Aust. J. Biol. Sci., 1976, **29**, 73–97

Haemoglobins of the Shark, *Heterodontus portusjacksoni* II.* Amino Acid Sequence of the α-Chain

A. R. Nash, W. K. Fisher 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 principal haemoglobin from the shark, *H. portusjacksoni* has been determined. The chain has 148 residues and is acetylated at the amino terminal. The soluble peptides obtained by tryptic and chymotryptic digestion of the protein or its cyanogen bromide fragments were isolated by gel filtration, paper ionophoresis and paper chromatography. The amino acid sequences were determined by the dansyl–Edman procedure. The insoluble 'core' peptide from the tryptic digestion contained 34 residues and required cleavage by several proteases before the sequence was established.

Compared with human α-chain there are 88 amino acid differences including the additional seven residues which appear on the amino terminal of the shark chain. There is also one deletion and one insertion. The chain contains no tryptophan but has four cysteinyl residues which is the highest number of such residues recorded for a vertebrate globin.

In the $\alpha_1\beta_2$ contact sites there are four changes in the oxyhaemoglobin form and six in the deoxy form. Nine of the 16 $\alpha_1\beta_1$ contact sites show variation while three of the haem contact sites have changed in comparison to the residues known to be involved in these interactions in horse haemoglobin α-chain.

Use of the sequence data to estimate a time of divergence of the shark from the main vertebrate line yielded the value of 410 ± 46 million years. The data, in general, support the palaeontological view that bony fishes arose before the elasmobranchs.

Introduction

The evolutionary origins and relationships of elasmobranchs are something of a mystery, due mostly to the lack of fossils. The cartilaginous skeletons of these animals are only rarely preserved and palaeontologists have had to rely for much of their data on isolated teeth and spines. On such evidence they have placed the earliest sharks in the Upper Silurian and Lower Devonian (Romer 1966). The exact relationship between the elasmobranchs, teleosts, cyclostomes and the higher vertebrates has never been fully elucidated.

The use of the amino acid sequences of orthologous proteins to delineate phylogenetic relationships is well established (Buettner-Janusch and Hill 1965; Fitch and Margoliash 1970; Barnabas *et al.* 1971; Dickerson 1971; Goodman *et al.* 1971; Buehner *et al.* 1973). The same information can be utilized for other purposes (see reviews by Fitch 1973; Arnheim 1973; Wu *et al.* 1974). Sequences from several elasmobranch proteins have already been used to approximate the times of divergence from ancestral stocks, including trypsin activation peptides (Bricteux-Gregoire *et al.* 1972, 1974), neurohypophyseal hormones (Acher 1974) and cytochrome *c* (Fitch and

* Part I, *Aust. J. Biol. Sci.*, 1974, **27**, 607–15.

Margoliash 1970; Dickerson 1971; Nolan *et al.* 1973). Unfortunately the information content of the smaller peptides is limited by their size, and deductions drawn from their comparison are tenuous (Margoliash *et al.* 1971). However, cytochrome *c* suffers no such limitations and phylogenetic trees derived from sequences of this protein should be valid. Nolan *et al.* (1973) concluded from their studies that the elasmobranchs, cyclostomes and teleosts are more closely related to each other than to the higher vertebrates and further speculated that these three groups diverged from the main vertebrate line before diverging from each other. This theory awaits supporting evidence.

In the first paper of this series (Nash and Thompson 1974) the separation and purification of haemoglobin from the Port Jackson shark, *Heterodontus portusiacksoni*, was reported. In this paper we wish to report on the amino acid sequence of the α-chain of the principal haemoglobin from this elasmobranch. These data will be used to investigate the phylogenetic relationships of the shark and to contribute to the study of molecular evolution.

Materials and Methods

The methods of high-voltage ionophoresis, peptide mapping, amino acid analysis, cyanogen bromide cleavage, sequence determination by the dansyl–Edman procedure and digestion with trypsin, chymotrypsin, thermolysin, elastase, Nagarse (subtilisin BPN′) and pepsin were substantially the same as have been described previously (Air and Thompson 1969, 1971; Beard and Thompson 1971).

The *S*-carboxymethylated α-chain of shark haemoglobin was prepared as described previously (Nash and Thompson 1974).

Separation of Cyanogen Bromide Fragments

Approximately 50 mg of globin was dissolved in 70% formic acid, treated with 50 mg of cyanogen bromide at 4°C for 24 h, freeze-dried, redissolved in 98% formic acid and diluted to give a 70% solution of formic acid. A column (75 by 2·5 cm) of Sephadex G50 Fine was packed and equilibrated with 70% formic acid and the protein loaded. The column was developed at a flow rate of approximately 15 ml/h and 2·5-ml fractions collected. The extinction of each fraction was measured at 280 nm.

Isolation of Tryptic Peptides

Each CNBr fraction was redissolved in 2 ml of 70% formic acid and small aliquots were taken for analysis and for one cycle of the dansyl–Edman degradation. After analysis fractions 2 and 3 were freeze-dried and digested with trypsin. Each digest was centrifuged, the insoluble pellet washed with 1% NH_4HCO_3 and the washings added to the soluble fraction. The soluble material was freeze-dried, dissolved in pyridine acetate buffer at pH 6·4, centrifuged, and the insoluble 'core' washed well with buffer. The soluble peptides from fractions 2 and 3 were finally separated on paper by combined ionophoresis and chromatography (Thompson *et al.* 1969).

Tryptic digests of α-chain were treated in a similar way.

Separation of Chymotryptic Peptides

Approximately 70 mg of α-chain was digested with chymotrypsin for 4 h at 37°C and loaded on a column (136 by 2·5 cm) of Sephadex G25 Superfine equilibrated with 0·05M ammonia solution. Fractions of 5·4 ml were collected and the extinction of each fraction measured at 220 nm.

Each peak fraction was freeze-dried, dissolved in 1 ml of pyridine acetate buffer, pH 6·4, and centrifuged to remove any insoluble peptides. The soluble peptides were fractionated by paper ionophoresis–chromatography and the peptides revealed with 0·02% ninhydrin in 95% ethanol.

Purification of the Tryptic 'Core' Peptides

After digestion of the α-chain or the CNBr2 fragment with trypsin, an insoluble 'core' was obtained when the freeze-dried digest was dispersed in pH 6·4 buffer. Amino acid analyses of these

cores gave variable results for residues such as glycine and isoleucine which suggested inhomogeneity, with possible contamination from incompletely digested α-chain, or core material from the small amounts of β-chain mixed with α-chain preparations.

In order to reduce this contamination the core from tryptic digests was digested with chymotrypsin and pepsin. No special limiting conditions for the digestions were used and digestion under the conditions of Air and Thompson (1969) for 3 h at 37°C gave soluble peptides, but the residual undigested core that remained insoluble at pH 6·4 yielded the most reproducible analyses.

Isolation and Sequence Determination of the Blocked Amino Terminal Fragment of the α-chain

Attempts to detect the amino terminal residue of both the intact chain and CNBr1 by the dansyl method were always unsuccessful. It was assumed therefore, that the chain had a blocked amino terminal. By staining peptide maps with the Cl_2–KI method of Reindel and Hoppe (1954), major peptides were detected in both tryptic and chymotryptic digests. The tryptic peptide was eluted and analysed. The blocked chymotryptic peptide was purified by loading a digest of CNBr1 onto a column (8 by 0·8 cm) of sulphonated polystyrene (Bio-Rad AG50-X2) in the acid form and eluting the required peptide with water. This peptide was digested with carboxypeptidase A, Nagarse and elastase. Removal of the blocking group was attempted by treatment with 1M anhydrous HCl in methanol (Kawasaki and Itano 1972) for 3 days at room temperature. After drying, the residual peptide material was dissolved in 60% pyridine and allowed to stand for 2 days to hydrolyse any ester groups. The peptides obtained by this treatment were recovered by removing the pyridine in a stream of nitrogen. After fractionation by paper ionophoresis at pH 1·8 and elution from the paper, the peptides were sequenced by the dansyl–Edman procedure.

Determination of the Blocking Group on the Amino Terminal of the α-chain

N-acetyl serine was prepared by the method of Narita (1958*b*) and *N*-formyl serine was synthesized by the method of Folsch *et al.* (1962). It should be noted that use of the method of Sheehan and Yang (1958) for the synthesis of *N*-formyl serine resulted, in this laboratory, in the formation of all possible formyl derivatives, the principal derivative being *O*,*N*-diformyl serine. Hydrazinolysis was performed by the method of Narita (1958*a*) and dansylation and chromatography of the hydrazides followed the method of Schmer and Kreil (1969). Samples of the α-chain were also digested with 3M H_2SO_4 and the liberated acetic acid determined by gas chromatography (Kang *et al.* 1967).

Attempts to isolate a single, blocked amino acid or peptide were made by digesting the α-chain with chymotrypsin and separating the peptides on a column of Sephadex G25 Superfine developed with 0·05M NH_3. The blocked chymotryptic peptide was known to elute in the first fraction from such a column so the tubes containing this fraction were bulked and freeze-dried. A column (12 by 0·8 cm) of cation-exchange resin in the acid form (Bio-Rad AG50-X2, 200–400 mesh) was prepared and fraction 1 was bound to the resin except for the blocked peptides which were eluted with water. The eluted acidic material was digested with elastase and the residual blocked acidic peptides repurified as before on the cation-exchange resin. The peptide material remaining after ion-exchange purification was dissolved in 0·05M NH_3 and loaded onto a column (127 by 2·5 cm) of Sephadex G25 Superfine. The column was developed with 0·05M NH_3 at a flow rate of 30 ml/h and 4-ml fractions were collected. After reading the extinction of each tube at 220 nm the peak tubes were bulked and freeze-dried. After drying, a portion of each peptide was taken for amino acid analysis and the remainder was prepared for mass spectrometry by permethylation according to the method of Morris *et al.* (1971). The mass spectra of the permethylated peptides was determined on an AEI MS9 Mass Spectrometer by drying a chloroform solution of the derivatized peptides onto the quartz tip of the injection probe prior to insertion of the probe into the ion source.

Dates of Divergence

Evolutionary rates of vertebrate α-chains were calculated using three methods. In the first, that of Air *et al.* (1971), amino acid differences were replaced by random evolutionary hits (REH) calculated from the equations of Holmquist *et al.* (1972) and Jukes and Holmquist (1972). The values obtained for the random evolutionary hits were corrected for the numbers of amino acids compared. The second method used was that proposed by Markussen and Volund (1974), while the third was that proposed by Dayhoff (1972) and required the use of accepted point mutations (PAM units). The paleontological estimates of the dates of divergence are largely those of Stahl (1974).

Having calculated the rate constants of the three methods, these were substituted back into the original equations to obtain a value which represented the period of time which has elapsed since the shark and the animal being compared last shared a common ancestor. The data used in these calculations are presented in Tables 5 and 6.

Phylogeny of the Lower Vertebrates

The minimum mutation distances (Fitch and Margoliash 1967) between a representative higher vertebrate (human) and the three lower vertebrates (carp, shark and lamprey) were calculated by comparison of their haemoglobin α-chains. Each deletion and insertion was treated as a single mutation. With four populations there are three possible tree forms for unrooted trees, i.e. trees for which there is no trunk. The minimum distances were fitted to these trees by the 'additive tree' method of Cavalli-Sforza and Edwards (1964). REH values were also inserted into the equations.

Results

As in previous papers, the tryptic peptides are numbered from the *N*-terminal (see Fig. 4) following the nomenclature for the human α-chain (Gerald and Ingram 1961).

Determination of the Blocking Group on the Amino Terminal of the α-Chain

Hydrazinolysis followed by dansylation of the α-chain yielded fluorescent spots with R_F values similar to both formyl hydrazide and acetyl hydrazide (Schmer and Kreil 1969). Digestion of the protein with 3M H_2SO_4 followed by gas chromatography showed the presence of acetic acid but estimation of the liberated acetic acid was largely unsuccessful. Wadke and Lowenstein (1975) have drawn attention to the problems inherent in the determination of acetate by gas–liquid chromatography. In addition to these difficulties formic acid cannot be detected by a flame ionization detector.

The initial attempts to utilize the potential of gas–liquid chromatography and mass spectrometry were also unsuccessful. However, in this case it was realized that a compound obtained by pronase digestion, initially thought to be *N*-acyl serine, was in fact a peptide and hence methylation of the free carboxyl group was not rendering the peptide sufficiently volatile for gas–liquid chromatography. Following digestion with chymotrypsin and elastase, coupled with ion-exchange purification, it was expected that a series of blocked peptides containing varying amounts of serine and threonine would be obtained. Gel filtration of the mixture on Sephadex G25 Superfine produced several fractions. Analysis of each fraction showed that only two of them contained predominantly serine and threonine. Other fractions after hydrolysis contained a large amount of glutamic acid and since a pyrrolidonyl peptide was known to be present as a contaminant in the chymotryptic digest the glutamic acid was thought to result from such a peptide in the elastase digest. The fractions which on hydrolysis contained serine and threonine in the molar ratio of approximately 2 : 1 would be expected to contain the peptide acetyl-Ser-Thr-Ser. After permethylation of the fractions the mass spectrum (Fig. 1) showed unequivocally that the blocking group was an acetyl group.

(i) *Separation of cyanogen bromide fragments*

Amino acid analyses of α-chain (Nash and Thompson 1974) showed the presence of a single methionyl residue so that cleavage with CNBr would be expected to give two fragments.

Separation of the cyanogen bromide fragments on Sephadex G50 is illustrated in Fig. 2. On analysis, fraction 1 appeared to contain all the amino acids expected in the intact protein and it was assumed that this peak represented uncleaved α-chain, possibly with methionine sulphoxide residues. Fraction 2 contained no methionine or homoserine and appeared to represent a fragment of approximately 109 residues.

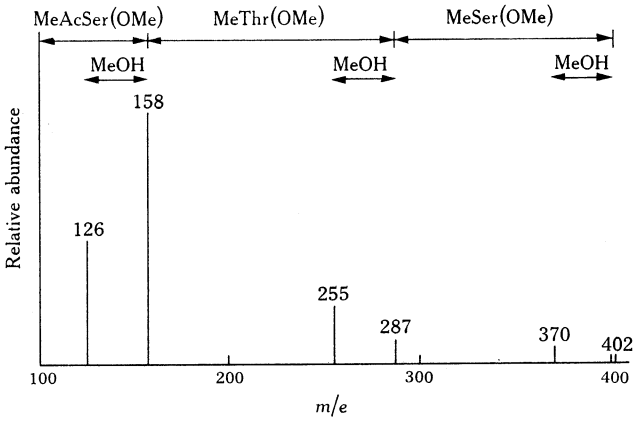


Fig. 1. Mass spectrum (above *m/e* 100) of a permethylated amino terminal peptide of shark α-chain. The spectrum was obtained on an AEI MS9 double focusing mass spectrometer. Source temperature 160°C, accelerator voltage 7 kV and an electron beam energy of 70 eV. The peptide has the sequence *N*-acetyl-Ser-Thr-Ser- and to simplify the figure the abbreviation Me denotes *N*-methylation while OMe denotes methylation of the side chain hydroxyl group.

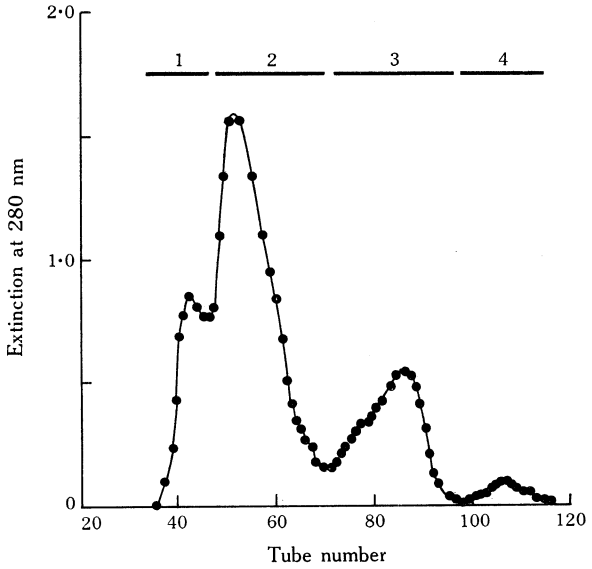


Fig. 2. Gel filtration of a cyanogen bromide digest of shark α-globin on a column of Sephadex G50 (75 by 2·5 cm) run in 70% formic acid. Fraction size 2·5 ml. Fractions bulked are shown by bars.

A peptide Phe-Thr-Val-Tyr was isolated from a tryptic peptide map of fraction 2. It occupied the same position on the fingerprint as the peptide αTp5*N* and had a similar sequence, lacking only the methionine at the *N*-terminal. Fraction 2 was therefore

assumed to be the major CNBr fragment originating from the carboxyl terminal of the protein. Fraction 3, on analysis, yielded two residues of arginine, one of lysine and one of homoserine. There was no *S*-carboxymethyl cysteine, proline or isoleucine. The fragment appeared to be of the order of 40 residues in length. After digestion with trypsin and subsequent fingerprinting only three major ninhydrin-positive spots were revealed, including free homoserine. A further major peptide was detected by the chlorine procedure of Reindel and Hoppe (1954). On the basis of this evidence fraction 3 was assumed to represent the amino terminal of the protein with a blocked α-amino group. Fraction 4 contained little peptide material and on analysis yielded all amino acids in trace amounts. It was not identified so far as its origin was concerned.

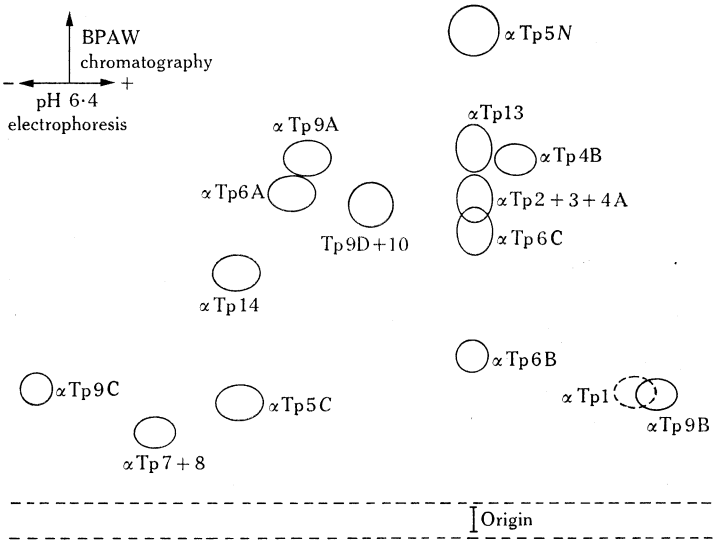


Fig. 3. Peptide map of the tryptic peptides of shark α-globin. Ionophoresis at pH 6·4 was followed by chromatography with butanol–pyridine–acetic acid–water (15 : 10 : 3 : 12 v/v) as indicated. The peptides are given the number corresponding to their position in the chain.

(ii) *Separation of the soluble tryptic peptides*

A peptide map of a digest of whole α-chain separated by paper ionophoresis–chromatography is shown in Fig. 3. All peptides except the insoluble core peptide, αTp11+12, are present. The alignment of the peptides relative to human α-chain is shown in Fig. 4. When the cyanogen bromide fragments were peptide-mapped separately, αTp2+3+4A and αTp6C were obtained pure as they occurred in separate fractions.

(iii) *Separation of the chymotryptic peptides on Sephadex G25 Superfine*

Separation of the chymotryptic peptides on Sephadex G25 Superfine is shown in Fig. 5. The principal peptides in each fraction are listed in the legend. The peptide αCh1 was ninhydrin-negative and could be detected only by the chlorine stain of Reindel and Hoppe (1954). Several of the longer chymotryptic peptides were

secondarily cleaved by the enzyme and often the smaller peptides were also detected. The positions of the peptides on a composite peptide map is shown in Fig. 6.

Amino Acid Composition and Sequence of Tryptic Peptides

The amino acid compositions of purified peptides are shown in Table 1.

The values for αTp11+12 are the means of two analyses, one of the chymotryptic digest of the tryptic core and the other of the peptic digest of the tryptic core. The total compositions are in good agreement with the most recent total analysis of the α-chain (Table 2). This analysis was chosen as it yielded the smallest value for tryptophan. The presence of trace amounts of tryptophan and the inflated values for glycine, isoleucine and glutamic acid reflect contamination by the β-chain which has a much higher content of these amino acids. Our previously reported analysis (Nash and Thompson 1974) reflected even higher levels of β-chain contamination.

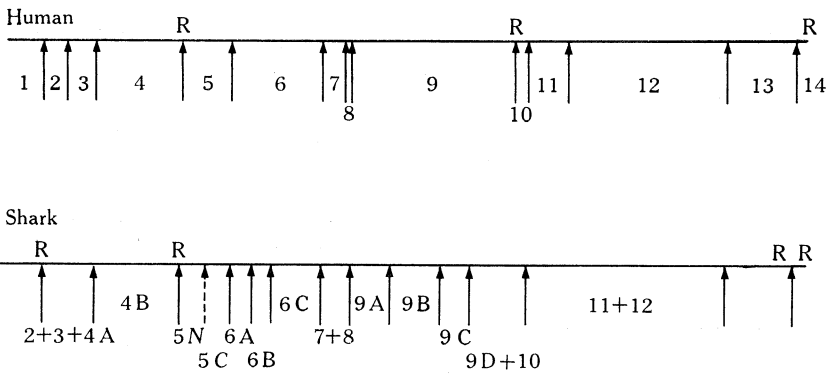


Fig. 4. Nomenclature for α-chain tryptic peptides of shark haemoglobin compared with the α-chain tryptic peptides of human haemoglobin. Where extra lysyl or arginyl residues have occurred the symbols A, B, C and D have been utilized. When a lysyl or arginyl residue has not occurred the peptide has been given all the numbers of the corresponding human peptides. The peptide αTp5 was not detected in its intact form and the two fragments corresponding to a split at the position indicated by the dotted arrow are referred to as αTp5*N* and αTp5*C*. The diagrams are scaled to show the relative number of residues in each peptide. ↑ Lysyl residue. R↑ Arginyl residue.

The peptides have been aligned by homology, supported by their allocations to areas of the chain by the cyanogen bromide fragments and by overlapping chymotryptic peptides. The residues are numbered from 1 to 148 beginning with the *N*-terminal serine. All residues identified by the dansyl–Edman method are printed in *italic* fount.

αTp1

This peptide had the sequence

$$\underset{1}{\textit{Ac-Ser-Thr-Ser-Thr-Ser-Thr-Ser-Asp-Tyr-Ser-Ala-Ala-Asp}}\text{-Arg}_{14},$$

where Ac represents an acetyl group. The amino acid composition of the peptide (Table 1) showed the presence of an arginyl residue as well as the abnormally high

content of serine and threonine. It could not be sequenced directly. There are only four arginyl residues in the α-chain, however, one of which is *C*-terminal. A second arginyl residue is *C*-terminal in αTp4B which has the terminating sequence -Ala-Leu-Ala-Arg. The third arginyl residue occurs in αTp13 which has the terminating sequence -Leu-Ser-Ser-Arg. A chymotryptic peptide (αCh2) with the sequence

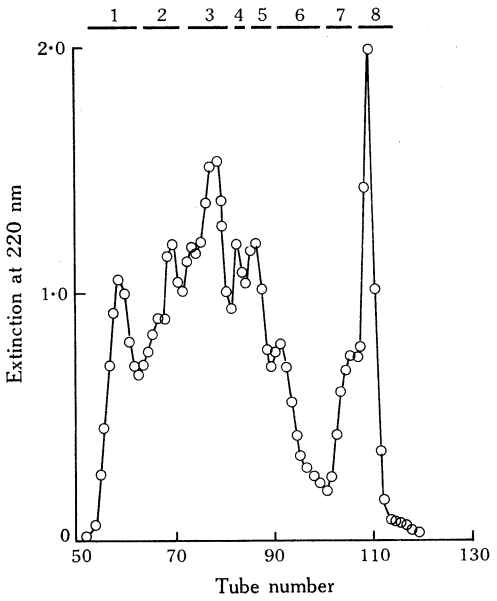


Fig. 5. Gel filtration of the peptides soluble at pH 8·2 after chymotryptic digestion of shark α-globin (70 mg) on a column (136 by 2·5 cm) of Sephadex G50 in 0·05M ammonia. Fraction size 5·4 ml. Tubes containing fractions combined for peptide recovery are shown as bars. The major peptides in each fraction are shown below where *N* and *C* refer to the *N*-terminal or *C*-terminal portion of subfragments of a particular peptide. The peptides are given the number corresponding to their position in the chain (see Table 4).

Fraction:

1	2	3	4	5	6	7	8
αCh14	αCh1	αCh5	αCh3	αCh9	αCh8	αCh7	αCh18
αCh14*N*	αCh2	αCh6	αCh4	αCh11	αCh10	αCh18*C*	αCh18*N*
	αCh19	αCh12		αCh15	αCh12*C*	αCh24	
	αCh23	αCh12*N*		αCh22*C*	αCh14*C*	αCh25	
		αCh13			αCh15*N*		
		αCh16			αCh15*C*		
		αCh17					
		αCh20					
		αCh21					
		αCh22					
		αCh22*N*					

Ser-Ala-Ala-Asp-Arg-Ala-Glu-Leu was therefore deduced to include the *C*-terminal portion of αTp1. By subtraction the *N*-terminal fragment of αTp1 must have had the composition Ac-(Ser_4 Thr_3 Asp Tyr). The chymotryptic peptide obtained by passage of the CNBr1 digest through the cation-exchange column had the composition $Asp_{1·0}$ $Thr_{3·0}$ $Ser_{3·8}$ $Tyr_{0·9}$. Digestion of portion of this peptide with carboxypeptidase A yielded only tyrosine as a free amino acid confirming a *C*-terminal tyrosine as expected from the specificity of chymotrypsin.

Digestion of further samples with subtilisin (Su) and elastase (El) gave a series of peptides which were mapped (Fig. 7) and sequenced to yield the following sequences:

Asp-Tyr	(El 4, Su 4)	
Ser-Asp-Tyr	(Su 3)	
Ser-Thr-Ser-Asp-Tyr	(El 6, Su 2)	
Ser-Thr-Ser	(El 1, Su 1)	
Ser-Thr-Ser-Asp	(El 2)	
Ser-Asp	(El 3)	
Ac-(Ser,Thr)$_{1-2}$	(El 5)	detected by Cl_2–KI.

This information was insufficient to give the complete sequence of the amino terminal portion of αTp1 although it suggested that Ac-(Ser,Thr)$_2$-Ser-Thr-Ser-Asp-Tyr was present.

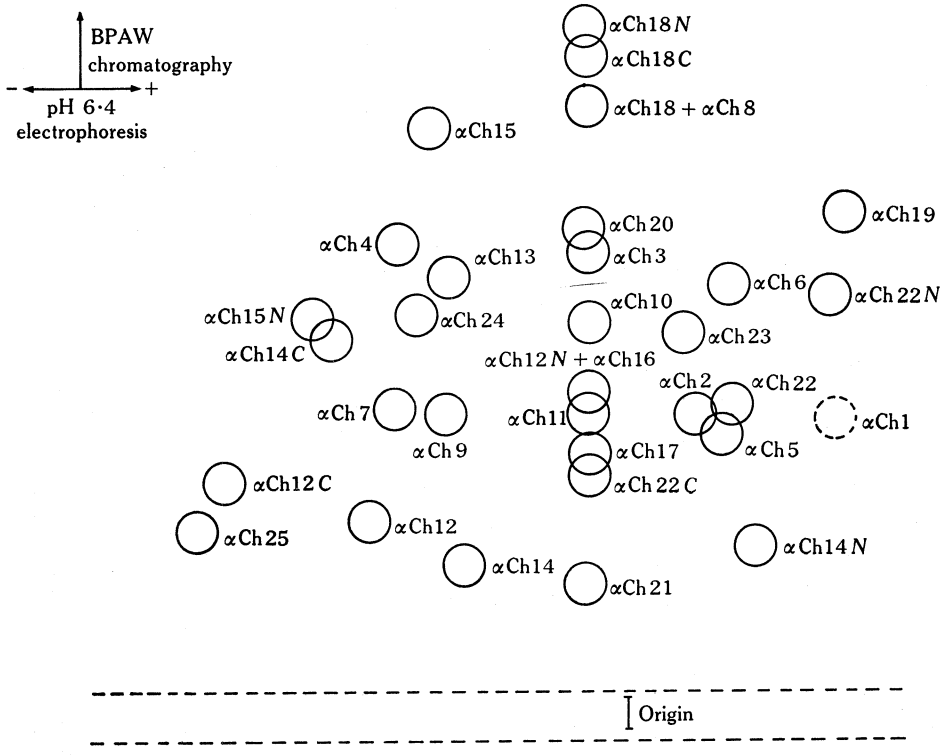


Fig. 6. Composite peptide map of the chymotryptic peptides of shark α-globin. Ionophoresis at pH 6·4 was followed by chromatography with butanol–pyridine–acetic acid–water (15 : 10 : 3 : 12 v/v) as indicated. The peptides are given the number corresponding to their position in the chain (see Table 4 and the legend to Fig. 5).

After treatment of αCh1 with methanolic HCl, *N*-terminal group estimation by the dansyl method showed that serine, threonine, aspartic acid and tyrosine were present, in descending order of abundance. Fractionation by paper ionophoresis at pH 1·8 gave a series of spots as revealed by dilute ninhydrin. The slowest-moving peptide was expected to be the largest. The dansyl–Edman procedure gave a clear

Table 1. Amino acid composition of shark α-globin tryptic peptides

Soluble peptides were purified by paper ionophoresis at pH 6·4 and paper chromatography. Hydrolysis was at 110°C for 24 h. Values are not corrected for losses during hydrolysis or due to incomplete hydrolysis and are given as moles per mole of peptide, with preferred values in parentheses. Detection by ninhydrin results in low recovery of the *N*-terminal residue

Amino acid	αTp1	αTp2+3+4A	αTp4B	αTp5*N*	αTp5*C*	αTp6A	αTp6B	αTp6C	αTp7+8	αTp9A	αTp9B	αTp9C	αTp9D+10	αTp11+12	Tp13	αTp14
Lys		1·0 (1)			1·0 (1)	1·0 (1)	1·0 (1)	0·9 (1)	1·0 (1)	1·1 (1)	1·1 (1)	1·0 (1)	1·0 (1)	1·2 (1)		
His									1·0 (1)		1·0 (1)	1·9 (2)	0·8 (1)	1·8 (2)	0·8 (1)	
Arg	0·8 (1)		1·0 (1)												0·8 (1)	1·0 (1)
SCMCys											0·8 (1)			0·9 (2)	0·7 (1)	
Asp	2·2 (2)		0·9 (1)				0·6 (1)	0·6 (1)			3·4 (3)			2·2 (3)	1·0 (1)	
Thr	3·1 (3)			0·8 (1)	0·9 (1)			1·0 (1)		0·6 (1)		0·5 (1)	0·9 (1)	2·0 (2)	1·1 (1)	
Ser	5·0 (5)	0·9 (1)				0·8 (1)		1·0 (1)					0·8 (1)	2·1 (2)	1·9 (2)	
Glu		1·0 (1)	3·1 (3)										1·0 (1)	4·0 (4)	1·1 (1)	
Pro								1·1 (1)						1·7 (2)		
Gly			1·0 (1)						1·2 (1)				1·0 (1)	0·8 (0)		
Ala	2·3 (2)	2·6 (3)	6·7 (6)		2·0 (2)			3·3 (3)	2·0 (2)	2·0 (2)	1·0 (1)		1·1 (1)	3·7 (4)		
Val			0·4 (1)	0·8 (1)						1·3 (2)				2·8 (3)	1·0 (1)	
Met				0·6 (1)												
Leu		1·9 (2)	1·6 (2)							1·2 (1)	2·1 (2)	1·0 (1)	1·8 (2)	4·4 (5)	1·9 (2)	
Ile								0·9 (1)						0·6 (0)		
Tyr	1·0 (1)			0·9 (1)		1·0 (1)	1·0 (1)							2·0 (2)		1·0 (1)
Phe			1·0 (1)	1·0 (1)		1·0 (1)		1·0 (1)					1·0 (1)	2·3 (2)	0·5 (1)	
Total	14	8	16	5	4	4	3	10	5	7	9	5	10	34	12	2

sequence for this peptide of *Ser-Thr-Ser-Thr-Ser-Thr-Ser-Asp-Tyr*. Other peptides were sequenced and corresponded to shorter peptides such as *Ser-Thr-Ser-Thr-Ser-Asp-Tyr*.

*αTp2+3+4*A

The sequence of this peptide was

Ala-Glu-Leu-Ala-Ala-Leu-Ser-Lys.
15 22

The peptide was neutral at pH 6·4 indicating the presence of glutamic acid. The overlap with αTp1 was provided by αCh2 which had the sequence *Ser-Ala-Ala-Asp*-Arg-*Ala-Glu-Leu*. This peptide was acidic at pH 6·4 confirming the presence of aspartyl and glutamyl residues. Further confirmation of the charge on the acidic residues was provided by a tryptic split of αCh2 which yielded the peptides *Ser-Ala-Ala-Asp*-Arg and *Ala-Glu-Leu* which were neutral and acidic respectively.

Table 2. Amino acid composition of shark α-globin

Samples were hydrolysed at 110°C for 24, 48 and 72 h under vacuum with 6M HCl containing 1 mg/ml phenol. Values are given as mole per mole of protein and are corrected for losses during hydrolysis and incomplete hydrolysis. Tryptophan was determined by the method of Beaven and Holiday (1952). Also shown is the amino acid composition determined by sequence data

Amino acid	Hydrolysate	Sequence	Amino acid	Hydrolysate	Sequence
Lysine	11·0	11	Alanine	24·8	26
Histidine	7·4	8	Valine	7·8	8
Arginine	4·1	4	Methionine	1·1	1
SCM cysteine	3·4	4	Isoleucine	1·4	1
Aspartic acid	11·9	12	Leucine	17·2	17
Threonine	11·9	12	Tyrosine	6·8	7
Serine	13·0	13	Phenylalanine	8·0	8
Glutamic acid	11·1	10	Tryptophan	0·3	0
Proline	3·7	3			
Glycine	3·6	3	Total		148

*αTp4*B

The sequence of this peptide was

Val-Leu-Ala-Gln-Asn-Ala-Glu-Ala-Phe-Gly-Ala-Glu-Ala-Leu-Ala-Arg.
23 38

The peptide Ala-Glu-Ala-Phe-Gly-Ala-Glu-Ala-Leu-Ala-Arg was also isolated from the tryptic digest, possibly as a result of chymotryptic activity, and this peptide was acidic at pH 6·4 confirming the glutamyl residues at positions 29 and 34. The overall mobility of αTp4B was quite low, almost neutral, so it was expected that probably both the first two acidic residues would be in the amide form. The peptide αCh5 was hydrolysed with thermolysin and yielded the peptides Ala-Gln-Asn and Ala-Phe, which were neutral, and Ala-Glu which was acidic at pH 6·4. These results

confirmed the presence of a glutaminyl and asparaginyl residue at positions 26 and 27 respectively, and the presence of a glutamyl residue at position 29. The chymocryptic peptide αCh4, which sequenced as *Ser-Lys-Val-Leu*, was found and provided the overlap between αTp2+3+4A and αTp4B. The peptides *Ala-Gln-Asn-Ala-Glu-Ala-Phe* (αCh5) and *Gly-Ala-Glu-Ala-Leu* (αCh6) were also isolated from the chymotryptic digestion.

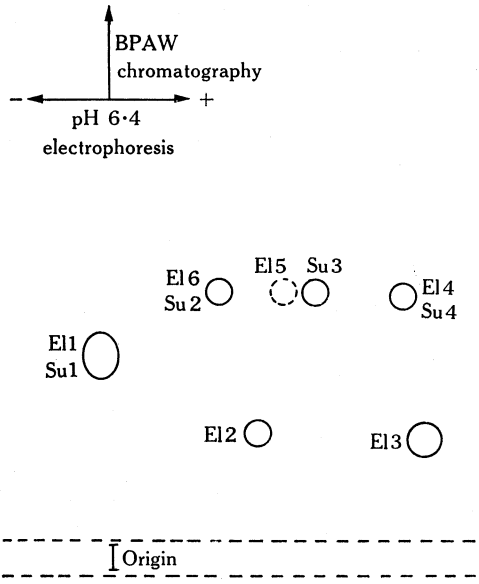


Fig. 7. Composite peptide map of the peptides obtained by subtilisin (Su) and elastase (El) digestion of tryptic peptide αTp1 of shark haemoglobin. Fractionation of the peptides as in Fig. 3.

αTp5N

This peptide had the sequence

Met-Phe-Thr-Val-Tyr.
39 43

The chymotryptic peptide Ala-Arg-Met-Phe (αCh7) was isolated and provided the overlap between αTp4B and αTp5*N*. An additional peptide, αCh8, was found and sequenced as *Thr-Val-Tyr*.

αTp5C

This peptide had the sequence

Ala-Ala-Thr-Lys.
44 47

This peptide was placed after αTp5*N* by homology as the intact peptide Met-Phe-Thr-Val-Tyr-Ala-Ala-Thr-Lys was never found in the tryptic digest. A thermolysin digest of the intact α-chain yielded a peptide *Val-Tyr-Ala* which excluded the possibility of a tryptophan residue at position 44 and also provided the overlap between αTp5*N* and αTp5*C*. The Tyr–Ala bond was apparently particularly susceptible to the action of chymotrypsin impurities in the trypsin used, or the inherent chymotryptic activity of trypsin.

*αTp6*A

This peptide had the sequence

Ser-Tyr-Phe-Lys.
48 51

Isolation of the chymotryptic peptide *Ala-Ala-Thr-Lys-Ser-Tyr* (αCh9) provided the overlap with αTp5*C*.

*αTp6*B

This peptide had the sequence

Asp-Thr-Lys.
52 54

The peptide was neutral at pH 6·4 confirming the presence of an aspartyl residue. The chymotryptic peptide Phe-Lys-Asp-Tyr (αCh10) was also neutral and provided the overlap with αTp6A.

*αTp6*C

This peptide had the sequence

Asp-Phe-Thr-Ala-Ala-Ala-Pro-Ser-Ile-Lys.
55 64

Both this peptide and the chymotryptic peptide *Lys-Asp-Phe* (αCh11) were neutral at pH 6·4 confirming the presence of an aspartyl residue at position 55. The tryptic peptide αTp6C was hydrolysed with elastase and a mixture was obtained containing

Asp-Phe
Phe-Thr
Thr-Ala-Ala
Thr-Ala-Ala-Ala-Pro-Ser
Ala-Pro-Ser
Ala-Pro-Ser-Ile-Lys
Ile-Lys.

The chymotryptic peptide Thr-Ala-Ala-Ala-Pro-Ser (αCh12*N*) was also found.

αTp7+8

This peptide had the sequence

Ala-His-Gly-Ala-Lys.
65 69

The chymotryptic peptides *Thr-Ala-Ala-Ala-Pro-Ser-Ile-Lys-Ala*-His (αCh12) and *Ile-Lys-Ala*-His (αCh12*C*) were found and provided the overlap with αTp6C. The susceptibility of the Ser–Ile bond to chymotryptic activity is unusual and cannot be explained.

*αTp9*A

This peptide had the sequence

Val-Val-Thr-Ala-Leu-Ala-Lys.
70 76

The low analysis values for valine can be attributed to the stability to acid hydrolysis of the Val–Val bond and the *N*-terminal position of these residues. The chymotryptic peptide *Gly-Ala-Lys-Val-Val-Thr-Ala-Leu* (αCh13) was found and provided the overlap with αTp7+8. The analysis for this peptide was $Gly_{0.6}\,Ala_{2.0}\,Val_{1.8}\,Leu_{1.1}\,Lys_{1.0}\,Thr_{0.9}$.

*αTp9*B

This peptide had the sequence

Ala-SCMCys-Asp-His-Leu-Asp-Asp-Leu-Lys.
77 85

A thermolysin digest of this peptide yielded three smaller fragments which were fractionated by ionophoresis at pH 6·4 followed by paper chromatography. The sequences and ionization of the peptides at pH 6·4 were as follows

Th1	*Ala-SCMCys-Asp-His*	Acidic	$m_{Asp} = 0{\cdot}57$
Th2	*Leu-Asp-Asp*	Acidic	$m_{Asp} = 0{\cdot}94$
Th3	*Leu-Lys*	Basic	$m_{Asp} = -0{\cdot}67$

The mobilities of the peptides at pH 6·4 relative to aspartic acid indicated the presence, in the parent tryptic peptide, of four acidic residues. The chymotryptic peptide αCh14 was isolated and had the sequence

Ala-Lys-Ala-SCMCys-Asp-His-*Leu-Asp-Asp-Leu-Lys-Thr*-His-*Leu*-His.

This peptide provided the overlap between αTp9A, αTp9B and αTp9C. A further chymotryptic peptide, αCh14*N*, was isolated and had the sequence

Ala-Lys-Ala-SCMCys-Asp-His-*Leu-Asp-Asp-Leu-Lys-Thr*-His.

This peptide corresponds to the *N*-terminal portion of αCh14. The *C*-terminal portion, Leu-His (αCh14*C*), was also found.

*αTp9*C

This peptide had the sequence

Thr-His-Leu-His-Lys.
86 90

The overlap between this peptide and αTp9B was provided by αCh14 as shown above. The low analysis for threonine is due to some destruction on hydrolysis coupled with the partial destruction by ninhydrin of *N*-terminal amino acids.

αTp9D+10

This peptide had the sequence

Leu-Ala-Thr-Phe-His-Gly-Ser-Glu-Leu-Lys.
91 100

A thermolysin digest of this peptide yielded three fragments which sequenced as

Th1 *Leu-Ala-Thr*
Th2 *Phe*-His-*Gly-Ser*-Glu
Th3 Leu-Lys.

Th2 was neutral at pH 6·4 confirming the presence of glutamic acid. The chymotryptic peptide αCh15 with the sequence *Lys-Leu-Ala-Thr-Phe* was isolated and provided the overlap with αTp9C. A further chymotryptic peptide, αCh16, was found which had the sequence His-*Gly-Ser-Glu-Leu* and was neutral at pH 6·4 and hence further confirmed the presence of glutamic acid.

αTp11+12

This peptide, the 'core', was difficult to free of partly digested chain and repeated analysis of 'purified' cores showed contamination with traces of β-chain. However, by digestion of the core with a variety of enzymes it was possible to obtain sufficient overlaps to yield the sequence of this peptide. The enzymes used were chymotrypsin, thermolysin and pepsin. Table 3 shows the proposed sequence for the core peptide together with the alignment of the fragments from the various enzyme digests. The peptides αCh18 and αCh18*N* were found to have a blocked amino terminal. They were obtained pure by passing fraction 8 of the chymotryptic digest (Fig. 4) through a small column (12 by 1 cm) of sulphonated polystyrene resin (Bio-Rad AG50 × X2) according to the method of Narita (1958*a*). The non-absorbed peptides were fractionated by paper ionophoresis at pH 3·5 and revealed by Cl_2–KI (Reindel and Hoppe 1954). Analysis of αCh18 yielded the values $Glu_{1.2}\,Ser_{1.0}\,Leu_{1.0}\,Tyr_{1.9}$, while αCh18*N* was composed of $Glu_{1.0}\,Tyr_{1.0}$ suggesting a pyrrolidonyl *N*-terminal group. After treatment of the former peptide with methanolic HCl (Kawasaki and Itano 1972), dansylation yielded glutamic acid as the amino terminal residue while the dansyl–Edman procedure gave the sequence Glu-Tyr-Leu-Ser-Tyr for αCh18. These results indicated that the *N*-terminal glutaminyl residue at position 107 had cyclized to form a pyrrolid-2-one-5-carboxyl residue. The proposed sequence for αTp11+12 is in good agreement with the analytical data although the yields of some acids, particularly *S*-carboxymethyl cysteine, are low. The presence of 0·9 residue of glycine, which is limited to three residues in the α-chain, appears to indicate contamination with core from the limited amount of β-chain present. The allocation of charge to the acidic residues was based on the mobility of the various peptides. Both αCh17 and Pe6 (residues 100–102) were neutral indicating the presence of an aspartyl residue at position 102, and an asparaginyl residue at position 105 is therefore indicated.

The peptides αTh18 (residues 113–114) and Pe16 (residues 113–116) were both acidic indicating the presence of a glutamyl residue at position 114. The peptides αCh20 and αCh21 were both neutral at pH 6·4 indicating the presence at positions

Table 3. Amino acid sequence of the core tryptic peptide comprising residues 101–134 of the α-chain of shark globin

The peptides obtained by cleavage with the enzymes chymotrypsin, pepsin and thermolysin are indicated. The residues identified by the dansyl–Edman method are printed in *italic* fount

Chymotryptic digest of tryptic core		101									110										120										130				134	
	-(Lys)-	Val-	Asp-	Pro-	Ala-	Asn-	Phe-	Gln-	Tyr-	Leu-	Ser-	Tyr-	SCMCys-	Leu-	Glu-	Val-	Ala-	Leu-	Ala-	Val-	His-	Leu-	Thr-	Glu-	Phe-	Ser-	Pro-	Glu-	Thr-	His-	SCMCys-	Ala-	Leu-	Asp-	Lys-	(Phe)-
		Val	*Asp*	*Pro*	*Ala*	*Asn*	*Phe*						*SCMCys*	*Leu*	*Glu*	*Val*	*Ala*	*Leu*								*Ser*	*Pro*	*Glu*	*Thr*	His				*Asp*	*Lys*	
																			Ala	*Val*	His	*Leu*	*Thr*	*Glu*	*Phe*						*SCMCys*	*Ala*	*Leu*			
	Lys	*Val*	*Asp*	*Pro*	*Ala*	*Asn*	*Phe*																								*SCMCys*	*Ala*	*Leu*	*Asp*	*Lys*	
Chymotryptic digest of intact α-chain				αCh17						αCh18					αCh19							αCh20						αCh21					αCh22			
	Lys	*Val*	*Asp*	*Pro*	*Ala*	*Asn*	*Phe*						*SCMCys*	*Leu*	*Glu*	*Val*	*Ala*	*Leu*								*Ser*	*Pro*	*Glu*	*Thr*	His						
								Gln	*Tyr*	*Leu*	*Ser*	*Tyr*							*Ala*	*Val*	His	*Leu*	*Thr*	*Glu*	*Phe*						*SCMCys*	*Ala*	*Leu*	*Asp*	*Lys*	*Phe*
								αCh18*N*																							αCh22*N*					
								Gln	Tyr																						*SCMCys*	*Ala*	*Leu*			
										αCh18*C*																								αCh22*C*		
										Leu	Ser	Tyr																						*Asp*	*Lys*	*Phe*
Peptic digest of tryptic core	*Lys*	*Val*	*Asp*	*Pro*	*Ala*	*Asn*	*Phe*			*Leu*	*Ser*	*Tyr*				*Val*	*Ala*	*Leu*							*Phe*	*Ser*	*Pro*	*Glu*	*Thr*	His	*SCMCys*	*Ala*	*Leu*			
	Lys	*Val*	*Asp*				*Phe*	*Gln*	*Tyr*				*SCMCys*	*Leu*	*Glu*				*Ala*	*Val*	His													*Asp*	*Lys*	
														Leu	*Glu*	*Val*	*Ala*		*Ala*	*Val*	His	*Leu*	*Thr*	*Glu*										*Asp*	*Lys*	*Phe*
															Glu	*Val*	*Ala*																			
Thermolysin digest of tryptic core		*Val*	*Asp*	*Pro*	*Ala*	*Asn*				*Leu*	*Ser*	*Tyr*	*SCMCys*									*Leu*	*Thr*	*Glu*												
							Phe	*Gln*	*Tyr*					*Leu*	*Glu*			*Leu*	*Ala*																	
Thermolysin digest of chymotryptic core							*Phe*	*Gln*	*Tyr*							*Val*	*Ala*			*Val*	His				*Phe*	*Ser*	*Pro*	Glu	Thr	His	SCMCys	Ala				
	Lys	*Val*	*Asp*	*Pro*	*Ala*	*Asn*				*Leu*	*Ser*	*Tyr*	*SCMCys*					*Leu*	*Ala*			*Leu*	*Thr*	*Glu*									*Leu*	Asp	Lys	
														Leu	*Glu*																					

123 and 127 of glutamyl residues. Position 133 is occupied by an aspartyl residue as αCh22 is acidic and αCh22*C* is neutral.

αTp13

This peptide had the sequence

Phe-Leu-Thr-Asn-Val-SCMCys-His-Glu-Leu-Ser-Ser-Arg.
135 146

A thermolysin digest of the peptide yielded the peptides *Leu*-Thr-Asn which was neutral at pH 6·4 and Val-SCMCys-His-Glu which was acidic. Position 138 is therefore occupied by asparagine and position 142 by glutamic acid. The chymotryptic peptide αCh22 had the sequence *SCMCys-Ala-Leu-Asp-Lys-Phe* and provided the overlap with αTp11+12.

αTp14

This peptide had the sequence

Tyr-Arg.
147 148

The chymotryptic peptide αCh24 had the sequence *Ser-Ser*-Arg-*Tyr* and provided the overlap with αTp13.

Complete Sequence

The complete amino acid sequence of the shark α-chain is shown in Table 4.

Date of Divergence

The times of divergence of the shark from eight vertebrates were calculated from the number of differences in amino acids in equivalent positions (Table 5) after alignment according to Dayoff (1972); or from the PAM method (Dayhoff 1972) or random evolutionary hits (Jukes and Holmquist 1972) shown in Table 6. Three separate methods were used and the results are presented in Table 7 and include the mean date of divergence for each method, with the 95% confidence interval about the mean. Values from the comparison between the shark, the carp, the sucker and the lamprey were not included in the estimation of the date of divergence. From Table 7 it appears that the shark last shared a common ancestor with the higher vertebrates approximately 410 million years ago. If one accepts the paleontological estimates of the times of divergence of the carp and lamprey, namely 425 and 470 million years respectively, then the elasmobranchs appear to post-date these animals.

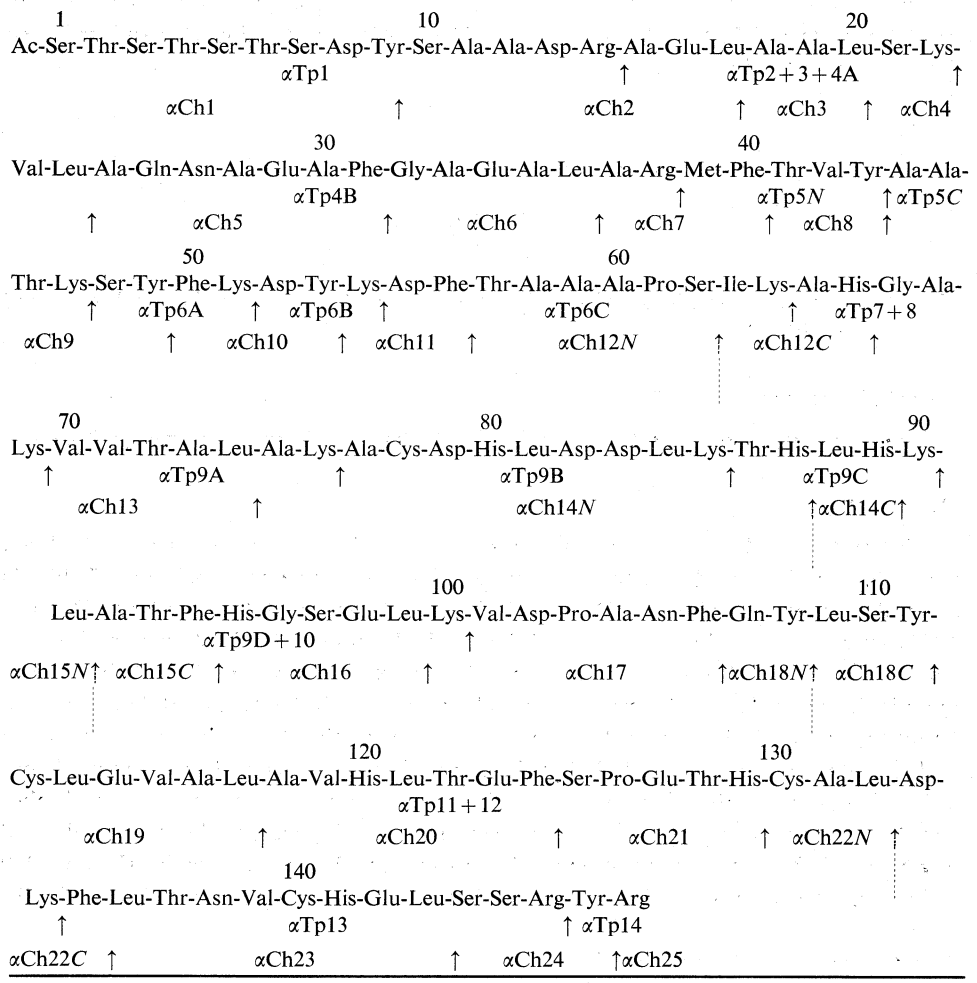
Construction of a Phylogenetic Tree

Using the minimum mutation distances between the human, carp, shark and lamprey it is possible to construct three unrooted phylogenetic trees (Figs 8A, 8B, 8C and Table 8). Cavalli-Sforza and Edwards (1967) and Kidd and Sgaramella-Zonta (1971) claim that the correct tree will contain no negative branch lengths and will have the minimum residual sum of squares. With these restrictions, tree B (Fig. 8B) appears to be the correct topology although tree C (Fig. 8C) satisfies the above

criteria when REH values are used instead of miminum mutation distances. It is possible, from paleontological evidence, to infer a root for these trees and they may be redrawn in the more usual form (Figs 8D, 8E). Substitution of mutation distances between the shark and any other mammal did not alter the form of the trees.

Table 4. Amino acid sequence of the α-chain of shark haemoglobin

Residues are numbered from the *N*-terminal residue and the positions of cleavage by trypsin and chymotrypsin shown by arrows. The dashed arrows indicate an incomplete cleavage. Where a split has occurred with trypsin that does not involve a basic amino acid residue, or where a partial split has occurred with chymotrypsin, the *N*-terminal portion is given the suffix *N* and the *C*-terminal portion the suffix *C* in *italic* fount. Ac represents an acetyl group



Discussion

The α-chain of the haemoglobin of *H. portusjacksoni* is demonstrably a vertebrate α-chain but it possesses several features which are relatively rare.

The α-amino group of the *N*-terminal serine is blocked by an acetyl group. The *N*-terminal serine of the α-chains of the carp (Hilse and Braunitzer 1968) and a

catostomid fish (Powers and Edmundson 1972) are both acetylated, as are the *N*-terminal residues of eel haemoglobin (Amano *et al.* 1968) and frog α-chains (Chauvet and Acher 1971; DeWitt and Ingram 1967). On the other hand, Fujiki *et al.* (1970) reported that about 80% of lamprey (*Lampetra* sp.) globin is formylated at the *N*-terminal proline. Among the lower vertebrates then, acylation of the amino terminal appears to be common, although the presence of carbohydrate cannot be excluded (Bookchin and Gallop 1968; Rahbar *et al.* 1969).

Table 5. Matrices of differences between α-globins

The matrix of observed amino acid differences between α-globin sequences compared over 143–148 residues is shown in the upper triangle. Minimum mutation distances are shown in the lower triangle. For the purposes of comparison the additional seven *N*-terminal residues in shark and eight *N*-terminal residues in lamprey chain are counted as only one amino acid difference (one genetic event)

	Human	Rabbit	Dog	Ox	Horse	Kangaroo	Chicken	Viper	Carp	Sucker	Shark	Lamprey
Human		25	23	17	18	27	35	50	71	68	82	95
Rabbit	27		27	25	25	37	43	52	74	71	78	99
Dog	25	29		28	27	33	44	54	70	69	83	96
Ox	20	28	31		18	26	38	53	68	66	78	96
Horse	22	30	31	23		29	40	52	70	67	80	96
Kangaroo	34	44	40	32	35		41	54	74	72	83	100
Chicken	45	55	55	50	51	54		57	75	70	86	104
Viper	74	74	77	75	74	78	81		76	75	86	102
Carp	96	96	95	95	97	104	104	110		14	89	100
Sucker	91	92	92	91	93	100	98	106	17		90	100
Shark	108	105	110	103	107	110	114	116	124	125		116
Lamprey	123	133	128	129	130	138	140	133	137	139	162	

Table 6. Matrices of differences between α-globins

The matrix of calculated point accepted mutations (PAM units per 100 links) compared over 143–148 residues is shown in the upper triangle. In the lower triangle the calculated number of random evolutionary hits per chain is shown. The amino acid differences in Table 5 were the bases for these values

	Human	Rabbit	Dog	Ox	Horse	Kangaroo	Chicken	Viper	Carp	Sucker	Shark	Lamprey
Human		18	16	12	13	19	25	35	50	48	59	72
Rabbit	43		19	18	18	26	30	37	52	50	56	76
Dog	41	45		20	19	23	31	38	50	49	59	73
Ox	38	49	53		13	18	27	38	48	47	56	73
Horse	46	60	56	55		20	28	37	50	47	57	73
Kangaroo	78	87	82	69	71		29	38	52	41	59	76
Chicken	109	132	123	131	121	142		40	53	50	61	79
Viper	120	276	290	271	276	265	299		54	53	61	78
Carp	278	239	279	318	312	361	318	424		10	63	76
Sucker	253	229	252	286	302	325	312	356	35		64	76
Shark	301	318	313	290	315	313	324	352	414	430		84
Lamprey	352	435	408	424	442	507	442	374	486	528	565	

Hydrazinolysis, using the method of Phillips (1963), will result in the formation of acyl hydrazides regardless of their position on the protein. Similarly, acid hydrolysis followed by determination of the liberated fatty acid by either gas–liquid chromatography or micro-enzymic methods (Stegink 1967) will not specify the site of attach-

Table 7. Dates of divergence of the shark

Estimates of the dates of divergence (millions of years) of the shark from other vertebrates using three different methods outlined in the text. Method 1, Air *et al.* (1971) using corrected REH values; method 2, Markusson and Volund (1974); method 3, Dayhoff (1972)

Divergence point	Method 1	Method 2	Method 3
Human/Shark	390	373	434
Rabbit/Shark	413	392	387
Dog/Shark	407	416	434
Ox/Shark	376	383	387
Horse/Shark	409	402	403
Kangaroo/Shark	407	416	434
Chicken/Shark	420	435	469
Viper/Shark	456	445	469
Mean ± 95% confidence interval	409 ± 45	407 ± 49	427 ± 65

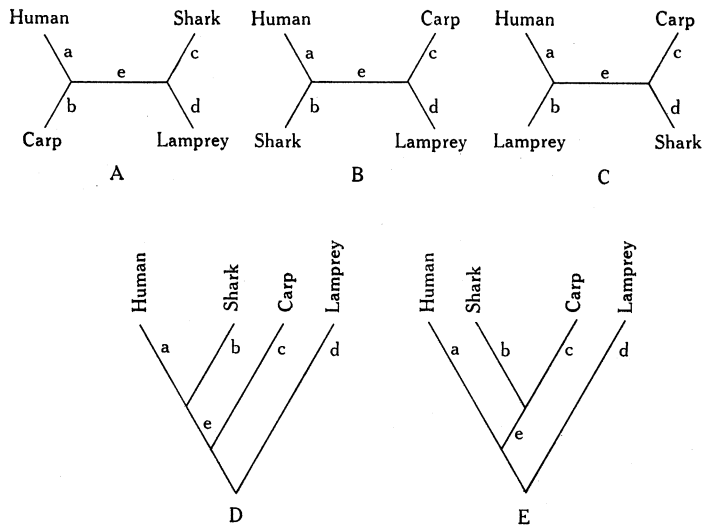


Fig. 8. A, B and C represent unrooted phylogenetic trees where a, b, c, d and e are the genetic distances along the branches. Branch lengths for each tree are found in Table 8. D and E are traditional representations of the phylogenetic relationships between the four animals. In these trees the trunk has been inferred from paleontological evidence.

ment to the protein of the acyl group. Jornvall (1967) has shown that acetate can bind non-covalently to globin chains in other than the *N*-terminal position and therefore the acylated *N*-terminal amino acid or peptide should be identified unambiguously. Mass spectrometry of a permethylated peptide fraction gave a positive identification

that an acetyl-Ser-Thr-Ser sequence is present in shark α-chain although no quantitative estimate of the acetyl blocking group was obtained.

It has been shown that blocking of the α-amino groups of the α-chains in haemoglobin reduces the Bohr effect by 25% (Kilmartin and Rossi-Bernardi 1959). Perutz *et al.* (1969) predicted, from X-ray studies of the haemoglobin molecule, that the *C*-terminal histidine of the β-chain was responsible for about 50% of the Bohr effect.

Table 8. Branch lengths of phylogenetic trees

Estimates for branch lengths of the trees shown in Fig. 8. Also shown are the residual error squares. Values given to the nearest integer. MMD represents minimum mutation distances, REH represents random evolutionary hits

Branch	Tree A		Tree B		Tree C	
	MMD	REH	MMD	REH	MMD	REH
a	40	77	17	63	38	58
b	56	201	71	238	53	182
c	74	252	52	187	71	232
d	88	313	85	299	85	294
e	−6	−33	4	9	2	24
Σ (error)^2	1	110	30	1482	42	784

Therefore a tetramer with blocked α-chains and a substitution in the *C*-terminal of the β-chain should exhibit no Bohr effect. This prediction has been substantiated by the work of Powers and Edmundson (1972) who found that the isolated cathodal components of the haemoglobin of the sucker fish satisfied the above criteria and indeed exhibited no alkaline Bohr effect. The acetyl group on the amino terminal residue of shark haemoglobin would be expected to reduce the Bohr effect. The β-chain of the shark terminates in the sequence Glu-Tyr-His yet despite the presence of the histidyl residue the intact haemoglobin exhibits little or no Bohr effect (unpublished data). One possible explanation is the absence of an acidic residue in the β-chain corresponding to the aspartyl residue at position 94 that forms a salt link with the *C*-terminal histidyl residue 146 in deoxygenated haemoglobins that exhibit the Bohr effect.

The shark α-chain, with 148 residues, is the longest α- or β-globin chain sequenced to date. The extra residues occur in the unique, repeating Ser-Thr sequence on the *N*-terminal end of the chain. The lamprey globin also has extra residues so that these extra residues seem to represent the primitive vertebrate condition and hence their presence in the shark could possibly place elasmobranchs phylogenetically closer to the Agnatha than to the Teleostii. Perutz *et al.* (1968) showed that the *N*-terminal valine of horse haemoglobin was free to interact with the aqueous environment, and hence one could assume that the *N*-terminal residues of shark α-globin are similarly free, particularly as these residues are polar. There is no proline in αTp1 and therefore the A helix could include all of the extra residues, or they may simply exist as a random coil as in lamprey globin (Hendrickson and Love 1971). In Table 9 the alignment of the amino acid residues in lamprey globin, horse α-, carp α- and shark α-chains is shown using the one-letter code for the amino acids and the alignments shown in Dayhoff (1972).

There are four cysteinyl residues in the shark α-chain, one more than the highest previously recorded for a vertebrate, namely the rat (Chua and Carrel 1974). Cys

Table 9. Comparison of the amino acid sequence of shark α-chain with that of horse α-, carp α- and lamprey globin

The helical regions are indicated and deletions are shown as dashes. For blocked amino terminal residues, *Ac* represents acetyl groups and *For* formyl groups. The interhelical regions are normally designated AB, CD, EF, FG, and GH, while NA denotes the *N*-terminal section and HC the *C*-terminal section. The one-letter code and alignment of the chains follows Dayoff (1972)

Globin		A helix		B helix	C helix
Horse α	V -	LSAADKTNVKAAWSKVG	G	HAGEYGAEALERMFLG	FPTTKTY
Carp α	*Ac* S -	LSDKDKAAVKIAWAKIS	P	KADDIGAEALGRMLTV	YPQTKTY
Shark α	*Ac* STSTSTSD -	YSAADRAELAALSKVLA	Q	NAEAFGAEALARMFTV	YAATKSY
Lamprey	*For* PIVDSGSVAP	LSAAEKTKIRSAWAPVY	S	NYETSGVDILVKFFTS	TPAAQEF

		D helix	E helix
Horse α	FPHF-DLS	H-----G	SAQVKAHGKKVA-DGLTLAVG
Carp α	FAHWADLS	P-----G	SGPVK-HGKKVIMGAVGDAVS
Shark α	FKDYKDFT	A-----A	APSIKAHGAKVV-TALAKACD
Lamprey	FPKFKGMT	SADELKK	SADVRWHAERII-NAVNDAVA

		F helix		G helix
Horse α	HLDD-LPGA	LSDLSNLHA	HKLRV	DPVNFKLLSHCLLSTLAVH
Carp α	KIDD-LVGG	LASLSELHA	SKLRV	DPANFKILANHIVVGIMFY
Shark α	HLDD-LKTH	LHKLATFHG	SELKV	DPANFQYLSYCLEVALAVH
Lamprey	SMDDTEKMS	MKDLSGKHA	KSFQV	DPQYFKVLA-VI-------

		H helix	
Horse α	LPNDF	TPAVHASLDKFLSSVSTVLTS	KYR
Carp α	LPGDF	PPEVHMSVDKFFQNLALALSE	KYR
Shark α	L-TEF	SPETHCALDKFLTNVCHGLSS	RYR
Lamprey	--ADT	VAAGDAGFEKLSMICILMLRS	AY-

112 (G11) is the only cysteine common to other vertebrate α-chains. Of the four cysteinyl residues, Cys 78 (E19), Cys 112 (G11) and Cys 140 (H16) are probably buried in the interior of the molecule (Perutz *et al.* 1965) and should play no part in the disulphide-bridge formation which is a characteristic of shark haemoglobins (Martin 1972; Nash and Thompson 1974; Fyhn and Sullivan 1975). The remaining residue, Cys 130 (H6), could be exposed and represent a reactive thiol in the shark haemoglobin. This has yet to be proved, however, as there are two cysteinyl residues in the β-chain and either or both of these residues may also be exposed and reactive in the native haemoglobin.

If one assumes that the shark α-chain has the same tertiary structure as that of the horse, then there are some interesting substitutions in the contact sites. In the important $\alpha_1\beta_2$ contact sites there are four substitutions in the oxyhaemoglobin form, at positions C3, C6, FG4 and G3. The substitutions at C6 (Thr→Ser) and FG4 (Arg→Lys) are relatively rare, glutamic acid and glutamine being the respective replacements in the lamprey. The remaining two positions are quite variable: C3, glutamine in carp and sucker but alanine in lamprey and shark; G3, alanine in shark, viper and carp but serine in sucker and glutamine in lamprey. In deoxyhaemoglobin there are six substitutions in the $\alpha_1\beta_2$ contact sites. In addition to the four already mentioned the conformational changes brought about by the loss of an oxygen molecule bring another two α-chain residues, C2 and CD2, into contact with the β2 chain. The CD2 site (Pro→Lys) is variable, with alanine in carp and sucker but serine in monotreme α-chains. It interacts in horse haemoglobin with a histidine (FG4) from the β-chain. It is an unusual substitution in that a polar residue is brought into a normally non-polar interacting site. The C2 substitution (Pro→Ala) has not been previously recorded and interaction with the *C*-terminal histidine (HC3) of β-chains is involved.

Of the 16 $\alpha_1\beta_1$ contacts listed by Perutz (1969), nine have substitutions in the shark: B11, B15, B16, C1, G10, G13, G14, GH2 and H6. Most of these sites are variable in different species but a change at H6 (Ala→Cys) has been recorded only once before, when the substitution was Ala→Met in the carp.

There are 19 haem contact sites in the horse α-chain. Of these, three vary in the shark α-chain: CD3 (His→Asp), CD4 (Phe→Tyr) and F7 (Leu→Phe). The last two are isopolar substitutions at sites where such substitutions have been recorded for other vertebrates. However, the substitution at CD3 is not isopolar and represents a change of charge at physiological pH. The site is on the surface of the molecule and the contribution of the interaction to the overall binding energy is probably quite small.

Perutz (1969) considers that tryptophan at A12 is hydrogen bonded to threonine at E16 and acts as a spacer between helices A and E. In the shark α-chain there is no tryptophan and a much smaller seryl residue is present at A12 while E16 is occupied by alanine, hence the tertiary structure may be expected to vary as a result of these changes.

Attention should be drawn to the insertion of a lysyl residue at position CD5 corresponding to that in carp α-chain and the absence of an insertion at E13 where there is one in carp α-chain, using the alignment of globin family sequences given by Dayhoff (1972). Two residues of glutamic acid occur in shark α-chain at positions usually occupied by hydrophobic residues. The first of these at H18, usually valine

in α-chains, is not invariably hydrophobic since aspartic acid is common in myoglobins. The second glutamyl residue at G13 has been invariably hydrophobic in globin chains but is not listed as an internal residue by Perutz *et al.* (1965).

A deletion occurs in the GH interhelical region of the α-chain corresponding to position GH2 which is usually occupied by a prolyl residue.

Inherent in the use of amino acid sequences to establish times of divergence is the assumption of a constancy of evolutionary rate, between taxa, in the protein being used. Romero-Herrera *et al.* (1973), and Wu *et al.* (1974) are among many workers who claim that the mutational rate not only varies between lineages but also varies within a single lineage. Consequently times of divergence estimated from data which include the chicken and viper α-chains, as this present study does, will have extremely large confidence intervals (cf. Whittaker and Thompson 1975). The ranges of the estimated times of divergence tend to overlap (e.g. shark and carp) to such a degree that precise relationships are impossible to define.

Either of the two proposed phylogenetic trees (Figs 8D, 8E) are in accord with the data. Obviously the sharks and teleosts have diverged from the main vertebrate line almost simultaneously but until more data become available it is impossible to delineate the true phylogeny. Clearly many more orthologous sequences from the representative animals are required before the full potential of the methods can be realized. Certainly none of the results of this study is inconsistent with the traditional viewpoint that bony fish arose prior to the elasmobranchs.

Acknowledgments

This work was supported in part by the Australian Research Grants Committee. The authors are indebted to Dr J. K. Allen and Dr R. H. Crozier for their helpful discussions and assistance with the computer programming, to Mr K. Bollinger for his assistance in capturing sharks, to Mr R. Mann for skilled technical assistance, and to Dr K. D. Barrow, Dr I. Brown and Mr M. Smythe for their help in identifying the *N*-terminal acetyl group.

References

Acher, R. (1974). *Angew. Chem. Int. Ed. Engl.* **13**, 186–97.
Air, G. M., and Thompson, E. O. P. (1969). *Aust. J. Biol. Sci.* **22**, 1437–54.
Air, G. M., and Thompson, E. O. P. (1971). *Aust. J. Biol. Sci.* **24**, 75–95.
Air, G. M., Thompson, E. O. P., Richardson, B. J., and Sharman, G. B. (1971). *Nature* (*London*) **229**, 391–4.
Amano, H., Hashimoto, K., and Matsuura, F. (1968). *Bull. Jap. Soc. Sci. Fish.* **34**, 937–43.
Arnheim, N. (1973). In 'The Antigens'. (Ed. M. Sela.) pp. 377–416. (Academic Press: New York.)
Barnabas, J., Goodman, M., and Moore, G. W. (1971). *Comp. Biochem. Physiol. B* **39**, 455–82.
Beard, J. M., and Thompson, E. O. P. (1971). *Aust. J. Biol. Sci.* **24**, 765–86.
Beaven, G. H., and Holiday, E. R. (1952). *Adv. Protein Chem.* **7**, 319–86.
Bookchin, R. M., and Gallop, P. M. (1968). *Biochem. Biophys. Res. Commun.* **32**, 86–93.
Bricteux-Gregoire, S., Schyns, R., and Florkin, M. (1972). *Comp. Biochem. Physiol. B* **42**, 23–39.
Bricteux-Gregoire, S., Schyns, R., and Florkin, M. (1974). *Biochem. Biophys. Acta* **351**, 87–91.
Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossman, M. G. (1973). *Proc. Nat. Acad. Sci. U.S.A.* **70**, 3052–4.
Buettner-Janusch, J., and Hill, R. L. (1965). *Science* **147**, 836–42.
Cavalli-Sforza, L. L., and Edwards, A. W. F. (1964). In *Proc. 11th Int. Congr. Genet.* **3**, 923–33.
Cavalli-Sforza, L. L., and Edwards, A. W. F. (1967). *Am. J. Hum. Genet.* **19**, 233–57.

Chauvet, J. P., and Acher, R. (1971). *Int. J. Protein Res.* **3**, 261–9.
Chua, C. G., and Carrell, R. W. (1974). *Biochim. Biophys. Acta* **365**, 328–34.
Dayhoff, M. O. (1972). 'Atlas of Protein Sequence and Structure'. (National Biomedical Research Foundation: Silver Spring.)
DeWitt, W., and Ingram, V. M. (1967). *Biochem. Biophys. Res. Commun.* **27**, 236–41.
Dickerson, R. E. (1971). *J. Mol. Evol.* **1**, 26–45.
Fitch, W. M. (1973). *Annu. Rev. Genet.* **7**, 343–80.
Fitch, W. M., and Margoliash, E. (1967). *Science* **155**, 279–84.
Fitch, W. M., and Margoliash, E. (1970). In 'Evolutionary Biology'. (Eds T. Dobzhansky, M. K. Hecht and W. C. Steere.) Vol. 4, pp. 67–109. (Appleton-Century-Crofts: New York.)
Folsch, G., Ryhage, R., and Stenhagen, E. (1962). *Ark. Kemi* **20**, 55–62.
Fujiki, H., Braunitzer, G., and Rudloff, V. (1970). *Hoppe-Seyler's Z. Physiol. Chem.* **351**, 901–2.
Fyhn, U. E. H., and Sullivan, B. (1975). *Comp. Biochem. Physiol. B* **50**, 119–29.
Gerald, P. S., and Ingram, V. M. (1961). *J. Biol. Chem.* **236**, 2155–6.
Goodman, M., Barnabas, J., Matsuda, G., and Moore, G. W. (1971). *Nature (London)* **233**, 604–13.
Hendrickson, W. A., and Love, W. E. (1971). *Nature (London)* **232**, 197–203.
Hilse, K., and Braunitzer, G. (1968). *Hoppe-Seyler's Z. Physiol. Chem.* **349**, 433–50.
Holmquist, R., Cantor, C., and Jukes, T. (1972). *J. Mol. Biol.* **64**, 145–61.
Jornvall, H. (1967). *Acta Chem. Scand.* **21**, 1805–11.
Jukes, T. H., and Holmquist, R. (1972). *J. Mol. Biol.* **64**, 163–79.
Kang, A. H., Bornstein, P., and Piez, K. A. (1967). *Biochemistry* **6**, 788–95.
Kawasaki, I., and Itano, H. A. (1972). *Anal. Biochem.* **48**, 546–56.
Kidd, K. K., and Sgaramella-Zonta, L. A. (1971). *Am. J. Hum. Genet.* **23**, 235–52.
Kilmartin, J. V., and Rossi-Bernadi, L. (1969). *Nature (London)* **222**, 1243–6.
Margoliash, E., Fitch, W. M., and Dickerson, R. E. (1971). In 'Biochemical Evolution and the Origin of Life'. (Ed. E. Schoffeniels.) (North Holland Publishing Co.: Amsterdam.)
Markussen, J., and Volund, A. (1974). *Int. J. Protein Res.* **6**, 79–86.
Martin, F. (1972). *Biochimie* **54**, 25–30.
Morris, H. R., Williams, D. H., and Ambler, R. P. (1971). *Biochem. J.* **125**, 189–201.
Narita, K. (1958*a*). *Biochim. Biophys. Acta* **28**, 184–91.
Narita, K. (1958*b*). *Biochim. Biophys. Acta* **30**, 352–9.
Nash, A. R., and Thompson, E. O. P. (1974). *Aust. J. Biol. Sci.* **27**, 607–15.
Nolan, C., Fitch, W. M., Uzzell, T., Weiss, L. J., and Margoliash, E. (1973). *Biochemistry* **12**, 4052–60.
Perutz, M. F. (1969). *Proc. R. Soc. B* **173**, 113–40.
Perutz, M. F., Kendrew, J. C., and Watson, H. C. (1965). *J. Mol. Biol.* **13**, 669–78.
Perutz, M. F., Muirhead, H., Cox, J. M., and Goaman, L. C. G. (1968). *Nature (London)* **219**, 131–9.
Perutz, M. F., Muirhead, H., Mazzarella, L., Crowther, R. A., Greer, J., and Kilmartin, J. V. (1969). *Nature (London)* **222**, 1240–3.
Phillips, D. M. P. (1963). *Biochem. J.* **86**, 397–401.
Powers, D. A., and Edmundson, A. B. (1972). *J. Biol. Chem.* **247**, 6694–707.
Rahbar, S., Blumenfeld, O., and Ranney, H. M. (1969). *Biochem. Biophys. Res. Commun.* **36**, 838–43.
Reindel, F., and Hoppe, W. (1954). *Chem. Ber.* **87**, 1103–7.
Romer, A. S. (1966). 'Vertebrate Paleontology'. 3rd Edn. (Chicago University Press.)
Romero-Herrera, A. E., Lehmann, H., Joysey, K. A., and Friday, A. E. (1973). *Nature (London)* **246**, 389–95.
Schmer, G., and Kreil, G. (1969). *Anal. Biochem.* **29**, 186–92.
Sheehan, J. C., and Yang, D. D. H. (1958). *J. Am. Chem. Soc.* **80**, 1154–8.
Stahl, B. J. (1974). 'Vertebrate History: Problems in Evolution'. (McGraw-Hill Inc.: New York.)
Stegink, L. D. (1967). *Anal. Biochem.* **20**, 502–16.
Thompson, E. O. P., Hosken, R., and Air, G. M. (1969). *Aust. J. Biol. Sci.* **22**, 449–62.
Wadke, M., and Lowenstein, J. M. (1975). In 'Methods in Enzymology'. Vol. XXXV. (Ed. J. M. Lowenstein.) (Academic Press: New York.)
Whittaker, R. G., and Thompson, E. O. P. (1975). *Aust. J. Biol. Sci.* **28**, 353–65.
Wu, T. T., Fitch, W. M., and Margoliash, E. (1974). *Annu. Rev. Biochem.* **43**, 539–66.

Manuscript received 9 October 1975