Biochem. J. (1975) **149**, 259-269
Printed in Great Britain

The Amino Acid Sequence of the α Chain of the Major Haemoglobin of the Rat (*Rattus norvegicus*)

By CHING GEH CHUA,* ROBIN W. CARRELL* and BERNARD H. HOWARD†
* *Department of Pathology, Christchurch Hospital, Christchurch, and*
† *Department of Biochemistry, Lincoln College, Canterbury, New Zealand*

(*Received* 11 *November* 1974)

1. A partial amino acid sequence of the α chain from the rat (Wistar, *Rattus norvegicus*) major haemoglobin is reported. The soluble tryptic peptides prepared from aminoethylated α-globin were separated by peptide 'mapping'. Sequencing of the tryptic peptides was carried out by the dansyl–Edman method and by the overlapping of smaller peptide fragments derived from secondary enzymic digestion. The insoluble 'core' peptides were further digested with chymotrypsin, thermolysin and pepsin to give smaller soluble peptides for sequencing. The tryptic peptides were ordered on the basis of their homology with the corresponding peptides of human α chain. 2. The proposed sequence is compared with that obtained by using an automated sequencer [Garrick *et al.* (1975) *Biochem. J.* **149**, 245–258]. The differences in sequence resulting from the two methods are discussed. 3. It is suggested that the externally situated cysteine (residue 13) is responsible for the observed inhibition of crystallization of rat haemoglobin at alkaline pH. 4. Detailed evidence for the sequence has been deposited as Supplementary Publication SUP 50047 (9 pages) at the British Library (Lending Division), Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from which copies can be obtained on the terms given in *Biochem. J.* (1975) **145**, 5.

The comparison of amino acid sequences of homologous proteins from different species is now well established as a discipline in the study of species phylogeny. In this context, the haemoglobin structures of many animals are now known and allow the globin phylogenetic tree to be constructed (Dayhoff, 1972). Current research effort is being concentrated on the mammalian haemoglobins.

Stein *et al.* (1971) reported that rat haemoglobin consists of six distinct haemoglobin fractions and crystallizes very rapidly below pH 8.6. Brada & Tobiska (1964) reported that rat haemoglobins also crystallize at low ionic strength.

In the investigation of the multiple haemoglobin system in the rat, the logical first step is a systematic structural study of its globin chains. As the structure–function relationship of the haemoglobin molecule is now well understood (Perutz *et al.*, 1965, 1968), the elucidation of the globin structures may provide some information about the characteristic crystallization of rat haemoglobins.

In the present paper, a partial amino acid sequence of the major rat α-globin is proposed.

Experimental

Animals

The rats used were Wistar albino *Rattus norvegicus* purchased from the University of Otago, Dunedin, New Zealand. The animals belonged to a colony which had a closed breeding history of 12 years. Only adult animals between the ages of 4 and 18 months were used.

Materials

Trypsin (twice crystallized), chymotrypsin (three times crystallized) and pepsin (twice crystallized) were purchased from Worthington Biochemical Corp., Freehold, N.J., U.S.A. Thermolysin (three times crystallized) was obtained from Calbiochem, Los Angeles, Calif., U.S.A.

All reagents used in the dansyl–Edman method were redistilled from analytical reagents. Polyamide sheets used for separation of Dns-amino acids were purchased from Cheng Chin Trading Co., Taipei, Taiwan. Thioglycollic acid (98%, w/w) was purchased from Pierce Chemical Co., Rockford, Ill., U.S.A. Reagents used for electrophoresis and chromatography were all analytical grade.

Sephadex gels used were purchased from Pharmacia Fine Chemicals, Uppsala, Sweden.

Preparation of the major haemoglobin fraction and the separation of the globin chains

Freshly prepared haemoglobin (500–1000 mg) was applied on to a DEAE-Sephadex A-50 column (2.6 cm × 40 cm). Two 0.05 M-Tris–HCl buffers at

pH 8.5 and pH 7.5 were prepared, with KCN added at 100 mg/litre. The haemoglobin fractions were separated by eluting the column with a linear pH gradient consisting of 500 ml of each buffer. Chain separation of the major haemoglobin fraction was carried out by using freshly prepared globin applied to a CM-cellulose column in the presence of 8 M-urea. The purity of the α chain was established by 6 M-urea–starch-gel electrophoresis. The experimental details have been previously published (Chua & Carrell, 1974).

Amino acid analysis of the α-globin

The amino acid composition of the globin was determined by the method of Hill *et al.* (1962), by using a Technicon NC-1 amino acid analyser with type C-2 Chromo-beads as the ion-exchange resin.

The cysteine residues were determined as cysteic acid after performic acid oxidation (Hirs, 1956).

Aminoethylation, tryptic digestion and peptide 'mapping' of the globin

The α-globin was aminoethylated by the method of Jones (1964). The protein was then desalted by passing it through a Sephadex G-25 (coarse grade) column (2.5 cm × 100 cm), and eluting with 0.5% (v/v) formic acid. This was followed by freeze-drying.

The globin was digested with trypsin and the insoluble 'core' peptides were precipitated by heat coagulation, freeze-dried and stored for later work. The soluble tryptic peptides were separated by high-voltage electrophoresis in pH 6.4 buffer (pyridine–water–acetic acid, 25:225:1, by vol.), followed by ascending chromatography with the top phase of the system 3-methylbutan-1-ol–pyridine–water (6:6:7, by vol.). The conditions for the tryptic digestion and the procedures for the peptide 'mapping' of the soluble peptides were as described by Beale (1967). Qualitative peptide 'maps' were prepared by using 1 mg of material, and 3 mg was used for the preparative 'maps'.

Peptide staining

The qualitative peptide 'maps' were stained with 0.2% ninhydrin in acetone followed by specific staining for methionine, aminoethylcysteine, tyrosine, tryptophan, histidine and arginine (Easley, 1965).

Preparative peptide 'maps' were stained with 0.02% ninhydrin solution.

Purification and amino acid analysis of the tryptic peptides

Peptides which did not separate cleanly at pH 6.4 were eluted with 5% (v/v) acetic acid and subjected to further electrophoresis at pH 3.5 (pyridine–acetic acid–water, 1:10:89, by vol.), followed by chromatography, as described above, if necessary. The purified peptides were then hydrolysed in 6 M-HCl (constant-boiling) at 105°C for 20 h before amino acid analysis (Beale, 1967).

Digestion of tryptic peptides with chymotrypsin, thermolysin and pepsin

Pure tryptic peptides were pooled from two or three preparative 'maps' and subjected to additional enzymic digestion: chymotrypsin (Beale, 1967), thermolysin (Lorkin *et al.*, 1970) and pepsin (Konigsberg *et al.*, 1963).

The peptides derived from each tryptic peptide were separated by paper electrophoresis at pH 6.4. A wider sheet of Whatman 3MM paper (54 cm × 28 cm) was used and the position of sample application depended on the charge of the peptides expected from the composition of the tryptic peptides. The electrophoresis was carried out for 50 min. The chromatography, staining and subsequent amino acid analysis of peptides were carried out as described above.

Sequencing of the tryptic peptides

The dansyl–Edman method was used (Gray & Smith, 1970; Gray, 1967*a*,*b*). The Dns-amino acids were identified on 5 cm-square polyamide sheets (Hartley, 1970). When aminoethylcysteine and methionine were detected in the amino acid analysis of a peptide, the Dns-aminoethylcysteine and the Dns-methionine could be detected in good yield only when 2% thioglycollic acid was included in the acid hydrolysis of the dansylated peptides (Matsubara & Sasaki, 1969).

With hydrophilic peptides, up to eight amino acid residues could be sequenced, but with hydrophobic peptides, only about four residues could be sequenced, because of the extensive loss of the hydrophobic peptides into the ethyl acetate used in the removal of the by-products of the Edman degradation.

The sequence of each of the tryptic peptides was established by the following procedures. (1) Amino acid analysis of the tryptic peptide provided the integral ratio of the amino acids present. (2) The tryptic peptide was sequenced as extensively as possible from the *N*-terminal end. (3) Long tryptic peptides were digested with other proteolytic enzymes to give smaller peptides. These peptides were subjected to amino acid analysis and then sequenced. (4) The *C*-terminal residue of each tryptic peptide was assumed to be lysine or arginine or aminoethylcysteine. (5) The sequence of each tryptic peptide was determined by the method of 'sequence overlap' of smaller peptides. When overlap of pep-

tides was not available, sequencing was achieved by comparing the total amino acid composition of the peptide fragments with that of the tryptic peptide.

Sequencing of the insoluble 'core' peptides

The insoluble 'core' peptides were digested with chymotrypsin (Carrell & Irvine, 1968) or with pepsin (Blackwell *et al.*, 1974) or with thermolysin (Lorkin *et al.*, 1970). The peptides were separated on paper (54cm×28cm) by electrophoresis for 50min at pH6.4, followed by ascending chromatography. Some peptides required further purification, and this was done by descending chromatography in butan-1-ol–acetic acid–water–pyridine (15:3:12:10, by vol.) (Waley & Watson, 1953).

The sequences of the 'core' peptides were established by the same procedures used for the tryptic peptides.

Assignment of amide residues

Asparagine and glutamine residues are hydrolysed to their acid forms during amino acid analysis and dansylation. Their presence in the sequence was determined by observing the electrophoretic mobility of the peptides at pH6.4 and assigning electric charge to the following amino acids: lysine (+1), arginine (+1), aminoethylcysteine (+1), histidine (+½), aspartic acid (−1), glutamic acid (−1), asparagine (0) and glutamine (0) (Offord, 1966).

Results

Peptide 'map' of soluble tryptic peptides

The two-dimensional separation of the soluble tryptic peptides from the aminoethylated α chain is shown in Fig. 1. The peptide pattern was similar to the peptide map prepared from human and sheep α chains (Beale, 1967). This was not at all surprising, as the mammalian globins are now well established as a goup of homologous proteins (Dayhoff, 1972).

After the amino acid composition of each tryptic peptide was compared with those from the human, they were assigned numbers by analogy (Lehmann & Huntsman, 1966). Peptides T9A and T9B are together homologous to the human tryptic peptide T9, and peptides T12A, T12B and T12C are together homologous to human tryptic peptide T12.

Tryptic peptide T13 is not detected in the peptide 'map'.

Amino acid analysis

Table 1 shows the amino acid composition for the α chain. This was obtained from a series of amino acid analyses of the globin after three different time-periods of acid hydrolysis.

Table 2 shows the amino acid compositions for the tryptic peptides eluted from the tryptic-peptide 'map' (Fig. 1). The value for the tryptic peptide T13 is inferred from the tentative sequence assigned for this peptide region.

Sequence of the rat major α chain

The proposed partial amino acid sequence of the rat major α chain and the essential sequencing information is shown in Fig. 2. Additional detailed sequencing information for the 'core' peptides is shown in Fig. 3.

Tryptic peptide T1 is acidic at pH6.4, indicating the presence of two aspartic acid residues. The presence of aspartic acid residues near the lysine residue inhibited

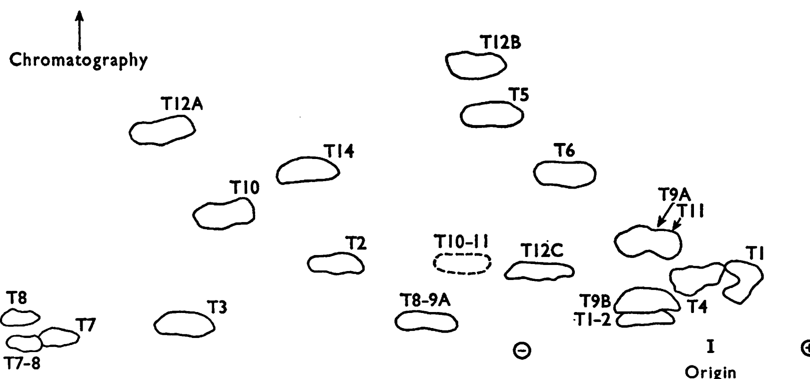


Fig. 1. *Pattern of soluble tryptic peptides from aminoethylated α chain of rat major haemoglobin*

Peptides are numbered homologously with human α chain. Peptides T3, T12A, T12B and T12C are absent from the peptide 'map' prepared with unmodified α chain.

Table 1. *Amino acid analysis of rat major α-globin*

No precaution was taken against oxidation. The results are expressed in molar proportions of the amino acids.

Time of hydrolysis ...	24h	48h	72h	Mean values	Integral values
Asp	12.7	13.0	12.5	12.7	13
Thr	7.7	7.3	7.0	8.0*	8
Ser	9.0	8.6	7.2	9.8*	10
Glu	7.2	7.1	7.4	7.2	7
Pro	5.6	5.8	5.4	5.9*	6
Gly	12.1	12.1	12.5	12.2	12
Ala	16.1	15.6	16.6	16.1	16
Val	8.5	9.4	9.1	9.0	9
Met	1.9	2.0	1.3	2.0†	2
Ile	2.7	3.0	2.8	2.8	3
Leu	16.0	15.0	16.3	15.8	16
Tyr	2.8	2.7	2.5	2.7	3
Phe	7.2	7.0	7.3	7.2	7
Lys	12.1	11.4	11.5	11.7	12
His	9.8	10.5	10.6	10.3	10
Arg	3.1	3.2	3.8	3.3	3
Cysteic acid	3.1	3.3		3.2‡	3
Trp§					1
Total					141

* Extrapolated value at zero time obtained from a plot of quantity liberated against time.
† Only the values for 24h and 48h were used.
‡ Obtained from performic acid-oxidized globin.
§ Deduced from peptide 'map' prepared from the globin: only one peptide was found to stain positively for tryptophan with the Ehrlich reagent.

the trypsin action at this site (Carrell *et al.*, 1974). This explained the low yields for tryptic peptides T1 and T2 and accounted for the presence of tryptic peptide T1-2 at the neutral region of the peptide 'map'.

Tryptic peptide T2 carries one positive charge at pH 6.4, so there is an asparagine residue present.

Tryptic peptide T3 was found only with the aminoethylated globin. Trypsin might be expected to cleave distally to an aminoethylcysteine residue, but no dipeptide of the composition (1 aspartic acid, 1 aminoethylcysteine) was found. This was presumably due to the steric hindrance exerted by the tryptophan residue. The chymotryptic digest yielded two major peptides. Peptide T3C1 stained positively for tryptophan and for bivalent sulphur and carries one positive charge. Amino acid analysis gave the composition: aspartic acid (0.8), aminoethylcysteine (1.0). Dansyl–Edman analysis gave Dns-aspartic acid in the first step and Dns-aminoethylcysteine in the second step. Because the aminoethylcysteine would account for the one positive charge that the peptide carries, it could be concluded that the *N*-terminus is asparagine. Another peptide, T3C2, also carries one positive charge and stained yellow initially with ninhydrin, which was consistent with it having an *N*-terminal glycine residue.

Tryptic peptide T4 stained poorly with ninhydrin and required purification with further electrophoresis at pH 3.5. Peptide T4Th2 carries one positive charge and therefore it has a glutamine residue. Peptide T4Th3 is neutral and must therefore have a glutamic acid residue whose ionization is partially suppressed by the *C*-terminal α-carboxyl group. Peptide T4Th4 stained yellow initially with ninhydrin: it carries two negative charges and therefore has two glutamic acid residues.

Tryptic peptide T5 has one positive charge at pH 6.4. Both peptides T5Th2 and T5Th5 stained yellow initially with ninhydrin. The staining characteristics are due to the occurrence of *N*-terminal proline or threonine. It is noteworthy that thermolysin cleaved the peptide bond involving the imino group of proline (peptide T5Th5).

Tryptic peptide T6 stained poorly with ninhydrin. Its sequence was determined by overlapping the chymotryptic and thermolytic peptides. Peptide T6Th6 is acidic and peptide T6Th7 is neutral, which indicated that aspartic acid is present in the former and glutamine is present in the latter peptide.

Tryptic peptide T7 required purification by twice-repeated electrophoresis at pH 6.4.

Tryptic peptide T8 was detected in the peptide 'map' at a position similar to that of peptide T8 from human α chain. Its presence in the globin was further supported by the detection of peptide T7-8, with the amino acid composition glycine (1.0), alanine (0.8), lysine (1.5), histidine (1.1), and of peptide T8-9A,

Table 2. *Amino acid analysis of the tryptic peptides eluted from the peptide 'map'* (*Fig.* 1)

The results are expressed in molar proportions of the amino acids. Values in parentheses give the nearest whole numbers.

	T1	T2	T3	T4	T5	T6	T7	T8	T9A	T9B	T10	T11	T12A	T12B	T12C	T13† 'Core' Th7	T13† 'Core' P9	T14	Sequence	Analysis§
Asp	2.0 (2)	1.0 (1)	0.8*(1)			1.1 (1)			1.0 (1)	3.4 (3)		2.1 (2)			2.0 (2)	0.8 (1)			14	13
Thr		0.6*(1)			1.9 (2)	0.6*(1)				1.1 (1)				1.1 (1)	0.9 (1)		1.0*(1)		8	8
Ser	1.0 (1)					3.2 (3)				2.2 (2)			0.8 (1)		1.0 (1)	1.3 (1)	1.0 (1)		10	10
Glu				4.4 (4)		1.0 (1)				1.2 (1)									6	7
Pro					0.7 (1)	1.0 (1)				0.9 (1)		0.8 (1)			1.7 (2)				6	6
Gly			1.3 (1)	5.0 (5)		0.9 (1)	1.0 (1)			1.0 (1)					1.2 (1)	0.9 (1)			11	12
Ala	1.1 (1)			1.1 (1)	2.3 (2)	1.2 (1)	0.8*(1)		2.8 (3)	3.6*(4)				1.1 (1)	1.8 (2)	0.7 (1)			17	16
Val	0.6*(1)					1.7 (2)			1.0*(1)	1.1 (1)		1.6*(2)		1.2 (1)		0.8 (1)			9	9
Met					0.4*(1)										0.5 (1)	0.6 (1)			3	2
Ile		1.0 (1)		0.9*(1)		0.9 (1)													3	3
Leu	1.0 (1)			0.7 (1)					1.2 (1)	4.3 (4)	0.8*(1)		1.0 (1)	2.6*(3)	1.3 (1)	1.9 (2)			15	16
Tyr				0.7 (1)		0.7 (1)												0.5*(1)	3	3
Phe					1.9 (2)	0.9 (1)						1.0 (1)	0.3*(1)		1.1 (1)	0.6*(1)			7	7
Lys	0.9 (1)	1.2 (1)	1.0 (1)		1.2 (1)	1.1 (1)	1.3 (1)	1.0 (1)	0.8 (1)	1.0 (1)		1.0 (1)			1.3 (1)		0.7 (1)		12	12
His				0.8 (1)		1.1 (1)	1.1 (1)			3.0 (3)			1.0 (1)		2.9*(3)				10	10
Arg				0.8 (1)							1.0 (1)							1.0 (1)	3	3
Cys‖			1.0 (1)										1.0 (1)	0.7 (1)					3	3
Trp			‡ (1)																1	1
																		Total	141	141

* *N*-Terminal amino acid partially destroyed by ninhydrin. The *N*-terminal amino acids were positively identified by dansylation.

† Tryptic peptide T13 was not found in the peptide 'map'. Two major peptide fragments obtained from the 'core' peptides were assumed to be derived from this tryptic peptide region.

‡ Peptide stained positively for tryptophan with the Ehrlich reagent.

§ Amino acid analysis for whole globin from Table 1.

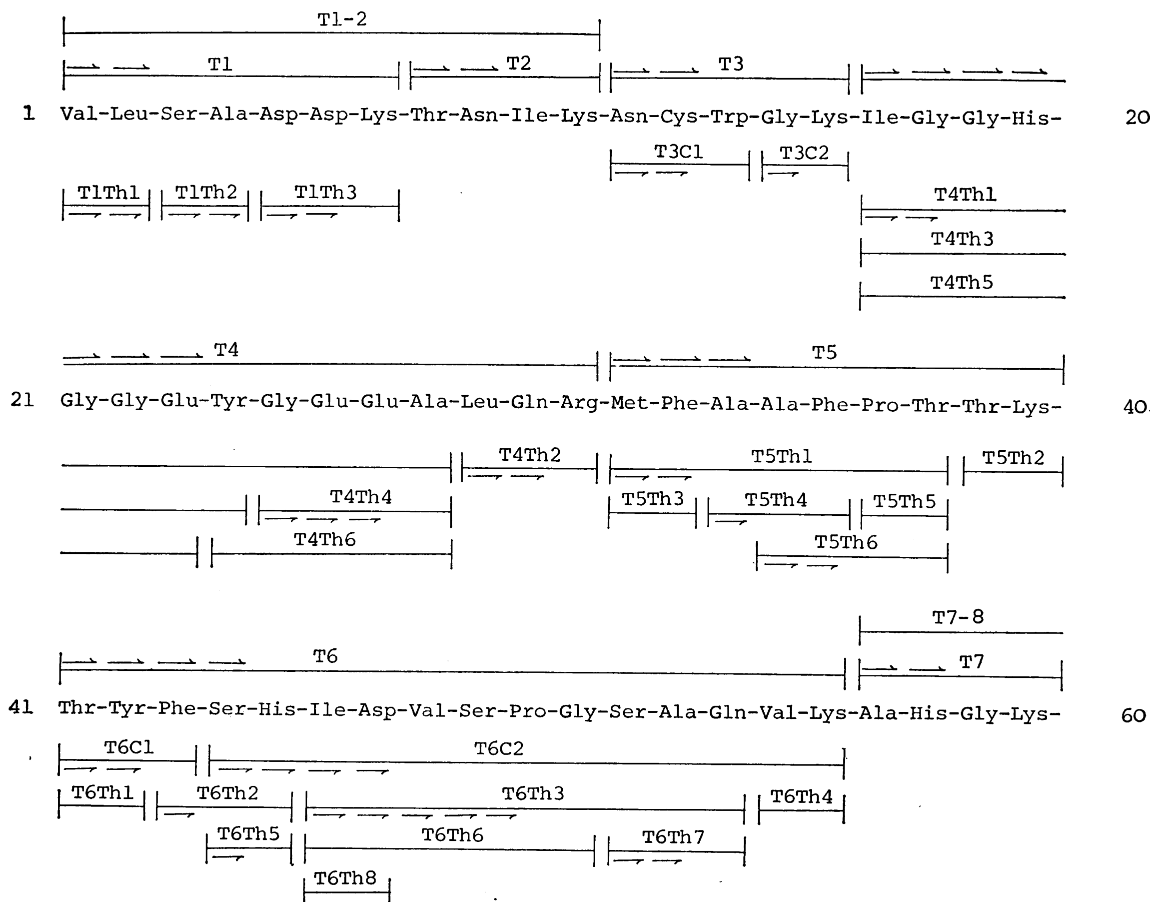
‖ As aminoethylcysteine.

with the amino acid composition aspartic acid (1.1), alanine (3.0), valine (0.8), leucine (0.7), lysine (1.7). The same peptide pattern is detected in the tryptic digest of human *α* chain, where two lysine residues also occur adjacent to each other in the homologous region (Lehmann & Huntsman, 1966).

Tryptic peptide T9A has an amino acid composition very similar to that of peptide T11. Their separation was only achieved with electrophoresis at pH 3.5 followed by both ascending and descending chromatography. Peptide T9A is a neutral peptide, therefore it has an aspartic acid residue.

Tryptic peptide T9B is a long peptide. It is one of the most hydrophilic peptides and is neutral. Digestion with thermolysin and pepsin failed to cleave peptides T9B/Th1 and T9B/P1 into smaller fragments. The predominantly polar nature of these peptides may be responsible for the inhibition of the enzyme actions. Both peptides T9B/Th1 and T9B/P1 carry approximately one negative charge at pH 6.4, indicating that each has two acidic residues and one amide residue. However, the unavailability of smaller peptide fragments for this peptide region prevented the precise assignment of the acidic and the amide residues. For the same reason, the peptide region (Glx.Pro.Val.Ala) could only be assigned by homology with sequences from mouse, dog and rabbit (Dayhoff, 1972). The peptide region (Leu-Ser-Thr-Leu-Ser-Asp) is also shown in parentheses in Fig. 2 to indicate that the sequence is deduced by homology. The information available from the smaller fragments could also support a sequence with the threonine and the aspartic acid residues being interchanged. Both peptides T9B/Th3 and T9B/P3 are acidic, therefore each has an aspartic acid residue.

Tryptic peptide T10 is similar to that of human *α* chain in amino acid composition and in its position in the peptide 'map'. The position of this peptide in the *α* chain is further supported by the presence of tryptic peptide T10-11, a product of the inhibition of tryptic cleavage at the arginine site. The aspartic acid residue in the tryptic peptide T11, which is two



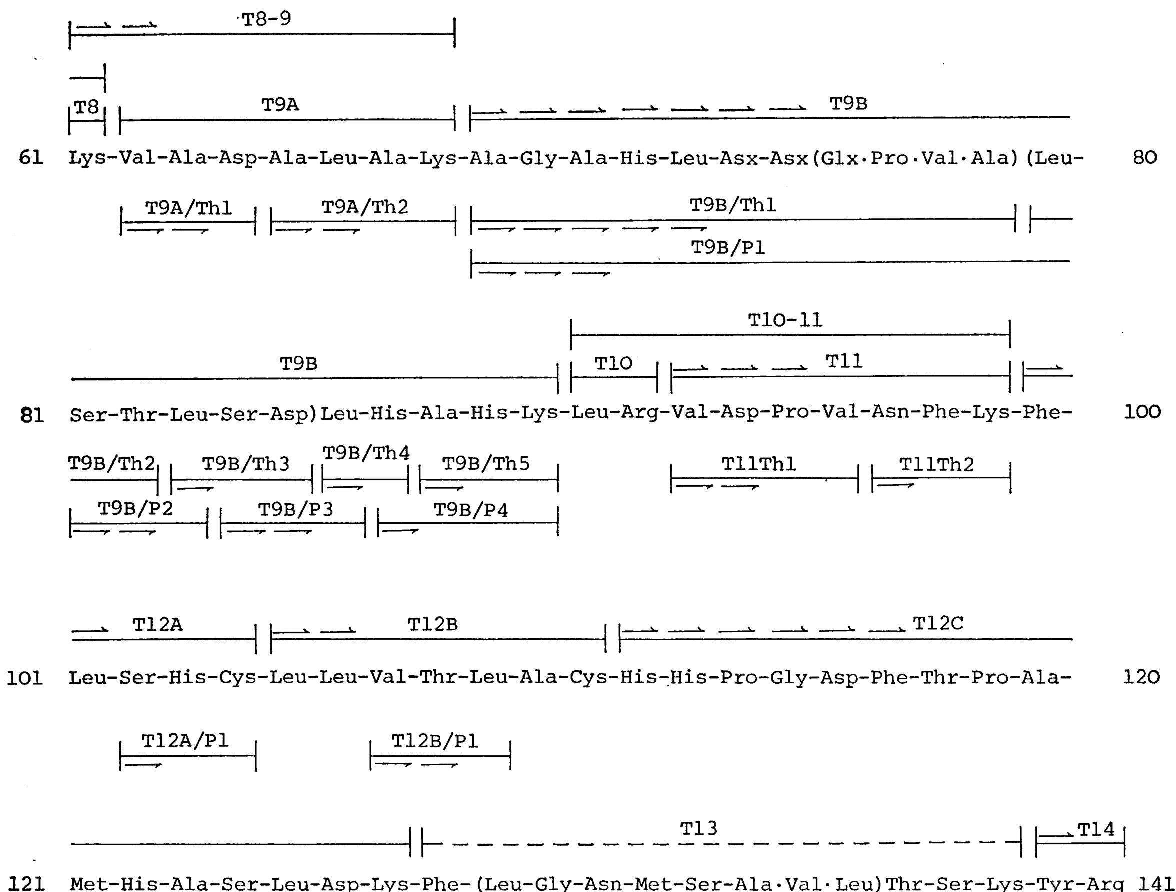


Fig. 2. *Proposed sequence of the rat major α chain* (*Wistar, Rattus norvegicus*)

Tryptic peptides from the peptide 'map' (Fig. 1) are enclosed within vertical bars above the sequence and positive dansyl–Edman identifications of the residues are indicated by ⇀. Further digestion of the tryptic peptide with chymotrypsin, thermolysin, or pepsin yielded smaller peptides denoted by C, Th or P respectively. Thus peptide T4Th1 indicates a thermolysin peptide obtained from tryptic peptide T4. These smaller peptide fragments are enclosed with vertical bars below the sequence. Amino acids identified by the dansyl–Edman method are indicated by ⇁. Tryptic peptide T13 was not found among the soluble tryptic peptides and is denoted by a broken line above the sequence. The sequence for tryptic peptide T13 was obtained from the 'core' peptides (see Fig. 3). The cysteine residues in the sequence were identified as aminoethylcysteine. Hyphens between amino acids indicate that the amino acid to the left has been positively identified either by the dansyl–Edman method or by quantitative amino acid analysis. An exception is tryptophan, which was identified by positive staining with Ehrlich reagent. Sequences in parentheses indicate that the composition but not the complete sequence has been determined experimentally. Median points within parentheses indicate that the amino acids to their left have been placed with at least 90% confidence by homology with other mammalian globins (Dayhoff, 1972). As discussed in the text, the regions of residues 129–136 should be regarded as tentative.

residues away from the arginine, is responsible for the inhibition. The amino acid composition of the peptide T10-11 is: aspartic acid (2.0), proline (1.1), valine (1.7), leucine (1.1), phenylalanine (1.2), lysine (1.1), arginine (1.1).

Tryptic peptide T11 was purified from peptide T9A by further electrophoresis at pH 3.5. Thermolysin digestion yielded two major peptides. Peptide T11Th1 is acidic, and so has an aspartic acid residue. Peptide T11Th2 carries one positive charge at pH 6.4, so has an asparagine residue.

Tryptic peptides T12A, T12B and T12C were obtained only from aminoethylated globin. They were present in lesser amount than other tryptic

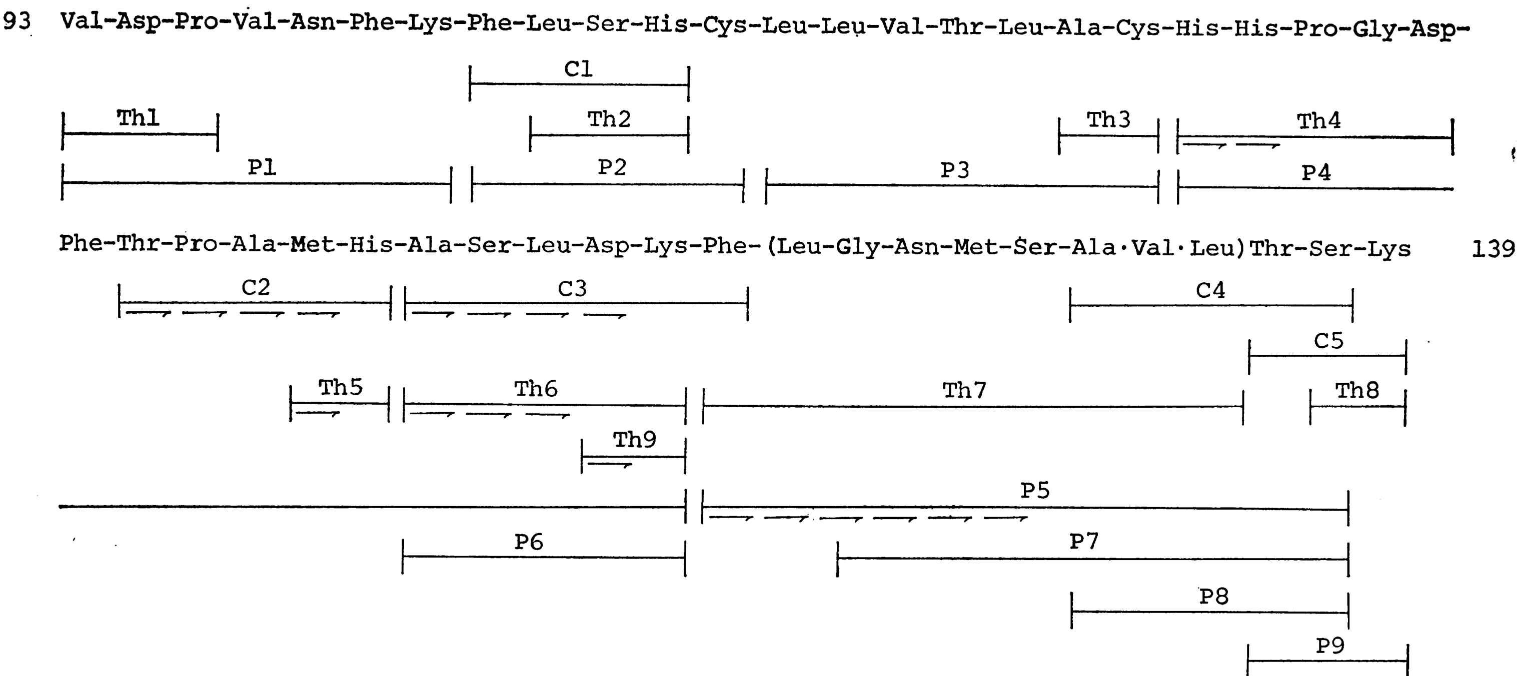


Fig. 3. *Proposed amino acid sequence for the 'core' peptides (residues 93–139) of rat major α chain*

The notations used are the same as for Fig. 2. Peptide 'map' and amino acid analysis of the peptide fragments are presented in Supplementary Publication SUP 50047.

peptides. This is because trypsin cleavage at the aminoethylcysteine site is not as efficient as at lysine or arginine sites (Plapp *et al.*, 1967; Wang & Carpenter, 1968). The penultimate aspartic acid residue in peptide T12C also inhibited tryptic cleavage at the lysine residue distal to it. All these four peptides (i.e. T12A, T12B, T12C, T13) constitute the insoluble 'core' peptides. Complete sequences for peptides T12A and T12B, and a partial sequence for peptide T12C, were determined after eluting them from the peptide 'map'. The remaining sequence for peptide T12C was obtained from the 'core' peptide.

Tryptic peptide T13 could not be found in the peptide 'map' despite extensive staining with ninhydrin and use of specific amino acid stains.

Tryptic peptide T14 has an amino acid composition and a position in the peptide 'map' similar to that of tryptic peptide T14 from human α chain. Dansylation of this peptide gave both Dns-tyrosine and bis-Dns-tyrosine.

The 'core' peptides, after digestion with chymotrypsin, thermolysin and pepsin, gave a series of similar peptide 'maps'. These peptide 'maps' are complicated owing to the non-specific nature of the enzymes. Nevertheless, a number of major peptide fragments could be purified by further electrophoresis and chromatography. Two of these fragments, peptides 'core' C2 and 'core' Th6, allowed the sequence for peptide T12C to be completed. Further, another two major peptide fragments, 'core' P5 and 'core' Th7, gave amino acid compositions which fitted them into the peptide-T13 region. Both peptides stained strongly for bivalent sulphur and each contains one methionine residue. The partial sequence of the peptide 'core' P5 was elucidated by the dansyl–Edman method; 2% thioglycollic acid was included in the 6M-HCl used for the sequencing of this peptide. However, the information is insufficient to establish conclusively that both peptides 'core' P5 and 'core' Th7 are indeed derived from the region for tryptic peptide T13 and therefore the amino acid sequence for this region is enclosed in parentheses to indicate its tentative nature.

Supplementary information

Detailed evidence for the partial amino acid sequence of the α chain has been deposited as Supplementary Publication SUP 50047 with the British Library (Lending Division), Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K. The information consists of peptide 'maps' of enzymic digests of tryptic peptides and amino acid analyses of the fragments. They are: (1) thermolysin digests of tryptic peptides T4, T5 and T6; (2) thermolysin and pepsin digests of tryptic peptide T9B; (3) thermolysin, chymotrypsin and pepsin digests of the 'core' peptides.

Discussion

Technique

To obtain a peptide 'map' with good peptide separation, not more than 3mg of whole globin digest should be applied on Whatman 3MM paper. This quantity would give 200nmol of each peptide. It was found that after staining the peptide 'map' with 0.02% ninhydrin, 60nmol of the *N*-terminal residue was destroyed by the ninhydrin. Consequently the amino acid analysis for the peptide eluted from the stained peptide 'map' gave a low value for the *N*-terminal amino acid. To provide an accurate interpretation of the amino acid analysis of the peptide, dansylation was performed to identify the *N*-terminal amino acid.

The interpretation of the amino acid analyses for long peptides such as tryptic peptides T4, T6, T9B and T12C required the consideration of the hydrolysis behaviour of the polypeptide, the destruction of serine, threonine and proline, and the resistance towards acid hydrolysis of peptide bonds involving valine, leucine and isoleucine (Tristram, 1966). For a final check, the amino acid composition of each long peptide must be compatible with the composition of its smaller peptide fragments (Supplementary Publication SUP 50047).

Tentative sequence in the region of residues 129–136

The proposed sequence for this peptide region was given in parentheses in Fig. 2 to indicate its tentative nature. There are three aspects of the sequence for this region which demand more investigation and also provoke some thought about the sequencing of homologous proteins. (1) The sequencing information for this peptide region was obtained from two major peptide fragments derived from the 'core' peptides; their positioning in the globin chain rested solely on their amino acid composition which made them fit best into the tryptic-peptide T13 region. But the lack of substantial sequence overlap between this peptide region and tryptic peptides T12C and T14 made the sequence proposed less conclusive. (2) The sequence proposed for this peptide region differs quite extensively from the homologous regions of other mammals closely related to the rat (Dayhoff, 1972):

Rabbit	Leu-Ala-Asn-Val-Ser-Thr-Val-Leu
Dog (α^1)	Phe-Ala-Ala-Val-Ser-Thr-Val-Leu
Dog (α^2)	Phe-Thr-Ala-Val-Ser-Thr-Val-Leu
Mouse	Leu-Ala-Ser-Val-Ser-Thr-Val-Leu
Rat	(Leu-Gly-Asn-Met-Ser-Ala·Val·Leu)
	129 130 131 132 133 134 135 136

Thus the homology of the globin family poses a question as to the validity of the proposed sequence for this region. On the other hand, one must ask how much should sequencing work be influenced by the guideline of homology. (3) The presence of methionine in this region increases the methionine content in the α chain to three residues, but only two methionine residues were obtained in the total amino acid analysis for the globin (Tables 1 and 2). Although a 30% oxidative destruction of methionine residues during acid hydrolysis is possible, this is not really a satisfactory explanation. Thus the sequence for the peptide region residues 129–136 requires further confirmation.

Comparison of the proposed sequence with that obtained by using the automated sequencer (Garrick et al., 1975)

The α chain of the rat major haemoglobin reported in this work is the same as the $^{I}\alpha$-globin reported by Garrick *et al.* (1975). Two different approaches have been used to obtain the pure major α-globin. In our work, the major haemoglobin fraction was first separated from the haemolysate by a DEAE-Sephadex column. The globin was then prepared and separated on a CM-cellulose column (Chua & Carrell, 1974). In the work reported by Garrick *et al.* (1975), the major α-globin was separated from the total rat globins by two-step column separation on CM-cellulose. Two different buffer gradient systems were used for elution.

Our proposed partial sequence agrees very well with that obtained from the sequencer (residues 1–60, 122–128), except in one residue at structural position 44. We have proposed a serine residue for this position, whereas asparagine was proposed by Garrick *et al.* (1975).

In the peptide regions containing residues 61–121 and 129–141, deduced by homology by Garrick *et al.* (1975), the following differences are observed when their sequences are compared with ours:

The differences demonstrated between the two sequences underline the problems that occur in making phylogenetic or mutation-rate deductions from published globin structures. As more sequences become available, the use of homologies has an increasing influence on the determination of sequences. This may occur electively to fill in undetermined areas, as discussed by Garrick *et al.* (1975), but there is also a more subtle bias. This arises from the use of homologies by sequence workers as a check on their own work. If a sequence is in expected agreement with homologies, it is accepted, but if in disagreement it is rechecked. Inevitably this will lead to a bias of results to the homologous norm. In this study we have endeavoured to avoid this bias, although obviously it has influenced us in labelling the region 129–136 as tentative.

Cysteine-13 and the properties of rat haemoglobin

The physiological pH of the rat erythrocyte is 7.4. The pK_a for the thiol group of cysteine is between 8.4 and 9.0, so at pH 7.4 the thiol group would be uncharged and the side chain of cysteine-13 would be favourably placed towards the interior of the molecule. Stein *et al.* (1971) reported that rat haemoglobins crystallized rapidly at low pH and the crystallization process could be prevented by raising the pH to 8.6. This phenomenon could be explained by the existence of cysteine-13 in the rat major α-chain. At pH 8.6, the thiol group of this cysteine would become ionized and would presumably swing to the exterior of the molecule. Crystallization could consequently be inhibited by the presence of the ionized side chain with its hydration shell, or more likely by the resultant small change in the tertiary structure. In support of this, we have observed that reaction of rat haemoglobins with iodoacetamide similarly prevents subsequent crystallization. The inhibition of crystallization in the rat is therefore a lesser case of the situation seen in alkali denaturation of haemoglobin, where ionization of internal cysteine residues produces major molecular unfolding (Perutz, 1974). In the rat, the ionization of a partially buried cysteine residue is presumably sufficient to prevent crystal formation.

	70	71	72	76	78	100	111	113	115	130	131	132	134
Chua *et al.*	Gly	Ala	His	Glx	Val	Phe	Cys	His	Gly	(Gly)	(Asn)	(Met)	(Ala)
Garrick *et al.*	Val	Glx	Asx	Leu	Gly	Leu	Asx	Gly	Ser	Ala	Ser	Val	Thr

The supporting evidence for our proposed sequence for residues 100 and 111 was published in Chua & Carrell (1974), and that for residues 44, 70, 71, 72, 76, 78, 113 and 115 is given in the Supplementary Publication SUP 50047. The tentative nature of our proposed sequence for residues 130, 131, 132 and 134 has been discussed above.

The problem of crystallization is one that only occurs *in vitro*, as it is confined to low-ionic-strength solution. In the rat erythrocyte the ionic concentration is presumably sufficient to prevent molecular aggregation at pH 7.4 (Brada & Tobiska, 1964).

Note Added in Proof (Received 2 April 1975)

Overall sequence: rat major α chain

The following is the consensus between ourselves and Garrick *et al.* (1975), concerning the differences in the sequences for the major α chain of the rat. (1) There appears to be polymorphism at residue 44 which can be occupied by either serine or asparagine (L. M. Garrick, R. L. Sloan, T. W. Ryan & M. D. Garrick, unpublished work). (2) Positions 13 and 111 as well as 104 are occupied by cysteine. (3) The favoured sequences for residues 70–72, 76, 78, 100, 113 and 115 are those reported in the present paper. (4) The favoured sequence for 130–134 is that of Garrick *et al.* (1975).

We acknowledge the courtesy of Dr. H. M. Ranney in allowing us to see the results of the investigations in her laboratory before publication (Garrick *et al.*, 1975). This work was supported in part by the Medical Research Council of New Zealand. C. G. C. was the recipient of a University Grants post-graduate scholarship.

References

Beale, D. (1967) *Biochem. J.* **103**, 129–140

Blackwell, R. Q., Wong, H. B., Wang, C.-L., Weng, M.-I. & Liu, C.-S. (1974) *Biochim. Biophys. Acta* **351**, 7–12

Brada, Z. & Tobiska, J. (1964) *Life Sci.* **3**, 77–80

Carrell, R. W. & Irvine, D. (1968) *Biochim. Biophys. Acta* **154**, 78–83

Carrell, R. W., Owen, M. C., Anderson, R. & Berry, E. (1974) *Biochim. Biophys. Acta* **365**, 323–327

Chua, C. G. & Carrell, R. W. (1974) *Biochim. Biophys. Acta* **365**, 328–334

Dayhoff, M. O. (ed.) (1972) *Atlas of Protein Sequence and Structure*, vol. 5, pp. D51–D85, National Biomedical Research Foundation, Washington, D.C.

Easley, C. W. (1965) *Biochim. Biophys. Acta* **107**, 386–388

Garrick, L. M., Sharma, V. S., McDonald, M. J. & Ranney, H. M. (1975) *Biochem. J.* **149**, 245–258

Gray, W. R. (1967*a*) *Methods Enzymol.* **11**, 139–151

Gray, W. R. (1967*b*) *Methods Enzymol.* **11**, 469–475

Gray, W. R. & Smith, J. F. (1970) *Anal. Biochem.* **33**, 36–42

Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822

Hill, R. J., Konigsberg, W., Guidotti, G. & Craig, L. C. (1962) *J. Biol. Chem.* **237**, 1549–1554

Hirs, C. H. W. (1956) *J. Biol. Chem.* **219**, 611–621

Jones, R. T. (1964) *Cold Spring Harbor Symp. Quant. Biol.* **29**, 297–308

Konigsberg, W., Goldstein, J. & Hill, R. J. (1963) *J. Biol. Chem.* **238**, 2028–2033

Lehmann, H. & Huntsman, R. G. (1966) *Man's Haemoglobins*, pp. 240–244, North-Holland Publishing Co., Amsterdam

Lorkin, P. A., Charlesworth, D., Lehmann, H., Rahbar, S., Tuchinda, S. & Lie-Injo, L. E. (1970) *Br. J. Haematol.* **19**, 117–125

Matsubara, H. & Sasaki, R. M. (1969) *Biochem. Biophys. Res. Commun.* **35**, 175–181

Offord, R. E. (1966) *Nature (London)* **211**, 591–593

Perutz, M. F. (1974) *Nature (London)* **247**, 341–344

Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13**, 669–678

Perutz, M. F., Muirhead, H., Cox, J. M. & Goaman, L. C. G. (1968) *Nature (London)* **219**, 131–139

Plapp, B. V., Raftery, M. A. & Cole, R. D. (1967) *J. Biol. Chem.* **242**, 265–270

Stein, S., Cherian, M. G. & Mazur, A. (1971) *J. Biol. Chem.* **246**, 5287–5293

Tristram, G. R. (1966) in *Techniques in Amino Acid Analysis* (Schmidt, D. I., ed.), pp. 61–72, Technicon Instruments Co. Ltd., Chertsey

Waley, S. G. & Watson, J. (1953) *Biochem. J.* **55**, 328–337

Wang, S.-S. & Carpenter, F. H. (1968) *J. Biol. Chem.* **243**, 3702–3710