The Amino Acid Sequence of Soybean *(Glycine max)* Leghemoglobin *c*

GUNNEL SIEVERS, MARJA-LIISA HUHTALA and NILS ELLFOLK

Department of Biochemistry, University of Helsinki, Unioninkatu 35, SF-00170 Helsinki 17, Finland

The amino acid sequences of leghemoglobins from soybean (*Glycine max*),[1,2] kidney bean (*Phaseolus vulgaris*),[3] broad bean (*Vicia faba*),[4] and yellow lupin (*Lupinus luteus*) [5] have been determined. All these proteins consist of single chains with molecular weight of about 15 000–16 000 daltons. The sequences of the leghemoglobins are closely similar, and they have features in common with those of animal origin. In this study the amino acid sequence of soybean leghemoglobin *c* ('fast' main component) is preliminarily reported. A detailed report will be given later.

Materials and methods. Leghemoglobin *c* and its apoprotein were prepared as previously described.[6,7] The apoprotein was digested with trypsin, chymotrypsin and thermolysin. The peptides obtained were separated by ion exchange chromatography, high voltage paper electrophoresis and paper chromatography.[8] Subtractive Edman and dansyl-Edman degradations, leucine aminopeptidase and carboxypeptidase A were used to elucidate the sequences of the peptides.[9] The *N*-terminal of the protein was determined by the Edman method and the *C*-terminal by hydrazinolysis.[9]

Results and discussion. The complete amino acid sequence of soybean leghemoglobin *c* is shown in Fig. 1 (p. 724). The *N*-terminal sequence of the protein is NH_2-Gly-Ala-Phe-Thr as determined by the Edman procedure. Hydrazinolysis of the protein gave equal amounts of lysine and phenylalanine.

Twenty-one tryptic peptides, one ditryptic peptide and free lysine were obtained from the tryptic hydrolysis of apoleghemoglobin *c*. The large number of tryptic peptides, as previously demonstrated by the fingerprint,[10] is a result of polymorphism at six positions (8, 20, 39, 79, 97, and 143) in the protein, giving six pairs of tryptic peptides in addition to eight single peptides. One Lys-Lys bond was found, at positions 140–141. The overlaps between the tryptic peptides were established with chymotryptic and thermolytic peptides. The polypeptide chain consists of 143 amino acids, one more than component *a*, giving an approximate molecular weight of 15 300 daltons for the apoprotein and 15 950 daltons for leghemoglobin *c*. The exact molecular weight cannot be calculated because of the polymorphism.

Leghemoglobin *c* contains two histidines at positions equivalent to those of other known leghemoglobins. In all leghemoglobins the distance between the histidines is 30 amino acids, two residues longer than in animal globins. Soybean leghemoglobin *c*, like component *a*,[1] kidney bean leghemoglobin *a* [3] and snake bean (*Vigna sinensis*) leghemoglobin,[11] has no sulfur-containing amino acids. This, however, is not an invariable rule for leghemoglobins, because broad bean leghemoglobin I contains two methionines [4] and lupine leghemoglobin I one methionine.[5]

The sequence of soybean leghemoglobin *c* differs completely from that of component *a* at only six positions (1, 31, 91, 114, 126 and 143). In addition, there is an insertion of lysine after position No. 140. At the positions of polymorphism one alternative is always identical to that of component *a*.

Leghemoglobin *c* has been reported to be resolved into two fractions on a DEAE-column,[12] and the amino acid composition of the fractions was found to differ slightly.[12] The difference, however, is not in congruence with the sequence elucidated here. The only difference we found between the fractions was in the *C*-terminal. The fraction first emerging from the DEAE-column has lysine and the later fraction phenylalanine as the *C*-terminal amino acid.

1. Ellfolk, N. and Sievers, G. *Acta Chem. Scand. 25* (1971) 3532.
2. Ellfolk, N. and Sievers, G. *Acta Chem. Scand. B 28* (1974) 1245.
3. Lehtovaara, P. and Ellfolk, N. *Eur. J. Biochem. 54* (1975) 577.
4. Richardson, M., Dilworth, M. J. and Scawen, M. D. *FEBS Lett. 51* (1975) 33.
5. Jegorov, C. A., Feigina, M. Iu., Kazakov, V. K., Shahparanov, M. J., Mitaleva, S. U. and Ovchinikov, Iu. A. *Bioorg. Khim. 2* (1976) 125.
6. Ellfolk, N. *Acta Chem. Scand. 14* (1960) 609.
7. Ellfolk, N. *Acta Chem. Scand. 15* (1961) 545.
8. Ellfolk, N. and Sievers, G. *Acta Chem. Scand. 26* (1972) 1155.
9. Ellfolk, N. and Sievers, G. *Acta Chem. Scand. 27* (1973) 3371.
10. Ellfolk, N. *Acta Chem. Scand. 16* (1962) 831.
11. Broughton, W. J. and Dilworth, M. J. *Biochem. J. 125* (1971) 1975.
12. Appleby, C. A., Nicola, N. A., Hurrell, J. G. R. and Leach, S. J. *Biochemistry 14* (1975) 4444.

Received August 5, 1977.

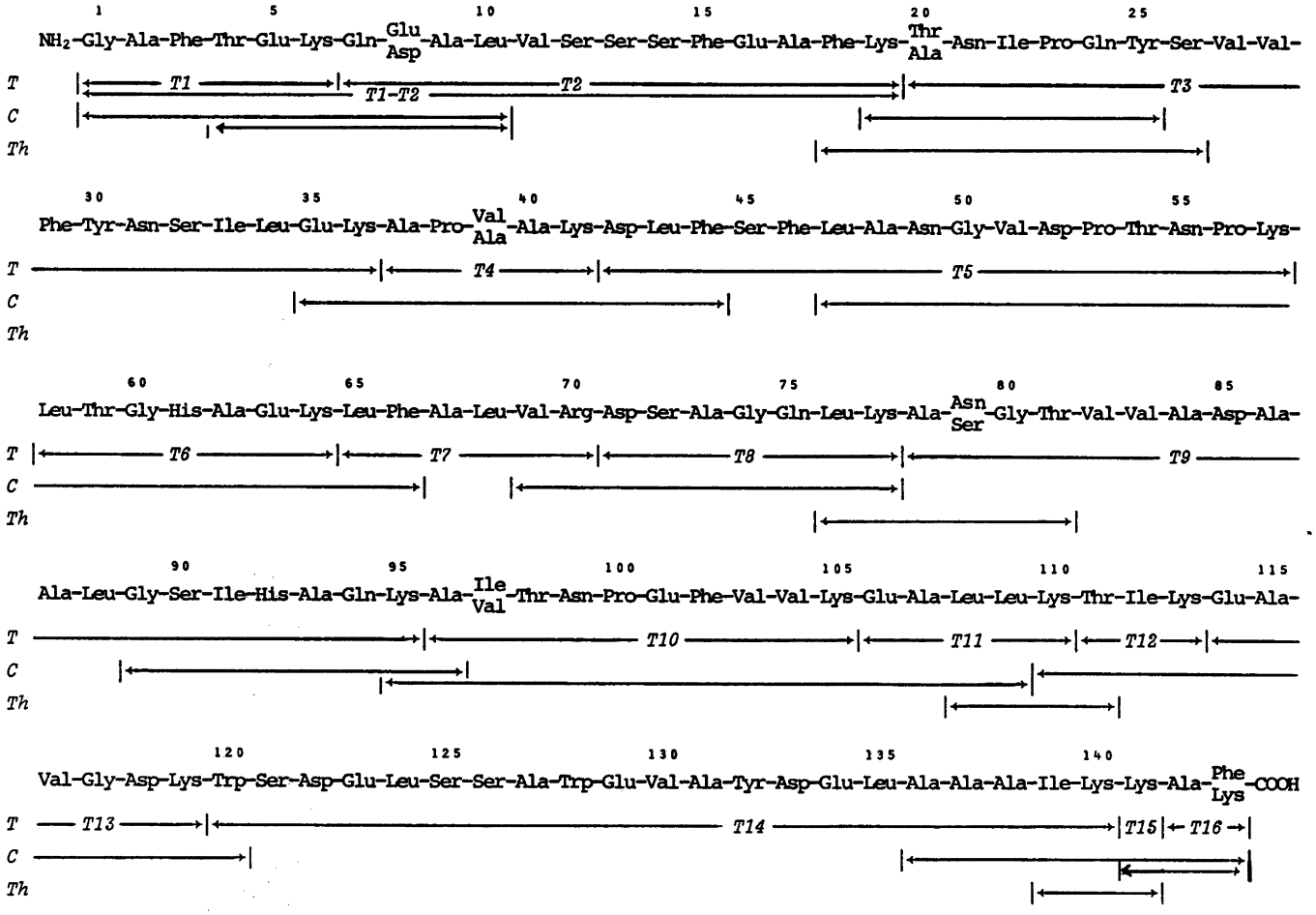

Fig. 1. The amino acid sequence of soybean leghemoglobin *c*. Tryptic peptides (*T*), chymotryptic peptides (*C*) and thermolytic peptides (*Th*) are indicated by horizontal lines.