

Globin and Globin Gene Structure of the Nerve Myoglobin of *Aphrodite aculeata**

(Received for publication, January 24, 1996, and in revised form, May 6, 1996)

Sylvia Dewilde‡, Mark Blaxter§, Marie-Louise Van Hauwaert‡, Jacques Vanfleteren¶, Eddy L. Esmans||, Mike Marden**, Nathalie Griffon**, and Luc Moens‡ ‡‡

From the ‡Department of Biochemistry, University of Antwerp (UIA), B-2610 Wilrijk, Belgium, the §Institute of Cell, Animal and Population Biology, Ashworth Laboratories, King's Buildings, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom, the ¶Department of Morphology, Systematics and Ecology, University of Ghent, B-9000 Ghent, Belgium, the ||Nucleoside Research and Mass Spectrometry Unit, University of Antwerp (RUCA-UIA), B-2020, Antwerp, Belgium, and **INSERM, Unité 299, Hôpital de Bicêtre, F 94275 Le Kremlin-Bicêtre, France

The globin of the nerve cord of the polychaete annelid *Aphrodite aculeata* was isolated and purified to homogeneity. The native molecule has a pI of 6.3 and acts as a dimer of two identical M_r 15,644.5 polypeptide chains as determined by electrospray mass spectrometry. It has an average affinity for oxygen ($P_{50} = 1.24$ torr) resulting from fast association ($k_{on} = 170 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$) and dissociation rates ($k_{off} = 360 \text{ s}^{-1}$). The partial primary structure of this nerve globin was determined at the protein level and completed and confirmed by translation of the cDNA sequence. The globin chain has 150 amino acid residues and a calculated M_r of 15,602.69 strongly suggesting that the amino terminus is acetylated. The absence of a leader sequence and the lack of Cys at the positions NA2 and H9 needed for the formation of the high M_r complexes found in extracellular annelid globins classify the *Aphrodite* globin with the cellular globin species. The *Aphrodite* nerve globin is unlikely to represent a separate globin family, as cDNA derived primers detect globin messenger RNA in muscle, gut, and pharynx tissue as well. The gene encoding this globin species is interrupted by a single intron, inserted at position G7.0. Comparison to other globin gene structures strongly suggest that introns can be lost independently, rather than simultaneously as a result of a single conversion event as suggested previously (Lewin, R. (1984) *Science* 226, 328).

Hemoglobin is widely distributed in ganglia and the nerve cord of invertebrates in many phyla. However, this "nerve globin" is by no means common and can be present or absent in closely related species (2–4). This erratic occurrence has led to the speculation that nerve globins might constitute a distinct globin subfamily, possibly with as yet unknown function. In molluscs and annelids, it mainly occurs intracellularly at millimolar concentration in glial cells surrounding the nerve cord (5–7). Kraus and Colacino (8) clearly demonstrated that neural

excitability in *Tellina alternata* (Bivalvia) is completely dependent on the oxygen stored on the globin in the nerves.

In molluscs and annelids, Mb¹-like tissue globins occur in addition to Hb-like molecules, respectively, intracellular in circulating "coelomocytes" and extracellular dissolved in the hemolymph. The molecular architecture of the latter forms is highly variable but all are built up from $M_r \sim 17,000$ globin chains (4, 9).

Previously, no nerve globins have been characterized in detail. Therefore it is unclear whether they are a unique, novel globin type selectively expressed in nerve tissue or a normal "myoglobin" type also occurring in other tissues and over-expressed in nerve tissue.

In the ventral nerve cord of the marine polychaete *Aphrodite aculeata*, a monomeric Hb of $M_r \sim 17,000$ with a hyperbolic oxygen dissociation curve ($P_{50} = 1.1$ mm Hg) is present (2, 3). These characteristics are similar with those of the nerve globin of the mollusc *Aplysia* and strongly resemble these of a myoglobin type molecule.

Vertebrate and plant globin genes contain, respectively, two introns (in the B- and G-helix) and three introns (in the B-, E-, and G-helix). The intron insertion position is conserved at B12.2, E15.0, and G7.0. The three intron/four exon pattern of plants is proposed to be ancestral and all other globin gene structures would be derived mainly by intron loss (1). Several invertebrate and protozoan globin genes have been characterized and it has become clear that the intron/exon pattern is less conserved than originally expected (10, 11). For example, the intron/exon pattern of the globin gene of the annelid, *Lumbricus terrestris*, is the same as in vertebrates whereas that of some nematode globin genes is plant-like. However, at least five different insertion positions for the central (E-helix) intron are documented in nonvertebrates. Therefore the evolution of the intron pattern in the globin gene family has become a subject of debate (10–15).

Here we describe the kinetics of ligand binding of purified *Aphrodite* nerve globin, and the primary structure of the protein as determined by protein and cDNA sequence analysis. The structure of the gene encoding the globin polypeptide was determined as well.

MATERIALS AND METHODS

Purification of *Aphrodite* Nerve Globin—Live specimens of *A. aculeata* were collected in the North Sea or obtained from the University Biological Supply Millport, Scotland. The brilliant red ventral nerve

* This work was supported by the Belgian National Science Foundation (NFWO) (to J. V. who is a Research Director and also acknowledged for Grants 20023.94 (to J. V. and L. M.) and G.2133.94 and 9.0008.93 (to E. L. E.)). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U46754.

‡‡ To whom correspondence should be addressed: Dept. of Biochemistry, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium. Tel.: 32-03-8202323; Fax: 32-03-8202248; E-mail: lmoens@uia.ua.ac.be.

This is an open access article under the CC BY license.

¹ The abbreviations used are: Mb, myoglobin; Hb, hemoglobin; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; RP-HPLC, reverse phase high performance liquid chromatography; ESMS, electrospray mass spectrometry.

cord was dissected and frozen at -80°C until use. After thawing the nerve globin was allowed to diffuse from the tissue into 50 mM Tris-HCl, pH 7.5, 2 mM phenylmethylsulfonyl fluoride and concentrated by ammonium sulfate precipitation (3). The material precipitating between 40 and 75% ammonium sulfate was freed of minor contaminants by gel permeation chromatography on a Zorbax GF450 HPLC column according to previous published methods (16). Final purity was checked by one- and two-dimensional electrophoresis (17, 18).

Protein Sequencing—Heme extraction was performed by acid acetone precipitation. Globin samples were carboxymethylated, maleylated, and cleaved separately with trypsin and endoproteinase Asp-N (19). The resulting peptide mixtures were separated by RP-HPLC using a Vydac C4 column developed with 0.1% trifluoroacetic acid/ CH_3CN . Peptides were sequenced in a ABI 471-B sequencer operated as recommended by the manufacturer. Globin sequence was partially reconstructed from relevant peptides using sperm whale myoglobin as a template (20, 21).

Mass Determination—Electrospray (ES) mass spectra were recorded on a VG Quattro II triple quadrupole mass spectrometer (VG Manchester, UK) equipped with a Kontron HPLC system (Kontron Instruments,

Milan Italy) consisting of a 325 pump, a HPLC autosampler 465, and a 332 HPLC detector. Tuning of the instrument was done by injecting a 20 pmol/ μl solution of myoglobin. The electrospray carrier solvent used was $\text{CH}_3\text{CN}/\text{H}_2\text{O}$ (50/50, v/v) containing 0.1% HCOOH . Samples were dissolved in the same solvent and injected via a Rheodyne loop injector of 100 μl . The flow rate of the carrier solvent was 40 μl . The capillary voltage was set at +3.81 KV. The source temperature was 80 $^{\circ}\text{C}$. The flow rate of the nebulizing gas and the drying gas were 2001/h and 3501/h, respectively. The cone voltage was 25 V. The mass spectrometer was scanned from mass to charge 500-1500. Spectra were recorded in the multichannel acquisition mode and by averaging 6 scans. The relative molecular mass (M_r) was calculated using Masslynx software.

Construction of Degenerate Oligonucleotide Primers—Two primers were designed based on the obtained protein sequence data. Primer APH1, CAYGGNGCNAARTTYATGGA a 20-mer with 128 redundancies, corresponding to the sense strand predicted by the peptide fragment HGAKFME. Primer APH2, GTRTTRTANACYTTNGTCCA, also a 20-mer with 128 redundancies, corresponding to the antisense strand predicted by the peptide fragment TNYVKTW (Fig. 4).

cDNA Sequencing—mRNA was isolated from different tissues with a Fast Track mRNA isolation Kit (Invitrogen). Reverse transcriptase-PCR was carried out with the Stratagene kit. First strand cDNA was synthesized using an oligo(dT) primer. A PCR reaction was then performed using the degenerate primers. The PCR was carried out for 30 cycles of 94 $^{\circ}\text{C}$ for 1 min, 50 $^{\circ}\text{C}$ for 1 min, and 72 $^{\circ}\text{C}$ for 1 min with *Taq* polymerase. The amplified fragment was blunt-end cloned into pBlue-script KS+, recombinants were confirmed by PCR, and the DNA sequence obtained from double stranded templates using T7 polymerase (Pharmacia Biotech Inc.) (22). Specific primers APH3 and APH4 were derived from the obtained DNA sequence (Fig. 5). APH3 and oligo(dT) primer were used in a PCR reaction to obtain the 3' end of the cDNA.

The 5' end was obtained with a 5' rapid amplification of cDNA ends experiment (RACE) (Life Technologies, Inc.) (23). First strand cDNA was synthesized with APH4 (specific antisense primer). A poly(C) tail was added to the 3' end of the cDNA with terminal deoxynucleotide

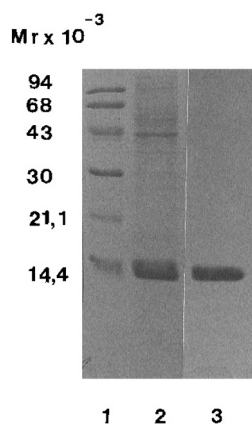


FIG. 1. Analysis of *Aphrodite* nerve myoglobin fractions by 15% SDS-PAGE. 1, markers; 2, 90% ammonium sulfate precipitated hemoglobin; 3, hemoglobin after gel permeation chromatography.

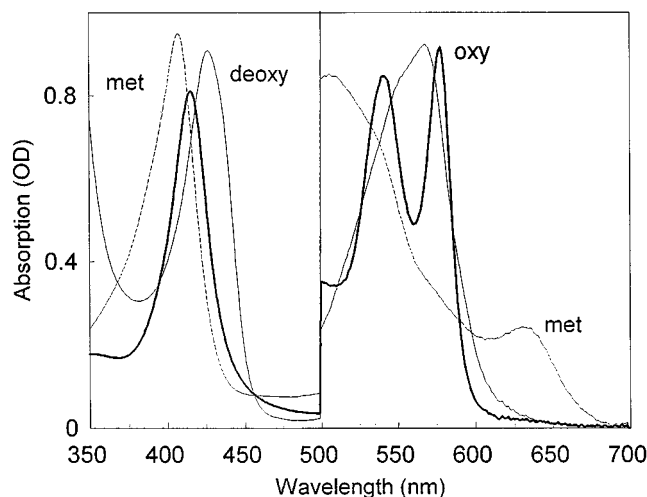


FIG. 2. Absorption spectra of the *Aphrodite* nerve myoglobin. For wavelengths above 500 nm, the absorption values are 10-fold enlarged.

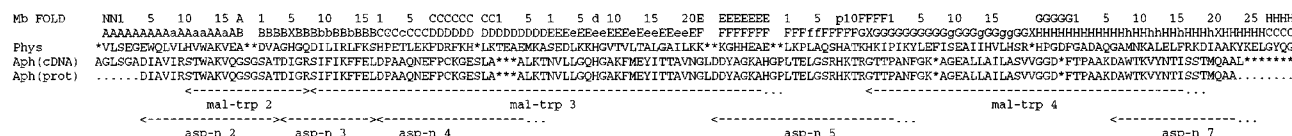


FIG. 4. Reconstruction of the *Aphrodite* nerve myoglobin and comparison with the cDNA derived sequence. Peptides were generated by protease Asp-N (*asp-n*) and tryptic digestion of the maleylated protein (*mal-trp*). *Aphrodite* nerve hemoglobin sequence (*Aph*) is aligned with sperm whale myoglobin (*Phys*).

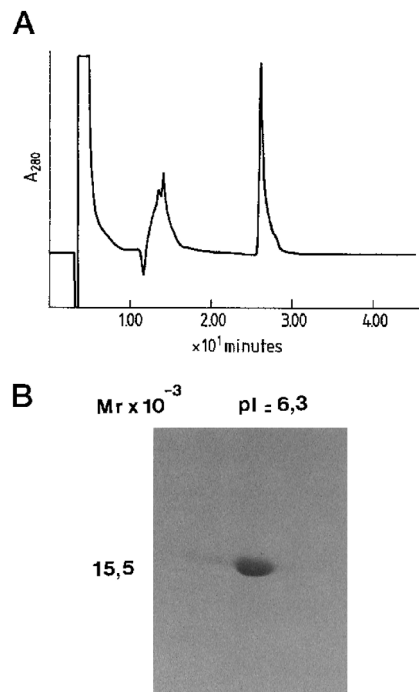


FIG. 3. Analysis of the dehemed and maleylated globin. A, by RP-HPLC; B, by two-dimensional electrophoresis.

transferase. The 5' end was then amplified using an oligo(dG) adaptor and the specific nested primer APH6 in a PCR of 30 cycles each consisting of 1 min at 94 °C, 1 min at 55 °C, and 2 min at 72 °C. The sequence was determined as described earlier.

Genomic DNA—Isolated gDNA was used as a template in a PCR with primers APH3 and APH8 and in a PCR with primers APH5 and APH10. Fragments were cloned and sequenced as described earlier.

Binding Kinetics—Ligand rebinding kinetics were measured after photolysis with 10-ns laser pulses at 532 nm (Quantel, France). Samples were 10 μ M in protein and equilibrated under 1 atm oxygen (1.2 mM at 20 °C) or 1 atm CO (1 mM at 20 °C). Detection of the kinetics was made at 436 nm, which is near the deoxy peak absorbance (Fig. 2).

For the ligand dissociation rates, the ligand replacement method was employed. For oxygen displacement by CO, samples were prepared under a mixed CO/oxygen atmosphere. For nearly equal partial pressures of the two gases, the CO form is dominant, due to the higher affinity. After photodissociation, a large fraction of exposed hemes will bind oxygen since it has the higher on-rate. A second phase on a much slower time scale is the replacement of oxygen by CO to return to the original stable CO form. With the previously determined on-rates for oxygen and CO, one can simulate the kinetics to determine the oxygen off-rate.

For the CO dissociation rate, an aliquot of the CO sample was injected into a cuvette containing buffer with 1–5 mM potassium ferri-cyanide. The kinetics of the CO to Met transition were followed by measuring the absorption spectra every 3 s (initially) with a HP8453 diode array spectrophotometer.

Structural and Evolutional Analysis of Globin Sequence—The primary structure of *Aphrodite* nerve globin was aligned with relevant vertebrate and 145 nonvertebrate globins by means of existing templates (20, 21, 24). Penalty scores were calculated manually.

Based on this alignment, phylogenetic trees were constructed using the TREECON software (25).

RESULTS AND DISCUSSION

Isolation and Characterization of Nerve Hemoglobin—The nerve globin of *Aphrodite* was purified by ammonium sulfate precipitation and gel permeation chromatography as described under "Materials and Methods." This results in a preparation essentially free of contaminants as judged by one-dimensional SDS-PAGE (Fig. 1).

The purified nerve globin shows absorption spectra in the oxy, deoxy, cyano-Met, and carbon monoxide forms essentially as reported previously (2). The spectrum of the deoxy derivative displays a shoulder at 550 nm, next to the maximum at 568 nm, confirming the earlier observations (2). Cytochrome b type spectral characteristics (maxima at 528 and 558 nm) as reported for the deoxy form of *Spisula solidissima* nerve globin (26) are absent (Fig. 2). The apparent M_r of the native protein as determined by HPLC permeation chromatography was 31,600 \pm 2,000. M_r estimation by SDS-PAGE yielded 15,500 \pm 400. After purification, extraction of the heme group, carboxymethylation, and maleylation, the nerve globin was subjected to RP-HPLC and a single globin chain was detected (Fig.

3A). Two-dimensional SDS-PAGE of this fraction revealed the presence of a single protein spot (Fig. 3B). Thus we conclude that the native protein is a homodimer. Similar conclusions were reached for the nerve globin of the bivalve species *T. alternata* and *S. solidissima* (3, 7). Interestingly Wittenberg *et al.* (2) calculated a M_r of 15,600 \pm 1,000 for native *Aphrodite* globin from $s_{20,w}$ = 1.7. Concentration-dependent association/dissociation might be the cause of this discrepancy.

Determination of the Primary Structure—Amino-terminal sequencing of approximately 1 nmol of globin yielded no phenylthiohydantoin-derivative signal, suggesting that the polypeptide was inaccessible for Edman degradation. Sequence analysis of overlapping internal fragments permitted the reconstruction of most of the polypeptide sequence as well as the construction of specific primers to amplify globin cDNA (Fig. 4). Full-length cDNA was isolated as described, and the sequence is presented in Fig. 5. It encompasses the entire coding region and confirms the amino acid sequence determined by Edman degradation. The initiation codon is preceded by at least 137 bases of untranslated sequence. No secretory leader sequence is present. This provides evidence that this globin is intracellular and a tissue or Mb type. The open reading frame extends for 150 codons and is followed by a 427-base long 3'-untranslated region ending with a polyadenylated tail. TATA box and polyadenylation signal are present.

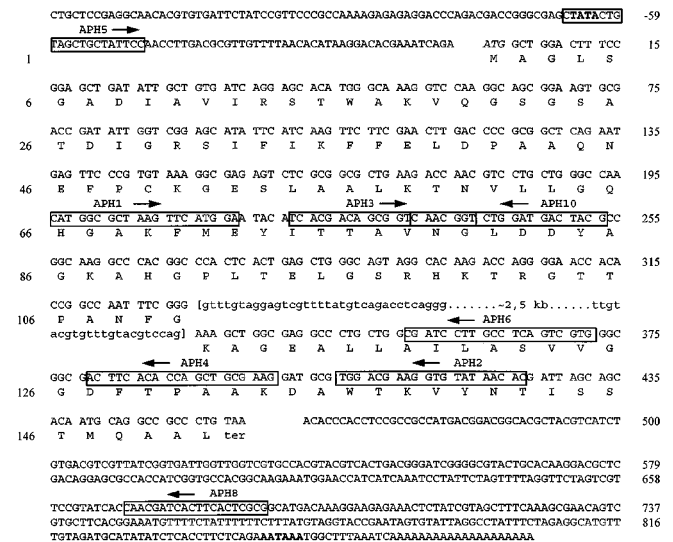


FIG. 5. Sequence of the *Aphrodite* myoglobin gene derived from cDNA and gDNA. Primers are boxed and named. Partial intron sequence is given in lower case letters. TATA box and polyadenylation sites are marked in bold.

TABLE I
Acetylated amino termini in nonvertebrate globins

| Species | Chain | Occurrence | Accession |
|------------------------------|-------|------------------------|--------------------|
| Echinodermata | | | |
| <i>Caudina arenicola</i> | c, d | Nucleated erythrocytes | GLBC-CAUAR, S15979 |
| <i>Paracaudina chilensis</i> | | Nucleated erythrocytes | S06134 |
| Mollusca | | | |
| <i>Anadara trapezia</i> | a, b | Nucleated erythrocytes | S06503, GGNKT |
| <i>Aplysia limacina</i> | | Radular muscle | GGGAA |
| <i>Aplysia kuroda</i> | | Radular muscle | GGGA2A |
| <i>Aplysia juliana</i> | | Radular muscle | GLB-APLJU |
| <i>Cerithidea izophorum</i> | | Radular muscle | GGGACR |
| <i>Dolabella auricularia</i> | | Triturative stomach | A25331 |
| <i>Bursatella leachii</i> | | Triturative stomach | A39023 |
| Unicellulars | | | |
| <i>Paramecium caudatum</i> | | | JQ1316 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------|------------------------|----------------------------|-------------------------------|-------------------------------|---------------------------|---------------------------|--------------------|----------------------------|------------------------------------|------------------------------|-------------------|-------------------------|-------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | NW1 | 5 | 10 | 15 | A | 1 | 5 | 10 | 15 | 1 | 5 | 10 | 15 | 20E | 1 | 5 | 10 | 15 | 20E | 1 | 5 | 10 | 15 | 20 | 25 | H |
| Mb fold | AAAAA | AAAAA | AAAAA | AAAAA | AAAAA | BBBBB | BBBBB | BBBBB | CCCCC | CCCCC | CCCCC | CCCCC | CCCCC | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD |
| Glyc fold | AAAAA | AAAAA | AAAAA | AAAAA | AAAAA | BBBBB | BBBBB | BBBBB | CCCCC | CCCCC | CCCCC | CCCCC | CCCCC | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD | DDDDD |
| Phys | *VISEGQVIVLHVAKVEA* | **** | DVAGHQDILIRLKFSPETLEKDFRKH* | LKTEASMKASEDLKKGVTVLALGAILKK* | KGRHAE* | **** | LKPLAQSHATKH* | IPIKYLEIFSEATIRVLSHR* | HPGDFGADAGQAMKALELFRKDIARKYKELGQYG | | | | | | | | | | | | | | | | | |
| Glyc MII | *GLSAAQRQVIAATWKTADAG* | **** | NGAGVGRDCLIKFLSAHPQMAAVFG* | TSG* | **** | ASDQVAAALGAKVLAQTGVAVSHL* | GD* | EGMVAQRAVGRVGRKGYGNKHKAQY* | FEFLGASLLSAMERHRTG* | RNNAAAKDMMAAAYADISGALISGLQS* | **** | | | | | | | | | | | | | | | |
| Glyc P1 | *HLTADQVAAAKWFEVSAGD* | G* | GAQLGLEMTKYFRENFMIFFG* | YSG* | RTEA* | LKHSKQLQHKGVIIDQTKAVAKH* | **** | DNARQMASTIALGVRRKGF* | DTRAEFFPALGMLDAMEKVF* | L* | **** | NRITLMAAYREISDADISGLQS* | **** | | | | | | | | | | | | | |
| Aph | AGLSGADIAVIRSTWAKVGS* | G* | SATDGRSIFIKFFELDPAAQNEFFCKGE* | SLAA* | LKTNVLLQGAKMEYITTAIVNGL* | DD* | AGKAHCPLELGSRIKTR* | **** | GTTANFGKAGALLAILASVVGSD* | FTPAKADMTKYNTISSTMQAL* | **** | | | | | | | | | | | | | | | |
| Tyl I | TDGILQRIKVKQWQVY* | SVGESRTDFADVNNFFRTNPD* | RSLEN* | **** | RVNG* | DNVISEPKARHVRVAGFDILISVL* | DK* | PVLQALAHYAAFKQFOT* | I* | **** | BKAFQDTHMPTQIARH* | **** | HGADIGWRACVASEQIVTIGTA* | **** | | | | | | | | | | | | |
| Lum I(d) | *ECVTEGLKVLQWASAFHIA* | HORVAFGLELWRGILREHPEIKAPF* | **** | RVRG* | DNVISEPKARHVRVAGFDILISVL* | DK* | PVLQALAHYAAFKQFOT* | I* | **** | BKAFQDTHMPTQIARH* | **** | HGADIGWRACVASEQIVTIGTA* | **** | | | | | | | | | | | | | |
| Tub I | *ECVTEGLKVLQWASAFHIA* | HORVAFGLELWRGILREHPEIKAPF* | **** | RVRG* | DNVISEPKARHVRVAGFDILISVL* | DK* | PVLQALAHYAAFKQFOT* | I* | **** | BKAFQDTHMPTQIARH* | **** | HGADIGWRACVASEQIVTIGTA* | **** | | | | | | | | | | | | | |

FIG. 6. Alignment of *Aphrodite* myoglobin with representative globin sequences. Mb, *Glyc fold*, myoglobin and *Glycera fold*, respectively; Phys, *Physeter catodon*; Glyc MII, *Glycera dibranchiata* globin MII, P1; Aph, *A. aculeata* nerve globin; Tyl I, *Tylorrhynchus heterochaetus* globin I; Lum I, *Lumbricus terrestris* globin I; Tub I, *Tubifex tubifex* globin I. In the heading of the alignment the major characteristics of the globin fold are summarized: helical notation (A to H), heme contacts (p = proximal; d = distal), internal (lower case) and surface (capital) positions.

TABLE II

Penalty scores of *Aphrodite* nerve myoglobin and some globin sequences against the vertebrate and nonvertebrate globin templates (20, 21)

The abbreviations used are: Phys; *P. catodon*; Glyc, *G. dibranchiata*; Tyl, *T. heterochaetus*; Lum, *L. terrestris*; Tub, *T. tubifex*.

| | A-motif | | BC-motif | | E-motif | | F-motif | | H-motif | | Total | |
|----------|---------|-----|----------|-----|---------|-----|---------|-----|---------|-----|-------|-----|
| | NV | V | NV | V | NV | V | NV | V | NV | V | NV | V |
| Phys | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.5 |
| Glyc MII | 0.0 | 0.0 | 0.0 | 0.7 | 0.7 | 2.2 | 0.0 | 1.2 | 0.0 | 0.0 | 0.7 | 4.1 |
| Glyc P1 | 0.0 | 0.0 | 1.0 | 2.4 | 0.5 | 0.5 | 1.0 | 1.5 | 0.7 | 0.0 | 3.2 | 4.4 |
| Tyl I | 1.0 | 2.0 | 1.0 | 1.0 | 0.9 | 2.2 | 0.7 | 1.0 | 1.5 | 2.7 | 5.1 | 8.9 |
| Lum I | 2.0 | 3.0 | 0.0 | 0.7 | 1.2 | 1.0 | 1.0 | 2.5 | 0.0 | 1.0 | 3.2 | 8.2 |
| Tub I | 1.0 | 2.0 | 0.0 | 0.7 | 0.2 | 1.0 | 0.0 | 2.2 | 0.0 | 2.0 | 1.2 | 7.9 |
| Aphr | 0.0 | 0.0 | 0.0 | 0.0 | 1.5 | 2.9 | 0.2 | 0.6 | 0.0 | 0.0 | 1.7 | 3.5 |

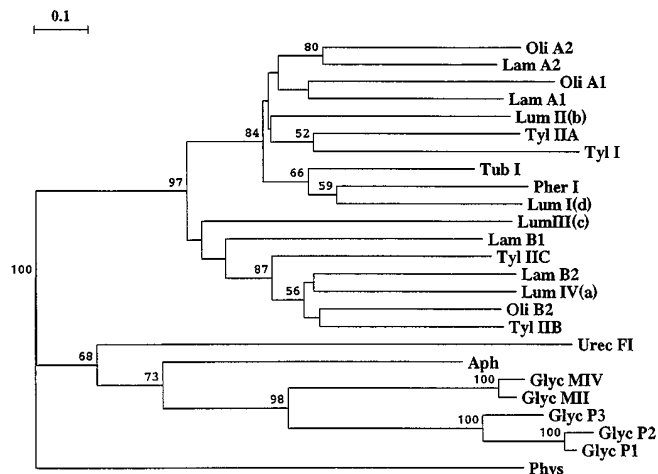


FIG. 7. Phylogenetic tree of annelid and annelid related globin sequences. *Aphrodite* myoglobin was aligned with 145 nonvertebrate globins (24) and an unrooted phylogenetic tree of the annelid and annelid-like globin sequences was constructed using the neighbor-joining method (TREECON, 25). The bootstrap values are given at each branch point. The sequences used are: *G. dibranchiata* (Glyc MII (GLB1), MIV (A33420), P1 (S13157), P2 (A36529), P3 (B36529)); *U. caupo* (Urec FI (A25537)); *T. heterochaetus* (Tyl I (GGWNS), IIA (A24811), IIB (GLB3), IIC (GGWN2C)); *L. terrestris* (Lum I (A29134), II (B28151), III (C28151), IV (A28151)), *Pheretima sieboldii* (Pher I (S06483)); *T. tubifex* (Tub I (S13836)); *Lameliobranchia* (Lam BI, BII, BIII (S08284), BIV)); *Oligobranchia* (Oli 1, 2, 3). *Physeter catodon* Mb is used as outgroup.

The M_r calculated for the deduced polypeptide is 15,602.69. Using ESMS, a M_r of $15,644.5 \pm 0.48$ was determined. Acetylation of the amino terminus would increase the mass by 43.03, which matches the difference of 41.81 ± 0.48 between measured and calculated M_r . Hence, we conclude that the amino terminus is likely blocked by an acetyl group. Modified amino termini are not uncommon in globins in general. *N*-Acetylated termini are found in nonvertebrate globins intracellularly in erythrocytes or muscles but never in extracellular globins (Table I). The meaning of this difference is unclear.

The protein and cDNA derived sequence of the *Aphrodite* nerve globin is confirmed by the primary structure derived from the globin gene sequence (Fig. 5).

m a b c d

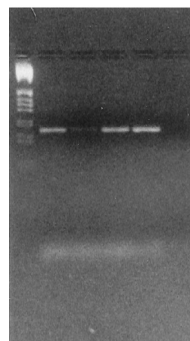


FIG. 8. Presence of globin mRNA in different tissues of *Aphrodite*. mRNA was prepared from gut (a), longitudinal muscle (b), nerve cord (c), and pharynx (d) tissue and used as template in a PCR reaction with APH3 and APH4 as primers. m, markers (BRL 1 Kb ladder).

Structural and Phylogenetic Aspects of Deduced Amino Acid Sequence—The alignment of the *Aphrodite* myoglobin sequence with the globin fold is unambiguous: (i) by the exclusion of polar residues from 33 out of the 33 invariant nonpolar sites (27), (ii) by the alignment of Pro-C2, which determines the folding of the BC corner, (iii) by the presence of the conserved heme-linked His(F8), and (iv) by the presence of the invariant Phe(CD1) (Fig. 6).

The *Aphrodite* nerve globin thus aligned matches both vertebrate and nonvertebrate globin templates quite well (20, 21). This is illustrated by the low penalty scores and proves that all major determinants of the globin fold are conserved (Table II).

Comparison with 145 nonvertebrate globin sequences reveals the highest similarity (31–32%) with the polymeric globins of the *Glycera* group (24, 28). All helical segments identified in sperm whale Mb and in the monomeric globin MII of *Glycera* (29) are also predicted to be present in the *Aphrodite* nerve globin. In addition, a 7-residue long α -helix, typical of β type globin chains, is present. The distal heme ligand is provided by His as in most globins, including the polymeric *Glycera* globins. The presence of Leu at that position in *Glycera* globin MII in MIV is exceptional (28). Most heme contacts are conserved, although some are unusual. Lys(CD4) and Arg(F7) replace the common residues Tyr or Phe, respectively. Similar

TABLE III

Kinetic and equilibrium constants for the reactions of ferrous *Aphrodite* myoglobin with oxygen and carbon monoxide compared to those of other globins

| Protein | k_{on} | k_{off} | K-assoc (k_{on}/k_{off}) | P50 (torr) 1.711 | K-diss (k_{off}/k_{on}) | I_{on} | I_{off} | L | Reference |
|---------------------|---------------|-----------|---------------------------------|------------------------|--------------------------------|---------------|-----------|------|------------|
| | $10^{-6}/M/s$ | s | μM | | μM | $10^{-6}/M/s$ | s | mM | |
| <i>Aphrodite</i> Mb | 170 | 360 | 0.47 | 1.24 | 2.1 | 21 | 0.1 | 21 | This paper |
| Whale Mb | 24 | 130 | 0.18 | 3.17 | 5.4 | 0.98 | 0.012 | 82 | 37 |
| Soybean Lb | 16 | 1.2 | 96.67 | 0.006 | 0.01 | 12.7 | 0.0078 | 1628 | 38 |
| <i>Glycera</i> HbA | 39 | 385 | 0.10 | 5.8 | 9.9 | 2.15 | 0.022 | 98 | 39 |
| <i>Glycera</i> HbB | 186 | 1800 | 0.10 | 5.7 | 9.7 | 26.8 | 0.042 | 638 | 39 |
| <i>Glycera</i> HbC | 190 | 2800 | 0.07 | 8.6 | 14.7 | 22.4 | 0.055 | 407 | 39 |

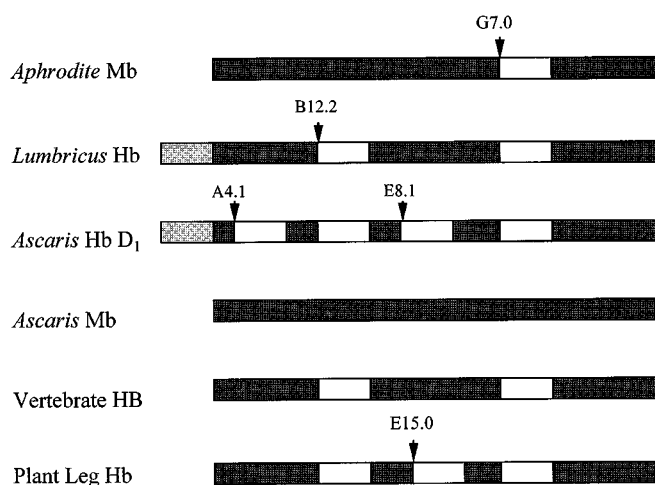


FIG. 9. Schematic representation of some representative globin genes. Open boxes are introns. Intron insertion position is given by the helix notification of the amino acid codon that is interrupted. B12.2 indicates that the intron is inserted in the codon of the amino acid at position B12 after the second base. E15.0 indicates interruption between codons E14 and E15.

changes were also seen in the globin species present in the perienteric fluid in *Ascaris*, and as these positions coincide with surface crevices, it was suggested that both basic amino acids could force their polar groups to reach into the solvent (30). Lys(CD4) and Lys(E10) might also form salt bridges with the heme propionates. The occupation of E11 by Phe is only shared in the *Calyptogenae soyocae* and *Lucina pectinaria* hemoglobins, two globins having a heme environment exceptionally rich in aromatic amino acids (31–33).

An evolutionary tree representing all sequenced annelid globin chains is depicted in Fig. 7. The *Urechis caupo* (Echiura), *Lameliibrachia* sp (Vestimentifera), and *Oligobrachia* sp (Pogonophora) were also included because they represent related phyla of coelomate worms. This tree clearly shows that there are two distinct clusters of globin chains. In fact, all sequences grouped with the upper cluster represent globin chains that form the giant polymeric globins ($M_r > 10^6$) of the extracellular fluid. They all have Cys occupying the positions NA2 and H9, these residues being required for aggregation of the composing chains into giant polymeric complexes (4, 21, 24). The lower cluster of globin chains is formed by intracellular globins: the *Aphrodite* nerve globin, and the globins from *Glycera* and *Urechis*, present in coelomocytes. Conceivably, a gene duplication of an ancestral globin gene initiated the separate evolution of the intra- and extracellular lines of globin evolution well before the divergence of the species mentioned.

Globin Expression in Different Tissues—Using nerve globin-specific primers the presence of nerve globin mRNA in different tissues was analyzed by PCR. Fig. 8 clearly illustrates that

globin mRNA can be detected in nerve, longitudinal muscle, gut, and pharynx tissue. The presence of the mRNA, however, does not guarantee the presence of the functional molecule. Next to the brilliant red nerve tissue, only the pharyngeal tissue is slightly pink and spectral analysis reveals the presence of an Hb-like molecule. Although we cannot entirely exclude the possibility that one or more isoforms might be detected by this experiment we feel it more likely that the same nerve globin species is expressed in other tissues as well. This proves that the nerve globin of *Aphrodite* must be considered as a normal Mb, with a novel expression site.

Functional Characteristics—The k_{on} and k_{off} values, respectively for O₂ and CO of the *Aphrodite* Mb were determined (Table III). The molecule shows rapid ligand rebinding. The oxygen and CO association rates are faster than those for whale or horse Mb, but are typical of legHb. The oxygen association rates are among the highest observed, approaching the diffusion limit. The dissociation rate for oxygen is also relatively high, resulting in an average oxygen affinity: less than legHb or *Glycera*, but more than whale Mb. The rate of autoxidation of 0.05/h at 37 °C is typical for Hb or Mb with this oxygen affinity, indicating no stability problems. Based on its intracellular location, its abundance, and its functional characteristics, the *Aphrodite* nerve globin most likely functions as an oxygen store. Similar functions are found in the gastrotrophic *Neodasy* (34).

Gene Structure—Using cDNA derived primers in PCR reactions, two overlapping, genomic DNA fragments were amplified covering the entire globin gene. They were cloned and partially sequenced.

Comparison of the globin gDNA with the cDNA reveals that the coding sequence is interrupted by a single 2.5-kilobase intron between bases 330 and 331 (Fig. 5). The standard splicing donor and acceptor sequences are present.

Based on the alignment of Fig. 6 it can be concluded that the intron insertion position is at the conserved position G7.0.

The structure of the *Aphrodite* globin gene is unusual. As mentioned before, globin genes mainly contain both the B and G helix introns together or neither of them. This intron configuration (B12.2 and G7.0) (Fig. 9) for example, occurs in the gene coding for the extracellular hemoglobin of the annelid, *Lumbricus* (12). The single G-helix intron in *Aphrodite* strongly suggests that the loss of individual introns can occur independently rather than simultaneously being caused by a single conversion event as suggested by Lewin (1). The presence of a single B-helix intron in the second domain of the *Pseudoterranova* globin gene confirms this view (35, 36).

Acknowledgments—S. D. and L. M. thank Dr. C. Van Broeckhoven, University of Antwerp, Neurogenetics Laboratory, for the generous use of molecular biology instrumentation. Dr. K. Cooreman is kindly acknowledged for the collection of live *Aphrodite* during campaigns of the Belgica in the North Sea. Dr. W. Van Dongen is acknowledged for excellent technical assistance in ESMS.

REFERENCES

1. Lewin, R. (1984) *Science* **226**, 328
2. Wittenberg, B. A., Briehl, R. W. & Wittenberg, J. B. (1965) *Biochem. J.* **96**, 363–371
3. Wittenberg, J. B. (1992) *Adv. Comp. Environ. Physiol.* **13**, 60–85
4. Vinogradov, S., Walz, D. A., Pohajdak, B., Moens, L., Kapp, O. H., Suzuki, T. & Trotman, C. (1993) *Comp. Biochem. Physiol.* **106**, 1–26
5. Kraus, D. W., Doeller, J. E. & Smith, P. R. (1988) *Biol. Bull. (Woods Hole)* **174**, 54–66
6. Kraus, D. W. & Doeller, J. E. (1988) *Biol. Bull. (Woods Hole)* **174**, 346–354
7. Doeller, J. E. & Kraus, D. W. (1988) *Biol. Bull. (Woods Hole)* **174**, 67–76
8. Kraus, D. W. & Colacino, J. M. (1986) *Science* **232**, 90–92
9. Vinogradov, S. N. (1985) *Comp. Biochem. Physiol.* **82**, 1–16
10. Moens, L., Vanfleteren, J., De Baere, I., Jellie, A. M., Tate, W. & Trotman, C. N. A. (1992) *FEBS Lett.* **312**, 105–109
11. Vinogradov, S. N., Walz, D. & Pohajdak, B. (1992) *Comp. Biochem. Physiol.* **103**, 759–773
12. Jhiang, S. M., Garey, J. R. & Riggs, A. F. (1988) *Science* **240**, 334–336
13. Dixon, B. & Pohajdak, B. (1992) *Trends Biochem. Sci.* **17**, 486–488
14. Stolz, A. & Doolittle, W. F. (1993) *Curr. Biol.* **3**, 215–217
15. Kao, W., Trewitt, P. & Bergtrom, G. (1994) *J. Mol. Evol.* **38**, 241–249
16. Blaxter, M. L., Vanfleteren, J., Xia, J. & Moens, L. (1994) *J. Biol. Chem.* **269**, 30181–30186
17. Laemmli, U. K. (1970) *Nature* **227**, 680–685
18. Hanash, S. M., Strahler, J. R., Neel, J. V., Hailat, N., Melhem, R., Keim, D., Zhu, X. X., Wagner, D., Gage, D. A. & Watson, J. T. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5709–5713
19. Allen, G. (1989) in *Laboratory Techniques in Biochemistry and Molecular Biology* (Burdon, R. H., and Van Knippenberg, P. H., eds) Vol. 9, Elsevier North Holland, Amsterdam
20. Bashford, D., Chothia, C. & Lesk, A. (1987) *J. Mol. Biol.* **196**, 199–216
21. Moens, L., Vanfleteren, J., Van de Peer, Y., Peeters, K., Kapp, O., Czelusniak, J., Goodman, M., Blaxter, M. & Vinogradov, S. (1996) *Mol. Biol. Evol.* **13**, 324–333
22. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8998–9002
23. Williams, S. A., Slatko, B. E., Moran, L. S. & DeSimone, S. M. (1986) *BioTechniques* **4**, 138–147
24. Kapp, O., Moens, L., Vanfleteren, J., Trotman, C., Suzuki, T. & Vinogradov, S. (1995) *Protein Sci.* **4**, 2179–2190
25. Van de Peer, Y. & De Wachter, R. (1994) *Comput. Appl. Biosci.* **10**, 569–570
26. Strittmatter, P. & Burch, H. B. (1963) *Biochim. Biophys. Acta* **78**, 562–563
27. Lesk, A. M. & Chothia, C. (1980) *J. Mol. Biol.* **136**, 225–270
28. Zafar, R. S., Chow, L., Stern, M. S., Scully, J. S., Sharma, P. R., Vinogradov, S. & Walz, D. (1990) *J. Biol. Chem.* **265**, 21843–21851
29. Arents, G. & Love, W. (1989) *J. Mol. Biol.* **210**, 149–161
30. De Baere, I., Liu, L., Moens, L., Van Beeumen, J., Gielens, C., Richelle, J., Trotman, C., Finch, J., Gerstein, M. & Perutz, M. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4638–4642
31. Hockenhull-Johnson, J. D., Stern, M. S., Martin, P., Dass, C., Desiderio, D. M., Wittenberg, J. B., Vinogradov, S. N. & Walz, D. A. (1991) *J. Protein Chem.* **10**, 609–622
32. Hockenhull-Johnson, J. D., Stern, M. S., Wittenberg, J. B., Vinogradov, S. N., Kapp, O. H. & Walz, D. A. (1993) *J. Protein Chem.* **12**, 261–277
33. Rizzi, M., Wittenberg, J. B., Coda, A., Fasano, M., Ascenzi, P. & Bolognesi, M. (1994) *J. Mol. Biol.* **244**, 86–99
34. Colacino, J. M. & Kraus, D. W. (1984) *Comp. Biochem. Physiol.* **79**, 363–369
35. Dixon, B., Walker, B., Kimmings, W. & Pohajdak, B. (1992) *J. Mol. Evol.* **35**, 131–136
36. Sherman, D. R., Kloek, A. P., Krishnan, B. R., Guinn, B. & Goldberg, D. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 11696–11700
37. Rohlfs, R. J., Mathews, A. J., Carver, T. E., Olson, J. S., Springer, B. A., Egeberg, K. D. & Sligar, S. G. (1990) *J. Biol. Chem.* **265**, 3168–3176
38. Appleby, C. A., Tjepkema, J. D. & Trinick, M. J. (1983) *Science* **220**, 951–953
39. Parkhurst, L. J., Sima, P. & Goss, D. J. (1980) *Biochemistry* **19**, 2688–2692