THE AMINO ACID SEQUENCE OF LEGHAEMOGLOBIN I FROM ROOT NODULES OF BROAD BEAN (*VICIA FABA* L.)

M. RICHARDSON, M. J. DILWORTH* and M. D. SCAWEN**
Department of Botany, University of Durham, Durham City, DH1 3LE, England

Received 12 September 1974
Revised version received 10 October 1974

1. Introduction

The leghaemoglobins are a group of proteins found in effective N_2-fixing nodules of legumes formed after infection with *Rhizobium* species. The amino acid sequences of soybean (*Glycine max* Merr.) leghaemoglobin I [1,2] and of kidney bean (*Phaseolus vulgaris* L.) leghaemoglobin (Lb) [3] have been published. Amino-acid composition data is available for a number of other legume Lb's (4). A number of analogies have been drawn between the sequences of Lb and those of haemoglobin (Hb), myoglobin (Mb), and other globins [5,6]. Since it is known that the type of Lb produced in a legume root nodule is genetically determined by the plant [7–9], and that the amino acid compositions of the Lb's show marked variation [4], it seems likely that Lb may be a useful protein for making phylogenetic comparisons between the legumes. On comparing the sequences of Lb from kidney bean and soybean, it is clear that considerable differences occur between tribes of the Fabaceae which are relatively closely related. We have determined the sequence of broad bean Lb I in order to indicate the sort of variability likely to occur between even more widely different tribes, and we now present its sequence.

* On sabbatical leave from Department of Soil Science and Plant Nutrition, University of Western Australia, Nedlands, Western Australia 6009.

** Present address: Department of Biochemistry, University of Newcastle-upon-Tyne, NE1 7RU, England.

2. Materials and methods

2.1. *Protein purification*

Nodules from field-grown *Vicia faba* cv. Early white plants inoculated with strain WU47 of *Rhizobium leguminosarum* were harvested at about 8 weeks, washed and homogenized under N_2 in 4 vol of 0.1 M potassium phosphate buffer (pH 7.6) containing 2.5 mM EDTA. After centrifugation under N_2, the Lb was fractionated out between 0.48 and 0.80 saturation with solid ammonium sulphate at 4°C, and the resulting precipitate dissolved in 0.1 M phosphate buffer (pH 7.6), and dialyzed overnight against distilled water. The centrifuged Lb solution was adsorbed onto a column (2.2 × 10 cm) of DEAE-cellulose (Whatman DE52) equilibrated with 3 mM phosphate buffer (pH 6.8) and non-adsorbed material (including interfering carbohydrate) washed through with 50 ml 3 mM phosphate buffer. Lb was then eluted with 50 mM phosphate buffer (pH 6.8), dialyzed free of phosphate, concentrated, and chromatographed on a column (2.2 × 48 cm) of DEAE-cellulose equilibrated with 13 mM sodium acetate buffer (pH 5.2) [10]. Lb I was recovered from the column as a green acetate complex [11] on elution with 13 mM acetate buffer (pH 5.2) at a flow rate of 45 ml/hr. After the pH had been adjusted to 7.0 with 1 M Tris buffer (pH 8.0), the Lb was concentrated to a small volume in a pressure cell fitted with an Amicon UM-10 membrane, and subjected to gel filtration on a column (2.5 × 90 cm) of Sephadex G-75 in water, at a flow rate of 45 ml/hr. The Lb recovered from this column had a single N-terminal amino-acid (glycine) and was used for

sequence determination. Haem was removed from it according to Broughton and Dilworth [4], the globin being dissolved in water and then freeze-dried before digestion.

2.2. *Sequence determination*

The apoprotein was cleaved with cyanogen bromide as described previously [12] and the resulting fragments separated by gel-filtration on Sephadex G-50. Peptides were obtained from the intact apoprotein and fragments by digestion with trypsin, thermolysin and chymotrypsin, and by cleavage with *N*-bromosuccinimide [12]. Mixtures of peptides were fractionated by gel filtration, high-voltage paper electrophoresis and paper chromatography [12].

The amino acid sequences of small fragments and peptides were determined by the dansyl-Edman procedure and by digestion with carboxypeptidase A [12]. Amino acid analyses were obtained using a Locarte amino acid analyzer.

Automated sequence analysis of the N-terminal region of the intact apoprotein was carried out on a Beckman model 890 C automatic sequencer using the fast-protein programme. Phenylthiohydantoin amino acids were identified by thin-layer chromatography, by gas-liquid chromatography, and by hydrolysis in constant-boiling hydriodic acid followed by dansylation as described previously [13].

3. Results and discussion

The complete amino acid sequence of broad bean Lb I is shown in fig.1. together with the details of the fragments and peptides from which it was deduced. The apoprotein contained 143 amino acids which corresponds to an approximate mol. wt of 15 500. A more accurate mol. wt cannot be calculated owing to the high degree of polymorphism exhibited by this protein.

The glycine N-terminus is a feature common to soybean Lb II, kidney bean Lb and broad bean Lb I; N-terminal valine appears only in soybean Lb I [1] and in Lb I from snake bean (*Vigna sesquipedalis*) [14].

In the overall sequence of broad bean Lb, there appear to be at least two insertions into the soybean sequence (residues 55 and 103; fig.2), one of which is also found in kidney bean Lb. Whether one measures the differences in terms of residues to be altered in the protein, or minimum base changes, the difference of the broad bean sequence from soybean or kidney bean Lb's is greater than the difference between kidney bean and soybean Lb's. The diversity is considerably greater than that found, for example, with plant cytochromes *c* [15]. It is noteworthy that there are two methionine residues in the broad bean sequence, both replacing leucine in the other sequences. The occurrence of methionine in Lb was first suggested for yellow lupin I and II [4].

There is an obvious similarity to Mb in that the molecule contains no cysteine, but does contain methionine. It is different from Hb which may contain both. Like soybean and kidney bean Lb's, the 'basic centre' near the probable haem-binding histidines is absent from broad bean Lb I. Histidine-101 in the broad bean sequence seems unlikely to be involved in haem binding in view of the other two sequences, and the constant positioning of His-61 and His-92 in all three. If nicotinic acid binding is a general property of Lb's, and the carboxyl group interaction with a basic residue is required, Lys-40 and His-61 are two of the positions suggested [16] which are constant in all three sequences, while Lys-64 is absent from *Phaseolus* Lb and polymorphic in broad bean Lb I.

Most of the key features of the soybean Lb I in relation to Mb and Hb appear to be present in broad bean and kidney bean Lb's, thus indicating the likelihood of some invariant regions in Lb's (fig.2). However, the leucine at position 88 considered by Ellfolk [5] to be analagous to the near invariant leucine at F4 for other globins (including kidney bean Lb) is not present in broad bean Lb I. The proline residues considered [1] to be near the beginning of the B, C, D, E and G helices are present, but with the exception of Pro-49. However, the wide variation between the three Lb sequences indicates that many residues in the molecule must be mutable without loss of function; this is presumed to be that of oxygen transport at high flux but at low absolute tension [17,18].

The polymorphism noted in soybean Lb I is more obvious still in the broad bean sequence, with examples at positions 4, 5, 22, 23, 34, 41, 59, 64, 71, 74, 77, 126, 128 and 134. The presence of these polymorphic positions is not an artifact of sequence determination as they have been recognized from families of peptides from more than one type of enzymatic digestion. Most

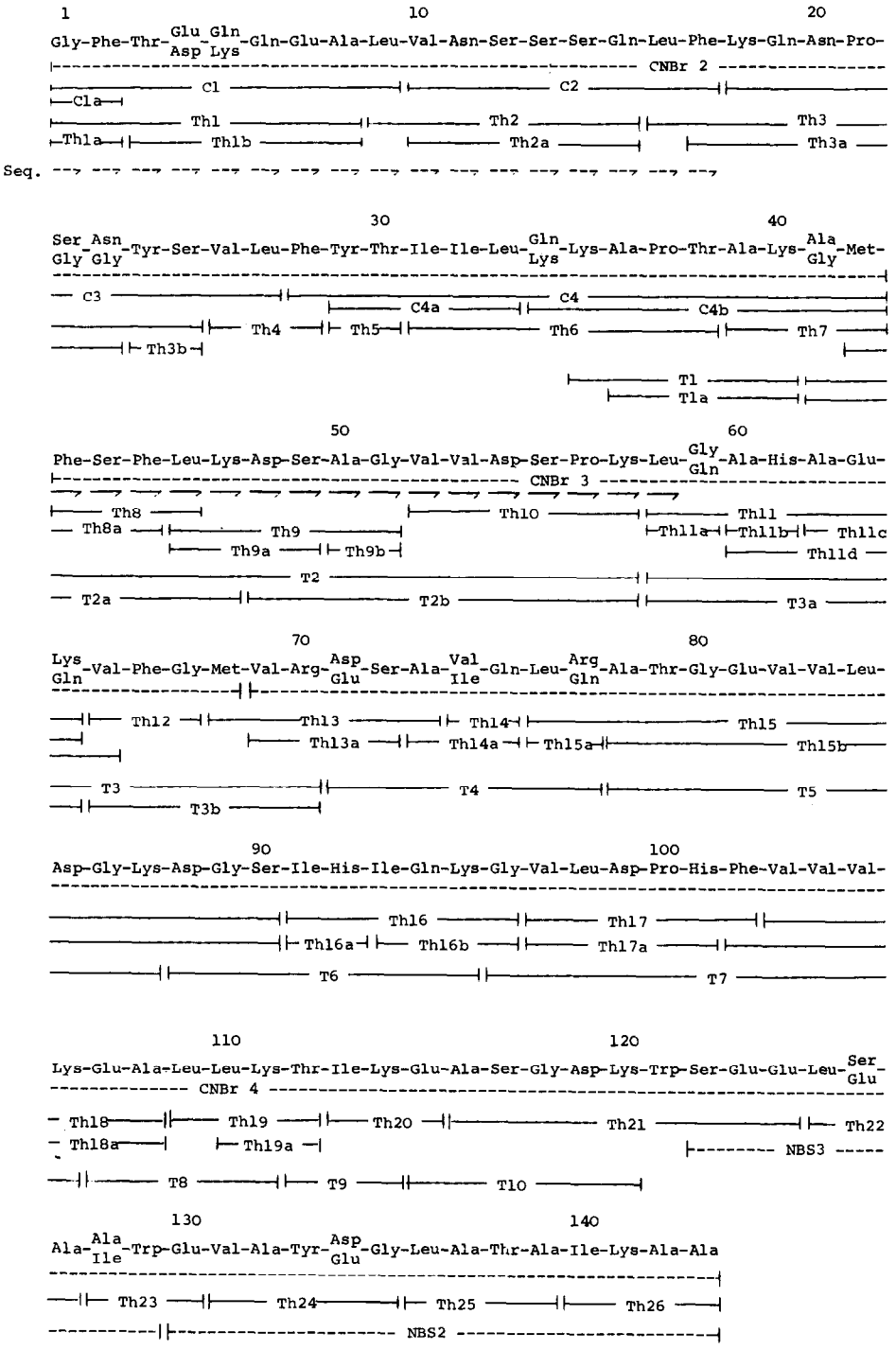

Fig.1. The amino acid sequence of leghaemoglobin I from broad bean (*Vicia faba*). ----CNBR---- fragments obtained by cleavage of protein with cyanogen bromide; ----NBS---- fragments obtained by cleavage with *N*-bromosuccinimide; --→ results from automated sequencer; —→ results from direct manual dansyl-Edman analysis of fragment; ———— peptides sequenced by dansyl-Edman analysis; C = chymotryptic peptides, Th = thermolytic peptides, T = tryptic peptides.

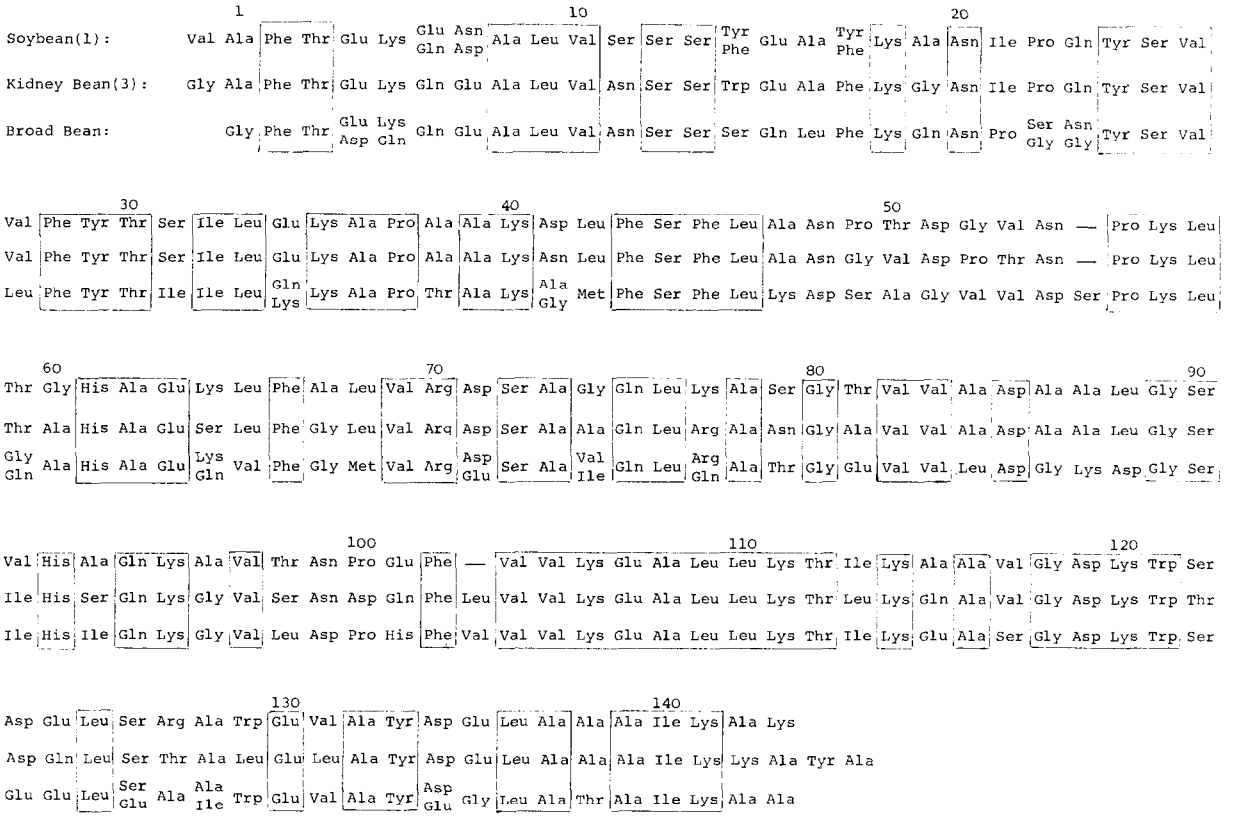

Fig. 2. Comparison of the amino acid sequences of leghaemoglobin from soybean (*Glycine max*) (1) kidney bean (*Phaseolus vulgaris*) (3) and broad bean (*Vicia faba*). Invariant regions of the sequences are indicated by the residues in boxes. The numbering of the residues is as adopted for the sequence of broad bean (fig. 1) and not the alternative schemes (1,3) which makes no allowance for the apparent insertions of residues at positions 55 and 103.

of such polymorphisms are not apparently detected by ion-exchange chromatography, but may be the reason for the multiple bands seen when purified Lb preparations are subjected to isoelectric focussing electrophoresis in polyacrylamide gels [9]. That they occur in a Lb preparation which may contain material from 600 different plants is not surprising in view of the obvious scope for differences between even closely related plants such as soybean and kidney bean.

With such differences between broad bean, kidney bean and soybean Lb's within the family Fabaceae of the Leguminales, even larger differences may be expected when sequence data is available for Lb's from the families Mimosaceae and Caesalpiniae.

Acknowledgement

We wish to thank Professor D. Boulter for the provision of certain facilities, and in particular for the use of the Beckman model 890 C automatic sequencer.

References

[1] Ellfolk, N. and Sievers, G. (1971) Acta Chem. Scand. 25, 3532–3534.
[2] Ellfolk, N. and Sievers, G. (1972) Acta Chem. Scand. 26, 1155–1165.
[3] Lehtovaara, P. and Ellfolk, N. (1974) FEBS Lett. 43, 239–240.

[4] Broughton, W. J. and Dilworth, M. J. (1973) Biochim. Biophys. Acta 317, 266–276.
[5] Ellfolk, N. (1972) Endeavour 21, 139–142.
[6] Dayhoff, M. O., Hunt, L. T., McLaughlin, P. J. and Jones, D. D. (1972) in: Atlas of Protein Structure (Dayhoff, M. O., ed.) Vol. 5, pp. 17–30, National Biomedical Research Foundation, Washington.
[7] Dilworth, M. J. (1969) Biochim. Biophys. Acta 184, 432–441.
[8] Cutting, J. A. and Schulman, H. M. (1971) Biochim. Biophys. Acta 229, 59–62.
[9] Broughton, W. J. and Dilworth, M. J. (1971) Biochem. J. 125, 1075–1081.
[10] Ellfolk, N. (1960) Acta Chem. Scand. 14, 609–616.
[11] Ellfolk, N. (1961) Acta Chem. Scand. 15, 975–984.
[12] Richardson, M. (1974) Biochem. J. 137, 101–112.
[13] Scawen, M. D., Ramshaw, J. A. M., Brown, R. H. and Boulter, D. (1974) Eur. J. Biochem. 44, 299–303.
[14] Appleby, C. A. and Dilworth, M. J. (1974) in: Dinitrogen Fixation (Burns, R. C., ed.) Vol 1, part 2 Wiley, in press.
[15] Ramshaw, J. A. M., Richardson, D. L., Meatyard, B. T., Brown, R. H., Richardson, M., Thompson, E. W. and Boulter, D. (1972) New Phytol. 71, 773–779.
[16] Appleby, C. A., Wittenberg, B. A. and Wittenberg, J. B. (1973) Proc. Natl. Acad. Sci., US 70, 564–568.
[17] Tjepkema, J. D. (1971) Ph. D. thesis, Univ. Michigan.
[18] Bergersen, F. J., Turner, G. L. and Appleby, C. A. (1973) Biochim. Biophys. Acta 292, 271–282.