Volume 231, number 2, 341–346 FEB 05794 April 1988

Amino acid sequence of hemoglobin I from root nodules of the non-leguminous *Casuarina glauca-Frankia* symbiosis

Alexander A. Kortt, Adam S. Inglis, Anthony I. Fleming* and Cyril A. Appleby +

*CSIRO, Division of Biotechnology, Parkville Laboratory, Melbourne, 3052 Victoria, *Department of Forestry, Australian National University, Canberra, ACT 2601* and *+CSIRO, Division of Plant Industry, Canberra, ACT 2601, Australia*

Received 4 March 1988

The amino acid sequence of hemoglobin I from nitrogen-fixing root nodules of the *Casuarina glauca-Frankia* symbiosis has been determined. The protein is composed of 151 amino acids including a single cysteine, a residue not found in the leghemoglobins. The molecular mass, including heme, is calculated to be 17856 Da. *C. glauca* hemoglobin I shows extensive sequence homology (43–52%) with other plant hemoglobins and this provides further evidence that hemoglobins from distant plant genera and animal hemoglobins share a common evolutionary origin.

Non-legume hemoglobin; Amino acid sequence; Plant hemoglobin sequence homology; Genetic origin; (*Casuarina glauca*)

1. INTRODUCTION

It is well known that hemoglobin occurs in nitrogen-fixing nodules of all legume-*Rhizobium* symbioses [1]. Structural studies [2,3], amino acid and gene sequence analyses [3–5] have confirmed that leghemoglobin is genetically related to the animal globins. On this basis it has been suggested that hemoglobin arose in ancient legumes following an act of horizontal gene transfer from the animal kingdom [4], a suggestion which assumed that the occurrence of hemoglobin in the plant kingdom was unique to legume nodules. This premise was proved false by the isolation of hemoglobin from the *Rhizobium*-induced nodules of the non-leguminous plant *Parasponia* (Ulmaceae) [6]. Both the amino acid sequence [7] and gene sequence [8] of this protein show considerable homology with leghemoglobin, indicating a common ancestral origin.

The finding of a common hemoglobin in two distantly related plant families nodulated by *Rhizobium* raised the question of just how widespread hemoglobin is in the plant kingdom [6]. Recent studies have confirmed Davenport's 1960 claim [9] that hemoglobin is indeed widely distributed in plants. It has been observed in nodules of phylogenetically-disparate species [8] which form nitrogen-fixing symbioses with the actinomycete *Frankia* [10,11].

Hemoglobin purified from the *Casuarina glauca-Frankia* symbiosis has physicochemical properties (molecular mass, optical spectra, kinetics of oxygen binding and oxygen affinity) similar to those of leghemoglobin and *Parasponia* hemoglobin, with which it is immunologically related [11]. These results suggest, but do not prove that *Casuarina* hemoglobin (and, by implication, other actinorhizal hemoglobins) shares a common origin with leghemoglobin and *Parasponia* hemoglobin. The amino acid sequence reported here establishes more firmly the relationship of *Casuarina* hemoglobin to other plant hemoglobins.

Correspondence address: A.A. Kortt, CSIRO, Division of Biotechnology, Parkville Laboratory, 343 Royal Parade, Parkville, 3052 Victoria, Australia

2. EXPERIMENTAL

C. glauca hemoglobin, hereafter referred to as *Casuarina*

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

hemoglobin, was isolated in low yield from fresh nodules of *C. glauca* Sieb ex Spreng inoculated with *Frankia* sp. and purified as in [11]. The *Casuarina* hemoglobin mixture obtained after sequential Sephacryl-200 and Sephadex G-75 chromatography [11] was further fractionated by preparative isoelectric focusing (IEF) as described [11] to yield hemoglobins I, II and III. Because of the low yields obtained for the separated hemoglobins, preliminary sequence information was obtained with the *Casuarina* hemoglobin mixture using ~0.8 mg of protein, and the complete amino acid sequence of the major component, hemoglobin I, was determined with ~0.75 mg (44 nmol) of protein.

The hemoglobin I sample (0.75 mg) was dissolved in 0.2 ml of 0.1 M Tris-HCl, 2% SDS, buffer, pH 8.5, reduced with dithiothreitol (0.01 M) at 50°C for 2 h under N_2 and alkylated with iodoacetic acid (0.04 M) for 1 h at 25°C in the dark. The reaction was stopped by the addition of an excess of dithiothreitol and dried under vacuum at 50°C. The reaction mixture was redissolved in 0.1 ml distilled water and the protein precipitated with 0.9 ml cold methanol, washed once with 1.0 ml of cold methanol and dried under N_2. The complete procedure was carried out in a 1.5 ml Eppendorf tube.

The *S*-carboxymethylated hemoglobin I was dissolved in 0.1 M ammonium bicarbonate buffer, pH 8.0, and aliquots (~12 nmol) were digested with trypsin (Worthington), α-chymotrypsin (Worthington) or *Staphylococcus aureus* V8 protease (Pierce) at 37°C for 4 h at an enzyme/substrate ratio of 1:50 (w/w). Aliquots (~15 nmol) of native *Casuarina* hemoglobin mixture were also digested with these three enzymes under the same conditions and the tryptic digests of native hemoglobin II and III were carboxymethylated immediately following digestion as described [12]. The digests were dried under vacuum at 50°C, redissolved in 0.25 ml of 0.1% (v/v) trifluoroacetic acid and the soluble peptides isolated by HPLC in 0.1% (v/v) trifluoroacetic acid on a Vydac 218TP54 column using an acetonitrile gradient.

Peptides (using 2–4 nmol) were sequenced either on a gas-phase sequencer or manually by a modified Edman procedure [13]. Residues were identified as the PTH-amino acid derivatives by HPLC as in [14]. Amino acid analysis was performed on a Waters HPLC amino acid analyser after hydrolysis of 1–2 nmol protein in 6 M HCl for 24 h at 110°C. Tryptophan was determined after hydrolysis of the protein in 4 M methane sulfonic acid/0.2% (w/v) tryptamine for 24 h at 115°C. The carboxyl-terminal residues were determined by amino acid analysis after digestion of the protein with carboxypeptidase Y (1%, w/w) in 0.1 M pyridine-acetate buffer, pH 5.6, at 37°C for 1 and 2 h.

3. RESULTS

The amino-terminal sequence of intact *Casuarina* hemoglobin I (residues 1–38) was determined using the gas-phase sequencer. The complete amino acid sequence of hemoglobin I was established by sequencing the peptides derived from tryptic, chymotryptic and staphylococcal protease digests of the protein as illustrated in fig.1. The tryptic peptides, isolated by RP-HPLC accounted for 147 residues of the molecule. The alignment of these tryptic peptides and the sequence of the two dipeptides not isolated, T11 and T14, was established with overlapping chymotryptic and staphylococcal protease peptides, and by homology, as shown in fig.1.

Tryptic peptides T1–T5 confirmed the gas-phase sequencer data and peptides T5, T6 and T7 were overlapped by S2 and C9 to establish the sequence to residue 58. Peptides T7, T8, T9 and T10 were located in the sequence by homology with other plant globins (fig.2) and peptide S3 provided the overlap to extend the sequence to residue 78. Peptides T10, T11 and T12 were overlapped by S5 and S6 overlapped T12 and T13 extending the sequence to residue 99. Peptides C15 and S6 connected T13, T14 and T15 and S7 connected T15 and T16 to extend the sequence to residues 119. Peptides T16 and T17 were overlapped by S7 and C17 and T17 and T18 were overlapped by S8. Finally peptides T18 and T19 were overlapped by C19, C21 and S10 to complete the sequence. Although S7 and S8 provided only a single residue overlap between T16, T17 and T18, the placements of these peptides are consistent with homology to other plant globins.

The carboxyl-terminal amino acid was identified as glutamic acid from the sequence data of peptide C21. Some difficulty was encountered in deducing the C-terminal residue by carboxypeptidase Y digestion. No major residue was released upon digestion of intact hemoglobin I with carboxypeptidase Y. A comparison of 1 h and 2 h digests, however, showed an increase in the yields of Glu, Lys, Met and Ala released consistent with the C-terminal sequence deduced from the peptide data.

The sequence determination showed that *Casuarina* hemoglobin I contains 151 amino acid residues corresponding to an M_r of 17856 (including the heme group) which is in agreement with the M_r of 17500 determined by SDS-PAGE [11]. The sequences of peptides T10 and C10 confirmed the presence of the single cysteine residue (identified as the PTH of carboxymethyl-*S*-cysteine) at position 71 which is located in the same region of the molecule, in a conserved sequence (fig.2), as that found in *Parasponia andersonii* hemoglobin I [8]. *Casuarina* hemoglobin I also contains 3 methionine residues in the C-terminal one-third of the molecule and Met-130 and Met-149 are homologous to corresponding

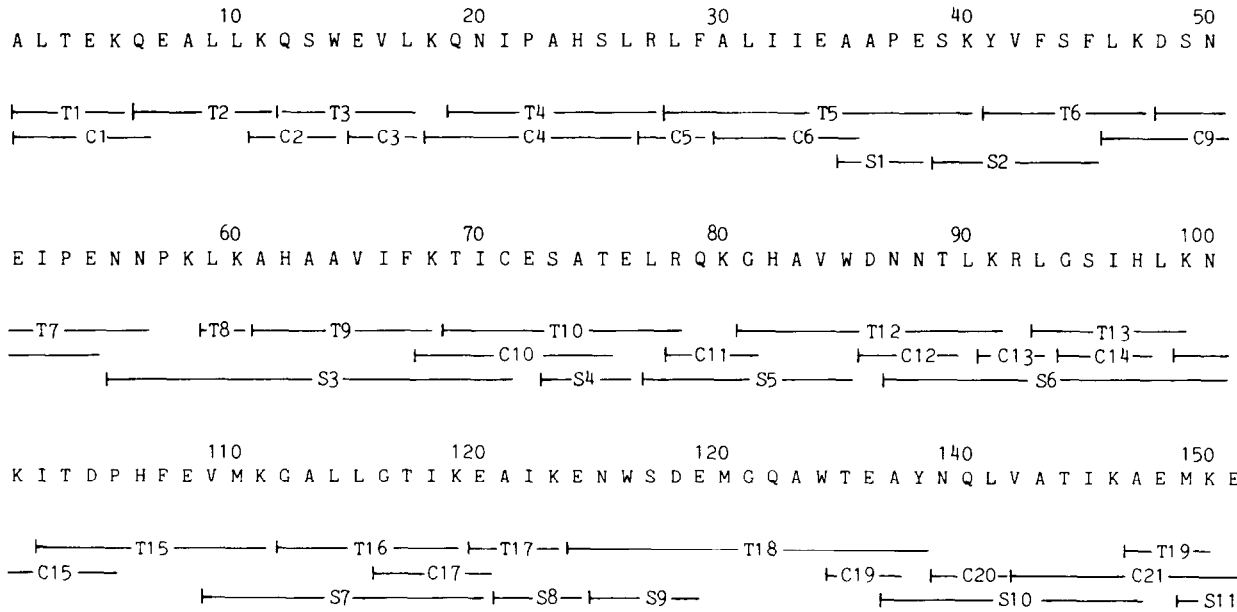

Fig.1. Primary structure of *Casuarina* hemoglobin I. The vertical bar indicates the amino-terminus of the peptide and the residues sequenced are indicated by the solid line. Peptides are numbered from the N-terminus of the protein where T, C and S signify tryptic, chymotryptic and staphylococcal protease peptides, respectively.

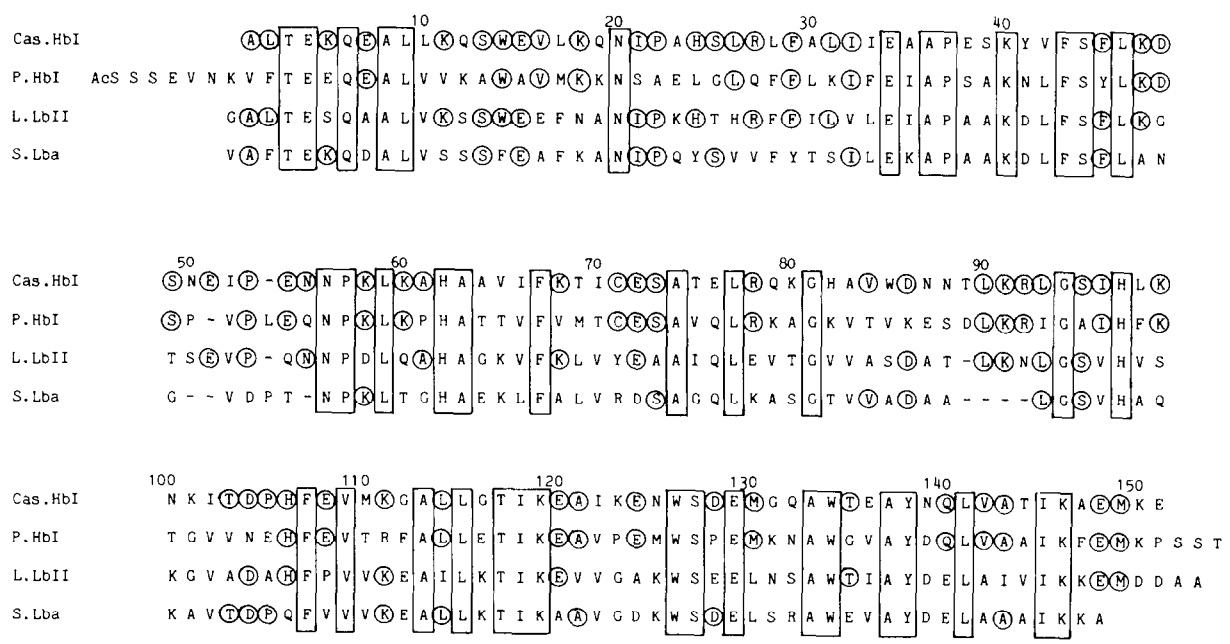

Fig.2. Sequence alignment indicating the homology between *Casuarina* hemoglobin I (Cas HbI), *P. andersonii* hemoglobin I (P.HbI) [7], lupin leghemoglobin II (L.LbII) [16] and soybean leghemoglobin a (S.Lba) [4]. Residues identical in the four plant hemoglobins are boxed. Ringed residues are identical in *Casuarina* hemoglobin I and one of the other hemoglobins. (——) Gaps introduced to optimize alignment.

Table 1

Amino acid composition of *Casuarina* hemoglobin I

Amino acid	Residues/mol	
	Analysis[a]	Sequence
Lys	16.3	17
His	4.2	5
Arg	3.6	3
Asx	13.7	13
Thr	7.6	8
Ser	7.6	8
Glx	21.4	22
Pro	5.2	5
Gly	6.1	5
Ala	16.5	17
1/2Cys	0.8	1
Val	5.9	6
Met	1.9	3
Ile	9.4	11
Leu	16.9	16
Tyr	1.8	2
Phe	5.4	5
Trp	3.3[b]	4

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

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

methionines in *P. andersonii* hemoglobin I [7]. The sequence composition is in agreement with the amino acid composition of the protein (table 1).

Although the *Casuarina* hemoglobin preparation obtained after Sephadex G-75 chromatography [11] is a mixture of hemoglobins, which is resolved into three major and many minor components on IEF [11], this mixture showed only very limited sequence heterogeneity. Comparisons of HPLC profiles of tryptic digests of *S*-carboxymethylated hemoglobins I, II and III, and the native hemoglobin mixture showed that these preparations produced very similar peptide patterns with only a few differences. Sequencing tryptic and chymotryptic peptides isolated from the hemoglobin mixture revealed heterogeneity at only seven positions within the sequence. Heterogeneity was found at positions 12 (Gln/Glu), 16 (Val/Ile), 23 (Ala/Gly), 33 (Ile/Leu), 101 (Lys/Gly), 131 (Gly/Glu) and 145 (Ile/Gly). This limited substitution of charged residues may account for some of the apparent heterogeneity of the hemoglobin mixture but it cannot, however, account for the large number of minor hemoglobin components observed on IEF [11]. It is possible that this heterogeneity

Fig.3. Comparison of the predicted α-helical segments of *Casuarina* hemoglobin I (Cas HbI) and the predicted and known α-helical segments of lupin leghemoglobin II (L.LbII). Boxes enclose the segments predicted as helices [15] and (——) designates the segments identified as helical from X-ray crystallographic studies [3].

may be due in part to non-enzymic glucosylation or other post-translational conjugation [11].

The secondary structure of *Casuarina* hemoglobin I predicted from the amino acid sequence [15] has a high content of α-helix (75%) and the predicted helical regions are compared with the known [3,16] and predicted [15] secondary structure of lupin hemoglobin II in fig.3. The predicted α-helical regions for *Casuarina* hemoglobin I correlate well with the known helical regions of lupin [16] and soybean leghemoglobins [2] determined by X-ray crystallographic studies. The method [15], however, failed to predict part of the B- and F-helices as noted previously [7] for other plant leghemoglobins.

4. DISCUSSION

The amino acid sequence of *Casuarina* hemoglobin I shows extensive homology with the sequence of hemoglobin I from *P. andersonii* [7], another non-legume, and the leghemoglobins from lupin [16] and soybean [4] (fig.2). Maximum homology was obtained with an alignment containing only a single residue deletion in the *Casuarina* and *P. andersonii* hemoglobins, 2 deletions in lupin leghemoglobin II and 7 deletions in the shorter soybean leghemoglobin a sequence (fig.2). In this alignment *Casuarina* hemoglobin I has 7 residues fewer at the amino-terminus than *P. andersonii* hemoglobin I has and one residue less than the two leghemoglobins have. These four hemoglobins show 27% homology with 41 identical positions, while only 44 positions in the *Casuarina* hemoglobin I sequence show no identity with a residue in one of the other 3 plant hemoglobin sequences (fig.2). When compared separately the homology between *Casuarina* hemoglobin I (with 151 residues) and *P. andersonii* hemoglobin I (161 residues), lupin leghemoglobin II (153 residues) and soybean leghemoglobin a (143 residues) is 52, 49 and 43%, respectively. This high level of homology between the hemoglobins of *Casuarina* and *Parasponia*, two widely separated genera, and the legumes is striking and is similar to the level of homology found between the hemoglobins within the legumes.

While IEF analysis of *Casuarina* hemoglobin suggested the presence of at least three major hemoglobin components with p*I* values in the range 5.3–5.9 [11], they appear to show only limited variation in their primary structures. Sequence analysis of the hemoglobin mixture indicated, at most, only 7 possible substitutions within 151 residues. In the case of the leghemoglobins from soybean and lupin, 12 changes within 143 residues are found between soybean Lba and Lbc1 [4] while lupin (*L. luteus*) LbI and LbII have 20 changes within 153 residues [16]. In contrast *Parasponia rigida* hemoglobins I (p*I* 6.15) and II (p*I* 5.64), non-legume hemoglobins, have only one amino acid substitution in 161 residues.

It has been noted that all known leghemoglobins contain 49 invariant residues [2]. Thirty-five of these residues occur in *Casuarina* hemoglobin with a further 6 showing a conservative amino acid substitution. A comparison of *Casuarina* hemoglobin, *Parasponia* hemoglobin and all known leghemoglobins reveals 31 invariant residues. This surprisingly high number, considering the phylogenetic distance which separates these plants, suggests that a strong conservative selection pressure has been active. This selection pressure presumably relates to hemoglobin function, which in legume nodules is the facilitation of oxygen diffusion to the endophyte [17]. Oxygen kinetic studies with *Casuarina* hemoglobin suggest that it may have a similar function [11]. The concentration of hemoglobin in *Casuarina* nodules (32 nmol/g fresh wt [11], 80 nmol/g fresh wt [18]) is consistent with such a function. While these volume averaged concentrations are lower than commonly found in legume nodules [18], the proportion of symbiotic tissue is also lower [10]; we assume that hemoglobin is present in the symbiotic tissue of *Casuarina* nodules in a similar concentration to that found in symbiotic tissue of legumes, where it has been proven to have a role in facilitated oxygen diffusion [1].

The similarity in primary structure and predicted tertiary structure (figs 2,3) between *Casuarina* hemoglobin and other plant hemoglobins (see [7]) further supports the proposal that all plant hemoglobins share a common evolutionary origin. Assuming the genes for hemoglobin have been inherited in these extant genera by a process of vertical descent, rather than horizontal transfer, the origin of hemoglobin in plants must be extremely primitive. Given that plant and animal

hemoglobins share the same origin [4,5], it seems probable that the hemoglobin gene has been present in both the plant and animal kingdoms since their divergence from a common ancestor. The presence of a central intron in the leghemoglobin and *Parasponia* hemoglobin genes [4,8], which has been taken to imply that these genes are more primitive than animal hemoglobin genes [4], supports this argument. It thus seems possible that the hemoglobin gene is widespread in the plant kingdom, a view reinforced by the recent finding of hemoglobin in root tissue of the non-symbiotic genera *Trema* and *Celtis* [19].

Acknowledgements: We thank J.E. Burns for technical assistance and F. Casagranda for gas-phase sequence analyses, P. Somers for HPLC analyses of PTH-amino acids and N. Bartone for amino acid analyses. A.I.F. was in receipt of a Commonwealth Postgraduate Research Award for much of this work.

REFERENCES

[1] Appleby, C.A. (1974) in: The Biology of Nitrogen Fixation (Quispel, A. ed.) pp.521–554, Elsevier, New York.
[2] 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.
[3] Arutyunyan, E.G. (1981) Mol. Biol. (USSR) 15, 27–44.
[4] Hyldig-Nielsen, J.J., Jensen, E.Ø., Paludan, K., Wilborg, O., Garret, R., Jørgensen, P. and Marcker, K. (1982) Nucleic Acids Res. 10, 689–701.
[5] Antoine, M. and Niessing, J. (1984) Nature 310, 795–798.
[6] Appleby, C.A., Tjepkema, J.D. and Trinick, M.J. (1983) Science 220, 951–953.
[7] Kortt, A.A., Burns, J.E., Trinick, M.J. and Appleby, C.A. (1985) FEBS Lett. 180, 55–60.
[8] Landsmann, J., Dennis, E.S., Higgins, T.J., Appleby, C.A., Kortt, A.A. and Peacock, W.J. (1986) Nature 324, 166–168.
[9] Davenport, H.E. (1960) Nature 186, 653–654.
[10] Tjepkema, J.D. (1983) Can. J. Bot. 61, 2924–2929.
[11] Fleming, A.I., Wittenberg, J.B., Wittenberg, B.A., Dudman, W.F. and Appleby, C.A. (1987) Biochim. Biophys. Acta 911, 209–220.
[12] Merrill, B.M., Lopresti, M.B., Stone, K.L. and Williams, K.R. (1987) Int. J. Peptide Protein Res. 29, 21–39.
[13] Peterson, J.D., Nehrlich, S., Dyer, P.E. and Steiner, D.F. (1972) J. Biol. Chem. 247, 4866–4871.
[14] Woods, E.F. and Inglis, A.S. (1984) Int. J. Biol. Macromol. 6, 277–283.
[15] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[16] Egorov, T.A., Kazakor, V.K., Shakhparaonov, M.I. and Feigina, M.Y. (1980) Bioorg. Chem. 6, 666–683.
[17] Appleby, C.A. (1984) Annu. Rev. Plant Physiol. 35, 443–478.
[18] Tjepkema, J.D. and Asa, D.J. (1987) Plant Soil 100, 225–236.
[19] Bogusz, D., Appleby, C.A., Landsmann, J., Dennis, E.S., Trinick, M.J. and Peacock, W.J. (1988) Nature 331, 178–180.