Aust. J. Biol. Sci., 1984, **37**, 191–203

Amino Acid Sequence of the Globin IIB Chain of the Dimeric Haemoglobin of the Bivalve Mollusc *Andara trapezia*

W. K. Fisher, A. T. Gilbert and E. O. P. Thompson

School of Biochemistry, University of New South Wales, P.O. Box 1, Kensington, N.S.W. 2033.

Abstract

The tryptic peptides of the *S*-carboxymethylated globin chain of a dimeric haemoglobin from *A. trapezia* were purified by high-performance liquid chromatography and their amino acid sequences determined by the dansyl–Edman method.

The cyanogen bromide fragments were purified or digested with chymotrypsin to give overlap fragments which were sequenced by similar methods.

In conjunction with some sequences of larger peptides derived by digestion with staphylococal protease, the complete sequence of 146 residues was deduced. The sequence has been compared with that previously determined for the 153 residue α-chain of the tetrameric haemoglobin from the same mollusc and shows 45% identity. In the E and F helical regions the homology is particularly noticeable with 75 and 100% identity respectively. The sequence of the dimeric haemoglobin chain of *A. trazezia* is very homologous (90% identity) to the published sequence of the similar haemoglobin from *A. broughtonii*. The homology when compared with other globin chains is much less, with approximately 20% identity in amino acid sequence.

Introduction

In a previous paper (Como and Thompson 1980*a*) the separation of the multiple haemoglobins of the bivalve mollusc *A. trapezia* was described. The major haemoglobin is tetrameric, of the usual $\alpha_2\beta_2$ type. Two minor polymorphic haemoglobins Hb IIa and Hb IIb are also present. These are dimeric with two identical chains in the dimers that are different from the α-and β-chains of the tetrameric haemoglobin. The differences in amino acid composition of the two polymorphic forms of the dimeric globin chain were limited to two or three amino acids in 145 residues.

The amino acid sequence of the 153 residues α-chain of the tetrameric haemoglobin was previously reported (Como and Thompson 1980*b*) and this paper presents the amino acid sequence of the dimeric globin chain Hb IIb, the form that occurs most abundantly in the molluscs in Sydney estuarine waters (Nicol and O'Gower 1967).

Peptides isolated from a variety of enzyme and cyanogen bromide digests have been purified and sequenced by the dansyl–Edman procedure to enable deduction of the sequence. The sequence has been compared with those previously determined for globin chains from other invertebrates or vertebrates, including the amino acid sequence of the dimeric globin chain from another arcid mollusc *A. broughtonii* (Furuta and Kajita 1983).

Materials and Methods

Globin Purification

Specimens of *A. trapezia* from estuaries in the Sydney area were bled from the pallial sinuses and the haemoglobins isolated and fractionated by gel filtration as previously described (Como and

0004-9417/84/040191$02.00

Thompson 1980*a*). The Hb II globin was prepared and purified by chromatography on carboxymethylcellulose in 8 M urea–thiol buffers as described previously (Fisher *et al.* 1980). The globin was normally carboxymethylated before recovery from the urea buffer.

Fractionation of Tryptic Digests

The tryptic digests were fractionated by high-performance liquid chromatography (HPLC) using a semi-preparative, reverse-phase μbondapack C-18 column developed for over 60 min with a linear gradient of 0·1% (v/v) triethylamine-trifluoroacetic acid, pH 3·0, to 70% (v/v) methanol in the starting solution, followed by pure methanol. Fractions were collected manually and the tubes corresponding to the largest peaks combined. After drying, the residual peptide material was further purified by peptide mapping on paper as previously described (Fisher and Thompson 1979).

Isolation of CNBr Fragments

The reaction with CNBr was done in 70% (v/v) formic acid with a large excess of CNBr (approximately 80-fold) for 16 h at room temperature. After removal of the reagents by freeze-drying the residue was dispersed in ammonia and loaded onto a Sephadex G50 column (140 by 2·5 cm) equilibrated with 1% (v/v) ammonia. The effluent was monitored at 230 and 280 nm and divided into eight pools from which the peptide material was recovered by freeze-drying. The fragments in the lower molecular weight fractions were purified by peptide mapping. The larger molecular weight fractions were digested by chymotrypsin and fractionated by peptide mapping.

Isolation of Radioactively Labelled Thiol-containing Peptides

The globin was dissolved in 7 M guanidine hydrochloride containing buffer, treated with dithiothreitol to ensure that oxidation of thiol groups was avoided and alkylated with [2-^{14}C]iodoacetic acid as described by Fisher and Thompson (1983). After recovery, samples of the labelled protein were digested separately with trypsin and chymotrypsin and the digests fractionated on a Sephadex G25 column (140 by 2·8 cm) equilibrated with 1% (v/v) ammonia. The effluent was monitored at 280 nm and by radioactivity. The radioactive peak was recovered after dialysis and freeze-drying. Further purification involved paper ionophoresis in 20% (v/v) formic acid combined with paper chromatography at right angles to the original direction of migration with butanol–pyridine–acetic acid–water (15 : 10 : 3 : 12 v/v) as solvent (Fisher and Thompson 1983).

Peptide Isolation, Amino Acid Analysis and Sequence Determination

The methods of enzyme digestion with trypsin, chymotrypsin, elastase, pepsin, papain and staphylococcal protease; peptide mapping, amino acid analysis and sequence determination were substantially the same as previously described (Air and Thompson 1969; Fisher and Thompson 1983). Because of inadequate facilities for quantitative amino acid analysis of all peptides generated, the hydrolysates were first examined by paper ionophoresis at pH 1·8 (Dreyer and Bynum 1967) while saving a small portion for quantitative analysis on the Beckman 121M analyser (0·5–2 nmol) if it later became essential. The hydrolysates in glass tubes were stored dry in plastic containers in the cold room and suffered no significant deterioration over a year.

Hydrazinolysis followed the procedure of Sleigh *et al.* (1969) using either no catalyst or hydrazine sulfate as catalyst with heating at 80°C for 24 h.

For the isolation of cysteic acid peptides from globin, performic acid oxidation and fractionation on sulfonated polystyrene were the same as previously described (Thompson and Fisher 1978).

Results

Tryptic (Tp), chymotryptic (Ch), and cyanogen bromide (CN) fragments are numbered sequentially from the *N*-terminal end. Staphylococcal protease (Sp), peptic (Pe) and thermolysin (Th) peptides which were investigated selectively to give overlaps, sequence confirmation, or for amide allocations, are indicated by the residue numbers in the peptide.

Amino Acid Composition and Terminal Residues

The amino acid composition of Hb IIa and Hb IIb, the polymorphic forms of the dimeric haemoglobin, were reported previously (Como and Thompson 1980*a*) with 144 and 145 residues, respectively. The differences were confined to residues of lysine, aspartic acid and glycine and were not sufficiently clear-cut to be certain differences without confirmatory amino acid sequence evidence. Differences in amide content are also possible.

The *N*-terminal residue was previously shown to be proline. Hydrazinolysis gave a clear result with leucine as the *C*-terminal residue (0·65 mol/mol) and only traces of glycine (0·07 mol/mol) and alanine (0·02 mol/mol). The uncatalysed reaction gave a higher yield than the catalysed reaction (0·46 mol/mol).

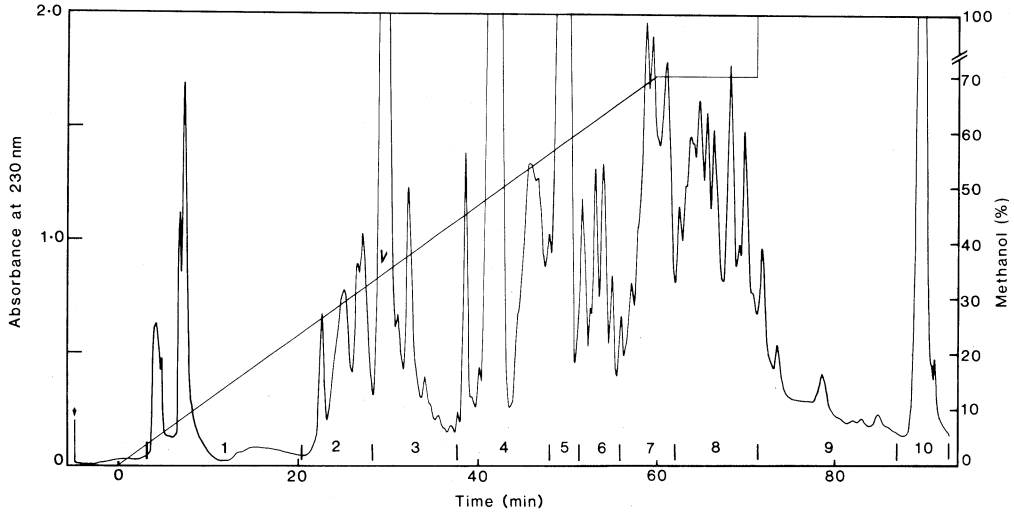


Fig. 1. High-performance liquid chromatography of a tryptic digest of 90 mg of the *S*-carboxymethylated dimeric globin of *A. trapezia* on a semi-preparative μbondapak C-18 column. Gradient elution was from 0·1% (v/v) triethylamine–trifluoracetic acid, pH 3·0, to 0·1% (v/v) triethylamine–fluoroacetic acid—70% (v/v) methanol, pH 3·0, over 60 min at a flow rate of 1·5 ml/min and measured at 230 nm. The elution was continued with the 70% methanol buffer followed by a final wash with pure methanol. The areas indicated 1–10 were bulked and concentrated for peptide map examination. The major peptides in each fraction are shown below, using the number corresponding to their position in the chain:

1	2	3	4	5	7	8	10
Tp 2,6,13,	Tp 2,3	Tp 4	Tp 9	Tp 18+19	Tp 1	Tp 1	Tp 11
16 (Lys)	Tp 3	Tp 5	Tp 12		Tp 1,2	Tp 7	
Tp 8 (Arg)	Tp 5,6	Tp 9	Tp 13+14		Tp 20		
Tp 10	Tp 15	Tp 15	Tp 14				
	Tp 17	Tp 18	Tp 19				
	Tp 18						

Sequence Analysis

The mixture of Hb IIa and Hb IIb obtained from the mixed population of mollusc in Sydney waters is predominantly Hb IIb (Nicol and O'Gower 1967) and because of the minor differences between the globins, the mixture was used for the initial sequence work. Differences in peptides could be revealed in the course of the work, but in any event it is easier to search for them systematically when the amino acid sequence is known, together with the fractionation procedures best suited to isolate particular peptides from the minor component, Hb IIa. The mixtures investigated contained 15–20% Hb IIa.

The fractionation of tryptic peptides is shown in Fig. 1. The principal peptides in each fraction are listed in the legend. Because of the large number of separated peaks in the

HPLC elution curve, no attempt was made to investigate each peak separately. Subsequent fractionation by paper techniques gave relatively clean separations with the larger peptides Tp 7 (22 residues) Tp 11 (29 residues) and Tp 20 (9 residues) being insoluble and remaining at the origin. As these peptides were in different HPLC fractions, they were recoverable separately.

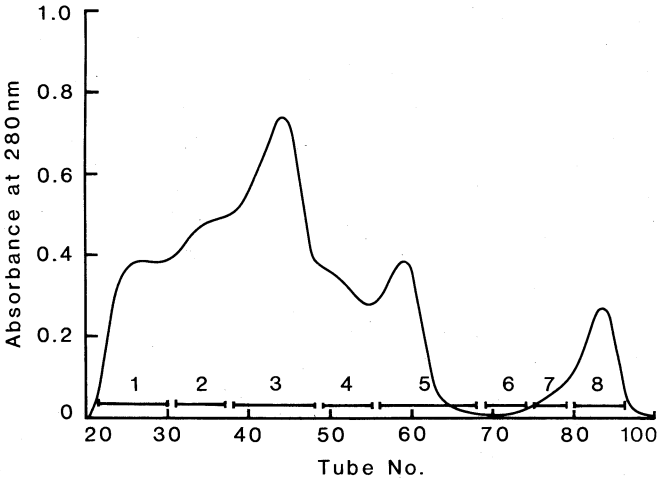


Fig. 2. Elution profile of CNBr fragments of *S*-carboxymethylated globin IIb of *A. trapezia* on a column of Sephadex G50 (120 by 2·3 cm) in 1% (v/v) ammonia. Fraction size was 5·5 ml. Fractions 6 and 7 were identified by absorption at 230 nm. Bulked fractions are shown by bars and CNBr fragments in each peak are indicated. The fragments in peak 6 and 7 were purified by peptide mapping as in Fig. 1 of the Accessory Publication; peak 8 contained no peptide material; the other fragments were deduced from chymotryptic peptides given by each fraction. The fragments in each fraction are shown below using the abbreviation CN with the number corresponding to their position in the chain:

1	2	3	4	5	6 and 7
CN 1	CN 1	CN 1	CN 2+3	CN 4	CN 2
CN 6	CN 3+4	CN 3+4	CN 3	CN 5	CN 4
CN 7	CN 7	CN 7	CN 4		CN 5
			CN 5		

Fractionation of CNBr fragments by gel filtration using formic acid for elution was not successful due to gel formation and poor resolution. With 1% ammonia for elution (Fig. 2), the resolution was insufficient to give pure components and dansyl terminal studies showed multiple components in each peak. Three fragments, CN 2 (3 residues), CN 4 (7 residues) and CN 5 (13 residues) were purified from the lower molecular weight fraction by peptide mapping on paper. The larger CNBr fragments, CN 1 (34 residues), CN 6 (44 residues) and CN 7 (28 residues) were difficult to separate by gel filtration or paper techniques, and the mixtures of fragments were digested with chymotrypsin before fractionation by peptide mapping. The Met-Thr bond was not completely cleaved and a longer fragment, CN 2+3, was found.

The peptides used in the deduction of the amino acid sequence of Hb IIb are shown in Table 1. Peptides for which quantitative amino acid analysis figures are available are underlined with a full line. These analyses and the basis for assignment of amide groups to particular residues are included in the supplementary data.*

*Supplementary data to this paper are deposited as Accessory Publication with the Editor-in-Chief, Editorial and Publications Service, CSIRO, P.O. Box 89, East Melbourne, Vic. 3002. The data deposited are as follows: amino acid analyses of underlined peptides; peptide maps for tryptic and chymotryptic digests of CNBr-treated *S*-carboxymethylated globin IIb; peptides and their net charge or other evidence for assignment of side chain amide and acidic groups in globin IIb. Copies may be obtained on request.

The sequence of Tp 1 given in the previous paper (Como and Thompson 1980*b*) was incomplete. Further analyses indicated a longer peptide and a different sequence beyond the first seven residues reported.

The sequence in peptide Tp 11, the longest peptide, could not be determined by the dansyl–Edman method alone, and two supplementary methods utilizing the single cysteine in the peptide chain were used. The cysteine was carboxymethylated with [2-^{14}C]iodoacetic acid and the labelled peptides in tryptic and chymotryptic digests were isolated by gel filtration, peptide mapping and radioautography. The amino acid composition of the tryptic peptide (origin) of the peptide map corresponded to residues 68–96, while the chymotryptic peptide corresponded to residues 76–97. The valine and isoleucine values were low due to incomplete hydrolysis. The composition of the chymotryptic peptide was

Lys	Asp	Thr	Ser	Glu	Ala	Val	Ile	Leu	Phe	CMCys
1·1	5·7	1·0	1·1	3·1	1·0	0·8	1·8	3·3	2·0	(+)
(1)	(5)	(1)	(1)	(3)	(1)	(2)	(2)	(3)	(2)	(1)

A partially split product of this peptide, corresponding to residues 81–97, was also found.

The dansyl–Edman degradation of the long chymotryptic peptide gave the sequence

$$\underline{\text{Ala-Leu-Glx-Asx-Phe-Ile-Asx-Glx-Leu-Asx}}$$

but did not reach the labelled cysteine residue.

In an alternative approach, a peptic digest of performic acid-oxidized globin was used, with isolation of the cysteic acid peptides. Two highly acidic peptides were isolated and sequenced. They had identical *N*-terminal residues and amino acid compositions, apart from a valine residue in one of them. Mobility at pH 6·4 indicated that all residues were acidic. The longer peptide obviously had a *C*-terminal valine residue extra, with the composition

$$CySO_3H_{1.0}Asp_{3.1}Thr_{1.2}Ser_{1.2}Ile_{0.9}Leu_{0.9}Val_{1.0}$$

and *N*-terminal sequence

$$\underline{\text{Asp-Ser-Thr-Asp-Asp-Leu-Ile-CySO}_3\text{H-Val}}$$

Thermolytic digests of the radioactive peptides gave the evidence for completion of the sequence and the assignment of amide groups as shown in Table 1. A staphylococcal protease peptide of 15 residues and the composition

$$Lys_{2.0}His_{1.0}Arg_{1.2}Asp_{1.1}Thr_{1.0}Ser_{1.0}Glu_{1.1}Gly_{1.1}Ala_{2.1}Val_{1.1}Ile_{1.8}Phe_{1.0}$$

gave an *N*-terminal sequence corresponding to residues 96–100 with the dansyl–Edman method, thus identifying it and providing the overlaps in the region 96–110.

Amino Acid Sequence of Dimeric Haemoglobin

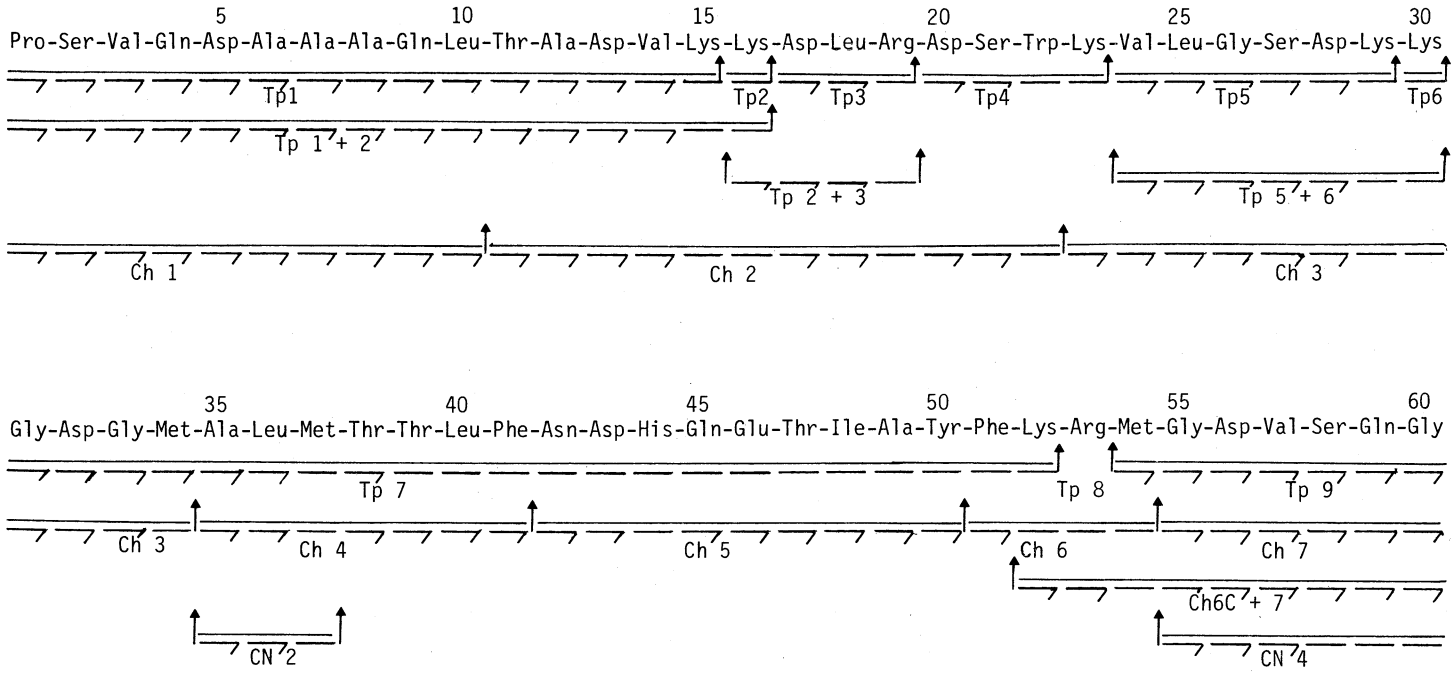
The complete amino acid sequence of Hb IIb of the dimeric haemoglobin of *A. trapezia* is shown in Table 1. Although the order of the fragments can be deduced from the peptides and overlap peptides, the strong homology to the α-chain and the globin chain from *A. broughtonii* (Furuta and Kajita 1983) makes this unnecessary.

A comparison of the amino acid composition expected from the sequence with that found by amino acid analysis (Como and Thompson 1980*a*) indicates satisfactory agreement, with an extra residue of lysine and serine in the sequence to give a total of 146 residues in the chain.

During the course of the work, some evidence of heterogeneity in some peptides was observed. Residue 139 was serine in a variant peptide. Amide differences were also apparent, as reflected in peptide mobilities. Thus, peptide Tp 7, which was large and insoluble, showed a variant that differed in amide content, as detected in chymotryptic

Table 1. Amino acid sequence of the dimeric globin chain from *Anadara trapezia* haemoglobin IIb

Residues are numbered from the *N*-terminal residue and the positions of enzyme cleavage shown by arrows. Where a secondary split has occurred within a large chymotryptic peptide, N and C represent the *N*-terminal and *C*-terminal fragments respectively. Residues identified by the dansyl–Edman procedure are underlined by an arrow. If no horizontal arrow is shown, the residue did not give a clear result, or was not investigated. The abbreviations Tp, Ch, Pe, Pa, Sp, Th are for trypsin, chymotrypsin, pepsin, papain, staphylococcal protease and thermolysin respectively. Peptides which were quantitatively analysed for amino acids are underlined with a full line. CN is the abbreviation for CNBr fragments



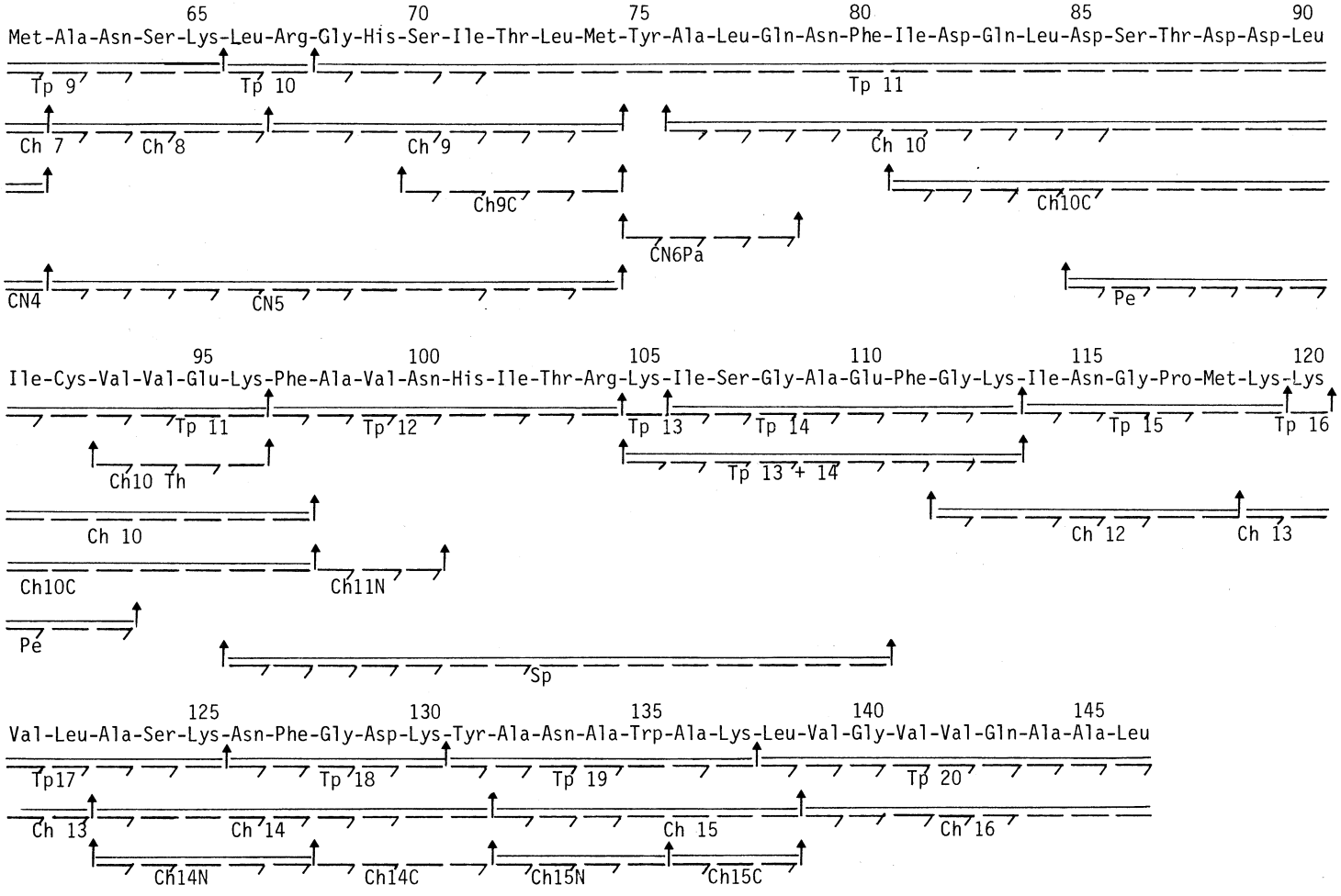

65 70 75 80 85 90
Met-Ala-Asn-Ser-Lys-Leu-Arg-Gly-His-Ser-Ile-Thr-Leu-Met-Tyr-Ala-Leu-Gln-Asn-Phe-Ile-Asp-Gln-Leu-Asp-Ser-Thr-Asp-Asp-Leu
Tp 9
Tp 10
Tp 11
Ch 7
Ch 8
Ch 9
Ch 10
Ch9C
Ch10C
CN6Pa
CN4
CN5
Pe
95 100 105 110 115 120
Ile-Cys-Val-Val-Glu-Lys-Phe-Ala-Val-Asn-His-Ile-Thr-Arg-Lys-Ile-Ser-Gly-Ala-Glu-Phe-Gly-Lys-Ile-Asn-Gly-Pro-Met-Lys-Lys
Tp 11
Tp 12
Tp 13
Tp 14
Tp 15
Tp 16
Ch10 Th
Tp 13 + 14
Ch 10
Ch 12
Ch 13
Ch10C
Ch11N
Pe
Sp
125 130 135 140 145
Val-Leu-Ala-Ser-Lys-Asn-Phe-Gly-Asp-Lys-Tyr-Ala-Asn-Ala-Trp-Ala-Lys-Leu-Val-Gly-Val-Val-Gln-Ala-Ala-Leu
Tp17
Tp 18
Tp 19
Tp 20
Ch 13
Ch 14
Ch 15
Ch 16
Ch14N
Ch14C
Ch15N
Ch15C

Table 2. Amino acid sequence of eight invertebrate globins aligned with myoglobin, α- and β-chains from human and shark haemoglobin, lamprey globin and leghaemoglobin from kidney bean

Alignment and sequences are from Dayhoff (1972) but to allow for deletions and insertions in the invertebrate chains additional gaps have been necessary in the interhelical regions: four in AB, five in EF and two in FG. Ac- and F- represent an acetyl and formyl group respectively. Sequences not in Dayhoff (1972) are taken from Fisher *et al.* (1977), Fisher and Thompson (1979) and Furuta and Kajita (1983). Only those residues different from those in *Anadara trapezia* IIb are shown for *A. broughtonii* II

	NA	A															AB					B														B
Globin	1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2	3	4	5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Human Mb				G	-	L	S	D	G	E	W	Q	L	V	L	N	V	W	G	K	V	E	A	-	-	-	-	D	I	P	G	H	G	Q	E	V	L	I	R	L	F	K	G
Human α				V	-	L	S	P	A	D	K	T	N	V	K	A	A	W	G	K	V	G	A	-	-	-	-	H	A	G	E	Y	G	A	E	A	L	E	R	M	F	L	S
Human β				V	H	L	T	P	E	E	K	S	A	V	T	A	L	W	G	K	V	-	-	-	-	-	-	N	V	D	E	V	G	G	E	A	L	G	R	L	L	V	V
Shark Mb								A	c	-	T	E	W	E	H	V	N	K	V	W	A	V	V	E	P	-	-	-	-	D	I	P	A	V	G	L	A	I	L	L	R	L	F	K	E
Shark α			A	c	-	S	T	S	T	S	T	S	D	-	Y	S	A	A	D	R	A	E	L	A	A	L	S	K	V	L	A	Q	-	-	-	-	N	A	E	A	F	G	A	E	A	L	A	R	M	F	T	V
Shark β													V	H	W	S	E	V	E	L	H	E	I	T	T	T	W	K	S	I	-	-	-	-	-	-	D	K	H	S	L	G	A	K	A	L	A	R	M	F	I	V
Lamprey	F	-	P	I	V	D	S	G	S	V	A	P	L	S	A	A	E	K	T	K	I	R	S	A	W	A	P	V	Y	S	-	-	-	-	N	Y	E	T	S	G	V	D	I	L	V	K	F	F	T	S
Aplysia Mb									A	c	-	S	-	L	S	A	A	E	A	D	L	A	G	K	S	W	A	P	V	F	A	-	-	-	-	N	K	N	A	N	G	A	D	F	L	V	A	L	F	E	K
Busycon Mb													G	-	L	D	G	A	Q	K	T	A	L	K	E	S	W	-	K	V	L	G	A	D	G	P	T	M	M	K	N	G	S	L	L	F	G	L	L	F	K	T
Anadara t.I-α			A	c	-	V	A	D	A	V	A	K	V	C	G	S	E	A	I	K	G	N	L	R	R	S	W	G	V	L	M	S	A	-	-	-	D	I	E	A	T	G	L	T	Y	L	A	N	L	F	T	L
A. trapezia IIB					P	S	V	Q	D	A	A	A	Q	L	T	A	D	V	K	K	D	L	R	D	S	W	K	V	L	G	S	-	-	-	-	D	K	K	G	D	G	M	A	L	M	T	T	L	F	N	D
A. broughtonii II								G																						I											N		V						A	
Chironomus IIβ													A	P	L	S	A	D	E	A	S	L	V	R	G	S	W	A	N	V	K	-	-	-	-	-	-	-	-	H	S	E	V	D	I	L	Y	Y	I	F	K	A
Chironomus III													-	-	L	S	A	D	Q	I	S	X	V	Z	A	S	F	D	K	V	K	G	-	-	-	-	-	-	D	P	V	G	-	-	I	L	Y	A	V	F	K	A
Chironomus VIIβ													S	P	L	T	A	D	E	A	S	L	V	Q	S	S	W	K	A	V	S	-	-	-	-	-	-	-	-	H	N	E	V	D	I	L	A	A	V	F	A	A
Glycera Hb													G	-	L	S	A	A	Q	R	Q	V	I	A	A	T	W	K	D	I	A	G	N	D	-	-	N	G	A	G	V	G	K	D	C	L	I	K	H	L	S	A
Leg Hb K. bean													G	A	F	T	E	K	Q	E	A	L	V	N	S	S	W	E	A	F	K	G	-	-	-	-	N	I	P	Q	Y	S	V	V	F	Y	T	S	I	L	E	K

Globin	C 1	2	3	4	5	6	C 7	CD 1	2	3	4	5	6	7	CD 8	D 1	2	3	4	5	6	D 7	E 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	E 20
Human Mb	H	P	E	T	L	E	K	F	D	K	F	K	H	L	K	S	E	D	E	M	K	A	S	E	D	L	K	K	H	G	A	T	V	L	-	T	A	L	G	G	I	L	K
Human α	F	P	T	T	K	T	Y	F	P	H	F	-	D	L	S	H	-	-	-	-	-	G	S	A	Q	V	K	G	H	G	K	K	V	A	-	D	A	L	T	N	A	V	A
Human β	Y	P	W	T	Q	R	F	F	E	S	F	G	D	L	S	T	P	D	A	V	M	G	N	P	K	V	K	A	H	G	K	K	V	L	-	G	A	F	S	D	G	L	A
Shark Mb	H	K	E	T	K	D	L	F	P	K	F	K	E	I	-	P	V	Q	Q	L	G	N	N	E	D	L	R	K	H	G	V	T	V	L	-	R	A	L	G	N	I	L	K
Shark α	Y	A	A	T	K	S	Y	F	K	D	Y	K	D	F	T	A	-	-	-	-	-	A	A	P	S	I	K	A	H	G	A	K	V	V	-	T	A	L	A	K	A	C	D
Shark β	Y	P	W	T	T	R	Y	F	G	N	L	K	E	F	T	A	-	-	-	-	-	C	S	Y	H	-	K	E	H	A	K	K	V	T	-	G	A	L	G	V	A	V	T
Lamprey	T	P	A	A	Q	E	F	F	P	K	F	K	G	M	T	S	A	D	E	L	K	K	S	A	D	V	R	W	H	A	E	R	I	I	-	N	A	V	N	D	A	V	A
Aplysia Mb	F	P	D	S	A	N	F	F	A	D	F	K	G	-	K	S	V	A	D	I	K	A	S	P	K	L	R	D	V	S	S	R	I	F	-	T	R	L	N	E	F	V	N
Busycon Mb	Y	P	D	T	K	K	H	F	K	H	F	D	D	A	T	-	F	A	A	M	D	T	T	G	V	G	K	A	H	G	V	A	V	F	-	S	G	L	G	S	M	I	C
Anadara t. I-α	R	P	D	T	K	T	Y	F	T	R	L	G	D	V	Q	K	G	-	-	-	K	A	N	S	K	L	R	G	H	A	I	T	L	T	-	Y	A	L	D	W	F	V	D
A. trapezia IIB	H	Q	E	T	I	A	Y	F	K	R	M	G	D	V	S	Q	G	-	-	-	M	A	N	S	K	L	R	G	H	S	I	T	L	M	-	Y	A	L	Q	N	F	I	D
A. broughtonii II	N					G					L		N										D																					
Chironomus IIβ	N	P	D	I	M	A	K	F	P	N	F	A	G	-	K	D	L	E	T	L	K	G	T	G	Q	F	A	T	H	G	G	R	I	V	-	G	F	V	S	E	I	V	A
Chironomus III	D	P	S	I	M	A	K	F	T	Q	F	A	G	-	K	D	L	E	S	I	K	G	T	A	P	F	H	T	E	A	N	R	I	V	-	G	F	F	S	K	L	I	G
Chironomus VIIβ	Y	P	D	I	Q/M	A	K	F	P/S	Q	F	A	G	-	K	D	L	A	S	I	K	D	T	G	A	F	A	T	H	A	T	R	I	V	-	S	F	L	S	E	V	I	A
Glycera Hb	H	P	Q	M	A	A	V	F	G	-	F	S	G	-	A	S	D	P	A	V	-	-	-	A	D	L	-	-	-	G	A	K	V	L	-	A	Z	I	G	V	A	V	S
Leg Hb K. bean	A	P	A	A	K	N	L	F	-	S	F	-	-	L	A	N	G	V	D	-	P	T	N	P	K	L	T	A	H	A	E	S	L	F	G	L	V	R	D	S	A	A	Q

Globin	EF 1	2	3	4	5	6	7	8	9	10	11	12	13	EF 14	F 1	2	3	4	5	6	7	8	FG 9	1	2	3	4	5	FG 6	7	G 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	G 19
Human Mb	K	K	G	H	-	-	-	H	E	A	-	-	-	E	I	K	P	L	A	Q	S	H	A	T	K	-	-	H	K	I	P	V	K	Y	L	E	F	I	S	E	C	I	I	Q	V	L	Q	S	K
Human α	H	V	D	D	-	-	-	M	P	N	-	-	-	A	L	S	A	L	S	D	L	H	A	H	K	-	-	L	R	V	D	P	V	N	F	K	L	L	S	H	C	L	L	V	T	L	A	A	H
Human β	H	L	D	N	-	-	-	L	K	G	-	-	-	T	F	A	T	L	S	E	L	H	C	D	K	-	-	L	H	V	D	P	E	N	F	R	L	L	G	N	V	L	V	C	V	L	A	H	H
Shark Mb	Q	K	G	K	-	-	-	H	S	T	-	-	-	N	V	K	E	L	A	D	T	H	I	N	K	-	-	H	K	I	P	P	K	N	F	V	L	I	T	N	I	A	V	K	V	L	T	E	M
Shark α	H	L	D	D	-	-	-	L	K	T	-	-	-	H	L	H	K	L	A	T	F	H	G	S	E	-	-	L	K	V	D	P	A	N	F	Q	Y	L	S	Y	C	L	E	V	A	L	A	V	H
Shark β	H	L	G	D	-	-	-	V	K	S	-	-	-	Q	F	T	D	L	S	K	K	H	A	E	E	-	-	L	H	V	D	V	E	S	F	K	L	L	A	K	C	F	V	V	E	L	G	I	L
Lamprey	S	M	D	D	T	E	K	M	S	-	-	-	-	-	M	K	D	L	S	G	K	H	A	K	S	-	-	F	Q	V	D	P	Q	Y	F	K	V	L	A	-	V	I	A	D	T	V	A	A	-
Aplysia Mb	D	A	A	N	A	G	K	M	S	A	-	-	-	M	L	S	Q	F	A	K	E	H	V	G	F	G	V	G	S	A	Q	F	E	N	V	R	S	M	F	P	G	F	V	A	S	V	A	A	-
Busycon Mb	S	I	D	D	D	D	C	V	B	G	-	-	-	L	A	K	K	L	S	R	N	H	L	A	-	-	-	R	G	V	S	A	A	D	F	K	L	L	-	E	A	V	F	K	Z	F	L	-	-
Anadara t. I-α	S	L	D	D	P	S	R	L	K	C	-	-	-	V	V	E	K	F	A	V	N	H	I	N	R	-	-	-	K	I	S	G	D	A	F	G	S	I	I	P	E	M	K	E	T	L	K	A	R
A.trapezia IIB	Q	L	D	S	T	D	D	L	I	C	-	-	-	V	V	E	K	F	A	V	N	H	I	T	R	-	-	-	K	I	S	G	A	E	F	G	K	I	N	G	P	M	K	K	V	L	A	S	K
A.broughtonii II				N					V																							A										I							
Chironomus IIβ	L	M	G	N	S	A	N	M	P	A	M	E	T	L	I	K	D	M	A	A	N	H	-	K	A	-	-	R	G	I	P	K	A	Q	F	N	E	F	R	A	S	L	V	S	Y	L	Q	S	K
Chironomus III	E	I	P	E	-	-	N	I	A	D	-	-	-	V	N	T	F	P	A	S	K	H	V	-	R	G	V	T	H	-	-	B	Z	-	L	B	B	F	R	A	G	F	V	S	Y	M	K	A	H
Chironomus VIIβ	L	S/M	G	N	E/A	S	N	A	A	A	V	Q	G	L	L	D	K	L	G	D	D	H	K	-	A	-	-	R	G	V	S	A	A	Q	F	G	E	F	R	L/T	A	L	V	A	Y	L	Q	A	H
Glycera Hb	H	L	G	D	Z	G	K	M	V	A	-	-	-	Q	M	K	A	V	G	V	R	H	K	G	Y	G	N	K	H	I	K	G	Q	Y	F	E	P	L	G	A	S	L	L	S	A	M	E	H	R
Leg Hb K. bean	L	R	A	N	-	-	G	A	V	V	-	-	-	A	D	A	A	L	G	S	I	H	S	Q	K	G	V	S	N	-	-	D	Q	-	F	L	V	V	K	E	A	L	L	K	T	L	K	Q	A

Globin	GH 1	2	3	GH 4	5	H 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	H 21	HC 1	2	(3	(25	H 26	HC 1	2	Mb 3	4)
Human Mb	H	P	G	D	F	G	A	D	A	Q	G	A	M	N	K	A	L	E	L	F	R	K	D	M	A	S	N	Y	K	E	L	G	F	Q	G
Human α	L	P	A	E	F	T	P	A	V	H	A	S	L	D	K	F	L	A	S	V	S	T	V	L	T	S	K	Y	R						
Human β	F	G	K	E	F	T	P	P	V	Q	A	A	Y	Q	K	V	V	A	G	V	A	N	A	L	A	H	K	Y	H						
Shark Mb	Y	P	S	D	M	T	G	P	M	Q	E	S	F	S	K	V	F	T	V	I	C	S	D	L	E	T	L	Y	K	E	A	D	F	Q	G
Shark α	L	-	T	E	F	S	P	E	T	H	C	A	L	D	K	F	L	T	N	V	C	H	E	L	S	S	R	Y	R						
Shark β	L	K	D	K	F	A	P	Q	T	Q	A	I	W	E	K	Y	F	G	V	V	V	D	A	I	S	K	E	Y	H						
Lamprey	-	-	-	-	-	-	-	-	G	D	A	G	F	E	K	L	S	M	(I	C)	I	L	M	L	R	S	A	Y	-						
Aplysia Mb	-	-	-	-	-	P	P	A	G	A	D	A	W	T	K	L	F	G	L	I	I	D	A	L	K	A	A	G	K						
Busycon Mb	-	-	D	E	A	T	Q	R	K	A	T	D	A	Q	K	D	A	D	G	A	L	L	T	M	L	I	K	A	H	V				
Anadara t. I-α	-	M	G	S	Y	S	D	D	V	G	A	A	W	V	Q	A	I	L	G	M	Q	N	A	V	L	S	A	L							
A.trapezia IIB	-	-	-	N	F	G	D	K	Y	A	N	A	W	A	K	L	V	G	V	V	Q	A	A	L											
A.broughtonii II																		A																	
Chironomus IIβ	-	-	V	S	W	N	D	S	L	G	A	A	W	T	N	G	L	D	N	V	F	N	M	M	F	S	-	Y	L						
Chironomus III	-	-	T	D	F	G	A	E	A	-	A	A	W	G	T	L	(F	D	T	F	A)	G	M	I	F	S	K	M	-						
Chironomus VIIβ	-	-	V	S	W	G	N	N	V	A	A	A	W	S	K	A	L	D	N	T	V F	A	I	V	V	P	R	L	-						
Glycera Hb	I	G	G	K	M	N	A	A	A	K	D	A	W	A	A	A	Y	A	D	I	S	G	A	L	I	S	-	-	-	-	-	G	L	Q	S
Leg Hb. K. bean	V	G	D	K	W	T	D	Q	L	S	T	A	L	E	L	A	Y	D	E	L	A	A	A	I	K	K	A	Y	A						

peptides Ch 42–50. One variant had a mobility indicating more aspartyl or glutamyl residues and this peptide had a *N*-terminal aspartic acid, compared with the slower variant with *N*-terminal asparagine.

These differences could reflect differences between the Hb IIa sequence, present in small amounts, and that of Hb IIb; or differences between individual animals in the pooled samples used; or deamidation during the fractionation experiments. Since deamidation has often been reported by others working on globin chains, particularly myoglobins (for summary, see Fisher and Thompson 1979) further comparative work is necessary to pinpoint the differences between Hb IIa and Hb IIb.

It is preferable to use procedures that avoid extremes of pH in the isolation of peptides for amide allocations, and the direct sequenator procedure used by Furuta and Kajita (1983) has obvious advantages.

Discussion

In Table 2, the sequence of the dimeric haemoglobin of *A. trapezia* is aligned with that of *A. broughtonii*, the α-chain of the tetrameric haemoglobin of *A. trapezia*, six other invertebrate globins and the myoglobin and α- and β-chains from human and shark haemoglobin; lamprey globin; and leghaemoglobin from kidney bean.

The number of differences between the dimeric haemoglobins of *A. trapezia* and *A. broughtonii* is 15 in 146 residues. Thus there is a 90% identity. The differences are spread throughout the chains. When compared with the α-chain of *A. trapezia*, 66 residues of the 146-residue chain are identical, or 45% identity. In some sections of the chain, the homology is much greater than this, with 100% identity in the F helix and 75% in the E helix. Otherwise, the differences are spread throughout the chain. Comparisons with the other globin chains shows less identity (of the order of 20%) in amino acid sequence.

The extension at the amino terminal end of the dimeric haemoglobin resembles that found in α-globins of the shark, lamprey and *A. trapezia* except that the amino group is not acylated. The shorter length of chain is largely due to a deletion of four residues at the *C*-terminus.

The alignment and known sequences in Table 2 were taken from Dayhoff (1972). For comparisons, it is assumed that the tertiary structure of haemoglobin chains are the same as that of the horse (Perutz 1976). In general, when the α-chain of *A. trapezia* haemoglobin I was compared with other globin chains (Como and Thompson 1980*b*), the helical regions A to H remained unchanged in length, but additional residues had to be put into the interhelical regions to preserve homology. The dimeric haemoglobin chains of *A. trapezia* and *A. broughtonii* have these additional residues as well as a high degree of homology with the α-chain. There are, apart from the four deleted *C*-terminal residues in *A. trapezia* globin IIb, three other deleted residues. One of these is in the A–B region and the other two in the GH interhelical region. With the additional information provided by the globin IIb sequence, the arrangement at the A–AB junction and at the GH–H junction has been adjusted slightly compared with the previous tabulation (Como and Thompson 1980*b*).

For the globin sequences shown in Table 2 there are only two invariant residues, the F8 histidine that binds the haem and CD1 phenylalanine that is a haem contact residue. The *A. trapezia* I-α and IIb globins have the E7 histidine and there is an additional residue of histidine in the globin IIb chain. Apart from these contact residues, several other residues in the Hb IIb globin match haem contact residues in either α-, β-, or myoglobin chains and the essential non-polar characteristics are preserved.

Globin IIb does not participate in tetramer formation with an unlike globin chain and only forms dimers with itself, so that it is not suprising to find that of the 34 amino acids involved in the $\alpha_1\beta_1$ contact sites in human α- and β-chains only four of them are in similar positions in the *A. trapezia* globin; while for the 20 amino acids in the $\alpha_1\beta_2$ contact site positions there is only one residue of similar location.

For residue positions with defined functional roles in mammalian α- and β-chains of haemoglobin in connection with the Bohr effect, binding of polyphosphates such as 2,3-diphosphoglycerate, or salt links between the chains in tetramers (Goodman *et al.* 1975), there are no identical positions when compared with human haemoglobin chains. This is in keeping with the absence of heterotropic interactions in the dimeric haemoglobin (Furuta *et al.* 1980).

Acknowledgments

This work was supported in part by the Australian Research Grants Scheme. We thank Mr R. G. Mann for the amino acid analyses and for assistance, together with M. Black, D. M. Webster and Dr A. R. Thompson, in the collection of specimens.

References

Air, G. M., and Thompson, E. O. P. (1969). Studies on marsupial proteins. II. Amino acid sequence of the β-chain of haemoglobin from the grey kangaroo, *Macropus giganteus*. *Aust. J. Biol. Sci.* **22**, 1437–54.

Como, P. F., and Thompson, E. O. P. (1980*a*). Multiple haemoglobins of the bivalve mollusc *Anadara trapezia*. *Aust. J. Biol. Sci.* **33**, 643–52.

Como, P. F., and Thompson, E. O. P. (1980*b*). Amino acid sequence of the α-chain of the tetrameric haemoglobin of the bivalve mollusc *Anadara trapezia*. *Aust. J. Biol. Sci.* **33**, 653–64.

Dayhoff, M. O. (1972). 'Atlas of Protein Sequence and Structure'. Vol. 5. Plus Supplements 1 (1973), 2 (1976), and 3 (1978). (National Biomedical Research Foundation: Silver Spring.)

Dreyer, W. J., and Bynum, E. (1967). High-voltage paper electrophoresis. In 'Methods in Enzymology'. (Ed. C. H. W. Hirs.) Vol. XI. pp. 32–9. (Academic Press: New York.)

Fisher, W. K., Koureas, D. D., and Thompson, E. O. P. (1980). Myoglobins of cartilaginous fishes. II. Isolation and amino acid sequence of myoglobin of the shark *Mustelus antarcticus*. *Aust. J. Biol. Sci.* **33**, 153–67.

Fisher, W. K., and Thompson, E. O. P. (1979). Myoglobin of the shark *Heterodontus portusjacksoni*: Isolation and amino acid sequence. *Aust. J. Biol. Sci.* **32**, 277–94.

Fisher, W. K., and Thompson, E. O. P. (1983). Amino acid sequence studies of sheep liver fructose-bisphosphatase. II. The complete sequence. *Aust. J. Biol. Sci.* **36**, 235–50.

Furuta, H., and Kajita, A. (1983). Dimeric hemoglobin of the bivalve mollusc *Anadara broughtonii*: Complete amino acid sequence of the globin chain. *Biochemistry* **22**, 917–22.

Furuta, H., Ohe, M., and Kajita, A. (1980). Ligand-dependent allosteric transformation of hemoglobins from the blood clam, *Anadara broughtonii*. *Biochim. Biophys. Acta* **625**, 318–27.

Goodman, M., Moore, G. W., and Matsuda, G. (1975). Darwinian evolution in the genealogy of haemoglobin. *Nature, Lond.* **253**, 603–8.

Nicol, P. I., and O'Gower, A. K. (1967). Haemoglobin variation in *Anadara trapezia*. *Nature, Lond.* **216**, 684.

Perutz, M. F. (1976). Structure and mechanism of haemoglobin. *Br. Med. Bull.* **32**, 195–208.

Sleigh, R., Hosken, R., Smith, M. B., and Thompson, E. O. P. (1969). Chemical characterization of a non-covalently linked peptide from plakalbumin. *Aust. J. Biol. Sci.* **22**, 239–46.

Thompson, E. O. P., and Fisher, W. K. (1978). Amino acid sequences containing half-cystine residues in ovalbumin. *Aust. J. Biol. Sci.* **31**, 433–42.

Webster, D. M., and Thompson, E. O. P. (1980). Position of the disulfide bond in ovalbumins of differing heat stability. Elimination of thiol-disulfide interchange as a mechanism for the formation of the ovalbumins. *Aust. J. Biol. Sci.* **33**, 269–78.

Manuscript received 26 March 1984, accepted 16 May 1984