

Cotranscription of genes encoding the small and large subunits of ribulose-1,5-bisphosphate carboxylase in the cyanobacterium *Anabaena* 7120

(*rbcS* gene sequence/promoter mapping/transcription regulation)

SANDRA A. NIERZWICKI-BAUER, STEPHANIE E. CURTIS*, AND ROBERT HASELKORN

Department of Biophysics and Theoretical Biology, University of Chicago, Chicago, IL 60637

Communicated by R. Stephen Berry, June 15, 1984

ABSTRACT In the filamentous cyanobacterium *Anabaena*, the gene for the small subunit (*rbcS*) of ribulose-1,5-bisphosphate carboxylase is linked to and transcribed together with the gene encoding the large subunit (*rbcL*) of the same enzyme. The two genes are separated by a spacer of 545 base pairs (bp) that does not contain an open reading frame. Both genes hybridize with a predominant 3.1-kilobase transcript that initiates 414 bp upstream from the *rbcL* coding region. The nucleotide sequence 14–8 bp preceding the transcription start site resembles a good *Escherichia coli* promoter, but the sequence in the –35 region does not. There is no obvious relationship between the sequence flanking the amino terminus of the cyanobacterial small subunit gene and the transit peptide of eukaryotic small subunit genes. The *Anabaena rbcS* gene is 61% homologous at the amino acid level with the gene from the unicellular cyanobacterium *Anacystis* and 37–38% homologous with the corresponding nuclear genes from eukaryotes. In contrast, the *Anabaena rbcL* gene is ≈86% homologous at the amino acid level with the *rbcL* genes from plant chloroplasts. Cotranscription satisfies one of the requirements for coordinate expression of the two genes whose products are needed in equimolar amounts in the mature enzyme. The *rbcL*–*rbcS* transcript is equally abundant in *Anabaena azollae* grown in the light or on fructose in the dark.

Ribulose-1,5-bisphosphate (RuP₂) carboxylase catalyzes the incorporation of CO₂ into carbohydrate in plant and algal chloroplasts as well as in photosynthetic prokaryotes. The enzyme contains eight identical large subunits of M_r = 56,000 and eight (nearly) identical small subunits of M_r ≈14,000 (1). In plants, a unique gene in chloroplast DNA encodes the large subunit, while a family of eight or more nuclear genes encodes the small subunit (2, 3). The nuclear genes are not identical and many or all are expressed, leading to some microheterogeneity in the small-subunit amino acid sequence (4). Moreover, the nuclear genes contain intervening sequences, some of which separate domains of the mature small-subunit protein (3). All of the eukaryotic small-subunit genes studied thus far contain an intervening sequence located between amino acid residues 2 and 3 of the mature protein and a leader region that codes for a transit peptide (5, 6). A precursor of the small subunit is synthesized on cytoplasmic ribosomes and then transported into the chloroplast accompanied by cleavage of the transit peptide at a Cys-Met bond to yield the mature protein (3).

In this report, we show that the cyanobacterium *Anabaena* contains a unique small-subunit gene located downstream from the large-subunit gene. There is no transit peptide in the small-subunit protein sequence nor are intervening sequences present in the prokaryotic gene. The genes

coding for the two subunits of RuP₂ carboxylase appear to be transcribed in a single mRNA in *Anabaena*.

MATERIALS AND METHODS

Isolation of DNA and Preparation of Southern Blots. DNA from *Anabaena* 7120 was prepared as described (7). *Escherichia coli* plasmid DNA was isolated as described by Davis *et al.* (8). Chromosomal and plasmid DNA were digested with restriction endonucleases, subjected to agarose gel electrophoresis, and transferred to nitrocellulose filters as described (9–11).

Isolation of RNA and Preparation of RNA Transfer Blots. Total RNA was isolated from *Anabaena* cells grown on complete medium as described (12). Aliquots containing 20 μg of the RNA were denatured with dimethyl sulfoxide and glyoxal, fractionated on 1% agarose gels, and transferred to nitrocellulose paper as described (12). Size markers, consisting of DNA fragments of known size, were similarly denatured and fractionated.

Homologous and Heterologous Hybridizations. The cDNA clone kindly provided by J. Bedbrook, pSSU160, contains 70% of the pea small subunit coding sequence (13) and was used as a heterologous hybridization probe to locate the *Anabaena* small-subunit gene. pSSU160 DNA was digested with *Hind*III and electrophoresed on a 5% acrylamide gel. A 230-base-pair (bp) fragment, containing the last 7 amino acids of the small-subunit transit peptide and the first 70 amino acids of the mature small-subunit coding sequence (13), was purified and nick-translated as described (9). This probe was hybridized with Southern blots containing *Hind*III-cut pAn600 and pAn606 (see Fig. 1). Filters were prehybridized at 40°C for 2 hr in 100 ml of 6× concentrated SSPE buffer (SSPE buffer = 0.15 M NaCl/0.01 M NaH₂PO₄/0.001 M Na₂EDTA, pH 7.0) containing 0.5% NaDodSO₄, Denhardt's solution (14), and 0.03 μg of pBR322 DNA per ml and then hybridized for 12 hr at 40°C in 10 ml of 6× concentrated SSPE buffer containing 0.5% NaDodSO₄, Denhardt's solution, and ≈1 × 10⁶ cpm of ³²P-labeled probe DNA. Filters were rinsed at 40°C three times for 1 hr each in 6× concentrated SSPE buffer/0.5% NaDodSO₄, air dried, and autoradiographed. Subsequently, filters were rewashed two times for 2 hr each in 6× concentrated SSPE buffer at 42°, 44°, 46°, and 50°C. Following each wash at the higher temperature, filters were air dried and autoradiographed.

In the homologous hybridizations, DNA fragments specific for the *Anabaena* large subunit and small subunit genes were purified and hybridized with RNA transfer blots or Southern blots. For the large-subunit probe, the plasmid pAn602 was digested with *Hind*III and *Hpa* I and the 1.0-kbp

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RuP₂, ribulose-1,5-bisphosphate; bp, base pair(s); kb, kilobase(s).

*Present address: Department of Genetics, North Carolina State University, Raleigh, NC 27650.

*Hind*III/*Hpa* I fragment internal to the large-subunit coding region (Fig. 1) was purified after electrophoresis on preparative agarose gels according to Maniatis *et al.* (15). A 0.95-kbp *Hind*III/*Xba* I fragment containing the entire small-subunit gene was isolated in an analogous manner from a *Hind*III and *Xba* I double digest of the plasmid pAn606 (Fig. 1).

DNA Sequence Analysis. *Anabaena* 7120 DNA fragments were prepared from plasmid pAn606, which contains a 1.5-kbp *Hind*III fragment (Fig. 1). DNA sequence analysis was carried out according to Maxam and Gilbert (16), with the exception of the piperidine cleavage step, which was modified as described by Smith and Calvo (17). The modified DNAs were analyzed on 5%, 8%, and 20% polyacrylamide gels (16) and autoradiographed with Kodak XAR-5 film using intensifier screens. All restriction sites except one were overlapped. The sequence at this site was determined by reading up to the labeled end on adjacent fragments by using 20% acrylamide gels (16). Sequencing ladders were transcribed manually, entered into an IBM computer, and translated by using the program of Queen and Korn (18).

S1 Nuclease Mapping and Primer Extension Assays. The start site of *rbcL* transcription was roughly determined by isolating *Hpa* I fragments containing sequences upstream from the *rbcL* coding region and subjecting these to S1 nuclease protection assays (see Fig. 1). The 450-bp *Hpa* I fragment that includes the 5' end of the coding region was fully protected, whereas the next *Hpa* I fragment upstream (400 bp) was only partially protected. To define transcription initiation more precisely, the 310-bp *Hinf*I fragment that overlaps the partially protected *Hpa* I fragment was labeled at its 5' ends and used for primer extension and S1 nuclease assays. For S1 nuclease protection, the *Hinf*I fragment was cut with *Hpa* I to yield a 280-bp fragment with the desired labeled end. The 280-bp fragment was hybridized with 120 μ g of RNA and treated with S1 nuclease as described (12). For the primer extension assay, the end-labeled *Hinf*I fragment was cut with *Dde* I to yield a 90-bp primer. The primer was hybridized with 120 μ g of RNA and extended with reverse transcriptase as described (12). The primer extension and S1 nuclease protection products were run on denaturing acrylamide gels with sequence ladders of the *Hinf*I fragment. See Fig. 6 for location of the sites used in these assays.

RESULTS AND DISCUSSION

Previous work by Heinhorst and Shively indicated that genes related to the large and small subunits of RuP₂ carboxylase of *Cyanophora* were located on the same restriction fragment of cyanelle (chloroplast) DNA of that organism (19). We also knew from preliminary experiments that the *Anabaena* large-subunit gene was transcribed to give a message large enough to include the small subunit and that the small-subunit gene sequence was not included in the nucleotides on the 5' side of the large-subunit coding region. We therefore directed our attention to the DNA on the 3' side of the large-subunit coding region—in particular, the 1.5-kbp *Hind*III fragment cloned in pAn606 (Fig. 1).

Fragments of pAn606, prepared by cutting with *Hind*III alone or with *Hind*III and *Xba* I, were fractionated by electrophoresis, blotted, and probed with the 230-bp fragment from the cloned pea small subunit gene (ref. 13; see *Materials and Methods*). Under low stringency conditions, the probe hybridized with the 1.5-kbp *Hind*III band and both *Hind*III-*Xba* I bands. When the blots were washed at increasing temperatures, the labeled band corresponding to the smaller *Hind*III-*Xba* I fragment (520 bp) disappeared, whereas the band corresponding to the larger *Hind*III-*Xba* I fragment (980 bp) remained (data not shown; see Fig. 1 for the map). Though washing blots at increasing temperatures allowed us to distinguish specific from nonspecific hybrid-

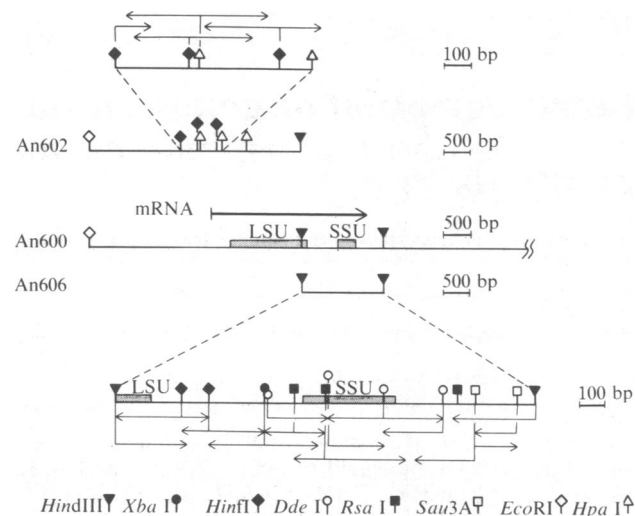


FIG. 1. Physical map of the region containing the *Anabaena* large-subunit (LSU) and small-subunit (SSU) genes. The 17.0-kbp *Eco*RI fragment of An600 had previously been shown to contain the large-subunit gene (7). The subclones An602 and An606 were used to determine the nucleotide sequences of the large-subunit 5' flanking region and the small-subunit gene, respectively, by using the strategies indicated. The thick bars represent the coding regions of the small-subunit and large-subunit genes. The two genes are transcribed into a message that initiates 414 bp upstream from the large-subunit coding region. The large-subunit and small-subunit genes are separated by a 545-bp spacer.

ization with cloned DNA, this procedure was not effective with total *Anabaena* DNA, at the DNA concentration used (4 μ g of DNA per lane). After determining the complete nucleotide sequence of the corresponding *Anabaena* DNA fragments (see below) it became apparent that, although the pea DNA probe and the *Anabaena* gene have average nucleotide homology of 55%, all of the hybridization observed is due to a single stretch in which 17 of 18 bp are identical.

We next determined the nucleotide sequence of the small-subunit gene following the strategy outlined in the lower portion of Fig. 1. Fig. 2 displays the nucleotide sequence of the noncoding strand of the small-subunit gene and its flanking regions. The *Anabaena* small-subunit coding sequence is 330 nucleotides long and terminates with a single ochre codon. The sequence flanking the open reading frame on the 5' side does not contain a typical prokaryotic ribosome-binding sequence (20). A possible prokaryotic transcription terminator, an inverted repeat followed by a stretch of T residues (21), was observed 32–63 bp beyond the ochre codon (underlined in Fig. 2).

The amino acid sequence of the *Anabaena* small subunit, determined by translating the gene sequence, is compared with the sequences of small subunits from other species in Fig. 3. The *Anabaena* small-subunit protein is 2 amino acid residues shorter than the *Anacystis* small subunit protein and 11–19 amino acid residues shorter than eukaryotic small-subunit proteins. As shown in Fig. 3, the amino acid residues at positions three and four of the eukaryotic mature small-subunit proteins are deleted in the *Anabaena* small subunit. Also, 12 amino acid residues have been deleted somewhere between positions 45–65. This deletion has been arbitrarily positioned at residues 54–65, since the exact location cannot be determined from the available data. Highly conserved regions in small-subunit proteins of *Anabaena* and other species, allowing conservative amino acid replacements, occur at positions 10–22, 66–76, and 98–116. Also, many individual or small groups of homologous amino acid residues are present (boxed, in Fig. 3). Small-subunit proteins of prokaryotes differ considerably from those of eukaryotes at the NH₂ ter-

-630 GGATAAGTAG AGGAGGCAGG GGAAGTAAAT AAGAATGACT ATGGACTATT GACCAATGAC TATTGGCAAC
 -560 TGACAACCTCCTATACA GTCATTGATA CATTITGTAA CTGATTGTTA ACAAAACGTT TAAAACTTTA
 -490 TGTAATAACA AATTTAAATA TGTAAGTTAA GAACTTTTCA AAGAATAACT TATGCCATTT CTTGATATAT
 -420 IGAGAGACAA GTTACAAATT ACGTGGTGTG CAATTTTTTC ATCTTGGCCT GATTACTCTA CTAAATATCC
 -350 GTC AAGTAAA TTGGCTCTTA GCTCGTCTCC TGTC AATAAAA GGAGGTCCGGC AAGAGTGCAG AAGCGGGAAT
 -280 GTGTGAAAAC TAACCCAATT CATTAAATAC CCCGAAATAT AGGGGAATCA TCTCATACTT TCCGTAAACC
 -210 GCGAAGGTCG TGAAGGGATA AAAGCAATTT AGTGGGTGAG AAGAACAGAT AAAAAGTAAA AGCGTTAACT
 -140 ATGCACTCCT AGATGAGCAA GACACTGGTG AAGAGGATTA CCACCTGACC TCTACCAAAA GATTATTCCT
 -70 GTTTTTCTCC TGCTGATAGG GAGGTAGGGC AATTGATAAG TAAAAAGAGT GACATCTTGG AAGGATAGAT

FIG. 6. Transcription initiation site of the *Anabaena rbcL* gene. The nucleotide sequence of the 630 bp preceding the *rbcL* coding region is shown. Numbering indicates the distance upstream from the *rbcL* coding region. Transcription initiation was mapped to nucleotide -414, indicated by the arrowhead (Fig. 5). Sequences -10 and -35 nucleotides from transcription initiation are underlined. The restriction sites used for the experiments of Fig. 5 are also indicated.

by the large-subunit coding region of 1428 bases (7), a spacer of 545 bases, the small-subunit coding region of 330 bases, and a 3' extension of 68 bases.

In *Cyanophora paradoxa* and both cyanobacteria studied to date, the *rbcS* and *rbcL* genes are linked and are probably cotranscribed. In contrast, these genes in plants are physically separated in the nuclear and chloroplast genomes, and *rbcS* is multicopy. The *rbcS* genes of plants contain a transit peptide required for transport of the small subunit into the chloroplast. The transit peptide is not encoded in the prokaryotic genes, as expected. In addition, there is no obvious relationship between the sequence flanking the terminus of the cyanobacterial *rbcS* genes and the transit peptide encoding regions of the plant genes. All of the plant *rbcS* genes sequenced to date also contain an insertion of 12 amino acid residues preceding the highly conserved region at residues 66-76, as well as an intervening sequence between residues 2 and 3 of the mature protein. It is possible that the 12 amino acid insert in the eukaryotic genes is necessary for chloroplast membrane crossing. This possibility can be tested by fusing the eukaryotic transit peptide coding sequence to the *Anabaena* small-subunit gene and examining the processing and transport behavior of the resulting polypeptide *in vitro*.

Mature RuP₂ carboxylase contains equimolar amounts of the large- and small-subunit protein. How this stoichiometry is obtained in plants is a mystery, since the total number of chloroplast *rbcL* genes vastly exceeds the number of nuclear *rbcS* genes. Cotranscription of the two genes in *Anabaena* makes it possible that equimolar amounts of the two proteins are produced by equal translation of the two messages. However, the *rbcL* gene has what appears to be a good ribosome-binding site just before its open reading frame (Fig. 6, A-G-G-A at residues -6 to -9), whereas the *rbcS* gene has none (Fig. 2). The distance between the genes, 545 bases, seems too great to expect ribosomes to read through from the *rbcL* message so the yield of small subunit obtained by translation of the 3-kb message could be low. It is also possible that processing and differential stability modify the relative levels of the two gene messages *in vivo*, but the problem of preventing nonspecific degradation of the RNA during isolation (Fig. 4) makes tests of those possibilities difficult.

We are grateful to Drs. S. J. Robinson and P. J. Lammers for helpful advice. This work was supported by Grant GM 21823 from the National Institutes of Health and Grant 5901-0410 from the U.S. Department of Agriculture, Science and Education Administration, through the Competitive Research Grants Office. S.A.N.-B. is a postdoctoral trainee supported by Grant HD 07136 and S.E.C. was a

postdoctoral fellow of the National Institute of General Medical Sciences.

1. McFadden, B. A. (1980) *Acc. Chem. Res.* **13**, 394-399.
2. Dunsmuir, P. & Bedbrook, J. (1983) in *Structure and Function of Plant Genomes*, eds. Ciferri, O. & Dure, L. (Plenum, New York), pp. 221-230.
3. Coruzzi, G., Broglie, R., Lamppa, G. & Chua, N.-H. (1983) in *Structure and Function of Plant Genomes*, eds. Ciferri, O. & Dure, L. (Plenum, New York), pp. 47-59.
4. Miziorko, H. M. & Lorimer, G. H. (1983) *Annu. Rev. Biochem.* **52**, 507-535.
5. Berry-Lowe, S. L., McKnight, T. D., Shah, D. M. & Meagher, R. B. (1982) *J. Mol. Appl. Genet.* **1**, 483-498.
6. Broglie, R., Coruzzi, G., Lamppa, G., Keith, B. & Chua, N.-H. (1983) *Bio/Technology* **1**, 55-61.
7. Curtis, S. E. & Haselkorn, R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1835-1839.
8. Davis, R., Roth, J. & Botstein, D. (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
9. Mazur, B. J., Rice, D. & Haselkorn, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 186-190.
10. Rice, D., Mazur, B. J. & Haselkorn, R. (1982) *J. Biol. Chem.* **257**, 13157-13163.
11. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503-517.
12. Tumer, N. E., Robinson, S. J. & Haselkorn, R. (1983) *Nature (London)* **305**, 337-342.
13. Bedbrook, J. R., Smith, S. M. & Ellis, R. J. (1980) *Nature (London)* **287**, 692-697.
14. Denhardt, D. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641-646.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 164-166.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
17. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255-2275.
18. Queen, C. L. & Korn, L. J. (1980) *Methods Enzymol.* **65**, 595-609.
19. Heinhorst, S. & Shively, J. M. (1983) *Nature (London)* **304**, 373-374.
20. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342-1346.
21. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319-353.
22. Shinozaki, K. & Sugiura, M. (1983) *Nucleic Acids Res.* **11**, 6957-6964.
23. Müller, K.-D., Salnikow, J. & Vater, J. (1983) *Biochim. Biophys. Acta* **742**, 78-83.
24. Martin, P. G. (1979) *Aust. J. Plant Physiol.* **6**, 401-408.
25. Coruzzi, G., Broglie, R., Cashmore, A. R. & Chua, N.-H. (1983) *J. Biol. Chem.* **258**, 1399-1402.