Symbiosis, 30 (2001) 41–55
Balaban, Philadelphia/Rehovot

A Glutathione-Dependent Formaldehyde Dehydrogenase Gene of Symbiotic Cyanobacterium *Anabaena azollae*

WILLIAM H. SHAW[1*], ANDREW MUIRHEAD[1], TONY ARIOLI[2], and JACEK PLAZINSKI[2]
[1]*School of Biological and Environmental Sciences, Northern Territory University, Darwin NT 0909, Australia, Tel. +61-8-89466714, Fax. +61-8-89466847, E-mail. bill.shaw@ntu.edu.au;*
[2]*PCB, Research School of Biological Sciences, The Australian National University, GPO Box 475, Canberra ACT 2601, Australia*

Received August 20, 2000; Accepted February 26, 2001

Abstract

A 3.2 kb fragment of an indigenous *Anabaena azollae* plasmid was isolated and fully sequenced. Sequence analysis identified an open reading frame of 1110-bp that showed high similarity (56–71%) with glutathione dependent formaldehyde dehydrogenase genes (*gdfaldh*) from various bacteria. The identity of the gene was confirmed by expressing the gene in *Escherichia coli* with a pET-32 (a) vector followed by enzyme assay. Adjacent to the *gdfaldh* a second gene was detected that showed high similarity (53%) with an S-formylglutathione hydrolase (*fgh*) gene from the methylotrophic bacterium *Paracoccus denitrificans*. The presence of *gdfaldh* and *fgh*-like genes adjacent to each other may suggest that the corresponding gene products interact in a common metabolic pathway involved in removing exogenous or endogenous formaldehyde.

Keywords: *Anabaena azollae*, *Azolla*, symbiosis, plasmids, GDFALDH

*The author to whom correspondence should be sent.

0334-5114/2001/$05.50

1. Introduction

Azolla is an aquatic fern that grows on the surface of fresh water ponds, lakes or streams. *Azolla* has a symbiotic association with *Anabaena azollae*, a nitrogen fixing cyanobacterium, found inside a cavity in the dorsal leaf of *Azolla* and at the apex of the stem. In the symbiotic association *Azolla* provides nutrients, including fixed carbon and a protective cavity in each leaf for *Anabaena* colonies; in exchange *Azolla* receives fixed atmospheric nitrogen (Peters and Meeks, 1989). The two partners normally remain together throughout both vegetative and sexual reproduction and as yet it is not possible to culture a symbiotic *Anabaena azollae* (Braun-Howland and Nierzwicki-Bauer, 1990). The symbiosis can confer high rates of nitrogen fixation and biomass production; hence *Azolla/Anabaena* is an effective green manure for flooded crops and is used as a fertiliser in rice growing areas of South East Asia (Lumpkin and Pucknet, 1980).

The presence of plasmids in free-living cyanobacteria is well documented; however only a few coding functions have been determined (Muro-Pastor et al., 1994; Nicholson and Laudenbach, 1995). Plazinski et al. (1991) have shown that *A. azollae* has plasmid-borne genes with similar sequences to *Rhizobium nod* box and *nodMN*. These genes are essential for nodulation in the legume-*Rhizobium* symbiosis (Djordjevic and Weinmen, 1991).

Methylotrophic bacteria are aerobic bacteria that utilise one-carbon compounds more reduced than formic acid as sources of carbon and assimilate formaldehyde as a major source of cellular carbon (Hanson and Hanson, 1996). It is known that a number of bacteria are capable of methylotrophic growth by utilising methanol (Hanson and Hanson, 1996) and that the genes encoding glutathione dependent formaldehyde dehydrogenase (*gdfaldh*) and S-formylglutathione hydrolase (*fgh*) are essential for growth on methanol (Ras et al., 1995; Harms et al., 1996). For non-methylotrophic organisms it has been suggested that GDFALDH and FGH have a possible role in the detoxification of exogenous or endogenous formaldehyde (Fernandez et al., 1995) by reactions 1 (GDFALDH) and 2 (FGH);

1) Glutathione + HCHO + NAD –> S-formylglutathione + NADH + H^+
2) S-formylglutathione + H_2O –> Glutathione + HCOOH

We present here evidence that an indigenous plasmid of *A. azollae* contains two genes for, glutathione dependent formaldehyde dehydrogenase (*gdfaldh*), and S-formylglutathione hydrolase (*fgh*). This may suggests that the *A. azollae* plasmid-encoded genes are involved in removing exogenous or endogenous formaldehyde.

2. Material and Methods

Material

The strain of *Azolla pinnata* (Chinese isolate) was grown in liquid medium BG-11 as described previously (Plazinski et al., 1991), at 22–28°C with a 16 h light (with light intensity 40 to 45 klx): 8 h dark cycle and 85% humidity. *A. azollae* cells were isolated from surface sterilised *Azolla* fronds as described by Plazinski et al. (1991).

The construction of an A. azollae genomic library

High molecular weight genomic DNA was isolated from *A. azollae* according to the protocol of Plazinski et al. (1991). This protocol results in the separation of plasmid DNA from chromosomal DNA. For the purpose of making the genomic library both fractions were combined. The DNA was partially digested with *Sau*3A and size fractionated in sucrose gradients. The cloning vector Lambda DashII (Stratagene, USA) was digested with *Bam*HI and *Hind*III (Promega, Australia) restriction enzymes. The vector arms were then annealed and ligated with partial *Sau*3A fragments of *A. azollae* genomic DNA in the size range of 12 to 15 kb. The ligation products were packaged using Gigapack Gold II (Stratagene, USA) and the viable phage propagated in the DNA methylation tolerant *Escherichia coli* strain K803 (Wood, 1966).

Isolation of DNA probe and screening of lambda library

The DNA probe used in this study was the insert derived from a clone, pAf22, described by Plazinski et al. (1991). The pAf22 clone insert is a 2.7 kb *Eco*RI DNA fragment from the smallest indigenous plasmid (45 kD in size) of the *A. azollae* present in *Azolla filiculoides* strain 134 (Plazinski et al., 1991). The DNA insert was isolated from the plasmid vector purified by using GENECLEAN kit (Bresatec, Australia) then radiolabelled by using 25 μCi of [α^{32}P]dCTP (Amersham, Australia) with a Random Primer Kit (Promega, Australia). The lambda library was propagated on plates according to the procedure described by Sambrook et al. (1989). Plaques were transferred to nylon Hybond N^+ filters as described by the Amersham handbook (Amersham Corp., 1989) and probed with the pAf22 insert as described by the Amersham handbook (Amersham Corp., 1989). Plaques showing strong hybridization were isolated and purified by repeating the procedure until all plaques showed hybridisation signals. Southern blots to nylon Hybond N^+ were done as described by the Amersham handbook (Amersham Corp., 1989). Hybridizations were conducted at 65°C for 16 hrs followed by 4 standard

washes as described by the Amersham handbook (Amersham Corp., 1989). Purification of Lambda DNA was carried out using the Qiagen method (Qiagen Inc. handbook, 1990). Bacterial plasmid DNA was isolated by a quick alkaline lysis procedure as described by Sambrook et al. (1989).

DNA digestions, electrophoresis and cloning

Restriction endonucleases (*Eco*RI, *Hind*III, *Spe*I, *Sau*3A and *Bam*HI) were used as recommended by the manufacturer (Promega, Australia). DNA cloning and electrophoresis were done according to the procedures described in Sambrook et al. (1989). Agarose gel electrophoresis of indigenous plasmid DNAs was done by the method described by Plazinski et al. (1991).

Nucleotide sequencing

The Erase-a-base™kit (Promega, Australia) was used to generate, by treatment with exonuclease III deletion, a series of overlapping subclones in both directions. Plasmid DNA was prepared using the Qiagen method (Qiagen Inc. handbook, 1990). Sequence reactions were carried out as detailed in protocols supplied with the *Taq* Dye Primer Cycle Sequencing Kit (Applied Biosystems, Australia). Electrophoresis and collection of sequence data were carried out using an Applied Biosystems Model 370A automatic sequencer under conditions recommended by the supplier.

Computer analysis

The comparison of gene sequences was carried out at the National Centre for Biotechnology Information (NCBI) with the Blast network service. Sequence data were analysed using the Australian National Genomic Information Service (ANGIS).

The 1399 bp nucleotide sequence of the GDFALDH gene reported in this study has been lodged under the GenBank accession number U89767. The nucleotide sequence of the FGH gene (accession number AF035558) has been published elsewhere (Shaw et al., 1998).

Construction of a GDFALDH expression vector, pAnAldh

In order to express the *Anabaena azollae gdfaldh* gene in *Escherichia coli* two restriction sites, upstream (*Bam*HI) and downstream (*Hind*III) of the *A. azollae* coding sequence, were generated using the corresponding oligonucleotide sequences as primers for the polymerase chain reaction (PCR). These were the

5' primer 5'-GGGGGATCCTTGCAAGTTAAAGCGGCA-3' and the 3' primer 5'-GGGAAGCTTGGGTTAAAAAGTCACCACACTG-3'. PCR was carried out using a Perkin Elmer 480 thermal cycler according to the following program: initial denaturing at 94°C for 5 min, 34 cycles of 94°C for 45 sec, 55°C for 45 sec and 72°C for 90 sec, 1 cycle of 94°C for 45 sec, 55°C for 45 sec and 72°C for 10 min. The PCR-amplified *Bam*HI-*Hind*III restriction fragment of *A. azollae* DNA was ligated into the corresponding sites of the expression vector pET-32a (+) (Novogene, USA) to generate a recombinant GDFALDH expression vector designated pAnAldh. The GDFALDH recombinant protein was expressed in the *E. coli* strain BL21 (DE3) being induced by IPTG as described by the manufacture (Novogene, USA). Induced and non induced cells were harvested by centrifugation and stored at –70°C.

Protein isolation and assay of enzyme activity

Soluble *E. coli* proteins were isolated using the B-PER™ Bacterial Protein Extraction Reagent as described by the manufacturer (Pierce, USA). GDFALDH enzyme activity was assayed by following the generation of NADH at 340 nm. The one ml reaction mixture consisted of 120 mM sodium phosphate buffer (pH 8.0), 1 mM glutathione, 1 mM formaldehyde, 1 mM NAD and 100 µl of crude protein extract. Assays were conducted at 35°C. Activities were calculated by using a molar absorption for NADH of 6.22×10^3 M^{-1} cm^{-1}.

3. Results and Discussion

The screening of the *A. azollae* genomic library extracted from *Azolla pinnata* with the *Anabaena* plasmid, pAf22, DNA insert resulted in a single plaque showing strong hybridization. Following purification of the plaque, lambda DNA was isolated and the insert (14 kb) was re-cloned into the vector pBluescript (Stratagene, USA). This plasmid was designated p22. A restriction map of the 14 kb fragment is shown in Fig. 1. The pAf22 plasmid insert hybridized exclusively to the smaller indigenous plasmid of *A. azollae* extracted from *A. pinnata* (data not shown) and with the 2.7 *Eco*RI fragment at the 3'end of the p22 plasmid insert thus confirming isolation of 14 kb of the *A. azollae* plasmid. A series of subclones of the 14 kb fragment were constructed. Preliminary sequencing of the subclones indicated that one clone p2232 contained sequences similar to two known genes. The clone, p2232, containing a fragment of *A. azollae* plasmid (3.2 kb) was completely sequenced using *Exo*III-generated overlapping subclones in both directions. Conflicts in sequence data were resolved by making several subclones (p223221 and p223226; see Fig. 1) of p2232 and sequencing them from both ends.

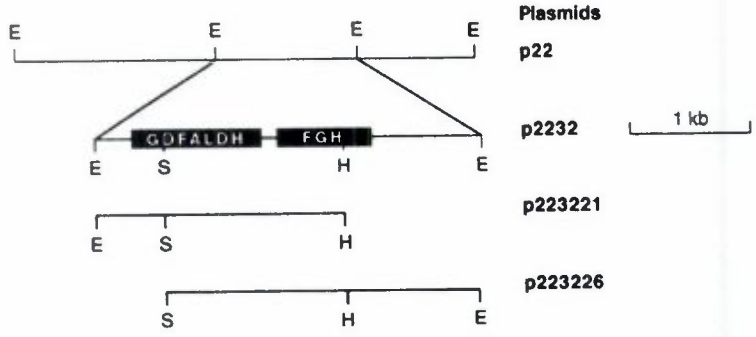

Figure 1. Restriction map of *Anabaena azollae* plasmid DNA inserts and location of genes identified by analysis of sequence data. E = *Eco*RI; H = *Hind*III; S = *Spe*I. GDFALDH = glutathione dependent formaldehyde dehydrogenase gene; FGH = S-formylglutathione hydrolase gene. Note the scale bar does not apply to p22.

Nucleotide sequence analysis

Analysis of sequence data revealed that the *Eco*R1 fragment of p2232 is 3261 base pairs long. Using Fickett's Testcode (Fickett, 1982), a pattern recognition program that measures the non-randomness of the base composition, a coding region approximately between 260 and 1360 nucleotides was detected. An open reading frame (ORF) was detected commencing with TTG (265), an alternative start codon for prokaryotes (Rice et al., 1992), and ending with the terminating codon TAA (1374). At eight bases (-8) prior to the start codon, there is a putative prokaryotic ribosome-binding site, AGGAGG (Shine and Delgarno, 1975). The nucleotide sequence and deduced amino acid sequence of the putative gene and its flanking nucleotide sequences are shown in Fig. 2. The codons TCG and CGA do not code for any of the 15 serine or 13 arginine residues and this is consistent with codon usage of other *Anabaena* genes (Tandeau de Marsac and Houmard, 1987). The G+C content, 41.1%, is also consistent with other reported Anabaena genes (Tandeau de Marsac and Houmard, 1987).

Amino acid sequence analysis

A search of all gene banks available on the NCBI Blast server network showed that the putative ORF had a very high similarity with a gene coding for glutathione-dependent formaldehyde dehydrogenase (*gdfaldh*) and found in a wide range of organisms. A dendogram of the 16 highest blastx scores (all GDFALDH/ class III alcohol dehydrogenase) is shown in Fig. 3. A comparison of the deduced amino acids of the *A. azollae* gene with the amino acid sequence of GDFALDHs from six other prokaryotes is given in Fig. 4. The alignment

```
AATTCTTGCCCCAGTGGAATTACACCCCTAAACCACAGTTACAGGTTATTTAATCCACTA      60

TCCTTAGTCTAAACTCCAGTGAAAAAATAAAGCCCGAGCACAAGGGGGGTTTCTCTTCTA     120

GGCATAGAAATTTACTATACAGATAGGGGATTTGACAAGCGCTCGCTGAAAAATTTTTAC     180

ATTTATCAACAAAAACAATATAAATTTATGTAACCATTTACTGATATACACTGATTGATA     240

          RBS           M  Q  V  K  A  A  V  A  Y  S  V  V
AGTTGCTATCAGGAGGAATATCTTTTGCAAGTTAAAGCGGCAGTAGCTTATAGTGTGGTT     300

 Q  P  L  T  I  E  T  V  E  L  E  G  P  Q  A  G  E  V  L  V
CAGCCCTTAACCATCGAAACGGTAGAACTTGAGGGACCACAAGCTGGGGAAGTATTAGTT     360

 E  I  K  A  S  G  V  C  H  T  D  A  Y  T  V  S  G  A  D  P
GAGATTAAAGCCAGCGGTGTTTGTCATACCGATGCTTACACGGTTTCTGGTGCTGATCCA     420

 E  G  L  F  P  A  I  L  G  H  E  G  A  G  V  V  V  E  V  G
GAAGGTTTATTCCCGGCGATTTTAGGACATGAAGGTGCGGGTGTGGTTGTAGAAGTGGGG     480

 A  D  V  K  S  V  K  P  G  D  H  V  I  P  L  Y  T  P  E  C
GCTGATGTAAAAAGCGTCAAACCCGGTGACCATGTCATTCCCTTATATACACCCGAATGT     540

 R  Q  C  E  Y  C  L  S  F  K  T  N  L  C  Q  A  I  R  V  T
CGTCAATGTGAATATTGTCTAAGTTTTAAAACTAACCTTTGTCAAGCCATTCGTGTGACT     600

 Q  G  R  G  L  M  P  N  A  T  S  R  F  S  I  V  G  K  M  I
CAAGGCAGGGGTTTAATGCCCAATGCTACTAGTCGTTTTAGTATCGTTGGCAAAATGATA     660

 H  H  Y  M  G  T  S  T  F  A  N  Y  T  V  L  P  E  K  A  V
CATCACTACATGGGAACATCAACCTTTGCTAATTATACGGTGTTACCAGAAAAAGCTGTG     720

 A  K  I  R  E  D  A  P  F  D  K  V  C  Y  I  G  C  G  V  T
GCTAAAATCCGTGAAGATGCGCCCTTTGATAAAGTTTGTTACATTGGCTGTGGAGTGACG     780

 T  G  I  G  A  V  I  Y  T  A  K  V  E  A  G  A  N  V  V  I
ACAGGTATAGGTGCAGTCATATATACAGCTAAAGTTGAAGCAGGTGCAAATGTAGTAATT     840

 S  G  L  G  G  I  G  L  N  I  I  Q  A  A  K  M  V  G  A  N
TCCGGTTTGGGCGGTATTGGTTTAAATATTATCCAAGCGGCAAAAATGGTAGGCGCAAAC     900

 M  I  V  G  V  D  I  N  P  K  K  R  A  L  A  E  K  L  G  M
ATGATTGTGGGAGTTGATATTAATCCTAAAAAACGCGCTTTAGCTGAAAAACTTGGCATG     960

 T  H  F  V  N  P  H  E  I  E  G  D  L  V  S  Y  L  I  D  L
ACACATTTTGTTAATCCCCATGAGATTGAAGGTGATTTAGTTTCTTATTTGATAGATTTA    1020

 T  K  G  G  A  D  Y  P  F  E  C  I  G  N  I  N  V  M  R  Q
ACTAAGGGTGGTGCAGATTATCCTTTTGAATGTATTGGCAATATCAATGTCATGCGTCAG    1080

 A  L  E  C  C  H  K  G  W  G  V  S  V  I  I  G  V  A  G  A
GCTTTAGAATGTTGTCATAAAGGTTGGGGTGTTAGTGTAATTATTGGAGTGGCAGGTGCA    1140

 G  Q  E  I  S  T  R  P  F  Q  L  V  T  G  R  V  W  K  G  S
GGACAAGAAATCAGCACTCGTCCATTTCAATTAGTAACTGGGCGCGTTTGGAAAGGTTCA    1200

 A  F  G  G  A  R  G  R  T  D  V  P  K  I  V  D  L  Y  M  N
GCTTTTGGTGGCGCAAGAGGCCGTACAGATGTGCCCAAAATTGTTGATTTATATATGAAT    1260

 G  Q  I  N  I  D  D  L  I  T  H  V  M  P  I  E  Q  I  N  H
GGGCAGATAAATATTGATGATTTAATCACTCATGTGATGCCAATTGAGCAAATTAATCAT    1320

 A  F  D  L  M  Y  R  G  E  S  I  R  S  V  V  T  F  *
GCCTTTGATTTAATGTACAGAGGAGAATCAATTCGCAGTGTGGTGACTTTTTAATAGGGA    1380

CATCCAATACATTTCAGTT     1399
```

Figure 2. Nucleotide sequence of the *Anabaena azollae gdfaldh* gene and deduced amino acid sequence of the protein. "RBS" is a potential ribosome binding site; * indicates stop codon: start codon is TTG.

Figure 3. Phylogenetic analysis of 17 glutathione dependent formaldehyde dehydrogenase (GDFALDH) amino acid sequences. The analysis was performed with the PAUP* program (Swofford, 1996). The neighbour joining method was used for the calculation. The numbers shown are the bootstrap values (1000 replicates). Sequences: P. denitrificans = *Paracoccus denitrificans* (P45382), R. sphaeroides = *Rhodobacter sphaeroides* (L02104), A. azollae = *Anabaena azollae* (U89767),

shows that 48% of the amino acids are conserved in all seven sequences and that this is increased to 66% when acceptable point mutations are considered. The deduced amino acid sequence is 71%, 71%, 58%, 58%, 57% and 56%, identical to *Paracoccus denitrificans, Rhodobacter sphaeroides, Haemophilus influenzae, Escherichia coli, Pasteurella piscicida* and *Synechocystis* sp., respectively. It is interesting to note that the amino acid sequence of GDFALDH from *P. denitrificans* and *R. sphaeroides* are highly homologous (88%) and both have an identical six amino acid insert at the same position, 265–270 (Fig. 4). The *P. denitrificans* gene has been confirmed by isolating the active protein and determining that the first 18 N-terminal amino acids are identical to those deduced from the sequenced gene (Ras et al., 1995). The *R. sphaeroides* gene has been confirmed by enzyme assay (Barber et al., 1996). The *E. coli* gene has been confirmed by enzyme assay of the expressed protein (Kummerle et al., 1996). In addition the first 47 amino acids of a purified GDFALDH gene product from *E. coli* (Gutheil et al., 1992) show strong similarity to the *E. coli* sequence (X73835) presented in Fig. 4. It should be noted that the *Synechocystis* sp., *H. influenzae* and *P. piscicida gdfaldh* genes were identified by their similarity to other reported *gdfaldh* genes. The *A. azollae* gene codes for a polypeptide of 369 amino acids with a calculated molecular weight of 39406 Daltons and this is similar to other reported prokaryotic GDFALDHs. The *E. coli* (Kummerle et al., 1996), and *P. piscicida* (Kim and Aoki, 1994) genes are both reported as plasmid-borne and both genes are the same length as the *A. azollae* and *Synechocystis* sp. genes.

A Block search (Henikoff and Henikoff, 1991) of the deduced amino acid sequence detected a zinc-containing alcohol dehydrogenase protein, prosite group (BLOOO59B). This is consistent with GDFALDH/class III alcohol dehydrogenase gene products. GDFALDHs are enzymes belonging to the zinc containing alcohol dehydrogenase family (Olle et al., 1994). These enzymes are reported as either dimers or tetramers with two zinc atoms per subunit (Vallee and Auld, 1990). One of the zinc atoms is essential for catalytic activity and is

Figure 3. Continuation.
E. coli = *Escherichia coli* (X73835), P. piscicida = *Paracoccus piscicda* (P39450), *Synechocystis* sp. (D90904), H. influenzae = *Haemophilus influenzae* (U32812), mouse A = *Mus musculus* (a60269), mouse B = *Mus musculus* (p28474), rat = *Rattus norvegicus* (p12711), human A = *Homo sapiens* (p11766), human B = *Homo sapiens* (g178132), horse = *Equus caballus* (p19854), lizard = *Uromastyx hardwickii* (p80467), octopus = *Octopus vulgaris* (a49662), fruit fly = *Drosophila melanogaster* (p46415), Arabidopsis = *Arabidopsis thaliana* (g1143388).

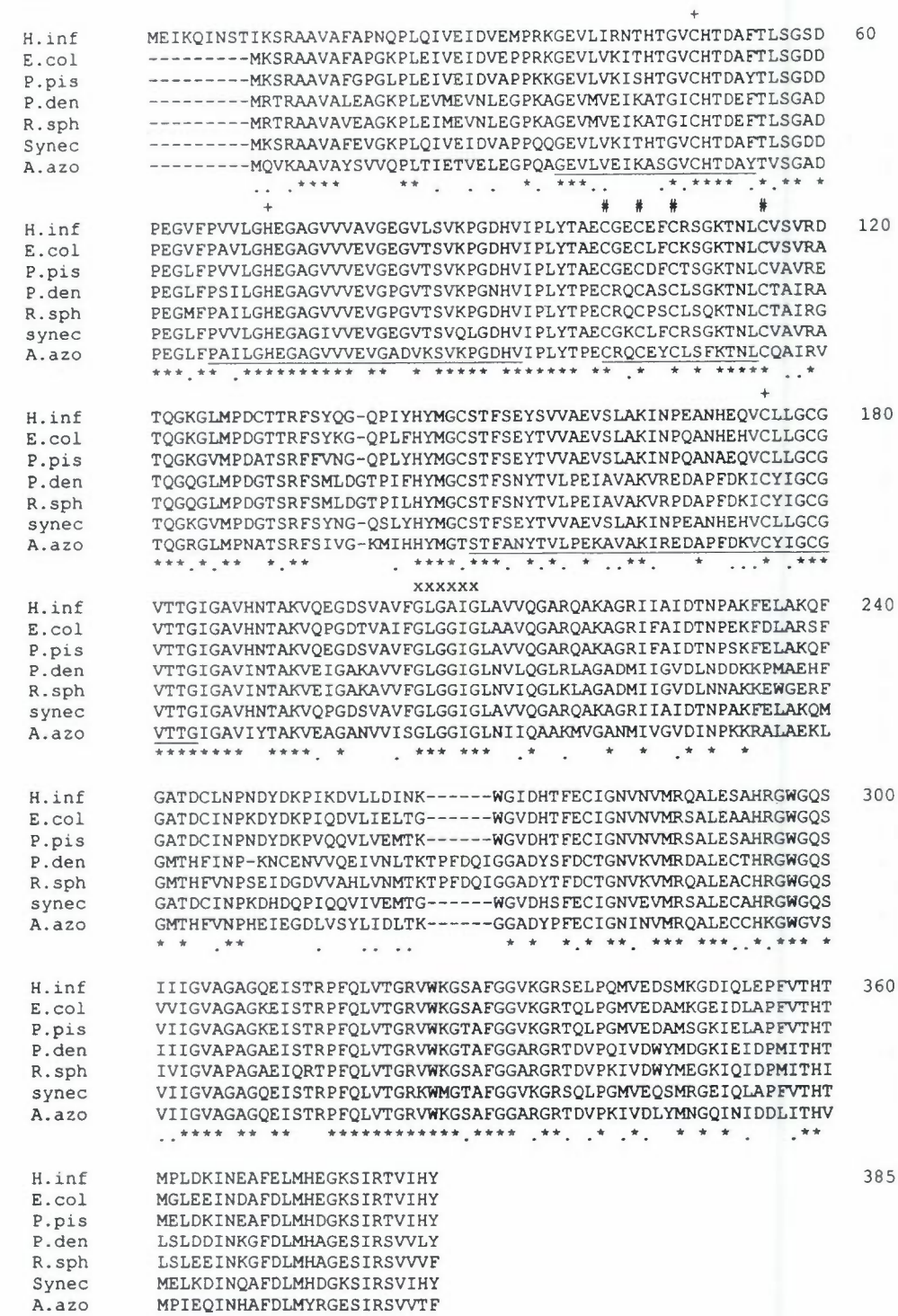
H.inf MEIKQINSTIKSRAAVAFAPNQPLQIVEIDVEMPRKGEVLIRNTHTGVCHTDAFTLSGSD 60
E.col ---------MKSRAAVAFAPGKPLEIVEIDVEPPRKGEVLVKITHTGVCHTDAFTLSGDD
P.pis ---------MKSRAAVAFGPGLPLEIVEIDVAPPKKGEVLVKISHTGVCHTDAYTLSGDD
P.den ---------MRTRAAVALEAGKPLEVMEVNLEGPKAGEVMVEIKATGICHTDEFTLSGAD
R.sph ---------MRTRAAVAVEAGKPLEIMEVNLEGPKAGEVMVEIKATGICHTDEFTLSGAD
Synec ---------MKSRAAVAFEVGKPLQIVEIDVAPPQQGEVLVKITHTGVCHTDAFTLSGDD
A.azo ---------MQVKAAVAYSVVQPLTIETVELEGPQAGEVLVEIKASGVCHTDAYTVSGAD
H.inf PEGVFPVVLGHEGAGVVVAVGEGVLSVKPGDHVIPLYTAECGECEFCRSGKTNLCVSVRD 120
E.col PEGVFPAVLGHEGAGVVVEVGEGVTSVKPGDHVIPLYTAECGECLFCKSGKTNLCVSVRA
P.pis PEGLFPVVLGHEGAGVVVEVGEGVTSVKPGDHVIPLYTAECGECDFCTSGKTNLCVAVRE
P.den PEGLFPSILGHEGAGVVVEVGPGVTSVKPGNHVIPLYTPECRQCASCLSGKTNLCTAIRA
R.sph PEGMFPAILGHEGAGVVVEVGPGVTSVKPGDHVIPLYTPECRQCPSCLSQKTNLCTAIRG
synec PEGLFPVVLGHEGAGIVVEVGEGVTSVQLGDHVIPLYTAECGKCLFCRSGKTNLCVAVRA
A.azo PEGLFPAILGHEGAGVVVEVGADVKSVKPGDHVIPLYTPECRQCEYCLSFKTNLCQAIRV
H.inf TQGKGLMPDCTTRFSYQG-QPIYHYMGCSTFSEYSVVAEVSLAKINPEANHEQVCLLGCG 180
E.col TQGKGLMPDGTTRFSYKG-QPLFHYMGCSTFSEYTVVAEVSLAKINPQANHEHVCLLGCG
P.pis TQGKGVMPDATSRFFVNG-QPLYHYMGCSTFSEYTVVAEVSLAKINPQANAEQVCLLGCG
P.den TQGQGLMPDGTSRFSMLDGTPIFHYMGCSTFSNYTVLPEIAVAKVREDAPFDKICYIGCG
R.sph TQGQGLMPDGTSRFSMLDGTPILHYMGCSTFSNYTVLPEIAVAKVRPDAPFDKICYIGCG
synec TQGKGVMPDGTSRFSYNG-QSLYHYMGCSTFSEYTVVAEVSLAKINPEANHEHVCLLGCG
A.azo TQGRGLMPNATSRFSIVG-KMIHHYMGTSTFANYTVLPEKAVAKIREDAPFDKVCYIGCG
xxxxxx
H.inf VTTGIGAVHNTAKVQEGDSVAVFGLGAIGLAVVQGARQAKAGRIIAIDTNPAKFELAKQF 240
E.col VTTGIGAVHNTAKVQPGDTVAIFGLGGIGLAAVQGARQAKAGRIFAIDTNPEKFDLARSF
P.pis VTTGIGAVHNTAKVQEGDSVAVFGLGGIGLAVVQGARQAKAGRIFAIDTNPSKFELAKQF
P.den VTTGIGAVINTAKVEIGAKAVVFGLGGIGLNVLQGLRLAGADMIIGVDLNDDKKPMAEHF
R.sph VTTGIGAVINTAKVEIGAKAVVFGLGGIGLNVIQGLKLAGADMIIGVDLNNAKKEWGERF
synec VTTGIGAVHNTAKVQPGDSVAVFGLGGIGLAVVQGARQAKAGRIIAIDTNPAKFELAKQM
A.azo VTTGIGAVIYTAKVEAGANVVISGLGGIGLNIIQAAKMVGANMIVGVDINPKKRALAEKL
H.inf GATDCLNPNDYDKPIKDVLLDINK------WGIDHTFECIGNVNVMRQALESAHRGWGQS 300
E.col GATDCINPKDYDKPIQDVLIELTG------WGVDHTFECIGNVNVMRSALEAAHRGWGQS
P.pis GATDCINPNDYDKPVQQVLVEMTK------WGVDHTFECIGNVNVMRSALESAHRGWGQS
P.den GMTHFINP-KNCENVVQEIVNLTKTPFDQIGGADYSFDCTGNVKVMRDALECTHRGWGQS
R.sph GMTHFVNPSEIDGDVVAHLVNMTKTPFDQIGGADYTFDCTGNVKVMRQALEACHRGWGQS
synec GATDCINPKDHDQPIQQVIVEMTG------WGVDHSFECIGNVEVMRSALECAHRGWGQS
A.azo GMTHFVNPHEIEGDLVSYLIDLTK------GGADYPFECIGNINVMRQALECCHKGWGVS
H.inf IIIGVAGAGQEISTRPFQLVTGRVWKGSAFGGVKGRSELPQMVEDSMKGDIQLEPFVTHT 360
E.col VVIGVAGAGKEISTRPFQLVTGRVWKGSAFGGVKGRTQLPGMVEDAMKGEIDLAPFVTHT
P.pis VIIGVAGAGKEISTRPFQLVTGRVWKGTAFGGVKGRTQLPGMVEDAMSGKIELAPFVTHT
P.den IIIGVAPAGAEISTRPFQLVTGRVWKGTAFGGARGRTDVPQIVDWYMDGKIEIDPMITHT
R.sph IVIGVAPAGAEIQRTPFQLVTGRVWKGSAFGGARGRTDVPKIVDWYMEGKIQIDPMITHI
synec VIIGVAGAGQEISTRPFQLVTGRKWMGTAFGGVKGRSQLPGMVEQSMRGEIQLAPFVTHT
A.azo VIIGVAGAGQEISTRPFQLVTGRVWKGSAFGGARGRTDVPKIVDLYMNGQINIDDLITHV
H.inf MPLDKINEAFELMHEGKSIRTVIHY 385
E.col MGLEEINDAFDLMHEGKSIRTVIHY
P.pis MELDKINEAFDLMHDGKSIRTVIHY
P.den LSLDDINKGFDLMHAGESIRSVVLY
R.sph LSLEEINKGFDLMHAGESIRSVVVF
Synec MELKDINQAFDLMHDGKSIRSVIHY
A.azo MPIEQINHAFDLMYRGESIRSVVTF

Figure 4.

coordinated by two cysteines and one histidine. The second zinc atom, which is not required for catalytic activity, is coordinated by either cysteine of histidine residues (Vallee and Auld, 1990). The deduced amino acid sequence presented in Fig. 4 shows possible catalytic zinc coordinating residues at cysteine (aa 49), histidine (aa 71), and cysteine (aa 175). Possible coordination sites for the second zinc atom are the cysteine residues at aa 101, 104, 107, and 115. In addition there is a conserved G-X-G-X-X-G region that is indicative of a putative nicotinamide adenine dinucleotide (NAD+) binding site (Bellamacina, 1996) and a conserved D (60) and R (119) residue, which have been reported, as essential for binding S-hydroxymethylglutathione (Holmquist et al., 1993; Engeland et al., 1993; Estonius et al., 1994).

GDFALDH enzyme activity

Only IPTG induced cells showed any measurable activity, 1.5×10^{-3} μmoles NADH/min. Non-induced cells, minus glutathione and minus formaldehyde control assays did not show any detectable activity.

Evidence for a second plasmid-borne gene

Further sequence analysis of clone p2232 using the Blast service located a stretch of DNA adjacent to *gdfalldh* that showed strong similarity with esterase (Shaw et al., 1998). The deduced amino acid sequence of the putative *A. azollae* gene was 64% identical and 78% similar to human esterase D (P10768). Similarly *E. coli*, (P33018), *H. influenzae* (P44556) and *Synechocystis* sp. (D90904) were 50%/66%, 45%/65% and 50%/63%, respectively. The deduced amino acid sequence also showed high similarity to S-formylglutathione hydrolase FGH (Shaw et al., 1998) from *P. denitrificans*

Figure 4. *A. azollae* GDFALDH deduced amino acid sequence and its alignment with other GDFALDH sequences. The analysis was performed with ClustalV (Higgins and Sharp, 1989). The conserved regions in all five sequences are indicated (*). Acceptable point mutations are indicated (.). The ligands to the zinc atom are indicated (+) for the catalytic zinc and (#) for the second zinc atom. The conserved prosite domains for zinc alcohol dehydrogenase are underlined. The domain involved in NAD binding is indicated (xxxxx). E. col = *Escherichia coli* (X73835), P. pis = *Pasteurella piscicida* (P39450), H. inf = *Haemophilus influenzae* (U32812), P. den = *Paracoccus denitrificans* (P45382), R. sph = *Rhodobacter sphaeroides* (L02104), Synec = *Synechocystis* sp. (D90904), A. azo = *Anabaena azollae* (U89767).

(53%/64%) (Harms et al., 1996). FGH is a thiol esterase that converts S-formylglutathione to glutathione and formate and it is interesting to note that Eiberg and Mohr (1986) and Apeshiotis and Bender (1986) have concluded that human esterase D (E.C. 3.1.1.1.) and S-formylglutathione hydrolase (E.C. 3.1.2.12) from human erythrocytes are identical. The reported esterase like gene from *H. influenzae*, *Synechocystis* sp. and *P. denitrificans* is in all cases close to their corresponding *gdfaldh* gene as found in *A. azollae* (Fig. 1). An analysis of the plasmid sequence data from *P. piscicida* (Kim and Aoki, 1994) showed that adjacent to the *gdfaldh* gene there is a partial esterase-like sequence that has been interrupted by a transposon sequence.

Possible role of plasmid-borne genes

The biochemical role of bacterial GDFALDH has not been generally established, however for *P. denitrificans* it has been demonstrated that mutants lacking the gene fail to grow on methanol, methylamine or choline thus implicating the enzyme in methylotrophic growth (Ras et al., 1995). The reported esterase-like genes from *H. influenzae* and *Synechocystis* sp. were both identified by their strong similarity with human esterase D gene. However, the esterase-like gene product from *P. denitrificans* has been shown to be the specific esterase FGH and *P. denitrificans* mutants lacking the gene fail to grow on methanol or methylamine thus indicating the enzyme is essential for methylotrophic growth. The sequence analysis of the *A. azollae* plasmid DNA fragment strongly suggests that the fragment encodes for two gene products the GDFALDH and FGH. The presence of *gdfaldh* and *fgh* adjacent to each other on the *A. azollae* plasmid may suggest that the corresponding gene products interact in a metabolic pathway involved in removing exogenous or endogenous formaldehyde.

The oxidation of formaldehyde to formate coupled with the formation of NADH could be utilised for energy generation via an electron transport chain. Formic acid is further oxidised to carbon dioxide that could be further utilised in the Calvin Benson cycle as shown in *P. denitrifcans* (Hanson, 1992).

For many living organisms formaldehyde is formed by the oxidation of methanol. Methanol is produced by plants and emitted to the atmosphere (Fall and Benson, 1996). The emitted methanol can support methylotrophic bacteria (Holland and Polacco, 1994). A possible way of methanol production in *Azolla* leaves is by pectin demethylation via pectic methylesterase within the cell walls (Fall and Benson, 1996). We speculate that methanol produced by the rapidly growing *Azolla* leaves, (*Azolla* is known to be capable of doubling its biomass every 3–4 days) is utilised by *A. azollae* as a possible energy source. The genus *Anabaena* has not been reported as methylotrophs but the presence of

a plasmid encoded *gdfaldh* and *fgh* genes may suggest that the cyanobacterium has acquired methylotrophic ability in order to utilise methanol produced by *Azolla*.

In summary, this paper describes the first reported isolation, sequencing and expression of a glutathione-dependent formaldehyde dehydrogenase (*gdfaldh*) gene from an indigenous plasmid of *A. azollae*. This conclusion is based on enzyme assays, Testcode and Prosite analysis together with Blast searches and a close alignment between the *A. azollae gdfaldh* gene and other bacterial gdfaldh genes (Fig. 4). The alignment presented in Fig. 4 shows many features that are characteristic for the GDFALDH-encoded genes from various sources and indicates an overall similarity of 48%, and allowing for acceptable mutations a similarity of 66%. The observation that several reported bacterial *gdfaldh* genes are adjacent to an esterase-like gene has led to the conclusion that both gene products are involved in a common metabolic pathway, the oxidation of endogenous or exogenous formaldehyde to formate.

Acknowledgments

This work was supported by the Australian Research Council Grant No. A09030443 and the Northern Territory University Research Committee.

REFERENCES

Amersham Corp. Handbook. 1989. Amersham Corp., Amersham, England.

Apeshiotis, F. and Bender, K. 1986. Evidence the S-formylglutathione hydrolase and esterase D polymorphisms are identical. *Human Genetics* **74**: 176–177.

Barber, R.D., Rott, M.C., and Donohue, T.J. 1996. Characterization of a glutathione-dependent formaldehyde dehydrogenase from *Rhodobacter sphaeroides*. *Journal of Bacteriology* **178**: 1386–1393.

Bellamacina, C.R. 1996. The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *Journal of the Federation of American Societies for Experimental Biology* **10**: 1257–1269.

Braun-Howland, E.B. and Nierzwicki-Bayer, S.A. 1990. *Azolla Anabaena* symbiosis: Biochemistry, ultrastructure, and molecular biology. In: *Handbook of Symbiotic Cyanobacteria*. A.N. Rai, ed. CRC Press Boca Raton, Florida, pp. 65–117.

Djordjevic, M.A. and Weinmen, J.J. 1991. Factors determining host recognition in the clover-*Rhizobium*. *Australian Journal of Plant Physiology* **18**: 543–557.

Eiberg, H. and Mohr, J. 1986. Identity of the polymorphisms for esterase D and S-formylglutathione hydrolase in red blood cells. *Human Genetics* **74**: 174–300.

Engeland, K., Hoog, J., Holmquist, B., Estonius, M., Jornvall, H., and Vallee, B. L. 1993. Mutation of Arg-115 of human class III alcohol dehydrogenase: a binding site required for formaldehyde dehydrogenase activity and fatty acid activation. *Proceedings of the National Academy of Science USA* **90**: 2491–1994.

Estonius, M., Hoog, J., Danielsson, O., and Jornvall, H. 1994. Residues specific for class III alcohol dehydrogenase. Site-directed mutagenesis of the human enzyme. *Biochemistry* **33**: 15080–15085.

Fall, R. and Benson, A. 1996. Leaf methanol – the simplest natural product from plants. *Trends in Plant Sciences* **1**: 296–300.

Fernandez, M.R., Biosca, J.A., Norin, A., Jornvall, H., and Pares, X. 1995. Class III alcohol dehydrogenase from *Saccharomyces cerevisiae*: structural and enzymatic features differ towards the human/mammalian forms in a manner consistent with functional needs in formaldehyde detoxication. *FEBS Letters* **370**: 23–26.

Fickett, J.W. 1982. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Research* **10**: 5303–5318.

Gutheil, W.G., Holmquist, B., and Vallee, B.L. 1992. Purification, characterization, and partial sequence of the glutathione-dependent formaldehyde dehydrogenase from *Escherichia coli*. *Biochemistry* **31**: 475–481.

Hanson, R.S. 1992. Introduction. In: *Methane and Methanol Utilizers*. J.C. Murrell and H. Dalton, eds. Plenum Press, New York, pp. 1–21.

Hanson, R.S. and Hanson, T.E. 1996. Methanotrophic bacteria (review). *Microbiological Reviews* **60**: 439–471.

Harms, N., Ras. J., Rreijnders, W.N.M., van Spanning, R.J.M., and Stouthamer, A.H. 1996. S-formylglutathione hydrolase of *Paracoccus denitrificans*. *Journal of Bacteriology* **178**: 6296–6299.

Henikoff, S. and Henikoff, J.D. 1991. Automated assembly of protein for database searching. *Nucleic Acids Research* **19**: 6565–6572.

Higgins, D.G. and Sharp, P.M. 1989. Fast and sensitive multiple sequence alignment on a microcomputer. *CABIOS* **5**: 151–153.

Holland, M.A. and Polacco, J.C. 1994. PPFMs and other covert contaminants: Is there more to plant physiology than just plants? *Annual Review of Plant Physiology and Plant Molecular Biology* **45**: 197–209.

Holmist, B., Moulis, J., Engeland, K., and Vallee, B.L. 1993. Role of arginine 115 in fatty acid activation and formaldehyde dehydrogenase activity in human classIII alcohol dehydrogenase. *Biochemistry* **32**: 5139–5144.

Kim, E. and Aoki, T. 1994. The transposon-like structure of IS26-tetracycline, and kanamycin resistance determinant derived from transferable R plasmid of fish pathogen, *Pasteurella piscicida*. *Microbiological Immunology* **38**: 31–38.

Kummerle, N., Feubert, H., and Kaulfers, P. 1996. Plamid-mediated formaldehyde resistance in *Escherichia coli*. *Antimicrobiological Agents in Chemotherapy* **40**: 2276–2279.

Lumpkin, T.A. and Plucknet, D.L. 1980. *Azolla*: botany, physiology and use as green manure. *Economic Botany* **34**: 111–153.

Muro-Pastor, A.M., Kuritz,T., Flores, E., Herrero, A., and Wolk, C.P. 1994. Transfer of a genetic marker from a megaplasmid of *Anabaena* sp. strain PCC 7120 to a megaplasmid of a different *Anabaena* sp. strain. *Journal of Bacteriology* **176**: 1093–1098.

Nicholson, M.L. and Laudenbach, D.E. 1995. Genes encoded on a cyanobacterial plasmid are transcriptionally regulated by sulfur availability and CysR. *Journal of Bacteriology* **177**: 2143–2150.

Olle, D., Atrian, S. Luque, T., Hjelmqvist, L., Gonzalez-Duarte, R., and Jornvall, H. 1994. Fundamental molecular differences between alcohol dehydrogenase classes. *Proceedings of the National Academy of Science USA* **91**: 4980–4984.

Peters, G.A. and Meeks, J.C. 1989. The *Azolla-Anabaena* symbiosis: Basic biology. *Annual Review of Plant Physiology and Plant Molecular Biology* **53**: 813–819.

Plazinski, J., Croft, L., Taylor, R., Zheng, Q., Rolfe, B.G., and Gunning, B.S.E. 1991. Indigenous plasmids in *Anabaena azollae*: their taxonomic distribution and existence of regions of homology with symbiotic genes of *Rhizobium*. *Canadian Journal of Microbiology* **37**: 171–181.

Qiagen Inc. Handbook. 1990. The Qiagenologist Application Protocols. 3rd Ed., Qiagen Inc. Chatsworth, California, USA.

Ras, J., Van Ophem, P.W., Reijnders, W.N.M., Van Spanning, R.J.M., Duine, J.A., Stouthamer, A.H., and Harms, N. 1995. Isolation, sequencing and mutagenesis of the gene encoding NAD-and glutathione-dependent formaldehyde dehydrogenase (GD-FALDH) from *Paracoccus denitrificans*, in which GD-FALDH is essential for methylotrophic growth. *Journal of Bacteriology* **177**: 247–251.

Rice, P.M., Elliston, K., and Gribskov, M. 1992. Sequence Project Management. In: *Sequence Analysis Primer*. M. Gribskov and J. Devereux, eds. Freeman Press, New York, pp. 1–59.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning A Laboratory Manual*. 2nd Ed., Cold Spring Harbour Lab. Press, New York.

Shaw, W.H., Arioli, T., and Plazinski, J. 1998. Cloning and sequencing of a S-formylglutathione hydrolase (FGH) gene from the cyanobacterium *Anabaena azollae* (accesion No AF035558) (PGR98-024). *Plant Physiology* **116**: 868.

Shine, J. and Dalgarno, L. 1975. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**: 34–38.

Swofford, D.L. 1996. PAUP: Phylogenic analysis using parsimony (and other methods), version 4.0.0.d60. Sinauer, Sunderland, Massachusetts.

Tandeau de Marsac, N. and Houmard, J. 1987. In: *The Cyanobacteria*. P. Fay, and C.Van Baalen, eds. Elsevier Press, Amsterdam, pp. 251–302.

Vallee, B.L. and Auld, D.S. 1990. Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry* **29**: 5647–5659.

Wood,W.B. 1966. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affected by restriction and modification of DNA. *Journal of Molecular Biology* **16**: 118–132.