

Sequence of a cDNA clone encoding part of the small subunit of the ribulose-1,5-bisphosphate carboxylase of *Nicotiana sylvestris*

Lothaire Pinck, Jacqueline Fleck*, Monique Pinck, Rafik Hadidane and Léon Hirth

Institut de Biologie Moléculaire et Cellulaire du CNRS, Université Louis Pasteur, 15, rue Descartes, 67084 Strasbourg Cédex, France

Received 19 January 1983; revision received 14 February 1983

The nucleotide sequence of the cDNA clone pSTV34 is reported. It codes for the 86 COOH-terminal amino acids of the protein coding for the small subunit of the ribulose bisphosphate carboxylase and has a 3' non-translated region 195 base pairs long. In the coding region this clone is 72% homologous with the corresponding cDNA clone from *Pisum sativum*.

<i>Plant</i>	<i>Nicotiana</i>	<i>Ribulose bisphosphate carboxylase</i>	<i>cDNA</i>	<i>Nucleotide sequence</i>
--------------	------------------	--	-------------	----------------------------

1. INTRODUCTION

To understand the molecular events that control gene expression during cellular differentiation, it is essential to have a molecular hybridization probe that can be used to quantify, characterize, and isolate the corresponding RNA or DNA sequences.

To understand the mechanism that regulates the expression of the gene coding for the small subunit of ribulose-1,5-bisphosphate carboxylase (review [1,2]) in callus cultures obtained from protoplasts cultivated in various physiological conditions, we used a cDNA probe corresponding to the mRNA of this small subunit. The construction of this probe, its identification by the hybridization-translation procedure, and the results obtained through its use have been reported [3]. Those results suggested that expression of this gene is regulated at the level of transcription. To complete the identification of this probe we report here the sequence of 556 base pairs of the insert in our clone (pSTV34), and we compare it with the same type of probe constructed from *Pisum sativum* [4].

2. MATERIALS AND METHODS

2.1. Materials

Restriction enzymes were purchased from BRL (Gaithersburg) and used according to the technical data sheets provided with the products. Alkaline phosphatase (calf intestine) was purchased from Boehringer (Mannheim) and the polynucleotide kinase from BRL. The reverse transcriptase was a gift from Dr Beard (Life Science Inc., USA). [γ - 32 P]ATP (3000 Ci/mmol) and [α - 32 P]dCTP (3000 Ci/mmol) were obtained from the Radiochemical Centre (Amersham).

2.2. Preparation of plasmids

The plasmid pSTV34, carrying sequences homologous to the mRNA coding for the small subunit of the ribulose bisphosphate carboxylase from *Nicotiana sylvestris*, was constructed and isolated as in [3]. The plasmid was then purified by the clear-lysate technique [5], followed by CsCl equilibrium centrifugation, isobutanol treatment and alcohol precipitation.

2.3. DNA sequencing

Plasmid pSTV34 was digested with appropriate

* To whom correspondence should be addressed

restriction endonucleases to obtain fragments each containing a part of the inserted sequence. These fragments were fractionated on 0.8% agarose or 6% polyacrylamide slab gels, and recovered by electrophoretic elution [6]. Fragments with staggered ends produced by *Bam*HI digestion were filled in with reverse transcriptase [7], the appropriate non-radioactive deoxynucleotides, and [α - 32 P] dCTP. The other fragments were alkaline-phosphatase-treated and 5'-end-labelled with [γ - 32 P]ATP and polynucleotide kinase [8].

After strand separation or additional cleavage with restriction endonucleases, the singly labelled fragments were sequenced by the Maxam and Gilbert procedure [8] with slight modifications [7]. Products of the chemical degradation reactions were separated on $0.05 \times 30 \times 40$ cm polyacrylamide slab gels with acrylamide concentrations of 8% or 12%. After electrophoresis the gels were fixed in 10% acetic acid, dried and autoradiographed, mostly without intensifying screens to improve band resolution. The sequences were determined on both strands.

3. RESULTS

Before analysis, the plasmid pSTV34 was first mapped with various restriction endonucleases. As it was obtained by inserting a blunt-ended ds-cDNA into the unique repaired *Sal*II site of

pBR322, the insert is flanked by two generated *Taq*I sites. In addition to these *Taq*I sites, there are two other internal *Taq*I sites, together with two *Hinf*I sites, two *Eco*RII sites, two *Alu*I sites, four *Dde*I sites, and single cleavage sites for *Ava*II, *Eco*RI, *Hae*I, *Hae*II, *Hae*III, and *Xho*II. In fig.1, we report only the restriction sites which were useful in developing our DNA sequencing strategies.

The complete nucleotide sequence together with the deduced amino acid sequence is given in fig.2. The insert in plasmid pSTV34 consists of 556 nucleotides. The first 258 at the 5'-end are a sequence coding for 86 amino acid residues. The C-terminal codon, for a tyrosine residue, is followed by the translation termination TAA codon and a 195-nucleotide-long untranslated region with a 103-nucleotide-long tail of poly(A).

We estimated in [3,9] by agarose-gel electrophoresis under denaturing conditions that the mRNA coding for the small subunit consisted of about 1000 nucleotides. It is clear from these results that the cloned cDNA is not the full length, but corresponds to ~60% of the total mRNA. We deduce that it covers the complete 3'-untranslated region and 71% of the region coding for the mature 12000 *M_r* small subunit peptide [10].

The 3'-untranslated region, which is very rich in T residues (46% of the total, with 17% of the total being T residues in TA sequences) does not contain

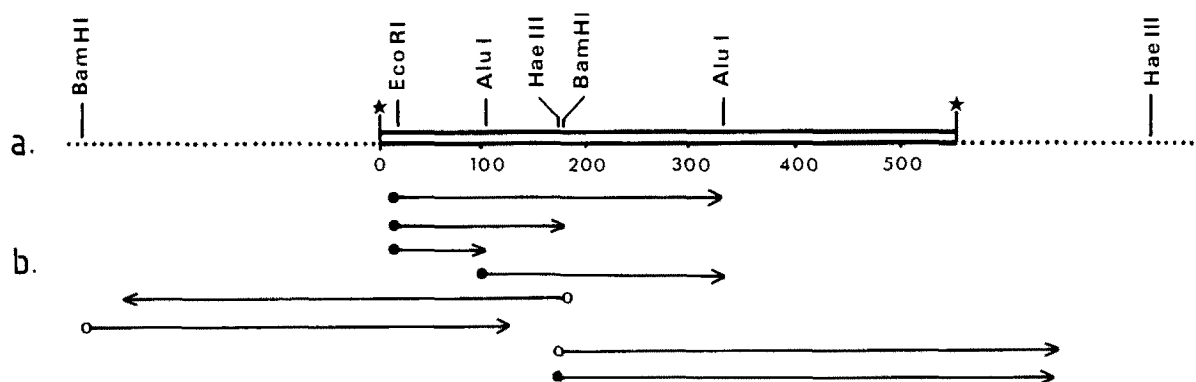


Fig.1. Strategy for sequencing the small-subunit cDNA clone. (a) Restriction map of the pSTV34 ds-cDNA inserted into the pBR322 unique repaired *Sal*II site, generating two *Taq*I sites (marked I* in the figure). This map shows only the restriction endonuclease sites used for the labelling and the sequencing. The rectangular box, drawn to scale in base pairs, represents the cDNA insert; the dotted lines represent the pBR322 sequences. (b) The arrows indicate both the direction of sequencing and the length of the sequence obtained from that particular site with 3'-end-labelled (○—○) or 5'-end-labelled (●—●) fragments.

```

      10      20      30      40      50      60
TGG GTT CCT TGC TGA TTC GAG ACT GAG CAC GGA TTT GTC TAT CGT GAA AAC AAC AAG
W V P C L E F E T E H G F V Y R E N N K

      70      80      90      100      110      120
TCA CCA GGA TAC TAT GAT GGC AGA TAC TGG ACC ATG TGG AAG CTA CCT ATG TTC GGA TGC
S P G Y Y D G R Y W T M W K L P M F G C

      130      140      150      160      170      180
ACT GAT GCC ACC CAA GTG TTG GCT GAG GTA GAA GAG GCG AAG AAG GCA TCA CCA CAG GCC
T A T Q V L A E V E E A K K A Y P Q A

      190      200      210      220      230      240
TGG ATC CGT ATC ATT GGA TTC GAC AAC GTG CGT CAA GTG CAG TGC ATC AGT TTC ATC GCC
W I R I I G F D N V R Q V Q C I S F I A

      250      260      270      280      290      300
TAC AAG CCA GAA GGC TAC TAA GTTTCATATTAGGACACCTACCCATTGTCTGACTTAGGGCAGTTTGT
Y K P E G Y

      310      320      330      340      350      360
TGAAATGTTACTTACGCTCTTTTCTTCTCTCCACAAAAATGATTATGTTTCTACTTCTATTCCGGGTATG
N=103

```

Fig.2. Nucleotide sequence and the corresponding amino acid sequence of pSTV34 cDNA insert. The nucleotide numbering system from the 5'-to-3' end is the same as in fig.1. The encoded amino acid sequence deduced from the nucleotide sequence by using the genetic code is indicated below the nucleotide sequence. The TAA terminator codon is boxed.

the classical AATAAA sequence but 3 times the sequence TAA, 26 base pairs away from the poly(A) tail.

The sequence of the insert was further compared to the corresponding sequence established in [4] for the P₂₀ cDNA clone of *P. sativum*. The amino acid sequences corresponding to the coding regions show considerable overall homology (72%; fig.3A,B). They have two major regions in common, one ending 6 amino acid residues before the C-terminus and containing 12 amino acids, and the other, containing 16 amino acids, ending 47 from the C terminus. A similar homology is observed between our insert (fig.3A) and the C-terminal portion of the amino acid sequence of the small subunit of *Spinacea oleracea* [11] (fig.3C).

In contrast to these striking similarities in the coding regions, there are far fewer homologies in the 3'-untranslated region. The 2 clones have very different nucleotide distributions and sequence

lengths (195 nucleotides for pSTV34 and 260 for P₂₀ cDNA). Except for a sequence of 6 nucleotides including the stop codon (TAAGTT) at the 5'-end and a sequence of 7 nucleotides (TGTTTGT) at the 3'-end of pSTV34 and 7 nucleotide residues away from the 3'-end of P₂₀ cDNA, the homologous sequences (1 of 12 nucleotides, 3 of 7 nucleotides, and 26 of 6 nucleotides) are scattered throughout this 3'-untranslated non-coding region.

4. DISCUSSION

The pSTV34 clone was first identified by estimating the *M_r* of the hybrid selected mRNA and by characterizing on SDS-polyacrylamide slab gels the polypeptide translated from this mRNA [3]. The identity of the clone, with > 70% homology between its coding region and the corresponding sequence in *P. sativum* [4], is now confirmed without ambiguity. Furthermore, this region shows the same amount of homology with the C-terminal portion of the amino acid sequence directly determined for the small subunit of *Spinacea oleracea* [11].

In contrast with the homogeneity in the coding regions, the 3'-untranslated regions are very different in the two clones, except for a common region of 12 nucleotides, 51 nucleotides upstream from the poly(A) tail in *Nicotiana sylvestris* and 173 nucleotides upstream from the poly(A) tail in *P. sativum* [4].

Another noteworthy feature is that the hexanucleotide AATAAA involved in the polyadenylation signal [12] is missing. However, as it is reported for an mRNA of zein [13], the AATAAT sequence present 22 nucleotides upstream the start of the poly(A) tail, may be considered as a variant of the classical AATAAA sequence.

Nevertheless, even though the insert in pSTV34

```

C WIPPLEFEVKDGFVYREHDKSPGYDGRYW KLPMEGGTDAQVNEVEVKAYPDAFVRFIGFBKREVQCISFIAYPAGY .COOH
A WPCLEFETEHRRNNRSGITM CQATLAVERAKKQWII DNVROGAYKEG .COOH
B WPCLEFELLKGFVYGEHNKSPRYDGRYWTNKLPMFGTTDAQVKELDEVVAAYPEAFVRVIGFNVRQVQCISFIAHTPEY .COOH

```

Fig.3. Comparison of the amino acid sequences deduced from the nucleotide sequence of the pSTV34 clone from *Nicotiana sylvestris* (A), the carboxy-terminal fragment of the P₂₀ cDNA clone [4] from *Pisum sativum* (B) and *Spinacea oleracea* (C) [11]. The fragments were aligned starting from the carboxy-terminal amino acid. The hatched areas are the homologous regions.

is not a full-length copy of the mRNA coding for the small subunit and does not contain any sequence region coding for the transit peptide as is the case for the 2 clones reported for *P. sativum* (P₂₀ cDNA [4], 13 amino acids; pSS15 [14], 100 amino acids), this clone will undoubtedly be a useful tool.

We have already used it as a probe to study the expression of the gene coding for the small subunit during cellular differentiation of protoplast cultures [3], and succeeded in detecting gene products which, depending on the state of differentiation, are often present only in minute quantities. With the help of this probe, we have also succeeded in isolating genomic fragments from *Nicotiana sylvestris*. Structural investigations are now in progress.

ACKNOWLEDGEMENTS

Part of this work was supported by the Centre National de la Recherche Scientifique, contrat 0781 (ATP 'Biologie Moléculaire Végétale'). Biohazards associated with the experiments described in this publication were examined in advance by the French National Control Committee. We are grateful to G. De Marcillac for help in computer analysis.

REFERENCES

- [1] Wildman, S.G. (1979) Arch. Biochem. Biophys. 196, 598–610.
- [2] Ellis, R.J. (1981) Carlsberg Res. Commun. 46, 259–278.
- [3] Vernet, T., Fleck, J., Durr, A., Fritsch, C., Pinck, M. and Hirth, L. (1982) Eur. J. Biochem. 126, 489–494.
- [4] Bedbrook, J.R., Smith, S.M. and Ellis, R.J. (1980) Nature 287, 692–697.
- [5] Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. USA 62, 1159–1166.
- [6] Wienand, U., Schwarz, Z. and Feix, G. (1979) FEBS Lett. 98, 319–323.
- [7] Smith, D.R. and Calvo, J.M. (1980) Nucleic Acids Res. 8, 2255–2273.
- [8] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
- [9] Lett, M.C., Fleck, J., Fritsch, C., Durr, A. and Hirth, L. (1980) Planta 148, 211–216.
- [10] Fleck, J., Durr, A., Lett, M.C. and Hirth, L. (1979) Planta 145, 279–285.
- [11] Martin, P.G. (1979) Aust. J. Plant Physiol. 6, 401–408.
- [12] Fitzgerald, M. and Shenk, V. (1981) Cell 24, 251–260.
- [13] Geraghty, D.E., Messing, J. and Rubenstein, I. (1982) EMBO J. 1, 1329–1335.
- [14] Broglie, R., Bellemare, G., Bartlett, S.G., Chua, N.H. and Cashmore, A.R. (1981) Proc. Natl. Acad. Sci. USA 78, 7304–7308.