Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 5682–5687, June 1996
Plant Biology

A new hemoglobin gene from soybean: A role for hemoglobin in all plants

(nonsymbiotic/leghemoglobin/evolution)

CAROL R. ANDERSSON*†‡, ERIK OSTERGAARD JENSEN§, DANNY J. LLEWELLYN*, ELIZABETH S. DENNIS*, AND W. JAMES PEACOCK*¶

*Commonwealth Scientific and Industrial Research Organization Plant Industry, G.P.O. Box 1600, Canberra, ACT 2601, Australia; †Division of Biochemistry and Molecular Biology, Australian National University, G.P.O. Box 4, Canberra ACT 2601, Australia; and §Laboratory of Gene Expression, University of Aarhus, Gustav Weids Vej 10, Aarhus C 8000 Denmark

Contributed by W. James Peacock, January 24, 1996

ABSTRACT **We have isolated a new hemoglobin gene from soybean. It is expressed in cotyledons, stems of seedlings, roots, young leaves, and in some cells in the nodules that are associated with the nitrogen-fixing *Bradyrhizobium* symbiont. This contrasts with the expression of the leghemoglobins, which are active only in the infected cells of the nodules. The deduced protein sequence of the new gene shows only 58% similarity to one of the soybean leghemoglobins, but 85–87% similarity to hemoglobins from the nonlegumes *Parasponia*, *Casuarina*, and barley. The pattern of expression and the gene sequence indicate that this new gene is a nonsymbiotic legume hemoglobin. The finding of this gene in legumes and similar genes in other species strengthens our previous suggestion that genomes of all plants contain hemoglobin genes. The specialized leghemoglobin gene family may have arisen from a preexisting nonsymbiotic hemoglobin by gene duplication.**

Despite the common perception of hemoglobin as a blood protein, the protein is also found in many invertebrates, bacteria, fungi, and in some higher plants. Plant hemoglobin was thought to be restricted to the nitrogen-fixing root nodules of legumes in the highly specialized symbiosis with strains of *Rhizobium* or *Bradyrhizobium* bacteria. The function of these legume hemoglobins (leghemoglobins) is the facilitated diffusion of oxygen to the respiring bacteroids within the root nodule (for review, see ref. 1).

More recently, hemoglobin and hemoglobin genes have been discovered in the nonlegume genera *Parasponia*, *Trema*, *Casuarina*, *Hordeum*, *Triticum*, and *Zea* (2–8). The finding of hemoglobin genes in non-nodulating plants (*Trema*, *Hordeum*, *Triticum*, and *Zea*) and in phylogenetically diverse plant genera, including monocots, reinforces the proposal that hemoglobin may be present in all plants (9). Two roles for plant hemoglobin outside the nitrogen-fixing symbiosis have been proposed: (*i*) as a facilitator of oxygen diffusion in rapidly respiring cells or (*ii*) as an oxygen sensor involved in switching plant metabolism to anaerobic pathways (9).

In general terms, there appear to be two different types of plant hemoglobins, the symbiotic (nodule specific and associated with a symbiotic interaction involving microorganisms) and the nonsymbiotic hemoglobins (expressed in non-nodule tissues such as roots and stems). The situation in the nodulating tree *Parasponia andersonii* is unusual in that a single gene encodes a protein that has both a symbiotic and nonsymbiotic role (3).

The symbiotic hemoglobin genes have been postulated to have arisen by duplication of preexisting, nonsymbiotic, hemoglobin genes (9). *Casuarina glauca* presents a possible example of such a gene duplication and divergence process in that it has both a single nonsymbiotic hemoglobin gene expressed in all tissues analyzed, and a different, symbiotic hemoglobin gene family expressed exclusively in the nodules induced by *Frankia*, an actinomycete (6, 7). Gene duplication with subsequent specialization is a common evolutionary strategy and recent evidence suggests that in legumes other nodule specific proteins have arisen by this mechanism, e.g., nodulin 26 (10).

If symbiotic hemoglobin genes were derived by duplication from nonsymbiotic genes, then where are the nonsymbiotic hemoglobins in legume species? Legumes have well-characterized, nodule-specific leghemoglobin gene families, but apart from a single report of a hemoglobin-like protein in the seeds of winged bean (11), expression of hemoglobin genes outside nodules has never been detected by either nucleic acid hybridization or reporter gene assays.

In this paper, we report the isolation and characterization of a nonsymbiotic hemoglobin gene from soybean and confirm that similar genes are present in other legumes. The gene is clearly related to nonsymbiotic plant hemoglobins on the basis of its predicted amino acid sequence and gene structure, including the promoter sequence; its expression pattern within the plant is unlike that of the symbiotic leghemoglobins.

MATERIALS AND METHODS

Isolation of Genomic DNA and Southern Blot Analysis. Isolation of genomic DNA and Southern blot analysis of hemoglobin genes from *Glycine max* cv Lincoln, *Pisum sativum* cv Greenfeast, *Pisum humile*, *Medicago sativa* cv R15 and *Trifolium repens* cv Haifa was as described (3). Probes were either the PCR fragment soyhb1f+4r2 or the leghemoglobin clones 13.0/1.6 and 7.5/2.0 (12), which together cover the entire coding region of soybean *lbps1*.

PCR Reactions. Degenerate primers were designed to conserved regions of plant nonsymbiotic hemoglobins (see Fig. 1) as follows: Hbexon1f (hemoglobin exon 1 forward primer), 5′-CGGAATTCGA(A/G)GA(A/G)(C/G)A(A/G)GA(A/G)GCI(T/C)TIGT; Hbexon2f (exon2 forward primer), 5′-CGGAATTCATITT(T/C)GA(A/G)ATIGCICC; Hbexon2r (exon 2 reverse primer), 5′-CGGGATCCGC(A/G)TGIII(T/C)TTIA(A/G)(T/C)TTIGG(A/G)TT; Hbexon4r2 (exon 4 reverse primer), 5′- CGGGATCCGC(T/C)TC(T/C)TTIATIGT(T/C)TC.

Template DNA was digested with *Hin*dIII or *Xho*I, followed by heat inactivation at 65°C for 20 min. The PCR reaction included *Taq* DNA polymerase buffer (Promega)/1 mM

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Data deposition: The sequence reported in this paper has been deposited in the GenBank data base (accession no. U47143).

‡Present address: Department of Biology, Texas A&M University, College Station, TX 77843.

¶To whom reprint requests should be addressed. e-mail: jpeacock@pi.csiro.au.

$MgCl_2$/200 μM each dNTP/1 μM each primer/250 ng digested genomic DNA/2.5 units *Taq* DNA polymerase (Promega) in a total reaction volume of 15 μl. Reactions were performed using a Corbett thermal cycler. The reaction incorporated stepwise reductions in annealing temperature ("touchdown PCR") from 50°C to 37°C, and slow ramp times from the annealing to the extension temperature. The denaturation times were 30 sec, with annealing and extension times of 1 min. PCR products were digested with *Bam*HI and *Eco*RI, purified from a polyacrylamide gel, and cloned into pBluescript SK (−) (Stratagene) or pGem 3Zf(−) (Promega). Plasmid inserts were sequenced using a Pharmacia T7 sequencing kit.

The end-labeled hemoglobin specific primer soyhb2f35 (soybean hemoglobin exon 2 forward 35 mer) 5′- CTCATTCTTGAGAGATTCAACGGTTCCTTTGGAGC was used to identify longer products in subsequent PCR reactions.

Isolation of Soybean Genomic Clones. A soybean genomic library in lambda EMBL4 was constructed from size fractionated, partial *Sau*3A digested genomic DNA and screened with a soyhb1f+4r2 probe using standard methods (13).

Northern Blot Analysis. Soybean tissue from shoots, leaves, stems, flowers (all stages), seed pods, roots, nodules, and germinating seeds were harvested, snap frozen in liquid nitrogen, and stored at −80°C. To eliminate the possibility that root tissue contained nodules, roots were harvested from seedlings grown from surface sterilized seeds sown in sterilized vermiculite/sand. Total RNA was isolated and analyzed by Northern blotting as described (3). Filters were probed with an antisense RNA probe (Amersham *in vitro* transcription kit) synthesized from soyhb1f+4r2, subsequently stripped and reprobed with an *A. thaliana* ubiquitin (14) antisense RNA probe. The hybridization signals were quantified using a PhosphorImager (Fujix) and the abundance of the soybean hemoglobin message expressed relative to ubiquitin. The size of transcripts were estimated relative to RNA standards (BRL).

```
             1                    hbexon1f>                    *hbexon2f>    60
    Tremahb  .......SSS EVDKVFTEEQ EALVVKSWAV MKKNSAELGL KFFLKIFEIA PSAKNLFSYL
     Parahb  .......SSS EVNKVFTEEQ EALVVKAWAV MKKNSAELGL QFFLKIFEIA PSAKNLFSYL
 Casnonsymh  ........ST LEGRGFTEEQ EALVVKSWSA MKPNAGELGL KFFLKIFEIA PSAQKLFSFL
      Soyhb  ........TT TLERGFSEEQ EALVVKSWNV MKKNSGELGL KFFLKIFEIA PSAQKLFSFL
    Ricehb1  ALVEDNNAV. ..AVSFSEEQ EALVLKSWAI LKKDSANIAL RFFLKIFEVA PSASQMFSFL
    Ricehb2  ALVEGNNGVS GGAVSFSEEQ EALVLKSWAI MKKDSANIGL RFFLKIFEVA PSGCQMFSFL
   Barleyhb  .......SAA EGAVVFSEEK EALVLKSWAI MKKDSANLGL RFFLKIFEIA PSARQMFPFL
    Soylbc2  .......... ...GAFTEKQ EALVSSSFEA FKANIPQYSV VFYTSILEKA PAAKDLFSFL
    Soylbc3  .......... ...GAFTDKQ EALVSSSFEA FKTNIPQYSV VFYTSILEKA PVAKDLFSFL
     Soylba  .......... ...VAFTEKQ DALVSSSFEA FKANIPQYSV VFYTSILEKA PAAKDLFSFL
    Soylbc1  .......... ...GAFTEKQ EALVSSSFEA FKANIPQYSV VFYNSILEKA PAAKDLFSFL
 Kidneybean  .......... ...GAFTEKQ EALVNSSWEA FKGNIPQYSV VFYTSILEKA PAAKNLFSFL
 Sesbaniaii  .......... ....GFTDKQ EALVNASYEA FKKNLPGHSV LFYSFILEKE PAAKGLFSFL
  Medicago1  .......... ....SFTDKQ EALVNSSYEA FKQNLSGYSV FFYTVILEKA PAAKGLFSFL
  Broadbean  .......... ....GFTEQQ EALVNSSSQL FKQNPSNYSV LFYTIILQKA PTAKAMFSFL
      Pealb  .......... ....GFTDKQ EALVNSSSE. FKQNLQGYAT LFYTIILEKA PAAKGLFSFL
   Lupin11b  .......... ...GVLTDVQ VALVKSSFEE FNANIPKNTH RFFTLVLEIA PGAKDLFSFL
   Cassymhb  .......... ....ALTEKQ EALLKQSWEV LKQNIPAHSL RLFALILEAA PESKYVFSFL

             61     <hbexon2r       *                                       120
    Tremahb  KDSPRPLEQN PKLKPHAMTV FVMTCESAVQ LRKAGKVTVR ESNLKRLGAI HFKNGVVNEH
     Parahb  KDSPVPLEQN PKLKPHATTV FVMTCESAVQ LRKAGKVTVK ESDLKRIGAI HFKTGVVNEH
 Casnonsymh  KDSNVPLERN PKLKSHAMSV FLMTCESAVQ LRKAGKVTVR ESSLKKLGAS HFKHGVADEH
      Soyhb  RDSTVPLEQN PKLKPHAVSV FVMTCDSAVQ LRKAGKVTVR ESNLKKLGAT HFRTGVANEH
    Ricehb1  RNSDVPLEKN PKLKTHAMSV FVMTCEAAAQ CGKAGKVTVR DTTLKRLGAT XLKYGVGXPX
    Ricehb2  RNSDVPLEKN PKLKTHGMSV FVMTCEAAAX CGKAGKVTVR DTTLKRLGAT ..........
   Barleyhb  RDSDVPLETN PKLKTHAVSV FVMTCEAAAQ LRKAGKITVR ETTLKRLGGT HLKYGVADGH
    Soylbc2  ..SNGVDPSN PKLTGHAEKL FGLVRDSAGQ LKANGTVV.A .DA..ALGSI HAQKAITDPQ
    Soylbc3  ..ANGVDPTN PKLTGHAEKL FGLVRDSAGQ LKASGTVV.I .DA..ALGSI HAQKAITDPQ
     Soylba  ..ANGVDPTN PKLTGHAEKL FALVRDSAGQ LKASGTVV.A .DA..ALGSV HAQKAVTDPQ
    Soylbc1  ..ANGVDPTN PKLTGHAEKL FALVRDSAGQ LKTNGTVV.A .DA..ALVSI HAQKAVTDPQ
 Kidneybean  ..ANGVDPTN PKLTAHAESL FGLVRDSAAQ LRANGAVV.A .DA..ALGSI HSQKGVNDSQ
 Sesbaniaii  KDSDGVPQNN PSLQAHAEKV FGLVHDAAGQ LRATGVVVLA .DA..SLGSV HVQKGVTDPH
  Medicago1  KDSAGV.QDS PQLQAHAEKV FGLVRDSASQ LRATGGVVLG .DA..ALGAI HIQKGVVDPH
  Broadbean  KDSAGVV.DS PKLGAHAEKV FGMVRDSAVQ LRATGEVVL. .DG..KDGSI HIQKGVLDPH
      Pealb  KDTAGV.EDS PKLQAHAEQV FGLVRDSAAQ LRTKGEVVLG .NA..TLGAI HAGKGVTNPH
   Lupin11b  KGSSEVPQNN PDLQAHAGKV FKLTYEAAIQ LQVNGAVASD .ATLKSLGSV HVSKGVVDAH
   Cassymhb  KDSNEIPENN PKLKAHAAVI FKTICESATE LRQKGHAVWD NNTLKRLGSI HLKNKITDPH

             121    <hbexon4r2                           169
    Tremahb  FEVTRFALLE TIKEAVP.EM WSPEMKNAWG EAYDQLVAAI KSEMKPSST
     Parahb  FEVTRFALLE TIKEAVP.EM WSPEMKNAWG VAYDQLVAAI KFEMKPSST
 Casnonsymh  FEVTKFALLE TIKEAVP.ET WSPEMKNAWG EAYDKLVAAI KLEMKPSS.
      Soyhb  FEVTKFALLE TIKEAVP.EM WSPAMKNAWG EAYDQLVDAI KSEMKPPSS
    Ricehb1  FEVVKFALLD TIKEEVPADM WSPAMKSAWS EAYDHLVAAI KQEMKPAE.
    Ricehb2  .......... .......... .......... .......... .........
   Barleyhb  FEVTRFALLE TIKEALPADM WGPEMRNAWG EAYDQLVAAI KQEMKPAE.
    Soylbc2  FVVVKEALLK TIKEAVG.DK WSDELSSAWE VAYDELAAAI KKAF.....
    Soylbc3  FVVVKEALLK TIKEAVG.DK WSDELSSAWE VAYDELAAAI KKAF.....
     Soylba  FVVVKEALLK TIKAAVG.DK WSDELSRAWE VAYDELAAAI KKA......
    Soylbc1  FVVVKEALLK TIKEAVG.GN WSDELSSAWE VAYDELAAAI KKA......
 Kidneybean  FLVVKEALLK TLKEAVG.DK WTDELSTALE LAYDELAAAI KKAYA....
 Sesbaniaii  FVVVKEALLK TLKEAAG.AT WSDEVSEAWE VAYDGLAAAI KKAMS....
  Medicago1  FVVVKEALLK TIKEAAG.DK WSEELSTAWE VAYDALATEI KKAMS....
  Broadbean  FVVVKEALLK TIKEASG.DK WSEELSAAWE VAYDGLATAI KAA......
      Pealb  FVVVKEALLQ TIKKASG.NN WSEELNTAWE VAYDGLATAI KKAMKTA..
   Lupin11b  FPVVKEAILK TIKEVVG.DK WSEELNTAWT IAYDELAIII KKEMKDAA.
   Cassymhb  FEVMKGALLG TIKEAIK.EN WSDEMGCAWT EAYNQLVATI KAEMKE...
               *
```

FIG. 1. Alignment of the predicted amino acid sequences of various plant hemoglobins using the GCG PILEUP program. The highly conserved residues involved in heme and ligand binding [the distal and proximal histidines, phenylanine CD1 and proline C2 (1)] are in bold. Intron positions are marked with asterisks. Conserved regions used to design degenerate PCR primers are overlined. Ricehb1 and ricehb2, expressed sequence tag sequences from rice, GenBank accession numbers D15507and D23324 or D25122 and D22678, respectively (X is indicated at positions of ambiguity in the entered DNA sequence); barleyhb (8); *Trema tomentosa* hemoglobin (tremahb) (4); *Parasponia andersonii* hemoglobin (parahb) (3), *Casuarina glauca* nonsymbiotic hemoglobin (casnonsymhb) (6); soybean nonsymbiotic hemoglobin (soyhb) (this study), lupin hemoglobin 1 (lupin1lb) (15); soybean lba and lbc1 leghemoglobins (soylba and soylbc1) (16); soybean lbc2 and lbc3 leghemoglobins (soylbc2 and soylbc3) (17); kidney bean leghemoglobin (kidneybean) (18); *Sesbania rostrata* leghemoglobin II (sesbaniaii) (19); *Medicago truncatula* leghemoglobin1 (medicago1) (20); broadbean leghemoglobin (broadbean) (21); pea leghemoglobin (pealb) (22), *Casuarina glauca* symbiotic hemoglobin (cassymhb) (7).

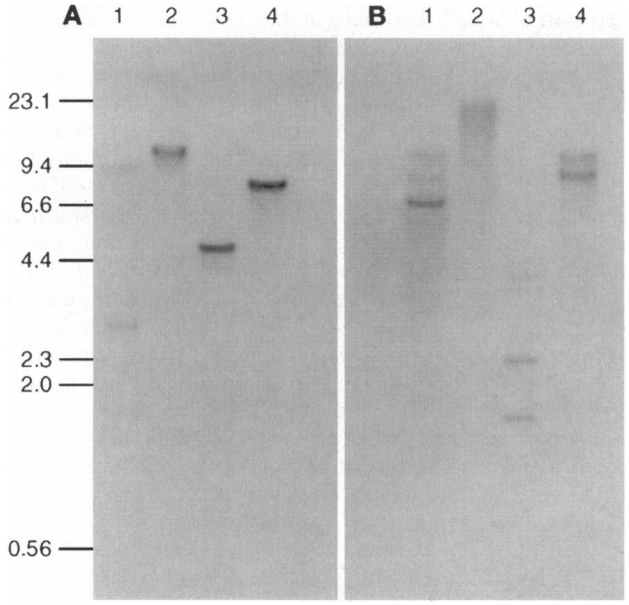

FIG. 2. Southern blot analysis of the soybean nonsymbiotic hemoglobin gene and leghemoglobin genes. *Glycine max* cv Lincoln genomic DNA was digested with *Eco*RI (lane 1), *Bam*HI (lane 2), *Hin*dIII, (lane 3) and *Bgl*II (lane 4). (*A*) Soybean nonsymbiotic hemoglobin (soybean 1f+4r2 PCR fragment probe). (*B*) Soybean leghemoglobins (*lbps1* 13.0/1.6 and 7.5/2.0 probes). Lambda *Hin*dIII size standards are in kb.

RESULTS

A New Hemoglobin Gene in the Soybean Genome. Short degenerate oligonucleotides designed from the conserved regions of known nonsymbiotic plant hemoglobins (Fig. 1) were used in a PCR reaction on soybean genomic DNA. Primers within the second exon (hbexon2f and hbexon2r) amplified a number of fragments including one of ≈110 bp, the size predicted from other plant hemoglobins. The 110-bp fragment had an open reading frame that contained residues known to be critical in hemoglobin structure and function, and was partially homologous to the known nonsymbiotic plant hemoglobins. A longer gene fragment was generated using degenerate primers derived from sequences in the first and last exons of nonsymbiotic hemoglobins (hbexon1f and hbexon4r2). As the genomic DNA was likely to include three introns, the hemoglobin-specific Soyhb2f35 probe was used to identify the correct product from this second PCR reaction.

The longer clone (1f+4r2) contained the original 110-bp fragment and was used to isolate corresponding genomic segments. The gene sequence (GenBank accession no. U47143) had a five amino acid N-terminal extension similar to that in the nonsymbiotic and *P. andersonii* hemoglobins (Fig. 1), and three introns at positions found in all plant hemoglobins.

The New Hemoglobin Gene Does not Hybridize to the Leghemoglobins; Homologues Are Present in Other Legumes. Southern blot analysis, using the 1f+4r2 PCR fragment and the previously described leghemoglobin gene probes, showed no hybridizing bands in common between these two classes of genes in soybean (Fig. 2) as expected from the low similarity between their nucleotide sequences (Table 1). Using the 1f+4r2 probe, *Eco*RI-digested soybean DNA displayed two hybridizing bands of about equal intensity, whereas the *Bam*HI and *Bgl*II lanes each had a single, more intense band. The *Hin*dIII digest produced two bands of different intensities. As the soybean 1f+4r2 probe does not contain an *Eco*RI site, we conclude that the soybean genome contains two copies of the new hemoglobin gene and, because there is a single *Bam*HI fragment, that these might be physically linked. Alternatively, and perhaps more likely, because soybean is an ancient tetraploid, the two copies may be present on homeologous chromosomes, giving rise to separate restriction fragments of equivalent length. It is unlikely that the two bands in the *Eco*RI and *Hin*dIII digests represent restriction fragment length polymorphisms of a single gene because of the inbred nature of the soybean cultivar.

Southern blot analysis of other legumes using the soybean 1f+4r2 probe at moderate hybridization stringency detected one or more cross-hybridizing bands in white clover, lucerne, and pea (data not shown), indicating that an homologous nonsymbiotic hemoglobin gene(s) is also present in the genomes of these species.

The Promoter of the New Hemoglobin Gene Does Not Contain the Nodulin Consensus Sequences. The symbiotic hemoglobin promoters contain two motifs that have been shown to be critical for nodule specific expression (the nodulin boxes, CTCTT and AAGAT) separated by six or seven nucleotides (23, 24). The nonsymbiotic (or dual function hemoglobin, in the case of the *P. andersonii* gene) hemoglobin genes from *C. glauca*, *Trema tomentosa*, and *P. andersonii* do not have these conserved motifs at this spacing (see Fig. 5). The critical CTCTT motif is absent from the nonsymbiotic genes. The promoter of the new soybean hemoglobin gene does not contain this motif either, but has the sequence CTCCC, identical in sequence and position to a motif in *P. andersonii* (see Fig. 5). Another motif shown to be critical for expression of the *P. andersonii* promoter (unpublished data), GAAGAG, is present in a similar position as GAAGGG is in the soybean promoter. These data suggest that the new soybean hemoglobin has a promoter that resembles promoters of the nonsymbiotic hemoglobins of other plant genera, more than it resembles the promoters of the symbiotic leghemoglobins in its own genome.

The New Soybean Hemoglobin Gene Is Expressed in Many Tissues. Northern blot analysis of total RNA isolated from various soybean tissues detected a nonsymbiotic hemoglobin mRNA of 830 nt. The message was readily detected in stems of mature plants and cotyledons of seedlings (Fig. 3). The gene was also expressed in roots, nodules, and young leaves of mature plants, but was virtually absent from older leaves, floral tissue, and seed pods. The level of expression in nodules was comparable to that in non-nodulated roots and considerably lower than that of leghemoglobin genes in the infected cells of nodules. The new hemoglobin's transcript was present at a

Table 1. Overall sequence similarity (%) between various plant hemoglobins

	Lupin1lb	Soylba	Cassymhb	Casnonsymhb	Parahb	Barleyhb	Soyhb
Lupin1lb		69	70	56	61	61	57
Soylba	72		67	61	62	61	58
Cassymhb	62	60		64	64	64	64
Casnonsymhb	59	54	56		84	84	87
Parahb	58	55	57	76		83	87
Barleyhb	56	60	56	69	69		85
Soyhb	58	53	56	76	78	69	

Numbers above the diagonal represent similarity between predicted hemoglobin amino acid sequences; numbers below the diagonal represent similarity between nucleotide sequences in the coding region (see Fig. 1 legend).

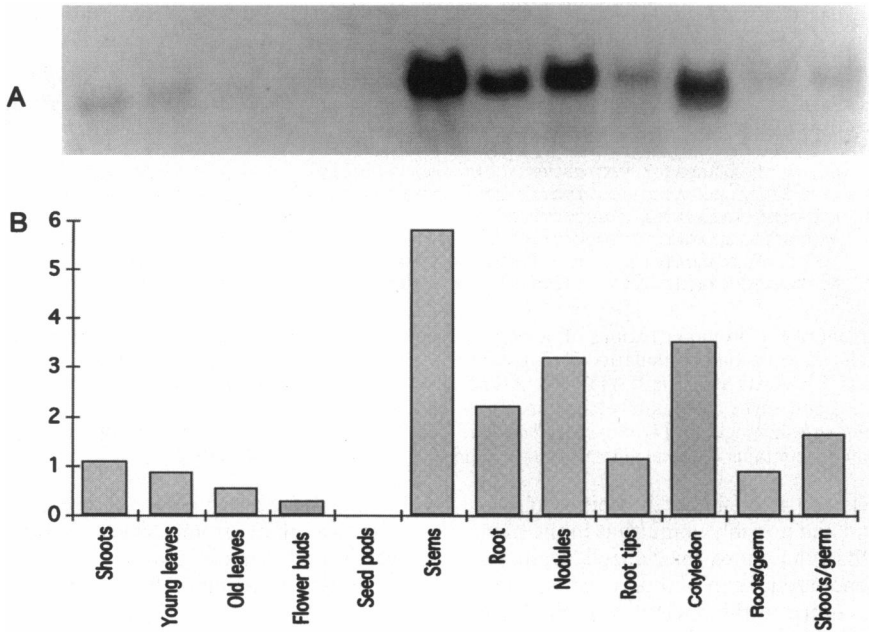

FIG. 3. Nonsymbiotic hemoglobin gene expression patterns in soybean seedlings and plants. Total RNA was isolated from soybean shoot tips, young leaves, old leaves, flower buds, whole seed pods, stems, roots, nodules, root tips, and cotyledons, roots, and shoots from germinating seeds. (*A*) Hybridization pattern with a riboprobe corresponding to the soybean PCR fragment 1f+4r2. (*B*) Rehybridization with an *Arabidopsis* ubiquitin probe provided levels of nonsymbiotic hemoglobin gene transcript relative to ubiquitin abundance.

lower concentration in root tips than in the remainder of the root (Fig. 3).

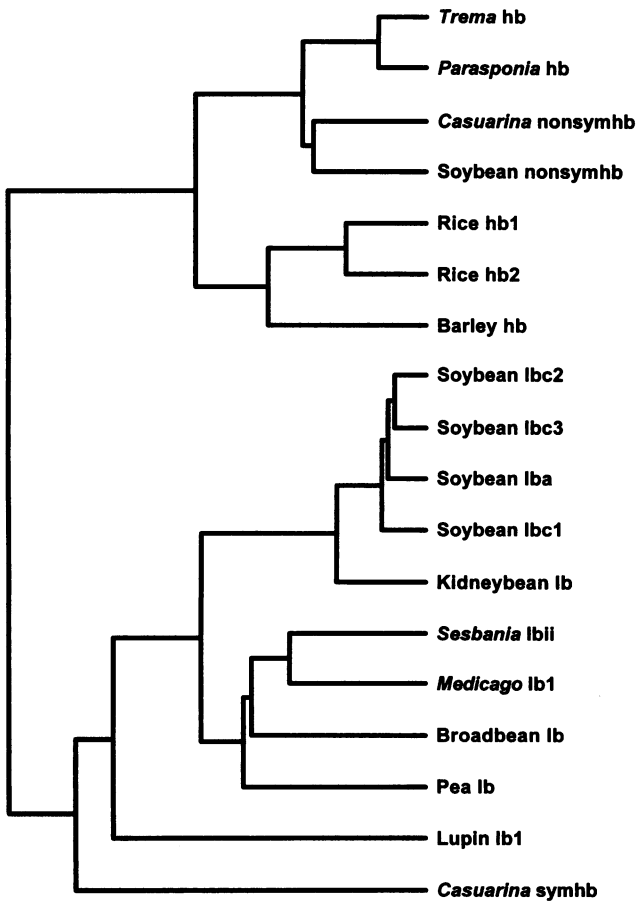

FIG. 4. Plant hemoglobin protein similarity tree. The similarity tree was constructed using the GCG PILEUP program. Sequences used are referenced in Fig. 1.

The Northern blot data are supported by *in situ* hybridization analyses that show that the soybean nonsymbiotic hemoglobin is expressed in both nodules and shoot tips. In nodules, the new hemoglobin gene is expressed in many cell types whereas leghemoglobins are present only in the bacteroid-containing cells (25).

DISCUSSION

The New Hemoglobin Gene Is Present in Soybean and Other Legumes and Encodes a "Nonsymbiotic" Protein. We have isolated a novel, expressed hemoglobin gene from soybean that is not a member of the leghemoglobin gene family. The gene has the characteristic four exon, three intron structure of all known plant hemoglobins. The predicted amino acid sequence is only 57–58% identical to leghemoglobin sequences (Table 1), but contains the critical heme and ligand binding residues of a functional hemoglobin. It resembles other nonsymbiotic hemoglobins in having an extension of five amino acids at the amino terminus (Fig. 1, Table 1). This new soybean gene is most abundantly expressed in stems and cotyledons of seedlings, and in roots, some nodule tissue, and young leaves of mature plants. Expression of the new soybean hemoglobin therefore differs considerably from that of the nodule-specific leghemoglobins and resembles that of the nonsymbiotic gene of *C. glauca* (7). Both the gene sequence and expression pattern indicate that we have isolated a nonsymbiotic soybean hemoglobin gene. Cross-hybridization of this soybean gene with sequences in the genomes of other legumes suggests that all legumes contain a similar nonsymbiotic hemoglobin in addition to their well-characterized leghemoglobin genes.

Evolution of Plant Hemoglobins. The isolation of this gene from soybean adds another important example to a growing list of plant hemoglobins whose expression is not associated with any symbiotic interactions with micro-organisms. The GenBank EMBL data base contains the 5′ sequences of two classes of rice expressed sequence tags with extensive homology to the nonsymbiotic hemoglobins (Fig. 1). The presence of expressed nonsymbiotic hemoglobin genes in a number of dicots, now including the legumes, as well as the monocots

```
soyhb        -184  TAAGCCACACAAATGGGAATGA-CTCCCCATTACA-ATGAAGGGCCAA--CTTTCATTTTCAATGAA-----TCCCACTATAAA  -110

Parahb       -109  TAAAAAACCCAAAGAT--ATGG-CTCCCCAATACCCT-GAAGAGTTACACACGA--------TCCCCATTTTTTCTACTATATA  -38
Tremahb      -117  AAAAAAACCCAAGGAG--ATGG-CTCTCCAGTACCCT-GAAGAGTTACATCTTATGATCTTATCCCCCATTTTTTCCATATATA  -38
Casnonsymhb -208  CAATTGACCCAAAGAA--ATGG-CTTTC---GACCCACGAAGAGCCGGAGCTATCCCTGTTACG---------TGCGCTATATA  -140

soylbc3      -141  AAGTTTTTGAAAAGATGA-TTGTCTCTTCACCATACCA---ATTGAT-------CACCCTCCTCCAACA---AGCCAAGAGAGAC  -71
soylba       -133  AAATTTTTTAAAAGATCG-TTGTTTCTTCTTCATCATGCTGATTGA--------CACCCTCC-------ACAAGCCAAGAGAAAC  -65
soylbc1      -128  AGGATTTTGAAAAGATCA-TTGGCTCTTCGTCATGCCG---ATTGA--------CACCCTCC-------ACAAGCCAAGAGAAAC  -63
soylbc2      -139  AAGATTTTGAAAAGATCATTTGGCTCTTCATCATGCCG---ATTGA--------CACCCTCC-------ACAAGCCAAGAGAAAC  -63
Cassymhb1    -490  ACTTCAATCCCAAGATGTCCTCTCTCTTATTGATATTTGAACAACAACAAAGATAAACAACCATTATCCCTACCAAGCAAGTAA  -408
Cassymhb1    -324  ACTTCAATCCCAAGATGTCCTCTCTCTTATTGATATTTGAACAACAACAAAGATAAACAACCTTTATCCTTACCAAGCAGGTAA  -241
```

FIG. 5. Sequence alignment of the promoter regions of hemoglobin genes from soybean, *Parasponia, Trema*, and *Casuarina*. Gaps, indicated by dashes, have been introduced to maximize similarity. The nodulin motifs, 5′-AAAGAT and 5′-CTCTT, are indicated in bold type, as is the putative nonsymbiotic motif 5′-GAAGAG. The TATA box is underlined. *Casuarina glauca* symbiotic hemoglobin 1 (7) and *C. glauca* hb nonsymbiotic hemoglobin (6) and soybean nonsymbiotic hemoglobin (this study) are numbered from the start codon, *Parasponia andersonii* hemoglobin (3), *Trema tomentosa* hemoglobin (4), Soybean *lba*, *lbc1*, *lbc2*, *lbc3* (27) are numbered from the transcription start. The *C. glauca* symbiotic hemoglobin promoter contains a direct repeat in this region and both copies are included.

barley, wheat, maize (8), and rice, shows that nonsymbiotic hemoglobin is widespread and possibly ubiquitous in the plant kingdom. It is likely that the more specialized symbiotic hemoglobins arose by gene duplication of a preexisting "nonsymbiotic" gene. The presence in the ancient tetraploid soybean of two gene clusters of leghemoglobins (26), and now of two nonsymbiotic hemoglobin genes (Fig. 2), is consistent with an origin for the leghemoglobins by a gene duplication event occurring before the hybridization of the diploid progenitors of *G. max*.

A similarity-tree based on pairwise comparisons of many of the known plant hemoglobin amino acid sequences was constructed using the Genetics Computer Group (GCG; Madison, WI) PILEUP program (Fig. 4). All the nonsymbiotic hemoglobins cluster together, including the newly characterized gene from soybean. Within the nonsymbiotic hemoglobins, the dicot and monocot sequences form separate subclusters. All of the leghemoglobin sequences cluster together, with those from plants with determinate, indeterminate and lupin-type nodules forming separate subgroups, consistent with the analysis of Marcker and Sandal (26). The symbiotic hemoglobins of one species are more similar to symbiotic hemoglobins of other species than to the nonsymbiotic hemoglobins within the same species. This analysis suggests that leghemoglobins arose before speciation within the legumes, the result of an initial gene duplication event that was followed by further duplications giving rise to the symbiotic leghemoglobin gene families within the present day legumes.

The *C. glauca* symbiotic hemoglobin (cassymhb) clusters with the leghemoglobins rather than with the nonsymbiotic hemoglobins. The *C. glauca* nonsymbiotic hemoglobin protein is 87% similar to the soybean nonsymbiotic hemoglobin and groups with it on a separate subbranch of the dendrogram (Fig. 4). The promoter motifs of these two nonsymbiotic genes are also similar (Fig. 5). Because the symbiotic hemoglobins (cassymhb and the leghemoglobins) are more similar to each other than to their putative progenitor nonsymbiotic hemoglobins in sequence, expression and regulation, they may have been derived from the same initial gene duplication before the divergence of *Casuarina* and the legumes. The similarities of the symbiotic and nonsymbiotic hemoglobins from *Casuarina* and soybean suggests that the species are more closely related than has previously been thought (28). A study based on the DNA sequences of the large subunit of the Rubisco gene has also placed the legumes close to *Casuarina* (29).

Because of the animal/plant cross-kingdom occurrence of hemoglobin, the conserved functional amino acid residues in the proteins and the conserved positions of the introns within the genes, hemoglobins must be very ancient. They presumably predate the divergence of plants and animals and existed well before the diversification of the Angiosperms. We predict that hemoglobin genes will also be present in the genomes of more primitive plant groups, such as bryophytes, pteridophytes, cycads, and gymnosperms.

Possible Functions of Nonsymbiotic Plant Hemoglobin. Two possible functions of nonsymbiotic plant hemoglobins have been proposed, as a sensor of oxygen concentration, or as a carrier in oxygen transport (9). Some hemoglobins are known to regulate associated enzymatic activities by changes in their conformation due to substrate binding (30, 31). Both the *C. glauca* and soybean nonsymbiotic hemoglobin genes have moderate levels of expression, and the proteins may therefore be more abundant than is expected for an oxygen sensor. The induction in barley of hemoglobin expression by low oxygen tension (8) suggests that hemoglobin may be a normal component of the anaerobic response in plants, presumably to facilitate oxygen diffusion at low oxygen concentrations and supports a role for nonsymbiotic hemoglobin in oxygen transport.

Nonsymbiotic hemoglobin may facilitate oxygen diffusion in rapidly dividing cells, such as those in the root meristem. Our Northern blot analyses show that expression of the soybean nonsymbiotic hemoglobin gene is higher in the root elongation zone than in the root tip, suggesting that any requirement for oxygen transport is not confined to meristematic cells. The high levels of expression in cotyledons and stems also suggests that nonsymbiotic hemoglobin gene expression does not correlate with cell division, but is perhaps associated with high levels of metabolic activity. Nonsymbiotic hemoglobin genes are all expressed in various metabolically active tissues such as developing seeds and roots (4, 8) or in the vascular tissues of leaves, stems and seedling cotyledons (ref. 8 and this study). These are all sites of intense short distance solute transfer, an energy demanding process. It is possible that the nonsymbiotic hemoglobin is facilitating intracellular diffusion of oxygen to the mitochondria in metabolically active cells in order to meet an increased demand for oxidative respiration.

The widespread distribution and expression of the nonsymbiotic hemoglobin genes implies that they play an important role(s) in the metabolic biochemistry of all plants, perhaps comparable with the oxygen binding roles of hemoglobins in animal systems.

We thank Janice Norman for excellent technical assistance, Fraser Bergersen and Inara Licis for providing soybean plants, Robert Hill for making information regarding the barley hemoglobin gene available before publication, and U. H. Jacobsen for providing unpublished *in situ* hybridization data showing the expression of the soybean nonsymbiotic hemoglobin gene in a number of cell types in soybean nodules.

1. Appleby, C. A. (1984) *Annu. Rev. Plant Physiol.* **35,** 443–478.
2. Appleby, C. A., Tjepkema, J. D. & Trinick, M. J. (1983) *Science* **220,** 951–953.

3. Landsmann, J., Dennis, E. S., Higgins, T. J. V., Appleby, C. A., Kortt, A. A. & Peacock, W. J. (1986) *Nature (London)* **324,** 166–168.
4. Bogusz, D., Appleby, C. A., Landsmann, J., Dennis, E. S., Trinick, M. J. & Peacock, W. J. (1988) *Nature (London)* **331,** 178–180.
5. Fleming, A. I., Wittenberg, J. B., Wittenberg, B. A., Dudman, W. F. & Appleby, C. A. (1987) *Biochim. Biophys. Acta* **911,** 209–220.
6. Christensen, T., Dennis, E. S., Peacock, J. W., Landsmann, J. & Marcker, K. A. (1991) *Plant Mol. Biol.* **16,** 339–344.
7. Jacobsen-Lyon, K., Jensen, E. O., Jorgensen, J.-E., Marcker, K. A., Dennis, E. S. & Peacock, W. J. (1995) *Plant Cell.* **7,** 213–223.
8. Taylor, E. R., Nie, X. Z., MacGregor, A. W. & Hill, R. D. (1994) *Plant Mol. Biol.* **24,** 853–862.
9. Appleby, C. A., Bogusz, D., Dennis, E. S. & Peacock, W. J. (1988) *Plant Cell Environ.* **11,** 359–367.
10. Miao, G.-H. & Verma, D. P. S. (1993) *Plant Cell* **5,** 781–794.
11. Hirano, H. (1989) *J. Protein Chem.* **8,** 115–130.
12. Wiborg, O., Hyldig-Nielsen, J. J., Jensen, E. O., Paludan, K. & Marcker, K. A. (1983) *EMBO J.* **2,** 449–452.
13. Davis, L. G., Dibner, M. D. & Battey, J. F. (1988) *Basic Methods in Molecular Biology* (Elsevier, New York).
14. Burke, T. J., Callis, J. & Vierstra, R. D. (1988) *Mol. Gen. Genet.* **213,** 435–443.
15. Konieczny, A. (1987) *Nucleic Acids Res.* **15,** 6742.
16. Hyldig-Nielsen, J. J., Jensen, E. L., Paludan, K., Wiborg, O., Garrett, R., Jorgensen, P. & Marcker, K. A. (1982) *Nucleic Acids Res.* **10,** 689–701.
17. Wiborg, O., Hyldig-Nielsen, J. J., Jensen, E. O., Paludan, K. & Marcker, K. A. (1982) *Nucleic Acids Res.* **10,** 3487–3494.
18. Lee, J. S. & Verma, D. P. S. (1984) *EMBO J.* **3,** 2745–2752.
19. Metz, B. A., Welters, P., Hoffman, H. J., Jensen, E. O., Schell, J. & de Bruijn, F. J. (1988) *Mol. Gen. Genet.* **214,** 181–191.
20. Gallusci, P., Dedieu, A., Journet, E. P., Huguet, T. & Barker, D. G. (1991) *Plant Mol. Biol.* **17,** 335–349.
21. Richardson, M., Dilworth, M. J. & Scawen, M. D. (1975) *FEBS Lett.* **51,** 33–37.
22. Lehtovaara, P., Lappalainen, A. & Ellfolk, N. (1980) *Biochim. Biophys. Acta* **623,** 98–106.
23. Ramlov, K. B., Laursen, N. G., Stougaard, J. & Marker, K. A. (1993) *Plant J.* **4,** 577–580.
24. Szczyglowski, K., Szabados, L., Fujimoto, S. Y., Silver, D. & de Bruijn, F. J. (1993) *Plant Cell* **6,** 317–332.
25. Nap, J.-P. & Bisseling, T. (1990) *Science* **250,** 948–954.
26. Marcker, K. A. & Sandal, N. N. (1987) in *Plant Molecular Biology,* eds. von Wettstein, D. & Chua, N.-H. (Plenum, New York), pp. 503–508.
27. Stougaard, J., Sandal, N. N., Gron, A., Kuhle, A. & Marcker, K. A. (1987) *EMBO J.* **6,** 3565–3569.
28. Cronquist, A. (1981) *An Integrated System of Classification of Flowering Plants* (Columbia Univ. Press, New York).
29. Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., *et al.* (1993) *Ann. Missouri Bot. Gard.* **80,** 528–580.
30. Zhu, H. & Riggs, A. F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5015–5019.
31. Sherman, D. R., Guinn, B., Perdok, M. M. & Goldberg, D. E. (1992) *Science* **258,** 1930–1932.