Primary Structure of the Major β -Chain of Rat Haemoglobins

By LAURA M. GARRICK, RICHARD L. SLOAN, THOMAS W. RYAN,
THOMAS J. KLONOWSKI and MICHAEL D. GARRICK*

*Departments of Medicine, Biochemistry and Pediatrics, State University of New York at Buffalo,
Buffalo, NY 14214, U.S.A.*

(Received 24 October 1977)

The amino acid sequence of the major β -chain, "β, from rat haemoglobins was established with an automated sequencer. Amino acid heterogeneities were found that appear to result from allelic variation at particular residues. We applied several new or unusual techniques in determining the sequence: (1) reaction of the polypeptide with dansylaziridine for detection of cysteine; (2) blockage of the *N*-terminal residue and the ϵ -amino group of lysine residues with 1-fluoro-2-nitro-4-trimethylammoniumbenzene iodide and subsequent identification of the modified lysine phenylthiohydantoin by absorbance at 420 nm; (3) identification of histidine phenylthiohydantoin by its blue fluorescence under long-wave u.v. light; (4) cleavage of the chain into two or three fragments and subsequent sequencing without purification [a detailed statement giving the major phenylthiohydantoins assigned at each step for each sequence run before their alignment in individual sequences has been deposited as Supplementary Publication SUP 50084 (10 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1978) 169, 5]; (5) separation of fragments produced by CNBr cleavage by cation-exchange chromatography; (6) peptide sequencing after attachment of the peptide to cytochrome *c*. The amino acid sequence was confirmed by amino acid compositions of the complete chain, of CNBr fragments 1 and 3, and of 11 purified tryptic peptides.

Rat haemoglobins are unusually heterogeneous compared with those of other mammals. In each individual adult rat, there are at least six, probably more, different haemoglobins. Two of the α -chains have been partially sequenced and found to be structurally different (Garrick *et al.*, 1975). We now describe the primary structure of the major β -chain, "β.

Materials and Methods

Purification of globin chains

Adult Wistar rats were bled from the orbital sinus. Cells were washed and lysed and globin was prepared as described previously (Garrick *et al.*, 1975). Globin was fractionated by a modification of the method of Clegg *et al.* (1966). We used two types of separation: (1) a preparative type, which involved about 1.1 g of globin, and (2) an analytical type, which used only about 40 mg of globin. Globin was dissolved in 8 M-urea, containing 5 mM- $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ and 0.05 M-2-mercaptoethanol (pH 7.05).

* Reprint requests should be addressed to M. D. Garrick, Department of Biochemistry, State University of New York at Buffalo, Buffalo, N.Y. 14214, U.S.A.

Samples were dialysed for 2 h against the same buffer and then placed on columns (4 cm $\times$ 23 cm for the preparative run; 0.7 cm $\times$ 15 cm for the analytical run) of CM-cellulose (Whatman CM-52; Reeve Angel, Clifton, NJ, U.S.A.) equilibrated with the same buffer. Chains were eluted with a linear gradient of increasing NaCl concentration in buffer (0–25 mM). For the preparative run, 8 litres were used for the gradient; for the analytical run, 400 ml was used. We made the second buffer from the first buffer by adding NaCl to give a final concentration of 25 mM. For the preparative run, 24 ml fractions were collected; for the analytical run, fractions were 3.2 ml. Pooled fractions were dialysed in Spectrapor 1 membrane tubing (Beckman Instruments, Science Essentials Operations, Mountainside, NJ, U.S.A.) against 0.1 M-acetic acid and freeze-dried.

Separation of fragments produced by CNBr cleavage

We used the method of Clegg *et al.* (1966), with a 0.7 cm $\times$ 15 cm column of CM-cellulose, to separate the fragments produced after CNBr cleavage. The starting buffer was the same as that described above for the separation of intact chains. A sample (approx 40 mg) was dissolved in starting buffer and dialysed

against starting buffer for 2h in Spectrapor 3 membrane tubing before chromatography. Fragments were eluted with a 400ml linear gradient of increasing NaCl concentration (0–75mM). Fraction size was 3.2ml. Since CNBr peptide 2 (residues 56–109) had no tryptophan or tyrosine residues, we could not rely on detection by A_{280} . Instead a sample (0.3ml) was taken from each fraction and mixed with 2.7ml of 6% (w/v) trichloroacetic acid. Turbidity was read at 420nm. Nevertheless peptide CB2 was not found.

Dansylaziridine treatment

The method for reaction of globin with dansylaziridine combines the method for aminoethylation (see below) with the method of Scouten *et al.* (1974) for reaction of cysteine with dansylaziridine. After dissolving 16mg of purified globin chain in 3ml of 8M-urea containing 0.55M-Tris/HCl, 5mM-EDTA and 0.28M-2-mercaptoethanol (pH8.6), we added 3.6mM-dansylaziridine (Pierce Chemical Co., Rockford, IL, U.S.A.) in 95% (v/v) ethanol to the sample in three 0.5ml portions at $\frac{1}{2}$ h intervals. After each addition the sample was flushed with N_2 and, after the final addition, stirred for 1h at 25°C. The sample was then dialysed in Spectrapor 1 tubing against 0.1M-acetic acid and freeze-dried.

Aminoethylation and succinylation

Globin was aminoethylated or succinylated (3-carboxypropionylated) as described previously (Garrick *et al.*, 1975) except that, after the reaction, the sample was dialysed in Spectrapor 1 tubing against 0.1M-acetic acid and freeze-dried.

Blockage with 1-fluoro-2-nitro-4-trimethylammonio-benzene iodide

The N-terminus and the ϵ -amino group of lysine residues were made to react by the method of Sutton *et al.* (1972). After dissolving 10mg of purified globin chain in 0.2ml of water, we added 0.8ml of dimethyl sulphoxide and adjusted the pH to 9.0 with 0.1M-NaOH. A 10-fold molar excess relative to nucleophilic groups of 1-fluoro-2-nitro-4-trimethylammonio-benzene iodide (Eastman Kodak Co., Rochester, NY, U.S.A.) was added in three portions at 1h intervals. The pH was maintained at 9.0 with 0.1M-NaOH, while the reaction proceeded at 37°C for 4h. The sample was then dialysed in Spectrapor 1 tubing against 0.1M-acetic acid and freeze-dried.

Cleavage of globin chains

With CNBr. Cleavage with CNBr has been previously described (Garrick *et al.*, 1975).

Cleavage between aspartic acid and proline residues. Samples were dissolved in 70% (v/v) formic acid (5mg/ml) and flushed with N_2 . Cleavage was at 50°C for 9h. After 5-fold dilution with water, samples were freeze-dried. This procedure is based on that of Bricker & Garrick (1974).

Cleavage after glutamic acid residues. To effect cleavage after glutamic acid residues, we used the method of Houmard & Drapeau (1972). The sample to be cleaved was dissolved in 50mM-ammonium acetate to yield a 1% (w/v) protein solution. *Staphylococcus aureus* proteinase (Miles Laboratories, Elkhart, IN, U.S.A.) was then added in a 1:30 (enzyme:globin) molar ratio. After 18h at 37°C, the sample was freeze-dried.

Tryptic digestion

A 5mg sample of purified chain was dissolved in 1.5ml of water; 1.5ml of 2% (w/v) $(NH_4)_2CO_3$ was then added, with stirring. Digestion was with 50 μ g of trypsin (treated with L-1-chloro-4-phenyl-3-L-tosylamidobutan-2-one; Worthington Biochemical Corp., Freehold, NJ, U.S.A.) for 3h at 22°C. The reaction was stopped by addition of a few drops of acetic acid and the sample was freeze-dried. Essentially the same procedure was applied to succinylated chain.

Peptide 'mapping'

Samples (5mg) that had been digested with trypsin were 'mapped' as previously described (Garrick *et al.*, 1975). Insoluble 'core' peptides were first removed by centrifugation at 14 000g for 5min.

Carboxypeptidase A digestion

Di-isopropyl phosphorofluoridate-treated carboxypeptidase A was purchased from Worthington. The digestion procedure was adapted from that of Ambler (1972), with an emphasis on minimizing yield of basic residues and increasing that for acidic residues: 200nmol of rat $^{11}\beta$ -globin or of $^{11}\beta$ -chain that had been cyanylated with 2-nitro-5-thiocyanobenzoic acid (Degani & Patchornik, 1971) and cleaved as described by Jacobson *et al.* (1973) was incubated with 150 μ g of enzyme in 0.3M-morpholine/acetate buffer, pH6.0, at 37°C. At selected intervals a 40nmol sample was removed, diluted 50-fold in 1M-acetic acid, freeze-dried and then subjected to amino acid analysis.

Amino acid analysis

Samples for analysis were dissolved in 6M-HCl containing 9mg of phenol/100ml and hydrolysed under N_2 at 110°C for the times indicated in the

Tables. Peptides from tryptic digests were hydrolysed for 18 h. Hydrolysates were flash-evaporated and their compositions determined with a Beckman 120C amino acid analyser that had been modified to the single-column system of the manufacturer.

Attachment of peptides to cytochrome c

Peptides were covalently attached to the ϵ -amino group of lysine residues in cytochrome *c* by a water-soluble-carbodi-imide technique (Garrick & Sloan, 1977).

Sequencing

A Beckman sequencer, model 890B, was used for sequence studies. Reagents and solvents for the sequencer were purchased from Pierce Chemical Co. About 300 nmol of protein was used for each run. A single-cleavage Quadrol [NNN'-tetrakis-(2-hydroxypropyl)ethylenediamine] program (Garrick *et al.*, 1975) was used. This program produced a repetitive yield of 94–97%, based on analysis of selected residues by g.l.c. The first degradation cycle was carried out without phenyl isothiocyanate; this procedure gave a more uniform coating of the protein film in the cup. We added 7 mM-dithioerythritol to the 1-chlorobutane to preserve serine and threonine residues (Hermodson *et al.*, 1970). The dried thiazolinone derivatives were converted into phenylthiohydantoin derivatives by heating at 80°C for 10 min in 1 M-HCl containing 0.2% (v/v) ethanethiol (Hermodson *et al.*, 1972). The products were then extracted into ethyl acetate.

The extracted residues were dried, and then dissolved in 10 μ l of ethyl acetate. These derivatives were identified by t.l.c., which was carried out on pre-coated silica-gel plates (E. Merck, Brinkman Instruments, Westbury, NY, U.S.A.; type 60-F-254) with fluorescent indicator and glass backing. For the first 15 residues, 1 μ l was usually spotted; for the next 10 residues, 2 μ l was spotted, and so on, with a 1 μ l increase for every 10 residues until identification was no longer possible. The solvent systems used were 1,2-dichloroethane/propionic acid/heptane (25:17:58, by vol.) (Jeppsson & Sjoquist, 1967) and chloroform/methanol (9:1, v/v) (Pataki, 1966). We allowed the solvent front to move only 10 cm, improving both resolution and sensitivity. The appearance of t.l.c. plates was generally equal to or superior to the photograph shown in Garrick & Sloan (1977). The quality declined only for the last 20% of the residues for a given run so that subjective identifications were unambiguous.

Aqueous phases were dried and tested for arginine phenylthiohydantoin by the phenanthraquinone spot test (Yamada & Itano, 1966) and for histidine phenylthiohydantoin by the isoamyl nitrite spot test

(Sanger & Tuppy, 1951). One of us (T. W. Ryan, unpublished work) noted that histidine phenylthiohydantoin (or a derivative thereof) also consistently appeared as a blue-fluorescent spot under long-wave u.v. light when the ethyl acetate phase was run on t.l.c. The spot remained at the origin in the first solvent system described and moved slightly away from the origin in the second solvent system. When the amount of histidine phenylthiohydantoin present was low, the blue fluorescence was not observed until the plate dried overnight. The blue-fluorescent spots were stable for over 1 year.

S-(α -Dansylaminoethyl)cysteine phenylthiohydantoin was identified in the ethyl acetate phase by a yellow-fluorescent spot on filter paper (Whatman 3MM) under long-wave u.v. light. This residue was also seen on the t.l.c. plates under long-wave u.v. light, as a yellow-gold spot that moved near threonine phenylthiohydantoin in the first solvent system and near aspartic acid phenylthiohydantoin in the second solvent system.

Lysine residues that had been treated with 1-fluoro-2-nitro-4-trimethylammonio benzene iodide were identified by the yellow of the phenylthiohydantoin in the aqueous phase, which was evaluated by A_{420} . ϵ -N-Succinyl-lysine α -N-phenylthiohydantoin was detected as a spot at the origin in the first solvent system and as an arc moving near glutamic acid phenylthiohydantoin in the second solvent system.

Leucine and isoleucine phenylthiohydantoin were sometimes hard to distinguish on t.l.c. and, when this was so, were hydrolysed in 6 M-HCl for 72 h and the free acids identified with the amino acid analyser.

Results

Separation of globin chains

Fig. 1 shows the separation of seven globin chains on CM-cellulose based on a modification of the method of Clegg *et al.* (1966). For this particular experiment, reticulocytes had been incubated with L-[³H]leucine. Radioactivity (c.p.m.) is plotted in Fig. 1 instead of A_{280} , because the combination of urea in the buffers used for separation, the pH of the buffers (7.05) and the duration of the run (33 h) causes a severe rise in the baseline for A_{280} , partially obscuring the smaller or latter peaks.

Partial sequences of the $^1\alpha$ - and $^{11}\alpha$ -chains have been published (Garrick *et al.*, 1975; Chua *et al.*, 1975). Sequence analyses of the $^0\alpha$ -, $^0\beta$ -, $^1\beta$ - and $^{11}\beta$ -chains have been begun, confirming their designation as α or β .

The peak for the $^{11}\beta$ -chain is partially resolved into two components, which could reflect the heterogeneity described below for position 87. For homozygous rats only one peak is observed.

Fig. 2. Separation of fragments after CNBr cleavage. Fragments from CNBr cleavage of the $^{\text{II}}$ β -chain were separated on CM-cellulose with a modification of the method of Clegg *et al.* (1966). A sample of each fraction was mixed with 6% (w/v) trichloroacetic acid and the turbidity read at 420nm (described in the Materials and Methods section). Peak designations were assigned from amino acid-composition data.

Fig. 3. Absorbance pattern for lysine phenylthiohydantoin modified by 1-fluoro-2-nitro-4-trimethylammoniumbenzene iodide

The pattern presented is from the sequence of the $^{\text{II}}$ β -chain after blockage and cleavage with CNBr. The yellow colour was monitored at 420nm. The numbers by each peak give the positions of lysine in the intact chain.

of their increased hydrophilic properties. They could easily be identified by their A_{420} . Fig. 3 shows such a pattern. The numbers by each peak correspond to the positions of lysine in the $^{\text{II}}$ β -chain. The first peak (fraction 2) has a very high absorbance that does not correspond to any lysine residues. This deep yellow in fraction 1, 2 or 3 is typical of almost every analysis performed. We do not have a definitive explanation for this. Perhaps excess of 1-fluoro-2-nitro-4-trimethylammoniumbenzene iodide that is not covalently attached is washed out during the first few cycles.

Histidine phenylthiohydantoin, the ethyl acetate phases of which had been spotted and run on t.l.c., gave a blue fluorescence under long-wave u.v. light. The intensity of the fluorescent spots increased with time over a period of a day and were stable for at

least a year. Using aqueous phases, we confirmed the presence of histidine phenylthiohydantoin by the isoamyl nitrite spot test (Sanger & Tuppy, 1951) and found a perfect correlation between the presence of blue-fluorescent spots and positive spot tests with isoamyl nitrite. The blue fluorescence was a more sensitive indicator of histidine phenylthiohydantoin, however, than the isoamyl nitrite spot test. For example, residue 92 of the $^{\text{II}}$ β -chain was identified as histidine phenylthiohydantoin by the presence of the blue fluorescent spot in position 37 of CNBr-treated chain; the isoamyl nitrite spot test, however, was not interpretable. This histidine residue was subsequently confirmed by data for position 2 after *S. aureus* proteinase had cleaved 1-fluoro-2-nitro-4-trimethylammoniumbenzene iodide-treated peptide $^{\text{II}}$ β CB2,3

(residues 56–146) on the carboxyl side of glutamic acid- β 90 residue.

Short peptides have been difficult to sequence in a spinning-cup sequencer because they will not remain in the cup, owing to their solubility in the solvents used. We approached this problem by covalent attachment, by the carbodi-imide reaction, of the C-terminus of the peptide to the ϵ -amino groups of the lysine residues in cytochrome *c* (Garrick & Sloan, 1977). Cytochrome *c* has a blocked *N*-terminus and will not be degraded by the Edman procedure (Edman & Begg, 1967). By using this approach, we were able to determine the sequence of the C-terminus of the $^{\text{II}}\beta$ -chain.

Sequencing strategies

For all our sequence work, we used a Beckman sequencer (model 890B) and the Quadrol buffer system. Sequential degradation from the *N*-terminus was based on the work of Edman & Begg (1967). Each run was done at least twice.

Sequencing of the intact chain enabled us to establish the *N*-terminal residues, 1–55, except residues 47 and 54. Examination of the β -chains from other species encouraged us to hope for the presence of an aspartic acid–proline sequence at residues 99–100 of the rat β -chains. We therefore blocked the *N*-terminus and treated the chains with formic acid, hoping to hydrolyse between aspartic acid and proline. We did find one major sequence, which by homology to other β -chains placed the aspartic acid–proline residues at positions 99–100. We obtained residues 100–124, except residue 123.

From the amino acid composition, we knew that there were three arginine residues in the $^{\text{II}}\beta$ -chain: data from the *N*-terminus showed that one was at position 30 and another at position 40. From the cleavage between aspartic acid and proline, we could locate the third arginine residue at position 104. After treating $^{\text{II}}\beta$ -globin with succinic anhydride, we used trypsin to cleave after the arginine residues. We were fortunate that the cleavage after arginine-30 was poor, so we recovered only residues 31–34 in poor yield before losing the sequence from this cleavage. Succinylation having blocked the *N*-terminus, we observed only two major sequences, one beginning after arginine-40 and the second after arginine-104. In this way, we determined residues 41–90, excluding 89, and 105–131, excluding residue 125.

The $^{\text{II}}\beta$ -chain also contained two methionine residues, as shown by amino acid analysis. From the analyses above, one was residue 55 and the second was residue 109. We therefore blocked the *N*-terminus and cleaved after the methionine residues with CNBr. We detected two major sequences, one beginning at residue 56, the other at residue 110, obtaining residues 56–92 and 110–131.

To obtain residues 91–99 and those at the C-terminus, we cleaved with CNBr after the two methionine residues and separated fragments by ion-exchange chromatography (Fig. 3). Since we did not locate peptide CB2, we blocked the *N*-terminus of peptide CB2,3 and cleaved after the glutamic acid residues at positions 90, 101 and 121 with the *S. aureus* proteinase (Houmard & Drapeau, 1972). Cleavage after glutamic acid-101 was poor. From these cleavages, we obtained residues 91–99 and 122–133.

For peptide CB3, we used *S. aureus* proteinase to cleave after glutamic acid-121 and attached the two peptides to cytochrome *c* (Garrick & Sloan, 1977). By sequencing, we identified residues 122–146 (and confirmed 110–121).

Two sequences running together are not at all difficult to interpret (Gray, 1968). From the unique sequences of the *N*-terminus and the cleavage between aspartic acid and proline and from the overlaps detected with the other cleavage methods, we were able to sort out and place residues unambiguously. A detailed statement giving for each run the major phenylthiohydantoins assigned for each step before their alignment in individual sequences has been deposited as Supplementary Publication SUP 50084 (10 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1978) 169, 5.

Sequence of the $^{\text{II}}\beta$ -chain

The sequence of the rat $^{\text{II}}\beta$ -chain is given in Fig. 4. Each run was done at least twice, and at least one of these runs represents material from a single rat. Four individual rats were used in these studies. Several heterogeneities exist, i.e. several positions where two amino acids are found instead of only one. We have not studied these heterogeneities extensively, but, accepting the assumption of allelic variation, the linkage appears to be proline at 48, histidine at 87, serine at 89 and serine at 123 versus alanine at 58, asparagine at 87, threonine at 89 and threonine at 123. These two groups seem to represent two alleles at one locus, since we have found rats that are apparently homozygous and heterozygous. Further work needs to be done on this problem, however. There are two cysteine residues, one at position 93 and the other at 125.

Amino acid compositions

To confirm the sequence obtained, we determined the amino acid composition of the complete chain, which is given in Table 1. The single rat whose $^{\text{II}}\beta$ -chain was analysed appears to be a homozygote for proline-58, histidine-87, serine-89 and serine-123. Table 2 gives the composition of CNBr peptides

Fig. 4. Amino acid sequence of the β -chain

Residues were placed by sequencer analysis of the following: —→, N-terminus; ----→, CNBr cleavage; ----→, tryptic digestion after succinylation; —x→, cleavage between aspartic acid and proline; ---→, cleavage after glutamic acid; by carboxypeptidase A digestion: ←—, for intact chain, ←---, for cyanylated cleaved chain.

Table 1. *Amino acid composition of rat β -chain*
Content (residues/146 residues)

	24h	48h	72h	Average	Sequence
Asp	17.9	17.0	17.1	17.3	17-18
Thr	3.3	3.0	2.9	3.5*	3-5
Ser	6.0	5.0	4.5	6.7*	5-7
Glu	7.9	7.9	7.9	7.9	8
Pro	5.7	5.3	5.2	5.4	4-5
Gly	15.1	14.3	14.6	14.7	14
Ala	16.3	17.1	16.4	16.6	17
Val	12.5	14.0	14.0	13.5	14
Met	1.6	1.3	0.8	2.0*	2
Ile	2.4	3.0	3.3	3.8*	4
Leu	13.0	14.8	15.5	16.9*	17
Tyr	2.6	2.7	2.6	2.6	3
Phe	7.7	6.9	6.7	7.1	7
His	10.3	9.9	10.0	10.1	9-10
Lys	12.9	12.1	11.8	12.3	12
Arg	3.0	3.1	2.8	3.0	3

* Extrapolated to zero or infinite time.

Table 2. *Amino acid compositions of β -chain CNBr peptides CB1 and CB3*
Abbreviation: ND, not determined.

	Content					
	CB1 (residues/55 residues)			CB3 (residues/37 residues)		
	24h	72h	Average	24h	72h	Average
Asp	7.4	7.0	7.2	0.5	0.5	0.5
Thr	1.9	1.8	1.9*	0.9	0.8	0.9*
Ser	3.5	3.1	3.7*	1.0	0.8	1.1*
Glu	3.2	3.2	3.2	3.0	3.1	3.0
Pro	1.5	1.6	1.6	1.0	1.4	1.2
Gly	5.9	ND	5.9	3.0	2.9	3.0
Ala	6.1	ND	6.1	7.2	8.0	7.6
Val	5.4	5.8	6.0*	4.2	4.7	4.9*
Hsr	1.0	1.0	1.0	—	—	—
Ile	1.0	1.1	1.1*	1.5	1.7	1.8*
Leu	6.6	5.7	6.2	3.4	3.3	3.4
Tyr	1.7	1.7	1.7	0.8	0.7	0.8
Phe	1.8	2.2	2.0	1.9	1.6	1.8
His	0.9	0.9	0.9	4.1	4.1	4.1
Lys	1.9	2.5	2.2	3.0	3.3	3.2
Arg	2.0	1.9	2.0	—	—	—

* Extrapolated to zero or infinite time.

CB1 and CB3. This rat appears to have threonine at position 123. We also obtained the amino acid compositions of 11 tryptic peptides from this chain. These are presented in Table 3. One expects and observes a decreased recovery of the *N*-terminal residue for each peptide, due to derivative formation by ninhydrin. Tryptic peptides 5 and 11-14 were not recovered. There is good agreement

between the composition and the sequence for all three Tables.

Carboxypeptidase A digestion

Table 4 summarizes the results, which indicate the presence of a *C*-terminal histidine residue preceded by tyrosine for the intact β -chain. For β -chain cleaved proximal to cysteine residues, the same residues are recovered as well as a terminal histidine (residue 92) preceded by leucine and glutamic acid (91 and 90 respectively). These are preceded by serine, leucine, alanine, phenylalanine, threonine and glycine, but their order is not adequately defined by the data. It is reasonable to postulate that proline-124 prevents digestion after cleavage between it and cysteine-125.

Discussion

We have presented here the amino acid sequence of the major β -globin chain from the rat. There seems to be allelic variation for the β -chain, which involves mostly changes that do not affect the charge of the protein. For this sequence work we used four rats and did not pool their globin, but rather kept the four samples separate. About 85% of the sequence was established for one particular rat. For certain regions of the chain, the different rats were compared. Although more extensive studies need to be made of this type of variation, two different alleles appear to exist for the β -chain. Similar observations have been reported for the β -chain of rabbit haemoglobin (Garrick *et al.*, 1974). For the major α -chain of the rat, a heterogeneity exists at residue 44 (both serine and asparagine are found) that seems to be allelic (Garrick *et al.*, 1975). We used Wistar rats, which are supposed to be inbred sufficiently to permit skin transplants. All this variation in so few animals is quite remarkable.

To obtain the sequences presented, we used several technological improvements that simplify and facilitate both gathering and interpreting data. To identify cysteine residues, we treated them with dansylaziridine, which served as a fluorescent tag. Ethyl acetate phases, which contained *S*-(2-dansyl-aminoethyl)cysteine phenylthiohydantoin, gave yellow or gold fluorescent spots under long-wave u.v. light on filter paper or t.l.c. plates. Thus the identification of this previously challenging residue became straightforward and simple.

We located two cysteine residues, at positions 93 and 125. It is unusual to have cysteine at the latter position, although this residue is found at position 112 in several primates and the dog, at 113 in the grey kangaroo, at 129 in the frog and at 126 in the chicken (Dayhoff, 1972, 1976). Position 125 of the β -chain is involved in the $\alpha_1\beta_1$ contact of the haemo-

Table 3. Amino acid compositions of ¹¹β-chain tryptic peptides

		Content (residues/peptide)										
Peptide	...	T1	T2	T3	T4	T6	T7	T8	T9	T10	T15	T16
Residues	...	1-8	9-17	18-30	31-40	60-61	62-65	66	67-76	77-82	133-144	145-146
Asp		1.0	1.2	3.4					3.1	2.2		
Thr		0.9			1.0							
Ser											1.0	
Glu		1.2		1.1	1.3							
Pro				1.1	1.0							
Gly			1.7	3.4			1.1		1.4		1.1	
Ala		1.0	1.7	1.1			0.7		1.1		4.2	
Val		0.6	1.1	2.2	1.5	0.5			0.7		2.3	
Ile									0.6			
Leu		1.1	1.3	1.1	1.6				1.1	2.0	1.1	
Tyr					0.5							0.6
Phe									0.8			
His		0.9					1.0			0.5	1.1	1.1
Lys		1.0	1.0			1.0	1.0	1.0	1.0	1.0	1.0	
Arg				1.0	0.8							

Table 4. Carboxypeptidase A digests of intact rat ¹¹β-chain and of ¹¹β-chain cleaved proximal to cysteine residues
Values in parentheses are from intact chain; others are from ¹¹β-chain after cyanylation and cleavage.
Amount of starting material: 40nmol.

		Residue (nmol)			
Time (h)	...	0.25	1	4	21
Thr	Nil	Nil	Nil	9.5	
Ser	Nil	Nil	7.3	26.5	
Glu	Nil	4.4	13.4	28.8	
Gly	Nil	Nil	Nil	17.6	
Ala	Nil	Nil	7.2	20.7	
Leu	4.0	7.0	18.6	40	
Tyr	3.4 (0.5)	8.6 (7.5)	13.0 (11.3)	23.9 (22.0)	
Phe	Nil	Nil	5.0	21.0	
His	12.9 (5.5)	15.8 (10.1)	27.8 (16.3)	53.7 (25.8)	

globin tetramer (Perutz *et al.*, 1968). Whether cysteine in this position affects the functional properties of the molecule remains to be seen. There are three cysteine residues in the major α-chain, ¹α (Chua *et al.*, 1975; Garrick *et al.*, 1975). Thus there are 10 cysteine residues per tetramer for some rat haemoglobins. This exceptionally high content could contribute to the instability of rat haemoglobins *in vitro* (Chua & Carrell, 1974).

We also observed that histidine phenylthiohydantoins from the ethyl acetate phase gave a blue-fluorescent spot under long-wavelength u.v. light after t.l.c. This observation was confirmed by the isoamyl nitrite spot test of Sanger & Tuppy (1951). Identification by blue fluorescence is much easier than drying down and spotting each aqueous phase, however, and is actually more sensitive than the

spot test. Inglis & Nicholls (1974) have independently noted the fluorescence of histidine phenylthiohydantoin and suggested that this property would aid identification.

Another approach that we often used was to recover several sequences at once instead of purifying fragments after cleavage (Gray, 1968). We deviated from this approach in one case: for cleavage after glutamic acid, we found it helpful for interpreting sequence data first to purify fragments from CNBr cleavage. Blockage of the *N*-terminus with 1-fluoro-2-nitro-4-trimethylammoniumbenzene iodide also simplified multiple sequencing. We usually obtained no more than two major residues for each position in this way. We sorted out the sequences by overlap with the *N*-terminal sequence and with other cleavages and by relying on homology to other β-chains. Homology was not needed, however. Amino acid compositions of intact chain, peptides CB1 and CB3 and 11 isolated tryptic peptides confirm the sequence presented here. Digestion of the *C*-termini of intact chains and globin cleaved before cysteine residues also confirmed regions of the sequence.

We thank Dr. Michael Apicella and the Core Laboratory of the Department of Medicine for making available the Beckman 890B sequencer and 120C amino acid analyser, Dr. Roy Oliver, President of Pierce Chemical Co., for suggesting the use of dansylaziridine, Dr. Vijay S. Sharma, Department of Medicine, University of California at San Diego, for providing us with the single-cleavage Quadrol program, and Dr. Robin M. Bannerman for his help and patience during this project. This work was supported by National Institutes of Health grants RR 05400 (to L. M. G.), AMHL 14923 (to M. D. G.) and RR 07066.

References

- Ambler, R. P. (1972) *Methods Enzymol.* **25B**, 262-272
- Bricker, J. & Garrick, M. D. (1974) *Biochim. Biophys. Acta* **351**, 437-441
- Chua, C. G. & Carrell, R. W. (1974) *Biochim. Biophys. Acta* **365**, 328-334
- Chua, C. G., Carrell, R. W. & Howard, B. H. (1975) *Biochem. J.* **149**, 259-269
- Clegg, J. B., Naughton, M. A. & Weatherall, D. J. (1966) *J. Mol. Biol.* **19**, 91-108
- Dayhoff, M. O. (ed.) (1972) *Atlas of Protein Sequence and Structure*, pp. D64-66, D77-78, National Biomedical Research Foundation, Washington
- Dayhoff, M. O. (ed.) (1976) *Atlas of Protein Sequence and Structure*, Suppl. 2, p. 219, National Biomedical Research Foundation, Washington
- Degani, Y. & Patchornik, A. (1971) *J. Org. Chem.* **36**, 2727-2728
- Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80-91
- Garrick, L. M., Sharma, V. S., McDonald, M. J. & Ranney, H. M. (1975) *Biochem. J.* **149**, 245-258
- Garrick, M. D. & Sloan, R. L. (1977) *Prep. Biochem.* **7**, 111-128
- Garrick, M. D., Hafner, R., Bricker, J. & Garrick, L. M. (1974) *Ann. N.Y. Acad. Sci.* **241**, 436-438
- Gray, W. R. (1968) *Nature (London)* **220**, 1300-1304
- Hermanson, M. A., Ericsson, L. H. & Walsh, K. A. (1970) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **29**, 728
- Hermanson, M. A., Ericsson, L. H., Titani, K., Neurath, H. & Walsh, K. A. (1972) *Biochemistry* **11**, 4493-4502
- Houmard, J. & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3506-3509
- Inglis, A. S. & Nicholls, P. W. (1974) *J. Chromatogr.* **97**, 289-292
- Jacobson, G. R., Schaffer, M. H., Stark, G. R. & Vanaman, T. C. (1973) *J. Biol. Chem.* **248**, 6583-6591
- Jeppsson, J. & Sjoquist, J. (1967) *Anal. Biochem.* **18**, 264-269
- Pataki, G. (1966) *Techniques of Thin-Layer Chromatography in Amino Acid and Peptide Chemistry*, pp. 157-160, Ann Arbor-Humphrey Science Publishers, Ann Arbor
- Perutz, M. F., Muirhead, H., Cox, J. M. & Goaman, L. C. G. (1968) *Nature (London)* **219**, 131-139
- Sanger, F. & Tuppy, H. (1951) *Biochem. J.* **49**, 463-481
- Scouten, W. H., Lubcher, R. & Baughman, W. (1974) *Biochim. Biophys. Acta* **336**, 421-426
- Sutton, D. A., Drewes, S. E. & Welz, U. (1972) *Biochem. J.* **130**, 589-595
- Yamada, S. & Itano, H. A. (1966) *Biochim. Biophys. Acta* **130**, 438-440