JOURNAL OF BACTERIOLOGY, Sept. 1988, p. 4406–4410
0021-9193/88/094406-05$02.00/0

Vol. 170, No. 9

Bacterial-Type Ferredoxin Genes in the Nitrogen Fixation Regions of the Cyanobacterium *Anabaena* sp. Strain PCC 7120 and *Rhizobium meliloti*

MARTIN E. MULLIGAN,† WILLIAM J. BUIKEMA, AND ROBERT HASELKORN*

Department of Molecular Genetics and Cell Biology, The University of Chicago, 920 East 58th Street, Chicago, Illinois 60637

Received 20 April 1988/Accepted 25 May 1988

The nucleotide sequence of a region located downstream of the *nifB* gene, both in the cyanobacterium *Anabaena* sp. strain PCC 7120 and in *Rhizobium meliloti*, has been determined. This region contains a gene (*fdxN*) whose predicted polypeptide product strongly resembles typical bacterial ferredoxins. Cyanobacteria have not previously been shown to contain bacterial-type ferredoxins. The presence of this gene suggests that nitrogen-fixing cyanobacteria have at least four distinct ferredoxins.

The filamentous cyanobacterium *Anabaena* sp. strain PCC 7120 will, if subjected to starvation for a source of fixed nitrogen, undergo a developmental change in which approximately every 10th cell differentiates to form a heterocyst (12). The heterocyst provides an anaerobic environment in which nitrogenase can function. Prior to expression of the nitrogen fixation (*nif*) genes, at least two DNA rearrangements must occur (10, 11). In one rearrangement, 11 kilobase pairs of DNA is excised from within the 3′ end of the *nifD* gene, thereby restoring the *nifD* reading frame and allowing expression of an intact *nifHDK* operon. The second rearrangement is now known to be the excision of approximately 55 kilobase pairs (J. W. Golden, C. D. Carrasco, M. E. Mulligan, G. J. Schneider, and R. Haselkorn, submitted for publication). We have sequenced the DNA on both sides of one of the breakpoints of the second rearrangement in the heterocyst chromosome and found that a second group of *nif* genes, consisting of the *nifB*, *nifS*, and *nifU* genes, is juxtaposed as a result of the rearrangement. The detailed description of this new group will be presented elsewhere. Here, we report the sequence of a bacterial-type ferredoxin which is found as part of this operon in *Anabaena* sp. strain PCC 7120. The presence of a bacterial-type ferredoxin gene adjacent to the *nifB* gene was first noted in *Rhizobium meliloti* (3), and we report the sequence of that gene also. Similar genes have recently been described in the same location in *Azotobacter vinelandii* (16) and *Bradyrhizobium japonicum* (7).

All manipulations described in this paper were performed by standard techniques (18). The *Anabaena* sp. strain PCC 7120 gene was subcloned from pAnH20.1 (10). The *R. meliloti* gene was subcloned from pRMB8.3R (3). Restriction digests were carried out following the recommendations of the manufacturer. DNA sequencing was performed both chemically (19) and by the chain termination method with both *Escherichia coli* DNA polymerase (24) and Sequenase (U.S. Biochemical Corp.) (27).

The organization of genes in the *Anabaena* sp. strain PCC 7120 heterocyst chromosome (i.e., after rearrangement has occurred) is shown in Fig. 1. Six *nif* genes have been identified. The genes for the major polypeptides of nitrogenase, *nifH*, *nifD*, and *nifK*, have been sequenced (17, 20, 21) and are transcribed as an operon (13). The *nifS* gene was originally identified on the basis of heteroduplex mapping as either the *nifV* or the *nifS* gene (23), and DNA sequence analysis has revealed the existence of the *nifB* and *nifU* genes (Fig. 1). Between the *nifB* gene and the *nifS* gene is a short open reading frame (*fdxN*) that can code for a polypeptide of 116 amino acids (M_r, 13,177). The DNA sequence of this reading frame is given in Fig. 2. The most remarkable feature of the predicted product of this gene is its similarity to bacterial ferredoxins, since bacterial-type ferredoxins in cyanobacteria have not previously been described. This similarity is reflected in the presence of the two cysteine clusters. The first cluster is of the type CXXCXXCXXXCP. This sequence is typically found in bacterial ferredoxins and other redox proteins that contain 4Fe:4S clusters. The second cluster is similar to the first except that it is split into two parts, CXXC and CXXXCP, with the two parts separated by nine amino acids. The split second cysteine cluster is a feature found in a number of ferredoxins that have been described in photosynthetic and nitrogen-fixing bacteria, and it is not restricted to these *nifB*-linked gene products (2, 9). The gene contains a directly repeated sequence in which 18 of 21 base pairs are identical; at the amino acid level, six of seven residues are identical. This repeating sequence is not found in other *nifB*-linked ferredoxins, and its significance is not known.

The *fdxN* gene in *Anabaena* sp. strain PCC 7120 is also affected by the DNA rearrangements that occur during heterocyst differentiation. In the vegetative cell chromosome, the 5′ and 3′ ends of the gene are separated by approximately 55 kilobase pairs of DNA (Golden et al., submitted). Under aerobic or anaerobic inducing conditions, this DNA is excised, thereby completing the *fdxN* gene. It is also probable that *fdxN* is transcribed as part of an operon comprising *nifB*, *fdxN*, *nifS*, and *nifU*. A complete transcriptional analysis of this operon will be presented elsewhere. However, the *Taq*I-*Tha*I fragment shown in Fig. 1 was used to analyze transcription through the rearrangement breakpoint that is located within *fdxN* (10). Transcription of this fragment occurs only after rearrangement has taken place (10).

* Corresponding author.

† Present address: Department of Biochemistry, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1B 3X9.

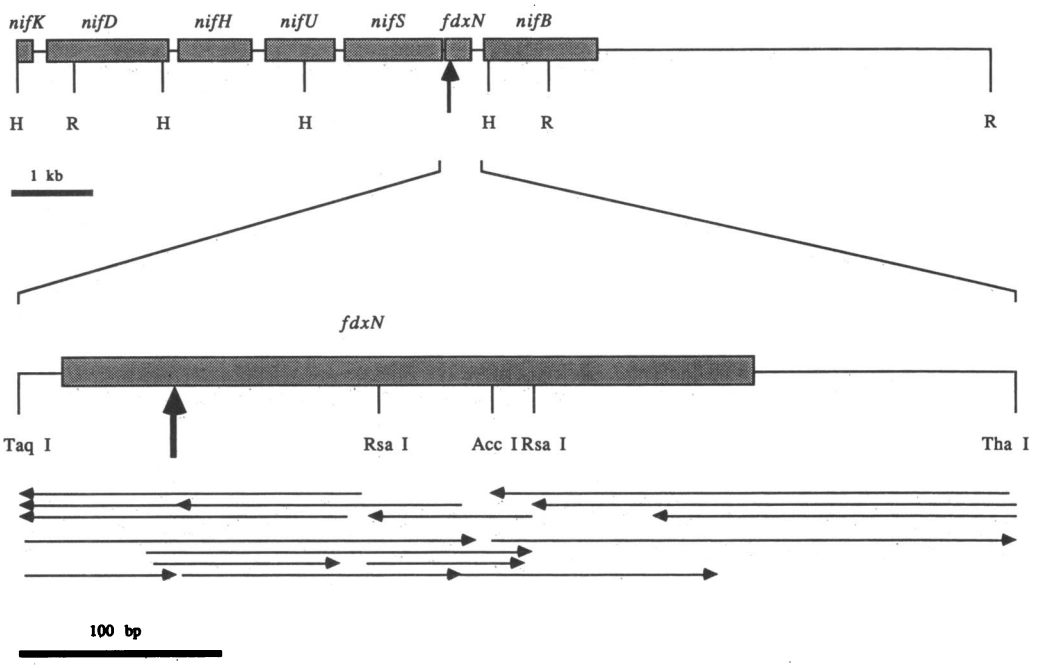

FIG. 1. Restriction map of the *Anabaena* sp. strain PCC 7120 heterocyst genome in the vicinity of the *fdxN* gene. The upper part of the figure shows the *nif* genes that are found next to *fdxN*. All are transcribed from right to left. The solid vertical arrow indicates the breakpoint of the large (55-kilobase excision) rearrangement. The restriction sites shown are for *Hin*dIII (H) and *Eco*RI (R). The lower part of the figure shows the strategy used to determine the sequence of the *Taq*I-*Tha*I fragment containing the *fdxN* gene. kb, Kilobase pair; bp, base pair.

We have also sequenced the corresponding region in *R. meliloti*. This region (Fig. 3) comprises four *fix* genes and two *nif* genes in addition to *fdxN*. The DNA sequence of the *fix* and *nif* genes has been determined previously (3, 4, 6); the sequence of the *fdxN* gene is given in Fig. 4. The gene can code for a small polypeptide of 64 amino acids (M_r, 6,821). The predicted polypeptide contains two cysteine clusters similar to those in *Anabaena* sp. strain PCC 7120. The first is

```
-10                    10          20          30          40
AGGAGAATAATC ATG GCT TAC ACT ATC ACC AGC CAA TGT ATT TCC TGC AAG CTC TGT TCG
             Met Ala Tyr Thr Ile Thr Ser Gln Cys Ile Ser Cys Lys Leu Cys Ser

 50          60          70          80          90          100
 TCT GTA TGC CCC ACT GGT GCA ATT AAA ATC GCT GAA AAC GGA CAG CAC TGG ATT GAC
 Ser Val Cys Pro Thr Gly Ala Ile Lys Ile Ala Glu Asn Gly Gln His Trp Ile Asp

   110         120         130         140         150         160
 TCC GAA CTG TGT ACA AAT TGC GTT GAT ACC GTC TAC ACA GTC CCA CAA TGT AAA GCT
 Ser Glu Leu Cys Thr Asn Cys Val Asp Thr Val Tyr Thr Val Pro Gln Cys Lys Ala

       170         180         190         200         210
 GGT TGT CCT ACT TGC GAT GGT TGC GTT AAA GTA CCT AGC GAT TAT TGG GAA GGC TGG
 Gly Cys Pro Thr Cys Asp Gly Cys Val Lys Val Pro Ser Asp Tyr Trp Glu Gly Trp

220         230         240         250         260         270
 TTT GCT AAC TAC AAC CGA GTT ATA GCG AAA TTG ACA AAA AAA CAA GAC TAT TGG GAA
 Phe Ala Asn Tyr Asn Arg Val Ile Ala Lys Leu Thr Lys Lys Gln Asp Tyr Trp Glu

 280         290         300         310         320         330
 CGT TGG TTT AAT TGT TAT TCT CAG AAA TTT TCC GAA CAG CTA CAA AAG CAT CAG GGT
 Arg Trp Phe Asn Cys Tyr Ser Gln Lys Phe Ser Glu Gln Leu Gln Lys His Gln Gly

   340         350         360         370 TaqI
 GAA ATC TTA GGG GTA TAA AG ATG AGT GTT ATT TAT CTC GAT
 Glu Ile Leu Gly Val  *     Met Ser Val Ile Tyr Leu Asp
```

FIG. 2. Nucleotide sequence and translation product of the *Anabaena* sp. strain PCC 7120 *fdxN* gene. The DNA sequence is numbered from the initiation codon of the gene. A possible ribosome-binding-site (26) sequence is underlined. The 5-base-pair core sequence (double underlined) of the large rearrangement (10) is located from residues 293 to 297. The conserved amino acids of the two cysteine clusters are double underlined. Two directly repeating sections of the predicted amino acid sequence are underlined. The stop codon is indicated by an asterisk. The first seven codons and corresponding amino acids of the *nifS* gene are also listed.

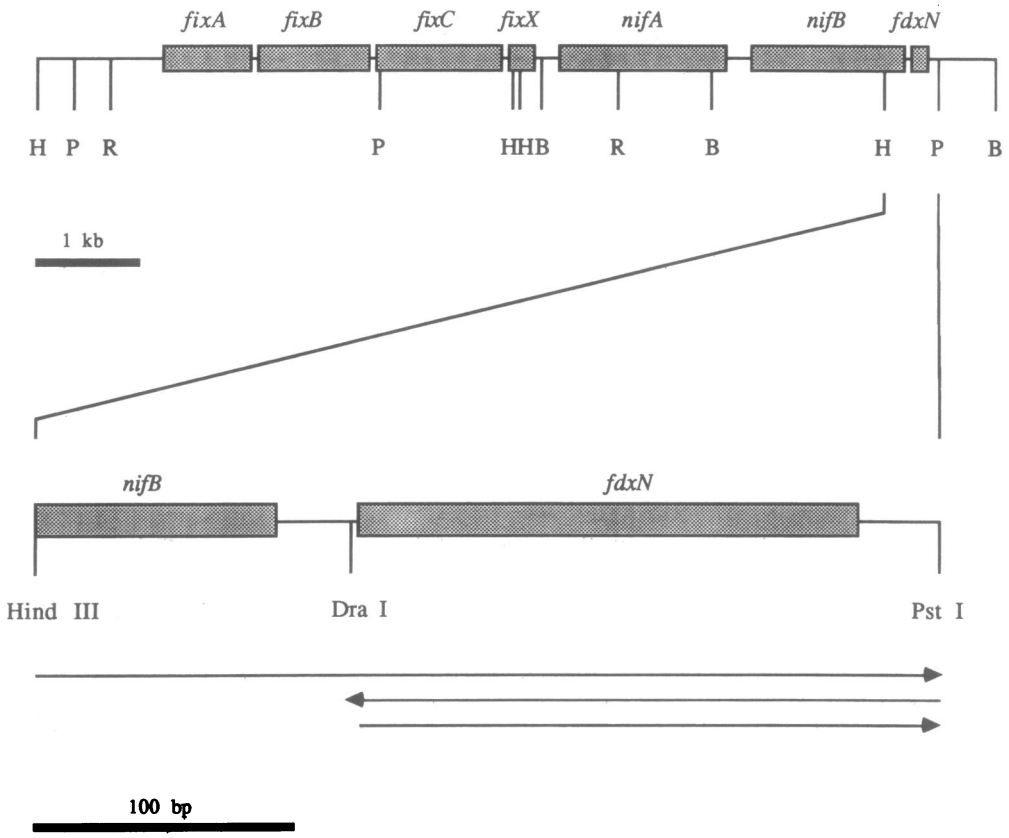

FIG. 3. Restriction map of *R. meliloti* in the vicinity of the *fdxN* gene. The upper part of the figure shows that four *fix* genes and two *nif* genes that are found next to *fdxN*. All are transcribed from right to left. This map has been adapted from maps in references 3 and 6. Restriction sites shown are for *Hin*dIII (H), *Eco*RI (R), *Pst*I (P), and *Bam*HI (B). The lower part of the figure shows the strategy used to determine the sequence of the gene. kb, Kilobase pair; bp, base pair.

a typical bacterial 4Fe:4S cluster; the second is also a split cluster in which the two parts are separated by eight amino acids.

Recently, genes coding for ferredoxinlike proteins have been described in *A. vinelandii* (*orf2*) (16) and *B. japonicum* (*frxA*) (7). The genes are located immediately downstream of the *nifB* gene in both of these organisms. The similarity between the *fdxN* gene products of *Anabaena* sp. strain PCC 7120 and *R. meliloti* and the other *nifB*-linked ferredoxins is shown in Fig. 5. All four genes are approximately 55% similar at both the DNA and protein levels. The *R. meliloti* and *B. japonicum* genes and their products are the most

```
-10   DraI              10              20              30              40
AAGGAGTTTAAA ATG GCC TTC AAG ATT ATT GCA TCT CAA TGC ACT CAG TGC GGC GCC TGC
             Met Ala Phe Lys Ile Ile Ala Ser Gln Cys Thr Gln Cys Gly Ala Cys

 50              60              70              80              90              100
 GAG TTT GAA TGT CCG CGG GGG GCG GTC AAT TTC AAA GGC GAG AAA TAC GTG ATT GAC
 Glu Phe Glu Cys Pro Arg Gly Ala Val Asn Phe Lys Gly Glu Lys Tyr Val Ile Asp

    110             120             130             140             150             160
 CCG ACC AAA TGC AAC GAA TGC AAG GGA GGA TTC GAC ACA CAG CAA TGC GCA TCG GTC
 Pro Thr Lys Cys Asn Glu Cys Lys Gly Gly Phe Asp Thr Gln Gln Cys Ala Ser Val

        170             180             190             200             210             220
 TGT CCG GTG TCG AAC ACC TGC GTC CCT GCC TGA GCTCCTACCA ATATGTTCCG ACAGAACTTC
 Cys Pro Val Ser Asn Thr Cys Val Pro Ala  *

   230          PstI
 GATTTTAAGG TCCTGCAG
```

FIG. 4. Nucleotide sequence and translation product of the *R. meliloti fdxN* gene. The DNA sequence given is that of the *Dra*I-*Pst*I fragment shown in Fig. 3, along with 6 base pairs upstream of the *Dra*I site, which was determined on one strand. The DNA sequence is numbered from the initiation codon. A possible ribosome-binding-site (26) sequence is underlined. The DNA sequence up to position 21 (double underlined) is identical to that already published (3). The conserved amino acids of the two cysteine clusters are double underlined. The stop codon is indicated by an asterisk.

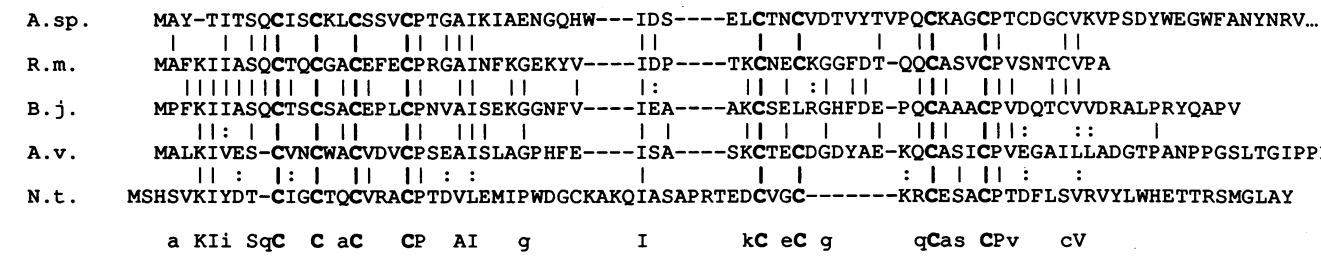

FIG. 5. Comparison of the predicted amino acid sequences of four *nifB*-linked ferredoxin genes and a chloroplast ferredoxin gene. The *nifB*-linked gene products are from *Anabaena* sp. strain PCC 7120 (A.sp.), *R. meliloti* (R.m.), *B. japonicum* (B.j.) (7), and *A. vinelandii* (A.v.) (16). The chloroplast gene product is that of the *psaC* gene of tobacco (N.t.) (14). The amino-terminal methionyl residues are included even though this residue is absent in many bacterial ferredoxins. The *Anabaena* sp. strain PCC 7120 and *A. vinelandii* sequences are truncated; there are no significant similarities in the unlisted regions. Identical residues are indicated by vertical lines between the sequences; conservative replacements are indicated by colons. Not all matches are marked. Alignments were based on the cysteinyl residues, and gaps were inserted as necessary. The conserved cysteinyl residues are in boldface. A consensus of all five sequences is given below the comparison.

similar, particularly at the 5′ and amino-terminal ends. All the proteins contain 17 or 18 amino acid residues between the first and second cysteine clusters, and all have a split second cysteine cluster with either eight or nine residues separating the two parts. The lengths of the *nifB*-linked gene products are variable; the *R. meliloti* product is the most typical of bacterial ferredoxins. The net charge, which is usually highly negative for bacterial ferredoxins, is +1 for the *Anabaena* sp. strain PCC 7120 gene product and 0 for the *R. meliloti* and *B. japonicum* gene products. Only the *A. vinelandii* gene product, with a net charge of −7, is typical of bacterial ferredoxins. The structure and evolution of bacterial ferredoxins is discussed extensively in some recent papers (2, 8, 9).

Figure 5 also indicates the similarities between the *nifB*-linked ferredoxins and a component of photosystem I in chloroplasts which is the product of the *psaC* (also called *frxA*) gene (14, 15, 28). The chloroplast protein has 25 residues separating the two cysteine clusters, and the second cluster is not split. Of the five gene products listed in Fig. 5, only the chloroplast protein has been purified. It has been isolated from a membrane fraction of chloroplasts (15, 28). Hydropathy profiles of the four *nifB*-linked ferredoxins are consistent with the possibility that they are also membrane associated (data not shown).

The function of the *nifB*-linked *fdxN* gene products is not known. The location of the genes among *nif* genes is suggestive of a role in nitrogen fixation. However, no role for the corresponding gene could be defined genetically in *B. japonicum* (7). Any proposed function in *Anabaena* spp. is complicated by the presence of two well-characterized soluble plant-type ferredoxins in *Anabaena variabilis* ATCC 29413. One of these is synthesized only in heterocysts and has been shown to transport reducing equivalents to nitrogenase in vitro, whereas the other is synthesized both in vegetative cells and in heterocysts but does not function in vitro in a nitrogenase assay (25). The gene coding for the former ferredoxin, *fdxH*, has recently been cloned and characterized in *Anabaena* sp. strain PCC 7120. It is located approximately 5 kilobase pairs downstream of the *nifK* gene; it is transcribed only under nitrogen-fixing conditions and in the same direction as *nifK* (H. Böhme and R. Haselkorn, Mol. Gen. Genet., in press).

It is likely that *Anabaena* sp. strain PCC 7120, as well as other nitrogen-fixing cyanobacteria, contains a minimum of four ferredoxins. Two are of the plant type, and two are of the bacterial type. The plant-type ferredoxins are encoded by the *petF* gene (1) and by the *fdxH* gene (Böhme and Haselkorn, in press). One of the bacterial ferredoxins is the *nifB*-linked ferredoxin that is encoded by the *fdxN* gene. The other is the 9-kilodalton component of photosystem I, which is encoded by the *psaC* gene. Neither this gene nor its product has yet been described in cyanobacteria, but since there is a great deal of similarity between cyanobacterial and chloroplast photosystems, the gene and its product probably exist. In view of this diversity of genes and gene products, we propose that ferredoxin genes, at least in cyanobacteria, be described by a dual system of nomenclature. We have adopted the prefix *fdx* to describe ferredoxins of both the plant and bacterial types. (The prefix *frx* has been used to describe only the genes for 4Fe:4S gene products in the chloroplast (5, 22) and the *nifB*-linked ferredoxin gene, *frxA*, in *B. japonicum* (7). In view of the likely presence in *Anabaena* sp. strain PCC 7120 of a *psaC* gene, which has also been called *frxA*, this name is not suitable for the *nifB*-linked ferredoxin.) Each locus is then named for the general function of its gene product in cellular metabolism. In addition, each is identified by an alternative name that indicates the specific function of its product in the cell and its relationship to other genes required for the same function. Hence, *fdxV* is another name for *petF*, the gene for the vegetative cell ferredoxin, and *fdxP* is another name for *psaC*, the gene for the 9-kilodalton component of photosystem I. Neither *fdxN* (for *nifB* linked) nor *fdxH* (for heterocyst specific) have had any specific name assigned to them yet.

This work was supported by Public Health Service grant GM21823 from the National Institutes of Health and by a grant from the Monsanto Company. W.J.B. was supported by a postdoctoral fellowship from the American Cancer Society.

LITERATURE CITED

1. **Alam, J., R. A. Whitaker, D. W. Krogmann, and S. E. Curtis.** 1986. Isolation and sequence of the gene for ferredoxin I from the cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. **168**:1265–1271.
2. **Bruschi, M., and F. Guerlesquin.** 1988. Structure, function and evolution of bacterial ferredoxins. FEMS Microbiol. Rev. **54**: 155–176.
3. **Buikema, W. J., J. A. Klingensmith, S. L. Gibbons, and F. M. Ausubel.** 1987. Conservation of structure and location of *Rhizobium meliloti* and *Klebsiella pneumoniae nifB* genes. J. Bacteriol. **169**:1120–1126.
4. **Buikema, W. J., W. W. Szeto, P. V. Lemley, W. H. Orme-Johnson, and F. M. Ausubel.** 1985. Nitrogen fixation specific

regulatory genes of *Klebsiella pneumoniae* and *Rhizobium meliloti* share homology with the general nitrogen regulatory gene *ntrC* of *K. pneumoniae*. Nucleic Acids Res. **13**:4539–4555.

5. **Dunn, P. J., and J. C. Gray.** 1988. Nucleotide sequence of the *frxB* gene in wheat chloroplast DNA. Nucleic Acids Res. **16**: 348.
6. **Earl, C. D., C. W. Ronson, and F. M. Ausubel.** 1987. Genetic and structural analysis of the *Rhizobium meliloti fixA*, *fixB*, *fixC*, and *fixX* genes. J. Bacteriol. **169**:1127–1136.
7. **Ebeling, S., J. D. Noti, and H. Hennecke.** 1988. Identification of a new *Bradyrhizobium japonicum* gene (*frxA*) encoding a ferredoxinlike protein. J. Bacteriol. **170**:1999–2001.
8. **Fukuyama, K., Y. Nagahara, T. Tsukihara, and Y. Katsube.** 1988. Tertiary structure of *Bacillus thermoproteolyticus* [4Fe-4S]ferredoxin. J. Mol. Biol. **199**:183–193.
9. **George, D. G., L. T. Hunt, L.-S. L. Yeh, and W. C. Barker.** 1985. New perspectives on bacterial ferredoxin evolution. J. Mol. Evol. **22**:20–31.
10. **Golden, J. W., M. E. Mulligan, and R. Haselkorn.** 1987. Different recombination site specificity of two developmentally regulated genome rearrangements. Nature (London) **327**:526–529.
11. **Golden, J. W., S. J. Robinson, and R. Haselkorn.** 1985. Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. Nature (London) **314**: 419–423.
12. **Haselkorn, R.** 1978. Heterocysts. Annu. Rev. Plant Physiol. **29**: 319–344.
13. **Haselkorn, R., J. W. Golden, P. J. Lammers, and M. E. Mulligan.** 1986. Developmental rearrangements of cyanobacterial nitrogen-fixation genes. Trends Genet. **2**:255–259.
14. **Hayashida, N., T. Matsubayashi, K. Shinozaki, M. Sugiura, K. Inoue, and T. Hiyama.** 1987. The identification and characterization of the 9 kd polypeptide gene found in tobacco chloroplast DNA. Curr. Genet. **12**:247–250.
15. **Høj, P. B., I. Svendsen, H. V. Scheller, and B. L. Møller.** 1987. Identification of a chloroplast-encoded 9-kDa polypeptide as a 2[4Fe-4S] protein carrying centers A and B of photosystem I. J. Biol. Chem. **262**:12676–12684.
16. **Joerger, R. D., and P. E. Bishop.** 1988. Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii*. J. Bacteriol. **170**:1475–1487.
17. **Lammers, P. J., and R. Haselkorn.** 1983. Sequence of the *nifD* gene coding for the α subunit of dinitrogenase from the cyanobacterium *Anabaena*. Proc. Natl. Acad. Sci. USA **80**:4723–4727.
18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
20. **Mazur, B. J., and C.-F. Chui.** 1982. Sequence of the gene coding for the β-subunit of dinitrogenase from the blue-green alga *Anabaena*. Proc. Natl. Acad. Sci. USA **79**:6782–6786.
21. **Mevarech, M., D. Rice, and R. Haselkorn.** 1980. Nucleotide sequence of a cyanobacterial *nifH* gene coding for nitrogenase reductase. Proc. Natl. Acad. Sci. USA **77**:6476–6480.
22. **Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S.-I. Aota, H. Inokuchi, and H. Ozeki.** 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature (London) **322**:572–574.
23. **Rice, D., B. J. Mazur, and R. Haselkorn.** 1982. Isolation and physical mapping of nitrogen fixation genes from the cyanobacterium *Anabaena* 7120. J. Biol. Chem. **257**:13157–13163.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
25. **Schrautemeier, B., and H. Böhme.** 1985. A distinct ferredoxin for nitrogen fixation isolated from heterocysts of the cyanobacterium *Anabaena variabilis*. FEBS Lett. **184**:304–308.
26. **Shine, J., and L. Dalgarno.** 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.
27. **Tabor, S. and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
28. **Wynn, R. M., and R. Malkin.** 1988. Characterization of an isolated chloroplast membrane Fe-S protein and its identification as the photosystem I Fe-S_A/Fe-S_B binding protein. FEBS Lett. **229**:293–297.