

(i) Conclusions

The system described for the integration of DNA fragments into the genome of *Synechococcus* sp. PCC 7942 was shown to function in an efficient and reliable manner as is evident from the results presented for the transfer of the *neo*, *aad*, *petE1* and *petF1* genes. A major advantage of this integration platform system is that by nature it is independent of the DNA insert to be transferred, unless the expression of the inserted DNA in *Synechococcus* sp. PCC 7942 has some deleterious effect on the cell. The integration event as well as the selection of the desired recombinants depends in principle on vector and target sequences only.

With the integration system, complementation and dominance studies of cloned genes can be undertaken without the complication of copy number differences of the complementing alleles that could arise when the cloned gene is present on a cyanobacterial (shuttle-)vector. Besides, the latter genetic make-up has shown to be highly unstable, at least during or shortly after transformation (Kuhlemeier *et al.*, 1985), since homogenotization between the two copies of genomic DNA occurred at high frequency. The isolation of a recombination deficient mutant proved unsuccessful yet. On the other hand, an indication for the stability of different alleles when both are present in the chromosome can be inferred from the existence of naturally occurring multigene families in the genome of *Synechococcus*, e.g. the three genes (*psbA*) coding for the Q_b protein of photosystem II (Golden *et al.*, 1986) and the two clusters encoding the phycocyanin subunit genes (*cpcA* and *cpcB*) (Kalla *et al.*, 1988). The actual stability of any intentionally integrated second allele of a particular gene into the genome of *Synechococcus* sp. PCC 7942, e.g. into the integration platform, still remains to be examined. In the context, however, of the experiments described in this paper no sign of instability of the integrated fragments was observed. The possibility to recover the

Fig. 8. Integration of an additional copy of the ferredoxin I gene (*petF1*) in R2-PIM9. **(Top)** Schematic representation of the platform region of R2-PIM9 before and after transfer of the ferredoxin gene in both orientations to the genomic integration platform. **(Bottom)** Southern analysis of the ferredoxin genes in R2-PIM9-Fd1 and R2-PIM9-Fd2. Four similar Southern blots were made containing *Eco*RV digests of chromosomal DNA from the acceptor strain R2-PIM9 (lane 1), from the wildtype strain *Synechococcus* sp. PCC7942 (lane 2) and from R2-PIM9-Fd1 and R2-PIM9-Fd2 (lane 3 and 4), two Ap^RKm^S transformants of R2-PIM9 with ferredoxin subclones in opposite orientations. These blots were hybridized with one of four different ^{32}P-labelled probes: total pUC18 DNA (pBR; panel A), the 1.5kb *Eco*RV *met1* fragment containing the *Bam*HI site (*met1*; panel B), the 370 bp *Sst*I-*Eco*RI fragment from the ferredoxin I gene (*petF1*; panel C) and the 1.3kb *Pst*I *neo* fragment from pUC4-K (*neo*; panel D).

platform region as an autonomously replicating plasmid, by excision from the chromosome, circularization and transformation to *E. coli*, should facilitate the check on the integrity of the DNA integrated.

Available special-purpose derivatives of pBR322, tailored for example for gene expression studies or promoter analysis in *E. coli*, can now be applied for *Synechococcus* sp. PCC 7942 without further modification. In this way the platform system in *Synechococcus* sp. PCC 7942 can be used for the analysis of gene regulation by the integration of promoter-reporter fusions and also for the expression of higher plant genes like *petE1*, *petF1* and others in order to study the synthesis and processing of their products in the heterologous host.

Acknowledgements

The authors thank Prof G.A. Van Arkel for his advice in the preparation of this manuscript.

References

Balbás P, Soberón X, Merino E, Zurita M, Lomeli H, Valle F, Flores N and Bolivar F (1986) Plasmid vector pBR322 and its special-purpose derivatives -- a review. Gene 50:3-40

Dente L and Cortese R (1987) pEMBL: a new family of single stranded plasmids for sequencing DNA. Methods Enzymol 152:111-119

Duyvesteyn MGC, Korsuize J, De Waard A, Vonshak A and Wolk CP (1983) Sequence specific endonucleases in strains of *Anabaena* and *Nostoc*. Arch Microbiol 134:276-281

Golden SS, Brusslan J and Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. EMBO J 5:2789-2798

Golden SS, Brusslan J and Haselkorn R (1987) Genetic engineering of the cyanobacterial chromosome. Methods Enzymol 153:215-231

Kalla RS, Lind LK, Lidholm J and Gustafsson P (1988) Transcriptional organization of the phycocyanin subunit gene clusters of the cyanobacterium *Anacystis nidulans* UTEX 625. J Bacteriol 170:2961-2970

Kolowsky KS, Williams JGK and Szalay AA (1984) Length of foreign DNA in chimeric plasmids determines the efficiency of its integration into the chromosome of the cyanobacterium *Synechococcus* R2. Gene 27:289-299.

Kolowsky KS and Szalay AA (1986) Double-stranded gap repair in the photosynthetic prokaryote *Synechococcus* R2. Proc Natl Acad Sci USA 83:5578-5582

Kuhlemeier CJ, Thomas AAM, Van der Ende, A., Van Leen RW, Borrias WE, Van den Hondel, CAMJJ and Van Arkel GA (1983) A host-vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. Plasmid 10:156-163

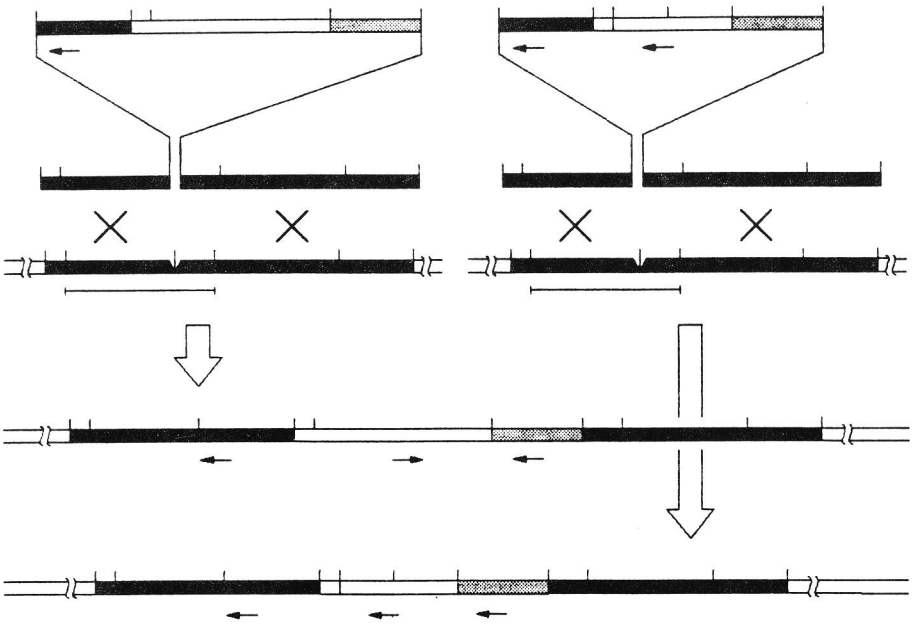
Kuhlemeier CJ, Hardon EM, Van Arkel GA and Van de Vate C (1985) Self-cloning in the cyanobacterium *Anacystis nidulans* R2: fate of a cloned gene after reintroduction. Plasmid 14:200-208

Kuhlemeier CJ and Van Arkel GA (1987) Host-vector systems for gene cloning in cyanobacteria. Methods Enzymol 153:199-215

Maniatis T, Fritsch EF and Sambrook J (1982) Molecular Cloning. A laboratory manual. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory
Perucho M, Hanahan D, Lipsich L and Wigler M (1980) Isolation of the chicken thymidine kinase gene by plasmid rescue. Nature (London) 285:207–210
Prentki P and Krisch HM (1984) In vitro mutagenesis with a selectable DNA fragment. Gene 29:303–313
Porter RD (1985) Transformation in cyanobacteria. Crit Rev Microbiol 13: 111–132
Tandeau de Marsac N, Borrias WE, Kuhlemeier CJ, Castets AM, Van Arkel GA and Van den Hondel CAMJJ (1982) A new approach for molecular cloning in Cyanobacteria: cloning of an *Anacystis nidulans met* gene using a Tn*901*–induced mutant. Gene 20:111–119
Tandeau de Marsac N and Houmard J (1987) Advances in cyanobacterial molecular genetics. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 251–302. Amsterdam: Elsevier
Thiel T and Wolk CP (1987) Conjugal transfer of plasmids to cyanobacteria. Methods Enzymol 153:232–243
Van den Hondel CAMJJ, Verbeek S, Van der Ende A, Weisbeek PJ, Borrias WE and Van Arkel GA (1980) Introduction of transposon Tn*901* into a plasmid of *Anacystis nidulans*: preparation for cloning in cyanobacteria. Proc Natl Acad Sci USA 77:1570–1574
Van der Plas J, De Groot RP, Woortman MR, Weisbeek PJ and Van Arkel GA (1986) Coding sequence of a ferredoxin gene from *Anacystis nidulans* R2 (*Synechococcus* PCC 7942). Nucl Acids Res 14:7804
Van der Plas J, De Groot R, Woortman M, Cremers F, Borrias M, Van Arkel G and Weisbeek P (1988) Genes encoding ferredoxins from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942: structure and regulation. Photosynth Res 18:179–204
Van der Plas J, Bovy A, Kruyt F, De Vrieze G, Dassen E, Klein B and Weisbeek P (1989) The gene for the precursor of plastocyanin from the cyanobacterium *Anabaena* sp. PCC 7937: isolation, sequence and regulation. Mol Microbiol 3:275–284
Vieira J and Messing J (1982) The pUC plasmids, an M13mp7–derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259–268
Williams JGK and Szalay AA (1983) Stable integration of foreign DNA into the chromosome of the cyanobacterium *Synechococcus* R2. Gene 24:37–51
Wolk CP, Vonshak A, Kehoe P and Elhai J (1984) Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen–fixing filamentous cyanobacteria. Proc Natl Acad Sci USA 81:1561–1565
Yanish–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

Chapter V

The small plasmid pUH24 from the cyanobacterium *Synechococcus* sp. PCC 7942: nucleotide sequence analysis, identification of replication and stability functions



J. van der Plas, R. Oosterhoff-Teertstra,
M. Borrias and P. Weisbeek
submitted

The small plasmid pUH24 from the cyanobacterium *Synechococcus* sp. PCC 7942: nucleotide sequence analysis, identification of replication and stability functions

Summary

The complete nucleotide sequence is presented for pUH24, the small plasmid of *Synechococcus* sp. PCC 7942. pUH24 consists of 7835 bp and has a G+C content of 59%. Analysis of the distribution of translation start and stop codons in the sequence shows the existence of 36 open reading frames potentially capable of encoding polypeptides of 50 or more amino acids. Based on several methods for searching coding regions we postulate that 8 of these open reading frames are actual coding sequences. The correlation between the outcome of the sequence analysis and properties of (shuttle-)vectors derived from pUH24 was studied. In this way a non-essential region has been identified that contains two functions, designated *pmaA* and *pmaB*, involved in the segregational stability of the plasmid. The minimal region of pUH24 fully capable of supporting autonomous replication consists of a 3.6 kb DNA fragment, which is almost entirely occupied by two overlapping genes most likely coding for essential replication proteins (*repA* and *repB*). The pUH24 origin of replication is presumably located in a 200 bp intergenic region on the minimal replicon fragment, as this sequence possesses several features commonly found in prokaryotic plasmid origins.

Introduction

One or several plasmids are found in about 60% of the cyanobacterial strains examined (for a review see: Tandeau de Marsac and Houmard, 1987). They vary in size from 1.3 kb to 130 kb. Although they were eventually thought to be encoding such varying functions as motility (Castets *et al.*, 1986), heterocyst formation (Reaston *et al.*, 1980), the production of gas vesicles (Walsby, 1977), or toxins (Hauman, 1981), none of these relationships could be established; so they all remain cryptic.

Several cyanobacterial host-vector systems based on endogenous cryptic plasmids were developed, mostly in the form of shuttle vectors as *Escherichia coli* replicons are not functional in the cyanobacterium and vice versa (Porter, 1985; Kuhlemeier and Van Arkel,

1987; Tandeau de Marsac and Houmard, 1987; Shestakov and Reaston, 1987). In these studies, most of the attention was directed to the physical properties of the constructed vectors: the ligation of the cyanobacterial replication functions to an *E. coli* replicon, the incorporation of suitable antibiotic resistance markers, enhancement of the number of unique restriction sites and reduction in size of the resulting hybrids. Except for the selection of a minimal fragment containing the functions essential for cyanobacterial replication (Laudenbach *et al.*, 1983, 1985; Schmetterer and Wolk, 1988), little or no attention has been paid to plasmid replication and (in)stability in cyanobacteria.

Plasmid instability can be of two types: segregational and structural (Primrose and Ehrlich, 1981). Segregational instability refers to the loss of the entire plasmid population from the cell. In addition to an efficient replication control mechanism, stable maintenance of a plasmid requires proper partitioning of plasmid molecules at cell division (Meacock and Cohen, 1980; Summers and Sherrat, 1985). Little is known about mechanisms of cyanobacterial plasmid replication, molecular control of copy numbers and segregational plasmid instability. Structural plasmid instability corresponds to rearrangements, most often deletions, of plasmid DNA sequences (for a review, see Ehrlich 1989). So far only two studies on structural plasmid instability in cyanobacteria have been carried out; the first involving recombination events during or shortly after introduction of plasmid constructs into *Synechococcus* sp. PCC 7942 (Kuhlemeier *et al.*, 1985) and the second involving Tn*5* insertions in cloning vectors for the same strain (Gendel, 1987).

Nucleotide sequence analysis of cyanobacterial plasmids is expected to give insight into their functional organization, improve and facilitate their use as cloning vectors and enlarge our understanding of plasmid replication and maintenance in cyanobacteria. The highly transformable strain *Synechococcus* sp. PCC 7942 (formerly called *Anacystis nidulans* R2) harbours two cryptic plasmids (Van den Hondel *et al.*, 1980), pUH24 (8 kb) and pUH25 (50 kb) of unknown copy number. In this paper we report the 7835 bp sequence of pUH24, the plasmid that is used in almost all of the many cloning vectors that were constructed for *Synechococcus* sp. PCC 7942 (Van den Hondel *et al.*, 1980; Kuhlemeier *et al.*, 1981, 1983; Sherman and van der Putte 1981; Tandeau de Marsac *et al.*, 1982; Gendel *et al.*, 1983a, 1983b; Golden and Sherman, 1983; Lau and Straus 1985; Friedberg and Seijffers 1986; Gallagher and Burke, 1987; Gruber *et al.*, 1987). At least eight major open reading frames were found, the largest of which occupies most of the essential 3.6 kb region shared by all functional cloning vectors for *Synechococcus* sp. PCC 7942. Replication and stability promoting functions were localized by segregation analysis and by studying the correlation between the physical maps and properties of pUH24 derived (shuttle-) vectors.

Table I. Strains and Plasmids

Strain/plasmid	Relevant properties	Reference/source
Escherichia.coli K12-803	*lac, gal, met, recA, hsdS, supE, supF*	Kourilsky P (Inst. Pasteur).
Escherichia coli GM113	*thr, leu, proA, his, thi, metB, thyA, deoB, dam3, lacY, galK, ara, mtl, tsx, phx, rpsL, supE*	Marinus MG
Escherichia coli JM101	*supE, thi,* Δ(*lac-proAB*), [F', *traD36, proAB, lacI*^{q} *Z* M15]	Yanish-Perron *et al.* (1985)
Escherichia coli PC2495	*recA, hsdS,* derivative of JM101	Van der Plas *et al.* (1988)
Synechococcus sp. PCC 7942	Wild type strain, harbouring plasmids pUH24 and pUH25	Van den Hondel *et al.* (1980)
Synechococcus. sp. strain R2-SPc	Strain PCC 7942 cured of pUH24	Kuhlemeier *et al.* (1983)
pUH24[a]	7.8 kb, small plasmid from *Synechococcus* sp. PCC 7942	Van den Hondel *et al.* (1980)
pCH1	12.4 kb, Ap^{R}, pUH24::Tn*901*	Van den Hondel *et al.* (1980)
pUC1	8.2 kb, Ap^{R}, deletion derivative of pCH1	Van den Hondel *et al.* (1980)
pUC12	9.8 kb, Ap^{R}, Sm^{R}, pUC1 + *aad* gene fragment from pRI1477S (=RSF1010)	Kuhlemeier *et al.* (1983)
pUC13	6.9 kb, Sm^{R}, deletion derivative of pUC12	Kuhlemeier *et al.* (1983)
pUC14	7.8 kb, Sm^{R} deletion derivative of pUC12	Kuhlemeier *et al.* (1983)
pBX3[b]	3.2 kb, Ap^{R}, deletion derivative of pBR322 with øX174 *ori*	Van der Ende *et al.* (1983)
pBX24[c]	11.0 kb, Ap^{R}, hybrid plasmid of pUH24 and pBX3	This study

[a] The source of pUH24 DNA for subcloning and sequencing was the chimeric plasmid pBX24 grown in *E. coli* K12-803.

[b] The 3.2 kb plasmid pBX3 is a deletion derivative of pBR322 called pBR322-4 with the øX174 (+) strand origin containing *Hin*dII fragment inserted into one of its *Hin*dII sites.

[c] pBX24 was made by inserting the 7.8 kb *Bam*HI linearized pUH24 into the unique *Bam*HI site of the 3.2 kb pBR322 derived vector pBX3. For restriction analysis the digestion patterns of pBX24 DNA extracted from *Synechococcus* sp. PCC 7942, *E. coli* K12-803 and *E. coli* GM113 (*dam*$^{-}$) were compared.

Materials and methods

(a) Bacterial strains, plasmids and growth conditions

The strains and plasmids used in this study are listed in Table I. The cyanobacterial strains were grown in the light in BG11 medium as described (Van den Hondel *et al.*, 1980).

(b) Plasmid segregation tests

The segregation of plasmids into daughter cells was analysed as follows. *Synechococcus* PCC 7942 cells harbouring pCH1 (Ap^R), pUC1 (Ap^R), pUC12 (Ap^R,Sm^R), pUC13 (Sm^R) or pUC14 (Sm^R) in stead of pUH24, were grown to the exponential growth phase under conditions selective for their respective plasmids (Kuhlemeier *et al.*, 1983), inoculated (1:100) into fresh BG11 medium without antibiotics and grown for 5 days (approximately 7 generations). These cultures were again diluted (1:100) and grown for 3–5 days and so on, until approximately 50 generations of unselected growth were reached. The cells were then diluted an plated on BG11 agar without antibiotics. At least 50 single colonies for each type of culture were subsequently tested for antibiotic resistance, by transferring them to plates with the selective antibiotics (Kuhlemeier *et al.*, 1983).

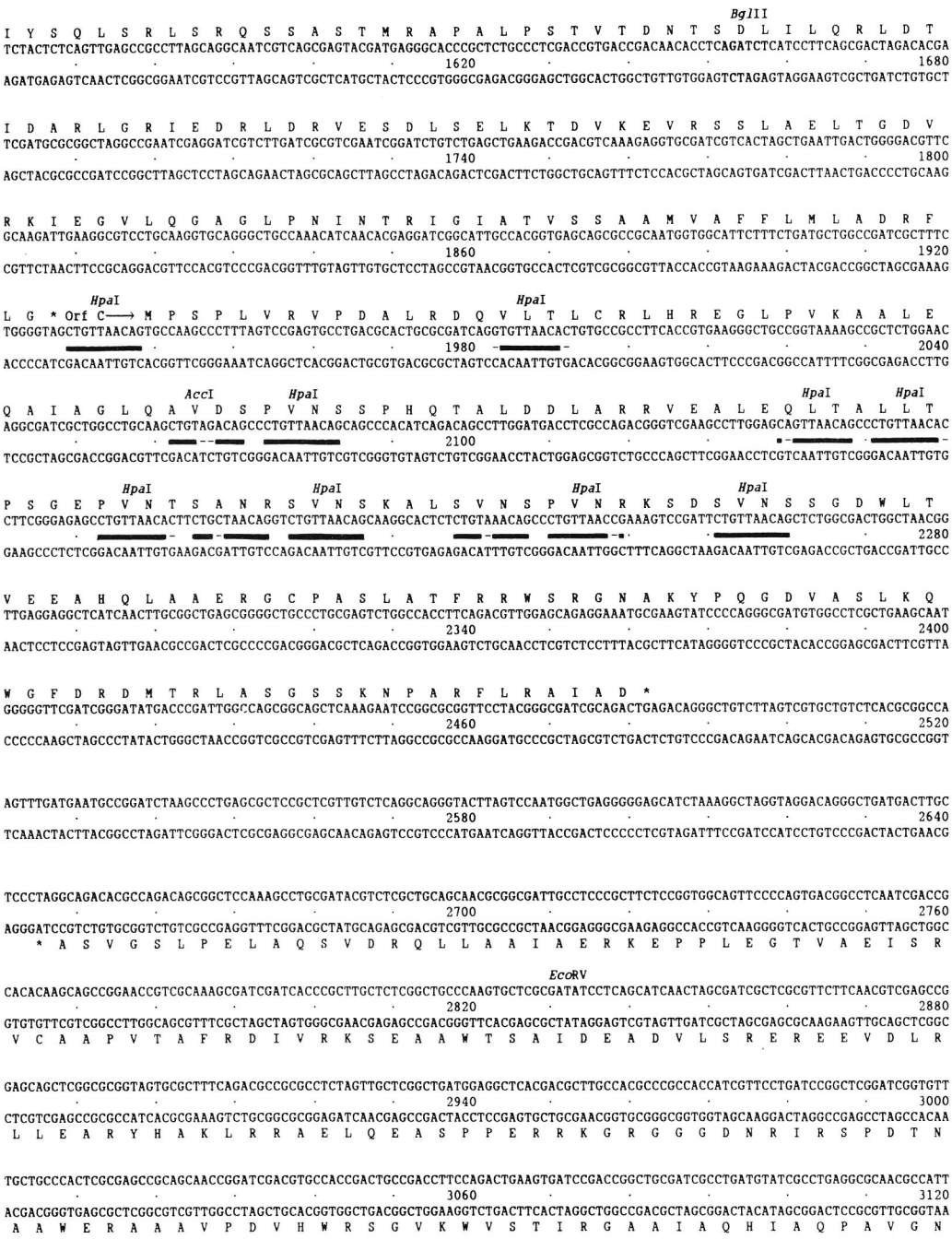
(c) DNA and RNA manipulations

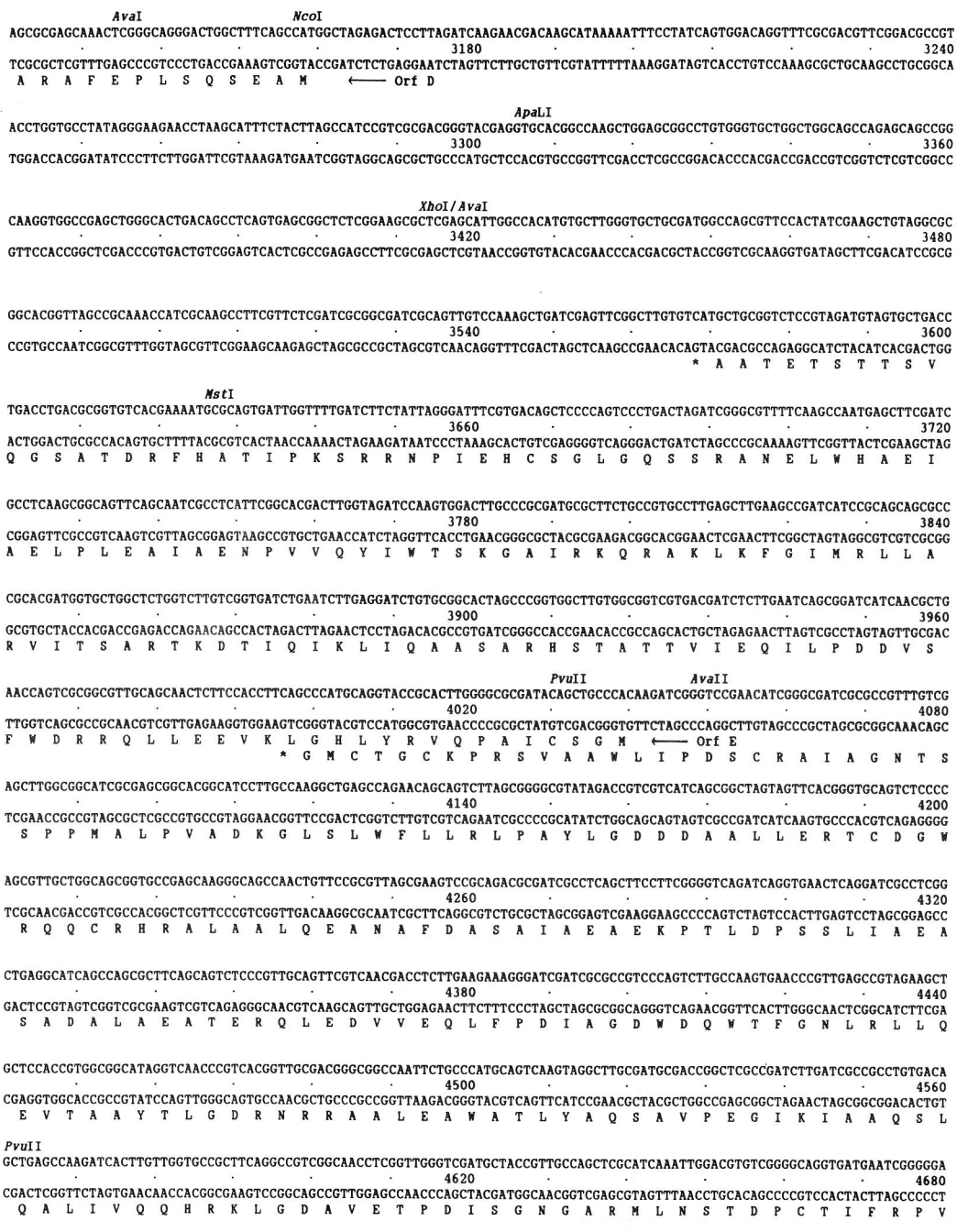
Basic techniques (e.g. plasmid and chromosomal DNA isolation, restriction endonuclease digestion, gel electrophoresis, restriction fragment purification, ligation, transformation, oligo-primed labelling, Southern hybridisation, RNA extraction and Northern analysis) were performed by standard methods (Maniatis *et al.*, 1985) or modified as described previously (Van der Plas *et al.*, 1988, 1989). Large scale preparations of plasmid DNA from *Synechococcus* sp. PCC 7942 were made according to Van den Hondel *et al.* (1980).

Fig. 1. Complete nucleotide sequence of pUH24 DNA. The sequence is numbered starting at the first base of the unique *Bam*HI site. The amino acid sequences of the predicted proteins encoded by the plasmid are shown in standard single letter code above (translation left to right) and below (translation right to left) the sequence. Asterisks indicate stopcodons. A family of 12 palindromic sequences (nt position 1900–2300; consensus: -CTGTTAACAG-) is indicated by solid lines (▬). The major direct repeats (⟵⟶) and inverted repeats (⟶ ⟵) in the presumed origin region (nt position 6625–6802) are marked. Two putative promoters as discussed in the text, located approximately at nt position 1286 and 6729, are underlined.

*Bam*HI

```
GGATCCGCGAGCGTGCTTGCTGACGAGTTGCACCGCTGCGGATCAGCTCATCCGCGAGGGACTGATCGAGGCGCGCCCGCGAGACGGCCTTCACAATGTCAGCCGCCCAGCGCTGCTCGG
                                                  60                                                  120
CCTAGGCGCTCGCACGAACGACTGCTCAACGTGGCGACGCCTAGTCGAGTAGGCGCTCCCTGACTAGCTCCGCGCGGGCGCTCTGCCGGAAGTGTTACAGTCGGCGGGTCGCGACGAGCC
  I  R  S  R  A  Q  Q  R  T  A  G  S  R  I  L  E  D  A  L  S  Q  D  L  R  A  R  S  V  A  K  V  I  D  A  A  W  R  Q  E  A

CTGCAGAGCGGGTCTGCGCGTTAGGGTCTGCAGGAATCGCCCACGGCGCTCACCTCCAAAAGCCGCCAGCGGTTCACAACAATCGAGTCGCGGAGATCGTCGTAGTCAGTCGAGACAAAGC
                                                 180                                                  240
GACGTCTCGCCCAGACGCGCAATCCCAGACGTCCTTAGCGGTGCCGCGAGTGGAGGTTTTCGGCGGTCGCCAAGTGTTGTTAGCTCAGCGCCTCTAGCAGCATCAGTCAGCTCTGTTTCG
  A  S  R  T  Q  A  N  P  D  A  P  I  A  M      ⟵ Orf H

GAACGCTCCAAGAGATCGCACCGGAACTGATCAGCTCCCACAGGCGATCGCGATCGGGATGTTCGAGCAAGGTGATCGTCGCTAGCAGTCCAGTTTCCCCGTTCTCCAGTCGCTCAGCTC
                                                 300                                                  360
CTTGCGAGGTTCTCTAGCGTGGCCTTGACTAGTCGAGGGTGTCCGCTAGCGCTAGCCCTACAAGCTCGTTCCACTAGCAGCGATCGTCAGGTCAAAGGGGCAAGAGGTCAGCGAGTCGAG

TGAGCGCCGTGGCAGTACCAACAACGGCTGAGGTTGAGGCGTCGTGATCGGTCAAAACGGGCAGCGGATCAGGTCGCTCGATATTGGAGAGGTCATCGGCGATGACTAGCTGATGACCGT
                                                 420                                                  480
ACTCGCGGCACCGTCATGGTTGTTGCCGACTCCAACTCCGCAGCACTAGCCAGTTTTGCCCGTCGCCTAGTCCAGCGAGCTATAACCTCTCCAGTAGCCGCTACTGATCGACTACTGGCA

                                                                    AvaI
CAATCAGTGCCGCGCTACTGCGATCACCGCTGTCAGGCTGTAAGTCGCAGGGTTTAAGGATGTGGGCAATGCTCGGGTGAGAGCCATTTGGTCTTACCTGTAGTGAGGGGCGGAAGTAGA
                                                 540                                                  600
GTTAGTCACGGCGCGATGACGCTAGTGGCGACAGTCCGACATTCAGCGTCCCAAATTCCTACACCCGTTACGAGCCCACTCTCGGTAAACCAGAATGGACATCACTCCCCGCCTTCATCT
                                                                                          *  R  Y  H  P  A  S  T  S

AGGCTTAGGCGGTACGATCGCCTTTTCGAGTTCGTTTCGGGTTTCAAGACTGCGGTGGGTTTCCCGCACGATCAGCCCGTGCAGCTCAGTGAGGAGCTGCAGCATCAGCAGCCCCCCTAG
                                                 660                                                  720
TCCGAATCCGCCATGCTAGCGGAAAAGCTCAAGCAAAGCCCAAAGTTCTGACGCCACCCAAAGGGCGTGCTAGTCGGGCACGTCGAGTCACTCCTCGACGTCGTAGTCGTCGGGGGGATC
  P  K  P  P  V  I  A  K  E  L  E  N  R  T  E  L  S  R  H  T  E  R  V  I  L  G  H  L  E  T  L  L  Q  L  M  L  L  G  G  L

GCGGATCTCTAGGCCGTCCAAGCGGAGCCATGCGCCACCTTGTCCACTGGTCTGAACCTCAAGACCAATCGAGCCGATCGAACTTCGCACCGGCTCATCGCTGCGGTCGGTGCCTGCCAG
                                                 780                                                  840
CGCCTAGAGATCCGGCAGGTTCGCCTCGGTACGCGGTGGAACAGGTGACCAGACTTGGAGTTCTGGTTAGCTCGGCTAGCTTGAAGCGTGGCCGAGTAGCGACGCCAGCCACGGACGGTC
  R  I  E  L  G  D  L  R  L  W  A  G  G  Q  G  S  T  Q  V  E  L  G  I  S  G  I  S  S  R  V  P  E  D  S  R  D  T  G  A  L

TCGTCGGGACGCTGCAACCGATGCGCCGATCACTTCCGCGATCGCTGCCAATCTGGGGAGGTCTGCAGGGACAGCACGCTCACTAGCAACCGTCACCACGCCAAACTCACCGCGCTCGCA
                                                 900                                                  960
AGCAGCCCTGCGACGTTGGCTACGCGGCTAGTGAAGGCGCTAGCGACGGTTAGACCCCTCCAGACGTCCCTGTCGTGCGAGTGATCGTTGGCAGTGGTGCGGTTTGAGTGGCGCGAGCGT
  R  R  S  A  A  V  S  A  G  I  V  E  A  I  A  A  L  R  P  L  D  A  P  V  A  R  E  S  A  V  T  V  V  G  F  E  G  R  E  C

GTCGATTACAAGGAACGCCGTTTGGCCTTGAGAGTCGCGACCGGCGATCGCTAGTGAACGTGTCTGGACGATCGCGCCGGTGTCGATGCTGGCCATTGCCCAGCCCATCATTTCTGGGTC
                                                1020                                                 1080
CAGCTAATGTTCCTTGCGGCAAACCGGAACTCTCAGCGCTGGCCGCTAGCGATCACTTGCACAGACCTGCTAGCGCGGCCACAGCTACGACCGGTAACGGGTCGGGTAGTAAAGACCCAG
  D  I  V  L  F  A  T  Q  G  Q  S  D  R  G  A  I  A  L  S  R  T  Q  V  I  A  G  T  D  I  S  A  M  A  W  G  M  M  E  P  D

ATCGCTGGTCAGTCGCAGCCCCCCGGCACTACTTGGAAAACCTTCGATATTTGGCATTTTTAGCCTCGCAATTTGCTGCAAGACTAGCTGAAACAGACAAGCGAAAACTTAGGCCCAGAC
                                                1140                                                 1200
TAGCGACCAGTCAGCGTCGGGGGGCCGTGATGAACCTTTTGGAAGCTATAAACCGTAAAAATCGGAGCGTTAAACGACGTTCTGATCGACTTTGTCTGTTCGCTTTTGAATCCGGGTCTG
  D  S  T  L  R  L  G  G  A  S  S  P  F  G  E  I  N  P  M      ⟵ Orf A

TGGCCTAGAGCTGCTTCAGAATTACCGCTTGTTCCCGCTAATTCACGCTTGTTCCTAACAGTTCACTGAACTATAGGTAAAACTCAATGAACAAACACTGAATGATAAGAAATTCAAATA
                                                1260                                                 1320
ACCGGATCTCGACGAAGTCTTAATGGCGAACAAGGGCGATTAAGTGCGAACAAGGATTGTCAAGTGACTTGATATCCATTTTGAGTTACTTGTTTGTGACTTACTATTCTTTAAGTTTAT
                                                       -10                                  -35

GATCGAGGTGGGGATAATTGCCTCTCTTTGCAGGTGATTGACGCTGATTGCAGACGGCTGGGAGGTGGTTTTTAGCGAGTAAATCGACGCTGATCAGACTTGGGGTAAGAGTCCACAGCC
                                                1380                                                 1440
CTAGCTCCACCCCTATTAACGGAGAGAAACGTCCACTAACTGCGACTAACGTCTGCCGACCCTCCACCAAAAATCGCTCATTTAGCTGCGACTAGTCTGAACCCCATTCTCAGGTGTCGG

                    Orf B ⟶          M  I  E  P  H  H  F  Y  E  A  A  L  A  A  F  S  Q  I  D  W  R  A  A  V  N
GACGCAGCAAAACTAGGCCGGCCCTTAACTAAACCCCAAGAAAAATGATTGAACCTCACCACTTCTACGAAGCAGCACTAGCCGCCTTCTCCCAAATCGATTGGCGGGCTGCAGTGAACA
                                                1500                                                 1560
CTGCGTCGTTTTGATCCGGCCGGGAATTGATTTGGGGTTCTTTTTACTAACTTGGAGTGGTGAAGATGCTTCGTCGTGATCGGCGGAAGAGGGTTTAGCTAACCGCCCGACGTCACTTGT
```


BglII
I Y S Q L S R L S R Q S S A S T M R A P A L P S T V T D N T S D L I L Q R L D T
TCTACTCTCAGTTGAGCCGCCTTAGCAGGCAATCGTCAGCGAGTACGATGAGGGCACCCGCTCTGCCCTCGACCGTGACCGACAACACCTCAGATCTCATCCTTCAGCGACTAGACACGA
1620
1680
AGATGAGAGTCAACTCGGCGGAATCGTCCGTTAGCAGTCGCTCATGCTACTCCCGTGGGCGAGACGGGAGCTGGCACTGGCTGTTGTGGAGTCTAGAGTAGGAAGTCGCTGATCTGTGCT
I D A R L G R I E D R L D R V E S D L S E L K T D V K E V R S S L A E L T G D V
TCGATGCGCGGCTAGGCCGAATCGAGGATCGTCTTGATCGCGTCGAATCGGATCTGTCTGAGCTGAAGACCGACGTCAAAGAGGTGCGATCGTCACTAGCTGAATTGACTGGGGACGTTC
1740
1800
AGCTACGCGCCGATCCGGCTTAGCTCCTAGCAGAACTAGCGCAGCTTAGCCTAGACAGACTCGACTTCTGGCTGCAGTTTCTCCACGCTAGCAGTGATCGACTTAACTGACCCCTGCAAG
R K I E G V L Q G A G L P N I N T R I G I A T V S S A A M V A F F L M L A D R F
GCAAGATTGAAGGCGTCCTGCAAGGTGCAGGGCTGCCAAACATCAACACGAGGATCGGCATTGCCACGGTGAGCAGCGCCGCAATGGTGGCATTCTTTCTGATGCTGGCCGATCGCTTTC
1860
1920
CGTTCTAACTTCCGCAGGACGTTCCACGTCCCGACGGTTTGTAGTTGTGCTCCTAGCCGTAACGGTGCCACTCGTCGCGGCGTTACCACCGTAAGAAAGACTACGACCGGCTAGCGAAAG
HpaI
HpaI
L G * Orf C⟶ M P S P L V R V P D A L R D Q V L T L C R L H R E G L P V K A A L E
TGGGGTAGCTGTTAACAGTGCCAAGCCCTTTAGTCCGAGTGCCTGACGCACTGCGCGATCAGGTGTTAACACTGTGCCGCCTTCACCGTGAAGGGCTGCCGGTAAAAGCCGCTCTGGAAC
1980
2040
ACCCCATCGACAATTGTCACGGTTCGGGAAATCAGGCTCACGGACTGCGTGACGCGCTAGTCCACAATTGTGACACGGCGGAAGTGGCACTTCCCGACGGCCATTTTCGGCGAGACCTTG
AccI
HpaI
HpaI
HpaI
Q A I A G L Q A V D S P V N S S P H Q T A L D D L A R R V E A L E Q L T A L L T
AGGCGATCGCTGGCCTGCAAGCTGTAGACAGCCCTGTTAACAGCAGCCCACATCAGACAGCCTTGGATGACCTCGCCAGACGGGTCGAAGCCTTGGAGCAGTTAACAGCCCTGTTAACAC
2100
TCCGCTAGCGACCGGACGTTCGACATCTGTCGGGACAATTGTCGTCGGGTGTAGTCTGTCGGAACCTACTGGAGCGGTCTGCCCAGCTTCGGAACCTCGTCAATTGTCGGGACAATTGTG
HpaI
HpaI
HpaI
HpaI
P S G E P V N T S A N R S V N S K A L S V N S P V N R K S D S V N S S G D W L T
CTTCGGGAGAGCCTGTTAACACTTCTGCTAACAGGTCTGTTAACAGCAAGGCACTCTCTGTAAACAGCCCTGTTAACCGAAAGTCCGATTCTGTTAACAGCTCTGGCGACTGGCTAACGG
2280
GAAGCCCTCTCGGACAATTGTGAAGACGATTGTCCAGACAATTGTCGTTCCGTGAGAGACATTTGTCGGGACAATTGGCTTTCAGGCTAAGACAATTGTCGAGACCGCTGACCGATTGCC
V E E A H Q L A A E R G C P A S L A T F R R W S R G N A K Y P Q G D V A S L K Q
TTGAGGAGGCTCATCAACTTGCGGCTGAGCGGGGCTGCCCTGCGAGTCTGGCCACCTTCAGACGTTGGAGCAGAGGAAATGCGAAGTATCCCCAGGGCGATGTGGCCTCGCTGAAGCAAT
2340
2400
AACTCCTCCGAGTAGTTGAACGCCGACTCGCCCCGACGGGACGCTCAGACCGGTGGAAGTCTGCAACCTCGTCTCCTTTACGCTTCATAGGGGTCCCGCTACACCGGAGCGACTTCGTTA
W G F D R D M T R L A S G S S K N P A R F L R A I A D *
GGGGGTTCGATCGGGATATGACCCGATTGGCCAGCGGCAGCTCAAAGAATCCGGCGCGGTTCCTACGGGCGATCGCAGACTGAGACAGGGCTGTCTTAGTCGTGCTGTCTCACGCGGCCA
2460
2520
CCCCCAAGCTAGCCCTATACTGGGCTAACCGGTCGCCGTCGAGTTTCTTAGGCCGCGCCAAGGATGCCCGCTAGCGTCTGACTCTGTCCCGACAGAATCAGCACGACAGAGTGCGCCGGT
2580
2640
TCAAACTACTTACGGCCTAGATTCGGGACTCGCGAGGCGAGCAACAGAGTCCGTCCCATGAATCAGGTTACCGACTCCCCCTCGTAGATTTCCGATCCATCCTGTCCCGACTACTGAACG
2700
2760
AGGGATCCGTCTGTGCGGTCTGTCGCCGAGGTTTCGGACGCTATGCAGAGCGACGTCGTTGCGCCGCTAACGGAGGGCGAAGAGGCCACCGTCAAGGGGTCACTGCCGGAGTTAGCTGGC
* A S V G S L P E L A Q S V D R Q L L A A I A E R K E P P L E G T V A E I S R
EcoRV
CACACAAGCAGCCGGAACCGTCGCAAAGCGATCGATCACCCGCTTGCTCTCGGCTGCCCAAGTGCTCGCGATATCCTCAGCATCAACTAGCGATCGCTCGCGTTCTTCAACGTCGAGCCG
2820
2880
GTGTGTTCGTCGGCCTTGGCAGCGTTTCGCTAGCTAGTGGGCGAACGAGAGCCGACGGGTTCACGAGCGCTATAGGAGTCGTAGTTGATCGCTAGCGAGCGCAAGAAGTTGCAGCTCGGC
V C A A P V T A F R D I V R K S E A A W T S A I D E A D V L S R E R E E V D L R
GAGCAGCTCGGCGCGGTAGTGCGCTTTCAGACGCCGCGCCTCTAGTTGCTCGGCTGATGGAGGCTCACGACGCTTGCCACGCCCGCCACCATCGTTCCTGATCCGGCTCGGATCGGTGTT
2940
3000
CTCGTCGAGCCGCGCCATCACGCGAAAGTCTGCGGCGCGGAGATCAACGAGCCGACTACCTCCGAGTGCTGCGAACGGTGCGGGCGGTGGTAGCAAGGACTAGGCCGAGCCTAGCCACAA
L L E A R Y H A K L R R A E L Q E A S P P E R R K G R G G G D N R I R S P D T N
TGCTGCCCACTCGCGAGCCGCAGCAACCGGATCGACGTGCCACCGACTGCCGACCTTCCAGACTGAAGTGATCCGACCGGCTGCGATCGCCTGATGTATCGCCTGAGGCGCAACGCCATT
3060
3120
ACGACGGGTGAGCGCTCGGCGTCGTTGGCCTAGCTGCACGGTGGCTGACGGCTGGAAGGTCTGACTTCACTAGGCTGGCCGACGCTAGCGGACTACATAGCGGACTCCGCGTTGCGGTAA
A A W E R A A A V P D V H W R S G V K W V S T I R G A A I A Q H I A Q P A V G N


AvaI
NcoI
3180
A R A F E P L S Q S E A M ⟵ Orf D
ApaLI
3300
XhoI / AvaI
3420
3540
* A A T E T S T T S V
MstI
3660
3780
A E L P L E A I A E N P V V Q Y I W T S K G A I R K Q R A K L K F G I M R L L A
3900
R V I T S A R T K D T I Q I K L I Q A A S A R H S T A T T V I E Q I L P D D V S
PvuII
AvaII
4020
F W D R R Q L L E E V K L G H L Y R V Q P A I C S G M ⟵ Orf E
* G M C T G C K P R S V A A W L I P D S C R A I A G N T S
4140
S P P M A L P V A D K G L S L W F L L R L P A Y L G D D D A A L L E R T C D G W
4260
R Q Q C R H R A L A A L Q E A N A F D A S A I A E A E K P T L D P S S L I A E A
4380
S A D A L A E A T E R Q L E D V V E Q L F P D I A G D W D Q W T F G N L R L L Q
4500
E V T A A Y T L G D R N R R A A L E A W A T L Y A Q S A V P E G I K I A A Q S L
PvuII
4620
Q A L I V Q Q H R K L G D A V E T P D I S G N G A R M L N S T D P C T I F R P V

```
CGTTGCTCCGCACCCGTCCTGCAATCTGCGCGATCGCACCTGGGGCGATGTGCCCAGCTCCGGCGAAGATCAGCACGGCATCAAATGGGCGATCGCCTACCCCTTGGGTGATCGACAAGC
                                                  4740                                                  4800
GCAACGAGGCGTGGGCAGGACGTTAGACGCGCTAGCGTGGACCCCGCTACACGGGTCGAGGCCGCTTCTAGTCGTGCCGTAGTTTACCCGCTAGCGGATGGGGAACCCACTAGCTGTTCG
  N   S   R   V   R   G   A   I   Q   A   I   A   G   P   A   I   H   G   A   G   A   F   I   L   V   A   D   F   P   R   D   G   V   G   Q   T   I   S   L   G

CAGCATCGATCACGCTCGTGAAGACAAGGTGGGGGCAACAGTTGATGAACATCTGCGGGTTGAGCTTCAGCGTTTCCAAAGCCCCCCGCTTATTGGCTTCGCTGTCGATGACCGGAACCG
                                                  4860                                                  4920
GTCGTAGCTAGTGCGAGCACTTCTGTTCCACCCCCGTTGTCAACTACTTGTAGACGCCCAACTCGAAGTCGCAAAGGTTTCGGGGGGCGAATAACCGAAGCGACAGCTACTGGCCTTGGC
  A   D   I   V   S   T   F   V   L   H   P   C   C   N   I   F   M   Q   P   N   L   K   L   T   E   L   A   G   R   K   N   A   E   S   D   I   V   P   V   P

GCTGACCGCAGGCTTCGGAGAGCTGTTGGGCGATCGTCGCTGCGGAATACAGGGACGGCCGCGACTGGCTCAGCTTCTGGCGGTCGCAGAAGACTGCTAAGCGCTTGCCCGCCTGAACCA
                                                  4980                                                  5040
CGACTGGCGTCCGAAGCCTCTCGACAACCCGCTAGCAGCGACGCCTTATGTCCCTGCCGGCGCTGACCGAGTCGAAGACCGCCAGCGTCTTCTGACGATTCGCGAACGGGCGGACTTGGT
  Q   G   C   A   E   S   L   Q   Q   A   I   T   A   A   S   Y   L   S   P   R   S   Q   S   L   K   Q   R   D   C   F   V   A   L   R   K   G   A   Q   V   M

TTTCGACCGCCATCGCCATGAGTTCGCGATCGCTGCCAACCTGATAGACCGGAGCACCGCTGAAGGGTTGGACGGTACTGCTGACCAGATATGCGGTTGATCCTGTCAGTGCTTCCAACA
                                                  5100                                                  5160
AAAGCTGGCGGTAGCGGTACTCAAGCGCTAGCGACGGTTGGACTATCTGGCCTCGTGGCGACTTCCCAACCTGCCATGACGACTGGTCTATACGCCAACTAGGACAGTCACGAAGGTTGT
  E   V   A   M   A   M   L   E   R   D   S   G   V   Q   Y   V   P   A   G   S   F   P   Q   V   T   S   S   V   L   Y   A   T   S   G   T   L   A   E   L   L

GGGCGATCGACACATCCGACAGCATCGCGTCAGCGGCGATAACCTGCCGCCCCCCTGCGAGTAGTCCAGCAATTTCAGCCATGACTTCGGGGCGATGTTTAGCGATCTCGGTCTTGCCGC
                                                  5220                                                  5280
CCCGCTAGCTGTGTAGGCTGTCGTAGCGCAGTCGCCGCTATTGGACGGCGGGGGGACGCTCATCAGGTCGTTAAAGTCGGTACTGAAGCCCCGCTACAAATCGCTAGAGCCAGAACGGCG
  A   I   S   V   D   S   L   M   A   D   A   A   I   V   Q   R   G   G   A   L   L   G   A   I   E   A   M   V   E   P   R   H   K   A   I   E   T   K   G   S

TCAGCAGGTGCTCAGCCCACAGTTCCACCTCATCGATGACGACGATCGCCCGTTGCCATCCGTGGGCGCTGAAGCGCAGCTTGCTGGCGGGGTGCATGGAGTCTAGGCAACCGCCAAGGC
                                                  5340                                                  5400
AGTCGTCCACGAGTCGGGTGTCAAGGTGGAGTAGCTACTGCTGCTAGCGGGCAACGGTAGGCACCCGCGACTTCGCGTCGAACGACCGCCCCACGTACCTCAGATCCGTTGGCGGTTCCG
  L   L   H   E   A   W   L   E   V   E   D   I   V   V   I   A   R   Q   W   G   H   A   S   F   R   L   K   S   A   P   H   M   S   D   L   C   G   G   L   G

CGATCATCCGAAGGTCGGAACCGGGAACGCATTCTTCAGCCCACGGCACCCCTAAGCGCTTGCCCAAGGCCGCGCCCAAGGTACGGCGATGCGTGATCAACACGACAGGAATGCCGGTAC
                                                  5460                                                  5520
GCTAGTAGGCTTCCAGCCTTGGCCCTTGCGTAAGAAGTCGGGTGCCGTGGGGATTCGCGAACGGGTTCCGGCGCGGGTTCCATGCCGCTACGCACTAGTTGTGCTGTCCTTACGGCCATG
  I   M   R   L   D   S   G   P   V   C   E   E   A   W   P   V   G   L   R   K   G   L   A   A   G   L   T   R   R   H   T   I   L   V   V   P   I   G   T   G

CAATGAGCGGCTGGAGATGCCCCGCGATCGCCTCTGTCTTGCCCGTCCCCATCGCAGCATTCAGGCAGACAAGGCGATGCTGTTCCGGCTGCGGGATGTAGTCGCCGTTGATGTACCCGC
                                                  5580                                                  5640
GTTACTCGCCGACCTCTACGGGGCGCTAGCGGAGACAGAACGGGCAGGGGTAGCGTCGTAAGTCCGTCTGTTCCGCTACGACAAGGCCGACGCCCTACATCAGCGGCAACTACATGGGCG
  I   L   P   Q   L   H   G   A   I   A   E   T   K   G   T   G   M   A   A   N   L   C   V   L   R   H   Q   E   P   Q   P   I   Y   D   G   N   I   Y   G   G

CGGTGTGTAAGCGGTTAGGGCTACGGGTCAGCCGCTGCCACTCGTTCGCGACCTCGCGCTGATATTCCCGCTGCTCTTGTCGAGTCAGCTTTGGCGGGCGGTAGCCGCCTTCTTTGGCCA
                                                  5700                                                  5760
GCCACACATTCGCCAATCCCGATGCCCAGTCGGCGACGGTGAGCAAGCGCTGGAGCGCGACTATAAGGGCGACGAGAACAGCTCAGTCGAAACCGCCCGCCATCGGCGGAAGAAACCGGT
  T   H   L   R   N   P   S   R   T   L   R   Q   W   E   N   A   V   E   R   Q   Y   E   R   Q   E   Q   R   T   L   K   P   P   R   Y   G   G   E   K   A   L

GCTTGGCCAGTGTGGCGATCGTGATGCCGCTGCCCTTGAAGCTGGCCCACTTCCGCTCTAGCTTGCGGTTGCCGTCGAACTTGGACGATTGCGCTGACCATGCAATCCAGTCATCTAACA
                                                  5820                                                  5880
CGAACCGGTCACACCGCTAGCACTACGGCGACGGGAACTTCGACCGGGTGAAGGCGAGATCGAACGCCAACGGCAGCTTGAACCTGCTAACGCGACTGGTACGTTAGGTCAGTAGATTGT
  K   A   L   T   A   I   T   I   G   S   G   K   F   S   A   W   K   R   E   L   K   R   N   G   D   F   K   S   S   Q   A   S   W   A   I   W   D   D   L   L

GGGCATCGCTGACTGACTGCAGGGCCATCCCGACAGCCACCCAGCTCTCGTAGTCGTCTGCGTAGCTGGCAGGTACATGGGCCAGCATGGTGCGGGCGATCGCGACCTCATCATCAGAAG
                                                  5940                                                  6000
CCCGTAGCGACTGACTGACGTCCCGGTAGGGCTGTCGGTGGGTCGAGAGCATCAGCAGACGCATCGACCGTCCATGTACCCGGTCGTACCACGCCCGCTAGCGCTGGAGTAGTAGTCTTC
  A   D   S   V   S   Q   L   A   M   G   V   A   V   W   S   E   Y   D   D   A   Y   S   A   P   V   H   A   L   M   T   R   A   I   A   V   E   D   D   S   A

                                                                                                        MstI
CTGGAGCACTGACGATCGGCTTATAGGATTTCGGCGCGGTCTGCCGGTCTTGCCGCTCCATTGCATCCAGCAATTCGGGCGGGGCCTCTGCAACCTCGATCTCCCAAGGTGCGCAGCCTT
                                                  6060                                                  6120
GACCTCGTGACTGCTAGCCGAATATCCTAAAGCCGCGCCAGACGGCCAGAACGGCGAGGTAACGTAGGTCGTTAAGCCCGCCCCGGAGACGTTGGAGCTAGAGGGTTCCACGCGTCGGAA
  P   A   S   V   I   P   K   Y   S   K   P   A   T   Q   R   D   Q   R   E   M   A   D   L   L   E   P   P   A   E   A   V   E   I   E   W   P   A   C   G   E

                                                                                                      HindIII
CCACCCAGCGATATGCGCCGGTCTCAGGGTGTTGACCAACGATGACTGATTGCAGCTTGTTCCAACGCAGCTCTAACAGATCACCGTCTGGGCTGTGAAGCTTCTGCGTGGCGATCGCAG
                                                  6180                                                  6240
GGTGGGTCGCTATACGCGGCCAGAGTCCCACAACTGGTTGCTACTGACTAACGTCGAACAAGGTTGCGTCGAGATTGTCTAGTGGCAGACCCGACACTTCGAAGACGCACCGCTAGCGTC
  V   W   R   Y   A   G   T   E   P   H   Q   G   V   I   V   S   Q   L   K   N   W   R   L   E   L   L   D   G   D   P   S   H   L   K   Q   T   A   I   A   A
```

```
CCCATCGCTCTTGGGGTACCTGATAGAGCCGTTGGTGGCGTCCTGGCTTGCCAGATGTATATGCAACTGTGGGCGGTAATTCTGTCAGCGTCAGCCCTTCTGGCAGTGCCTCATGAGCCG
                                                  6300                                                  6360
GGGTAGCGAGAACCCCATGGACTATCTCGGCAACCACCGCAGGACCGAACGGTCTACATATACGTTGACACCCGCCATTAAGACAGTCGCAGTCGGGAAGACCGTCACGGAGTACTCGGC
  W  R  E  Q  P  V  Q  Y  L  R  Q  H  R  G  P  K  G  S  T  Y  A  V  T  P  P  L  E  T  L  T  L  G  E  P  L  A  E  H  A  T

TTGGGCCATCGAAGTCAACTGCGATTAACCCACCGCTTGCGGGGCCAGTAAGCAGACCGATTCCGGTAACGCGCCGATCGCTGCGGATACGTTCAGCAACAGCAGCATGGTCAAGCGGGT
                                                  6420                                                  6480
AACCCGGTAGCTTCAGTTGACGCTAATTGGGTGGCGAACGCCCCGGTCATTCGTCTGGCTAAGGCCATTGCGCGGCTAGCGACGCCTATGCAAGTCGTTGTCGTCGTACCAGTTCGCCCA
  P  G  D  F  D  V  A  I  L  G  G  S  A  P  G  T  L  L  G  I  G  T  V  R  R  D  S  R  I  R  E  A  V  A  A  H  D  L  P  N

                                    HpaI
TATTTTGCCAGCCAGGGTCAAATGGTGCCTTGCGACTGTTAACGACGCAGAAGCGCCAGTTTGGGGGCAGCCGGTCAAGTTGATCAATTAATTCGCAGCGGGTCGGGATCTGCGACCTTT
                                                  6540                                                  6600
ATAAAACGGTCGGTCCCAGTTTACCACGGAACGCTGACAATTGCTGCGTCTTCGCGGTCAAACCCCCGTCGGCCAGTTCAACTAGTTAATTAAGCGTCGCCCAGCCCTAGACGCTGGAAA
  N  Q  W  G  P  D  F  P  A  K  R  S  N  V  V  C  F  R  W  N  P  P  L  R  D  L  Q  D  I  L  E  C  R  T  P  I  Q  S  R  Q

                  ←- ——  ——→  ←- ——  ——→
GCAGATGATTGGCGATAAATGCCATACTTCTGTTGAGTTATTTTTCCTGACCTCCCCTGGGCAAAGTTTGCCGACCTGGCCTGGGGGGGGTTTTGCTTTGATTGCCGATAACTATAGCTG
                  ‹▪ ▪▪▪▪▪▪▪▪▪▪▪▪▪▪ ▪ ▪▪▪ · ▪▪▪ ▪ ▪▪▪▪▪▪▪▪▪▪▪▪▪▪ ▪›                                  6720
CGTCTACTAACCGCTATTTACGGTATGAAGACAACTCAATAAAAAGGACTGGAGGGGACCCGTTTCAAACGGCTGGACCGGACCCCCCCCAAAACGAAACTAACGGCTATTGATATCGAC
  L  H  N  A  I  F  A  M  ←—— Orf F                                                                    -10

GAATGAATACGGCAGGTGGCAAGGCTGCCTCGAATTGATCGGCAAGCCAGAGCGCGATCGCCCCTGGCTGTGATGAGACTTAGACGCCGGACGCTGTCGAGATCCGAACGATCGGACGGT
                                                  6780                                                  6840
CTTACTTATGCCGTCCACCGTTCCGACGGAGCTTAACTAGCCGTTCGGTCTCGCGCTAGCGGGGACCGACACTACTCTGAATCTGCGGCCTGCGACAGCTCTAGGCTTGCTAGCCTGCCA
            -35                                                   * V  G  S  A  T  S  I  R  V  I  P  R  H

GATCGACTACGCAGCAGCCGAAGCTGTGCTCAGCCTTGGCGCGGTAAAAGCCCGTCTCGAACTGCACCTCGCTAGTTAGCTGAGGCGCGGGCGTCAGGCGTCCCAGCTCCAGAGGTAGGT
                                                  6900                                                  6960
CTAGCTGATGCGTCGTCGGCTTCGACACGAGTCGGAACCGCGCCATTTTCGGGCAGAGCTTGACGTGGAGCGATCAATCGACTCCGCGCCCGCAGTCCGCAGGGTCGAGGTCTCCATCCA
  D  V  V  C  C  G  F  S  H  E  A  K  A  R  Y  F  G  T  E  F  Q  V  E  S  T  L  Q  P  A  P  T  L  R  G  L  E  L  P  L  N

                                                                                        BglII
TGCCGCGTGCAACTAGGTAGTAGTAGCTGCCGCTCAGCCGTGCGTCGATCGCAACCTCCAGCGACCAAGGGTTGAACTCATCGCGGCTGTTGACTTGCAGATCTGCCACCAACTGACGGA
                                                  7020                                                  7080
ACGGCGCACGTTGATCCATCATCATCGACGGCGAGTCGGCACGCAGCTAGCGTTGGAGGTCGCTGGTTCCCAACTTGAGTAGCGCCGACAACTGAACGTCTAGACGGTGGTTGACTGCCT
  G  R  A  V  L  Y  Y  Y  S  G  S  L  R  A  D  I  A  V  E  L  S  W  P  N  F  E  D  R  S  N  V  Q  L  D  A  V  L  Q  R  I

                                    NotI
TAGGCAGCTCACTCTCAGGCGACGCGAGGATCAGTCCAGGCTCTTGATCGAGCAGGCGGCCGCCGATCGTCGCCTGACTTCTCAACAACGCGATCGCCGCACCCAAGCTCGGAAGCGTGA
                                                  7140                                                  7200
ATCCGTCGAGTGAGAGTCCGCTGCGCTCCTAGTCAGGTCCGAGAACTAGCTCGTCCGCCGGCGGCTAGCAGCGGACTGAAGAGTTGTTGCGCTAGCGGCGTGGGTTCGAGCCTTCGCACT
  P  L  E  S  E  P  S  A  L  I  L  G  P  E  Q  D  L  L  R  G  G  I  T  A  Q  S  R  L  L  A  I  A  A  G  L  S  P  L  T  I

TGCCAGTCTCATCCGTCGTCTCGTTCAGGTGCTCCGGTGCGAACAAGGCCAAGCCATCGGGTGCGGGTGCTCCGTTGGCTGAACCTTCGATTAGGGCAACCAGTGCTGCCACTTCAGCCG
                                                  7260                                                  7320
ACGGTCAGAGTAGGCAGCAGAGCAAGTCCACGAGGCCACGCTTGTTCCGGTTCGGTAGCCCACGCCCACGAGGCAACCGACTTGGAAGCTAATCCCGTTGGTCACGACGGTGAAGTCGGC
  G  T  E  D  T  T  E  N  L  H  E  P  A  F  L  A  L  G  D  P  A  P  A  G  N  A  S  G  E  I  L  A  V  L  A  A  V  E  A  S

                                                                  AflII
ATAGGCAGGCATCGCCCAGCTCAGCGAAGCCTTGCTGCAGGCCAGCGAGATCATCGTTGCTTAAGGACTCCGGCGTTGCCACCAGCCCTAGGCCGAAGCGCTGCGGCTTGTACTTCGCGA
                                                  7380                                                  7440
TATCCGTCCGTAGCGGGTCGAGTCGCTTCGGAACGACGTCCGGTCGCTCTAGTAGCAACGAATTCCTGAGGCCGCAACGGTGGTCGGGATCCGGCTTCGCGACGCCGAACATGAAGCGCT
  L  C  A  D  G  L  E  A  F  G  Q  Q  L  G  A  L  D  D  N  S  L  S  E  P  T  A  M  ←—— Orf G

GCGAGTCTTCGACATAGGAGAAGGTGATCTCGCCACCTTGCCGGATCGGGCGAGCCTCACCGACTAAGCCGTGGCGACGGTCGAAGCCTCGCGATAGTCGTTCAGCGGTCGGATGTTCGC
                                                  7500                                                  7560
CGCTCAGAAGCTGTATCCTCTTCCACTAGAGCGGTGGAACGGCCTAGCCCGCTCGGAGTGGCTGATTCGGCACCGCTGCCAGCTTCGGAGCGCTATCAGCAAGTCGCCAGCCTACAAGCG
                                                  * A  T  A  V  T  S  A  E  R  Y  D  N  L  P  R  I  N  A

AATCCGCAGGACGCCACGAGCCGGAGCACTAAAGCGTTGCTGCAGGATGCGATCGCCGGTAGATTCCAGTGCTGCCGGAAGATCCGAGGTGCTCATCGCACGGGCGATCACTTGTGCCGG
                                                  7620                                                  7680
TTAGGCGTCCTGCGGTGCTCGGCCTCGTGATTTCGCAACGACGTCCTACGCTAGCGGCCATCTAAGGTCACGACGGCCTTCTAGGCTCCACGAGTAGCGTGCCCGCTAGTGAACACGGCC
  I  R  L  V  G  R  A  P  A  S  F  R  Q  Q  L  I  R  D  G  T  S  E  L  A  A  P  L  D  S  T  S  M  A  R  A  I  V  Q  A  P

                                    HindIII
AGTGCGGCCTTGGATACCAGCCGCTCGCAGTGCCCAAGCCAAGCTTTCTTGGCCATCAAAAGCTCTGGCGACGGCGTTAGCGATGGGCGACTGCCGCTCACGGGTCACATGGATCTGGGT
                                                  7740                                                  7800
TCACGCCGGAACCTATGGTCGGCGAGCGTCACGGGTTCGGTTCGAAAGAACCGGTAGTTTTCGAGACCGCTGCCGCAATCGCTACCCGCTGACGGCGAGTGCCCAGTGTACCTAGACCCA
  T  R  G  Q  I  G  A  A  R  L  A  W  A  L  S  E  Q  G  D  F  A  R  A  V  A  N  A  I  P  S  Q  R  E  R  T  V  H  I  Q  T

CACGCTTGGAGCCTCCCGGCTGGCTAGCTCCTCAA
                  7835
GTGCGAACCTCGGAGGGCCGACCGATCGAGGAGTT
 V  S  P  A  E  R  S  A  L  E  E  L
```

Table II. Restriction sites on pUH24

Enzyme	Site	#	Location				
AatII	(GACGTC)	1	1752				
AccI	(GTMKAC)	1	2064				
AflII	(CTTAAG)	1	7380				
ApaLI	(GTGCAC)	1	3305				
AseI	(ATTAAT)	1	6567				
AvaII	(GGWCC)	1	4048				
BamHI	(GGATCC)	1	1				
BspHI	(TCATGA)	1	6351				
BssHII	(GCGCGC)	1	71				
MstII	(CCTNAGG)	1	3102				
EcoNI	(CCTN$_5$AGG)	1	4111				
Eco0109I	(RGGNCCY)	1	6081				
EcoRV	(GATATC)	1	2830				
NcoI	(CCATGG)	1	3156				
NotI	(GCGGCCGC)	1	7136				
NspHI	(RCATGY)	1	3427				
SfiI	(GGCCN$_5$GGCC)	1	5756				
Tth111I	(GACN$_3$GTC)	1	6810				
XhoI	(CTCGAG)	1	3413				
EagI	(CGGCCG)	2	4976	7137			
BglII	(AGATCT)	2	1652	7059			
HindIII	(AAGCTT)	2	6218	7721			
MstI	(TGCGCA)	2	3625	6110			
NaeI	(GCCGGC)	2	1457	3356			
NheI	(GCTAGC)	2	321	7823			
PvuII	(CAGCTG)	2	4030	4559			
AflIII	(ACRYGT)	3	1018	3427	4651		
AvaI	(CYCGRG)	3	552	3133	3413		
AvrII	(CCTAGG)	3	716	2643	7407		
KpnI	(GGTACC)	3	4007	5516	6255		
XmnI	(GAAN$_4$TTC)	3	1116	5035	5425		
BclI	(TGATCA)	4	268	1411	5494	6561	
BglI	(GCCN$_5$GGC)	4	5757	5766	6334	6670	
NciI	(CCSGG)	4	1102	3905	5421	7815	
AlwNI	(CAGN$_3$CTG)	5	3139	3675	4991	5669	5995
ClaI	(ATCGAT)	6	1536	1680	2791	4388	4805
			5312				
Eco47III	(AGCGCT)	7	109	2550	3409	4335	5020
			5455	7417			
GdiII	(YGGCCG)	7	1906	2515	3310	3366	4486
			4976	7137			
BalI	(TGGCCA)	8	1050	2329	2428	3422	3448
			5755	5764	7730		
PstI	(CTGCAG)	9	121	148	697	903	1549
			2692	5897	7355	7600	
Sau96I	(GGNCC)	9	1192	1460	4048	5804	5902
			5959	6082	6363	6402	
AhaII	(GRCGYC)	10	398	1752	1812	2910	3233
			6277	6803	6930	6937	7570
BanI	(GGYRCC)	10	829	1613	3245	4007	4216
			4582	5445	5516	6255	6504
BstNI	(CCWGG)	10	2372	3242	4719	6282	6492
			6656	6675	6680	6783	7116
HgiAI	(GWGCWC)	10	2822	3305	5092	5288	6004
			6866	7229	7267	7584	7649
HpaI	(GTTAAC)	10	1931	1985	2076	2141	2153
			2175	2199	2232	2253	6518
EspI	(GCTNAGC)	11	352	2304	4121	4561	4988
			5016	5279	5290	6868	6992
			7339				
MaeIII	(GTNAC)	12	932	1635	1772	2741	3615
			3663	3925	4469	4555	6426
			7784	7799			

Enzyme	Site	#	Location				
NruI	(TCGCGA)	12	288	995	2826	3011	3221
			3289	4093	5064	5686	5980
			7435	7529			
XhoII	(RGATCY)	12	1	723	1652	1730	2536
			3767	3887	6586	6820	7059
			7640	7791			
HaeII	(RGCGCY)	13	109	164	363	1875	2550
			3409	3835	4335	5020	5345
			5455	6532	7417		
RsaI	(GTAC)	14	375	612	1603	2577	3239
			3298	4008	5115	5481	5517
			5633	5953	6256	7430	
ScrFI	(CCNGG)	14	1102	2372	3242	3905	4719
			5421	6282	6492	6656	6675
			6680	6783	7116	7815	
SduI	(GDGCHC)	14	1612	2822	3305	3376	4730
			5092	5288	6004	6866	7229
			7267	7584	7649	7710	
Cfr10I	(RCCGGY)	15	809	1000	1036	1457	2018
			3076	3356	4527	4917	5513
			5639	6043	6137	6550	7615
EaeI	(YGGCCR)	15	1050	1906	2329	2428	2515
			3310	3366	3422	3448	4486
			4976	5755	5764	7137	7730
HincII	(GTYRAC)	15	1931	1985	2076	2141	2153
			2175	2199	2232	2253	4364
			4461	6151	6374	6518	7049
MseI	(TTAA)	15	534	1465	1932	1986	2077
			2142	2154	2176	2200	2233
			2254	6385	6519	6568	7381
NlaIII	(CATG)	15	749	3157	3428	3568	4001
			4500	5057	5240	5375	5859
			5956	5966	6352	6466	7788
StyI	(CCWWGG)	15	716	2101	2131	2643	3156
			4116	4782	5394	5464	5476
			6104	6874	7025	7407	7688
NspBII	(CMGCKG)	18	33	186	422	507	2432
			2662	3942	4030	4168	4212
			4559	5097	5191	5672	5708
			5787	6576	7543		
HinfI	(GANTC)	19	154	204	992	1429	1699
			1725	2247	2324	2447	3167
			3878	3938	4670	5379	5722
			6419	7386	7443	7622	
NlaIV	(GGN$_2$CC)	21	1	744	829	1613	2458
			2666	2774	3245	4007	4047
			4216	4582	4914	5417	5445
			5516	6081	6255	6401	6504
			7808				
MaeI	(CTAG)	27	322	466	717	729	923
			1011	1164	1205	1453	1518
			1671	1692	1776	2614	2644
			2847	2922	3162	3685	3901
			4174	5383	5818	6912	6973
			7408	7824			
SecI	(CCN$_2$GG)	34	160	367	716	1101	1864
			2101	2131	2371	2372	2643
			3156	4116	4315	4447	4606
			4719	4782	5340	5394	5441
			5464	5476	6104	6492	6655
			6656	6680	6782	6874	7025
			7407	7509	7644	7688	

Table II. *continued*

Enzyme	Site	#	Location				
*Hpa*II	(CCGG)	39	261	810	1001	1037	1103
			1458	2019	2451	2534	2724
			2772	2878	2983	3027	3077
			3357	3906	4528	4740	4912
			4918	5089	5421	5514	5605
			5640	6044	6138	6423	6551
			6807	7234	7390	7481	7581
			7616	7635	7677	7816	
*Pvu*I	(CGATCG)	40	285	291	615	795	879
			1005	1029	1678	1767	1910
			2044	2408	2470	2789	2851
			3084	3519	3528	3716	4063
			4267	4390	4711	4770	4951
			5067	5164	5323	5545	5776
			5977	6013	6232	6435	6775
			6829	7006	7144	7171	7610
*Hae*III	(GGCC)	41	86	732	984	1051	1192
			1202	1456	1460	1695	1907
			2052	2330	2384	2429	2516
			2746	3311	3325	3367	3423
			3449	4487	4594	4977	5398
			5468	5756	5765	5804	5903
			5960	6083	6364	6403	6678
			7138	7247	7360	7411	7686
			7731				
*Dde*I	(CTNAG)	43	353	360	388	604	685
			1188	1567	1581	1649	1738
			2305	2480	2495	2540	2547
			2567	2580	2593	2836	3103
			3172	3264	3277	3389	4122
			4140	4274	4305	4321	4562
			4989	5017	5280	5291	5452
			6143	6799	6869	6920	6993
			7094	7340	7504		
*Hha*I	(GCGC)	47	71	73	110	136	165
			364	492	752	863	952
			1034	1686	1876	1973	2454
			2551	2891	2901	2916	3108
			3122	3410	3477	3626	3792
			3836	4022	4068	4336	4395
			4708	5021	5346	5354	5456
			5472	5696	5851	6034	6111
			6135	6431	6533	6772	6879
			6925	7418			
*Taq*I	(TCGA)	51	66	202	229	303	438
			626	789	798	962	1043
			1124	1323	1404	1537	1629
			1681	1702	1723	2125	2407
			2754	2792	2873	3032	3414
			3465	3518	3551	3715	4078
			4389	4617	4792	4806	4905
			5043	5167	5313	5720	5835
			6096	6369	6750	6817	6843
			6897	7005	7128	7287	7449
			7521				
*Alu*I	(AGCT)	54	45	273	356	468	683
			695	1166	1209	1741	1778
			1927	2060	2259	2439	2885
			3316	3372	3469	3545	3668
			3711	3811	4031	4081	4277
			4437	4560	4634	4736	4863
			4941	4992	5358	5727	5760
			5800	5820	5923	5944	5999
			6174	6189	6219	6716	6862
			6918	6945	6985	7086	7186
			7338	7722	7741	7826	

Enzyme	Site	#	Location				
*Fnu*DII	(CGCG)	57	6	53	72	78	137
			208	289	491	877	951
			996	1033	1687	1719	1974
			2455	2513	2701	2827	2859
			2892	2915	3012	3123	3222
			3290	3478	3523	3609	3786
			3968	4023	4067	4094	4244
			4265	4394	4709	4980	5065
			5186	5471	5543	5687	5695
			5981	6035	6430	6773	6880
			6926	6964	7042	7103	7169
			7436	7530			
*Mae*II	(ACTG)	74	61	226	266	326	331
			347	373	485	497	649
			687	766	838	959	1090
			1199	1259	1265	1297	1552
			1569	1788	1936	1970	1991
			2139	2269	2479	2731	2739
			3046	3062	3144	3208	3380
			3391	3534	3629	3675	3732
			3964	4136	4191	4237	4345
			4358	4404	4504	4839	4984
			5013	5117	5147	5300	5768
			5868	5892	5896	6008	6165
			6306	6344	6378	6406	6515
			6537	6901	7072	7112	7204
			7301	7627	7708	7770	
*Mbo*I	(GATC)	107	2	41	64	214	254
			269	286	292	314	406
			427	502	616	670	724
			796	868	880	1006	1030
			1321	1412	1653	1679	1707
			1716	1731	1768	1853	1911
			1977	2045	2409	2471	2537
			2790	2794	2852	2980	2991
			3030	3070	3085	3176	3520
			3529	3549	3643	3688	3717
			3768	3822	3873	3888	3930
			3947	4043	4064	4268	4294
			4310	4387	4391	4538	4544
			4571	4712	4748	4771	4790
			4808	4952	5068	5139	5165
			5264	5324	5402	5495	5546
			5777	5978	6014	6098	6199
			6233	6436	6562	6587	6757
			6776	6821	6830	6841	7007
			7060	7109	7126	7145	7172
			7369	7466	7484	7611	7641
			7666	7792			

There are no restriction sites for the following enzymes in pUH24:

*Apa*I	(GGGCCC)	*Ppu*MI	(RGGWCCY)
*Asu*II	(TTCGAA)	*Rsr*II	(CGGWCCG)
*Ava*III	(ATGCAT)	*Sac*I	(GAGCTC)
*Ban*II	(GRGCYC)	*Sac*II	(CCGCGG)
*Bsp*MII	(TCCGGA)	*Sal*I	(GTCGAC)
*Bst*EII	(GGTNACC)	*Sca*I	(AGTACT)
*Bst*XI	($CCAN_6TGG$)	*Sma*I	(CCCGGG)
*Dra*I	(TTTAAA)	*Sna*BI	(TACGTA)
*Dra*III	($CACN_3GTG$)	*Spe*I	(ACTAGT)
*Eco*RI	(GAATTC)	*Sph*I	(GCATGC)
*Mlu*I	(ACGCGT)	*Spl*I	(CGTACG)
*Nar*I	(GGCGCC)	*Ssp*I	(AATATT)
*Nde*I	(CATATG)	*Stu*I	(AGGCCT)
*Pfl*MI	($CCAN_5TGG$)	*Xba*I	(TCTAGA)

Single letter code:

B = C, G or T	M = A or C	V = A, C or G
D = A, G or T	N = A, C, G or T	W = A or T
H = A, C or T	R = A or G	Y = C or T

DNA sequences were determined by the dideoxy chain termination method of Sanger *et al.* (1980) with ^{32}P–dATP or ^{35}S–dATP (Amersham, Little Chalfont, UK) as radiolabel, with both specific and standard sequencing primers. Templates for sequencing were single–strand DNAs of M13mp8, M13mp9 (Messing and Vieira, 1982) or *in vivo* packaged pEMBL8, –9, –18 and –19 (Dente and Cortese, 1987) clones in *E. coli* host strains JM101 and PC2495. One stretch refractory to the enzymatic dideoxy method was sequenced chemically according to Maxam and Gilbert (Maxam and Gilbert, 1980). pUH24 DNA was sequenced several times in both directions. Compilation and analysis of the sequences were performed with the computer programs MICROGENIE (Queen and Korn, 1984) and the Staden programs (Staden, 1984, 1987).

Results

(a) Nucleotide sequencing

For shotgun sequencing, pUH24 DNA was partially digested with *Sau*3AI, *Taq*I and *Pst*I, and cloned into M13mp8 or M13mp9. The nucleotide sequences determined for 80 of such clones were assembled into long consensus sequences with aid of the Staden software (1987). Subsequently, remaining gaps were filled by sequencing selected DNA fragments, or by using specific synthetic sequence primers. pUH24 appeared to consists of 7835 bp and to have a G+C content of 59%. Its complete nucleotide sequence is presented in Fig. 1. Also shown are the predicted amino–acid sequences of postulated open reading frames (ORFs), the most significant sequence repeats and the location of some restriction sites. pUH24 DNA contains a unique *Bam*HI site (Van den Hondel *et al.*, 1980). This site is used as a reference point to present the nucleotide sequence and the maps of the plasmid (Figs. 1, 2 and 3). They are given in the clockwise direction according to earlier circular maps of pUH24 and its derivatives from our laboratory (Van den Hondel *et al.*, 1980; Kuhlemeier *et al.*, 1981, 1983; Kuhlemeier and van Arkel, 1987).

(b) Restriction analysis

A circular restriction map of pUH24, deduced from the sequence, is shown in Fig. 2. The nucleotide sequence showed the presence of a remarkably great number of GATC sequences (Table II). However, the number of 104 GATC sites matches the predicted frequency (98) of this tetranucleotide for pUH24 when calculated from the Markov predictor p(GAT) x p(ATC) / p(AT) (McClelland, 1985). In this approach, differences between observed and predicted tetranucleotide frequencies are more likely to be due to effects at the tetranucleotide level than to lower–order effects such as codon usage. On the other hand, the observed subset of 40 sites forming the recognition site for *Pvu*I,

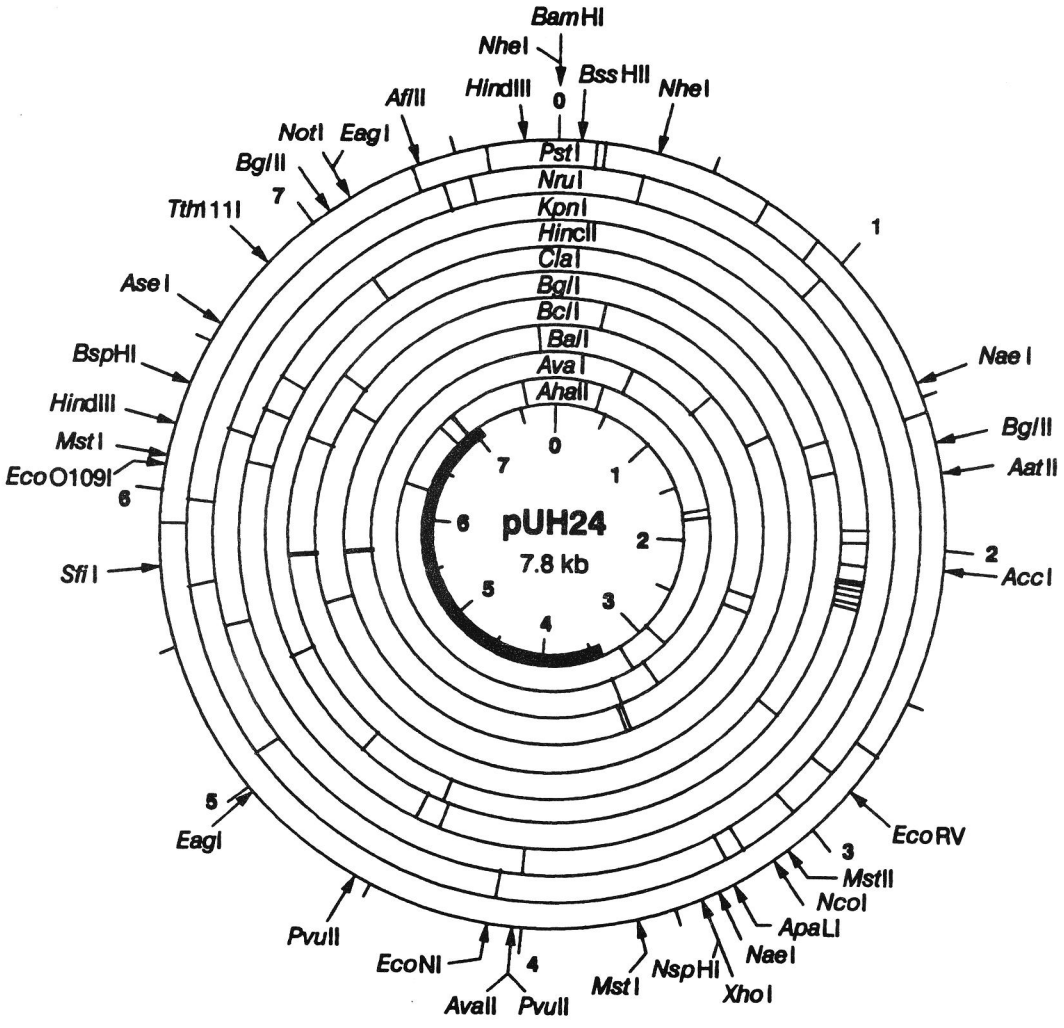


Fig. 2. Circular restriction maps of pUH24. For a number of enzymes with a hexanucleotide recognition sequence the cleavage map is given on concentric rings. Around the perimeter of these circular maps appear the sites for enzymes which cut pUH24 only once or twice. The numbers inside and outside the circles are the distances in kilobases from the unique *Bam*HI site. The precise coordinates of the restriction cuts are listed in Table II. The inner section given in bold indicates the smallest replicon fragment in use (Friedberg and Seijffers, 1986)

CGATCG (Table II), is significantly larger than the number of 14 that would be predicted from its Markov "best predictor". An explanation for this irregularity is not at hand, though the same anomaly is characteristic for chromosomal DNA of *Synechococcus* sp PCC 7942 and PCC 6301, where 137 sites are found in 46.5 kb of compiled genomic sequences whereas only 32 are predicted (Van der Plas, unpublished).

In *E. coli* the *dam* methylase methylates the adenine residue in its recognition sequence GATC, thereby preventing cleavage of these sites by the restriction endonuclease *Mbo*I (site: GATC; Barbeyron *et al.*, 1984). pUH24 DNA derived from *Synechococcus* sp. PCC 7942 could not be cleaved by *Mbo*I in spite of the presence of 104 such sites detected by sequence analysis. This absence of cleavage is an indication for *dam*-type methylation to occur in strain PCC 7942. Additional evidence that pUH24 DNA from strain PCC 7942 is methylated in the adenine of GATC comes from restriction analysis with *Cla*I (ATCGAT) and *Bcl*I (TGATCA) of pBX24 DNA extracted from strain PCC 7942 and from *E. coli* dam^+ and dam^- cells (not shown). From the nucleotide sequence 6 sites are predicted for *Cla*I (nt positions 1536, 1680, 2791, 4388, 4805, 5312; Table II) and 4 sites for *Bcl*I (268, 1411, 4083, 1067; Table II). When the DNA was derived directly from strain PCC 7942 or from Dam^+ *E. coli* the sites having overlap with the *dam* methylase recognition site, i.e. 4 out of 6 *Cla*I sites (nt positions 1680, 2791, 4388, 4805) and all 4 *Bcl*I sites, appeared resistant to cleavage, but were fully cleavable when the DNA was extracted from *E. coli* dam^-.

(c) Searching for genes in pUH24

As nothing is known about the proteins potentially encoded by pUH24, a combination of several computer gene-search methods was applied in order to reliably select those ORFs which most likely represent real protein coding sequences.

(i) Presumptive ORFs

Coding regions can be identified by the fact that coding for a protein puts constraints on a sequence. The most obvious constraint is that for a long open reading frame (ORF). This criterion can be used to delimit the regions and frames of potential genes. Almost every gene has an ORF of at least 50 codons (Stormo, 1987). A search for reading frames starting with an ATG or GTG codon and terminating after 50 or more coding triplets with any of the 3 stop codons, revealed 36 potential coding sequences (Fig. 3 part B, open and shaded boxes), 13 of which have more than 100 codons (every ORF of over 100 codons generally represents a gene; Stormo, 1987). The following cues were subsequently used to discriminate between coding and non-coding ORFs.

Proteins with different functions and from all species share a characteristic amino acid composition. This fact is the basis of a "gene search" method by Staden (1984); by looking for positional base preferences (counting the proportion of each base at each codon position for each possible reading frame), ORFs are rated for their likelihood of being translated. Applied to the nucleotide sequence of pUH24 (Fig. 3A), eight of the major ORFs (Fig. 3B) with more than 100 codons, designated Orf A – Orf H, were identified as putative coding sequences. Orf E and Orf F have the same orientation and overlap for 15

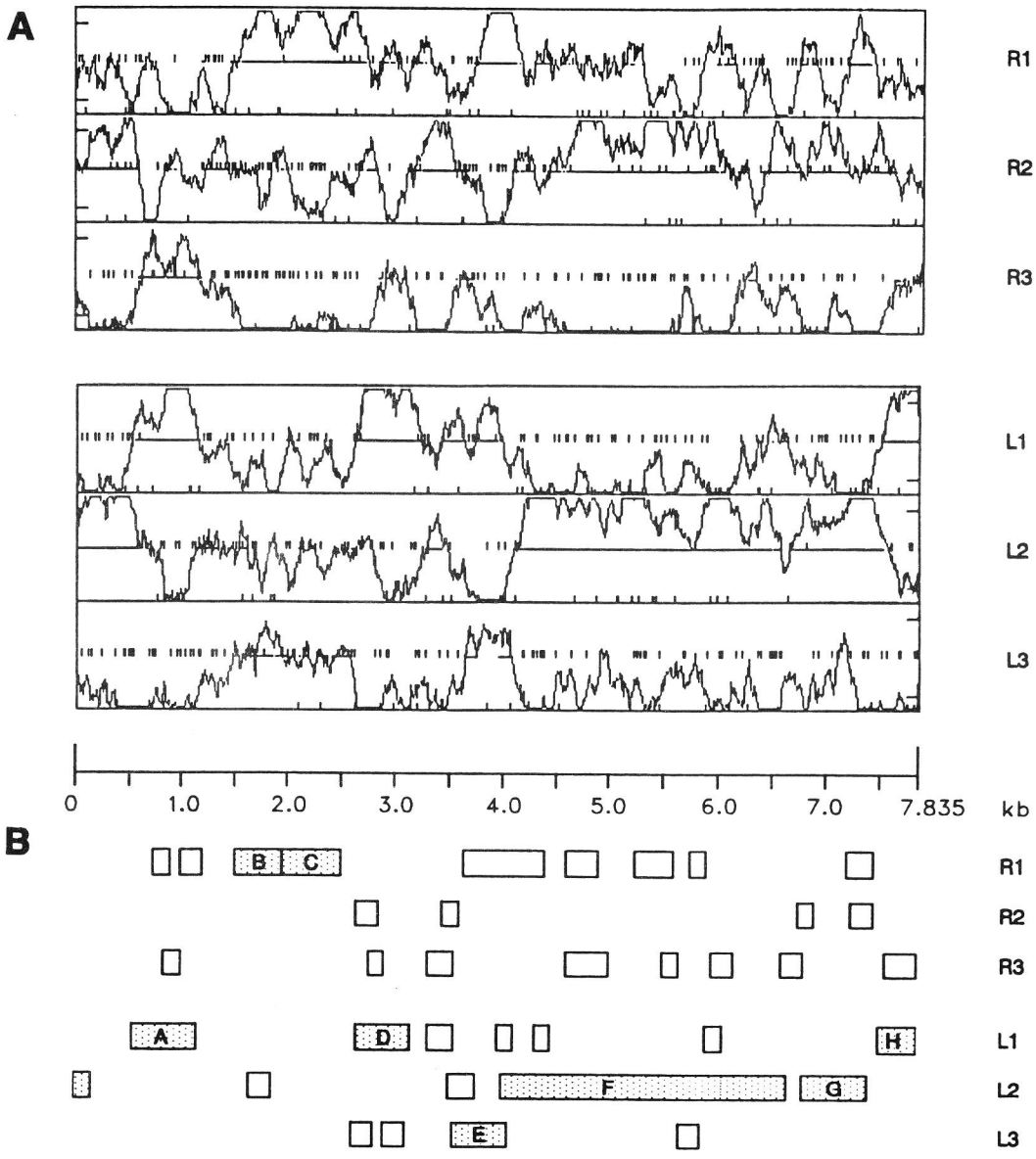


Fig. 3. Gene predictions from the pUH24 plasmid DNA sequences. **(A)** The positional base preference method of Staden (1984) was applied to pUH24 sequence in both orientations. The 3 parallel boxes represent the 3 translational reading frames of the DNA. Points above the midline are likely to be protein-coding, whereas those below are not. Vertical bars on the mid-lines indicate possible stopcodons, those on the abscissae represent possible ATG startcodons (GTG codons are not given). **(B)** All 36 potential ORFs of more than 100 codons are indicated by rectangles, and arranged according to their respective rightward (R1-R3) and leftward (L1-L3) reading frames. The rectangles labelled A - H represent the eight putative coding sequences, as predicted by the combination of three different gene search methods (Staden, 1984; Bibb *et al.*, 1984; Kolaskar and Reddy, 1985).

codons; Such overlapping genes have been found before in *Synechococcus* sp. PCC 7942 (Golden and Stearns, 1988). The alternative for Orf E in the opposite strand (Fig. 3: frame R1) is less likely because its mRNA would be largely complementary to that of Orf F, which for both ORFs could lead to inhibition of translation.

The results presented were confirmed by the outcome of the "FRAME" method, based on the relationship between overall base composition and codon usage (Bibb *et al.*, 1984). Bacterial genes that code for proteins appear to possess a codon usage characteristic of their overall base composition. Codon degeneracy is effected mainly by base variation in the third codon position and to a lesser extend in the first codon position. This results in different but predictable non-random distributions of nucleotides within codons, permitting the recognition of protein-coding sequences in a wide range of bacterial species (Bibb *et al.*, 1984). The pUH24 DNA has a G+C content of 59.4 %, which is not very far from the value of 55.0% determined for the genomic DNA of *Synechococcus* sp. PCC 7942 (Houmard and Tandeau de Marsac, 1988). ORFs of *Synechococcus* sp. PCC 7942 and closely related strains tend to have a biased codon usage such that coding sequences have a high G+C content of about 60% in the first and 65% in the third position of the codon triplet and a markedly lower percentage of about 42% G+C in the second position, as derived from the reported coding sequences for 37 genes (Van der Plas, unpublished; see Addendum). This observed position-specific difference in G+C composition was used in a "FRAME" analysis (Bibb *et al*, 1984) of the 36 ORFs with a length of 150 bp or more found in pUH24. This resulted in the same 8 ORFs (Fig. 3 and Table III), that were also found by the Staden analysis. The G+C content in the first, second and third position of the codons for each of the proposed ORFs is presented after the codon usage data in Table IV.

As an alternative approach, a combined search on content and on signal was performed as described by Kolaskar and Reddy (1985). These authors found as a characteristic of prokaryotic coding sequences the pronounced presence or absence of 13 specific triplets. These codons can be used as markers, to distinguish coding DNA sequences from non-coding sequences. This search on content is combined with a search rule for protein synthesis initiation sites. The quality of translation initiation sites appeared not to be entirely determined by the presence of a good SD sequence (Shine and Dalgarno, 1974) 4-9 bases upstream of the ATG or GTG startcodon. At least for *E. coli* it has been shown that also a larger region including the beginning of the coding sequence itself is important for the initiation process (Scherer *et al.*, 1980). Therefore the search on signal was based on a weight matrix for nucleotide positions from -18 to +18 (Kolaskar and Reddy, 1985). This combined search routine was applied for open reading frames in pUH24 larger than 60 nucleotides and resulted in the detection of 8 potential protein coding

sequences, 7 of which coincide with the assignments discussed above. Orf D was not detected by this approach. The additional ORF detected could be rejected for its aberrant G+C distribution (58%-66%-66% in codon position 1,2 and 3 respectively). The proposed arrangement of genes in pUH24 is summarized in Figure 3A.

(ii) Translational signals

Table III lists the DNA sequences immediately preceding the most likely initiation codons of the proposed ORFs. Additional data, particularly N-terminal protein sequence data, are usually required to decide unambiguously on the position of a true translation initiation site. Four of the eight predicted ORFs start with GTG. Although the vast majority of coding sequences known for cyanobacterial genes have an ATG startcodon, we consider a GTG startcodon feasible, for there is now indication for at least 5 such startcodons in cyanobacterial genes (Conley *et al.*, 1988; Golden and Stearns, 1988; Reddy *et al.*, 1988; Yasui *et al.*, 1988; Meng *et al.*, 1989). All initiation codons listed in Table III are preceded by sequences resembling ribosome-binding sites observed in *E. coli* (Shine and Dalgarno, 1974). These sequences complementary to the 3' end of 16S rRNA of *Synechococcus* sp. PCC 6301 (Tomioka and Sugiura, 1983) are thought to play a role in initiation of translation. Their presence was used as an additional criterion for identification of potential protein-coding regions.

Table III. Ribosome binding sites preceding the initiation codons of the proposed protein-coding sequences in pUH24

Orf	Position	Translational initiation signals[a]	Codons	Mol wt[b]	Gene
A	1137- 577	CTAGTCTTGCAGCAAATTGCGAGGCTAAAAATG	187	19975	
B	1485-1928	AGGCCGGCCCTTAACTAAACCCCAAGAAAAATG	147	16312	*pmaA*
C	1938-2481	GCCGATCGCTTTCTGGGGTAGCTGTTAACAGTG	181	19633	*pmaB*
D	3159-2646	TGTCGTTCTTGATCTAAGGAGTCTCTAGCCATG	171	18762	
E	4040-3569	CGCGATCGCCCGATGTTCGGACCCGATCTTGTG	157	17603	*repB*
F	6625-3997	GGAGGTCAGGAAAAATAACTCAACAGAAGTATG	876	95446	*repA*
G	7402-6802	TACAAGCCGCAGCGCTTCGGCCTAGGGCTGGTG	200	21283	
H	163-7506	AACCGCTGGCGGCTTTTGGAGGTGAGCGCCGTG	164	17852	

[a] Bases exhibiting complementarity (allowing G-U pairing) to the 3' end of the 16S rRNA (AUUCCUCCA-5') known for the closely related strain *Synechococcus* sp. PCC 6301 are underlined, as are the putative initiation codons.

[b] Protein Mol wts are calculated from the nucleotide sequence and include the N-terminal methionine, although this residue is usually absent from mature proteins.

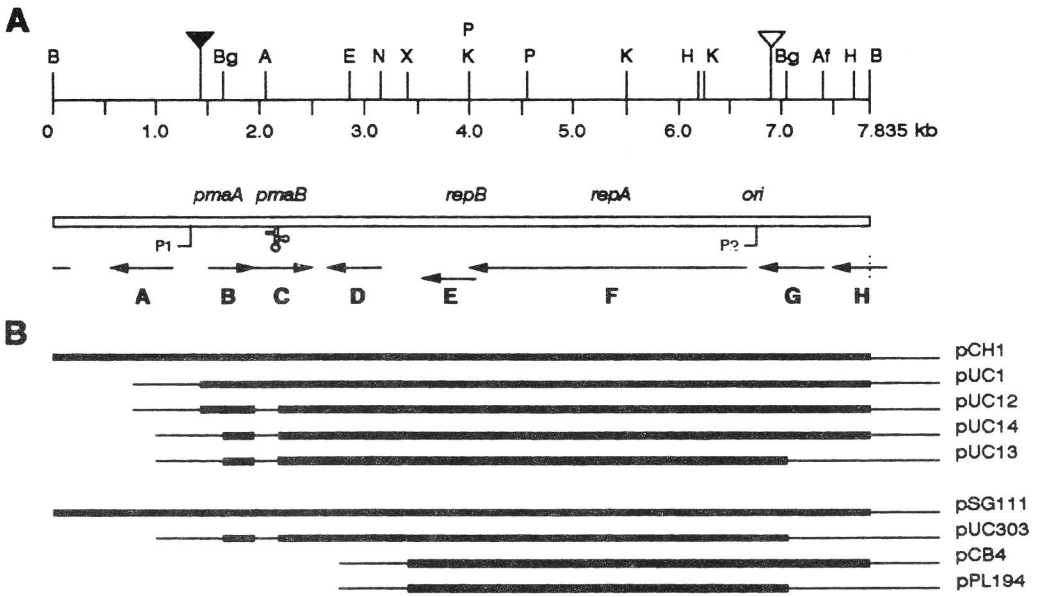


Fig. 4. Organization of the pUH24 genome. **(A)** Distances in kb, corresponding to the numbering of Fig. 1, are shown on the top line as are some convenient restriction sites; the site of insertion of Tn*901* is marked by small solid triangle for pCH1 and an open triangle for pCH4 (Van den Hondel, 1980). Letters indicate restriction sites as followes: A, *Acc*I; Af, *Af*III; B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RV; H, *Hin*dIII; K, *Kpn*I; N, *Nco*I; P, *Pvu*II; X, *Xho*I. The position of the putative genes, promoters and secondary structures discussed in this paper, are indicated on the second line. **(B)** Alignment of the cyanobacterial vectors used for the segregation analysis (see Table I) and of four functional *Synechococcus* sp. PCC 7942 – *E. coli* shuttle-vectors: pSG111 (Golden and Sherman, 1983), pUC303 (Kuhlemeier *et al.*, 1983), pCB4 (Gendel *et al.*, 1983a) and pPL194 (Friedberg and Seijffers, 1986). Heavy lines represent the pUH24 sequences present in the cloning vehicles, whereas thin lines indicate the presence of non-pUH24 derived parts (Note that the latter were not drawn to scale!).

(iii) Transcriptional signals

Some putative cyanobacterial regulatory sequences present on cloned fragments are capable of initiating transcription in *E. coli*. Moreover, regulatory sequences of some *E. coli* and other foreign genes introduced into *Synechococcus* sp. PCC 7942 can be recognized in vivo by the cyanobacterial polymerase (Tandeau de Marsac and Houmard, 1987). This suggests that at least some promoters and RNA polymerases in cyanobacteria resemble their *E. coli* counterparts. In contrast to the situation in *E. coli*, in cyanobacteria no consensus promoter sequence could be derived as yet for the RNA polymerase recognition sites upstream of the mapped transcription initiation points (Houmard and Tandeau de Marsac, 1987). Although in many instances an *E. coli* –10 box (Hawley and McClure, 1983) has been found upstream of cyanobacterial genes, a –35

region going with it is only rarely encountered. In the pUH24 sequence two *E. coli* promoter like sequences with both a good matching -10 and -35 region are present (see Fig. 1), one 130 bp upstream of Orf A (TTCAGT --21bp-- TATAGT; nt. position 1302-1270) and one 80 bp upstream of Orf F (TTGCCA --20bp--TATAGT; nt. position 6742-6711). These might be functional *Synechococcus* sp. PCC 7942 promoters. A mRNA of approximately 1200 b, specific for Orf A was detected (Van der Plas *et al.*, unpublished results), but whether this transcript is initiated from the putative promoter found upstream of Orf A is unknown, as mapping of the startsite has not been performed yet.

Examination of the sequence in the regions where termination sites would be expected revealed one substantial G+C-rich dyad symmetry region followed by a A/T-rich sequence, located between Orf A and Orf G (Fig. 1; centred at nt position 6670). It may represent a Rho-independent termination signal, since terminators of this type appear to be widely conserved in eubacteria (Higgins, 1986). This sequence, capable of forming a stable hairpin structure (ΔG=-24 kCal/mol) when transcribed into mRNA, is located between the putative promoter at position 6742-6711 and the Orf F coding sequence starting at nt position 6625. Therefore this structure could also play a role in the regulation (attenuation) of the expression from the putative Orf F promoter (Higgins, 1986).

(iv) Databank searches of the putative pUH24 protein sequences

After having determined the likely coding capacity of the various ORFs present in pUH24, we analysed each of the predicted amino acid sequences to determine any similarities to previously characterised proteins. No significant sequence similarities with the NBRF/PIR (release 19.0, 12/88), NBRF/NEW (release 37, 12/88) and Swiss-Prot (release 10.0, 3/89) databases were detected using the FASTP program (Lipman and Pearson, 1985).

(d) Replication region

The results of the sequence analysis were correlated with the composition and properties of a range of (shuttle-)vectors derived from pUH24 (Fig. 4). The smallest contiguous DNA segment from pUH24 actually shown to be capable of supporting replication in *Synechococcus* is the 3.6 kb *Xho*I-*Bgl*II restriction fragment used in the construction of pPL194 (Friedberg and Seijffers, 1986). This fragment, when present in hybrid vectors, is able to transform *Synechococcus* sp. PCC 7942 cured of its indigenous small plasmid (strain also called: R2-SPc) and thus plasmid replication occurs in the absence of other segments of pUH24. Tn*5* Insertions (obtained in *E. coli*) distributed throughout the somewhat larger 4.4 kb *Xho*I-*Bam*HI replicon fragment used in the construction of pCB4 (Gendel *et al.*, 1983a) appeared to be unstable and were lost after transformation of R2-SPc, indicating that interruption of the cyanobacterial replicon resulted in loss of the

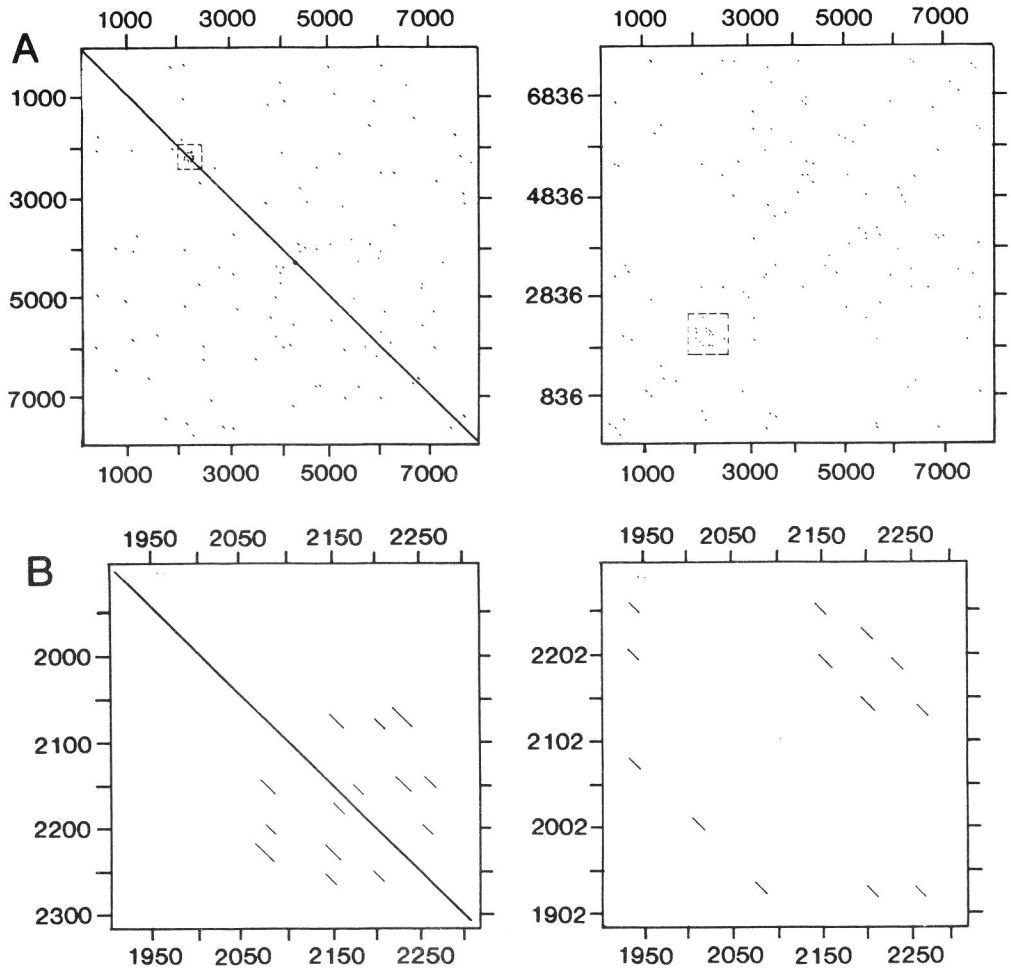


Fig. 5. Dot-matrix analysis for direct and indirect repeats. **(A)** The entire nucleotide sequence of pUH24 was compared with itself (left panel) or its complement (right panel) in order to detect direct and indirect repeats respectively. **(B)** The region 1900 to 2300, showing the highest amount in both direct and indirect repeats (left and right panel, respectively), was analysed in more detail. The results indicate the presence of a family of palindromic sequences in this particular section (see also Fig. 1). The parameters for the dot-matrix analysis shown were 95% matching with a minimum of 12 bases, for stretches at a maximal distance of 450 nucleotides and no loopouts allowed (Queen and Korn, 1984).

ability to transform R2-SPc (Gendel, 1987). Based on these observations the minimal replicon has to be confined to the contiguous 3.6 kb *Xho*I-*Bgl*II fragment, accommodating the overlapping Orf F and Orf E (Fig. 4), which may form a single operon and encode proteins necessary for replication of pUH24. The proteins encoded by the remaining ORFs (A, B, C, D, G and H) do not seem to contain essential replication functions, although they may be involved in other plasmid propagation-related processes, like maintenance or copy control.

All known prokaryotic plasmids contain a DNA locus, denoted *ori*, that is required in *cis* for initiation of plasmid replication. Two small regions of the 3.6 kb *Xho*I-*Bgl*II fragment are not occupied by open reading frames: the 250 bp sequence between the *Xho*I site (nt position 3413) and the stopcodon of ORF E (nt position 3569), and the 200 bp intergenic region separating Orf F and Orf G (nt position 6625-6802). It seems likely that one of these regions is the origin of replication, although it is not a priori impossible that the origin is in a coding sequence (ΦX174; Baas *et al.*, 1976). The best candidate for this function is the region between Orf F and Orf G as this contains a prominent inverted repeat, two adjacent 20 bp stretches of directly repeated nucleotide sequence (overlapping the dyad structure) and is rich in adenine-thymine pairs. These are features commonly found in replication origins of prokaryotic plasmids (reviewed by Scott, 1984).

(e) Plasmid stability

Most naturally occurring plasmids are rarely lost from their host cells (Sherrat, 1986). In addition to a replication control mechanism, the stable maintenance of a plasmid requires either random segregation in combination with a high copy number, or a specific system for the proper partitioning of (low copy number) plasmids at cell division (Austin *et al.*, 1981; Summers and Sherrat, 1985). In order to secure accurate partitioning, some random segregating multicopy plasmids contain functions to reverse multimerization, caused by the host site-specific recombination systems (Garnier *et al.*, 1987; Summers and Sherrat, 1988).

In order to investigate the presence and location of stability promoting functions on pUH24, segregation analyses were performed with a number of non-shuttle derivatives of pUH24 (Fig. 4). The plasmids pCH1 (Van den Hondel *et al.*, 1980), consisting of pUH24 with Tn*901* (Ap^R) inserted at approximately nt position 1450, and pUC1, a derivative of pCH1 with the transposon immobilised by deletion of a hybrid *Bam*HI fragment (Van den Hondel *et al.*, 1980), are stably maintained even after 50 generations of growth without antibiotic selection (frequency of plasmid loss < 0.05%). Spontaneous loss of plasmids, however, was found in cells harbouring pUC12, a pUC1 derivative made by exchanging a few very small *Hpa*I fragments for the Sm^R marker from RSF1010 (Kuhlemeier *et al.*, 1983).

The frequency of plasmid loss was about 0.25%, which means that approximately one in 400 cells produced a plasmid-free descendant per generation. From pUC12 two smaller plasmids, pUC13 and pUC14, were made by deleting a *Bgl*II fragment (pUC13) or a *Bgl*II-*Bam*HI fragment (pUC14), with the Ap^R marker (the remainder of Tn*901*) and some extra pUH24 DNA (Kuhlemeier *et al.*, 1983). These plasmids, although still able to replicate in *Synechococcus* sp. PCC 7942, were readily lost during prolonged growth in BG11 medium without antibiotic selection (Sm^R). Approximately 1 in 20 cells failed in one growth cycle to pass on plasmid to one of both daughter cells (frequency of plasmid loss approximately 5%). About 30 generations were required to obtain cultures in which 50% of the cells lacked these small plasmids.

When these data on segregational stability are correlated with the physical and genetic maps of the pUH24 derivatives (Fig. 4B), it follows that the severe instability caused by the conversion of pUC12 to pUC13 and pUC14 must be caused by the deletion of the first part of Orf B. Obviously this ORF encodes a protein involved in the stable maintenance of pUH24. Therefore, we called this gene *pmaA* (*pma* stands for plasmid maintenance). The instability of pUC12 can either be explained by insertion of the Sm^R marker into the sequence of Orf C, or by the concomitant removal of some of the small fragments created by a cluster of 9 closely spaced *Hpa*I sites (Fig. 1 and Table II: nt position 1931-2253). These *Hpa*I sites (-GTTAAC-) form the core of 12 larger (10bp) palindromic sequences (-CTGTTAACAG-) that were detected by dot-matrix analysis of the pUH24 sequence for direct and inverted repeats (Fig. 5.). Together these *Hpa*I motifs are capable of forming even larger secondary structures. Therefore, the role in plasmid maintenance of this locus (named *pmaB*) could either be accomplished by the presumed protein encoded by Orf C and/or by these special sequence elements coinciding with Orf C. The latter elements might function e.g. as a resolution site for site-specific monomerization or as protein binding sites for the titration of regulatory proteins; for both mechanisms there are examples in literature (Scott, 1984; Summers and Sherrat, 1988). It should be noted that during the construction of pUC12 the deletion did remove most but not all of the *Hpa*I-motifs (Kuhlemeier *et al.*, 1983). Complete removal might have a more severe effect on plasmid stability.

Discussion

Sequence features

The complete nucleotide sequence of the *Synechococcus* sp. PCC 7942 plasmid pUH24 was determined and analysed for its composition, open reading frames and expression signals. The results obtained were correlated with segregational analyses and previously

published data on the construction of pUH24 derived vectors. The circular plasmid is 7835 nucleotides in length and has a G+C composition of 59%, which is close to the 55% G+C reported for *Synechococcus* sp. PCC 7942 DNA (Houmard and Tandeau de Marsac, 1987).

In addition to the similarity in G+C content, the excess of *PvuI* sites found both in the pUH24 sequence and in genomic sequences of *Synechococcus* sp. PCC 7942 indicates that both DNAs are closely related. This conclusion is relevant in the context of the observation that a plasmid with the same size and restriction patterns as pUH24, was found not only in closely related *Synechococcus* strains of different geographical origin (Van den Hondel *et al.*, 1979), but also in *Cyanobium* sp. PCC 6707 which possesses a 13% higher chromosomal G+C content (Tandeau de Marsac and Houmard, 1987). This suggests interspecific, or even intergeneric, transmission of pUH24 might occur in nature.

pUH24 DNA from *Synechococcus* sp. PCC 7942 is extensively modified in the sequence GATC. This probably is the result of the action of an equivalent of the *Dam* methylase of *Escherichia coli*. Many strains of *Escherichia coli* contain a site-specific methylase that transfers a methyl group from S-adenosylmethionine to the N^6 position of the adenine residues in the sequence GATC (*dam* methylase). This modification may well be important in DNA repair-synthesis and signal which of the two strands is to be used as template (Modrich, 1989). In addition other biological functions are possible, e.g. in gene expression, replication, or chromosomal segregation (Messer and Noyer-Weidner, 1988). *Dam*-like methylation has also been reported for 5 other cyanobacterial strains (Barbeyron *et al.*, 1984).

The DNA sequence fully characterizes pUH24 in terms of its restriction sites and genetic content, such that the exact length and coding potential of every fragment is known. This knowledge should aid in designing cloning strategies and should facilitate the construction of more effective cyanobacterial vectors.

ORFs

Eight putative ORFs were identified in the pUH24 DNA sequence. Three of them, possibly four when Orf C is included, appear to be implicated in specific plasmid maintenance (replication and stability) related functions, whereas for the other ORFs no function could be assigned yet. The mere fact that functional proteins generally are long, and therefore have long ORFs, is fairly indicative of coding regions. Coding regions tend to show a "codon bias" which is not found in non-coding sequences, and through the years several algorithms exploiting this phenomenon have been designed (e.g. Bibb *et al.*, 1984; Staden, 1984; Kolaskar and Reddy, 1985). When combined, the analyses based on length, base/position preferences and codon usage, supplemented with a search on translational

Table IV. Codon usage and positional base preferences of the proposed genes on pUH24

aa	codon	total	A	B	C	D	E	F	G	H
Arg (R)	CGA	21	5	3	3	2	1	5	—	2
	CGC	59	6	4	2	5	3	25	5	9
	CGG	47	1	2	5	8	5	19	2	5
	CGT	23	2	1	2	4	1	7	2	4
	AGA	10	1	—	3	—	3	1	1	1
	AGG	7	1	3	1	1	—	1	—	—
Leu (L)	CTA	25	3	4	2	2	—	8	5	1
	CTC	27	5	1	2	2	2	10	3	2
	CTG	95	8	8	8	6	8	38	13	6
	CTT	25	3	3	2	—	2	12	1	2
	TTA	13	—	—	4	—	—	8	1	—
	TTG	31	2	2	3	2	1	14	6	1
Ser (S)	TCA	17	1	3	1	2	2	7	1	—
	TCC	17	2	1	1	1	1	8	2	1
	TCG	31	3	4	2	3	—	10	4	5
	TCT	18	2	2	5	2	2	3	—	2
	AGC	61	3	4	10	4	6	22	7	5
	AGT	18	6	1	1	1	2	4	2	1
Thr (T)	ACA	15	1	—	4	—	2	7	1	—
	ACC	44	6	4	2	1	6	19	—	6
	ACG	29	2	4	1	1	2	11	7	1
	ACT	19	2	1	1	3	3	5	2	2
Pro (P)	CCA	22	3	1	2	2	1	12	—	1
	CCC	24	1	2	1	1	1	14	3	1
	CCG	42	2	—	2	5	3	22	4	4
	CCT	30	4	1	7	2	1	7	6	2
Ala (A)	GCA	65	7	7	3	5	4	22	9	8
	GCC	76	4	5	8	7	4	36	5	7
	GCG	75	7	2	6	9	7	28	8	8
	GCT	66	2	3	6	9	3	29	4	10
Gly (G)	GGA	15	1	—	2	—	1	8	3	—
	GGC	72	9	3	5	5	3	31	12	4
	GGG	28	5	3	2	2	2	14	—	—
	GGT	26	5	1	—	2	1	13	2	2
Val (V)	GTA	11	2	—	3	1	—	4	1	—
	GTC	43	2	3	2	4	3	22	3	4
	GTG	39	5	5	3	3	4	12	3	4
	GTT	34	3	1	6	5	1	14	4	—
Ile (I)	ATA	1	—	—	—	1	—	—	—	—
	ATC	94	11	7	2	6	11	40	9	8
	ATT	17	1	3	—	1	3	6	—	3
Lys (K)	AAA	7	—	1	1	1	1	3	—	—
	AAG	46	2	2	5	4	7	24	1	1
Asn (N)	AAC	41	1	4	7	2	1	16	6	4
	AAT	11	1	—	2	1	2	5	—	—
Gln (Q)	CAA	43	3	3	3	1	4	21	3	5
	CAG	57	3	2	5	5	5	25	5	7

Table IV. ***continued***

aa	codon	total	A	B	C	D	E	F	G	H
His (H)	CAC	24	3	2	1	2	2	11	3	—
	CAT	15	—	—	2	1	3	8	—	1
Glu (E)	GAA	54	7	5	3	4	5	26	2	2
	GAG	79	7	3	5	13	5	25	14	7
Asp (D)	GAC	53	7	4	5	2	3	28	1	3
	GAT	57	2	7	6	6	2	20	9	5
Tyr (Y)	TAC	17	1	2	—	1	2	7	4	—
	TAT	11	—	—	1	—	—	9	—	1
Cys (C)	TGC	20	1	—	2	—	—	14	3	—
	TGT	5	—	—	—	1	2	2	—	—
Phe (F)	TTC	28	1	3	3	—	2	14	5	—
	TTT	14	2	2	—	2	1	4	1	2
Met (M)	ATG	29	5	4	1	1	1	16	—	1
	GTG	4			1		1		1	1
Trp (W)	TGG	36	2	1	3	4	3	20	1	2
End (U)	TAA	2	1	—	—	—	—	—	1	—
	TAG	3	—	1	—	1	—	—	—	1
	TGA	3	—	—	1	—	1	1	—	—
Total		2091	188	148	182	172	158	877	201	165
G + C content at the three positions in the codons used:										
1st position		66%	68%	60%	63%	73%	58%	66%	68%	71%
2nd position		50%	52%	45%	52%	53%	48%	49%	48%	56%
3rd position		66%	63%	64%	58%	65%	66%	68%	70%	65%

initiation signals, are thought to be accurate (Stormo, 1987). As a further check for the validity of the selected reading frames, each ORF was examined for the unbalanced presence of codons known to be used rarely in *Synechococcus* species. No unusual codon usage could be seen for any of the presumptive ORFs (compare Table IV and Addendum). Unambiguous identification of the ORFs proposed for pUH24 as factual protein coding sequences, however, should ultimately come from protein sequencing or from expression studies. Examination of the polypeptides encoded on and expressed by the shuttle vector pDF30 in *E. coli* minicells revealed no gene products from the pUH24 related pDF3 (Friedberg and Seijffers, 1983). Efforts made to express pUH24 *in vitro* with *E. coli* cell-free extracts capable of coupled transcription and translation, resulted only in a very weak (fourfold) stimulation of protein synthesis (Gruber *et al.*, 1987). The 30 kDa protein poorly produced by the cell-free system programmed with pUH24, could not be detected with hybrid vectors consisting of pUH24 fused to an *E. coli* replicon either at the *Bam*HI site or at the *Xho*I site. Therefore, the location of the sequence encoding this 30 kDa protein remained obscure (Gruber *et al.*, 1987). According to the sequence of pUH24

(Fig. 1) a protein of this size could either be the product the major Orf in Frame R1 (Fig. 3; 239 amino acids and not expected to be coding in *Synechococcus* sp. PCC 7942), or represent an incomplete product of Orf F. Most likely, the *E. coli* cell-free system is not very well suited for the job and should the cyanobacterial transcription and/or translation machinery be used in order to acquire significant data on protein expression from pUH24.

Stability

Segregation analysis was performed with a series of non-shuttle derivatives of pUH24. In this way the possibility of inadvertent complementation was avoided that might occur in shuttle vectors by *E. coli* replicon encoded functions. A non-essential region of pUH24 was identified containing two functions involved in the segregational stability of the plasmid in a population of dividing cells. The DNA sequence of the *pmaA* locus codes for a protein of 147 amino acids. Deletion of the first 56 triplets from the coding sequence resulted in severe plasmid-instability, as in each generation of unselected growth one in 20 cells failed to pass on plasmid to its progeny. The *pmaB* locus, located adjacent to the 5' site of *pmaA*, potentially encodes for a protein, but also contains twelve 10bp long palindromic elements with the potential of forming extensive secondary structures. Deletion of eight of these palindromic elements from the *pmaB* locus resulted in loss of plasmid once in 400 cell divisions. This *pmaB* locus could function as a target for site-specific resolution of multimers (Austin *et al.*, 1981; Summers and Sherratt, 1984, 1988). In the absence of the complete *pmaB* function a reduced number of independent plasmid molecules would then be available for partitioning, thus resulting in segregational instability. A functional protein coding sequence can not be excluded however, as an open reading frame occupying the same region has also been identified. Either *pmaA* or *pmaB* could encode the presumed site-specific resolvase acting on the secondary structures just described (Garnier *et al.*, 1987), or they could represent proteins associated with copy number control (Scott, 1984; Summers and Sherratt, 1985). Another conceivable function of *pmaA* and/or *pmaB* is that they are genes for proteins actively involved in partitioning, e.g. by specific membrane association or by specific association to chromosomal nucleoids (Sherratt, 1986).

It could be argued that in practice useful vectors derived from pUH24 can do without the *pma* loci, since cells carrying such plasmids usually are grown under antibiotic pressure. However, when plasmids are unstable by defects in their copy control they can severely decrease the viability of their host cells. Copy mutants can only be isolated if there is at least some kind of regulation left, because uncontrolled plasmid replication is likely to be lethal to the host (Uhlin and Nordstrom, 1978). Besides, after initial growth in

liquid or on solid media, cells may be subjected to lowered antibiotic pressure because of turnover/shielding effects by the high density of cells. Under these conditions plasmid-less cells might be able to survive and eventually thrive better than plasmid-containing cells because of the absence of the metabolic burden of plasmid replication and expression. Therefore it seems advisable to consider the stability of cyanobacterial (shuttle) vectors before applying them e.g. for gene or promoter analysis.

Replication

Little is known about plasmid replication in cyanobacteria, not even for the extensively used plasmid pUH24. A transposon-tagged derivative of pUH24 is unable to replicate in *E. coli* and none of the common cloning vectors from *E. coli* is maintained in *Synechococcus* sp. PCC 7942 (Kuhlemeier and Van Arkel, 1987). Actually, no heterogeneric plasmid replication has been established in cyanobacteria, except in one case, as vectors based on plasmid pDU1 from *Nostoc* sp. strain 7524 are also found to replicate in several *Anabaena* species (Thiel and Wolk, 1987). These data suggest that the mechanisms which control plasmid replication differ significantly between cyanobacteria and *E. coli*, and also among cyanobacteria.

The minimal region of pUH24 fully capable of supporting its autonomous replication consists of a 3.6 kb DNA fragment, considerably larger than the 1.3 kb of plasmid pDU1 for *Anabaena* sp. strain M-131 (Schmetterer and Wolk, 1988). The 3.6 kb *Synechococcus* replicon is almost entirely occupied by two overlapping ORFs. They most likely represent two genes that are essential for replication (*repA* and *repB*). The pUH24 origin of replication is presumably found just upstream of *repA* in a 200 bp intergenic region on the mini replicon, as the sequence found there possesses several features commonly found in prokaryotic plasmid origins (Scott, 1984).

Since the complete nucleotide sequence of the pUH24 replicon is now available, investigation can be directed further at the functional characterization of the detected ORFs and at the mechanisms of cyanobacterial plasmid replication and maintenance.

Acknowledgements

The authors thank Dr S Bron (University of Groningen) and Prof GA van Arkel for their valuable comments and suggestions for improvement of the manuscript, Drs AJ Andriesse, MA Huynen and SA Langeveld for their assistance in the computer analyses. Finally the many students are gratefully acknowledged, who contributed to this sequencing project as part of their training in molecular genetics.

References

Austin S, Ziese M, Sternberg N (1981) A novel role for site-specific recombination in maintenance of bacterial replicons. Cell 25:729-736

Barbeyron T, Kean K and Forterre P (1984) DNA adenine methylation of GATC sequences appeared recently in the *Escherichia coli* lineage. J Bacteriol 160:586-590

Baas PD, Jansz HS and Sinsheimer RL (1976) Bacteriophage ΦX174 DNA synthesis in a replication deficient host: determination of the origin of ΦX DNA replication. J Mol Biol 102:633-656

Bibb MJ, Findlay PR and Johnson MW (1984) The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene 30:157-166

Castets AM, Houmard J and Tandeau de Marsac N (1986) Is cell motility a plasmid-encoded function in the cyanobacterium *Synechocystis* 6803. FEMS Microbiol Lett 37:277-281

Conley PB, Lemaux PG and Grossman A (1988) Molecular characterization and evolution of sequences encoding light-harvesting components in the chromatically adapting cyanobacterium *Fremyella diplosiphon*. J Mol Biol 199:447-465

Dente L and Cortese R (1987) pEMBL: a new family of single stranded plasmids for sequencing DNA. Methods Enzymol 152:111-119

Ehrlich SD (1989) Illegitimate recombination in bacteria. In: Berg D and Howe M, (eds) Mobile DNA, pp 799-832. Washington: American Society for Microbiology

Friedberg D and Seijffers J (1983) A new hybrid plasmid capable of transforming *Escherichia coli* and *Anacystis nidulans*. Gene 22:267-275

Friedberg D and Seijffers J (1986) Controlled gene expression utilising Lambda phage regulatory signals in a cyanobacterium host. Mol Gen Genet 203:505-510

Gallagher ML and Burke Jr WF (1987) Construction of hybrid plasmid vectors for cloning in *Escherichia coli*, *Bacillus subtilis*, and the cyanobacterium *Anacystis nidulans*. Curr Microbiol 16:43-47

Garnier T, Saurin W and Cole ST (1988) Molecular characterization of the resolvase gene, *res*, carried by a multicopy plasmid from *Clostridium perfringens*: common evolutionary origin for prokaryotic site-specific recombinases. Mol Microbiol (1987) 1:371-376

Gendel SM (1987) Instability of Tn*5* inserts in cyanobacterial cloning vectors. J Bacteriol 169:4426-4430

Gendel S, Straus N, Pulleyblank D and Williams J (1983a) Shuttle cloning vectors for the cyanobacterium *Anacystis nidulans*. J Bacteriol 156:148-154

Gendel S, Straus N, Pulleyblank D and Williams J (1983b) A novel shuttle cloning vector for the cyanobacterium *Anacystis nidulans*. FEMS Microbiol Lett 19:291-294

Golden SS, Sherman LA (1983) A hybrid plasmid is a stable cloning vector for the cyanobacterium *Anacystis nidulans* R2. J Bacteriol 155:966-972

Golden SS and Stearns GW (1988) Nucleotide sequence and transcript analysis of three photosystem II genes from the cyanobacterium *Synechococcus* sp. PCC 7942. Gene 67:85-96

Gruber MY, Glick BR, Thompson JE and Gendel SM (1987) In vitro expression of a cyanobacterial plasmid. Curr Microbiol 15:265-268

Hauman JH (1981) Is a plasmid(s) involved in the toxicity of *Microcystis aeruginosa*? In Carmichael WW (ed) The water environment (Algal Toxins and Health, vol 20), pp 97-102. New York: Plenum Press

Hawley DK and McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucl Acids Res 11:2237-2255

Higgins CF (1986) The regulation of gene expression in *Escherichia coli* and its bacteriophage. In: Rigby PWJ (ed) Genetic Engineering vol 5, pp 2-59. London: Academic Press

Houmard J and Tandeau de Marsac N (1988) Cyanobacterial genetic tools: current status. Methods Enzymol 167:808-847

Kolaskar AS and Reddy BVB (1985) A method to locate protein coding sequences in DNA of prokaryotic systems. Nucleic Acids Res 13:185-194

Kuhlemeier CJ, Borrias WE, Van den Hondel CAMJJ and Van Arkel GA (1981) Construction and characterization of two hybrid plasmids capable of transformation to *Anacystis nidulans* R2 and *Escherichia coli* K12. Mol Gen Genet 184:249–254

Kuhlemeier CJ, Thomas AAM, Van der Ende A, Van Leen RW, Borrias WE, Van den Hondel CAMJJ and Van Arkel GA (1983) A host–vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. Plasmid 10:156–163

Kuhlemeier CJ, Hardon EM, Van Arkel GA and Van de Vate C (1985) Self–cloning in the cyanobacterium *Anacystis nidulans* R2: fate of a cloned gene after reintroduction. Plasmid 14:200–208

Kuhlemeier CJ and van Arkel GA (1987) Host–vector systems for gene cloning in cyanobacteria. Methods Enzymol 153:199–215

Lau RH and Straus NA (1985) Versatile shuttle cloning vectors for the universal cyanobacterium *Anacystis nidulans* R2. FEMS Microbiol Lett 27:253–256

Laudenbach DE, Straus NA, Gendel S and Williams JP (1983) The large endogenous plasmid of *Anacystis nidulans*: mapping, cloning and localization of the origin of replication. Mol Gen Genet 192:402–407

Laudenbach DE, Straus NA and Williams JP (1985) Evidence for two distinct origins of replication in the large endogenous plasmid of *Anacystis nidulans* R2. Mol Gen Genet 199:300–305

Lipman DJ and Pearson WR (1985) Rapid and sensitive protein similarity searches. Science 227:1435–1441

Maniatis T, Fritsch EF and Sambrook J (1982) Molecular Cloning: A laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory

Maxam AM and Gilbert W (1980) Sequencing end–labelled DNA with base–specific chemical cleavages. Methods Enzymol 65:400–560

Meacock PA and Cohen SN (1980) Partitioning of bacterial plasmids during cell division: a cis–acting locus that accomplishes stable plasmid inheritance. Cell 20:529–542

Meng BY, Shinozaki K and Sugiura M (1989) Genes for the ribosomal proteins S12 and S7 and elongation factors EF–G and EF–Tu of the cyanobacterium, *Anacystis nidulans*: structural homology between 16S rRNA and S7 mRNA. Mol Gen Genet 216:25–30

Messing, J and Vieira J (1982) A new pair of M13 vectors for selecting either strand of a double–digest restriction fragment. Gene 19:269–276.

Porter RD (1985) Transformation in cyanobacteria. Crit Rev Microbiol 13:111–132

Primrose SB and Ehrlich SD (1981) Isolation of plasmid deletion mutants and a study of their instability. Plasmid 6:193–201

Queen C and Korn LJ (1984) A comprehensive sequence analysis program for the IBM personal computer. Nucl Acids Res 12:581–599

Reaston J, Van den Hondel CAMJJ, Van der Ende A, Van Arkel GA, Stewart WDP and Herdman M (1980) Comparison of plasmids from the cyanobacterium *Nostoc* PCC 7524 with two mutant strains unable to form heterocysts. FEMS Microbiol Lett 9:185–188

Reddy KJ, Bullerjahn GS, Sherman DM and Sherman LA (1988) Cloning, nucleotide sequence, and mutagenesis of a gene (*irp*A) involved in iron–deficient growth of the cyanobacterium *Synechococcus* sp. Strain PCC 7942. J Bacteriol 170:4466–4476

Sanger F, Coulson AR, Barrell BG, Smith AJH and Roe BA (1980). Cloning in single–stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143:161–178

Scott JR (1984) Regulation of plasmid replication. Microbiol Rev 48:1–23.

Sherman LA and van der Putte P (1981) Construction of a hybrid plasmid capable of replication in the cyanobacterium *Anacystis nidulans*. J Bacteriol 150:410–413

Sherrat D (1986) Control of plasmid maintenance. Symposia of the Society for General Microbiology 39:239–250

Shestakov SV, Reaston J (1987) Gene–transfer and host–vector systems of cyanobacteria. Oxford Surveys of Plant Molec Cell Biol 4:137–166

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

Smetterer G and Wolk CP (1988) Identification of the region of cyanobacterial plasmid pDU1 necessary for replication in *Anabaena* sp. strain M–131. Gene 62:101–109

Staden R (1984) Measurement of the effects that coding for a protein has on a DNA sequence and their use for finding genes. Nucl Acids Res 12:551–567
Staden R (1987) Computer handling of DNA sequencing projects. In: Bishop MJ and Rawlings (eds) Nucleic acid and protein sequence analysis: a practical approach, pp 173–217. Oxford: IRL Press
Stormo GD (1987) Identifying coding sequences. In: Bishop MJ and Rawlings (eds) Nucleic acid and protein sequence analysis: a practical approach, pp 231–258. Oxford: IRL Press
Summers DK and Sherrat DJ (1984) Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. Cell 36:1097–1103
Summers DK and Sherrat DJ (1985) Bacterial plasmid stability. BioEssays 2:209–211
Summers DK and Sherrat DJ (1988) Resolution of ColEI dimers requires a DNA sequence implicated in the three dimensional organization of the cer site. EMBO J 7(3):851–858
Tandeau de Marsac N, Borrias WE, Kuhlemeier CJ, Castets AM, Van Arkel GA and Van den Hondel CAMJJ (1082) A new approach for molecular cloning in Cyanobacteria: cloning of an *Anacystis nidulans met* gene using a Tn*901*–induced mutant. Gene 20:111–119
Tandeau de Marsac N and Houmard J (1987) Advances in cyanobacterial molecular genetics. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 251–302, Amsterdam: Elsevier
Thiel T and Wolk CP (1987) Conjugal transfer of plasmids to cyanobacteria. Methods Enzymol 153:232–243
Tomioka N and Sugiura M (1983) The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue–green alga, *Anacystis nidulans*. Mol Gen Genet 191:46–50
Uhlin BE and Nordstrom K (1978) A runaway–replication mutant of plasmid R1drd–19: temperature–dependent loss of copy number control. Mol Gen Genet 165:167–179
Van den Hondel CAMJJ, Keegstra W, Borrias WE and Van Arkel GA (1979) Homology of plasmids in strains of unicellular cyanobacteria. Plasmid 2, 323–333
Van den Hondel CAMJJ, Verbeek S, Van der Ende A, Weisbeek P, Borrias WE and Van Arkel GA (1980) Introduction of transposon Tn901 into a plasmid of *Anacystis nidulans*: preparation for cloning in cyanobacteria. Proc Natl Acad Sci USA 77:1570–1574
Van der Ende A, Teertstra R and Weisbeek PJ (1983) n' Protein activator sites of plasmid pBR322 are not essential for its DNA replication. J Mol Biol 167:751–756
Van der Plas J, De Groot R, Woortman M, Cremers F, Borrias M, Van Arkel G and Weisbeek P (1988) Genes encoding ferredoxins from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942: structure and regulation. Photosynth Res 18:179–204
Van der Plas J, Bovy A, Kruyt F, De Vrieze G, Dassen E, Klein B and Weisbeek P (1989) The gene for the precursor of plastocyanin from the cyanobacterium *Anabaena* sp. PCC 7937: isolation, sequence and regulation. Mol Microbiol 3:275–284
Walsby AE (1977) Absence of gas vesicle protein in a mutant of *Anabaena flos–aqua*. Arch Microbiol 114:167–170
Yanish–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
Yasui A, Takao, M, Oikawa A, Kiener A, Walsh CT and Eker APM (1988) Cloning and characterization of a photolyase gene from the cyanobacterium *Anacystis nidulans*. Nucl Acids Res 16:4447–4463

Codon usage in *Synechococcus* sp. PCC 7942 and closely related strains

The codon usage in 37 genes from *Synechococcus* sp. PCC 7942, *Synechococcus* sp. PCC 6301 and *Anacystis nidulans* 1402-1/SAUG has been determined (Table I and II). Such data are useful in several instances, like in the identification of potential coding sequences (codon bias) and in the construction of synthetic oligonucleotide probes based on amino acid sequence data (Staden, 1984; Bibb *et al.*, 1984; Bryant and Tandeau de Marsac, 1988). The G+C% for each of the three positions of the codon triplet has been calculated from the codon frequencies (Table I). Generally, a markedly lower G+C% has been found in the second position (overall: 60%-42%-65% G+C). This bias is used in Chapter V for the identification of protein-coding sequences by the "Frame" method (Bibb *et al.*, 1984).

Table I. Codon Usage and positional base preferences in genes from *Synechococcus* sp. PCC 7942 and closely related strains

aa	codon	gene																			
		apcA	*apcB*	*apcC*	*atpA*	*atpB*	*atpC*	*atpD*	*atpE*	*atpF*	*atpG*	*atpH*	*atpI*	*cpbA*	*cpcA*	*cpcB*	*fus*	*Gene1*	*irpA*	*petF*	*petF1*
Arg (R)	CGA	—	—	1	—	—	1	2	2	1	1	—	—	1	—	—	1	2	1	1	—
	CGC	3	6	2	10	8	11	7	4	8	3	1	4	10	1	7	11	2	10	—	—
	CGG	2	—	2	4	7	8	2	2	1	2	—	1	4	1	2	12	2	3	1	—
	CGT	7	4	1	10	8	8	1	4	4	3	1	3	3	4	4	11	3	5	—	1
	AGA	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	2	—	1	—	—
	AGG	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—
Leu (L)	CTA	3	—	—	—	2	2	2	1	—	1	—	3	4	—	—	2	1	1	—	—
	CTC	2	5	2	4	4	7	4	3	7	6	3	12	14	2	2	12	4	9	3	2
	CTG	6	7	3	29	19	19	8	3	9	4	6	15	10	6	7	21	7	10	3	2
	CTT	2	—	—	1	3	1	2	—	1	2	2	4	—	1	—	4	3	4	—	—
	TTA	—	—	—	—	—	—	1	1	—	1	—	1	1	—	—	2	3	2	2	—
	TTG	—	5	2	14	18	16	10	7	10	7	2	11	13	7	5	9	3	8	1	3
Ser (S)	TCA	—	2	—	1	1	—	2	1	—	—	—	2	—	1	1	5	1	2	2	—
	TCC	2	1	—	3	2	3	1	—	1	—	2	6	5	2	2	7	—	—	1	3
	TCG	5	3	—	11	18	4	1	—	1	—	—	3	5	2	3	11	—	2	1	2
	TCT	—	1	—	5	2	1	4	—	4	1	2	2	2	2	1	5	—	4	1	1
	AGC	2	5	2	11	2	7	4	1	5	1	1	3	8	7	4	8	2	11	1	—
	AGT	1	3	—	—	1	4	3	3	—	—	—	2	7	1	—	1	1	3	1	—
Thr (T)	ACA	—	—	—	1	2	1	—	1	—	2	—	—	3	—	—	2	1	2	1	—
	ACC	7	9	4	16	21	7	3	3	2	—	3	10	10	5	6	20	2	11	4	8
	ACG	1	—	1	5	5	6	1	—	3	1	—	1	3	3	2	14	1	5	1	3
	ACT	1	1	2	2	6	2	3	3	2	1	—	2	8	2	1	4	2	11	2	—
Pro (P)	CCA	—	—	1	—	1	2	1	3	—	1	—	1	5	—	—	3	1	5	1	—
	CCC	2	2	2	6	5	4	2	1	—	1	2	4	8	1	1	9	—	2	2	—
	CCG	1	1	1	11	13	4	1	—	—	—	1	5	10	3	—	12	1	5	2	1
	CCT	2	—	—	3	5	1	—	1	—	—	—	6	6	2	2	8	—	4	1	1
Ala (A)	GCA	2	5	—	7	5	4	4	8	8	3	2	4	13	6	10	6	5	12	—	3
	GCC	4	5	—	17	11	18	9	7	9	9	—	7	12	2	4	20	2	14	5	3
	GCG	2	1	1	18	13	6	4	3	6	5	6	9	8	9	6	9	4	9	3	—
	GCT	9	10	2	19	9	12	5	3	9	15	10	3	15	9	11	24	4	13	4	6
Gly (G)	GGA	1	—	1	—	—	—	1	6	—	1	—	2	4	—	—	1	1	4	1	—
	GGC	9	7	2	19	27	7	7	5	1	3	4	11	22	4	9	24	—	11	3	4
	GGG	—	—	—	—	4	—	1	—	1	1	1	1	4	1	—	5	4	6	2	—
	GGT	6	5	1	23	16	9	1	—	4	1	5	3	12	8	4	28	3	3	3	2
Val (V)	GTA	2	—	—	2	1	—	1	5	1	—	—	1	1	—	—	4	—	—	1	—
	GTC	8	8	—	17	15	5	5	6	2	5	2	4	6	2	3	26	4	11	3	6
	GTG	2	3	—	11	20	11	7	—	1	1	3	9	5	2	7	18	3	6	2	2
	GTT	3	—	2	12	9	9	4	3	5	4	1	4	6	3	3	23	3	10	2	1
Ile (I)	ATA	—	—	—	—	—	—	—	2	—	—	—	—	—	1	—	—	—	—	—	—
	ATC	7	10	5	27	28	6	5	3	6	1	3	10	18	4	5	32	2	8	2	3
	ATT	5	1	—	18	4	6	4	—	3	5	2	6	9	3	6	16	4	5	1	2
Lys (K)	AAA	3	4	1	18	14	2	5	3	4	2	1	3	17	2	3	24	3	2	—	2
	AAG	3	2	4	7	11	8	2	2	1	5	—	3	10	6	2	16	—	3	3	1
Asn (N)	AAC	2	4	1	13	12	10	3	—	3	1	1	6	14	7	7	15	2	4	—	1
	AAT	1	1	2	1	3	3	3	3	2	1	—	3	8	3	1	3	—	5	—	1
Gln (Q)	CAA	3	5	2	12	15	11	8	5	8	11	2	1	17	3	4	14	2	21	5	4
	CAG	1	—	3	20	8	12	3	2	6	12	1	6	8	—	—	10	3	18	5	—

Table I. ***continued***

aa	codon	gene																	sum	freq
		petG	*phr*	*ppc*	*psbA1*	*psbA2*	*psbA3*	*psbC1*	*psbC2*	*psbD1*	*psbD2*	*rbcL*	*rbcS*	*rps7*	*rps12*	*trxM*	*tufA*	*wcaA*		
Arg (R)	CGA	1	1	16	—	—	—	—	2	2	2	—	—	1	—	—	—	—	39	7%
	CGC	1	14	45	7	5	6	8	1	1	2	11	3	10	8	—	11	6	247	42%
	CGG	—	9	15	1	1	1	1	2	8	8	3	—	2	4	2	3	4	120	20%
	CGT	—	7	6	6	7	6	8	3	5	4	13	5	6	4	1	8	1	175	30%
	AGA	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	7	1%
	AGG	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	1%
Leu (L)	CTA	2	4	8	—	—	—	—	8	—	—	—	—	—	1	1	—	1	47	4%
	CTC	2	15	40	5	4	4	11	10	2	2	10	3	2	—	2	5	6	230	20%
	CTG	1	23	61	15	17	14	25	11	17	17	22	3	6	3	2	14	11	456	41%
	CTT	1	3	8	1	—	—	1	3	—	—	1	—	—	1	—	—	—	49	4%
	TTA	6	—	1	1	—	—	1	4	2	2	—	—	—	—	—	—	—	31	3%
	TTG	—	8	24	9	11	14	11	10	19	19	10	1	4	2	1	8	9	311	28%
Ser (S)	TCA	3	3	6	1	—	—	2	—	2	2	—	—	2	1	1	—	—	44	7%
	TCC	1	1	7	3	2	3	8	4	—	—	5	—	5	1	1	2	4	88	14%
	TCG	1	3	16	11	10	10	9	3	11	11	10	2	3	2	—	6	5	185	29%
	TCT	1	—	6	1	2	1	5	3	1	1	2	—	1	1	—	1	1	65	10%
	AGC	2	11	19	11	12	12	5	2	6	6	2	5	1	1	2	3	7	192	30%
	AGT	4	7	9	1	—	—	1	2	—	—	—	—	1	—	3	—	1	60	9%
Thr (T)	ACA	3	3	5	—	—	—	—	5	—	—	—	—	—	2	1	1	1	37	6%
	ACC	1	6	26	13	12	13	18	2	8	8	24	1	7	6	2	16	13	327	53%
	ACG	—	6	24	1	3	3	4	4	9	9	4	1	—	1	4	11	8	148	24%
	ACT	4	2	9	3	4	3	—	8	—	—	1	2	—	2	1	5	3	102	17%
Pro (P)	CCA	—	5	7	1	—	—	2	4	3	3	1	—	1	—	1	—	1	54	11%
	CCC	—	11	19	5	8	7	6	5	4	3	7	5	2	3	2	8	4	153	31%
	CCG	—	13	24	5	5	6	7	4	9	9	9	1	2	3	1	5	6	181	37%
	CCT	3	6	7	5	3	3	8	1	—	1	5	2	1	2	3	6	2	100	20%
Ala (A)	GCA	1	10	6	11	18	17	7	15	10	10	7	1	3	1	—	7	9	240	21%
	GCC	4	14	23	7	6	6	12	16	7	7	8	—	10	4	2	6	6	296	26%
	GCG	2	21	20	4	4	5	7	5	15	15	10	—	1	2	3	6	7	249	22%
	GCT	4	17	17	8	7	7	16	12	3	3	19	2	5	—	5	23	7	347	31%
Gly (G)	GGA	3	1	6	—	—	—	1	6	1	1	—	—	2	—	1	—	1	46	5%
	GGC	7	13	27	15	16	16	29	13	11	11	18	3	2	5	—	18	14	397	46%
	GGG	1	8	7	—	—	—	2	1	13	13	4	—	—	—	1	2	1	84	10%
	GGT	7	7	11	19	17	17	28	15	8	8	23	1	3	5	2	21	13	342	39%
Val (V)	GTA	2	—	4	1	2	2	1	2	5	5	—	—	1	—	3	—	1	48	6%
	GTC	2	5	21	4	1	2	15	10	3	3	13	2	1	4	4	11	10	249	32%
	GTG	—	2	26	14	14	14	11	7	15	15	10	2	5	3	2	13	7	273	35%
	GTT	6	7	10	7	10	9	5	7	1	1	4	1	3	3	8	12	4	205	26%
Ile (I)	ATA	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	0%
	ATC	2	14	23	18	20	19	20	6	5	5	21	7	3	4	2	26	4	384	63%
	ATT	11	11	28	6	2	3	10	9	10	10	1	—	—	4	3	5	5	218	36%
Lys (K)	AAA	4	7	10	2	1	1	5	5	1	1	17	2	4	8	5	13	14	213	56%
	AAG	4	7	16	—	—	—	5	2	3	3	9	3	5	5	1	8	4	164	44%
Asn (N)	AAC	2	7	19	19	19	18	11	4	14	14	13	2	6	3	3	8	5	273	75%
	AAT	9	7	12	2	2	3	5	5	—	—	1	1	1	—	—	1	—	93	25%
Gln (Q)	CAA	9	14	29	9	7	7	7	6	7	7	8	5	4	2	3	5	4	286	53%
	CAG	3	30	50	1	4	4	2	5	4	4	4	2	2	3	3	6	4	249	47%

Table I. *continued*

aa	codon	gene																			
		apcA	*apcB*	*apcC*	*atpA*	*atpB*	*atpC*	*atpD*	*atpE*	*atpF*	*atpG*	*atpH*	*atpI*	*cpbA*	*cpcA*	*cpcB*	*fus*	*Gene1*	*irpA*	*petF*	*petF1*
His (H)	CAC	—	—	—	1	5	—	1	—	—	—	—	2	1	—	—	7	—	1	—	1
	CAT	—	—	—	1	—	—	1	1	1	1	—	3	1	1	—	1	—	3	1	—
Glu (E)	GAA	10	4	2	21	20	18	6	9	10	8	4	8	12	5	4	35	2	10	4	5
	GAG	4	3	1	13	16	4	3	1	7	4	—	7	5	2	3	23	3	9	2	1
Asp (D)	GAC	9	7	—	18	22	7	3	2	2	4	—	5	10	5	9	24	—	4	4	8
	GAT	1	3	1	7	6	7	8	7	5	6	1	—	15	4	4	22	—	12	4	4
Tyr (Y)	TAC	8	10	1	14	8	7	2	—	—	2	1	6	6	7	5	12	5	—	2	4
	TAT	—	2	—	1	4	—	—	1	2	—	—	3	6	2	—	6	3	5	3	1
Cys (C)	TGC	1	1	1	1	—	1	—	—	—	—	—	—	4	—	—	4	1	1	1	3
	TGT	—	—	—	—	1	—	—	—	—	—	—	—	—	1	3	—	—	2	3	1
Phe (F)	TTC	2	1	2	7	14	5	2	1	2	2	3	9	11	3	3	18	2	2	1	1
	TTT	1	—	2	4	3	4	5	2	1	3	—	10	10	3	3	5	3	9	1	1
Met (M)	ATG	3	4	2	8	7	5	2	1	1	3	2	6	8	1	4	18	4	1	2	1
	GTG																1		1		
Trp (W)	TGG	—	—	1	—	—	—	—	2	1	1	—	5	12	1	—	4	4	8	—	—
End (U)	TAA	1	1	—	—	—	—	—	—	—	—	1	—	1	—	—	1	—	—	—	1
	TAG	—	—	1	1	1	1	1	—	1	1	—	1	—	1	1	—	—	1	—	—
	TGA	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1	—	1	—
Total:		162	162	68	506	485	317	181	138	172	159	82	262	451	164	174	695	119	357	106	100
position 1:	A	36	44	25	128	116	67	38	25	32	23	13	55	123	45	41	175	24	74	18	22
	C	34	30	20	112	103	91	45	32	46	48	19	70	102	25	29	138	31	102	25	12
	G	72	61	13	204	194	117	69	65	71	70	39	78	150	62	77	293	38	135	43	45
	T	20	27	10	62	72	42	29	16	23	18	11	59	76	32	27	89	26	46	20	21
	total	162	162	68	506	485	317	181	138	172	159	82	262	451	164	174	695	119	357	106	100
position 2:	A	46	46	19	148	145	90	49	36	52	58	12	57	131	48	43	213	23	98	33	34
	C	38	41	14	125	119	75	41	34	45	39	28	65	113	49	50	159	24	101	31	31
	G	32	31	15	79	74	56	29	30	26	17	13	35	91	29	33	112	26	71	18	11
	T	46	44	20	154	147	96	62	38	49	45	29	105	116	38	48	211	46	87	24	24
	total	162	162	68	506	485	317	181	138	172	159	82	262	451	164	174	695	119	357	106	100
position 3:	A	25	21	9	63	61	41	33	48	32	31	10	26	79	18	22	102	23	63	19	15
	C	68	81	24	184	184	105	58	36	48	38	26	99	159	52	67	249	28	99	32	47
	G	30	29	22	152	160	104	46	23	49	47	22	83	105	45	42	183	39	97	28	16
	T	39	31	13	107	80	67	44	31	43	43	24	54	108	49	43	161	29	98	27	22
	total	162	162	68	506	485	317	181	138	172	159	82	262	451	164	174	695	119	357	106	100
%G+C pos. 1		65%	56%	49%	62%	61%	66%	63%	70%	68%	74%	71%	56%	56%	53%	61%	62%	58%	66%	64%	57%
%G+C pos. 2		43%	44%	43%	40%	40%	41%	39%	46%	41%	35%	50%	38%	45%	48%	48%	39%	42%	48%	46%	42%
%G+C pos. 3		60%	68%	68%	66%	71%	66%	57%	43%	56%	53%	59%	69%	59%	59%	63%	62%	56%	55%	57%	63%

Table I. ***continued***

aa	codon	gene																	sum	freq
		petG	*phr*	*ppc*	*psbA1*	*psbA2*	*psbA3*	*psbC1*	*psbC2*	*psbD1*	*psbD2*	*rbcL*	*rbcS*	*rps7*	*rps12*	*trxM*	*tufA*	*wxA*		
His (H)	CAC	—	5	15	10	10	10	9	7	7	7	15	2	1	3	—	10	—	130	74%
	CAT	—	6	9	1	—	—	3	7	1	1	1	1	—	—	1	—	—	46	26%
Glu (E)	GAA	8	15	39	4	5	5	10	9	6	6	24	3	8	3	3	23	13	381	58%
	GAG	3	8	52	12	12	12	8	1	12	12	5	8	3	1	1	9	5	275	42%
Asp (D)	GAC	5	11	18	6	5	6	13	7	6	6	25	2	3	3	2	20	9	290	56%
	GAT	11	16	27	4	3	2	5	5	3	3	6	3	3	1	6	7	8	230	44%
Tyr (Y)	TAC	2	8	23	12	12	12	11	10	7	7	12	6	4	2	—	8	5	241	73%
	TAT	6	6	13	2	2	2	3	4	1	1	2	—	1	2	2	1	4	91	27%
Cys (C)	TGC	—	4	6	2	3	3	2	2	2	2	5	4	—	2	—	—	3	59	68%
	TGT	1	—	4	2	1	1	—	—	1	1	2	—	—	—	2	1	1	28	32%
Phe (F)	TTC	3	8	25	23	26	26	21	15	21	21	19	7	4	—	1	9	10	330	65%
	TTT	5	10	18	5	2	2	8	8	16	16	5	1	1	1	2	3	3	176	35%
Met (M)	ATG	2	6	16	14	13	13	10	5	10	10	12	4	9	2	3	13	2	227	98%
	GTG		1					1											4	2%
Trp (W)	TGG	4	17	13	10	10	10	16	10	14	14	9	2	1	—	2	1	—	172	100%
End (U)	TAA	1	—	—	1	1	1	1	1	—	—	1	1	—	—	—	—	—	14	38%
	TAG	—	1	—	—	—	—	—	—	1	1	—	—	1	1	1	1	1	19	51%
	TGA	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4	11%
		171	485	1054	361	361	361	462	343	353	353	473	112	157	125	108	410	278	10827	
position 1:	A	48	94	218	90	88	88	95	59	66	66	105	28	37	38	30	110	67	2451	23%
	C	23	166	359	72	71	68	98	79	70	70	110	32	40	37	22	81	50	2562	24%
	G	66	156	314	116	120	120	171	131	119	119	176	28	53	35	43	178	115	3956	37%
	T	34	69	163	83	82	85	98	74	98	98	82	24	27	15	13	41	46	1858	17%
	total	171	485	1054	361	361	361	462	343	353	353	473	112	157	125	108	410	278	10827	
position 2:	A	67	148	332	85	83	83	98	78	73	73	143	41	46	37	31	120	76	2995	28%
	C	28	121	222	79	84	84	111	91	82	82	112	17	43	31	27	103	77	2616	24%
	G	31	99	187	74	72	72	102	59	72	72	90	23	29	29	16	68	52	1975	18%
	T	45	117	313	123	122	122	151	115	126	126	128	31	39	28	34	119	73	3241	30%
	total	171	485	1054	361	361	361	462	343	353	353	473	112	157	125	108	410	278	10827	
position 3:	A	43	63	139	31	34	33	38	67	39	39	58	12	26	18	19	49	45	1494	14%
	C	34	147	356	160	161	163	199	114	104	104	208	52	61	49	23	161	106	3886	36%
	G	21	163	365	97	104	106	119	70	160	160	121	29	44	32	27	106	74	3120	29%
	T	73	112	194	73	62	59	106	92	50	50	86	19	26	26	39	94	53	2327	21%
	total	171	485	1054	361	361	361	462	343	353	353	473	112	157	125	108	410	278	10827	
%G+C pos. 1		52%	66%	64%	52%	53%	52%	58%	61%	54%	54%	60%	54%	59%	58%	60%	63%	59%	60%	
%G+C pos. 2		35%	45%	39%	42%	43%	43%	46%	44%	44%	44%	43%	36%	46%	48%	40%	42%	46%	42%	
%G+C pos. 3		32%	64%	68%	71%	73%	75%	69%	54%	75%	75%	70%	72%	67%	65%	46%	65%	65%	65%	

Table II. Listing of the genes for which the codon usage was analyzed

no	gene[a]	gene product	strain	reference
1	*apcA*	Allophycocyanin α subunit	PCC 6301	Houmard *et al.*, 1986
2	*apcB*	Allophycocyanin ß subunit	PCC 6301	Houmard *et al.*, 1986
3	*apcC*	$L_C^{7.8}$ linker polypeptide	PCC 6301	Houmard *et al.*, 1986
4	*atpA*	α subunit ATPase F_1	PCC 6301	Cozens and Walker, 1987
5	*atpB*	ß subunit ATPase F_1	PCC 6301	Cozens and Walker, 1987
6	*atpC*	Γ subunit ATPase F_1	PCC 6301	Cozens and Walker, 1987
7	*atpD*	δ subunit ATPase F_1	PCC 6301	Cozens and Walker, 1987
8	*atpE*	ε subunit ATPase F_1	PCC 6301	Cozens and Walker, 1987
9	*atpF*	*b* subunit ATPase F_0	PCC 6301	Cozens and Walker, 1987
10	*atpG*	*b'* subunit ATPase F_0	PCC 6301	Cozens and Walker, 1987
11	*atpH*	*c* subunit ATPase F_0	PCC 6301	Cozens and Walker, 1987
12	*atpI*	*a* subunit ATPase F_0	PCC 6301	Cozens and Walker, 1987
13	*cpbA*	Membrane carotenoprotein	PCC 7942	Reddy *et al.*, 1989
14	*cpcA*	Phycocyanin α subunit	PCC 7942	Lau RH *et al.*, 1987
15	*cpcB*	Phycocyanin ß subunit	PCC 7942	Lau PCK *et al.*, 1987
16	*fus*	Elongation factor EF-G	PCC 6301	Meng *et al.*, 1989
17	Gene1	ATPase associated protein	PCC 6301	Cozens and Walker, 1987
18	*irpA*	Iron regulated protein	PCC 7942	Reddy *et al.*, 1988
19	*petF*	Ferredoxin apoprotein	PCC 6301	Cozens and Walker, 1987
20	*petF1*	Ferredoxin I apoprotein	PCC 7942	Van der Plas *et al.*, 1986
21	*petG*	Flavodoxin apoprotein	PCC 7942	Laudenbach *et al.*, 1988
22	*phr*	Photolyase	1402-1/SAUG	Yasui *et al.*, 1988
23	*ppc*	Phosphoenolpyruvate carboxylase	PCC 6301	Katagiri *et al.*, 1985
24	*psbA1*	D1 protein of PS II	PCC 7942	Golden *et al.*, 1986
25	*psbA2*	D1 protein of PS II	PCC 7942	Golden *et al.*, 1986
26	*psbA3*	D1 protein of PS II	PCC 7942	Golden *et al.*, 1986
27	*psbC1*	Chlorophyll *a*-binding protein of PS II (CP43)	PCC 7942	Golden and Stearns, 1988
28	*psbC2*	Chlorophyll *a*-binding protein of PS II (CP43)	PCC 7942	Laudenbach and Straus, 1988
29	*psbD1*	D2 protein of PS II	PCC 7942	Golden and Stearns, 1988
30	*psbD2*	D2 protein of PS II	PCC 7942	Golden and Stearns, 1988
31	*rbcL*	Ribulose-bisphosphate carboxylase large subunit	PCC 6301	Shinozaki *et al.*, 1983
32	*rbcS*	Ribulose-bisphosphate carboxylase small subunit	PCC 6301	Shinozaki and Sugiura, 1983
33	*rps12*	Ribosomal protein S12	PCC 6301	Meng *et al.*, 1989
34	*rps7*	Ribosomal protein S7	PCC 6301	Meng *et al.*, 1989
35	*trxM*	Thioredoxin *m* apoprotein	PCC 7942	Muller and Buchanan, 1989
36	*tufA*	Elongation factor EF-Tu	PCC 6301	Meng *et al.*, 1989
37	*woxA*	33-kDa Mn-stabilizing protein of PS II	PCC 7942	Kuwabara *et al.*, 1987

[a] nomenclature according to Houmard and Tandeau de Marsac 1988

References

Bibb MJ, Findlay PR and Johnson MW (1984) The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene 30:157-166

Bryant DA and Tandeau de Marsac N (1988) Isolation of genes encoding components of photosynthetic apparatus. Methods Enzymol 167:755-765

Cozens AL and Walker JE (1987) The organization and sequence of the genes for ATP synthase subunits in the Cyanobacterium Synechococcus 6301; Support for an endosymbiotic origin of chloroplasts. J Mol Biol 194:359-383

Golden SS, Brusslan J and Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. EMBO J 5:2789-2798

Golden SS and Stearns GW (1988) Nucleotide sequence and transcript analysis of three photosystem II genes from the cyanobacterium *Synechococcus* sp. PCC 7942. Gene 67:85-96

Houmard J, Mazel D, Moguet C, Bryant DA and Tandeau de Marsac N (1986) Organization and nucleotide sequence of genes encoding core components of the phycobilisomes from *Synechococcus* 6301. Mol Gen Genet 205:404-410

Houmard J and Tandeau de Marsac N (1988) Cyanobacterial genetic tools: current status. Methods Enzymol 167:808-847

Katagiri F, Kodaki T, Fujita N, Izui K and Katsuki H (1985) Nucleotide sequence of the phosphoenolpyruvate carboxylase gene of the cyanobacterium, *Anacystis nidulans*. Gene 38:265-269

Kuwabara T, Reddy KJ and Sherman LA (1987) Nucleotide sequence of the gene from the cyanobacterium *Anacystis nidulans* R2 encoding the Mn stabilizing protein involved in photosystem II water oxidation. Proc Natl Acad Sci USA 84:8230-8234

Lau PCK, Condie JA, Alvarado-Urbina G and Lau RH (1987) Nucleotide sequence of phycocyanin α-subunit gene of cyanobacterium *Anacystis nidulans* strain R2. Nucl Acids Res 15:2394

Lau RH, Alvarado-Urbina G and Lau PCK (1987) Gene 52:21-29. Phycocyanin ß-subunit gene of *Anacystis nidulans* R2: cloning, nucleotide sequencing and expression in *Escherichia coli*. Gene 52:21-29

Laudenbach DE, Reith ME and Straus NA (1988) Isolation, sequence analysis, and transcriptional studies of the flavodoxin gene from *Anacystis nidulans* R2. J Bacteriol 170:258-265

Laudenbach DE and Straus NA (1988) Characterization of a cyanobacterial iron stress-induced gene similar to *psbC*. J Bacteriol 170:5019-5026

Meng BY, Shinozaki K and Sugiura M (1989) Genes for the ribosomal proteins S12 and S7 and elongation factors EF-G and EF-Tu of the cyanobacterium, *Anacystis nidulans*: structural homology between 16S rRNA and S7 mRNA. Mol Gen Genet 216:25-30

Muller EGD and Buchanan BB (1989) Thioredoxin is essential for photosynthetic growth. J Biol Chem 264:4008-4014

Reddy KJ, Bullerjahn GS, Sherman DM and Sherman LA (1988) Cloning, nucleotide sequence, and mutagenesis of a gene (*irp*A) involved in iron-deficient growth of the cyanobacterium *Synechococcus* sp. Strain PCC 7942. J Bacteriol 170:4466-4476

Reddy KJ, Masamoto K, Sherman DM and Sherman LA (1989) DNA sequence and regulation of the gene (*cbpA*) encoding the 42-kilodalton cytoplasmic membrane carotenoprotein if the cyanobacterium *Synechococcus* sp. strain PCC 7942. J Bacteriol 171:3486-3493

Shinozaki K, Yamada C, Takahata and Sugiura M (1983) Molecular cloning and sequence analysis of the cyanobacteria gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc Natl Acad Sci USA 80:4050-4054

Shinozaki K and Sugiura M (1983) The gene for the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase is located close to the gene for

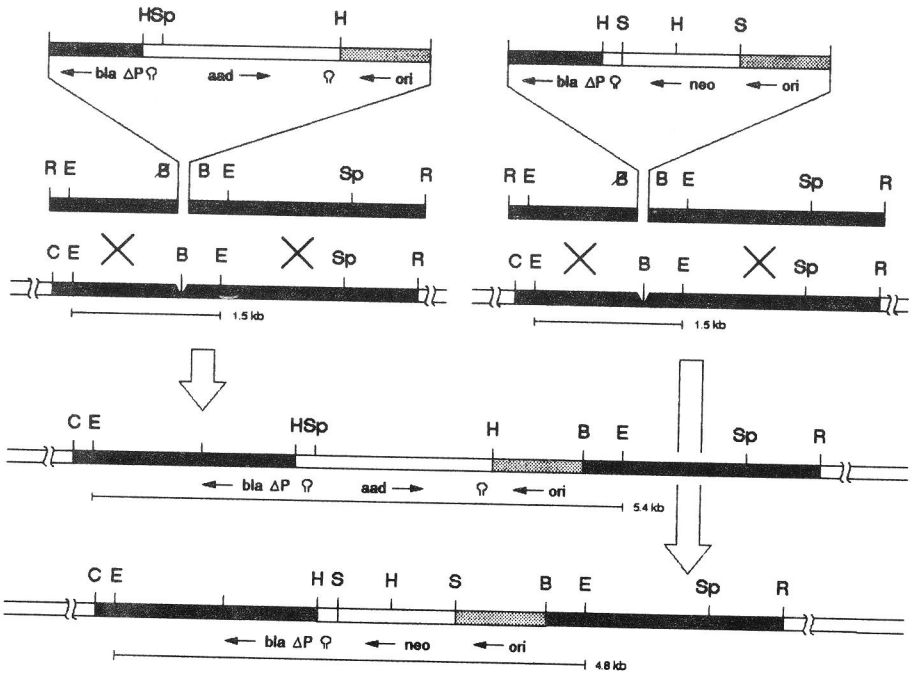
the large subunit in the cyanobacterium *Anacystis nidulans* 6301. Nucl Acids Res 11:6957–6964

Staden, R (1984) Measurement of the effects that coding for a protein has on a DNA sequence and their use for finding genes. Nucl Acids Res 12:551–567

Yasui A, Takao, M, Oikawa A, Kiener A, Walsh CT and Eker APM (1988) Cloning and characterization of a photolyase gene from the cyanobacterium *Anacystis nidulans*. Nucl Acids Res 16:4447–4463

Chapter VI

Summary
Samenvatting



Summary

In this thesis the cloning and characterization of cyanobacterial genes encoding ferredoxin and plastocyanin is described. The dynamic regulation of these small metalloproteins by the availability of Fe^{3+} and Cu^{2+}, respectively, was studied. In addition, an effective integrating shuttle system was developed for *Synechococcus* sp. PCC 7942 and applied for the transfer of ferredoxin and plastocyanin genes to *Synechococcus* sp. PCC 7942. Finally, the complete nucleotide sequence of pUH24, the small cryptic plasmid of *Synechococcus* sp. PCC 7942, was determined and analysed for its genetic content. In combination with segregation analysis of pUH24 derived vectors this led to the identification of sequences essential for replication and stable maintenance of pUH24.

Chapter I gives a brief general introduction of several aspects of gene analysis and regulation in cyanobacteria, especially focused on genes involved in photosynthesis.

Chapter II describes the cloning of the ferredoxin genes from *Anabaena variabilis* sp. PCC 7937 and *Synechococcus* sp. PCC 7942. The gene encoding ferredoxin I (*petF1*) from the filamentous cyanobacterium *Anabaena* sp. PCC 7937 was cloned by low stringency hybridization with the ferredoxin cDNA from the higher plant *Silene pratensis*. The *petF1* gene from the unicellular cyano bacterium *Synechococcus* sp. PCC 7942 in its turn was cloned by low stringency hybridization with the *petF1* gene from *Anabaena* sp. PCC 7937. Both the *Synechococcus* sp. PCC 7942 and the *Anabaena* Sp. PCC 7937 *petF1* genes encode a small acidic protein, consisting of 98 amino acid residues, with a molecular mass of about 10.5 kDa. One copy of the *petF* genes was found in each organism. The second *Anabaena* gene, coding for an other [2Fe-2S]-type ferredoxin, obviously is too distantly related to be detected. A single transcript of about 630 bases was found for *Synechococcus* sp. PCC 7942.

Iron limitation was known to induce in cyanobacteria a shift from the synthesis of ferredoxin to the production of the alternative electron carrier flavodoxin, as was concluded from protein purification and reconstitution experiments. However, it had not been established whether flavodoxin can fully replace ferredoxin *in vivo*. The presence of ferredoxin, even though at a very reduced level, might be essential for certain specific functions. To determine the level at which the regulation of the ferredoxin/flavodoxin interchange occurs, both the ferredoxin protein content and the level of mRNA encoding ferredoxin were studied.

The ferredoxin protein content of *Synechococcus* sp. PCC 7942 indeed was found to be strongly reduced under iron-limited growth conditions. In Western blot analysis the samples from cells grown in the presence of iron reproducibly gave signals approximately 30 times stronger than the samples from iron-limited cultures. Electron microscopic analysis of cryo sections of iron-limited cells and iron-replete cells with gold-labelled ferredoxin antibodies, showed the same degree in ferredoxin reduction as was been found in the Western blots.

The ferredoxin mRNA level in iron-limited *Synechococcus* sp. PCC 7942 cells, as determined by Northern hybridisation analysis of total RNA, was found to be only 2-3 times lower than in cells grown in standard medium. This slight decrease in the amount of ferredoxin transcript found under iron limitation does not account for the significant reduction in ferredoxin protein observed. Therefore the main regulation of the ferredoxin content is probably effected on the level of translation of ferredoxin messenger and/or degradation of the ferredoxin (apo-)protein.

Although ferredoxin expression can be strongly reduced by iron stress, the ferredoxin function seems to be indispensable, as *Synechococcus* sp. PCC 7942 appeared refractory to yield a mutant lacking the *petF1* gene. Such a mutant would be very useful as the host strain for studies on structure and function of the four plant-type ferredoxin genes isolated in our laboratory. In repeated attempts to eliminate the ferredoxin gene via homologous recombination by replacement for a kanamycin resistance gene, those cells seem to be selected in which a partially heterozygous situation secures the cell of both the necessary *petF1* and *neo* gene products. These results are interpreted as an indication that the *petF1* ferredoxin is essential to the cell, even under conditions of iron-limitation, when the synthesis of flavodoxin is induced. The precise genetic make-up of these transformants is not yet understood. The intact and the impaired gene copy probably reside on different copies of the chromosome, because *Synechococcus* sp. PCC 7942 contains several genome equivalents per growing cell. Further evidence regarding the indispensability of ferredoxin may

come from conditional ferredoxin mutants, in which expression of the *petF1* gene is directed by a controllable promoter.

In **Chapter III** the isolation of the gene for plastocyanin (*petE1*) from the filamentous cyanobacterium *Anabaena* sp. PCC 7937 by hybridization with two sets of mixed oligonucleotide probes, derived from conserved regions in the protein, is reported. The gene encodes a protein precursor of 139 amino acids. The amino-terminal extension of 34 residues shows all the characteristics of a signal peptide and is most likely involved in translocation of preplastocyanin over the thylakoid membrane, into the thylakoid lumen where plastocyanin is functional. At present, however, nothing is known about the mechanism of protein targeting to the functionally distinct cytoplasmic membrane and intracytoplasmic thylakoid membrane of cyanobacteria. The inferred signalpeptide sequence did not reveal a direct clue why its specifically directs its passenger protein to the thylakoid lumen. Isolation of the gene for plastocyanin now provides a tool to investigate protein routing to and transport over the thylakoid membrane.

Southern hybridization analysis of *Synechococcus* sp. PCC 7942 chromosomal DNA with the *Anabaena* sp. PCC 7937 plastocyanin gene as probe gave no evidence for the presence of *petE1* specific sequences. Moreover, attempts in the past to purify plastocyanin or to detect plastocyanin immunologically for *Anacystis nidulans* (former designation of *Synechococcus* sp. PCC 6301 and closely related strains like *Synechococcus* sp. PCC 7942) failed. We conclude from the combined results that *Synechococcus* sp. PCC 7942 has no genetic information to synthesize plastocyanin. This fact and the genetic tools available for *Synechococcus* sp. PCC 7942 make the strain an interesting host for the analysis of targeting and functioning of *Anabaena* sp. PCC 7937 plastocyanin. As a first step in this direction we successfully transferred the *Anabaena* sp. PCC 7937 gene to the genome of *Synechococcus* sp. PCC 7942 (See also Chapter IV).

The copper-containing plastocyanin can be replaced by the iron-haem protein cytochrome c_{553}. Many cyanobacteria, among them *Synechococcus* sp. PCC 7942, have cytochrome c_{553} instead of plastocyanin. Other cyanobacterial species (e.g. *Anabaena variabilis*) express either plastocyanin, or cytochrome c_{553}, or both. The plastocyanin content of the latter organisms depends on the availability of copper to the cell· under growth conditions with insufficient Cu^{2+}, cytochrome c_{553} is produced as an alternative and additional electron carrier for plastocyanin. To investigate the level at which regulation occurs, the levels of plastocyanin mRNA were determined by Northern analysis. A single specific transcript of about 740 b was detected.

Regulation of the plastocyanin content of *Anabaena* sp. PCC 7937 was found to occur by influencing the steady state levels of plastocyanin transcripts, as the amount of the specific mRNA in cells grown under Cu^{2+}-limitation is approximately 30-fold lower than that in cells grown in complete medium. This points to regulation of transcription initiation and/or differential stability of the mRNA under different conditions for growth. The *Anabaena* sp. PCC 7937 plastocyanin gene transferred to the genome of *Synechococcus* sp. PCC 7942, however, showed no Cu^{2+} dependent transcription into plastocyanin mRNA in its new host.

Chapter IV details the development of a genomic integration system for *Synechococcus* sp. PCC 7942 based on pBR sequences. *Synechococcus* sp. PCC 7942 recipient strains were constructed for the chromosomal integration of DNA fragments cloned in any Ap^R vector derived from pBR322. The construction was based on the incorporation of specific recombination targets, the so-called "integration platforms", into the chromosome. The location of these platforms is the *met1* gene of *Synechococcus* sp. PCC 7942, which can be inactivated without affecting the viability of the host strain when grown on methionine-supplemented medium. The platforms consist of an incomplete *bla* gene (Ap^S) and the pBR322 origin of replication, separated from each other by a functional antibiotic resistance gene: the *aad* gene (Sm^R) in the case of strain *Synechococcus* R2-PIM8 (Ap^SSm^R), and the *neo* gene (Km^R) in the case of strain *Synechococcus* R2-PIM9 (Ap^SKm^R). Recombination between a pBR-derived donor plasmid and such a chromosomal platform in both the *bla* and the *ori* sequences, results with high frequency in restoration of the *bla* gene and in concomitant replacement of the chromosomal marker (Sm^R or Km^R) by the insert of the donor plasmid ("insert replacement"). The integration into the platform depends on recombination between pBR sequences only and is therefore independent of the DNA insert to be transferred, unless of course the expression of the inserted DNA in *Synechococcus* sp. PCC 7942 has some deleterious effect on the cell. The desired recombinants are easily found by selection for a functional *bla* gene (Ap^R) and subsequent screening for absence of the chromosomal antibiotic marker.

Transformations of the platform strains *Synechococcus* R2-PIM8 (Ap^SSm^R) and R2-PIM9 (Ap^SKm^R) were performed with plasmids pUC7-K (Ap^R Km^R) and pUC7-S (Ap^R Sm^R), respectively. Insert replacement with these donor plasmids will result in substitution of the *aad* gene of the PIM8 platform by the *neo* gene from pUC7-K, and of the *neo* gene of the PIM9 platform by the *aad* gene from pUC7-S. The accuracy of these transfers could be measured by challenging the finally selected transformants for their new antibiotic resistances. The Ap^R colonies showing loss of the original platform-sited antibiotic resistance had acquired the new antibiotic resistance from

the donor plasmids in 100% of the cases. This showes that the integration system functions accurately.

The structure of the integration platform allows the reisolation of the sequences cloned into the platform (plasmid rescue). In the transformed cells the cloned DNA is flanked by a functional *bla* gene and a complete pBR origin of replication. This enables the recovery of the platform region as an autonomously replicating plasmid, after excision from the chromosome, circularization and transformation to *E. coli*.

The platform integration system was used to bring into *Synechococcus* sp. PCC 7942 the plastocyanin gene (*petE1*) from *Anabaena* sp. PCC 7937. *Synechococcus* sp. PCC 7942 itself does not contain a gene for plastocyanin (Chapter III). Extensive hybridisation analysis showed that the plastocyanin gene had been integrated successfully and in the expected way into the genome of *Synechococcus* sp. PCC 7942. Furthermore, the plastocyanin gene was specifically transcribed into mRNA. Antibodies elicited by a synthetic amino-terminal fragment of *Anabaena* plastocyanin are presently used to investigate the plastocyanin expression at the protein level and the routing of plastocyanin in its new host.

The platform system can also be used for the construction of partial diploids. This was shown by the transfer of the gene encoding ferredoxin I (*petF1*) from *Synechococcus* sp. PCC 7942, for which an endogenous genomic copy is already present. From extensive hybridisation analysis of the selected transformants it appeared that the ferredoxin gene has been correctly integrated at the expected site into the genome of *Synechococcus* sp. PCC 7942. So, even in the presence of a homologous resident gene the desired integration of an additional copy of this gene at the site of the platform can easily be obtained.

In this way a copy of the ferredoxin gene can be integrated that is expressed from a controllable promoter, allowing the subsequent deletion or mutation of the resident ferredoxin gene, in order to obtain conditional ferredoxin mutants. Such experiments may provide more evidence on the indispensability of ferredoxin and also can offer opportunities to study structure-function relationships of ferredoxin, e.g. in its interactions with nitrate reductase.

In **Chapter V** the first complete sequence of a cyanobacterial plasmid is presented. Analysis of the nucleotide sequence of pUH24, the small cryptic plasmid from *Synechococcus* sp. PCC 7942 gives the opportunity to gain insight into the functional organization of the plasmid and will facilitate its use in the construction of improved cloning vectors. Eight major open reading frames are postulated to be actual coding

sequences. The proteins they encode show no significant sequence similarities to previously characterized proteins of known or unknown function. The largest of these coding sequences almost fully occupies the essential 3.5 kb region minimally shared by all reported functional cloning vectors for *Synechococcus* sp. PCC 7942. This fact could explain the relatively large contiguous fragment needed for replication.

Except for the selection of a convenient fragment containing the functions essential for replication, little or no attention has been paid before to plasmid (in)stability related to vector sequences in cyanobacteria. Two plasmid stability related functions are found on pUH24. In the first place, a region has been detected that consists of 12 palindromic elements (10bp) with the potential of forming extensive secondary structures. These structures could e.g. function as a target for site-specific resolution of multimers, as the same region has been associated with the segregational stability of the plasmid in a population of dividing cells. Deletion of part of this locus resulted in loss of plasmid from approximately 1 in 400 cells per generation of unselected growth. However, a potential open reading frame overlapping this region has also been identified, which means that the role of a functional protein can not be excluded completely. In the second place, just in front of the ORF mentioned above another ORF coding for a stability protein is present, as deletion of the first 56 triplets from this coding sequence resulted in severe plasmid instability.

In the functional analysis of cloned *Synechococcus* DNA the use of the pUH24 based shuttle vectors is limited by the lack of a recombination-deficient strain of *Synechococcus* sp. PCC 7942, because recombination between a cloned gene on a autonomously replicating vector and its chromosomal counterpart leads to homogenotization with high frequency. With the genomic integration system no stability problems were thus far encountered. Besides, complementation and dominance studies of cloned genes can now be undertaken by genomic integration in one of the platform strains without the complication of copy number differences of the complementing alleles which could arise when the cloned gene is present on a cyanobacterial (shuttle-)vector.

Samenvatting

In het eerste deel van dit proefschrift wordt, na een algemene inleiding (Hoofdstuk I), de klonering en analyse beschreven van cyanobacteriële genen coderend voor ferredoxine en plastocyanine, beide elektronen carriers uit de fotosynthetische elektronen transportketen (Hoofdstuk II en III). Ferredoxine is een klein ijzer–zwavel eiwit (±10 kDa) en plastocyanine is een ongeveer even groot koper bevattend eiwit. De regulatie in de cel van het gehalte aan deze elektronen transport eiwitten door de hoeveelheid beschikbare ijzer, respectievelijk koper ionen in het kweekmedium komt uitgebreid aan de orde.

Het ferredoxine gen (*petF1*) van *Anabaena* sp. PCC 7937 werd gekloneerd met behulp van laag stringente DNA hybridisatie met een plante ferredoxine cDNA als probe. Het overeenkomstige gen van *Synechococcus* sp. PCC 7942 werd vervolgens verkregen met het *Anabaena* gen als probe. Van beide cyanobacteriële genen kon slechts één kopie worden aangetoond. Sterke reductie van het gehalte aan ferredoxine in *Synechococcus* sp. PCC 7942 gegroeid onder ijzer gebrek, werd aangetoond met behulp van ferredoxine specifieke antilichamen. Zoals reeds bekend gaan cyanobacteriën onder dergelijke omstandigheden over tot het aanmaken van flavodoxine, een alternatief voor ferredoxine. Het gehalte aan ferredoxine mRNA nu, bleek niet in overeenkomstig sterke mate af te nemen, zodat de regulatie van ferredoxine voornamelijk moet liggen op het niveau van translatie van ferredoxine mRNA en/of van afbraak van het (apo)ferredoxine eiwit. Een duidelijke aanwijzing dat ferredoxine essentieel is voor de cel en niet volledig door flavodoxine kan worden vervangen, vormt het zonder resultaat blijven van alle pogingen om het ferredoxine gen te inactiveren.

Het plastocyanine gen (*petE1*) werd geïsoleerd uit *Anabaena* sp. PCC 7937 door hybridisatie met twee sets oligonucleotide probes, gesynthetiseerd op basis van de

aminozuurvolgordes van geconserveerde gebieden in het plastocyanine eiwit. Het plastocyanine gen bleek voor een precursor eiwit te coderen met een aminoterminale extensie van 34 residuen. Dit karakteristieke signaalpeptide is naar alle waarschijnlijkheid betrokken bij transport van het (apo)plastocyanine over de thylakoid membraan naar het thylakoid lumen, het cel compartiment waarin plastocyanine functioneel is. Op dit moment is nog geheel onduidelijk met behulp van welke informatie de cel onderscheid maakt tussen eiwitten bestemd voor het thylakoid lumen en voor het periplasma. Met het plastocyanine gen is nu een hulpmiddel voorhanden om de adressering van eiwit aan het thylakoid lumen te bestuderen.

Het koper bevattende plastocyanine kan functioneel worden vervangen door cytochroom c_{553}. Veel cyanobacteriën, waaronder *Synechococcus* sp. PCC 7942, bevatten cytochroom c_{553} in plaats van plastocyanine. Andere soorten cyanobacteriën (bv. *Anabaena variabilis*) brengen plastocyanine, cytochroom c_{553} of beide tegelijk tot expressie. Het plastocyanine gehalte van de laatste categorie organismen wordt dan bepaald door de hoeveelheid beschikbare koper ionen in het medium. Deze regulatie door koper speelt zich af op het niveau van transcriptie initiatie en/of mRNA stabiliteit, aangezien het gemeten gehalte aan plastocyanine transcript een factor 30 lager is in koper gehongerde cellen van *Anabaena* sp. PCC 7937.

Het tweede deel van dit proefschrift is gewijd aan verbetering van de middelen voor het uitvoeren van gen analyse in *Synechococcus* sp. PCC 7942. Er wordt beschreven hoe een efficiënt systeem voor inbouw van "vreemd" DNA in het genoom van de cyanobacterie *Synechococcus* sp. PCC 7942 werd ontwikkeld en hoe dit vervolgens werd toegepast voor de overdracht van ferredoxine en plastocyanine genen naar *Synechococcus* sp. PCC 7942 (Hoofdstuk IV). Tenslotte volgt de bepaling en analyse van de volledige nucleotidenvolgorde van pUH24, het kleine cryptische plasmide van *Synechococcus* sp. PCC 7942. Door de resultaten van deze genetische analyse te combineren met de experimentele eigenschappen van afgeleide plasmiden, konden gebieden worden aangewezen die betrokken zijn bij de vermeerdering en de stabiele handhaving van pUH24 (Hoofdstuk V).

Veel van de plasmide vectoren die momenteel gebruikt worden in het recombinant onderzoek zijn afgeleid van het alom bekende pBR322. De meeste van deze plasmiden hebben de origin van replicatie (*ori*) en het ampicilline resistentie (Ap^R) veroorzakend gen (*bla*) gemeen. Dit gegeven vormde het uitgangspunt voor de aanleg van een zogenaamd recombinatie platform in een niet essentieel chromosomaal gen (*met1*) van *Synechococcus* sp. PCC 7942. Een dergelijk platform bestaat uit een inkompleet *bla* gen (Ap^S) plus het *ori* gebied, van elkaar gescheiden door een fragment met een functioneel

antibioticum resistentie gen (bv Km^R). Recombinatie nu tussen een pBR-afgeleid plasmide en een dergelijk chromosomaal platform resulteert zeer frequent in herstel van het *bla* gen (Ap^R) en in vervanging van de chromosomale antibioticum resistentie marker door het insert van het pBR-afgeleide donor plasmide (in dit voorbeeld Km^S). Deze integratie van een insert in het platform berust slechts op recombinatie tussen van pBR afkomstige sequenties en is daarom onafhankelijk van het DNA fragment wat dient te worden overgebracht. De gewenste transformanten worden gevonden door selectie op een functioneel *bla* gen (Ap^R), gevolgd door screening op de afwezigheid van de chromosomale antibioticum resistentie. Gen transfer met dit systeem is efficiënt en betrouwbaar, zoals bleek uit tests waarin het te integreren DNA fragment een nieuwe eenvoudig selecteerbare eigenschap bevatte (nl. een derde antibioticum resistentie naast het pBR afkomstige *bla* gen en de tot het platform behorende marker).
Een bijkomend voordeel van de aanwezigheid van de origin van replicatie in het platform is dat het de reisolatie van de in het platform geïntegreerde sequenties mogelijk maakt. Digestie van chromosomaal DNA met een geschikt restrictie enzym wat niet in platform met insert knipt, circularisatie en transformatie van *E. coli* (Ap^R), leidt tot de reisolatie van autonoom replicerende plasmiden bestaande uit het platform plus insert met enig flankerend chromosomaal DNA.

Het platform integratie systeem werd gebruikt om het plastocyanine gen (*petE1*) van *Anabaena* sp. PCC 7937 naar *Synechococcus* sp. PCC 7942 over te brengen. *Synechococcus* sp. PCC 7942 zelf bevat geen plastocyanine gen. Uitgebreide hybridisatie analyse wees uit dat het plastocyanine gen op de juiste wijze in het genoom van *Synechococcus* sp. PCC 7942 was geïntegreerd. Verder bleek dat het plastocyanine gen in *Synechococcus* sp. PCC 7942 op gering niveau werd afgeschreven in mRNA, maar dat dit niet werd gereguleerd door koper zoals in *Anabaena* sp. PCC 7937. Antilichamen opgewekt tegen een synthetisch amino-terminaal peptide van het *Anabaena* plastocyanine worden momenteel gebruikt om de expressie van het plastocyanine gen op eiwit niveau te onderzoeken en tevens om na te gaan of het eiwit ook in de nieuwe gastheer zijn uiteindelijk plaats van bestemming, het thylakoid lumen bereikt.

Het platform systeem kan ook worden toegepast om partiële diploïden te maken. Dit werd aangetoond door de integratie van een extra ferredoxine I gen in *Synechococcus* sp. PCC 7942. Hybridisatie analyse van de geselecteerde transformanten wees uit dat het extra ferredoxine gen op de juiste plaats en op de juiste wijze was geïntegreerd. Dus zelfs in de aanwezigheid van het oorspronkelijke gen kan de gewenste integratie van een extra genkopie in het platform eenvoudig worden bewerkstelligd. Op deze

wijze zou een ferredoxine gen onder controle van een reguleerbare promoter kunnen worden aangebracht, zodat de daaropvolgende mutatie van het originele ferredoxine gen conditionele ferredoxine mutanten kan opleveren. Dergelijke experimenten zouden meer inzicht kunnen verschaffen in de onmisbaarheid van ferredoxine en zouden ook de bestudering mogelijk moeten maken van structuur-functie relaties in het ferredoxine eiwit, bijvoorbeeld in zijn interacties met nitraat reductase.

In het laatste experimentele hoofdstuk wordt de eerste volledige sequentie gepresenteerd van een cyanobacterieel plasmide. Deze analyse van de nucleotidenvolgorde van pUH24, het kleine cryptische plasmide van *Synechococcus* sp. PCC 7942, verschaft inzicht in de functionele organisatie van het plasmide en zal zeer nuttig zijn bij de constructie van verbeterde kloneringsvectoren gebaseerd op pUH24. Acht van de grootste aanwezige open reading frames (ORF's) vertegenwoordigen waarschijnlijk eiwit coderende sequenties. De eiwitten die ze coderen vertonen geen significante overeenkomst in aminozuurvolgorde met eerder gekarakteriseerde eiwitten. De langste van deze coderende sequenties beslaat het grootste deel van het 3,5 kb grote gebied wat alle functionele van pUH24 afgeleide kloneringsvectoren gemeenschappelijk hebben. Dit grote vrijwel zeker voor een essentiële replicatie functie coderende gen zou de verklaring zijn voor het relatief lange ononderbroken fragment wat nodig is in het minimale replicon. Naast dit grote open reading frame bevat het minireplicon nog een tweede kleiner open reading frame in dezelfde oriëntatie en gedeeltelijk overlappend met het terminale deel van het eerstgenoemde ORF. De origin van replicatie bevindt zich vermoedelijk vlak voor het grote open reading frame.

Behalve aan de selectie van een geschikt fragment met alle voor replicatie in cyanobacteriën benodigde functies, werd in het verleden vrijwel geen aandacht besteed aan plasmide (in)stabiliteit gerelateerd aan vector sequenties. Mede aan de hand van de sequentie analyse konden nu twee functies worden gelokaliseerd die betrokken zijn bij segregationele plasmide stabiliteit. In de eerste plaats werd een gebied gevonden dat bestaat uit 12 palindromische elementen (10 bp) welke ook onderling uitgebreide secundaire structuren kunnen vormen. Deze structuren, zouden bijvoorbeeld kunnen dienen voor membraan associatie of als doelwit voor site-specifieke resolutie van multimeren, beide functies direct of indirect betrokken bij de verdeling (partitioning) van plasmiden in delende cellen. Deletie van een deel van dit gebied resulteerde onder niet selectieve omstandigheden per generatie in plasmide verlies in 1 op de 400 cellen. Er ligt echter ook een potentieel open reading frame in hetzelfde gebied, zodat een rol voor een functioneel eiwit niet volledig kan worden uitgesloten. In de tweede plaats werd een open reading frame coderend voor

een stabiliteit gerelateerd eiwit aangetroffen wat ligt net voor het bovengenoemde ORF. Deletie van de eerste 56 tripletten van de coderende sequentie resulteerde in zware instabiliteit, waarbij onder niet selectieve omstandigheden per generatie ongeveer 1 op de 20 cellen zijn plasmide inhoud kwijt raakte.

Het gebruik van op pUH24 gebaseerde shuttle vectoren voor de functionele analyse van gekloneerd *Synechococcus* sp. PCC 7942 DNA, is tot nog toe beperkt gebleven vanwege het ontbreken van een recombinatie deficiënte mutant. Recombinatie tussen een gekloneerd gen op een autonoom replicerende vector en zijn chromosomale equivalent, leidt namelijk met zeer hoge frequentie tot eliminatie van eventueel aanwezige verschillen tussen beide kopieën. Met het genomische integratie systeem werden dergelijke stabiliteits problemen tot dusver niet waargenomen. Verder kunnen complementatie en dominantie studies van gekloneerde genen met behulp van het integratie systeem worden uitgevoerd, zonder complicerende verschillen in kopie aantal tussen de complementerende allelen, die kunnen optreden wanneer een van beide aanwezig is in een cyanobacteriële (shuttle-)vector.

Nawoord

"Deze reis is ten einde, het einddoel is bereikt." Bij het verschijnen van dit boekje als neerslag daarvan, wil ik een ieder bedanken die mij heeft vergezeld, me een eindje op weg heeft geholpen of gewoon een stukje mee op is gelopen.

Enkele van de vele metgezellen wil ik met name bedanken. Mijn promotores Peter Weisbeek en Gerard van Arkel dank ik voor het in mij gestelde vertrouwen en voor de positieve wijze waarop ze mijn ellenlange manuscripten steeds weer hebben opgepakt. Alle collega's dank ik voor de prettige samenwerking, in het bijzonder Mies Borrias die samen met Geert de Vrieze veel heeft bijgedragen aan de realisatie van mijn wilde chromosomale constructies in *Anacystis*. Johan Hageman en Simon Langeveld wil ik noemen voor hun geduldige introductie in het werken met de PC. Mijn bijzondere dank gaat verder uit naar Martin Woortman, Mark Tuyl, Ed Schmidt, Rolf de Groot, Frank Kruyt, Marc van Dijk, Han Hegeman, René Pieterse, Bart Klein en Edwin Dassen, die in het kader van hun doctoraalstudie een belangrijke bijdrage hebben geleverd aan het onderzoek.

Mijn ouders, die mij steeds hebben gestimuleerd datgene te gaan studeren wat ik graag wilde en die met veel enthousiasme mijn vorderingen hebben gevolgd, dank ik voor hun belangstelling en meeleven.

Tenslotte, Korry, jij zorgde samen met Floor, Hedwig en Jaap als mijn thuisbasis voor het gezonde tegenwicht en je hebt er voor gewaakt dat de eindbestemming niet uit het oog werd verloren.

Curriculum vitae

De schrijver van dit proefschrift werd op 18 augustus 1955 geboren te Landsmeer. In 1973 behaalde hij het diploma Gymnasium ß aan de Christelijke Scholengemeenschap te Emmen. In 1975 werd een begin gemaakt met de studie Biologie aan de Rijksuniversiteit te Groningen, in aansluiting op twee jaar Farmacie studie aan dezelfde instelling. In september 1979 werd het kandidaatsexamen (B4) afgelegd. Het doctoraalexamen, met als hoofdvak Moleculaire Microbiologie (Prof. Dr. W.N. Konings en Dr. K.J. Hellingwerf) en bijvak Moleculaire Genetica (Prof. Dr. G. Venema en Dr. S. Bron) werd in augustus 1983 behaald. Van september 1983 tot februari 1988 was hij als wetenschappelijk assistent verbonden aan de vakgroep Moleculaire Celbiologie, sectie Moleculaire Genetica van de Rijksuniversiteit te Utrecht. Binnen deze vakgroep werd het in dit proefschrift beschreven onderzoek verricht, onder begeleiding van Prof. Dr. P.J. Weisbeek en Prof. Dr. G.A. van Arkel. Sinds april 1988 is hij als moleculair bioloog werkzaam bij de afdeling Microbiologie van het Instituut CIVO-Analyse, Hoofdgroep Voeding en Voedingsmiddelen TNO, te Zeist.

List of publications

Van der Plas J, Hellingwerf KJ, Seijen HG, Guest JR, Weiner JH and Konings WN (1983) Identification and localization of enzymes of the fumarate and nitrate respiration systems of *Escherichia coli* by crossed immunoelectrophoresis. J Bacteriol 153:1027–1037.

Van der Plas J, de Groot RP, Weisbeek PJ and van Arkel GA (1986) Coding sequence of a ferredoxin gene from *Anabaena variabilis* ATCC29413. Nucl Acids Res 14:7803.

Van der Plas J, de Groot RP, Woortman MR, Weisbeek PJ and van Arkel GA (1986) Coding sequence of a ferredoxin gene from *Anacystis nidulans* R2 (*Synechococcus* PCC7942). Nucl Acids Res 14:7804.

Van der Plas J, De Groot R, Woortman M, Cremers F, Borrias M, Van Arkel G and Weisbeek P (1988) Genes encoding ferredoxins from *Anabaena* sp. PCC 7937 and *Synechococcus* sp. PCC 7942: structure and regulation. *Photosynth Res* 18:179–204.

Van der Plas J, Bovy A, Kruyt F, De Vrieze G, Dassen E, Klein B and Weisbeek P (1988) The gene for the precursor of plastocyanine from the cyanobacterium *Anabaena* sp. PCC 7937: isolation, sequence and regulation. Mol Microbiol 3:275–284.

Van der Plas J, Hegeman H, De Vrieze G, Tuyl M, Borrias M and Weisbeek P. Genomic integration system based on pBR sequences for the cyanobacterium *Synechococcus* sp. PCC 7942: transfer of genes encoding plastocyanin and ferredoxin. Accepted for publication in Gene.

Van der Plas J, Oosterhoff–Teertstra R, Borrias M and Weisbeek P. The small plasmid pUH24 from the cyanobacterium *Synechococcus* sp. PCC 7942: nucleotide sequence analysis, identification of replication and stability functions. submitted.

geschenk vd Plaw, 10/89