Phytochemistry, 1981. Vol. 20, pp. 413–415. © Pergamon Press Ltd. Printed in England. 0031-9422/81/030413-03 $02.00/0

AMINO ACID SEQUENCE OF THE SMALL SUBUNIT OF RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE OF *PISUM SATIVUM*

ISHAQ A. H. TAKRURI*, DONALD BOULTER* and R. JOHN ELLIS†

* Department of Botany, University of Durham, Durham, U.K.; † Department of Biological Sciences, University of Warwick, Coventry, U.K.

(*Received* 13 *October* 1980)

Key Word Index—*Pisum sativum*; Leguminosae; pea; amino acid sequence; small subunit; ribulose-1,5-bisphosphate carboxylase; confirmation of nucleic acid sequence cDNA.

Abstract—The amino acid sequence of the small subunit of ribulose-1,5-bisphosphate carboxylase from pea consists of a single polypeptide chain of 123 residues with a calculated MW of *ca* 14480. The *N*-terminus was 'ragged' and both methionine and glutamine were determined in residue position 1. No heterogeneity was found even though two isofocussing variants were observed. The amino acid sequence confirms the nucleic acid sequence of cDNA of mRNA determined independently.

INTRODUCTION

Ribulose-1,5-bisphosphate carboxylase (EC 4.1.1.39) is a CO_2-fixing enzyme of chloroplasts which consists of eight similar large (LSU) and eight similar small (SSU) subunits. The pea SSU is encoded in the nuclear genome and synthesized in the cytoplasm as a larger molecule; about 50 residues are lost after passage through the chloroplast envelope before the subunits are assembled into the enzyme [1, 2]. Analysis by electrofocussing reveals that the pea small subunit contains two charge variants.

Bedbrook *et al.* [3] have determined the nucleic acid sequence of cDNA encoding the pea SSU. The present work, carried out independently, in which the amino acid sequence of the mature SSU has been determined directly, confirms the nucleic acid sequence determined by these authors. The results suggest that the pea SSU molecule charge-variants do not represent the products of different structural genes.

The complete amino acid sequence (Fig. 1) was determined, there being 123 residues, giving a calculated MW from the amino acid sequence of 14480. The *N*-terminus was found to be 'ragged', i.e. both methionine and glutamine were determined in residue position 1. In most preparations, the yield of methionine at the *N*-terminus equalled the yield of valine in position 2/3, suggesting that protein molecules lacking methionine at the *N*-terminus, i.e. those having glutamine instead, had probably become blocked by cyclization.

The major problem encountered in attempting to determine the amino acid sequence was the insolubility of the SSU after using sodium dodecyl sulphate to separate the large and small subunits. This problem was overcome by dissociating the protein into subunits by using the maleylation reaction [4] whereby the lysine residues were derivatized.

RESULTS AND DISCUSSION

The CNBr fragments were ordered by their *N*- and *C*-terminal amino acids thus: X1 had the same *N*-terminus as the intact SSU and homoserine at its *C*-terminus, the latter residue shown to be present by digestion with carboxypeptidase A. The second fragment was identified by the presence of homoserine at the *C*-terminus and tryptophan at the *N*-terminus, and the third fragment has the same *C*-terminal residue, tyrosine, as the intact protein.

The amino acid sequence is presented in Fig. 1. All the enzymes used (see legend, Fig. 1) showed their expected specificities apart from the following results. A complete set of tryptic peptides was obtained and their sequence determined. However, trypsin cleaved the maleylated SSU poorly between residues 35 and 36 presumably because of the presence of the derivatized lysine residue at position 36. A complete set of peptides produced by the protease from *Staphylococcus* was not separated, but the sequence of seven *Staphylococcus* peptides provided much of the confirmatory overlapping sequence. This enzyme also gave an unexpected cleavage at position 52/53 which resulted in peptide S6/7. Chymotrypsin gave many peptides of which C3 and C4 were used to confirm that part of the amino acid sequence given by T3, T4 and T5. All of the tryptic peptide sequences were overlapped with peptides obtained from staphylococcal or chymotryptic digests, or by automatic sequence analysis.

No heterogeneity was observed during the determination of the amino acid sequence and the occurrence of two isoelectric focussing variants may be due to the removal of the *N*-terminal methionine from a proportion of molecules; the glutamine residue so exposed could then cyclize, and thus lose a positive charge. A similar explanation has been put forward to account for the occurrence of two charge variants of SSU in spinach [5], but in tobacco the occurrence of such variants has been attributed to different structural genes [6]. The amino acid sequence is in complete agreement with the DNA sequence which was obtained completely independently [3].

Lastly, it is of interest to note that the pea amino acid sequence contains an amino acid triplet (positions 68–70) which is absent from the spinach molecule [5]. This triplet contains a methionine residue which gives rise to an extra

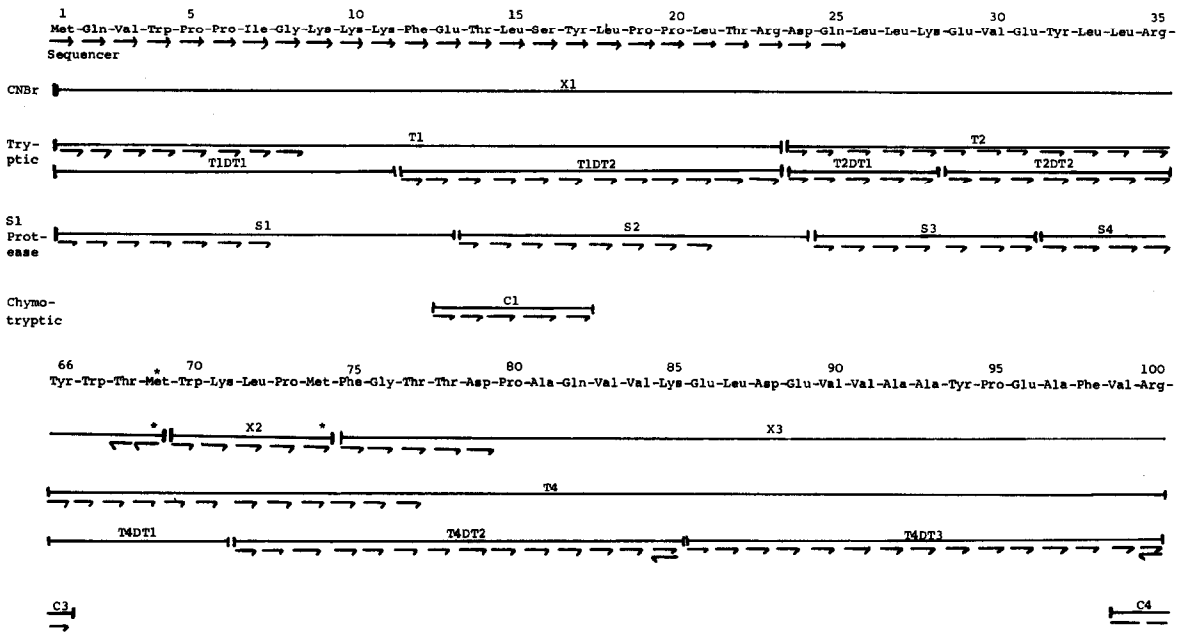


Fig. 1. Amino acid sequence of small subunit. X1, X2 and X3 represent the CNBr fragments. * indicates homoserine was identified as the derivative from CNBr cleavage. S, T and C represent peptides obtained by a protease from *Staphylococcus* [8], tryptic [11] and chymotryptic [11] digestion, respectively, of carboxymethylated and maleylated protein. Peptides are numbered on the basis of their order within the complete sequence. De-maleylated tryptic peptides are designated by DT, e.g. T2DT1 is the first sub-peptide of demaleylated tryptic peptide T2. Arrows →, ⇀, ← represent automatic sequencing, manual Edman degradation and carboxypeptidase A and B digestion, respectively (see ref. [10] for methods). Residues that were unambiguously identified are indicated with solid arrows. Others (e.g. weaker spot) with broken arrows.

fragment on cyanogen bromide cleavage. Otherwise, the pea and spinach molecules show considerable homology; 89 residues are the same in both species. Such homology is not unexpected since SSU from pea can assemble with LSU from spinach to form a protein with the same electrophoretic mobility as the native spinach enzyme. [7]

EXPERIMENTAL

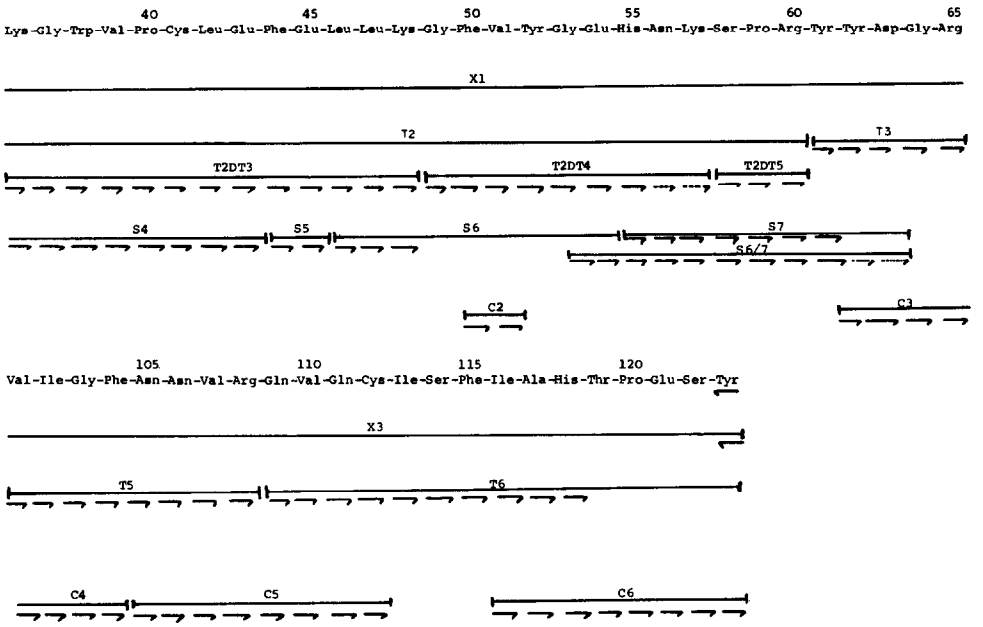
Ribulose-1,5-bisphosphate carboxylase was purified from young leaves of *Pisum sativum* (cv Feltham First) by a method based on that of Covey and Taylor [9]. After carboxymethylation [10], the protein was maleylated [4] and the large and small subunits were separated by chromatography on Sephadex G-75 equilibrated in 50 mM NH_4HCO_3 [11]. The fractions containing SSU were lyophilized.

Twenty-five *N*-terminal residues (Fig. 1) were determined by analysing 5 mg of SSU using a Beckman automatic sequencer with the fast protein QUADROL programme 122974. The SSU was cleaved chemically with cyanogen bromide [12] and the resulting fragments, X1–X3, were separated on a column of Sephadex G-75 [11].

Separate enzymic digests (see Fig. 1) were performed on the intact SSU and the resulting peptides separated by Sephadex G-75 (superfine grade) column chromatography, followed, where necessary in order to obtain pure peptides, by paper electrophoresis at pH 6.5 [11]. Tryptic peptides T1, T2 and T4 after separation were de-maleylated [13] to expose the amino group of lysine and then sub-digested with trypsin again. These peptides were purified on a column of Bio Gel P-4 [14] and by paper electrophoresis at pH 6.5 [11]. The observed mobilities of the peptides from enzymic digests were in agreement with those calculated from the proposed sequence (data not shown). Tryptophan-containing peptides were also identified by positive staining with Ehrlich's reagent [11].

N-Terminal residues were determined by the dansyl Edman method [15] and pure peptides were sequenced by the 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate (DABTIC) method [16]. *C*-Terminal amino acids were sequenced by using either carboxypeptides A [11] or B [17].

Acknowledgements—This work was supported by a Science Research Council Grant to R. J. E. and D. B. We thank Mr. Alan Guest for technical assistance.



REFERENCES

1. Highfield, P. E. and Ellis, R. J. (1978) *Nature* **271,** 420.
2. Smith, S. M. and Ellis, R. J. (1979) *Nature* **278,** 662.
3. Bedbrook, J., Smith, S. M. and Ellis, R. J. (1980) *Nature* **287,** 692.
4. Sugiyama, T. and Akazawa, T. (1970) *Biochemistry* **9,** 4499.
5. Martin, P. G. (1979) *Aust. J. Plant Physiol.* **6,** 401.
6. Kung, S.-D. (1976) *Science* **191,** 429.
7. Chua, N.-H. and Schmidt, G. W. (1980) in *Photosynthetic Carbon Assimilation* (Siegelman, H. W. and Hind, G., eds.) p. 325. Plenum Press, New York.
8. Houmard, J. and Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **59,** 3506.
9. Covey, S. N. and Taylor, S. C. (1980) *FEBS Letters* (in press).
10. Milne, P. R. and Wells,, J. R. E. (1970) *J. Biol. Chem.* **245,** 1566.
11. Takruri, I. A. H., Haslett, B. G., Boulter, D., Andrew, P. W. and Rogers, L. J. (1978) *Biochem. J.* **173,** 459.
12. Steers, E., Craven, G. R. and Anfinsen, C. B. (1965) *J. Biol. Chem.* **240,** 2478.
13. Kootstra, A. and Bailey, G. S. (1978) *Biochemistry* **17,** 2504.
14. Takruri, I. A. H. and Boulter, D. (1979) *Phytochemistry* **18,** 1481.
15. Gray, W. R. and Hartley, B. S. (1963) *Biochem. J.* **89,** 378.
16. Chang, J. W., Brauer, D. and Wittmann-Liebold, B. (1978) *FEBS Letters* **93,** 205.
17. Ambler, R. P. (1972) *Methods Enzymol.* **25,** 143.

NOTE ADDED IN PROOF

Martin (personal communication) has re-investigated the tryptic peptide commencing at position 66 and has found that there is no difference between spinach and pea; i.e. the omission of a tri-peptide from spinach was an error.