THE AMINO ACID SEQUENCE OF SOYBEAN LEGHAEMOGLOBIN c_2

John G. R. HURRELL* and Sydney J. LEACH
The Russell Grimwade School of Biochemistry, University of Melbourne, Parkville, Victoria, 3052, Australia

Received 6 June 1977

1. Introduction

The leghaemoglobin found in the root nodules of the soybean plant (*Glycine max* cultivar Lincoln) after rhizobial infection was originally fractionated into two major and two minor components [1,2]. Recently, the second of the major components, leghaemoglobin c, has been found to consist of two distinct molecular species, leghaemoglobin c_1 and c_2, separable by elution from a DEAE cellulose column using a continuous acetate gradient at pH 5.2 [3]. Since the amino acid sequence of the first major component, leghaemoglobin a, is known [4] and it has been well characterized both structurally and functionally [5–8] we felt it would be of value to elucidate the primary structure of a second major component, leghaemoglobin c_2, as a step towards establishing a function (if any) for the multiplicity of leghaemoglobin species found in the root nodules of most legumes.

2. Materials and methods

Soybean plants were grown as previously described [9] and isolated by gradient elution (10 to 50 mM acetate) at pH 5.2 from a DEAE cellulose column [3]. The leghaemoglobin c_2 fraction was pooled and rechromatographed on a shallow gradient (25–35 mM acetate) to ensure removal of the c_1 component.
The protein (65 mg) was dehaemed by the HCl-acetone method [10] and succinylated in an auto-pH titrator at pH 9.2 by the addition of eight lots of finely-divided succinic anhydride (80 mg each) at 30-min intervals. After de-salting by passage down a Sephadex G-25 column equilibrated in 0.1% NH_4HCO_3, 20.5 mg was cleaved at the single arginine residue in the protein [3] by digestion with trypsin (1 h, pH 8.2, 37°C, 1% enzyme/protein (w/w)). Cleavage of another sample of the succinylated protein (43 mg) at the asparaginyl-glycyl bond was carried out by the hydroxylamine method [11] extended for 20 h. The fragments from both cleavages were separated by chromatography on a Sephadex G-50 (fine) column (100 × 3 cm) equilibrated in 1% NH_4HCO_3.

Small peptides were generated by digesting the apoprotein with trypsin [3], staphylococcal protease [12] and thermolysin [13] using established procedures. Peptides were isolated by high voltage electrophoresis at pH 6.5 in the first dimension on Whatman 3 MM paper and descending chromatography with BAWP (*n*-butanol–actetic acid–water–pyridine, 30:6:24:20) in the second dimension. The strip containing the neutral peptides was cut out and re-electrophoresed at pH 3.5 after sewing onto a new sheet of paper. Peptide spots were located by spraying the peptide maps with either ninhydrin or fluorescamine [14], cut from the paper, eluted, the peptides hydrolyzed in 6 M HCl (containing 1% phenol) and their amino acid compositions determined by analysis on a Beckman 120 B Amino Acid Analyzer.

The C-terminal sequence of the apoprotein was determined by carboxypeptidase (A+B) digestion [15] and subsequent identification and quantitation of the amino acids released was by means of the Analyzer. The sequences of the N-terminal regions of the apoprotein and the large peptide fragments were obtained by automated sequence analysis using an Illitron Protein Sequenator [Illinois Tool Co., Ill.]

*Present Address: Research and Development Division, Commonwealth Serum Laboratories, Parkville, Victoria, 3052, Australia

using the program of Edman and Begg [16] suitably modified for this instrument, and replacing the 1-chlorobutane with a mixture of 1,2-dichloromethane and benzene (3:1) for the extraction of the thiazolinones [17]. Aliquots of the thiazolinones were converted to the phenylthiohydantoin and identified by thin-layer chromatography (TLC) [17]. When desirable the remainder of each sample was dried, hydrolyzed by hydriodic acid (0.2 ml, in vacuo, 120°C, 24 h) and the amino acid identified and quantitated by amino acid analysis.

Small peptides were sequenced by a manual Edman procedure [18]. The thiazolinones were converted to the phenylthiohydantoin and identified by two-dimensional TLC on silica gel [19] and/or by the subtractive procedure [20].

3. Results and discussion

A summary of the results of sequenator analysis of apoleghaemoglobin c_2 and its fragments is seen in Table 1. Automated sequenator analysis provided the sequence of 68 of the 141 residues of leghaemoglobin c_2. Microheterogeneity was observed at residue 5 where glutamic and aspartic acids were both identified in the molar ratio of 4 : 6.5. This was observed in both of the sequence analyses of the apoprotein summarised in Table 1.

The complete sequence of leghaemoglobin c_2 is shown in fig.1. The sequence information lacks overlap peptides in the region (110–139). As there is only one amino acid substitution between leghaemoglobins a and c_2 in this region, however, positioning of the peptides within this region was carried out by comparison of the two sequences. Overall, the sequences of the two proteins differ only at 10 positions. Most of the substitutions are conformationally conservative [21] which is consistent with the conformational identity for the two proteins indicated by circular dichroism studies [5], and the high degree of immunological cross-reactivity [12]. The close sequence homology indicates that the genetic event which gave rise to the separate leghaemoglobin a and c_2 genes was either a relatively recent one, or, that accepted changes within the leghaemoglobin gene of a plant species are restricted by functional interaction of leghaemoglobin with other proteins in the cell. While slight differences in oxygen-binding affinities for the multiple leghaemoglobin components of soybean have been observed [23], the amino acid sequence adds no support to the suggestion that multiple forms of leghaemoglobins in the root nodules of legumes serve different biological functions [24].

Acknowledgements

This work was supported by a grant to S. J. L. from the Australian Research Grants Committee and the Research and Graduate Studies Committee of the

Table 1
Summary of automated sequenator analyses on apoleghaemoglobin c_2 and its fragments

Protein or fragment	Number of residues identified	Position of sequence in protein	Repetitive yield (%)
Apoprotein	21	1–21	82.6
Apoprotein	36	1–37[a]	91.0
S-T-(1)[b]	20	71–91	89.5
S-H-(1)[c]	12	50–63	83.9

[a]Residue 35 (Glu) was not identified due to an electrical fault in the instrument.
[b]S-T(1) was the fragment generated by tryptic cleavage of succinylated leghaemoglobin c_2. This was a mixture of two peptides (1–71) and (72–141), the N-terminal of peptide (1–71) being blocked by a succinyl group. The addition of a succinyl group to the ε-amino side-chain of lysine 77 prevented its identification in the sequenator.
[c]S-H(1) was a fragment generated by hydroxylamine cleavage of succinylated leghaemoglobin c_2. Lysine 57 was not identified because of side-chain succinylation.

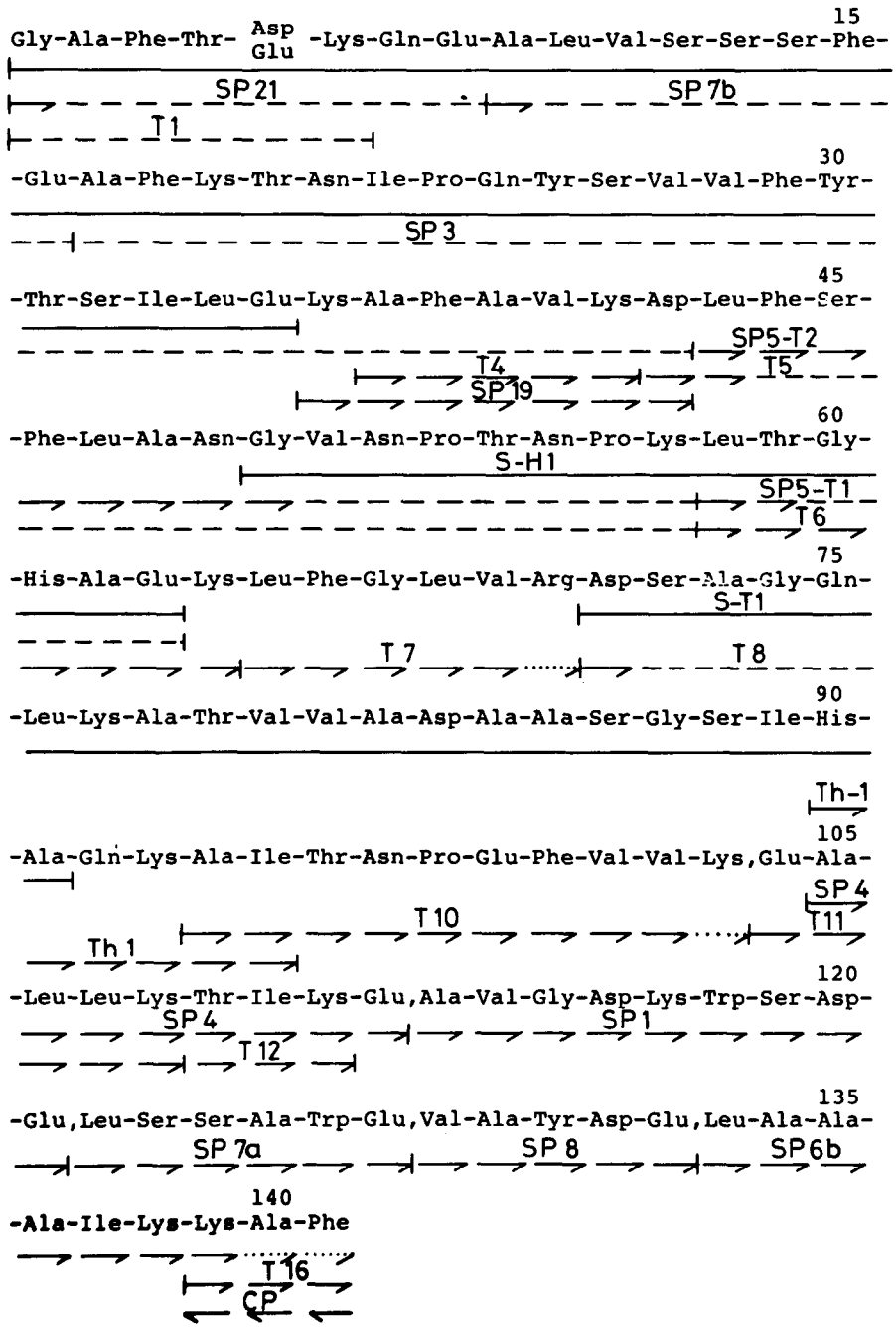

Fig.1. The complete amino acid sequence of soybean leghaemoglobin c_2. Regions of the sequence determined by automated sequenator analysis are indicated by solid lines. Residues directly identified as the PTH derivative and/or by the subtractive procedure are indicated thus ———>. Residues occurring at the carboxyl-termini of peptides and identified by deduction are identified thus · · · >. Residues identified by digestion with carboxypeptidase (A+B) are indicated thus <CP——. Peptides isolated and identified by their amino acid compositions are indicated by dashed lines. Peptides generated by digestion with trypsin, staphylococcal protease and thermolysin are prefixed respectively T, SP and Th. Large peptides, sequenced by sequenator analysis, were generated from succinylated apoleghaemoglobin c_2 by hydroxylamine digestion (S-HI) and trypsin digestion (S-TI).

University of Melbourne. J. G. R. H. was supported by the Commonwealth Serum Laboratories, Parkville, Victoria and a Commonwealth Post-graduate Research Award. The authors wish to acknowledge Dr C. A. Appleby for his assistance in preparing leghaemoglobin c_2. The assistance of A. S. Inglis, A. C. Fairchild and R. Goulet in the automated sequence analyses is gratefully acknowledged.

References

[1] Thorogood, E. (1957) Science 126, 1011–1012.
[2] Ellfolk, N. (1960) Acta Chem. Scand. 14, 607–616.
[3] Appleby, C. A., Nicola, N. A., Hurrell, J. G. R. and Leach, S. J. (1975) Biochemistry 14, 4444–4450.
[4] Ellfolk, N. and Sievers, G. (1971) Acta Chem. Scand. 25, 3532–3534.
[5] Nicola, N. A., Minasian, E., Appleby, C. A. and Leach, S. J. (1975) Biochemistry 14, 5141–5149.
[6] Ellfolk, N. and Sievers, G. (1975) Biochim. Biophys. Acta 405, 213–227.
[7] Wittenberg, B. A., Wittenberg, J. B. and Appleby, C. A. (1973) J. Biol. Chem. 248, 3178–3182.
[8] Wittenberg, J. B., Bergersen, F. J., Appleby, C. A. and Turner, G. L. (1974) J. Biol. Chem. 249, 4057–4066.
[9] Appleby, C. A. (1969) Biochim. Biophys. Acta 189, 267–279.
[10] Antonini, E. and Brunori, M. (1971) Haemoglobin and Myoglobin in their Reactions with Ligands, pp. 124–126, North-Holland, Amstedam.
[11] Bornstein, P. (1970) Biochemistry 9, 2408–2420.
[12] Houmard, J. and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[13] Vigna, R. A., Gurd, L. J. and Gurd, F. R. N. (1974) J. Biol. Chem. 249, 4144–4148.
[14] Mendez, E. and Lai, C. Y. (1975) Anal. Biochem. 65, 281–292.
[15] Ambler, R. P. (1972) Methods in Enzymology 25, 143–154.
[16] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[17] Inglis, A. S. and Burley, R. W. (1977) FEBS Lett. 73, 33–37.
[18] Van Eerd, J-P. and Takahashi, K. (1976) Biochemistry 15, 1171–1180.
[19] Jacobs, J. W. and Niall, H. D. (1975) J. Biol. Chem. 250, 3629–3636.
[20] Konigsberg, W. (1972) Methods in Enzymology 25, 326–332.
[21] Tanaka, S. and Scheraga, H. A. (1976) Macromolecules 9, 168–182.
[22] Hurrell, J. G. R., Nicola, N. A., Broughton, W. J., Dilworth, M. J., Minasian, E. and Leach, S. J. (1976) Eur. J. Biochem. 66, 389–399.
[23] Appleby, C. A. (1962), Biochim. Biophys. Acta 60, 226–235.
[24] Fuchsman, W. H., Barton, C. R., Stein, M. M., Thompson, J. T. and Willett, R. M. (1976) Biochem. Biophys. Res. Comm. 68, 387–392.