STUDIES ON MARSUPIAL PROTEINS

V.* AMINO ACID SEQUENCE OF THE α-CHAIN OF HAEMOGLOBIN FROM THE GREY KANGAROO, *MACROPUS GIGANTEUS*

By J. M. BEARD† and E. O. P. THOMPSON†

[Manuscript received February 17, 1971]

Abstract

The amino acid sequence of the α-chain of haemoglobin from *M. giganteus* has been determined. The soluble peptides formed by tryptic digestion were isolated by gel filtration, ion-exchange chromatography, paper ionophoresis, and chromatography. The amino acid sequences were determined by the "dansyl"–Edman procedure. Incomplete hydrolysis of one bond resulted in a large insoluble-core peptide containing 40 amino acid residues. The sequence of this peptide was deduced from the sequences of smaller peptides resulting from further digestion with thermolysin and papain. Maleylation of the α-globin before tryptic digestion gave three large fragments which assisted in assigning tryptic peptides to specific areas of the molecule. A special procedure involving maleylation of a chymotryptic digest of globin was used to isolate peptides containing arginine which provided overlap sequences of tryptic peptides.

The number of changes in sequence is compared with those known in other α-globin chains and may be used for evolutionary rate calculations. The number of differences between the α-globin, β-globin, and myoglobin chains of kangaroo are reported.

Of the amino acid residues in the α-chain of horse haemoglobin which are in contact with the haem group or the β-chains (listed by Perutz 1969), those involving α-haem showed only two changes whereas the α_1–β_2 contacts are identical with those residues in the α-chain of kangaroo haemoglobin. For those residues in contact between α_1–β_1 chains there is more variation, with five changes in the 16 residues as compared with horse haemoglobin.

I. INTRODUCTION

In a previous paper (Thompson, Hosken, and Air 1969) the separation of the α- and β-chains of globin from the grey kangaroo *Macropus giganteus* was reported. The β-chains were polymorphic and the sequences of these chains have been published (Air and Thompson 1969). The *N*-terminal sequence of the α-globin has been reported (Beard and Thompson 1970) and in this paper the complete amino acid sequence is given.

* Part IV, *Aust. J. biol. Sci.*, **24**, 75–95.

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

The methods used were similar to those employed for the β-chain so far as soluble tryptic peptides were concerned. Fractionation of these peptides involved gel filtration, ion-exchange chromatography on sulphonated polystyrene resin, paper ionophoresis, and paper chromatography. The sequences were determined by the "dansyl"–Edman procedure.

For the "insoluble core" peptides a variety of techniques was used. Amino acid analysis and *N*-terminal sequence analysis of the washed insoluble core showed that it was a 40-residue peptide containing two lysine residues. That is, two potential tryptic peptides remained joined by a lysyl bond incompletely cleaved by the enzyme. The sequence of this peptide was deduced from the sequences of smaller peptides obtained after further digestion with thermolysin and papain. Labelling the single cysteine residue in the α-chain with [^{14}C]iodoacetate was useful in studying the fractionation of enzyme digests since this cysteine residue is present in the insoluble core peptide fraction.

Maleylation of the α-globin before tryptic digestion facilitated the selective cleavage of the two arginyl bonds (one arginine residue is *C*-terminal) so that only three maleylated fragments were obtained. The isolated peptides gave information on the location of tryptic peptides in the α-globin chain; a special procedure involving maleylation of a chymotryptic digest of kangaroo globin was used to isolate peptides containing arginine which provide overlap sequences.

II. Materials and Methods

The *S*-carboxymethylated (SCM) α-chain of kangaroo haemoglobin was prepared as described previously (Thompson, Hosken, and Air 1969). Globin was prepared from two different grey kangaroos—one a β-I homozygote and the other a β-I-II heterozygote. The α-chain prepared from the heterozygote was more heavily contaminated with β-chain but was adequate for use in sequence studies as the β-sequence was already known (Air and Thompson 1969).

For work on the insoluble-core peptides the α-globin was prepared from the homozygote β-I animal as this reduced the contamination with β-globin, which also gives several insoluble tryptic peptides.

The methods of tryptic and chymotryptic digestion, high-voltage paper ionophoresis, peptide mapping, amino acid analysis, and sequence determination by the dansyl–Edman procedure were the same as previously described (Air and Thompson 1969, 1971).

(*a*) *Fractionation of Tryptic Peptides*

(i) *Gel Filtration*

The SCM-globin was not completely solubilized by 4 hr tryptic digestion at 37°C. The mixture, pH 8·7, was filtered and the insoluble material discarded. The supernatant was adjusted to pH 6 and the material which precipitated removed by centrifugation and washing with pyridine–acetate buffer (pH 6·4) and water. The supernatant and washings were freeze-dried and redissolved in 0·1M NH_3. A small amount of insoluble material was removed by centrifugation and the peptide mixture loaded on to a 2·3 by 120 cm column of Sephadex G25 in 0·01M NH_3. The extinction of the eluate at 230 and 280 nm was measured and seven fractions bulked and freeze-dried. Each fraction was examined by dissolving a portion in pyridine–acetate buffer (pH 6·4) for peptide mapping by paper techniques.

(ii) *Ion-exchange Chromatography*

Fractionation of soluble peptides on sulphonated polystyrene resin (Aminex MS, Q-15 blend; Bio-Rad Laboratories) was carried out as described by Air and Thompson (1969).

(iii) *Paper Ionophoresis and Chromatography*

Preparative high-voltage ionophoresis was carried out as described by Air and Thompson (1969). If a two-dimensional separation was required on a preparative scale each zone located after the first dimension was sewn on to another sheet of Whatman 3MM paper. The paper was wetted carefully from each side of the band allowing the solvent to wet the area containing the peptide by capillarity, thus concentrating the zone. A second pH could then be used for fractionation.

Paper chromatography using butanol–acetic acid–water–pyridine, 15 : 3 : 12 : 10 (Thompson, Hosken, and Air 1969) was used as the first, second, or third dimension where necessary. Elution of peptides from these large strips was done with pyridine–acetate buffer, pH 6·4 (Air and Thompson 1969, 1971).

(*b*) *Further Digestion of Tryptic Peptides*

All digestions were at 37°C for 3–4 hr.

Thermolysin (Chugai Boyaki Co., Osaka, Japan) digestions were carried out in 0·2M ammonium acetate, pH 8·5, containing 2·5 mM $CaCl_2$ (Ambler and Meadway 1968). Peptide (0·5 μmole) was dissolved in 0·5 ml buffer and 0·05 mg enzyme added in an equal volume of buffer. After digestion the mixture was freeze-dried and then redissolved in water for repeated redrying. This was necessary to remove ammonium acetate prior to fractionation by paper ionophoresis.

Digestions with chymotrypsin (Worthington) were carried out in 1% ammonium bicarbonate, pH 8·7, containing 0·1% phenol, using 0·2 mg enzyme per μmole peptide in a volume of 1 ml.

Papain (Sigma Chemical Co.) was supplied as a suspension containing 2·7 mg enzyme per millilitre of 0·05M sodium acetate, pH 4·5. Prior to use, the enzyme was activated by adding 10 μl of this suspension to 2·5 ml 10% pyridine containing 0·005M ethylenediamine tetraacetic acid (EDTA) and 0·05M mercaptoethanol, pH 7·4, and incubating at 37°C for 1 hr. 0·2 ml of activated enzyme solution was added to the peptide (10 mg) suspended in 0·8 ml 10% pyridine–0·05M mercaptoethanol titrated to pH 7 with 1M acetic acid, and digestion allowed to proceed for 4 hr at 37°C.

After drying, the digests were dissolved in 50% pyridine (50 μl/2 mg), centrifuged to remove any insoluble material, and fractionated by the paper techniques already described. Peptides were detected with 0·02% ninhydrin and the paper washed with acetone before elution with 50% pyridine using the technique of Sanger and Tuppy (1951). For sequence determination the dansyl–Edman procedure was used. The coupling of phenylisothiocyanate with peptide was done on the 50% pyridine eluate and recoupling, after drying, ensured complete reaction. In this way loss of α-amino groups during drying in the presence of traces of ninhydrin are minimized.

(*c*) *^{14}C-Labelling of α-Globin*

The α-globin (40 mg) dissolved in 8M urea–0·07M Tris buffer (5 ml) containing 0·001M EDTA and 50 μl mercaptoethanol was stood at 4°C for 16 hr under nitrogen. The solution was dialysed against three changes of 8M urea–0·001M EDTA at pH 3.25 under nitrogen before carboxymethylating the thiol groups with [2-^{14}C]iodoacetate (2 μmoles) at pH 8·5 for 15 min. To ensure complete carboxymethylation the protein was treated with additional mercaptoethanol (50 μl) at room temperature for 1 hr and cold iodoacetic acid (135 mg) added. Protein was recovered after dialysis against water and freeze-drying. The specific activity of the ^{14}C-labelled globin was 4×10^5 counts per minute per milligram.

(*d*) *Tryptic Digestion of Maleylglobin*

Maleylated [^{14}C]-*S*-carboxymethylated globin (325 mg including 5 mg [^{14}C]iodoacetate-labelled α-globin) was prepared by the method of Butler *et al.* (1969). The reaction was carried out at room temperature in 0·2M borate, pH 9·0, and the pH was kept constant by the manual addition of 5M NaOH as required. The α-globin was not completely soluble at the start of the reaction but dissolved as maleylation proceeded.

Tryptic digestion was carried out as for unmodified globin, except that an enzyme substrate ratio of 5% (w/w) was used. Enzyme was added in three lots over a total digestion time of 24 hr.

(*e*) *Fractionation of Tryptic Peptides Derived from Maleylated Globin*

The freeze-dried digestion mixture was demaleylated by incubation in 0·2M acetic acid, pH 3·0, for 6 hr at 60°C (Butler *et al.* 1969). During this time the peptide mixture partly dissolved. The fraction insoluble at pH 3·0 was removed by centrifugation and washed with 0·2M acetic acid. In a small-scale experiment using [^{14}C]-*S*-carboxymethylated globin it was found that 85–90% of the radioactivity was in the material insoluble at pH 3·0 after demaleylation. This insoluble SCM-cysteine-containing fraction (from 130 mg globin) was digested with chymotrypsin and the peptides formed were isolated and examined by paper techniques as described above.

Another portion of the pH 3·0 insoluble demaleylated material (from 50 mg globin) was digested with trypsin at 37°C. After 16 hr fresh trypsin was added and the digestion continued for a further 16 hr. The lyophilized digest was fractionated by solubility at pH 6·4 in 10% pyridine–acetate buffer. The insoluble fraction was dissolved in 8M urea–50% formic acid and the centrifuged solution loaded on to a Sephadex G100 column and eluted with 10% formic acid. The main retarded fraction was subjected to amino acid and *N*-terminal sequence analysis. The tryptic peptides soluble at pH 6·4 were studied by the usual methods.

The tryptic peptide fraction soluble at pH 3·0 during demaleylation was loaded on to a column of Sephadex G50 in 10% formic acid and the fractions obtained examined for amino acid composition. The bulked fractions were dried, redigested with trypsin for 3 hr at 37°C, and the peptides obtained fractionated and characterized by paper techniques.

(*f*) *Isolation of Arginine Peptides from a Chymotryptic Digest of Kangaroo Globin*

The tryptic peptides from maleylated globin had *C*-terminal arginine residues and in order to obtain peptides linking these together unfractionated globin was digested with chymotrypsin and a selective procedure (Fruchter *et al.* 1970) used to concentrate those peptides containing arginine. The globin (400 mg) was digested in 20 ml 0·1M sodium phosphate (pH 8·7) containing 0·1% phenol with 8 mg chymotrypsin at 37°C for 4·5 hr. As there was a considerable precipitate the mixture was heated in a boiling water-bath for 15 min and after cooling further digested overnight at 37°C with 4 mg chymotrypsin. The digest was made 8M in urea, titrated to pH 9, and maleylated by the gradual addition over 30 min of maleic anhydride (400 mg) while maintaining the stirred solution at pH 9 with 6N KOH. After completion of the reaction the solution was adjusted to pH 3 and the precipitate removed by centrifugation. The supernatant was added to a column of sulphonated polystyrene (30 by 1·8 cm diam., Bio-Rad AG50W-X2, 200–400 mesh) which was in the acid form and equilibrated with 8M urea at pH 3. After washing with 8M urea solution at pH 3 till no further elution of maleylated peptides (absorbing at 254 nm) was evident the column was washed with water and the adsorbed maleylated arginine- and histidine-containing peptides (positively charged at pH 3) were eluted with 1N NH_3. The eluate was freeze-dried and the peptides demaleylated by heating at 60°C for 20 hr in 5 ml 0·2N acetic acid, pH 2·75. Insoluble material was removed by centrifugation and aliquots of all non-adsorbed and adsorbed fractions hydrolysed and analysed for their arginine content. The arginine was present only in the soluble portions of the demaleylated peptide fraction which had been bound on the column.

Samples of this fraction (50 µl) were fractionated by ionophoresis at pH 6·4 followed by chromatography in the second dimension using butanol–pyridine–acetic acid–water and the peptides containing arginine revealed by their fluorescence after treatment with phenanthraquinone (Yamada and Itano 1966). Preparative isolation of the peptides containing arginine was done as previously described.

III. Results

The nomenclature of Gerald and Ingram (1961) has been used to describe the α-chain tryptic peptides. The numbering of the peptides from the *N*-terminal end follows that of the human α-chain as far as possible. The symbols A and B have been added where an extra lysine has occurred in the kangaroo chain and when a lysine

has not occurred in the kangaroo chain the peptide has been given all the numbers of the corresponding human peptides (see Fig. 1).

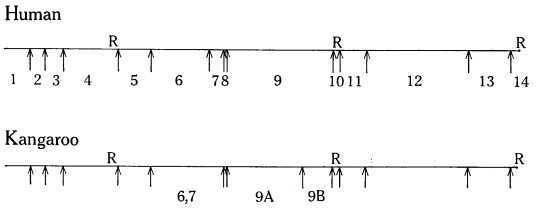

Fig. 1.—Nomenclature for α-chain tryptic peptides of kangaroo haemoglobin compared with α-chain tryptic peptides of human haemoglobin. The diagrams are scaled to show the relative number of the residues in each peptide. ↑ Lysine residue. R↑ Arginine residue.

(*a*) *Fractionation of Soluble Tryptic Peptides*

The peptide map of the whole digest obtained by paper ionophoresis–chromatography is shown in Figure 2. Separation of tryptic peptides by gel filtration

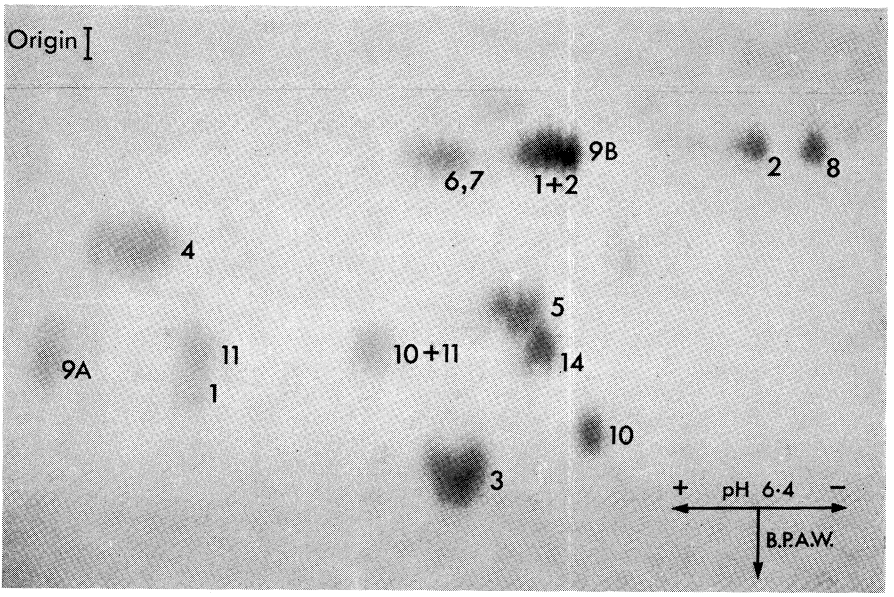

Fig. 2.—Peptide map of α-chain of kangaroo globin digested with TPCK-trypsin. Ionophoresis at pH 6·4 was followed by chromatography with butanol–pyridine–acetic acid–water (15 : 10 : 3 : 12 v/v) as indicated. The peptides are given the identification number corresponding to their position in the chain as shown in Figure 1. Unnumbered spots are peptides resulting from chymotryptic cleavages or tryptic peptides derived from β-chain contamination.

on Sephadex G25 is shown in Figure 3. The peptides present in each fraction are listed in the legend to the figure. The globin used for the characterization of the soluble tryptic peptides had been prepared from a β-heterozygote (Thompson, Hosken, and

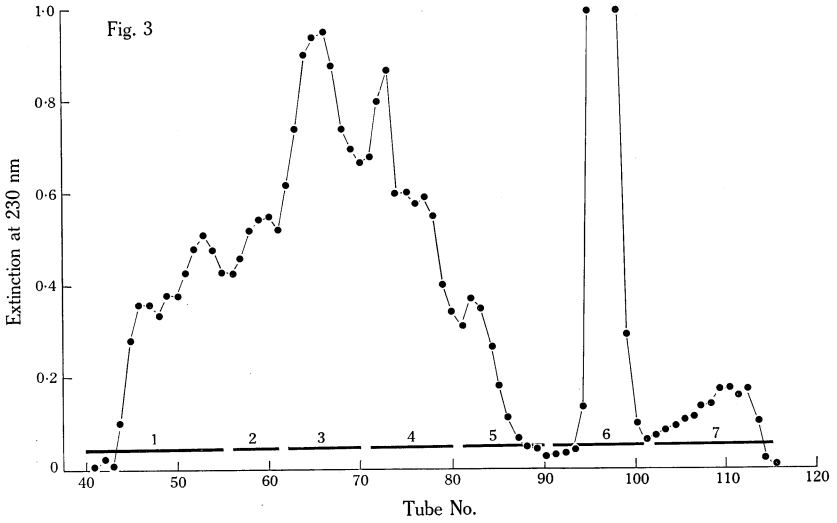

Fig. 3.—Gel filtration of peptides soluble at pH 6·4 after tryptic digestion of the α-chain of kangaroo globin (350 mg) on a column (120 by 2·3 cm) of Sephadex G25 in 0·01M NH_3. An aliquot (200 μl) of each fraction was diluted in 3 ml water to measure the extinction at 230 nm. Flow rate 48 ml/hr. Fraction size 4 ml. Tubes containing fractions combined for peptide recovery are shown as bars. Fraction 7 contained phenol and no peptides. The peptides present in the other fractions are as follows:

Fraction 1	Fraction 2	Fraction 3	Fraction 4	Fraction 5	Fraction 6
αTp9A	αTp1+2*	αTp1	αTp5	αTp2	αTp3
	αTp4	αTp1+2	αTp9B	αTp5*	αTp14
		αTp6,7	αTp11	αTp8	
		αTp10+11		αTp10	

* Indicates major proportion of peptide is in another fraction.

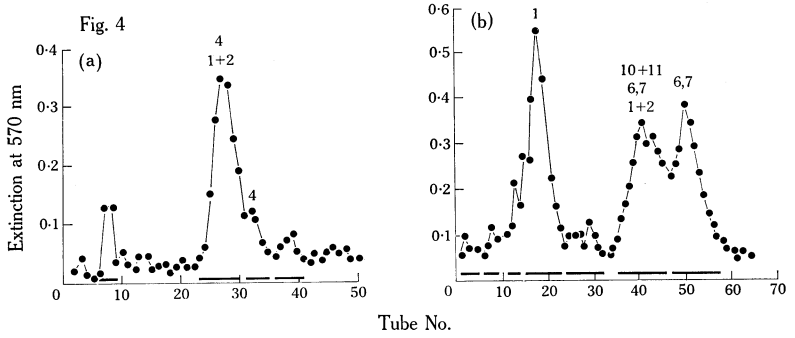

Fig. 4.—Separation of tryptic peptides on a column (15 by 0·9 cm diam.) of sulphonated polystyrene resin (Aminex MS, Q-15) after preliminary fractionation on Sephadex G25. (*a*) Fraction 2 from Figure 3. The column was pumped at 32 ml/hr. Fractions of 5·3 ml were collected and aliquots (100 μl) taken for alkaline hydrolysis and peptide determination with ninhydrin. The contents of the tubes indicated by bars were combined for further characterization. The numbers correspond to the α-chain tryptic peptides (Fig. 1) present in each peak. (*b*) Fraction 3 from Figure 3. Fractions of 4 ml were collected and assayed as above.

Air 1969) and the α-chain was contaminated by the adjacent β-II chain of the order of 10%. This results in the presence of some β-globin tryptic peptides.

Fractions 2 and 3 from Sephadex G25 were further fractionated on sulphonated polystyrene resin. The elution profiles are shown in Figure 4. Each peak obtained was further purified by paper ionophoresis at pH 6·4. Fraction 2 gave only one major peak containing αTp4 together with a small amount of αTp6,7. Fraction 3 gave three main peaks containing αTp1, αTp10+11, αTp1+2, and αTp6,7 readily purified by ionophoresis at pH 6·4.

Peptides in the remaining fractions obtained from Sephadex G25 were purified using paper techniques. The soluble peptides fractionated on Sephadex included all tryptic peptides from the globin except for αTp12 and αTp13 which make up residues 100–139 inclusive.

(*b*) *Isolation of Insoluble-core Peptide αTp12+13*

For characterization of this region a separate digest was made using the same conditions as before except that SCM-α-globin prepared from a β-I homozygote was used. In this way the α-chain could be obtained after cellulose chromatography which was almost completely free of β-contamination (Thompson, Hosken, and Air 1969). A small amount of [^{14}C]-SCM-α-chain was included in the digest in order to monitor the single cysteine residue which is located in αTp12+13. After digestion the major proportion of the counts was soluble at pH 8·7. The undigested globin was removed by centrifugation and the supernatant freeze-dried. The freeze-dried material was suspended in pH 6·4 buffer and the insoluble material purified by centrifugation and washing several times with buffer and water.

The insoluble material analysed as αTp12+13 and gave a single sequence using the dansyl–Edman technique beginning *Leu-Leu-Ser-*, indicating the presence of a single peptide and not a mixture of αTp12 and αTp13. Guidotti, Hill, and Konigsberg (1962) reported that the Lys-Leu bond between αTp12 and αTp13 in human α-chain was resistant to tryptic digestion, probably due to the aspartyl residue preceding the lysine and our results support this finding. Some radioactivity, however, was also present in the pH 6·4 soluble fraction which could have been due to partial solubility of αTp12+13 or due to some αTp12 in solution. This was not investigated further.

(*c*) *Amino Acid Composition of Tryptic Peptides*

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

The total composition shows some differences from that given by Thompson, Hosken, and Air (1969). For eight amino acids one residue more or less than the analytical value was present. As mentioned earlier the α-chain isolated from the α-β-I+α-β-II heterozygote animal was contaminated with some β-II globin and this is reflected in the analytical values for those residues where the α- and β-chains differ significantly. Better analytical values were obtained for seven analyses of α-chain from the α-β-I homozygote. The values are not given in detail but they were in agreement with the composition derived from the sequence data except for aspartic acid, glutamic acid, histidine, and serine, the values for which may still reflect slight β-globin contamination.

TABLE 1

AMINO ACID COMPOSITION OF KANGAROO α-GLOBIN TRYPTIC PEPTIDES

Soluble peptides were purified by gel filtration, ion-exchange chromatography, and paper ionophoresis at pH 6·4. Hydrolysates were made at 115–120°C for 24 hr. 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

Amino Acid	αTp1	αTp2	αTp3	αTp4	αTp5	αTp6,7	αTp8	αTp9A	αTp9B	αTp10	αTp11	αTp12 +13*	αTp14	Total in Protein
Lysine	1·0(1)	1·0(1)	1·0(1)		1·1(1)	1·2(1)	1·0(1)	0·9(1)	1·0(1)		0·7(1)	2·4(2)		11
Histidine		1·0(1)		0·9(1)	1·0(1)	2·7(3)		0·9(1)	1·8(2)			2·9(3)		12
Arginine				1·0(1)						1·1(1)			1·0(1)	3
SCM-cysteine												0·9(1)		1
Aspartic acid	1·4(1)					1·1(1)		3·0(3)	1·1(1)		2·0(2)	2·4(2)		10
Threonine					2·3(3)	0·8(1)		0·9(1)				2·7(4)		9
Serine	0·6(1)				1·2(1)	1·5(2)		0·7(1)	0·6(1)			2·4(4)		10
Glutamic acid				3·1(3)		2·0(2)		2·1(2)				1·7(1)		8
Proline					1·1(1)	0·9(1)		1·0(1)			1·0(1)	1·0(1)		5
Glycine		1·0(1)	1·1(1)	4·0(4)		2·2(2)		2·0(2)				1·9(1)		11
Alanine	1·7(2)		1·0(1)	2·9(3)		2·2(2)		3·0(3)	1·1(1)			5·8(6)		18
Valine	1·1(1)	0·8(1)		0·9(1)		0·2		1·0(1)			1·9(2)	4·2(4)		10
Isoleucine			1·0(1)			0·9(1)		1·9(2)				0·4		4
Leucine	1·0(1)			1·0(1)		1·0(1)		3·0(3)	2·0(2)	1·0(1)		7·9(8)		17
Tyrosine				0·8(1)		0·7(1)							0·7(1)	3
Phenylalanine					2·0(2)	1·8(2)					1·0(1)	3·0(3)		8
Tryptophan†			+(1)											1
Total in peptide	7	4	5	15	9	20	1	21	8	2	7	40	2	141

* This was the peptide material insoluble at pH 6·4 after tryptic digestion. Shows some contamination with β-globin.

† Detected by Ehrlich reagent.

(d) *Amino Acid Sequences of Tryptic Peptides*

Quantitative data is not obtained with the dansyl–Edman procedure but the results were sufficiently clear-cut to leave no doubt as to the correct sequence. Most peptides or fragments derived from them by further proteolytic digestion were degraded completely. As previously (Air and Thompson 1971), residues identified by the dansyl–Edman procedure have been printed in *italic* fount. Each peptide is discussed in turn starting from the *N*-terminal end. The evidence for alignment of the peptides is given later.

αTp1, αTp2, and αTp3

Evidence for the sequence of the first 16 residues has been given in an earlier paper (Beard and Thompson 1970). The sequence is shown below:

Val-Leu-Ser-Ala-Ala-Asp-Lys-Gly-His-Val-Lys-Ala-Ile-Trp-Gly-Lys

αTp4

The sequence of this peptide is

Val-Gly-Gly-His-*Ala-Gly-Glu-Tyr-Ala-Ala-Glu-Gly-Leu-Glu-Arg*

The sequence of the first 13 residues was obtained from the degradation of the whole peptide although no clear result was obtained for residue 4. Chymotryptic digestion gave two peptides resulting from cleavage after the tyrosyl residue, each of which was degraded completely. The whole peptide was acidic at pH 6·4 and the mobilities of the chymotryptic peptides confirmed the presence of three glutamic acid residues in αTp4.

αTp5

This peptide was degraded completely. The histidyl residue can be assigned to position 3 as all other residues were definitely identified. The sequence is

Thr-Phe-His-*Ser-Phe-Pro-Thr-Thr-Lys*

αTp6,7

The sequence of this peptide is

Thr-Tyr-Phe-Pro-His-*Phe-Asp-Leu-Ser*-His-*Gly-Ser-Ala-Gln-Ile-Gln-Ala*-His-*Gly-Lys*

Definite results were obtained for the first nine residues during degradation of the whole peptide although the histidyl residue gave less clear-cut identification. Thermolysin digestion gave four major fragments. These were sequenced completely and could be ordered from the *C*-terminal lysine residue and the known *N*-terminal sequence. Partial splitting of bonds involving the amino groups of the alanyl residues gave the neutral peptides *Ala-Gln* and *Ile-Gln*. Amide residues were assigned using ionophoretic mobilities at pH 6·4 of each of the thermolysin peptides.

The points of cleavage by thermolysin are shown below. Dotted arrows indicate minor points of attack:

Thr-Tyr-Phe-Pro-His-Phe-Asp-Leu-Ser-His-Gly-Ser-Ala-Gln-Ile-Gln-Ala-His-Gly-Lys

αTp8

Lys. This spot had the same mobility at pH 6·4 both before and after hydrolysis and was therefore free lysine.

αTp9A

Ile-Ala-Asp-Ala-Leu-Gly-Gln-Ala-Val-Glu-His-*Ile-Asp-Asp-Leu-Pro-Gly-Thr-Leu*-Ser-*Lys*

Direct degradation gave

Ile-Ala-Asx-Ala-Leu-Gly-Glx-Ala-Val-Glx-

The peptide was digested with thermolysin to give five main products. All amides could be assigned from the mobilities of these at pH 6·4. Since the lysine residue had to be *C*-terminal in this tryptic peptide the known *N*-terminal sequence allowed the complete ordering of the thermolysin peptides. Each thermolysin peptide was degraded by the dansyl–Edman method to complete the sequence of αTp9A. The sequence following shows the action of thermolysin on this peptide. There was some hydrolysis, approximately 20%, of the glutaminylalanyl bond:

Ile-Ala-Asp-Ala-Leu-Gly-Gln-Ala-Val-Glu-His-Ile-Asp-Asp-Leu-Pro-Gly-Thr-Leu-Ser-Lys

αTp9B

The sequence is

Leu-Ser-Asp-Leu-His-Ala-His-Lys

The mobility of the whole peptide at pH 6·4 indicated residue 3 was aspartic acid and not asparagine. This was confirmed by isolation of an acidic peptide, *Leu-Ser-Asp*, after digestion with thermolysin.

αTp10

The sequence of this dipeptide is *Leu-Arg*. Tryptic digestion after the arginyl residue was incomplete and a basic peptide, αTp10+11, was also recovered from tryptic digests.

αTp11

The sequence is

Val-Asp-Pro-Val-Asn-Phe-Lys

Sequence degradation of the whole peptide gave

Val-Asx-Pro-Val-Asx-Phe-Lys

This peptide was neutral at pH 6·4 indicating the presence of one asparaginyl residue and one aspartyl residue. These were assigned to the positions given after digestion of αTp11 with thermolysin. An acidic peptide *Val-Asp-Pro* and a neutral peptide *Val-Asn* were present in the digest.

αTp12 and αTp13

The lysyl-phenylalanyl bond linking these two peptides was incompletely digested with trypsin. The sequence of this region of the α-chain was determined

using the αTp12+13 tryptic peptide which was insoluble at pH 6·4 after freeze-drying. It was also obtained from a tryptic digest of the demaleylated fragment αM3. The complete sequence is:

Leu-Leu-Ser-His-Cys-Leu-Leu-Val-Thr-Phe-Ala-Ala-His-Leu-Gly-Asp-Ala-Phe-Thr-Pro-Glu-Val-His-Ala-Ser-Leu-Asp-Lys-Phe-Leu-Ala-Ala-Val-Ser-Thr-Val-Leu-Thr-Ser-Lys

The *N*-terminal sequence found by direct degradation was

Leu-Leu-Ser-His-CMC-Leu-Leu-Val-Thr-Phe-

Results became less certain on further degradation. The rest of the sequence was determined using the fragments obtained after digestion with thermolysin or papain.

After fractionation of the thermolysin digest the chromatograms were stained with ninhydrin to give the patterns shown in Figure 5. The spots from one paper

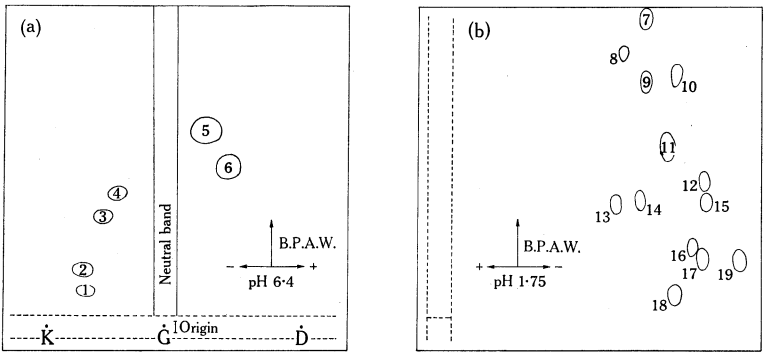

Fig. 5.—Fractionation of peptides formed after digestion of αTp12+13 (2 mg) with thermolysin. (*a*) Ionophoresis at pH 6·4 was followed by chromatography with butanol–pyridine–acetic acid–water (15 : 10 : 3 : 12 v/v) as indicated. The band containing the neutral peptides was cut out before staining. The ionophoretic mobilities of marker amino acids lysine (*K*), aspartic acid (*D*), and glycine (*G*) are shown. (*b*) Fractionation of the neutral band from Figure 5(*a*). It was sewn on to another sheet of paper for ionophoresis at pH 1·75 as indicated.

(2 mg fractionated) were hydrolysed with 6N HCl and the analyses are shown in Table 2. In some cases the amino terminal residues give lower values due to destruction by the ninhydrin used to locate the peptides. The peptides isolated from the enzyme digests were sequenced by the techniques already described. The results are summarized in Table 3. The amide groups were assigned from the mobilities at pH 6·4.

The data in Table 3 show two lysyl sequences, one of which is obviously *C*-terminal, and peptides derived from the known *N*-terminal sequence given above. The peptide sequences are given below:

	Leu-Leu-Ser-His-CMC-Leu-Leu-Val-Thr-Phe	(*N*-terminal)
Th12	Leu-Asp-Lys-Phe	
Th 4	Val-Leu-Thr-Ser-Lys	(*C*-terminal)
Th 6	Leu-Gly-Asp-Ala	
Th 5	Phe-Thr-Pro-Glu	
Th13	Ala-Val-Ser-Thr	
Th 9	Leu-Ala-Ala	
Th 2	Ala-His	
Th 3A	Val-His	
Th16	Ala-Ser	

TABLE 2

AMINO ACID COMPOSITIONS OF THERMOLYSIN PEPTIDES FROM αTp12+13

The peptides were fractionated by paper ionophoresis and chromatography (Fig. 5). Hydrolyses were with 6N HCl for 24 hr at 115–120°C. Values are not corrected for losses during hydrolysis or due to reaction of *N*-terminal amino acid with ninhydrin and are given as moles per mole peptide

Amino Acid	Th1	Th2	Th3A*	Th4	Th5	Th6	Th7	Th8	Th9	Th10†	Th11	Th12	Th13	Th14	Th15	Th16	Th17	Th18	Th19
Lysine	1·0			1·0								1·0							1·0
Histidine		0·8	0·9														1·0	1·0	
Cysteic acid‡																	0·6	0·6	
Aspartic acid						1·0						1·0							1·0
Threonine				0·8	1·0			0·8			0·6		0·7	0·8					
Serine	0·6			0·7									0·8	0·8		0·6	0·8	1·0	
Glutamic acid					1·2														
Proline					0·9														
Glycine						1·0													
Alanine		1·0				1·0			2·0	1·3			0·9		1·0	1·0			
Valine			1·0	1·0				1·0			1·0		1·0	1·0					
Leucine				1·0		0·8	1·0	1·0	1·0	1·0		1·0					1·0	1·1	0·9
Phenylalanine					0·9					0·4		0·9							

* Peptide Th3A was further purified by paper ionophoresis at pH 1·75.

† Peptide Th10 was a mixture of at least two peptides.

‡ Cysteic acid was from oxidation of *S*-carboxymethylcysteine during hydrolysis.

There is a limited number of possible overlaps in these sequences. Digestion with papain, which shows a different specificity (cf. Thompson, Sleigh, and Smith 1971),

TABLE 3

SEQUENCES OF THERMOLYSIN PEPTIDES FROM αTp12+13

Peptide	Sequence*	Peptide	Sequence*
Th1	*Ser*-Lys	Th9	*Leu*-Ala-Ala
Th2	*Ala-His*	Th10	Mixture—no sequence deduced
Th3A	*Val-His*	Th11	*Val*-Thr
Th4	*Val-Leu-Thr-Ser*-Lys	Th12	*Leu-Asp-Lys-Phe*
Th5	*Phe-Thr-Pro-Glu* (peptide acidic at pH 6·4, therefore glutamic acid)	Th13	*Ala-Val-Ser-Thr*
		Th14	*Val-Ser-Thr*
Th6	*Leu-Gly-Asp-Ala* peptide acidic at pH 6·4, therefore aspartic acid)	Th15‡	Ala-Ala
		Th16	*Ala*-Ser
Th7	*Leu*	Th17§	*Leu-Ser-His*-CMC
Th8†	*Val-Leu-Thr* and *Leu-Val-Thr*	Th18§	*Leu-Ser-His-CMC*
		Th19	*Leu-Asp-Lys*

* Residues identified by the dansyl–Edman procedure are shown in italics.

† Peptide Th8 contained both sequences shown. The first two steps of the dansyl–Edman procedure gave two dansyl-amino acids but only DNS-Thr in step 3. The ratio of intensities suggested two parts *Val-Leu-Thr* to one part *Leu-Val-Thr*.

‡ Mobility at pH 1·75 was the same as that of an authentic sample of Ala-Ala, and less than that of amino acid.

§ Peptides Th17 and Th18 gave the same sequence and one of them is due to partial oxidation of the *S*-carboxymethylcysteine side chain during the preparation.

TABLE 4

SEQUENCES OF PAPAIN PEPTIDES FROM αTp12+13*

Peptide	Sequence†
Pa1	*Ser-Lys*
Pa2	*Asp-Ala-Phe-Thr-Pro-Glu-Val*-His
Pa3	*Leu-Leu*-Ser
Pa4	*Thr-Val-Leu-Thr*
Pa5	*Phe-Ala*
Pa6	*Ala-Ala-Val-Ser*
Pa7	*Ala-His-Leu-Gly-Asp-Ala-Phe*-Thr-Pro-Glu-Val-His
Pa8	*Ala*-Ser-*Leu*-Asp

* The peptides were fractionated by ionophoresis at pH 6·4 and chromatography. Peptides neutral at pH 6·4 were further purified by paper ionophoresis at pH 1·75.

† Residues identified by the dansyl–Edman procedure are shown in italics. Amides were assigned from the mobilities of peptides at pH 6·4.

gave peptides which enabled most of the sequence to be deduced. The papain peptides identified are shown in Table 4. Not all peptides produced were isolated in sufficient yield for complete characterization.

From the sequence of Pa7, which incorporates Pa2, linking of Th2, Th6, Th5, and Th3A in that order is revealed. Peptide Pa4 overlaps Th13 and Th4 with supporting evidence from Pa6 which includes the remainder of Th13. Peptide Pa6 is consistent with an overlap of Th9 and Th13. Peptide Pa8 establishes a linkage between peptides Th16 and Th12. The major sequences in αTp12+13 now reduce to

1. Leu-Leu-Ser-His-CMC-Leu-Leu-Val-Thr-Phe (*N*-terminal)
2. Leu-Ala-Ala-Val-Ser-Thr-Val-Leu-Thr-Ser-Lys (*C*-terminal)
3. Ala-Ser-Leu-Asp-Lys-Phe
4. Ala-His-Leu-Gly-Asp-Ala-Phe-Thr-Pro-Glu-Val-His
5. Phe-Ala

These sequences account for all the residues in αTp12+13 as determined by amino acid analysis (Table 1) provided that the *N*-terminal phenylalanyl residue in sequence 5 is the same as that present in one of the two peptide sequences (1 and 3) with *C*-terminal phenylalanine residues. There is a total of three phenylalanine residues present in αTp12+13.

Tryptic digestion of maleylated α-globin gave three fragments, one of which (αM3) included αTp12+13. The amino acid compositions and sequences of some chymotryptic peptides derived from αM3 gave the information necessary to complete the sequence of αTp12+13. Two neutral peptides

Ala-Ala-His-Leu-Asp-Ala-Phe-Thr-Pro-Glu-Val-His

and

Ala-Ala-His-Leu-Gly-Asp-Ala-Phe-Thr-Pro-Glu-Val-His-Ala-Ser-Leu-Asp-Lys-Phe

showed that sequence 3 follows sequence 4.

The Phe-Ala sequence 5 must overlap sequence 1 or sequence 3 so that the phenylalanyl residue is common. Since an Ala-Ala-His sequence must be present, sequence 5 must precede sequence 4+3 thus having the phenylalanyl residue in common with sequence 1.

The sequence of αTp12+13, including the peptides formed during digestion with thermolysin and papain, is shown in Figure 6.

It will be noted that in the analysis of αTp12+13 (Table 1) the values for glutamic acid, glycine, and isoleucine were higher than in the sequence finally deduced. This was due in part to some contamination with insoluble-core material from a small amount of β-chain. Some β-chain thermolysin peptides were detected (e.g. Leu-Gly-Asn, Leu-Ala-Glu-His, Phe-Gly-Lys) during fractionation of digests.

αTp14

This peptide contained only tyrosine and arginine and must therefore be Tyr-Arg. Arginine was known to be *C*-terminal in the α-globin following its release by carboxypeptidase B (Moon and Thompson, unpublished data).

(*e*) *Ordering of Tryptic Peptides*

There were three fragments expected from a tryptic digest of maleylated *S*-carboxymethylated α-chain since there are three arginyl residues, one of which was known to be *C*-terminal. After demaleylation of the tryptic digest there was an

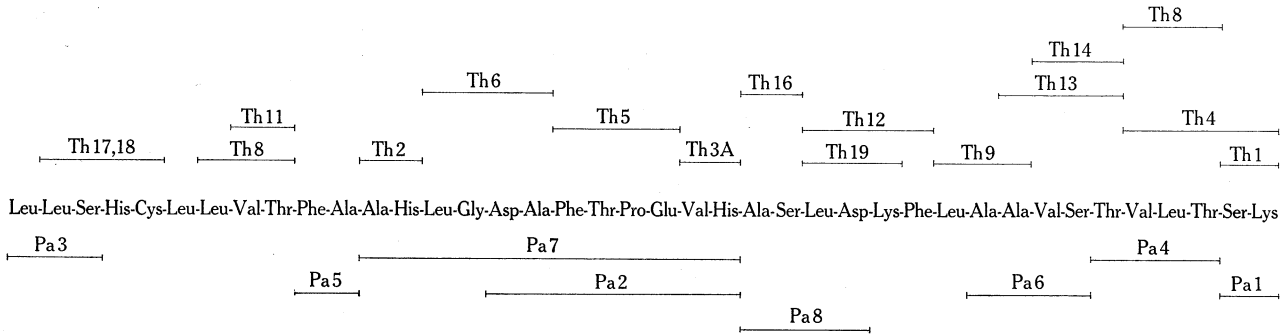

Fig. 6.—The amino acid sequence of the core tryptic peptide comprising residues 100–139 of the α-chain of kangaroo globin. The peptides obtained by cleavage with the enzymes thermolysin (Th) and papain (Pa) are indicated together with the number assigned to each peptide.

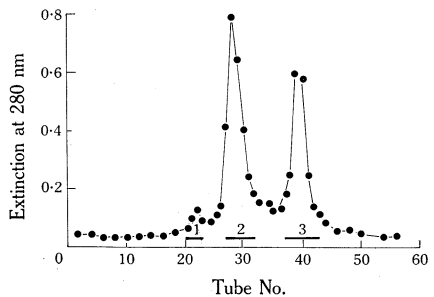

Fig. 7.—Fractionation of soluble tryptic peptides derived from maleylated α-globin (130 mg). The peptides soluble at pH 3·0 after demaleylation were loaded on to a column (1 by 130 cm) of Sephadex G50 in 10% formic acid. Flow rate 10 ml/hr. Fraction size 2·5 ml. The tubes indicated by bars were combined for further characterization. Fraction 3 was mainly phenol. Fraction 1 contained αM2 representing residues 32–92; fraction 2 contained αM1 representing residues 1–31 of the α-chain.

insoluble peptide, αM3, which from an experiment with [^{14}C]-*S*-carboxymethylated α-chain was known to contain about 90% of the radioactive counts. The two soluble fragments (αM1 and αM2) were partially separated by Sephadex G50 gel filtration (Fig. 7). The fragments are numbered according to their relative position in the chain.

Digestion of the αM2 fragment with trypsin followed by two-dimensional fractionation of the tryptic peptides as used for the α-globin (Fig. 2) showed the presence of peptides αTp5; αTp6,7; αTp8; αTp9A; αTp9B; αTp10; and a larger peptide comprising αTp8+9A due to incomplete hydrolysis of a Lys-Lys bond. The amino acid composition of each of these peptides was identical with that already known and the sum of the amino acid residues in these single peptides was in good agreement with the amino acid composition of αM2 (Table 5).

TABLE 5

AMINO ACID COMPOSITION OF TRYPTIC PEPTIDES FROM MALEYLATED α-GLOBIN

The peptides were purified as described in the text, then hydrolysed under vacuum with 6N HCl at 115–120°C for 24 hr. Values are given as moles per mole of peptide. The compositions determined from the sum of the compositions of the constituent tryptic peptides are also shown

Amino Acid	αM1		αM2		αM3‡	
Lysine	2·9	3	5·0	5	3·4	3
Histidine	1·8	2	6·6	7	2·6	3
Arginine	0·7	1	0·6	1	1·0	1
SCM-cysteine	0	0	0	0	0·8	1
Aspartic acid	1·1	1	5·7	5	4·1	4
Threonine*	0	0	3·8	5	3·2	4
Serine*	1·4	1	3·0	5	2·2	4
Glutamic acid	2·3	3	4·8	4	2·6	1
Proline	0	0	3·2	3	2·0	2
Glycine	5·3	6	4·8	4	2·6	1
Alanine	5·8	6	7·1	6	7·0	6
Valine	3·0	3	1·8	1	5·8	6
Isoleucine	0·9	1	3·1	3	0·9	0
Leucine	2·0	2	6·9	7	8·0	8
Tyrosine	0·7	1	1·1	1	1·0	1
Phenylalanine	0	0	4·0	4	3·9	4
Tryptophan†	+	1	—	0	—	0
Total		31		61		49

* Uncorrected for decomposition.

† Detected by Ehrlich's reagent.

‡ The sample analysed was further purified by gel filtration on Sephadex G50 in 50% formic acid.

The amino acid composition of the other fraction from the gel filtration was quite different from that of αM2. Some contamination with αM2 was apparent from the low non-stoicheiometric content of phenylalanine and threonine, which are present in αM2 but absent from αM1. αM1 was the major zone obtained from this fraction by ionophoresis at pH 6·4 followed by chromatography in butanol–pyridine–acetic acid–water. The composition of the purified peptide is shown in Table 5. The *N*-terminal sequence was *Val-Leu-Ser-Ala-Ala-*. Tryptic digestion of the

Sephadex fraction followed by peptide mapping as described gave the peptides αTp1, αTp2, αTp3, and αTp4 as major products together with some αTp1+2 which results from incomplete hydrolysis by trypsin of an Asp-Lys-Gly bond (Beard and Thompson 1970). ↑

The insoluble peptide (αM3) was also digested with trypsin and the pH 6·4 insoluble material loaded on to Sephadex G100. The main retarded peak analysed as αTp12+13 and had the *N*-terminal sequence

Leu-Leu-Ser-His-*CMC-Leu-Leu-Val-*

αTp14 was recovered from the fraction soluble at pH 6·4 by paper chromatography. The expected αTp11 which is neutral was not readily purified but evidence for its presence in αM3 was obtained by isolation of the peptide *Val-Asp-Pro-Val-Asn-Phe* from a chymotryptic digest of αM3. Other peptides from αTp12, 13, and 14 were also isolated from the chymotryptic digest.

The sum of the compositions of αTp11, αTp12, αTp13, and αTp14 is

Lys_3 His_3 Arg_1 Asp_4 Thr_4 Ser_4 Glu_1 Pro_2 Gly_1 Ala_6 CMC_1 Val_6 Leu_8 Tyr_1 Phe_4

Amino acid analysis of the αM3 fragment after gel filtration on Sephadex G50 in 50% formic acid gave a similar composition although the values of some residues, e.g. glutamic acid and isoleucine, were higher than expected. This is due to the small amount of β-chain in the α-chain used for maleylation. The insoluble fraction would contain maleylated β-chain tryptic peptides involving residues 30–146.

The fragment αM1 is clearly the *N*-terminal fragment of the α-chain since it gives tryptic peptides that were previously obtained from the *N*-terminal portion of the α-chain, which had been released by selective cleavage at the single tryptophan residue (Beard and Thompson 1970). Both αM2 and αM3 have *C*-terminal arginine residues due to arginine being the *C*-terminal residue in the α-chain. A knowledge of the sequences around the arginine residues of the α-chain is necessary for the correct order to be deduced.

After chymotryptic digestion of kangaroo globin the peptides containing arginine were concentrated. The α-chain peptides, in addition to free arginine derived from splitting the *C*-terminal Tyr-Arg sequence, included an acidic peptide *Ala-Ala-Glu-Gly-Leu-Glu-Arg-Thr-Phe* (as well as a fragment of this peptide *Glu-Arg-Thr-Phe*). This sequence links αTp4 to αTp5; hence αM1 to αM2. The other α-chain arginine peptides were *Arg-Val-Asp* and Arg-Val-Asp-Pro-Val-Asn-Phe. The longer peptide was obtained in a much smaller yield due to a secondary cleavage during demaleylation. Both were neutral peptides indicating an aspartyl residue. Only the latter peptide was obtained when α-globin was digested with chymotrypsin and the non-maleylated fragments separated by paper techniques. This sequence is consistent with αM3 following αM2, the peptide αTp11 representing the *N*-terminal portion of αM3. As mentioned earlier a peptide αTp10+11 had been isolated from a tryptic digest of α-globin.

In addition to these sequences derived from the α-chain the chymotryptic digest of the whole globin gave the peptides *Thr-Ser-Arg-Phe*, *Ser-Arg-Phe*, and *Gly-Arg-Leu-Leu* derived from the β-chain, the sequence of which was already known.

The ordering of the tryptic peptides within the αM1, αM2, and αM3 fragments was completed from the composition and partial sequence of chymotryptic peptides

TABLE 6

COMPLETE AMINO ACID SEQUENCE OF THE α-CHAIN OF KANGAROO HAEMOGLOBIN

Residues are numbered from the *N*-terminal residue and the positions of cleavage by trypsin are indicated by arrows. A dotted arrow indicates limited cleavage. The 27 residues which differ from those present in human α-chain are shown in capital letters

```
                5                  10                  15                  20                  25
Val-Leu-Ser-ALA-Ala-Asp-Lys-GLY-HIS-Val-Lys-Ala-ILE-Trp-Gly-Lys-Val-Gly-GLY-His-Ala-Gly-Glu-Tyr-ALA-Ala-Glu-
          αTp1          ↑     αTp2     ↑     αTp3     ↑           αTp4

       30                  35                  40                  45                  50                  55
GLY-Leu-Glu-Arg-THR-Phe-HIS-Ser-Phe-Pro-Thr-Thr-Lys-Thr-Tyr-Phe-Pro-His-Phe-Asp-Leu-Ser-His-Gly-Ser-Ala-Gln-ILE-GLN-
             ↑           αTp5           ↑                 αTp6,7

          60                  65                  70                  75                  80
ALA-His-Gly-Lys- Lys -ILE-Ala-Asp-Ala-Leu-GLY-GLN-Ala-Val-GLU-His-ILE-Asp-Asp-LEU-Pro-GLY-THR-Leu-Ser-LYS-
                ↑ αTp8 ↑                                  αTp9A                                          ↑

       85                  90                  95                  100                 105                 110
Leu-Ser-Asp-Leu-His-Ala-His-Lys-Leu-Arg-Val-Asp-Pro-Val-Asn-Phe-Lys-Leu-Leu-Ser-His-Cys-Leu-Leu-Val-Thr-PHE-Ala-Ala-His-
          αTp9B            ↑ αTp10 ↑          Tp11          ↑

      115                 120                 125                 130                 135                 140
Leu-GLY-ASP-ALA-Phe-Thr-Pro-GLU-Val-His-Ala-Ser-Leu-Asp-Lys-Phe-Leu-Ala-ALA-Val-Ser-Thr-Val-Leu-Thr-Ser-Lys-Tyr-Arg
          αTp12                                          ↑           αTp13                          ↑ αTp14
```

derived from α-globin. For αM1 the order of the peptides was already known (Beard and Thompson 1970) except for αTp4 which contains arginine and must be *C*-terminal. A chymotryptic peptide

Gly-Lys-Val-Gly-Gly-His-Ala-Gly-Glu-Tyr

confirmed this.

The *N*-terminal peptide of αM2 is αTp5. Two chymotryptic peptides, one

His-Ser-Phe-Pro-Thr-Thr-Lys-Thr-Tyr

and the other lacking the histidine residue, linked αTp5 to αTp6,7.

A large peptide with the composition

Lys_2 His_2 Asp_3 Thr_1 Glu_2 Pro_1 Gly_3 Ala_4 Val_1 Ile_2 Leu_3

and *N*-terminal sequence *Ala-His-Gly-Lys-* linked αTp6,7; αTp8; and αTp9A in that order. As αTp10 contains arginine it must be *C*-terminal so that αTp9B follows αTp9A. This was confirmed by the isolation of the chymotryptic peptides *Ser-Lys-Leu-Ser-*(Asp,Leu,His) and *Ala-His-Lys-Leu*.

αM3 gives three major products on tryptic digestion and the order of these is easily deduced. The *N*-terminal peptide is αTp11, previously mentioned as being linked to αTp10 from αM2. The long peptide αTp12+13 is linked to αTp14, which contains the arginine, by a chymotryptic peptide *Thr-Ser-Lys-Tyr*.

The complete amino acid sequence of the α-chain is shown in Table 6.

IV. Discussion

A major problem in the determination of the amino acid sequence of α-globin chains has been the insoluble core present in tryptic digests. In the case of kangaroo α-globin, however, the insoluble core, representing αTp12+13, does not precipitate during digestion at pH 8·7 but becomes insoluble on adjusting the pH of the solution to 6·4. It can then be recovered in good yield and adequate purity by washing with pH 6·4 buffer.

The sequence of αTp12+13 was determined from the structure of peptides obtained by further digestion with the enzymes thermolysin and papain, which have different specificities. Peptides formed by papain, with a specificity for bonds of the type

Phe-X-Y
↑

or where valine or leucine replaces the phenylalanyl residue (Berger and Schechter 1970), help to order the thermolysin peptides, where the bonds preceding residues with aromatic or hydrophobic side-chains have been hydrolysed (Ambler and Meadway 1968). Thermolysin has also been found useful for further digestion of soluble tryptic peptides from the α-chain. In particular the sequence of αTp9A was comparatively easy to determine using this enzyme which gave fewer peptides than papain and pepsin gave in work on human (Konigsberg and Hill 1962) and monkey haemoglobins (Matsuda *et al.* 1970*a*).

The technique of maleylation (Butler *et al.* 1969) proved extremely helpful in allocating particular tryptic peptides to specific areas of the globin chain. The homology between kangaroo α-globin and the α-globins of other species whose amino

acid sequence is known (Dayhoff 1969) is sufficiently marked that the ordering of tryptic peptides is not a difficult task; nevertheless, it was considered worthwhile to gather evidence to confirm the order. During demaleylation of a peptide mixture at pH 2·75 an aspartyl-prolyl bond was hydrolysed. Piszkiewicz, Landon, and Smith (1970) have recently drawn attention to the lability of this sequence.

The length of the peptide chain and the arrangement of the amino acids in the α-chain of kangaroo globin follows the pattern evident in haemoglobins of other mammalian species. Comparison of this amino acid sequence with that of horse α-globin, which has a known three-dimensional structure, shows few substitutions that may have an effect on the structure.

The glycine residue at position 25 which has been one of the seven residues constant in all known globin chains (Dayhoff 1969; Perutz 1969) is alanine in kangaroo. However, Braunitzer (1970) has stated in a preliminary communication on an insect haemoglobin that there are now only three invariant positions in globins. The isoleucine residue at position 13 in kangaroo is considerably bulkier than the common alanine residue in this position. It might be expected that this change would be compensated for by a change in other contact areas of the molecule such as at positions 121 and 125 as seen in the model of horse α-globin (Perutz, personal communication). There is no change, however, at these positions in kangaroo, although a change from leucine to phenylalanine at residue 109 may be significant. Isoleucine also occurs in chicken α-globin at residue 13 (Matsuda *et al.* 1970*b*) but here residues 109, 121, and 125 are the same as in horse α-globin.

From the model of horse oxyhaemoglobin Perutz *et al.* (1968) have listed many "contact sites". These are defined as distances between residues of 4Å or less, short enough for interactions to occur. Of the 19 contacts between α-chain residues and the haem group in horse oxyhaemoglobin 17 residues are identical in kangaroo α-globin. The changes are threonine for methionine at residue 32 and isoleucine for valine at residue 62, both of which are conservative.

For the contact sites between subunits the vital $\alpha_1\beta_2$ contact sites utilize α-residues that are identical to those in horse oxyhaemoglobin whereas there was one change in the β-chain. For the $\alpha_1\beta_1$ contact sites there are five changes in the 16 residues of α-chain involved compared with 7 out of 18 residues in the β-chain. These would not be expected to interfere significantly with the structure. For the α-chain the substitution of hisitidine for leucine at residue 34 is the most significant change; the other changes at residues 35, 107, and 111 being the same as in the α-chain of humans and other species. The presence of glycine rather than proline at residue 114 is in line with the glycyl residue found in this position in all β-globins so far sequenced.

The total number of sequence differences between the α-globins of kangaroo and man is 27 in agreement with previous evidence that the α-chain has evolved more slowly than the β-chain (Zuckerkandl and Pauling 1965) which had 38 changes (Air and Thompson 1970). The total number of differences between all α-globins of known amino acid sequence are shown in Table 7. Only α-globins where the sequence has been determined directly or are listed by Dayhoff (1969) as 90% certain of being correct are included. This table is similar to one given by Air *et al.* (1971) but revised to include the complete 141 residue sequence of the α-chain of kangaroo. The additional data have no significant effect on the calculations of the time in millions of years for a change in one amino acid residue per 100 amino acid residues. This infor-

mation and that gained from the sequences of the β-chain and myoglobin from kangaroo have been used to calculate 129 million years as the approximate time since divergence of marsupials from eutherian mammals.

TABLE 7

NUMBER OF AMINO ACID SEQUENCE DIFFERENCES BETWEEN α-GLOBINS

	Human	Monkey*	Horse	Bovine	Rabbit	Kangaroo	Chicken	Carp†
Human	0	4	18	17	25	27	35	71
Monkey*	4	0	16	16	25	26	35	71
Horse	18	16	0	18	25	29	40	70
Bovine	17	16	18	0	25	26	38	68
Rabbit	25	25	25	25	0	37	44	74
Kangaroo	27	26	29	26	37	0	41	74
Chicken	35	35	40	38	44	41	0	75
Carp†	71	71	70	68	74	74	75	0

* Rhesus monkey.

† Sequences are of 141 residues except for carp (143 residues).

The sequences of three globin chains from the kangaroos have now been reported. The α- and β-chains are from the grey kangaroo *Macropus giganteus*, while the myoglobin is from the red kangaroo *Megaleia rufa*. The sequence of the

TABLE 8

NUMBER OF DIFFERENCES IN AMINO ACID SEQUENCE BETWEEN GLOBIN CHAINS OF SEVERAL SPECIES

Globins Compared*	Kangaroo	Horse	Bovine
α-Globin–β-globin	92	84	87
α-Globin–myoglobin	118	117	113
β-Globin–myoglobin	118	118	115

* The globins were aligned over 156 residues with deletions or insertions as shown in Dayhoff (1969). Each of the 156 positions were compared.

β-chain of the red kangaroo has probably only one change (β56 Ala to Gly) from that of the grey species. It was of interest to compare the sequences of these chains after aligning them over 156 residues as in Dayhoff (1969). The comparison (Table 8)

shows that the α-globin and β-globin chains show the same number of different residues when compared with myoglobin.

This result is similar to the relationship in bovine and horse globins using the sequences given by Dayhoff (1969). For the amide residues of horse β-globin the data of Smith and Chung (1970) and the predictions for residues 94, 99, 102, and 108 made by Perutz (1969) have been used. The sequence of bovine myoglobin was reported by Han *et al.* (1970). The relatively constant differences between the pairs of globins is evidence for common dates of gene duplication to form α-, β-, and myoglobins. These differences are much greater in number than the differences between the individual globins, e.g. β-globin, of different species, in line with globin differentiation having occurred before the differentiation of the species compared.

V. Acknowledgments

We are indebted to Professor G. B. Sharman and his colleagues for samples of blood and to Miss J. M. Salter, Miss W. K. Fisher, and Mr. R. G. Whittaker for skilled technical assistance. This work was supported in part by the Australian Wool Board and the Australian Research Grants Committee.

VI. References

Air, G. M., and Thompson, E. O. P. (1969).—*Aust. J. biol. Sci.* **22**, 1437.

Air, G. M., and Thompson, E. O. P. (1971).—*Aust. J. biol. Sci.* **24**, 75.

Air, G. M., Thompson, E. O. P., Richardson, B. J., and Sharman, G. B. (1971).—*Nature, Lond.* **229**, 391.

Ambler, R. P., and Meadway, R. J. (1968).—*Biochem. J.* **108**, 893.

Beard, J. M., and Thompson, E. O. P. (1970).—*Aust. J. biol. Sci.* **23**, 185.

Berger, A., and Schechter, I. (1970).—*Phil. Trans. R. Soc.* B **257**, 249.

Braunitzer, G. (1970).—*Biochem. J.* **119**, 31P.

Butler, P. J. G., Harris, J. I., Hartley, B. S., and Leberman, R. (1969).—*Biochem. J.* **112**, 679.

Dayhoff, M. O. (1969).—"Atlas of Protein Sequence and Structure." (National Biomedical Research Foundation: Silver Spring.)

Fruchter, R. G., Jackson, S. A., Mole, L. E., and Porter, R. R. (1970).—*Biochem. J.* **116**, 249.

Gerald, P. S., and Ingram, V. M. (1961).—*J. biol. Chem.* **236**, 2155.

Guidotti, G., Hill, R. J., and Konigsberg, W. (1962).—*J. biol. Chem.* **237**, 2184.

Han, K., Dantrevaux, M., Chaila, X., and Biserte, G. (1970).—*Eur. J. Biochem.* **16**, 465.

Konigsberg, W., and Hill, R. J. (1962).—*J. biol. Chem.* **237**, 2547.

Matsuda, G., Maita, T., Igawa, N., Ota, H., and Miyauchi, T. (1970*a*).—*Int. J. Prot. Res.* **2**, 13.

Matsuda, G., Takei, H., Wu, K. C., Mizuno, K., and Shiozawa, T. (1970*b*).—Abstr. 8th Int. Congr. Biochem., Switzerland. p. 4.

Perutz, M. F. (1969).—*Proc. R. Soc.* B **173**, 113.

Perutz, M. F., Muirhead, H., Cox, J. M., and Goaman, L. C. G. (1968).—*Nature, Lond.* **219**, 131.

Piszkiewicz, G., Landon, M., and Smith, E. L. (1970).—*Biochem. biophys. Res. Commun.* **40**, 1173.

Sanger, F., and Tuppy, H. (1951).—*Biochem. J.* **49**, 463.

Smith, D. B., and Chung, W. P. (1970).—*Can. J. Biochem.* **48**, 1160.

Thompson, E. O. P., Hosken, R., and Air, G. M. (1969).—*Aust. J. biol. Sci.* **22**, 449.

Thompson, E. O. P., Sleigh, R. W., and Smith, M. B. (1971).—*Aust. J. biol. Sci.* **24**, 525.

Yamada, S., and Itano, H. A. (1966).—*Biochim. biophys. Acta* **130**, 538.

Zuckerkandl, E., and Pauling, L. (1965).—In "Evolving Genes and Proteins". (Eds. V. Bryson and H. J. Vogel.) p. 97. (Academic Press: New York.)