The amino acid sequence of hemoglobin I from *Parasponia andersonii*, a nonleguminous plant

Alexander A Kortt, John E. Burns, Michael J. Trinick+ and Cyril A Appleby+

CSIRO Division of Protein Chemistry, Parkville (Melbourne) 3052, Victoria and +*CSIRO Division of Plant Industry, Canberra 2601, Australia*

Received 22 October 1984

The complete amino acid sequence of the hemoglobin I from nitrogen-fixing root nodules of the nonleguminous plant, *Parasponia andersonii*, has been determined This dimeric protein consists of two identical polypeptide chains of 155 amino acids and shows extensive sequence homology with other hemoglobins Homology between the hemoglobin I of *P andersonii* and the leghemoglobins of lupin and soybean nodules is 41 and 39%, respectively The predicted secondary structure of *P andersonii* hemoglobin I has a high content of α-helix, except for the E-helix, similar helices were predicted as those in the leghemoglobins The close homology of the sequences provides evidence that this nonleguminous hemoglobin shares the same genetic origin as the legume and animal hemoglobins

Hemoglobin *Nonleguminous plant* Parasponia andersonii *Amino acid sequence* *Sequence homology* *Predicted secondary structure* *Genetic origin*

1. INTRODUCTION

In higher plants, hemoglobin was generally thought to occur only in the nitrogen-fixing root nodules of legumes [1,2]. Recently, however, a hemoglobin was shown to occur [3] in the nitrogen-fixing root nodules of *Parasponia*, a nonleguminous plant of the Ulmaceae family [4,5]. This plant was nodulated by strains of *Rhizobium* that also nodulate certain members of the Leguminosae [3–5]. A dimeric hemoglobin was purified from the nitrogen-fixing root nodules of *Parasponia andersonii* and the oxygen affinity was found to be sufficiently high (unpublished) to allow *P. andersonii* hemoglobin to function in a similar way to that of legume hemoglobin during symbiotic nitrogen fixation [3].

DNA sequencing studies suggest that legume hemoglobins and animal globins share the same genetic origin and it has been suggested that the gene for legume hemoglobin was transferred to legumes from another eukaryote outside the plant kingdom, relatively recently in evolutionary history [6,7]. However, the identification of hemoglobin in *Parasponia* [3], *Casaurina* and other nonlegume species [8–10] questions this proposal and suggests that hemoglobin genes may have persisted in many plant families since divergence of the plant and animal lines some 1000 million years ago [1,7]. Structural studies on *P. andersonii* hemoglobin I were undertaken to elucidate its relationship to other hemoglobins. We report here the complete amino acid sequence of *P. andersonii* hemoglobin I and compare it with other legume hemoglobin sequences.

2. EXPERIMENTAL

The major hemoglobin I was isolated from fresh nodules from *P. andersonii* Planch, infected with *Rhizobium* strain CP283 and purified to homogeneity as in [5]. A minor component (hemoglobin II) was not examined in this study. The protein was dehemed as described [11]. Separate aliquots of the native protein (~50 nmol) in 0.05 M ammonium bicarbonate, pH 8, were

Published by Elsevier Science Publishers B V (Biomedical Division)
00145793/85/$3 30

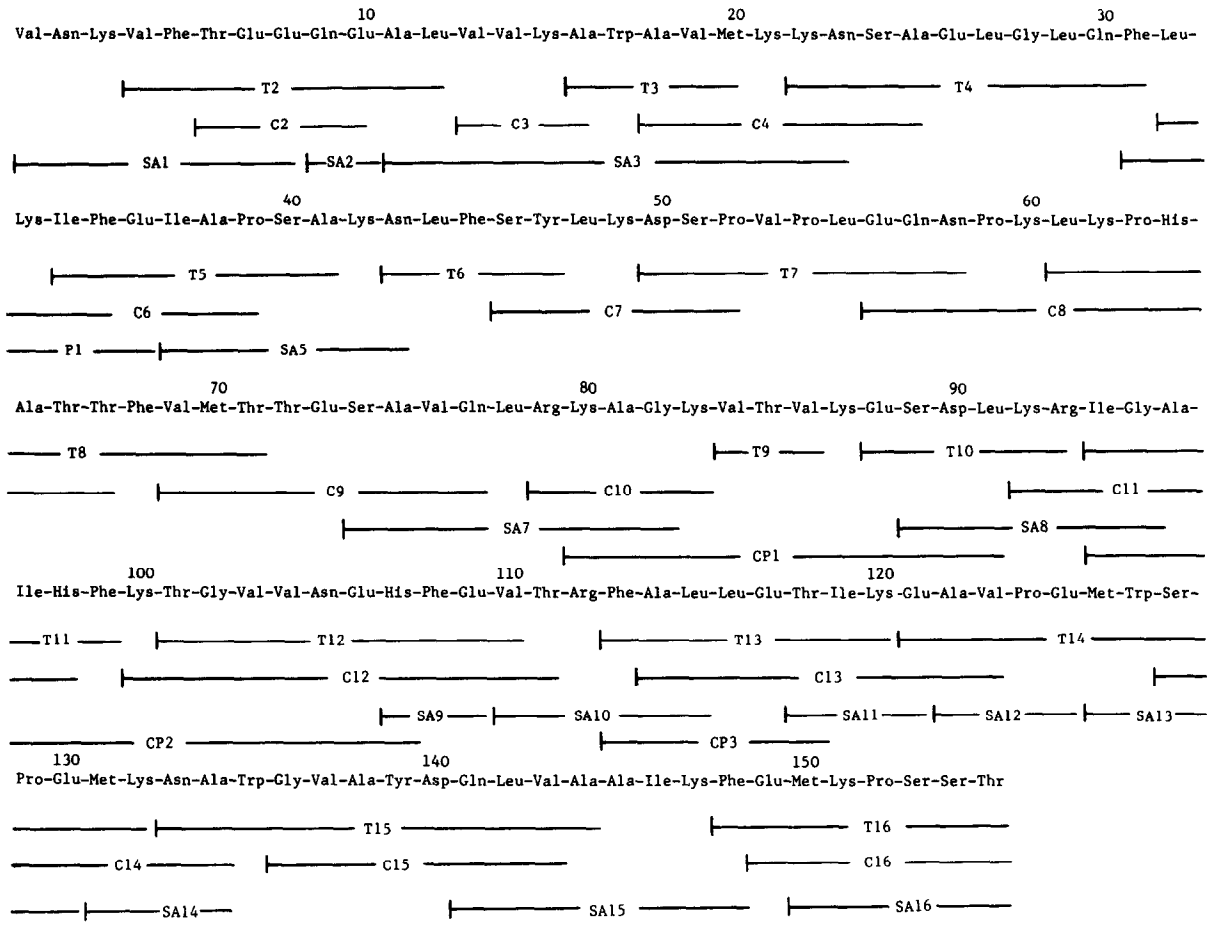

Fig 1 Primary structure of hemoglobin I from nitrogen-fixing root nodules of *P. andersonii* Residues sequenced are indicated by the solid lines, T, tryptic peptides; C, chymotryptic peptides, SA, *S aureus* V8 protease peptides; CP, clostripain peptides; P, peptic peptides

digested with trypsin (Worthington), chymotrypsin (Worthington), clostripain (Calbiochem-Behring) and *Staphylococcus aureus* V8 protease (Pierce) at 37°C for 4 h at an enzyme:substrate ratio of 1:50 (w/w). The dehemed protein, which was insoluble in 0.05 M ammonium bicarbonate, pH 8, was digested only by trypsin to yield soluble peptides; the other proteases failed to digest the insoluble dehemed protein. By contrast the dehemed protein was soluble at pH 1.7, and was digested with pepsin (Worthington) at 37°C for 4 h using an enzyme:substrate ratio of 1:50 (w/w). The digests were lyophilized and redissolved in 0.05% (v/v) trifluoroacetic acid. Insoluble material was removed by centrifugation and the soluble peptides were isolated by HPLC in 0.05% (v/v) trifluoroacetic acid on a Waters μBondapak C_{18} column using an acetonitrile gradient.

Peptides were sequenced manually by a modified Edman procedure [12] using 50% pyridine as the coupling buffer and extracting with *n*-heptane–ethyl acetate (2:1, v/v) instead of benzene. Residues were identified as PTH-amino acid derivatives by HPLC gradient elution on a Zorbax C_{18} column (Du Pont). Amino acid analysis of the protein and peptides was performed on a modified Beckman 120C analyser after hydrolysis in 6 N HCl for 24 h at 110°C. The secondary structures of *P. andersonii* hemoglobin I (P.HbI), soybean leghemoglobin *a* (S Lb*a*) and lupin leghemoglobin II (L.LbII) were predicted from the amino acid sequences as in [13].

3. RESULTS AND DISCUSSION

The complete amino acid sequence of *P. andersonii* hemoglobin I is shown in fig.1. Sequence data necessary for the unequivocal determination of the primary structure are illustrated. The polypeptide chain contains 155 amino acid residues corresponding to an M_r of 18044 (including the heme group) which is consistent with the M_r of 36000 determined in the ultracentrifuge for the dimeric ferric hemoglobin by the meniscus depletion sedimentation equilibrium method (G. Lilley, personal communication). The sequence is in agreement with the amino acid composition of the protein (table 1).

The same tryptic peptides were isolated from digests of dehemed and native protein and they accounted for 148 residues of the molecule. The amino terminal tripeptide, residue 80(Lys) and the tripeptide, Ala[81]-Gly[82]-Lys[83], were not found in either tryptic digest. Chymotryptic peptides isolated from the native protein accounted for 150 residues and the only peptide not isolated was the amino-terminal pentapeptide (fig.1). Automatic sequence analysis of the native hemoglobin provided sequence data up to residue 16 [14] which indicated that this hemoglobin was homologous with the leghemoglobins. Peptides T2, C3 and SA1 confirmed this sequence. Tryptic peptides T2, T3 and T4 were overlapped by C4 and SA3 to establish the sequence to residue 31. Peptic peptide, P1, derived from a digest of the dehemed protein provided the overlap of peptides T4, T5 and C6 to extend the sequence to residue 41. Peptides T6, T7 and T8 were aligned by the sequences of SA5, C7 and C8 extending the sequence to residue 71. Peptides C9, C10 and CP1 were overlapped with peptide SA7 and CP1 aligned T9 and T10 to extend the sequence to residue 91. Overlap peptide SA8 connected CP1 and CP2, and peptide CP2 aligned peptides T11, T12, C11 and C12. Peptide SA10 connected T12 and T13 and peptide C13 connected T13 and T14 to extend the sequence to residue 132. The sequence was completed by aligning T15 and T16 with overlap peptides C14, SA14 and SA15. The carboxyl terminal amino acid was identified as threonine from the composition and sequence data of 3 different cleavages. Upon digestion of *P. andersonii* hemoglobin I with carboxypeptidase Y, Thr was the first residue released.

Table 1

Amino acid composition of *P. andersonii* hemoglobin I

Amino acid	Residues per polypeptide	
	Analysis[a]	Sequence
Lys	16.9	18
His	2 7	3
Arg	2.8	3
Asx	8.8	9
Thr	8 3[b]	10
Ser	8.5[b]	9
Glx	20.4	20
Pro	8.5	8
Gly	5 0	5
Ala	15.6	16
½Cys	0	0
Val	16.7	16
Met	4.6	5
Ile	5.7	6
Leu	12.7	13
Tyr	1.7	2
Phe	9.5	9
Trp	2.4[c]	3
Total	150 8	155

[a] Samples hydrolysed for 24 h at 110°C under vacuum in 5.7 N HCl

[b] Values obtained by extrapolation to zero hydrolysis time

[c] Determined after hydrolysis of sample in 4 M $MeSO_3H$ containing 0.2% (w/v) tryptamine for 24 h at 115°C under vacuum

The sequence of *P. andersonii* hemoglobin I shows extensive homology with the sequence of lupin leghemoglobin II [15] and soybean leghemoglobin *a* [6] (fig.2). Maximum homology was obtained with an alignment containing 3 single residue deletions in the *P. andersonii* hemoglobin I sequence, 2 deletions in the lupin leghemoglobin II sequence and 7 deletions in the soybean leghemoglobin *a* sequence (fig.2). In this alignment these 3 plant hemoglobins show 30% homology with 51 identical positions. When compared separately the homology between *P. andersonii* hemoglobin I with 155 residues and lupin leghemoglobin II with 153 residues is 41%, and that with soybean leghemoglobin *a* with 143 residues is 39%. This compares with 54% homology between lupin leghemoglobin II and

soybean leghemoglobin *a*. All the known legume hemoglobin (leghemoglobin) sequences contain 50 invariant residues [16]; 38 of these invariant residues occur in the *P. andersonii* (nonlegume) hemoglobin I sequence and a further 7 invariant residues of the leghemoglobins are replaced by conservative amino acid substitutions including 3 methionines replacing Phe[20], Leu[70] and Leu[131] in

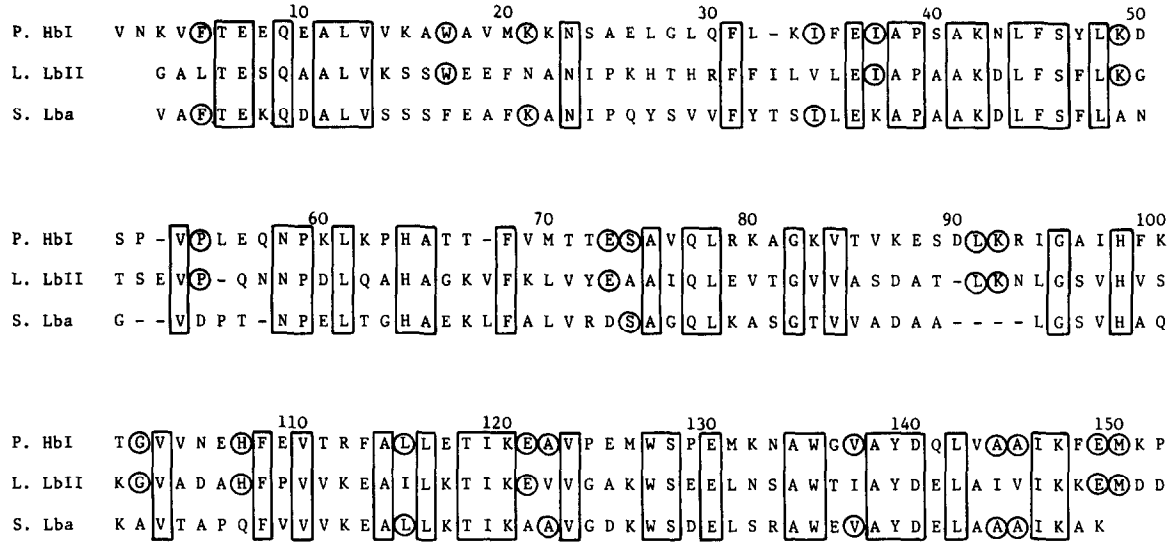

Fig.2. Comparison of the amino acid sequences of hemoglobin I from *P andersonii* (P.HbI) (this work), lupin leghemoglobin II (L LbII) [15] and soybean leghemoglobin *a* (S.Lba) [6] Boxed residues are identical in the 3 hemoglobin sequences. Ringed residues are identical in *P. andersonii* hemoglobin I and only one of the leghemoglobins. (– – –) Deletions introduced to maximize homology

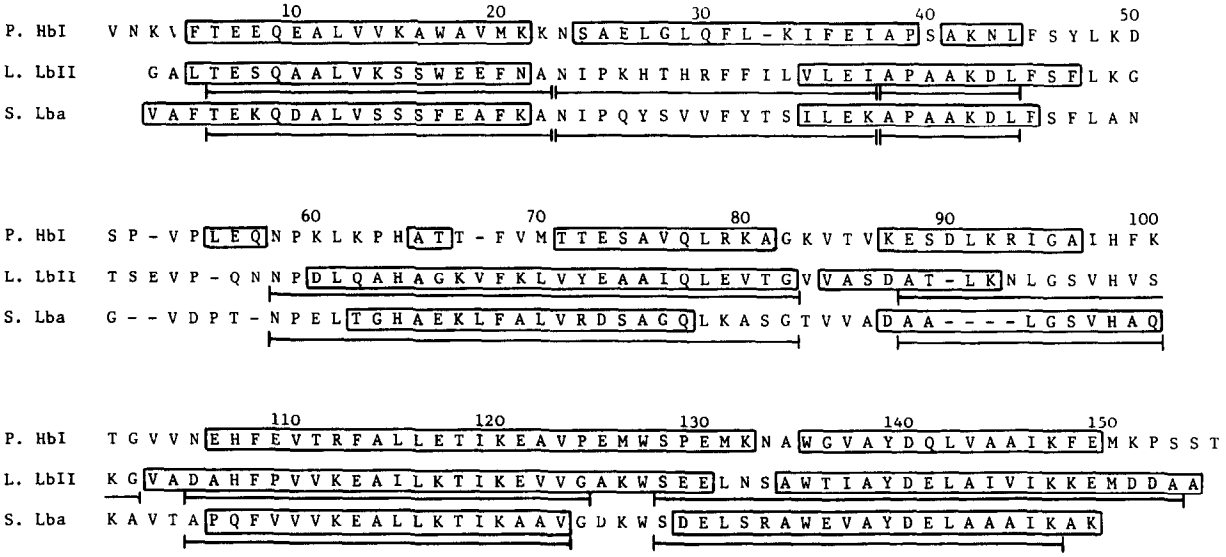

Fig.3. Comparison of the predicted α-helical segments of *P. andersonii* hemoglobin I (P.HbI) and the predicted and known α-helical segments of lupin leghemoglobin II (L LbII) and soybean leghemoglobin *a* (S.Lb*a*) Boxes enclose the segments predicted as helices and (⊢—⊣) designates the segments identified as helices from X-ray crystallographic studies [16,19].

the leghemoglobins. *P. andersonii* hemoglobin I contains 5 methionine residues and Met^{150} is identical with the single methionine in lupin leghemoglobin II (fig.2). The sequence of *P. andersonii* hemoglobin I shows almost no homology with a purported partial amino acid sequence of hemoglobin from *P. rigida* (a closely related species) derived from DNA hybridization and sequencing studies [17]. Presumably this DNA-derived sequence is not of the hemoglobin.

Structural studies [18] have shown that hemoglobins from various sources are highly helical molecules. Soybean [16] and lupin [19] leghemoglobins contain 7 helical regions which represent ~77% of the structure and these regions are arranged in the well known myoglobin fold [20]. The secondary structure of *P. andersonii* hemoglobin I predicted from the amino acid sequence [13] has a high content of α-helix (67%) with 10% β-sheet and 9% β-turns and the helical regions are illustrated in fig.3. The secondary structures of soybean leghemoglobin *a* and lupin leghemoglobin II predicted from their amino acid sequences are also shown in fig.3 and they are compared with the known structures from X-ray crystallographic studies [16,19]. The predicted α-helical regions of soybean leghemoglobin *a* and lupin leghemoglobin II show a good correlation with the experimentally determined helical regions except that a large part of the B-helix was not predicted by this method in both leghemoglobins and part of the F-helix was not predicted in lupin leghemoglobin II. The helical regions predicted for *P. andersonii* hemoglobin I are in good agreement with the known and predicted helical structures of the leghemoglobins. The method predicted the B-helix region in *P. andersonii* hemoglobin I but not the region of the E-helix which contains the distal histidine residue (His^{64}). The failure to predict the E-helix can be attributed to an extra proline residue at position 63 in this region of the *P. andersonii* hemoglobin I molecule. These structural comparisons indicate that *P. andersonii* hemoglobin I has the characteristic globin structure and that there is nothing unusual about this nonlegume hemoglobin to distinguish from the other well known hemoglobins.

The location of two of the three introns of soybean leghemoglobin genes in exactly the same position as found in animal hemoglobin genes [6] together with the similarity in amino acid sequences and tertiary structures provides strong evidence that soybean leghemoglobins and animal hemoglobins share the same genetic origin. Although the *P. andersonii* hemoglobin I gene structure remains to be determined the amino acid sequence homology between *P. andersonii* hemoglobin I and the leghemoglobin provides strong evidence that these proteins have the same genetic origin. This finding suggests that all plant hemoglobins may eventually be shown to have the same evolutionary origin.

ACKNOWLEDGEMENTS

We thank the analytical group at the CSIRO Division of Protein Chemistry for amino-terminal sequence determination (A.S. Inglis), HPLC analysis of PTH amino acids (C. Roxburgh, N. Cook and P. Sommers) and amino acid analyses (N. Bartone). We also thank G. Lilley for the ultracentrifuge experiment.

REFERENCES

[1] Appleby, C.A. (1974) in: The Biology of Nitrogen Fixation (Quispel, A. ed.) pp 521–554, Elsevier, Amsterdam, New York.
[2] Coventry, D.R., Trinick, M.J. and Appleby, C.A. (1976) Biochim Biophys. Acta 420, 105–111
[3] Appleby, C.A., Tjepkema, J.D. and Trinick, M.J. (1983) Science 220, 951–953.
[4] Trinick, M.J. (1980) New Phytol. 85, 37–45.
[5] Trinick, M.J. and Galbraith, J. (1980) New Phytol. 86, 17–26.
[6] Hyldig-Nielsen, J.J., Jensen, E.Ø., Paludan, K., Wiborg, O., Garrett, R., Jørgensen, P. and Marcker, K. (1982) Nucleic Acids Res 10, 689–701.
[7] Appleby, C.A. (1984) Annu Rev Plant Physiol. 35, 443–478.
[8] Davenport, H.E. (1960) Nature 186, 653–654.
[9] Tjepkema, J.D. (1983) Can. J. Bot. 61, 2924–2929.
[10] Fleming, A.I (1984) in: AIAS Occasional Publ. no.12 (Proc 7th Aust. Legume Nod. Conf.) (Kennedy, I.R. and Copeland, L. eds) p.85, Aust. Inst. Agric. Sci., Sydney.
[11] Yonetani, T. (1967) J. Biol. Chem. 242, 5008–5013.
[12] Peterson, J D., Nehrlich, S., Dyer, P.E. and Steiner, D.F. (1972) J. Biol. Chem. 247, 4866–4871.

[13] Garnier, J., Osguthorpe, D J. and Robson, B (1978) J. Mol Biol. 120, 97–120
[14] Appleby, C.A , Chen, Y.P , Dudman, W.F , Higgins, T.J , Inglis, A.S., Kortt, A.A., Sutton, W D. and Trinick, M.J. (1984) in: Advances in Nitrogen Fixation Research, Proc 5th Int. Symp Nitrogen Fixation (Veeger, C and Newton, W E eds) p.493, Nijhoff, The Hague.
[15] Egorov, T.A., Kazakov, V.K , Shakhparonov, M.I. and Feigina, M Y. (1980) Bioorg. Chem. 6, 666–683.
[16] Ollis, D L., Appleby, C.A , Colman, P.M., Cutten, A.E., Guss, J.M., Venkatappa, M P. and Freeman, H.C. (1983) Aust J. Chem 36, 451–468.
[17] Gresshoff, P.M , Newton, S , Mohapatra, S S , Scott, K F , Howitt, S , Price, G.D , Bender, G.L , Shine, J and Rolfe, B.G (1984) in Advances in Nitrogen Fixation Research, Proc 5th Int Symp Nitrogen Fixation (Veeger, C and Newton, W E. eds) pp.483–489, Nijhoff, The Hague
[18] Love, W.E., Klock, P A., Lattman, E.E , Padlan, E.A , Ward, K B and Hendrickson, W.A (1971) Cold Spring Harbor Symp Quant. Biol 36, 349–357
[19] Harutyunyan, E G., Kuranova, I P , Vainshtein, B.K and Steigemann, W (1980) Kristallografiya 25, 80–103.
[20] Kendrew, J C , Dickerson, R.E., Strandberg, B.E , Hart, B E., Davies, D R , Phillips, D C. and Shore, V C. (1960) Nature 185, 422–427