Molecular Cloning and Characterization of a Moss (*Ceratodon purpureus*)
Nonsymbiotic Hemoglobin Provides Insight into the Early Evolution of Plant Nonsymbiotic Hemoglobins

Verónica Garrocho-Villegas and Raúl Arredondo-Peter
Laboratorio de Biofísica y Biología Molecular, Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México

Nonsymbiotic hemoglobins (nsHbs) are widespread in plants including bryophytes. Bryophytes (such as mosses) are among the oldest land plants, thus an analysis of a bryophyte nsHb is of interest from an evolutionary perspective. However, very little is known about bryophyte nsHbs. Here, we report the cloning and characterization of an *nshb* gene (*cerhb*) from the moss *Ceratodon purpureus*. Sequence analysis showed that *cerhb* is interrupted by 3 introns in identical position as all known plant *nshb* genes, which suggests that the ancestral *nshb* gene was interrupted by 3 introns. Expression analysis showed that *cerhb* expresses in protonemata and gametophytes growing in normal conditions and that it overexpresses in protonemata subjected to osmotic (sucrose), heat-shock, cold-, and nitrate-stress conditions. Also, modeling of the *Ceratodon* nsHb (CerHb) tertiary structure suggests that CerHb is hexacoordinate and that it binds O$_2$ with high affinity. Comparative analysis of the predicted CerHb with native rice Hb1 and soybean leghemoglobin structures revealed that the major evolutionary changes that probably occurred during the evolution of plant Hbs were 1) a hexacoordinate to pentacoordinate transition at the heme prosthetic group, 2) a length decrease at the CD-loop and N- and C-termini regions, and 3) the compaction of the protein into a globular structure.

Introduction

Nonsymbiotic hemoglobins (nsHbs) are plant proteins that reversibly bind O$_2$ and other gaseous ligands. These proteins have been detected in seeds, roots, stems, leaves, and flowers from plants growing in normal conditions (Taylor et al. 1994; Andersson et al. 1996; Lira-Ruan et al. 2001; Ross et al. 2004); however, levels of nsHbs increase when plants are subjected to stress conditions, such as microaerobiosis and anoxia (Taylor et al. 1994; Trevaskis et al. 1997; Lira-Ruan et al. 2001). Based on O$_2$ affinity and sequence similarity nsHbs are classified into class 1 and class 2 (nsHbs-1 and nsHbs-2, respectively) (Trevaskis et al. 1997). The O$_2$ affinity of nsHbs-1 and nsHbs-2 is very high and high, respectively (Arredondo-Peter et al. 1997; Duff et al. 1997; Trevaskis et al. 1997). The very high O$_2$ affinity of nsHbs-1 results from an extremely low O$_2$ dissociation rate constant, which suggests that in vivo these proteins do not release O$_2$ after oxygenation and therefore have functions other than O$_2$ transport (Arredondo-Peter et al. 1998). Recent evidence showed that a possible function of nsHbs-1 is to modulate the levels of NO and redox potentials (Igamberdiev et al. 2001; Ross et al. 2004); however, levels of nsHbs were generated by adding 10% sucrose (Umeda et al. 1994), and 85 mM Ca(NO$_3$)$_2$ (Ohwaki et al. 2005) to the growth medium, respectively. Cold-stress conditions were generated by incubating protonemata at 4 °C for 24 h (Takahashi et al. 1994).

Materials and Methods

Plant Material and Growing Conditions

*Ceratodon purpureus* protonemata were kindly provided by Dr Ralf Reski (Albert-Ludwigs-Universität, Freiburg, Germany) as an axenic culture. Protonemata were cultured in Knop medium (Reski and Abel 1985) and grown in a plant growing chamber (Biotronette Plant Growth Chamber, Lab-Line Instruments, Melrose Park, IL) at 25 °C with 16/8-h light/dark periods. *Ceratodon purpureus* gametophytes were obtained after culturing protonemata for 8 weeks. Normal and stressed (see below) plants were harvested, immediately frozen in liquid N$_2$, and stored at −70 °C until used.

The (*nshb*) gene expression in plants subjected to stress conditions was evaluated in 2-week-old protonemata by duplicate. Osmotic-, salt-, and nitrate-stress conditions were generated by adding 10% sucrose (Umeda et al. 1994), 250 mM NaCl (Umeda et al. 1994), and 85 mM Ca(NO$_3$)$_2$ (Ohwaki et al. 2005) to the growth medium, respectively. Cold-stress conditions were generated by incubating protonemata at 4 °C for 24 h (Takahashi et al. 1994).
Dark-stress conditions were generated by incubating the protonemas in darkness for 24 h inside a plant growth chamber (Lira-Ruan et al. 2001). Flooding-stress conditions were generated by fully covering the protonemas with distilled water for 24 h (Taylor et al. 1994; Lira-Ruan et al. 2001). Heat-shock stress conditions were generated by incubating the protonemas at 42 °C for 2 h (Higo K and Higo H 1993). Control protonemas (i.e., those growing under normal conditions) were cultivated for 2 weeks, and control and stressed protonemas were harvested at the same time.

Total DNA and Poly(A⁺) RNA Isolation

Total DNA was isolated from ~100 mg of *C. purpureus* protonemas using the ceteryltrimethylammonium bromide (CTAB) method (Crose and Amorese 1978). Poly(A⁺) RNA was isolated from plant tissues using the QuickPrep Micro mRNA Purification kit (Amersham Biosciences, Little Chalfont Buckinghamshire, United Kingdom) following the manufacturer’s instructions. Total DNA and poly(A⁺) RNA were quantitated by spectrophotometry assuming 1 A₂₆₀ = 50 or 40 µg/ml for DNA or RNA, respectively (Ausubel et al. 1995).

Polymerase Chain Reaction Amplification, Gene Cloning, and DNA Sequencing

Primers were designed for polymerase chain reaction (PCR) to amplify the *C. purpureus* nsHb gene using sequences at the start and stop codons of the *C. purpureus* nsHb cDNA (GenBank accession number AF309562). The primer sequences (degenerated for the NcoI and EcoRI restriction sites [underlined]) were (sense) primer CerHb/ATG-

<table>
<thead>
<tr>
<th>ATG-</th>
<th>CAGTCG-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>RI:</td>
</tr>
</tbody>
</table>

and (antisense) primer CerHb/TAA-EcoRI: 5'-GAAATCTTACTGAGCAGCCTC-3'. Total *C. purpureus* DNA (~1 µg) was used as a template for PCR amplification. PCR components and concentrations were 1 µM of each sense and antisense primer, 400 µM of each deoxynucleotide triphosphate, and 1 U of *Tag* DNA polymerase (Invitrogen, Carlsbad, CA) in 1X PCR buffer containing 4 mM MgCl₂ in a final volume of 25 µl. PCR amplification was carried out for 35 cycles at 55 °C for annealing using a theramicycler (Minicycler, MJ Research, Watertown, MA). PCR products were detected in a 1.2% agarose gel after staining with ethidium bromide, isolated from the gel using the GeneClean kit (Q-BIOgene, Carlsbad, CA), and cloned into the pCR2.1 cloning vector (Invitrogen) following the manufacturer’s instructions. Insert DNA was fully sequenced in both orientations at the Molecular Biology Facility of the Cell Physiology Institute of the National Autonomous University of México.

Southern Blot Analysis

Southern blot analysis of the total *C. purpureus* DNA was done as described by Arredondo-Peter et al. (1997) using the *C. purpureus* nsHb gene as a probe. Membranes were prehybridized and hybridized at 55 °C and washed at high stringency (60 °C) in 2× standard saline citrate (SSC)/0.1% sodium dodecyl sulfate (SDS) twice for 5 min each and 0.5× SSC/0.1% SDS twice for 15 min each.

Gene Expression Analysis by Reverse Transcriptase–Polymerase Chain Reaction

Expression of the *C. purpureus* nsHb gene was examined in protonemas and gametophytes (4 and 10 weeks old, respectively) and in protonemas grown under stress conditions (see above) by reverse transcriptase–polymerase chain reaction (RT-PCRs) using the GeneAmp RNA PCR kit (Applied Biosystems, Branchburg, NJ) as described by Arredondo-Peter et al. (1997). Both reverse transcription and PCR amplification were performed using the CerHb/ATG-NcoI and CerHb/TAA-EcoRI primers and 100 ng of template poly(A⁺) RNA. Amplification conditions were identical to those used for the amplification of the *C. purpureus* nsHb gene (see above). After gel electrophoresis, PCR fragments were stained with ethidium bromide, and the gel was photodocumented, scanned, and analyzed using the ImageJ 1.36b software (http://rsb.info.nih.gov/ij/). The relative abundance of the CerHb transcripts in control and differentially treated protonemas was calculated from the intensity of each band using as reference the intensity detected in nsHb amplicon observed in protonemas grown under normal conditions.

In Silico Analysis and Molecular Modeling

Multiple sequence alignment and cluster analysis of the *C. purpureus* nsHb and selected plant Hbs were performed using the Neighbor-Joining method of the ClustalX program (Thompson et al. 1997). The tertiary structure of the *C. purpureus* nsHb was predicted by homology modeling using the crystal structure of the rice Hb1 (Hargrove et al. 2000) (PDB ID 1D8U) as a template as described by Gopalasubramaniam, Garrocho-Villegas, et al. (2008).

Results and Discussion

Cloning of a *C. purpureus* nsHb Gene

The nsHbs evolved alongside land plants after bryophyte-like organisms colonized land about 450 MYA. Thus, it is likely that these proteins played roles in plant adaptation to the terrestrial environment (Ross et al. 2002; Garrocho-Villegas et al. 2007). To characterize a bryophyte nsHb, we designed PCR primers from the *C. purpureus* Hb cDNA (GenBank accession number AF309562) and used them to amplify fragments of 540 and 950 bp in length from template *C. purpureus* DNA (supplementary fig. S1, Supplementary Material online). These PCR fragments were cloned and sequenced. Blast analysis (Altschul et al. 1990) with sequences from the GenBank database revealed that sequence of the 540-bp fragment had no similarity to plant and nonplant Hbs and that it corresponded to a nonspecific amplification. However, the coding sequence of the 950-bp fragment was identical to that from the *C. purpureus* Hb cDNA, thus...
showing that it corresponded to the *C. purpureus* hb (*cerhb*) gene. The CerHb cDNA and gene sequences were compared to identify exon and intron sequences. The *cerhb* gene has 4 exons interrupted by 3 introns (supplementary fig. S2, Supplementary Material online), which are located at identical positions as all the known plant *nshb* and *leghemoglobin* (*lbs*). This observation indicates that the ancestral *nshb* gene was interrupted by 3 introns located in identical positions as those in modern *nshb* and *lb* genes.

Characterization of the CerHb Protein Sequence

The *cerhb* gene codes for a predicted protein of 177 amino acid residues in length (supplementary fig. S2, Supplementary Material online) with a calculated molecular mass of 19,535 Da. Sequence analysis showed that CerHb contains distal (H86) and proximal (H121) His and Phe B10 (F52), which are highly conserved in plant Hbs. However, in *Ceratodon* and *Physcomitrella* nsHbs (see below), the CD1 position is occupied by Tyr (Y66) instead of Phe, which is conserved in evolved nsHbs (supplementary fig. S3, Supplementary Material online). This observation shows that Tyr in ancestral nsHbs occupied the CD1 position and was replaced by Phe during the evolution of nsHbs.

Arredondo-Peter et al. (2000) identified and partially characterized an nsHb sequence from the moss *P. patens*. Multiple sequence alignment showed that moss (*C. purpureus* and *P. patens*) nsHbs are the longest known plant Hbs and that the size of plant nsHbs decreased over time, mostly due to a loss of amino acid residues at the N-terminal (pre-helix A) region (supplementary fig. S3, Supplementary Material online) (Ross et al. 2002). Also, sequence alignment revealed that the *C. purpureus* and *P. patens* nsHbs are 80% identical and that the highest sequence variability occurs at the N-termini regions (i.e., within the first 35 amino acid residues). The identity value of moss nsHbs is similar to that observed between evolved nsHbs. For example, the identity value between rice and maize nsHb1s is approximately 80%. This observation suggests that nsHbs evolved at approximately the same rate and that major changes during plant nsHb evolution occurred at the N-terminal region (see below).

Identification of the cerhb Gene Copy Number

The number of *nshb* gene copies varies with each species in higher plants. For example, single and multiple *nshb* gene copies exist in barley (Taylor et al. 1994) and rice (Arredondo-Peter et al. 1997; Lira-Ruan et al. 2002), respectively. Also, the search for homologs to the globin fold in the SUPERFAMILY database (Gough et al. 2001) identified several Hb sequences in the *P. patens* genome (data not shown). When total *C. purpureus* DNA was subjected to Southern blotting with the *cerhb* probe, 2 hybridizing fragments of 4 and 6–7 kb in length were detected in the DNA digested with EcoRI (fig. 1). This observation suggests that at least 2 copies of the *cerhb* gene exist in the *C. purpureus* genome. The existence of cDNAs coding for CerHb indicates that *cerhb* is functional in *C. purpureus*.

Expression of the cerhb Gene in *C. purpureus* Grown under Normal and Stress Conditions

The nsHb transcripts and proteins have been reported to exist in several organs from plants growing under normal and stress conditions (Taylor et al. 1994; Andersson et al. 1996; Arredondo-Peter et al. 1997; Trevaskis et al. 1997; Lira-Ruan et al. 2001; Ross et al. 2001). We used RTPCR and specific primers for CerHb to evaluate the *cerhb* gene expression in *C. purpureus* gametophytes and protonemases growing under normal and stress conditions. We selected stress conditions that others have evaluated for *nshb* gene expression in diverse higher plants; also, these stress conditions have been postulated to be primary plant adaptations during land colonization (i.e., osmoregulation, osmoprotection, temperature tolerance, and nutrient availability) (Rensing et al. 2008). We amplified transcripts of the expected molecular size (531 bp) from the *C. purpureus* gametophytes and protonemases grown under normal and stress conditions. Figure 2a shows that the *cerhb* gene expresses in *C. purpureus* organs. Analysis of the relative abundance for the CerHb transcripts showed that *cerhb* expression is higher in gametophytes than in protonemases. Also, the *cerhb* gene over- and underexpresses in protonemases grown in sucrose-, heat-shock-, cold- and nitrate-, and salt-, dark- and flooding-stress conditions, respectively (fig. 2b). These observations suggest that higher levels of CerHb could be required for differentiating gametophytes and that CerHb functions during the plant response to osmotic (sucrose), temperature (heat and cold), and high levels of nitrate stresses but not in salt, dark, and flooding stresses.
Predicted Tertiary Structure of CerHb

Nothing is known about the tertiary structure of bryophyte nsHbs. Understanding the structural properties of nsHbs is important to clarifying the role these proteins play in plant organs. In silico methods exist to predict with high reliability the tertiary structure of proteins from template structures (Sañez-Rivera et al. 2004; Gopalasubramaniam, Garrocho-Villegas, et al. 2008). Predicting a structure yields insights into potential evolutionary pathways for nsHbs. Because CerHb and rice Hb1 are approximately 50% identical, we predicted the tertiary structure of CerHb using rice Hb1 as a template. Figure 3a shows that the predicted CerHb folds into the globin fold and that its structure is quite similar to that of native rice Hb1, including the positions of helices E and F, where distal and proximal His are located, respectively. However, the structures differ in that the N- and C-termini regions are longer in CerHb than in rice Hb1, the CD-loop folds differently, and an (2 turns) a-helix exists within the prehelix A region of predicted CerHb, which may facilitate protein translocation. Ross et al. (2002) detected leader peptidase-like sites in moss nsHbs and suggested that the ancestor to plant nsHbs was translocated to cellular organelles. The existence of an a-helix at the prehelix A region in predicted CerHb (fig. 3a) is consistent with the possibility that CerHb (and other bryophyte nsHbs) may function inside cell organelles. If this is correct, it is possible that nsHbs became cytoplasmic during evolution (Ross et al. 2002). The examination of the amino acid residues that are essential for ligands binding showed that the positions of the proximal and distal His, Phe B10, Tyr/Phe CD1, and Tyr 158 are quite similar in predicted CerHb and native rice Hb1 (fig. 3b). These observations suggest that heme–Fe in CerHb is hexacoordinate and that the kinetic properties of CerHb and rice Hb1 are similar. Thus, it is likely that, because of a low O2 dissociation rate constant, the O2 affinity of CerHb is rather high, and was, as well, in ancestral nsHb.

The Structural Evolution of CerHb and Plant nsHbs

Comparative analysis of the CerHb, rice Hb1, and soybean leghemoglobin a (Lba) tertiary structures (i.e., primitive to evolved plant Hbs) reveals major evolutionary changes that probably occurred during plant Hbs evolution (fig. 4). These changes include 1) a hexacoordinate to pentacoordinate transition at the heme prosthetic group (Hoy et al. 2007; Gopalasubramaniam, Kovacs, et al. 2008), 2) a length decrease at the CD-loop and N- and C-termini regions, and 3) the compaction of the protein into a globular structure. Some of these structural changes have been described for the nsHb to Lb transition and for the specialization of a symbiotic function for Lb (Gopalasubramaniam, Kovacs, et al. 2008). Thus, the pattern of structural evolution for nsHbs has apparently been conserved for 450 Myr, which allowed nsHbs to specialize and help plants adapt to and diversify within the land environment.
**Supplementary Material**

Supplementary figures S1–S3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

The authors wish to express their gratitude to Dr Ralf Reski (Albert-Ludwigs-Universität, Freiburg, Germany) for kindly providing the *C. purpureus* protonemas, to an anonymous reviewer for useful suggestions to improve the contents of this article, and to Miss Gillian Klucas for corrections made to improve the English language. This work was funded by Consejo Nacional de Ciencia y Tecnología (CoNaCyT project no. 42873Q), México. V.G.-V. is a postdoctoral fellow financed by CoNaCyT (IdAP 9272).

**Literature Cited**


Geoffrey McFadden, Associate Editor

Accepted April 7, 2008