Detection, PCR-amplification and characterization of a 4 intron-containing non-symbiotic hemoglobin gene from the moss *Physcomitrella patens*

Fernando Martínez-Ocampo, Consuelo Vázquez-Limón, Raúl Arredondo-Peter*

Laboratorio de Biofísica y Biología Molecular, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Col. Chamilpa, 62210 Cuernavaca, Morelos, México

*Author for Correspondence: Raúl Arredondo-Peter, email: ra@uaem.mx Received 14 Mar 2012; Accepted 12 Apr 2012; Available Online 13 Apr 2012

Abstract

It has been well established that land plant non-symbiotic hemoglobin (*nshb*), symbiotic hemoglobin (*shb*) and leghemoglobin (*lb*) genes are interrupted by 3 introns (IVS-I, IVS-II and IVS-III). Here, we report the detection, PCR-amplification and DNA sequencing of a 4 intron-containing *nshb* (*nshb2*) gene from the *Physcomitrella patens* DNA. Sequence analysis revealed that *nshb2* is interrupted by introns IVS-I, IVS-II and IVS-III, and by an extra intron (pre-IVS-I) interrupting coding sequences upstream to IVS-I. Phylogenetic analysis showed that *P. patens* *nsHb2* clusters with bryophyte *nsHbs*, however *nsHb2* is divergent from *P. patens* *nsHb1*. Modeling of the tertiary structure of *P. patens* *nsHb2* showed that this protein could fold into the myoglobin-fold. The existence of intron pre-IVS-I in *P. patens* *nshb2* but not in other known land plant *nshbs*, *shbs* and *lbs* suggests that this intron inserted after gene duplication that originated *nshb2*. If this is correct intron insertion possibly occurred in land plant *nshbs* other than *P. patens* *nshb2*. This will be verified as novel *nshbs* become available from newly sequenced land plant genomes.

Keywords: Gene structure; Hemoglobin; Intron; Non-symbiotic; *Physcomitrella patens*; Plants

Abbreviations: Hb, hemoglobin; IVS, intervening sequence (intron); Lb, leghemoglobin; Mb, myoglobin; nsHb, non-symbiotic hemoglobin; ORF, open reading frame; sHb, symbiotic hemoglobin; tHb, truncated hemoglobin

1. Introduction

Globins (Hbs) are a large family of proteins that have been detected in organisms from the three kingdoms of life [1]. In land plants three types of Hbs exist: symbiotic Hbs (*sHbs*), non-symbiotic Hbs (*nsHbs*) and truncated Hbs (*tHbs*) [2-4]. Symbiotic Hbs (or leghemoglobins (*Lbs*) when isolated from legume species) are only localized in nODULES of N₂-fixing land plants [5-9]. In contrast, *nsHbs* and *tHbs* are widely distributed in land plants and are localized in tissues from symbiotic and non-symbiotic plant organs [10-16]. Phylogenetic analysis of land plant Hbs suggests that these proteins vertically evolved and diversified with land plants, that the *nsHb*, *sHb* and *Lb* and *tHb* lineages evolved from different ancestors previously to the emergence of Archaeplastida (i.e. algae and land plants), and that *sHbs* and *Lbs* originated from a *nsHb* ancestor [4, 17, 18].

During the last decades the availability of sequenced genomes permitted the identification and characterization of *hb* genes from several land plants, ranging from primitive bryophytes to evolved monocots and dicots. Known land plant *nshb*, *shb* and *lb* genes are interrupted by 3 introns (IVS-I, IVS-II and IVS-III) mostly conserved at positions B12.2, E15.0 and G7.0 (i.e. intron position within the myoglobin (Mb)-fold and phase). Thus, the apparent gene structure for land plant *nshbs*, *shbs* and *lbs* is 4 exons/3 introns. In contrast, intron position and number is variable in known land plant *thbs* [18]. During the search for *nshb* gene copies in the genome of the moss *Physcomitrella patens* at the COSSMOS database (see subsection 2.1.) we detected a *nshb* (*nshb2*) that is interrupted by 4 introns. This observation suggested that gene structure for *nshb2* is 5 exons/4 introns. Here, we report the detection, PCR-amplification and characterization of 4 intron-containing *nshb2* from the *P. patens* DNA. Results from this work revealed that *nshbs* with gene structure other than 4 exons/3 introns exist in land plants.

2. Experimental

2.1. Database search and mapping of the *nshb2* gene in the *P. patens* genome

*Physcomitrella patens* *nshb* (*nshb1*) gene and protein sequences (GenBank accession number EF028055 and ABK20873, respectively) were used as probes to search for *nshb* gene copies in the *P. patens* genome at the COSMOS database (http://www.cosmoss.org/, last accessed on September 2011) using the Blast tool [19]. Putative *nshb* sequences were subjected to a FUGE analysis (http://www-cryst.bioc.cam.ac.uk, last accessed on September 2011) to determine the most similar globin structure and presence of proximal His at the Mb-fold position F8. Putative globins had to satisfy the following criteria: length higher than 100 amino acids, a FUGE Z score higher than 6 (which corresponds to 99% specificity [20]) with globin structures, and the presence of proximal His at position F8. Scaffolds containing copies of the *nshb* gene were used for mapping *nshbs*. This included the detection of open reading frames (ORFs) ~10-16 and ~5-9 Kb
2.2. Plant growth, isolation of total DNA, PCR-amplification and DNA sequencing

Physcomitrella patens var. Gransden was grown essentially as described by Garrocho-Villegas et al. [21]. Briefly, protonemas were cultured in Knop medium in a plant growth chamber (Biotronette Plant Growth Chamber, Lab-Line) at 20°C with 16/8 h light/dark periods for 1 to 2 months. Plants were harvested immediately frozen in liquid N₂ until used. Total DNA was isolated from 100 mg of P. patens protonemas using the cetyltrimethylammonium bromide (CTAB) method [22]. Primers were designed for PCR-amplification using sequences at the start and stop codons from the nsh2 gene detected at the COSMoss database (see subsection 2.1.). Primer sequences were: (sense) primer P. patens Hb2fwd: 5'-ATGGCATCATCGATTGGCG-3' and (antisense) primer P. patens Hb2revs: 5'-TCAAGTATGAGATTTAGCTGCTGC-3'. Total P. patens DNA (1.6 ng) was used as template for PCR-amplification. PCR components and concentrations were 14 pmol of each sense and antisense primer, 0.2 mM of each dNTP (Invitrogen), and 1 U of Taq DNA polymerase (Invitrogen) in 1X PCR buffer containing 2 mM MgCl₂ in a final volume of 25 µl. PCR-amplification was carried out for 35 cycles at 58°C for annealing using a thermalcycler (Minicycler, MJ Research). 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) and sequenced by PCR in both orientations using the sense and antisense primers at the Institute of Biotechnology of the National Autonomous University of México.

2.3. Sequence alignment and phenetic analysis

Pairwise and multiple sequence alignment of P. patens nsHb1 and nsHb2, and P. patens nsHb2 and selected Archaeplastida nsHbs and Lbs, respectively, were performed using the ClustalX program [23]. Archaeplastida nsHb and Lb sequences were obtained from the GenBank database (http://www.ncbi.nlm.nih.gov, last accessed on September 2011) using the accession numbers reported by Garrocho-Villegas et al. [4], Fernández et al. [24] and Vinogradov et al. [18]. Sequence alignment was manually verified using the procedure described by Kapp et al. [25] based on the Mb-fold [26]. Phenogram was obtained from aligned sequences using the MEGA program (version 5.05) [27] and the Neighbor Joining method [28] with 100000 bootstrap replicates.

2.4. Molecular modeling and analysis of the predicted P. patens nsHb2 tertiary structure

The tertiary structure of P. patens nsHb2 was modeled using the automated mode of the I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/, last accessed on November 2011) [29-31] and the crystal structure of rice Hb1 [32] (PDB ID 1D8U) as a template. Model reliability was evaluated using the Verify3D program (http://nihserver.mbi.ucla.edu/Verify3D/, last accessed on November 2011) [33]. Model was edited using the VMD program [34] and Adobe Photoshop® software. Distance of amino acid residues at the heme prosthetic group were calculated using the distance tool of the SwissPDBViewer program as described by Gopalasubramaniam et al. [35].

3. Results and Discussion

3.1. Detection and mapping of nshb2 in the P. patens genome

The search for copies of nshb genes at the COSMoss database using P. patens nshb1 sequences (GenBank accession number EF028055 and ABK208783) as probe showed that a number of scaffolds containing hh-like sequences exist into the P. patens genome. After sequence evaluation only nshb sequences from scaffold_71 and scaffold_660 satisfied the criteria to be considered authentic globins (see subsection 2.1.). Scaffold_71 contains the P. patens nshb1 sequence used as probe, and scaffold_660 contains a copy (nshb2) of nshb1. Mapping of nshb1 and 2 revealed that these genes are located at the – and + strands from scaffolds _71 and _660, respectively, that genes coding for a PAP2_haloperoxidase and hypothetical proteins are flanking to nshb1, and that no ORFs exist ~16 and 5 Kb up- and downstream to nshb2, respectively (Figure 1). Translation of nshb1 and 2 into predicted nsHb1 and 2 polypeptides and sequence comparison between nshb1 and 2 genes revealed that nshb1 and 2 are interrupted by 3 and 4 introns, respectively. Gene nshb1 is interrupted by IVS-I, IVS-II and IVS-III previously recognized in known land plant nshb1, nshb2 and lbs. Gene nshb2 is interrupted by IVS-I, IVS-II and IVS-III identically to those in nshb1, and by an extra intron interrupting coding sequences upstream to IVS-I (see subsection 3.2.). Because sequences deposited in genome databases occasionally result from sequencing and assembly...
3.2. PCR-amplification, DNA sequencing and analysis of P. patens nshb2

Total DNA isolated from P. patens protonemas was used as template for PCR-amplification in combination with primers designed from the nshb2 sequence detected in scaffold_660 (see subsection 2.2.). Figure 2 shows that a single fragment of 1,832 bp in size was amplified from the P. patens total DNA. This fragment was isolated and sequenced in both orientations. Sequence comparison showed that 1,832 bp fragment and nshb2 from scaffold_660 are identical, thus indicating that nshb2 exists in the P. patens DNA. Sequence analysis showed that nshb2 is interrupted by introns at positions 61, 341, 613 and 1,519 (Figure 3). Introns at positions 341, 613 and 1,519 interrupt nshb2 identically to IVS-I, IVS-II and IVS-III from known land plant nshbs, including to P. patens nshb1. Intron at position 61 interrupts nshb2 at coding sequences located upstream to IVS-I, thus we named this intron as pre-IVS-I. Sequences flanking the exon/intron boundaries to pre-IVS-I are tg/gt and ag/ag. With the exception of the 5’-end of exon/intron boundary in IVS-II (tg/gt), which is identical to the 5’-end of exon/intron boundary in pre-IVS-I, the exon/intron boundaries among pre-IVS-I and IVS-I, IVS-II and IVS-III are different (Figure 3). However, exon/intron boundaries in IVS-I, IVS-II and IVS-III (ga/gt and ag/gg, tg/gt and ag/ac, and ag/gt and ag/gt, respectively) from P. patens nshb2 and other land plant nshbs, such as P. patens nshb1 (GenBank accession number EF028055), rice hb1 [36] and Parasponia hb [37], are highly conserved.

3.3. Sequence alignment and phenetic analysis of P. patens nsHb2 and selected Archaeplastida hemoglobins

Sequence alignment showed that nsHb1 and nsHb2 are 63 and 78% identical and similar, respectively, and that these proteins contain the highly conserved distal (H84) and proximal (H119) histidine. Also, pairwise sequence alignment showed that intron pre-IVS-I from nshb2 is located at position preA11.1, and that introns IVS-I, IVS-II and IVS-III are located in nsHb1 and 2 at positions B12.2, E15.0 and G7.0, respectively (Figure 4). Phylogenetic analysis of selected Archaeplastida nHb and Lbs showed that P. patens nshb2 clusters with bryophyte nHbs, however nshb2 is divergent to P. patens nsHb1 and Ceratodon purpureus nHb (Figure 5). The above observations indicate that P. patens nsHb1 and 2 are conserved and evolved from a common ancestor. However, the existence of intron pre-IVS-I in nshb2 but not in nsHb1 and other known land plant nshbs suggests that pre-IVS-I inserted into nshb2 after gene duplication that originated nshb2 (see section 4).

3.4. Molecular modeling and analysis of the predicted P. patens nsHb2 tertiary structure

The crystal structure of few land plant nsHbs has been reported [17, 32, 38, 39], but no crystallographic analysis has been performed on bryophyte nHbs. However, the predicted structure of C. purpureus and P. patens nsHbs was reported by Garrocho-Villegas et al. [21] and is deposited in the Caspur database (http://mi.caspur.it/PMDB/, ID: PM0074985), respectively. Elucidating the tertiary structure of nHbs is important to understand the evolution and clarify the function of these proteins in land plants. We modeled the tertiary structure of predicted P. patens nshb2 using the automated mode of the I-TASSER server and the crystal structure of rice Hb1 [32] (PDB ID 1DSU) as a template. Evaluation using the Verify 3D program indicated that model for P. patens nshb2 is reliable, as values along the sequence were higher than 0. Figure 6 shows that the predicted structure of P. patens nsHb1 and 2 are highly similar, and that P. patens nshb2 could fold into the Mb-fold. Minor differences between the predicted structure of P. patens nsHb1 and 2 were detected at the conformation of pre-helix A and CD- and GH-loop. However, a major difference was that length of helix H is longer in predicted P. patens nsHb2 than nsHb1. The examination of amino acids that are essential for binding of ligands to the Fe-heme showed that position of proximal and distal His are similar among P. patens nsHb1 and 2 and hexacoordinate land plant nHbs, such as rice Hb1 [32] and maize Hb [40] (not shown). This observation suggests that Fe-heme in P. patens nsHb1 and 2 is hexacoordinate.

4. Conclusions

Based on available sequences, for many years it was known that land plant nshb, shbs and lbs are interrupted by 3 introns. As result from these observations it was postulated that the ancestor to land plant nshbs was interrupted by 3 introns and that the 4 exons/3 introns gene structure

---

Global J. Biochem. 2012. 3: 12
www.simplex-academic-publishers.com

© 2012 Simplex Academic Publishers. All rights reserved.
Figure 3. Nucleotide and predicted protein sequence of *P. patens* nshb2 amplified by PCR (Figure 2). Coding (exon) and non-coding (intron) sequences are shown in upper and lowercase letters, respectively. Sequences flanking the exon/intron boundaries are underlined. Distal (H84) and proximal (H119) His residues are shown with black background. The *P. patens* nshb2 gene sequence was deposited in the GenBank database under the accession number JQ692084.
conserved in land plant nshbs, shbs and lbs [4, 21, 41]. These postulates were confirmed over the last years as novel nshbs, shbs and lbs became available from newly sequenced land plant genomes. However, our search in land plant genomes revealed the existence of a nshb gene (nshb2) interrupted by 4 introns in P. patens (see subsections 3.1. and 3.2.). This observation indicated that gene structure other than 4 exons/3 introns may exist in land plants. Introns interrupting P. patens nshb2 correspond to canonical IVS-I, IVS-II and IVS-III from known land plant nshbs, shbs and lbs plus pre-IVS-I. Thus, an intriguing question is: what was the origin of intron pre-IVS-I in P. patens? One possibility is that intron pre-IVS-I already existed in an ancestral algal nshb and conserved into first land plant nshbs. Other possibility is that nshb2 inserted into the P. patens genome by horizontal gene transfer from a 4 intron-containing hb donor. This implicates that hb donor was interrupted by introns pre-IVS-I, IVS-I, IVS-II and IVS-III. However, a candidate for a 4 intron-containing hb donor is not known. We consider the above possibilities unlikely and favor the following most parsimonious hypothesis: a 3 intron (IVS-I, IVS-II and IVS-III)-containing nshb1 duplicated into the P. patens genome originating nshb2, which was subsequently interrupted by insertion of intron pre-IVS-I thus resulting in 4 intron (pre-IVS-I, IVS-I, IVS-II and IVS-III)-containing P. patens nshb2. If this hypothesis.

Figure 3. (Continued)
**Figure 4.** Sequence alignment of *P. patens* nsHb1 and 2. Conserved amino acid residues, and distal and proximal His residues are shown with asterisks and black background, respectively. Intron (IVS) position within the Mb-fold is indicated in parenthesis and with gray background. Helix regions were calculated from the structure of rice Hb1 [32].

**Figure 5.** Phenetic relationships of *P. patens* nsHb1 and 2 (arrow) and selected Archaeplastida nsHbs and Lbs. Phenogram was constructed from Hb sequences reported by Garrocho-Villegas et al. [4], Fernández et al. [24] and Vinogradov et al. [18].
is correct, it is possible that intron insertion similar to that postulated for *P. patens* *nsHb2* has occurred in other land plant *nshb*s. This hypothesis will be verified as novel *nshb*s become available from newly sequenced land plant genomes.

**Acknowledgements**

Work in the authors’ laboratory has been financed by SEP-PROMEP (grant number UAEMor-PTC-01-01/PTC23) and Consejo Nacional de Ciencia y Tecnología (CoNaCyT grant numbers 25229N and 42873Q), México. C. V.-L. is a postdoctoral fellow supported by CoNaCyT.

**References**


Cite this article as: