Specific Cyanylation and Cleavage at Cysteine-104 of Human Haemoglobin α -Chain

A NOVEL APPROACH TO THE PROBLEM OF THE α -CHAIN TRYPTIC CORE IN THE STUDY OF HAEMOGLOBIN VARIANTS BY 'FINGERPRINTING' METHODS

By RODERICK CASEY and ANDREW LANG

M.R.C. Abnormal Haemoglobin Unit, Department of Biochemistry, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QR, U.K.

(Received 22 August 1974)

1. A new approach to the analysis, by 'fingerprinting', of the tryptic core region of human haemoglobin α -chain is described. 2. The α -chain is cyanylated at its single cysteine residue (α 104) and then split, by exposure to mild alkali, at the *N*-peptide bond of the resulting β -thiocyanoalanine residue. 3. The two cleavage fragments, α 1–103 and α 104–141, are separated by gel filtration, and the fragment α 104–141, which contains all the residues of the α -chain tryptic core, is digested with pepsin. 4. Preparative 'fingerprints' of these peptic peptides yield eight major peptides, which provide complete sequence information for the whole region α 104–141. 5. The utility of the method is demonstrated by repeating the determination of the substitution in haemoglobin Hopkins-2, a known α -chain core variant in which histidine- α 112 (G19) is replaced by an aspartic acid residue.

The analysis of human haemoglobin variants by the classical 'fingerprinting' method of Ingram (1958) was limited, in the early stages, both by the lack of a simple method for separating the α - and β -chains and by the large amount of insoluble material remaining after tryptic digestion (the tryptic 'core'; Hunt & Ingram, 1958). The difficulty in globin chain separation was overcome by the application of a specific cation-exchange chromatography method (Clegg *et al.*, 1966) which, combined with *S*- β -aminoethylation and tryptic digestion of the isolated chains (Jones, 1964), meant that all of the tryptic peptides from the β - (and δ -) chains could be examined by 'fingerprinting'. This technique, when applied to the α -chain, however, results in the α -chain tryptic core peptides (from residues α 105–139) remaining as a poorly defined streak at the point of application to the 'fingerprint'.

One approach to overcome this difficulty has been the secondary digestion of the tryptic core from the unmodified, or aminoethylated, α -chain with α -chymotrypsin before 'fingerprint' analysis of the chymotryptic fragments (Baglioni, 1962; Carrell & Irvine, 1968). It is, however, necessary to wash the core before chymotryptic digestion to remove any traces of the soluble tryptic peptides; this, in our hands, leads to serious losses of material.

It seemed desirable to us to obtain intact a pure preparation of the region corresponding to the α -chain core, and that one way to do this would be to split the α -chain at its single cysteine residue (cysteine- α 104). The introduction of 2-nitro-5-thiocyanobenzoic acid by Degani *et al.* (1970) has led

to subsequent methods for the *S*-cyanylation and consequent cleavage of proteins, in high yields and under mild conditions, at the *N*-peptide bond of cysteine residues (Jacobson *et al.*, 1973). We therefore applied the *S*-cyanylation and cleavage procedure to human haemoglobin α -chain, which provided us with a pure preparation of fragment α 104–141 (containing all of the residues which constitute the α -chain core). Digestion of this fragment with pepsin, combined with subsequent 'fingerprinting' of the peptic peptides, provided complete sequence information for the α -chain core and enabled us readily to re-determine the structure of haemoglobin Hopkins-2, a known α -chain variant (Ostertag *et al.*, 1972).

Experimental

Materials

Urea (crystallized, extra pure) and ninhydrin (analytical grade) were obtained from E. Merck, Darmstadt, Germany. Dithiothreitol (A grade) was obtained from Calbiochem, London W.1, U.K. Trizma base (Tris) (reagent grade) and 5,5'-dithiobis-(2-nitrobenzoic acid) were products of Sigma (London) Chemical Co., Kingston-upon-Thames, Surrey, U.K. CNBr was obtained from Eastman Kodak Co., Rochester, N.Y., U.S.A., ethyleneimine from Koch-Light Laboratories Ltd., Colnbrook, Bucks, U.K., and fluorecamine from Roche Diagnostics, Nutley, N.J., U.S.A. Dansyl chloride, dansyl-amino acids and guanidinium chloride (AristaR

grade) were obtained from BDH Chemicals Ltd., Poole, Dorset, U.K. Sephadex G-25 (coarse grade), G-75 and A-50 (medium grade), and Blue Dextran 2000 marker were obtained from Pharmacia (G.B.) Ltd., London W.4, U.K.; CM-cellulose (Whatman CM23) was obtained from Reeve Angel and Co. Ltd., London E.C.4, U.K. L-[U- 14 C]Leucine (210 mCi/mmol) and Na 14 CN (59.3 mCi/mmol) were obtained from The Radiochemical Centre, Amersham, Bucks., U.K.

Trypsin treated with 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one (batch TRTPCK 335768), α -chymotrypsin (batch DC1 0BF) and pepsin (batch PM 9FD) were obtained from Worthington Chemical Corp., Freehold, N.J., U.S.A. The modified trypsin was dissolved in 1 mM-HCl (2 mg of protein/ml of 1 mM-HCl) and stored in small portions at -20°C until required; α -chymotrypsin and pepsin solutions (2 mg/ml of water) were freshly prepared immediately before use.

Polyamide thin-layer plates for the separation of dansyl-amino acids were from the Cheng-Chin Trading Co. Ltd., Taipei, Taiwan.

Other reagents not described above were of analytical grade if available; isoamyl alcohol was BDH AnalaR grade, which consists principally of 3-methylbutan-1-ol with a smaller proportion of 2-methylbutan-1-ol. Pyridine was redistilled from solid ninhydrin before use (Hill & Delaney, 1967); ethyl acetate and light petroleum (b.p. $100-120^{\circ}\text{C}$), used for the recrystallization of 2-nitro-5-thiocyanobenzoic acid, were redistilled and stored over molecular sieves type 4A (from BDH). Acetone used for making up stock solutions of fluorescamine and dansyl chloride was dried over anhydrous Na $_2$ SO $_4$ and stored over Linde molecular sieves type 3A and 4A.

*Preparation of α -chains from Hb*A and Hb Hopkins-2*

Blood was obtained by venepuncture from normal individuals (Hb A homozygotes) and from an individual heterozygous for the α -chain variant Hb Hopkins-2. Haemolysates of saline-washed erythrocytes were prepared in the usual way (Lehmann & Huntsman, 1966) and Hb Hopkins-2 was purified from the appropriate haemolysate by chromatography on DEAE-Sephadex A-50 (Huisman & Dozy, 1965). Globin was prepared from normal haemolysates and from isolated Hb Hopkins-2 by acid-acetone precipitation (Winterhalter & Huehns, 1964) and separated into its component chains by chromatography on a column (23 cm $\times$ 0.9 cm) of CM-cellulose, by using buffers containing 8 M-urea as described by Clegg *et al.* (1966), but with

the difference that 2-mercaptoethanol was replaced by 0.3 mM-dithiothreitol (Kilmartin & Rossi-Bernardi, 1971). The flow rate was 60 ml/h and 10 ml fractions were collected.

α -Chains were occasionally S-cyanylated immediately after CM-cellulose chromatography, in which case the cyanylation reagent (see below) was added at this stage; generally, however, the isolated α -chains were freed of urea, salts and excess of thiol by passage through a column (100 cm $\times$ 2.5 cm) of Sephadex G-25 (coarse grade) equilibrated with aq. 0.5% (v/v) formic acid and recovered by freeze-drying. The α -chains from Hb Hopkins-2 were subjected to a second cycle of CM-cellulose chromatography and desalting to remove any traces of residual normal α -chains.

Uniform labelling of α -chains with [U- 14 C]leucine

For later use in obtaining information as to the efficiency of the cyanylation and cleavage process, α -chains from normal individuals were uniformly labelled with ^{14}C at leucine residues by incubation of a reticulocyte-rich fraction of fresh blood with [U- 14 C]leucine as described previously (Lang *et al.*, 1972). After a 3–5 h incorporation period the cells were washed free of excess of label with ice-cold iso-osmotic (0.9%) NaCl and lysed by the addition of 5 vol. of water. After removal of erythrocyte debris by centrifugation, globin was prepared and [^{14}C]leucine-labelled α -chains were obtained as described above. The specific radioactivity of the α -chains was 0.95 $\mu\text{Ci}/\text{mmol}$ of protein.

Synthesis of 2-nitro-5-thiocyanobenzoic acid and 2-nitro-5-thio[^{14}C]cyanobenzoic acid

The cyanylation reagent, 2-nitro-5-thiocyanobenzoic acid, was prepared as the free acid from NaCN and 5,5'-dithiobis-(2-nitrobenzoic acid) (Degani & Patchornik, 1971); the uncorrected melting point of the recrystallized product was $160-161^{\circ}\text{C}$. 2-Nitro-5-thio[^{14}C]cyanobenzoic acid was synthesized in exactly the same manner, by using Na 14 CN (0.81 mmol/mol of NaCN), to a specific radioactivity of 32 mCi/mol.

Radioactivity counting procedure

Samples (10–100 μl in 0.5 M-NH $_3$ or 0.2 M-acetic acid) were counted for radioactivity in 15 ml of scintillator with the composition Triton X-100, 500 ml; toluene, 1 litre; PPO, 5 g (Hunt *et al.*, 1968) with an ICN Tracerlab Corumatic 200 liquid-scintillation spectrometer. Counting efficiency for ^{14}C was 86.9%.

* Abbreviations: Hb, haemoglobin; PPO, 2,5-diphenyloxazole; Tp, tryptic peptide.

S-Cyanylation and cleavage of isolated α -chains

S-Cyanylation and subsequent cyclization and cleavage of α -chains were carried out as described by Jacobson *et al.* (1973). Isolated α -chains (50–130 mg; 3.3–8.6 μ mol) were reduced for 2 h at 37°C in 10 ml of 0.01 M-dithiothreitol–0.2 M-Tris–6 M-guanidinium chloride, final pH 8.2 (adjusted with acetic acid); subsequent cyanylation was achieved by reaction with a fivefold excess of 2-nitro-5-thiocyanobenzoic acid or 2-nitro-5-thio[14 C]cyanobenzoic acid over total thiol for 15 min at room temperature (22°C), and the reaction was stopped by the addition of an equal volume of acetic acid. The S-cyanylated α -chains were freed of excess of reagents by passage through a column (100 cm $\times$ 2.5 cm) of Sephadex G-25 (coarse grade) equilibrated with aq. 50% (v/v) acetic acid and recovered by freeze-drying. Any traces of acetic acid were removed by a second freeze-drying from water. The S-cyanylated α -chains were split by incubation for 16 h at 37°C in 4–6 ml of 0.1 M-sodium borate–6 M-guanidinium chloride, final pH 9.0.

S-Cyanylation immediately after separation of α - and β -chains

If only a small amount of globin was available, it was considered advisable to cyanylate immediately after chromatographic separation of α - and β -chains, thus minimizing losses through transfer. Sufficient solid Tris was added to the solution of α -chains, in 8 M-urea–0.3 mM-dithiothreitol–0.03 M- Na_2HPO_4 , final pH 6.7, to make the final Tris concentration 0.2 M. The pH was adjusted to 8.2 with acetic acid and the appropriate amount of 2-nitro-5-thio[14 C]cyanobenzoic acid added; after 15 min reaction time an equal volume of acetic acid was added and the S-cyanylated α -chains were dialysed exhaustively against 50% acetic acid. The subsequent cleavage procedure was exactly as described above.

Separation of the cleavage fragments

The cleavage incubation mixture was cooled to room temperature and adjusted to pH 4 with acetic acid. Excess of borate was removed by centrifugation (3200 g, 10 min) and the supernatant was applied to a column (180 cm $\times$ 2.5 cm) of Sephadex G-75 (total bed volume 880 ml) equilibrated with 0.2 M-acetic acid at 4°C. The column was eluted with 0.2 M-acetic acid at 4°C at a flow rate of 20–28 ml/h and fractions (5–7 ml) were collected. The absorbance at 275 nm of a 3 ml portion of each fraction was measured by using a Zeiss recording spectrophotometer model DMR 21, and the radioactivity of alternate fractions was determined as described above; appropriate fractions were pooled and freeze-dried.

Estimation of the degree of cleavage

The degree of cleavage was calculated from the elution profile from Sephadex G-75 of the fragments from [14 C]leucine-labelled α -chains. The areas under the three peaks of radioactivity corresponding to unsplit α -chains, fragment α 1–103 and fragment α 104–141 (see below) were measured and then divided by the known leucine content of each polypeptide. The values for radioactivity thus derived were therefore a measure of the relative amount of each polypeptide species present.

Enzymic hydrolysis of cleavage products

(i) *Trypsin*. The residual unsplit α -chains and the two cleavage fragments α 1–103 and α 104–141 were dissolved in water (0.7 μ mol/ml of water). Sufficient modified trypsin solution (in 1 mM-HCl) was added to give a ratio 0.3 mg of enzyme/ μ mol of substrate, followed by enough 0.5 M- NH_4HCO_3 to give a final concentration of 0.08 M. Incubation at 37°C for 2 h was sufficient to clarify the initially turbid suspensions from the cleavage fragments, whereas the digest of the intact α -chain was still turbid after 2 h. Insoluble material was removed from the latter by centrifugation (3200 g, 10 min) and hydrolysis terminated in all three cases by freeze-drying.

(ii) *α -Chymotrypsin*. The fragment α 104–141 was hydrolysed with α -chymotrypsin exactly as described for trypsin except that the enzyme solution was in water, not 1 mM-HCl. The final solution was clear.

(iii) *Pepsin*. The fragment α 104–141 was dissolved in 0.1 M-HCl (0.7 μ mol/ml of 0.1 M-HCl). Then 0.3 mg of pepsin/ μ mol of substrate was added (in water); the mixture was incubated for 2 h at 37°C, whereupon the clear solution was freeze-dried.

Peptide 'maps' ('fingerprints')

The freeze-dried digests were dissolved in 0.5 M- NH_3 (0.5 μ mol/100 μ l) and applied, in 10 μ l portions, as a 1 cm-diameter spot, 15 cm (8 cm for peptic peptides) from the end and 4 cm from the long edge of a sheet (34 cm $\times$ 57 cm) of Whatman 3MM paper; about 0.5 μ mol was applied for preparative and about 0.1 μ mol for diagnostic 'maps'. The sample was dried in a current of cool air and the spot sharpened to a band at the origin, by careful application of buffer, before electrophoresis in the long dimension of the paper. Electrophoresis of tryptic digests was performed nominally at pH 6.4 (pyridine–acetic acid–water, 25:1:224, by vol.) at 60 V/cm for 1 h as described by Romero-Herrera & Lehmann (1974), under which conditions the pH of buffer on the paper is actually 5.5, owing to the extraction of pyridine from the buffer by toluene

coolant. Chymotryptic and peptic peptides were separated at (nominally) pH 3.5 in the buffer pyridine-acetic acid-water (1:10:89, by vol.) (60 V/cm; 1 h and 1.5 h respectively).

The peptides were then separated in the second dimension by ascending chromatography in the upper phase of the solvent system pyridine-isoamyl alcohol-water (6:6:7, by vol.) (Baglioni, 1961) for 20 h at room temperature. Where necessary, some peptic peptides were stitched to a fresh sheet of paper and further purified by electrophoresis at pH 6.4 (60 V/cm, 1.5 h).

Location of peptides on 'fingerprints'

Peptides were detected on diagnostic 'fingerprints' by staining at 60°C with 0.2% ninhydrin in acetone [containing 1% (v/v) pyridine] and ninhydrin-negative peptides, containing *N*-terminal 2-iminothiazolidine-4-carboxylic acid, were located by using the chlorine-*o*-toluidine reagent of Reindel & Hoppé (1954). Specific amino acids were detected as follows: (i) arginine by the phenanthrenequinone reaction (Yamada & Itano, 1966); (ii) histidine by the Pauly reaction as described by Smith (1969); (iii) tyrosine as described by Jepson & Smith (1953).

Preparative-scale 'fingerprints' were stained with fluorescamine (Udenfriend *et al.*, 1972; Weigle *et al.*, 1972) or with 0.02% (w/v) ninhydrin in acetone at room temperature. For detection with fluorescamine, a stock solution of the reagent in dry acetone (1 mg/100 ml) was diluted with 50 vol. of 1% (v/v) pyridine in acetone and the paper immediately dipped and allowed to dry at room temperature. The fluorescent derivatives of all peptides with a free primary amino group were visible on examination of the 'fingerprint' under a long-wavelength u.v. lamp.

Amino acid-composition studies

Peptides were eluted from preparative 'fingerprints' or from electrophoretograms with 0.5 M-NH₃, dried and hydrolysed, in sealed evacuated tubes at 108°C for 18 h, with 0.5 ml of constant-boiling HCl containing 0.1% (w/v) phenol (Sanger & Thompson, 1963). A small amount of the isolated fragment α 104-141 was also hydrolysed in a similar manner. When methionine-containing peptides which had been eluted from paper were hydrolysed, 0.01% (w/v) dithiothreitol was added to the acid to protect against oxidative losses. After removal of the excess of HCl *in vacuo*, the amino acid compositions of the dried hydrolysates were determined by using a Locarte amino acid analyser.

Determination of N-terminal residues

The dansyl chloride method of Gray & Hartley (1963) was used to identify the *N*-terminal amino acid of certain peptides. The peptide (1-10 nmol) was dissolved in 20 μ l of 0.2 M-triethylamine and dried *in vacuo* over P₂O₅ to remove traces of NH₃. The dry material was redissolved in 10 μ l of 0.2 M-triethylamine-acetic acid, pH 9.5; 10 μ l of dansyl chloride (2.5 mg/ml in dry acetone) was added and the mixture incubated for 1 h at 37°C. After drying *in vacuo*, dansyl-peptides were hydrolysed with 6 M-HCl in sealed tubes for 4 h at 108°C (Gros & Labouesse, 1969). Excess of HCl was removed *in vacuo* over NaOH pellets. The dried hydrolysates were dissolved in 1-2 μ l of aq. 50% (v/v) pyridine and applied to polyamide thin-layer plates (7.5 cm $\times$ 7.5 cm) (Woods & Wang, 1967). Identification of the *N*-terminal dansyl-amino acid was made by two-dimensional chromatography against marker dansyl-amino acids in the solvent systems described by Hartley (1970).

Results

Separation and identification of the cleavage fragments and estimation of the degree of splitting

Fig. 1(a) shows the elution profile from Sephadex G-75 of the α -chain cleavage fragments obtained by using 2-nitro-5-thio[¹⁴C]cyanobenzoic acid, which introduces an atom of ¹⁴C into each *N*-terminal iminothiazolidine ring of the fragment α 104-141, according to Fig. 2. The distribution of the radioactivity in Fig. 1(a) shows that all the ¹⁴C radioactivity is eluted with the material designated as peak III. The three peaks were identified from 'fingerprints' of their tryptic hydrolysates. The soluble tryptic peptides from peak I gave a 'fingerprint' characteristic of unmodified α -chain, and its elution position from the Sephadex G-75 column was consistent with it being an α -chain monomer. The 'fingerprint' of peak II was essentially the same except that peptide α TpXIV, the α -chain C-terminal tryptic peptide (see Fig. 3), was absent and an additional peptide, of amino acid composition Ser 0.9, Leu 1.9, His 1.0 (α 100-103), was present; this is consistent with peak II representing fragment α 1-103, the larger cleavage product (Fig. 2). The 'map' from peak III showed two peptides, one of which was α TpXIV in good yield; the other was, from its amino acid composition (Table 1), fragment α 104-139 in low yield (11% relative to peptide α TpXIV). Peak III was probably therefore fragment α 104-141; this was confirmed by amino acid analysis of the whole fragment (Table 1).

Thus, although the unsplit α -chain and fragment α 1-103 overlap in their elution positions after gel

Fig. 1. Sephadex G-75 elution profile of α -chain cleavage fragments

(a) Pattern obtained by using radioactive cyanylation reagent and unlabelled α -chain. (b) Pattern obtained by using $[\text{U-}^{14}\text{C}]$ leucine-labelled α -chain and unlabelled cyanylation reagent. In the interests of clarity, certain non-crucial points have been omitted. Conditions of chromatography were described in the Experimental section. Fractions were pooled as indicated by the bars; peak I = residual α -chain; peak II = $\alpha 1-103$; peak III = $\alpha 104-141$. $\circ$, E_{275} ; $\blacktriangle$, ^{14}C radioactivity (d.p.s.).

filtration on Sephadex G-75, fragment $\alpha 104-141$ can be obtained in pure form.

The efficiency of cleavage was estimated by using $[\text{U-}^{14}\text{C}]$ leucine-labelled α -chain and unlabelled reagent. The Sephadex G-75 elution profile (Fig. 1b) showed that most of the radioactivity is in the cleavage products, and the average yield of splitting was estimated, as described in the Experimental section, as 85–95%; the apparent disparity between E_{275} and ^{14}C radioactivity (d.p.s.) in Fig. 1(b) reflects the different tyrosine and tryptophan contents of the intact α -chain and the two fragments.

Cyanylation immediately after separation of α -chains from β -chains

It was possible to cyanylate α -chains efficiently in the buffer containing 8M-urea directly after the separation of α - and β -chains as described in the Experimental section. Dithiothreitol was used as the reducing agent during chromatographic separation of α - and β -chains, because a much lower molar excess is required, compared with 2-mercaptoethanol, to maintain cysteine thiol groups in the reduced state (Cleland, 1964) and subsequently less 2-nitro-5-thio $[\text{U-}^{14}\text{C}]$ cyanobenzoic acid was required for cyanylation. In fact, conditions pertaining immediately after such separation of α - from β -chains closely approximate to those described by Degani & Patchornik (1974) as being optimal for cyanylation. The use of 2-nitro-5-thio $[\text{U-}^{14}\text{C}]$ cyanobenzoic acid meant that α -chain from as little as 20 mg of globin could be taken through the entire procedure and the fragment $\alpha 104-141$ could be readily detected after gel filtration on Sephadex G-75.

Further sequence analysis of the fragment $\alpha 104-141$

Having obtained the pure fragment $\alpha 104-141$, it was necessary to decide how best to obtain sequence information from it when dealing with a variant haemoglobin α -chain. The *N*-terminal 2-iminothiazolidine ring, which the fragment possesses, precludes sequence analysis by Edman degradation, presumably because it is unreactive to phenyl isothiocyanate (Jacobson *et al.*, 1973); although it is possible that mass-spectrometric sequence analysis of the permethylated peptide could provide limited information, it will not at present provide complete sequence information for a 38-residue peptide (see, for instance, Morris *et al.*, 1971).

It was therefore clear that the fragment $\alpha 104-141$ would require further splitting to obtain detailed sequence information. Since digestion with trypsin (see above) was unlikely to be informative, hydrolysis with α -chymotrypsin and pepsin was examined.

(1) *Further fragmentation with α -chymotrypsin.* The fragment $\alpha 104-141$ was initially insoluble in 0.08M- NH_4HCO_3 , but went into solution after 5 min treatment with α -chymotrypsin. The 'fingerprint' of the peptides obtained after 2 h digestion (Fig. 4) showed, on staining with ninhydrin, five distinct peptides plus a poorly defined streak. A further peptide became apparent after staining with the chlorine-*o*-tolidine reagent; the lack of reaction of this peptide with ninhydrin suggested that it possessed *N*-terminal 2-iminothiazolidine-4-carboxylic acid. The amino acid compositions of the seven peptides (one of them free arginine) allowed each to be unambiguously assigned to a specific region within

Fig. 2. Diagrammatic representation of the cyanylation and cleavage of α -chain with 2-nitro-5-thio[^{14}C]cyanobenzoic acid (based on Degani & Patchornik, 1974)

The numbers above the sequence refer to the amino acid residue positions within the α -chain, $\alpha 104$ being the sole cysteine residue. * represents ^{14}C .

Fig. 3. Amino acid sequence of the tryptic core region of human haemoglobin α -chain

The arrows represent the bonds hydrolysed by: T, trypsin, C, α -chymotrypsin, and P, pepsin, under the conditions outlined in the text.

Table 1. Amino acid compositions of the fragment α 104–141 and the tryptic peptides α 104–139 and α 140–141

For details see the Experimental section. 2-Iminothiazolidine-4-carboxylic acid, the acid-hydrolysis product from the *N*-terminal iminothiazolidinecarboxylate residue, gives very little colour on reaction with ninhydrin (Degani & Patchornik, 1974) and cannot be detected by routine amino acid analysis of α 104–141.

Fragment ... Amino acid	Content (residues/molecule)		
	α 104–141	α 104–139	α 140–141
Asp	1.2 (1)	0.9 (1)	
Thr	3.8 (4)	4.0 (4)	
Ser	3.8 (4)	3.9 (4)	
Glu	1.2 (1)	1.3 (1)	
Pro	2.1 (2)	1.6 (2)	
Ala	6.1 (6)	6.2 (6)	
Val	3.9 (4)	4.1 (4)	
Leu	7.2 (7)	6.6 (7)	
Tyr	0.9 (1)		1.0 (1)
Phe	1.8 (2)	1.9 (2)	
His	2.0 (2)	1.9 (2)	
Lys	2.0 (2)	1.8 (2)	
Arg	1.0 (1)		1.0 (1)

the sequence α 104–141, as shown in Figs. 3 and 4. [The amino acid compositions and points of chymotryptic hydrolysis were, in fact, essentially those reported by Carrell & Irvine (1968) and Baglioni (1962) except for the sequence around position cysteine- α 104.] A serious disadvantage in using α -chymotrypsin, however, is the very variable and generally poor recovery of peptide α 129–136, the peptide that streaks in Fig. 4.

(II) *Further fragmentation with pepsin.* The ‘fingerprint’ of the peptic peptides from fragment α 104–141 showed 13 spots on staining with ninhydrin, and a further spot could be located with chlorine-*o*-tolidine (Fig. 5). Amino acid-composition studies (Table 2) indicated that 11 of these spots contained pure peptide; one spot contained fragments α 117–122 and α 118–125, which could be separated by electrophoresis at pH 6.4; another spot contained residues α 130–131 plus α 132–133, which could be separated by extending the time of chromatography; finally, there was a small quantity of free leucine (α 129) and free phenylalanine (α 117) as demonstrated by amino acid composition and dansylation. With two exceptions, the amino acid composition alone

Fig. 4. ‘Fingerprint’ of the peptides obtained by digestion of fragment α 104–141 with α -chymotrypsin

The conditions used for ‘fingerprinting’ were as described in the Experimental section. Electrophoresis was carried out at pH 3.5. Numbers refer to the sequence that each peptide represents (see Fig. 3). Peptide α 104–109 possesses an *N*-terminal 2-iminothiazolidine-4-carboxylic acid residue and was detected by using chlorine-*o*-tolidine. * = origin.

Fig. 5. 'Fingerprint' of the peptic peptides of fragment $\alpha 104-141$

The conditions of 'fingerprinting' were as described in the Experimental section; electrophoresis was carried out at pH 3.5. Staining was carried out with 0.2% (w/v) ninhydrin in acetone followed by chlorine-*o*-toluidine. The major peptides are denoted by the cross-hatching, and the numbering refers to the sequence that each peptide represents. Peptide $\alpha 134-136$ gave a brown colour on staining with ninhydrin, indicative of *N*-terminal threonine, whereas peptide $\alpha 118-125$ gave a yellow colour which presumably arises from the presence of an *N*-terminal Thr-Pro sequence. Peptide $\alpha 104-106$ (containing an *N*-terminal 2-iminothiazolidine-4-carboxylic acid residue) was not detected with ninhydrin but could be located by using chlorine-*o*-toluidine and contained ^{14}C radioactivity when 2-nitro-5-thio[^{14}C]cyanobenzoic acid had been used for cyanylation. The arrow indicates the peptide missing from the corresponding 'fingerprint' of Hb Hopkins-2; the broken lines denote the positions of the three new peptides found in this variant. * = origin.

allowed assignment of all the peptides to specific sequences in the region $\alpha 104-141$. The ambiguities $\alpha 123-125/129-131$ (Ser, Ala, Leu) and $\alpha 107-109/134-136$ (Thr, Val, Leu) were readily resolved by dansylation. Eight of the peptides, indicated in Fig. 5 by cross-hatching, represented the whole of fragment $\alpha 104-141$ and could be obtained uncontaminated directly from the 'fingerprint'. The remainder represent further splits of these eight major peptides, which have been used to advantage in this laboratory to provide confirmatory information in the determination of the structure of variant α -chains.

It was concluded that pepsin was rather more useful than α -chymotrypsin for the production of small fragments from sequence $\alpha 104-141$, which could then be used in the analysis of a variant.

Amino acid substitution of Hb Hopkins-2

The α -chain from purified Hb Hopkins-2 was eluted earlier than normal α -chains during CM-

cellulose chromatography, in a position that suggested an increase of one net negative charge per α -chain. The 'fingerprint' of the soluble tryptic peptides from the unmodified α -chain, however, was identical with that from normal α -chains, indicating that the substitution was in the α -chain core.

The 'fingerprint' of the peptic peptides from Hb Hopkins-2 $\alpha 104-141$ clearly showed the absence of residues $\alpha 110-116$ and the presence of three additional peptides (Fig. 5). The relative recoveries and amino acid compositions of these latter (Table 3) made it clear first that histidine- $\alpha 112$ had been replaced by a residue of aspartic acid and that the variant was Hb Hopkins-2 (Ostertag *et al.*, 1972) and, secondly, that the Leu-Pro bond $\alpha 114-115$ had been almost quantitatively hydrolysed by pepsin. Pepsin has previously been reported to split at bonds involving proline (see, e.g., Hill, 1965); the extensive hydrolysis in this instance must somehow result from the His-Asp replacement, since it was not observed to take place when fragment $\alpha 104-141$ from normal α -chains was used.

Table 2. Amino acid compositions of the peptic peptides from the fragment α 104-141

For details see the Experimental section. The recoveries are expressed relative to those of peptides α 107-109 and α 126-128, which were obtained in the highest yield. The absolute yield of peptide α 107-109, based on the amount of starting globin, was 23%. ITC, iminothiazolidinecarboxylic acid residue.

Peptide ... Amino acid	Content (residues/molecule)																
	104-106*	107-109	110-116	117-122	117-125	118-125	117+129	123-125	126-128	129-131	129-133	130-131	132-133	134-136	137-141		
Asp									0.9								
Thr		0.9															
Ser										1.0	1.7	0.9	0.8	1.0	0.9		
Glu			1.0														
Pro			0.9														
Ala			2.9														
Val																	
ITC																	
Leu	(1)†	1.0															
Tyr	(2)†		1.0														
Phe																	
His			1.0														
Lys																	
Arg																	
Recovery(%)	69	100	79	10	8	53	11	5, 4	41	100	39	33	9	14	49	76	

* This peptide could be located with chlorine-*o*-tolidine, but not with ninhydrin. It came from the only area of the 'fingerprint' to contain radioactivity and was therefore assumed to possess an *N*-terminal [^{14}C]iminothiazolidinecarboxylic acid residue.
† The specific radioactivity (^{14}C d.p.m./nmol of leucine) in this peptide was 22, whereas that of the parent peptide α 104-141 was 55 d.p.m. of ^{14}C /nmol of peptide. It follows therefore that this peptide is 2-iminothiazolidine-4-carboxyl-L-Leu rather than 2-iminothiazolidine-4-carboxyl-L-Leu.

Table 3. *Amino acid composition of the three new peptides from Hb Hopkins-2*

For details see text. 'Recovery (%)' is relative to that of peptide α 110-113.

Amino acid	Content (residues/molecule)			Expected for peptide α^A 110-116
	Peptide 110-113	Peptide 114-116	Peptide 110-116	
Asp	1.0		1.1*	(-)*
Glu		0.9	1.1	(1)
Pro		1.2	1.1	(1)
Ala	1.8	0.9	2.6	(3)
Leu	1.0		1.0	(1)
His			0*	(1)*
Recovery (%)	100	80	29	

* The substitution of Hb Hopkins-2 is therefore histidine to aspartic acid at α 112; although asparagine is converted into aspartic acid during acid hydrolysis preceding amino acid analysis, the possibility of a histidine-asparagine replacement could be ruled out on the basis of the relative elution positions of the whole haemoglobin from DEAE-Sephadex and of the α -chain from CM-cellulose. Comparison of the electrophoretic mobilities of peptides 110-113 and 110-116 at pH 3.5 (Fig. 5) with those of the neutral peptide 129-133 and the acidic peptide 114-116 establishes that they are acidic and therefore also indicates that the substitution of Hb Hopkins-2 is histidine to aspartic acid, rather than to asparagine.

Discussion

The insolubility of the α -chain core is the result of a number of contributory factors. If the thiol group of cysteine- α 104 is not blocked by some means, disulphide-bond formation will lead to the production of insoluble material during digestion. The most satisfactory means of blocking is aminoethylation (Jones, 1964) by using ethyleneimine (Raftery & Cole, 1963), which introduces an additional point for tryptic hydrolysis (Lindley, 1956), thus producing peptides α TpXIIa and α TpXIIb (Fig. 3). Neither peptide α TpXIIb, however, nor α TpXIII normally appears on a 'fingerprint' of the soluble tryptic peptides from aminoethylated α -chains, presumably because they are insoluble under the conditions of 'fingerprinting'; thus the intrinsic relative insolubilities of the tryptic peptides themselves contribute towards the formation of 'core'. A third contribution stems from the fact that the close proximity of the β -carboxyl group of aspartate- α 126 markedly inhibits the rate of tryptic hydrolysis at the C-peptide bond of lysine- α 127, resulting in the formation of insoluble peptide α TpXIIb-XIII (α 105-139). If the charge on this carboxyl group is eliminated by coupling the carboxylate function, by means of a water-soluble carbodi-imide, to a nucleophile such as glycnamide,

the rate of tryptic hydrolysis at lysine- α 127 is greatly enhanced (R. T. Jones, personal communication); evaluation of the usefulness of this approach to 'fingerprinting' indicates, as might be expected, that the insolubility of peptide α TpXIII still precludes its appearance on a 'fingerprint' of the tryptic peptides from aminoethylated 'glycineamidated' α -chain.

At present, probably the most straightforward way to examine critically the α -chain core region by 'fingerprinting' methods is that, outlined above, of 'fingerprinting' peptic peptides from the fragment α 104-141. All of the relevant peptic peptides can be obtained in pure form with relative ease, and the 'fingerprint' of the peptic peptides from fragment α 104-141 has proved to be a very diagnostic one in which a substitution which results in an alteration of the electrophoretic or chromatographic mobility of a given peptide would be readily detected. The method was used to determine the substitution in Hb Hopkins-2, a known α -chain core variant (Ostertag *et al.*, 1972). An indication of the resolution of the technique is provided by the good separation of Val-Thr-Leu from Thr-Val-Leu and of Leu-Ala-Ser from Ala-Ser-Leu.

The lower limit of material used will depend on the sensitivity of the amino acid analyser available. We have found it possible, by using 2-nitro-5-thio[14 C]cyanobenzoic acid, to obtain satisfactory results from 20 mg of globin, when using an amino acid analyser capable of detecting about 5 nmol of amino acid without any special modification. The limiting factor will be the detection of fragment α 104-141 after chromatography on Sephadex G-75; if 2-nitro-5-thio[14 C]cyanobenzoic acid is not used, some 40-50 mg of globin may well be required. The fact that cyanylation can be achieved directly after globin chain separation is a great help, since it lessens the losses of α -chain likely to be incurred during desalting and drying before reaction.

We confirm the findings of Jacobson *et al.* (1973) as to the apparent lack of serious side reactions during cyanylation and cleavage. No deamidation of tryptic peptides from fragment α 1-103 was detected by 'fingerprinting', and amino acid analysis of peptide α TpV, from fragment α 1-103, showed 80% recovery of methionine after anaerobic acid hydrolysis in the presence of 0.01% dithiothreitol.

The cleavage reaction was considered to be essentially complete. When a small amount of unsplit α -chain was present after the use of 2-nitro-5-thio[14 C]cyanobenzoic acid, as in Fig. 1(a), the radioactivity elution profile showed that no radioactivity was in the intact α -chain, suggesting either that cyanylation was incomplete, possibly as a result of incomplete reduction of cysteine- α 104 before cyanylation, or that some β -elimination of 14 C cyanide had occurred during the cleavage step

(Degani & Patchornik, 1974). Occasionally a small shoulder of material precedes the α -chain monomer from the Sephadex G-75 column (see Fig. 1b). This shoulder was considered to be disulphide-linked α -chain dimer, since (i) the 'fingerprint' of the soluble tryptic peptides from the unmodified protein was that of α -chain and (ii) the cysteine in this material was available for aminoethylation (as judged by the appearance of peptide α TpXIIa on the 'fingerprint' of the soluble tryptic peptides from the aminoethylated protein) only if the material was first reduced with dithiothreitol under denaturing conditions; denaturation in the absence of dithiothreitol followed by treatment with ethyleneimine did not result in aminoethylation.

We are indebted to Professor H. Lehmann for the sample of haemoglobin Hopkins-2 used in this work.

References

- Baglioni, C. (1961) *Biochim. Biophys. Acta* **48**, 392–396
 Baglioni, C. (1962) *Biochim. Biophys. Acta* **65**, 389–393
 Carrell, R. W. & Irvine, D. (1968) *Biochim. Biophys. Acta* **154**, 78–83
 Clegg, J. B., Naughton, M. A. & Weatherall, D. J. (1966) *J. Mol. Biol.* **19**, 91–108
 Cleland, W. W. (1964) *Biochemistry* **3**, 480–482
 Degani, Y. & Patchornik, A. (1971) *J. Org. Chem.* **36**, 2727–2728
 Degani, Y. & Patchornik, A. (1974) *Biochemistry* **13**, 1–11
 Degani, Y., Neumann, H. & Patchornik, A. (1970) *J. Amer. Chem. Soc.* **92**, 6969–6971
 Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 59P
 Gros, C. & Labouesse, B. (1969) *Eur. J. Biochem.* **7**, 463–470
 Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822
 Hill, R. L. (1965) *Advan. Protein Chem.* **20**, 37–107
 Hill, R. L. & Delaney, R. (1967) *Methods Enzymol.* **11**, 339–351
 Huisman, T. H. J. & Dozy, A. M. (1965) *J. Chromatogr.* **19**, 160–169
 Hunt, J. A. & Ingram, V. M. (1958) *Biochim. Biophys. Acta* **28**, 546–549
 Hunt, R. T., Hunter, A. R. & Munro, A. J. (1968) *J. Mol. Biol.* **36**, 31–45
 Ingram, V. M. (1958) *Biochim. Biophys. Acta* **28**, 539–545
 Jacobson, G. R., Schaffer, M. H., Stark, G. R. & Vanaman, T. C. (1973) *J. Biol. Chem.* **248**, 6583–6591
 Jepson, J. B. & Smith, I. (1953) *Nature (London)* **172**, 1100–1101
 Jones, R. T. (1964) *Cold Spring Harbour Symp. Quant. Biol.* **29**, 297–308
 Kilmartin, J. V. & Rossi-Bernardi, L. (1971) *Biochem. J.* **124**, 31–45
 Lang, A., White, J. M. & Lehmann, H. (1972) *Nature (London) New Biol.* **240**, 268–271
 Lehmann, H. & Huntsman, R. G. (1966) *Man's Haemoglobins*, 1st ed., p. 286, North-Holland Publishing Co., Amsterdam
 Lindley, H. (1956) *Nature (London)* **178**, 647–648
 Morris, H. R., Williams, D. H. & Ambler, R. P. (1971) *Biochem. J.* **125**, 189–201
 Ostertag, W., von Ehrenstein, G. & Charache, S. (1972) *Nature (London) New Biol.* **237**, 90–93
 Raftery, M. A. & Cole, R. D. (1963) *Biochim. Biophys. Res. Commun.* **10**, 467–472
 Reindel, F. & Hoppé, W. (1954) *Chem. Ber.* **87**, 1103–1107
 Romero-Herrera, A. E. & Lehmann, H. (1974) *Biochim. Biophys. Acta* **336**, 318–323
 Sanger, F. & Thompson, E. O. P. (1963) *Biochem. Biophys. Acta* **71**, 468–471
 Smith, I. (ed.) (1969) *Chromatographic and Electrophoretic Techniques*, 3rd edn., vol. 1, p. 370, Heinemann, London
 Udenfriend, S., Stein, S., Böhlen, P., Dairman, W., Leimgruber, W. & Weigle, M. (1972) *Science* **178**, 871–872
 Weigle, M., Debernado, S. L., Tengi, J. P. & Leimgruber, W. (1972) *J. Amer. Chem. Soc.* **94**, 5927–5928
 Winterhalter, K. H. & Huehns, E. R. (1964) *J. Biol. Chem.* **239**, 3699–3705
 Woods, K. R. & Wang, K.-T. (1967) *Biochim. Biophys. Acta* **133**, 369–370
 Yamada, S. & Itano, H. A. (1966) *Biochim. Biophys. Acta* **130**, 538–540