Expression of a gene encoding a novel ferredoxin in the cyanobacterium *Synechococcus* 6301

Alison L. COZENS* and John E. WALKER†

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

A gene was discovered in the cyanobacterium *Synechococcus* 6301 that encodes a protein highly related to members of the [2Fe-2S] ferredoxin family found in chloroplasts and cyanobacteria. It follows a cluster of seven genes encoding subunits of the cyanobacterial ATP synthase complex. It is transcribed as a monocistronic mRNA of 408 nucleotide residues. Transcription starts at a site 55 bp upstream of the initiator methionine codon. Transcriptional initiation and termination signals with sequences similar to those found in *Escherichia coli* are not present. Comparison of the predicted sequence of the ferredoxin protein with those of other cyanobacterial and plant ferredoxins shows an average sequence identity of about 40 %. Twelve amino acid residues are invariant, including the four cysteine residues that provide ligands for the [2Fe-2S] cluster. The deduced *Synechococcus* ferredoxin protein sequence has a C-terminal extension of eight amino acid residues relative to most other 2Fe-2S ferredoxins except for those from halobacteria, which also have a C-terminal extension. The sequence of the *Synechococcus* protein is most closely related to ferredoxins from the two complex cyanobacteria *Chlorogloeopsis fritschii* and *Mastigocladus laminosus*. The deduced protein sequence is not that of the major soluble ferredoxin that has been isolated from *Synechococcus* 6301 and is reported in the accompanying paper [Wada, Masui, Matsubara & Rogers (1988) *Biochem. J.* 252, 571–575]. So it appears to be a novel [2Fe-2S] ferredoxin and *Synechococcus* 6301 contains at least two [2Fe-2S] ferredoxins, which may have different roles *in vivo*.

INTRODUCTION

Ferredoxins are non-haem iron–sulphur proteins that transfer electrons in a wide variety of metabolic reactions (Cammack *et al.*, 1985). The best defined role of [2Fe-2S] ferredoxins is as an electron carrier between photosystem I and ferredoxin–NADP⁺ reductase (Shin & Arnon, 1965), but they are known to participate in a number of other important reactions, including nitrogen fixation, nitrate and nitrite reduction, sulphite reduction, glutamate synthesis, reductive carboxylic acid cycle and fatty acid metabolism (reviewed by Cammack *et al.*, 1981). They can be classified into groups based both on their contents of iron and inorganic sulphur atoms and on their amino acid sequences. For example, chloroplast and cyanobacterial ferredoxins, the subject of the present work, have a [2Fe-2S] cluster and are acidic proteins of related sequence with an *M_r* of about 11 000 (Matsubara & Hase, 1983). Sequences of ferredoxins have been determined in a wide range of plants and cyanobacteria (Matsubara & Hase, 1983), and high-resolution structures have been determined by X-ray crystallography of the proteins from two cyanobacteria, *Spirulina platensis* (Fukuyama *et al.*, 1980; Tsukihara *et al.*, 1981) and *Aphanothece sacrum* (Tsutsui *et al.*, 1983). These studies demonstrate that the three-dimensional structures of the two proteins are very similar and that many of the most highly conserved amino acid residues are involved in forming the iron–sulphur pocket. These include four conserved cysteine residues that provide ligands for the two iron atoms.

All plants and cyanobacteria appear to contain a major soluble [2Fe-2S] ferredoxin, ferredoxin I, but many species contain a second ferredoxin, which can be distinguished from the former by amino acid sequence (Hase *et al.*, 1976, 1977b, 1982; Dutton *et al.*, 1980) and may also differ in charge and midpoint redox potential (Cammack *et al.*, 1977; Hutson *et al.*, 1978) or in their activity in biological reactions *in vitro* (Dutton *et al.*, 1980; Hutson *et al.*, 1978; Wada *et al.*, 1981). It has been suggested that where two or more ferredoxins are present in the same organism they may perform different functions in the cell (Shin *et al.*, 1977). In the case of the two soluble ferredoxins isolated from natural populations of *Microcystis aeruginosa*, the relative proportions of ferredoxins I and II depended upon growth conditions, and both were replaced by flavodoxin under conditions of iron deficiency (Cohn *et al.*, 1985).

As reported in the present paper, in the course of studies of two operons encoding subunits of ATP synthase in the cyanobacterium *Synechococcus* 6301 (Cozens & Walker, 1987) we have discovered that a gene following one operon encodes a protein that is highly homologous to the [2Fe-2S] ferredoxin family. However, this protein is not the major soluble ferredoxin isolated from this micro-organism. In order to prove that the gene is expressed, we have demonstrated the presence of the corresponding mRNA in the organism, and, by further transcript mapping experiments, the site of initiation of transcription has been identified. However, the function of the protein is unknown. Its predicted protein sequence contains 12 amino acid residues that

* Present address: Department of Microbiology and Immunology, University of California at San Francisco, San Francisco, CA 94143, U.S.A.

† To whom correspondence should be addressed.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00725.

are maintained in all known ferredoxins in the [2Fe-2S] family, among them the four cysteine residues that provide the ligands for the iron-sulphur centre. The predicted protein has one unusual feature, a C-terminal extension of eight amino acid residues. This has not been found previously in the chloroplast and cyanobacterial ferredoxins, but a similar feature has been noted previously in [2Fe-2S] ferredoxins from halobacteria. An unusual membrane-bound ferredoxin has been described in the cyanobacterium *Aphanizomenon flos-aquae* (Cohn *et al.*, 1985). This protein has a C-terminal extension of six amino acid residues when compared with ferredoxin I from the same species.

MATERIALS AND METHODS

Cloning and DNA sequence analysis

A 1.6 kb *KpnI* fragment hybridizing to the C-terminal part of the gene for the γ -subunit of *Synechococcus* 6301 ATP synthase was isolated from the bacteriophage λ S5, a λ 2001 recombinant containing a 20 kb fragment of *Synechococcus* genomic DNA (Cozens & Walker, 1987). This DNA fragment was purified from a *KpnI* digest of the λ recombinant by electrophoresis on low-melting-point (LMP) agarose (Wieslander, 1979). Portions of the DNA fragment were digested with the restriction enzymes *AluI* and *RsaI*; both recognize a 4 bp restriction site, and cut leaving a blunt-ended fragment. These fragments were cloned into M13mp8 that had previously been cut with *SmaI*. Sequences in M13 clones were determined by the dideoxy chain-termination method of Sanger *et al.* (1977) as modified by Biggin *et al.* (1983). The sequence was completed by using specific sequencing primers 17 nucleotide residues in length, made with an Applied Biosystems 380B DNA synthesizer. Compressions were resolved by replacing dGTP by dITP (Mills & Kramer, 1979) in the sequencing reactions. The DNA sequence was determined at least once on each strand of the DNA.

Data analysis

DNA sequences were compiled in a database by using the computer program DBAUTO (Staden, 1982b) and were analysed with ANALYSEQ (Staden, 1985). A modified version of the positional-base-preference method (Staden, 1984) was used to predict the positions of open reading frames (potential genes). The protein sequences deduced from DNA sequences were compared with the database of the Protein Information Resource (P.I.R., National Biomedical Research Foundation, Georgetown University Medical Center, N.W. Washington, DC, U.S.A.) by using the program FASTP (Lipman & Pearson, 1985). Homologous sequences were analysed further by pairwise comparison with DIAGON (Staden, 1982a).

DNA hybridization

High- M_r genomic DNA was prepared by the method of Marmur (1961) from cells of *Synechococcus* 6301 harvested during exponential growth (Cozens & Walker, 1987). Portions (0.5 μ g) of this DNA were digested with restriction endonucleases and fractionated on a 0.8% agarose gel (200 ml) by electrophoresis for 18 h at 25 mA. DNA was transferred from the gel to nitrocellulose according to the method of Southern (1975). Filters were prehybridized at 65 °C for 1 h in a solution

containing 6 $\times$ SSC (0.9 M-NaCl/0.09 M-sodium citrate), 0.1% Ficoll, 0.1% bovine serum albumin (fraction V), 0.1% polyvinylpyrrolidone, 0.5% Sarkosyl and yeast RNA (50 μ g/ml). Hybridizations were carried out in more of the same solution to which had been added 10% (w/v) dextran sulphate and a single-stranded 'prime-cut' probe (Farrell *et al.*, 1983).

The hybridization probe was prepared from an M13mp8 bacteriophage as described by Biggin *et al.* (1983), except that the labelled fragment was liberated by digestion with *EcoRI*. The fragment employed covers nucleotide residues 7711 to 8148 in Fig. 1, and includes the whole of the proposed ferredoxin gene. Filters were washed three times in a solution of 6 $\times$ SSC, under non-stringent conditions, each for 20 min, at 65 °C. They were autoradiographed with pre-flashed film and an intensifying screen at -70 °C for 2-8 days.

RNA preparation

Synechococcus 6301 was grown photosynthetically as described previously (Cozens & Walker, 1987). It was harvested by centrifugation at the mid-exponential stage of growth. Cells (1 g wet wt.) were resuspended in lysis buffer (0.15 M-NaCl/0.1 M-EDTA/0.05 M-Tris/HCl buffer, pH 8.0) (10 ml) containing lysozyme (1 mg/ml) and incubated at 37 °C for 1 h. SDS (10%, w/v) was added to a final concentration of 1% and proteinase K to 1 mg/ml, and the cells were incubated at 60 °C for 10 min. They were cooled briefly on ice, and NaClO₄ (5 M) was added to give a final concentration of 1 M. The mixture was extracted with an equal volume (12 ml) of chloroform/isopentanol-1-ol (24:1, v/v), then with phenol (12 ml) equilibrated with TE buffer (10 mM-Tris/HCl buffer, pH 8.0, containing 0.1 mM-EDTA). The nucleic acids were precipitated with ethanol and resuspended in TE buffer (1 ml). The DNA was removed by incubation at 37 °C for 30 min with RQ1 RNAase-free DNAase 1 (Promega Biotech, Madison, WI, U.S.A.) (10 units) followed by phenol extraction and ethanol precipitation as before. The RNA pellet was dissolved in TE buffer (100 μ l).

S1-nuclease mapping

Single-stranded probes complementary to the ferredoxin mRNA were prepared by the 'prime-cut' method (Farrell *et al.*, 1983). Probes (50000 c.p.m.) and *Synechococcus* RNA (1 μ g) were co-precipitated with ethanol and resuspended in 80% (v/v) formamide annealing buffer (Berk & Sharp, 1978). The RNA was denatured at 80 °C for 10 min and then hybridized at 45 °C for 16 h. The RNA-DNA duplexes were digested with S1 nuclease (Bethesda Research Laboratories, Gaithersburg, MD, U.S.A.) (800 units) in a solution (400 μ l) containing 30 mM-sodium acetate buffer, pH 4.5, 0.25 M-NaCl, 1 mM-ZnSO₄ and 5% (v/v) glycerol, by incubation at 15 °C for 1 h. The nucleic acids were precipitated with ethanol and analysed by electrophoresis in a 6% polyacrylamide gel containing 7 M-urea.

Primer extension analysis

'Prime-cut' probes (50000 c.p.m.) were prepared and annealed to *Synechococcus* RNA as described for S1-nuclease mapping. The annealed mixture was precipitated with ethanol and resuspended in a solution (10 μ l) containing 0.1 M-Tris/HCl buffer, pH 8.3, 0.15 M-KCl, 10 mM-MgCl₂, 10 mM-dithiothreitol and 1 mM each of

Fig. 1. DNA sequence of a segment of the *Synechococcus* 6301 genome

It encodes the 3' end of the gene encoding the γ -subunit of ATP synthase, a gene encoding a protein homologous to [2Fe-2S] ferredoxins, and the 5' end of an unidentified open reading frame, URF2. The boxed sequence is a potential ribosome-binding site (Shine & Dalgarno, 1974), and the two opposed horizontal broken arrows indicate the position of a possible rho-factor-independent transcription terminator able to form a stem-loop structure. The vertical arrows indicate the transcriptional initiation site at position 7733 and the termination site at position 8141. The numbering starts at a *Sau*3A site at the 5' end of the fragment of *Synechococcus* 6301 DNA inserted in λ S1 (Cozens, 1986; Cozens & Walker, 1987).

Fig. 2. Arrangement of genes in the locus containing the ferredoxin gene

The letter F denotes the gene encoding ferredoxin; a, b, c, α , γ and δ indicate ATP synthase subunits encoded in the genes. b' is a diverged and duplicated form of b. I is related to the *E. coli* *uncl* gene (Gay & Walker, 1981). URF2 encodes another protein of unknown function.

the four unlabelled deoxynucleoside triphosphates dATP, dCTP, dGTP and dTTP. The probe was extended with reverse transcriptase (Super RT; Anglian Biotechnology, Colchester, Essex, U.K.) (5 units) for 30 min at 42 °C. The reaction products were analysed by electrophoresis in the presence of 7 M-urea in a 6% polyacrylamide gel.

RESULTS AND DISCUSSION

DNA sequence and gene arrangement

The DNA sequence shown in Fig. 1 has been reported previously, but without description and comment (Cozens & Walker, 1987). It is part of a continuous sequence of 8890 bp containing an operon (*atp1*) encoding seven of the subunits of ATP synthase (see Fig. 2). A complete open reading frame (bases 7788–8101) was strongly predicted by the positional-base-preference method (Fig. 3). This potentially encodes an acidic protein of 105 amino acid residues with an M_r of 11132. It is preceded by a sequence similar to an *Escherichia coli* ribosome-binding site (Shine & Dalgarno, 1974). It is followed by an incomplete open reading frame, URF2 (starting at base 8265), of unknown function.

Fig. 3. Gene predictions performed on the *Synechococcus* 6301 DNA by the positional-base-preference method of Staden (1984)

The three boxes represent the translational reading frames of the DNA. The scale on each ordinate represents the relative probability of coding. Points above the midline are likely to be protein-coding, points below are not. Vertical bars on the abscissae represent possible ATG start codons, those on the midlines potential stop codons.

Fig. 4. S1-nuclease mapping of the 5' end of the ferredoxin mRNA

Tracks marked M contain a *HpaII* digest of pBR322 DNA labelled with [α - 32 P]dCTP. The sizes of marker bands are given in bp on the left. Probes 1 and 2 refer to the probes described in the text. Tracks a contain probe alone. Tracks b contain probe digested with S1 nuclease in the absence of RNA, tracks c probes digested with S1 nuclease in the presence of *Synechococcus* RNA. The sizes of protected fragments are given on the right in nucleotide residues.

The genes in the *atp1* operon are co-transcribed and transcription terminates around nucleotides 7718–7726 (see Fig. 1), immediately following the gene for the γ -subunit (A. L. Cozens & J. E. Walker, unpublished work). This region contains an inverted repeat sequence in the DNA, which could form a stem-loop structure in the mRNA, followed by a run of T residues. This is similar to the rho-factor-independent transcriptional terminators found in *E. coli* (Rosenberg & Court, 1979). Thus the ferredoxin gene appears to be transcribed separately from the ATP synthase genes.

Transcription of the ferredoxin gene

In order to determine the location of transcriptional initiation of the ferredoxin gene, S1-nuclease mapping experiments were performed. Two internally radioactively labelled single-stranded DNA molecules complementary to the mRNA were employed as hybridization probes. They have a common 3' end and differ in length; they are referred to as probe 1 (nucleotides 7711–7877) and probe 2 (nucleotides 7711–7900). The sizes of labelled DNA fragments protected by the ferredoxin mRNA were determined by comparison with marker fragments of known length. Thereby, the transcriptional initiation

Fig. 5. Primer-extension analysis of the 5' end of the ferredoxin mRNA

Track M contains the marker as in Fig. 4. Track a contains the probe 3 treated with reverse transcriptase in the absence of RNA, track b contains probe 3 extended with reverse transcriptase in the presence of *Synechococcus* RNA. The size of the largest extended product is given on the right in nucleotide residues.

site was located between nucleotides 7727 and 7733 (Fig. 4). A more accurate and independent determination of the transcriptional start site was provided by primer-extension analysis. As shown in Fig. 5, probe 3 (nucleotides 7786–7976) was extended by 53 nucleotide residues with reverse transcriptase and so the transcriptional initiation site is located precisely at the G residue at nucleotide 7733.

The DNA sequence upstream of the start site of the ferredoxin mRNA does not show any obvious homology to *E. coli* promoter consensus sequences (Pribnow, 1978). This is in contrast with the sequences found upstream of several other cyanobacterial genes (Tumer *et al.*, 1983; Curtis & Haselkorn, 1984; Nierzwicki-Bauer *et al.*, 1984; Mulligan *et al.*, 1984; Shinozaki & Sugiura, 1985), including the *atp* operons of *Synechococcus* 6301 (Cozens & Walker, 1987), where sequences homologous to *E. coli* –10 boxes are present. However, it is clear that, although some cyanobacterial genes are preceded by promoter sequences that can be recognized by *E. coli* RNA polymerase (Bryant *et al.*, 1985), others use promoters unrelated to *E. coli* promoter sequences, for example the genes encoding proteins involved in nitrogen fixation and assimilation in *Anabaena* 7120 (Tumer *et al.*, 1983).

Fig. 6. S1-nuclease mapping of the 3' end of ferredoxin mRNA

Tracks M contain marker as in Fig. 4. The probes 4 and 5 are described in the text. Tracks a contain probe alone, tracks b probe digested with S1 nuclease in the absence of RNA and tracks c probes digested with S1 nuclease in the presence of *Synechococcus* RNA. The sizes of protected fragments are given on the right in nucleotide residues.

The region between the 3' end of the ferredoxin gene and the 5' end of URF2 contains neither inverted repeat sequences nor sequences recognizable as potential promoters. However, S1-nuclease mapping indicates that the ferredoxin gene is not co-transcribed with URF2. As shown in Fig. 6, probe 4 (nucleotides 7972–8300) gives rise to a fragment of 170 nucleotide residues that is protected by *Synechococcus* mRNA. However, probe 5 (nucleotides 7972–8148) is digested by S1 nuclease to give protected fragments ranging in size from approx. 160 to 175 nucleotide residues presumably including the 170-nucleotide-residue fragment protected by probe 4. It seems likely that probe 5 forms a less stable duplex with the ferredoxin mRNA, so that the nuclease can nibble into the DNA–RNA duplex. These results are therefore consistent with termination of the ferredoxin gene transcription at nucleotide 8141.

Number of ferredoxins in *Synechococcus* 6301

The major soluble ferredoxin (ferredoxin I) has been isolated from *Synechococcus* 6301 and its protein sequence has been determined (S. Buckle, personal communication; Wada *et al.*, 1988). It is almost identical with the protein sequence encoded by the ferredoxin I gene from the closely related cyanobacterium *Anacystis nidulans* R2 (van der Plas *et al.*, 1986b; Reith *et al.*, 1986). However, the ferredoxin gene described in the

present paper clearly does encode a ferredoxin and the predicted protein sequence is much more closely related to ferredoxins from the complex cyanobacteria *Chlorogloeopsis fritschii* (Takahashi *et al.*, 1982) and *Mastigocladus laminosus* (Hase *et al.*, 1978). In addition, this predicted protein is more similar to ferredoxin II from *Nostoc* than it is to ferredoxin I from the same organism (Hase *et al.*, 1982). For these reasons we refer to it as *Synechococcus* 6301 ferredoxin II. It also has an unusual feature, a C-terminal extension of eight amino acid residues. However, it is not known that this extension is present in the mature protein, although similar extensions are present in the [2Fe-2S] ferredoxins found in halobacteria (Hase *et al.*, 1977a, 1980).

The *Synechococcus* 6301 ferredoxin II gene described here is only distantly related to the plant and cyanobacterial ferredoxin I genes that have been described elsewhere. These are from the white campion (*Silene platensis*) (Smeekens *et al.*, 1985) and from the cyanobacteria *Anabaena variabilis* (van der Plas *et al.*, 1986a), *Anabaena* 7120 (Alam *et al.*, 1986) and *Anacystis nidulans* R2 (van der Plas *et al.*, 1986b; Reith *et al.*, 1986). These genes are greater than 60% identical in nucleotide sequence and will cross-hybridize (van der Plas *et al.*, 1986a,b). However, the ferredoxin II gene is less than 50% homologous to these genes and it is therefore unlikely that the *Synechococcus* 6301 ferredoxin I gene

Fig. 7. Hybridization of the ferredoxin gene with *Synechococcus* 6301 genomic DNA

An *EcoRI*–*AluI* fragment (1338 bp) containing the *Synechococcus* 6301 ferredoxin gene was employed. *Synechococcus* 6301 genomic DNA was digested with *Bam*HI (track a), *Kpn*I (track b) and *Pst*I (track c). The sizes of bands are shown in kb on the right-hand side. The 1.6 kb *Kpn*I fragment corresponds to that which extends from position 7220 to 8885 in the sequenced DNA; the 3.5 kb *Pst*I fragment extends from position 5280 to 8799. A *Bam*HI site at position 7360 is consistent with the presence of the hybridizing fragment, which is larger than 1.6 kb. This fragment is probably the same as a 1.9 kb fragment detected in the *Bam*HI digests of both λ S1 and λ S5 (Cozens, 1986).

would cross-hybridize with ferredoxin II DNA under the experimental conditions used. This view was confirmed by Southern hybridizations of the isolated ferredoxin gene with digests of *Synechococcus* 6301 DNA. The results obtained (Fig. 7) are consistent with hybridization with the known gene and there is no evidence for a second closely related gene. These results suggest that the ferredoxin II gene resulted from a gene duplication before the divergence of the cyanobacterial genera *Synechococcus* and *Anabaena*.

Structural implications

A striking difference between the deduced protein sequence for the *Synechococcus* ferredoxin and related proteins (Fig. 8) is that the *Synechococcus* protein has a C-terminal extension of eight amino acid residues. In the known structures of [2Fe-2S] ferredoxins both N- and C-termini are located on the surface of the protein. So extension of the C-terminal part of the chain could be achieved without grossly interfering with the folding of the rest of the polypeptide chain.

Earlier studies of ferredoxin sequences from 38 different species of plants and cyanobacteria showed that 18 amino acid residues, including the four co-ordinated cysteine residues, are invariant (Matsubara & Hase, 1983). The other invariant residues are mostly in the regions surrounding the iron-sulphur centre. However, six of these hitherto invariant residues are not conserved in the *Synechococcus* 6301 ferredoxin (see Fig. 7). The remaining 12 invariant amino acid residues are found in the C-terminal two-thirds of the protein chain, and include the four co-ordinated cysteine residues. One group of six invariant amino acid residues is localized from residues 39 to 50. This segment and residues 79–82, which includes two additional invariant amino acid residues, are thought to be important in stabilizing the iron-sulphur-cluster cavity (Fukuyama *et al.*, 1980). The importance of the six invariant residues in the segment from residues 39 to 50 is emphasized by the fact that they are also conserved in [2Fe-2S] ferredoxins in halobacteria, as are Gly-57 and Cys-80 (Hase *et al.*, 1977a, 1980).

We thank Dr. S. Buckle, Dr. T. Hase and Dr. L. Rogers for providing us with unpublished experimental data. A.L.C. was supported by a Medical Research Council research studentship.

Fig. 8. Alignment of amino acid sequences of ferredoxins

The sequences of ferredoxins from *Aphanothece sacrum* and *Spirulina platensis* and that from *Synechococcus* 6301 are shown. Conserved amino acid residues are boxed. Asterisks mark residues conserved in all known cyanobacterial or plant chloroplast ferredoxins. The Λ symbols indicate additional residues conserved in all ferredoxins other than *Synechococcus* ferredoxin II.

REFERENCES

- Alam, J., Whitaker, R. A., Krogmann, D. W. & Curtis, S. E. (1986) *J. Bacteriol.* **168**, 1265–1271
- Berk, A. J. & Sharp, A. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1274–1278
- Biggin, M., Gibson, T. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965
- Bryant, D. A., Dubbs, J. M., Fields, P. I., Porter, R. D. & de Lorimer, R. (1985) *FEBS Lett.* **29**, 343–349
- Cammack, R., Rao, K. K., Barger, C. P., Hutson, K. G., Andrew, P. W. & Rogers, L. J. (1977) *Biochem. J.* **168**, 205–209
- Cammack, R., Rao, K. K. & Hall, D. O. (1981) *BioSystems* **14**, 57–80
- Cammack, R., Rao, K. K. & Hall, D. O. (1985) *Physiol. Vég.* **23**, 649–658
- Cohn, C. L., Alam, J. & Krogmann, D. W. (1985) *Physiol. Vég.* **23**, 659–667
- Cozens, A. L. (1986) Ph.D. Thesis, University of Cambridge
- Cozens, A. L. & Walker, J. E. (1987) *J. Mol. Biol.* **194**, 359–383
- Curtis, S. E. & Haselkorn, R. (1984) *Plant Mol. Biol.* **3**, 249–258
- Dutton, J. E., Rogers, L. J., Haslett, B. G., Takuro, I. A. H., Gleaves, J. T. & Boulter, D. (1980) *J. Exp. Bot.* **31**, 379–391
- Farrell, P. J., Deininger, P. L., Bankier, A. & Barrell, B. G. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1565–1569
- Fukuyama, K., Hase, T., Matsumoto, S., Tsukihara, T., Katsube, Y., Tanaka, N., Kakudo, M., Wada, K. & Matsubara, H. (1980) *Nature (London)* **286**, 522–524
- Gay, N. J. & Walker, J. E. (1981) *Nucleic Acids Res.* **9**, 3919–3926
- Hase, T., Wakabayashi, S., Matsubara, H., Kerscher, L., Oesterheld, D., Tao, K. K. & Hall, D. O. (1976) *J. Biochem. (Tokyo)* **83**, 1657–1670
- Hase, T., Wakabayashi, S., Matsubara, H., Kerscher, L., Oesterheld, D., Rao, K. K. & Hall, D. O. (1977a) *FEBS Lett.* **77**, 308–310
- Hase, T., Wada, K. & Matsubara, H. (1977b) *J. Biochem. (Tokyo)* **82**, 277–286
- Hase, T., Wakabayashi, S., Matsubara, H., Rao, K. K., Hall, D. O., Widner, H., Gysi, J. & Zuber, H. (1978) *Phytochemistry* **17**, 1863–1867
- Hase, T., Wakabayashi, S., Matsubara, H., Mevarech, M. & Werber, M. M. (1980) *Biochim. Biophys. Acta* **623**, 139–145
- Hase, T., Matsubara, H., Hutber, G. N. & Rogers, L. J. (1982) *J. Biochem. (Tokyo)* **92**, 1347–1355
- Hutson, K. G., Rogers, L. J., Haslett, B. G., Boulter, D. & Cammack, R. (1978) *Biochem. J.* **172**, 465–477
- Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441
- Marmur, J. (1961) *J. Mol. Biol.* **3**, 208–218
- Matsubara, H. & Hase, T. (1983) in *Proteins and Nucleic Acids in Plant Systematics* (Jensen, U. & Fairbrothers, D. E., eds.), pp. 168–181, Springer-Verlag, Berlin and Heidelberg
- Mills, D. R. & Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2232–2235
- Mulligan, B., Schultes, N., Chern, L. & Bogorad, L. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2693–2697
- Nierzwicki-Bauer, S. A., Curtis, S. E. & Haselkorn, R. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5961–5965
- Pribnow, D. (1978) in *Biological Regulation and Development* (Goldberger, R., ed.), pp. 219–277, Plenum Press, New York
- Reith, M. E., Laudenbach, D. E. & Strauss, N. A. (1986) *J. Bacteriol.* **168**, 1319–1324
- Rosenberg, M. & Court, P. (1979) *Annu. Rev. Genet.* **13**, 319–353
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Shin, M. & Arnon, D. I. (1965) *J. Biol. Chem.* **240**, 1405–1411
- Shin, M., Sukenobu, M., Oshino, R. & Kitazume, Y. (1977) *Biochim. Biophys. Acta* **460**, 85–93
- Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342–1346
- Shinozaki, K. & Sugiura, M. (1985) *Mol. Gen. Genet.* **200**, 27–32
- Smeekens, S., van Buisbergen, J. & Weisbeek, P. (1985) *Nucleic Acids Res.* **13**, 3179–3194
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
- Staden, R. (1982a) *Nucleic Acids Res.* **10**, 2951–2961
- Staden, R. (1982b) *Nucleic Acids Res.* **10**, 4731–4751
- Staden, R. (1984) *Nucleic Acids Res.* **12**, 521–538
- Staden, R. (1985) *Genet. Eng.* **7**, 67–114
- Takahashi, Y., Hase, T., Matsubara, H., Hutber, G. N. & Rogers, L. J. (1982) *J. Biochem. (Tokyo)* **92**, 1363–1368
- Tsukihara, T., Fukuyama, K., Nakamura, M., Katsube, Y., Tanaka, N., Kakudo, M., Wada, K., Hase, T. & Matsubara, H. (1981) *J. Biochem. (Tokyo)* **90**, 1763–1773
- Tsutsui, T., Tsukihara, T., Fukuyama, K., Katsube, Y., Hase, T., Matsubara, H., Nishikawa, Y. & Tanaka, N. (1983) *J. Biochem. (Tokyo)* **94**, 299–302
- Tumer, N. E., Robinson, S. J. & Haselkorn, R. (1983) *Nature (London)* **306**, 337–342
- van der Plas, J., de Groot, R. P., Weisbeek, P. J. & van Arkel, G. A. (1986a) *Nucleic Acids Res.* **14**, 7803
- van der Plas, J., de Groot, R. P., Woortman, M. R., Weisbeek, P. J. & van Arkel, G. A. (1986b) *Nucleic Acids Res.* **11**, 7804
- Wada, K., Matsubara, H., Chain, R. K. & Arnon, D. T. (1981) *Plant Cell Physiol.* **22**, 278–281
- Wada, K., Masui, R., Matsubara, H. & Rogers, L. J. (1988) *Biochem. J.* **252**, 571–575
- Wieslander, L. (1979) *Anal. Biochem.* **98**, 305–309

Received 23 November 1987; accepted 9 February 1988