

The Primary Structure of Globin and Linker Chains from the Chlorocruorin of the Polychaete *Sabella spallanzanii**

Received for publication, August 2, 2000, and in revised form, April 2, 2001
Published, JBC Papers in Press, April 6, 2001, DOI 10.1074/jbc.M006939200

Alberto Pallavicini^{‡§}, Enrico Negrisolò^{‡§}, Roberto Barbatò[¶], Sylvia Dewilde^{||**},
Anna Ghiretti-Magaldi^{‡‡}, Luc Moens^{||}, and Gerolamo Lanfranchi^{‡§§}

From the [‡]Dipartimento di Biologia and Centro di Ricerca Interdipartimentale per le Biotechnologie Innovative Biotechnology Center, Università di Padova, Via Ugo Bassi 58/B, 35131 Padova, Italy, the [¶]Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale Amedeo Avogadro, Corso Borsalino 54 I-15100 Alessandria, Italy, the ^{‡‡}Dipartimento di Biologia and Consiglio Nazionale delle Ricerche Center for the Study of Metalloproteins, Università di Padova, via Ugo Bassi 58/B, 35131 Padova, Italy, and the ^{||}Department of Biochemistry, University of Antwerp UIA, Antwerp, Belgium

Annelid hemoglobins are organized in a very complex supramolecular network of interacting polypeptides, the structure of which is still not wholly resolved. We have separated by two-dimensional electrophoresis the 4-MDa chlorocruorin of *Sabella spallanzanii* and identified its components by amino-terminal sequencing. This work reveals a high rate of heterogeneity of constituent chains in a single animal as well as in the *Sabella* population. Using a cDNA library prepared from the hematopoietic tissue of this worm, we have isolated and fully sequenced most globin and linker cDNAs. The primary structure features of these polypeptides have been characterized by comparison with model globin and linker sequences.

Chlorocruorins (Chls)¹ are giant extracellular oxygen-binding heme proteins found in four marine polychaete families. They contain an altered heme with a formyl group substituting the canonical three-vinyl group and consequently appear as greenish red (1). The Chls have in electron micrographs the hexagonal bilayer (HBL) appearance and size characteristic of annelid and vestimentiferan extracellular Hbs, a similar sedimentation coefficient of ~60 S (molecular mass ~3, 600 kDa), and an abnormally low iron content of 0.23 weight % (2–6). The HBL Hbs, including the Chls, have almost identical three-dimensional structures as determined by cryoelectron microscopy at about 3-nm resolution (7–11).

All HBL Hbs and Chls are composed of two type of chains, heme-containing 16–17-kDa globin chains and nonglobin linker chains of 25–32 kDa in an approximate 2:1 molar ratio (1). The structural hierarchy within the HBL Hb of *Lumbricus*

terrestris was determined recently by x-ray diffraction analysis at 5.5 Å resolution (12). This Hb, composed of 144 globin and 36 linker chains, is arranged in dodecameric substructures. Twelve trimeric linker complexes project triple-stranded helical coiled-coil “spokes” toward the center of the complex; interdigitation of these spokes seems crucial for stabilization. The resulting complex of linker chains forms a scaffold on which 12 Hb dodecamers assemble.

The nature of the disulfide-bonded globin subunits composing the dodecameric structures is different among the different classes. Monomers and disulfide-bonded trimers or monomers and disulfide-bonded dimers are found in oligochaete/polychaete and leeches/vestimentiferans Hbs, respectively (5, 13, 14). In contrast the *Eudistylia vancouverii* Chl (15, 16) consists of two types of globin subassemblies, a dodecamer formed by the noncovalent association of three disulfide-bonded trimers and a tetramer formed by the noncovalent association of disulfide-bonded dimers.

In contrast to annelid Hbs, in which a wealth of sequence information of globin chains is available, only a single primary structure of a Chl globin chain has been published (17). Globin chain E of *Sabellastarte indica* shows 27–49% sequence identity with the annelid globin chains. Five cysteines, crucial for the subunit formation, are present. Two adjacent cysteines just preceding A1 and one at position H11 are conserved as in all annelid globin chains of type I (18). Two other cysteines occur at position E8 and within the corner between the G and H helices, as frequently seen in other annelid Hbs as well (19).

The Chl of *Sabella spallanzanii*, a marine fan worm polychaete formerly known as *Spirographis spallanzanii*, was chosen as a model molecule to identify the set of polypeptides that form the complete protein. The globin and linker polypeptides that compose its supramolecular structure have been resolved by two-dimensional gel electrophoresis. Three globin and three linker mRNAs and some variants have been isolated from *S. spallanzanii* cDNA libraries by screening with specific antibodies and have been completely sequenced and characterized.

MATERIALS AND METHODS

Purification and Gel Separation of *S. spallanzanii* Globin and Linker Chains—Live specimens of *S. spallanzanii* were collected in the Bay of Napoli. When Chl was prepared from pools of animals, the collected hemolymph was first centrifuged (10 min at 1,000 × g) and then frozen at –20 °C in 0.1 M Tris/HCl, pH 7.0, 35 mM CaCl₂, and 20% sucrose. Tablets of the Complete protease inhibitor kit (Roche) were added to prevent proteolytic activity. The Chl was sedimented by ultracentrifugation (3.5 h at 300,000 × g) at 4 °C, and the obtained pellet was resuspended in the same buffer and stored at –20 °C. The Chl concen-

* This work was supported in part by the Italian Ministry of University and of Scientific and Technological Research MURST Grant-Cofinanziamento Protocol 9805192993-002 (to A. G.-M.). 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 sequences reported in this paper have been submitted to the DDBJ/GenBank™/EBI Data Bank with accession numbers AJ131283 (globin 1), AJ131284 (globin 2), AJ131285 (globin 3), AJ131900 (linker 1), AJ131901 (linker 2) and AJ131286 (linker 3).

§ Both authors contributed equally to this work.

** Post-doctoral fellow of the Fund for Scientific Research-Flanders (FWO).

§§ To whom correspondence should be addressed. Tel.: 390-4982-76221; Fax: 390-4982-76280; E-mail: lanfra@cibi.unipd.it.

¹ The abbreviations used are: Chl, chlorocruorin; HBL, hexagonal bilayer; 2-DE, two-dimensional gel electrophoresis; ORF, open reading frame; Glb, globin; L, linker; pI, isoelectric point.

tration was determined using $E_{1\text{cm}}^{1\%} = 2.18$ for Chl at 280 nm. Absorption spectra were measured with an 8452A Hewlett Packard spectrophotometer. For single *S. spallanzanii* specimens, the crowns were dissected and immediately frozen in liquid nitrogen. The Chl was purified by running directly the hemolymph obtained from each crown on a fast protein liquid chromatography Superose 6 (Amersham Pharmacia Biotech) column in 0.1 M Tris/HCl, pH 7.0. The final purity was checked by transmission electron microscopy, spectroscopy, and polyacrylamide gel separation.

One- and two-dimensional electrophoresis (2-DE) were performed according to Fling and Gregerson (20) and Görg *et al.* (21). The first dimension was isoelectric focussing on a 4.0–7.0 Immobiline pH gradient. A 15% polyacrylamide/SDS slab gel electrophoresis was used as second dimension.

Protein Sequencing—Globin and linker chains were separated by 2-DE, and the band pattern was transferred to a polyvinylidene fluoride membrane (Millipore). Selected spots were sequenced in an Applied Biosystems ABI 471-B sequencer operating as recommended by the manufacturer.

cDNA Library Construction—The monolayer hematopoietic tissue of *S. spallanzanii* (22) was prepared from living animals, immediately frozen in liquid nitrogen, and kept at -80°C . The extraction of total RNA was performed according to Chomczynsky and Sacchi (23) using the tissues of a total of seven specimens. Poly(A)⁺ mRNA was purified by oligo(dT) affinity chromatography (24). The first strand cDNA was synthesized with an oligo(dT)-*NotI* primer using the Copy kit (Invitrogen). The double-stranded blunt-ended cDNA was ligated to *BstXI* nonpalindromic adaptors, *NotI*-digested, and finally directionally cloned into a *BstXI-NotI*-cut pcDNAII plasmid vector (Invitrogen). The recombinant vectors were electroporated into *Escherichia coli* strains *InvαF'* and *Top10F'* (Invitrogen).

Production of Anti-Chl Antibodies—Purified Chl was separated into its components by SDS-polyacrylamide gel electrophoresis (20), and the gel was stained with acid-free Coomassie. Three main bands were cut out from the gel and electro-eluted as described previously (25). After checking the purity by a second SDS-polyacrylamide gel electrophoresis run, the purified proteins were emulsified with complete Freund's adjuvant and injected into rabbits. Animals were booster-injected every second week over a period of 2 months. Blood was collected, and the serum was separated by centrifugation and stored at -80°C . As controls, sera were collected from the same rabbits before the immunization protocol. The specificity of the different antisera was checked by immunoblotting according to standard protocols (26).

Screening and Sequencing of *Sabella* Globin and Linker cDNA Clones—In a first round of screening, 100,000 bacteria colonies were grown on 10 nylon filters laid on square Luria Bertani broth plates (27). The colonies were then replicated twice on nitrocellulose filters and re-grown on Luria Bertani broth plates containing isopropyl β -D-thiogalactopyranoside as an inducer for the expression of the recombinant proteins. The clones containing the globin and linker cDNAs were detected by immunological screening using the specific antibodies at a dilution of 1:1,000. Two rounds of subscreening were performed to select single positive clones. Both strands of the cDNA inserts were sequenced by a primer-walking strategy using the fluorescent BigDye™ terminators chemistry (PE Biosystems), and the sequencing reactions were analyzed on an ABI-377 automated DNA sequencer (PE Biosystems). The sequences were assembled using the SeqMan II program from the Lasergene software package (DNASTar, Madison, WI). Full-length sequences of linker 3 and globin 1 were completed by 5' and 3' rapid amplification of cDNA ends on the *S. spallanzanii* cDNA. A primer designed on *BstXI* adaptor was used for the 5' rapid amplification of cDNA ends, whereas the oligo(dT)-*NotI* was used as primer for the 3' rapid amplification of cDNA ends reaction. One μl of adaptor-ligated cDNA was mixed with 200 μM dNTPs, 200 nM primers, 2.5 units of AmpliTaq GOLD (PerkinElmer Life Sciences), and 1.75 mM MgCl_2 . After the activation of AmpliTaq GOLD at 94°C for 15 min, 35 cycles of amplification were performed using the following steps: denaturation at 94°C for 30 s, annealing at 50°C for 1 min, and extension at 72°C for 1 min. The same protocol was adopted for the isolation of globin 3 cDNA, because immunoscreening identified no positive clones. In this case, a degenerated reverse primer coupled to the *BstXI* adaptor primer was designed on the amino terminus of globin D. A 180-base pair fragment was obtained and sequenced. On this sequence two nested forward primers were designed for reaching the 3' end of mRNA using the oligo(dT)-*NotI* as reverse primer.

Sequence Analysis—Several programs, available at the ExPASy molecular biology server, were used to analyze the new sequences. The ORFs codified by the cDNAs were obtained using the Translate pro-



FIG. 1. 2-DE patterns of *S. spallanzanii* chlorocruorin. A, representative 2-DE pattern of Chl from a single *S. spallanzanii* specimen. Two groups of components could be identified. The low molecular mass group consists of at least 10 different globin polypeptides. In the higher molecular mass linker group a total of eight spots could be detected. B, 2-DE pattern of Chl pooled from 500 *S. spallanzanii* specimens. A higher number of spots are detectable. A–M, globin spots; a–p, linker spots; Mw, molecular mass of marker proteins.

gram. Then the ORFs were manually aligned to the polypeptide amino termini sequenced during this work or with globins and linkers available from the literature and data bases to identify the export signal peptides. Successively, the pI and molecular mass were computed from the cDNA sequences with the program Compute pI/Mw and compared with the experimental data obtained from the 2-DE gel analysis. The pI and Mw measured with the two protocols are perfectly comparable. Amino acid compositions of the mature polypeptides were determined with the program ProtParam. Finally, the linker chains were aligned using the ClustalW program (28), and the globins were manually aligned according to the tertiary structure template of invertebrate globins (19).

RESULTS AND DISCUSSION

Determination of the Chl Polypeptide Components—A representative 2-DE separation of the Chl extracted from a single specimen is shown in Fig. 1A. A complex pattern of multiple spots is clearly detectable. They can be divided into three groups: heavy linkers (molecular mass ~ 35 kDa), light linkers (molecular mass ~ 31 kDa), and globin chains (Glb) (molecular mass ~ 14.4 kDa). This pattern is comparable with that obtained for the Chl of *E. vancoverii*, in which two groups of linkers (L_{1a-f} and L_{2a-d} , 10 chains) and six globin chains were detected by electron spray ionization mass spectrometry (16).

TABLE I
Primary structure of *Sabella* globins and linkers

Amino-terminal sequences of linker and globin chains of *S. spallanzanii* chlorocruorin. The 2-DE spots were blotted onto polyvinylidene fluoride, stained, and sequenced. Linker and globin sequences are named according to the spot lettering of Fig. 1B. The total number of amino acids determined for each polypeptide is indicated in brackets. X, undetermined amino acid.

Polypeptide	pI	Sequence
Linker <i>f</i> (36)	4.81	XXXWGSXLTTARVDSQEMRLDQLAKQLDWXEGXXWV
Linker <i>g</i> (28)	4.90	XXXWGSXLTTARVDSQEMRLDQLAKQLD
Globin <i>B</i> (44)	5.14	EAEEXMEDRQEVVKAWAMWXAEEFTGRRVXXXXXXFQRLFEEKP
Globin <i>C</i> (25)	5.40	EAEEXMEDRQEVVKAWAMWXAEEFT
Globin <i>D</i> (23)	5.47	XNILQRLKVKHQWQEAFGYADDR
Globin <i>E</i> (44)	5.66	WNILQRLKVKHQWSQAFGFTDEREDLNTDVVRYIFQVKPEAEAK
Globin <i>J</i> (32)	6.08	EGCCSMGDRQEVLNAAEWALWSAEYTGRRVMIA

Although the solution of the three-dimensional structure of the HBL Hb of *L. terrestris* offers a splendid explanation for the structural hierarchy within the molecule, no rationale is presented for the globin and linker chain multiplicity (12). The observed high cooperativity of the HBL Hbs (Hill coefficient $n_{50} > 3$) (29–32) as well as the aggregation into trimers, tetramers, and dodecamers definitively need different globin types (1) (*a*, *b*, *c*, and *d* in *L. terrestris*). The formation of the coiled coils in the linker scaffold complex also probably needs structurally different linker chain types (1) (L_{1-4} in *L. terrestris*).

A comparison of the globin and linker chains of a single animal (Fig. 1A) with that of a pool of 500 *S. spallanzanii* specimens (Fig. 1B) reveals a high degree of heterogeneity. Sixteen spots in the range of ~31 kDa and thirteen in the range of 14.4 kDa are clearly distinguishable. Amino-terminal sequencing of the major spots and alignment with published sequences confirm their identification as linkers and globin chains (Table I). There are three possible explanations for this heterogeneity: (i) the described variations (see below) suggest the presence of multiple copies of the same gene such that allelic as well as nonallelic variations can occur, (ii) post-translational modifications, or (iii) artificial modifications might be induced by the extraction and separation procedures used (21). It should be considered however that in invertebrates as well as in vertebrates, multiple copies of globin genes are a rule rather than an exception (33–35). Therefore this multiplicity could most likely be explained by the necessity to synthesize huge amounts of the oxygen carrier (35). A final conclusion on the exact number of globin and linker chains in *S. spallanzanii* must wait for a careful analysis of the Chl of single animal by electron spray ionization mass spectrometry.

Characterization of Specific Antibodies against Chl Components—For the molecular cloning of globin and linker cDNAs, we produced specific antibodies as described under “Materials and Methods.” Immunological tests on Chl Western blots reveal that the produced antibodies are efficient and specific for the detection of the polypeptides for which they were developed (Fig. 2). No cross-reaction was found between antibodies for globins and linkers. Some cross-reaction, however, was observed for the two classes of linkers (heavy and light) with the corresponding antisera, probably because of the high level of similarity among the members of the two classes.

Cloning and Sequencing of *Sabella* Globin and Linker cDNAs—A cDNA library was constructed using mRNA extracted from the hematopoietic tissue of *S. spallanzanii*. Immunological screening detected many recombinant clones that were sequenced. Six complete cDNAs, coding for three globin and three linker chains, have been identified and are presented in Fig. 3, A and B, respectively. Each cDNA represents the consensus sequence of at least three independent clones except for Glb3, for which only two positive clones were obtained (see “Materials and Methods”). The main features of the cDNA

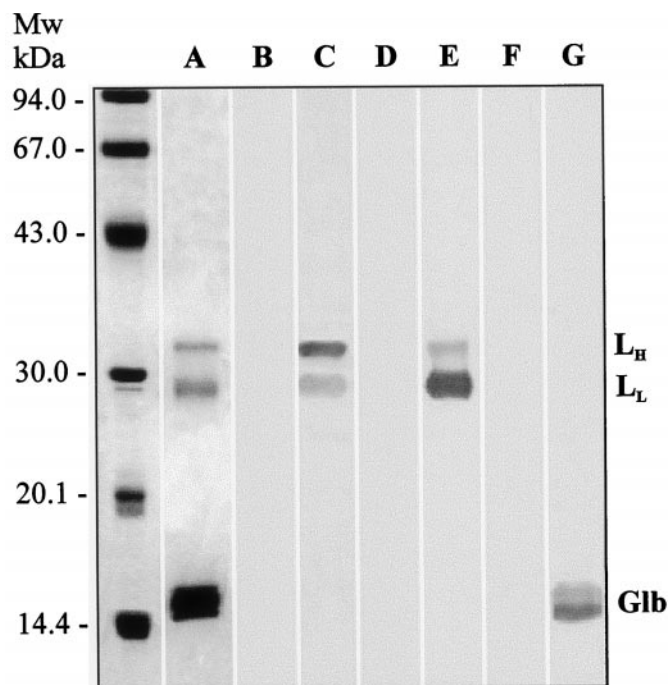


FIG. 2. Characterization of antibodies against *Sabella* Chl components. Chl chains were separated by SDS-polyacrylamide gel electrophoresis, blotted, and revealed by Ponceau red (lane A), pre-immune sera (lanes B, D, and F), and immunostaining with antibodies raised against heavy linkers (L_H), light linkers (L_L), and globins (lanes C, E, and G, respectively). Mw, molecular mass of protein standards.

sequences are summarized in Table II.

The ORF coded by the Glb1 cDNA can be identified as the polypeptide *J* in 2-DE (Fig. 1B) because the corresponding amino-terminal sequences are identical. Two variants for Glb2 cDNA were obtained that differ for two transversions in the coding region. The first (G versus T) is placed in the third base of the codon 66 and does not change the coded amino acid (Leu). The second (A versus C), however, produces a variation of the ORF (codon 92) with the presence of Ala instead of Asp in six of seven independent Glb2 cDNA clones. The signal peptide of Glb2 was identified by aligning the amino-terminal sequences obtained from the Glb B and C 2-DE spots (Fig. 1B; Table I). The cDNA for Glb3 is about 200 bases shorter than the cDNAs for Glb1 and Glb2. The signal peptide was identified aligning the ORF with the amino-terminal of the Glb D polypeptide. For this cDNA we found also some variants that differ both in the coding and in the 3'-noncoding regions. The C→T transition in the coding region (codon 42) produces Pro-Leu variants that have the same pI but vary slightly in molecular mass. One transversion and two transitions are also present in the 3'-noncoding regions of these two variants.

The cDNAs of L1 and L2 share 99% identity and code for

A

Table A: cDNA sequences and deduced ORFs of the S. spallanzanii Chl components. It lists sequences for Glb1, Glb2, Glb3, and linker chains (L1-L3) with their corresponding amino acid translations. Signal peptides and polyadenylation sites are underlined.

B

Table B: cDNA sequences and deduced ORFs of the S. spallanzanii Chl components. It lists sequences for linker chains (L1-L3) with their corresponding amino acid translations. Signal peptides and polyadenylation sites are underlined.

FIG. 3. cDNA sequences and deduced ORFs of the S. spallanzanii Chl components. A, globin chains; B, linker chains. Signal peptides and polyadenylation sites are underlined. ORF amino acids and variant nucleotides are in bold.

the same ORF (Fig. 3B). They are identical in the 5'-noncoding regions, whereas some transitions and transversions are accumulated in the coding and in the 3'-noncoding regions. The most important differences are a six-base deletion (5'→GAATA→3') and the insertion of a single T in L2. These two genes could have originated in a very recent gene duplication, because the observed differences are regularly present in all the independent clones. The amino-terminal amino acid sequences deduced from L1 and L2 cDNAs match with the amino-terminal sequences obtained, at the protein level, from the spots f and g in the 2-DE of Chl components (Fig. 1B; Table D). The cDNA for L3 is markedly shorter than those for L1 and L2; nevertheless it codes for an ORF that is two amino acids longer. No amino terminus was available to determine the signal peptide of L3. However, a putative signal can be identi-

fied by aligning the ORF with the amino terminus of Linker LAV1 of Lamellibrachia sp. (36). The amino acid compositions of the globin and linker chains deduced from the cDNA sequences were computed using the program ProtParam. The data are summarized in Fig. 4. Glb1 and Glb2 have a rather similar composition even though they show some peculiar differences (e.g. Glu is more abundant in Glb2). The amino acid composition of the other two globins. This is particularly evident for Ser residues. The percentage of Ser in Glb3 is twice that in Glb1 and Glb2. As mentioned above, L1 and L2 cDNAs both globin and linker chains show a high percentage of neg-

TABLE II
Main features of linker and globin cDNAs of *S. spallanzani chlorocruorin*

L_{AA} , length expressed as the total number of amino acids; mP, mature protein without signal peptide. Molecular mass and pI were determined by computer analysis of the cDNA sequences.

cDNA	length	ORF L_{AA}	mP L_{AA}	mP molecular mass	mP pI
	bp			Da	
Linker 1	1872	246	222	24120.73	4.80
Linker 2	1867	246	222	24120.73	4.80
Linker 3	1352	248	233	26170.35	4.96
Globin 1	1086	165	149	16793.11	5.79
Globin 2, var. Ala	1177	166	149	16877.26	5.62
Globin 2, var. Asp	1177	166	149	16921.27	5.38
Globin 3, var. Leu	914	159	144	16011.16	5.49
Globin 3, var. Pro	914	159	144	15995.12	5.49

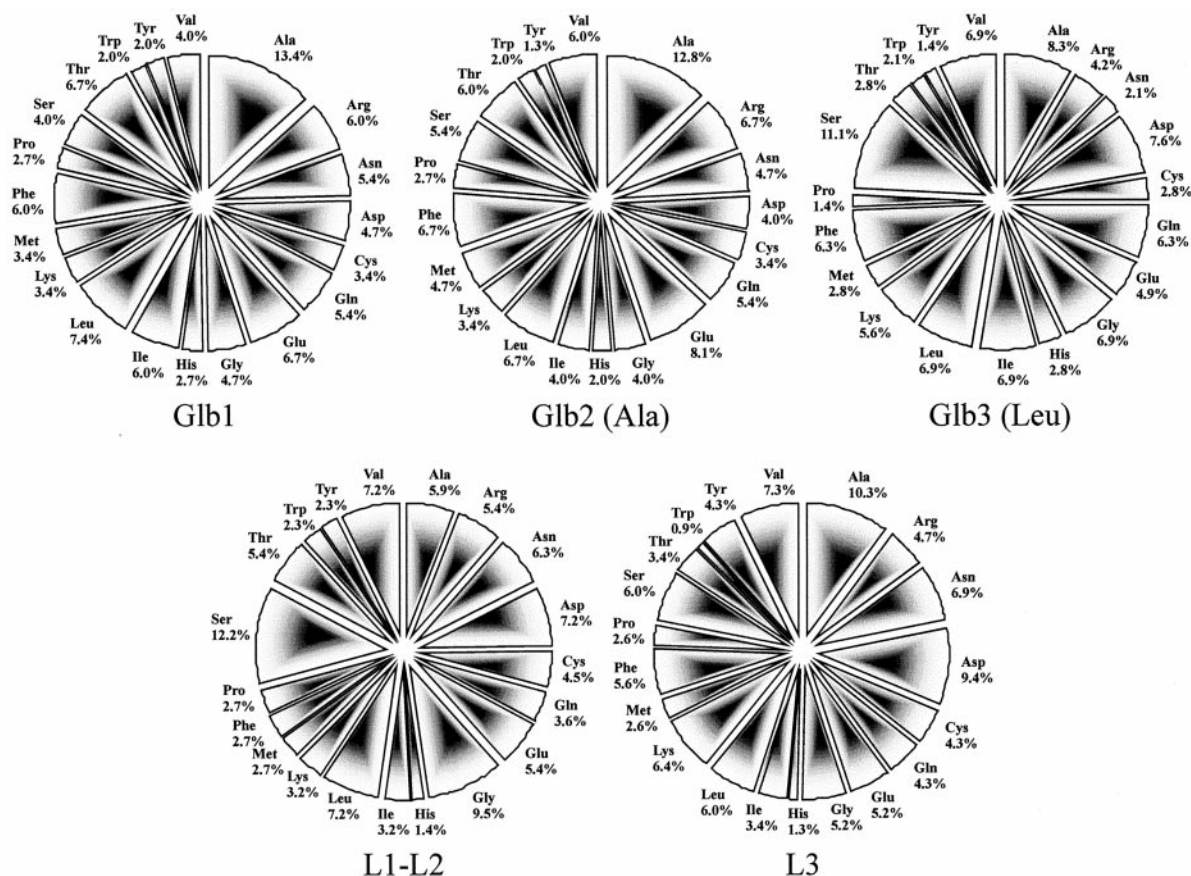


FIG. 4. Amino acid composition of *S. spallanzanii* chlorocruorin globin and linker chains. Data are expressed as percentages. For Glb2 and Glb3, the data refer to the Ala and Leu variants, respectively.

actively charged residues that explains the low pI calculated for all the chains (Table II).

Sequence Alignment and Characterization—To characterize the primary structure of the *Sabella* globin chains, we have aligned them with the annelid, pogonophoran, and vestimentiferan globin sequences available in our data base including a globin chain from the Chl of *S. indica* and globin references (17–19). A representative selection of this alignment is presented in Fig. 5. The alignment is unambiguous because of the presence of the globin landmarks A12-Trp, C2-Pro, CD1-Phe, E7-His, F8-His and H8-Trp. All three novel globin chains fit the nonvertebrate globin template quite well, resulting in low penalty scores (19). Glb3 has a hydrophobic residue at the surface positions A6 and CD2. Because similar hydrophobic substitutions occur in other annelid globin sequences, it might be that they represent specific adaptations for the aggregation into high molecular mass complexes. No specific adaptations can be localized to harbor the formyl group on the heme ring. The

positions of cysteine residues in annelid globin chains are strictly conserved because of their role in the formation of disulfide-bonded subunits (18). On the basis of the pattern of these residues, two types of chains can be distinguished. Type I has absolutely conserved cysteines at positions NA2 and H11 and a less conserved one at position GH4. Type II displays the same pattern with an additional cysteine at position NA1. As such, Glb3 can be classified as type I and Glb1 and Glb2 can be classified as type II. In all annelid-like globin chains studied thus far, the cysteines at positions NA2 and H11 are involved in an intrachain disulfide bridge linking the NA terminus to the H-helix and leaving the two other cysteine residues free for the formation of inter-chain bridges (12, 16). As such, based on the alignment it can be concluded that the globin chains of *S. spallanzanii* are similar to the other annelid, pogonophoran, and vestimentiferan globin sequences. Therefore, the *Sabella* globin primary structure is not sufficiently informative to explain how these Hbs have acquired the possibility of harboring

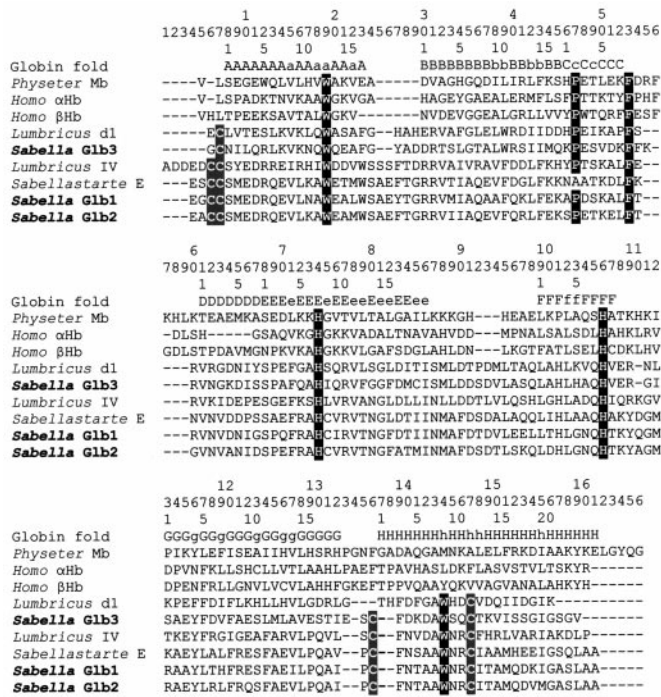


FIG. 5. Alignment of globin chains. The sequences have been aligned in the coding portions using as a reference (19) the tertiary structure template of invertebrate globins (Globin fold). *Physeter Mb* (P02185), *Physeter catodon* myoglobin; *Homo αHb* (P01922), *Homo sapiens α-chain*; *Homo βHb* (P02023), *Homo sapiens β-chain*; *Lumbricus d1* (U55073), *L. terrestris* Hb d1; *Sabella G1b3* (AJ131285), *S. spallanzanii* globin 3; *Lumbricus IV* (P13579), *L. terrestris* Hb IV; *Sabellastarte E* (D58418), *S. indica* globin; *Sabella G1b1* (AJ131283), *S. spallanzanii* globin 1; *Sabella G1b2* (AJ131284), *S. spallanzanii* globin 2 are shown (the accession numbers from the Swiss-Prot, NCBI, and EMBL data banks are in parentheses). Black bars indicate the position of amino acids that are important for the globin fold, and the gray bars indicate the cysteine residues that characterize the two groups of annelid globins (see text for details).

the modified heme group.

A similar alignment was carried out for the linker sequences translated from the cloned cDNAs with two other linkers of *Lumbricus* and *Lamellibrachia* (Fig. 6). The most remarkable feature of the linker chains, including these of *S. spallanzanii*, is a conserved 38–39-residue segment containing a repeating pattern of cysteinyl residues ((Cys-X₅₋₆)₃-Cys-X₅-Cys-X₁₀-Cys) (36–38). This pattern is identical with the cysteine-rich repeats of the ligand-binding domain of the low density lipoprotein receptors of man and *Xenopus laevis*. The sequence Asp-Gly-Ser-Asp-Glu, characteristic of the low density lipoprotein receptor repeats, is less conserved in *S. spallanzanii* L1, L3, and *Lamellibrachia* LAV1 than in *L. terrestris* L1. Nevertheless, the pattern Asn-Gly-X-Asp-Glu is easily recognizable in the *S. spallanzanii* linker sequences (38). Two other cysteines at positions 129 and 228 are also conserved together with a Leu, a Gly, and a Tyr residue at positions 31, 122, and 182, respectively. The suggestion that linker chains resulted from gene duplication of a heme-containing chain with a three exon/two intron structure and that the first exon of domain 1 and the last exon of domain 2 have been lost during evolution can not be confirmed (37). It is more likely that the cysteine-rich motif of the low density lipoprotein receptor and linker chains represents a multipurpose protein-binding unit of ancient origin that has been incorporated into diverse unrelated proteins by the process of exon shuffling (38).

In this paper we have presented the primary structure of three globin and three linker chains composing the Chl of *S. spallanzanii* and thereby significantly increase the number of

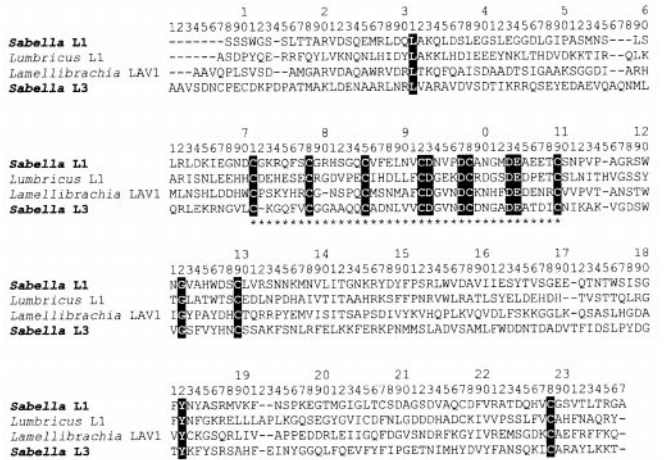


FIG. 6. Alignment of linker chains. Sabella L1 (AJ131900), *S. spallanzanii* linker 1; *Lumbricus* L1 (A46587), *L. terrestris* linker 1; *Lamellibrachia* LAV1 (P16222), *Lamellibrachia* sp. linker LAV1; *Sabella* L3 (AJ131286), *S. spallanzanii* linker 3 are shown (the accession numbers from the Swiss-Prot, NCBI, and EMBL data banks are in parentheses). Black bars indicate the position of invariant amino acids, and the asterisks underline the conserved cysteine-rich segment (see text for details).

annelid, pogonophoran, and vestimentiferan sequences available today. A detailed evolutionary analysis of our data in combination with that available from the literature is presented in an accompanying paper (39).

Acknowledgments—We thank Dr. Lucio Cariello Director of the Stazione Zoologica “A. Dohrn” (Napoli, Italy) for providing the specimens of *Sabella spallanzanii* and Drs. Silvia Trevisan, Beniamina Pacchioni, and Rosanna Zimbello for help in sequencing. Marie-Louise Van Hauwaert is acknowledged for skillful technical assistance.

REFERENCES

- Lamy, J. N., Green, B. N., Toulmond, A., Wall, J. S., Weber, R. E., and Vinogradov, S. N. (1996) *Chem. Rev.* **96**, 3113–3124
- Terwilliger, R. C., Garlick, R., Terwilliger, N. B., and Blair, D. P. (1975) *Biochim. Biophys. Acta* **400**, 302–309
- Terwilliger, R. C., Terwilliger, N. B., and Schabtach, E. (1976) *Comp. Biochem. Physiol. A* **55**, 51–55
- Vinogradov, S. N., Kapp, O. H., and Ohtsuki, M. (1982) in *Electron Microscopy of Proteins* (Harris, J., ed) Vol. 3, pp. 135–163, Academic Press, New York
- Vinogradov, S. N., Standley, P. R., Mainwaring, M. G., Kapp, O. H., and Crewe, A. V. (1985) *Biochim. Biophys. Acta* **828**, 43–50
- Vinogradov, S. N., Sharma, P. K., and Walz, D. A. (1991) *Comp. Biochem. Physiol. B* **98**, 187–194
- De Haas, D., Taveau, J. C., Boisset, N., Lambert, O., Vinogradov, S. N., and Lamy, J. N. (1996) *J. Mol. Biol.* **255**, 140–153
- De Haas, D., Boisset, N., Taveau, J. C., Lambert, O., Vinogradov, S. N., and Lamy, J. N. (1996) *Biophys. J.* **70**, 1973–1984
- De Haas, D., Zal, F., Lallier, F., Toulmond, A., and Lamy, J. N. (1996) *Proteins Struct. Funct. Genet.* **26**, 241–256
- De Haas, D., Zal, F., You, V., Lallier, F., Toulmond, A., and Lamy, J. N. (1996) *J. Mol. Biol.* **264**, 111–120
- De Haas, D., Kuchumov, A. R., Taveau, J.-C., Boisset, N., Vinogradov, S. N., and Lamy, J. N. (1997) *Biochemistry* **36**, 7330–7338
- Royer, W. E., Strand, K., van Heel, M., and Hendrickson, W. A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7107–7111
- Vinogradov, S. N. (1985) in *Respiratory Pigments in Animals* (Lamy, J., Truchot, J. P., and Gilles, R., eds) pp. 9–20, Springer Verlag, Berlin
- Zal, F., Lallier, F., Green, B. N., Vinogradov, S. N., and Toulmond, A. (1996) *J. Biol. Chem.* **271**, 8875–8881
- Qabar, A. N., Stern, M. S., Walz, D. A., Chiu, J. T., Timkovich, R., Wall, J. S., Kapp, O. H., and Vinogradov, S. N. (1991) *J. Mol. Biol.* **222**, 1109–1129
- Green, B. N., Braswell, E. H., Kuchumov, A. R., Walz, D. A., Moens, L., and Vinogradov, S. N. (1998) *Biochemistry* **37**, 6598–6605
- Suzuki, T., Hirao, Y., and Vinogradov, S. (1995) *Biochim. Biophys. Acta* **1252**, 189–193
- Zal, F., Suzuki, T., Kawasaki, Y., Childress, J. J., Lallier, F. H., and Toulmond, A. (1997) *Proteins* **29**, 562–574
- Kapp, O. H., Moens, L., Vanfleteren, J., Trotman, C., Suzuki, T., and Vinogradov, S. (1995) *Protein Sci.* **4**, 2179–2190
- Fling, S. P., and Gregerson, D. S. (1986) *Anal. Biochem.* **155**, 83–88
- Görg, A., Postel, W., and Günter, S. (1988) *Electrophoresis* **9**, 531–546
- Mezzasalma, V., Tognon, G., and Ghiretti-Magaldi, A. (1992) *Anim. Biol.* **1**, 85–96
- Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69**, 1408–1412

25. Barbato, R., Shipton, C. A., Giacometti, G. M., and Barber, J. (1991) *FEBS Lett.* **290**, 162–166
26. Bjerrum, O. J., and Heegaard, N. H. H. (1988) *CRC Handbook of Immunoblotting of Proteins: Technical Descriptions*, Vol. 1, CRC Press, Inc., Boca Raton, FL
27. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, 2nd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
28. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
29. Antonini, E., Rossi-Fanelli, A., and Caputo, A. (1962) *Arch. Biochem. Biophys.* **97**, 336–342
30. Antonini, E., Rossi-Fanelli, A., and Caputo, A. (1962) *Arch. Biochem. Biophys.* **97**, 343–350
31. Imai, K., Sharma, P. K., and Vinogradov, S. N. (1996) *Comp. Biochem. Physiol. B* **113**, 613–618
32. Imai, K., and Yoshikawa, S. (1985) *Eur. J. Biochem.* **147**, 453–463
33. Dickerson, R. E., and Geis, I. (1983) *Hemoglobin: Structure, Function, Evolution and Pathology*, Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA
34. Rashid, A. K., Van Hauwaert, M.-L., Haque, M., Siddiai, A. H., Lasters, I., De Mayer, M., Griffon, N., Marden, M. C., Dewilde, S., Clauwaert, J., Vinogradov, S. N., and Moens, L. (1997) *J. Biol. Chem.* **272**, 2992–2999
35. Dewilde, S., Blaxter, M., Van Hauwaert, M.L., Van Houtte, K., Pesce, A., Griffon, N., Kiger, L., Marden, M. C., Vermeire, S., Vanfleteren, J., Esmans, E., and Moens, L. (1998) *J. Biol. Chem.* **273**, 32467–32474
36. Suzuki, T., Takagi, T., and Ohta, S. (1990) *J. Biol. Chem.* **265**, 1551–1555
37. Suzuki, T., Takagi, T., and Gotoh, T. (1990) *J. Biol. Chem.* **265**, 12168–12177
38. Suzuki, T., and Riggs, A. (1993) *J. Biol. Chem.* **268**, 13548–13555
39. Negrisolò, E., Pallavicini, A., Barbato, R., Dewilde, S., Ghiretti Magaldi, A., Moens, L., and Lanfranchi, G. (2001) *J. Biol. Chem.* **276**, 26391–26397

**The Primary Structure of Globin and Linker Chains from the Chlorocruorin of the
Polychaete *Sabella spallanzanii***

Alberto Pallavicini, Enrico Negrisolò, Roberto Barbato, Sylvia Dewilde, Anna
Ghiretti-Magaldi, Luc Moens and Gerolamo Lanfranchi

J. Biol. Chem. 2001, 276:26384-26390.

doi: 10.1074/jbc.M006939200 originally published online April 6, 2001

Access the most updated version of this article at doi: [10.1074/jbc.M006939200](https://doi.org/10.1074/jbc.M006939200)

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 35 references, 9 of which can be accessed free at
<http://www.jbc.org/content/276/28/26384.full.html#ref-list-1>