Aust. J. Biol. Sci., 1985, **38**, 221–36

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

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 amino acid sequence of the β-chain of the principal haemoglobin from *A. trapezia* has been determined. The sequence was deduced from the sequences of tryptic peptides, which were fractionated using high-performance liquid chromatography and peptide mapping. Additional sequence data, particularly for the large tryptic peptides, was obtained from enzyme digests of both cyanogen bromide fragments and large citraconyltryptic peptides.

The β-chain has 151 residues which is longer than all the other sequenced haemoglobin chains except the α-chain of *A. trapezia*, which is 153 residues in length. The residues corresponding to those normally in the D helix are absent in this β-chain. The additional residues are contributed by an extension of the *N*-terminal region, which was also found to be acetylated.

Comparison of the β-chain amino acid sequence with that of the α-chain of *A. trapezia*, the dimeric chain of *A. trapezia*, and the dimeric chain of *A. broughtonii* showed 53% identity in each case. In the E and F helices, the homology is particularly noticeable. There is 100% homology in the F helix of all four chains. The dimeric globin of *A. trapezia* also shows 100% homology with the β-chain in the E helix, while the α-chain shows 75%.

If the tertiary structure of the α- and β-chains of *A. trapezia* haemoglobin is the same as that of horse haemoglobin, then there are many changes in the $\alpha_1\beta_2$ contact site residues.

Introduction

In a previous paper (Como and Thompson 1980*a*) the separation of the haemoglobins of the bivalve mollusc *Anadara trapezia* was reported. The major haemoglobin is tetrameric, of the usual $\alpha_2\beta_2$ type. Also present are two dimeric, polymorphic haemoglobins with identical chains that are different from the α- and β-chains of the tetrameric haemoglobin.

The amino acid sequence of the α-chain of the tetramer has been determined (Como and Thompson 1980*b*), as has that of the chain of the HbIIb dimer (Fisher *et al.* 1984). In this paper, the complete amino acid sequence of the β-chain of the tetrameric haemoglobin is presented.

The peptides isolated from enzyme digests of the β-chain and its cyanogen bromide fragments or from fragments obtained by tryptic digestion of citraconylated β-globin have been purified and sequenced by the dansyl-Edman procedure. The sequence of the β-chain has been compared with other *Anadara* spp. globin chain sequences, and the functional residues of the $\alpha_1\beta_2$ contact sites of other tetrameric haemoglobins compared with those present in *A. trapezia* globin chains.

0004-9417/85/030221$02.00

Materials and Methods

Globin Purification

A. trapezia specimens were collected from estuarine waters in the Sydney and Gosford areas. They were bled from the pallial sinuses and the haemoglobins fractionated by gel filtration as previously described (Como and Thompson 1980*a*). The α- and β-chains of HbI were separated by ion-exchange chromatography on DEAE-Sepharose (Como and Thompson 1980*a*). The β-globin was *S*-carboxymethylated according to the method of Thompson *et al.* (1969).

Fractionation of Tryptic Digests

Tryptic peptides were fractionated by high-performance liquid chromatography (HPLC) using a semi-preparative, μbondapak C18 column developed over 60 min with a linear gradient of 0·1% (v/v) triethylamine–trifluoroacetic acid, pH 2·5, to 70% (v/v) methanol in the starting buffer. This was followed by a 100% methanol wash. Fractions were collected manually and tubes corresponding to the largest peaks combined. After drying, the residual peptide material was further purified by peptide mapping as described by Fisher and Thompson (1979).

Isolation of Cyanogen Bromide Fragments

S-Carboxymethyl β-globin (200 mg) in 5% (v/v) acetic acid (35 ml) was first reacted with thioglycollic acid, 5 ml, 50% (w/w) purified free of thiolesters (White 1960), at 37°C for 36 h following the method of Houghten and Li (1983), to reduce any methionine sulfoxide groups and so facilitate more efficient cleavage with cyanogen bromide (CNBr). The reduced globin was subsequently reacted, in 70% (v/v) formic acid, with twice its weight of CNBr for 16 h at room temperature. After removal of the reagents by freeze-drying, the peptide material was dissolved in 100% formic acid and loaded onto a column of Sephadex G50 equilibrated in 50% (v/v) formic acid. The fractionation was monitored at 280 nm and the effluent divided into six pools from which fragments were recovered by freeze-drying. The fragments in the lowest molecular weight fraction were purified by peptide mapping. The larger molecular weight fractions were digested with chymotrypsin and the peptides were purified by peptide mapping.

Isolation of Fragments after Arginyl-bond-specific Trypsin Cleavage

β-Globin was citraconylated according to the method of Henschen and Lottspeich (1975). The citraconylated globin was digested with trypsin for 3 h and the peptide mixture was then acidified to pH 2·5 for 16 h, desalted by gel filtration on Sephadex G10 in 5% (v/v) formic acid and freeze-dried. The soluble peptides were fractionated using HPLC in combination with peptide mapping as described above, the HPLC column being developed to 25% (v/v) isopropanol in the starting buffer, instead of 70% methanol.

The insoluble citraconyl-tryptic peptides were fractionated by gel-filtration on Sephadex G50 equilibrated in 50% (v/v) formic acid. The separation was monitored at 280 nm and each peak was recovered by freeze-drying. The largest molecular weight fragment was digested with pepsin. The other fragments were digested with chymotrypsin with the exception of the smallest fragment, which was sequenced directly. The peptides from the digested fragments were purified by peptide mapping.

N-*Terminal Determination*

The β-globin did not give a dansyl derivative corresponding to a residue with a free α-amino group, and this was consistent with the presence of an *N*-terminal blocking group. A CNBr fragment, CN1, similarly did not give a dansyl derivative. After digestion of CN1 with chymotrypsin, the *N*-terminal chymotryptic fragment was purified on a column of sulfonated polystyrene as previously described (Fisher and Thompson 1979) and its composition determined by amino acid analysis of a hydrolysate. The chymotryptic peptide was further digested with thermolysin and the smaller *N*-terminal peptide repurified on the sulfonated polystyrene column. The non-adsorbed fraction was examined by amino acid analysis of a hydrolysate. A portion of this fraction was subjected to hydrazinolysis and the free amino acid identified by paper ionophoresis at pH 1·8, by dansylation and by amino acid analysis. The complete sequence of the *N*-terminal thermolytic peptide, and the nature of the blocking group, was determined by mass spectrometry (Fisher and Thompson 1979) of its trimethylsilyl (TMS) derivative.

Peptide Isolation, Amino Acid Analysis and Sequence Determination

The methods of enzyme digestion with trypsin, chymotrypsin, thermolysin and pepsin, of peptide mapping, of amino acid analysis and of sequence determination were substantially the same as described by Air and Thompson (1969) and Fisher and Thompson (1983).

Portions of the peptides generated were routinely hydrolysed and subjected to ionophoresis at pH 1·8 (Dreyer and Bynum 1967) for determination of the amino acid composition. A small portion of the hydrolysate was saved for quantitative amino acid analysis on a Beckman 121M analyser (0·5–2 nmol) if later required.

Hydrazinolysis followed the procedure of Sleigh *et al.* (1969) using no catalyst at 80°C for 24 h.

Results

Tryptic (Tp), chymotryptic (Ch), thermolytic (Th), peptic (Pe), citraconyl (Cit) and cyanogen bromide (CN) fragments are numbered sequentially from the *N*-terminal end (Table 1).

Terminal Residues

The blocked *N*-terminal chymotryptic peptide had an amino acid composition of $T_{0.8}S_{1.0}E_{0.9}A_{1.1}V_{1.0}L_{1.0}$. The smaller *N*-terminal peptide resulting from thermolytic digestion of the chymotryptic peptide had a composition of $S_{1.0}T_{0.8}$. The hydrazinolysis product of this dipeptide was identified, by dansylation and amino acid analysis, as threonine.

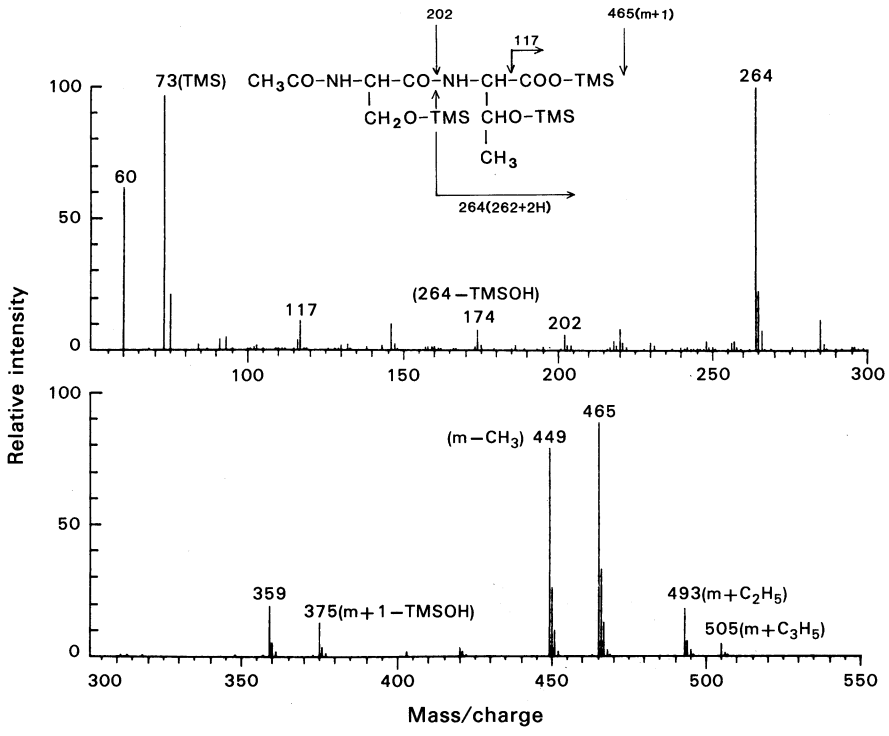


Fig. 1. Mass spectrograph trace of the blocked amino-terminal peptide isolated from a sample of *A. trapezia* *S*-carboxymethylated β-chain haemoglobin which had been subjected to successive digestions with chymotrypsin and thermolysin. Inset refers to the diagrammatic representation of some of the mass spectrographic split products of the peptide acetyl-Ser-Thr, which had been derivatized with $Si(CH_3)_3$(TMS). Additional explanations of mass spectrographic peaks are also given.

Mass spectrometric evidence (Fig. 1) indicated that the peptide was blocked with an acetyl group, and had the sequence acetyl-serylthreonine.

The *C*-terminal residue was identified, by hydrazinolysis of the *S*-carboxymethylated β-globin, as leucine. This agrees with the results of Furuta *et al.* (1977) for the β-chain of *A. broughtonii.*

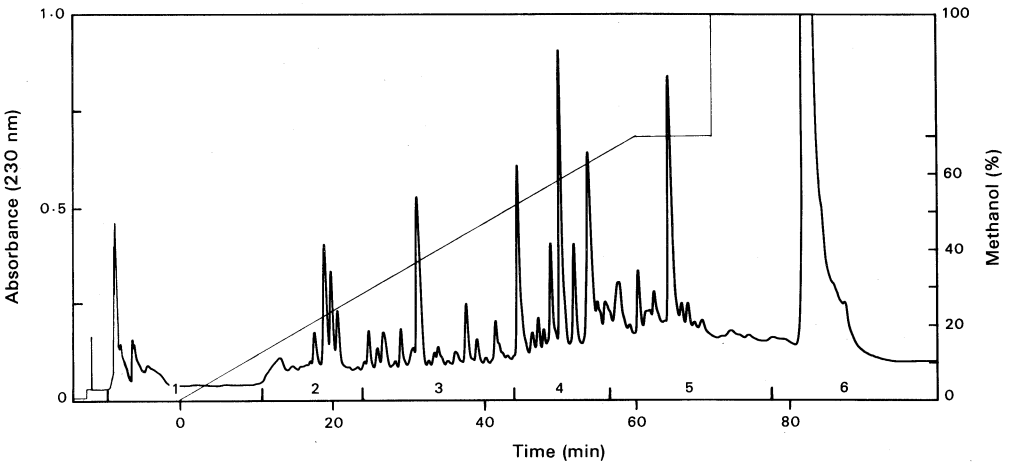


Fig. 2. High-performance liquid chromatography of a tryptic digest of 75 mg of the *S*-carboxymethylated β-globin of *A. trapezia* on a semi-preparative μbondapak C-18 column. Gradient elution was from 0·1% (v/v) triethylamine–trifluoroacetic acid, pH 2·5, to 0·1% (v/v) triethylamine–trifluoroacetic acid–70% (v/v) methanol, pH 2·5, over 60 min at a flow rate of 1·7 ml/min and measured at 230 nm. The elution was continued with the 70% methanol buffer followed by a final wash with 100% methanol. The areas indicated 1–6 were bulked and concentrated for further purification by peptide mapping. The major peptides in each fraction are shown below, using the number corresponding to their position in the chain:

1	2	3	4	5	6
Tp4	Tp5 + 6	Tp2	Tp14	Tp1	Tp3
Tp5	Tp6	Tp8	Tp17N	Tp17C	Tp10
Tp9	Tp7	Tp13			Tp10 + 11
Tp11	Tp7 + 8				
Tp15	Tp12				
Tp16					

Sequence Analysis

The fractionation of tryptic peptides by HPLC is shown in Fig. 2. Because of the large number of peaks in the elution curve, no attempt was made to investigate each peak separately. Subsequent fractionation by paper techniques gave relatively pure peptides with the large peptides [Tp1 (17 residues), Tp3 (23 residues), Tp10 (22 residues)] being insoluble and remaining at the origin. Tp17 cleaved (chymotryptic-like activity) at tryptophan (residue 140) and its *N*-terminal and *C*-terminal peptides were recovered in different HPLC fractions. Peptides Tp3 and Tp10 were both present in the same fraction and were partially separated by paper electrophoresis in 20% (v/v) formic acid followed by descending chromatography, at right-angles to the original direction of migration, with butanol–pyridine–acetic acid–water (15 : 10 : 3 : 12 v/v) as solvent.

The fractionation of CNBr fragments is shown in Fig. 3. Two fragments, CN2 (6 residues) and CN3 (2 residues) were purified from the lower molecular weight fraction by peptide mapping. The larger CNBr fragments CN1 (31 residues), CN4 (37 residues), CN5 (52 residues) and CN6 (23 residues) were not purified to homogeneity, but after digestion with chymotrypsin the peptides were fractionated by peptide mapping.

The HPLC elution peaks of the soluble citraconyl–tryptic peptides were bulked into several fractions which were then purified further by peptide mapping in a similar manner to the tryptic peptides. The insoluble citraconyltryptic peptides were fractionated by gel filtration on G50 Sephadex. A poor separation of these peptides was obtained, most likely due to the presence of tryptophan, which has a retarding effect, in two of the peptides. Cit2 (29 residues) would be expected to separate from Cit5 (22 residues) and

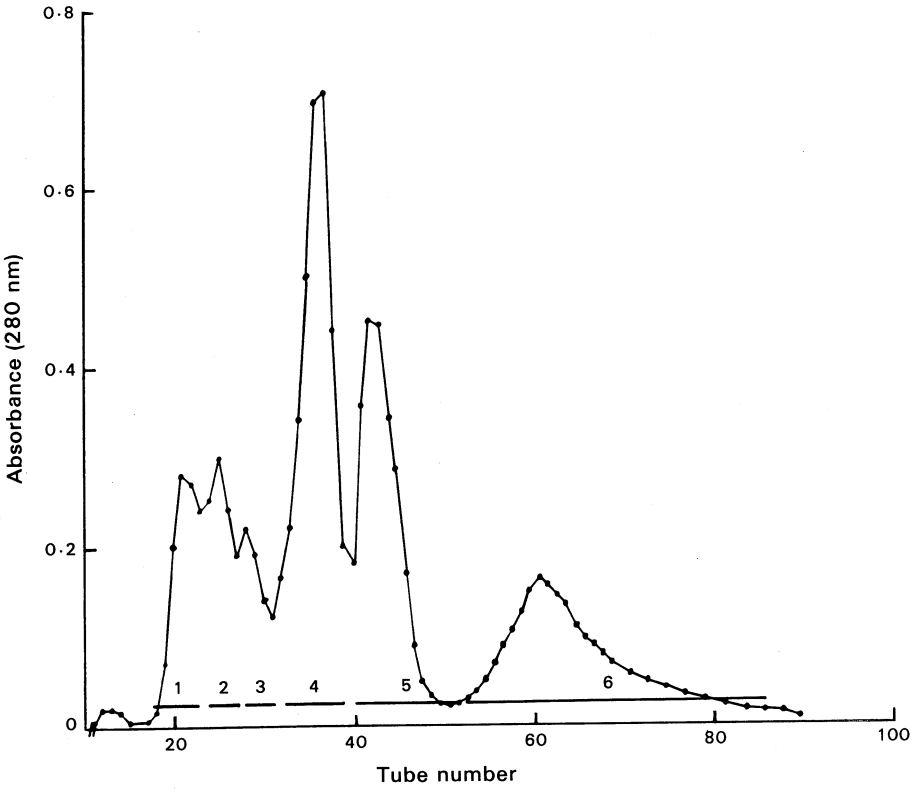


Fig. 3. Elution profile of CNBr fragments of *S*-carboxymethylated β-globin of *A. trapezia* on a column of Sephadex G50 (140 by 2·4 cm) in 50% (v/v) formic acid. Fraction size was 8 ml. Bulked fractions are shown by bars and CNBr fragments in each peak are indicated. The fragments in peak 6 were purified by peptide mapping; the other fragments were deduced from chymotryptic peptides given by each fraction. The fragments in each fraction are shown below using the abbreviations CN with the number corresponding to their position in the chain.

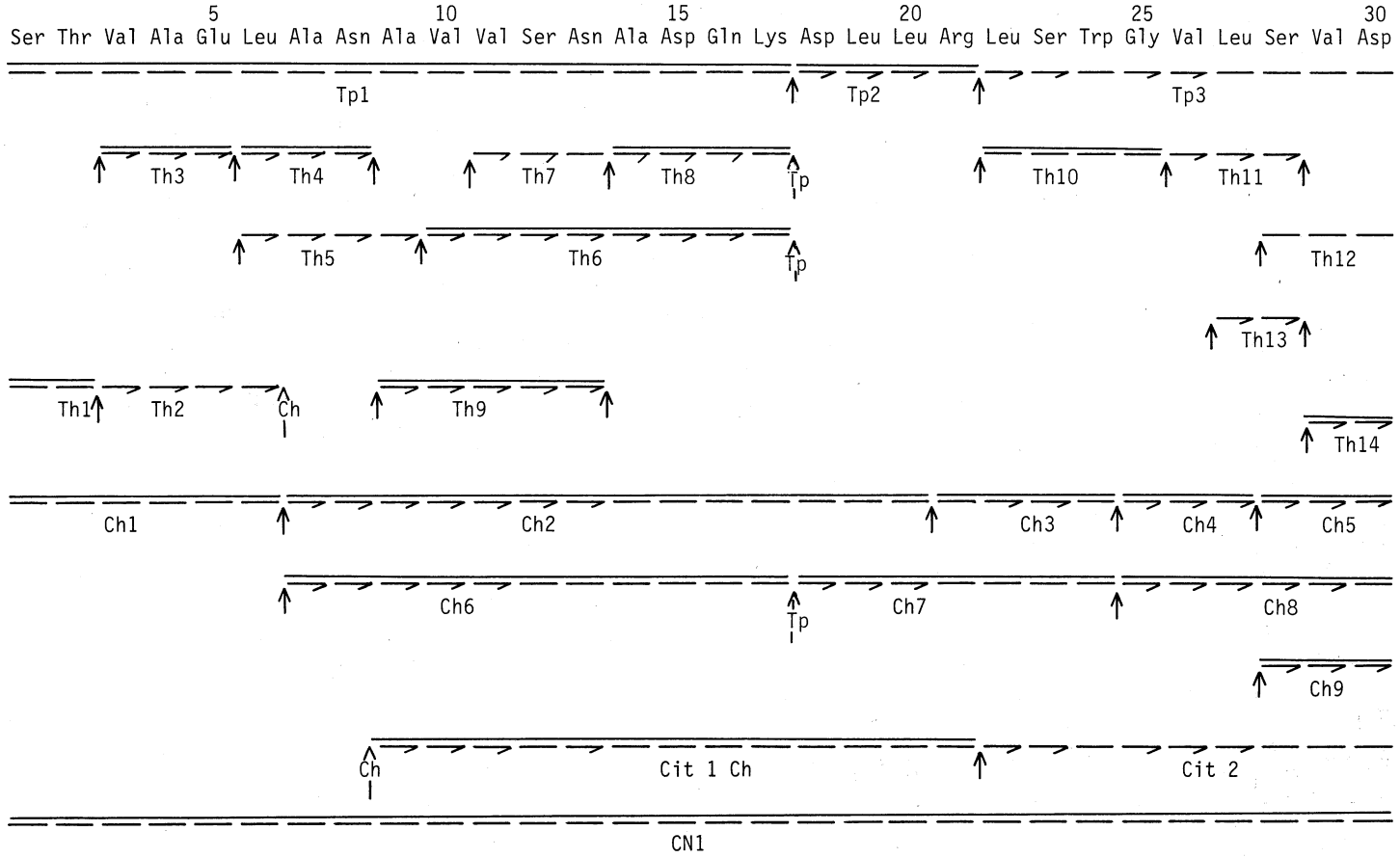
1	2	3	4	5	6
CN4	CN5	CN4	CN1	CN6	CN2
CN5	(CN6)	CN5	(CN4)		CN3

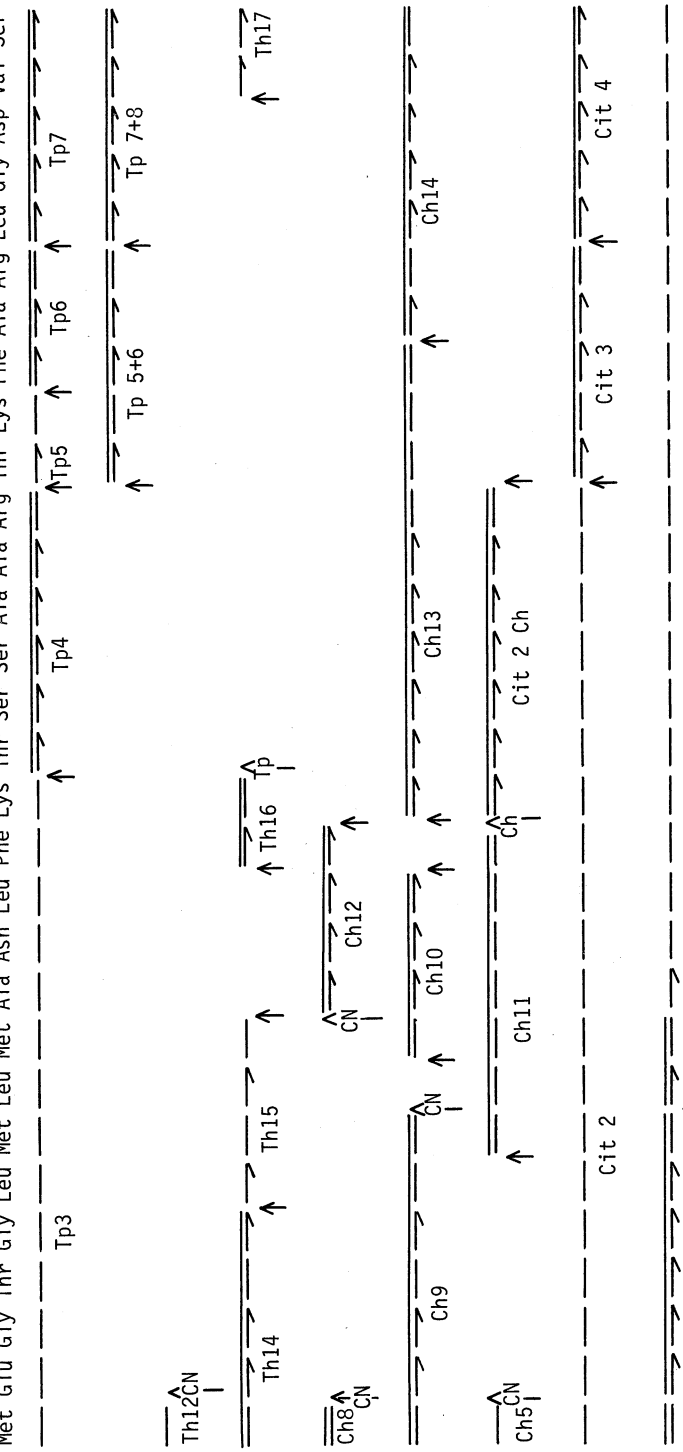
Cit9 (24 residues) which would probably elute together. However, both Cit2 and Cit9 contain tryptophan and so Cit2 was found in a partially resolved peak with Cit5, while Cit9 eluted separately and was identified by its *N*-terminal sequence. The mixed fractions containing Cit2 and Cit5 were digested with chymotrypsin and the peptides purified by peptide mapping.

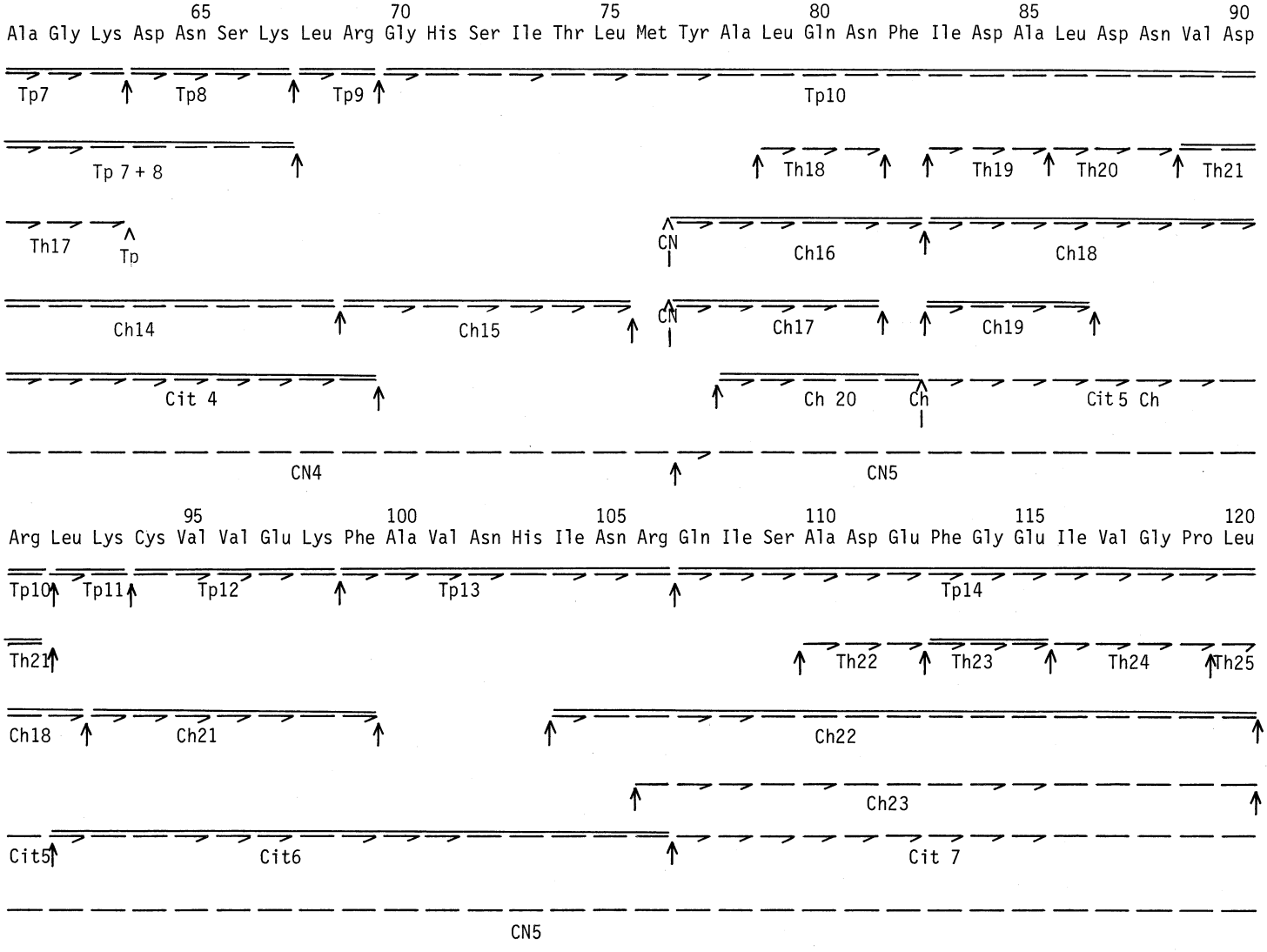
The peptides used in the deduction of the primary structure of the β-globin chain are shown in Table 1. Peptides for which quantitative amino acid analysis figures are available are underlined with an unbroken line. These analyses and the basis for the

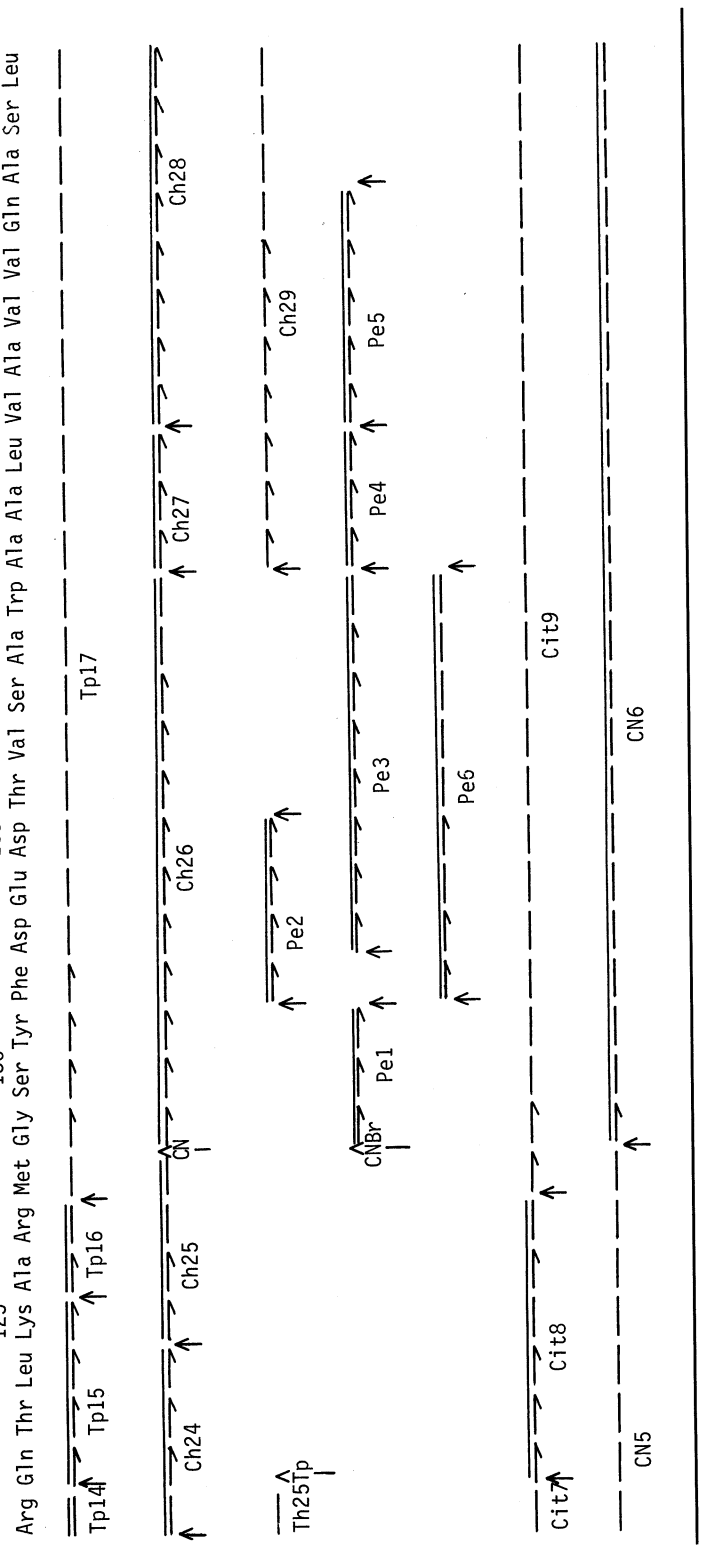
Table 1. Amino acid sequence of the β-globin chain from haemoglobin I of *A. trapezia*

Residues are numbered from the *N*-terminal residue and the positions of enzyme cleavage 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. The abbreviations Tp, Ch, Pe, Th are for trypsin, chymotrypsin, pepsin and thermolysin respectively. CN is the abbreviation for CNBr fragments and Cit is the abbreviation for citraconyl fragments. Peptides which were quantitatively analysed for amino acids are underlined




35
40
45
50
55
60
Met Glu Gly Thr Gly Leu Met Leu Met Ala Asn Leu Phe Lys Thr Ser Ser Ala Ala Arg Thr Lys Phe Ala Arg Leu Gly Asp Val Ser
Tp3
Tp4
Tp5
Tp6
Tp7
Tp 5+6
Tp 7+8
Th12
CN
Th14
Th15
Th16
Tp
Th17
Ch8
CN
Ch12
Ch9
CN
Ch10
Ch13
Ch14
Ch
Ch11
Cit 2 Ch
Ch5
CN
Cit 2
Cit 3
Cit 4
CN1
CN2
CN3
CN4


65 70 75 80 85 90
Ala Gly Lys Asp Asn Ser Lys Leu Arg Gly His Ser Ile Thr Leu Met Tyr Ala Leu Gln Asn Phe Ile Asp Ala Leu Asp Asn Val Asp
Tp7
Tp8
Tp9
Tp10
Tp 7 + 8
Th18
Th19
Th20
Th21
Th17
Tp
CN
Ch16
Ch18
Ch14
Ch15
CN
Ch17
Ch19
Cit 4
Ch 20
Ch
Cit 5 Ch
CN4
CN5
95 100 105 110 115 120
Arg Leu Lys Cys Val Val Glu Lys Phe Ala Val Asn His Ile Asn Arg Gln Ile Ser Ala Asp Glu Phe Gly Glu Ile Val Gly Pro Leu
Tp10
Tp11
Tp12
Tp13
Tp14
Th21
Th22
Th23
Th24
Th25
Ch18
Ch21
Ch22
Ch23
Cit5
Cit6
Cit 7
CN5


125 130 135 140 145 150
Arg Gln Thr Leu Lys Ala Arg Met Gly Ser Tyr Phe Asp Glu Asp Thr Val Ser Ala Trp Ala Ala Leu Val Ala Val Val Gln Ala Ser Leu
Tp14
Tp15
Tp16
Tp17
Ch24
Ch25
CN
Ch26
Ch27
Ch28
Th25Tp
Pe2
Ch29
CNBr
Pe1
Pe3
Pe4
Pe5
Pe6
Cit9
Cit7
Cit8
CN5
CN6

assignment of amide groups to particular residues are included in the supplementary data.*

Acetylated Tp1 was located by normal ninhydrin staining of peptide maps, probably due to a reactive amino group on the lysyl residue. Its sequence was obtained largely from peptides generated after digestion of Tp1 with thermolysin, as indicated in Table 1. Confirmatory and overlap evidence was also provided by chymotryptic peptides from CN1.

Tp3, one of the large, insoluble tryptic peptides, sequenced only for the first five residues. Its sequence was deduced from overlapping chymotryptic peptides derived originally from cyanogen bromide fragments (CN1 and 2) or citraconyl peptides (Cit2). Thermolytic peptides also provided confirmatory evidence. There was some evidence (Ch9) which indicated that the Met-Glu (residues 31–32) bond was not always cleaved. The reason for this is uncertain, but may have been due to oxidation of the methionine involved, since the peptide was obtained from a CNBr cleavage of globin which had not been reduced with thioglycollic acid.

The sequences of Tp10 and Tp17 were deduced from chymotryptic, thermolytic and peptic digests. Confirmatory evidence came from digestions of CNBr and citraconyl fragments. The order of residues 149 and 150 was confirmed by isolation of the peptide Ala-Ser from a subtilopeptidase digest of CN6.

Amino Acid Sequence of the β-chain of Haemoglobin I

The complete amino acid sequence of the β-chain of HbI of the tetrameric haemoglobin of *A. trapezia* is shown in Table 1. Although the order of fragments can be deduced from the peptides and overlap peptides, the strong homology with the α-chain and the dimeric globins of both *A. trapezia* and *A. broughtonii* (Furuta and Kajita 1983; Fisher *et al.* 1984) makes this less essential.

A comparison of the amino acid composition expected from the sequence with that found by amino acid analysis (Como and Thompson 1980*a*) indicated agreement except for one extra residue of alanine and serine and one less residue of aspartic acid in the sequence to give a total of 151 residues in the chain. The amino acid analyses were repeated to see if better agreement for these values was obtained with the samples of globin used for the sequence work. Since any methionine sulfoxide in the hydrolysate can contaminate the aspartic acid peak, some samples were oxidized with performic acid before hydrolysis. After hydrolysis, the aspartic acid, serine and alanine values were 19·4, 13·2 (corrected for decomposition) and 19·0 residues per mole respectively, in better agreement with the sequence data. The other amino acids were in agreement with the sequence except for glycine, which at 10·0 residues was higher than the 9·2 of the previous analyses and the 9 obtained from the sequence. No satisfactory explanation for this high glycine value can be given. Some variation in the glycine and alanine values was expected following the isolation, during peptide purification, of two chymotryptic peptides corresponding to residues 7–19. One of these had glycine *N*-terminal and the other alanine *N*-terminal and the suspected polymorphism at

*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 are as follows: amino acid analyses of underlined peptides; peptide maps for tryptic, chymotryptic, thermolytic, peptic and citraconyltryptic digests of *S*-carboxymethylated β-globin or its fragments, peptide map for the separation of the small CNBr fragments, CN2 and CN3; peptides and their net charge or other evidence for assignment of side-chain amide and acidic groups in β-globin.

residue 7 was supported by the amino acid analyses of each peptide. When such large numbers of molluscs are used to obtain sufficient globin for sequence work, the occurrence of polymorphism in some residues is not unexpected. A rise of glycine values, however, should be accompanied by a fall in alanine, and this was not observed in the total analysis or the analysis of βTp1. We have checked the analyses of smaller peptides carefully, particularly where gaps or insertions have occurred in the sequence when compared with other globin chains. No evidence for an additional glycine residue in the sequence (Table 1) was found.

Discussion

In Table 2, the sequence of the β-chain of the tetrameric haemoglobin of *A. trapezia* is aligned, according to Dayhoff (1972), with that of the α-chain of the tetramer, as well as with the sequence of the dimeric haemoglobins from both *A. trapezia* and *A. broughtonii*. Comparisons of the other three chains with the β-chain, including gaps, deletions or additions, revealed 53% identity of amino acid residues for each chain. There is 100% identity of residues in the F helix of all four chains. The dimeric globin of *A. trapezia* also shows 100% identity with the β-chain in the E helix, while the α-chain shows 75% identity.

The β-chain is 151 amino acid residues in length, which is longer than the 146 residues of mammalian β-chains. It is longer than mammalian α-chains (141 residues) and the dimeric chains of the haemoglobins (146 residues) of *A. trapezia*. The increased size of the β-chain derives from an extension at the *N*-terminal end. An extended *N*-terminal region has only been seen previously in lamprey globin, the α-chains of shark and *A. trapezia* and the dimeric globins from *A. trapezia* and *A. broughtonii* (Fisher *et al.* 1984). It has not been reported previously for any β-globin chain. Preliminary X-ray analysis data of haemoglobins from another mollusc, *Scapharca inaequivalvis* (Royer and Love, personal communication), has indicated an additional helix, pre-A, at the *N*-terminal end of all the chains of the *S. inaequivalvis* globins studied, with some 3–4 residues normally in the A helix incorporated in the pre-A helix. The extended *N*-terminal region of the β-chain of *A. trapezia* found in this study is consistent with the presence of this pre-A helix, particularly since the residues involved are predominantly helical-forming (Chou and Fasman 1978).

The presence of an *N*-terminal blocking group has not previously been reported for a β-chain. Only the α-chains of the shark *Heterodontus portusjacksoni* (Nash *et al.* 1976) and *A. trapezia* and some myoglobins have demonstrated this feature. In vertebrates, the *N*-terminal amino groups of the β-chains are important in the binding of organophosphates. Since acetylation renders such groups unavailable in *A. trapezia*, and since in *A. trapezia* there are β-chain changes in all the basic residues which participate in organophosphate binding in mammalian haemoglobins, then such heterotropic ligands would be absent and the oxygen affinity of the mollusc haemoglobin molecule would be expected to be enhanced.

It is unlikely that the tetrameric haemoglobin of *A. trapezia* would demonstrate a Bohr effect because the β-chain *C*-terminal histidine and the βFG1 aspartic acid residues involved are absent. Also the *N*-terminal of the α-chain, which contributes to the Bohr effect in other species, is blocked. There are changes in all the residues involved in the Bohr effect in mammals.

Table 2. Amino acid sequences of the three globin chains (α-, β-, and IIB) of *A. trapezia* plus that of the IIB chain of *A. broughtonii*

The alignment of sequences follows Dayhoff (1972). Deletions common to the four chains, when compared with the globin chains of many species (Fisher *et al.* 1984), are indicated only as non-consecutive numbers in the interhelical regions and the D helix, where these deletions occur. Dots indicate identity with the β-chain sequence. Gaps are indicated by dashes

														NA	A												A	AB	
												1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2
A. trapezia β	Ac	S	T	V	A	E	L	A	N	A	V	V	S	N	A	D	Q	K	D	L	L	R	L	S	W	G	V	L	-	S	V
A. trapezia α		-	Ac	.	.	D	A	V	A	K	.	C	G	S	E	A	I	.	G	N	.	.	R	.	.	.	.	.	M	.	A
A. trapezia II B		-	-	P	S	V	Q	D	A	.	A	Q	L	T	.	.	V	.	K	D	.	.	D	.	.	K	.	.	G	.	-
A. broughtonii II		-	-	P	S	V	Q	G	A	.	A	Q	L	T	.	.	V	.	K	D	.	.	D	.	.	K	.	I	G	.	-

	B															B	C						C	CD						CD	D
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2	3	4	5	6	7	1	2	3	4	5	6	7	8	1
A. trapezia β	D	M	E	G	T	G	L	M	L	M	A	N	L	F	K	T	S	S	A	A	R	T	K	F	A	R	L	G	D	V	S	A
A. trapezia α	.	I	.	A	.	.	.	T	Y	L	.	.	.	.	T	L	R	P	D	T	K	.	Y	.	T	.	.	.	.	.	Q	K
A. trapezia II B	.	K	K	.	D	.	M	A	.	.	T	T	.	.	N	D	H	Q	E	T	I	A	Y	.	K	.	M	.	.	.	.	Q
A. broughtonii II	.	K	K	.	N	.	V	A	.	.	T	T	.	.	A	D	N	Q	E	T	I	G	Y	.	K	.	.	.	N	.	.	Q

			D	E																	E	EF							
	2	6	7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	1	2	3	4	5	6	7	8	9
A. trapezia β	G	K	D	N	S	K	L	R	G	H	S	I	T	L	M	Y	A	L	Q	N	F	I	D	A	L	D	N	V	D	R	L	K
A. trapezia α	.	.	A	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	D	W	.	V	.	S	.	.	D	P	S	.	.	.
A. trapezia II B	.	M	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	.	.	S	T	.	D	.	I
A. broughtonii II	.	M	A	.	D	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	.	.	N	T	.	D	.	V

		EF	F								F	FG			FG	G													
	10	14	1	2	3	4	5	6	7	8	9	1	2	4	5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A. trapezia β	C	V	V	E	K	F	A	V	N	H	I	N	R	Q	I	S	A	D	E	F	G	E	I	V	G	P	L	R	Q	T	L	K
A. trapezia α	.	.	.	.	.	.	.	.	.	.	.	.	.	K	.	.	G	.	A	.	.	S	.	I	P	E	M	K	E	.	.	.
A. trapezia II B	.	.	.	.	.	.	.	.	.	.	.	T	.	K	.	.	G	A	.	.	.	K	.	N	.	.	M	.	.	V	.	A
A. broughtonii II	.	.	.	.	.	.	.	.	.	.	.	T	.	K	.	.	.	A	.	.	.	K	.	N	.	.	I	K	K	V	.	A

		G	GH				GH	H																				H	HC		
	18	19	1	2	3	4	5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	1	2		
A. trapezia β	A	R	M	G	S	Y	F	D	E	D	T	V	S	A	W	A	A	L	V	A	V	V	Q	A	S	L						
A. trapezia α	.	.	.	.	.	.	-	S	D	.	V	G	A	.	.	V	Q	A	I	L	G	M	.	N	A	V	L	S	A	L		
A. trapezia II B	S	K	-	-	-	N	.	G	D	K	Y	A	N	.	.	.	K	.	.	G	.	.	.	.	A	.							
A. broughtonii II	S	K	-	-	-	N	.	G	D	K	Y	A	N	.	.	.	K	.	.	.	.	.	.	.	A	.							

The β-chain of *A. trapezia* does not possess residues corresponding to a D helix. The absence of a D helix has been a characteristic of only α-chains but has previously been documented in a β-chain for shark (Fisher *et al.* 1977). The dimeric haemoglobins of both *A. trapezia* and *A. broughtonii* also lack the D-helix residues. These observations raise some interesting questions about quaternary haemoglobin structure. In vertebrate globins, the α-chain cannot form stable tetramers, but the β-chain can, although these do not show cooperativity (Dickerson and Geis 1983). The D helix was therefore thought to be important in tetramer formation, but the sequence evidence for the β-chain of shark or *A. trapezia* globin does not support this hypothesis.

Two other characteristics of the β-chain of vertebrate haemoglobin are the presence of histidine at NA2 and the deletion of residue A16. The β-chain of *A. trapezia* does not have a histidine residue at NA2, but it does have the deletion at A16. The dimeric globins of *A. trapezia* and *A. broughtonii* do not have either of these features.

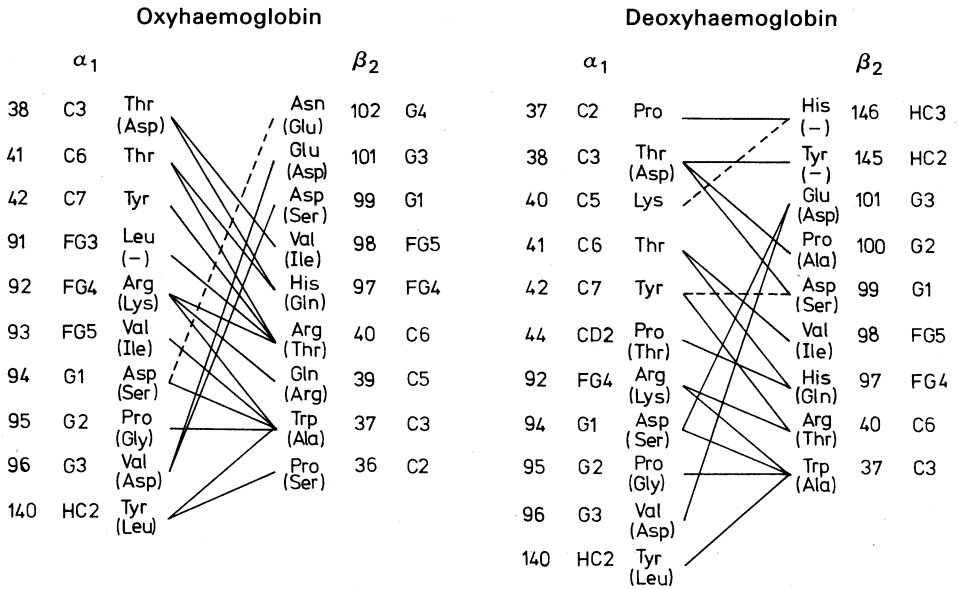


Fig. 4. The $\alpha_1\beta_2$ contacts in oxy- and deoxyhaemoglobin of horse and *A. trapezia*. Residues in horse haemoglobin are shown as given by Bolton and Perutz (1970) and any different residues in *A. trapezia* haemoglobin chains are shown in parentheses.

X-ray studies of haemoglobin from different species have revealed that the molecule has a highly conserved three-dimensional structure in spite of large variability in primary structure. If one assumes that the *A. trapezia* β-chain has the same tertiary structure as that of the horse, then there have been some marked changes in contact sites. There are 16 haem contact sites in horse β-chain (Perutz 1976). The essential CD1 Phe, E7 His and F8 His contacts have been maintained in all the globins of *A. trapezia*, including the β-chain. Most of the other contact sites are either unchanged or have been replaced with similar residues. However, some of the haem contacts have been substituted with markedly different residues and these include C7 Phe to Lys, E10 Lys to Thr, E11 Val to Leu, F7 Leu to Asn and G4 Asn to Glu. The changes at E10, E11 and F7 are common to all the *A. trapezia* chains.

Of the 18 β-chain residues involved in the $\alpha_1\beta_1$ contacts (Perutz 1969), only two (GH2 Gly and GH5 Phe) are identical to those of horse haemoglobin. The $\alpha_1\beta_1$ contact sites are variable in different species and are unchanged during oxygenation and deoxygenation of the haemoglobin molecule.

None of the residues involved in $\alpha_1\beta_2$ contact sites in horse β-globin (Bolton and Perutz 1970) are present in the β-chain of *A. trapezia*. Many of the β-chain residues have been substituted by amino acids with quite different side-chains (see Fig. 4), e.g. C3 Trp to Ala and C6 Arg to Thr. Of the α-chain contact residues, two in the oxy (C6 and C7) and four in the deoxy state (C2, C5, C6 and C7) are identical to those of horse haemoglobin. Most of the other α-chain contact residues in horse haemoglobin have been substituted by quite different residues in the α-globin of *A. trapezia*. The contact sites FG3, HC2 and HC3 are absent. If the $\alpha_1\beta_2$ contact sites involve the same residue positions in *A. trapezia* as in horse, then there are fewer contacts present, by virtue of the deletions occurring in both chains. It is unlikely that many of the sites involved in the $\alpha_1\beta_2$ contacts in horse haemoglobin could fulfil this role in *A. trapezia* haemoglobin because they are markedly different. It is possible that the actual sites of contact between the unlike chains are afforded by residues in different positions to those of horse haemoglobin. Alternatively, there may be compensating changes in the complementary residues that are involved in the contacts. The understanding of these matters will be advanced when the three-dimensional structure of mollusc haemoglobins is available.

Acknowledgments

This work was supported in part by the Australian Research Grants Scheme. We thank Dr W. K. Glenn for helpful advice, and Mr R. G. Mann for amino acid analyses, Dr A. M. Duffield for the mass spectrographic analyses and Mr S. S. Moore for identification of phenylthiohydantoins.

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.

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

Chou, P. Y., and Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**, 45–148.

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.)

Dickerson, R. E., and Geis, I. (1983). 'Hemoglobin: Structure, Function, Evolution and Pathology.' (Benjamin/Cummings Publ. Co.: Menlo Park, California.)

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., Gilbert, A. T., and Thompson, E. O. P. (1984). Amino acid sequence of the globin IIb chain of the dimeric haemoglobin of the bivalve mollusc *Anadara trapezia*. *Aust. J. Biol. Sci.* **37**, 191–203.

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.
Fisher, W. K., and Thompson, E. O. P. (1983). Amino acid sequence studies on sheep liver fructose-bisphosphatase. II. The complete sequence. *Aust. J. Biol. Sci.* **36**, 235–50.
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.
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.
Henschen, A., and Lottspeich, F. (1975). Amino acid sequence of human fibrin. *Z. Physiol. Chem.* **356**, 1985–8.
Houghton, R. A., and Li, C. H. (1983). Reduction of sulfoxides in peptides and proteins. In 'Methods in Enzymology'. (Eds C. H. W. Hirs and S. N. Timasheff.) Vol. 91. pp. 549–59. (Academic Press: New York.)
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. (1969). The haemoglobin molecule. *Proc. R. Soc.* B**173**, 113–40.
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., Hosken, R., and Air, G. M. (1969). Studies of marsupial proteins. I. Polymorphism of haemoglobin of the grey kangaroo, *Macropus giganteus*. *Aust. J. Biol. Sci.* **22**, 449–62.
White, F. H. (1960). Regeneration of enzymatic activity by air-oxidation of reduced ribonuclease with observations on thiolation during reduction with thioglycolate. *J. Biol. Chem.* **235**, 383–9.

Manuscript received 29 March 1985, accepted 5 June 1985

Note Added in Proof

Since the submission of this paper we have received a copy of a manuscript entitled 'The cooperative dimeric and tetrameric chain hemoglobins are novel assemblages of myoglobin folds' by W. E. Royer, W. E. Love and F. F. Fenderson [*Nature (London)*, **316**, 277–80]. The X-ray crystallographic structure of the haemoglobins from *Scapharca inaequivalvis* indicated an additional helix involving the amino-terminal extension as well as a different set of subunit contacts. The contacts involve the E and F helices in contrast to vertebrate haemoglobins, where the $\alpha_1\beta_2$ contacts involve mainly helices C and G, and the FG corner. This would explain the strong conservation of the E and F helical residues (Table 2) and the lack of conservation of the $\alpha_1\beta_2$ residues (Fig. 4).

The crystals of the haemoglobins from *S. inaequivalvis* were prepared by E. Chiancone and co-workers, who have recently completed the amino acid sequence of its dimeric chain, which shows 15 differences with the dimeric chain of *A. trapezia*. We wish to thank Professor Chiancone for a copy of a manuscript by R. Petruzzeli, B. M. Goffredo, D. Barra, F. Bossa, A. Boffi, D. Verzili, F. Ascoli and E. Chiancone entitled 'Amino acid sequence of the cooperative homodimeric hemoglobin from the mollusc *Scapharca inaequivalvis* and topology of the intersubunit contacts'. (*FEBS Lett.*, in press).