Aust. J. Biol. Sci., 1977, **30**, 1–19

α-Chain Sequence of Newt Haemoglobin (*Taricha granulosa*)

Michael Coates,[A] *B. Brimhall,*[B] *Peter Stenzel,*[B] *Mark Hermodson,*[C] *David Gibson,*[D] *R. T. Jones*[B] *and Thomas Vedvick*[E]

[A] Department of Zoology, University of Adelaide, G.P.O. Box 498D, Adelaide, S.A. 5001.
[B] Department of Biochemistry, University of Oregon Medical School, 3181 S.W. Sam Jackson Park Road, Portland, Oregon, 97201, U.S.A.
[C] Division of Medical Genetics, University of Washington Medical School, Seattle, Washington, 98195, U.S.A.
[D] Department de Biochimie, Centre Hospitalier Universitaire, Sherbrooke, Quebec, Canada.
[E] Department of Pathology, University of California, La Jolla, California, 92037, U.S.A.

Abstract

The amino acid sequence of the α-chain of the major haemoglobin of a newt, *T. granulosa*, has been determined. The chain is 142 residues long and has an extra methionine at its *N*-terminus when compared with human α-chain. Most of the tryptic peptides were sequenced by a combination of the subtractive Edman method and by deduction from the compositions of overlapping fragments produced by various enzymic treatments. The sequence of two 'core' regions was obtained by automatic sequencing of large peptides produced by trypsin cleavage at arginine residues only after blockage of lysine residues by citraconylation; by cleavage between aspartic acid and proline residues with 70% formic acid, and by cyanogen bromide cleavage at methionine residues.

The sequence of *T. granulosa* α-chain is compared with those of representative species from the other classes of vertebrates. The differences in α-chain between the classes of vertebrates are compared with the differences in this protein between an equal number of orders of mammals. This comparison allows us to conclude that the major functional and conformational features of α-chain have been conserved since the divergence of the classes of jawed vertebrates.

Introduction

This paper gives the complete amino acid sequence of the α-chain of the major haemoglobin component from a newt, *Taricha granulosa*. With the completion of this sequence it is now possible to compare α-chain from all but one of the extant classes of vertebrates. Complete α-chain amino acid sequences have been determined for a number of eutherian mammals (Dayhoff 1972), as well as for two marsupials (Mammalia) (Beard and Thompson 1971; Stenzel 1974), two monotremes (Mammalia) (Whittaker *et al.* 1973; Whittaker and Thompson 1974), a bird (Aves) (Matsuda *et al.* 1972), a viper (Reptilia) (Duguet *et al.* 1974), a newt (Amphibia) (this publication), two bony fishes (Actinopterygii) (Hilse and Braunitzer 1968; Powers and Edmundson 1972), and a shark (Elasmobranchi) (Nash *et al.* 1976). Amino acid sequences have not been determined for the haemoglobins of any of the lungfishes or the Coelacanth (Crossopterygii) and the haemoglobins of lampreys (Agnatha) can not be identified as either α- or β-chains (Braunitzer and Fujuki 1969; Li and Riggs 1970). Comparisons among the classes of vertebrates should give additional information about the evolution of haemoglobin, or at least serve as a check on the generality of conclusions about the evolution of α-chain in particular, which at present are

based for the most part on comparisons of mammalian haemoglobins (Dayhoff 1972; Goodman *et al.* 1975).

Furthermore, inspection of the sequence of α-chain of newt may permit a partial explanation, at the structural level, of the low Bohr effect observed in this animal (Coates and Metcalfe 1971) and in other newts and salamanders (Sullivan 1974).

Methods

Preparation of Haemolysate

In April or May of four consecutive years, newts (*T. granulosa*) were collected in the Nehalem River area of the Oregon coast, U.S.A.

Sodium heparin (0·25 ml, 1000 i.u./ml) was injected into the body cavity of each newt. After 30 min the heads were cut off and the blood drained into small beakers containing buffered physiological saline to which EDTA and heparin had been added. The red cells were centrifuged down and washed twice with physiological saline.

Physiological saline containing 1% saponin was added to the washed red cells. After 15 min the mixture was centrifuged at 8000 rev/min. The clear haemolysate was dialysed against a buffer appropriate for the ensuing chromatography. [If water alone is used for haemolysis, the nuclei are lysed and a thick gel (probably a nucleic acid–haemoglobin complex) is formed from which it is difficult to separate the haemoglobin.] A haemoglobin yield of approximately 20 mg per newt was obtained. Friedman (1974) has found the haemoglobin concentration in the blood of *T. granulosa* from southern Vancouver Island to be approximately 8·5 g/100 ml.

Electrophoresis of the haemolysate, and of the various chromatographic zones of haemoglobin, was carried out by the starch gel method of Huisman (1963) using a tris–EDTA–borate buffer at pH 8·3.

Globin Preparation and Chain Separation

Globin was prepared by precipitation with cold acid–acetone (Anson and Mirsky 1930). The α-chain could be separated from the β-chains by chromatography on carboxymethyl (CM)-cellulose using a linear gradient of two pyridine–formate buffers (Dintzis 1961).

Electrophoresis of the globin chains, before and after chromatography, was carried out in urea and mercaptoethanol by the method of Schneider (1974). After chromatography the chains were aminoethylated with ethylene imine in 8 M urea (Jones 1964). Determination of the *N*-terminal residue of the α-chain was made by the Stark cyanate method (Stark 1972).

Formation and Purification of Tryptic Peptides

In order to obtain the tryptic peptides, the entire aminoethylated (AE) α-chain, or fragments of the AE α-chain, were hydrolysed with trypsin (Worthington, TPCK) in a tetramethyl ammonium hydroxide-buffered solution at pH 8–9 at room temperature using 1 part by weight of trypsin to 100 parts protein for 3 h (Baglioni 1965).

Final purification of the tryptic peptides and of smaller peptides derived from them was carried out as described by Jones (1970), using either a 0·9 by 15 cm or a 0·9 by 35 cm column of Aminex A5 resin for the initial separation, followed by a 0·9 by 60 cm column of Aminex AG50W-X2 (200–400 mesh, Bio-Rad Laboratories) or a 0·9 by 16 cm column of Aminex AG50W-X4 for rechromatography of each zone. A linear gradient of two pyridine–acetate buffers (0·2 M, pH 3·1, and 2·0 M, pH 5·0) at a flow rate of 30 ml/h was used for these columns. In some cases a 0·6 by 60 cm column of anion-exchange resin, Aminex AG1-X2, was used with the same apparatus except that the eluting buffers were 1% collidine–acetate and 0·1–0·5 M acetic acid as described by Guest *et al.* (1967).

Amino Acid Analysis

For amino acid analysis, each peptide, or whole α-chain, was hydrolysed in an evacuated ampoule in 6 M HCl containing 9 mg phenol/100 ml for 22–72 h at 110°C and run on a Beckman–Spinco Model 120A analyser.

Sequence Determination of Peptides

The substractive Edman method (Konigsberg 1972) was used in the manual determination of the amino acid sequences of the tryptic peptides. For automatic Edman degradation in the region from residues 42 to 90, the Beckman Sequencer Model 890 C was employed according to the method of Hermodson *et al.* (1972). For the small peptide (residues 65–93, CB-2) a peptide program was employed. This program is identical to the usual program except for two additional procedures. First, after coupling and vacuum drying of the buffer solution, enough benzene was introduced into the sequenator cup to just cover the peptide film. A delay of 60 s was programmed at this stage to allow time for the benzene to precipitate the peptide after which the usual extraction with benzene proceeded. Seconds after cleavage and vacuum drying of the acid, enough chlorobutane to just cover the peptide film was introduced and was completely evaporated under vacuum to leave a dried film on the wall of the cup. The usual extraction of the product with chlorobutane then followed. These modifications in the program greatly reduce extractive losses of small peptides, allowing degradation to within 3–5 residues of the carboxyl terminus of most peptides. For automatic Edman degradation in the T-14 region, the Illitron Model 9001 Sequencer was used as described by Gibson (1974).

Enzymic digestions of the tryptic peptides were carried out by several conventional methods using thermolysin (Matsubara *et al.* 1965), carboxypeptidases A and B (Ambler 1972), and chymotrypsin and papain (Smyth 1967). Subtilisin BPN′ was used at pH 8·2 and 37°C for 2 h. To distinguish between aspartic acid and asparagine and glutamic acid and glutamine, aminopeptidase (microsomal) was used by the method of Light (1972). For cleavage at aspartic acid residues, a solution of the tryptic peptide in 0·5 M acetic acid was hydrolysed overnight in an evacuated ampoule at 110°C.

Formation and Purification of Chymotryptic Peptides

The AE α-chain was digested with chymotrypsin overnight at room temperature (1 part enzyme to 100 parts protein) at pH 8·5. The insoluble portion was removed by centrifigation and the supernatant adjusted to pH 3 and rotary evaporated to about 2 ml. The peptides were separated by chromatography on Aminex 50W-X2 and purified, in most cases, by rechromatography on the anion-exchange resin Aminex 1-X2 as described above.

Formation and Purification of Large Peptides

In order to cleave the α-chain at the arginine residues only, the lysines were blocked by citraconylation, the α-chain digested with trypsin and the lysines then deblocked by the method of Singhai and Atassi (1971).

For cleavage at methionyl residues, the large arginyl peptide (residues 42–142) or the entire AE α-chain, was first reduced by treatment for 2 days at 37°C in a solution of 30% 2-mercaptoethanol–5 M guanidine hydrochloride–0·1 M NH_4HCO_3 (Weihung and Korn 1972). The reduced peptide was desalted by passage through Sephadex G25 in 0·2 M acetic acid. Then to each 100 mg of lyophilized peptide was added 1 ml of 70% formic acid and 100 mg of solid cyanogen bromide. After 24 h at room temperature the solution was diluted 10-fold with water and applied to a column of Sephadex G50 for fractionation.

The method of Aksoy *et al.* (1972) was used to split the α-chain between aspartic acid and proline residues. 100 mg of lyophilized α-chain was dissolved in 1 ml of 70% formic acid containing 40 mg of 2-mercaptoethanol. This solution was incubated at 50°C for 24 h. It was then diluted 10-fold with water and applied to a Sephadex G50 column for fractionation.

The large peptides produced by the above methods were fractionated by passage through a 5 by 200 cm column of Sephadex G50 Fine in 7% formic acid (Gross 1967). The effluent from this column was monitored continuously at 280 nm.

Results

Purification of α-Chain

Starch gel electrophoresis of the haemolysate at pH 8·6 showed three bands: a small band in the position of human Hb F, a small band in the position of human Hb A, and a large band somewhat faster than Hb A (Fig. 1). We could achieve no clear-cut separation of these haemoglobins on chromatography with the usual resins

and buffers. On DEAE-Sephadex with tris–HCl buffers (Huisman and Dozy 1965) the haemoglobin was completely retarded. On Amberlite IRC-50 (Jones and Schroeder 1963) it was all eluted at the front except when Developer No. 3 was used starting at 5°C with stepwise increase in temperature. Two zones were obtained by this procedure but they showed little difference by gel electrophoresis. CM-Sephadex and CM-cellulose with varying gradients of phosphate buffers gave only slight separation.

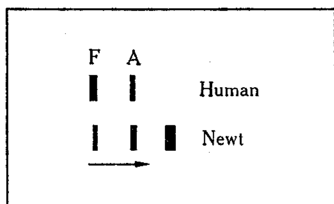


Fig. 1. Diagram of starch gel electrophoresis at pH 8·6 of newt haemolysate compared with human haemoglobins A and F.

Dintzis' (1961) method of chain separation of the globin gave three zones (Fig. 2). Zone I contained relatively pure α-chain. Electrophoresis in urea–mercaptoethanol showed a single band for this zone. It was concluded that all of the haemoglobins had the same α-chain. Both zones II and III contained mixtures of β-chains. The amino acid composition of the α-chain as indicated by analysis of the whole chain and by adding the residues of the tryptic peptides is given in Table 1.

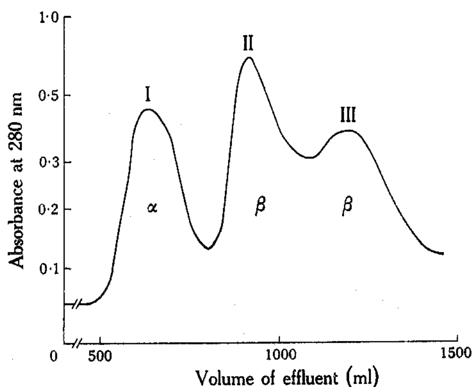


Fig. 2. Chain separation of newt globin on CM-cellulose with pyridine–formate buffers.

The *N*-terminal amino acid residue of the α-chain was shown by the Stark procedure to be methionine. No amino acid residue could be recovered when the same procedure was applied to the β-chain components, indicating that the β-chains have blocked *N*-termini.

Composition of Tryptic Peptides

Tryptic hydrolysis of the AE α-chain and fractionation of the resulting peptides produced 18 different tryptic peptides. Rechromatography resulted in purification of all but one of the tryptic peptides (T-14). The amino acid compositions of the AE α-chain tryptic peptides are given in Table 2. The composition of T-14 was obtained during sequencing.

Table 1. Amino acid analyses of *T. granulosa* AE α-chain and overlapping peptides

A, arginyl peptide; T, tryptic peptide; Ch, chymotryptic peptide

Amino acid	AE α-chain		Arginyl peptides				Tryptic peptide T-2,3	Chymotryptic peptides			
	Found	Sum of peptides	A-1 (res. 1–32)		A-3 (res. 42–142)			Ch-3,4	Ch-4,5	Ch-16,17	Ch-17,18
			F[A]	T[B]	F	T	F	F	F	F	F
Lys	12·9	13	4·0	4	9·0	9	2·0	1·0	1·0	1·1	1·2
His	11·7	12	2·9	3	8·3	9	1·0		2·1		
AE-Cys	2·3	2			2·1	2					
Arg	3·0	3	1·1	1	1·1	1					
Trp	tr[C]	1	0·2	1				0·1			
Asp	15·1	15	3·0	3	12·0	12	1·9	1·0	1·2	1·0	
Thr	6·8	7	1·8	2	2·6	3		1·6			1·0
Ser	10·0	10	1·2	1	7·8	8	0·9			1·0	1·0
Glu	6·5	6	3·6	4	2·2	2	1·0		2·0		
Pro	4·9	5			4·0	4					
Gly	7·1	7	2·0	2	4·9	5			1·2		
Ala	15·0	15	3·9	4	10·1	10	1·1		1·0	1·0	
Val	10·9	11	1·3	1	9·5	10	0·8	1·0			
Met	4·5	5	0·4	1	2·6	3					
Ile	4·3	5	0·8	1	4·1	4			0·7		
Leu	16·0	16	3·4	3	12·0	12	1·1		1·1		
Tyr	3·0	3			2·6	3					0·8
Phe	6·2	6	1·1	1	4·0	4				0·9	
Total	140·2	142·0	30·7	32	98·9	101	9·8	4·7	10·3	5·0	4·0

[A] F = ratio of amino acid residues found on amino acid analysis.
[B] T = theoretical value of amino acid residues.
[C] Trace.

Table 2. Amino acid composition of tryptic peptides from *T. granulosa* α-chain

Values given are molar ratios of amino acids produced by hydrolysis with hydrochloric acid. Values of 0·2 or less have been omitted

Amino acid	T-1 1–2[A]	T-2 3–8	T-3 9–12	T-4 13–19	T-5 20–32	T-6 33–41	T-7 42–7	T-8 48–61	T-9 62	T-10 63–82[B]	T-11 83	T-12 84–7	T-13 88–101	T-14 102–15[B]	T-15 116–24	T-16 125–8	T-17 129–40	T-18 141–2
Lys	1·0	1·0	0·9	1·0			1·0	1·1	1·0		1·0	1·0	1·2	1		1·1	1·2	
His			1·1	1·0	1·0			2·0		1·1			0·9	2	2·0		0·9	
AE-Cys										0·9					1·0			
Try				0·1[C]														
Arg					0·9	1·0												1·0
Asp		1·0	1·1	1·1				1·0		3·9		1·0	3·0	1		1·1	1·1	
Thr				2·1		2·0	1·0								1·0		1·0	
Ser		0·9				1·0		3·8		1·2		1·0				0·9	1·0	
Glu		1·0			3·0			1·1					0·9					
Pro						1·0	1·0						2·2		0·9			
Gly					2·2			2·0		1·2				1			1·1	
Ala		1·0			3·2	1·0	1·0			4·6			2·1	1	1·0			
Val			0·8							2·0			0·9	2	1·0	1·0	2·7	
Met	1·0					1·0				0·6			0·7	1				
Ile				1·0						2·1				2				
Leu		1·0			2·0	1·0		2·0		2·1		1·0	1·0	3	1·0		1·9	
Tyr							1·0								1·0			1·0
Phe					1·0	1·0	1·0	1·0					1·1				0·9	
Total	2·0	5·9		6·3	13·1	9·0	6·0	14·0	1·0	19·7	1·0	4·0	14·1	14·0	8·9	4·1	11·8	2·0

[A] Residue numbers in chain. One residue extra at *N*-terminus as compared with human α-chain.
[B] As determined by sequenator. [C] Much lost on acid hydrolysis.

Sequencing of Tryptic Peptides

The sequencing of the tryptic peptides is detailed in Table 3. Most of the tryptic peptides were sequenced manually by stepwise removal of amino acid residues from the *N*-terminus by the Edman method or from the *C*-terminus by carboxypeptidase, and by cleaving the peptides by enzymic or chemical methods, and deducing the sequence from the amino acid compositions of overlapping peptides.

However, two of the tryptic peptides, T-10 and T-14, could not be obtained in amounts sufficient for manual sequencing due to their insolubility in the various aqueous chromatography buffers used. The sequences of these peptides were determined through partial sequencing, by the automatic sequenator, of larger fragments of the α-chain.

Formation and Purification of Large Peptides

In order to facilitate the sequencing of the residues contained within T-10 and T-14, and to help in the ordering of the tryptic peptides, the AE α-chain was fragmented into large peptides by several chemical methods.

Fig. 3 shows the separation on Sephadex G50 of the three arginyl peptides resulting from citraconylation and tryptic hydrolysis before deblocking of the lysine and AE cysteine residues. Peak III corresponded to the T-6 tryptic peptide and this peptide was subsequently named A-2. Peak II, peptide A-1, when further digested with trypsin yielded the tryptic peptides T-1, T-2, T-3, T-4, and T-5. A third peptide, A-3, was in peak I. Tryptic peptides T-7 to T-18 were isolated from this fragment, with the exception of T-14 which was again lost due to insolubility. The separation by column chromatography of the tryptic peptides from A-1 and A-3 are shown in Figs 4 and 5 respectively. The amino acid compositions of A-1 and A-3 are given in Table 1.

The arginyl peptide A-3 was further broken down by cleavage at its three methionyl residues with cyanogen bromide, giving peptides CB-1 to CB-4. Passage through the Sephadex G50 column gave five zones, three of which contained mixtures of incompletely cleaved peptides. A fourth zone contained peptides CB-2 and CB-4. These were separated by recycling five times through the Sephadex column. Likewise, a fifth zone contained CB-1 and CB-3. After recycling five times, CB-1 was recovered but CB-3 had disappeared. Its composition was deduced by subtracting the tryptic peptides of CB-1 from those of the zone containing both CB-1 and CB-3 before recycling.

Treatment of the AE α-chain with 70% formic acid, which cleaves proteins between aspartic acid and proline, resulted in two peptides, Asp-Pro-1 and Asp-Pro-2, which were separated on Sephadex G50 (Fig. 6). Tryptic hydrolysis of the Asp-Pro-2 peptide yielded tryptic peptides T-15, T-16, T-17, T-18, and part of T-13 (Pro, Ala, Asn, Phe, Pro, Lys). T-14 was again lost due to insolubility.

Sequencing of T-10 and T-14

Automatic sequencing gave the first 26 residues of the arginyl peptide A-3. These had the same sequence as T-7, followed by T-8, followed by Lys (which is the same as T-9), followed by the sequence Val-Met-Gly-Ala-Leu. Automatic sequencing of the cyanogen bromide peptide CB-2 gave the sequence Gly-Ala-Leu followed by the rest of the T-10 sequence, followed by Lys (which is the same as T-11), followed by T-12, followed by the first three residues of T-13.

Table 3. Sequence of tryptic peptides from *T. granulosa* α-chain

Fragments of tryptic peptides pertinent to sequencing are indicated by lines under each tryptic peptide. The lines are labelled with abbreviations of the method used in producing the fragment. Ac, 0·5 M acetic acid; ApM, aminopeptidase (microsomal); CB, cyanogen bromide; Ch, chymotrypsin; Hse, homoserine; Pa, papain; S, subtilisin; T, tryptic peptide; Th, thermolysin; — residue removed by carboxypeptidase; — residue removed by Edman degradation. The numbers of the first and last resides in each peptide are shown

No.	Sequence
T-1	1 2 Met-Lys Met is *N*-terminus of α-chain.
T-2	3 8 Leu-Ser-Ala-Glu-Asp-Lys → → → Ac ApM shows Glu and Asp.
T-3	9 12 His-Asn-Val-Lys → → ApM shows Asn.
T-4	13 19 Thr-Thr-Trp-Asp-His-Ile-Lys Th-1 Th-2 Ac-1 Ac-2 Ac-3 Ch-1 Ch-2 ApM shows Asp and Trp.
T-5	20 32 Gly-His-Glu-Glu-Ala-Leu-Gly-Ala-Glu-Ala-Leu-Phe-Arg Th-1 Th-2 Th-4 → → → → → → S Ch ← → Pa ApM shows 3 Glu.
T-6	33 41 Met-Phe-Thr-Ser-Leu-Pro-Ala-Thr-Arg ← ← Ch-1 Ch-2 → → → → →

Table 3 (*continued*)

No.	Sequence
T-7	42 47 Thr–Tyr–Phe–Pro–Ala–Lys Th-1 Th-2 Ch-1 Ch-2 Automatic sequenator on A-3 confirms T-7, T-8 and T-9 as written.
T-8	48 61 Asp–Leu–Ser–Glu–Gly–Ser–Ser–Phe–Leu–His–Ser–His–Gly–Lys Ch-1 Ch-2 Ch-4 Pa-1 Pa-2 Pa-3 Ch-3 ApM shows Glu and Asp.
T-9	62 Lys
T-10	63 82 Val–Met–Gly–Ala–Leu–Ser–Asn–Ala–Ala–Val–Ala–His–Ile–Asp–Asp–Ile–Asp–Ala–Leu–Cys The automatic sequenator on A-3 and CB-2 shows sequence as written.
T-11	83 Lys
T-12	84 87 Leu–Ser–Asp–Lys Ac-1 Ac-2 Ac-3 ApM shows Asp.
T-13	88 101 His–Ala–Gln–Asp–Leu–Met–Val–Asp–Pro–Ala–Asn–Phe–Pro–Lys Th-1 Th-2 CB-2+trypsin Hse Ac-1 Ac-2 CB-2+Ac-1 CB-2+Ac-2 ApM on Th-1 shows Gln and Asp; ApM on Ac-2 shows Asn (res. 98). Automatic sequenator on Asp-Pro-2 (see Table 4):

Table 3 (*continued*)

No.	Sequence
T-14	102 115 Leu-Ala-His-Asn-Ile-Leu-Val-Val-Met-Gly-Ile-His-Leu-Lys Ch-4 + trypsin → → → Automatic sequenator on Asp-Pro-2: → → → → → → → → → → →
T-15	116 124 Ala-His-Leu-Thr-Tyr-Pro-Val-His-Cys ← ← Th-1 Th-2 Th-3 Ch-1 Ch-2 → → → →
T-16	125 128 Ser-Val-Asp-Lys → → ApM shows Asp.
T-17	129 140 Phe-Leu-Asp-Val-Val-Gly-His-Val-Leu-Thr-Ser-Lys Th-1 Th-2 → → Ac-1 Ac-2 Ac-3 → → → Ch → ApM shows Asp.
T-18	141 142 Tyr-Arg →

Automatic sequencing of the peptide Asp-Pro-2 gave the sequence of the last six residues of T-13, followed by the first 11 residues of T-14, which ended in the sequence -Met-Gly-Ile. Tryptic hydrolysis of the cyanogen bromide peptide CB-4 yielded tryptic peptides T-15, T-16, T-17, T-18 and a peptide which on manual sequencing proved to have the sequence Gly-Ile-His-Leu-Lys. These then are the last four residues of T-14 (see Table 3).

Ordering of the Tryptic Peptides

The sequence of newt α-chain was then completed by ordering the sequenced tryptic peptides. The proposed sequence is given in Table 4. The rationale behind the ordering is as follows: the three large arginyl peptides, A-1, A-2 and A-3, are ordered

first. A-1 contains the tryptic peptides T-1, T-2, T-3, T-4 and T-5; A-2 is T-6; and A-3 contains T-7 to T-18 (see Table 4). Cyanogen bromide cleavage of the whole AE α-chain yielded the overlapping peptide between A-2 and A-3 labelled CB-5 in Table 4. This peptide contained all but the *N*-terminal Met of T-6, T-7 to T-9 and the first two residues of T-10. This showed that A-2 is joined to the *N*-terminus of A-3. Both A-1 and A-2 have Met at the *N*-terminus, thus either could be the *N*-terminal peptide of the chain. However, tryptic hydrolysis of the peptide Asp-Pro-2 yielded tryptic peptides associated with A-3 only, thus A-1 can not be at the *C*-terminus and must be at the *N*-terminus of the chain (see Table 4).

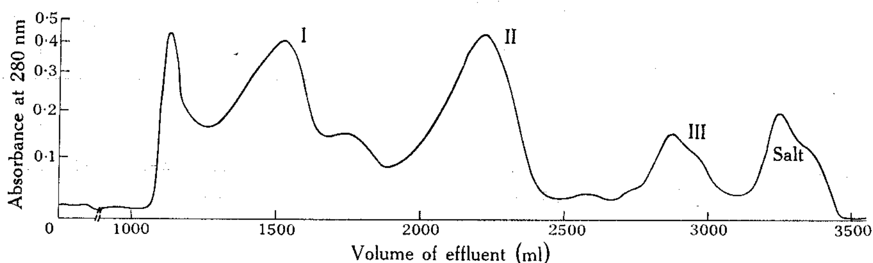


Fig. 3. Separation on Sephadex G50 of fragments from cleavage of newt AE α-chain at arginyl residues. Peak I = peptide A-3, peak II = peptide A-1, peak III = peptide A-2.

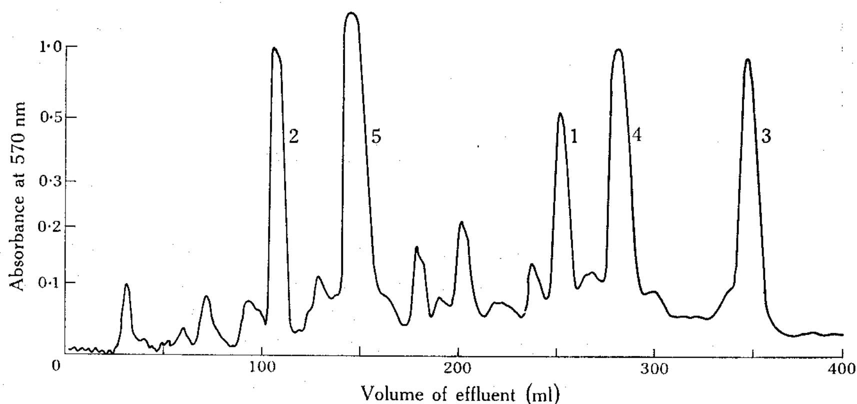


Fig. 4. Separation of tryptic peptides from arginyl peptide A-1 (residues 1–32) on A5 resin using pyridine–acetate buffers, followed by ninhydrin reaction with peptides. Numbers (1–5) refer to tryptic peptides (T1–T5) found in peaks.

Of the tryptic peptides within A-1, T-1 (Met-Lys) is the only one with methionine and therefore is the *N*-terminal tryptic peptide of the newt α-chain. T-5 is the only one with arginine and is therefore the *C*-terminal tryptic peptide of A-1. The overlapping chymotryptic peptides Ch-3,4 and Ch-4,5 order T-3 and T-4 (see Tables 1 and 4). Finally the discovery of a joined-together tryptic peptide T-2,3 completed the ordering of the tryptic peptides in A-1 (Table 4).

Tryptic peptide T-6, which is A-2, follows T-5. Two automatic sequenator degradations of peptide A-3 showed the order of tryptic peptides between residues 42 and 80 to be T-7–T-8–T-9–T-10 (Tables 3 and 4). Tryptic digestion of the peptide CB-2 yielded the tryptic peptides T-10 without Val-Met, the same peptide with Lys attached to its *C*-terminus, T-12, and part of T-13 up to its Met residue, thus ordering

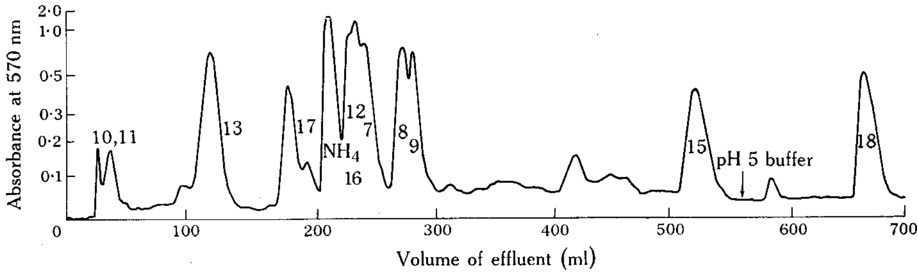


Fig. 5. Separation of tryptic peptides from arginyl peptide A-3 (residues 42–142) as for Fig. 6. Numbers (7–18) again refer to tryptic peptides (T7–T18) found in peaks.

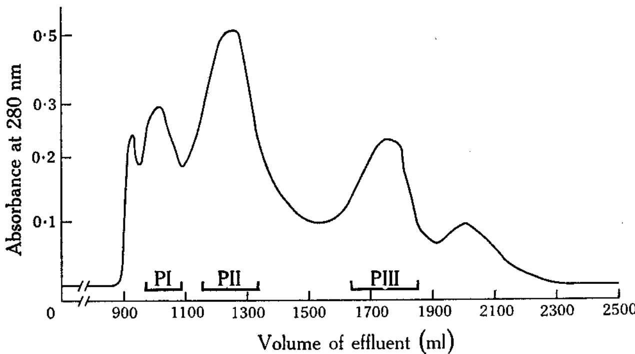


Fig. 6. Separation on Sephadex G50 of fragments from cleavage of newt AE α-chain by 70% formic acid at Asp-Pro linkage.
PI = uncleaved AE α-chain,
PII = peptide Asp-Pro-1,
PIII = peptide Asp-Pro-2.

the tryptic peptides up to T-13 (Table 4). Starting with the peptide Asp-Pro-2, the automatic sequenator has also sequenced the residues contained in CB-3, showing it to contain part of T-13 and part of T-14. CB-4 contained the rest of T-14 at its *N*-terminus. At its *C*-terminus is placed Tyr-Arg (T-18) since this is the only arginine in A-3. Peptides T-16 and T-17 were placed by the overlapping chymotryptic peptides Ch-16,17 and Ch-17,18 (Tables 1 and 4). This leaves T-15 which was placed between T-14 and T-16.

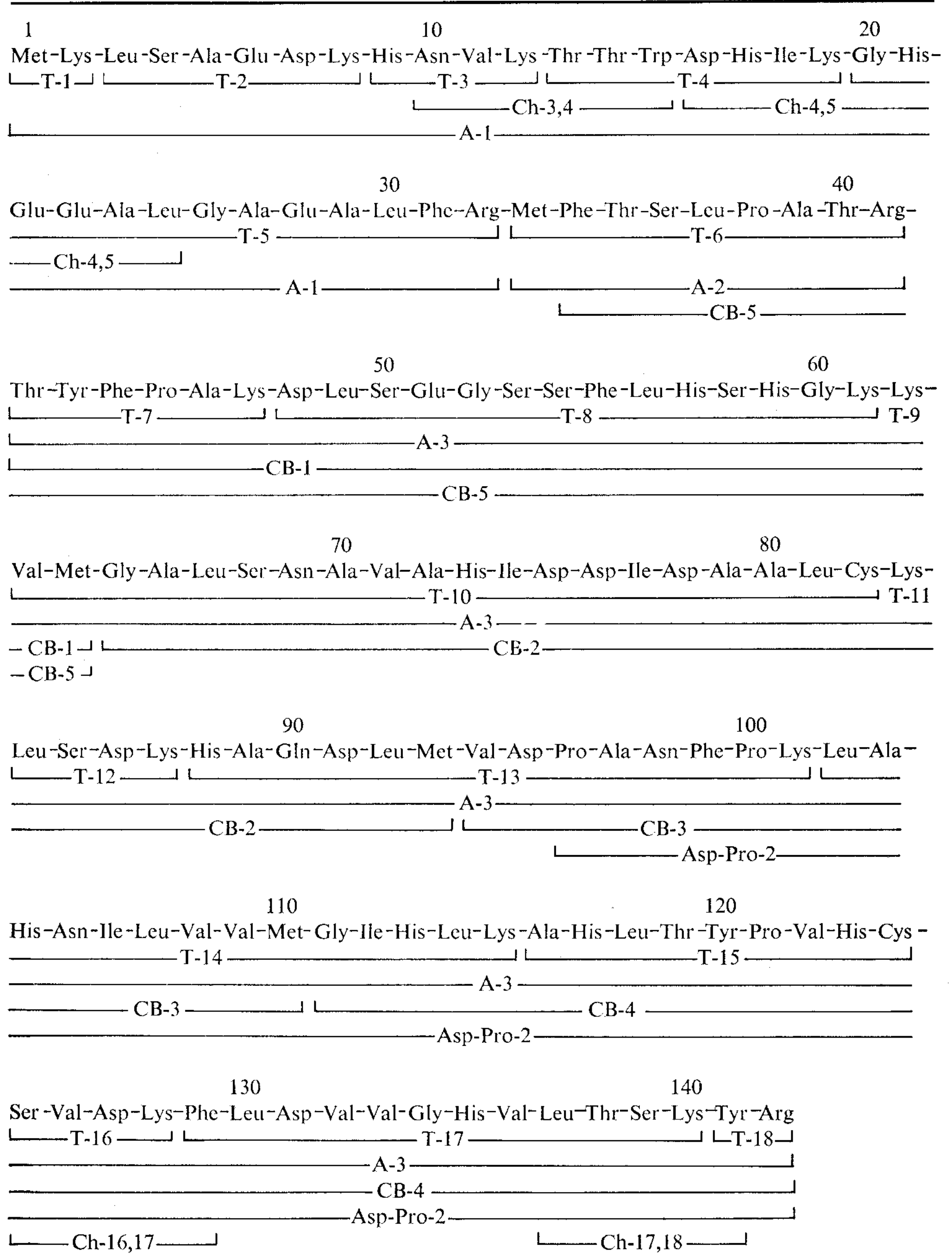
Discussion

In Table 5 is an alignment of sequences of α-chain from representative species from six classes of vertebrates. The length of the alignment is 150, rather than the 141 positions of human α-haemoglobin, because of the addition of seven *N*-terminal amino acids in the shark and additions at positions 47 and 65 in the fish. Table 6 is a matrix of amino acid differences between these sequences.

A notable feature of the matrix of differences is the relatively small number of differences between the human sequence and the others. This is particularly striking in the case of human *v.* chicken with only 35 differences, whilst viper *v.* chicken has

Table 4. Proposed amino acid sequence of *T. granulosa* α-chain with tryptic and overlapping peptides indicated

T, tryptic peptide; Ch, chymotryptic peptide; Asp-Pro, peptide produced by splitting the chain between aspartic acid and proline; A, arginyl peptide; CB, cyanogen bromide peptide. CB peptides 1–4 were produced by the splitting of A-3, and CB-5 was produced by the splitting of entire AE α-chain



1 10 20
Met-Lys-Leu-Ser-Ala-Glu-Asp-Lys-His-Asn-Val-Lys-Thr-Thr-Trp-Asp-His-Ile-Lys-Gly-His-
T-1 T-2 T-3 T-4
Ch-3,4 Ch-4,5
A-1

30 40
Glu-Glu-Ala-Leu-Gly-Ala-Glu-Ala-Leu-Phe-Arg-Met-Phe-Thr-Ser-Leu-Pro-Ala-Thr-Arg-
T-5 T-6
Ch-4,5
A-1 A-2
CB-5

50 60
Thr-Tyr-Phe-Pro-Ala-Lys-Asp-Leu-Ser-Glu-Gly-Ser-Ser-Phe-Leu-His-Ser-His-Gly-Lys-Lys-
T-7 T-8 T-9
A-3
CB-1
CB-5

70 80
Val-Met-Gly-Ala-Leu-Ser-Asn-Ala-Val-Ala-His-Ile-Asp-Asp-Ile-Asp-Ala-Ala-Leu-Cys-Lys-
T-10 T-11
A-3
CB-1 CB-2
CB-5

90 100
Leu-Ser-Asp-Lys-His-Ala-Gln-Asp-Leu-Met-Val-Asp-Pro-Ala-Asn-Phe-Pro-Lys-Leu-Ala-
T-12 T-13
A-3
CB-2 CB-3
Asp-Pro-2

110 120
His-Asn-Ile-Leu-Val-Val-Met-Gly-Ile-His-Leu-Lys-Ala-His-Leu-Thr-Tyr-Pro-Val-His-Cys-
T-14 T-15
A-3
CB-3 CB-4
Asp-Pro-2

130 140
Ser-Val-Asp-Lys-Phe-Leu-Asp-Val-Val-Gly-His-Val-Leu-Thr-Ser-Lys-Tyr-Arg
T-16 T-17 T-18
A-3
CB-4
Asp-Pro-2
Ch-16,17 Ch-17,18

Table 5. Alignment of the sequences of the α-chains of man, chicken, viper (*Vipera aspis*), newt (*Taricha granulosa*), bony fish (*Catostomus clarkii*), and shark (*Heterodontus portusjacksoni*)

Residues of the A, B, C, E, F, G and H helical regions in human α-chains are indicated. *c* at the bottom of the alignment indicates a 'certain convergence' (see text for explanation). * indicates a blocked *N*-terminal

α-Chain	Sequence
	A helix: 10–25; B helix: 27–41; C helix: 42–49
	1 2 3 4 5 6 7 8 9 | 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 | 6 | 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 | 2 3 4 5 6 7 8 9 (positions 1–49)
Human	– – – – – – – 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 F P T T K T Y
Chicken	– – – – – – – V L S N A D K N N V K G I F T K I A G H A E E Y G A E T L E R M F I G F P T T K T Y
Viper	– – – – – – – V L S E D D K N R V R T S V G K N P E L P G E Y G S E T L T R M F A A H P T T K T Y
Newt	– – – – – – M K L S A E D K H N V K T T W D H I K G H E E A L G A E A L F R M F T S L P A T R T Y
Bony fish	– – – – – – – *S L S D K D K A D V K I A W A K I S P R A D E I G A E A L G R M L T V Y P Q T K T Y
Shark	*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 Y A A T K S Y
	c at positions 29, 30
	E helix: 60–80; F helix: 89–97
	5 0 1 2 3 4 5 6 7 8 9 | 6 0 1 2 3 4 5 6 7 8 9 7 0 1 2 3 4 5 6 7 8 9 8 0 | 1 2 3 4 5 6 7 8 | 9 9 0 1 2 3 4 5 6 7 | (positions 50–97)
Human	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 H V D D M P N A L S A L S D L H A
Chicken	F P H F – D L S H G S A Q I K G H G K K V A – L A I T N A I E H A D D I S G A L S K L S D L H A
Viper	F P H F – D L S S G S P N L K A H G K K V I – D A L D N A V E G L D D A V A T L S K L S D L H A
Newt	F P A K – D L S E G S S F L H S H G K K V M – G A L S N A V A H I D D I D A A L C K L S D K H A
Bony fish	F A H W A D L S P G S G P V K – H G K K V I M G A I G D A V T K F D D L L G G L A S L S E L H A
Shark	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 H L D D L K T H L H K L A T F H G
	c at positions 61, 63, 65, 71, 75, 87
	G helix: 103–121; H helix: 127–147
	10 8 9 0 1 2 | 11 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 12 0 1 | 2 3 4 5 6 | 13 7 8 9 0 1 2 3 4 5 6 7 8 9 14 0 1 2 3 4 5 6 7 | 8 9 15 0 (positions 98–150)
Human	H K L R V D P V N F K L L S H C L L V T L A A H 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
Chicken	H K L R V D P V N F K L L G Q C F L V V L V A H L P A E L A P K V H A S L D K F L C A V G T V L T A K Y R
Viper	Q K L R V D P A N F K I L S Q C L L S T L A N H R N P E F G P A V L A S V D K F L C N V S E V L E S K Y R
Newt	Q D L M V D P A N F P K L A H N I L V V M G I H L K A H L T Y P V H C S V D K F L D V V G H V L T S K Y R
Bony fish	S K L R V D P S N F K I L A N C I T V V I M F Y L P G D F P P E V H A S V D K F F Q N L A L A L G Q K Y R
Shark	S E L K V D P A N F Q Y L S Y C L E V A L A V H 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
	c at positions 98, 105, 109, 111, 114, 117, 126, 132, 134, 142

57 differences. Since the fossil record indicates that the common ancestor of birds and reptiles is more recent than the common ancestor of the bird–reptile and mammalian lines (Young 1962), we would expect fewer differences between chicken and viper than between either of these species and man. The matrix suggests that the α-chain sequence of man has converged towards the other sequences.

If we assume that all the sequences in Table 5 have a common ancestor, 'certain' convergences can be identified because the branching pattern of the phylogenetic tree of the sequences is known from an independent source, the fossil record (Young 1962). Sites at which sequences have identical amino acids which could not have occurred through inheritance from a common ancestor are labelled with a *c* in Table 5. This method underestimates the number of convergences because not all convergences can be identified with certainty. Of the 18 such sites in Table 5 only one does not involve the human sequence as identical with another species through a possible convergence. This supports the suggestion that the human sequence is convergent with the species from other classes.

It is pertinent to ask if the patterns of evolution of α-chain over the time span represented by the sequences in Table 5 are similar to those in α-chain sequences from an equal sized group of representative eutherian mammals where the time span of evolution is much shorter. Accordingly, we have compared the number of substitutions at those sites identified with function in human haemoglobin with the number of substitutions in the rest of the sequence for both groups. Further, we have compared the number of sites at internal and external helical positions which are occupied by hydrophobic residues only with the number of internal and external helical sites which do not have hydrophobic residues in all representative species for each group. The group of mammalian sequences used are man, dog, mouse, rabbit, horse, and bovine α-chains (Dayhoff 1972), which represent six orders of eutherian mammals.

Goodman *et al.* (1975) have listed the sites identified with function in both α- and β-chains of human haemoglobin. There are 70 such sites in α-chain. Comparisons of various (a) mammalian and (b) vertebrate α-chains are shown in the tabulation below.

	Functional sites (70)	Other sites (71)	*P*[A]
(a) Man, dog, mouse, rabbit, horse and bovine α-chains			
Sites occupied by:			
a single amino acid	54	40	<0·01
more than one amino acid	16	31	
(b) Man, chicken, viper, newt,[B] bony fish[B] and shark[B] α-chains			
Sites occupied by:			
a single amino acid	28	8	<0·001
more than one amino acid	42	63	
Sites occupied by:			
one or two amino acids	43	25	<0·01
more than two amino acids	27	46	

[A] χ^2 test with one degree of freedom.
[B] Additions can not be counted as either functional or 'other'.

Fig. 1 in Perutz *et al.* (1965) indicates the orientation, with respect to the inside or outside of the globin molecule, of residues in the helical regions of haemoglobin. Of

the 108 helical sites in α-haemoglobin, 22 were disregarded because we could not be certain from the figure whether they were internal or external. Of the remaining 86 helical sites, 45 were judged to be internal and 41 external. Perutz *et al.* (1965)

Table 6. Matrix of differences between haemoglobin α-chains

The animals are the same as those in Table 5. The additional seven *N*-terminal residues in shark are counted as one amino acid difference

	Man	Chicken	Viper	Newt	Shark
Man	0				
Chicken	35	0			
Viper	50	57	0		
Newt	63	63	73	0	
Shark	82	86	86	88	0
Bony fish[A]	68	70	75	79	90

[A] From Powers and Edmundsen (1972).

point out that the hydrophobic nature of many internal helical sites tends to be conserved. Comparisons of various (a) mammalian and (b) vertebrate α-chains are shown in the tabulation below.

	Internal helical sites (45)	External helical sites (41)	P[A]
(a) Man, dog, mouse, rabbit, horse and bovine α-chains			
Sites occupied by:			
hydrophobic residues only	22	9	<0·01
hydrophobic and hydrophilic residues	23	32	
(b) Man, chicken, viper, newt, bony fish and shark α-chains			
Sites occupied by:			
hydrophobic residues only	19	3	<0·001
hydrophobic and hydrophilic residues	26	38	

[A] χ^2 test with one degree of freedom.

These tests indicate that the characteristics of the sites identified as functional and internal in human α-chain have been conserved to about the same extent over the longer time span required for the divergence of the classes of jawed vertebrates as they have for the shorter time span during which the orders of mammals diverged. That the functional sites have been conserved to a relatively large extent during the longer, as well as the shorter, time span supports the suggestion that most of the major features of α-haemoglobin function originated near the origin of jawed vertebrates (Coates 1975; Goodman *et al.* 1975). Twenty-nine residues of newt α-chain are different from any of those found in analogous positions of other reported animal α-chain sequences. They are: newt 2-Lys, 9-His, 17-His, 19-Lys, 25-Leu, 31-Phe, 37-Leu, 41-Arg, 46-Ala, 47-Lys, 51-Glu, 54-Ser, 55-Phe, 57-His, 58-Ser, 64-Met, 82-Cys, 91-Asp, 93-Met, 100-Pro, 101-Lys, 111-Gly, 112-Ile, 115-Lys, 117-His, 120-Tyr, 121-Pro, 131-Asp, 132-Val. The cysteine residue occupies a different position than has been reported previously for other animals. Newt 82-Cys is present in the residues surrounding the haem group whereas human α-chain has serine (human

F2-Ser). Of the residues listed above, only two correspond to human haem contacts: newt 47-Lys (human CD4-Phe) and newt 87-Lys (human F7-Leu). Among the $\alpha_1\beta_2$ contacts, the only substitution not previously reported is newt 93-Met (human FG4-Arg). The *N*-terminus of the newt α-chain (Met-Lys-Leu) is more analogous to a β-chain with Val-His-Leu than it is to other α-chains.

As argued above the major conformational features of α-chain have been strongly conserved since the divergence of the classes of vertebrates. It is thus surprising to find a proline at G6α in newt (newt 100-Pro), where only lysine or glutamic acid are found in other vertebrates. Its presence has been checked both by manual and automatic sequencing. A proline at this site could be expected to cause a bend in the G helix at this point. This may be reflected in the substitution of a hydrophilic lysine at G7α (newt 101-Lys) where only hydrophobic residues have existed before.

With regard to the low Bohr effect in newt haemoglobin Perutz (personal communication) states that: "the additional residue at the *N*-terminus would inhibit the normal contribution of the α-NH_3 to the alkaline Bohr effect". However, he further points out that Lys 2α (newt) can form a salt bridge with Asp H13α (newt 131-Asp) which replaces Ala in human α-chain (human 130-Ala). The main source of the alkaline Bohr effect is His 146β (Perutz and Ten Eyck 1972), but as we have not yet completed the amino acid sequence of newt β-chain we are not able to say if this residue and its interactions are present in newt haemoglobin.

Acknowledgments

We thank Marie Duerst, James Joyce, and Dr John Black, for their assistance. This work was largely supported by Grant AM 17850 from the U.S. Public Health Service and by an A.R.G.C. grant to Dr M. Coates.

References

Aksoy, M., *et al.* (1972). Hemoglobin Istanbul: substitution of glutamine for histidine in a proximal histidine (F8(92)β). *J. Clin. Invest.* **51**, 2380–7.

Ambler, R. P. (1972). Carboxypeptidases A and B. *Methods Enzymol.* **25**, 262–72.

Anson, M. L., and Mirsky, A. E. (1930). Protein coagulation and its reversal. *J. Gen. Physiol.* **13**, 469–76.

Baglioni, C. (1965). Abnormal human hemoglobins. X. A study of hemoglobin Lepore Boston. *Biochim. Biophys. Acta* **97**, 37–46.

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

Braunitzer, G., and Fujiki, H. (1969). Zur Evolution der Vertebraten. Die Konstitution und Tertiärstruktur des Hämoglobins des Flussneunauges. *Naturwissenschaften* **56**, 322–3.

Coates, M. L. (1975). Evolution of blood oxygen transport in the newt, *Taricha. J. Mol. Evol.* **6**, 285–307.

Coates, M. L., and Metcalfe, J. (1971). Hemoglobin function in the vertebrates: an evolutionary model. *Respir. Physiol.* **11**, 94–103.

Dayhoff, M. O. (1972). 'Atlas of Protein Sequence and Structure'. (National Biomedical Research Foundation: Silver Spring.)

Dintzis, H. M. (1961). Assembly of the peptide chains of hemoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **47**, 247–61.

Duguet, M., Chauvet, J. P., and Archer, R. (1974). Phylogeny of hemoglobins: the complete amino acid sequence of an α-chain of viper (*Vipera aspis*) hemoglobin. *FEBS Lett.* **47**, 333–7.

Friedman, G. B. (1974). The annual cycle of red blood cell count and haemoglobin level in the urodele *Taricha granulosa* on southern Vancouver Island. *Can. J. Zool.* **52**, 487–94.
Gibson, D. (1974). Structural studies on normal horse immunoglobin light chains. Detection of k-type n-terminal sequences. *Biochemistry* **13**, 2776–84.
Goodman, M., Moore, G. W., and Matsuda, G. (1975). Darwinian evolution in the genealogy of haemoglobin. *Nature (London)* **253**, 603–8.
Gray, W. R. (1972). Sequence analysis with dansyl chloride. *Methods Enzymol.* **25**, 333–44.
Gross, E. (1967). The cyanogen bromide reaction. *Methods Enzymol.* **11**, 238–55.
Guest, J. R., Carlton, B. C., and Yanofsky, C. (1967). The amino acid sequence of the A protein (α subunit) of the Tryptophan synthetase of *Escherichia coli.* I. Tryptic peptides. *J. Biol. Chem.* **242**, 5397–412.
Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972). Application of sequenator analysis to the study of proteins. *Biochemistry* **11**, 4493–502.
Hilse, V. K., and Braunitzer, G. (1968). Die Aminosäuresequenz der α-Ketten der beiden Hauptkomponenten des Karpfenhämoglobins. *Hoppe-Seylers Z. Physiol. Chem.* **349**, 433–50.
Huisman, T. H. J. (1963). Normal and abnormal human hemoglobins. *Adv. Clin. Chem.* **6**, 231–361.
Huisman, T. H. J., and Dozy, A. M. (1965). Studies on the heterogeneity of hemoglobin. IX. The use of tris (hydroxymethyl) aminomethane–HCl buffers in the anion-exchange chromatography of hemoglobins. *J. Chromatogr.* **19**, 160–9.
Jones, R. T. (1964). Chromatography of human hemoglobin. Factors influencing chromatography and differentiation of similar hemoglobins. *Cold Spring Harbor Symp. Quant. Biol.* **29**, 297–308.
Jones, R. T. (1970). Automatic peptide chromatography. *Methods Biochem. Anal.* **18**, 205–58.
Jones, R. T., and Schroeder, W. A. (1963). Structural studies of aminoethylated hemoglobins by automatic peptide chromatography. *J. Chromatogr.* **10**, 421–31.
Konigsberg, W. (1972). Subtractive Edman degradation. *Methods Enzymol.* **25**, 326–32.
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.
Light, A. (1972). Leucine aminopeptidase in sequence determination of peptides. *Methods Enzymol.* **25**, 253–62.
Matsubara, H., Sasaki, R. M., Singer, A., and Jukes, T. H. (1965). Specific nature of hydrolysis of insulin and tobacco mosaic virus protein by thermolysin. *Arch. Biochem. Biophys.* **115**, 324–31.
Matsuda, G., Takei, H., Wu, K. C., Shiozawa, T., and Ota, Y. (1972). Sequence studies on the tryptic peptides and the chymotryptic peptides from the α-polypeptide chain of AII component of the chicken hemoglobin. *Int. J. Pept. Protein Res.* **4**, 291–302.
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., Kendrew, J. C., and Watson, H. C. (1965). Structure and function of haemoglobin. II. Some relations between polypeptide chain configuration and amino acid sequence. *J. Mol. Biol.* **13**, 669–78.
Perutz, M. F., and Ten Eyck, L. F. (1972). Stereochemistry of cooperative effects in hemoglobin. *Cold Spring Harbor Symp. Quant. Biol.* **36**, 295–310.
Powers, D. A., and Edmundson, A. B. (1972). Multiple hemoglobins of catostomid fish. II. The amino acid sequence of the major α-chain from *Catostomus clarkii* hemoglobins. *J. Biol. Chem.* **247**, 6694–7070.
Schneider, R. G. (1974). Differentiation of electrophoretically similar hemoglobins—such as S, D, G, and P; or A_2, C, E, and O—by electrophoresis of the globin chains. *Clin. Chem. (N.Y.)* **20**, 1111–5.
Singhai, R. P., and Atassi, M. Z. (1971). Immunochemistry of sperm whale of myoglobin. IX. Specific interaction of peptides obtained by cleavage at arginine peptide bonds. *Biochemistry* **10**, 1756–62.
Smyth, D. G. (1967). Techniques in enzymic hydrolysis. *Methods Enzymol.* **11**, 214–31.
Stark, G. R. (1972). Use of cyanate for determining NH_2-terminal residues in proteins. *Methods Enzymol.* **25**, 103–20.
Stenzel, P. (1974). Opossum Hb chain sequence and neutral mutation theory. *Nature (London)* **252**, 62–3.
Sullivan, B. (1974). Amphibian hemoglobins. *Chem. Zool.* **9**, 77–122.

Weihung, R. R., and Korn, E. D. (1972). Acanthamoeba action. Comparison of the peptide that contains 3-methylhistidine and a peptide that contains Nε-methyllysine. *Biochemistry* **11**, 1538–43.

Whittaker, R. G., Fisher, W. K., and Thompson, E. O. P. (1973). Studies on monotreme proteins. II. Amino acid sequence of the α-chain in haemoglobin from the echidna, *Tachyglossus aculeatus aculeatus*. *Aust. J. Biol. Sci.* **26**, 877–88.

Whittaker, R. G., and Thompson, E. O. P. (1974). Studies on monotreme proteins. V. Amino acid sequence of the α-chain of haemoglobin from the platypus, *Ornithorhynchus anatinus*. *Aust. J. Biol. Sci.* **27**, 591–605.

Young, J. Z. (1962). 'The Life of Vertebrates'. 2nd Edn. (Oxford University Press.)

Manuscript received 1 September 1976