A Partial Amino Acid Sequence for Sheep Haemoglobin A

By D. BEALE

Medical Research Council Abnormal Haemoglobin Research Unit,
Department of Biochemistry, University of Cambridge

(Received 13 July 1966)

Amino acid analysis and terminal-group analysis of tryptic and chymotryptic peptides from sheep haemoglobin A have enabled a partial amino acid sequence to be worked out. By comparing this partial sequence with the known amino acid sequences of human haemoglobins A and F as well as horse slow haemoglobin the most probable sequence of sheep haemoglobin has been deduced.

Two allelic haemoglobins, A and B, are found among adult sheep (Evans, King, Cohen, Harris & Warren, 1956; Huisman, van Vliet & Sebens, 1958) and give rise to the genotypes AA, AB and BB. These two adult haemoglobins have similar α -polypeptide chains, but differ in their non- α -chains (Muller, 1961; van Vliet & Huisman, 1964; Beale & Lehmann, 1965a; Beale, Lehmann, Drury & Tucker, 1966) by having at least five differences in amino acid sequence (Beale, 1966a). This is in contrast with the allelic human haemoglobins, which generally differ in one amino acid residue (see Beale & Lehmann, 1965b), although recently a variant with two amino acid differences in the β -chain has been reported (Bookchin, Nagel, Ranney & Jacobs, 1966). The differences in amino acid sequence between the human haemoglobin variants and normal haemoglobin A can be interpreted as arising from the change of a single nucleotide base in a codon of messenger RNA (Beale & Lehmann, 1965b) and each of the differences in sequence between the two sheep alleles can also be interpreted in this manner (Beale, 1966a).

Under conditions of environmental stress due to artificial anaemia, AA and AB sheep gradually produce a third type of haemoglobin, C, in place of haemoglobin A. However, in AB sheep the proportion of haemoglobin B is unaltered and in BB sheep no haemoglobin C is produced (Huisman *et al.* 1958; Blunt & Evans, 1963; van Vliet & Huisman, 1964).

It therefore appeared to be of general interest to investigate the primary structures of the sheep haemoglobins, and the present paper describes how a partial sequence for haemoglobin A was obtained.

MATERIALS

Trypsin. This was a twice-crystallized salt-free preparation from Worthington Biochemical Corp., Freehold, N.J.,

U.S.A. The chymotryptic activity in the sample was assessed by examining the peptides released from human haemoglobin A (Beale, 1966b).

Chymotrypsin. This was a three-times-crystallized sample from Worthington Biochemical Corp.

Leucine aminopeptidase and carboxypeptidase A. These were both preparations from Worthington Biochemical Corp.

Solvents. Pyridine was of A.R. grade and 3-methylbutan-1-ol was of 'milk-testing' grade from Hopkin and Williams Ltd., Chadwell Heath, Essex. The proportion of isomers tended to vary slightly from batch to batch, which slightly altered the R_F of some peptides. Each batch was always tested against peptides of known R_F .

METHODS

Preparation and tryptic digestion of globin. Sheep haemoglobin A (1.2 ml.; 5%, w/v, in water), which had been purified by paper electrophoresis, was poured with stirring into 24 ml. of acetone containing 1.5% (v/v) of HCl at 0° (Ingram & Stretton, 1962). The precipitated globin was washed with 3 × 12 ml. of acetone and the residual solvent removed in a vacuum desiccator.

The globin (60 mg.) was dissolved in 6 ml. of water and the pH brought to 8.2 with 1.2 ml. of 0.5 M-NH₄HCO₃. Trypsin (0.6 ml.; 2 mg./ml. in water) was added and digestion carried out at 38° for 2 hr. The pH was then lowered to 6.4 with N-acetic acid and the insoluble 'core' peptides were precipitated and coagulated by heating the digest in a water bath at 90° for 5 min. The 'core' was spun down, washed with 3 × 1 ml. of water and dried in a vacuum desiccator to give 18 mg. of material. The supernatant containing the soluble tryptic peptides was freeze-dried to yield 41 mg. of material, which was then dissolved in 0.6 ml. of water.

Separation and amino acid analysis of the tryptic peptides. A sample (0.05 ml.) of the above solution (equivalent to 5 mg. of digested globin) was applied to a sheet of Whatman 3MM paper (22 cm. × 54 cm.) along a 2 cm. line by using a Pasteur pipette and warm air-drier to minimize spreading. The line of application was 15 cm. from a short edge and 4 cm. from a long edge of the paper sheet (Fig. 1a). Two such papers were prepared. After saturating with pH 6.4 buffer (pyridine-water-acetic acid, 25:225:1, by vol.) the

Bioch. 1967, 103

Fig. 1. Separation of the tryptic peptides for amino acid analysis. (a) Peptides are applied to a 2 cm. line (A) and separated by electrophoresis. (b) Separated peptide bands (P) are washed into small compact areas by pipetting on 2% formic acid from both sides. (c) A paper sheet of equivalent area is sewn on to the electrophoretograms ready for ascending chromatography.

papers were hung in a Michl-type electrophoretic tank (Michl, 1951; Ryle, Sanger, Smith & Kitai, 1955) with the lines of application nearest the anode. Electrophoresis was then carried out with a potential gradient of 50 v/cm. for 1.5 hr. with toluene as the coolant. The electrophoretograms were dried in a cabinet at 60° and the separated peptide bands washed into small compact spots by carefully pipetting on 2% formic acid at both ends of the bands (Fig. 1b). After again drying the papers, a sheet of equivalent area was sewn to each one along the long edge farthest from the separated peptides (Fig. 1c). Ascending chromatography was then carried out for 24 hr. with the top phase of the system 3-methylbutan-1-ol-pyridine-water (6:6:7, by vol.). After drying the chromatograms at 60° they were dipped into 0.02% (w/v) ninhydrin in acetone and hung in a cabinet at 80° for 5 min. to reveal the peptide spots. An initial yellow colour was taken as indicating *N*-terminal glycine and an initial brown colour as *N*-terminal serine.

Each peptide spot was eluted with constant-boiling HCl into a 0.1 ml. capillary tube (Sanger & Tuppy, 1951), which was then sealed at both ends and placed in an oven at 105° for 20 hr. (Clegg, Naughton & Weatherall, 1965). The capillary tube was opened and washed out with 1 ml. of pH 2.2 buffer for automatic amino acid analysis (Spackman, Stein & Moore, 1958).

Separation of tryptic peptides for sequence analysis. The remainder of the solution of tryptic peptides (0.5 ml., equivalent to 50 mg. of digested globin) was applied to a sheet of Whatman 3MM paper (33 cm. x 54 cm.) along a 25 cm. line, 15 cm. from a short edge of the paper (Fig. 2a). Electrophoresis was carried out at pH 6.4 as described above, after which strips (3 mm. wide) were cut from the electrophoretogram (Fig. 2a), developed with 0.2% (w/v) ninhydrin in acetone and tested for methionine (Toennies & Kolb, 1953), tyrosine (Acher & Crocker, 1952), tryptophan (Smith, 1953), arginine (Jepson & Smith, 1953) and histidine

(Whitehead & Beale, 1959). By using the developed strips as markers the areas of the electrophoretogram containing the separated peptides were each cut out and sewn to sheets of Whatman 3MM paper (44 cm. x 33 cm.) along lines 4 cm. from a short edge (Fig. 2b). Chromatography in 3-methylbutan-1-ol-pyridine-water was then carried out as described above. Strips were cut from the chromatograms, developed and used as markers as before. The areas of the chromatograms containing the separated peptides were cut out and each was eluted with *N*-acetic acid. Eluates were freeze-dried and dissolved in 0.5 ml. of water.

Some peptides required further purification by using electrophoresis at pH 3.5 (pyridine-acetic acid-water, 1:10:89, by vol.) (Michl, 1951). This was carried out in a similar manner to that described for electrophoresis at pH 6.4 except that the peptide sample was applied to a 12 cm. line and 'white spirit' or Varsol was used as the coolant.

Amino acid analysis of purified peptides. A sample (0.1 ml.) of each peptide solution (usually containing 0.05–0.1 μ mole of peptide) was freeze-dried in a 1 ml. tube with a neck restriction and 0.2 ml. of constant-boiling HCl was added. The tube was then frozen and evacuated, sealed at the restriction and placed in an oven at 105° for 20 hr. Each hydrolysate was freeze-dried with 2 x 0.2 ml. of water and analysed for amino acids on an automatic analyser (Spackman *et al.* 1958).

Digestion of tryptic peptides with chymotrypsin. Most of the purified tryptic peptides that contained tyrosine, tryptophan, phenylalanine, methionine or leucine were digested with chymotrypsin as follows. About 0.2–0.3 ml. of purified peptide solution was adjusted to pH 8.2 with 0.02 ml. of 0.5 M-NH₄HCO₃, and 0.04 ml. of chymotrypsin solution (1 mg./ml. in water) was added. After digestion at 38° for 4 hr. the digest was freeze-dried and the chymo-

Fig. 2. Separation of the tryptic peptides for sequence analysis. (a) Peptides are applied to a 25 cm. line (A) and separated by electrophoresis. A strip (M) is cut out and developed in ninhydrin to locate the peptide bands. By using the strips as a marker, the areas of the electrophoretogram containing the separated peptides (P) are cut out. (b) The cut-out areas are sewn to a sheet of Whatman 3MM paper for ascending chromatography. The paper at the backs of the sewn-on areas is cut away.

tryptic peptides were separated by paper electrophoresis on sheets of Whatman 3MM paper (54cm. x 18cm.). The pH used (6.4, 3.5 or 2.0) depended on the type of chymotryptic peptide (basic, acidic or neutral) expected from the amino acid composition of the tryptic peptide. When several neutral chymotryptic peptides were expected chromatography in 3-methylbutan-1-ol-pyridine-water was used as a second dimension to electrophoresis. Peptides were detected, eluted and freeze-dried as above, and then dissolved in each case in 0.2 ml. of water.

Amino acid analysis of chymotryptic peptides. A sample (0.1 ml.) of each chymotryptic peptide was freeze-dried in a 0.5 ml. tube with a restricted neck and 0.1 ml. of constant-boiling HCl was added. The tube was evacuated, sealed at the restriction and placed in an oven at 105° for 20 hr. The hydrolysate was freeze-dried, dissolved in 0.1 ml. of water, freeze-dried once more and applied to a sheet of Whatman 3MM paper (54cm. x 22cm.) for electrophoresis in a Michl-type tank at pH 2.0 with a potential gradient of 100 v/cm. for 20 min. Standard amino acid solutions and hydrolysed paper blanks were also run at the same time and the electrophoretograms were developed in 0.2% (w/v) ninhydrin in acetone. Visual comparison of the colour intensities of amino acid spots in the hydrolysates and standard solutions gave an approximate measure of amino acid composition. The paper blanks showed only traces of serine, glycine and leucine.

N-Terminal-group analysis. The fluorodinitrobenzene method as modified by Ambler (1963) was used. A sample (0.1 ml.) of peptide solution together with 0.1 ml. of 0.2 M-N-ethylmorpholine was adjusted to pH 8.4 with acetic acid, and 0.4 ml. of 1% (v/v) 1-fluoro-2,4-dinitrobenzene in ethanol was added. The reaction mixture was incubated at 38° for 4 hr. and then excess of reagent extracted with

ether. The aqueous phase was freeze-dried and dinitrophenol removed by sublimation *in vacuo*. The DNP-peptide was hydrolysed with 0.1 ml. of constant-boiling HCl in an evacuated tube at 105° for 12 hr. and the liberated DNP-amino acid extracted with ether. The DNP-amino acid was identified by electrophoresis at pH 9.0 [1% (w/v) $(\text{NH}_4)_2\text{CO}_3$] or by chromatography in 2-methylbutan-2-ol saturated with 2N-NH₃ (Fraenkel-Conrat, Harris & Levy, 1955) and 1.5 M-sodium phosphate buffer, pH 6.0 (Levy, 1954).

C-Terminal groups of tryptic and chymotryptic peptides. In tryptic peptides containing lysine or arginine these residues were presumed to be C-terminal. In chymotryptic peptides containing tyrosine, tryptophan or phenylalanine these residues were presumed to be C-terminal. For other peptides the C-terminal residue was found by using carboxypeptidase A. A sample (0.1 ml.) of peptide solution was brought to pH 8.4 with 0.02 ml. of 0.5 M-NH₄HCO₃, and 0.01 ml. of a solution of carboxypeptidase A (1 mg./ml. in water) was added. Digestion was carried out at 38° for 1 hr. and the digest then freeze-dried. Liberated amino acids were identified by electrophoresis at pH 2.0 and blank digestions were always run at the same time. However, no significant amounts of amino acid were found in the blanks.

Digestions with leucine amino peptidase. Purified peptide solution (0.1 ml.) was brought to pH 8.4 with 0.02 ml. of 0.5 M-NH₄HCO₃, and 0.01 ml. of leucine aminopeptidase solution (1 mg./ml. in 0.2 M-MgCl₂) was added. Digestion was carried out at 38° and samples (0.02 ml.) of the digest were taken after $\frac{1}{4}$, $\frac{1}{2}$ and 1 hr., acidified with acetic acid and freeze-dried. Released amino acids were identified by paper electrophoresis at pH 2.0 and blank digestions were always run at the same time.

Table 1. Amino acid compositions of the tryptic peptides from sheep haemoglobin A compared with those from human haemoglobin A

Peptide	Amount (μ mole)	Composition (moles/mole)													No. of residues			
		Asp	Thr	Ser	Glu	Pro	Gly	Ala	Val	Met	Leu	Tyr	Phe	Lys		His	Arg	Trp
15T (2)	0.04	1.1	—	0.9	—	—	—	2.2	1.0	—	0.9	—	—	1.0	—	—	—	7
α Tp1	—	1	—	1	—	1	—	1	1	—	1	—	—	1	—	—	—	7
7T (1)	0.05	2.0	—	2.3	—	—	—	2.1	1.8	—	0.9	—	—	2.0	—	—	—	11
α Tp1 + α Tp2	—	2	1	1	—	1	—	1	2	—	1	—	—	2	—	—	—	11
5T (1)	0.08	0.9	—	1.2	—	—	—	—	1.1	—	—	—	—	1.0	—	—	—	4
5T (2)	0.06	1.0	—	0.9	—	—	—	—	1.0	—	—	—	—	1.0	—	—	—	4
α Tp2	—	1	1	—	—	—	—	—	1	—	—	—	—	1	—	—	—	4
21T (1)	0.12	—	—	—	—	—	1.2	2.0	—	—	—	—	—	1.0	—	—	+	(5)†
21T (2)	0.08	—	—	—	—	—	1.0	2.0	—	—	—	—	—	1.0	—	—	+	(5)†
α Tp3	—	—	—	—	—	—	1	2	—	—	—	—	—	1	—	—	1	5
11T (1)	0.10	1.1	—	—	2.1	—	4.2	3.7	0.8	—	1.1	0.7	—	—	—	1.0	—	15
11T (2)	0.05	1.1	—	—	1.9	—	4.1	3.9	1.0	—	1.2	0.8	—	—	—	1.0	—	15
α Tp4	—	—	—	—	3	—	3	4	1	—	1	1	—	—	1	1	—	15
28T (1)	0.09	—	2.2	1.3	—	0.9	—	—	—	0.7	0.8	—	1.8	1.0	—	—	—	9
28T (2)	0.04	—	2.1	1.0	—	1.0	—	—	—	0.9	0.8	—	1.9	1.0	—	—	—	9
α Tp5	—	—	2	1	—	1	—	—	—	1	1	—	2	1	—	—	—	9
26T (1)	0.11	1.0	0.9	2.2	0.9	0.9	1.3	1.2	1.1	—	1.0	0.7	1.8	1.0	1.8	—	—	16
26T (2)	0.06	1.1	0.9	1.8	1.0	1.1	1.2	1.0	0.9	—	1.0	0.8	1.9	1.0	1.8	—	—	16
α Tp6	—	1	1	2	1	1	1	1	1	—	1	1	2	1	2	—	—	16
6T (1)	0.09	—	—	—	1.0	—	2.2	—	—	—	—	—	—	1.0	0.8	—	—	5
6T (2)	0.05	—	—	—	1.1	—	2.0	—	—	—	—	—	—	1.0	0.9	—	—	5
α Tp7	—	—	—	—	—	—	2	—	—	—	—	—	—	1	1	—	—	4
α Tp8	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
20T (1)	0.08	—	1.2	—	—	—	—	3.2	0.9	—	1.1	—	—	1.0	—	—	—	7
20T (2)	0.04	—	1.0	—	—	—	—	3.1	1.1	—	0.9	—	—	1.0	—	—	—	7
13T (2)	0.03	3.7	1.1	2.3	—	0.8	2.2	1.9	1.0	—	4.7	—	—	1.0	3.7	—	—	22
α Tp9	—	6	1	2	—	1	—	7	3	1	4	—	—	1.0	3	—	—	29
22T (1)	0.08	—	—	—	—	—	—	—	—	—	1.1	—	—	—	—	1.0	—	2
α Tp10	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1	—	2
19T (1)	0.04	1.9	—	—	—	0.8	—	—	2.2	—	2.1	—	—	1.0	—	0.9	—	9
19T (2)	0.02	2.1	—	—	—	0.8	—	—	1.9	—	2.0	—	—	1.0	—	0.8	—	9
α Tp10 + α Tp11	—	2	—	—	—	1	—	—	2	—	2	—	—	1	—	1	—	9
16T (2)	0.03	2.2	—	—	—	0.9	—	—	1.8	—	1.1	—	—	1.0	—	—	—	7
α Tp11	—	2	—	—	—	1	—	—	2	—	1	—	—	1	—	—	—	7
24T (1)	—	—	—	—	—	—	—	—	—	—	—	0.8	—	—	—	1.0	—	2
α Tp14	0.11	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	—	2

12T (1)	0.07	—	1.1	—	1.9	—	—	—	1.2	—	0.8	—	—	1.0	—	—	7
12T (2)	0.04	—	1.0	—	2.1	—	—	—	1.1	—	0.9	1.0	—	1.0	—	—	7
βTp_1		—	1	—	2	1	—	—	—	1	—	1	—	1	—	—	8
27T (1)	0.09	—	1.2	—	—	—	2.2	0.9	2.3	0.9	—	—	0.9	1.0	—	+	(9) ⁺
27T (2)	0.06	—	1.0	—	—	—	1.9	1.0	2.1	1.0	—	—	1.0	1.0	—	+	(9) ⁺
βTp_2		—	1	1	—	—	1	—	2	1	—	1	—	1	—	1	9
4T* (1)	0.21	—	—	—	—	—	—	1.0	—	1.0	—	—	—	1.0	—	—	2
10T (1)	0.12	0.9	—	—	2.1	—	2.3	1.8	2.1	1.8	—	0.9	—	—	1.0	—	11
10T (2)	0.07	1.1	—	—	1.9	—	2.1	1.9	2.0	1.9	—	1.1	—	—	1.0	—	11
βTp_3		2	—	—	2	—	3	1	1	3	—	1	—	—	1	—	13
29T (1)	0.08	—	1.2	—	0.9	1.0	—	1.7	—	1.7	—	1.6	0.6	—	1.0	+	(10) ⁺
29T (2)	0.04	—	0.8	—	1.0	1.1	—	1.9	—	1.9	—	1.8	0.7	—	1.0	+	(10) ⁺
βTp_4		—	1	—	1	1	—	2	—	2	—	2	1	—	1	1	(10)
18T (1)	0.12	3.6	—	2.2	1.1	—	1.3	0.9	3.2	0.9	0.8	1.0	—	2.7	1.0	0.8	19
18T (2)	0.07	3.8	—	2.0	1.0	—	1.1	1.0	3.1	1.0	0.8	1.0	—	2.7	1.0	0.9	19
βTp_5		3	1	2	1	2	2	1	1	1	1	1	—	3	1	—	19
4T* (1)	0.21	—	—	—	—	—	—	1.0	—	1.0	—	—	—	1.0	—	—	2
βTp_6		—	—	—	—	—	—	1	—	1	—	—	—	1	—	—	2
3T (1)	0.07	—	—	—	—	—	1.2	—	0.8	—	—	—	—	1.0	0.9	—	4
3T (2)	0.05	—	—	—	—	—	1.1	1.0	1.0	—	—	—	—	1.0	0.9	—	4
βTp_7		—	—	—	—	—	1	—	1	—	—	—	—	1	1	—	4
2T (1)	0.04	—	—	—	—	—	1.1	0.8	0.8	—	—	—	—	2.0	0.8	—	5
$\beta\text{Tp}_7 + \beta\text{Tp}_8$		—	—	—	—	—	1	1	1	—	—	—	—	2	1	—	5
1T (1)	0.10	—	—	—	—	—	—	—	—	—	—	—	—	1.0	—	—	1
βTp_8		—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
17T (2)	0.04	3.7	—	2.3	1.1	—	1.2	—	—	2.1	—	2.8	—	0.9	1.0	0.8	16
βTp_9		3	—	1	—	—	2	1	2	1	—	4	—	1	1	1	16
14T (2)	0.02	3.6	—	2.2	1.0	—	1.3	—	—	2.2	—	2.7	—	0.8	2.0	0.9	17
$\beta\text{Tp}_8 + \beta\text{Tp}_9$		3	—	1	—	—	2	1	2	1	—	4	—	1	2	1	17
8T (1)	0.04	2.8	0.8	0.9	3.2	0.9	0.7	1.1	1.0	1.1	—	3.2	—	1.9	1.0	1.7	(22) ⁺
$\beta\text{Tp}_{10} + \beta\text{Tp}_{11}$		3	2	1	2	1	1	1	1	1	—	3	—	2	1	2	(22) ⁺
9T (1)	0.03	1.1	1.2	1.4	2.8	0.7	1.2	1.0	1.0	0.9	—	0.9	1.8	1.0	1.7	—	16
βTp_{13}		—	1	—	3	2	—	2	1	1	—	—	1	1	1	—	12
$\beta_{117-120}$		—	—	—	—	—	1	—	—	—	—	—	1	1	1	—	4
23T (1)	0.09	0.9	—	—	—	—	1.2	4.1	2.9	0.9	—	0.9	—	—	1.0	1.0	12
23T (2)	0.06	1.1	—	—	—	—	1.2	3.8	3.1	1.0	—	1.0	—	—	0.9	1.0	12
βTp_{14}		1	—	—	—	—	1	4	3	1	—	1	—	1	1	—	12
25T (1)	0.11	—	—	—	—	—	—	—	—	—	—	—	0.8	—	1.0	—	2
βTp_{15}		—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	2

* Peptide included twice since 2 moles were released/mole of protein.

+ Tryptophan detected with Ehrlich reagent and presumed to be one residue.

Cysteic acid also present, presumably arising from one residue of cysteine.

§ One residue of cysteine also present.

β -Chain: 12T-27T-4T-10T-29T-18T-4T-3T-1T-17T-8T—core—9T-23T-25T (also, 2T is 3T-1T and 14T is 1T-17T).

Partial amino acid sequences of tryptic peptides. The results of the treatment of tryptic peptides with leucine aminopeptidase are given in Table 2, and the results of chymotryptic digestion of tryptic peptides are given in Table 3.

Peptide 15T had *N*-terminal valine, and leucine aminopeptidase released in order valine, leucine, serine and alanine. Chymotrypsin gave the peptides (Val,Leu) and (Ala₂Ser,Asp,Lys), the latter having *N*-terminal serine. The partial

sequence was therefore Val-Leu-Ser-Ala-(Ala,Asp)-Lys.

Peptide 5T had *N*-terminal serine and an electrophoretic mobility at pH 6.4 that suggested that its one aspartic acid residue was present as asparagine. Leucine aminopeptidase released in order serine and asparagine, then valine and lysine. The sequence was therefore Ser-Asn-Val-Lys.

Peptide 21T had *N*-terminal alanine and chymotrypsin gave the peptides (Ala₂Trp) and (Gly,Lys). The sequence was Ala-Ala-Trp-Gly-Lys.

Peptide 11T had *N*-terminal valine, and chymotrypsin gave the peptides (Val,Gly₃Ala₂Asp,Tyr), (Gly,Ala₂Glu,Leu), which had *N*-terminal glycine

Table 2. *Amino acids released after treatment of tryptic peptides with leucine aminopeptidase*

Peptide	Duration of treatment.....	Amino acids released		
		$\frac{1}{4}$ hr.	$\frac{1}{2}$ hr.	1 hr.
15T		Val+	Val+	Val+
		Leu+	Leu+	Leu+
			Ser+	Ser+
				Ala+
5T		Ser±	Ser+	Ser+
			Asn+	Asn+
				Val±
				Lys±
26T		Thr±	Thr+	Thr+
			Tyr±	Tyr+
			Phe±	Phe±
6T		Gly+	Gly+	Gly+ +
		His±	His+	His+
				Glu±
				Lys±
16T		Val+	Val+	Val+
			Asp±	Asp+
27T		Ala+	Ala+ +	Ala+ +
			Val+	Val+
				Thr+
				Gly±
12T		Met±	Met+	Met+
			Leu+	Leu+
			Thr±	Thr+
			Ala±	Ala+
				Glu±
10T		Val+	Val+	Val+
			Asp±	Asp+
				Glu+
3T		Ala+	Ala+	Ala+
		His±	His+	His+
			Gly±	Gly+
			Lys±	Lys+
17T		Val+	Val+	Val+
		Leu+	Leu+	Leu+
		Asp±	Asp+	Asp+ +
				Phe±

Table 3.. *Peptides from the chymotryptic digestion of tryptic peptides*

Peptide	Amino acid composition	Electrophoresis	Other details
15T1C	Val (+), Leu (+)	pH 3.5	
15T2C	Lys (+), Ala (++), Ser (+), Asp (+)	pH 3.5	<i>N</i> -Terminal Ser
21T1C	Ala (++), (Trp)	pH 6.4	Trp by Ehrlich reagent
21T2C	Lys (+), Gly (+)	pH 6.4	Yellow with ninhydrin
11T1C	Val (+), Gly (+++), Ala (++), Asp (+), Tyr (+)	pH 6.4	
11T2C	Gly (+), Ala (++), Leu (+), Glu (+)	pH 6.4	Yellow with ninhydrin; <i>C</i> -terminal Leu
11T3C	Glu (+), Arg (+)	pH 6.4	
28T1C	Met (+)	pH 2.0	Chromatography needed
28T2C	Met (+), Phe (+)	pH 2.0	Chromatography needed
28T3C	Leu (+)	pH 2.0	
28T4C	Leu (+), Ser (+), Phe (+)	pH 2.0	Chromatography needed; <i>N</i> -terminal Leu
28T5C	Lys (+), Thr (++), Pro (+)	pH 2.0	
20T1C	Ala (+++), Val (+), Leu (+)	pH 6.4	<i>C</i> -Terminal Leu
20T2C	Lys (+), Thr (+)	pH 6.4	
26T1C	Thr (+), Tyr (+)	pH 6.4	Chromatography needed
26T2C	Phe (+)	pH 6.4	Chromatography needed
26T3C	His (+), Pro (+), Phe (++)	pH 6.4	<i>N</i> -Terminal Phe
26T4C	His (+), Pro (+), Phe (+)	pH 6.4	<i>C</i> -Terminal Phe
26T5C	Asp (+), Ser (+), Leu (+), His (+)	pH 6.4	<i>N</i> -Terminal Asp; <i>C</i> -terminal His
26T6C	Lys (+), Gly (+), Ala (+), Val (+), Leu (+), His (+), Ser (++), Asp (+)	pH 6.4	<i>N</i> -Terminal Asp
26T7C	Lys (+), Gly (+), Ala (+), Val (+), Ser (+), Glu (+)	pH 6.4	<i>N</i> -Terminal Gly
12T1C	Met (+)	pH 2.0	
12T2C	Met (+), Leu (+)	pH 2.0	
12T3C	Leu (+)	pH 2.0	
12T4C	Lys (+), Ala (+), Thr (+), Glu (++)	pH 2.0	<i>N</i> -Terminal Thr
27T1C	Gly (+), Ala (++), Val (+), Thr (+), Phe (+)	pH 2.0	
27T2C	Trp (+)	pH 2.0	Trp by Ehrlich reagent
27T3C	Lys (+), Gly (+)	pH 2.0	Yellow with ninhydrin
10T1C	Gly (++), Ala (+), Val (++), Leu (+), Glu (++) , Asp (+)	pH 6.4	<i>C</i> -Terminal Leu
10T2C	Arg (+), Ala (+)	pH 6.4	
29T1C	Leu (+)	pH 2.0	
29T2C	Leu (+), Val (++), Tyr (+)	pH 2.0	<i>N</i> -Terminal Leu
29T3C	Val (++), Tyr (+)	pH 2.0	Chromatography needed
29T4C	(Trp), Val (++), Tyr (+), Pro (+)	pH 2.0	<i>N</i> -Terminal Val
29T5C	Pro (+), (Trp)	pH 2.0	Chromatography needed
29T6C	Arg (+), Thr (+), Glu (+)	pH 2.0	<i>N</i> -Terminal Thr
18T1C	Phe (+++), Glu (+), Ser (+)	pH 6.4	<i>N</i> - and <i>C</i> -Terminal Phe
18T2C	His (+), Gly (+), Ala (+), Val (+), Leu (+), Ser (+), Asp (+++), Met (+)	pH 6.4	<i>N</i> -Terminal Gly; <i>C</i> -terminal Met
18T3C	Ala (++), Asp (+), Lys (+)	pH 6.4	<i>N</i> -Terminal Ala
23T1C	Gly (+), Ala (+++), Val (+++), Leu (+), Asp (+)	pH 3.5	<i>C</i> -Terminal Leu
23T2C	Arg (+), His (+), Ala (+)	pH 3.5	<i>N</i> -Terminal Ala

and *C*-terminal leucine, and (Glu,Arg). The partial sequence was Val-(Gly₃,Ala₂,Asp)-Tyr-Gly-(Ala₂,Glu)-Leu-Glu-Arg.

Peptide 28T had *N*-terminal methionine and gave the chymotryptic products Met, (Met,Phe), Leu, (Leu,Ser,Phe), which had *N*-terminal leucine,

and (Pro,Thr₂,Lys). Therefore the partial sequence was Met-Phe-Leu-Ser-Phe-(Pro,Thr₂)-Lys.

Peptide 26T had *N*-terminal threonine and gave numerous products (Table 3) with chymotrypsin that could best be represented by the partial sequence Thr,Tyr-Phe-(Pro,His)-Phe-Asp-(Ser,-

Leu)-His-Gly-(Ser,Ala,Glu,Val)-Lys. The electrophoretic mobilities of the tryptic peptide 26T and the chymotryptic peptide 26T7C suggested that the glutamic acid residue was present as glutamine. Leucine aminopeptidase released threonine, tyrosine and phenylalanine.

Peptide 6T had *N*-terminal glycine, and leucine aminopeptidase released glycine and histidine followed by glutamic acid and lysine. The sequence was therefore Gly-His-(Gly,Glu)-Lys.

Peptide 20T had *N*-terminal valine and gave two chymotryptic peptides, (Ala₃,Val,Leu) and (Thr,Lys), the former having *C*-terminal leucine. The sequence was Val-Ala-Ala-Ala-Leu-Thr-Lys.

Peptide 16T had *N*-terminal valine, and leucine aminopeptidase released only valine and a little aspartic acid, most probably owing to the presence of a proline residue immediately following the aspartic acid (see Hill, Swenson & Schwartz, 1960). The partial sequence was therefore likely to be Val-Asp-Pro-(Val,Asn,Phe)-Lys.

Peptide 12T had *N*-terminal methionine and with chymotrypsin gave the peptides (Met,Leu) and (Glu₂,Thr,Ala,Lys), the latter having *N*-terminal threonine. Leucine aminopeptidase released methionine, then leucine, threonine and alanine, followed by a little glutamic acid. The sequence was therefore Met-Leu-Thr-Ala-Glu-Glu-Lys.

Peptide 27T had *N*-terminal alanine and with leucine aminopeptidase gave alanine and valine in the ratio 2:1 followed by threonine and a little glycine. The chymotryptic products were (Gly-Ala₂,Val,Thr,Phe), (Gly,Lys) and Trp. The partial sequence appeared to be Ala-(Ala,Val)-Thr-Gly-Phe-Trp-Gly-Lys.

Peptide 10T had *N*-terminal valine, and leucine aminopeptidase released valine followed by aspartic acid and glutamic acid. Chymotrypsin gave the peptides (Gly₂,Ala,Val₂,Leu,Glu₂,Asp) and (Ala,-Arg) in which the former had *C*-terminal leucine. The partial sequence was therefore Val-Asp-Glu-(Gly₂,Ala,Val,Glu)-Leu-Ala-Arg.

Peptide 29T had *N*-terminal leucine and gave several chymotryptic peptides (Table 3) that suggested a sequence Leu-Leu-Val-Val-Tyr-Pro-Trp-Thr-Gln-Arg. The electrophoretic mobility of the peptide at pH 6.4 indicated that the glutamic acid residue was present as glutamine.

Peptide 18T had *N*-terminal phenylalanine and gave the chymotryptic peptides (Phe₃,Glu,Ser), which had *N*- and *C*-terminal phenylalanine, (His,Gly,Ala,Val,Leu,Ser,Asp₃,Met), which had *N*-terminal glycine and *C*-terminal methionine, and (Ala₂,Asp,Lys), which had *N*-terminal alanine. The electrophoretic mobility of the last peptide at pH 6.4 suggested that the aspartic acid was present as asparagine. The partial sequence was Phe-

(Phe,Glu,Ser)-Phe-Gly-(His,Ala,Val,Leu,Ser,Asp₃)-Met-Ala-(Asn,Ala)-Lys.

Peptide 3T had *N*-terminal alanine, and leucine aminopeptidase released alanine and histidine followed by glycine and lysine. The sequence was therefore Ala-His-Gly-Lys.

Peptide 17T had *N*-terminal valine, and leucine aminopeptidase released valine and leucine with a little aspartic acid. Although this peptide contained phenylalanine and leucine, chymotrypsin attacked it extremely slowly, suggesting that these residues might be situated next to acidic ones.

Peptide 23T had *N*-terminal valine, and leucine aminopeptidase released valine and alanine with a little glycine. Chymotrypsin gave the two peptides (Gly,Ala₃,Val₃,Leu,Asp) and (His,Ala,Arg), the former having *C*-terminal leucine. The partial sequence was Val-(Val₂,Ala₃,Gly,Asn)-Leu-Ala-His-Arg. The electrophoretic mobility of the peptide suggested that the aspartic acid was present as asparagine.

Most probable amino acid sequence of sheep haemoglobin A. Fig. 5 shows the partial sequence obtained by arranging the tryptic peptides in order along the α - and non- α -chains, and positioning the residues of unknown sequence according to the known sequences of human haemoglobin A and horse slow haemoglobin (Konigsberg, Goldstein & Hill, 1963; Hill & Konigsberg, 1962; Matsuda, Gehring-Müller & Braunitzer, 1963). In the α -chain the sequence of residues 69–82 is less certain than that of the others. In the β -chain residues 66–75 were positioned by also taking into consideration the known sequence of the γ -chain of human haemoglobin F (Schroeder, Shelton, Shelton & Cormick, 1962). The insoluble 'core' peptides that form residues 100–139 of the α -chain and residues 104–115 of the β -chain have not yet been separated and analysed. However, by subtracting the total amino acid compositions of the soluble peptides in each chain from the known amino acid composition of each whole chain (Huisman, Reynolds, Dozy & Wilson, 1965) a tentative amino acid composition for the insoluble peptides can be deduced.

The α -chain of sheep haemoglobin A appears to contain 141 amino acid residues as does the human α -chain. In Fig. 5 residue numbers are given in parentheses whereas residue numbers in helical and non-helical segments (see Perutz, 1965) are indicated without parentheses.

The β -chain of sheep haemoglobin A appears to contain 145 residues, one residue less than the human β -chain. This deleted residue is the second one from the *N*-terminus of the chain so that from there onwards the residue number of a particular amino acid is always one less than that of an equivalent amino acid in the human β -chain. However, the residue numbers in helical and

Fig. 5. Probable amino acid sequence of sheep haemoglobin A. T, Hydrolysis by trypsin; t, partial hydrolysis by trypsin; C, hydrolysis by chymotrypsin; c, partial hydrolysis by chymotrypsin. Residues underlined represent insoluble 'core'.

larly aspartic acid, leucine peptide bonds generally showed considerable resistance to chymotryptic action. Only one histidine peptide bond was significantly broken with chymotrypsin and breaks after other residues were generally insignificant during the rather short digestion time used.

The action of leucine aminopeptidase varied appreciably. With some peptides amino acids were released too rapidly to deduce their sequence with any reliability. In other cases the order of some of the released amino acids could be found and in a few cases a considerable portion of the *N*-terminal sequence could be ascertained. Acidic residues generally retarded and sometimes stopped the action of leucine aminopeptidase, and proline residues seemed to be quite resistant. However, with additional information such as *N*-terminal groups and points of chymotryptic action a study of the release of amino acids from tryptic peptides by leucine aminopeptidase generally provided useful information about *N*-terminal sequences.

Amino acid sequence of sheep haemoglobin A. It is unusual for the β -chain of a mammalian haemoglobin to have only 145 residues, although this has been reported for bovine haemoglobin (D. Babin & W. A. Schroeder, unpublished work, cited by Zuckerkandl & Pauling, 1965). The mammalian β -chain is believed to have evolved from a primitive α -chain by way of the γ -chain by means of gene duplication (Ingram, 1961), and during this process the addition of a residue at position 2 presumably occurred. However, in the sheep β -chain it is position 2 from which the residue seems to be missing. Since the γ -chain of sheep foetal haemoglobin also has a residue missing from position 2 (D. Beale, unpublished work) as does the bovine γ -chain (D. Babin & W. A. Schroeder, unpublished work, cited by Zuckerkandl & Pauling, 1965), it would appear that this characteristic was established in early evolution, presumably owing to a coincidental deletion during the divergence of the chains (see Zuckerkandl & Pauling, 1965).

An examination of the probable sequence of the peptide chains of sheep haemoglobin A shows that residues can be positioned at the nine invariant sites, in accordance with the observations of Perutz, Kendrew & Watson (1965) for other haemoglobin and myoglobin chains (i.e. B6 glycine, C2 proline, C4 threonine, CD1 phenylalanine, E7 histidine, F4 leucine, F8 histidine, H10 lysine, H23 tyrosine). Basic residues can be positioned at B12, E5, H10, G19 and H24 and acidic residues at A4 and B8, again in accordance with the observations of these authors, and hydrophobic residues can be positioned at the internal sites listed by them. Residues can also be positioned in accordance with those listed by Perutz (1965) for possible polar and non-polar contacts between unlike chains.

It seems likely that the α -chains of sheep haemoglobin A and human haemoglobin A differ in sequence at 18 positions and that the β -chains differ at 26 positions. This gives a mean difference between chains of 22, or 11 mutations/chain. If a mutational change arises on average every 8 million years (Zuckerkandl & Pauling, 1962) it would mean that the evolutionary divergence of the sheep and human chains from a common ancestral chain has taken some 90 million years.

Sheep haemoglobin B more readily dissociates into its sub-units and has a lower oxygen affinity than sheep haemoglobin A (Shreffler & Vinograd, 1962; Huisman *et al.* 1958), presumably owing to structural differences between these two haemoglobins. However, it is difficult to see how the different properties of these haemoglobins can arise from the structural differences so far reported (Beale, 1966a), namely β CD8 asparagine, β E2 proline, β E19 methionine, β E20 lysine and β H22 lysine. None of these positions lies near the haem group or has been associated with contact between sub-units (see Perutz, 1965), although a proline at E2 instead of alanine would cause a slight change in secondary structure since proline cannot be satisfactorily positioned in an α -helix. Alternatively, an important structural difference might possibly occur in the region G7–G18, which has not yet been examined but is a likely area of sub-unit contact (Perutz, 1965).

Note added in proof. Since this paper went to press Wilson, McDaniel, Dobbs & Huisman (1966) and Boyer *et al.* (1966) have also reported sequences for β -chain of sheep haemoglobin A. The techniques used by these workers were quite different from those employed in this paper, but the agreement between the sequences is very good.

The author thanks Dr H. Lehmann for valuable advice and discussion and for the purification of the haemoglobin, Sir Alan Drury and Dr Elizabeth Tucker for providing the haemoglobin sample, and Mr D. Irvine for assistance with the amino acid analyses.

REFERENCES

- Acher, R. & Crocker, C. (1952). *Biochim. biophys. Acta*, **9**, 704.
- Ambler, R. P. (1963). *Biochem. J.* **89**, 349.
- Beale, D. (1966a). *Biochim. biophys. Acta*, **127**, 239.
- Beale, D. (1966b). In *Man's Haemoglobins*, p. 299. Ed. by Lehmann, H. & Huntsman, R. G. Amsterdam: North-Holland Publishing Co.
- Beale, D. & Lehmann, H. (1965a). *Biochem. J.* **97**, 35F.
- Beale, D. & Lehmann, H. (1965b). *Nature, Lond.*, **207**, 259.
- Beale, D., Lehmann, H., Drury, A. & Tucker, E. M. (1966). *Nature, Lond.*, **209**, 1099.
- Blunt, M. H. & Evans, J. V. (1963). *Nature, Lond.*, **200**, 1215.

- Bookchin, R. M., Nagel, R. L., Ranney, M. H. & Jacobs, A. S. (1966). *Biochem. biophys. Res. Commun.* **23**, 122.
- Boyer, S. H., Hathaway, P., Pascasio, F., Orton, C., Bordley, J. & Naughton, M. A. (1966). *Science*, **153**, 1539.
- Braunitzer, G., Gehring-Müller, R., Hilschmann, N., Hilse, K., Hobom, G., Rudloff, V. & Wittmann-Liebold, B. (1961). *Hoppe-Seyl. Z.* **325**, 283.
- Clegg, J. B., Naughton, M. A. & Weatherall, D. J. (1965). *Nature, Lond.*, **207**, 945.
- Evans, J. V., King, J. W. B., Cohen, B. L., Harris, H. & Warren, F. L. (1956). *Nature, Lond.*, **178**, 849.
- Fraenkel-Conrat, H., Harris, J. I. & Levy, A. L. (1955). *Meth. biochem. Anal.* **2**, 339.
- Hill, R. J. & Konigsberg, W. (1962). *J. biol. Chem.* **237**, 3151.
- Hill, R. L., Swenson, R. T. & Schwartz, H. C. (1960). *J. biol. Chem.* **235**, 3182.
- Huisman, T. H. J., Reynolds, C. A., Dozy, A. M. & Wilson, J. B. (1965). *J. biol. Chem.* **240**, 2455.
- Huisman, T. H. J., van Vliet, G. & Sebens, T. (1958). *Nature, Lond.*, **182**, 171.
- Ingram, V. M. (1961). *Nature, Lond.*, **189**, 704.
- Ingram, V. M. & Stretton, A. O. W. (1962). *Biochim. biophys. Acta*, **62**, 456.
- Jepson, J. B. & Smith, I. (1953). *Nature, Lond.*, **172**, 1100.
- Konigsberg, W., Goldstein, J. & Hill, R. J. (1963). *J. biol. Chem.* **238**, 2028.
- Levy, A. L. (1954). *Nature, Lond.*, **174**, 126.
- Matsuda, G., Gehring-Müller, R. & Braunitzer, G. (1963). *Biochem. Z.* **338**, 669.
- Miehl, H. (1951). *Mh. Chem.* **82**, 489.
- Muller, C. J. (1961). Ph.D. Thesis: University of Assen, The Netherlands. Assen: Gorcum and Co.
- Neil, G. L., Neimann, C. & Hein, G. E. (1966). *Nature, Lond.*, **210**, 903.
- Perutz, M. F. (1965). *J. molec. Biol.* **13**, 646.
- Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965). *J. molec. Biol.* **13**, 669.
- Ryle, A. P., Sanger, F., Smith, L. F. & Kitai, R. (1955). *Biochem. J.* **60**, 541.
- Sanger, F. & Tuppy, H. (1951). *Biochem. J.* **49**, 463.
- Schroeder, W. A., Shelton, J. R., Shelton, J. B. & Cormick, J. (1962). *Proc. nat. Acad. Sci., Wash.*, **48**, 284.
- Shreffler, D. C. & Vinograd, J. (1962). *Biochim. biophys. Acta*, **65**, 101.
- Smith, I. (1953). *Nature, Lond.*, **171**, 43.
- Spackman, D. H., Stein, W. H. & Moore, S. (1958). *Analyt. Chem.* **30**, 1190.
- Toennies, G. & Kolb, J. J. (1953). *Analyt. Chem.* **23**, 823.
- van Vliet, G. & Huisman, T. H. J. (1964). *Biochem. J.* **93**, 401.
- Whitehead, J. K. & Beale, D. (1959). *Clin. chim. Acta*, **4**, 710.
- Wilson, J. B., Edwards, W. C., McDaniel, M., Dobbs, M. M. & Huisman, T. H. J. (1966). *Arch. Biochem. Biophys.* **115**, 385.
- Zuckerkindl, E. & Pauling, L. (1962). In *Horizons in Biochemistry*, p. 198. Ed. by Kasha, M. & Pullman, B. New York: Academic Press Inc.
- Zuckerkindl, E. & Pauling, L. (1965). In *Evolving Genes and Proteins*, p. 97. Ed. by Bryson, V. & Vogel, H. J. New York: Academic Press Inc.