Biochem. J. (1988) 255, 541–545 (Printed in Great Britain)

N-Terminal amino acid sequence of the deep-sea tube worm haemoglobin remarkably resembles that of annelid haemoglobin

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 giant tube worm *Lamellibrachia*, belonging to the phylum Vestimentifera, contains two extracellular haemoglobins, an M_r 3000000 haemoglobin and an M_r 440000 haemoglobin. The former has a hexagonal bilayer structure and consists of six polypeptide chains (AI–VI); a study of its haem content shows that not all of the chains contain haem. The M_r 440000 haemoglobin consists of four haem-containing chains (BI–IV). We isolated most of the chains by reverse-phase chromatography and determined the amino acid sequences of the 21–45 *N*-terminal residues. Eight chains (AI–IV and BI–IV) showed significant homology with haem-containing chains of annelid giant haemoglobin. The highest homology was found between *Lamellibrachia* chain AI and *Tylorrhynchus* chain I; surprisingly, 18 out of the 20 *N*-terminal residues are identical. On the other hand, chain AV, with an unusual M_r of 32000, showed a rather different sequence and is likely to be a non-haem chain which might act as a linker protein in the assembly of the haem-containing chains. From these results, we conclude that the tube worm M_r 3000000 haemoglobin is highly homologous with annelid haemoglobin.

INTRODUCTION

The giant tube worms *Riftia* and *Lamellibrachia* which are found in hydrothermal vents or cold seeps at a depth of 600–2500 m (Corliss *et al.*, 1979; Kennicutt *et al.*, 1985; Ohta & Laubier, 1987), present an interesting taxonomic question. Although they have been placed in the phylum Vestimentifera, there is some controversy about their relationship with the other phyla (Southward, 1975; Jones, 1981, 1985).

The tube worms contain abundant extracellular haemoglobin, which is compatible with their high oxygen demand (Arp & Childress, 1981). The haemoglobin also has a special ability to bind sulphide, which it transports to internal bacterial symbionts (Arp & Childress, 1983; Childress *et al.*, 1987). The electron microscopic appearance and electrophoretic pattern of *Riftia* and *Lamellibrachia* sp. haemoglobins were shown to be similar to those of annelid haemoglobin (Terwilliger *et al.*, 1980, 1985). Here we show that the amino acid sequences of all the constituent chains of *Lamellibrachia* haemoglobin are highly homologous with those of annelid giant haemoglobin.

MATERIALS AND METHODS

Lamellibrachia sp. (50–60 cm long) was collected from the cold-seep area located off Sagami Bay (34° 59.90′, 139° 13.60′) at depth of 1160 m, southeast of Hatsushima, Japan, by a Japanese submersible SHINKAI 2000 during November of 1987 (dive nos. 314 and 315) (Ohta, 1987). As soon as the animals were brought to the surface, blood was removed by cutting the worms and stored at −80 °C or −40 °C until used.

Thawed samples were centrifuged at 15000 rev./min at 2 °C to remove insoluble materials, and the resultant haemoglobin solution was applied to a column of Sepharose CL-4B equilibrated with 0.1 M-phosphate buffer (pH 7.2).

Electron microscopy was carried out on the whole blood according to the method of Terwilliger *et al.* (1980). Like other tube worm haemoglobins (Terwilliger *et al.*, 1980, 1985), *Lamellibrachia* haemoglobin showed a hexagonal bilayer structure (T. Suzuki, unpublished results).

SDS/polyacrylamide-gel electrophoresis (SDS/PAGE) was carried out in 15% acrylamide gel containing 0.087% bisacrylamide, 0.375 M-Tris/HCl (pH 8.8) and 0.1% SDS. The sample was incubated in 0.75% SDS at 100 °C for 5 min in the presence of 2-mercaptoethanol before electrophoresis.

The haem content of *Lamellibrachia* haemoglobin was estimated from haem analysis (pyridine-haemochrome method) and amino acid analysis of protein (Suzuki & Gotoh, 1986*b*).

The constituent polypeptide chains were separated by h.p.l.c. After removal of haem, the sample was reduced with 10 mM-dithiothreitol at 37 °C for 2 h in 6 M-guanidinuim chloride and 0.2 M-Tris/HCl (pH 8.5) and applied to the column. The column (Cosmosil $5C_{18}$-300, 4.6 mm × 150 mm, Nakarai) was equilibrated with 30% acetonitrile in 0.1% trifluoroacetic acid and eluted with a linear gradient of 30–60% acetonitrile in 0.1% trifluoracetic acid over 40 min at a flow rate of 1 ml/min.

The amino acid sequence of the chain was determined by an automated sequencer (Applied BioSystems 477A Protein Sequencer). Carboxymethylated chains (500–1500 pmol) (Suzuki & Gotoh, 1986*a*) were applied.

Abbreviation used: PAGE, polyacrylamide-gel electrophoresis.

* To whom all correspondence should be addressed.

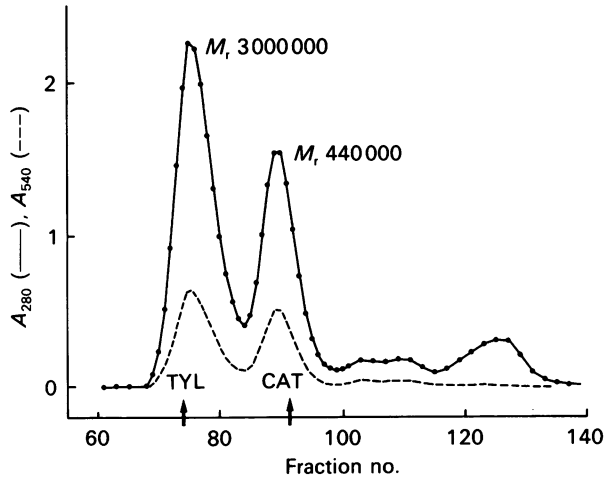

Fig. 1. Gel filtration of *Lamellibrachia* haemoglobin on a Sepharose CL-4B column

The column (3 cm × 117 cm) was equilibrated and eluted with 0.1 M-phosphate buffer (pH 7.2). The M_r markers are *Tylorrhynchus* haemoglobin (TYL, M_r 3000000) and catalase (CAT, 240000). The M_r were estimated to be about 3000000 and 440000, respectively. Fraction size, 4 ml/tube.

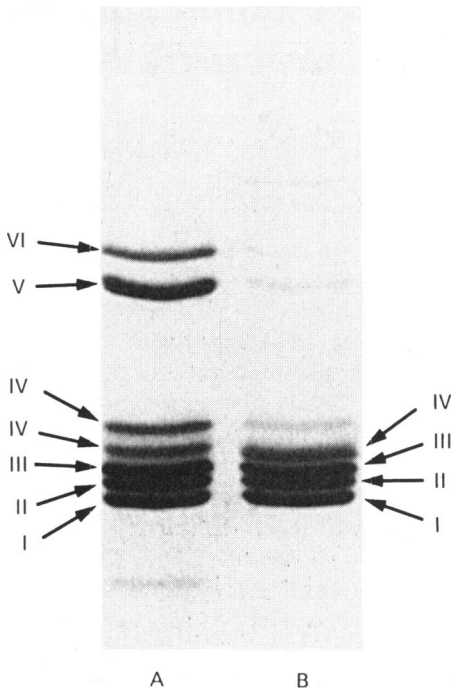

Fig. 2. SDS/PAGE of *Lamellibrachia* haemoglobins in the presence of 2-mercaptoethanol

Nomenclature of the chains was based on the M_r (Vinogradov *et al.*, 1977). Lane A, *Lamellibrachia* M_r 3000000 haemoglobin; lane B, M_r 440000 haemoglobin.

RESULTS AND DISCUSSION

Lamellibrachia haemoglobin produced two peaks, a 3000000-M_r haemoglobin and a 440000-M_r haemoglobin, on a gel filtration column (Fig. 1). The elution profile was the same as for *Riftia* haemoglobin (Arp, 1986). Since annelid extracellular haemoglobin consists of a single component of M_r (3–4) × 10^6, the M_r 440000 haemoglobin is unique to tube worm.

Fig. 2 shows the SDS/PAGE of *Lamellibrachia* haemoglobin under reducing conditions. The M_r 3000000 haemoglobin gave seven bands (I, II, III, IV′, IV, V and VI) corresponding to M_r of 16000, 17000, 17500, 19000, 20500, 32000 and 36500, respectively. It is noted that bands V and VI are about double the size of the other chains. On the other hand, the M_r 440000 haemoglobin gave four bands (I, II, III and IV) corresponding to M_r of 16000, 17000, 17500 and 18500, respectively.

The M_r 440000 haemoglobin contained, like vertebrate haemoglobins and myoglobins, one haem per 17600 g of protein, suggesting that all four chains contain haem. On the other hand, the M_r 3000000 haemoglobin contained one haem per 21500 g of protein. This value is similar to that of annelid giant haemoglobin and suggests that not all of the chains contain haem.

We succeeded in isolating most of the constituent polypeptide chains of *Lamellibrachia* haemoglobin by reverse-phase h.p.l.c. Fig. 3(*a*) shows an h.p.l.c. pattern for the M_r 3000000 haemoglobin. The eight chains eluted were named AV-1, AV-2, AIII, AII, AI-1, AI-2, AII, AIV and AIV′ on the basis of their electrophoretic mobility. The chain corresponding to band VI in Fig. 2 was not recovered. Heterogeneity was found between chains AI and AV. Fig. 3(*b*) shows an h.p.l.c. pattern for the M_r 440000 haemoglobin. Four major chains (BI–IV) were eluted and corresponded exactly to the four bands on SDS/PAGE.

The 21–45 *N*-terminal residues of *Lamellibrachia* carboxymethylated chains were determined by an automated sequencer. Chains AI-1 and AI-2 had the same sequence up to residue 41, while only one replacement was detected between chains AV-1 and AV-2 (Ser at position 7 for AV-1, and Ala for AV-2). We judged that chain AIV′ is a hetero-type of chain AIV, because their sequences were very similar (83% identity), although chain AIV′ lacked two residues (Ser-Gly) at the *N*-terminus. Chains AII and BII, and also chains AIII and BIII may be identical, since the 24 *N*-terminal residues and M_r on SDS/PAGE for both chains are identical. We prepared band VI in Fig. 2, which could not be recovered by h.p.l.c., by extraction from SDS/PAGE and applied it to the sequencer. However, no reliable results were obtained.

Fig. 4 summarizes the sequencing results, and Table 1 shows the homologies (% identity) between the sequences. The sequence of the smallest chain (chain I) of annelid giant haemoglobin from the polychaete *Tylorrhynchus heterochaetus* is also shown for comparison. Two invariable residues, Cys at position 13 and Trp at 25, are conserved not only in eight *Lamellibrachia* chains (AI–IV and BI–IV) of myoglobin-like size, but also in all the haem-containing chains of annelid giant haemoglobins (Suzuki *et al.*, 1982, 1985*a*,*b*; Suzuki & Gotoh, 1986*a*; Garlick & Riggs, 1982; Shishikura *et al.*, 1987). In *Tylorrhynchus* haemoglobin, Cys-13 participates in the formation of an intra-chain disulphide bridge. Eight *Lamellibrachia* chains showed homology with each other (20–60% identity) and with haem-containing chains of annelid haemoglobin. The highest homology (about 60% identity) was found between *Lamellibrachia* chains AI and AIII (or BIII), between chains BI and BIII (or

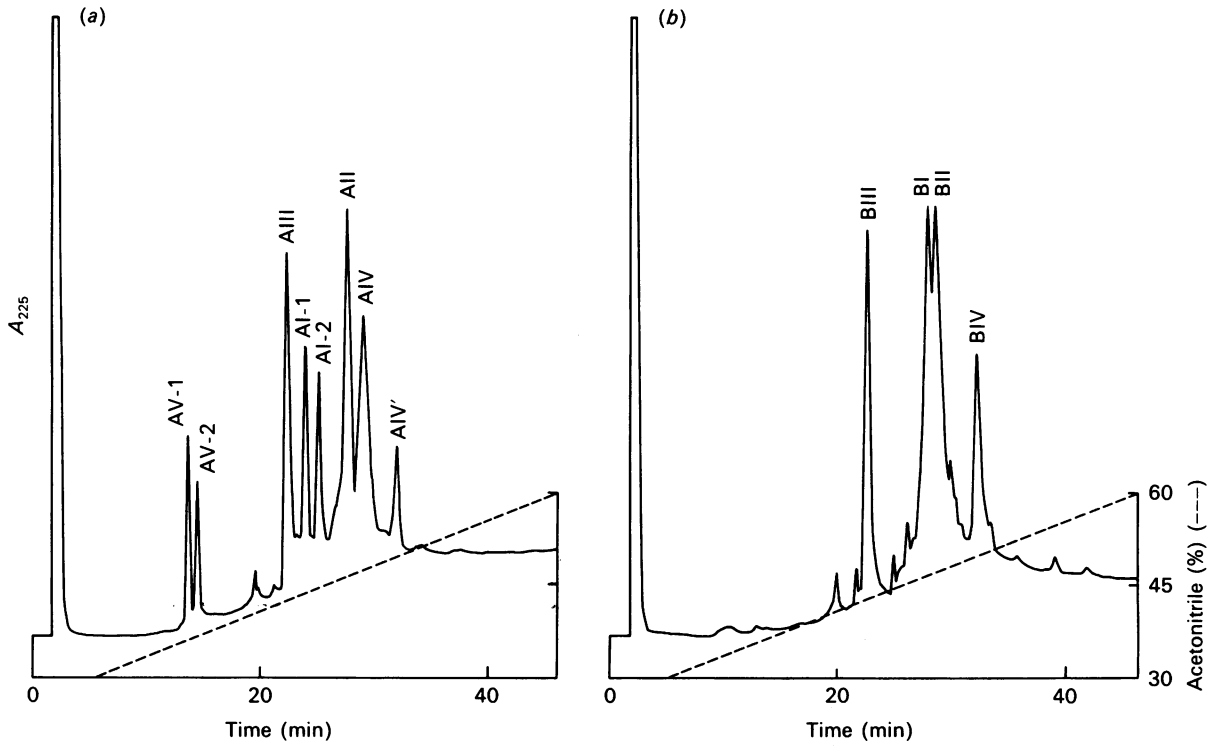

Fig. 3. Separation of the constituent polypeptide chains of *Lamellibrachia* haemoglobin by reverse-phase h.p.l.c.

Nomenclature of the chains was based on the electrophoretic mobility in Fig. 2. (*a*), *Lamellibrachia* M_r 3000000 haemoglobin; (*b*), M_r 440000 haemoglobin.

```
                                    10                  20                  30                  40                  50
Annelid (Mr 3300000)
  Tyl.I                              T D C G I L Q R I K V K Q Q W A Q V Y - - S V G E S R T D F A I D V F N N F F R T N P
Tube worm (Mr 3000000)               * * * *  * * * * * * * * * * *  * *   *  *   *  * *                 *     *
  AI                                 T D C G M L Q R I K V K Q Q W A S V Y - - S S G I A R E D F G E A I W K A V F A L A P
  AII                          S S N S C T T E D R R E M Q L M W A N V W S A Q
  AIII                               Y E C G P L Q R L K V K R Q W A E A Y - - G S G N D
  AIV                  S G N V A E A P K H Y H C S Y E D A D I V M R E W Y H V W - - G S G
  AIV'                   S V A N A P K H N H C S Y E D A E I V M S E W Y H V W - - G S G
  AV                                 A A V Q P L S V S D A M G A R V D A Q - - A W R V D R L T K Q F Q A I S D A A D T S I
Tube worm (Mr 440000)
  BI                                   D C N I L Q R L K V K M Q W A K A Y - - G F G A E R A K F G N S L W T S I F N Y A
  BII                          S S N S C T T E D R R E M Q L M W A N V W S A Q F T G R R X A I A Q A V F K D L F A N V P
  BIII                               Y E C G P L Q R L K V K R Q W A E A Y - - G S G N D R E E F G E F I W T T V F K D A P
  BIV                          S K F C S E G D A R I V I K Q W N Q I Y - - N
```

Fig. 4. Alignment of *N*-terminal amino acid sequences of *Lamellibrachia* chains with that of *Tylorrhynchus* chain I (Suzuki *et al.*, 1982)

The boxed residues indicate two invariable residues in eight haem-containing chains of *Lamellibrachia* haemoglobin and *Tylorrhynchus* chain I. The residues indicated by asterisk are identical between *Tylorrhynchus* chain I and *Lamellibrachia* chain AI. Heterogeneity was found at position 17 (S and A) in chain AV. X, unidentified.

AIII) and between chain AI and *Tylorrhynchus* chain I. These results imply that *Lamellibrachia* chains AI–IV and BI–IV are haem-containing chains and are consistent with the haem content of the M_r 440000 haemoglobin. Compared with the other chains, chain AIV has an *N*-terminal extension of 8–11 residues, but such an extension is also found in several invertebrate haemoglobins (Gilbert & Thompson, 1985). On the other hand, *Lamellibrachia* chain AV with an unusual M_r of 32000 showed a rather different sequence with only 5–10% of the residues identical. Chain AV also lacked the two residues Cys-13 and Trp-25, which are conserved in all haem-containing chains. Therefore, chain AV is likely to be a non-haem chain. As yet, sequence data for the unusual non-haem chains present in annelid giant haemoglobin have not been reported.

We noticed that the homology between *Lamellibrachia* chain AI (the smallest chain) and the polychaete *Tylorrhynchus* chain I is extremely high for the *N*-terminal 20 residues; 18 (90%) out of 20 residues are identical. This

Table 1. Sequence homologies (% identity) between the partial sequences of *Lamellibrachia* chains and *Tylorrhynchus* chain I

Since we judged that *Lamellibrachia* chains AII and BII, and also chains AIII and BIII are identical (see text), chains BII and BIII sequenced up to about 40 residues were used for construction of the matrix.

	*Tyl.*I	*Lam.*AI	AII	AIII	AIV	AIV′	AV	*Lam.*BI	BII	BIII
*Lam.*AI	59									
AII	28	28								
AIII	42	61	18							
AIV	23	27	25	27						
AIV′	23	27	29	27	83					
AV	7	5	5	12	5	5				
*Lam.*BI	44	49	18	56	24	24	8			
BII	28	28	100	18	25	29	5	18		
BIII	42	61	18	100	27	27	12	56	18	
BIV	30	25	24	25	33	33	0	26	24	25

homology is just comparable with that (85%) between the smallest chains of the polychaetes, *Tylorrhynchus* chain I and *Arenicola* chain I (Sgourous *et al.*, 1986), for the *N*-terminal 20 residues, but is rather higher than that (35%) between the smallest chains of the polychaete and oligochaete, *Tylorrhynchus* chain I and *Lumbricus* chain I (Shishikura *et al.*, 1987). Therefore, it may be concluded that the tube worm *Lamellibrachia* is closely related to the polychaetes.

Lamellibrachia M_r 3000000 haemoglobin shares many characteristics with annelid giant haemoglobin. Both haemoglobins are mainly composed of four haem-containing chains of 'myoglobin-like' size. In *Tylorrhynchus* haemoglobin, the four chains are in equimolar proportions and a tetramer is supposed to be their minimum structural entity (Suzuki & Gotoh, 1986*b*). Vinogradov and his co-workers (Vinogradov *et al.*, 1977, 1986; Mainwaring *et al.*, 1986) postulated that the two additional chains with unusual M_r (32000–36000), at least one of which may correspond to a non-haem chain, act as 'linker proteins' in the assembly of the tetrameric subunit. Although there is no sequence data for the non-haem chains of annelid haemoglobin, *Lamellibrachia* chain AV, which we isolated and sequenced partially, must correspond to the 'linker protein'.

The M_r 440000 haemoglobin, which is not a dissociated product of the M_r 3000000 haemoglobin, appears to be a prototype of the 3000000 haemoglobin, since it also consists of four haem-containing chains and lacks only the 'linker protein'.

Many taxonomists agree that the worm-like animals belonging to the phyla Vestimentifera, Pogonophora and Annelida are closely related, although there is some controversy about detailed points (Southward, 1975; Terwilliger *et al.*, 1980). In this respect, our sequence studies with *Lamellibrachia* haemoglobin have an important suggestion to make regarding the taxonomical position of the vestimentiferan tube worms. It is well known that the phylogenetic tree constructed from globin sequences shows a good correlation with that from classical taxonomy (Goodman *et al.*, 1975). So far, no sequence data are available for the haemoglobin from Pogonophora, but sequences for several annelid haemoglobins have been reported (Suzuki & Gotoh, 1986; Shishikura *et al.*, 1987). In conclusion, we propose that the tube worms should be placed in the phylum Annelida as a member of the polychaetes, because their haemoglobin sequence had extremely strong homology with those of the polychaetes. The tube worms probably evolved in the deep sea from other polychaetes, taking on an unique outward appearance, such as the very long trunk region and the absence of a mouth, gut and anus (Jones, 1981), which are adaptations for symbiosis with sulphide-oxidizing bacteria (Childress *et al.*, 1987).

REFERENCES

Arp, A. J. (1986) in Invertebrate Oxygen Carriers (Linzen, B., ed.), pp. 129–132, Springer-Verlag, Berlin

Arp, A. J. & Childress, J. J. (1981) Science **213**, 342–344

Arp, A. J. & Childress, J. J. (1983) Science **219**, 295–297

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

Garlick, R. L. & Riggs, A. F. (1982) J. Biol. Chem. **257**, 9005–9015

Gilbert, A. T. & Thompson, E. O. P. (1985) Aust. J. Biol. Sci. **38**, 221–236

Goodman, M., Moore, G. W. & Matsuda, G. (1975) Nature (London) **253**, 603–608

Jones, M. L. (1981) Science **213**, 333–336

Jones, M. L. (1985) Biol. Soc. Wash. Bull. **6**, 117–158

Kennicutt, M. C., II, Brooks, J. M., Bidigare, R. R., Fay, R. R., Wade, T. L. & McDonald, T. J. (1985) Nature (London) **317**, 351–353

Mainwaring, M. G., Lugo, S. D., Fingal, R. A., Kapp, O. H. & Vinogradov, S. N. (1986) J.Biol.Chem. **261**, 10899–10908

Ohta, S. & Laubier, L. (1987) Earth Planet. Sci. Lett. **83**, 329–342

Ohta, S. (1987) JAMSTECTR Deepsea Res. **3**, 51–60

Sgourous, J., Kleinschmidt, T. & Braunitzer, G. (1986) in Invertebrate Oxygen Carriers (Linzen, B., ed.), pp. 73–76, Springer-Verlag, Berlin

Shishikura, F., Snow, J. W., Gotoh, T., Vinogradov, S. N. & Walz, D. A. (1987) J. Biol. Chem. **262**, 3123–3131

Southward, E. C. (1975) Symp. Zool. Soc. London **36**, 235–251

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., 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
Terwilliger, R. C., Terwilliger, N. B. & Schabtach, E. (1980) Comp. Biochem. Physiol. 65B, 531–535
Terwilliger, R. C., Terwilliger, N. B., Bonaventura, C., Bonaventura, J. & Schabtach, E. (1985) Biochim. Biophys. Acta **829**, 27–33
Vinogradov, S. N., Shlom, J. M., Hall, B. C., Kapp, O. H. & Mizukami, H. (1977) Biochim. Biophys. Acta **492**, 136–155
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 31 May/25 July 1988; accepted 1 August 1988