Amino Acid Sequence of the Monomer Subunit of the Extracellular Hemoglobin of the Earthworm, Pheretima hilgendorfi

Author: Shishikura, Fumio

Source: Zoological Science, 13(4) : 551-558

Published By: Zoological Society of Japan

URL: https://doi.org/10.2108/zsj.13.551

ZOOLOGICAL SCIENCE **13**: 551–558 (1996)

Amino Acid Sequence of the Monomer Subunit of the Extracellular Hemoglobin of the Earthworm, *Pheretima hilgendorfi*

Fumio Shishikura

Department of Biology, Nihon University School of Medicine, Ohyaguchi-kami machi, Itabashi-ku, Tokyo 173, Japan

ABSTRACT—The complete amino acid sequence of the monomer subunit of *Pheretima hilgendorfi* hemoglobin was determined: It consists of 140 amino acid residues, including a disulfide bond but no methionine, and has a molecular weight of 16,107 Da. Using computed analyses (amino acid maximum homology) with known sequences of monomer subunits of earthworm's hemoglobins, 115 (82%) were found to be identical with those in the corresponding positions of chain I (monomer subunit) of *Pheretima sieboldi* hemoglobin; 81 residues (55%), 71 residues (47%), and 66 residues (43%) were found to be in identical positions of the sequences of chain I of *Lumbricus terrestris* hemoglobin, chain I of *Tubifex tubifex* hemoglobin and chain I of *Tylorrhynchus heterochaetus* hemoglobin. Orthologous sequence data of monomer globins that belong to the strain A of annelid hemoglobins are discussed as useful clues for investigation of the divergence pattern of *Pheretima* species.

INTRODUCTION

Annelid hemoglobin is a giant extracellular protein comprised of two main heme-containing subunits, a monomer and a disulfide-bonded trimer, and a linker including two or three chains (Vinogradov, 1985). To date, several chains of earthworm hemoglobin have been sequenced, monomer globins in particular, from oligochaetes (Shishikura *et al.*, 1987; Suzuki, 1989a; Stern *et al.*, 1990) as well as polychaetes (Suzuki *et al.*, 1982). However, there has been no phylogenetical study on the primary structures of hemoglobins from a single genus of annelid.

The *Pheretima* group contains over 750 nominal species from south Asia and is the dominant component of earthworm fauna (Easton, 1984). In Japan, there are as many as 155 species of the genus *Pheretima* (Kobayashi, 1941), including such cosmopolitan species as *Pheretima hilgendorfi* and *Pheretima communissima* in the central and northern regions. It is very difficult to demonstrate the phylogenetical relationship among the large members of *Pheretima* by using only morphological classification. Recently, the mitochondrial DNA of earthworms (from *Lumbricus terrestris*) has been sequenced (Boore and Brown, 1995), and it is hoped to be used for analyzing genetic divergence patterns among the species. On the contrary, appropriate protein molecules for use in the investigation should be selected by a method by which the orthologous relationship between homologous genes can be detected (Sugita and Shishikura, 1995). Hemoglobin molecules are good materials for this purpose since a previous study showed that earthworm hemoglobin constitutes two strains of globins, in which an orthologous relationship is expected to show evidence for the molecular evolution of annelid hemoglobins (Gotoh *et al.*, 1987). However, little is known about the molecular divergence pattern of species belonging to the genus of *Pheretima*, even primary globin structures, except for the sequence of *Pheretima sieboldi* hemoglobin (Suzuki, 1989a). The main purposes of this study are to establish the amino acid sequence of the monomer subunit of *Pheretima hilgendorfi* hemoglobin and obtain a clue for analyzing the phylogenetical relationship among the *Pheretima* group.

In computed analyses among the known sequences, the identity of 82% of the amino acid sequence was found in the two monomer globins from *Pheretima hilgendorfi* (this study) and *Pheretima sieboldi* (Suzuki, 1989a). This value, as expected, is the highest degree of sequence similarity among known sequences of annelids (Goodman *et al.*, 1988; Vinogradov *et al.*, 1993; Kapp *et al.*, 1995). Analyzing the primary structure of homologous sequences of hemoglobins can provide evolutionary implications clarifying the phylogenetical relationship among the *Pheretima* group as well as annelids.

MATERIALS AND METHODS

Materials

Live earthworms *Pheretima hilgendorfi* were obtained from a local supplier. Formic acid, *5,5'*-dithio-*bis* 2-nitorbenzoic acid, *tri-n*-butyl phosphine, *4*-vinylpyridine and *p*-amidinophenylmethanesulfonyl fluoride-HCl (*p*-APMSF) were purchased from Nacalai Tesque (Kyoto, Japan). Methylethylketone, trifluoroacetic acid (TFA), acetonitrile,

TPCK-trypsin (type XIII; from bovine pancreases), V8 protease (from *Staphylococcus aureus* strain V8), and pepsin (from porcine stomach mucosa) were obtained from Sigma Chemical Co. Lysyl endopeptidase (*Achromobactor* protease I) was purchased from Wako Pure Chemicals Co. (Tokyo, Japan). Separation columns, Superose 12, PepRPC 5/5 and ProRPC 5/10 were purchased from Pharmacia Biotech., and placed in a Fast Protein Liquid Chromatography (FPLC) system (Pharmacia Biotech., Sweden). Amino acid analyses and sequencing analyses were performed, respectively, using a Shimadzu amino acid analyzer LC-10A and a gas phase protein sequencer PPSQ-10 (Shimadzu Co., Kyoto, Japan).

Preparation of hemoglobin

The *Pheretima hilgendorfi* hemoglobin was prepared as previously described (Shlom and Vinogradov, 1973) with slight modifications (Shishikura *et al.*, 1987). The blood was withdrawn into 50 mM Tris-HCl, pH 8.0, containing 1 mM EDTA (final concentration) and two kinds of serine proteinase inhibitors (final concentrations; 2.5 mM *p*-aminobenzamidine and 2.5 mM *p*-APMSF). The hemoglobin was converted to the cyanomet form by the addition of 2% $K_3Fe(CN)_6$, 0.5% KCN, and 0.1% $NaHCO_3$ (Moss and Ingram, 1968). Heme was removed from the hemoglobin using methylethylketone as described (Teale, 1959).

Purification of monomer globin

Monomer globin of *Pheretima hilgendorfi* was separated and purified by the method described previously (Shishikura *et al.*, 1987) with a slight modification. Final purification of globin was achieved by gel-filtration of the apoprotein in 35% formic acid on Superose 12.

SDS-polyacrylamide gel electrophoresis

SDS-polyacrylamide gel electrophoresis was conducted in a 10-20% acrylamide gradient gel containing 0.08% *N,N'*-methylenebisacrylamide, 0.375 M Tris-HCl (pH 8.8) and 0.1% SDS (Laemmli, 1970). The protein standards used were bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (29 kDa) and β-lactoglobulin (18.4 kDa). The unreduced hemoglobin was electrophoresed and stained by the method of Granath (1988) to detect heme containing protein in the SDS-gels.

Protein modification

Reduction and alkylation (*S*-pyridylethylation or *S*-carboxymethylation) of globin were performed separately by methods described elsewhere (Friedman *et al.*, 1970; Waxdal *et al.*, 1968). Finally, the remaining reagents in the reaction mixture were removed completely from the sample by gel-filtration on a Superose 12 column in 35% formic acid or a reversed-phase column.

Enzymatic cleavages

Lysyl endopeptidase digestion was performed essentially as described (Jekel *et al.*, 1983). Briefly, a sample (15-20 nmoles) of the *S*-pyridylethyladed (or *S*-carboxymthylated) protein was first dissolved in 8 M urea and incubated at 37°C for 30 min, followed by the addition of 0.5 M ammonium bicarbonate at a final concentration of 4 M urea in 0.1 M ammonium bicarbonate. Lysyl endopeptidase digestion of the sample was performed at an enzyme/substrate ratio of 1/30 (mol/mol) for 4 hr at 37°C in 0.1 M ammonium bicarbonate, pH 8.2, containing 4 M urea. Two peptides (designated as K3 and K5) derived from the monomer globin were digested further with trypsin at an enzyme/substrate ratio of 1/200 (by mass) at 37°C for 2 hr in 0.1 M Tris-HCl, pH 8.0, containing 2 M urea. To obtain the overlapping peptides, the monomer globin (20 nmoles) was digested with the V8 protease at a ratio of 1:100 (w/w, enzyme/substrate) for 48 hr at 37°C in 0.1 M Tris-HCl, pH 8.0, containing 1 M urea. The intact globin (10-20 nmoles) was also digested with pepsin at an enzyme/substrate ratio of 1/200 (by mass) at 37°C for 3 hr in 5% formic acid.

Peptide separation

All peptides derived from monomer globin were separated using a reversed-phase column (PepRPC HR 5/5 or ProRPC HR 5/10) in a 0.1% TFA buffered gradient to 60% acetonitrile in 0.08% TFA. Flow rates were maintained at 0.3 ml/min. All fractions were monitored at 214 nm and 280 nm. Re-chromatography of selected peptides was performed as previously described (Shishikura *et al.*, 1987). The presence of free cysteine residues in the peptide fractions of pepsin digestion as well as lysyl endopeptidase digestion of intact protein were examined by the method of Ellman (Ellman, 1959).

Amino acid analysis

Samples (4 nmoles or 10-20% of peptide fractions from reversed phase column chromatography) were hydrolyzed in 6 N HCl-1% phenol at 105°C for 22 hr, and dried using a Pico-Tag workstation. The amino acids in the hydrolyzates were analyzed by a on-line post-column derivatization to form OPA (*o*-phthalaldehyde)-amino acids using a Shimadzu Amino Acid Analyzer.

Amino acid sequence determination

Sequence analysis was performed using a Shimadzu gas phase protein sequencer, PPSQ-10, equipped with a PTH-10 amino acid analyzer. PTH-derivatives from the sequencer were separated and quantitated.

Prediction of secondary structure

Prediction of conformation was computed by the method of Chou and Fasman (1974). Computations were made using the DNASIS programs (NBRF-Protein Sequence Database, version 23, 1995) of Hitachi Software Engineering Co., Ltd (Yokohama, Japan).

RESULTS

Figure 1A shows the gel-filtration chromatography of whole *Pheretima hilgendorfi* hemoglobin on a Superose 12 column in 35% formic acid. The monomer fraction (indicated by the horizontal line) was rechromatographed under the same conditions (Fig. 1B). The purity of the monomer globin was examined on SDS-polyacrylamide gel electrophoresis in the presence or absence of a reducing agent (Fig. 2). An apparent molecular weight of the purified monomer globin is estimated to be 15,500 +/– 500 based on SDS-polyacrylamide gel electrophoresis.

The strategies used to establish the complete amino acid sequence of the monomer globin of *Pheretima hilgendorfi* are summarized in Fig. 3 and the data to support this figure are provided in Fig. 4 and Tables 1-4. The globin is composed of 140 amino acid residues and contains no methionine. The molecular mass was calculated to be 16,107 Da. All overlaps were quantitatively confirmed by at least two amino acids except for residues 42 to 47 and 53 to 64. By sequencing the peptide fraction containing a half-cystine residue, cytsteines at position 2 and 130 were revealed to be linked by an interchain disulfide bond.

Figure 5 shows that the monomer globin sequence is aligned with that previously reported for *Pheretima sieboldi*, and those for two species of Oligochaeta, including *Lumbricus terrestris* and *Tubifex tubifex*, and that for a marine polychaete *Tylorrhynchus heterochaetus*. When the globin sequence of *Pheretima hilgendorfi* is compared with those of known sequences, there are 54 invariant amino acids among the four

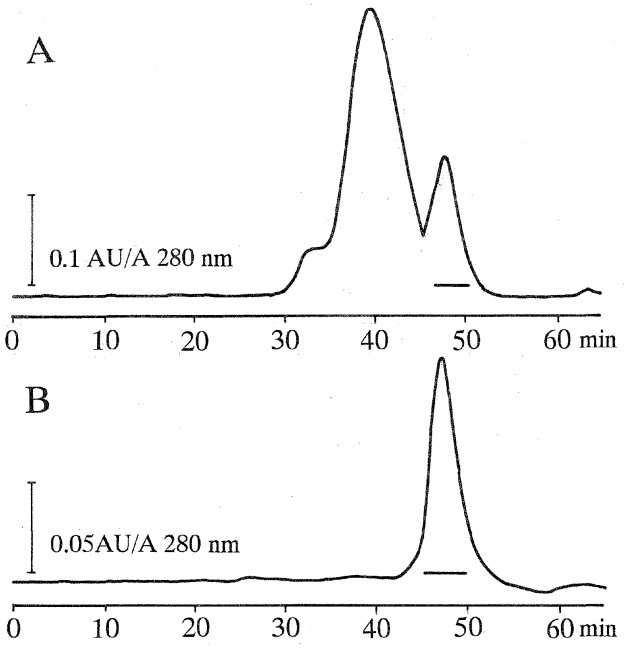

Fig. 1. A; Superose 12 chromatography of whole *Pheretima hilgendorfi* hemoglobin in 35% formic acid at a flow rate of 0.3 ml/min. Fractions for *Pheretima hilgendorfi* monomeric globin were pooled as indicated by the horizontal line. B; Rechromatography of *Pheretima hilgendorfi* monomeric globin on Superose 12. Fractions for the purified monomer globin were pooled as indicated by the horizontal line.

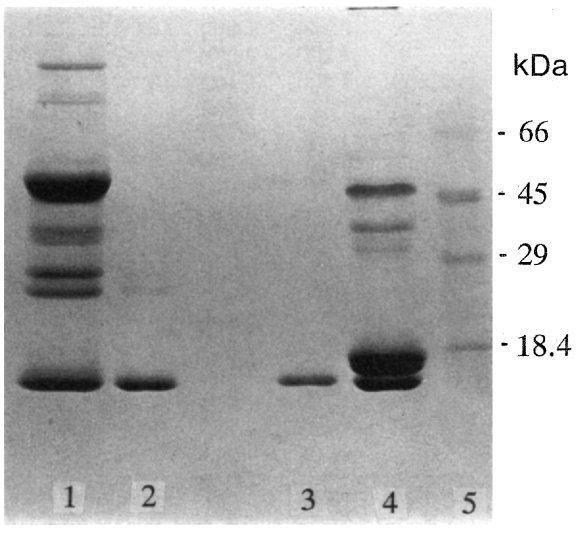

Fig. 2. SDS-polyacrylamide gel electrophoresis of *Pheretima hilgendorfi* hemoglobin. Lanes 1 and 4 are the unreduced and reduced whole hemoglobin, respectively; lanes 2 and 3 are the unreduced and reduced monomeric globin, respectively; lane 5 is standard proteins (bovine serum albumin, ovalbumin, carbonic anhydrase and β-lactoglobulin).

globins except for the *Tylorrhynchus* globin. As for the identities of amino acid residues, there are 115 residues (82%) identical to the globin sequence of *Pheretima sieboldi*, 81 residues (55%) to *Lumbricus*, 71 residues (47%) to *Tubifex* and 66 residues (43%) to *Tylorrhynchus*.

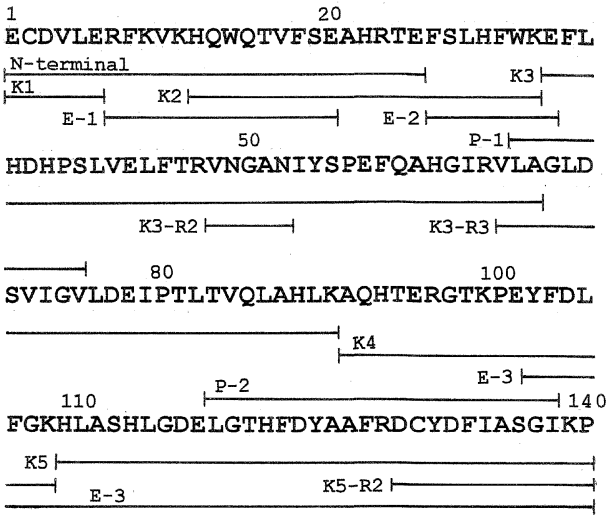

Fig. 3. The complete amino acid sequence of the monomer subunit of *Pheretima hilgendorfi* hemoglobin deduced from overlapping peptides and fragments. The residues marked with lines are those identified by Edman degradation. Vertical lines represent the beginning and the end of sequencing. Peptide nomenclatures are as follows: lysyl endopeptidase, K; V8 protease, E; pepsin, P; trypsin after lysyl endopeptidase, K-R.

DISCUSSION

During the course of this study, the biochemical properties of *Pheretima hilgendorfi* hemoglobin reported in previous studies (Ochiai, 1979; Ochiai and Enoki, 1979) were mostly confirmed: two-heme containing subunits, a monomer and a disulfide-linked trimer and a linker comprising two or three chains. Subunits containing heme were identified by the method of Granath (1988) in the SDS-polyacrylamide gel electrophoresis without denaturation (data not shown). As for the molecular weight of monomer globin, the previous study (Ochiai and Enoki, 1979) estimated it to be 13,100 +/– 200 (non reduced form) and 13,700 (reduced form) based on SDS-polyacrylamide gel electrophoreses. After sequenced, however, the value was calculated to be 16,107 Da, which is comparable to those of other monomer globins of *Pheretima sieboldi* (16,911 Da; Suzuki, 1989), *Lumbricus* (16,750 Da; Shishikura *et al.*, 1987), *Tubifex* (16,286 Da; Stern *et al.*, 1990) and *Tylorrhynchus* (16,327 Da; Suzuki *et al.*, 1982).

The complete amino acid sequence of the monomer globin of *Pheretima hilgendorfi*, which is the second report on the primary structure of the *Pheretima* group, is aligned with that of the homologous globins of four oligochaetes and one polychaete (Fig. 5). Several interesting similarities and differences were observed: The chain length of the globin sequence of *Pheretima hilgendorfi* is a 140-residue protein and others are 141 residues for *Pheretima sieboldi*, 142 residues for *Lumbricus*, 141 residues for *Tubifex* and 139 residues for *Tylorrhynchus*. There are 54 invariant amino acid residues among the four monomer globins of oligochaetes except for the marine polychaete, *Tylorrhynchus*. These invariants include the relative positions of the two histidine

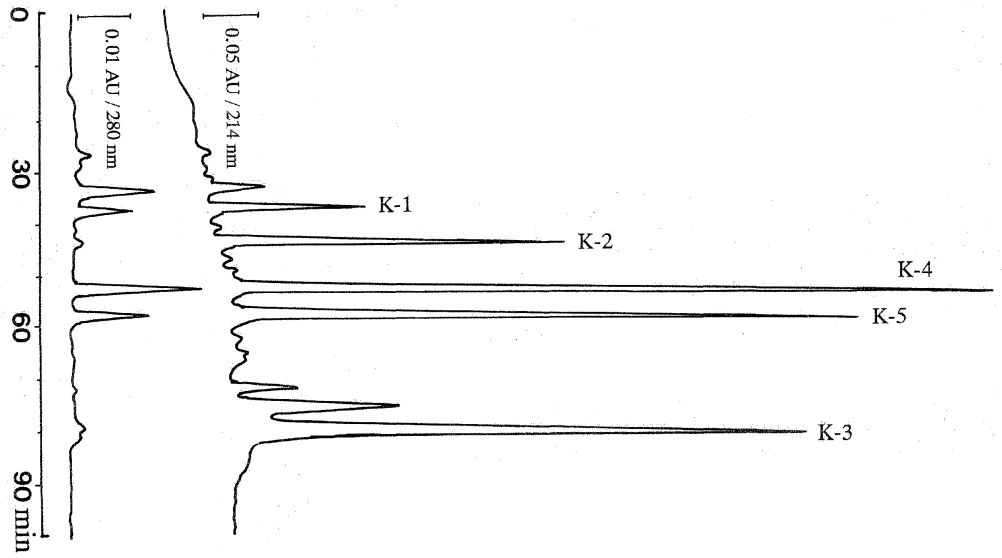

Fig. 4. Elution profile of the reversed-phase FPLC separation of the lysyl-endopeptidase digested peptides of the monomer subunit of *Pheretima hilgendorfi* hemoglobin. The column (Pep RPC 5/5, Pharmacia) was eluted with a linear gradient of 0-60% acetonitrile in 0.1% TFA at a flow rate of 0.3 ml/min.

Table 1. Amino acid composition of the monomer subunit of *Pheretima hilgendorfi* hemoglobin

Amino acid	Number of residues/mol	
	acid hydorlysates	amino acid sequence
Asp	11.4	10
Asn		2
Thr	7.5	8
Ser	6.5	6
Glu	16.3	11
Gln		5
Gly	9.7	9
Ala	10.6	10
Val	8.4	9
Pro	4.8	5
Met	0.0	0
Ile	5.7	6
Leu	14.9	15
Tyr	3.7	4
Phe	10.4	12
Lys	6.8	7
His	11.4	11
Trp	n.d.	2
Arg	6.7	6
Cys	1.7*	2
Total		140

Samples were hydrolyzed for 24-, 48-, and 72-hr at 105°C in 6N HCl-1% Phenol, dried and analyzed. Extrapolated or average values estimated from 24-, 48-, and 72-hr hydrolysates. n.d.: not determined. * Determined as cysteic acid after oxidation with performic acid.

Table 2. Amino acid compositions of the Lysyl-endopeptidase digested peptides, K1, K2, K3, K4, and K5, from the monomeric subunit of *Pheretima hilgendorfi* hemoglobin

Amino acid	K1	K2	K3	K4	K5
Asx	1.0(1)		5.3(5)	1.1(1)	4.1(4)
Thr		2.0(2)	2.9(3)	2.0(2)	1.1(1)
Ser		2.2(2)	2.6(3)		2.0(2)
Glx	2.2(2)	4.0(4)	6.1(6)	3.1(3)	1.3(1)
Pro			2.8(3)	1.0(1)	0.9(1)
Gly			4.3(4)	2.0(2)	3.1(3)
Ala		0.6(1)	3.7(4)	0.6(1)	4.1(4)
Cys	n.d.(1)				n.d (1)
Val	1.0(1)	1.0(1)	5.5(6)		
Met					
Ile			3.6(4)		1.8(2)
Leu	1.0(1)	1.2(1)	9.0(9)	1.0(1)	3.0(3)
Tyr			0.5(1)	0.7(1)	1.7(2)
Phe	1.0(1)	2.8(3)	2.8(3)	2.0(2)	2.8(3)
His		3.3(3)	4.4(4)	1.2(1)	3.0(3)
Lys	1.1(1)	1.1(1)	1.3(1)	2.0(2)	1.0(1)
Arg	1.2(1)	1.2(1)	2.1(2)	1.0(1)	1.1(1)
Trp		n.d.(2)			
Total	(9)	(21)	(58)	(18)	(32)
Position	1-9	12-32	33-90	91-108	109-140

Samples were hydrolyzed for 22 hr at 105°C in 6N HCl-0.1% phenol, dried and analyzed by a Shimadzu PC10 amino acid analyzer. Analyses are expressed as a ratio to Valine (K1 and K2), or Leucine (K3, K4, and K5). Values in parentheses are taken from amino acid sequence data. n.d.: not determined.

residues (residues 61 and 93) for heme coordination. The monomer globin of *Pheretima hilgendorfi* contains two cysteine residues linked at positions 2 and 130. These data, taken together with the comparison of partial amino-terminal sequences of the heme containing chains of trimer of *Pheretima hilgendorfi* (data not shown) suggest that the *Pheretima hilgendorfi* hemoglobin's constituents can be separated into two strains, strain A and B by Gotoh (Gotoh *et*

Table 3. Amino acid compositions of the trypsin digested fragments of K3 and K5 derived from the monomeric subunit of *Pheretima hilgendorfi* hemoglobin

	K3			K5	
Amino acid	K3-R1	K3-R2	K3-R3	K5-R1	K5-R2
Asx	1.8(1)	2.0(2)	2.2(2)	2.2(2)	2.1(2)
Thr	1.1(1)		2.0(2)	1.0(1)	
Ser	1.5(1)	1.2(1)	1.2(1)	1.2(1)	1.5(1)
Glx	2.2(2)	2.2(2)	2.1(2)	1.2(1)	0.5(0)
Pro	1.0(1)	0.9(1)	1.0(1)		1.0(1)
Gly	0.2(0)	1.8(2)	1.5(2)	2.2(2)	1.4(1)
Ala		1.5(2)	1.4(2)	3.2(3)	0.6(1)
Cys					n.d.(1)
Val	1.3(1)	1.0(1)	3.7(4)		
Met					
Ile		1.7(2)	1.8(2)		2.0(2)
Leu	3.0(3)		6.0(6)	3.0(3)	
Tyr		0.5(1)		0.8(1)	0.6(1)
Phe	1.6(2)	0.9(1)		2.0(2)	1.0(1)
His	1.9(2)	1.5(1)	1.5(1)	3.1(3)	
Lys	0.5(0)		1.1(1)		1.5(1)
Arg	1.0(1)	1.2(1)		1.2(1)	
Trp					
Total	(15)	(17)	(26)	(20)	(12)
Position	33-47	48-64	65-90	109-128	129-140

Samples were hydrolyzed for 22 hr at 105°C in 6N HCl-0.1% phenol, dried and analyzed by a Shimadzu PC10 amino acid analyzer. Analyses are expressed as a ratio to Leucine (K3-R1,K3-R3, and K5-R1), Valine (K3-R2), or Isoleucine (K5-R2). Values in parentheses are taken from the amino acid sequence data. n.d.: not determined.

al., 1987) as previously set up in the study of tetrameric *Lumbricus terrestris* hemoglobin, in which the earthworm's hemoglobin is, in general, akin to the divergence of the α- and β-globins of vertebrate hemoglobins. The monomer globin sequence of *Pheretima hilgendorfi*, therefore, belongs to the A-strain that already contains the sequences of chains I and IIA of *Tylorrhynchus* (Suzuki *et al.*, 1982; Suzuki and Gotoh, 1986), chains I and *b* of *Lumbricus* (Shishikura *et al.*, 1987; Fushitani *et al.*, 1988) and chain I of *Tubifex* (Stern *et al.*, 1990).

One of the most significant characteristics of this comparison of the alignment of the four globins of oligochaetes is the presence of a missing residue at position 20 in the amino-terminal region of *Pheretima hilgendorfi* (Fig. 5). This missing residue can not be found in the other homologous parts of known monomer sequences of oligochaetes or polychaetes. The secondary structure of monomer globin of *Pheretima hilgendorfi* shows that the missing residue corresponds to a part of the connection between A and B helices (Fig. 5). However, it is still an unresolved question why the deletion of amino acids can occur in the sequence of monomer globin of *Pheretima hilgendorfi* and not in that of other unrelated species of annelids.

The 82% amino acid sequence identity between the two *Pheretima* species indicates high similarity of each lineage. According to the hypothesis of two strains of globins (Gotoh *et al.*, 1987), the evolutionary divergence pattern of species depends on finding orthologous subunits in many homologues. In the cases of the two species of *Pheretima*, each of the monomer globins is sharing orthologous relationship and shown to have a closer divergence pattern than the globins belong to other strain. This is true in recent studies on sea cucumber hemoglobins (McDonald *et al.*, 1992) as well as on arthropod hemocyanins (Sugita and Shishikura, 1995). Especially, the sea cucumber C globin from *Caudina* (*Molpadia*) *arenicola* (McDonald *et al.*, 1992) shares a 60% sequence identity with the sequenced D globin from the same species (Mauri *et al.*, 1991), however, has a greater sequence identity of 93.6% with a globin from a different species of sea cucumber, *Paracaudina chilensis* (Suzuki, 1989b). Similar type of relationships are seen in hemocyanin molecules of horseshoe crabs: Although only a 57% sequence identity was found in immunological unrelated hemocyanin subunits, a hemocyanin subunit, HT6, from *Tachypleus tridentatus* shares a 97.3% identity with the sequenced hemocyanin subunit, HR6, from a different species *Carcinoscorpius rotundicauda*, since HT6 and HR6 have been shown to be immunological related hemocyanin subunits (Sugita and Shishikura, 1995). Hence, if attention is paid to the use of globin sequence data of orthologous relationship, we can deduce a reliable evolutionary divergence pattern of the *Pheretima* group. Therefore, a study of the molecular evolution of the genus *Pheretima* is in progress.

A more recent study on monomer globin of *Pheretima communissima* shows that only four variants (at the 1st, the 16th, the 21st and the 25th positions) and no deletion of amino acid residue were observed as compared with both first 40 amino-terminal amino acid sequences of orthologous globins of *Pheretima hilgendorfi* and *Pheretima communissima* (data not shown).

ACKNOWLEDGMENTS

The author is gratefully acknowledged to Dr. M. Esumi, First Department of Pathology, Nihon University School of Medicine, for her generosity in allowing the use of DNASIS and helping set of the equipments for analyzing amino acid compositions and amino acid sequences. Thanks are also to the staff of the Department of Biology, Nihon University School of Medicine for their cooperation.

REFERENCES

Bashford D, Chothia C, Lesk AM (1987) Determinants of a protein fold unique features of the globin amino acid sequences. J Mol Biol 196: 199–216

Boore JL, Brown WM (1995) Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris*. Genetics 141: 305–319

Chou PY, Fasman GD (1974) Prediction of protein conformation. Biochemistry 13: 222–245

Easton EG (1984) Earthworms (Oligochaeta) from islands of the south-western Pacific, and a note on two species from Papua New Guinea. New Zealand J Zool 11: 111–128

Ellman GL (1959) Tissue sulfhydryl groups. Arch Biochem Biophys 82: 70–77

Table 4. Summary of quantitative sequence analysis of the monomer subunit of *Pheretima hilgendorfi*. The intact globin and peptides correspond to those identified in Fig. 4 and Tables 1-3

No. of residues	Amino acid identified from intact globin and peptides (estimated recovery, pmol)			No. of residues	Amino acid identified from intact globin and peptides (estimated recovery, pmol)		
	Intact:	**K1:**					
1	E (314)	E (264)		71	S (20)		
2	Pe-C	Pe-C		72	V (97)		
3	D (170)	D (146)		73	I (88)		
4	V (325)	V (277)		74	G (63)		
5	L (263)	L (275)		75	V (74)		
6	E (235)	E (191)	**E-1:**	76	L (72)		
7	R (108)		R (132)	77	D (45)		
8	F (321)		F (892)	78	R (42)		
9	K (262)		K (602)	79	I (53)		
10	V (233)		V (687)	80	P (36)		
11	K (202)	**K2:**	K (586)	81	T (17)		
12	H (66)	H (116)	H (177)	82	L (34)		**P-2:**
13	Q (149)	Q (189)	Q (442)	83	T (13)		T (118)
14	W (125)	W (113)	W (368)	84	V (26)		V (154)
15	Q (95)	Q (150)	Q (370)	85	Q (26)		Q (117)
16	T (37)	T (72)	T (176)	86	L (26)		L (159)
17	V (130)	V (134)	V (310)	87	A (24)		A (138)
18	F (132)	F (149)	F (334)	88	H (10)		H (80)
19	S (16)	S (26)	S (42)	89	L (22)		L (154)
20	E (63)	E (92)	E (51)	90	K (8)	**K4:**	K (83)
21	A (123)	A (97)		91		A (488)	A (124)
22	H (39)	H (40)		92		Q (360)	Q (112)
23	R (51)	R (41)		93		H (105)	H (75)
24	T (22)	T (32)		94		T (105)	T (53)
25	E (70)	E (55)	**E-2:**	95		E (267)	E (154)
26		F (72)	F (736)	96		R (96)	R (35)
27		S (9)	S (94)	97		G (233)	G (84)
28		L (39)	L (459)	98		T (69)	T (25)
29		H (16)	H (125)	99		K (235)	K (32)
30	**P-1:**	F (39)	F (413)	100		P (200)	P (39)
31	W (3167)	W (7)	W (212)	101	**E3:**	E (152)	E (39)
32	K (2852)	K (14)	K (254)	102	Y (1032)	Y (185)	Y (38)
33	E (2542)	**K3:** E (256)	E (55)	103	F (1184)	F (204)	F (61)
34	F (3139)	F (283)		104	D (591)	D (94)	
35	L (2805)	L (233)		105	L (863)	L (140)	
36	H (927)	H (72)		106	F (978)	F (153)	
37	D (1544)	D (95)		107	G (680)	G (100)	
38	H (935)	H (51)		108	K (722)	K (68)	**K5**
39	P (1538)	P (120)		109	H (305)		H (167)
40	S (204)	S (19)		110	L (700)		L (461)
41	L (652)	L (99)		111	A (707)		A (432)
42		V (97)		112	S (101)		S (59)
43		E (66)		113	H (180)		H (84)
44		L (95)		114	L (391)		L (218)
45		F (80)		115	G (323)		G (166)
46		T (36)		116	D (270)		D (113)
47		R (30)	**K3-R2:**	117	E (260)		E (163)
48		V (66)	V (175)	118	L (280)		L (165)
49		N (46)	N (96)	119	G (238)		G (131)
50		G (28)	G (90)	120	T (130)		T (58)
51		A (44)	A (97)	121	H (86)		H (51)
52		N (50)	N (96)	122	F (251)		F (141)
53		I (38)		123	D (171)		D (78)
54		Y (33)		124	Y (199)		Y (105)
55		S (7)		125	A (202)		A (113)
56		P (22)		126	A (261)		A (178)
57		E (22)		127	F (207)		F (126)
58		F (20)		128	R (97)	**K5-R2:**	R (47)
59		Q (23)		129	D (127)	D (48)	D (63)
60		A (26)		130	Pe-C	Pe-C	Pe-C
61		H (9)		131	Y (150)	Y (40)	Y (78)
62		G (18)		132	D (80)	D (43)	D (53)
63		I (18)		133	F (91)	F (47)	F (45)
64	**K3-R3:**	R (9)		134	I (58)	I (40)	I (29)
65	V (199)	V (18)		135	A (65)	A (35)	A (30)
66	L (172)	L (20)		136	S (8)	S (6)	S (4)
67	A (184)	A (17)		137	G (38)	G (20)	G (15)
68	G (135)			138	I (31)	I (26)	I (11)
69	L (139)			139	K (16)	K (24)	K (1)
70	D (94)			140	P (5)	P (7)	P (0.5)

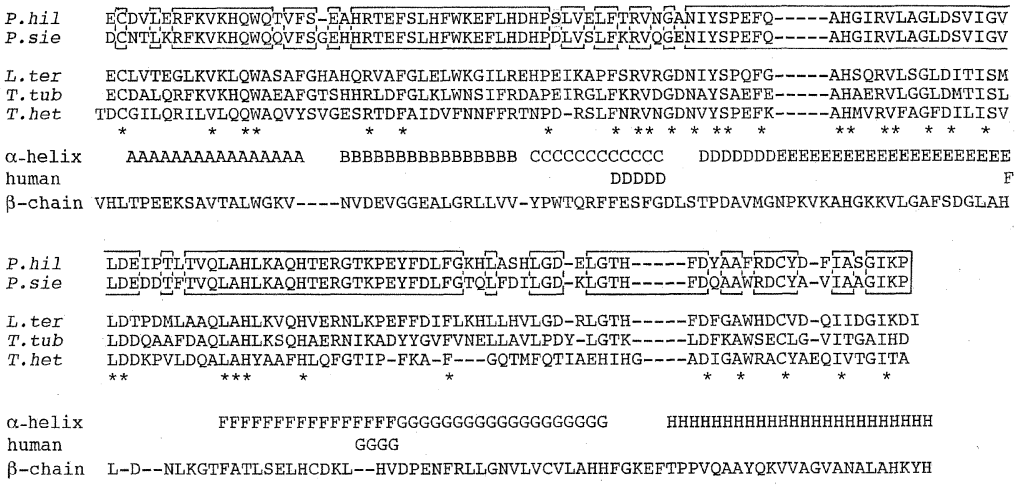

Fig. 5. Alignment of the amino acid sequences of five monomer globins from *Pheretima hilgendorfi* (*P. hil*), *Pheretima sieboldi* (*P. sie*) (Suzuki, 1989a), *Lumbricus terrestris* (*L. ter*) (Shishikura *et al.*, 1987), *Tubifex tubifex* (*T. tub*) (Stern *et al.*, 1990) and *Tylorrhynchus heterochaetus* (*T. het*) (Suzuki *et al.*, 1982). The alignment of *P. hil* and *P. sie* was achieved by a computer program (amino acid maximum homology by Needleman and Wunsch, 1970) of DNASIS. The previous data on alignment of annelid monomer globins (Stern *et al.*, 1990) were also used. Assignment of helical segments (A, B, ..., and H) are taken from combination with computed prediction of secondary structure from Chou and Fasman (1974) and the recent crystallographic analysis of human β-chain (Bashford *et al.*, 1987). The 34 residues conserved in the five globins of annelids are shown by asterisks. Boxes indicate the identical residues between the two species of *Pheretima*.

Friedman M, Krull LH, Cavins JF (1970) The chromatographic determination of cystine and cysteine residues in proteins as *S*-β-(4-pyridylethyl)cysteine. J Biol Chem 245: 3868–3871

Fushitani K, Matsuura MSA, Riggs AF (1988) The amino acids sequences of chains *a*, *b*, and *c* that form the trimer subunit of the extracellular hemoglobin from *Lumbricus terrestris*. J Biol Chem 263: 6502–6517

Goodman M, Pedwaydon J, Czelusniak J, Suzuki T, Gotoh T, Moens L, Shishikura F, Walz D, Vinogradov S (1988) An evolutionary tree for invertebrate globin sequences. J Mol Evol 27: 236–249

Gotoh T, Shishikura F, Snow JW, Ereifej KI, Vinogradov SN, Walz DA (1987) Two globin strains in the giant annelid extracelluar haemoglobins. Biochem J 241: 441–445

Granath WOJr (1988) Identification of hemeproteins in the hemolymph of *Biomphalaria glabrata* (Gastropda) separated sodium dodecyl sulfate-polyacrylamide gel electrophoresis. J Invertebr Pathol 52: 43–48

Jekel PA, Weijer WJ, Beintema JJ (1983) Use of endoproteinase Lys-C from *Lysobacter enzymogenes* in protein sequence analysis. Anal Biochem 134: 347–354

Kapp OH, Moens L, Vanfleteren J, Trotman CNA, Suzuki T, Vinogradov SN (1995) Alignment of 700 globin sequences: Extent of amino acid substitution and its correlation with variation in volume. Protein Sci 4: 2179–2190

Kobayashi S (1941) An outlook of distribution of terrestrial oligochaetes in the western regions of Japan (in Japanese). Doubutsugaku Zasshi 53: 371–384

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685

Mauri F, Omnaas J, Davidson L, Whitfill C, Kitto B (1991) Amino acid sequence of a globin from the sea cucumber *Caudina* (*Molpadia*) *arenicola*. Biochim Biophys Acta 1078: 63–67

McDonald GD, Davidson L, Kitto B (1992) Amino acid sequence of the coelomic C globin from the sea cucumber *Caudina* (*Molpadia*) *arenicola*. J Protein Chem 11: 29–37

Moss B, Ingram VM (1968) Hemoglobin synthesis during amphibian metamorphosis. I. Chemical studies on the hemoglobins from the larval and adult stages of *Rana catesbeiana*. J Mol Biol 32: 481–492

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48: 443–453

Ochiai T (1979) The giant hemoglobin molecules and their constituents in some species from the genus *Pheretima*. Sci Rep Niigata Univ Ser D (Biol) 16: 19–27

Ochiai T, Enoki Y (1979) Earthworm (*Pheretima communissima* and *Pheretima hilgendorfi*) hemoglobins: Giant assemblies and subunits. Biochim Biophys Acta 579: 442–451

Shishikura F, Snow JW, Gotoh T, Vinogradov SN, Walz DA (1987) Amino acid sequence of the monomer subunit of the extracellular hemoglobin of *Lumbricus terrestris*. J Biol Chem 262: 3123–3131

Stern MS, Vinogradov SN, Sharma PK, Ereifej K, Walz DA (1990) Amino acid sequence of the monomer subunit of the giant extracellular hemoglobin of the aquatic oligochaete, *Tubifex tubifex*. Eur J Biochem 194: 67–73

Sugita H, Shishikura F (1995) A case of orthologous sequences of hemocyanin subunits for an evolutionary study of horseshoe crabs: Amino acid sequence comparison of immunologically identical subunits of *Carcinoscorpius rotundicauda* and *Tachypleus tridentatus*. Zool Sci 12: 661–667

Suzuki T (1989a) Amino acid sequence of the monomer subunit of the giant multisubunit hemolgobin from the earthworm *Pheretima sieboldi*. Eur J Biochem 185: 127–134

Suzuki T (1989b) Amino acid sequence of a major globin from the sea cucumber *Paracaudina chilensis*. Biochim Biophys Acta 998: 292–296

Suzuki T, Gotoh T (1986) The complete amino acid sequence of giant multisubunit hemoglobin from the polychaete *Tylorrhynchus heterochaetus*. J Biol Chem 261: 9257–9267

Suzuki T, Takagi T, Gotoh T (1982) Amino acid sequence of the smallest polypeptide chain containing heme of extracellular hemoglobin from the polychaete *Tylorrhynchus heterochaetus*. Biochim Biophys Acta 708: 253–258

Teale FWJ (1959) Cleavage of the haem-protein link by acid methylethylketone. Biochim Biophys Acta 35: 543

Vinogradov SN (1985) The structure of invertebrate extracellular hemoglobins (erythrocruorins and chlorocruorins). Comp Biochem

Physiol 82B: 1–15
Vinogradov SN, Walz DA, Pohajdak B, Moens L, Kapp OH, Suzuki T, Trotman CNA (1993) Adventitious variability? The amino acid sequences of nonvertebrate globins. Comp Biochem Physiol 106B: 1–26
Waxdal MJ, Konigsberg WH, Henley WL, Edleman GM (1968) The covalent structure of a human γG-immunoglobulin. II. Isolation and characterization of the cyanogen bromide fragments. Biochemistry 7: 1959–1966

(Received March 11, 1996 / Accepted April 19, 1996)