Aust. J. Biol. Sci., 1980, **33**, 653–64

Amino Acid Sequence of the α-Chain of the Tetrameric Haemoglobin of the Bivalve Mollusc *Anadara trapezia*

P. F. Como and E. O. P. Thompson

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

Abstract

The amino acid sequence of the α-chain of the tetrameric haemoglobin of the bivalve mollusc *A. trapezia* shows 153 amino acid residues, longer than invertebrate globins previously sequenced. The amino terminal residue is acetylated as shown by nuclear magnetic resonance and mass spectrographic analysis of an *N*-terminal peptide. There are additional residues at the amino terminus, similar to the α-globin of the shark and lamprey globin. In common with other invertebrate globin sequences, when compared with vertebrate globins it is necessary to insert extra gaps in interhelical regions of the vertebrate alignment to preserve homology.

The amino acid sequence has been determined following digestion with trypsin, chymotrypsin, thermolysin, pepsin and cyanogen bromide. Sequences of purified peptides were determined by the dansyl–Edman procedure. The sequence has strong homology to vertebrate α-chains rather than β-chains in the region C2 to CD1. Three D-helical residues are missing, which is a characteristic of α-chains and possibly of invertebrate globins. The sequence has been compared with those of α-, β- or myoglobin chains of vertebrates, and with invertebrate and plant globins.

Introduction

In a previous paper (Como and Thompson 1980) the separation of the multiple haemoglobins of the bivalve mollusc *Anadara trapezia* was described and the amino acid compositions of the different globin chains were given. The major haemoglobin is tetrameric, of the usual $\alpha_2\beta_2$ type. The α- and β-chains were separated and in this paper the determination of the amino acid sequence of the α-chain is presented.

Peptides isolated from a variety of enzyme digests have been purified and sequenced by the dansyl–Edman procedure. Several large insoluble peptides, as well as the large number of differences in amino sequence from the globin chains of vertebrates and invertebrates, necessitated the use of a range of proteolytic enzymes to obtain sufficient overlapping peptides to confirm the sequence deduced. The sequence has been compared with those previously determined for globin chains from other invertebrates or vertebrates.

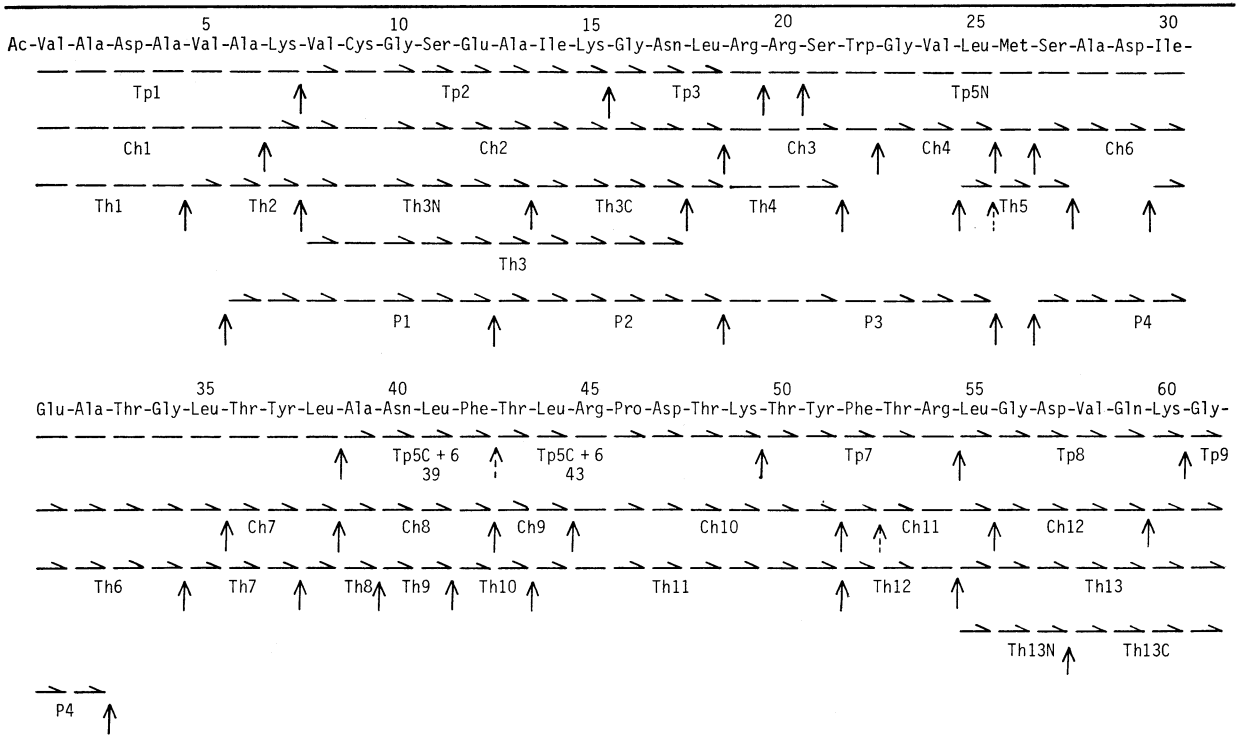
Materials and Methods

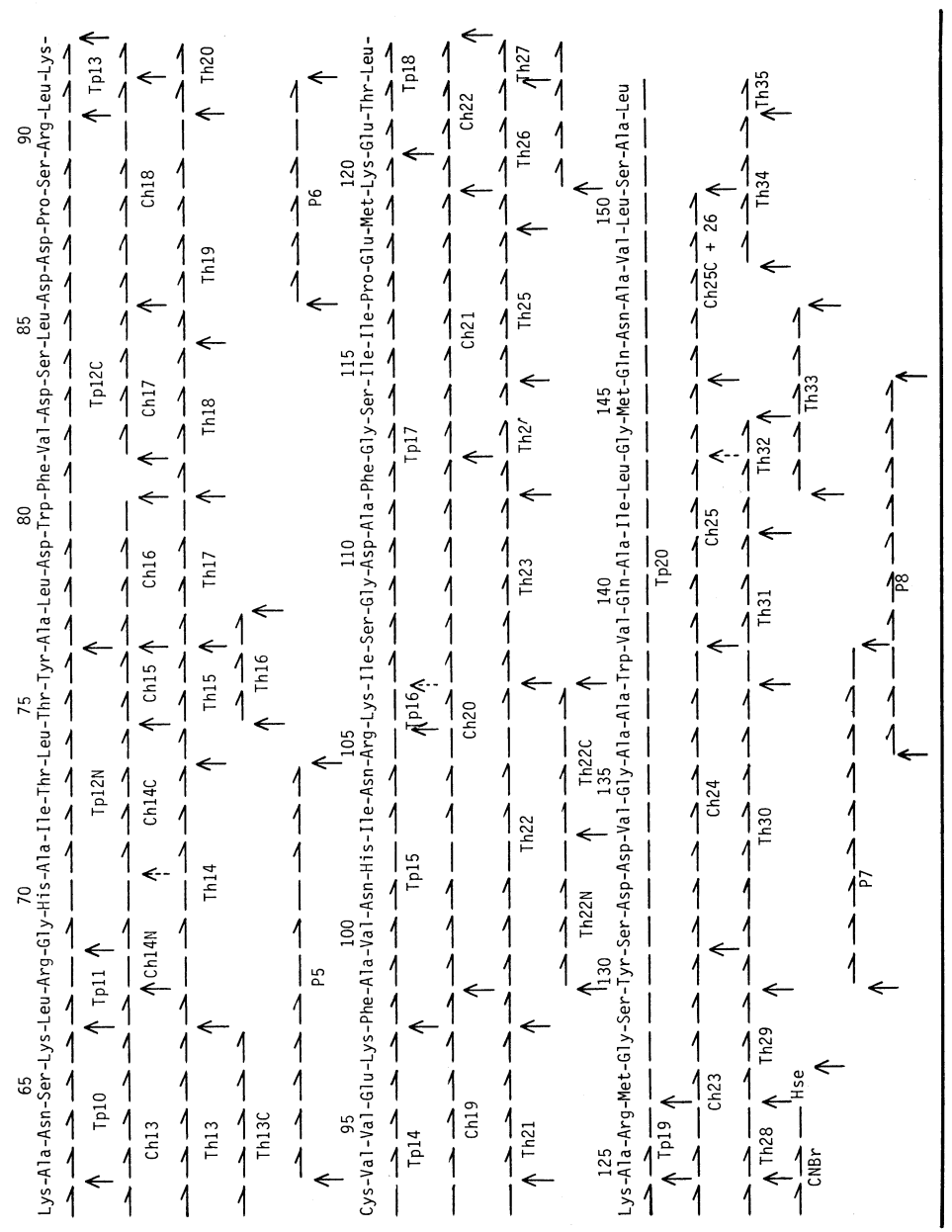
The methods of gel filtration, peptide mapping, amino acid analysis, cyanogen bromide cleavage, sequence determination by the dansyl–Edman procedure, and digestion with trypsin, chymotrypsin, thermolysin and pepsin were substantially the same as previously described (Air and Thompson 1969, 1971; Beard and Thompson 1971; Fisher and Thompson 1979).

The α-chain of HbI of *A. trapezia* was prepared as previously described (Como and Thompson 1980) in the carboxymethylated form.

Table 1. Amino acid sequence of the α-globin chain from *Anadara trapezia* haemoglobin I

Residues are numbered from the *N*-terminal residue and the positions of enzyme cleavage shown by arrows. Where a split has occurred with trypsin that does not involve a basic amino acid residue, N and C represent the *N*-terminal and *C*-terminal fragments respectively. Similarly for other enzymes the positions of cleavage by proteolytic enzymes are shown by vertical arrows. Residues identified by the dansyl–Edman procedure are underlined with an arrow. If no horizontal arrow is shown, the residue did not give a clear result, or was not investigated. Ac represents an acetyl group




65
70
75
80
85
90
Lys-Ala-Asn-Ser-Lys-Leu-Arg-Gly-His-Ala-Ile-Thr-Leu-Thr-Tyr-Ala-Leu-Asp-Trp-Phe-Val-Asp-Ser-Leu-Asp-Asp-Pro-Ser-Arg-Leu-Lys-
Tp10
Tp11
Tp12N
Tp12C
Tp13
Ch13
Ch14N
Ch14C
Ch15
Ch16
Ch17
Ch18
Th13
Th14
Th15
Th16
Th17
Th18
Th19
Th20
Th13C
P5
P6
95
100
105
110
115
120
Cys-Val-Val-Glu-Lys-Phe-Ala-Val-Asn-His-Ile-Asn-Arg-Lys-Ile-Ser-Gly-Asp-Ala-Phe-Gly-Ser-Ile-Ile-Pro-Glu-Met-Lys-Glu-Thr-Leu-
Tp14
Tp15
Tp16
Tp17
Tp18
Ch19
Ch20
Ch21
Ch22
Th21
Th22
Th23
Th2′
Th25
Th26
Th27
Th22N
Th22C
125
130
135
140
145
150
Lys-Ala-Arg-Met-Gly-Ser-Tyr-Ser-Asp-Asp-Val-Gly-Ala-Ala-Trp-Val-Gln-Ala-Ile-Leu-Gly-Met-Gln-Asn-Ala-Val-Leu-Ser-Ala-Leu
Tp19
Tp20
Ch23
Ch24
Ch25
Ch25C + 26
Th28
Th29
Th30
Th31
Th32
Th33
Th34
Th35
CNBr
Hse
P7
P8

For determination of the blocking group on the amino terminal of the chain the methods of Fisher and Thompson (1979) and Fisher *et al.* (1980) were used.

Hydrazinolysis followed the procedure of Sleigh *et al.* (1969) using hydrazine sulfate as catalyst at 80°C for 16 h. The *C*-terminal residue was identified in the dried reaction mixture on the amino acid analyser.

Results

Tryptic (Tp), chymotryptic (Ch) and thermolysin (Th) peptides are numbered sequentially from the *N*-terminal end. Peptic peptides are numbered similarly for those peptides investigated selectively to give overlaps or confirmatory sequences.

Amino Acid Composition and Sequence of Peptides

Amino acid compositions of tryptic peptides, including *N*- and *C*-terminal portions of some peptides split by chymotryptic-like activity, were determined. Satisfactory analyses were obtained for most peptides, including chymotryptic, thermolytic and peptic peptides used to establish overlaps for the tryptic peptides.

The sequence of the tryptic peptides is shown in Table 1. Also shown in Table 1 are the designation and sequence of peptides obtained by digestion with the other proteolytic enzymes.*

In Table 1 each residue identified by the dansyl–Edman procedure is underlined by an arrow. The residues are numbered from the *N*-terminal residue as 1, but homology indicates that residue 8 corresponds to the *N*-terminal residue of most globin chains. The sequences of most soluble peptides were readily determined and only the *N*- and *C*-terminal tryptic peptides are discussed. Side-chain amide and acidic groups were assigned wherever possible from the ionophoretic mobility of small peptides at pH 6·4.

Tp1

This peptide had the sequence

$$\underset{1}{\text{Ac-Val}}\text{-Ala-Asp-Ala-Val-Ala-}\underset{7}{\text{Lys}}$$

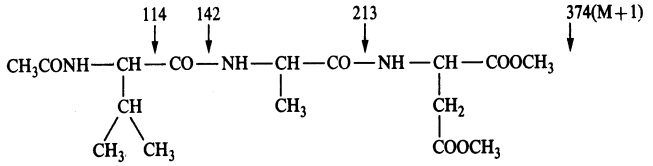
where Ac represents an acetyl group on the amino terminal. The peptide was acidic at pH 6·4, indicating an aspartyl residue as well as the blocking group. The peptide stained weakly with ninhydrin due to the presence of lysine.

The sequence was determined from thermolytic fragments. After 5 h digestion those adsorbed to sulfonated polystyrene and purified after elution with ammonia were Val-Ala-Lys and Ala-Lys.

The non-adsorbed acidic peptide analysed as $Asp_{1.0}Ala_{1.7}Val_{1.0}$. To determine the sequence and the identity of the blocking residue the acidic fraction from a 24-h thermolytic digest was used to try and reduce the size of the peptide to within the molecular weight limits of the instrument for mass spectrometry. The composition

* Supplementary data to this paper are deposited with, and can be obtained from, the Editor-in-Chief, Editorial and Publications Service, CSIRO, P.O. Box 89, East Melbourne, Vic. 3002. The data deposited contain the following material: amino acid analyses of tryptic peptides; peptide maps for tryptic, chymotryptic and thermolytic digests; amino acid analyses of some chymotryptic, thermolytic and peptic peptides; peptides and their net charge used for assignment of side chain amide and acidic groups.

in this case was $Asp_{1\cdot0}Ala_{1\cdot3}Val_{1\cdot0}$ and the proton n.m.r. spectrum was consistent with the presence of an acetyl moiety (Whittaker *et al.* 1979). From the mass spectrograph the most abundant peaks (*cf.* Fisher *et al.* 1980), apart from the small M+1, corresponded to the following molecular weight fragments:



Tp20

This peptide was the *C*-terminal fragment, insoluble at pH 6·4, with the sequence

127 135 145
Met-Gly-Ser-Tyr-Ser-Asp-Asp-Val-Gly-Ala-Ala-Trp-Val-Gln-Ala-Ile-Leu-Gly-Met-
153
Glu-Asn-Ala-Val-Leu-Ser-Ala-Leu

Peptide Ch23 gave the overlap with the preceding tryptic peptide. The remaining sequence was deduced from chymotryptic, thermolytic and peptic peptides (Table 1) that were isolated from digests of the α-globin or the material insoluble at pH 6·4. The soluble peptides sequenced cleanly and there was ample confirmatory data. The aspartyl residues at positions 132 and 133 were allocated on the basis of electrophoretic mobility at pH 6·4 of Ch24, P7 and Th30. The neutral peptides αTh31 and αTh33 showed the presence of glutaminyl and asparaginyl residues in them.

As carboxypeptidase-A digestion of α-globin released not only leucine but also alanine in poor yield (Como and Thompson 1980), the *C*-terminal residue was confirmed by the release of leucine (0·6 residue/mole) by hydrazinolysis.

That leucine is the *C*-terminal residue of the α-chain of the major haemoglobin was reported previously by Furuta *et al.* (1977).

Amino Acid Sequence of the α-Chain of HbI of A. trapezia

The complete amino acid sequence of the α-chain is shown in Table 1, together with the peptides used in deducing the sequence. The sequence gave 153 residues which compares favourably with the amino acid analysis estimate of 152 residues (Como and Thompson 1980). The sequence showed three variations from the analytical values for particular amino acids, with one residue less for aspartic acid and one residue more for threonine and leucine.

Discussion

With 153 residues the α-chain of the tetrameric haemoglobin I of *A. trapezia* is the longest invertebrate globin so far reported. The amino terminal is blocked by acetylation as is the case in many lower vertebrate globin chains, such as in eel haemoglobin (Amano *et al.* 1968*a*); α-chains of the carp (Hilse and Braunitzer 1968), frog (Chauvet and Acher 1971; Flavin *et al.* 1976), sucker fish (Powers and Edmundson 1972) and shark (Nash *et al.* 1976); and myoglobins of the mollusc

Table 2. Amino acid sequences of seven invertebrate globins aligned with myoglobin, α- and β-chains from human being and shark; lamprey globin; and leghaemoglobin A from kidney and soy 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 inter-helical regions: four in AB, five in EF and two in FG. Ac- and F- represent an acetyl and formyl group respectively. Sequencies not in Dayhoff (1972) are taken from Fisher *et al.* (1977) or Fisher and Thompson (1979)

Globin	NA	A																A	AB				AB	B													B
	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 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
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
Leg Hb S. bean											V	A	F	T	E	K	Q	D	A	L	V	S	S	S	F	E	A	F	K	A	-	-	-	-	N	I	P	Q	Y	S	V	V	F	Y	T	S	I	L	E	K

Globin	C 1	2	3	4	5	6	7	C CD 1	2	3	4	5	6	7	8	CD D 1	2	3	4	5	6	7	D 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 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
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
Leg Hb S.bean	A	P	A	A	K	D	L	F	-	S	F	-	-	L	A	N	G	V	D	-	P	T	N	P	K	L	T	G	H	A	E	K	L	F	A	L	V	R	D	S	A	G	Q

(Table 2 contd overleaf)

Table 2 (*contd*)

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	F 8	FG 9	1	2	3	4	FG 5	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 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
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
Leg Hb S. bean	L	K	A	S	-	-	G	T	V	V	-	-	-	A	D	A	A	L	G	S	V	H	A	Q	K	A	V	T	N	-	-	P	Q	-	F	-	V	V	K	E	A	L	L	K	I	I	K	A	A

Globin	GH 1	2	3	4	GH H 5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	H HC 21	1	2	(H HC Mb 3	(25	26	1	2	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 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						
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					
Leg Hb.S.bean	V	G	D	K	W	S	K	E	L	S	R	A	W	E	V	A	Y	D	E		L	A	A	A	I	K	A	K	-						

Aplysia limacina (Tentori *et al.* 1973), tuna and skipjack (Amano *et al.* 1968*b*) and shark (Fisher and Thompson 1979; Fisher *et al.* 1980).

The extra residues in the chain occur in an extension at the *N*-terminal end of the chain similar to that previously found in lamprey globin (Braunitzer and Fujiki 1969; Li and Riggs 1970) and shark α-chain (Nash *et al.* 1976). In Table 2 a comparison of the amino acid sequence of the α-chain of *A. trapezia* is made with human globin chains, with other lower vertebrate globins and with invertebrate globins. The alignment and sequences are taken from Dayhoff (1972). In general, the helical regions A–H remain unchanged, but additional residues must be put in the interhelical regions to preserve homology. Thus *A. trapezia* myoglobin has additional residues in the AB interhelical region as found also for myoglobin of *Busycon canaliculatum* (Bonner and Laursen 1977) and globin of *Glycera dibranchiata* (Imamura *et al.* 1972); in the EF region as in *Aplysia limacina* and *Busycon canaliculatum* myoglobins and lamprey globin.

The D-helix does not appear to be formed in the α-globin of *A. trapezia* in line with other α-globin sequences. It is this feature as well as the strong homology of the sequence in the C2–CD1 region,

C2 CD1
Pro-Asp-Thr-Lys-Thr-Tyr-Phe,

to other α-chain sequences that is most suggestive that it is an α-chain and not a β-chain with anomolous electrophoretic behaviour and amino acid content.

There are only two histidine residues in the chain and they are associated with the usual haem-binding residues, at E7 and F8. The only common residues in the sequence shown in Table 2 are F8 His and CD1 Phe. *A. trapezia* α-globin exhibits 26 identical residues in comparison with *Busycon canaliculatum*; and 31 in comparison with *Aplysia limacina* myoglobin, whereas the two gastropod molluscs have only 19 identical residues. The sequence shows 31 ± 2 identical residues compared with vertebrate α-, β- or myoglobin chains and, with the exception of the 31 identical residues in *Aplysia limacina* myoglobin, 23 ± 3 identical residues compared with invertebrate or plant globins out of a total of 136–153 residues. It is apparent that the globin chains of plants and lower vertebrates exhibit individual differences as far as insertions and deletions are concerned, while retaining the myoglobin fold that originated early in evolutionary history (Vainshtein *et al.* 1975).

If the tertiary and quaternary structure of the tetrameric haemoglobin of *A. trapezia* is the same as that of the horse, apart from slight extensions of interhelical regions, then there have been a considerable number of changes in residues involved in haem contacts (Perutz 1976) and interactions with the β-chain (Perutz 1976). Thus in haem contacts, 11 out of 16 residues known to be involved in the binding are different, although the hydrophobic character of the side chains is preserved. Of the residues in α-chain involved in $\alpha_1\beta_2$ contact sites (Bolton and Perutz 1970), only two are identical in the oxy-contact and four in the deoxy-contact site. This number of changes of α-chain residues in the $\alpha_1\beta_2$ contact sites is greater than found with tetrameric haemoglobins of lower vertebrates, such as shark α-chain (Fisher *et al.* 1977). It will be interesting to examine the corresponding changes in the β-chains when the amino acid sequence of the β-chain is known.

Acknowledgments

This work was supported in part by the Australian Research Grants Committee. We are indebted to Mr R. G. Mann for amino acid analyses, and to Drs K. D. Barrow and A. M. Duffield for the n.m.r. and mass spectrograph analyses.

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.

Air, G. M., and Thompson, E. O. P. (1971). Studies on marsupial proteins. IV. Amino acid sequence of myoglobin from the red kangaroo, *Megaleia rufa*. *Aust. J. Biol. Sci.* **24**, 75–95.

Amano, H., Hashimoto, K., and Matsuura, F. (1968*a*). *N*-terminal amino acid of eel haemoglobin. *Bull. Jap. Soc. Sci. Fish.* **34**, 937–43.

Amano, H., Hashimoto, K., and Matsuura, F. (1968*b*). *N*-terminal structure of fish myoglobin. *Bull. Jap. Soc. Sci. Fish.* **34**, 955–8.

Beard, J. M., and Thompson, E. O. P. (1971). Studies on marsupial proteins. V. Amino acid sequences of the α-chain of haemoglobin from the grey kangaroo, *Macropus giganteus*. *Aust. J. Biol. Sci.* **24**, 765–86.

Bolton, W., and Perutz, M. F. (1970). Three dimensional Fourier synthesis of horse deoxyhaemoglobin at 2·8 Å resolution. *Nature (London)* **228**, 551–2.

Bonner, A. G., and Laursen, R. A. (1977). The amino acid sequence of a dimeric myoglobin from the gastropod mollusc, *Busycon canaliculatum* L. *FEBS Lett.* **73**, 201–3.

Braunitzer, G., and Fujiki, H. (1969). Zur evolutionen der vertebraten. Die konstitution und tertiärstruktur des hamoglobins des flussneunauges. *Naturwissenschaften* **56**, 322–3.

Chauvet, J. P., and Acher, R. (1971). Isolation of a frog haemoglobin (*Rana esculenta*). Purification, amino acid composition, and terminal sequences of the α and β chains. *Int. J. Protein Res.* **3**, 261–9.

Como, P. F., and Thompson, E. O. P. (1980). Multiple haemoglobins of the bivalve mollusc *Anadara trapezia*. *Aust. J. Biol. Sci.* **32**, 000–00.

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.)

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., Nash, A. R., and Thompson, E. O. P. (1977). Haemoglobins of the shark, *Heterodontus portusjacksoni*. III. Amino acid sequence of the β-chain. *Aust. J. Biol. Sci.* **30**, 487–506.

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.

Flavin, M., Blouquit, Y., and Rosa, J. (1976). Structural study of the α-chain of one haemoglobin from the adult salamander, *Pleurodeles waltii*. *FEBS Lett.* **67**, 52–7.

Furuta, H., Ohe, M., and Kajita, A. (1977). Subunit structure of haemoglobins from erythrocytes of the blood clam, *Anadara broughtonii*. *J. Biochem. (Tokyo)* **82**, 1723–30.

Hilse, K., and Braunitzer, G. (1968). Hämoglobine. XVI. Die aminosäuresequenz der α-Ketten der beiden Hauptkompanenten des karpfenhämoglobins. *Z. Physiol. Chem.* **349**, 433–50.

Imamura, T., Baldwin, T. O., and Riggs, A. (1972). The amino acid sequence of the monomeric hemoglobin component from the bloodworm *Glycera dibranchiata*. *J. Biol. Chem.* **247**, 2785–97.

Li, S. L., and Riggs, A. (1970). The amino acid sequence of hemoglobin V from the lamprey *Petromyzon marinus*. *J. Biol. Chem.* **245**, 6149–69.

Nash, A. R., Fisher, W. K., and Thompson, E. O. P. (1976). Haemoglobins of the shark, *Heterodontus portusjacksoni*. II. Amino acid sequence of the α-chain. *Aust. J. Biol. Sci.* **29**, 73–97.

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

Powers, D. A., and Edmundson, A. B. (1972). II. The amino acid sequence of the major α-chain from *Catostomus clarkii* hemoglobins. *J. Biol. Chem.* **247**, 6694–707.

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.

Tentori, L., Vivaldi, G., Carta, S., Marinucci, M., Massa, A., Antonini, E., and Brunori, M. (1973). The amino acid sequence of myoglobin from the mollusc *Aplysia limacina*. *Int. J. Pept. Protein Res.* **5**, 187–200.

Vainshtein, B. K., Harutyunyan, E. H., Kuranova, I. P., Borisov, V. V., Sosfenov, N. I., Pavlovsky, A. G., Grebenko, A. O., and Konareva, N. V. (1975). Structure of leghaemoglobin from lupin root nodules at 5 Å resolution. *Nature (London)* **254**, 163–4.

Whittaker, R. G., Moss, B. A., and Appleby, C. A. (1979). Determination of the blocked *N*-terminal of soybean leghemoglobin b. *Biochem. Biophys. Res. Commun.* **89**, 552–8.

Manuscript received 28 July 1980, accepted 10 September 1980