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, 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

Manuscript Title:	Neuroglobins: pivotal proteins associated with emerging neural systems and precursors of metazoan globin diversity
Manuscript No:	JBC/2012/407601 [R1]
Manuscript Type:	Regular Paper
Date Submitted by the Author:	15 Nov 2012
Complete List of Authors:	Christophe Lechauve, Muriel Jager, Laurent Laguerre, Laurent Kiger, Gaelle Correc, Cedric Leroux, Serge Vinogradov, Mirjam Czjzek, Michael Marden, and Xavier Bailly
Keywords:	Animal models; Confocal microscopy; Crystal structure; Evolution; Hemoglobin ; Immunohistochemistry; Phylogenetics; metazoans; neuroglobin

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

Christophe Lechauve[1], Muriel Jager[2], Laurent Laguerre[3], Laurent Kiger[4], Gaëlle Correc[5], Cédric Leroux[3], Serge Vinogradov[6], Mirjam Czjzek[3, 5], Michael C Marden[4], and Xavier Bailly[3*]

[1]INSERM - UMR S 968. CNRS, UMR_7210, Institut de la Vision UPMC/CHNO des Quinze-Vingts.
17, rue Moreau - 75012 Paris. France

[2]Université Pierre et Marie Curie (Paris 06), UMR 7138 CNRS UPMC MNHN IRD, 7 quai St Bernard, 75005 Paris, France

[3]UPMC-CNRS, FR2424, Station Biologique de Roscoff, 29680, Roscoff, France.

[4]INSERM U779, 94275 Le Kremlin-Bicêtre, France

[5]UPMC-CNRS, UMR 7139, Marine Plants and Biomolecules, Station Biologique de Roscoff, 29680, Roscoff, France.

[6]Department 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 mechano-sensory 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 (O_2) 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 O_2 binding affinities resemble those of vertebrate Mbs and their function is considered to be O_2 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 acoel 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 acoel (Fig. 1A), thus occupying a phylogenetic position either preceding the deuterostome-protostome split or branching at the base of deuterostomes (23) (24). This hermaphroditic 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 immunocytochemical 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, mollusc 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 (O_2 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 N_2. 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 O_2 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% $MgCl_2$ 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 (2µg/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 (10mM 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 N_2-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 $P6_222$. These starting phases were used to build the initial model using ARP/*w*ARP 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. (*Symsagittifera 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 hmmalign 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 O_2 and CO binding and of O_2 dissociation are similar to those of vertebrate Mbs, and consequently so is its O_2 binding affinity (Table 1). The rate of heme autoxidation under pure O_2 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 O_2 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 O_2 (44). Overall, these observations are consistent with an *in vivo* function involving reversible binding of the diatomic ligand rather than a redox reaction with O_2 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 O_2 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 SrNbg 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 hymenopter *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 *Hellobdella 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 theses 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 meristoyilation 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. Sub-cellular 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 O_2 binding affinity is similar to that of Mb, is likely to be involved in O_2 storage, and thus provision of O_2 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 O_2 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 O_2 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 O_2 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 O_2 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., Kurtzhals, 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*

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., Plachetzki, 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., Shapiro, 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 *Symsagittifera 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 O_2 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 O_2 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 (107M), as obtained with the program ESPRIPT (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.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 hemerythin) 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. O_2 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. O_2 solubility coefficient was 1.82 x 10^{-6} moles/liter at 25 °C and for CO solubility coefficient was 1.36 x 10^{-6} moles/liter at 25 °C. ° O_2 affinity was estimated equal to 1.8 Torr.

FIGURE 1

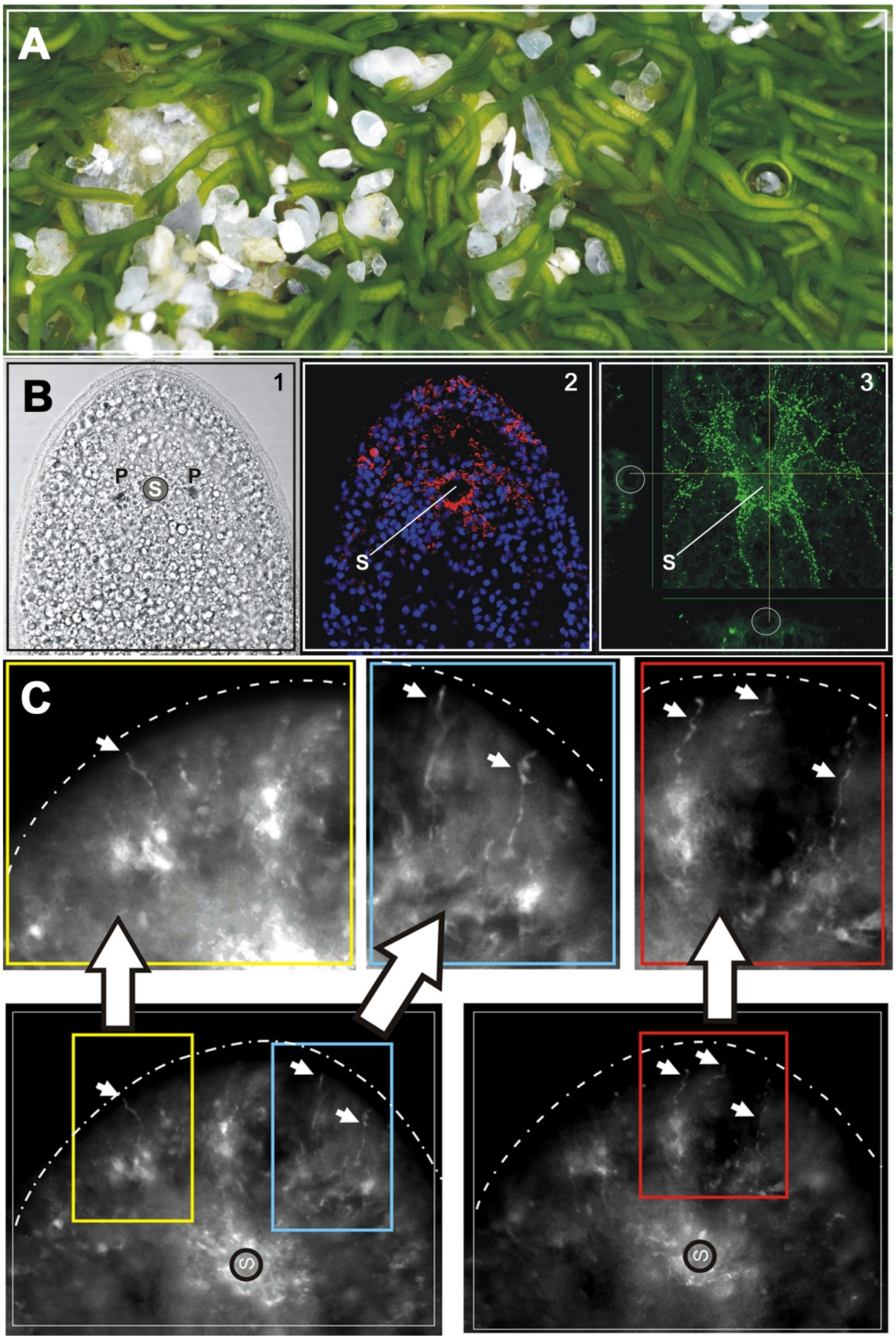

FIGURE 2

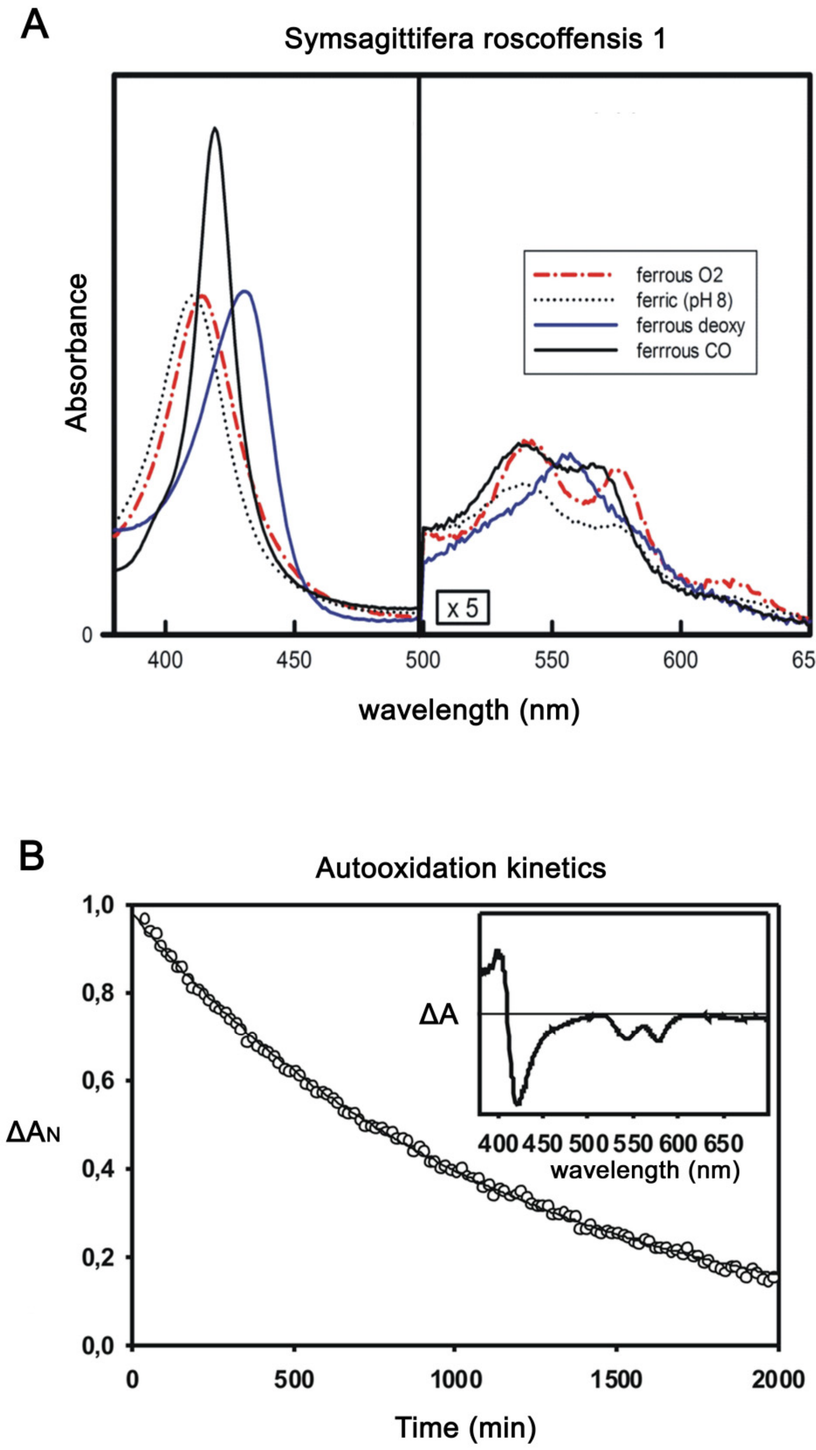
A
Symsagittifera roscoffensis 1
Absorbance
ferrous O2
ferric (pH 8)
ferrous deoxy
ferrrous CO
x 5
0
400
450
500
550
600
650
wavelength (nm)
B
Autooxidation kinetics
1,0
0,8
0,6
0,4
0,2
0,0
ΔAN
0
500
1000
1500
2000
Time (min)
ΔA
400
450
500
550
600
650
wavelength (nm)

FIGURE 3

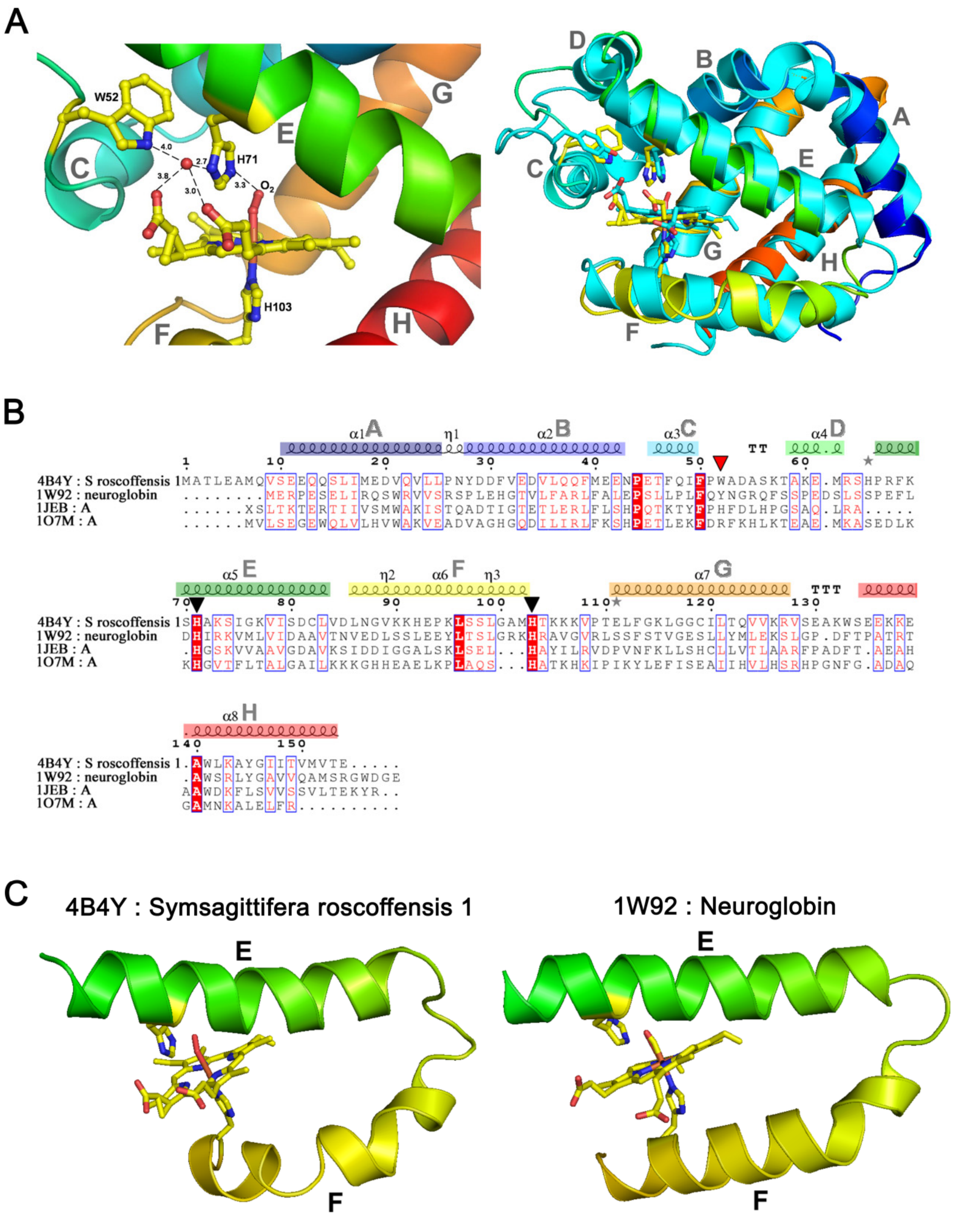
A
W52
4.0
2.7
H71
3.8
3.0
3.3
O_2
H103
C
E
G
F
H
D
B
A
C
E
G
H
F
B
α1 A
η1
α2 B
α3 C
TT
α4 D
1
10
20
30
40
50
60
4B4Y : S roscoffensis 1 MATLEAMQVSEEQQSLIMEDVQVLLPNYDDFVEDVLQQFMEENPETFQIFPWADASKTAKE.MRSHPRFK
1W92 : neuroglobin ........MERPESELIRQSWRVVSRSPLEHGTVLFARLFALEPSLLPLFQYNGRQFSSPEDSLSSPEFL
1JEB : A ......XSLTKTERTIIVSMWAKISTQADTIGTETLERLFLSHPQTKTYFPHFDLHPGSAQ.LRA.....
1O7M : A ......MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAE.MKASEDLK
α5 E
η2
α6 F
η3
α7 G
TTT
70
80
90
100
110
120
130
4B4Y : S roscoffensis 1 SHAKSIGKVISDCLVDLNGVKKHEPKLSSLGAMHTKKKVPTELFGKLGGCILTQVVKRVSEAKWSEEKKE
1W92 : neuroglobin DHIRKVMLVIDAAVTNVEDLSSLEEYLTSLGRKHRAVGVRLSSFSTVGESLLYMLEKSLGP.DFTPATRT
1JEB : A .HGSKVVAAVGDAVKSIDDIGGALSKLSEL...HAYILRVDPVNFKLLSHCLLVTLAARFPADFT.AEAH
1O7M : A KHGVTFLTALGAILKKKGHHEAELKPLAQS...HATKHKIPIKYLEFISEAIIHVLHSRHPGNFG.ADAQ
α8 H
140
150
4B4Y : S roscoffensis 1 .AWLKAYGIITVMVTE.....
1W92 : neuroglobin .AWSRLYGAVVQAMSRGWDGE
1JEB : A AAWDKFLSVVSSVLTEKYR..
1O7M : A GAMNKALELFR..........
C
4B4Y : Symsagittifera roscoffensis 1
E
F
1W92 : Neuroglobin
E
F

FIGURE 4

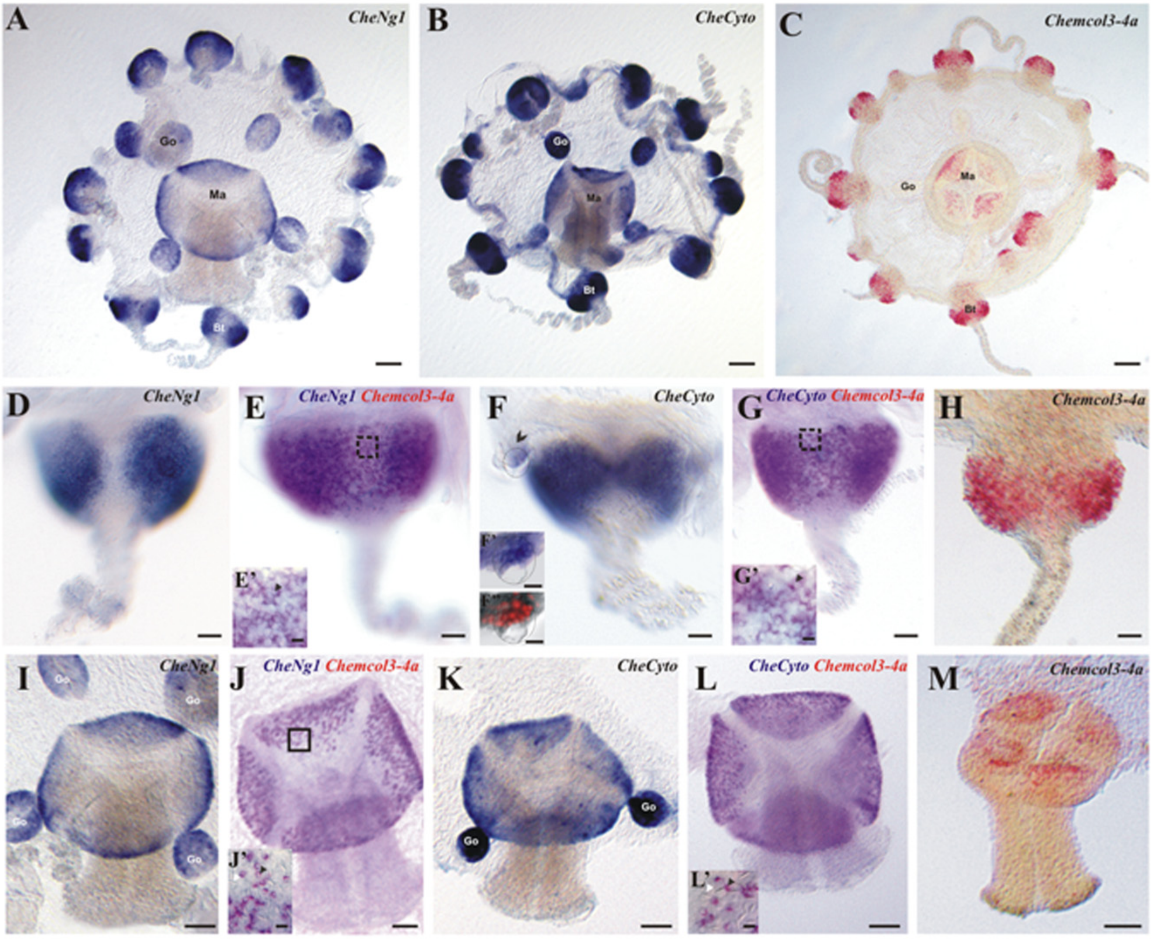
A
CheNg1
Go
Ma
Bt
B
CheCyto
Go
Ma
Bt
C
Chemcol3-4a
Go
Ma
Bt
D
CheNg1
E
CheNg1 Chemcol3-4a
E'
F
CheCyto
F'
G
CheCyto Chemcol3-4a
G'
H
Chemcol3-4a
I
CheNg1
Go
Go
Go
Go
J
CheNg1 Chemcol3-4a
J'
K
CheCyto
Go
Go
L
CheCyto Chemcol3-4a
L'
M
Chemcol3-4a

FIGURE 5

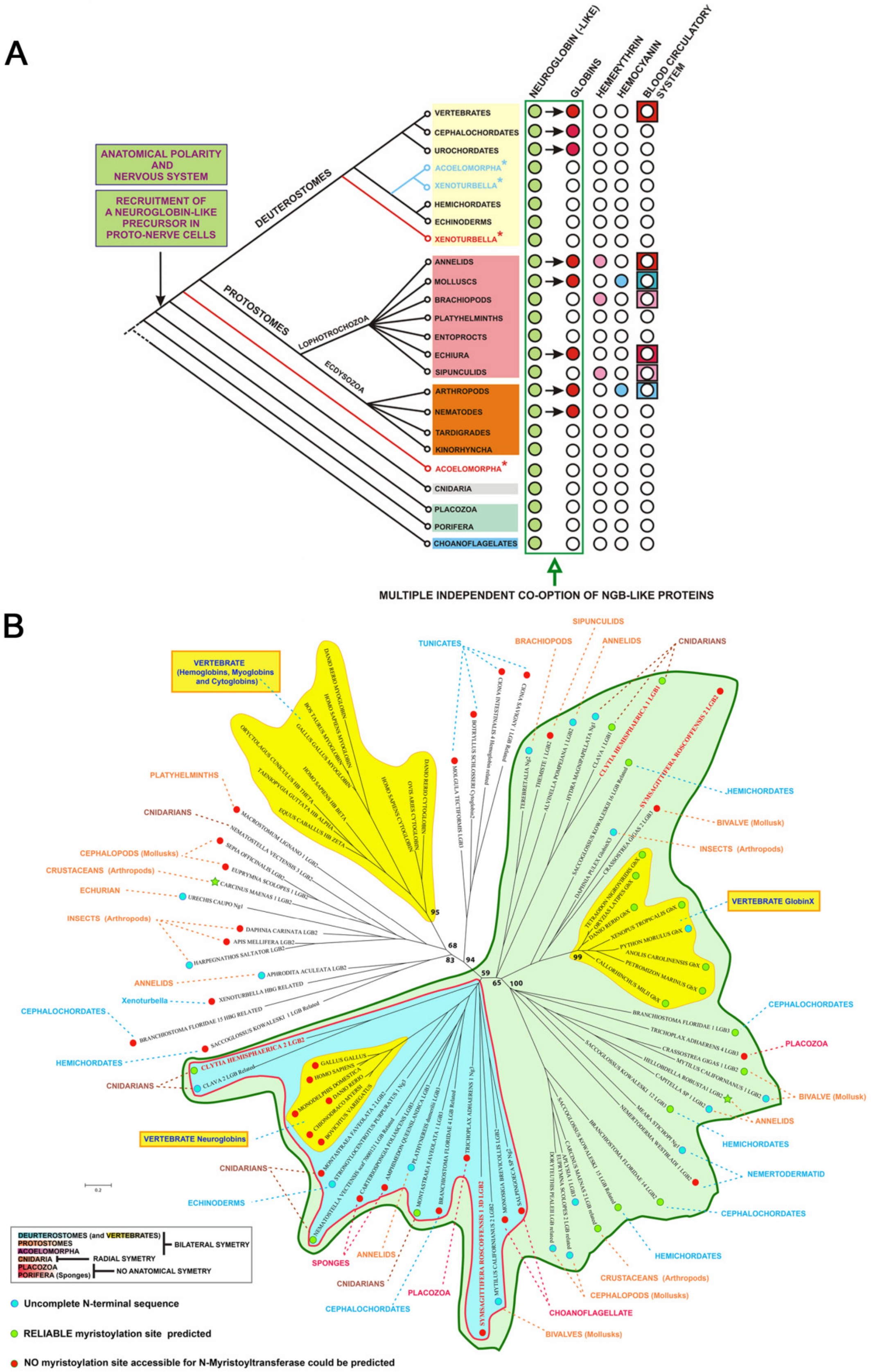
A
NEUROGLOBIN (-LIKE)
GLOBINS
HEMERYTHRIN
HEMOCYANIN
BLOOD CIRCULATORY SYSTEM
ANATOMICAL POLARITY AND NERVOUS SYSTEM
RECRUITMENT OF A NEUROGLOBIN-LIKE PRECURSOR IN PROTO-NERVE CELLS
DEUTEROSTOMES
PROTOSTOMES
LOPHOTROCHOZOA
ECDYSOZOA
VERTEBRATES
CEPHALOCHORDATES
UROCHORDATES
ACOELOMORPHA*
XENOTURBELLA*
HEMICHORDATES
ECHINODERMS
XENOTURBELLA*
ANNELIDS
MOLLUSCS
BRACHIOPODS
PLATYHELMINTHS
ENTOPROCTS
ECHIURA
SIPUNCULIDS
ARTHROPODS
NEMATODES
TARDIGRADES
KINORHYNCHA
ACOELOMORPHA*
CNIDARIA
PLACOZOA
PORIFERA
CHOANOFLAGELATES
MULTIPLE INDEPENDENT CO-OPTION OF NGB-LIKE PROTEINS
B
VERTEBRATE (Hemoglobins, Myoglobins and Cytoglobins)
VERTEBRATE GlobinX
VERTEBRATE Neuroglobins
TUNICATES
BRACHIOPODS
SIPUNCULIDS
ANNELIDS
CNIDARIANS
PLATYHELMINTHS
CNIDARIANS
CEPHALOPODS (Mollusks)
CRUSTACEANS (Arthropods)
ECHURIAN
INSECTS (Arthropods)
ANNELIDS
Xenoturbella
CEPHALOCHORDATES
HEMICHORDATES
CNIDARIANS
HEMICHORDATES
BIVALVE (Mollusk)
INSECTS (Arthropods)
CEPHALOCHORDATES
PLACOZOA
BIVALVE (Mollusk)
ANNELIDS
HEMICHORDATES
NEMERTODERMATID
CEPHALOCHORDATES
HEMICHORDATES
CRUSTACEANS (Arthropods)
CEPHALOPODS (Mollusks)
CHOANOFLAGELLATE
BIVALVES (Mollusks)
PLACOZOA
CEPHALOCHORDATES
CNIDARIANS
ANNELIDS
SPONGES
ECHINODERMS
CNIDARIANS
CLYTIA HEMISPHAERICA 1 LGB1
SYMSAGITTIFERA ROSCOFFENSIS 2 LGB2
CLYTIA HEMISPHAERICA 2 LGB2
SYMSAGITTIFERA ROSCOFFENSIS 1 3D LGB2
95
68
83
94
59
65
100
99
0.2
DEURTEROSTOMES (and VERTEBRATES)
PROTOSTOMES
ACOELOMORPHA
CNIDARIA
PLACOZOA
PORIFERA (Sponges)
BILATERAL SYMETRY
RADIAL SYMETRY
NO ANATOMICAL SYMETRY
● Uncomplete N-terminal sequence
● RELIABLE myristoylation site predicted
● NO myristoylation site accessible for N-Myristoyltransferase could be predicted

TABLE 1

	SrNgb1	Human Mb	human Ngb (25°C)
$k_{on}CO$ (/µM/s)	0.35	0.65	40
$k_{on}O_2$ (/µM/s)	7	15	170
$k_{off}O_2$ (/s)	35+/-5	27	0.7
K O_2 (µM)	5.0	1.8	0.004
P_{50} O_2 (Torr)	**2.8**	**1**	**6.8**
$k_{on}His$ (/s)			1800
$k_{off}His$ (/s)			0.6
K His			3000

TABLE S1 Data collection, phasing and refinement statistics on globin crystals, space group $P6_222$

Beamline at ESRF	ID23-I	BM30A
Wavelength (Å)	0.93	1.7389
Unit cell parameters in Å and degrees (°)	a=b=97.07, c=140.10, α=β=90, γ=120	a=b=97.07, c=140.10, α=β=90, γ=120
Resolution range (Å)	30.06–2.3 (2.36-2.30)[a]	50.06–3.2 (3.29-3.20)[a]
No. of observations	57993 (3605)	60276 (3599)
No. of unique reflections	12137 (688)	11480 (675)
Completeness (%)	91.7 (77.4)	93.3 (77.4)
$<I/\sigma(I)>$	16.6 (2.2)	14.9 (2.0)
Redundancy	4.8 (5.2)	5.2 (5.3)
R_{sym} (%)[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		
R_{cryst} (%)	21.4	
R_{free} (%)[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] $R_{sym} = \Sigma \mid I-I_{av} \mid / \Sigma \mid I \mid$, where the summation is over all symmetry-equivalent reflections.

[c] R_{free} 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)
-------------------------------------MGCALSGAQDPPVSEECSPLDDG
LDLNRETTLGSNGRTTEPFPLSGAQKELIRGSWEILHK---DIARVGIIVFIRLFETHPE
-CKDVFFLFR-D--ID--DFQQLKMSKELQAHGLRVMSFIEKSVARM---DQ-EPKLHHL
AFELGRSHCRYK----APPKYYEYIGIQFIQAAQPILKEAWT-PETEKAWEGLFQYLAAT
MRRGFYKEQKATGKN---------------------------------------------
----------------------
>CALLORHINCHUS MILII GbX (from Droge et al 2011)
------------------------------------------------------MGCAIS
GPGQYPASGREDVVAVASLSLSDRQTQLVKETWRLVQE---DIAKVGIIMFVRLFETHPE
-CKDAFFLFR-D--ID--DLQQLRKSKGLRAHGLRVMSFIEKTVARL---DQ-EDRLQQL
XLELGKSHFRYS----AAPKYYPYVGNEFICAVQPILKEKWT-AEVEEAWKGLFHYLTSV
MKKGYQDEERGSCPREKPKHGPNSV-----------------------------------
----------------------
>DANIO RERIO GbX (from Droge et al 2011)
-----------------------------------------------------MGCAISG
SGLTAGAPEIRPGEEETPAGLTTNHIRLIKESWRLIQE---DIAKVGIIMFVRLFETHPE
-CKDVFFLFR-D--VE--DLERLRTSRELRAHGLRVMSFIEKSVARL---DQ-LERLETL
ALELGKSHYRYN----APPKYYGYVGAEFICAVRPILKDRWT-PELEEAWKTLFQYVTSI
MREGFLEEERNKRSNTQTSSRERPDKRSTAI-----------------------------
----------------------
>ORYZIAS LATIPES GbX (from Droge et al 2011)
------------------------------------------------------MGCAIS
GLAAKTDLAERSREDAAVEHPNEEQIQMIKDSWKVIRD---DIAKVGIIMFVRLFETHPE
-CKDVFFLFR-D--VE--DLERLRTNRELRAHGLRVMSFIEKSVARL---DQ-PERLEAL
AVELGKSHYHYN----APPKYYNYVGAEFICAVQPILKEQWT-TELEKAWQTLFQFVTAL
MKQGYQEE-SARQRQLA-TSPKDRLDKRNTAL----------------------------
----------------------
>PETROMIZON MARINUS GbX (from Droge et al 2011)
--------------------------------MGCTVSTDERTGAQSSSQGQSQASRKQQ
QPEQQRAAGEGHQPPGPPQAPSESQRRLVRDSWLALQC---DIARVGVIMFVRLFETHPE
-CKDVFYQFR-D--CE--DLQKLKMNKQLQAHGLRVMSFIEKSVARL---EQ-ECVLEQL
IVEMGRKHYKYN----ASPKYYSFVGIEFIATVQPFLQEKWT-NEVEDAWQCLFRYIAAV
MKRGYLEEEAASNGVNTANYDRGQGNHGATAM----------------------------
----------------------
>PYTHON MORULUS GbX (from Droge et al 2011)
------------------------------------------------------------
------------------------QKELIRESWKILHK---NITRVGIIVFIRLFETHPE
-CKDVFFLFR-D--ID--DLQQLKMNKELQAHGLRVMSFIEKSVARL---DQ-EGKLEVL
AFELGRSHFRYK----APPNYYEYIGIQFIQAVQPILKEDWT-LEVEKAWK---------
------------------------------------------------------------
----------------------
```

```
>TETRAODON NIGROVIRIDIS GbX (from Droge et al 2011)
------------------------------------------------MGCAISSLGAKA
EFGDRSAEEEDAAAAAAVVYPREDQIQMIKDSWKVIRD---DIAKVGIIMFVRLFETHPE
-CKDVFFLFR-D--VE--DLERLRSSRELRAHGLRVMSFIEKSVARL---DQ-QDRLEAL
AVELGKSHYHYN----APPKYYSYVGAEFICAVQPILKERFT-SELEEAWKTLFQYVTGL
MRKGHQEEGSRQRHLALPPKDGPEKRTSAL------------------------------
----------------------
>XENOPUS TROPICALIS GbX (from Droge et al 2011)
---------------------------------------------------MGCILSSLG
WQWRDSLDHTETSPLLPTLNLSEQQQLLVESWRLIQH---DIAKVGVILFVRLFETHPE
-CKDVFFLFR-D--VD--DLQALRANKDLRAHGLRVLSFVEKSVARI---AD-CARLEEL
ALELGRSXYRYN----APPRYYQYVGTEFISAVCPMLHDKWT-AEVEEAWKGLFAYICTV
MERGYQEEE-RRHSDGRSLIDGLQGNKGLI------------------------------
----------------------
>VERTEBRATE MYOGLOBINS
----
>HOMO_SAPIENS_MYOGLOBIN_gi|4885477|ref|NP_005359.1|
------------------------------------------------------------
------------------MGLSDGEWQLVLNVWGKVEA---DIPGHGQEVLIRLFKGHPE
-TLEKFDKFK-H--LK--SEDEMKASEDLKKHGATVLTALGGILKK-------KGHHEAE
IKPLAQSHATKHK---IPVKYLEFISECIIQVLQSKHPGDFG-ADAQGAMNKALELFRKD
MASNYKELGFQG------------------------------------------------
----------------------
>DANIO_RERIO_MYOGLOBIN_gi|41053652|ref|NP_956880.1|
------------------------------------------------------------
----------------------MADHDLVLKCWGAVEA---DYAANGGEVLNRLFKEYPD
-TLKLFPKFS-G--IS---QGDLAGSPAVAAHGATVLKKLGELLKA-------KGDHAAL
LKPLANTHANIHK---VALNNFRLITEVLVKVMAEKAGLDAA-GQGALRRVMDAVIGDID
GYYKEIGFAG--------------------------------------------------
----------------------
>BOS_TAURUS_MYOGLOBIN_gi|27806939|ref|NP_776306.1|_
------------------------------------------------------------
------------------MGLSDGEWQLVLNAWGKVEA---DVAGHGQEVLIRLFTGHPE
-TLEKFDKFK-H--LK--TEAEMKASEDLKKHGNTVLTALGGILKK-------KGHHEAE
VKHLAESHANKHK---IPVKYLEFISDAIIHVLHAKHPSDFG-ADAQAAMSKALELFRND
MAAQYKVLGFHG------------------------------------------------
----------------------
>GALLUS_GALLUS_MYOGLOBIN_gi|268607704|ref|NP_001161224.1|_
------------------------------------------------------------
------------------MGLSDQEWQQVLTIWGKVEA---DIAGHGHEVLMRLFHDHPE
-TLDRFDKFK-G--LK--TPDQMKGSEDLKKHGATVLTQLGKILKQ-------KGNHESE
LKPLAQTHATKHK---IPVKYLEFISEVIIKVIAEKHAADFG-ADSQAAMKKALELFRND
MASKYKEFGFQG------------------------------------------------
----------------------
>VERTEBRATE CYTOGLOBINS

>DANIO_RERIO_CYTOGLOBIN_gi|23308615|ref|NP_694484.1|_
------------------------------------------------------------
-----MEGDGGVQLTQSPDSLTEEDVCVIQDTWKPVYA---ERDNAGVAVLVRFFTNFPS
-AKQYFEHFR-E--LQ--DPAEMQQNAQLKKHGQRVLNALNTLVENL---RD-ADKLNTI
FNQMGKSHALRHK---VDPVYFKILAGVILEVLVEAFPQCFSPAEVQSSWSKLMGILYWQ
MNRVYAEVGWEN------------------------------------------------
----------------------
>OVIS_ARIES_CYTOGLOBIN_gi|261244962|ref|NP_001159664.1|_
------------------------------------------------------------
--MEKVPGEMEIERRERSEELSEAERKAVQATWARLYA---NCEDVGVAILVRFFVNFPS
-AKQYFSQFK-H--ME--EPLEMERSPQLRKHACRVMGALNTVVENL---HD-PEKVSSV
LALVGKAHALKHK---VEPVYFKILSGVILEVIAEEFASDFP-PETQRAWAKLRGLIYSH
VTAAYKEVGWVQ------------------------------------------------
----------------------
>HOMO_SAPIENS_CYTOGLOBIN_gi|20987458|gb|AAH29798.1|_
```

```
------------------------------------------------------------
--MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYA---SCEDVGVAILVRFFVNFPS
-AKQYFSQFK-H--ME--DPLEMERSPQLRKHACRVMGALNTVVENL---HD-PDKVSSV
LALVGKAHALKHK---VEPVYFKILSGVILEVVAEEFASDFP-PETQRAWAKLRGLIYSH
VTAAYKEVGWVQ------------------------------------------------
----------------------
>VERTEBRATE HEMOGLOBINS

>EQUUS_CABALLUS_HB_ZETA_gi|167621441|ref|NP_001108014.1|_hemoglobin_subunit
_zeta_
------------------------------------------------------------
------------------MSLTKAERTMVVSIWGKISM---QADAVGTEALQRLFSSYPQ
-TKTYFPHF-----------DLHEGSPQLRAHGSKVAAAVGDAVKSI-------DNVAGA
LAKLSELHAYILR---VDPVNFKFLSHCLLVTLASRLPADFT-ADAHAAWDKFLSIVSSV
LTEKYR------------------------------------------------------
----------------------
>HOMO_SAPIENS_HB_BETA_gi|256028940|gb|ACU56984.1|_beta-globin_
------------------------------------------------------------
-----------------MVHLTPEEKSAVTALWGKV-----NVDEVGGEALGRLLVVYPW
-TQRFFESFG-D--LS--TPDAVMGNPKVKAHGKKVLGAFSDGLAHL-------DNLKGT
FATLSELHCDKLH---VDPENFRLLGNVLVCVLAHHFGKEFT-PPVQAAYQKVVAGVANA
LAHKYH------------------------------------------------------
----------------------
>TAENIOPYGIA_GUTTATA_HB_ALPHA_gi|323668297|ref|NP_001191172.1|_hemoglobin_s
ubunit_alpha-A_
------------------------------------------------------------
------------------MVLSAGDKSNVKAVFGKIGG---QADEYGADALERMFATYPQ
-TKTYFPHF-----------DLGKGSAQVKGHGKKVAAALVEAANNV-------DDLAGA
LSKLSDLHAQKLR---VDPVNFKLLGQCFLVVVATRNPSLLT-PEVHASLDKFLCAVGTV
LTAKYR------------------------------------------------------
----------------------
>ORYCTOLAGUS_CUNICULUS_HB_THETA_gi|284005393|ref|NP_001164887.1|_theta_1_gl
obin_
------------------------------------------------------------
------------------MALSAAERALLRALWKKLGS---NVGVYATEALERTLEAFPR
-TKIYFSHM-----------DLSPGSAQVRAHGRKVADALTLAADHL-------DDLPGA
LSALSDLHVRTLR---VDPHHFGLLGHCLLVTLARHYPGDFG-PAMHASVDKFLHHVISA
LTSKYR------------------------------------------------------
----------------------
>VERTEBRATE NEUROGLOBINS

>HOMO_SAPIENS_gi|10864065|ref|NP_067080.1|_neuroglobin_[Homo_sapiens]
------------------------------------------------------------
-----------------------P--ELIRQSWRAVSR---SPLEHGTVLFARLFALEPD
-LLPLFQYNC-R--QFS-SPEDCLSSPEFLDHIRKVMLVIDAAVTNV---ED-LSSLEEY
LASLGRKHRAVG----VKLSSFSTVGESLLYMLEKCLGPAFT-PATRAAWSQLYGAVVQA
MSRGWDG-----------------------------------------------------
----------------------
>DANIO_RERIO_gi_18859087_ref_NP_571928.1__neuroglobin_[Danio_rerio]
------------------------------------------------------------
-------------------KLSEKDKGLIRDSWESLGK---NKVPHGIVLFTRLFELDPA
-LLTLFSYST-N--CG--DAPECLSSPEFLEHVTKVMLVIDAAVSHL---DD-LHTLEDF
LLNLGRKHQAVG----VNTQSFALVGESLLYMLQSSLGPAYT-TSLRQAWLTMYSIVVSA
MTRG--------------------------------------------------------
----------------------
>GALLUS_GALLUS_gi_154816290_gb_ABS87379.1__neuroglobin_[Gallus_gallus]
------------------------------------------------------------
-------------------MLSRTQQALIRESWRRVSG---SPVQHGVVLFSRLFDLDPD
-LLPLFQYNC-K--RFA-SPQECLAAPEFLDHIRKVMLVIDAAVSHL---ED-LPCLEEY
LCNLGKKHQAVG----VKVESFSTVGESLLYMLENCLGAAFS-PDVREAWIELYGAVVKA
MQR---------------------------------------------------------
```

```
----------------------
>MONODELPHIS_DOMESTICA_gi_78486584_ref_NP_001030592.1__neuroglobin_[Monodel
phis_domestica]
------------------------------------------------------------
------------------RRLSGPEQELIRESWQKVNS---NPLQHGMILFTRLFDLEPD
-LLPLFQYNC-R--QFS-SPQDCLSSPEFLDHIRKVMLVIDAAVTHV---EN-LSSLEEY
LTNLGKKHKAVG----VKLSSFSTVGESLLYMLEQCLGSTFT-VTMKEAWTQLYGAVVQA
MSRGW-------------------------------------------------------
----------------------
>CHIONODRACO_MYERSI_gi|270341071|emb|CAR57914.1|_neuroglobin_[Chionodraco_m
yersi]
------------------------------------------------------------
-------------------KLSGKDKELIRGSWESLGK---NKVPHGVVMFSRLFELDPE
-LLTLFHYTT-N--CG--STQDCLSSPEFLEHVTKVMLVIDAAVSHL---DD-LPSLEDF
LLNLGRKHQAVG----VNTQSFAEVGESLLYMLQCSLGQAYT-APLRQAWLNLYSIVVAA
MM----------------------------------------------------------
----------------------
>BOVICHTUS_VARIEGATUS_gi_270341067_emb_CAR57912.1__neuroglobin_[Bovichtus_v
ariegatus]
------------------------------------------------------------
-------------------KLSGKDKELIRGSWESLGK---NKVPHGVVMFSRLFELDPE
-LLTLFHYTT-N--CG--STQDCLSSPEFLEHVTKVMLVIDAAVSNL---DD-LPSLEDF
LLNLGGKHQAVG----VNTQSFAEVGESLLYMLQCSLGQAYT-APLRQAWLNLYSIVVAA
MSR---------------------------------------------------------
----------------------
>SIPUNCULA

>THEMISTE_1_LGB2_gnl|ti|2001595631_Tl_MM2_10H06__From_transcriptome_availab
le_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
----------MNQAPEDQTKFTPDELMALRLSWNAWVA--GDLAGXGFDMFVKMFEQRKE
-TKEVFPFAR-G--N---DAXQMHQIINVLFHVTKFMKNIDEVVKNA---DR-LEDVVSM
LRRVEGRHGHQGHN--VPSAYFPFLGAAMH-TLIKANYKSYD-SKLDDCWVPLWNFMNNE
MTTGQEVYRGGKI-----------------------------------------------
----------------------
>THEMISTE_2_LGB2_gnl|ti|2001595657_Tl_MM2_11B08_From_transcriptome_availabl
e_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
----------------------------LRLSWNAWVA--GDLVGRGFDSFVKMFEQKKE
-TKEVFSFAR-G--N---DAKQMQQSSNVLFHVTRVMKNIDEVVKHA---DR-LEDVVSM
LRQVGGRHGHQGHN--VPSAYFPYLGEAMR-TLIKANYKAYD-SKLDDCWVRLWDFINKQ
MTTGQEYTEEEKS-----------------------------------------------
----------------------
>THEMISTE_3_LGB2_gnl|ti|2001594772_Tl_MM2_01H11_From_transcriptome_availabl
e_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
----------MNLDLEDQCRFDPEELINLRLSWNAWVA--GDLPGRGFDMFTKMFEQRKE
-TKEVFSFVR-G--S---DAHQMQQSSKFLFHVTRVMKNIDEVVKHA---DR-LEDVVSM
LRQVGGRHGSSHHN--VPSAYFPFLGIALR-NLMSQSYKGYD-SKLDDAWTRLWGFMTSQ
MMYG--------------------------------------------------------
----------------------
>BRACHIOPODA

>TEREBRETALIA_Ng2_gnl|ti|1987128689_Tt_MM1_02C05__From_transcriptome_availa
ble_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
--------------------FTTQEKNRLRSSWRAFVA--EDPAKRGYQMFMKMFENRPD
-TKKFFK-FV-Q--KD--ALGTMKSSTSLVFHTGRVAKYIQLVCDCL---DD-PTEAVPL
LRQLGGKHGCGPNNFNVDATYFPELGKAMRQLMESSGLQGYD-QELDALWEKLYTWITKR
MEEGM-------------------------------------------------------
----------------------
```

```
>ANNELIDA

>ALVINELLA_POMPEJANA_1_LGB2_gi|301587212|gb|FP491105.1|FP491105_FP491105_Al
vinella_pompejana_whole_body_library_Alvinella_pompejana_cDNA_clone_MH0AAB5
9YJ14_5',_mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nuc
est/
------------------------------------------------------------
----------------------------------YRLS---DISERGMDIFVRLFELHPV
-YKSYFQKLR-D--V---DIEDLRQSGKLRVHSTSVMKSITDLVETL---DH-PPDLRDM
AIKIAHPHFDRG----VRPSQYRELFAIILEYLKDKAKVVFN-DEAEAAWQKLFDYVLDI
TAAVMDLQIEKMG-----------------------------------------------
----------------------
>ALVINELLA_POMPEJANA_2_LGB2_gi|223839268|gb|GO218263.1|GO218263_CAGB27239.r
ev_CAGB_Alvinella_pompejana_Normalized_library_RN05_posterior_end_Alvinella
_pompejana_cDNA_clone_CAGB27239_3',_mRNA_sequence_From_ESTs_stored_at_http:
//www.ncbi.nlm.nih.gov/nucest/
------------------------------------------------------------
---------MASYKPDPRCPLTERQLYSITKSWKAINR---EMASTAVNMFIRLLEHDGI
--RSFFTKFK-D--HK--TVAELRASKVFESHALMVISVIDDVITNL---DD-MDYVMSL
LQATGESHSIKFKN--FNPDFLWNVEGAFLWAVKETLGDRYT-ISIENIYTITIRYILQS
LHDAFTKHRERQNSTNNDCEKTNLLNQELSTADRKT------------------------
----------------------
>APHRODITA_ACULEATA_LGB2_gi|1491803|gb|AAC47259.1|nerve myoglobin
[Aphrodita aculeata]
------------------------------------------------------------
-------------------glsgaDIAVIRSTWAKVQG-sgSATDIGRSIFIKFFELDPA
-AQNEFPCKG-E------SLAALKTNVLLGQHGAKFMEYITTAVNGL---dDYAGKAHGP
LTELGSRHKTRG----TTPANFGKAGEALlailasvvggdft-paakdawtkvyntisst
mqa---------------------------------------------------------
----------------------
>CAPITELLA_SP_1_LGB2_From_transcriptome_available_at_ftp://ftp.ncbi.nlm.nih
.gov/pub/TraceDB/
------------------------------------------------------------
-----------------DQLLTPEEIVLVRVTWEQLKT-nlTLANLGKKVFLRIFNLKPD
-IKKLFPFSD-V------WGDDLIRHPKFVLHSERFMLVVDCCVQNL---ECIKSEhGEM
LANLGRAHVNYkG---FSRENFEVFMKAIWYVWYHQLKDSMD-SEVECAWKKLLLfiivq
qragydaekeappnglsfllqt
>CAPITELLA_SP_2_LGB2_From_transcriptome_available_at_ftp://ftp.ncbi.nlm.nih
.gov/pub/TraceDB/
----------------------------------------------------mgnaqpfv
scaggqppdqkpsnipqHEFLTQNQKLGAKSTWEFLCH-tsTPTERGMRVFLRIFEIAPV
-TKTLFPFKD-M------PNEDLHRNSLFKGHATRFMKSVEFTMQNL---DALDVIvNPT
LVSIGNKHVHiKG---FHPDYLDTFQTALMDIWDEDLGKKFS-KETKEAWIKIFALITRK
VFEGfqeettrfrpplpyegkq
>CAPITELLA_SP_3_LGB3_gnl|ti|1068987169_BGYZ80398.g1_From_transcriptome_avai
lable_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
-------------------LFTDRQKAIITKTWRHMGN---DLTGRGSKVFLKIFNLHPE
-VKQLFPSLK-N--D---NEDQLLKNPCFRGHASRFMQSVGAVVENL---DS-PGDLSPL
LIDLGRKHVLFGG---FTPEYFAAFTEGMMCIWSEELGKGFT-DEVSVAWKTVFDFIMSQ
LQDGYA----------------
>PLATHYNEREIS_dumerilii_LGB3_454_sequencing_of_a_normalized_cDNA_library_fr
om_Platynereis_heads_and_mixed_larval_stages)_(normalization_and_sequencing
_was_done_by_the_company_Agowa)_Kindely_provided_by_Kristin_Tessmar_Raible_
from_Max_F._Perutz_Laboratories/_University_of_Vienna,_Austria.
------------------------------------------------------------
-------------------PLNDNQKELIKKLWEIVRE--GDIESTGVSMYMHMFTLQPE
-VMKMFSFVP-K--NVT-NPEDLKNSARFLRHARNLMTTVTTAVDML---GD-MEGLSEV
LVDLGRRHKKYR----AKTEHFPIVGRSLTHAISAATGDAFT-PETAEAFTAFYGVVTH-
------------------------------------------------------------
----------------------
```

```
>HELLOBDELLA_ROBUSTA1_LGB2_Helobdella_robusta_leech_jgi|Helro1|171404|fgene
sh4_pg.C_scaffold_20000312_From_transcriptome_available_at_ftp://ftp.ncbi.n
lm.nih.gov/pub/TraceDB/
------------------------------------------------------MGANGF
KKVKSEPLLNNLNYLNNDDVLTIREKQLVRESWTLLSI---KLKSLGKQVFLRIFELRPS
-TKNLFPFKT-V------WGDKLIKHPLFLTHSKRFVKVIGCVVDRL---DYLQEECAQP
LIELGKKHVSIEG---FLPDYYDVYIRAIISIWKQELKDVYT-NELSEAWHKVLVYIVSK
LKEGYETEWKVATYFNPQ------------------------------------------
----------------------
>ECHIURA

>URECHIS_CAUPO_Ng1_gnl|ti|1987440364_Uc_MM1_01H01_From_transcriptome_availa
ble_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
--------------------VSLQNIALVLKSWAILKK---DLLSNGAALFLALFEAYPD
-YKDLFKQFP-G--L---RLEDLTRMPPLRALGATFMHSLGSMVDNL---GD-LECVVEL
LRERTHTHWERG----IRFEHFQNVFDLLPAFLKSKLGYNFD-DATGAACAAASVMITVL
QAELKTL-----------------------------------------------------
----------------------
>PLATYHELMINTHES

>MACROSTOMUM_LIGNANO_1_LGB2_Ml_aW_013_G17_c_From_transcriptome_available_at
_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
-------------------MLTADETQLILSGWNQAMK---DAKGLGLDIFLTLFEMFPQ
-HQELFrDFK-G--K---SRAELEKMPKLRAHGLRVVNTLDGAIQSL---DD-MEVCVSS
LELIGASHKGRN----MNAGHFEDLNKALDVVFTRRLGAAYT--DNKAVWMKLlegvipv
Iqrgm-------------------------------------------------------
----------------------
>MACROSTOMUM_LIGNANO_2_LGB2__Ml_aW_008_G17_c_From_transcriptome_available_a
t_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
-------------------MLNEVEKKIILSGWQQAIK---DKKALGMDVFMTLFEMFPQ
-HQELFrDFK-G--K---SRAELEKMPKMRAHALRVVNTLDGAIQSL---DD-MEVCASS
LELIGASHKSHH----LSAKHFEDLNAALAVVFERRLGKAFV--DNKAVWVKLLQGIipv
Iqrgl-------------------------------------------------------
----------------------
>CHOANOFLAGELLATE

>MONOSIGA_BREVICOLLIS_LGB2_gi|167520949|ref|XP_001744813.1|_hypothetical_pr
otein_[Monosiga_brevicollis_MX1]
----------------------------------------------------------MS
DRRRSSGSSEDEAHAPDYPSFDPVAVKLARKQWKRVVQ---LVPNWHEVFFSYLFERAPY
-ARTLFPFD----------VDRLQGNSSLAEHAKRVGQALETALQGL---FE-YYSLVEV
LEKLGRRHFKYG----VEPEHIDLFEETFYKTLAIGLGKKWN-PEARRAWEIVCGLILSP
IRTGILQARTKANHLRAKEAERKRQLEMAAARLEGRVASSGVQFSSDTERSRRSTAASAT
ATPHCGLKSNRFSSNSLRKTIL
>SALPINGOECA_SP_Ng2_gi|326430027|gb|EGD75597.1|hypothetical protein
PTSG_06664 [Salpingoeca sp. ATCC 50818]
------------------------------------------------------------
------------------MRLDMEQLKIALGSWTAVVE---LVPTWHEVFFAELFQAHPE
-TERLLYSS-------------DKSKSWNERHMARVGKSVGDVIKSL---SN-YDDVIEH
LTALGTRHARYG----LHVDQLDLFINAFLWTLGAGLGDSWD-HSVKKAWMHVLPFILSP
LKSGLVVARTLRNDYNTSGCLRCRRLLIPLHGRRLRPITVSLV-----------------
----------------------
>PLACOZOA

>TRICHOPLAX_ADHAERENS_1_Ng3_fgeneshTA2gi|196007506|ref|XP_002113619.1|hypot
hetical protein TRIADDRAFT_57230 [Trichoplax adhaerens]
------------------------------------------------------------
----MDQAQTDSVQTPPQPSLTEEQKAIIRENWQDVEE---NMSEVGLYLFSKLFTIAPE
```

```
-YREVFPFE------------TTTDNVRLRVHATGVMKTVGKAVQNL---DQ-FSELQSA
LSTLGQFHHRKA----IKFENFQAVGQALIQTLSDKLQENFT-PEVHEAWSKTFDMITAA
MKSGMN------------------------------------------------------
----------------------
>TRICHOPLAX_ADHAERENS_2_LGB3_gi|196012120|ref|XP_002115923.1|hypothetical
protein TRIADDRAFT_59832 [Trichoplax adhaerens]
------------------------------------------------------------
------------------MAPTAQDLQTIRETWALVAP---DLKKHGTVLFLRLFEQHPD
-VQRLFEKIK-D--V---PHDQLATNENFVFHTTRVMETIDHAVKGI---DN-LPALTVL
LKQLGSSHAQYN----VKKEYFKIGLRISEF-----------------------------
------------------------------------------------------------
----------------------
>TRICHOPLAX_ADHAERENS_3_Ng1_gi|196001583|ref|XP_002110659.1|hypothetical
protein TRIADDRAFT_54901 [Trichoplax adhaerens]
------------------------------------------------------------
---------------------------------------------MVLVNNYSLIKLSPA
-TKIYFHGVD-F--EK--RDSYLAKNTFLRNHAARFMEAINVIIGQD---MD-IFSVESY
FRVVGSKHHSYN----LKLEHVQDISDAFLEMARNALKKKFT-KSTEAAWRSFFQMVTDA
IKNGIMKAQNRN------------------------------------------------
----------------------
>TRICHOPLAX_ADHAERENS_4_LGB3_gi|196016934|ref|XP_002118316.1|_hypothetical_
protein_TRIADDRAFT_62364_[Trichoplax_adhaerens]
--------------------------------------------------------MTKI
DSENHVKSKNVVTGNDIKSYLNYQERQAIIDSWNAIST---EKQKYGTILFLKLFELEPR
-VKSLFTIFD-F--NE--PLEDIIQSPHFRSHAMRFMQSLETGVLMG---FD-KESCDFL
FKSLGSRHHFYD----LKSEFLDVIPECILHTIKKGCGNNWS-NETADAWKIATKVLCEL
FREGLETKPKK-------------------------------------------------
----------------------
>PORIFERA

>CARTERIOSPONGIA_FOLIASCENS_LGB3_gi|241971149|gb|GO083496.1|GO083496_DMPR57
15864_Carteriospongia_foliascens_DMP_cDNA_Library_Carteriospongia_foliascen
s_cDNA,_mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nuces
t/
------------------------------------------------------------
----------------MSHPLSSEELKLVEKSWALIQE--FGLQEAGMVMFKRLFEMSPG
-LQNLFPFA---------GDSLSDDNEGMKHHSFIVMTSLDDCLKIV---HD-VPKLKDE
LISLGAVHHIQG----VTSEHFAPVGSALIWALEAALKEKFT-PEVKAAWLALYKVVQSL
MEQGMEEGI---------------------------------------------------
----------------------
>AMPHIMEDON_QUEENSLANDICA_LGB3_gi|340378768|ref|XP_003387899.1|_PREDICTED:_
neuroglobin-like_[Amphimedon_queenslandica]
------------------------------------------------------------
-------------------SLTSAQVALIESTWKVVKK---DLQGAGNIMFLKLFQIDVS
-VRDKFPFRD-V------PYEELEDSESFLKHSLQVMETIDLAITLLLG-GE-MEKLVEA
LVDLGMAHAMQG----LKPEDFDHVGEALVHALGVALGKEFN-DEAKKAWTLLYSVVTAK
MKEG--------------------------------------------------------
----------------------
>CNIDARIA

>NEMATOSTELLA_VECTENSIS_1_LGB3_gi|156408000|ref|XP_001641645.1|_predicted_p
rotein_[Nematostella_vectensis]
------------------------------------------------------------
-----------TENTLTSVPLSTRRKKLVRESWELIEP---VKITIGKRLFTRLFDVNPN
-MQDTFPNFK-G--K---ELKDILNSRSLYLHAKRVMVAVENAVTVL---DD-AETFESY
LINLGGRHLPWG----VTKDHFGVVGEAFIWALQDVLGEGCT-SDVAEAWIDLYGYIVQA
MLEG--------------------------------------------------------
----------------------
>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
```

```
------------------------------------------------------------
MHGVIEEGLLQLERINPITGLSAREVAVVKQTWNLVKP---DLMGVGMRIFKSLFEAFPA
-YQAVFPKFS-D--V---PLDKLEDTPAVGKHAISVTTKLDELIQTL---DE-PANLALL
ARQLGEDHIVLK----VNKPMFKSFGKVLVRLLENDLGQRFS-SFASRSWHKAYDVIVEY
IEEGLQQSYKQDPVTGITDAEKVLVQESWDLLKPDLLGLG--------------------
----------------------
>NEMATOSTELLA_VECTENSIS_3_LGB2_gi|162098183|gb|FC288989.1|FC288989_CAGN4759
.rev_CAGN_Nematostella_vectensis_Nemve_mixed_stages_unfert_eggs_to_primary_
polyps_Nematostella_vectensis_cDNA_clone_CAGN4759_3',_mRNA_sequence
------------------------------------------------------------
---------------NPQNAFSAADIQAIQGTWALAK---PDLMGKGAMVFKQLFTEHGY
--QPLFSNLA-Q--Y---EITGLEGSPELNTHARNVMAQLDTLVGSL---QN-SIELGQS
LAQLGKDHVPRK----VNRVHFKDFAEHFIPLMKADLGDEFT-PLAESAWKKAFDVMIAT
IEQGQRARRSVATFLTNPVA----------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_4_LGB2_gi|162112725|gb|FC303536.1|FC303536_CAIC1367
1.fwd_CAIC_Nematostella_vectensis_Nemve_whole_embryos_normalized_Nematostel
la_vectensis_cDNA_clone_CAIC13671_5',_mRNA_sequence
------------------------------------------------------------
------------------------------------------------------------
----CFKAFN-K--V---SLEDLEKSPFLKAHATSVMSAINEVVCNL---DE-VEILGIL
LEKIGFSHARRE----IRRIHFENLAKVVVAYLRQALGSHLT-EEGADAWRKALCVMIDI
IEKGSTSERW--------------------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_5_LGB3_gi|156405932|ref|XP_001640985.1|predicted
protein [Nematostella vectensis]
------------------------------------------------------------
--MGCGSSTFKPPREPVKIPLSVAQKYLVRETWETIEQ---HSKAVGKKTFLRFFEMNPD
-YQKLFPEFA-T--L---DQVELEQANALHGHAERVMKAVENAVSAM---DD-AESFAAY
LENLGARHKARA----LKPAYLDAMQVAYTDTIQDLLKTQWT-DGTAEAWNKLFRFIADT
MKHGLSS-----------------------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_scaffold_7000121_LGB3_from_Bailly_and_Vinogradov_20
08
------------------------------------------MGCGASKTLTTPHGTEEH
LTKKSQSENGNQSFVGNRPRLTERQIKLVQDTWRLLIP---SQKKTAMIFYLKLFTLDPI
-FKEVFS-FH-T--E---NEGQLEQDERFLFQSRKFMEMINSAVDRL---ND-ISLLVMI
LKSLGEVHWTKF---KIKPEYYEPVGKALIYSISKGLGSLFN-DEIGEAWQAMYDLMSGA
MISGTKAVQARSQNSL--------------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_scaffold_3000224_LGB3_from_Bailly_and_Vinogradov_20
08
---------------------------------------------------MGCGASSTV
RPFFIRQPASDTENTLTSVPLSTRRKKLVRESWELIEP---VKITIGKRLFTRLFDVNPN
-MQDTFPNFK-G--K---ELKDILNSRSLYLHAKRVMVAVENAVTVL---DD-AETFESY
LINLGGRHLPWG----VTKDHFGVVGEAFIWALQDVLGEGCT-SDVAEAWIDLYGYIVQA
MLEGLQQAKKGR------------------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_scaffold_76000030_LGB2_from_Bailly_and_Vinogradov_2
008
-----------------------------------------------MGCASSLATQTKL
LKGHLPTCETQYLTKVDQLPLTETQKYYIKQSWMGLES---NKGELGIEIFLRLFSENPT
-LQLMFPEFR-E--YS--TLEELKESRSLQGHTKRVMKVVENAVNSL---ED-GHALMEY
LQELGRRHKTRQ--IKPTVSNLQEISQAINETFEENLGIKWT-VEIAESWKLLLDYVMAM
IIRGLRSP----------------------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_scaffold_50000067_LGB2_from_Bailly_and_Vinogradov_2
008
-------------------------------------------------MGCVVSKNPST
VAKIVPGGGEEKLFETSRIPLDAKETQLVRKTWAILGD---RQVEVGKSLFLRFFEEHPT
-SKDLFPEFR-N--I---SNEKIAESPALYGHARRVMKSVDNAVASI---EN-VQVYSAY
```

```
LYELGTRHQTRQ----LSEEQLKFMGGAFLFAMRLHLRKEWS-RATSKAWEKIFSFMADA
MMRGCKG-----------------------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_scaffold_42000019_LGB2_from_Bailly_and_Vinogradov_2
008
--------------------------------------------------------MGCG
SSVVTQGNMPHLCGLKLECDMTYEQKYLIRETWKFLEV---SKKEIGVSVYKRFLNMHPG
-LQTYFSEFK---------HIKIDNINGSHGHPRRLLMAIDNAVTAL---GD-SDSFSAY
LVELGRRHHGMNFRPGPTHFNDLRKCFLSVIEEILATASLWD-FQVEEAWNRLFDSITAM
ILRGIQLAKV--------------------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_scaffold_5000153_LGB3_from_Bailly_and_Vinogradov_20
08
------------------------------------------------------------
--MGCGSSTFKPPREPVKIPLSVAQKYLVRETWETIEQ---HSKAVGKKTFLRFFEMNPD
-YQKLFPEFA-T--L---DQVELEQANALHGHAKRVMKAVENAVSAM---DD-AESFAAY
LENLGARHKARA----LKPAYLDAMQVAYTDTIQDLLKTQWT-DGTAEAWNKLFRFIADT
MKHGLSS-----------------------------------------------------
----------------------
>NEMATOSTELLA_VECTENSIS_scaffold_141000032_LGB3_from_Bailly_and_Vinogradov_
2008
------------------------------------------------------------
---MGCSSSLSQANLPRTMPLSEAQKYLVRETWETIEP---QKQTVGKKAFLRFFDMNPD
-YQNLFPEFK-S--L---SYEELQKANALHGHAKRVMKAVENAVMSI---DD-VMSFSAY
LEELGRRHKTRA----LKPSYLEAMHGALMDTLRNLLQSQWT-EETAEAWNKLFSFISTT
MVRGLQSRD---------------------------------------------------
----------------------
>CLYTIA_HEMISPHAERICA_1_LGB1_gi|294376172|gb|FP931337.1|FP931337_FP931337_C
lytia_hemisphaerica_library_Clytia_hemisphaerica_cDNA_clone_SA0AAB120YK22_5
',_mRNA_sequence
--------------------------------------------------MGAICSHKTL
VSKSFVYIKSIQKSSNIKVPLSAKEVNILKKTWPPVKN---NWLKICLVAFERWFTLYPQ
-LRYMFRSLP-E--DI--QFEDLFKTDALKMHVDKVRDVLELLIKKI---DN-VEELVNT
LVDFGRQHHMLG----AEQRYATALAASFQYGICMIMDVD---SSVENAWDSLLRFVMDL
LKLGMRMEKEAQEKESLNKGYNTEELLEKAQDGGEALDENENSAPMPLALINEDDSTSSV
RTFSCR----------------
>CLYTIA_HEMISPHAERICA_2_LGB2_gi|294400477|gb|FP945496.1|FP945496_FP945496_C
lytia_hemisphaerica_library_Clytia_hemisphaerica_cDNA_clone_SA0AAB133YH06_5
',_mRNA_sequence
--------------------------------------------------------MGSS
GSCLKFTMKPNGKVDAAPADYTEAEINIVQKQWSVAMR---NLDSVGYKLFAKLFQYKDI
MAKFDFAVAN-N--L---DFQKSMNDARLSFHIRRVFHTINTVVSCL---ND-GDFVASQ
LEHVGAIHAEYG----LQATHLARFKDVMLETLEEAFKEGFQ-EDSKTAWSKIVDAIAKY
MLKGIDKAKEQTKTSELSSDDGDGIKMEQ-------------------------------
----------------------
>CLAVA_1_LGB1_Locus_11167_Transcript_3/8_Confidence_0.360_Kindly_provided_b
y_Stefano_Piraino_from_Universita_del_Salento_Italy
-------------------------------------------------MGSKLCLAIHK
SVARSTQPSIHTKYGNICIPLTKKDIKKLRKSWGIMKM---NWYKICLVVFDRWFSMYPE
-LRNMFKSFS-S--VL--TLQELLASRTMKSHIQQLEELVEDLLYKV---DD-PTDFIET
VIQFGEGHHKLG------AKQLYATSLAAAFQYAICISLNLD-RDTENAWDSLFRFLMDC
LKFGMRNEMHKETTSPLDKLEVQDG-----------------------------------
----------------------
>CLAVA_2_Ng_Locus_28538_Transcript_4/6_Confidence_0.667_Kindly_provided_by_
Stefano_Piraino_from_Universita_del_Salento_Italy
------------------------------------------------------------
------------------------EVGILRKQWSIMMS---DADANGIMLFTALFAINGH
-LRTKFAFGD-V--YGF-DCKTTFGDERLLGHLRGVFSSVEMLINSLVQ-NT-FYESLGK
IKELGKVHSLAG----IHLSDLQAFKKAMIQTLKKELKSEFD-DKALKAWSKFIDIVIEA
FDVNDDTDDEKTQPIGK-------------------------------------------
----------------------
```

```
>HYDRA_MAGNIPAPILLATA_Ng1_gi|221117935|ref|XP_002162062.1|_PREDICTED:_simil
ar_to_neuroglobin_[Hydra_magnipapillata]
------------------------------------------------------------
--------------------LSGKEIETLKKSWTTAKQ---FWNEICTCAFSRWFSTYPE
-IQSKFGVYG-D--NL--TMNEVLASESLCIHIRKSVELIEIIIKKV---DE-RHELSEY
LIELGKLHHKFG----AEQKYATALGSSFVFAISQICPNIDM--ITEGAWDSLFKYIVT-
------------------------------------------------------------
----------------------
>MONTASTRAEA_FAVEOLATA_1_LGB3_gi|282539129|gb|GW263294.1|GW263294_CCHW10651
.b1_CCHW_Montastraea_faveolata_heat/dark/disease-
stressed_adult_+_6_day_old_larvae_Montastraea_faveolata_cDNA_clone_CCHW1065
1_5',_mRNA_sequence
MGCGGSKAIKNRTAPAPVAEQTQSPGLKETSPAKQHRQQPEGKRTEETVEETAPADGYGR
QGTEGDKQATEEQSENVEGPITQEQISLVQDTWKLVNG---DLEQVGVEFYTRLFKENPE
-LLQMFSFRD-L--ANS-TEDAMRTDDRFKRQGLVTMQHVDLAVASL---SD-LGSIVPA
LKDLGARHSMYK----VEEHHFGPVG----------------------------------
------------------------------------------------------------
----------------------
>MONTASTRAEA_FAVEOLATA_2_LGB2_gi|282548677|gb|GW271367.1|GW271367_CCHW15436
.b1_CCHW_Montastraea_faveolata_heat/dark/disease-
stressed_adult_+_6_day_old_larvae_Montastraea_faveolata_cDNA_clone_CCHW1543
6_5',_mRNA_sequence
------------------------------------------------------------
-----------MERSKESDGLTDLQIEMIRSSWEKVTP---NKKHHGQLLFHKLFEIAPE
-MTDLFPFG------------DDFTKPQFTTHALNIMNALDHAIQNL---DN-PDVLIPK
LRELGQMHAGFE----LTIKEFQHVGEALIWVLATGLGDDFT------------------
------------------------------------------------------------
----------------------
>ACOELA

>SYMSAGITTIFERA_ROSCOFFENSIS_2_LGB2_From_the_present_study_:_illumina_libra
ry_made_by_Genoscope,_French_Sequencing_Center
-----------------------------------------------MQLNNIQKCLPFL
FPKRNSHMSDYFAFEGLPPVPNSSESEEIRKSWKKIEL---DAAKLGIAVFVGLFERYPE
-IQASFSKIA-N--V---NKSSLNSNVMLHAHSLHIMHMIGKLIQLL---NE-PENLLSK
VVELGERHFDRK----ANDELLQYFCPAY-VEAMAKKGQWK--KKTIIAWEKFFDFIRAA
MVHGLKKRKGHSSISNTTSAANTAAEKNHNSPSSSQ------------------------
----------------------
>SYMSAGITTIFERA_ROSCOFFENSIS_1_3D_LGB2_From_the_present_study_:_illumina_li
brary_made_by_Genoscope,_French_Sequencing_Center
------------------------------------------------------------
------------MATLESMQVSEEQQSLIMEDVQVLLP---NYDDFVEDVLQQFMEENPE
-TFQIFPWAD-A--SK--TAKEMRSHPRFKSHAKSIGKVISDCLVDL---NG-VKKHEPK
LSSLGAMHTKKK----VPTELFGKLGGCILTQVVKRVSEAKWSEEKKEAWLKAYGIITVM
VTE---------------------------------------------------------
----------------------
>NEMERTODERMATIDA

>MEARA_STICHOPI_Ng1_transcript_1_Kindly_provided_by_Andreas_Hejnol_from_SAR
S_International_Center_for_Marine_Molecular_Biology
------------------------------------------------------------
---------------------------------------------SRKILVCRLFESHPQ
-VRDIFEHSE---------DKDLVPADIIKAHGMRIMGLLGRFVSNLESNVD-DNLLGQP
IHDLGRKHVDFK----APPYLFDVFALQIQHIIKSKLEQVWT-DEIGDAWKVMFDIIVFN
LKSGQNQEMQDRGITV--------------------------------------------
----------------------
>NEMERTODERMA_WESTBLADI_1_LGB2_gnl|ti|2247831830_Nw_MM1_39E05_From_transcri
ptome_available_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
------------------MGLTETQRVLIKQSWKKVIS--GGRVEAGWVLFSKFFTSSPE
-AQNYFKAFK-G--K---PLSELQNNTQMKGHVLRVINYITDIVDTL---EV-DEMREEM
```

```
SINIGRTHGRRA----IPAEMFQCLKPAVFGTVIDLNGGSLS-DEAAAAWGLLWEALVLC
VLEGMKKPGQPGMGQTHSHADQIIGELLVTHPAHVVALYITNYLLLFIKYLHHEPK----
----------------------
>NEMERTODERMA_WESTBLADI_2_LGB2_gnl|ti|2247830945_Nw_MM1_30C08_From_transcri
ptome_available_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
----------------------------------------------------------MG
SYLSGLLWNSVPDIPDELTKLTPSEKNALVDSWMLVID--DGLQQGGIKLFLKFFTLDPA
-ARPYFTKFL-K--L---SDDELRESKMLRAHVIRVMNTLNSLVDGL---DD-PELFVEL
SKFIGRTHYRHN----IDTEHLVNLGEAVIWLVEQSNGGRLP-PAVEDSWQRLWERITHF
ILAGEEEESKTDSGGETGVSVEQNSHVIMSTGVQSNSLNFHHIFFNNNCPYCIPQLT---
----------------------
>NEMERTODERMA_WESTBLADI_3_LGB1_gnl|ti|2247830237_Nw_MM1_22H08_From_transcri
ptome_available_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
------------------MALTDSQQFLVKKVWDKVVA--DGYHEPGLIFMNKFFTDTPE
-AQNYFWFVK-D--R---PLYELRENIHMEAHLTRGLSHITCTINKV---AE-PTVCADM
WRNLGRTHTAHT----VPFKMFDLMKPAFVNMVIQCCGGMVD-QDTLVALSLYWDFIMRC
VKEGMEEAAAKVWSDSTTLNGSLYLKIINTPLLYCPCHGRTVSDIITRCMLRVAII----
----------------------
>XENOTURBELLIDA

>XENOTURBELLA_HEMOGLOBIN_RELATED_S02970_XB_MM1_31H06_From_transcriptome_ava
ilable_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
---------------MVFAGLSGGEVAAVKSSFGKVTD----LAGTGAELFLKYFQARPS
-AQNYFKWKG-Q------SLDAVKADAYLQTHGNRVMSTLKTVIGSI---ED-AGKCEAL
ITQLGNLHTGRN----VAAADFEALFPIIAGVLSEKCGGALA-GDASSGWQKLYNEIVRL
IDAGQKTPAAGGARAEFLAAIGAN------------------------------------
----------------------
>MOLLUSCA

>SEPIA_OFFICINALIS_LGB2_Sequence_has_been_provided_by_L.Bonnaud_and_Y.Bassa
glia_from_an_ESTs_library_built_from_Sepia_officinalis_embryos_by_the_Genos
cope/CEA_(project_AP07/08_n?07-http://www.genoscope.cns.fr/spip/Collection-
d-ESTs-d-embryons-de.html_)ADY0AAA118YM16CM1
-------------------------------------------------MGALFSVIDAL
IRKIGIMSPLDDEEVDSKTGLKVKEKRAMTQSWKYISP---NLKAEGIHFFNMLFTEHPD
-YIDYFPSFR-G--K---KLEEFNTKASFMAHAKNVFYAVTLIVDTL---DD-ADELVEI
LLKTGRNHHRRS----VPLSPFQGRAIVFEKIMTERLGKAIP-PLGKGLVDKSIDCS---
------------------------------------------------------------
----------------------
>EUPRYMNA_SCOLOPES_1_LGB2_gi|84449782|gb|DW284378.1|DW284378_UI-S-HH0-aea-
p-02-0-UI.s1_UI-S-HH0_Euprymna_scolopes_cDNA_clone_UI-S-HH0-aea-p-02-0-
UI_3',_mRNA_sequence_
From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/
-------------------------------------------------MGALLSVIDAV
IRKIGIMSPRDDEEVDSKTGLKVKEKRAITQSWKHVSQ---NLKQESMNFFNLLFTEHPA
-YIEYFPAFC-G--K---KLEEFNTKPKFIAHAKNVFYALILVVDSL---DE-PDELVEI
LLKTGRDHHNRG----VPMAAFHNLAIVFDKFLTIRLGNNYT-PLAKESWTKALTVVNAV
IEKGIEDGIDCSECRYRERNRRST------------------------------------
----------------------
>DORYTEUTHIS_PEALEII_LGB_related_gi|342663023|gb|JK329464.1|JK329464_oy57h0
1.y1_Woods_Hole_Squid_Stellate_Ganglia_cDNA_Library_Doryteuthis_pealeii_cDN
A,_mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/
------------------------------------------------------------
VRADGLPYPAPPPPTDPRLPLSPLQVFKLKKSWKGIKR---SIELTGVEMFVRMFRTQPG
-LKNLFKDFR-E--LE--TDDEMRENEALEKHATLVMNTLDDAITNI---EN-VDLVLDL
LHRIGKSHLRFQG---FNVEYFWLAEQPLLDAIKITLGDRYS-DNMDIIYKLVIRFLLTE
VTKGARVDVSST------------------------------------------------
----------------------
```

```
>EUPRYMNA_SCOLOPES_2_LGB_related_gi|84436671|gb|DW271268.1|DW271268_UI-S-
GS1-acj-k-06-0-UI.s1_UI-S-GS1_Euprymna_scolopes_cDNA_clone_UI-S-GS1-acj-k-
06-0-
UI_3',_mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest
/
------------------------------------------------------------
------------------------------------------------------------
-----------I--E---TDDQMRENEALEKHATLVMNTIDEVIANI---EN-VDLVLDL
LHRIGKSHCRFQG---FNVEYFWLAEQPLIDAIKITLGDRYS-DNMNIIYKLVIRFLLTE
VTKGARANVSST------------------------------------------------
----------------------
>MYTILUS_CALIFORNIANUS_1_LGB2_gi|145897418|gb|ES403000.1|ES403000_MUT03-
C03.x1d-t_SHGC-
MUT_Mytilus_californianus_cDNA_5',_mRNA_sequence_From_ESTs_stored_at_http:/
/www.ncbi.nlm.nih.gov/nucest/
------------------------------------------------------------
---------------------GELATDTIRSTWPLLSD---DIERTGIKVFLRIFYEEPK
-IRNVFKRFG-E--M---DEIELRRSPIFKEHAYRFMRVVDDLVDNM---DPPKAHIQQN
LMMLGAKHATFEG---FRIEYFEAYSESLIDVWEYTIGEEFI-PEVRESWTEFFDYLVKY
MCQGYNVFTNET------------------------------------------------
----------------------
>MYTILUS_CALIFORNIANUS_2_LGB2_gi|223024721|gb|FL490027.1|FL490027_Mg_Nor01_
51M10_Nor01_Mytilus_galloprovincialis_cDNA_3',_mRNA_sequence_From_ESTs_stor
ed_at_http://www.ncbi.nlm.nih.gov/nucest/
------------------------------------------------------------
------------------QTMESSTISAIKNTAAALEE---HSSQLGLKTYELLYEEHPE
-VKATFNKS----------EYSTEADAVPKDLAVKVGKAIATFAKHC---DD-LDSQKEL
FGGIAIKHVKLGG---ILEEQYSAMEGSFLKAVKEVLGDAAT-DEVLDHWKKGFHFLAEH
IRGLEAEK----------------------------------------------------
----------------------
>SPISULA_SOLIDISSIMA_LGB2_gi|76058055|emb|CAJ31107.1|_nerve_hemoglobin_[Spi
sula_solidissima]
------------------------------------------------------------
-------------DPCPVTKLTKAEKDAVANSWAALKQ---DWKTIGADFFVKLFETYPN
-IKAYFKSFD-N--M---DMSEIKQSPKLRAHSINFCHGLNSFIQSL---DE-PDVLVIL
VQKLTVNHFRRK----IAVDRFQEAFALYVSYAQDHAKFD---DFTAAAWTKTLKVVADV
IG----------------------------------------------------------
----------------------
>APLYSIA_1_LGB3_CNSN01-C-002803-
501_From_transcriptome_available_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/
------------------------------------------------------------
-------------------------IFSLKQSWKGIKR---KMEDTGVDMFVRLFKTSSH
-LQTMFQGFE-D--IR--SDDELRSNEALEYHATLVMTTLDDAITHI---DN-YDFVKQL
LSKTGASHVKFAG---FKSTNFLAIKGPFLEAVRVTLGDRYT-ENMQNIYTIAIIFILET
LQQGMEEALEKAGSSEVAQGHIRV------------------------------------
----------------------
>APLISIA_2_LGB2_CNSN01-F-074342-
501_From_transcriptome_available_at_ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/

----------------------------------------------------MGAEQSLS
CFPCTSVRRCRRRHCQLSKYLNPRQVQLVQCTWSILKM---DLTTLGVTVFLNFFETEPD
-LKMMFPKMI-RM-NESNQLEWDVDKDMLQKHAVTVMEGLGAAVETL---ND-SQFLNTV
LIALGQTHEKRN----IKPNMLKRMWPSMHVGLGAVLGEGYT-KEVSEAWCKLYSYICLQ
MQIGMENPN---------------------------------------------------
----------------------
>LIOLOPHURA_JAPONICA_Ng1_gi|47115693|sp|Q7M416.1|GLB1_LIOJA RecName:
Full=Globin-1; AltName: Full=Myoglobin I
------------------------------------------------------------
------------------gISADQAKALKDDIAVVAQ---NPNGCGKALFIKMFEMNPG
-WVEKFPAWK-G--K---SldEIKASDKITNHGGKVINELANWINNI-------NSASGI
LKSQGTAHKGRS----IGIEYFENVLPVIDATFAQQMGGAYT-AAMKDALKAAWTgvivp
```

```
gmkagy------------------------------------------------------
----------------------
>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://ww
w.ncbi.nlm.nih.gov/nucest/
------------------------------------------------------------
---------------MTSTEFTDEQKQLVKTTWNIVRE---DISKVGVITFLRLFEKFPD
-VQDLFVPFR-G--L---NSEELRQNVGLREHGMRVMGTIEKCITRI---DQ-PDKLTSM
LESLGEKHVVFD----TKIEYFDLLSPQLIQAITTRHWSVD--SQREQAWNNFYCTL---
------------------------------------------------------------
----------------------
>CRASSOSTREA_GIGAS_1_LGB2_gi|318048551|gb|FQ663445.1|FQ663445_FQ663445_Cras
sostrea_gigas_library_(Genoscope_-
_CEA)_Crassostrea_gigas_cDNA_clone_WY0AAA49YL07FM1,_mRNA_sequence_From_ESTs
_stored_at_http://www.ncbi.nlm.nih.gov/nucest/
---------------------------------------------------------MGC
DASKSTDVIENEEIESDTFDFTDTQIDTIRSTWPILAC---DMVDIGSKVFLKIFIDEPK
-LKYAFPSFS-D--M---EENELLRHPPFIDHVTRFMQIIDYLVENL---DQQNSDFHQA
LLMLGAKHATYPG---FQVSHFTVFNKALLEAWESAIGEEFI-PEVQNCWTQLFAYIMRY
IVQGYELYFSEC------------------------------------------------
----------------------
>ARTHROPODA

>APIS_MELLIFERA_LGB2_gi|118150510|ref|NP_001071291.1|globin 1 [Apis
mellifera]
---------------------------------------------------------MGT
FLRFLGISSSDDNRIDQATGLTERQKKLVQNTWAVVRK---DEVASGIAVMTAFFKKYPE
-YQRYFTAFM-D--T---PLNELPANKRFQAHCAGVITALNNVIDFL---HD-PGLMEAS
LIGLVERHKKRG----QTKEEFQNLKEVMLEVLRQALGKQYT-PEVAEAWNKTLDMMFGK
IYQ---------------------------------------------------------
----------------------
>CARCINUS_MAENAS_1_LGB2_gi|299757081|emb|CBN88274.1|_hemoglobin_[Carcinus_m
aenas]
---------------------------------------MGAVLSVVWGWLSPGTQVGAV
TFPEEGSLGPEADVPDKTTGLTLRHRTAIYRTWDLVRP---NPKLHGINLFLTMFQEEPV
-LQTRFKGFA-G--K---SIEELKNSKRLAAHGTTVVMAITAMVDNL---ED-VSVLVEL
LKNTGANHRDRG----VPKGDFELLAPVLVRFLKDNLGSAWS-PVAEEAWTQAMKVINAV
IFTSYDACSILS------------------------------------------------
----------------------
>CARCINUS_MAENAS_2_LGB_related_gi|84413153|gb|DW250579.1|DW250579_Cm_mx1_36
e01_SP6_Green_Shore_Crab_Multiple_Tissue,_Normalized_Carcinus_maenas_cDNA_c
lone_Cm_mx1_36e01_5'_similar_to_gb|EAL38715.1|_ENSANGP00000028536__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
VKVIDLPEPPEPPPPDPRLPLTARQRFNIIKSWKGIAR---AIEPTGVNMFVKLFENHSE
-LITFFTKFR-Q--LR--TRDEQAESLELAEHATIVMNSIDEGIKAM---DN-VDFFFDL
LHQIGASHRKIPG---FKKEYFWKIEHPFLEAVRLTLGDRYT-DNMDNIYRITIKLLIET
VVRGYELAELKE------------------------------------------------
----------------------
>DAPHNIA_PULEX_GlobinX3_gi|321478927|gb|EFX89883.1|hypothetical protein
DAPPUDRAFT_231989 [Daphnia pulex]
------------------------------------------------------------
----------------------------------RLEK---DIAQVGIIVFINLFETHPD
-MQSVFLPFT----------GVVLDDLKKSKLLSEHALRVMGAVQRAVHRLQEPEKLHAF
LSELGRKHEKNG----AKLEYIDYIGPQFLCAIRPILGDDRT-LETEKAWTLLLDYMTAA
MKESLVEARNASAAESSKPLTLPPSSSSSSSAATDD------------------------
----------------------
```

```
>DAPHNIA_CARINATA_LGB2_gi|290774776|gb|GR506052.1|GR506052_LAaa_0006_F05_Ga
mogenetic_water_flea_(Daphnia_carinata)_cDNA_library_Daphnia_carinata_cDNA_
5',_mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/
------------------------------------------------------------
-----------------MDTLKTVNVAAVQNTWAIVKK---DLNTHAPQFYVALLTAHPE
-YQPMFPTIA-N--V---PAGELLNNPALKTLSVNVLTKLSELIDCMG----NPDALQGQ
LVDLANQHKQRG----TTRAHFDNLSKVLIDFLAAKLGGEFT-PEARQAWTATMQGINTV
VEASS-------------------------------------------------------
----------------------
>HARPEGNATHOS_SALTATOR_LGB2_gi|307192580|gb|EFN75768.1|_Globin_[Harpegnatho
s_saltator]
------------------------------------------------------------
----------------EKTGMSEKQKKLVQNTWAIVRK---DDVSSGLAIMNAFFTRYPE
-YQDQFKSFK-G--I---PFEELSKNKKFQAHCVSVIAGLSNVIDHI---HN-PELMEAS
LINLAERHKNRG----QTREHFQNLRYVLEDLIPSVFGKQYT-QEVQEAWKKMFDYLF--
------------------------------------------------------------
----------------------
>TUNICATA

>Ciona_intestinalis_Gb4 HBG related NP_001027701.1
-----------------------MQSMSTPNPNNQCVNGPNYFCSCCCGNSVCNSNWLTC
VTAGSVAPTVSTTVPPADEGLKRSDIINIQDSWNTLKGF--GYETVGMLVLHRLFNDAPQ
-TRYLFSQLS-LSSNESFTLEQMRNNSRVVYHANRVARAVGRLVDLI---ELP-TNFTDH
LVWLGQRHAYHG----VAPVNFDYMGPVLLETIKVNLELPSD-SPTLSAWAKAYGVIKNG
IKDAIIATYAEG------------------------------------------------
----------------------
>Ciona_savigny_Gb1 LGB3 (from Hoffmann et al 2012)
------------------------------------------------------------
------------------MEMNAQEIQDVRDSWKRLCAD--GEKTVGLMLMQKLFNTYPE
-SIKVFSRLG-ITNKAIITIDDLSTNSAASRHAESLTSRIGTLVDLM---HNT-HEFKEC
STEVGEIHIKYG----VTAEHVDILGNVLLSVICDSQGLSKS-SDLYLCWTKTWEGIAKY
VKIGLQQ-----------------------------------------------------
----------------------
>MOLGULA_TECTIFORMIS_LGB3_gi|68107250|gb|CJ414980.1|CJ414980_CJ414980_Molgu
la_tectiformis_unpublished_cDNA_library,_larva_Molgula_tectiformis_cDNA_clo
ne_mtlv020d18_5',_mRNA_sequence_From_ESTs_stored_at_http://www.ncbi.nlm.nih
.gov/nucest/
------------------------------------------------------------
-----------------MGGLTQEEIQAVRESWAAVIKK-VGVTETGLAVLHRLFADVPE
-TKTLFYNIG-LDSYDTITMDQLKDNKKIRNHALRVVTSISNVLKNI---KN-GEKLAEL
FKKLGEIHKKNK----VPPEYFDKMGNVLLAVLTSVLEVDLC-HPTSLAWGKVYGILRDQ
------------------------------------------------------------
----------------------
>BOTRYLLUS_SCHLOSSERI_Cytoglobin2_gi|322520564|gb|JG298067.1|JG298067_CCAO1
145.b1_CCAO_Botryllus_schlosseri_total_asexual_and_embryonic_development_Bo
tryllus_schlosseri_cDNA_clone_CCAO1145_5',_mRNA_sequence_From_ESTs_stored_a
t_http://www.ncbi.nlm.nih.gov/nucest/
------------------------------------------------------------
--------------MEEKSLFTAEESDAIKSSWSKMTE--NGVSTAGRLILCRLFQDVPE
-VTTLFYRLG-ISGDVSYTLESLSQKSQFNSHAKRLSLALDAIVKKL---DD-TAFITQK
CTDLGKAHEKHN----VKPHYFDLLGQVLVKVICQALNLTES-HPTVKAWIKAYGAISTG
LKNGLTSELYLS------------------------------------------------
----------------------
>CEPHALOCHORDA
---------------------------
>Branchiostoma floridae GB1 LGB3  XP_002608549.1 hypothetical protein
BRAFLDRAFT_98913
--------------------------------------------MGAFLTKPFSLVGRLL
WKVLFSWWVKQIETPSDVTGLTPTQSRLVKESWKMFLS---KKRENGFVIFRVLFTDYPV
-TRKLFKGVE-Q--LDLDAPGQLESSITLRAHVTRFMHSFDTYMESL---DDP-EDLKQL
LYDTGKSHLIHD----IKPEYFDVLETVLMKSLRIVFGSKLT-PQLEEAWQTAYSHLKVT
```

```
IKQGLEDAIQKRDQADTSVVVTVE------------------------------------
----------------------
>Branchiostoma floridae Gb14 LGB2  XP_002610160.1 hypothetical protein
BRAFLDRAFT_77082
---------------------------------------MGANMGCSNSKKMSHESESAN
SGDSTPPKSSTPSALDERLPLTQKQKFLLLKSWKGVAR---QISQCGKTMLIRLFKDDPQ
-LMAVFNQKF-R-HLRERDADVLYQDAILDAHAATVMEALHEAITHL---DDS-VFVMKV
LHDVGKMHQRYN----VDPSVFLKVEKPFLTAVSEVLGDRYT-KNMEEIYTITIKFILAT
LSEGATMELTEDEQKNLGRLWRPPGRVHKFVRPEKVAAIVDAQSEENGVH----------
----------------------
>Branchiostoma floridae Gb15 HGB related GenBank: CBL51564.1
------------------------------------------------------------
------------------MGLTSEDKSAVLDSWAKMSGP--TFQDAGEKVFLLLLKTDS-
-TKALFPKFR-D-----IPYDQLAGHPDVRDHGGKVMQVLDDFIKGL---DNG---GDGA
VQKVGLLHKGVG----VSHDNINLMKPVLMTLLGELGCSSA-----AGAWENLWARFMDV
HRTCY-------------------------------------------------------
----------------------
>Branchiostoma floridae Gb4 LGB3 XP_002589215.1 hypothetical protein
BRAFLDRAFT_74626  XP_002589215.1
----------------------------------------------------MGTIADGE
GTELNGYGGEKEPGGGHGGPLTQEQVHGIKETWAILAQ---DPVERGVDLFMKIFEEDPD
-LKKLFYFAD-D------GRELSREDQRMRSHGERVMEAVGGAVDSL---GDL-TAVVPV
LTELGALHHKYG----VQPSYFDTVGAALIYILETNLGDKLT-PNIRQGWVLVYGIVGAT
MKKGMQQAMDHQNMAKTRP-----------------------------------------
----------------------
>HEMICHORDATA

>Saccoglossus kowalevskii Gb1 NGB2 neuroglobin-like protein NP_001161601.1
------------------------------------------------------------
-------------MADPVTTLTSDEVAAIKSSWSAVYD---KKKESGVTLFVKLFTENPS
-FKSQFGYMS-G-V----ADGDMKTLPALENHGVKVMDRINEWMGNL----TNGAELVKQ
LKHLGTTHIALK----VTEDNFNAMDSVLMYTLQEQGGSAFT-PAAKAAWQKAWGVMKSV
IVGALKG-----------------------------------------------------
----------------------
>Saccoglossus kowalevskii Gb12 LGB3 Saccoglossus kowalevskii scaffold 38407
in Acorn worm assembly (see Hoffmann et al 2012)
-----------------------------------------------------------M
GCTPSINERDFQTPVDDKHLLDDRQKRIVRKTWRPLAN---DMTENGQKIFINIFESHPE
-IKYMFPTRD-I-E----GRDNLSANPHFRMHSSRFMQSVGAAIDNL---NDLDNALRPL
LVKLAKTHVRFK---GFKPDYFDAFEEAMLSVWQEELGQRFT-TEVEESWKLLFFYIKDC
LKEGYDIAMNEKTSGELNNSDFINQ-----------------------------------
----------------------
>Saccoglossus kowalevskii GB15 LGB related scaffold 38908 in Acorn worm
assembly (see Hoffmann et al 2012)
----------------------------------------------MGCTVSTKPDLYSQ
NGDADPYKKKKVAPVDSRLPLTARQKFQITKSWKGIAR---NMENTGKSMFMRLFQSNIE
-LKNMFTGFE-E-FD---DLEDMRESQQLENHASLVMYTIDEAIASI---DDI-DFVVEL
LGKIGRTHTRTD----FNPQLFWRIEQPFLSAVKETLEDRYT-KNIEEIYKITFRFIVDA
LIDGVVAGVNERKAEAEAEAEAKAEAEDEQDNEMMKKTMHNGDKEDNCIESSKQKDITVI
ASTKQR----------------
>Saccoglossus kowalevskii Gb16 LGB 3 scaffold 38908 in Acorn worm assembly
(see Hoffmann et al 2012)
------------------------------------------------------------
-------MGCLISSNGEHPMITKEQTKILTSTWHSIHG---DLEKIGLLMFMGMFDNYPE
-TRQFFGLSG-G-SI---VLEDPAVIQKIREHGLRFMTTARKLVMNL---DDK-DKFDRI
LLDLGRRHHGYK----ADVDLIEVFGQQFIASIQPTLKDNWN-PAVGEAWEQLFKCVSSR
MKDGFLQAQSSPSNTELLK-----------------------------------------
------
>BALANOGLOSSUS_CLAVIGERUS_Cytoglobin2_gi|311137357|gb|FN985678.1|FN985678_F
N985678_dmp027_Balanoglossus_clavigerus_cDNA_clone_dmp027P0033D09,_mRNA_seq
uence_From_ESTs_stored_at_http://www.ncbi.nlm.nih.gov/nucest/
```

```
----------------------------------------------------------MG
SIIGYFWASSDGNVPDPGTTLTPNERRAIRDTWNAIYE---NKGENGVAFFMKLFEAHPV
-YKKLFKNLE-G-ID---DMAKLSKHPRMKAHGLRVMASFNSFIENL---DD-AEVLVQL
LVDIGISHSKHK----VTEEQFNVSNHFLSYFLRSLFLFAFI-YLAIPFPNFVVVLIYLF
VRLSLLSVYLSV------------------------------------------------
----------------------
>ECHINODERMATA

>STRONGYLOCENTROTUS_PURPURATUS_1_Ng3_Strpur_D1_(147-305)
------------------------------------------------------------
--------------------LDERQKTLVENTWKTLEK---NTELYGSIMFAKLTTDHPD
-IGKLFPFGG-K--NL--TYGELLVDPDVRVHGKRVIETLGSVVEDL---DD-MELVIQI
LEDLGQRHNAYN----AKKTHIIAVGGALLFTIEEALGAGFT-PEVKAAWAAVYNIVSDT
MSTKL-------------------------------------------------------
----------------------
>KINORHYNCHA

>ECHINODERES_HORNI_Globin_NeuroBlast_KINORHYNCHA__GNL|TI|1987172666_EH_MM1_
21H06__FROM_TRANSCRIPTOME_AVAILABLE_AT_FTP://FTP.NCBI.NLM.NIH.GOV/PUB/TRACE
DB/
--------------------------------------------------------MGCG
QSLESQGDGGAADIDSNKCPLDRRQIFELTKSWRMISK---NMESAGVMMFVKLFTTTNE
-INHMFAHLI-T--EDG-KVRNSEAQKQLEHHAAMVMYTLDESIQAF---TD-LDVLFEL
TTRIGQQHRKMDG---FSANYFWKIEDPFLTAVKQTLDDKYA-AELGHIYRRPIRFILKL
LYNGYMKKD---------------------------------------------------
----------------------
>ENTOPROCTA

>PEDICELLINA_SP_LGB1_ENTOPROCT__GNL|TI|1987415660_PC_MM1_28H03__FROM_TRANSC
RIPTOME_AVAILABLE_AT_FTP://FTP.NCBI.NLM.NIH.GOV/PUB/TRACEDB/
------------------------------------------------------------
-------------IRDRLLSLDSNEMMAMQLMWNDWVG--GDAAARGFDSFILMFKQHPE
-TQTLFDFAR-G--C---SAAQMQNSSRLLFHVTRVIHNITDVMDHL---HE-LEEVVPK
LRQLGGRHGTNPGCYNVPGAYFPWLGDAMKELMQGKGGWG---DSHNKAWAFVWSFICNS
LTSGQAIYGGGSCPAKKA------------------------------------------
----------------------
>TARDIGRADA

>RICHTERSIUS_CORONIFER_HEMOGLOBIN_RELATED_TARDIGRADA_GNL|TI|1987149271_RC_M
M1_21B09__FROM_TRANSCRIPTOME_AVAILABLE_AT_FTP://FTP.NCBI.NLM.NIH.GOV/PUB/TR
ACEDB/
----------------------------------------------------------MG
GRASRILSSGSANVPDKVTGLSDVEKRAIQENWRIVYR---DLKGNGVELFVRYFSRYPE
-YKDAFESLR-D--I---RMQDIGKSHKLRAHSVQVMQYGTKPRKAR---YA-AGTFYGE
PLNL---HTTTVDF-------------------LAFKLGHNFGHVHKTAWNKMLELVMTV
IQQGMTAEADDSKYKENQTAAVRQSEHEELTLTNLTETT---------------------
----------------------
```

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    ---------- ---------- ---------- -------MGC ALSGAQDPPV SEECSPLDDG LDLNRETTLG
CALLORHINCHUS MILII G    ---------- ---------- ---------- ---------- ---------- ----MGCAIS GPGQYPASGR
DANIO RERIO GbX (from    ---------- ---------- ---------- ---------- ---------- ---MGCAISG SGLTAGAPEI
ORYZIAS LATIPES GbX (    ---------- ---------- ---------- ---------- ---------- ----MGCAIS GLAAKTDLAE
PETROMIZON MARINUS Gb    ---------- ---------- ---------- --MGCTVSTD ERTGAQSSSQ GQSQASRKQQ QPEQQRAAGE
PYTHON MORULUS GbX (f    ---------- ---------- ---------- ---------- ---------- ---------- ----------
TETRAODON NIGROVIRIDI    ---------- ---------- ---------- ---------- --------MG CAISSLGAKA EFGDRSAEEE
XENOPUS TROPICALIS Gb    ---------- ---------- ---------- ---------- ---------- -MGCILSSLG WQWRDSLDHT
VERTEBRATE MYOGLOBINS    ----...... .......... .......... .......... .......... .......... ..........
HOMO_SAPIENS_MYOGLOBI    ---------- ---------- ---------- ---------- ---------- ---------- ----------
DANIO_RERIO_MYOGLOBIN    ---------- ---------- ---------- ---------- ---------- ---------- ----------
BOS_TAURUS_MYOGLOBIN_    ---------- ---------- ---------- ---------- ---------- ---------- ----------
GALLUS_GALLUS_MYOGLOB    ---------- ---------- ---------- ---------- ---------- ---------- ----------
VERTEBRATE CYTOGLOBIN    .......... .......... .......... .......... .......... .......... ..........
DANIO_RERIO_CYTOGLOBI    ---------- ---------- ---------- ---------- ---------- ---------- -----MEGDG
OVIS_ARIES_CYTOGLOBIN    ---------- ---------- ---------- ---------- ---------- ---------- --MEKVPGEM
HOMO_SAPIENS_CYTOGLOB    ---------- ---------- ---------- ---------- ---------- ---------- --MEKVPGEM
VERTEBRATE HEMOGLOBIN    .......... .......... .......... .......... .......... .......... ..........
EQUUS_CABALLUS_HB_ZET    ---------- ---------- ---------- ---------- ---------- ---------- ----------
HOMO_SAPIENS_HB_BETA_    ---------- ---------- ---------- ---------- ---------- ---------- ----------
TAENIOPYGIA_GUTTATA_H    ---------- ---------- ---------- ---------- ---------- ---------- ----------
ORYCTOLAGUS_CUNICULUS    ---------- ---------- ---------- ---------- ---------- ---------- ----------
VERTEBRATE NEUROGLOBI    .......... .......... .......... .......... .......... .......... ..........
HOMO_SAPIENS_gi|10864    ---------- ---------- ---------- ---------- ---------- ---------- ----------
DANIO_RERIO_gi_188590    ---------- ---------- ---------- ---------- ---------- ---------- ----------
GALLUS_GALLUS_gi_1548    ---------- ---------- ---------- ---------- ---------- ---------- ----------
MONODELPHIS_DOMESTICA    ---------- ---------- ---------- ---------- ---------- ---------- ----------
CHIONODRACO_MYERSI_gi    ---------- ---------- ---------- ---------- ---------- ---------- ----------
BOVICHTUS_VARIEGATUS_    ---------- ---------- ---------- ---------- ---------- ---------- ----------
SIPUNCULA                .......... .......... .......... .......... .......... .......... ..........
THEMISTE_1_LGB2_gnl|t    ---------- ---------- ---------- ---------- ---------- ---------- ----------
THEMISTE_2_LGB2_gnl|t    ---------- ---------- ---------- ---------- ---------- ---------- ----------
THEMISTE_3_LGB2_gnl|t    ---------- ---------- ---------- ---------- ---------- ---------- ----------
BRACHIOPODA              .......... .......... .......... .......... .......... .......... ..........
TEREBRETALIA_Ng2_gnl|    ---------- ---------- ---------- ---------- ---------- ---------- ----------
ANNELIDA                 .......... .......... .......... .......... .......... .......... ..........
ALVINELLA_POMPEJANA_1    ---------- ---------- ---------- ---------- ---------- ---------- ----------
ALVINELLA_POMPEJANA_2    ---------- ---------- ---------- ---------- ---------- ---------- ---------M
APHRODITA_ACULEATA_LG    ---------- ---------- ---------- ---------- ---------- ---------- ----------
CAPITELLA_SP_1_LGB2_F    ---------- ---------- ---------- ---------- ---------- ---------- ----------
CAPITELLA_SP_2_LGB2_F    ---------- ---------- ---------- ---------- ---------- --mgnaqpfv scaggqppdq
CAPITELLA_SP_3_LGB3_g    ---------- ---------- ---------- ---------- ---------- ---------- ----------
PLATHYNEREIS_dumerili    ---------- ---------- ---------- ---------- ---------- ---------- ----------
HELLOBDELLA_ROBUSTA1_    ---------- ---------- ---------- ---------- ---------- ----MGANGF KKVKSEPLLN
ECHIURA                  .......... .......... .......... .......... .......... .......... ..........
URECHIS_CAUPO_Ng1_gnl    ---------- ---------- ---------- ---------- ---------- ---------- ----------
PLATYHELMINTHES          .......... .......... .......... .......... .......... .......... ..........
MACROSTOMUM_LIGNANO_1    ---------- ---------- ---------- ---------- ---------- ---------- ----------
MACROSTOMUM_LIGNANO_2    ---------- ---------- ---------- ---------- ---------- ---------- ----------
CHOANOFLAGELLATE         .......... .......... .......... .......... .......... .......... ..........
MONOSIGA_BREVICOLLIS_    ---------- ---------- ---------- ---------- ---------- --------MS DRRRSSGSSE
SALPINGOECA_SP_Ng2_gi    ---------- ---------- ---------- ---------- ---------- ---------- ----------
PLACOZOA                 .......... .......... .......... .......... .......... .......... ..........
TRICHOPLAX_ADHAERENS_    ---------- ---------- ---------- ---------- ---------- ---------- ----MDQAQT
TRICHOPLAX_ADHAERENS_    ---------- ---------- ---------- ---------- ---------- ---------- ----------
TRICHOPLAX_ADHAERENS_    ---------- ---------- ---------- ---------- ---------- ---------- ----------
TRICHOPLAX_ADHAERENS_    ---------- ---------- ---------- ---------- ---------- ------MTKI DSENHVKSKN
PORIFERA                 .......... .......... .......... .......... .......... .......... ..........
CARTERIOSPONGIA_FOLIA    ---------- ---------- ---------- ---------- ---------- ---------- ----------
AMPHIMEDON_QUEENSLAND    ---------- ---------- ---------- ---------- ---------- ---------- ----------
CNIDARIA                 .......... .......... .......... .......... .......... .......... ..........
NEMATOSTELLA_VECTENSI    ---------- ---------- ---------- ---------- ---------- ---------- ----------
NEMATOSTELLA_VECTENSI    ---------- ---------- ---------- ---------- ---------- ---------- MHGVIEEGLL
NEMATOSTELLA_VECTENSI    ---------- ---------- ---------- ---------- ---------- ---------- ----------
NEMATOSTELLA_VECTENSI    ---------- ---------- ---------- ---------- ---------- ---------- ----------
```

```
NEMATOSTELLA_VECTENSI   ---------- ---------- ---------- ---------- ---------- ---------- --MGCGSSTF
NEMATOSTELLA_VECTENSI   ---------- ---------- ---------- ---------- --MGCGASKT LTTPHGTEEH LTKKSQSENG
NEMATOSTELLA_VECTENSI   ---------- ---------- ---------- ---------- ---------- -MGCGASSTV RPFFIRQPAS
NEMATOSTELLA_VECTENSI   ---------- ---------- ---------- ---------- -------MGC ASSLATQTKL LKGHLPTCET
NEMATOSTELLA_VECTENSI   ---------- ---------- ---------- ---------- ---------M GCVVSKNPST VAKIVPGGGE
NEMATOSTELLA_VECTENSI   ---------- ---------- ---------- ---------- ---------- ------MGCG SSVVTQGNMP
NEMATOSTELLA_VECTENSI   ---------- ---------- ---------- ---------- ---------- ---------- --MGCGSSTF
NEMATOSTELLA_VECTENSI   ---------- ---------- ---------- ---------- ---------- ---------- ---MGCSSSL
CLYTIA_HEMISPHAERICA_   ---------- ---------- ---------- ---------- ---------- MGAICSHKTL VSKSFVYIKS
CLYTIA_HEMISPHAERICA_   ---------- ---------- ---------- ---------- ---------- ------MGSS GSCLKFTMKP
CLAVA_1_LGB1_Locus_11   ---------- ---------- ---------- ---------- ---------M GSKLCLAIHK SVARSTQPSI
CLAVA_2_Ng_Locus_2853   ---------- ---------- ---------- ---------- ---------- ---------- ----------
HYDRA_MAGNIPAPILLATA_   ---------- ---------- ---------- ---------- ---------- ---------- ----------
MONTASTRAEA_FAVEOLATA   MGCGGSKAIK NRTAPAPVAE QTQSPGLKET SPAKQHRQQP EGKRTEETVE ETAPADGYGR QGTEGDKQAT
MONTASTRAEA_FAVEOLATA   ---------- ---------- ---------- ---------- ---------- ---------- ----------
ACOELA                  .......... .......... .......... .......... .......... .......... ..........
SYMSAGITTIFERA_ROSCOF   ---------- ---------- ---------- ---------- -------MQL NNIQKCLPFL FPKRNSHMSD
SYMSAGITTIFERA_ROSCOF   ---------- ---------- ---------- ---------- ---------- ---------- ----------
NEMERTODERMATIDA        .......... .......... .......... .......... .......... .......... ..........
MEARA_STICHOPI_Ng1_tr   ---------- ---------- ---------- ---------- ---------- ---------- ----------
NEMERTODERMA_WESTBLAD   ---------- ---------- ---------- ---------- ---------- ---------- ----------
NEMERTODERMA_WESTBLAD   ---------- ---------- ---------- ---------- ---------- --------MG SYLSGLLWNS
NEMERTODERMA_WESTBLAD   ---------- ---------- ---------- ---------- ---------- ---------- ----------
XENOTURBELLIDA          .......... .......... .......... .......... .......... .......... ..........
XENOTURBELLA_HEMOGLOB   ---------- ---------- ---------- ---------- ---------- ---------- ----------
MOLLUSCA                .......... .......... .......... .......... .......... .......... ..........
SEPIA_OFFICINALIS_LGB   ---------- ---------- ---------- ---------- ---------M GALFSVIDAL IRKIGIMSPL
EUPRYMNA_SCOLOPES_1_L   ---------- ---------- ---------- ---------- ---------M GALLSVIDAV IRKIGIMSPR
DORYTEUTHIS_PEALEII_L   ---------- ---------- ---------- ---------- ---------- ---------- VRADGLPYPA
EUPRYMNA_SCOLOPES_2_L   ---------- ---------- ---------- ---------- ---------- ---------- ----------
MYTILUS_CALIFORNIANUS   ---------- ---------- ---------- ---------- ---------- ---------- ----------
MYTILUS_CALIFORNIANUS   ---------- ---------- ---------- ---------- ---------- ---------- ----------
SPISULA_SOLIDISSIMA_L   ---------- ---------- ---------- ---------- ---------- ---------- ----------
APLYSIA_1_LGB3_CNSN01   ---------- ---------- ---------- ---------- ---------- ---------- ----------
APLISIA_2_LGB2_CNSN01   ---------- ---------- ---------- ---------- ---------- --MGAEQSLS CFPCTSVRRC
LIOLOPHURA_JAPONICA_N   ---------- ---------- ---------- ---------- ---------- ---------- ----------
CRASSOSTREA_GIGAS_2_L   ---------- ---------- ---------- ---------- ---------- ---------- ----------
CRASSOSTREA_GIGAS_1_L   ---------- ---------- ---------- ---------- ---------- -------MGC DASKSTDVIE
ARTHROPODA              .......... .......... .......... .......... .......... .......... ..........
APIS_MELLIFERA_LGB2_g   ---------- ---------- ---------- ---------- ---------- -------MGT FLRFLGISSS
CARCINUS_MAENAS_1_LGB   ---------- ---------- ---------- ---------M GAVLSVVWGW LSPGTQVGAV TFPEEGSLGP
CARCINUS_MAENAS_2_LGB   ---------- ---------- ---------- ---------- ----MGCHIT KNKKNKETEE VKVIDLPEPP
DAPHNIA_PULEX_GlobinX   ---------- ---------- ---------- ---------- ---------- ---------- ----------
DAPHNIA_CARINATA_LGB2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
HARPEGNATHOS_SALTATOR   ---------- ---------- ---------- ---------- ---------- ---------- ----------
TUNICATA                .......... .......... .......... .......... .......... .......... ..........
Ciona_intestinalis_Gb   ---------- ---------- ---MQSMSTP NPNNQCVNGP NYFCSCCCGN SVCNSNWLTC VTAGSVAPTV
Ciona_savigny_Gb1 LGB   ---------- ---------- ---------- ---------- ---------- ---------- ----------
MOLGULA_TECTIFORMIS_L   ---------- ---------- ---------- ---------- ---------- ---------- ----------
BOTRYLLUS_SCHLOSSERI_   ---------- ---------- ---------- ---------- ---------- ---------- ----------
CEPHALOCHORDA           ---------- ---------- -------... .......... .......... .......... ..........
Branchiostoma florida   ---------- ---------- ---------- ---------- ----MGAFLT KPFSLVGRLL WKVLFSWWVK
Branchiostoma florida   ---------- ---------- ---------- ---------M GANMGCSNSK KMSHESESAN SGDSTPPKSS
Branchiostoma florida   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Branchiostoma florida   ---------- ---------- ---------- ---------- ---------- --MGTIADGE GTELNGYGGE
HEMICHORDATA            .......... .......... .......... .......... .......... .......... ..........
Saccoglossus kowalevs   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Saccoglossus kowalevs   ---------- ---------- ---------- ---------- ---------- ---------M GCTPSINERD
Saccoglossus kowalevs   ---------- ---------- ---------- ---------- ------MGCT VSTKPDLYSQ NGDADPYKKK
Saccoglossus kowalevs   ---------- ---------- ---------- ---------- ---------- ---------- -------MGC
BALANOGLOSSUS_CLAVIGE   ---------- ---------- ---------- ---------- ---------- --------MG SIIGYFWASS
ECHINODERMATA           .......... .......... .......... .......... .......... .......... ..........
STRONGYLOCENTROTUS_PU   ---------- ---------- ---------- ---------- ---------- ---------- ----------
KINORHYNCHA             .......... .......... .......... .......... .......... .......... ..........
ECHINODERES_HORNI_Glo   ---------- ---------- ---------- ---------- ---------- ------MGCG QSLESQGDGG
ENTOPROCTA              .......... .......... .......... .......... .......... .......... ..........
PEDICELLINA_SP_LGB1_E   ---------- ---------- ---------- ---------- ---------- ---------- ----------
TARDIGRADA              .......... .......... .......... .......... .......... .......... ..........
RICHTERSIUS_CORONIFER   ---------- ---------- ---------- ---------- ---------- --------MG GRASRILSSG


                        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                80         90        100        110        120        130        140
VERTEBRATE GlobinX      .......... .......... .......... .......... .......... .......... ..........
ANOLIS CAROLINENSIS G   SNGRTTEPFP LSGAQKELIR GSWEILHK-- -DIARVGIIV FIRLFETHPE -CKDVFFLFR -D--ID--DF
CALLORHINCHUS MILII G   EDVVAVASLS LSDRQTQLVK ETWRLVQE-- -DIAKVGIIM FVRLFETHPE -CKDAFFLFR -D--ID--DL
```

```
DANIO RERIO GbX (from       RPGEEETPAG LTTNHIRLIK ESWRLIQE-- -DIAKVGIIM FVRLFETHPE -CKDVFFLFR -D--VE--DL
ORYZIAS LATIPES GbX (       RSREDAAVEH PNEEQIQMIK DSWKVIRD-- -DIAKVGIIM FVRLFETHPE -CKDVFFLFR -D--VE--DL
PETROMIZON MARINUS Gb       GHQPPGPPQA PSESQRRLVR DSWLALQC-- -DIARVGVIM FVRLFETHPE -CKDVFYQFR -D--CE--DL
PYTHON MORULUS GbX (f       ---------- ----QKELIR ESWKILHK-- -NITRVGIIV FIRLFETHPE -CKDVFFLFR -D--ID--DL
TETRAODON NIGROVIRIDI       DAAAAAAVVY PREDQIQMIK DSWKVIRD-- -DIAKVGIIM FVRLFETHPE -CKDVFFLFR -D--VE--DL
XENOPUS TROPICALIS Gb       ETSPLLPTLN LSEQQQQLLV ESWRLIQH-- -DIAKVGVIL FVRLFETHPE -CKDVFFLFR -D--VD--DL
VERTEBRATE MYOGLOBINS       .......... .......... .......... .......... .......... .......... ..........
HOMO_SAPIENS_MYOGLOBI       --------MG LSDGEWQLVL NVWGKVEA-- -DIPGHGQEV LIRLFKGHPE -TLEKFDKFK -H--LK--SE
DANIO_RERIO_MYOGLOBIN       ---------- --MADHDLVL KCWGAVEA-- -DYAANGGEV LNRLFKEYPD -TLKLFPKFS -G--IS---Q
BOS_TAURUS_MYOGLOBIN_       --------MG LSDGEWQLVL NAWGKVEA-- -DVAGHGQEV LIRLFTGHPE -TLEKFDKFK -H--LK--TE
GALLUS_GALLUS_MYOGLOB       --------MG LSDQEWQQVL TIWGKVEA-- -DIAGHGHEV LMRLFHDHPE -TLDRFDKFK -G--LK--TP
VERTEBRATE CYTOGLOBIN       .......... .......... .......... .......... .......... .......... ..........
DANIO_RERIO_CYTOGLOBI       GVQLTQSPDS LTEEDVCVIQ DTWKPVYA-- -ERDNAGVAV LVRFFTNFPS -AKQYFEHFR -E--LQ--DP
OVIS_ARIES_CYTOGLOBIN       EIERRERSEE LSEAERKAVQ ATWARLYA-- -NCEDVGVAI LVRFFVNFPS -AKQYFSQFK -H--ME--EP
HOMO_SAPIENS_CYTOGLOB       EIERRERSEE LSEAERKAVQ AMWARLYA-- -SCEDVGVAI LVRFFVNFPS -AKQYFSQFK -H--ME--DP
VERTEBRATE HEMOGLOBIN       .......... .......... .......... .......... .......... .......... ..........
EQUUS_CABALLUS_HB_ZET       --------MS LTKAERTMVV SIWGKISM-- -QADAVGTEA LQRLFSSYPQ -TKTYFPHF- ----------
HOMO_SAPIENS_HB_BETA_       -------MVH LTPEEKSAVT ALWGKV---- -NVDEVGGEA LGRLLVVYPW -TQRFFESFG -D--LS--TP
TAENIOPYGIA_GUTTATA_H       --------MV LSAGDKSNVK AVFGKIGG-- -QADEYGADA LERMFATYPQ -TKTYFPHF- ----------
ORYCTOLAGUS_CUNICULUS       --------MA LSAAERALLR ALWKKLGS-- -NVGVYATEA LERTLEAFPR -TKIYFSHM- ----------
VERTEBRATE NEUROGLOBI       .......... .......... .......... .......... .......... .......... ..........
HOMO_SAPIENS_gi|10864       ---------- ---P--ELIR QSWRAVSR-- -SPLEHGTVL FARLFALEPD -LLPLFQYNC -R--QFS-SP
DANIO_RERIO_gi_188590       ---------K LSEKDKGLIR DSWESLGK-- -NKVPHGIVL FTRLFELDPA -LLTLFSYST -N--CG--DA
GALLUS_GALLUS_gi_1548       ---------M LSRTQQALIR ESWRRVSG-- -SPVQHGVVL FSRLFDLDPD -LLPLFQYNC -K--RFA-SP
MONODELPHIS_DOMESTICA       --------RR LSGPEQELIR ESWQKVNS-- -NPLQHGMIL FTRLFDLEPD -LLPLFQYNC -R--QFS-SP
CHIONODRACO_MYERSI_gi       ---------K LSGKDKELIR GSWESLGK-- -NKVPHGVVM FSRLFELDPE -LLTLFHYTT -N--CG--ST
BOVICHTUS_VARIEGATUS_       ---------K LSGKDKELIR GSWESLGK-- -NKVPHGVVM FSRLFELDPE -LLTLFHYTT -N--CG--ST
SIPUNCULA                   .......... .......... .......... .......... .......... .......... ..........
THEMISTE_1_LGB2_gnl|t       MNQAPEDQTK FTPDELMALR LSWNAWVA-- GDLAGXGFDM FVKMFEQRKE -TKEVFPFAR -G--N---DA
THEMISTE_2_LGB2_gnl|t       ---------- --------LR LSWNAWVA-- GDLVGRGFDS FVKMFEQKKE -TKEVFSFAR -G--N---DA
THEMISTE_3_LGB2_gnl|t       MNLDLEDQCR FDPEELINLR LSWNAWVA-- GDLPGRGFDM FTKMFEQRKE -TKEVFSFVR -G--S---DA
BRACHIOPODA                 .......... .......... .......... .......... .......... .......... ..........
TEREBRETALIA_Ng2_gnl|       ---------- FTTQEKNRLR SSWRAFVA-- EDPAKRGYQM FMKMFENRPD -TKKFFK-FV -Q--KD--AL
ANNELIDA                    .......... .......... .......... .......... .......... .......... ..........
ALVINELLA_POMPEJANA_1       ---------- ---------- ----YRLS-- -DISERGMDI FVRLFELHPV -YKSYFQKLR -D--V---DI
ALVINELLA_POMPEJANA_2       ASYKPDPRCP LTERQLYSIT KSWKAINR-- -EMASTAVNM FIRLLEHDGI --RSFFTKFK -D--HK--TV
APHRODITA_ACULEATA_LG       ---------g lsgaDIAVIR STWAKVQG-s gSATDIGRSI FIKFFELDPA -AQNEFPCKG -E------SL
CAPITELLA_SP_1_LGB2_F       -------DQL LTPEEIVLVR VTWEQLKT-n lTLANLGKKV FLRIFNLKPD -IKKLFPFSD -V------WG
CAPITELLA_SP_2_LGB2_F       kpsnipqHEF LTQNQKLGAK STWEFLCH-t sTPTERGMRV FLRIFEIAPV -TKTLFPFKD -M------PN
CAPITELLA_SP_3_LGB3_g       ---------L FTDRQKAIIT KTWRHMGN-- -DLTGRGSKV FLKIFNLHPE -VKQLFPSLK -N--D---NE
PLATHYNEREIS_dumerili       ---------P LNDNQKELIK KLWEIVRE-- GDIESTGVSM YMHMFTLQPE -VMKMFSFVP -K--NVT-NP
HELLOBDELLA_ROBUSTA1_       NLNYLNNDDV LTIREKQLVR ESWTLLSI-- -KLKSLGKQV FLRIFELRPS -TKNLFPFKT -V------WG
ECHIURA                     .......... .......... .......... .......... .......... .......... ..........
URECHIS_CAUPO_Ng1_gnl       ---------- VSLQNIALVL KSWAILKK-- -DLLSNGAAL FLALFEAYPD -YKDLFKQFP -G--L---RL
PLATYHELMINTHES             .......... .......... .......... .......... .......... .......... ..........
MACROSTOMUM_LIGNANO_1       ---------M LTADETQLIL SGWNQAMK-- -DAKGLGLDI FLTLFEMFPQ -HQELFrDFK -G--K---SR
MACROSTOMUM_LIGNANO_2       ---------M LNEVEKKIIL SGWQQAIK-- -DKKALGMDV FMTLFEMFPQ -HQELFrDFK -G--K---SR
CHOANOFLAGELLATE            .......... .......... .......... .......... .......... .......... ..........
MONOSIGA_BREVICOLLIS_       DEAHAPDYPS FDPVAVKLAR KQWKRVVQ-- -LVPNWHEVF FSYLFERAPY -ARTLFPFD- ---------V
SALPINGOECA_SP_Ng2_gi       --------MR LDMEQLKIAL GSWTAVVE-- -LVPTWHEVF FAELFQAHPE -TERLLYSS- ----------
PLACOZOA                    .......... .......... .......... .......... .......... .......... ..........
TRICHOPLAX_ADHAERENS_       DSVQTPPQPS LTEEQKAIIR ENWQDVEE-- -NMSEVGLYL FSKLFTIAPE -YREVFPFE- ----------
TRICHOPLAX_ADHAERENS_       --------MA PTAQDLQTIR ETWALVAP-- -DLKKHGTVL FLRLFEQHPD -VQRLFEKIK -D--V---PH
TRICHOPLAX_ADHAERENS_       ---------- ---------- ---------- -----MVLVN NYSLIKLSPA -TKIYFHGVD -F--EK--RD
TRICHOPLAX_ADHAERENS_       VVTGNDIKSY LNYQERQAII DSWNAIST-- -EKQKYGTIL FLKLFELEPR -VKSLFTIFD -F--NE--PL
PORIFERA                    .......... .......... .......... .......... .......... .......... ..........
CARTERIOSPONGIA_FOLIA       ------MSHP LSSEELKLVE KSWALIQE-- FGLQEAGMVM FKRLFEMSPG -LQNLFPFA- --------GD
AMPHIMEDON_QUEENSLAND       ---------S LTSAQVALIE STWKVVKK-- -DLQGAGNIM FLKLFQIDVS -VRDKFPFRD -V------PY
CNIDARIA                    .......... .......... .......... .......... .......... .......... ..........
NEMATOSTELLA_VECTENSI       -TENTLTSVP LSTRRKKLVR ESWELIEP-- -VKITIGKRL FTRLFDVNPN -MQDTFPNFK -G--K---EL
NEMATOSTELLA_VECTENSI       QLERINPITG LSAREVAVVK QTWNLVKP-- -DLMGVGMRI FKSLFEAFPA -YQAVFPKFS -D--V---PL
NEMATOSTELLA_VECTENSI       -----NPQNA FSAADIQAIQ GTWALAK--- PDLMGKGAMV FKQLFTEHGY --QPLFSNLA -Q--Y---EI
NEMATOSTELLA_VECTENSI       ---------- ---------- ---------- ---------- ---------- ----CFKAFN -K--V---SL
NEMATOSTELLA_VECTENSI       KPPREPVKIP LSVAQKYLVR ETWETIEQ-- -HSKAVGKKT FLRFFEMNPD -YQKLFPEFA -T--L---DQ
NEMATOSTELLA_VECTENSI       NQSFVGNRPR LTERQIKLVQ DTWRLLIP-- -SQKKTAMIF YLKLFTLDPI -FKEVFS-FH -T--E---NE
NEMATOSTELLA_VECTENSI       DTENTLTSVP LSTRRKKLVR ESWELIEP-- -VKITIGKRL FTRLFDVNPN -MQDTFPNFK -G--K---EL
NEMATOSTELLA_VECTENSI       QYLTKVDQLP LTETQKYYIK QSWMGLES-- -NKGELGIEI FLRLFSENPT -LQLMFPEFR -E--YS--TL
NEMATOSTELLA_VECTENSI       EKLFETSRIP LDAKETQLVR KTWAILGD-- -RQVEVGKSL FLRFFEEHPT -SKDLFPEFR -N--I---SN
NEMATOSTELLA_VECTENSI       HLCGLKLECD MTYEQKYLIR ETWKFLEV-- -SKKEIGVSV YKRFLNMHPG -LQTYFSEFK ---------H
NEMATOSTELLA_VECTENSI       KPPREPVKIP LSVAQKYLVR ETWETIEQ-- -HSKAVGKKT FLRFFEMNPD -YQKLFPEFA -T--L---DQ
NEMATOSTELLA_VECTENSI       SQANLPRTMP LSEAQKYLVR ETWETIEP-- -QKQTVGKKA FLRFFDMNPD -YQNLFPEFK -S--L---SY
CLYTIA_HEMISPHAERICA_       IQKSSNIKVP LSAKEVNILK KTWPPVKN-- -NWLKICLVA FERWFTLYPQ -LRYMFRSLP -E--DI--QF
CLYTIA_HEMISPHAERICA_       NGKVDAAPAD YTEAEINIVQ KQWSVAMR-- -NLDSVGYKL FAKLFQYKDI MAKFDFAVAN -N--L---DF
CLAVA_1_LGB1_Locus_11       HTKYGNICIP LTKKDIKKLR KSWGIMKM-- -NWYKICLVV FDRWFSMYPE -LRNMFKSFS -S--VL--TL
CLAVA_2_Ng_Locus_2853       ---------- ----EVGILR KQWSIMMS-- -DADANGIML FTALFAINGH -LRTKFAFGD -V--YGF-DC
HYDRA_MAGNIPAPILLATA_       ---------- LSGKEIETLK KSWTTAKQ-- -FWNEICTCA FSRWFSTYPE -IQSKFGVYG -D--NL--TM
```

```
MONTASTRAEA_FAVEOLATA    EEQSENVEGP ITQEQISLVQ DTWKLVNG-- -DLEQVGVEF YTRLFKENPE -LLQMFSFRD -L--ANS-TE
MONTASTRAEA_FAVEOLATA    -MERSKESDG LTDLQIEMIR SSWEKVTP-- -NKKHHGQLL FHKLFEIAPE -MTDLFPFG- ----------
ACOELA                   .......... .......... .......... .......... .......... .......... ..........
SYMSAGITTIFERA_ROSCOF    YFAFEGLPPV PNSSESEEIR KSWKKIEL-- -DAAKLGIAV FVGLFERYPE -IQASFSKIA -N--V---NK
SYMSAGITTIFERA_ROSCOF    --MATLESMQ VSEEQQSLIM EDVQVLLP-- -NYDDFVEDV LQQFMEENPE -TFQIFPWAD -A--SK--TA
NEMERTODERMATIDA         .......... .......... .......... .......... .......... .......... ..........
MEARA_STICHOPI_Ng1_tr    ---------- ---------- ---------- -----SRKIL VCRLFESHPQ -VRDIFEHSE ---------D
NEMERTODERMA_WESTBLAD    --------MG LTETQRVLIK QSWKKVIS-- GGRVEAGWVL FSKFFTSSPE -AQNYFKAFK -G--K---PL
NEMERTODERMA_WESTBLAD    VPDIPDELTK LTPSEKNALV DSWMLVID-- DGLQQGGIKL FLKFFTLDPA -ARPYFTKFL -K--L---SD
NEMERTODERMA_WESTBLAD    --------MA LTDSQQFLVK KVWDKVVA-- DGYHEPGLIF MNKFFTDTPE -AQNYFWFVK -D--R---PL
XENOTURBELLIDA           .......... .......... .......... .......... .......... .......... ..........
XENOTURBELLA_HEMOGLOB    -----MVFAG LSGGEVAAVK SSFGKVTD-- --LAGTGAEL FLKYFQARPS -AQNYFKWKG -Q------SL
MOLLUSCA                 .......... .......... .......... .......... .......... .......... ..........
SEPIA_OFFICINALIS_LGB    DDEEVDSKTG LKVKEKRAMT QSWKYISP-- -NLKAEGIHF FNMLFTEHPD -YIDYFPSFR -G--K---KL
EUPRYMNA_SCOLOPES_1_L    DDEEVDSKTG LKVKEKRAIT QSWKHVSQ-- -NLKQESMNF FNLLFTEHPA -YIEYFPAFC -G--K---KL
DORYTEUTHIS_PEALEII_L    PPPPTDPRLP LSPLQVFKLK KSWKGIKR-- -SIELTGVEM FVRMFRTQPG -LKNLFKDFR -E--LE--TD
EUPRYMNA_SCOLOPES_2_L    ---------- ---------- ---------- ---------- ---------- ---------- -I--E---TD
MYTILUS_CALIFORNIANUS    ---------- -GELATDTIR STWPLLSD-- -DIERTGIKV FLRIFYEEPK -IRNVFKRFG -E--M---DE
MYTILUS_CALIFORNIANUS    --------QT MESSTISAIK NTAAALEE-- -HSSQLGLKT YELLYEEHPE -VKATFNKS- ---------E
SPISULA_SOLIDISSIMA_L    ---DPCPVTK LTKAEKDAVA NSWAALKQ-- -DWKTIGADF FVKLFETYPN -IKAYFKSFD -N--M---DM
APLYSIA_1_LGB3_CNSN01    ---------- -----IFSLK QSWKGIKR-- -KMEDTGVDM FVRLFKTSSH -LQTMFQGFE -D--IR--SD
APLISIA_2_LGB2_CNSN01    RRRHCQLSKY LNPRQVQLVQ CTWSILKM-- -DLTTLGVTV FLNFFETEPD -LKMMFPKMI -RM-NESNQL
LIOLOPHURA_JAPONICA_N    ---------g ISADQAKALK DDIAVVAQ-- -NPNGCGKAL FIKMFEMNPG -WVEKFPAWK -G--K---Sl
CRASSOSTREA_GIGAS_2_L    -----MTSTE FTDEQKQLVK TTWNIVRE-- -DISKVGVIT FLRLFEKFPD -VQDLFVPFR -G--L---NS
CRASSOSTREA_GIGAS_1_L    NEEIESDTFD FTDTQIDTIR STWPILAC-- -DMVDIGSKV FLKIFIDEPK -LKYAFPSFS -D--M---EE
ARTHROPODA               .......... .......... .......... .......... .......... .......... ..........
APIS_MELLIFERA_LGB2_g    DDNRIDQATG LTERQKKLVQ NTWAVVRK-- -DEVASGIAV MTAFFKKYPE -YQRYFTAFM -D--T---PL
CARCINUS_MAENAS_1_LGB    EADVPDKTTG LTLRHRTAIY RTWDLVRP-- -NPKLHGINL FLTMFQEEPV -LQTRFKGFA -G--K---SI
CARCINUS_MAENAS_2_LGB    EPPPPDPRLP LTARQRFNII KSWKGIAR-- -AIEPTGVNM FVKLFENHSE -LITFFTKFR -Q--LR--TR
DAPHNIA_PULEX_GlobinX    ---------- ---------- ----RLEK-- -DIAQVGIIV FINLFETHPD -MQSVFLPFT ----------
DAPHNIA_CARINATA_LGB2    -------MDT LKTVNVAAVQ NTWAIVKK-- -DLNTHAPQF YVALLTAHPE -YQPMFPTIA -N--V---PA
HARPEGNATHOS_SALTATOR    ------EKTG MSEKQKKLVQ NTWAIVRK-- -DDVSSGLAI MNAFFTRYPE -YQDQFKSFK -G--I---PF
TUNICATA                 .......... .......... .......... .......... .......... .......... ..........
Ciona_intestinalis_Gb    STTVPPADEG LKRSDIINIQ DSWNTLKGF- -GYETVGMLV LHRLFNDAPQ -TRYLFSQLS -LSSNESFTL
Ciona_savigny_Gb1 LGB    --------ME MNAQEIQDVR DSWKRLCAD- -GEKTVGLML MQKLFNTYPE -SIKVFSRLG -ITNKAIITI
MOLGULA_TECTIFORMIS_L    -------MGG LTQEEIQAVR ESWAAVIKK- VGVTETGLAV LHRLFADVPE -TKTLFYNIG -LDSYDTITM
BOTRYLLUS_SCHLOSSERI_    ----MEEKSL FTAEESDAIK SSWSKMTE-- NGVSTAGRLI LCRLFQDVPE -VTTLFYRLG -ISGDVSYTL
CEPHALOCHORDA            .......... .......... .......... .......... .......... .......... ..........
Branchiostoma florida    QIETPSDVTG LTPTQSRLVK ESWKMFLS-- -KKRENGFVI FRVLFTDYPV -TRKLFKGVE -Q--LDLDAP
Branchiostoma florida    TPSALDERLP LTQKQKFLLL KSWKGVAR-- -QISQCGKTM LIRLFKDDPQ -LMAVFNQKF -R-HLRERDA
Branchiostoma florida    --------MG LTSEDKSAVL DSWAKMSGP- -TFQDAGEKV FLLLLKTDS- -TKALFPKFR -D-----IPY
Branchiostoma florida    KEPGGGHGGP LTQEQVHGIK ETWAILAQ-- -DPVERGVDL FMKIFEEDPD -LKKLFYFAD -D------GR
HEMICHORDATA             .......... .......... .......... .......... .......... .......... ..........
Saccoglossus kowalevs    ---MADPVTT LTSDEVAAIK SSWSAVYD-- -KKKESGVTL FVKLFTENPS -FKSQFGYMS -G-V----AD
Saccoglossus kowalevs    FQTPVDDKHL LDDRQKRIVR KTWRPLAN-- -DMTENGQKI FINIFESHPE -IKYMFPTRD -I-E----GR
Saccoglossus kowalevs    KVAPVDSRLP LTARQKFQIT KSWKGIAR-- -NMENTGKSM FMRLFQSNIE -LKNMFTGFE -E-FD---DL
Saccoglossus kowalevs    LISSNGEHPM ITKEQTKILT STWHSIHG-- -DLEKIGLLM FMGMFDNYPE -TRQFFGLSG -G-SI---VL
BALANOGLOSSUS_CLAVIGE    DGNVPDPGTT LTPNERRAIR DTWNAIYE-- -NKGENGVAF FMKLFEAHPV -YKKLFKNLE -G-ID---DM
ECHINODERMATA            .......... .......... .......... .......... .......... .......... ..........
STRONGYLOCENTROTUS_PU    ---------- LDERQKTLVE NTWKTLEK-- -NTELYGSIM FAKLTTDHPD -IGKLFPFGG -K--NL--TY
KINORHYNCHA              .......... .......... .......... .......... .......... .......... ..........
ECHINODERES_HORNI_Glo    AADIDSNKCP LDRRQIFELT KSWRMISK-- -NMESAGVMM FVKLFTTTNE -INHMFAHLI -T--EDG-KV
ENTOPROCTA               .......... .......... .......... .......... .......... .......... ..........
PEDICELLINA_SP_LGB1_E    ---IRDRLLS LDSNEMMAMQ LMWNDWVG-- GDAAARGFDS FILMFKQHPE -TQTLFDFAR -G--C---SA
TARDIGRADA               .......... .......... .......... .......... .......... .......... ..........
RICHTERSIUS_CORONIFER    SANVPDKVTG LSDVEKRAIQ ENWRIVYR-- -DLKGNGVEL FVRYFSRYPE -YKDAFESLR -D--I---RM


                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                150        160        170        180        190        200        210
VERTEBRATE GlobinX       .......... .......... .......... .......... .......... .......... ..........
ANOLIS CAROLINENSIS G    QQLKMSKELQ AHGLRVMSFI EKSVARM--- DQ-EPKLHHL AFELGRSHCR YK----APPK YYEYIGIQFI
CALLORHINCHUS MILII G    QQLRKSKGLR AHGLRVMSFI EKTVARL--- DQ-EDRLQQL XLELGKSHFR YS----AAPK YYPYVGNEFI
DANIO RERIO GbX (from    ERLRTSRELR AHGLRVMSFI EKSVARL--- DQ-LERLETL ALELGKSHYR YN----APPK YYGYVGAEFI
ORYZIAS LATIPES GbX (    ERLRTNRELR AHGLRVMSFI EKSVARL--- DQ-PERLEAL AVELGKSHYH YN----APPK YYNYVGAEFI
PETROMIZON MARINUS Gb    QKLKMNKQLQ AHGLRVMSFI EKSVARL--- EQ-ECVLEQL IVEMGRKHYK YN----ASPK YYSFVGIEFI
PYTHON MORULUS GbX (f    QQLKMNKELQ AHGLRVMSFI EKSVARL--- DQ-EGKLEVL AFELGRSHFR YK----APPN YYEYIGIQFI
TETRAODON NIGROVIRIDI    ERLRSSRELR AHGLRVMSFI EKSVARL--- DQ-QDRLEAL AVELGKSHYH YN----APPK YYSYVGAEFI
XENOPUS TROPICALIS Gb    QALRANKDLR AHGLRVLSFV EKSVARI--- AD-CARLEEL ALELGRSXYR YN----APPR YYQYVGTEFI
VERTEBRATE MYOGLOBINS    .......... .......... .......... .......... .......... .......... ..........
HOMO_SAPIENS_MYOGLOBI    DEMKASEDLK KHGATVLTAL GGILKK---- ---KGHHEAE IKPLAQSHAT KHK---IPVK YLEFISECII
DANIO_RERIO_MYOGLOBIN    GDLAGSPAVA AHGATVLKKL GELLKA---- ---KGDHAAL LKPLANTHAN IHK---VALN NFRLITEVLV
BOS_TAURUS_MYOGLOBIN_    AEMKASEDLK KHGNTVLTAL GGILKK---- ---KGHHEAE VKHLAESHAN KHK---IPVK YLEFISDAII
GALLUS_GALLUS_MYOGLOB    DQMKGSEDLK KHGATVLTQL GKILKQ---- ---KGNHESE LKPLAQTHAT KHK---IPVK YLEFISEVII
VERTEBRATE CYTOGLOBIN    .......... .......... .......... .......... .......... .......... ..........
DANIO_RERIO_CYTOGLOBI    AEMQQNAQLK KHGQRVLNAL NTLVENL--- RD-ADKLNTI FNQMGKSHAL RHK---VDPV YFKILAGVIL
```

```
OVIS_ARIES_CYTOGLOBIN        LEMERSPQLR KHACRVMGAL NTVVENL--- HD-PEKVSSV LALVGKAHAL KHK---VEPV YFKILSGVIL
HOMO_SAPIENS_CYTOGLOB        LEMERSPQLR KHACRVMGAL NTVVENL--- HD-PDKVSSV LALVGKAHAL KHK---VEPV YFKILSGVIL
VERTEBRATE HEMOGLOBIN        .......... .......... .......... .......... .......... .......... ..........
EQUUS_CABALLUS_HB_ZET        DLHEGSPQLR AHGSKVAAAV GDAVKSI--- ----DNVAGA LAKLSELHAY ILR---VDPV NFKFLSHCLL
HOMO_SAPIENS_HB_BETA_        DAVMGNPKVK AHGKKVLGAF SDGLAHL--- ----DNLKGT FATLSELHCD KLH---VDPE NFRLLGNVLV
TAENIOPYGIA_GUTTATA_H        DLGKGSAQVK GHGKKVAAAL VEAANNV--- ----DDLAGA LSKLSDLHAQ KLR---VDPV NFKLLGQCFL
ORYCTOLAGUS_CUNICULUS        DLSPGSAQVR AHGRKVADAL TLAADHL--- ----DDLPGA LSALSDLHVR TLR---VDPH HFGLLGHCLL
VERTEBRATE NEUROGLOBI        .......... .......... .......... .......... .......... .......... ..........
HOMO_SAPIENS_gi|10864        EDCLSSPEFL DHIRKVMLVI DAAVTNV--- ED-LSSLEEY LASLGRKHRA VG----VKLS SFSTVGESLL
DANIO_RERIO_gi_188590        PECLSSPEFL EHVTKVMLVI DAAVSHL--- DD-LHTLEDF LLNLGRKHQA VG----VNTQ SFALVGESLL
GALLUS_GALLUS_gi_1548        QECLAAPEFL DHIRKVMLVI DAAVSHL--- ED-LPCLEEY LCNLGKKHQA VG----VKVE SFSTVGESLL
MONODELPHIS_DOMESTICA        QDCLSSPEFL DHIRKVMLVI DAAVTHV--- EN-LSSLEEY LTNLGKKHKA VG----VKLS SFSTVGESLL
CHIONODRACO_MYERSI_gi        QDCLSSPEFL EHVTKVMLVI DAAVSHL--- DD-LPSLEDF LLNLGRKHQA VG----VNTQ SFAEVGESLL
BOVICHTUS_VARIEGATUS_        QDCLSSPEFL EHVTKVMLVI DAAVSNL--- DD-LPSLEDF LLNLGGKHQA VG----VNTQ SFAEVGESLL
SIPUNCULA                    .......... .......... .......... .......... .......... .......... ..........
THEMISTE_1_LGB2_gnl|t        XQMHQIINVL FHVTKFMKNI DEVVKNA--- DR-LEDVVSM LRRVEGRHGH QGHN--VPSA YFPFLGAAMH
THEMISTE_2_LGB2_gnl|t        KQMQQSSNVL FHVTRVMKNI DEVVKHA--- DR-LEDVVSM LRQVGGRHGH QGHN--VPSA YFPYLGEAMR
THEMISTE_3_LGB2_gnl|t        HQMQQSSKFL FHVTRVMKNI DEVVKHA--- DR-LEDVVSM LRQVGGRHGS SHHN--VPSA YFPFLGIALR
BRACHIOPODA                  .......... .......... .......... .......... .......... .......... ..........
TEREBRETALIA_Ng2_gnl|        GTMKSSTSLV FHTGRVAKYI QLVCDCL--- DD-PTEAVPL LRQLGGKHGC GPNNFNVDAT YFPELGKAMR
ANNELIDA                     .......... .......... .......... .......... .......... .......... ..........
ALVINELLA_POMPEJANA_1        EDLRQSGKLR VHSTSVMKSI TDLVETL--- DH-PPDLRDM AIKIAHPHFD RG----VRPS QYRELFAIIL
ALVINELLA_POMPEJANA_2        AELRASKVFE SHALMVISVI DDVITNL--- DD-MDYVMSL LQATGESHSI KFKN--FNPD FLWNVEGAFL
APHRODITA_ACULEATA_LG        AALKTNVLLG QHGAKFMEYI TTAVNGL--- dDYAGKAHGP LTELGSRHKT RG----TTPA NFGKAGEALl
CAPITELLA_SP_1_LGB2_F        DDLIRHPKFV LHSERFMLVV DCCVQNL--- ECIKSEhGEM LANLGRAHVN YkG---FSRE NFEVFMKAIW
CAPITELLA_SP_2_LGB2_F        EDLHRNSLFK GHATRFMKSV EFTMQNL--- DALDVIvNPT LVSIGNKHVH iKG---FHPD YLDTFQTALM
CAPITELLA_SP_3_LGB3_g        DQLLKNPCFR GHASRFMQSV GAVVENL--- DS-PGDLSPL LIDLGRKHVL FGG---FTPE YFAAFTEGMM
PLATHYNEREIS_dumerili        EDLKNSARFL RHARNLMTTV TTAVDML--- GD-MEGLSEV LVDLGRRHKK YR----AKTE HFPIVGRSLT
HELLOBDELLA_ROBUSTA1_        DKLIKHPLFL THSKRFVKVI GCVVDRL--- DYLQEECAQP LIELGKKHVS IEG---FLPD YYDVYIRAII
ECHIURA                      .......... .......... .......... .......... .......... .......... ..........
URECHIS_CAUPO_Ng1_gnl        EDLTRMPPLR ALGATFMHSL GSMVDNL--- GD-LECVVEL LRERTHTHWE RG----IRFE HFQNVFDLLP
PLATYHELMINTHES              .......... .......... .......... .......... .......... .......... ..........
MACROSTOMUM_LIGNANO_1        AELEKMPKLR AHGLRVVNTL DGAIQSL--- DD-MEVCVSS LELIGASHKG RN----MNAG HFEDLNKALD
MACROSTOMUM_LIGNANO_2        AELEKMPKMR AHALRVVNTL DGAIQSL--- DD-MEVCASS LELIGASHKS HH----LSAK HFEDLNAALA
CHOANOFLAGELLATE             .......... .......... .......... .......... .......... .......... ..........
MONOSIGA_BREVICOLLIS_        DRLQGNSSLA EHAKRVGQAL ETALQGL--- FE-YYSLVEV LEKLGRRHFK YG----VEPE HIDLFEETFY
SALPINGOECA_SP_Ng2_gi        --DKSKSWNE RHMARVGKSV GDVIKSL--- SN-YDDVIEH LTALGTRHAR YG----LHVD QLDLFINAFL
PLACOZOA                     .......... .......... .......... .......... .......... .......... ..........
TRICHOPLAX_ADHAERENS_        -TTTDNVRLR VHATGVMKTV GKAVQNL--- DQ-FSELQSA LSTLGQFHHR KA----IKFE NFQAVGQALI
TRICHOPLAX_ADHAERENS_        DQLATNENFV FHTTRVMETI DHAVKGI--- DN-LPALTVL LKQLGSSHAQ YN----VKKE YFKIGLRISE
TRICHOPLAX_ADHAERENS_        SYLAKNTFLR NHAARFMEAI NVIIGQD--- MD-IFSVESY FRVVGSKHHS YN----LKLE HVQDISDAFL
TRICHOPLAX_ADHAERENS_        EDIIQSPHFR SHAMRFMQSL ETGVLMG--- FD-KESCDFL FKSLGSRHHF YD----LKSE FLDVIPECIL
PORIFERA                     .......... .......... .......... .......... .......... .......... ..........
CARTERIOSPONGIA_FOLIA        SLSDDNEGMK HHSFIVMTSL DDCLKIV--- HD-VPKLKDE LISLGAVHHI QG----VTSE HFAPVGSALI
AMPHIMEDON_QUEENSLAND        EELEDSESFL KHSLQVMETI DLAITLLLG- GE-MEKLVEA LVDLGMAHAM QG----LKPE DFDHVGEALV
CNIDARIA                     .......... .......... .......... .......... .......... .......... ..........
NEMATOSTELLA_VECTENSI        KDILNSRSLY LHAKRVMVAV ENAVTVL--- DD-AETFESY LINLGGRHLP WG----VTKD HFGVVGEAFI
NEMATOSTELLA_VECTENSI        DKLEDTPAVG KHAISVTTKL DELIQTL--- DE-PANLALL ARQLGEDHIV LK----VNKP MFKSFGKVLV
NEMATOSTELLA_VECTENSI        TGLEGSPELN THARNVMAQL DTLVGSL--- QN-SIELGQS LAQLGKDHVP RK----VNRV HFKDFAEHFI
NEMATOSTELLA_VECTENSI        EDLEKSPFLK AHATSVMSAI NEVVCNL--- DE-VEILGIL LEKIGFSHAR RE----IRRI HFENLAKVVV
NEMATOSTELLA_VECTENSI        VELEQANALH GHAERVMKAV ENAVSAM--- DD-AESFAAY LENLGARHKA RA----LKPA YLDAMQVAYT
NEMATOSTELLA_VECTENSI        GQLEQDERFL FQSRKFMEMI NSAVDRL--- ND-ISLLVMI LKSLGEVHWT KF---KIKPE YYEPVGKALI
NEMATOSTELLA_VECTENSI        KDILNSRSLY LHAKRVMVAV ENAVTVL--- DD-AETFESY LINLGGRHLP WG----VTKD HFGVVGEAFI
NEMATOSTELLA_VECTENSI        EELKESRSLQ GHTKRVMKVV ENAVNSL--- ED-GHALMEY LQELGRRHKT RQ--IKPTVS NLQEISQAIN
NEMATOSTELLA_VECTENSI        EKIAESPALY GHARRVMKSV DNAVASI--- EN-VQVYSAY LYELGTRHQT RQ----LSEE QLKFMGGAFL
NEMATOSTELLA_VECTENSI        IKIDNINGSH GHPRRLLMAI DNAVTAL--- GD-SDSFSAY LVELGRRHHG MNFRPGPTHF NDLRKCFLSV
NEMATOSTELLA_VECTENSI        VELEQANALH GHAKRVMKAV ENAVSAM--- DD-AESFAAY LENLGARHKA RA----LKPA YLDAMQVAYT
NEMATOSTELLA_VECTENSI        EELQKANALH GHAKRVMKAV ENAVMSI--- DD-VMSFSAY LEELGRRHKT RA----LKPS YLEAMHGALM
CLYTIA_HEMISPHAERICA_        EDLFKTDALK MHVDKVRDVL ELLIKKI--- DN-VEELVNT LVDFGRQHHM LG----AEQR YATALAASFQ
CLYTIA_HEMISPHAERICA_        QKSMNDARLS FHIRRVFHTI NTVVSCL--- ND-GDFVASQ LEHVGAIHAE YG----LQAT HLARFKDVML
CLAVA_1_LGB1_Locus_11        QELLASRTMK SHIQQLEELV EDLLYKV--- DD-PTDFIET VIQFGEGHHK LG------AK QLYATSLAAA
CLAVA_2_Ng_Locus_2853        KTTFGDERLL GHLRGVFSSV EMLINSLVQ- NT-FYESLGK IKELGKVHSL AG----IHLS DLQAFKKAMI
HYDRA_MAGNIPAPILLATA_        NEVLASESLC IHIRKSVELI EIIIKKV--- DE-RHELSEY LIELGKLHHK FG----AEQK YATALGSSFV
MONTASTRAEA_FAVEOLATA        DAMRTDDRFK RQGLVTMQHV DLAVASL--- SD-LGSIVPA LKDLGARHSM YK----VEEH HFGPVG----
MONTASTRAEA_FAVEOLATA        -DDFTKPQFT THALNIMNAL DHAIQNL--- DN-PDVLIPK LRELGQMHAG FE----LTIK EFQHVGEALI
ACOELA                       .......... .......... .......... .......... .......... .......... ..........
SYMSAGITTIFERA_ROSCOF        SSLNSNVMLH AHSLHIMHMI GKLIQLL--- NE-PENLLSK VVELGERHFD RK----ANDE LLQYFCPAY-
SYMSAGITTIFERA_ROSCOF        KEMRSHPRFK SHAKSIGKVI SDCLVDL--- NG-VKKHEPK LSSLGAMHTK KK----VPTE LFGKLGGCIL
NEMERTODERMATIDA             .......... .......... .......... .......... .......... .......... ..........
MEARA_STICHOPI_Ng1_tr        KDLVPADIIK AHGMRIMGLL GRFVSNLESN VD-DNLLGQP IHDLGRKHVD FK----APPY LFDVFALQIQ
NEMERTODERMA_WESTBLAD        SELQNNTQMK GHVLRVINYI TDIVDTL--- EV-DEMREEM SINIGRTHGR RA----IPAE MFQCLKPAVF
NEMERTODERMA_WESTBLAD        DELRESKMLR AHVIRVMNTL NSLVDGL--- DD-PELFVEL SKFIGRTHYR HN----IDTE HLVNLGEAVI
NEMERTODERMA_WESTBLAD        YELRENIHME AHLTRGLSHI TCTINKV--- AE-PTVCADM WRNLGRTHTA HT----VPFK MFDLMKPAFV
XENOTURBELLIDA               .......... .......... .......... .......... .......... .......... ..........
XENOTURBELLA_HEMOGLOB        DAVKADAYLQ THGNRVMSTL KTVIGSI--- ED-AGKCEAL ITQLGNLHTG RN----VAAA DFEALFPIIA
MOLLUSCA                     .......... .......... .......... .......... .......... .......... ..........
```

```
SEPIA_OFFICINALIS_LGB     EEFNTKASFM AHAKNVFYAV TLIVDTL--- DD-ADELVEI LLKTGRNHHR RS----VPLS PFQGRAIVFE
EUPRYMNA_SCOLOPES_1_L     EEFNTKPKFI AHAKNVFYAL ILVVDSL--- DE-PDELVEI LLKTGRDHHN RG----VPMA AFHNLAIVFD
DORYTEUTHIS_PEALEII_L     DEMRENEALE KHATLVMNTL DDAITNI--- EN-VDLVLDL LHRIGKSHLR FQG---FNVE YFWLAEQPLL
EUPRYMNA_SCOLOPES_2_L     DQMRENEALE KHATLVMNTI DEVIANI--- EN-VDLVLDL LHRIGKSHCR FQG---FNVE YFWLAEQPLI
MYTILUS_CALIFORNIANUS     IELRRSPIFK EHAYRFMRVV DDLVDNM--- DPPKAHIQQN LMMLGAKHAT FEG---FRIE YFEAYSESLI
MYTILUS_CALIFORNIANUS     YSTEADAVPK DLAVKVGKAI ATFAKHC--- DD-LDSQKEL FGGIAIKHVK LGG---ILEE QYSAMEGSFL
SPISULA_SOLIDISSIMA_L     SEIKQSPKLR AHSINFCHGL NSFIQSL--- DE-PDVLVIL VQKLTVNHFR RK----IAVD RFQEAFALYV
APLYSIA_1_LGB3_CNSN01     DELRSNEALE YHATLVMTTL DDAITHI--- DN-YDFVKQL LSKTGASHVK FAG---FKST NFLAIKGPFL
APLISIA_2_LGB2_CNSN01     EWDVDKDMLQ KHAVTVMEGL GAAVETL--- ND-SQFLNTV LIALGQTHEK RN----IKPN MLKRMWPSMH
LIOLOPHURA_JAPONICA_N     dEIKASDKIT NHGGKVINEL ANWINNI--- ----NSASGI LKSQGTAHKG RS----IGIE YFENVLPVID
CRASSOSTREA_GIGAS_2_L     EELRQNVGLR EHGMRVMGTI EKCITRI--- DQ-PDKLTSM LESLGEKHVV FD----TKIE YFDLLSPQLI
CRASSOSTREA_GIGAS_1_L     NELLRHPPFI DHVTRFMQII DYLVENL--- DQQNSDFHQA LLMLGAKHAT YPG---FQVS HFTVFNKALL
ARTHROPODA                .......... .......... .......... .......... .......... .......... ..........
APIS_MELLIFERA_LGB2_g     NELPANKRFQ AHCAGVITAL NNVIDFL--- HD-PGLMEAS LIGLVERHKK RG----QTKE EFQNLKEVML
CARCINUS_MAENAS_1_LGB     EELKNSKRLA AHGTTVVMAI TAMVDNL--- ED-VSVLVEL LKNTGANHRD RG----VPKG DFELLAPVLV
CARCINUS_MAENAS_2_LGB     DEQAESLELA EHATIVMNSI DEGIKAM--- DN-VDFFFDL LHQIGASHRK IPG---FKKE YFWKIEHPFL
DAPHNIA_PULEX_GlobinX     GVVLDDLKKS KLLSEHALRV MGAVQRAVHR LQEPEKLHAF LSELGRKHEK NG----AKLE YIDYIGPQFL
DAPHNIA_CARINATA_LGB2     GELLNNPALK TLSVNVLTKL SELIDCMG-- --NPDALQGQ LVDLANQHKQ RG----TTRA HFDNLSKVLI
HARPEGNATHOS_SALTATOR     EELSKNKKFQ AHCVSVIAGL SNVIDHI--- HN-PELMEAS LINLAERHKN RG----QTRE HFQNLRYVLE
TUNICATA                  .......... .......... .......... .......... .......... .......... ..........
Ciona_intestinalis_Gb     EQMRNNSRVV YHANRVARAV GRLVDLI--- ELP-TNFTDH LVWLGQRHAY HG----VAPV NFDYMGPVLL
Ciona_savigny_Gb1 LGB     DDLSTNSAAS RHAESLTSRI GTLVDLM--- HNT-HEFKEC STEVGEIHIK YG----VTAE HVDILGNVLL
MOLGULA_TECTIFORMIS_L     DQLKDNKKIR NHALRVVTSI SNVLKNI--- KN-GEKLAEL FKKLGEIHKK NK----VPPE YFDKMGNVLL
BOTRYLLUS_SCHLOSSERI_     ESLSQKSQFN SHAKRLSLAL DAIVKKL--- DD-TAFITQK CTDLGKAHEK HN----VKPH YFDLLGQVLV
CEPHALOCHORDA             .......... .......... .......... .......... .......... .......... ..........
Branchiostoma florida     GQLESSITLR AHVTRFMHSF DTYMESL--- DDP-EDLKQL LYDTGKSHLI HD----IKPE YFDVLETVLM
Branchiostoma florida     DVLYQDAILD AHAATVMEAL HEAITHL--- DDS-VFVMKV LHDVGKMHQR YN----VDPS VFLKVEKPFL
Branchiostoma florida     DQLAGHPDVR DHGGKVMQVL DDFIKGL--- DNG---GDGA VQKVGLLHKG VG----VSHD NINLMKPVLM
Branchiostoma florida     ELSREDQRMR SHGERVMEAV GGAVDSL--- GDL-TAVVPV LTELGALHHK YG----VQPS YFDTVGAALI
HEMICHORDATA              .......... .......... .......... .......... .......... .......... ..........
Saccoglossus kowalevs     GDMKTLPALE NHGVKVMDRI NEWMGNL--- -TNGAELVKQ LKHLGTTHIA LK----VTED NFNAMDSVLM
Saccoglossus kowalevs     DNLSANPHFR MHSSRFMQSV GAAIDNL--- NDLDNALRPL LVKLAKTHVR FK---GFKPD YFDAFEEAML
Saccoglossus kowalevs     EDMRESQQLE NHASLVMYTI DEAIASI--- DDI-DFVVEL LGKIGRTHTR TD----FNPQ LFWRIEQPFL
Saccoglossus kowalevs     EDPAVIQKIR EHGLRFMTTA RKLVMNL--- DDK-DKFDRI LLDLGRRHHG YK----ADVD LIEVFGQQFI
BALANOGLOSSUS_CLAVIGE     AKLSKHPRMK AHGLRVMASF NSFIENL--- DD-AEVLVQL LVDIGISHSK HK----VTEE QFNVSNHFLS
ECHINODERMATA             .......... .......... .......... .......... .......... .......... ..........
STRONGYLOCENTROTUS_PU     GELLVDPDVR VHGKRVIETL GSVVEDL--- DD-MELVIQI LEDLGQRHNA YN----AKKT HIIAVGGALL
KINORHYNCHA               .......... .......... .......... .......... .......... .......... ..........
ECHINODERES_HORNI_Glo     RNSEAQKQLE HHAAMVMYTL DESIQAF--- TD-LDVLFEL TTRIGQQHRK MDG---FSAN YFWKIEDPFL
ENTOPROCTA                .......... .......... .......... .......... .......... .......... ..........
PEDICELLINA_SP_LGB1_E     AQMQNSSRLL FHVTRVIHNI TDVMDHL--- HE-LEEVVPK LRQLGGRHGT NPGCYNVPGA YFPWLGDAMK
TARDIGRADA                .......... .......... .......... .......... .......... .......... ..........
RICHTERSIUS_CORONIFER     QDIGKSHKLR AHSVQVMQYG TKPRKAR--- YA-AGTFYGE PLNL---HTT TVDF------ ----------


                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                 220        230        240        250        260        270        280
VERTEBRATE GlobinX        .......... .......... .......... .......... .......... .......... ..........
ANOLIS CAROLINENSIS G     QAAQPILKEA WT-PETEKAW EGLFQYLAAT MRRGFYKEQK ATGKN----- ---------- ----------
CALLORHINCHUS MILII G     CAVQPILKEK WT-AEVEEAW KGLFHYLTSV MKKGYQDEER GSCPREKPKH GPNSV----- ----------
DANIO RERIO GbX (from     CAVRPILKDR WT-PELEEAW KTLFQYVTSI MREGFLEEER NKRSNTQTSS RERPDKRSTA I---------
ORYZIAS LATIPES GbX (     CAVQPILKEQ WT-TELEKAW QTLFQFVTAL MKQGYQEE-S ARQRQLA-TS PKDRLDKRNT AL--------
PETROMIZON MARINUS Gb     ATVQPFLQEK WT-NEVEDAW QCLFRYIAAV MKRGYLEEEA ASNGVNTANY DRGQGNHGAT AM--------
PYTHON MORULUS GbX (f     QAVQPILKED WT-LEVEKAW K--------- ---------- ---------- ---------- ----------
TETRAODON NIGROVIRIDI     CAVQPILKER FT-SELEEAW KTLFQYVTGL MRKGHQEEGS RQRHLALPPK DGPEKRTSAL ----------
XENOPUS TROPICALIS Gb     SAVCPMLHDK WT-AEVEEAW KGLFAYICTV MERGYQEEE- RRHSDGRSLI DGLQGNKGLI ----------
VERTEBRATE MYOGLOBINS     .......... .......... .......... .......... .......... .......... ..........
HOMO_SAPIENS_MYOGLOBI     QVLQSKHPGD FG-ADAQGAM NKALELFRKD MASNYKELGF QG-------- ---------- ----------
DANIO_RERIO_MYOGLOBIN     KVMAEKAGLD AA-GQGALRR VMDAVIGDID GYYKEIGFAG ---------- ---------- ----------
BOS_TAURUS_MYOGLOBIN_     HVLHAKHPSD FG-ADAQAAM SKALELFRND MAAQYKVLGF HG-------- ---------- ----------
GALLUS_GALLUS_MYOGLOB     KVIAEKHAAD FG-ADSQAAM KKALELFRND MASKYKEFGF QG-------- ---------- ----------
VERTEBRATE CYTOGLOBIN     .......... .......... .......... .......... .......... .......... ..........
DANIO_RERIO_CYTOGLOBI     EVLVEAFPQC FSPAEVQSSW SKLMGILYWQ MNRVYAEVGW EN-------- ---------- ----------
OVIS_ARIES_CYTOGLOBIN     EVIAEEFASD FP-PETQRAW AKLRGLIYSH VTAAYKEVGW VQ-------- ---------- ----------
HOMO_SAPIENS_CYTOGLOB     EVVAEEFASD FP-PETQRAW AKLRGLIYSH VTAAYKEVGW VQ-------- ---------- ----------
VERTEBRATE HEMOGLOBIN     .......... .......... .......... .......... .......... .......... ..........
EQUUS_CABALLUS_HB_ZET     VTLASRLPAD FT-ADAHAAW DKFLSIVSSV LTEKYR---- ---------- ---------- ----------
HOMO_SAPIENS_HB_BETA_     CVLAHHFGKE FT-PPVQAAY QKVVAGVANA LAHKYH---- ---------- ---------- ----------
TAENIOPYGIA_GUTTATA_H     VVVATRNPSL LT-PEVHASL DKFLCAVGTV LTAKYR---- ---------- ---------- ----------
ORYCTOLAGUS_CUNICULUS     VTLARHYPGD FG-PAMHASV DKFLHHVISA LTSKYR---- ---------- ---------- ----------
VERTEBRATE NEUROGLOBI     .......... .......... .......... .......... .......... .......... ..........
HOMO_SAPIENS_gi|10864     YMLEKCLGPA FT-PATRAAW SQLYGAVVQA MSRGWDG--- ---------- ---------- ----------
DANIO_RERIO_gi_188590     YMLQSSLGPA YT-TSLRQAW LTMYSIVVSA MTRG------ ---------- ---------- ----------
GALLUS_GALLUS_gi_1548     YMLENCLGAA FS-PDVREAW IELYGAVVKA MQR------- ---------- ---------- ----------
MONODELPHIS_DOMESTICA     YMLEQCLGST FT-VTMKEAW TQLYGAVVQA MSRGW----- ---------- ---------- ----------
CHIONODRACO_MYERSI_gi     YMLQCSLGQA YT-APLRQAW LNLYSIVVAA MM-------- ---------- ---------- ----------
```

```
BOVICHTUS_VARIEGATUS_     YMLQCSLGQA YT-APLRQAW LNLYSIVVAA MSR------- ---------- ---------- ----------
SIPUNCULA                 .......... .......... .......... .......... .......... .......... ..........
THEMISTE_1_LGB2_gnl|t     -TLIKANYKS YD-SKLDDCW VPLWNFMNNE MTTGQEVYRG GKI------- ---------- ----------
THEMISTE_2_LGB2_gnl|t     -TLIKANYKA YD-SKLDDCW VRLWDFINKQ MTTGQEYTEE EKS------- ---------- ----------
THEMISTE_3_LGB2_gnl|t     -NLMSQSYKG YD-SKLDDAW TRLWGFMTSQ MMYG------ ---------- ---------- ----------
BRACHIOPODA               .......... .......... .......... .......... .......... .......... ..........
TEREBRETALIA_Ng2_gnl|     QLMESSGLQG YD-QELDALW EKLYTWITKR MEEGM----- ---------- ---------- ----------
ANNELIDA                  .......... .......... .......... .......... .......... .......... ..........
ALVINELLA_POMPEJANA_1     EYLKDKAKVV FN-DEAEAAW QKLFDYVLDI TAAVMDLQIE KMG------- ---------- ----------
ALVINELLA_POMPEJANA_2     WAVKETLGDR YT-ISIENIY TITIRYILQS LHDAFTKHRE RQNSTNNDCE KTNLLNQELS TADRKT----
APHRODITA_ACULEATA_LG     ailasvvggd ft-paakdaw tkvyntisst mqa------- ---------- ---------- ----------
CAPITELLA_SP_1_LGB2_F     YVWYHQLKDS MD-SEVECAW KKLLlfiivq qragydaeke appnglsfll qt........ ..........
CAPITELLA_SP_2_LGB2_F     DIWDEDLGKK FS-KETKEAW IKIFALITRK VFEGfqeett rfrpplpyeg kq........ ..........
CAPITELLA_SP_3_LGB3_g     CIWSEELGKG FT-DEVSVAW KTVFDFIMSQ LQDGYA---- ---------- --........ ..........
PLATHYNEREIS_dumerili     HAISAATGDA FT-PETAEAF TAFYGVVTH- ---------- ---------- ---------- ----------
HELLOBDELLA_ROBUSTA1_     SIWKQELKDV YT-NELSEAW HKVLVYIVSK LKEGYETEWK VATYFNPQ-- ---------- ----------
ECHIURA                   .......... .......... .......... .......... .......... .......... ..........
URECHIS_CAUPO_Ng1_gnl     AFLKSKLGYN FD-DATGAAC AAASVMITVL QAELKTL--- ---------- ---------- ----------
PLATYHELMINTHES           .......... .......... .......... .......... .......... .......... ..........
MACROSTOMUM_LIGNANO_1     VVFTRRLGAA YT--DNKAVW MKLlegvipv Iqrgm----- ---------- ---------- ----------
MACROSTOMUM_LIGNANO_2     VVFERRLGKA FV--DNKAVW VKLLQGIipv Iqrgl----- ---------- ---------- ----------
CHOANOFLAGELLATE          .......... .......... .......... .......... .......... .......... ..........
MONOSIGA_BREVICOLLIS_     KTLAIGLGKK WN-PEARRAW EIVCGLILSP IRTGILQART KANHLRAKEA ERKRQLEMAA ARLEGRVASS
SALPINGOECA_SP_Ng2_gi     WTLGAGLGDS WD-HSVKKAW MHVLPFILSP LKSGLVVART LRNDYNTSGC LRCRRLLIPL HGRRLRPITV
PLACOZOA                  .......... .......... .......... .......... .......... .......... ..........
TRICHOPLAX_ADHAERENS_     QTLSDKLQEN FT-PEVHEAW SKTFDMITAA MKSGMN---- ---------- ---------- ----------
TRICHOPLAX_ADHAERENS_     F--------- ---------- ---------- ---------- ---------- ---------- ----------
TRICHOPLAX_ADHAERENS_     EMARNALKKK FT-KSTEAAW RSFFQMVTDA IKNGIMKAQN RN-------- ---------- ----------
TRICHOPLAX_ADHAERENS_     HTIKKGCGNN WS-NETADAW KIATKVLCEL FREGLETKPK K--------- ---------- ----------
PORIFERA                  .......... .......... .......... .......... .......... .......... ..........
CARTERIOSPONGIA_FOLIA     WALEAALKEK FT-PEVKAAW LALYKVVQSL MEQGMEEGI- ---------- ---------- ----------
AMPHIMEDON_QUEENSLAND     HALGVALGKE FN-DEAKKAW TLLYSVVTAK MKEG------ ---------- ---------- ----------
CNIDARIA                  .......... .......... .......... .......... .......... .......... ..........
NEMATOSTELLA_VECTENSI     WALQDVLGEG CT-SDVAEAW IDLYGYIVQA MLEG------ ---------- ---------- ----------
NEMATOSTELLA_VECTENSI     RLLENDLGQR FS-SFASRSW HKAYDVIVEY IEEGLQQSYK QDPVTGITDA EKVLVQESWD LLKPDLLGLG
NEMATOSTELLA_VECTENSI     PLMKADLGDE FT-PLAESAW KKAFDVMIAT IEQGQRARRS VATFLTNPVA ---------- ----------
NEMATOSTELLA_VECTENSI     AYLRQALGSH LT-EEGADAW RKALCVMIDI IEKGSTSERW ---------- ---------- ----------
NEMATOSTELLA_VECTENSI     DTIQDLLKTQ WT-DGTAEAW NKLFRFIADT MKHGLSS--- ---------- ---------- ----------
NEMATOSTELLA_VECTENSI     YSISKGLGSL FN-DEIGEAW QAMYDLMSGA MISGTKAVQA RSQNSL---- ---------- ----------
NEMATOSTELLA_VECTENSI     WALQDVLGEG CT-SDVAEAW IDLYGYIVQA MLEGLQQAKK GR-------- ---------- ----------
NEMATOSTELLA_VECTENSI     ETFEENLGIK WT-VEIAESW KLLLDYVMAM IIRGLRSP-- ---------- ---------- ----------
NEMATOSTELLA_VECTENSI     FAMRLHLRKE WS-RATSKAW EKIFSFMADA MMRGCKG--- ---------- ---------- ----------
NEMATOSTELLA_VECTENSI     IEEILATASL WD-FQVEEAW NRLFDSITAM ILRGIQLAKV ---------- ---------- ----------
NEMATOSTELLA_VECTENSI     DTIQDLLKTQ WT-DGTAEAW NKLFRFIADT MKHGLSS--- ---------- ---------- ----------
NEMATOSTELLA_VECTENSI     DTLRNLLQSQ WT-EETAEAW NKLFSFISTT MVRGLQSRD- ---------- ---------- ----------
CLYTIA_HEMISPHAERICA_     YGICMIMDVD ---SSVENAW DSLLRFVMDL LKLGMRMEKE AQEKESLNKG YNTEELLEKA QDGGEALDEN
CLYTIA_HEMISPHAERICA_     ETLEEAFKEG FQ-EDSKTAW SKIVDAIAKY MLKGIDKAKE QTKTSELSSD DGDGIKMEQ- ----------
CLAVA_1_LGB1_Locus_11     FQYAICISLN LD-RDTENAW DSLFRFLMDC LKFGMRNEMH KETTSPLDKL EVQDG----- ----------
CLAVA_2_Ng_Locus_2853     QTLKKELKSE FD-DKALKAW SKFIDIVIEA FDVNDDTDDE KTQPIGK--- ---------- ----------
HYDRA_MAGNIPAPILLATA_     FAISQICPNI DM--ITEGAW DSLFKYIVT- ---------- ---------- ---------- ----------
MONTASTRAEA_FAVEOLATA     ---------- ---------- ---------- ---------- ---------- ---------- ----------
MONTASTRAEA_FAVEOLATA     WVLATGLGDD FT-------- ---------- ---------- ---------- ---------- ----------
ACOELA                    .......... .......... .......... .......... .......... .......... ..........
SYMSAGITTIFERA_ROSCOF     VEAMAKKGQW K--KKTIIAW EKFFDFIRAA MVHGLKKRKG HSSISNTTSA ANTAAEKNHN SPSSSQ----
SYMSAGITTIFERA_ROSCOF     TQVVKRVSEA KWSEEKKEAW LKAYGIITVM VTE------- ---------- ---------- ----------
NEMERTODERMATIDA          .......... .......... .......... .......... .......... .......... ..........
MEARA_STICHOPI_Ng1_tr     HIIKSKLEQV WT-DEIGDAW KVMFDIIVFN LKSGQNQEMQ DRGITV---- ---------- ----------
NEMERTODERMA_WESTBLAD     GTVIDLNGGS LS-DEAAAAW GLLWEALVLC VLEGMKKPGQ PGMGQTHSHA DQIIGELLVT HPAHVVALYI
NEMERTODERMA_WESTBLAD     WLVEQSNGGR LP-PAVEDSW QRLWERITHF ILAGEEEESK TDSGGETGVS VEQNSHVIMS TGVQSNSLNF
NEMERTODERMA_WESTBLAD     NMVIQCCGGM VD-QDTLVAL SLYWDFIMRC VKEGMEEAAA KVWSDSTTLN GSLYLKIINT PLLYCPCHGR
XENOTURBELLIDA            .......... .......... .......... .......... .......... .......... ..........
XENOTURBELLA_HEMOGLOB     GVLSEKCGGA LA-GDASSGW QKLYNEIVRL IDAGQKTPAA GGARAEFLAA IGAN------ ----------
MOLLUSCA                  .......... .......... .......... .......... .......... .......... ..........
SEPIA_OFFICINALIS_LGB     KIMTERLGKA IP-PLGKGLV DKSIDCS--- ---------- ---------- ---------- ----------
EUPRYMNA_SCOLOPES_1_L     KFLTIRLGNN YT-PLAKESW TKALTVVNAV IEKGIEDGID CSECRYRERN RRST------ ----------
DORYTEUTHIS_PEALEII_L     DAIKITLGDR YS-DNMDIIY KLVIRFLLTE VTKGARVDVS ST-------- ---------- ----------
EUPRYMNA_SCOLOPES_2_L     DAIKITLGDR YS-DNMNIIY KLVIRFLLTE VTKGARANVS ST-------- ---------- ----------
MYTILUS_CALIFORNIANUS     DVWEYTIGEE FI-PEVRESW TEFFDYLVKY MCQGYNVFTN ET-------- ---------- ----------
MYTILUS_CALIFORNIANUS     KAVKEVLGDA AT-DEVLDHW KKGFHFLAEH IRGLEAEK-- ---------- ---------- ----------
SPISULA_SOLIDISSIMA_L     SYAQDHAKFD ---DFTAAAW TKTLKVVADV IG-------- ---------- ---------- ----------
APLYSIA_1_LGB3_CNSN01     EAVRVTLGDR YT-ENMQNIY TIAIIFILET LQQGMEEALE KAGSSEVAQG HIRV------ ----------
APLISIA_2_LGB2_CNSN01     VGLGAVLGEG YT-KEVSEAW CKLYSYICLQ MQIGMENPN- ---------- ---------- ----------
LIOLOPHURA_JAPONICA_N     ATFAQQMGGA YT-AAMKDAL KAAWTgvivp gmkagy---- ---------- ---------- ----------
CRASSOSTREA_GIGAS_2_L     QAITTRHWSV D--SQREQAW NNFYCTL--- ---------- ---------- ---------- ----------
CRASSOSTREA_GIGAS_1_L     EAWESAIGEE FI-PEVQNCW TQLFAYIMRY IVQGYELYFS EC-------- ---------- ----------
ARTHROPODA                .......... .......... .......... .......... .......... .......... ..........
```

```
APIS_MELLIFERA_LGB2_g      EVLRQALGKQ YT-PEVAEAW NKTLDMMFGK IYQ------- ---------- ---------- ----------
CARCINUS_MAENAS_1_LGB      RFLKDNLGSA WS-PVAEEAW TQAMKVINAV IFTSYDACSI LS-------- ---------- ----------
CARCINUS_MAENAS_2_LGB      EAVRLTLGDR YT-DNMDNIY RITIKLLIET VVRGYELAEL KE-------- ---------- ----------
DAPHNIA_PULEX_GlobinX      CAIRPILGDD RT-LETEKAW TLLLDYMTAA MKESLVEARN ASAAESSKPL TLPPSSSSSS SAATDD----
DAPHNIA_CARINATA_LGB2      DFLAAKLGGE FT-PEARQAW TATMQGINTV VEASS----- ---------- ---------- ----------
HARPEGNATHOS_SALTATOR      DLIPSVFGKQ YT-QEVQEAW KKMFDYLF-- ---------- ---------- ---------- ----------
TUNICATA                   .......... .......... .......... .......... .......... .......... ..........
Ciona_intestinalis_Gb      ETIKVNLELP SD-SPTLSAW AKAYGVIKNG IKDAIIATYA EG-------- ---------- ----------
Ciona_savigny_Gb1 LGB      SVICDSQGLS KS-SDLYLCW TKTWEGIAKY VKIGLQQ--- ---------- ---------- ----------
MOLGULA_TECTIFORMIS_L      AVLTSVLEVD LC-HPTSLAW GKVYGILRDQ ---------- ---------- ---------- ----------
BOTRYLLUS_SCHLOSSERI_      KVICQALNLT ES-HPTVKAW IKAYGAISTG LKNGLTSELY LS-------- ---------- ----------
CEPHALOCHORDA              .......... .......... .......... .......... .......... .......... ..........
Branchiostoma florida      KSLRIVFGSK LT-PQLEEAW QTAYSHLKVT IKQGLEDAIQ KRDQADTSVV VTVE------ ----------
Branchiostoma florida      TAVSEVLGDR YT-KNMEEIY TITIKFILAT LSEGATMELT EDEQKNLGRL WRPPGRVHKF VRPEKVAAIV
Branchiostoma florida      TLLGELGCSS A-----AGAW ENLWARFMDV HRTCY----- ---------- ---------- ----------
Branchiostoma florida      YILETNLGDK LT-PNIRQGW VLVYGIVGAT MKKGMQQAMD HQNMAKTRP- ---------- ----------
HEMICHORDATA               .......... .......... .......... .......... .......... .......... ..........
Saccoglossus kowalevs      YTLQEQGGSA FT-PAAKAAW QKAWGVMKSV IVGALKG--- ---------- ---------- ----------
Saccoglossus kowalevs      SVWQEELGQR FT-TEVEESW KLLFFYIKDC LKEGYDIAMN EKTSGELNNS DFINQ----- ----------
Saccoglossus kowalevs      SAVKETLEDR YT-KNIEEIY KITFRFIVDA LIDGVVAGVN ERKAEAEAEA EAKAEAEDEQ DNEMMKKTMH
Saccoglossus kowalevs      ASIQPTLKDN WN-PAVGEAW EQLFKCVSSR MKDGFLQAQS SPSNTELLK- ---------- ----------
BALANOGLOSSUS_CLAVIGE      YFLRSLFLFA FI-YLAIPFP NFVVVLIYLF VRLSLLSVYL SV-------- ---------- ----------
ECHINODERMATA              .......... .......... .......... .......... .......... .......... ..........
STRONGYLOCENTROTUS_PU      FTIEEALGAG FT-PEVKAAW AAVYNIVSDT MSTKL----- ---------- ---------- ----------
KINORHYNCHA                .......... .......... .......... .......... .......... .......... ..........
ECHINODERES_HORNI_Glo      TAVKQTLDDK YA-AELGHIY RRPIRFILKL LYNGYMKKD- ---------- ---------- ----------
ENTOPROCTA                 .......... .......... .......... .......... .......... .......... ..........
PEDICELLINA_SP_LGB1_E      ELMQGKGGWG ---DSHNKAW AFVWSFICNS LTSGQAIYGG GSCPAKKA-- ---------- ----------
TARDIGRADA                 .......... .......... .......... .......... .......... .......... ..........
RICHTERSIUS_CORONIFER      ---LAFKLGH NFGHVHKTAW NKMLELVMTV IQQGMTAEAD DSKYKENQTA AVRQSEHEEL TLTNLTETT-


                           ....|....| ....|....| ....|....| ....|....| ..
                                  290        300        310        320
VERTEBRATE GlobinX         .......... .......... .......... .......... ..
ANOLIS CAROLINENSIS G      ---------- ---------- ---------- ---------- --
CALLORHINCHUS MILII G      ---------- ---------- ---------- ---------- --
DANIO RERIO GbX (from      ---------- ---------- ---------- ---------- --
ORYZIAS LATIPES GbX (      ---------- ---------- ---------- ---------- --
PETROMIZON MARINUS Gb      ---------- ---------- ---------- ---------- --
PYTHON MORULUS GbX (f      ---------- ---------- ---------- ---------- --
TETRAODON NIGROVIRIDI      ---------- ---------- ---------- ---------- --
XENOPUS TROPICALIS Gb      ---------- ---------- ---------- ---------- --
VERTEBRATE MYOGLOBINS      .......... .......... .......... .......... ..
HOMO_SAPIENS_MYOGLOBI      ---------- ---------- ---------- ---------- --
DANIO_RERIO_MYOGLOBIN      ---------- ---------- ---------- ---------- --
BOS_TAURUS_MYOGLOBIN_      ---------- ---------- ---------- ---------- --
GALLUS_GALLUS_MYOGLOB      ---------- ---------- ---------- ---------- --
VERTEBRATE CYTOGLOBIN      .......... .......... .......... .......... ..
DANIO_RERIO_CYTOGLOBI      ---------- ---------- ---------- ---------- --
OVIS_ARIES_CYTOGLOBIN      ---------- ---------- ---------- ---------- --
HOMO_SAPIENS_CYTOGLOB      ---------- ---------- ---------- ---------- --
VERTEBRATE HEMOGLOBIN      .......... .......... .......... .......... ..
EQUUS_CABALLUS_HB_ZET      ---------- ---------- ---------- ---------- --
HOMO_SAPIENS_HB_BETA_      ---------- ---------- ---------- ---------- --
TAENIOPYGIA_GUTTATA_H      ---------- ---------- ---------- ---------- --
ORYCTOLAGUS_CUNICULUS      ---------- ---------- ---------- ---------- --
VERTEBRATE NEUROGLOBI      .......... .......... .......... .......... ..
HOMO_SAPIENS_gi|10864      ---------- ---------- ---------- ---------- --
DANIO_RERIO_gi_188590      ---------- ---------- ---------- ---------- --
GALLUS_GALLUS_gi_1548      ---------- ---------- ---------- ---------- --
MONODELPHIS_DOMESTICA      ---------- ---------- ---------- ---------- --
CHIONODRACO_MYERSI_gi      ---------- ---------- ---------- ---------- --
BOVICHTUS_VARIEGATUS_      ---------- ---------- ---------- ---------- --
SIPUNCULA                  .......... .......... .......... .......... ..
THEMISTE_1_LGB2_gnl|t      ---------- ---------- ---------- ---------- --
THEMISTE_2_LGB2_gnl|t      ---------- ---------- ---------- ---------- --
THEMISTE_3_LGB2_gnl|t      ---------- ---------- ---------- ---------- --
BRACHIOPODA                .......... .......... .......... .......... ..
TEREBRETALIA_Ng2_gnl|      ---------- ---------- ---------- ---------- --
ANNELIDA                   .......... .......... .......... .......... ..
ALVINELLA_POMPEJANA_1      ---------- ---------- ---------- ---------- --
ALVINELLA_POMPEJANA_2      ---------- ---------- ---------- ---------- --
APHRODITA_ACULEATA_LG      ---------- ---------- ---------- ---------- --
CAPITELLA_SP_1_LGB2_F      .......... .......... .......... .......... ..
CAPITELLA_SP_2_LGB2_F      .......... .......... .......... .......... ..
```

```
CAPITELLA_SP_3_LGB3_g     .......... .......... .......... .......... ..
PLATHYNEREIS_dumerili     ---------- ---------- ---------- ---------- --
HELLOBDELLA_ROBUSTA1_     ---------- ---------- ---------- ---------- --
ECHIURA                   .......... .......... .......... .......... ..
URECHIS_CAUPO_Ng1_gnl     ---------- ---------- ---------- ---------- --
PLATYHELMINTHES           .......... .......... .......... .......... ..
MACROSTOMUM_LIGNANO_1     ---------- ---------- ---------- ---------- --
MACROSTOMUM_LIGNANO_2     ---------- ---------- ---------- ---------- --
CHOANOFLAGELLATE          .......... .......... .......... .......... ..
MONOSIGA_BREVICOLLIS_     GVQFSSDTER SRRSTAASAT ATPHCGLKSN RFSSNSLRKT IL
SALPINGOECA_SP_Ng2_gi     SLV------- ---------- ---------- ---------- --
PLACOZOA                  .......... .......... .......... .......... ..
TRICHOPLAX_ADHAERENS_     ---------- ---------- ---------- ---------- --
TRICHOPLAX_ADHAERENS_     ---------- ---------- ---------- ---------- --
TRICHOPLAX_ADHAERENS_     ---------- ---------- ---------- ---------- --
TRICHOPLAX_ADHAERENS_     ---------- ---------- ---------- ---------- --
PORIFERA                  .......... .......... .......... .......... ..
CARTERIOSPONGIA_FOLIA     ---------- ---------- ---------- ---------- --
AMPHIMEDON_QUEENSLAND     ---------- ---------- ---------- ---------- --
CNIDARIA                  .......... .......... .......... .......... ..
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
NEMATOSTELLA_VECTENSI     ---------- ---------- ---------- ---------- --
CLYTIA_HEMISPHAERICA_     ENSAPMPLAL INEDDSTSSV RTFSCR---- ---------- --
CLYTIA_HEMISPHAERICA_     ---------- ---------- ---------- ---------- --
CLAVA_1_LGB1_Locus_11     ---------- ---------- ---------- ---------- --
CLAVA_2_Ng_Locus_2853     ---------- ---------- ---------- ---------- --
HYDRA_MAGNIPAPILLATA_     ---------- ---------- ---------- ---------- --
MONTASTRAEA_FAVEOLATA     ---------- ---------- ---------- ---------- --
MONTASTRAEA_FAVEOLATA     ---------- ---------- ---------- ---------- --
ACOELA                    .......... .......... .......... .......... ..
SYMSAGITTIFERA_ROSCOF     ---------- ---------- ---------- ---------- --
SYMSAGITTIFERA_ROSCOF     ---------- ---------- ---------- ---------- --
NEMERTODERMATIDA          .......... .......... .......... .......... ..
MEARA_STICHOPI_Ng1_tr     ---------- ---------- ---------- ---------- --
NEMERTODERMA_WESTBLAD     TNYLLLFIKY LHHEPK---- ---------- ---------- --
NEMERTODERMA_WESTBLAD     HHIFFNNNCP YCIPQLT--- ---------- ---------- --
NEMERTODERMA_WESTBLAD     TVSDIITRCM LRVAII---- ---------- ---------- --
XENOTURBELLIDA            .......... .......... .......... .......... ..
XENOTURBELLA_HEMOGLOB     ---------- ---------- ---------- ---------- --
MOLLUSCA                  .......... .......... .......... .......... ..
SEPIA_OFFICINALIS_LGB     ---------- ---------- ---------- ---------- --
EUPRYMNA_SCOLOPES_1_L     ---------- ---------- ---------- ---------- --
DORYTEUTHIS_PEALEII_L     ---------- ---------- ---------- ---------- --
EUPRYMNA_SCOLOPES_2_L     ---------- ---------- ---------- ---------- --
MYTILUS_CALIFORNIANUS     ---------- ---------- ---------- ---------- --
MYTILUS_CALIFORNIANUS     ---------- ---------- ---------- ---------- --
SPISULA_SOLIDISSIMA_L     ---------- ---------- ---------- ---------- --
APLYSIA_1_LGB3_CNSN01     ---------- ---------- ---------- ---------- --
APLISIA_2_LGB2_CNSN01     ---------- ---------- ---------- ---------- --
LIOLOPHURA_JAPONICA_N     ---------- ---------- ---------- ---------- --
CRASSOSTREA_GIGAS_2_L     ---------- ---------- ---------- ---------- --
CRASSOSTREA_GIGAS_1_L     ---------- ---------- ---------- ---------- --
ARTHROPODA                .......... .......... .......... .......... ..
APIS_MELLIFERA_LGB2_g     ---------- ---------- ---------- ---------- --
CARCINUS_MAENAS_1_LGB     ---------- ---------- ---------- ---------- --
CARCINUS_MAENAS_2_LGB     ---------- ---------- ---------- ---------- --
DAPHNIA_PULEX_GlobinX     ---------- ---------- ---------- ---------- --
DAPHNIA_CARINATA_LGB2     ---------- ---------- ---------- ---------- --
HARPEGNATHOS_SALTATOR     ---------- ---------- ---------- ---------- --
TUNICATA                  .......... .......... .......... .......... ..
Ciona_intestinalis_Gb     ---------- ---------- ---------- ---------- --
Ciona_savigny_Gb1 LGB     ---------- ---------- ---------- ---------- --
MOLGULA_TECTIFORMIS_L     ---------- ---------- ---------- ---------- --
BOTRYLLUS_SCHLOSSERI_     ---------- ---------- ---------- ---------- --
CEPHALOCHORDA             .......... .......... .......... .......... ..
Branchiostoma florida     ---------- ---------- ---------- ---------- --
```

```
Branchiostoma florida   DAQSEENGVH ---------- ---------- ---------- --
Branchiostoma florida   ---------- ---------- ---------- ---------- --
Branchiostoma florida   ---------- ---------- ---------- ---------- --
HEMICHORDATA            .......... .......... .......... .......... ..
Saccoglossus kowalevs   ---------- ---------- ---------- ---------- --
Saccoglossus kowalevs   ---------- ---------- ---------- ---------- --
Saccoglossus kowalevs   NGDKEDNCIE SSKQKDITVI ASTKQR---- ---------- --
Saccoglossus kowalevs   ---------- ---------- ------.... .......... ..
BALANOGLOSSUS_CLAVIGE   ---------- ---------- ---------- ---------- --
ECHINODERMATA           .......... .......... .......... .......... ..
STRONGYLOCENTROTUS_PU   ---------- ---------- ---------- ---------- --
KINORHYNCHA             .......... .......... .......... .......... ..
ECHINODERES_HORNI_Glo   ---------- ---------- ---------- ---------- --
ENTOPROCTA              .......... .......... .......... .......... ..
PEDICELLINA_SP_LGB1_E   ---------- ---------- ---------- ---------- --
TARDIGRADA              .......... .......... .......... .......... ..
RICHTERSIUS_CORONIFER   ---------- ---------- ---------- ---------- --
```