Hemoglobin (Hb) occurs in circulating red blood cells, neural tissue, and body wall muscle tissue of the nemertean worm, Cerebratulus lacteus. The neural and body wall tissue each express single major Hb components for which the amino acid sequences have been deduced from cDNA and genomic DNA. These 109-residue globins form the smallest stable Hbs known. The globin genes have three exons and two introns with splice sites in the highly conserved positions of most globin genes. Alignment of the sequences with those of other globins indicates that the A, B, and H helices are about one-half the typical length. Phylogenetic analysis indicates that shortening results in a small tendency of globins to group together regardless of their actual relationships. The neural and body wall Hbs in situ are half-saturated with O$_2$ at 2.9 and 4.1 torr, respectively. The Hill coefficient for the neural Hb in situ, --2.9, suggests that the neural Hb self-associates in the deoxy state at least to tetramers at the 2–3 mM (heme) concentration estimated in the cells. The Hb must dissociate upon oxygenation and dilution because the weight-average molecular mass of the Hb$_2$O$_2$ in vitro is only about 18 kDa at 2–3 μM heme concentration. Calculations suggest that the Hb can function as an O$_2$ store capable of extending neuronal activity in an anoxic environment for 5–30 min.

Hemoglobin was first observed in neural tissue of invertebrates by Lankester (3) who noted the brilliant crimson color of the ganglia of the polychaetous annelid Aphrodite aculeata. Neural Hbs have since been recorded in or associated with nervous tissue of other annelids (4–7), molluscs (5, 8–12), arthropods (13, 14), and nemerteans (15, 16). The functional significance of Hb$^+$ in neural tissue has been addressed in very few animals. Chalazonitis et al. (17) correlated the oxygenation state of neural Hb in Aplysia diplepis with the electrical activity of the neural ganglia and found that firing activity was proportional to the degree of oxygenation of the Hb. The neural tissue Hbs of the clams Tellina alternata and Spisula solidissima can extend the time of O$_2$ delivery to nerves during anoxic periods by as much as 30 min by acting as an O$_2$ store (12, 18).

Recently, De Wilde et al. (19) isolated and determined the amino acid sequence of the Hb from the neural tissue of A. aculeata. They found that the Hb was dimeric and had a relatively high O$_2$ affinity and that the sequence and gene structure clearly showed it to be a member of the globin family.

Many nemertean worms express intracellular Hbs in red blood cells, body wall muscle tissue, and neural tissue. We report here the amino acid sequences, gene structure, and oxygen equilibria in situ for both the body wall and the neural tissue Hbs of the marine nemertean, Cerebratulus lacteus. We also propose a role of the neural hemoglobin as an oxygen store. We have used the amino acid sequences of the globins in maximum parsimony analyses to address possible phylogenetic relationships. These Hbs are very small, yet stable, unlike the artificially truncated mini-Mbs (20–22). This finding should make the C. lacteus Hbs particularly valuable for studies of folding and stability.

**EXPERIMENTAL PROCEDURES**

**Animals and Tissue Preparation**

Eighty-four specimens of *C. lacteus* were purchased from the Department of Marine Resources, Marine Biological Laboratories, Woods Hole, MA, and maintained at 4 °C in the laboratory in 0.45 m of filtered seawater. Animals used for oxygen-binding experiments were kept a maximum of 2 weeks. Tissue samples needed for protein, RNA, and DNA experiments were dissected from animals within 72 h of delivery, immediately frozen in a dry ice/ethanol bath, and stored at −80 °C. Tissues were prepared for microscopy by relaxing the animals in MgCl$_2$ isotonic to seawater and fixing the head and trunk segments in 2.5% glutaraldehyde in 200 mM phosphate buffer, pH 7.4, made isotonic (0.58 m) by addition of NaCl. Segments were postfixed in 1% osmium tetroxide in the same buffer, dehydrated in an ethanol series, and embedded in Epon 812 resin. Ultrathin sections were stained with methylene blue and examined by light microscopy.

**Isolation and Purification of Hemoglobins**

Hb-containing tissues were ground in liquid N$_2$, then transferred to 50 mM Tris acetate, 10 mM EDTA, pH 7.5, at 0 °C, followed by centrifugation at 12,000 × g for 10 min at 4 °C. Brain and lateral nerve tissue extracts are referred to as neural Hb, and body wall muscle extracts are referred to as body wall Hb. Tissue extracts were stored at 0 °C if used immediately or refrozen in liquid N$_2$ and stored at −80 °C. Hbs were purified first by size-exclusion chromatography with G-75 SF Sephadex (Sigma) in a column (12 × 450 mm) equilibrated at 4 °C with 50 mM Tris acetate, 10 mM EDTA, pH 7.4 (measured at room temperature). Hb-containing fractions from this column were pooled, dialyzed against 10 M Tris acetate, pH 7.4, and dialyzed against the appropriate buffer used in high performance liquid chromatography; RP-HPLC, reverse phase HPLC; STPD, standard temperature and pressure, dry; PCR, polymerase chain reaction; bp, base pair; kb, kilobase pair.
The diagram is not drawn to scale.

globin sequence was determined from the pP8 and p4-7 PCR-generated constructs and from the p42 4.5-kb determined from the p1-3 and TA9 PCR-generated constructs and from genomic clone fragments equilibrated with 10 mM Tris acetate, pH 8.0, at room temperature. Hbs purified by RP-HPLC on a Synchropak RP300 C-18 column (4.6 mm) by HPLC anion-exchange chromatography on a Synchropak AX300 exclusion column, TSK-300 mm, Bio-Rad), on an LDC Milton Roy CCM automated HPLC system. Extracts and molecular size standards from 0 to 80% acetonitrile containing 0.1% trifluoroacetic acid over 70 min at a flow rate of 0.8 ml/min. Hb fractions, identified by absorbance at 415 nm, were pooled, dialyzed against 10 mM Tris, pH 8.0, and lyophilized. These Hb samples were resuspended in 0.1% trifluoroacetic acid in water and further purified by RP-HPLC on a Synchropak RP300 C-18 column (4.6 × 250 mm) equilibrated with 0.1% trifluoroacetic acid in water at room temperature. Globins were eluted by establishing a series of linear gradients from 0 to 80% acetonitrile containing 0.1% trifluoroacetic acid over 70 min at a flow rate of 0.8 ml/min.

Neural tissue extracts from four animals were analyzed on a size-exclusion column, TSK-30 (4.6 × 300 mm, Bio-Rad), on an LDC Milton Roy CCM automated HPLC system. Extracts and molecular size standards (Sigma, catalog number MW-GF-70) were run at room temperature on the column equilibrated with 50 mM Tris acetate, pH 7.5, over 15 min at a flow rate of 1.0 ml/min. The standards were bovine serum albumin, ovalbumin, β-lactoglobulin, horse heart myoglobin, and cytochrome c. Protein standards and Hb samples were monitored at 235 and 415 nm, respectively.

**Amino Acid Composition and Sequence Analyses**

Purified globins were hydrolyzed in 6 N HCl, sealed in vacuo in glass tubes, and hydrolyzed at 110°C for 24 or 72 h. Amino acid compositions were determined with a Beckman model 121MB analyzer. Amino-terminal sequences of the globins were determined with a model 477A protein sequencer (Applied Biosystems, Inc.). Both instruments are at the University of Texas Microanalysis Facility.

**Preparation and Amplification of Neural Globin mRNA**

Neural tissue was ground in liquid N2 and dissolved in a solution containing 3 N LiCl, 6 M urea, and 0.2% sodium deoxycholate at 0°C. Brain homogenate was extracted with phenol/chloroform/isoamyl alcohol (25:24:1 v/v/v), and the RNA was precipitated with ethanol and resuspended in diethylpyrocarbonate-treated water, and storage at −80°C. Poly(A)+ RNA was isolated from the total RNA by oligo(dT) affinity chromatography and reverse-transcribed using Moloney-murine leukemia virus reverse transcriptase (Promega). This was followed by extraction with phenol/chloroform/isoamyl alcohol, ethanol precipitation of the cDNA, resuspension in diethylpyrocarbonate-treated water, and storage at −80°C. The cDNA was used as a template to amplify globin cDNA in the polymerase chain reaction (PCR™, Hoffmann-La Roche). Primers were oligo(dT) and a redundant oligonucleotide based on the NH2-terminal residues of the globin (VNWAAV = 5′-GTGAAATGC/TGAGTCGCCT-3′). The reaction conditions were 100 mM Tris-HCl (pH 8.3 at 25°C), 50 mM KCl, 2 mM MgCl2, 1 μg/μl bovine serum albumin, 0.25 μM oligo(dT), 1 μM redundant primer, 200 μM dNTPs, and 2.5 units of AmpliTaq DNA polymerase (Perkin-Elmer) in a volume of 100 μl. Reactants were heated to 98°C for 2 min prior to the addition of enzyme, and amplification was performed for 30 cycles (94°C, 1.5 min; 45°C, 1 min; 72°C, 1.5 min (cycles 1–10), and then 2.5 min (cycles 11–30). PCR products were cloned into the pCR-1000 vector (TA Cloning system, Invitrogen, San Diego).

**Genomic DNA Isolation and Sequencing**

A genomic DNA library was constructed in the Lambda Dash II vector (Stratagene) from BamHI-restricted, size-selected genomic DNA that came from the sperm of a single animal. The library was screened with a 345-bp AccI fragment from the neural globin cDNA PCR product (see “Results”). A genomic clone containing a 25-kb insert that hybridized to the globin cDNA probe was isolated from the library, and several enzyme restriction fragments were subcloned into the Smal site of pBluescript (Stratagene) for sequencing. Fig. 1 shows the positions of the various clones used to determine the organization and sequence of the genes for neural and body wall globins.

**Body Wall Globin**—The 6-kb HpaI clone (HpaI-6) contained 5.1 kb of a body wall globin gene that was truncated in the first intron of the gene and lacked exon 1. Other clones (SpaI-7, 7 kb; ClaI-3, 3 kb; and ClaI-7, 7 kb) provided additional sequence information (see Fig. 1). The missing exon I region was isolated from genomic DNA by the use of two PCR amplification protocols. The first utilized a redundant primer within exon I and a non-redundant primer from within intron I of the body wall globin gene to amplify the intervening sequence from a template of Xhol-digested genomic DNA yielding fragment TA9 (see Fig. 1). The reaction mixture consisted of 20 mM Tris-HCl (pH 8.8 at 25°C), 10 mM KCl, 10 mM (NH4)2SO4, 2 mM MgSO4, 0.1% Triton X-100, 2.4 μM redundant primer, 0.2 μM non-redundant primer, 400 μM dNTPs, and 2.5 units of AmpliTaq DNA polymerase. The region from intron I to the region 5′ to exon I of the body wall globin gene was amplified in a second PCR reaction using a set of two nested primers from intron I and two primers from the PCR in vitro Single Site Amplification and Cloning System (catalog number TAK R015, Panvera Corp., Madison, WI) in a reaction mix given by the manufacturer. HindIII-digested genomic DNA was used as the template for the reaction which yielded fragment p41 (see Fig. 1). All PCR products were subcloned into pUC19 or pCR™II vector (TA cloning system, Invitrogen, San Diego, CA) for sequencing performed by the dyeoxy method using Sequenase 2.0 (U. S. Biochemical Corp.).

**Neural Globin**—A second screening of the genomic DNA library, using the 581-bp cDNA for neural globin, yielded a clone containing an HpaI digestion fragment of 4.5 kb that contained the complete sequence of the neural globin gene except for the first two codons of exon 1 and the adjoining 5′ upstream sequence (clone p42, see Fig. 1). The missing region was obtained with the PCR in vitro Single Site Amplification and Cloning System with a pair of non-redundant, nested primers located in the region of exon 2 and a sample of PstI-digested genomic DNA yield-
ing fragment pP8 (see Fig. 1). A second pair of non-redundant, nested primers from the exon 3 sequence was used to amplify exon 3 and intron 2 of the neuronal globin gene from a sample of XbaI-digested genomic DNA yielding fragment p4-7 (see Fig. 1). The PCR products and fragment p4-2 were cloned and sequenced as described above.

Phylogenetic Analysis of Globin Sequences

Alignment—Amino acid sequences of globins were obtained from the literature. Sequences of 22 globins from vertebrate and invertebrate animals (including Cerebratulus), bacteria, protists, algae, plants, and fungi were aligned by eye in a series of steps as follows. First, the 84 conserved residue positions identified by Kapp et al. (23) were shared by all members of the globin family were used for a core alignment that included the conserved CD1 Phe and F8 His. The template of Vinogradov et al. (24) was used to facilitate the identification of conserved hydrophilic or hydrophobic residues at external and internal locations, respectively. This template is a version of template I of Bashford et al. (25) extended to include invertebrates. This core comprises the shared parts of helices A, B, C, E, F, G, and H. The second step was to fill in the interhelical regions by using known X-ray structures of individual globins where possible (10 out of 22 globins). The helical regions, as identified in the X-ray analyses, were aligned with a minimum number of insertions or deletions in order to maintain the overall tertiary structure. Gaps, where residues were confined almost entirely to the loops between helices. Finally, the external/buried/internal classification for individual residues of Fermi and Perutz (26), originally developed for vertebrate myoglobins and hemoglobins, was taken as an additional template. This template includes both helical and interhelical regions, unlike those of Bashford and Vinogradov (24, 25), which carry only helical information. The Fermi template helps to identify and preserve the chemical patterns such as the arrangement of polar and non-polar amino acids in interhelical loops. All interhelical regions were expanded just enough to accommodate the longest sequence in these regions. The resulting alignment needed a total of 179 positions for the 22 globins with ~110 positions common to a majority of sequences.

Analysis—The aligned sequences were analyzed for relationships by employing the maximum parsimony method. All programs involved in this analysis are part of the PHYLIP Program Package, version 3.5, of Felsenstein (27). The original sequences were bootstrapped either 100 or 1,000 times using the program SEQBOOT (28). The replicate data sets so obtained were analyzed by the program PROTPARS which calculated a corresponding number of unrooted parsimony trees with the use of both the global rearrangement option and the option for a randomized input order of sequences. Finally, the CONSENSE program was used to generate a single majority-rule consensus tree from that population of 100 or 1,000 parsimony trees, with calculated bootstrap support values at each node. The human α-globin sequence was taken as outgroup for rooting the tree because the monophyly of vertebrate myoglobins and hemoglobins is established.

Oxygen Equilibria

The oxygen binding properties of the neural and body wall Hbs were determined in situ by thin layer microspectrophotometry at 15 °C (29, 30). Hb solutions for in vitro experiments were prepared by homogenizing tissues on ice in 50 mM Tris buffers at various pH values followed by centrifugation at 12,000 × g at 4 °C for 10 min to remove cellular debris. Tissue for in situ experiments was prepared by cutting a piece of the brain or body wall muscle and rinsing it in a solution of 0.45 mM filtered seawater at 0 °C buffered with 50 mM Tris at the appropriate pH and containing 20 mM KCN to inhibit cellular respiration. Thin sections were removed from the pieces with a razor blade and placed in a fresh buffer at 0 °C. All oxygen binding experiments were performed at 15 °C. All pH measurements were made at room temperature. Absorbance data from 12 wavelength (nm) pairs (410/420, 425, 430, 435, 440; 415/425, 430, 435, 440; 420/430, 435; 425/435) were used in a two-wavelength method for determining fractional saturation values. The tissue slice pathlength, measured optically, varied between 80 and 1100 µm. Mean absorbance changes were 0.050 and 0.183 at 415 and 430 nm, respectively. Between 5 and 7 experiments were performed under each set of conditions. Typically, 40–60 data points were averaged for each of 6 oxygen pressures in each experiment. The combined data for the experiments under each set of conditions were fitted by nonlinear least squares (31) directly to the Hill equation, y = KPⁿ/(1 + KPⁿ) to estimate n and P(O₂) = (nth root of K). Comparison of the absorption spectra before and after an experiment suggests that Met-Hb formation during an experiment is no more than about 5%.

RESULTS AND DISCUSSION

Hemoglobin Expression—Examination of C. lacteus neural ganglia and lateral nerve cords in fresh tissue shows that the cells surrounding axon bundles are uniformly bright red (HC in Fig. 2, A and B). Microspectrophotometry of this tissue gives the characteristic spectrum of HbO₂, which is not detectable within the axonal bundles (AX), although some Hb-containing cells appear to have processes that enter the axonal bundles.
Cells within the body wall also express Hb, but at differing concentrations, being most concentrated in the trunk region surrounding the gonadal sacs where it may aid in the delivery of oxygen to the developing gametes.

The concentration of Hb in the neural ganglia of the brain was estimated by dissecting out the brains from two animals and analyzing the entire extracts by size-exclusion HPLC. Integration of the absorption peaks at 415 nm yielded 1.1 and 1.7 nmol of Hb for the two extracts. We assumed a millimolar extinction coefficient of 131 at 415 nm, the value for human HbO2 (32). The volume of the brain shown in Fig. 2A would be completely contained within a sphere of radius 600 μm and volume 0.9 ml. This, then, provides an upper limit to the brain volume and provides a minimal concentration of 1.2–1.9 mM. The actual concentration must be substantially higher because the volume of 0.9 ml includes parts of the foregut, rhynchocoel, and the axonal tissue itself, none of which have Hb. We conclude that the concentration is at least 2–3 mM and could be significantly higher. This estimate is comparable to the 2.1–2.8 mM values for myoglobin found in the muscles of seals, porpoises, and tuna (33–35).

Structure of Globins and Globin Genes—A single major globin protein fraction was isolated from neural tissue extracts in a three-step chromatographic procedure (Fig. 3, A–C). The body wall Hb also gave a single major peak in the first two chromatographic steps, but RP-HPLC of the Hb fraction showed several minor peaks (Fig. 3D), raising the possibility of additional Hbs in this tissue. NH2-terminal sequences of neural and body wall globins were determined (43 and 45 residues, respectively) and found to share 79% amino acid sequence identity (Fig. 4). The first eight residues of both globins are identical (VNWAAVVD), allowing us to construct a redundant oligonucleotide primer directed at both globin genes (see “Experimental Procedures”). Use of this primer and oligo(dT) permitted the amplification and cloning of a 581-bp cDNA fragment that encodes the complete sequence of a globin from the neural tissue (Fig. 5). Five clones were completely sequenced and found to be identical in the coding region, but they possessed...
some differences in the 3' non-coding region. The amino acid sequence determined by direct sequencing was DFY at residues 9–11 but was EL-STOP in the PCR-generated cDNA clones (Figs. 4 and 5). We conclude that the EL-STOP sequence is the result of errors in amplification and/or cloning because the neural globin genomic DNA sequence (Fig. 6) matches the direct determination of residues 1–42. An additional difference is that the cDNA sequence has Leu at position 56, whereas the genomic DNA gives Ile for this position.

The amino acid sequence of the body wall globin was deduced from the genomic DNA (Fig. 7). This sequence matches the first 45 NH2-terminal residues determined directly (Fig. 4). The gene contains three exons separated by two introns. Intron 1 separates exons 1 and 2, splitting a Glu codon (residue 13) corresponding to position B12 in both neural and body wall globins. Intron 2 separates exons 3 and 2 between codons for His-80 (G6) and Asn-81 (G7) in neural globin and Thr-80 (G6) and His-81 (G7) in body wall globin. The positions of the 5' and 3' splice junctions are identical in both genes (Figs. 6 and 7) and match those found for typical globins (36). The two globin sequences are 78% identical and are clearly derived by duplication of an ancestral gene.

The amino acid compositions of neural and body wall globins are compared in Table I with those deduced from DNA. The agreement is within 5–7% for the neural globin except for Cys and Trp, which are partially destroyed by acid hydrolysis, and Val. The low Val value presumably results from difficulty in cleaving the Val-Val bond between residues 6 and 7. The analysis of body wall globin shows greater deviations with unexplained high Ser and low Ala values.

Protein Structure—The chains of neural and body wall Hbs have only 109 residues which makes them the smallest naturally occurring Hbs known. Unlike the proteolytically generated mini-Mb of 108 residues (20–22), the mini-Hbs of C. lacteus are stable. Paramecium Hb (37) is slightly larger (116 residues).

![Fig. 5. Sequence of the 581-bp cDNA encoding the C. lacteus neural globin. The first 17 nucleotides (underlined) correspond to the redundant primer (see “Experimental Procedures”). Splice junction positions predicted from sequence alignment and subsequently verified from genomic clones are marked with arrows. The single-letter amino acid designations for residues of the globin chain are listed below their corresponding codons. A putative polyadenylation signal in the 3' non-coding region is boxed. Asterisks mark nucleotides that are C or T in genomic DNA.](http://www.jbc.org/)

![Fig. 6. Genomic DNA sequence encoding the three exons of C. lacteus neural globin. Splice junction sequences are underlined. Numbers to the right are the corresponding nucleotide numbers in the GenBank file. Coding regions are in capital letters, and single-letter amino acid designations are used for the globin sequence.](http://www.jbc.org/)
residues) than the nemertean Hbs and autoxidizes (38) 10-fold faster at pH 8 than sperm whale Mb (153 residues). The

C. lacteus Hbs are sufficiently stable to permit the HbO₂ to survive the lengthy purification procedures and to make possible

the measurement of O₂ equilibria over a period of several hours. In contrast, the mini-myoglobins (20–22) are rapidly

oxidized in air, and O₂ equilibrium measurements are impossible. We examine here how the C. lacteus chains differ from

those of other Hbs and how these chains may achieve stability despite their short length.

The Alignment Reveals Amino- and Carboxyl-terminal Deletions—

The alignment of 22 globins (Fig. 8) shows an impressive conservation of hydrophobic residues at 37 internal positions throughout the globin family (i.e. the 34 columns of residues with blue backgrounds plus the CD1, E7, and F8 columns in red in Fig. 8) This pattern of 37 non-polar amino acids in globins at solvent-inaccessible locations, identified by Gerstein et al. (68), was applied to a sequence alignment by Kapp et al. (23). It allows a much improved refinement of the alignment of all globins. These residues at solvent-inaccessible positions are a major factor for maintaining the universal globin fold. The C. lacteus sequences have 34 of these 37 positions retained and lack only A8, B6, and H19 because of deletions (see Fig. 8). Kapp et al. (23) have also identified 84 positions conserved in all 700 globin sequences examined (boxed regions in Fig. 8). According to our alignment, the C. lacteus globins have all of these conserved positions except for 7 residues deleted in regions B5–B7 and H16–H19. Despite their shortness, the globins of this nemertean have clearly retained most of the 37 solvent-inaccessible and the 84 core residue positions. However, large deletions appear at both ends of C. lacteus globins. The alignment shows that greater parts of the A and B helix, as well as from the H helix, are missing. Still, all six

residues) than the nemertean Hbs and autoxidizes (38) 10-fold faster at pH 8 than sperm whale Mb (153 residues). The C. lacteus Hbs are sufficiently stable to permit the HbO₂ to survive the lengthy purification procedures and to make possible

the measurement of O₂ equilibria over a period of several hours. In contrast, the mini-myoglobins (20–22) are rapidly

oxidized in air, and O₂ equilibrium measurements are impossible. We examine here how the C. lacteus chains differ from

those of other Hbs and how these chains may achieve stability despite their short length.

The Alignment Reveals Amino- and Carboxyl-terminal Deletions—

The alignment of 22 globins (Fig. 8) shows an impressive conservation of hydrophobic residues at 37 internal positions throughout the globin family (i.e. the 34 columns of residues with blue backgrounds plus the CD1, E7, and F8 columns in red in Fig. 8) This pattern of 37 non-polar amino acids in globins at solvent-inaccessible locations, identified by Gerstein et al. (68), was applied to a sequence alignment by Kapp et al. (23). It allows a much improved refinement of the alignment of all globins. These residues at solvent-inaccessible positions are a major factor for maintaining the universal globin fold. The C. lacteus sequences have 34 of these 37 positions retained and lack only A8, B6, and H19 because of deletions (see Fig. 8). Kapp et al. (23) have also identified 84 positions conserved in all 700 globin sequences examined (boxed regions in Fig. 8). According to our alignment, the C. lacteus globins have all of these conserved positions except for 7 residues deleted in regions B5–B7 and H16–H19. Despite their shortness, the globins of this nemertean have clearly retained most of the 37 solvent-inaccessible and the 84 core residue positions. However, large deletions appear at both ends of C. lacteus globins. The alignment shows that greater parts of the A and B helix, as well as from the H helix, are missing. Still, all six

residues) than the nemertean Hbs and autoxidizes (38) 10-fold faster at pH 8 than sperm whale Mb (153 residues). The C. lacteus Hbs are sufficiently stable to permit the HbO₂ to survive the lengthy purification procedures and to make possible

the measurement of O₂ equilibria over a period of several hours. In contrast, the mini-myoglobins (20–22) are rapidly

oxidized in air, and O₂ equilibrium measurements are impossible. We examine here how the C. lacteus chains differ from

those of other Hbs and how these chains may achieve stability despite their short length.

The Alignment Reveals Amino- and Carboxyl-terminal Deletions—

The alignment of 22 globins (Fig. 8) shows an impressive conservation of hydrophobic residues at 37 internal positions throughout the globin family (i.e. the 34 columns of residues with blue backgrounds plus the CD1, E7, and F8 columns in red in Fig. 8) This pattern of 37 non-polar amino acids in globins at solvent-inaccessible locations, identified by Gerstein et al. (68), was applied to a sequence alignment by Kapp et al. (23). It allows a much improved refinement of the alignment of all globins. These residues at solvent-inaccessible positions are a major factor for maintaining the universal globin fold. The C. lacteus sequences have 34 of these 37 positions retained and lack only A8, B6, and H19 because of deletions (see Fig. 8). Kapp et al. (23) have also identified 84 positions conserved in all 700 globin sequences examined (boxed regions in Fig. 8). According to our alignment, the C. lacteus globins have all of these conserved positions except for 7 residues deleted in regions B5–B7 and H16–H19. Despite their shortness, the globins of this nemertean have clearly retained most of the 37 solvent-inaccessible and the 84 core residue positions. However, large deletions appear at both ends of C. lacteus globins. The alignment shows that greater parts of the A and B helix, as well as from the H helix, are missing. Still, all six

residues) than the nemertean Hbs and autoxidizes (38) 10-fold faster at pH 8 than sperm whale Mb (153 residues). The C. lacteus Hbs are sufficiently stable to permit the HbO₂ to survive the lengthy purification procedures and to make possible

the measurement of O₂ equilibria over a period of several hours. In contrast, the mini-myoglobins (20–22) are rapidly

oxidized in air, and O₂ equilibrium measurements are impossible. We examine here how the C. lacteus chains differ from

those of other Hbs and how these chains may achieve stability despite their short length.

The Alignment Reveals Amino- and Carboxyl-terminal Deletions—

The alignment of 22 globins (Fig. 8) shows an impressive conservation of hydrophobic residues at 37 internal positions throughout the globin family (i.e. the 34 columns of residues with blue backgrounds plus the CD1, E7, and F8 columns in red in Fig. 8) This pattern of 37 non-polar amino acids in globins at solvent-inaccessible locations, identified by Gerstein et al. (68), was applied to a sequence alignment by Kapp et al. (23). It allows a much improved refinement of the alignment of all globins. These residues at solvent-inaccessible positions are a major factor for maintaining the universal globin fold. The C. lacteus sequences have 34 of these 37 positions retained and lack only A8, B6, and H19 because of deletions (see Fig. 8). Kapp et al. (23) have also identified 84 positions conserved in all 700 globin sequences examined (boxed regions in Fig. 8). According to our alignment, the C. lacteus globins have all of these conserved positions except for 7 residues deleted in regions B5–B7 and H16–H19. Despite their shortness, the globins of this nemertean have clearly retained most of the 37 solvent-inaccessible and the 84 core residue positions. However, large deletions appear at both ends of C. lacteus globins. The alignment shows that greater parts of the A and B helix, as well as from the H helix, are missing. Still, all six

residues) than the nemertean Hbs and autoxidizes (38) 10-fold faster at pH 8 than sperm whale Mb (153 residues). The C. lacteus Hbs are sufficiently stable to permit the HbO₂ to survive the lengthy purification procedures and to make possible
A-helical core residues are maintained in the nemertine globins. A tryptophan occurring position 3 of this conserved core forms a hydrophobic anchor that makes contacts with the E, G, and H helices. This position, Trp-3 in the A-helical core residues are maintained in the nemertine globins, correspond to human α chain residue Trp-14 (A12) which forms contacts with Ala-63 (E12), Leu-66 (E15), Thr-67 (E16), Val-70 (E19), Leu-105 (G12), and Leu-109 (G16) (see Ref. 69). The alignment suggests that very similar contacts are likely in C. lacteus Hb: Val-49 (E12), Ile-52 (E15), Asn-53 (E16), Cys-90 (G16). The alignment preserves the propensity for a hydrogen bond-forming residue at position E16 in 15/22 of the chains shown in Fig. 8. The six-residue core of the A helix identified by Kapp et al. (23) is part of the 13-residue core found in six high resolution x-ray studies (70). The six-residue core is conserved in all globins including that of C. lacteus. We conclude that C. lacteus globins have retained the minimal core of residues in helix A required for the Mb fold.

The truncated A helix (residues 1–7) may be particularly important for the stability of C. lacteus Hbs. The artificial mini-myoglobin (20–22) is unstable and rapidly autoxidizes. It comprises residues 32–139 of sperm whale Mb and so lacks the entire A helix and has a shortened COOH-terminal H helix. The x-ray structure of the monomeric mini-myoglobin (20–22) is unstable and rapidly autoxidizes and lacks the entire A helix and has a shortened COOH-terminal H helix. The COOH-terminal residue of C. lacteus corresponds to residue 138 of sperm whale Mb; the mini-Mb extends only one residue beyond this position (E16). The artificial mini-myoglobin (20–22) is unstable and rapidly autoxidizes. It comprises residues 32–139 of sperm whale Mb and so lacks the entire A helix and has a shortened COOH-terminal H helix. The x-ray structure of the monomeric mini-myoglobin (20–22) is unstable and rapidly autoxidizes. It comprises residues 32–139 of sperm whale Mb and so lacks the entire A helix and has a shortened COOH-terminal H helix. The COOH-terminal residue of C. lacteus corresponds to residue 138 of sperm whale Mb; the mini-Mb extends only one residue beyond this position (E16).
residue further. It appears, therefore, that the presence of an A helix, albeit shortened, in *C. lacteus* Hb and its absence in mini-myoglobin may be correlated with the difference in stability.

Most globins have a B helix that is almost twice as long as that in *C. lacteus*. How might the apparent gap between helices A and B in *C. lacteus* globins be accommodated? The gap between residues 8 and 9 of the *C. lacteus* globins corresponds to residues B1 to B7 in the human α chain. *Chironomus* III globin has a gap that corresponds to human α chain helical residues B1 and B4. The high resolution x-ray structure of the very similar *Chironomus* III globin (61) shows one way in which such apparent gaps can be readily accommodated. Residues 12–19 in *Chironomus* III Hb form a 3_10 helix. Eight residues of a 3_10 helix would be about 4 Å longer than an eight-residue α helix. Thus, helical segments could “stretch” to maintain essential contacts. There is no necessary correlation between globin chain length and stability, provided the essential features of the myoglobin fold are conserved.

The alignment of the *C. lacteus* sequence (Fig. 8) places some limitations on the minimal requirements for folding. Jennings and Wright (71) have identified 20 residues in helices A, B, G, and H in sperm whale apoMb for which the amide protons become protected within 5 ms during the initial folding and helix formation. *C. lacteus* globin lacks residues in five of these helical positions: A7–A9 and H19–H20. Therefore, residues at these positions do not appear to be required for folding.

**Heme Pocket**—The structural stability of Hbs depends strongly on the affinity for heme. Ferrous Mbs and Hbs show little tendency to lose heme because of the covalent bond between the proximal His and the Fe^{2+} atom. This bond is much weaker in ferric Hb (Met-Hb). The lower retention of the hemin in Met-Mb and Met-Hb then depends more on the hydrophobicity of the heme pocket and, in aquo-Met-Hb, on the water molecule that is anchored to the iron and to distal residues (72). Additional stabilization is contributed by electrostatic contacts with the propionic acid groups on the heme. Table II shows the residues of the heme pocket in sperm whale Mb and in the β chain of human Hb. Without exception, all residues in the positions that are hydrophobic in both these chains are also hydrophobic in *C. lacteus* globins.

The presence of Thr at E11 in *C. lacteus* Hbs is unusual and may contribute substantially to the stability of the Hbs. The mutation Val-68 → Thr (E11) in sperm whale Mb (72) results in a very large increase in stability by slowing the rate of hemin loss from the ferric Mb because the water molecule bound to the Fe^{3+} of the hemin is stabilized by hydrogen bonds to both the distal His-64 (E7) and Thr-68 (E11). The distal His at E7 of most vertebrate Hbs is replaced by Gln in the Hbs of *C. lacteus*. This substitution is not, however, associated with any significant increase in the rate of hemin loss in sperm whale Mb (72). So, the corresponding substitution in *C. lacteus* globins should not be expected to alter their rate of hemin loss significantly. Glutamine is found in the distal E7 position in virtually all of the prokaryotic and protist globins as well as many invertebrate globins. This suggests that Gln might be the ancestral residue at this position.

Additional stabilization of the hemin in *C. lacteus* globins is
provided by the interaction of residue Lys-47 (E10) corresponding to human β Lys-66 (E10) which is known to form an electrostatic link with the propionic acid group of pyrrole IV of the heme (26). The Arg at F7 and Lys or Arg at C7 also should provide stabilization in the globins by propionic acid group interaction.

Hydrophilicity—The hydrophilicity patterns of Fig. 9 show the close similarity of the body wall and neural Hbs of C. lacteus. The neural Hb has, however, substantially more hydrophilicity than the body wall Hb. This difference can be quantified by examining the 24 positions where the amino acid residues differ (Fig. 4). The sums of the hydrophilicity values at these positions for body wall and neural Hbs are 7.9 and 39.4, respectively. Why the Hbs should differ in this way is puzzling. Perhaps it reflects differences in the nature of the subunit interfaces involved in self-association.

Fig. 10 shows a residue-by-residue comparison of the hydrophobicity of body wall Hb and sperm whale Mb. The alignment is identical to that in Fig. 8. The high degree of correspondence evident between the two Hbs provides substantial support for the overall alignment. The least satisfactory correspondence is in helix A where Trp of position 14 (sperm whale Mb) is used as an anchor (see arrow). Comparison of the two patterns in the EF region suggests an alternative to that in Fig. 8: the Ser-Ala-Asp- sequence (residues 59–61 of C. lacteus neural or body wall globin) corresponds better in hydrophilicity to sperm whale Mb if it is moved two residues to the right (see Fig. 8).

Phylogenetic Analysis of Globin Sequences—The 22 aligned globin sequences (Fig. 8) were analyzed as described under “Experimental Procedures.” Fig. 11 summarizes the results. Consensus tree A is based on 1,000 bootstrap replicates, and the consensus trees B–D each are based on 100 replicates of the original data set of the 22 sequences.

Tree A, rooted with human α-globin, shows the consensus topology for all sequences (179 positions, as in Fig. 8). This tree has several important features as follows. 1) Three major clades of globins are present (besides vertebrate globins), and their monophyletic origins are well supported (nodal bootstrap confidence is shown in bold, 95%, yeasts and bacteria (99% confidence), plants (99%), and the truncated globins of protozoans, the green alga Chlamydomonas and the cyanobacterium Nostoc (100%). This finding of three clades is in excellent agreement with the results of a phylogenetic analysis of Moens et al. (74). The surprising similarities between yeast and bacterial globins and between all four truncated sequences were explained by these authors with two horizontal gene transfers as follows: one from a common ancestor of both yeasts to the prokaryotic ancestor of Escherichia, Vitreocella, and Alcaligenes; and the other from the common ancestor of ciliate protozoans and the green algae to the cyanobacterial lineage that evolved into Nostoc. The latter hypothesis is consistent with the extensive intracellular symbiosis between ciliates and algae that occur. 2) A gradient of decreasing vertebrate similarities from human α chain to Petromyzon, the sperm whale Mb, and Caudina globins exists (see also trees B–D) that leads into...
Nemertean Mini-hemoglobins

| Table II | Comparison of heme pocket residues of C. lacteus neural and body wall globins with those determined for sperm whale myoglobin and human \( \beta \)-globin |

| Position | Helical designation | Sperm whale myoglobin | Human \( \beta \)-globin | Cerebratulus globins |
|----------|---------------------|-----------------------|---------------------|---------------------|
| Distal   | B10                 | L                     | L                   | Y                   |
| Distal   | B13                 | L                     | L                   | L                   |
| Distal   | C4                  | T                     | T                   | Y                   |
| Proximal | C7                  | K                     | F                   | K\*                 |
| Distal   | CD1                 | F                     | F                   | F                   |
| Distal   | CD3                 | R                     | S                   | F                   |
| Distal   | CD4                 | F                     | F                   | F                   |
| Distal   | E7                  | H                     | H                   | Q                   |
| Distal   | E10                 | T                     | K                   | K                   |
| Distal   | E11                 | V                     | V                   | T                   |
| Proximal | E14                 | A                     | A                   | Y                   |
| Distal   | E15                 | L                     | F                   | I                   |
| Proximal | F1                  | L                     | F                   | A                   |
| Proximal | F4                  | L                     | L                   | L                   |
| Proximal | F7                  | S                     | L                   | R                   |
| Proximal | F8                  | H                     | H                   | H                   |
| Proximal | FG2                 | H                     | L\*                 |                     |
| Proximal | G4                  | Y                     | N                   | E                   |
| Proximal | G5                  | L                     | F                   | F                   |
| Distal   | G6                  | I                     | L                   | A                   |
| Distal   | H12                 | L                     | V                   | L                   |
| Proximal | H15                 | F                     | V                   | L                   |
| Proximal | H19                 | I                     | L                   |                     |

a Helical designations are those for sperm whale Mb. Heme contact residues for sperm whale Mb and human Hb are from Refs. 26 and 72.

b Body wall globin has Arg at C-7.

c FG3 in the human \( \beta \)-chain.

The attraction of long branches to each other, and the different evolutionary rates among the branches can all result in false groupings. The length of sequences is certainly another variable that can give biased trees. Gaps are treated as the 21st character state in the program PROTPARS that we used, with each deletion being three changes away from all other character states (i.e. codons for all 20 amino acids). Therefore, short sequences with more matches of gaps in homologous positions to each other than to longer, non-deleted proteins, have a surplus of unique 21st state identities that might contribute to group them together to the exclusion of longer sequences. The effect of sequence length on the apparent topology is examined in trees B–D.

Tree B shows the result of shortening all sequences to 93 positions (E1–H19). This truncation removes length-variable NH₃- and COOH termini as well as the AB, CD, and D helix regions and minimizes the number of gaps involved in the analysis. The strength of the monophyletic support of the yeast/bacteria, the plant, and the “truncation” clades was not affected by this procedure, but the bootstrap value for the Cerebratulus truncation linkage increased from ~40 to ~59%. This finding shows that both the bootstrap confidence and the relative arrangement of lineages can be manipulated by artificial changes in sequence length.

Tree C takes this approach one step further. Here, the topology was determined by using only the 84 common core positions described by Kapp et al. (23) (see boxes in Fig. 8). This procedure confines the sequence data to the parts A10–A15, B5–C7, CD1, CD2, E1–E20, F1–F19, G4–G17, and H6–H19. This set of positions completely lacks any interhelical sequences except for CD1 and CD2. The result of this is that the bootstrap confidence for two major units (plants and bacteria/yeasts) drops below the 95% significance cutoff (bootstrap values in lightface), and the apparent relationship of Cerebratulus to other short sequences disappears. The sequences have become too short and too conservative to reveal any phylogenetic differentiation. This finding suggests that most of the phylogenetically informative positions are apparently located within interhelical and not in helical regions, especially in the EF, FG, and GH loops (compare trees B and C). This conclusion makes clear that phylogenetic differentiation using globins relies on the interhelical segments rather than on the functionally important but conservative blocks of helical residues.

Tree D displays the result of artificially shortening Glycera globin from 147 to 109 residues with no change in the other sequences. The shortening of Glycera globin was accomplished by introducing deletions in positions homologous to those found in the five short globins. The result of this procedure is to move the Glycera globin even closer to the other short globins except for Cerebratulus which is now displaced to an insignificant level of correspondence. A similar result was obtained by artificial truncation of Urechis globin which also displaced Cerebratulus and gave the following grouping: Cerebratulus (Urechis\*, short sequences)\#5\#6. Shortening Chironomus yielded a (((Cerebratulus, Chironomus)\#8 short sequences)\#9) topology (data not shown). These results demonstrate a difficulty in comparing homologous sequences of significantly different lengths. We conclude that the apparent relationship of the Cerebratulus sequence to other short sequences is a result of a common shortness, not common ancestry. Furthermore, the alignment in Fig. 8 makes clear that deletions in four miniglobins (Chlamydomonas, Nostoc, Tetrahymena, and Paramecium) occurred independently of those in Cerebratulus because the positions of the deletions differ.

\#3 Only the starred sequences are shortened. Bootstrap confidences are given as superscripts.
Hemoglobin Function—Data on oxygen equilibria measured in situ by microspectrophotometry (Fig. 12, A and B, and Table III) show that the neural tissue Hb has a higher oxygen binding affinity ($P_{50} = 2.9$ torr) than that of the body wall Hb ($P_{50} = 4.1$ torr) at 15 °C. Oxygen binding by the neural Hb in situ exhibits little or no dependence upon extracellular pH within the narrow range of pH 7.3–7.9 ($\Delta \log P_{50}/\Delta \text{pH}$, 0.1). Intracellular pH was not measured. Therefore, these results could mean that the intracellular pH was relatively constant and independent of extracellular pH.

Our results show that the oxygen equilibria for body wall Hb in vivo and in vitro at pH 7.4 are identical within the errors of measurement. The spectrum of the body wall Hb in vitro indicates that Met-Hb formation is no more than about 5% during the long 2–5-h O$_2$ binding experiments which indicates substantial stability. Under similar conditions, the truncated mini-Mbs would be 100% oxidized. Some Met-Hb was present initially, but we cannot accurately quantify this in the tissues because the extinction coefficients are unknown and because other absorbing substances are present. Although no in vitro O$_2$ binding measurements were made with the neural Hb, it proved to be relatively stable throughout the lengthy chromatographic preparation.

Oxygen binding by both Hbs is clearly cooperative. The mean Hill coefficient, 2.9, for O$_2$ binding by neural Hb at an estimated concentration of ~2–3 mM (heme) in situ is only slightly lower than that for human Hb. This means that the neural Hb must self-associate, probably at least to tetramers. The estimated molecular weight obtained by size-exclusion HPLC for HbO$_2$ is only 17,700 ± 2,000 obtained at a mean concentration across the peak of 2–5 μM. This value is the weight-average molecular weight expected in a mixture of monomers ($M_r \approx 12,000$) and dimers if the weight fraction of dimers were 0.48. These data can be interpreted in terms of an oxygenation-linked and concentration-dependent dissociation of presumed Hb tetramers: Hb$_4 \rightleftharpoons 2$ Hb$_2 \rightleftharpoons 4$ Hb. Oxygenation-linked dissociation processes are widespread in Hbs of both vertebrates and invertebrates (76–79) where the oligomeric Hbs are always more tightly associated in the deoxygenated state.

Our results show that the oxygen equilibria for body wall Hb in situ is similar to that of the neural Hb, but the data in Fig. 12B show a distinct difference. The slopes of the Hill plots for the body wall Hb in situ and in vitro show a pronounced decrease to 1.0–1.4 above 50% saturation ($\log (y/(1-y)) = 0$). Such a drop may result from dissociation of the oligomer, Hb$_4 \rightleftharpoons 2$ Hb$_2 \rightleftharpoons 4$ Hb, at earlier stages of oxygenation than occurs with the neural Hb. The maximal Hill coefficient approaches 2.9 for both in situ and in vitro body wall Hbs between $y = 0.3$ and $y = 0.5$.

Cerebratulus lives in intertidal and subtidal sediment in shallow water and mud flats along the North Atlantic coast, quite often deep below the surface in the anoxic zone where it burrows in search of prey. The relatively high concentration of high affinity Hb within the neural tissue suggests that the Hb may function as an O$_2$ store when the animal encounters anoxic conditions during burrowing. Oxygen storage has been demonstrated for the neural Hb of the clams T. alternata and S. solidissima, which live in very similar habitats (12, 18), and has been suggested for the neural Hb of A. aculeata (19).
oxygen store could also be useful under normoxic conditions to augment the oxygen supply to nervous tissues (29, 80). Such tissues may exhibit bursts of activity during which oxygen demand could temporarily exceed oxygen transport from the environment. A nearby auxiliary oxygen source would be valuable under such circumstances. We used a simple model to...
calculate the capacity of the potential O₂ store in the lateral nerves of *Cerebratulus*. Storage capacity was then divided by the potential O₂ demand to give estimates of the maximum time the stored O₂ would last. We assumed that no O₂ was lost by diffusion (see last paragraph) and that the O₂ consumption rate was constant and homogeneous throughout the nerves.

The lateral nerves of *Cerebratulus* are roughly cylindrical with nerve cell bodies and Hb restricted to an outer shell and axons to a central core (Fig. 2B). The volume of a cylindrical nerve cord (shell plus core) was 9.5 cm³, giving O₂ storage capacity 1.3–1.9 nmol or 1.9–2.4 nmol/ml, giving the oxygen storage capacity 1.3–1.9 nmol or 1.9–2.4 nmol/ml.

The large Hill coefficient of both neural and body wall Hb could function as a useful O₂ store to extend or augment neuronal activity under both anoxic and normoxic conditions. The presence of high affinity, low molecular weight Hbs in the body wall and nervous tissues could also serve to facilitate O₂ transport to and within these tissues.

Acknowledgments—We thank Peter Smith, Elizabeth Balser, and Pat Behrens for assistance with the experiments; Alfred Wheeler and Edward Ruppert for the use of equipment; and Janet Young, Gwen Gage, Kristina Schlegel, and David Green for assistance with graphics. We thank Sandra Smith of the University of Texas Microsequencing Facility for automated sequencing of proteins.

**REFERENCES**

1. Vandergron, T. L., and Colacino, J. M. (1989) Am. Zool. 29, 59
2. Vandergron, T. L., Riggs, C. K., Colacino, J. M., and Riggs, A. (1990) Am. Zool. 31, 72
3. Lankester, E. R. (1872) Proc. R. Soc. Lond. B Biol. Sci. 21, 70–81
4. Marwell, C. (1960) Arch. Biochem. Biophys. 89, 194–201
5. Wittenberg, J. B., Brown, P. R., and Wittenberg, B. A. (1965) Biochim. Biophys. Acta 109, 518–529
6. Wittenberg, B. A., Biriehl, R. W., and Wittenberg, J. B. (1965) Biochem. J. 96, 363–371
7. Weber, R. E. (1978) in *Physiology of Annelids* (Mills, P. J., ed) pp. 393–446, Academic Press, London
8. Kennedy, D. (1960) J. Gen. Physiol. 44, 277–299
9. Arvanitaki, A., and Chalazonitis, N. (1960) Biochim. Biophys. Acta 51, 1–83
10. Strittmatter, P., and Burch, H. B. (1963) Biochim. Biophys. Acta 78, 562–563
11. Schindelmeiser, I., Kuhlmann, E., and Nolte, A. (1979) Biochim. Biophys. Acta 562–563
12. Kraus, D. W., and Colacino, J. M. (1986) Science 232, 90–92
13. Fox, H. M. (1955) Proc. R. Soc. Lond. B Biol. Sci. 143, 203–214
14. Fox, H. M. (1957) Nature 179, 148
The Mini-hemoglobins in Neural and Body Wall Tissue of the Nemertean Worm, *Cerebratulus lacteus*

Thomas L. Vandergon, Claire K. Riggs, Thomas A. Gorr, James M. Colacino and Austen F. Riggs

*J. Biol. Chem.* 1998, 273:16998-17011.
doi: 10.1074/jbc.273.27.16998

Access the most updated version of this article at [http://www.jbc.org/content/273/27/16998](http://www.jbc.org/content/273/27/16998)

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 77 references, 15 of which can be accessed free at [http://www.jbc.org/content/273/27/16998.full.html#ref-list-1](http://www.jbc.org/content/273/27/16998.full.html#ref-list-1)