Neuroglobins, pivotal proteins associated with emerging neural systems and precursors of metazoan globin diversity.

Christophe Lechauve, Muriel Jager, Laurent Laguerre, Laurent Kiger, Gaëlle Correc, Cédric Leroux, Serge Vinogradov, Mirjam Czjzek, Michael C Marden, Xavier Bailly

To cite this version:

Christophe Lechauve, Muriel Jager, Laurent Laguerre, Laurent Kiger, Gaëlle Correc, et al.. Neuroglobins, pivotal proteins associated with emerging neural systems and precursors of metazoan globin diversity.. Journal of Biological Chemistry, American Society for Biochemistry and Molecular Biology, 2013, 288 (10), pp.6957-67. 10.1074/jbc.M112.407601 . hal-01020584

HAL Id: hal-01020584
https://hal.sorbonne-universite.fr/hal-01020584
Submitted on 8 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Neuroglobins: pivotal proteins associated with emerging neural systems and precursors of metazoan globin diversity

Christophe Lechauve¹, Muriel Jager², Laurent Laguerre³, Laurent Kiger⁴, Gaëlle Correc⁵, Cédric Leroux³, Serge Vinogradov⁶, Mirjam Czjzek³,⁵, Michael C Marden⁴, and Xavier Bailly³∗

¹INSERM - UMR S 968. CNRS, UMR_7210, Institut de la Vision UPMC/CHNO des Quinze-Vingts. 17, rue Moreau - 75012 Paris. France
²Université Pierre et Marie Curie (Paris 06), UMR 7138 CNRS UPMC MNHN IRD, 7 quai St Bernard, 75005 Paris, France
³UPMC-CNRS, FR2424, Station Biologique de Roscoff, 29680, Roscoff, France.
⁴INSERM U779, 94275 Le Kremlin-Bicêtre, France
⁵UPMC-CNRS, UMR 7139, Marine Plants and Biomolecules, Station Biologique de Roscoff, 29680, Roscoff, France.
⁶Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, Michigan 48201, USA

To whom correspondence should be addressed: Xavier Bailly, Station Biologique de Roscoff, 29680, Roscoff, France, Tel: + 33 2 98 29 23 29; Fax: + 33 2 98 29 23 24; E-mail: xavier.bailly@sb-roscoff.fr

Key words: Neuroglobin, structure, nervous system evolution, globin evolution, acoel, cnidarian

Background: Neuroglobins are expressed in vertebrate neurons.

Results: Neuroglobins are located in acoel and medusa neural systems (two basal animals) but also ubiquitous in metazoan transcriptomes.

Conclusion: Neuroglobin has been recruited early in neural cell prototypes and later co-opted in hemoglobin-based blood systems.

Significance: Universality of neuroglobin sheds new light on the origin and evolution of globins.
SUMMARY

Neuroglobins, previously thought to be restricted to vertebrate neurons, were detected in the brain of a photosymbiotic acoel, *Symsagittifera roscoffensis*, and in neuro-sensory cells of the jellyfish *Clytia hemispherica*. For *S. roscoffensis*, a member of a lineage that originated either at the base of the Bilateria or of the deuterostome clade, we report the ligand-binding properties, crystal structure at 2.3 Å and specific brain immuno-cytochemical pattern. Furthermore, we describe in situ hybridizations of two neuroglobins specifically expressed in differentiating nematocytes (neuro-sensory cells) and in statocytes (ciliated mechanosensory cells) of the nervous system of *C. hemisphaerica*, a member of the early-branching animal phylum Cnidaria. In silico searches using these neuroglobins as queries revealed the presence of previously unidentified neuroglobin-like sequences in most metazoan lineages. Since neural systems are ubiquitous in Metazoa (except Porifera and Placozoa), the constitutive expression of neuroglobin-like proteins in an acoel and a cnidarian, two metazoans with simple body plans, strongly supports the notion of an intimate association of neuroglobins with the evolution of animal neural systems and hints at the preservation of a vitally important function. Neuroglobins were probably recruited in the first proto-neurons in early metazoans, from globin precursors we identified in choanoflagellates, sponges or placozoans, and were strongly conserved concomitantly with nervous system evolution. Since the origin of neuroglobins predates the origins of other metazoan globins, it is likely that neuroglobin gene duplication followed by co-option and subfunctionalization led to the emergence of polyphyletic families of globins in protostomes and deuterostomes (i.e convergent evolution).

Interest in the structure, function and evolutionary relationships of circulating hemoglobins (Hbs) and intracellular myoglobins (Mbs) of animals dates back to the first 3D structural determination of these proteins in the 1960s (1) (2) (3). The large range of animal globins and the extensive occurrence of globins in prokaryotes (4) is now recognized. Prominent among the recently described metazoan globins is vertebrate neuroglobin (Ngb) (5) which is expressed in neurons of the central and peripheral nervous systems. The in vivo function of Ngb remains undefined despite a major effort over the last decade. Suggested functions include oxygen (O₂) supply in hypoxia and ischemia (6), scavenging of reactive oxygen free radicals (7), protection from apoptosis (8), redox-regulated nitrite reductase activity (9) and involvement in respiratory chain function (10). In murine models of human neuropathology, Ngb is also expressed in reactive astrocytes, a subtype of glia cells in the nervous system (11).

Recently, Blank and collaborators (12) demonstrated that the functional hexacoordinated Globin X (GbX) protein of the
cypriniform adult Zebrafish is located in nervous central system and retina, suggesting a neural-based function but contradicting a previous result obtained from the other cypriniform *Carassius auratus* GbX showing that mRNA GbX was not detected in brain and eye but in other tissues (muscle, heart, gut, liver) (13). Thought to be restricted to vertebrate, GbX-like sequences have been recently in silico identified in other deuterostomes and in protostomes supporting an early emergence of this gene family in metazoan evolution (14). However further cellular investigations must be perform for assigning a non ambiguous neural function of the GbX-like sequences so far identified in metazoans.

In protostomes, globins have been observed in the nerve tissue of certain annelids, molluscs and a nematode (15), but have not been phylogenetically linked to vertebrate Ngbs or other deuterostome globins. Their O₂ binding affinities resemble those of vertebrate Mb and their function is considered to be O₂ storage and thus protection against hypoxia (16), (17).

Recent phylogenomic analyses of vertebrate globins have demonstrated that they can be separated into two groups, one derived from vertebrate-specific duplications (Cytoglobins, Globin E, Globin Y, the Hb chains and Mb), and another resulting from duplications preceding the emergence of chordates (Ngb, HbX) (18), (19), (20). The most recent molecular phylogenetic analysis of globin sequences from the five major groups comprising the deuterostomes, i.e. cephalochordates, echinoderms, hemichordates, urochordates and vertebrates, suggests that all deuterostome globins occur in four clades (21). Despite the fact that a molecular analysis of metazoan globins (including echinoderm and cnidarian globins) suggested an ancestral connection to the nervous system (22), Ngbs have not been reported in deep branching metazoan lineages, and evolutionary patterns of emergence of metazoan globin lineages are still unresolved.

We have employed the discovery of the hitherto unknown Ngbs in an acelo and a cnidarian, that exhibit simple morphological organizations characteristic of ancestral Bilateria / bilaterian Deuterostomes and Radiata, to further clarify the origin of globins in metazoan lineages.

*Symsagittifera roscoffensis* is a photosymbiotic acelo (Fig. 1A), thus occupying a phylogenetic position either preceding the deuterostome-protostome split or branching at the base of deuterostomes (23) (24). This hermaphrodite marine flatworm has a simple body plan with a digestive syncytium (no epithelially-lined gut), a ventral mouth, a muscle system, a nervous system with a simple central brain, but no excretory or blood circulatory systems (25).

We report the discovery of Ngb-like sequences in EST libraries from *S. roscoffensis*, the cloning and purification of a Ngb, its immunocytotoxic localization within neural cells, its ligand binding properties and crystal structure. We examined the sites of expression of putative Ngbs in the jellyfish *Clytia hemispherica* (Cnidaria, Anthozoa), which, like...
the “higher” animals (the Bilateria), exhibits a complex body organization, including striated musculature, reproductive organs and a specialized nervous system (26). In situ hybridization experiments using two specific Ngb-like probes highlight differentiating neural cell type called nematocytes or stinging cells (mechanoreceptors, i.e. neuro-sensory cells) and statocystes (gravito-sensors) in the jellyfish Clytia hemispherica.

A broad in silico transcriptome survey revealed expressed Ngb-like proteins in most of the metazoan phyla, ranging from animals with no symmetrical body plan (sponges, placozoans) to complex bilaterians, through the symmetrically radial cnidarians. Based on Ngb conservation throughout metazoans and recent biomedical studies underlying the irreversible detrimental effects of Ngb dysfunction in neurons, we assume that Ngb played a crucial role early in the subsequent evolution of metazoan nervous systems and brains in metazoan exhibiting more complex body-plan. Indeed, Ngb appears as a key and central partner in neurone physiology as a neuroprotectant preventing the oxidative damages and neurodegenerescence as illustrated in Alzheimer’s disease transgenic mice models (27).

Our data and results suggest that an ancestral globin-like gene was recruited in emerging proto-neural cells and system in the first diploblastic animals (ancestors of extant cnidarians such as sea anemones, corals or medusa) and specifically evolved as a neural globin. A natural corollary is a novel scenario for metazoan globin evolution, namely, the independent emergence of globins such as extracellular annelid Hbs, molluse and arthropods Hbs, and vertebrate Hbs, via functional shifts from Ngb copies early during metazoan radiation and concomitantly with increasing body plan complexity and the development of blood circulatory systems.

**EXPERIMENTAL PROCEDURES**

Expression, purification and characterization of S. roscoffensis Ngb - The coding sequence of S. roscoffensis Ngb i.e. (SrNgb1), (ID number European Nucleotide Archive HE972520) was amplified by PCR and subsequently cloned into a pET-3a cloning vector (Invitrogen). The construct was transformed into E. coli BL21DE3 for protein expression in auto-inducible medium (28). The protein was purified with an Akta purifier system (GE Healthcare), due to the low pI of the S. roscoffensis globin the samples were loaded on a 5 ml HiTrap DEAE FF column (GE Healthcare) equilibrated with Tris HCl 50mM pH 8.5, and eluted at a concentration of 25mM NaCl. The obtained samples were loaded on a desalting Sephadex G-25 column (GE Healthcare) suspended in PBS, pH 7.4 and the material was finally purified on a Superose 12 HR 16/50 (Amersham Biosciences) column equilibrated with PBS, pH 7.4. Finally, ferric and ferrous spectra UV/visible spectra (O₂ and CO) were measured with a Cary model 400 spectrophotometer.
Autoxidation kinetics and ligand rebinding of SrNgb1 - Full spectra were measured versus time on HP 8453 diode-array spectrophotometer. The sample was first thoroughly deoxygenated in a sealed optical cuvette under a stream of N₂. Then a slight excess of sodium dithionite was added to reduce the globin heme moiety. Finally the cuvette was equilibrated under air to obtain the oxy reduced species and to allow the depletion of the residual unreacted dithionite. Ligand recombination kinetics were measured at a single wavelength after photodissociation by 10 ns pulses at 532 nm, as previously described (29). Samples in sealed cuvettes were equilibrated under various fractions of CO or oxygen. A mixed atmosphere of both CO and O₂ was used to study the oxygen to CO replacement reaction after photolysis of CO.

Immuno-cyto-localization with SrNgb1 and RF-amide antibodies - Acoel flatworms were collected in Roscoff (Brittany, France) and anesthetized with 7% MgCl₂ and fixed during 45 min in 4% PFA at 4°C. Animals were then washed with phosphate buffer pH 7.4, permeabilized with 0.1% Triton X-100 in PBS 3 times for 15 min at room temperature. They were then incubated with 5% BSA, 0.1% Triton and 0.05% Tween 20 in PBS for 2-3 hours at room temperature and then incubated overnight at 4°C alternatively with 1/700 polyclonal S. roscoffensis anti-Ngb, produced against whole recombinant protein by Eurogentec (Speedy 28-day polyclonal packages) or with anti-RF-amide (courtesy of Thomas Leitz, Kaiserslautern). The next day, acoels were washed three times for 15 min in PBS and incubated with the appropriate secondary antibodies. They were then incubated for 10 min in a DAPI solution (2 µg/ml in PBS), washed 3 times in PBS and mounted on a glass slide for microscope observation. Image acquisition of fluorescence labeling was monitored with a confocal microscope (Leica sp5) equipped with a 20x objective and using Leica LAS-AF software.

Animal collection and in situ hybridization - Medusae were obtained in Paris by culture of C. hemisphaerica colonies in artificial seawater (Reef Crystals®) established from polyps provide by Evelyn Houliston (Villefranche-sur-Mer) as previously described (30). Medusae were left unfed for one day before fixation. They were fixed for 40 min at 4°C in 3.7% PFA, 0.2% glutaraldehyde, PBT 1X (10 mM Na2HPO₄, 150 mMNaCl, pH7.5, 0.1% Tween 20). DIG-labelled antisense RNA probes synthesis and whole-mount in situ hybridizations were carried out as previously described (31) The only modification to the in situ protocol was an acetic anhydride treatment before hybridization. Alkaline phosphatase activity was revealed using NBT/BCIP (NitroBlueTetrazolium/Bromochlorylindolophosphate, blue staining) or fast red TR-naphthol reagent ® (Sigma, red staining). After postfixation and DAPI staining (32), samples were mounted in Citifluor®. Double in situ hybridizations were performed as described in (33). DIC images were obtained with an
Olympus BX61 microscope using Q-imaging Camera with Image Pro plus® software (Media Cybernetics).

Protein crystallization - All crystallization experiments were carried out at 292 K. Initial crystallization trials were performed with the PACT, JCSG+, PEG I and PEG II Suites (Qiagen) that is a total of 384 conditions in four 96-well plates from Corning. The trials were set up using a Cartesian crystallization robot, and the sitting drops were made by mixing 300 nl of protein (13 mg/mL in 30 mM PBS buffer pH 7.5, 100 mM NaCl) with 150 nl of reservoir solution. A single hit was identified in the PEG II screen, containing 1M LiCl, 0.1 M Na acetate and 30% (w/v) PEG 6000. Subsequently, this crystallization condition was optimized in 24-well Linbro plates by the hanging-drop vapour-diffusion method, screening ranges from 0.6 to 1.0 M LiCl and 30% to 39% PEG 6000. These drops were prepared on siliconized cover slips by mixing 2 ml of protein with 1 ml of well solution. The drops were equilibrated against reservoir solutions of 0.75 ml volume. Best crystals were obtained for 32% PEG 6000, 1.0 M LiCl and 0.1 M Na acetate. For cryo-protection, 5% glycerol was added to the crystal drop solution before flash-freezing the crystals in the gaseous N2-stream at 100 K.

Data collection and X-ray diffraction analysis - X-ray diffraction data were first collected from globin crystals at 100 K on beamline ID23-I at the ESRF (Grenoble, France) using an ADSC Quantum 4R CCD detector. All crystals were flash-cooled in a liquid nitrogen stream. The crystals were rotated through 120° with a 0.5° oscillation range per frame at a wavelength of 0.933 Å. All raw data were processed using the program XDS and the resultant data were merged and scaled using the program XSCALE (34). Models for structure solution by molecular replacement were selected by a sequence search using BLAST against the PDB sequence database. However, all attempts, to solve the structure of this globin by molecular replacement performed with the program AMORE (35), using various neuroglobin or myoglobin models stayed unsuccessful. A second data set was therefore collected at the Fe absorption edge at a wavelength of 1.7387 Å on beamline BM30A, covering an angular section of 90° with an oscillation range of 1.0°. The data treatment was performed with XDS in the same way as for the native data set. All further data collection statistics are given in Table S1.

Crystal structure determination and refinement - The iron atom substructure solution was calculated with SHELXD (36) followed by phasing and density modification performed with SHELXE, using the graphical interface HKL2MAP (37) and the resulting electron density map was displayed with Coot (38). Both possible enantiomorph space groups were tried and the phasing procedure allowed a selection of a clear and contrasted structure solution in P6222. These starting phases were used to build the initial model using ARP/wARP and REFMAC as part of the CCP4
suite (39), and switching to the higher resolution data at 2.3 Å. Roughly 70% of the helices were constructed by the automatic procedure. The subsequent manual adjustment and model building was carried out with Coot and alternated with refinement cycles using REFMAC. Water molecules were added automatically with the REFMAC-ARP/wARP option and visually verified, one by one, using Coot. The final model contained residues ranging from 6 to 154, the prosthetic heme group, 98 water molecules and an oxygen ligand bound to the iron atom. The asymmetric unit contains one globin molecule leading to a Matthews coefficient of 4.9 and a solvent content of 74.9%. The phasing and final refinement statistics are given in Table S1.

(Synsagittifera roscoffensis neuroglobin PDB accession number: 4B4Y)

**Phylogenomics and molecular phylogeny** - The identification of Ngb-like / putative neural globin sequences was performed using *S. roscoffensis* Ngb1 and vertebrate Ngb sequences queries in blastp searches of the non-redundant nucleotide database maintained by NCBI (http://www.ncbi.nlm.nih.gov/nucest/) and of non-annotated ESTs databases from various metazoans, deposited and archived at (ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/).

A multiple alignment of a representative subset of Ngb-like sequences has been automatically generated with HMMER v3.0 package (40) using the hmmlalign program and the Globin (PF0042) raw HMM as a guide. Molecular phylogenetic analysis was carried out using the Maximum likelihood approach with PhyML software (41) with LG option as model of amino acid substitution, NNI moves option for the tree topology search operation and SH-like support option for the default branch support. The tree topology (Newick format) was edited with MEGA5.1 (42).

In addition to be accessible in Mat&Met / Results paragraphs and supplementary data, the molecular phylogeny analysis has been performed and deposited using respectively the BioSide software and his dedicated website at http://www.bioside.org. Indeed, in order to be easily traceable and reproducible by anyone that would like to replay the molecular phylogeny procedure, a file including the original multiple alignment of sequences and PhyML setups are available and detailed following the permalink http://www.bioside.org/workflow/BS1211140001 (id number is BS121114001) or at http://www.bioside.org/community. Prediction of N-terminal myristoylation of Ngb-like sequences was performed with the program The MYR Predictor, a web-service available at http://mendel.imp.ac.at/myristate/SUPLpredictor.htm. This program calculates whether a protein is predicted as myristoylated with reliable/twilight zone confidence or not.

**RESULTS and DISCUSSION**

*The Ngb-like protein 1 of S roscoffensis is a functional neuroglobin* - SrNgb1 is expressed in the brain and nervous system of *S. roscoffensis* (Fig. 1B; 1C). The acoel brain is formed by a
layer of neuronal cell bodies surrounding a central neuropile, embedding the statocyst, a gravity sensor (25). The SrNgb1 signal mainly occurs in the anterior tip (“head”) where photoreceptors and frontal sensory organs collect environmental information. The signal surrounds the statocyst and the photoreceptors and is superimposable with the anti-RFamide antibody pattern (Fig. 1B) and the serotonergic nervous system (43). Constitutive expression of SrNgb1 during embryogenesis and in juvenile and adult stages indicates its implication throughout nervous system development and in maintenance of brain activity.

The spectroscopic properties of purified SrNgb1 (UV and visible absorption spectra of the ferrous and ferric forms) indicate that in the absence of external ligands it is pentacoordinated, in contrast to vertebrate Ngbs in which a sixth coordination bond is formed with a distal histidine (Fig. 2A). The rate constants of O2 and CO binding and of O2 dissociation are similar to those of vertebrate Mbs, and consequently so is its O2 binding affinity (Table 1). The rate of heme autoxidation under pure O2 at 25 °C is slow (first order rate 0.053 h⁻¹; Fig. 2B), which is not surprising in view of the fact that there is a well established inverse relationship between O2 affinity and autoxidation rate for pentacoordinated globins. This reaction is much slower than those observed for vertebrate Ngbs, probably due to a higher capacity of the hexacoordinated form for transferring an electron to molecular O2 (44). Overall, these observations are consistent with an in vivo function involving reversible binding of the diatomic ligand rather than a redox reaction with O2 as a terminal electron acceptor.

The structure of S. roscoffensis neuroglobin - The structural model consists of 149 residues (including Ala6 to Glu154) that bind a heme b prosthetic group, with a bond between the heme iron and the proximal histidine (H103), the distal ligand being an O2 molecule (Fig. 3A). The tertiary structure corresponds to the classical globin fold, consisting of eight helices (A to H, Fig. 3A), the heme binding cleft formed by helices E and F. Despite being deoxy-pentacoordinated, SrNgb1 shares certain structural features with vertebrate Ngbs that are quite different from classical Hb and Mb structures. Although the identity of SrNgb1 with mouse Ngb is only 19% (Fig. 3B), all of the conserved globin-fold residues (45) are present, including the heme ligand residues E7His and F8His. The C and D helix regions most closely resemble those described in murine Ngb (46). The Trp residue at position 52 in SrNgb1 (Fig. 3A) may present a barrier to ligand exit and entry by forming a stable hydrogen bond to one of the heme propionates (distance 2.8Å; Fig. 3A). This interaction is reinforced by a water molecule located near by (heme-propionate-O2D/HOH, distance 3.8Å; HOH117), which is further hydrogen bonded to the distal histidine (ND1, 2.7Å) and the second propionate group of the heme (HOH/heme-O2D, 3.0Å). In murine Ngb, residues Lys67 and Tyr44 form a similar hydrogen bonding network involving a water molecule also binding to the distal His(46). Structural equivalence is provided by superimposition of
HOH117 with its murine counterpart, and by superimposition of the Tyr44 OH-group in murine Ngb with the Trp52 NH-group in SrNgb1. Moreover, Tyr44 in murine Ngb and Trp52 in SrNgb1 are at equivalent positions in the sequence alignment (Fig. 3B). SrNgb1 also shares with murine Ngb the high flexibility of the connection between helices E and F (data not shown). SrNgb1 displays a unique feature in that helix F is bent by the presence of a proline (Pro94) (Fig. 3C). This could be analogous to the transition of human Ngb structures that is triggered by a disulfide bond in the CD region (47). The closest match to SrNgb in the PDB database was ferrous CO-bound murine Ngb (1W92). Overall, the SrNgb1 structural sequence matches Ngbs and plant Hbs, with a slightly better Z-score (48) than to Mbs (data not shown).

In the cnidarian medusa Clytia hemisphaerica, two globins (CheNgb1 and CheNgb2) are expressed in differentiating neuro-sensory cells (nematocytes) - Nematocytes exhibit many characteristics of neuro-sensory cells, including mechano-sensitive cilia, neurite-like outgrowths and synapses. They contain a single-use dart specialized for killing prey. Nematogenesis (the generation of nematocytes) in Cnidaria is used as a model for non-bilaterian neurogenesis (26), (49), as these neural cells are continuously generated throughout larval and adult life.

The CheNgb1 and CheNgb2 genes are mainly expressed in the nematogenic ectoderm of tentacle bulbs and manubrium (compare Fig. 4A, 4B and 4C). In the tentacle bulbs, their expression patterns are crescent-shaped and interrupted on the external side of the bulb (Fig. 4D-4F, blue staining), thus exactly matching the expression of minicollagen 3-4a (mcol3-4a, Fig. 4H, red staining). The latter belongs to a family of small collagen-like proteins known in hydrozoans to be a major component of the nematocyst wall (33). Double in situ hybridizations revealed extensive co-expression of mcol3-4a with both CheNgb1 (purple color in Fig. 4E) and CheNgb2 (purple color in Fig. 4G), indicating that both genes are expressed in differentiating nematoblasts over a large time window.

CheNgb2 mRNA was also detected in the statocysts (Fig. 4F arrowhead, 4F’ and 4F’’), the equilibration organs arranged regularly around the rim of the bell of the animal. CheNgb2-expressing cells are located in the basal epithelium of the statocyst, near the bell margin and interpreted as ciliated mechano-sensory cells (figure 4F’ and 4F’’).

CheNgb1 and CheNgb2 transcripts were also abundant in the proximal part of the manubrium ectoderm and mimicked the expression pattern of minicollagen, with which they are co-expressed as demonstrated by double in situ hybridization. CheNgb1 and CheNgb2 were also localized in the female gonad in an unidentified cell type (not germ line cells) (fig 4A and 4B).

Neuroglobins are ubiquitously expressed in Metazoa - Using SrNgb1 as an in silico probe
for blasting genomic resources, we identified 50 or so transcripts never described so far from different phyla (Table S2A-B) mostly related to other neuroglobins / neuroglobin-like sequences according to classical blastp searches against the NCBI non-redundant nucleotide database. After a cross verification systematically conducted with the Panther predictive tool, all the new globins fall into the Panther Leghemoglobin-related family that encompass 14 subfamilies including Neuroglobin, Globin X, Non-Symbiotic Hemoglobin and Leghemoglobin. None of the new sequences we found are related to the Panther Hemoglobin-family that encompasses vertebrate Hemoglobin, Cytoglobin or Myoglobin.

The taxonomic distribution of the neuroglobin-related sequences suggests broad conservation throughout metazoan evolution (Fig. 5A, Table S2A-B). They were detected in non-symmetrical body plan basal metazoans with neither nervous system nor circulatory system, i.e. in the metazoan lineages Porifera (the sponges Amphimedon queenslandica and Carterospongia foliascens) and Placozoa (Trichoplax adherens). In the radially symmetrical cnidarians which have a simple nervous system but no circulatory blood system, Ngb-like sequences were present in Anthozoa (the coral Montastraea faveolata and the sea anemones Anemona viridis and Nematostella vectensis) and Hydrozoa (Clytia hemisphaerica and Hydra magnipapillata). No other types of globin (neither homologs of circulating Hbs nor Mb-like globins) were detected in these basal metazoans. In protostomes, expressed Ngb-like sequences were found in (1) cephalopod mollusks such as the cuttlefish Sepia officinalis and Euprimna scolopes and the squid Dorytuthis paeleii, (2) many arthropods such as the hymenopteran Apis mellifera (bee), the crustacean Carcinus maenas (green shore crab) and Daphnia pulex (a common species of water flea) or the social insects Harpegnathos saltator (ant), (3) the sipunculid Themiste sp. (the peanut worm), (4) the brachiopod Terebratalia transversa (the common lampshell), (5) various annelids such as the polychaetes Alvinella pompejana (Pompeii worm from deep-sea hydrothermal vents) or the hirudinea Helobdella robusta (leech). Expressed Ngb-like sequences were also identified in so called “minor phyla” such as platyhelminthes, tardigrads, kinorhynchs, and nemertodermatids (a sister group of acoels) (Table S2A-B). In deuterostomes, Ngb-like sequences were identified in all phyla preceding the emergence of vertebrates: in the echinoderms Strongilocentrotus purpuratus and Paracentrotus lividus (sea urchins), the hemichordates Saccoglossus kowalevskii (acorn worm) and Balanoglossus clavigerus; the cephalochordate Branchiostoma lanceolatum (amphioxus, also known as the lancelets), and the urochordates Molgula tectiformis and Botrylus schlosseri (tunicates).

Vertebrate species have a single Ngb gene copy while many of the other metazoans have several copies, indicating gene duplication events correlated with subfunctionalization. The existence of a second S. roscoffensis or cnidarian Ngb sequence (Table S2A-B)
illustrates classical cases of diversification by a gene duplication event. The molecular unrooted phylogenetic tree (Fig. 5B) clearly shows that Vertebrate Hbs, Mbs, and Cybs clearly form a distinct monophyletic group (Fig. 5B), in agreement with earlier results (50) (21). Vertebrate Ngbs and GbXs are included into a group of functional neuroglobins and neuroglobin-related sequences that harbors the neuroglobins characterized in this study i.e. the Ngb duplicates of *S. roscoffensis* and of *C. haemispherica*. The presence of Vertebrate GbXs sequences in this group supports a likely connection of these proteins with neural systems. The cluster that contains the choanoflagellates leghemoglobin-related sequences (the closest living unicellular relative to metazoan (38)), the poriferan and the vertebrate Ngbs sequences likely represents the ancestral Ngb lineage with plesiomorphic characteristics. Indeed we noticed in Blast results that choanoflagellates and poriferans, cnidarians and *S. roscoffensis* neuroglobins produced significant alignments with protists, especially with the unicellular green algae *Micromonas* and the diatom (unicellular brown algae) *Thalassiosira* globin that both exhibit Leghemoglobin-related signature according Panther prediction system. These findings are in agreement that metazoan globins were likely inherited from a unicellular eukaryote globin.

The second cluster with SrNgb2, CheNgb1 and Vertebrates GbX represents another cluster of neuroglobin-related sequences. The other sequences diagnosed as putative neuroglobin-related proteins (with a leghemoglobin-related signature) that do not cluster specifically within the Ngb group (includingreflect primary sequence divergence and likely species-specific functional diversification. Further exploratory approaches such as gene or protein expression localization will be required for formally establishing the involvement of these proteins (including the so-called GbX) in the nervous system.

We also noticed, when the coding sequences we recovered were complete, that some neuroglobin-related sequences exhibited a meristoylation site and some not, with no clear pattern in the phylogenetic tree (Fig. 5B). Vertebrates Hbs, Mbs, and Cybs clearly form a distinct monophyletic group (Fig. 5B).

It is clear that our molecular phylogeny of Ngb-like sequences is inevitably based on a heteroclite subset of paralogous and orthologous Ngb-like sequences, but as transcriptomes do not reveal 100% of transcripts and especially cryptically expressed genes (those with a low number of corresponding transcripts), the number of Ngb-like proteins is likely to be significantly underestimated. In other words, more sequences with more functional data from more taxa will refine the phylogenetic relationship among Ngb-related sequences.

*Neuroglobin is likely an early constitutive actor in nervous systems and brain evolution* - It is clear that Ngb-like proteins are ubiquitous in metazoans (Fig. 5A). The emergence of neural structures in metazoans represented an innovation resulting in functions such as
interneuronal and neuro-muscular transmission, allowing feeding, reproduction, vision and complex behaviours like predation (51). Although the origin of nerve cells remains unknown, the Cnidaria, whose name derived from cnidocytes (i.e. nematocytes), occupy a key position in Metazoan with respect to early nervous system evolution (52). Together with the ctenophores, the Cnidaria form the Coelenterata, the sister group of eumetazoans (Bilateria) (53). It is assumed that transduction of chemical and mechanical stimuli in nematocytes are hallmarks of primitive nerve cells and that nematocytes are thus representative of ancestral sensory cells that preceded the differentiation of neuronal cell types in animal evolution (54). The unequivocal expression of Ngbs in nematocytes of the jellyfish Clytia hemispherica appears to be a robust indication of the essential role of these proteins in early evolution of the nervous system. The fact that acoel and jellyfish statocysts (the sensory organs measuring pressure) are respectively and specifically targeted by Ngb antibody and Ngb probes illustrates the intimate connection of Ngb with nerve nets and transmission of information. We assume that an original exaptation, i.e. the recruitment of a globin by proto-nervous cells and proto-nervous circuitry, laid the foundations for elaborated nervous systems and brains in the first metazoans displaying anatomical polarity (radial then bilateral symmetry) and differentiated nervous systems. Neuroglobin precursors are likely homologous to those identified in unicellular eukaryotes (choanoflagellates) and simple metazoans (sponges and placozoans) devoid of neural cells, but possessing the basic genetic toolkit encoding proteins homologous to those involved in nervous system development in higher animals (55,56) (Fig. 5A).

The deleterious effects on nerve cells of Ngb silencing (57) (10) and the conservation of this protein throughout metazoan evolution underline the pivotal function of Ngbs in development and physiology of neurons. Subcellular expression of Ngb in mitochondria of neuronal cells in regions of the brain with high metabolic activity (58) (10) is an indicator of the implication of Ngb in cellular homeostasis in extant organisms and, by extension, in early emerging metazoan neuronal cells. The Ngb-like sequences of certain cnidarians, protostomes and deuterostomes exhibit a predicted N-terminal myristoylation site indicating a possible interaction with membranes, putatively including those of the mitochondria (Fig.5B). The presence of such a site has already been described for the globin expressed in the gills of the crab Carcinus maenas mentioned above, a Ngb-like protein (Leghemoglobin-related family) according to Panther prediction (59).

In the core of the globin-fold, hexacoordination of the heme iron atom leads to a high autoxidation rate, suggesting that hexacoordinated vertebrate Ngbs are involved in redox metabolism connected to oxidative phosphorylation either with electron carriers or with reactive oxygen species produced by the mitochondria (60). Our results show that some Ngbs, such as SrNgb1, can be functionally
pentacoordinated. SrNgb1, whose O₂ binding affinity is similar to that of Mb, is likely to be involved in O₂ storage, and thus provision of O₂ during periods of hypoxia. This proposal is in agreement with the most likely roles of nerve Hbs in the annelid (Aphrodite aculeata), the clams (Spisula solidissima and Tellina alternata) and the nemertean (Cerebratulus lacteus), which have been established to be the provision of O₂ to the metabolically highly active neural cells and thus protection under hypoxic conditions (15), (61), (62), (16).

It remains to be determined which form of coordination (penta- or hexa-) of metazoan Ngbs was associated with neofunctionalization and which was the ancestral state. It is pertinent to note that human Ngb exists as an equilibrium between the two forms, with the hexacoordinated form being dominant (~99:1) (9).

Neuroglobins could also be precursors of the metazoan globin repertoire - The results of our survey highlight the presence of putative Ngbs proteins in radial and bilateral animals irrespective of the presence or absence of a blood circulatory system and of the respiratory protein employed (hemocyanin in mollusks and arthropods, hemerythrin in sipunculids and brachiopods, hemoglobin in other metazoans). The presence of Ngb in ice fish, where circulating Hb has disappeared from the blood circulatory system, is not paradoxical as claimed by Cheng (63), but illustrates the separate evolutionary pathways of Ngbs and O₂ binding Hbs, the mandatory constitutive expression of Ngb in the nervous system, and a clear case of disadaptation, i.e. loss of the circulating oxygen carrier.

Assuming that the ancestral bilaterian body plan was very simple with a nervous system but no blood circulatory system, it is obvious that the presence of Ngb predates the emergence of circulatory Hb. Given that Ngbs are ancestral and constitutively expressed in all metazoans (Fig. 5A), the sporadic presence of O₂ binding Hb in individual metazoan lineages strongly suggests that they are polyphyletic. In other words, the emergence of circulating Hb in metazoans is likely due to convergent evolution. The globin lineages other than Ngb found in many metazoan groups have probably emerged as the result of functionalization (64) and cooption of a Ngb-like globin in early metazoans. Indeed, most of the metazoan transcriptomes checked in this study exhibit multiple Ngb-like paralogs, likely originating from gene duplication events.

CONCLUSION

We demonstrate the presence of a functional Ngb in neural cells of the acoel S. roscoffensis and expression of homologous Ngbs specific to neuro-sensory cells (differentiating nematoblasts) in the cnidarian jellyfish Clytia hemispherica. These results suggest that the first globins expressed in early bilaterians and symmetrically radial cnidarians were specifically linked to the metazoan nervous system. The pentacoordination of SrNgb1 vis a
vis the hexacoordination of the vertebrate Ngbs may be due to differences in function, with the acoel Ngb playing an O₂ storage role providing neuroprotection during hypoxic periods. This interpretation is supported by reports of the functions of “nerve globin” in several protostomes.

Extensive in silico mining of genomic data using SrNgb1 as a probe revealed the occurrence of expressed Ngb-like sequences in most metazoan phyla, including sponges and Placozoa, basal metazoans lacking neural and circulatory systems. Our results clearly demonstrate that the emergence of Ngb in metazoans chronologically preceded the emergence of other globin families. Consequently, we propose a novel scenario for metazoan globin evolution, based on two broad and complementary statements. On the one hand, our experimental and in silico results suggest that an ancestral globin-like gene was recruited in the emerging proto-neural system in the ancestor of Bilateria and diploblastic animals (ancestors of extant cnidarians such as sea anemones, corals or medusa) to become a functional Ngb. On the other hand, metazoan globins other than Ngbs, such as annelid, mollusc, arthropod, and vertebrate Hbs, likely originated independently from early Ngbs, via co-option of duplicated Ngb genes and functionalization during metazoan radiation, concomitant with increasing body plan complexity and the emergence of blood circulatory systems.

Access to multiple ontogenetic stages of emerging marine models, for which genomic resources and molecular tools are increasingly available (65), will be of a prime importance for functional genomic exploration using Ngbs as key developmental markers in animal lineages exhibiting complex nervous tissues (cephalopods), elaborated social behavior (ants), or subject to anthropogenically-induced stresses or diseases (corals, mussels, oysters).
REFERENCES

1. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) *Nature* 181, 662-666
2. Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. (1960) *Nature* 185, 416-422
3. Zuckerkandl, E., and Pauling, L. (1965) *J Theor Biol* 8, 357-366
4. Vinogradov, S. N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Guertin, M., Gough, J., Dewilde, S., Moens, L., and Vanfleteren, J. R. (2005) *Proceedings of the National Academy of Sciences of the United States of America* 102, 11385-11389
5. Burmester, T., Weich, B., Reinhardt, S., and Hankeln, T. (2000) *Nature* 407, 520-523
6. Burmester, T., Gerlach, F., and Hankeln, T. (2007) *Adv Exp Med Biol* 618, 169-180
7. Li, W., Wu, Y., Ren, C., Lu, Y., Gao, Y., Zheng, X., and Zhang, C. (2010) *Proteins* 79, 115-125
8. Brittain, T., Skommer, J., Henty, K., Birch, N., and Raychaudhuri, S. (2010) *IUBMB Life* 62, 878-885
9. Tiso, M., Tejero, J., Basu, S., Azarov, I., Wang, X., Simplaceanu, V., Frizzell, S., Jayaraman, T., Geary, L., Shapiro, C., Ho, C., Shiva, S., Kim-Shapiro, D. B., and Gladwin, M. T. (2011) *J Biol Chem* 286, 18277-18289
10. Lechauve, C., Augustin, S., Cwerman-Thibault, H., Bouaita, A., Forster, V., Celier, C., Rustin, P., Marden, M. C., Sahel, J. A., and Corral-Debrinski, M. (2012) *Biochim Biophys Acta* 1823, 2261-2273
11. DellaValle, B., Hempel, C., Kurtzhal, J. A., and Penkowa, M. (2010) *Glia* 58, 1220-1227
12. Blank, M., Wollberg, J., Gerlach, F., Reimann, K., Roesner, A., Hankeln, T., Fago, A., Weber, R. E., and Burmester, T. (2011) *PloS one* 6, e25292
13. Roesner, A., Fuchs, C., Hankeln, T., and Burmester, T. (2005) *Mol Biol Evol* 22, 12-20
14. Droge, J., and Makalowski, W. (2011) *Biol Direct* 6, 54
15. Wittenberg, B. A., Briehl, R. W., and Wittenberg, J. B. (1965) *Biochem J* 96, 363-371
16. Geuens, E., Dewilde, S., Hoogewijs, D., Pesce, A., Nienhaus, K., Nienhaus, G. U., Olson, J., Vanfleteren, J., Bolognesi, M., and Moens, L. (2004) *IUBMB Life* 56, 653-656
17. Hundahl, C., Fago, A., Dewilde, S., Moens, L., Hankeln, T., Burmester, T., and Weber, R. E. (2006) *Febs J* 273, 1323-1329
18. Hoffmann, F. G., Opazo, J. C., and Storz, J. F. (2010) *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14274-14279

19. Storz, J. F., Opazo, J. C., and Hoffmann, F. G. (2011) *IUBMB Life* **63**, 313-322

20. Hoffmann, F. G., Opazo, J. C., and Storz, J. F. (2012) *Mol Biol Evol* **29**, 303-312

21. Hoffmann, F. G., Opazo, J. C., Hoogewijs, D., Hankeln, T., Ebner, B., Vinogradov, S. N., Bailly, X., and Storz, J. F. (2012) *Mol Biol Evol* **29**, 303-312

22. Bailly, X., and Vinogradov, S. N. (2008) *The bilaterian sea urchin and the radial starlet sea anemone globins share strong homologies with vertebrate neuroglobins.* Springer Verlag, New York

23. Bailly, X. (2009) *Cahiers de Biologie Marine* **50**, 106-107

24. Lowe, C. J., and Pani, A. M. (2011) *Curr Biol* **21**, R151-153

25. Bery, A., Cardona, A., Martinez, P., and Hartenstein, V. (2010) *Dev Genes Evol* **220**, 61-76

26. Houliston, E., Tsuyoshi Momose, T., and Manuel, M. (2010) *Trends in Genetics* **26**, 159-167

27. Khan, A. A., Mao, X. O., Banwait, S., Jin, K., and Greenberg, D. A. (2007) *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19114-19119

28. Studier, F. W. (2005) *Protein Expr Purif* **41**, 207-234

29. Uzan, J., Dewilde, S., Burmester, T., Hankeln, T., Moens, L., Hamdane, D., Marden, M. C., and Kiger, L. (2004) *Biophys J* **87**, 1196-1204

30. Chevalier, S., Martin, A., Leclere, L., Amiel, A., and Houliston, E. (2006) *Dev Genes Evol* **216**, 709-720

31. Jager, M., Queinnec, E., Le Guyader, H., and Manuel, M. (2011) *Evodevo* **2**, 12

32. Chiori, R., Jager, M., Denker, E., Wincker, P., Da Silva, C., Le Guyader, H., Manuel, M., and Queinnec, E. (2009) *PloS one* **4**, e4231

33. Denker, E., Manuel, M., Leclere, L., Le Guyader, H., and Rabet, N. (2008) *Developmental biology* **315**, 99-113

34. Kabsch, W. (1988) *Journal of Applied Crystallography* **21**, 916-924

35. Navaza, J. (2001) *Acta Crystallogr D Biol Crystallogr* **57**, 1367-1372

36. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1772-1779

37. Pape, T., and Schneider, T. R. (2004) *Journal of Applied Crystallography* **37**, 843-844

38. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A., and Noble, M. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2288-2294
39. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 1131-1137

40. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) *Nucleic Acids Res* **32**, D138-141

41. Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) *Syst Biol* **59**, 307-321

42. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) *Mol Biol Evol* **28**, 2731-2739

43. Semmler, H., Chiodin, M., Bailly, X., Martinez, P., and Wanninger, A. (2010) *Dev Growth Differ* **52**, 701-713

44. Brunori, M., and Vallone, B. (2007) *Cell Mol Life Sci* **64**, 1259-1268

45. Bashford, D., Chothia, C., and Lesk, A. M. (1987) *J Mol Biol* **196**, 199-216

46. Vallone, B., Nienhaus, K., Brunori, M., and Nienhaus, G. U. (2004) *Proteins* **56**, 85-92

47. Hamdane, D., Kiger, L., Dewilde, S., Green, B. N., Pesce, A., Uzan, J., Burmester, T., Hankeln, T., Bolognesi, M., Moens, L., and Marden, M. C. (2003) *J Biol Chem* **278**, 51713-51721

48. Holm, L., and Rosenstrom, P. (2010) *Nucleic Acids Res* **38**, W545-549

49. Galliot, B., Quiquand, M., Ghila, L., de Rosa, R., Miljkovic-Licina, M., and Chera, S. (2009) *Developmental biology* **332**, 2-24

50. Burmester, T., and Hankeln, T. (2009) *J Exp Biol* **212**, 1423-1428

51. Marlow, H. Q., Srivastava, M., Matus, D. Q., Rokhsar, D., and Martindale, M. Q. (2009) *Dev Neurobiol* **69**, 235-254

52. Gaillot, B. (2010) *A Key Innovation in Animal Evolution, the Emergence of Neurogenesis: Cellular and Molecular Cues from Cnidarian Nervous Systems*, Science Publishers & CRC Press, New York

53. Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Queinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D. J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Worheide, G., and Manuel, M. (2009) *Curr Biol* **19**, 706-712

54. Miljkovic-Licina, M., Dominique Gauchat, D., and Galliot, B. (2004) *BioSystems* **76**, 75–87

55. Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E., Mitros, T., Richards, G. S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N. H., Stanke, M., Adamska, M., Darling, A., Degnan, S. M., Oakley, T. H., Plachetzk, D. C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S. F., Goodstein, D. M., Harris, C., Jackson, D. J., Leys, S. P., Shu, S., Woodcroft, B. J., Vervoort, M., Kosik, K. S., Manning, G., Degnan, B. M., and Rokhsar, D. S. (2010) *Nature* **466**, 720-726
56. Srivastava, M., Begovic, E., Chapman, J., Putnam, N. H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M. L., Signorovitch, A. Y., Moreno, M. A., Kamm, K., Grimwood, J., Schmutz, J., Shaprio, H., Grigoriev, I. V., Buss, L. W., Schierwater, B., Dellaporta, S. L., and Rokhsar, D. S. (2008) Nature 454, 955-960
57. Ye, S. Q., Zhou, X. Y., Lai, X. J., Zheng, L., and Chen, X. Q. (2009) Acta Pharmacol Sin 30, 913-918
58. Hundahl, C. A., Allen, G. C., Hannibal, J., Kjaer, K., Rehfeld, J. F., Dewilde, S., Nyengaard, J. R., Kelsen, J., and Hay-Schmidt, A. (2010) Brain Res 1331, 58-73
59. Ertas, B., Kiger, L., Blank, M., Marden, M. C., and Burmester, T. (2011) J Biol Chem 286, 3185-3193
60. Fago, A., Mathews, A. J., Moens, L., Dewilde, S., and Brittain, T. (2006) FEBS Lett 580, 4884-4888
61. Kraus, D. W., and Colacino, J. M. (1986) Science 232, 90-92
62. Doeller, J. E., and Kraus, D. W. (1988) Biol. Bull 174, 67-76
63. Cheng, C. H., di Prisco, G., and Verde, C. (2009) IUBMB Life 61, 184-188
64. Ganfornina, M. D., and Sanchez, D. (1999) Bioessays 21, 432-439
65. Nosengo, N. (2011) Nature 470, 444-445

FOOTNOTES

XB received support from Europôle Mer, a research consortium on marine science and technology in Brittany. Crystal structure determination was performed at the crystallography platform of the Station Biologique de Roscoff, supported by the ‘Region Bretagne’, the Centre National de la Recherche Scientifique’ and ‘Université Pierre et Marie Curie, Paris 06’. We are indebted to the staff of the European Synchrotron Radiation Facilities (ESRF, Grenoble, France), beamline ID23-I and BM30A, for technical support during data collection and treatment. Confocal microscopy was performed in the imaging facility platform “Merimage” at the Station Biologique de Roscoff. Spectroscopic studies were supported by Inserm and the University Paris 11. All authors are indebted to Ian Probert from the Marine Resource Center of the Station Biologique de Roscoff for discussions and critical reading of the manuscript.

FIGURE LEGENDS

FIGURE 1. A: photograph of a colony of the symbiotic acoel Symagittifera roscoffensis (4 to 5 mm long) at low tide. The green color is due to the presence of about 50000 photosymbionts (the
unicellular green algae *Tetraselmis convolutae*) harbored within each adult acoel. B: (1) Light micrograph of the anterior tip of a juvenile with the statocyst (S) flanked by two photoreceptors (P). Cilia are visible on the periphery of the head; (2) Composite confocal image showing the red SrNgb1-antibody signal surrounding the statocyst with peripheral extensions, DAPI stained nuclei appear in blue; (3) Confocal image illustrating RF-amide stained *Symsagittifera roscoffensis* nervous system. C: Magnification of the extremity of the anterior tip (“the head”). Arrows indicate fiber-like structures labelled with SrNgb1 anti-body occurring at the same place of the frontal glands (frontal sensors). These fiber-like structures are superimposable with serotonergic nervous system and especially neurites.

**FIGURE 2.** A: UV and visible spectra for SrNgb1. In the inset is shown the partially oxygenated spectrum measured under an oxygen tension of 10 Torr. The dashed lines refer to the maximum absorption for the fully oxy and deoxy spectra. B: Autoxidation of SrNgb1 in 50 mM Tris-HCl 0.2 mM EDTA 10 U SOD and catalase at pH 8.0 under 1 atm O₂ at 25 °C. In the inset is shown the variation of absorption occurring during the redox kinetics. (right panel) Autoxidation of SrNgb1 in 50 mM Tris-HCl 0.2 mM EDTA 10 U SOD and catalase at pH 8.0 under 1 atm O₂ at 25 °C. The inset shows the variation of absorption occurring during the redox kinetic.

**FIGURE 3.** A: Ribbon representation of SrNgb1 crystal structure (4B4Y): (left panel) close up view of the heme binding pocket of SrNgb1 highlighting the hydrogen bonding network involving the distal heme binding position and a tightly bound water molecule (HOH117); (right panel) 3D structural super imposition with murine Ngb (1W92). B: Multiple sequence alignment based on the structural superimposition with murine Ngb (1W92), bovine hemoglobin (1JEB) and sperm whale myoglobin (1O7M), as obtained with the program ESPRIP (http://espript.ibcp.fr/ESPRIP/cgi-bin/ESPRIP.cgi). The conserved histidines (axial heme ligands) are marked by black triangles. The red triangle marks a Trp residue involved in the tight binding of a water molecule in the distal heme pocket. The eight helices that form the classical globin fold are numbered from A to H and color coded from blue (N-terminus) to red (C-terminus) in the same manner as in the ribbon representation of SrNgb1 in Figure 1b. C: Extract of the SrNgb1 crystal structure highlighting the relative orientations of the heme-ligand containing helixes E and F. A proline at position 94 in helix F leads to a discontinuous and bent helix F in SrNgb1 (4B4Y). The same structural extract showing the relative orientations of helixes E and F in murine Ngb (1W92), where Helix F is continuous and straight.
FIGURE 4. Expression patterns of two Clytia globin genes CheNg1 and CheCyto in several territories of the medusa. A-C: Whole-mount in situ hybridizations for CheNg1, CheCyto (Blue= NBT/BCIP development) and Chemcol3-4a (red=Fast red development). D-H: All bulbs have the same orientation, proximal area on the top and tentacle on the bottom. Crescent shaped expression patterns in the ectodermal layer of a tentacle bulb. All of them are interrupted on the external side of the bulb but sometimes the continuity of the staining on the inner face is visible (E, G). E and G: purple staining indicates expression of two genes: in each case, minicollagen 3-4a staining was revealed first with fast red and then the other probe was revealed using NBT/BCIP. E’ and G’: details of the staining in nematoblasts (black arrowhead). F’ and F’’: Higher magnification of a statocyst delimited by the dotted circle: in situ hybridization (in black and white) and DAPI counterstaining (in red) merged after conversion of the in situ staining in grey scale. I-L: Gene expression in the manubrium views (mouth on the bottom). The signal is concentrated in the ectoderm layer at the base of the manubrium. J’ and L’: detailed views of double-stained cells corresponding to nematoblasts (black arrowhead), note that there is no signal in mature nematocytes (white arrowhead). go: gonad, ma: manubrium, tb: tentacle bulb. Scale bars: A-C: 100 µm ; D-H: 25µm ; I-M: 50µm ; E’, G’, J’, L’:5µm ; F’, F’’: 10µm.

FIGURE 5. A: Schematic and consensual representation of metazoan phylogeny illustrating the presence /absence of Ngbs, other globins, the two other respiratory proteins (hemocyanin and hemerythrin) and blood circulatory systems. The sporadic presence of globins in certain metazoan lineages can be explained by independent functional shifts from Ngb-like proteins (i.e convergent evolution). Acoelomorphs and Xenoturbella are represented in two alternative phylogenetic positions, reflecting the ongoing debate as to their affiliations. B: Unrooted molecular phylogeny based on multiple alignments of a subset of 84 sequences that comprise 138 amino acids of Ngbs, LGB-related (Ngb-like), Hb, Mb and Cyb sequences from diverse phyla. Dots indicate a possible miristoylated Ngb-like (green) or not (red). Vertebrates globins are in the yellow clusters: Hemoglobin, myoglobin and cytoglobins appear clearly as an invention of Vertebrates while vertebrate Ngbs and GbXs are imbedded within the large green group where functional neuroglobin of Symasagittifera roscoffensis and Clytia haemispherica occur (respective names are in bold red). The blue cluster that includes Ngb-like sequences from Choanoflagellates, Porifera (sponges), Placozoa, some Cnidaria, some protostome and deuterostomes including vertebrate Ngbs (yellow cluster) likely represents the plesiomorphic members of Ngbs.
**TABLE 1.** O₂ and CO binding data. Experimental conditions: 50 mM Tris-HCl 100 mM NaCl, 5 mM DTT, pH 8.0. Human Ngb experimental conditions: 100 mM potassium phosphate, 2 mM DTT, pH 7.0. O₂ solubility coefficient was $1.82 \times 10^{-6}$ moles/liter at 25 °C and for CO solubility coefficient was $1.36 \times 10^{-6}$ moles/liter at 25 °C. ° O₂ affinity was estimated equal to 1.8 Torr.
**FIGURE 2**

A. 

**Symagittifera roscoffensis 1**

![Graph of absorbance vs. wavelength (nm)](image)

- Ferrous O2
- Ferric (pH 8)
- Ferrous deoxy
- Ferrous CO

B. 

**Autooxidation kinetics**

![Graph of ΔA vs. time (min)](image)

- ΔA vs. wavelength (nm)
TABLE 1

|                      | SrNgb1 | Human Mb | human Ngb (25°C) |
|----------------------|--------|----------|------------------|
| \( k_{on} \text{CO} \) (\( \mu \text{M/s} \)) | 0.35   | 0.65     | 40               |
| \( k_{on} \text{O}_2 \) (\( \mu \text{M/s} \)) | 7      | 15       | 170              |
| \( k_{off} \text{O}_2 \) (s)             | 35+/-5 | 27       | 0.7              |
| \( K \text{ O}_2 \) (M)                 | 5.0    | 1.8      | 0.004            |
| \( P_{50} \text{O}_2 \) (Torr)           | 2.8    | 1        | 6.8              |
| \( k_{on} \text{His} \) (s)             |        |          | 1800             |
| \( k_{off} \text{His} \) (s)             |        |          | 0.6              |
| \( K \text{ His} \)                     |        |          | 3000             |
|                                | Beamline at ESRF | ID23-I | BM30A |
|--------------------------------|-----------------|--------|-------|
| Wavelength (Å)                 | 0.93            | 1.7389 |       |
| Unit cell parameters           | a=b=97.07, c=140.10, |       |       |
|                               | α=β=90, γ=120   |       |       |
| Resolution range (Å)           | 30.06–2.3 (2.36-2.30)       | 50.06–3.2 (3.29-3.20) |
| No. of observations            | 57993 (3605)     | 60276 (3599) |
| No. of unique reflections      | 12137 (688)      | 11480 (675) |
| Completeness (%)               | 91.7 (77.4)      | 93.3 (77.4) |
| <I/σ (I)>                      | 16.6 (2.2)       | 14.9 (2.0) |
| Redundancy                     | 4.8 (5.2)        | 5.2 (5.3) |
| Rsym (%)^b                     | 4.8 (43.3)       | 7.6 (56.7) |

**Phasing statistics**

Anomalous difference (CC in %, given by ShelxE) 35.21

Figure of merit 0.335

**Refinement statistics**

Rcryst (%) 21.4

Rfree (%)^c 25.1

Esu based on Free R value 0.17

Overall B factor (Å^2) 38.1

Protein 38.0

Heme 33.4

Solvent 41.2

Rms deviation in bond lengths (Å) 0.029

Rms deviation in bond angles (°) 2.46

^a Values for the highest resolution shell are given in parenthesis.

^b Rsym = Σ | I-Iav | / Σ | I |, where the summation is over all symmetry-equivalent reflections.

^c Rfree values were calculated on 5% of the data (904 reflections) that were set aside in the minimization steps.
TABLE S2A

Below are the primary sequences (in fasta format) used for molecular phylogeny.

In order to see the original alignment used to perform the molecular phylogeny this file has to be opened with any sequence editor.

New LegHemoglobin related (including Ngb-like) sequences found in various phyla are highlighted in yellow (with supplemental information concerning their origin).

For each sequences except for vertebrate globins, Panther prediction are mentioned in the single line description such as Ng for neuroglobin-like, LGB for LegHemoglobin-related, Cytoglobin for cytoglobin-like, HGB for Hemoglobin related, and GlobinX for globinX-like.

Note that for each of these sequences the score of Panther prediction hit is specified: 1 (the score of the Panther hit is better than E-3, but worse than E-11 (protein is evolutionarily related but function may have diverged), 2 (the score of the Panther hit is better than E-11, but worse than E-23 (molecular function likely to be the correct but biological process/pathway less certain), 3 (the score of the Panther hit is better than E-23 (very likely to be a correct functional assignment).

>VERTEBRATE GlobinX

>ANOLIS CAROLINENSIS GbX (from Droge et al 2011)
-------------------------------------MGCAISGATNCAGAVEGTCQ motivation
LDLNRTELGLSNGRTPTEPFSQGAKELIRGSWEILHK-----DIARVGIIVFIRLTFETHPE
-CKDVFFLFR-D-ID-DFQQLKSRELQALRHGLRVMFIEKSVARM-----DQ-EPKLHL
AFELGRSHFRYK----APKYYEYGIGIQRQGAAQPILKEAWT-PETERKAMEGLFQYLAAT
MREGFYMKEQKATOKL--------------------------

>CALLORHINCHUS MILII GbX (from Droge et al 2011)
------------------------------------------------------MGCAIS
GPQYYYLPAGREGIDVAVASLSDRQTQLKETWLQVQ-----DIAKVGIIFVRLTFETHPE
-CKDAAFLFR-D-VE-DLQLRSKSLRHLRVMSFIEKSVARL-----DQ-EDRLQQL
XLELGKSHFRYS----APKYYYVGNEFACVQPILKEWRT-AVEEANKLHFLYTSV
MKGKYQDEERGCPREKPGHPNSV--------------------------

>DANIO RERIO GbX (from Droge et al 2011)
-----------------------------------------------------MGCAISG
SGLTAGAPEIRFQEEPAGILTMSHREKKESDKEQWVQ-----DIAKVGIIMFVRLTFETHPE
-CKDVFFLFR-D-VE-DLQLRSKSLRHLRVMSFIEKSVARL-----DQ-EPKLHL
AELGKSHFRYS----APKYYYVGNEFICAVQPILKEAWT-AVEEANKLHFLYTSV
MKGKYQDEERGCPREKPGHPNSV--------------------------

>ORYZIAS LATIPES GbX (from Droge et al 2011)
------------------------------------------------------MGCAIS
GPQYYYLPAGREGIDVAVASLSDRQTQLKETWLQVQ-----DIAKVGIIFVRLTFETHPE
-CKDAAFLFR-D-VE-DLQLRSKSLRHLRVMSFIEKSVARL-----DQ-EDRLQQL
XLELGKSHFRYS----APKYYYVGNEFACVQPILKEWRT-AVEEANKLHFLYTSV
MKGKYQDEERGCPREKPGHPNSV--------------------------

>PETROMIZON MARINUS GbX (from Droge et al 2011)
--------------------------------MGCTSTDEHTCAQSSSSQGQASRQKQ
QPEEQRAAGEEQPPGPGQQPASEQSRVLRSWALQ-----DIARVGIIVFIRLTFETHPE
-CKDVFFLFR-D-CE-DLQLKSMLMKQALHLRVMSFIEKSVARL-----EQ-ECVLEQL
IVEMRGKHYKYN----APKYYSYVGIFIEFLATVQFPLQEKWT-NEVEDAWCLFLRYAAT
MKGRGELLEAAASNQTNAYQDGQGNHATAM--------------------------

>PYTHON MORULUS GbX (from Droge et al 2011)
----------------------------------------QKELIRESKILH-----NITRVGIIVFIRLTFETHPE
-CKDVFFLFR-D-ID-DLQLMKMLKQALHLRVMSFIEKSVARL-----DQ-EGKLEVL
AFELGKSHFRYS----APKYYYEGIGIQRQGAAQPILKEWRT-LEVEKAWK
--------------------

29
>TETRAODON NIGROVIRIDIS GbX (from Droge et al 2011)

--------------------------------------------MGCAISSLGAKA
EFGRSAAEEDAAAAAVYPREDQIQMIKDSWKVIRD-----DIAYGVIIMFVRLFETHPE
-CKDVFLFR-D--VE--DLERSSRELRAHGLRVSMIEKSVARAI--DQ-QDRLREAL
AVELGKSHYHNY-----APFYYSSYGAEFIAVQPIRKERT-SELEEAJTLFQQVTG
LRKGHEGQRHRHLPPKDPGKRTSA-------------------

>XENOPUS TROPICALIS GbX (from Droge et al 2011)

---------------------------------------------MGCAISSLG
WQWRRSDLHTETSPFLPLNLSEQQQLQLVESWRLIQH-----DIAYGVIIMFVRLFETHPE
-CKDVFLFR-D--VD--DLQARANKDLRAHGLRVLIPVEKSVAR----AD-CARLEEL
ALELGRSXRYRN------APFPRYQQYGTEFISAVCPMLHDKWT-AEVEAAWKLGFAYICTV
MERGYQEEE-RRHDGRSLIDGLQGNKGLI-------------------

>VERTEBRATE MYOGLOBINS

----

>HOMO_SAPIENS_MYOGLOBIN_gi|4885477|ref|NP_005359.1|

------------------------------------------MGCAISSLGQWQLVLNVWGKVEA-----DIPGHQEVIRLFGKHPE
-TLKEFKFK-H--IK--SEDEMKEALDKKHATVLALGLIK--------KGHEAE
IKPLAGSHATKHK------IPVKYLEFISECIQQLQSHFDAGF-ADAQGAMKALELFRKD
MAYYKELFGQH-------

>DANIO_RERIO_MYOGLOBIN_gi|41053652|ref|NP_956880.1|

------------------------------------------MGCLSSLGWQWQLVLNVWGKVEA-----DIPGHQEVIRLFGKHPE
-TLLEFKFK-G--IK--SEDEMKEALDKKHATVLALGLIK--------KGHEAE
VHLAEHANKHK------IPVKYLEFISEIQLQSHFDAGF-ADAQGAMKALELFRKD
MAYYKELFGQH-------

>BOS_TAURUS_MYOGLOBIN_gi|27806939|ref|NP_776306.1|

------------------------------------------MGCAISSLGQWQLVLNVWGKVEA-----DIPGHQEVIRLFGKHPE
-TLKEFKFK-H--IK--SEDEMKEALDKKHATVLALGLIK--------KGHEAE
VHLAEHANKHK------IPVKYLEFISEIQLQSHFDAGF-ADAQGAMKALELFRKD
MAYYKELFGQH-------

>GALLUS_GALLUS_MYOGLOBIN_gi|268607704|ref|NP_001161224.1|

------------------------------------------MGCAISSLGQWQLVLNVWGKVEA-----DIPGHQEVIRLFGKHPE
-TLLEFKFK-G--IK--SEDEMKEALDKKHATVLALGLIK--------KGHEAE
VHLAEHANKHK------IPVKYLEFISEIQLQSHFDAGF-ADAQGAMKALELFRKD
MAYYKELFGQH-------

>VERTEBRATE CYTOGLOBINS

----

>DANIO_RERIO_CYTOGLOBIN_gi|23308615|ref|NP_694484.1|

------------------------------------------MEKVPGEMEIERRERSEELSEAERKAVQATWARLYA-----DIAYGHEVLRMFLHPE
-TLLEFKFK-G--IK--TPEQMGSEDALDHGHTVLQGLKILKQ--------KGHEAE
VHLAEHANKHK------IPVKYLEFISEIQLQSHFDAGF-ADAQGAMKALELFRKD
MAYYKELFGQH-------

>OVIS_ARIES_CYTOGLOBIN_gi|261244962|ref|NP_001159664.1|

------------------------------------------MEKVPGEMEIERRERSEELSEAERKAVQATWARLYA-----DIAYGHEVLRMFLHPE
-TLLEFKFK-G--IK--TPEQMGSEDALDHGHTVLQGLKILKQ--------KGHEAE
VHLAEHANKHK------IPVKYLEFISEIQLQSHFDAGF-ADAQGAMKALELFRKD
MAYYKELFGQH-------

>HOMO_SAPIENS_CYTOGLOBIN_gi|20987458|gb|AAH29798.1|

------------------------------
>VERTEBRATE HEMOGLOBINS

| Species                  | Accession          | Type          | Description                  |
|--------------------------|--------------------|---------------|------------------------------|
| EQUUS_CABALLUS_HB_ZETA   | gi|167621441|ref|NP_001108014.1|_hemoglobin_subunit_zeta|
| HOMO_SAPIENS_HB_BETA     | gi|256028940|gb|ACU56984.1|_beta-globin|
| TAENIOPYGIA_GUTTATA_HB  | gi|323668297|ref|NP_001191172.1|_hemoglobin_subunit_alpha-A|
| ORYCTOLAGUS_CUNICULUS_HB| gi|284005393|ref|NP_001164887.1|_theta_1_gobin|
| JBC                      |                   |               |                              |

>VERTEBRATE NEUROGLOBINS

| Species                  | Accession          | Type          | Description                  |
|--------------------------|--------------------|---------------|------------------------------|
| HOMO_SAPIENS             | gi|108640645|ref|NP_067080.1|_neuroglobin_[Homo_sapiens]|
| DANIO_RERIO              | gi_18859087_ref    | NP_571928.1|_neuroglobin_[Danio_rerio]|
| GALLUS_GALLUS            | gi_154816290_gb    | ABS87379.1|_neuroglobin_[Gallus_gallus]|

---
>MONODELPHIS_DOMESTICA_gi_78486584_ref_NP_001030592.1__neuroglobin_[Monodelphis_domestica]
>----------------------
>-RLSGLFQEQELRESWQKVNS--NPLQHGMILFTRLDFLED
-LLPLFAQVEQ-R-OFS-SPQDCLSSEPFLDHIRKVMLVIDAHTVH---EN-LSSLEEEY
LTNLGKKHKAVG----VKLSSFSTVESGLLYMLEQCLGSPFTT-VTMKEAWTQLYGAVVQA
MSRGW
>----------------------
>CHIONODRACO_MYERSI_gi|270341071|emb|CAR57914.1|_neuroglobin_[Chionodraco_myersi]
>--------------KLSGKDELIRGSWESLGK---NKVPHGVMFSRFLED
-LLTLFHYTT-N--CG--STQDCLSSEPFLHEVTKMLVIDAASPLDLEF
LNLNGKHKAVG----VNTQSFSAEVGLYMLQLGSPFALT-YPLQAWLNLYSIVVAAMM
>--------------
>BOVICHTUS_VARIEGATUS_gi_270341067_emb_CAR57912.1__neuroglobin_[Bovichtus_variegatus]
>--------------KLSGKDELIRGSWESLGK---NKVPHGVMFSRFLED
-LLTLFHYTT-N--CG--STQDCLSSEPFLHEVTKMLVIDAASPLDLEF
LNLNGKHKAVG----VNTQSFSAEVGLYMLQLGSPFALT-YPLQAWLNLYSIVVAAMR
>--------------
>SIPUNCULA
>THEMISTE_1_LGB2_gnl|ti|2001595631_Tl_MM2_10H06__From_transcriptome_available_at ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
----------MNQAPEDQTKTFTPDEMALRSSWNAVA--GDLAGXFDMFVKMFEQKKE
-TKEVFPPFAR-G-N~~~DAXQMQHGIINVFHVTFMKNIDEVVKNADDR-LEDVSM
LARVEGHGHHQGNNVPSAYFPFLGAMAMI-LLLKANYKSYD-SKLDCWVPLNFMNNE
MTTGQEQVYRGK
>THEMISTE_2_LGB2_gnl|ti|2001595657_Tl_MM2_11B08__From_transcriptome_available_at ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
----------LRLSNGWNAVA--GDLVGRFDSFVKMFEQKKE
-TKEVFSSFAR-G-N----DAXQMQSNSVNFHTRVMKNDENVKNA---DR-LEDVSM
LRQVGGRRHGGHHQNNVPSAYFPVLGERAM-LLIKANYKAYD-SKLDCWVRLDFINK
MTTQEQYEEIKES
>THEMISTE_3_LGB2_gnl|ti|2001594772_Tl_MM2_01H11__From_transcriptome_available_at ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
----------MNLDLEDQRFDFPEELNRLRWSWNAVA--GDLFGGRFDMFTKMFQKKE
-TKEVFSSFVR-G-S~~~DAXQMQSSKFKFHTRVMKNDENVKHA---DR-LEDVSM
LRQVGGGHRGSSHNNVPSAYFPFLGIALR-NLMSQSYKGD-SKLDWTRAWGFTMSQ
MMYG
>BRACHIOPODA
>TEREBRETALIA_Ng2_gnl|ti|1987128689_Tt_MM1_02C05__From_transcriptome_available_at ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
----------FTTQEKNLRSSWRAFVA--EDPAKRYQMFMMENRFD
-TKFFFK-FV-Q-KD---ALGTMKSTLSLFHTRGVRAYQIQCVCDCL---DD-PTEAVPL
LRQVGGKHCNPCNNVFDATYFPFLGAMMAR-QMSESSGLQGYD-QELDALWEOFYTWIKR
MEEGM
>TRICHOPLAX_ADHAERENS_2_LGB3_gi|196012120|ref|XP_002115923.1|hypothetical protein TRIADDRAFT_59832 [Trichoplax adhaerens]--------------------------MAPAQDLQTIRETWALVF-----DLKKHGTVLFLRLEFQHPD-VQRLFEKIKD----V---PHDQLATNENFHTTRVMETIDHAVKGI---DN-LPALTVL
LKQLGSSHAQYN---VKKEYFKIGLRISEF----------

>TRICHOPLAX_ADHAERENS_3_Nq1_gi|196001583|ref|XP_002110659.1|hypothetical protein TRIADDRAFT_54901 [Trichoplax adhaerens]--------------------------MVLVDNYSLIKLSPA-TKIYFHGVD-F---EK-RDSYLAKNTFLRNASFAIFIVIQGD-----MD-IPSVESY
FRVVGSKHHYSN---IKLEHVOQDISDAFLERNAKKFT-KSTEWSRSGQMTDA
IKNGIMKAQQRN---

>TRICHOPLAX_ADHAERENS_3_LGB3_gi|196016934|ref|XP_002118316.1|_hypothetical_protein_TRIADDRAFT_62364_[Trichoplax_adhaerens]--------------------------MTKIDSENHVKSNVTVNDIKSYLNYQERQAIIDSWNAIST---EKQKYGTILFLKLFQIDVS-VKSLFTIFD-F---PLEDIQSPIHFRSHAMRFMQSLETGVLMLGFD-KESCDFL
FKSLGSRHHFYMID-LKSEPFLDIPEICILTHIKKCGCNWMT-NETADMWIAVKVLCEL
FREGLETKPK---

>PORIFERA

>CARTERIOSPONGIA_FOLIASCENS_LGB3_gi|241971149|gb|GO083496.1|GO083496_DMPR57_15864_Carteriospongia_foliascens_DMP_cDNA_Library_Carteriospongia_foliascens_cDNA,_mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nuccore/--------------------------MVLVNNYSLIKLSFA-TKIYFHGVD-F---EK-RDSYLAKNTFLRNASFAIFIVIQGD-----MD-IPSVESY
FRVVGSKHHYSN---IKLEHVOQDISDAFLERNAKKFT-KSTEWSRSGQMTDA
IKNGIMKAQQRN---

>AMPHIMEDON_QUEENSLANDICA_LGB3_gi|340378768|ref|XP_003387899.1|_PREDICTED:_neuroglobin-like_[Amphimedon_queenslandica]--------------------------SLTSAQVALIESTWKVVK----DLQGAGMVFMRKLFEMSG-LIQNLFFFA---GDSLSDDDNEGMKHHSFIVMTSLDDCLKIV----HD-VFKLKDE
LISLGAVHHIQG----VTSEHFAPVGSALWALEALKKEFKT-PEVKAWLALYKVQSL
MEQGMEEG---MEQGMEEG---

>NEMATOSTELLA_VECTENSIS_1_LGB3_gi|156408000|ref|XP_001641645.1|_predicted_protein_[Nematostella_vectensis]--------------------------TMKIDSENHVKSNVTVNDIKSYLNYQERQAIIDSWNAIST---EKQKYGTILFLKLFQIDVS-VKSLFTIFD-F---PLEDIQSPIHFRSHAMRFMQSLETGVLMLGFD-KESCDFL
FKSLGSRHHFYMID-LKSEPFLDIPEICILTHIKKCGCNWMT-NETADMWIAVKVLCEL
FREGLETKPK---

>CNIDARIA

>NEMATOSTELLA_VECTENSIS_2_LGB2_gi|162098184|gb|FC288990.1|FC288990_CAGN4759.fwd_CAGN_Nematostella_vectensis_Nemve_mixed_stages_unfert_eggs_to_primary_polyps_Nematostella_vectensis_cDNA_clone_CAGN4759_5',_mRNA_sequence

---
LYELGTRHQTRQ----LSEEQLKFMGGAFLFAMRLHLRKEWS-RATSKAWEKIFSFMADA
MMRGCKG-----------------------------------------------

>NEMATOSTELLA_VECTENSIS_scaffold_42000019_LGB2_from_Bailly_and_Vinogradov_2008

-------------------------------------------------------- MGCG
SSVVTQGMPLCHLGKLECMDTYEQYLIRETWFKEV--SKKEIGVSYKRFPLMMHPG
-LQTYSEFK--------HKKININGSHGHPRLMAIDNVALT---GD-SDFSAY
LEVELGGRHNMNFRGPHTFNDRKFLSLVEEILATASLWD-FQVEEAWNLFLDSITAM
ILRGQIALKV--------

>NEMATOSTELLA_VECTENSIS_scaffold_5000153_LGB3_from_Bailly_and_Vinogradov_2008

---------------------------------------------------------MGCG
SSVVTQGMPLCHLGKLECMDTYEQYLIRETWFKEV--SKKEIGVSYKRFPLMMHPG
-LQTYSEFK--------HKKININGSHGHPRLMAIDNVALT---GD-SDFSAY
LEVELGGRHNMNFRGPHTFNDRKFLSLVEEILATASLWD-FQVEEAWNLFLDSITAM
ILRGQIALKV--------

>NEMATOSTELLA_VECTENSIS_scaffold_141000032_LGB3_from_Bailly_and_Vinogradov_2008

---------------------------------------------------------MGCG
GSSVVTQGMPLCHLGKLECMDTYEQYLIRETWFKEV--SKKEIGVSYKRFPLMMHPG
-LQTYSEFK--------HKKININGSHGHPRLMAIDNVALT---GD-SDFSAY
LEVELGGRHNMNFRGPHTFNDRKFLSLVEEILATASLWD-FQVEEAWNLFLDSITAM
ILRGQIALKV--------

>CLYTIA_HEMISPHERAERICA_1_LGB1_gi|294376172|gb|FP931337.1|FP931337_FP931337_CLytia_hemisphaerica_library_Clytia_hemisphaerica_cDNA_clone_SA0AAB120YK22_5',_mRNA_sequence

-------------------------------------------------MGAICSHKTL
VSQSFVYIKSQSNVPLSAKEVNILKTLTFPMFVKN---NLWKICLVAFERWFTLFPQ
-LRYMFRSLP-E-DI--QFEDLFKDALKMHDVTVLRELLIKKI--DN-VELLENT
LVDFQGQRHHMLG--AEQRYATALAAFLQGICMIMDVA--SSVENAWSLLLRFVMDL
LKLGMRMMEKEAEKESLNKYGNTGEELEAKQDGGEALDENENSAPLALINEDDSTSS
MVRGLQSRD--------

>CLYTIA_HEMISPHERAERICA_2_LGB2_gi|294400477|gb|FP945496.1|FP945496_FP945496_CLytia_hemisphaerica_library_Clytia_hemisphaerica_cDNA_clone_SA0AAB133YH06_5',_mRNA_sequence

--------------------------------------------------------MGSS
GSCLFKTMKPGNKFDAAPHTAEAEINIVQKWOWVAM---NLDSVYKLFKALFLQYKI
MAKDFPAN-N---DFQKSMNDRLSFHRVFHTINTVSLC--ND-GDFVASQ
LEHYGAIMAEY---IQATHLARFKDVMLETLEEAFKEGFQ-EDSKTAWSKIDVIAKY
MLKGDIAKEQTQKSELSSDDGQMIKQE--------

>CLAVA_1_LGB1_Locus_11167_Transcript_3/8_Confidence_0.360_Kindly_provided_by_Stefano_Piraino_from_Universita_del_Salento_Italy

-------------------------------------------------MGSKLCLAIHK
SVARSTQPSIHTKYNICPILTSDKDIKKLHSGSWIMKML-NWYKICLVVFDRWFMSYPE
-LRNMFKS-S-VL-TLQGKMMGMLQQLEELVDDLYK---DD-PDFBIET
VIQFEGGHHKLQ------AKQVLAYSLAIAFFQAICISLNLDRDNTENAWSLRFMLDC
LKFQMRNEMKETTSPDLKLEVQDG--------

>CLAVA_2_Ng_Locus_28538_Transcript_4/6_Confidence_0.667_Kindly_provided_by_Stefano_Piraino_from_Universita_del_Salento_Italy

--------------------------------------------------------EVGILRQWSMMS---DADANGIMFTALFAINGH
-LRTKAFGVD-V-YGP-DCCTTFGDERLLHLHRGFSSVEMINSVLQ-NF-FYELSGLK
IEKELKVHSLAG------IHLSDLQAFKKAMIQTTLKELKSEFD-DKALKAWSKFIDIVIEA
FDVDDDOIDDEKTOPGK---
>HYDRA_MAGNIPAPILLATA_Ng1_gi|221117935|ref|XP_002162062.1|_PREDICTED:_simil
ar_to_neuroglobin_[Hydra_magnipapillata]
------------------------------------------------------------
--------------------LSGKEIETLKKSWTTAKQ---FWNEICTCAFSRF
YDG---AEQKYATALGGSSVFAISQICFNIDM---ITEGAWDSLFKYIVT--

> MONTASTREA_FAVEOLATA_1_LGB3_gi|282539129|gb|GW263294.1|GW263294_CCHW10651
.b1_CCHW_Montastrea_faveolata_heat/dark/disease-
stressed_adult_+6_day_old_larvae_Montastrea_faveolata_cd
AClone_CCHW10651_5',_mRNA_sequence
MGCGGSGAIAKKNRTAPAPVAEQTQSPGLKETSPAKQHRQQPEGKRTEETVE
TEGAWDSLFKYIVT------------------

> MONTASTREA_FAVEOLATA_2_LGB2_gi|282548677|gb|GW271367.1|GW271367_CCHW15436
.b1_CCHW_Montastrea_faveolata_heat/dark/disease-
stressed_adult_+6_day_old_larvae_Montastrea_faveolata_cd
AClone_CCHW15436_5',_mRNA_sequence
---------------MERSKEGTDTLQIEMIRSSWEKTF---NKK<span style='color:transparent'></span>HHGQLLFKHLFEIAPE
-<span style='color:transparent'>MTDLFPFG</span>-----------DDFTKPQFTTHALNIMALDHAIQNLDN-PDVLPK
LRELQMQHAGFE---ITIKEFQHVGEALINVVLATGLD---

> ACOELA

> SYMSAGITTIFERA_ROSCOFFENSIS_2_LGB2_From_the_present_study:_illumina_lib
ary_made_by_Genoscope_French_Sequencing_Center
-----------------------------------------------MQLNNIQKCLPLF
FPKRNHMSDYDFAFEPFVPPVNSSESEERKSWKQIEEL---DAAKLGIAVFGVLF
FERYEIQASFSKIA-N--NKSSLNSVMLAHS</span>HMGHMIGKLQIPLL---NE-PENLLSK
VELGERHDFRK----ANDELLQYFCAY-VEAMAKKGQWK--KKTI</span>WEEKDDFIRAA
MVGLKRRKGGHSSNTTSAANTAAEKKNNSSSQQ----------------------

> SYMSAGITTIFERA_ROSCOFFENSIS_1_3D_LGB2_From_the_present_study:_illumina_l
rary_made_by_Genoscope_French_Sequencing_Center

--------MATLEMVQYVEQQSLIMEDVQVLLF-<span style='color:transparent'>NYDDFVFEDVELQOFMENPE</span>
-TQIFPFWAD-A-SK-TAKERMSHPRFSKSHAKSKGVSIDCLVDL---NG-VKKHEF
LSSLGAMHTKTK---VFTELF</span>GLGCGILQQVKKRVS</span>AEKWSEEKEWL</span>KAYGIITVM
VTE------------------

> NEMERTODERMATIDA

> MEARA_STICHOPI_Ng1_transcript_1_Kindly_provided_by_Andreas_Hejnol_from_SAR
SInternational_Center_for_Marine_Molecular_Biology
---------------------------------------------SRKILVCRLFESHFQ
-VRDEFHSE------DKDLVPADIAKAHMGRIMGLGRFVSNLESNVD-DNLLGQP
IHDLKRGHKDFK---APYLFDFVAFQIQHIKSKLEQVWT-DEIGDAKWMDIIVFN
LKSQONQMFDQGTV-------

>NEMERTODERMA_WESTBLADI_1_LGB2_gnl|ti|2247831830_Nw_MM1_39E05_From_transcri
ptome_available_at_ftpsftp.ncbi.nlm.nih.gov/pub/TraceDB/

------------------------MGTLTQVRVLIKSQKKVIS-GGRVEAGWLF</span>SKFFTSSE
-AQNYFKAEK-G--PLSEQLONNTQMKGHVLRVinYITDIV</span>TL---EV-DEMREEM
>CRASSOSTREA_GIGAS_2_LGB3_gi|313365341|gb|HS221690.1|HS221690_CCTS16119.b1_CCTS_Crassostrea_gigas_mixed_adult_tissues_library_4_normalized_Crassostrea_gigas_cDNA_clone_CCTS16119_5', mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/

---------------MTSTEFTDEQKQLVKTTWNIVRE---DISKVGVITFLRLFEKFPD-VQDLFVPFR-G-I---NSEELEQRNVQYHLMVGMGIHIEKCITRI----DQ-PAKTSMLMLGEKHHVVPD---TIKIEYFDDLPSQLIQAIATRHWSDV--SQREQWNNFYC---

>CRASSOSTREA_GIGAS_1_LGB2_gi|318048551|gb|FQ663445.1|FQ663445_FQ663445_Crassostrea_gigas_library_(Genoscope_-_CEA)_Crassostrea_gigas_cDNA_clone_WY0AAA49YL07FM1,_mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/

---------MGC_DASKSTDVIENEEIESDTPDFTDQIQTDTIRTWPIALAC-DMVDIGSKVFLKIFIDEK-LKYAFFSFS-D---EENELLRHPPFIDHVRMFOIQIDYVENL---DQONSDFHQA_LMLGAKHATYFPF---QVSHTFVNFKALLEAWEASEIGEEFIP-EVQNCWVTQLFAYIMRY_IVQGYELFSEC---

>ARTHROPODA

>APIS_MELLIFERA_LGB2_gi|118150510|ref|NP_001071291.1|globin 1 [Apis mellifera]

---------MGT_FLRFGISSSDDNRIQDAATGQLTERKVQVQNTWAVVRK---DEVASGIAVMATTAFKYYK_LYYRQYTMF-G---ENELPANKRFQAHCVTIALNNVIDFL---HD-PGLMEAS_LIGLVERHKKRG---QTKEFQNLKEVMMILRQALGKQYT-PEVAAEWNKTLDMMFGK_IVY---------

>CARCINUS_MAENAS_1_LGB2_gi|299757081|emb|CBN88274.1|hemoglobin [Carcinus maenas]

----------------------------------MGAVLSVVGWLSPTQVGAV_TFPEEGSGLPGEADVDDKTGTGLLTRKARTAYRTWDLRVP---NPKLHGINLFLTMQFEPPF-IQTRFKQFAG-K---SIEELKNSSKRLAAGVTVMVATMVNLD---ED-VSLLLKLNATAHNRDG---VFKDFELLAVFLRLVDNLGSAWS-PVAAEWNKTLDMMFGK_IVY---------

>CARCINUS_MAENAS_2_LGB_related_gi|84413135|gb|DW250579.1|DW250579_Cm_mx1_36e01_SP6_Green_Shore_Crab_Multiple_Tissue, Normalized Carcinus_maenas_cDNA_c lone_Cm_mx1_36e01_5' similar_to_gb|EAL8715.1|_ENSGPF00000028536_Anophele s_gambiae_str._PEST._Score_=_156_bits_(395),_Expect_=_4e37, mRNA_sequence_F rom_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/

---------MGCHITKNKKNKETEE.VKVIDLQEPPEPPPPDPRLPLTRARQFNNIKSWKGIAR--AIEPTGVNMFFKQLENHSE-LITFFKFR-O--LR--TRDELASELEAEHAVIMNSIDEIKAM---DN-VDFFDL_LHIGASHKHFPG---FKKEYFWKIEHFPLEAVRGLLDYRT-DNMDNIYRITIKLLETV_VVRGYNLAEKEL---

>DAPHNIA_PULEX_GlobinX3_gi|321478927|gb|EFX89883.1|hypothetical protein DAPPUDRAFT_231989 [Daphnia pulex]

---------RLEK---DIAQVGIIVFINLFEPD-MQSVLFPT--------GVLDDLKKSSKLSEHALVRMGAVQRAVRLQEPKHAF_LSELRKHEKNG---AKLEYIDYI GFQFCLAIRPILGDRT-LETEKAWTULLDMYTAA_MKESLVEARNASAESSKKPLTLPSSSSSSSAAATD---
Daphnia Carinata LGB2 gi|290774776|gb|GR506052.1|GR506052_LAaa_0006_F05_Gametogenic_water_flea_(Daphnia_carinata)_cDNA_library_Daphnia_carinata_cDNA_5', mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/

>Harpeggnathos_Saltator_LGB2_gi|307192580|gb|EFN75678.1|_Globin_[Harpeggnathos_saltator]

>Tunicata

>Ciona_intestinalis_Gb4 HBG related NP_001027701.1

>Ciona_savigny_Gb1 LGB3 (from Hoffmann et al 2012)

>Molgula Tectiformis LGB3 gi|68107250|gb|CJ414980.1|CJ414980_CJ414980_Molgula_tectiformis_unpublished_cDNA_library,_larva_Molgula_tectiformis_cDNA_clone_mtlv020d18_5', mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/

>Botryllus Schlosseri Cytoglobin2 gi|322520564|gb|JG298067.1|JG298067_CCAO1145.1_CCAO_Botryllus_schlosseri_total_asexual_and_embryonic_development_Botryllus_schlosseri_cDNA_clone_CCA01145_5', mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/

>Cephalochorda

>Branchiostoma floridae GB1 LGB3 XP_002608549.1 hypothetical protein BRLDRAFT_98913

--------------------------
TABLE S2B

Multiple alignments with emphasis on informative sites selected for the molecular phylogeny analysis (underlined in yellow)

| 10  | 20  | 30  | 40  | 50  | 60  | 70  |
|-----|-----|-----|-----|-----|-----|-----|
| Vertebrate GlobinX |  |  |  |  |  |  |
| Anolis Carolinensis G |  |  |  |  |  |  |
| Callosobruchus Mili G |  |  |  |  |  |  |
| Danio Rerio Gbx (from |  |  |  |  |  |  |
| Oryzias Latipes Gbx |  |  |  |  |  |  |
| Petromyzon Marinus Gb |  |  |  |  |  |  |
| Python Morulus Gbx |  |  |  |  |  |  |
| Tetraodon Nigrovirid |  |  |  |  |  |  |
| Xenopus Tropicalis Gb |  |  |  |  |  |  |
| Vertebrate Myoglobin |  |  |  |  |  |  |
| Homo Sapiens Myoglobin |  |  |  |  |  |  |
| Bos Taurus Myoglobin |  |  |  |  |  |  |
| Gallus Gallus Myoglobin |  |  |  |  |  |  |
| Vertebrate Cytochrome C |  |  |  |  |  |  |
| Homo Sapiens gi | 10864 |  |  |  |  |  |
| Gallus Gallus gi | 188590 |  |  |  |  |  |
| Monodelphis Domestica |  |  |  |  |  |  |
| Chionodraco Myersi gi |  |  |  |  |  |  |
| Bovichus Variegatus |  |  |  |  |  |  |
| Sipuncula |  |  |  |  |  |  |
| Themiste 1 LGB2_gnl |  |  |  |  |  |  |
| Themiste 2 LGB2_gnl |  |  |  |  |  |  |
| Themiste 3 LGB2_gnl |  |  |  |  |  |  |
| Brachiopoda |  |  |  |  |  |  |
| Terebratalia Mg2_gnl |  |  |  |  |  |  |
| Annelida |  |  |  |  |  |  |
| Alvinella Pompejana 1 |  |  |  |  |  |  |
| Alvinella Pompejana 2 |  |  |  |  |  |  |
| Aphroditia Aculeata LG |  |  |  |  |  |  |
| Capitella sp 1 LGB2_F |  |  |  |  |  |  |
| Capitella sp 2 LGB2_F |  |  |  |  |  |  |
| Capitella sp 3 LGB2_F |  |  |  |  |  |  |
| Platyneris dumerilli |  |  |  |  |  |  |
| Helobdella Robustula |  |  |  |  |  |  |
| Echiura |  |  |  |  |  |  |
| Urechis Caulo Mg1_gnl |  |  |  |  |  |  |
| Platyhelminthes |  |  |  |  |  |  |
| Macrostomum Linnavo 1 |  |  |  |  |  |  |
| Macrostomum Linnavo 2 |  |  |  |  |  |  |
| Chondorhynchus |  |  |  |  |  |  |
| Monosiga Breviscullels |  |  |  |  |  |  |
| Salpingoeca sp Mg2_gi |  |  |  |  |  |  |
| Placozoa |  |  |  |  |  |  |
| Trichoplax Adhaerens |  |  |  |  |  |  |
| Trichoplax Adhaerens |  |  |  |  |  |  |
| Trichoplax Adhaerens |  |  |  |  |  |  |
| Forsteria |  |  |  |  |  |  |
| Carteria Folia |  |  |  |  |  |  |
| Aphytmidion Queensland |  |  |  |  |  |  |
| Cnidaria |  |  |  |  |  |  |
| Nematostella vectensi |  |  |  |  |  |  |
| Nematostella vectensi |  |  |  |  |  |  |
| Nematostella vectensi |  |  |  |  |  |  |

JBC
| Species                        | Locus   |
|-------------------------------|---------|
| Branchiostoma florida         |         |
| Cephalochorda                 |         |
| Botryllus schlosseri          |         |
| Molgula tectiformis_l         |         |
| Ciona savigny_gb1              |         |
| LGB                            |         |
| Ciona intestinalis_gb          |         |
| Tunica                        |         |
| Daphnia pulex_globinX          |         |
| Crassostrea gigas_1            |         |
| LGB                            |         |
| Crassostrea gigas_2            |         |
| Mollusca                      |         |
| Xenoturbella                  |         |
| Symsagittifera roscof         |         |
| Nemertodermatida               |         |
| Meara stichopi_ng1             |         |
| Tetrahalocystis jonesi         |         |
| Mytilus californianus          |         |
| Spisula solidissima_l          |         |
| Aplysia_1_lgb3_cnsn01          |         |
| Aplysia_2_lgb2_cnsn01          |         |
| Licophora japonica_n           |         |
| Crassostrea gigas_2_l          |         |
| Crassostrea gigas_1_l          |         |
| Artiophoda                     |         |
| Ciona intestinalis_gb          |         |
| Ciona savigny_gb1              |         |
| Molgula tectiformis_l          |         |
| Botryllus schlosseri           |         |
| Cephalochorda                  |         |
| Branchiostoma florida          |         |
| Taxon                          | Sequence                        |
|-------------------------------|---------------------------------|
| Branchiostoma florida        | DAQSEENQVH (M)                  |
| Branchiostoma florida        | (H)                            |
| Branchiostoma florida        | (M)                            |
| HEMICHORDATA                 |                                 |
| Saccoglossus kowalev         | NGDKEQDNCIE SSQKMDITVI ASTKQR---|
| Saccoglossus kowalev         |                                |
| ECHINODERMATA                |                                 |
| STRONGYLOCENTROTUS_PU        |                                 |
| KIMORRHYNCHA                 |                                 |
| ECHINODERES_HORNI_Glo        |                                 |
| ENTOPROCTA                   |                                 |
| PEDICELLINA_SP_LGB1_E        |                                 |
| TARDIGRADA                   |                                 |
| RICHTERSIUS_CORONIFER        |                                 |