Biochem. J. (1990) **266**, 221–225 (Printed in Great Britain)

Primary structure of a constituent polypeptide chain (AIII) of the giant haemoglobin from the deep-sea tube worm *Lamellibrachia*

A possible H_2S-binding site

Tomohiko SUZUKI,* Takashi TAKAGI† and Suguru OHTA‡

* Department of Biology, Faculty of Science, Kochi University, Kochi 780, Japan, † Biological Institute, Faculty of Science, Tohoku University, Sendai 980, Japan, and ‡ Ocean Research Institute, University of Tokyo, Tokyo 164, Japan

The deep-sea tube worm *Lamellibrachia*, belonging to the Phylum Vestimentifera, contains two giant extracellular haemoglobins, a 3000 kDa haemoglobin and a 440 kDa haemoglobin. The former consists of four haem-containing chains (AI–AIV) and two linker chains (AV and AVI) for the assembly of the haem-containing chains [Suzuki, Takagi & Ohta (1988) Biochem. J. **255**, 541–545]. The tube-worm haemoglobins are believed to have a function of transporting sulphide (H_2S) to internal bacterial symbionts, as well as of facilitating O_2 transport [Arp & Childress (1983) Science **219**, 295–297]. We have determined the complete amino acid sequence of *Lamellibrachia* chain AIII by automated or manual Edman sequencing. The chain is composed of 144 amino acid residues, has three cysteine residues at positions 3, 74 and 133, and has a molecular mass of 16620 Da, including a haem group. The sequence showed significant homology (30–50 % identity) with those of haem-containing chains of annelid giant haemoglobins. Two of the three cysteine residues are located at the positions where an intrachain disulphide bridge is formed in all annelid chains, but the remaining one (Cys-74) was located at a unique position, compared with annelid chains. Since the chain AIII was shown to have a reactive thiol group in the intact 3000 kDa molecule by preliminary experiments, the cysteine residue at position 74 appears to be one of the most probable candidates for the sulphide-binding sites. A phylogenetic tree was constructed from nine chains of annelid giant haemoglobins and one chain of vestimentiferan tube-worm haemoglobin now determined. The tree clearly showed that *Lamellibrachia* chain AIII belongs to the family of strain A of annelid giant haemoglobins, and that the two classes of Annelida, polychaete and oligochaete, and the vestimentiferan tube worm diverged at almost the same time. H.p.l.c. patterns of peptides (Figs. 4–7), amino acid compositions of peptides (Table 2) and amino acid sequences of intact protein and peptides (Table 3) have been deposited as Supplementary Publication SUP 50154 (13 pages) at the British Library Document Supply Centre, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in Biochem. J. (1990) **265**, 5.

INTRODUCTION

The deep-sea giant tube worms *Riftia* and *Lamellibrachia*, belonging to the Phylum Vestimentifera (Jones, 1985), are found in hydrothermal vents or cold seeps at a depth of 600–2500 m (Corliss *et al.*, 1979; Kennicutt *et al.*, 1985). These invertebrate animals are sustained by mutual symbiosis with sulphide-oxidizing bacteria (Childress *et al.*, 1987), and their blood, containing abundant haemoglobin, is believed to have a function of transporting sulphide (H_2S) to internal bacterial symbionts (Arp & Childress, 1983), as well as of facilitating O_2 transport compatible with high O_2 demand (Arp & Childress, 1981). However, the sulphide-binding sites are not known as yet, because of the lack of data for primary and tertiary structure.

The tube worm *Lamellibrachia*, living in the cold-seep area of the upper bathyal depth of Sagami Bay, Japan, has two giant extracellular haemoglobins, a 3000 kDa haemoglobin and a 440 kDa haemoglobin. Previously we have isolated most of the chains of *Lamellibrachia* haemoglobins by reverse-phase chromatography, and determined the *N*-terminal sequences and subunit structure (Suzuki *et al.* 1988*a*, 1989). In the present paper, we report the complete amino acid sequence of a constituent chain (AIII) of *Lamellibrachia* 3000 kDa haemoglobin. It is shown that the sequence shows significant homology with those of annelid giant haemoglobins, and has three cysteine residues at positions 3, 74 and 133. The possibility that Cys-74 is a sulphide-binding site is also discussed.

MATERIALS AND METHODS

The tube worm *Lamellibrachia* sp. (undescribed) was collected from the cold-seep area located off Sagami Bay, Japan, at a depth of 1160 m by the Japanese submersible *SHINKAI 2000*. The constituent polypeptide chain AIII of *Lamellibrachia* 3000 kDa haemoglobin was prepared by reverse-phase chromatography (Suzuki *et al.*, 1988*a*).

Methods of carboxymethylation of cysteine residues, enzymic digestion, amino acid analysis and manual

* To whom correspondence should be addressed.

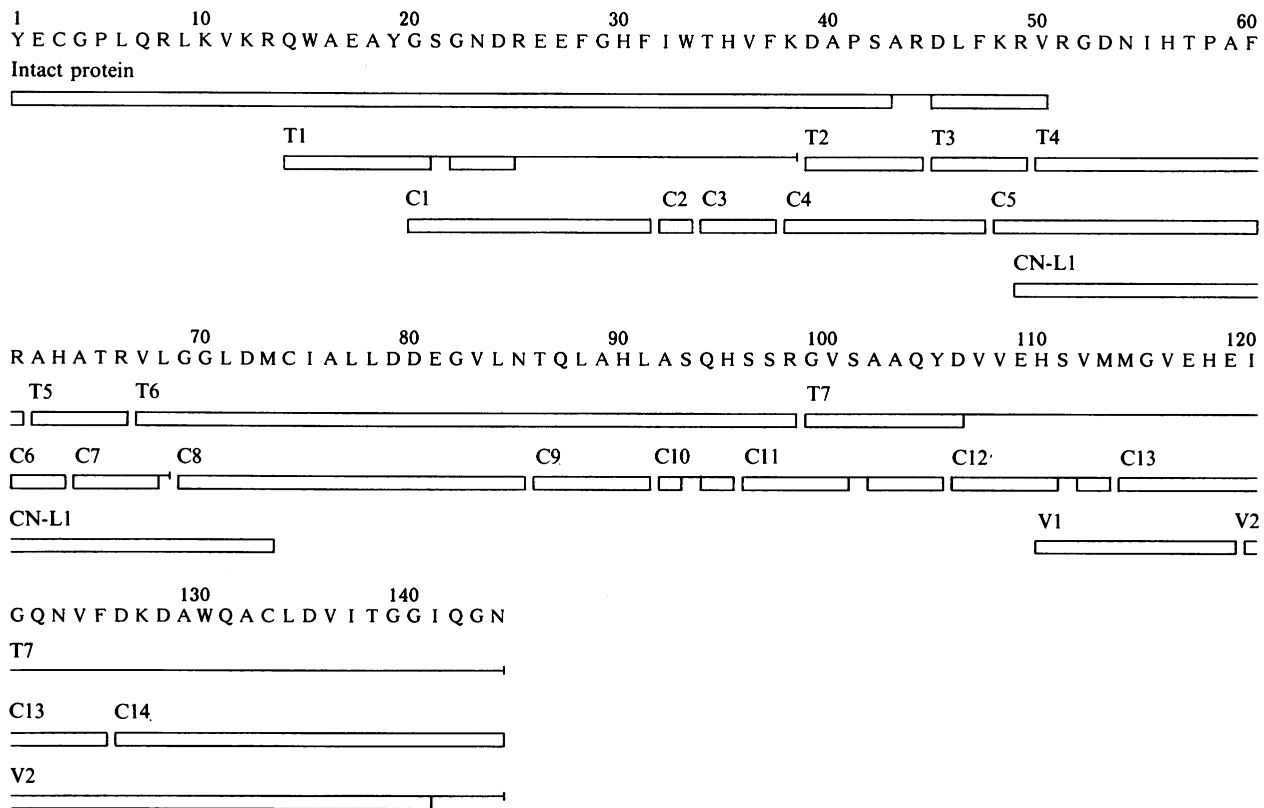

Fig. 1. Summary of data to establish the amino acid sequence of *Lamellibrachia* haemoglobin chain AIII

Automated sequencer or manual Edman degradation (▭) was employed for sequence determination. Key: T, a tryptic peptide; C, a chymotryptic peptide; V, an *S. aureus* V8 proteinase peptide; CN-L, a peptide cleaved with CNBr and lysyl endopeptidase.

Edman sequencing were the same as described previously (Suzuki & Gotoh, 1986*a*).

The AIII chain (20 nmol) was digested with trypsin in 0.1 M-ammonium bicarbonate at 37 °C for 2 h (Fig. 4 in Supplementary Publication SUP 50154). To obtain overlap peptides, the chain (18–20 nmol portions) was digested with chymotrypsin (in 0.1 M-ammonium bicarbonate at 37 °C for 3 h) (Fig. 5 in Supplementary Publication SUP 50154), *Staphylococcus aureus* V8 proteinase (in 0.1 M-ammonium bicarbonate at 37 °C for 22 h) (Fig. 6 in Supplementary Publication SUP 50154) and CNBr [in 70 % (v/v) formic acid at room temperature for 13 h]. The CNBr-cleavage products were digested further with lysyl endopeptidase in 37.5 mM-Tris/HCl buffer, pH 8.8, containing 3 M-urea at 30 °C for 5 h (Fig. 7 in Supplementary Publication SUP 50154).

The peptides were purified on a reverse-phase column [TSK Gel ODS-120T (4.6 mm × 150 mm) (Tosoh, Tokyo, Japan) or Cosmosil $5C_{18}$-300 (4.6 mm × 150 mm) (Nacalai Tesque, Kyoto, Japan)] with a linear gradient of 0–80 % (v/v) acetonitrile over 120 min at a flow rate of 1 ml/min. Amino acid compositions of the peptides are shown in Table 2 (in Supplementary Publication SUP 501154).

The *N*-terminal 50 residues of chain AIII were determined directly by use of an automated sequencer (Applied BioSystems 477A protein sequencer). The peptides obtained by digestion with enzymes or CNBr were sequenced by the automated sequencer (for the peptides T6, T7, C13, C14 and V2) or the manual Edman method (Table 3 in Supplementary Publication SUP 50154).

The peptides were numbered from the *N*-terminus. In this paper T indicates a tryptic peptide, C a chymotryptic peptide, V an *S. aureus* V8 proteinase peptide and CN-L a peptide cleaved with CNBr and lysyl endopeptidase.

Native *Lamellibrachia* 3000 kDa haemoglobin was labelled with 5-({2-[(iodoacetyl)amino]ethyl}amino)-naphthalene-1-sulphonic acid (IAEDANS) in 50 mM-Tris/HCl buffer, pH 7.2, at 4 °C for 3 h (Duke *et al.*, 1976; Mornet *et al.*, 1988). The reaction was stopped by addition of 2-mercaptoethanol, and the haemoglobin solution was immediately subjected to SDS/PAGE (Suzuki *et al.*, 1988*a*). The labelled chain was observed on the SDS/PAGE gel under u.v. light.

RESULTS AND DISCUSSION

Lamellibrachia 3000 kDa haemoglobin is composed of three types of subunits: two monomers (chains AIII and AIV), a disulphide-bonded dimer of chains AI and AII, and two linker subunits (chains AV and AVI) (Suzuki *et al.*, 1988*a*, 1989). Chains AI–AIV have a 'myoglobin-like' size and each contains a haem group, but chains AV and AVI with an unusual molecular size of 32–36 kDa are proposed to be linker chains for the assembly of haem-containing chains (Vinogradov *et al.*, 1986). Since most of the annelid 3000–4000 kDa haemoglobins consist of a monomer, a disulphide-bonded trimer and linker chains (Vinogradov, 1985), the subunit structure of *Lamellibrachia* haemoglobin is similar to that of an

Table 1. Matrix for sequence homologies (percentage identity) between the haem-containing chains of annelid and tube-worm haemoglobins

The percentage above the diagonal was obtained for the 136 residues common to all chains (see Fig. 2). The values below the diagonal show the Poisson corrected values by the equation: $-\ln(\%\ \text{identity}/100)$. Key: Tyl., *Tylorrhynchus*; Lum., *Lumbricus*; Phe., *Pheretima*; Lam., *Lamellibrachia*.

	Tyl.I	Tyl.IIA	Tyl.IIB	Tyl.IIC	Lum.I	Lum.II	Lum.III	Lum.IV	Phe.I	Lam.AIII
Tyl.I		43.4	26.5	30.1	34.6	36.0	28.7	31.6	42.6	39.0
Tyl.IIA	0.84		31.6	35.3	40.4	46.3	33.1	31.6	41.2	49.3
Tyl.IIB	1.33	1.15		50.0	35.3	36.0	34.6	47.8	31.6	29.4
Tyl.IIC	1.20	1.04	0.69		33.1	33.1	39.0	46.3	34.6	30.9
Lum.I	1.06	0.91	1.04	1.11		46.3	30.9	28.7	50.7	41.2
Lum.II	1.02	0.77	1.02	1.11	0.77		31.6	29.4	38.2	47.1
Lum.III	1.25	1.11	1.06	0.94	1.17	1.15		39.7	31.6	33.8
Lum.IV	1.15	1.15	0.74	0.77	1.25	1.22	0.92		30.1	30.9
Phe.I	0.85	0.89	1.15	1.06	0.68	0.96	1.15	1.20		39.7
Lam.AIII	0.94	0.71	1.22	1.17	0.89	0.75	1.08	1.17	0.92	

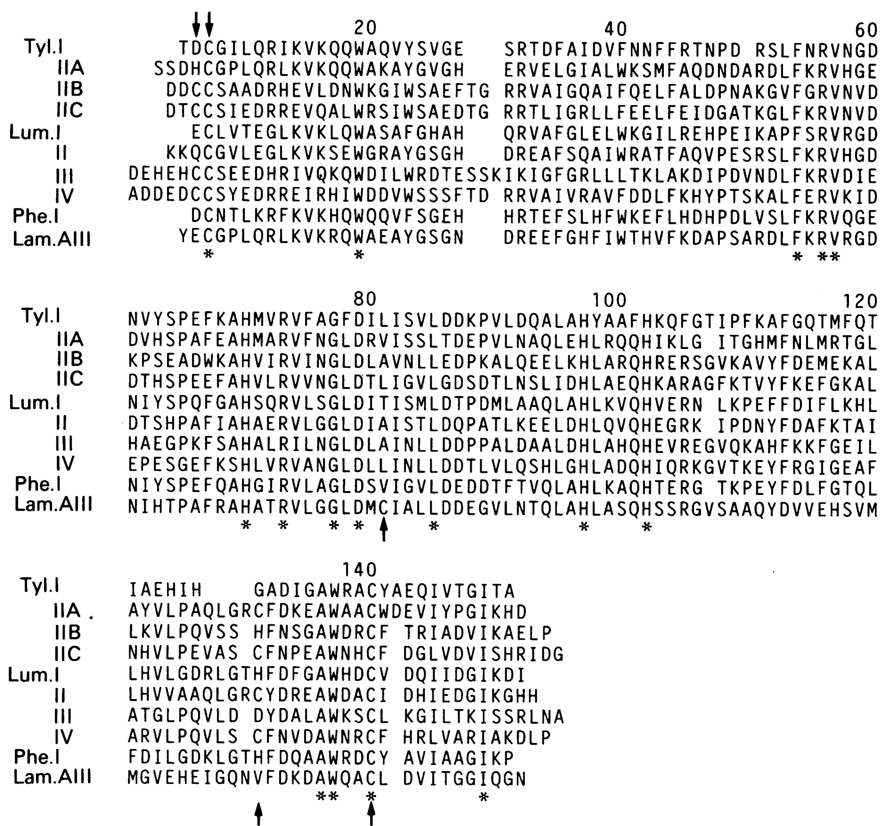

Fig. 2. Alignment of the amino acid sequence of *Lamellibrachia* haemoglobin chain AIII with those of annelid haem-containing chains

Several deletions are used to obtain significant correspondence between the sequences. Asterisks (*) indicate the 16 invariant residues in the ten globins. Positions containing cysteine residues are indicated by the arrows. Key: Tyl., *Tylorrhynchus*; Lum., *Lumbricus*; Phe., *Pheretima*; Lam., *Lamellibrachia*.

annelid giant haemoglobin, except for a difference in a disulphide-bonded dimer or trimer.

We have determined the complete amino acid sequence of the monomeric chain AIII of *Lamellibrachia* haemoglobin. The sequence was determined by automated or manual Edman degradation of intact protein and peptides derived by cleavage with trypsin, chymotrypsin, *S. aureus* V8 proteinase and CNBr–lysyl endopeptidase. The strategies used to establish the complete sequence are summarized in Fig. 1. The alignment of the peptides is supported by at least three overlapped residues. As shown in Fig. 1, *Lamellibrachia* chain AIII is composed of 144 amino acid residues, has three cysteine residues at positions 3, 74 and 133, and has a calculated molecular mass of 16620 Da, including a haem group. The hydropathy profile (Kyte & Doolittle, 1982) of *Lamelli-*

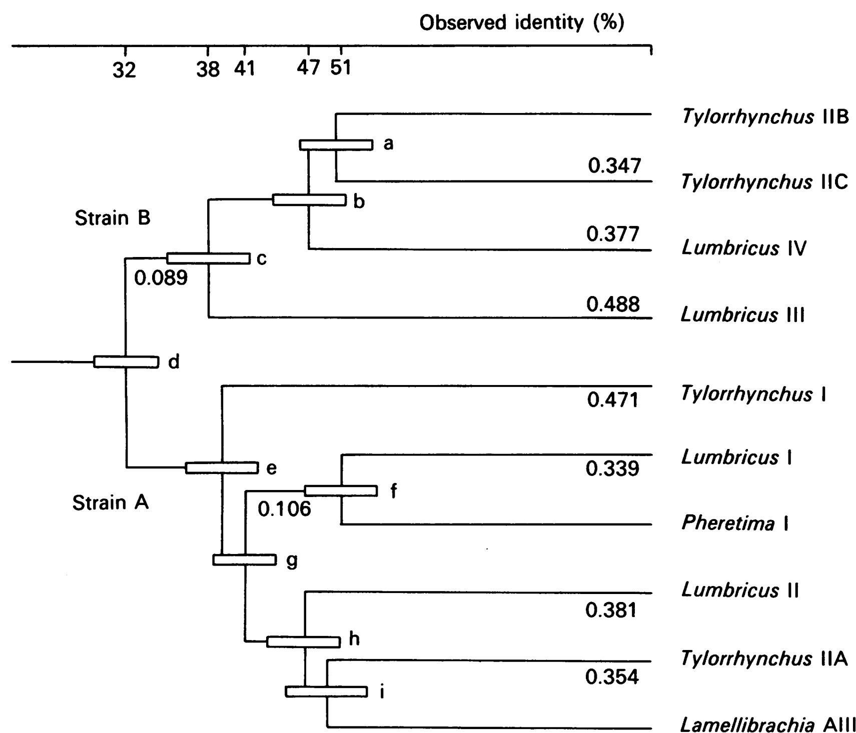

Fig. 3. Phylogenetic tree for the ten haem-containing chains from annelid and tube-worm giant haemoglobins

The tree was constructed from the Poisson corrected values in Table 1 by an unweighted pair-group clustering method. Standard errors at the branching points a–i are 0.043, 0.039, 0.045, 0.035, 0.040, 0.042, 0.035, 0.039 and 0.044 respectively.

brachia chain AIII resembled those of annelid and human haemoglobins, indicating that they have a similar globin folding.

The amino acid sequence of *Lamellibrachia* chain AIII is aligned with those of nine haem-containing chains of annelid giant haemoglobins from *Tylorrhynchus* (Suzuki *et al.*, 1982, 1985*a*,*b*; Suzuki & Gotoh, 1986*a*), *Lumbricus* (Shishikura *et al.*, 1987; Fushitani *et al.*, 1988) and *Pheretima* (Suzuki, 1989). A matrix for sequence homologies (percentage identity) between the chains is shown in Table 1. As shown in Fig. 2 and Table 1, the sequence of *Lamellibrachia* chain AIII shows significant homology (30–50% identity) with those of annelid giant haemoglobins. In the ten globin sequences, 16 residues {Cys-7, Trp-19, Phe-54, Arg-56, Val-57, His-70 [distal (E7)-His], Arg-73, Gly-77, Asp-79, Leu-85, His-97 [haem-binding proximal(F8)-His], His-103, Ala-136, Trp-137, Cys-140 and Ile-149} appear to be invariant.

In annelid giant haemoglobins, cysteine residues play a particularly important role in the subunit assembly of the giant molecule. In fact, the giant haemoglobin from the polychaete *Tylorrhynchus* dissociates completely into constituent chains in the presence of a reducing agent in the absence of any other protein denaturant (Suzuki *et al.*, 1983). The cysteine residues of annelid giant haemoglobins are restricted to the positions 6, 7, 131 and 140 in our alignment (Fig. 2), and they are all participating in either intra- or inter-chain disulphide bridges (Suzuki *et al.*, 1988*b*; Fushitani *et al.*, 1988). Thus there is no free cysteine residue in the haem-containing chains of annelid giant haemoglobins. According to the '192-chain' model (Suzuki & Gotoh, 1986*b*), *Tylorrhynchus* haemoglobin has 288 disulphide bonds in the molecule, that is 96 interchain bonds and 192 intrachain bonds. Here it should be noted that the intrachain disulphide bridge that is formed between Cys-7 and Cys-140 is conserved in all annelid chains.

On the other hand, *Lamellibrachia* chain AIII has three cysteine residues, at positions 7, 81 and 140 (Fig. 2). Judged from the structural homology, it is very likely that Cys-7 and Cys-140 in *Lamellibrachia* chain AIII form the intrachain disulphide bridge, as in annelid chains. If so, the remaining Cys-81 must be a free cysteine residue, since the AIII chain exists as a 'monomer' in the intact molecule (Suzuki *et al.*, 1989).

Arp and co-workers (Arp & Childress, 1983; Arp *et al.*, 1987) have shown that both 3000 kDa and 400 kDa haemoglobins of the deep-sea tube worm *Riftia*, a phylogenetically closely related species to *Lamellibrachia*, have (a) special sites(s) for binding sulphide, in order to supply it to internal bacterial symbionts. Therefore *Riftia* haemoglobins have two physiologically important functions, transport of O_2 and sulphide. Although the subunit structure, chain composition and amino acid sequence of *Riftia* haemoglobins are not available, data for amino acid composition, SDS/PAGE pattern and electron-microscopic appearance strongly suggest a structural similarity between *Riftia* and *Lamellibrachia* haemoglobins (Terwilliger *et al.*, 1980, 1985). In addition, a preliminary observation suggested that both of the *Lamellibrachia* haemoglobins bind sulphide (Suzuki *et al.*, 1989).

Arp *et al.* (1987) proposed that the sulphide is bound via thiol–disulphide exchange at the disulphide bridges in *Riftia* haemoglobins. However, this is unlikely, because *Lumbricus* giant haemoglobin has no ability to bind

sulphide (Arp *et al.*, 1987), and the tube-worm haemoglobins share many characteristics with annelid giant haemoglobins, including disulphide-bridge configuration (Terwilliger *et al.*, 1980, 1985; Suzuki *et al.*, 1988*a*,*b*; Fushitani *et al.*, 1988; the present work).

Alternatively, we propose that a free cysteine residue, which is not present in any haem-containing chains of annelid haemoglobins, is responsible for sulphide-binding ability. In fact, *Lamellibrachia* chain AIII has a free cysteine residue at a unique position (Cys-81 in Fig. 2). Our preliminary experiment with the use of a fluorescent labelling (Mornet *et al.*, 1988) suggested that all chains but AI have reactive thiol groups in intact *Lamellibrachia* 3000 kDa haemoglobin. Since the tube-worm 400–440 kDa haemoglobin, which is composed of only haem-containing chains (Suzuki *et al.*, 1988*a*) and contains no linker chains, also has the ability to bind sulphide (Arp *et al.*, 1987), the free cysteine residues in haem-containing chains appear to be among the most probable candidates for sulphide-binding sites. Such a cysteine residue in the deep-sea tube-worm haemoglobins may have been acquired after its divergence from polychaetes and oligochaetes, as a high-molecular adaption for symbiosis with sulphide-oxidizing bacteria.

Recently the deep-sea tube worms *Riftia* and *Lamellibrachia* were placed on a new Phylum, Vestimentifera, on the basis of their unique outward appearance, such as the very long trunk region and absence of a mouth, gut and anus (Jones, 1985). However, all biochemical data on tube-worm haemoglobins suggest that they are probably members of the Phylum Annelida (Terwilliger *et al.*, 1980, 1985; Suzuki *et al.*, 1988*a*, 1989). Now that we have obtained the complete amino acid sequence of a constituent chain of *Lamellibrachia* haemoglobin, we have constructed a phylogenetic tree for annelid and tube-worm haemoglobins, by an unweighted pair-group clustering method (Nei *et al.*, 1985). The result is shown in Fig. 3. Standard errors are given at each branching point, to help evaluation of the tree (Nei *et al.*, 1985).

Fig. 3 clearly shows that there are two globin strains A and B, as suggested by Gotoh *et al.* (1987), for annelid and tube-worm haemoglobin chains. *Lamellibrachia* chain AIII is most closely related to *Tylorrhynchus* (polychaete) chain IIA and *Lumbricus* (oligochaete) II, with about 50 % identical residues, and therefore these three chains may correspond to a 'homologous' chain. It may be noted that this degree of homology (50 % identity) is slightly higher than that (44 % identity) between human α- and β-globins, which were separated by gene duplication about 450 million years ago (Goodman *et al.*, 1975, 1988). As the phylogenetic tree constructed from amino acid sequences of a homologous globin shows a good correlation with that from classical taxonomy (Goodman *et al.*, 1975, 1988), it may be suggested from Fig. 3 that *Lamellibrachia*, *Tylorrhynchus* and *Lumbricus* diverged from their ancestors at almost the same time, if we consider the standard errors at branching points h and i. We propose from these results that the deep-sea giant tube worms *Lamellibrachia* and *Riftia* should be placed in the Phylum Annelida, in place of Vestimentifera, but a new Class may be needed for the tube worms with unique appearance.

This work was partly supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan (to T. S.). We thank Dr. T. Gotoh and Dr. T. Furukohri for their encouragement during this work.

REFERENCES

Arp, A. J. & Childress, J. J. (1981) Science **213**, 342–344
Arp, A. J. & Childress, J. J. (1983) Science **219**, 295–297
Arp, A. J., Childress, J. J. & Vetter, R. D. (1987) J. Exp. Biol. **128**, 139–158
Childress, J. J., Felbeck, H. & Somero, G. N. (1987) Sci. Am. **256**, 106–112
Corliss, J. B., Dymond, J., Gordon, L. I., Edmond, J. M., von Herzen, R. P., Ballard, R. D., Green, K., Williams, D., Bainbridge, A., Crane, K. & van Andel, T. H. (1979) Science **203**, 1073–1083
Duke, J., Takashi, R., Ue, K. & Morales, M. F. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 302–306
Fushitani, K., Matsuura, M. S. A. & Riggs, A. F. (1988) J. Biol. Chem. **263**, 6502–6517
Goodman, M., Moore, G. W. & Matsuda, G. (1975) Nature (London) **253**, 603–608
Goodman, M., Pedwaydon, J., Czelusniak, J., Suzuki, T., Gotoh, T., Moens, L., Shishikura, F., Walz, D. & Vinogradov, N. (1988) J. Mol. Evol. **27**, 236–249
Gotoh, T., Shishikura, F., Snow, J. W., Ereifej, K. & Vinogradov, S. N. (1987) Biochem. J. **241**, 441–445
Jones, M L. (1985) Biol. Soc. Wash. Bull. **6**, 117–158
Kennicutt, M. C., Brooks, J. M., Bidigare, R. R., Fay, R. R., Wade, T. L. & McDonald, T. J. (1985) Nature (London) **317**, 351–353
Kyte, J. & Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132
Mornet, D., Audemard, E. & Derancourt, J. (1988) Biochem. Biophys. Res. Commun. **154**, 564–571
Nei, M., Stephens, J. C. & Saitou, N. (1985) Mol. Biol. Evol. **2**, 66–85
Shishikura, F., Snow, J. W., Gotoh, T., Vinogradov, S. N. & Walz, D. A. (1987) J. Biol. Chem. **262**, 3123–3131
Suzuki, T. (1989) Eur. J. Biochem. **185**, 127–134
Suzuki, T. & Gotoh, T. (1986*a*) J. Biol. Chem. **261**, 9257–9267
Suzuki, T. & Gotoh, T. (1986*b*) J. Mol. Biol. **190**, 119–123
Suzuki, T., Takagi, T. & Gotoh, T. (1982) Biochim, Biophys. Acta **708**, 253–258
Suzuki, T., Takagi, T., Furukohri, T. & Gotoh, T. (1983) Comp. Biochem. Physiol. B **75**, 567–570
Suzuki, T., Furukohri, T. & Gotoh, T. (1985*a*) J. Biol. Chem. **260**, 3145–3154
Suzuki, T., Yasunaga, H., Furukohri, T., Nakamura, K. & Gotoh, T. (1985*b*) J. Biol. Chem. **260**, 11481–11487
Suzuki, T., Takagi, T. & Ohta, S. (1988*a*) Biochem. J. **255**, 541–545
Suzuki, T., Kapp, O. H. & Gotoh, T. (1988*b*) J. Biol. Chem. **263**, 18524–18529
Suzuki, T., Takagi, T., Okuda, K., Furukohri, T. & Ohta, S. (1989) Zool. Sci. **6**, 915–926
Terwilliger, R. C., Terwilliger, N. B. & Schabtach, E. (1980) Comp. Biochem. Physiol. B **65**, 531–535
Terwilliger, R. C., Terwilliger, N. B., Bonaventura, C., Bonaventura, J. & Schabtach, E. (1985) Biochim. Biophys. Acta **829**, 27–33
Vinogradov, S. N. (1985) Comp. Biochem. Physiol. B **82**, 1–15
Vinogradov, S. N., Lugo, S. D., Mainwaring, M. G., Kapp, O. H. & Crewe, A. V. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 8024–8038

Received 3 July 1989/24 August 1989; accepted 8 September 1989