The Amino Acid Sequence of the γ Chain of Human Fetal Hemoglobin*

W. A. SCHROEDER, J. ROGER SHELTON, JOAN BALOG SHELTON, JEAN CORMICK, AND RICHARD T. JONES

From the Division of Chemistry and Chemical Engineering,† California Institute of Technology, Pasadena

Received April 15, 1963

The 146 amino acid residues of the γ chain of human fetal hemoglobin have been placed in sequence. The fetal hemoglobin for this investigation was isolated chromatographically from umbilical cord blood. The α and γ chains were separated prior to the determination of sequence. For the determination of the sequence, peptides were produced by individual hydrolyses with trypsin, chymotrypsin, or pepsin. The sequence of the individual peptides was determined largely through the application of the Edman degradation. The differences between the γ chains of human fetal hemoglobin and the β chains of human adult hemoglobin are responsible for the differences in the properties of the two molecules.

The pronounced differences in the properties and characteristics of human adult hemoglobin (hemoglobin A) and human fetal hemoglobin (hemoglobin F) are so well known that documentation is unnecessary. The dissimilarities obviously must stem from unlike chemical structures. Yet, despite these differences, hemoglobins A and F are basically alike in gross structure: both contain four polypeptide chains in pairs (Rhinesmith *et al.*, 1957; Schroeder and Matsuda, 1958). The α chains of hemoglobin A have the N-terminal sequence val–leu– and the β chains have val–his–leu– (Rhinesmith *et al.*, 1957; Rhinesmith *et al.*, 1958). The α chains of hemoglobin F also have the N-terminal sequence val–leu– but the γ chains have N-terminal glycine (Schroeder and Matsuda, 1958). The identity of the N-terminal sequence of the α chain of both hemoglobins led Schroeder and Matsuda (1958) to suggest that the sequence of the α chains was identical in hemoglobins A and F. Indirect data (Hunt, 1959; Jones *et al.*, 1959) supported this conclusion which has now been substantiated by the determination of the sequence of the α^F chain (Schroeder *et al.*, 1963). The dissimilarity in hemoglobins A and F, thus, must lie in the non-α chains of each. Knowledge of the sequence of the γ chain is necessary if the properties of hemoglobin F and hemoglobin A are eventually to be understood in terms of their structures.

A brief description of almost the entire sequence of the γ chain has already been presented (Schroeder *et al.*, 1961a, 1962b). It is the purpose of this paper to present the detailed results of the experiments which lead to the formulation of the sequence of the γ chain, although a few uncertainties still exist.

During the period that the sequence of the γ chain was being determined, advances were made in the elaboration of methods for the determination of sequence. A more direct approach to the solution of specific problems is now possible. The results of these advances are apparent if all the data are considered chronologically. Despite the fact that some initial conclusions were somewhat tentative because of the quality of the data, the more positive results from later experiments never vitiated the earlier conclusions. As the total sequence became apparent, experiments were frequently designed solely to substantiate a proposed sequence. Confusing initial results frequently were readily interpretable in the light of later results. The final sequence was attained through devious paths that hindsight shows could have been much shortened.

* This investigation was supported in part by a grant (H-2558) from the National Institutes of Health, United States Public Health Service.

† Contribution No. 2996.

EXPERIMENTAL

Source of Blood Samples.—Umbilical cord blood was obtained immediately after birth from apparently normal healthy infants most of whom were of Caucasian descent. In most instances, Alsever's solution was used as the anticoagulant in volume equal to that of the blood. The sample was then refrigerated until the hemoglobin solution was prepared. During the course of the investigation, blood from about a dozen infants was used.

Preparation of Hemoglobin Solutions.—Solutions of oxyhemoglobin were prepared from blood by methods that have been described in detail (Allen *et al.*, 1958; Clegg and Schroeder, 1959). The solution thus prepared was dialyzed against chromatographic developer at 4° until used. With rare exceptions the hemoglobin solution was prepared from the cells within a few days after they had been obtained and was used for its intended purpose almost immediately after dialysis against chromatographic developer was complete.

Chromatographic Isolation of Hemoglobin F.—The presence of several hemoglobin components in umbilical cord hemoglobin requires some procedure of separation. The chromatographic procedures of Allen *et al.* (1958) have been used to isolate the main component (which has been termed F_{II}[1]) for further investigation. Figure 5 of Allen *et al.* (1958) is typical of the chromatographic results that were obtained. During the course of the investigation there have been some refinements in procedure over those of Allen *et al.* (1958) as described by Jones and Schroeder (1963). The following procedure has proved to be most convenient.

A 3.5 × 35–cm column of Amberlite IRC-50 was equilibrated at 6° with 15 liters of Developer No. 4 (Allen *et al.*, 1958) in the course of several days. A 2-g sample of cord blood hemoglobin at a concentration of approximately 60 mg/ml[2] was then placed on the column and layered with developer (Jones and Schroeder, 1963). Development was carried out with Developer No. 4 at a flow rate of 60 ml/hour until the minor component designated as hemoglobin F_I had emerged from the column and the emergence of the front of hemoglobin F was incipient. Next, the upper

[1] The hemoglobin component termed F_{II}, because it is by far the major component of cord blood hemoglobin, must necessarily be identified with what is commonly called "fetal hemoglobin" or "hemoglobin F." Throughout the remainder of this paper "hemoglobin F" will be used to designate the purified hemoglobin F_{II} unless otherwise stated.

[2] This concentration has proved to be most satisfactory for large-scale columns. Concentrations of 90–100 mg/ml sometimes gave badly skewed zones.

portion of the column that contained hemoglobin A was removed. Then the remainder of the column was warmed to 40° for 10 minutes by circulating water through the jacket. At the end of 10 minutes the hemoglobin F was removed in 10–15 minutes at a flow rate of 750 or more ml/hour. The effluent was chilled in ice.

Prior to the repacking and re-equilibration of the column, the hemoglobin A was also eluted at 40° from that portion of the column that had been removed.

Concentration of Hemoglobin Solutions.—Hemoglobin F in the chromatographic effluent was concentrated by high speed centrifugation (Vinograd and Hutchinson, 1960). The concentrated solution was freed of salts by dialysis against water at 4°. From 2 g of cord blood hemoglobin *ca.* 1 g of hemoglobin F was commonly isolated.

Preparation of Globin. The procedure for the preparation of globin differed little from that of Anson and Mirsky (1930). In a typical experiment, 400 mg of hemoglobin F in 5 ml of water was added dropwise with vigorous stirring to a mixture of 400 ml of acetone and 8 ml of 2 N hydrochloric acid that had been cooled to −15°. The mixture was stirred for 30 minutes. After the globin had settled the supernatant solution was decanted as completely as possible. The globin was centrifuged at 4° and was washed three times with 50 ml of acetone at 4°. It was dissolved in 30 ml of water and dialyzed against water at 4°. Visking tubing with the designations 18/32 or 32/32 was impervious to the passage of globin during dialysis.

Separation of the α and γ Chains.—The α and γ chains were separated by chromatography according to Wilson and Smith (1959) or by precipitation according to Hayashi (1961).

The chromatographic procedure was carried out without significant modification.

In a typical separation by precipitation, 500 mg of globin F was dissolved in 250 ml of 8 M urea. Equal portions were placed in ten centrifuge tubes and to each was added an equal volume of 2 M trichloroacetic acid in 8 M urea. The mixture was allowed to stand for 5 hours with occasional stirring before the precipitate of γ chains was centrifuged. The α chains were recovered from the supernatant solution by dialysis against water (during which procedure they precipitated) and subsequent lyophilization. The precipitate of γ chains in each centrifuge tube was washed twice with 50 ml of M trichloracetic acid in 8 M urea. The washings were discarded. The precipitated γ chains were suspended in 30 ml of water and, after dialysis against water, were lyophilized.

This procedure is unsatisfactory unless the concentration of globin F is approximately 1 mg/ml of precipitating solution and unless the precipitate of γ chains is washed twice with precipitating solvent equal in volume to the original. Either procedure is applicable to globin F or carbamidomethylglobin F (only the γ chains are carbamidomethylated).

The physical nature of the product from the two procedures is different. The lyophilized material from the chromatographic separation is voluminous and of low density as are most lyophilized preparations. The product from precipitation is, after lyophilization, a compact, friable material.

Reaction with Iodoacetamide.—Hemoglobin F was treated with iodoacetamide in the same proportions and under the same conditions that Goldstein *et al.* (1961) applied to hemoglobin A.

Oxidation Procedure.—The oxidation of methionine to the sulfone and of cysteine to cysteic acid either in the γ chain prior to further degradation or in certain peptides after isolation followed exactly the description of Hirs (1956) for oxidation at −10° with the exception that the test tube with a side arm was dispensed with. The material and reagents were cooled in separate containers, rapidly combined, and returned to the cooling bath. No attempt was made to remove chloride from the material that was being oxidized. There has been little evidence of destruction of tyrosine or the formation of chlorotyrosine.

Enzymatic Hydrolyses.—Several enzymes have been applied to a variety of substrates to produce the desired peptides for the derivation of the sequence. The basic information about these hydrolysates is presented in Table I.

When hemoglobin F itself was used, the solution was taken to *p*H 8 and the protein was denatured at 90° for 4 minutes. The isolated γ chains were not denatured before enzymatic degradation. When the γ chain or its derivatives had been prepared chromatographically, the material dissolved to form a solution of about *p*H 5. When the *p*H was changed to 8 prior to the reaction, precipitation occurred at approximately *p*H 6. Even vigorous stirring did not prevent the precipitate from coagulating to rather large particles at *p*H 8. When the precipitation procedure had been used to prepare the γ chain or its derivatives, the product was not soluble in water but formed a fine suspension that did not alter in character when the *p*H was changed to 8. Despite these physical differences, the course of the enzymatic hydrolysis seemed to be little influenced. The original concentration of substrate was 1–2%. The enzyme-substrate ratio was as follows: trypsin, 1:200; chymotrypsin, 1:100; and pepsin, 1:100. During short hydrolyses with trypsin, an equal amount of enzyme was added after 60 minutes of reaction; in the longer hydrolyses it was added after 3 and 10 hours of hydrolysis. In the chymotryptic hydrolyses, enzyme was added after

TABLE I
CONDITIONS FOR ENZYMATIC HYDROLYSES

Substrate	Enzyme	Temperature	*p*H	Time of Hydrolysis (hours)
Hemoglobin F	Trypsin[a]	40°	8	1.5
γ chains	Trypsin[a]	40°	8	1.5
Carbamidomethyl γ chains	Trypsin[a]	Room temperature	8	24
Oxidized γ chains	Trypsin[a]	Room temperature	8	24
γ chains	α-Chymotrypsin[b]	40°	8	3
Oxidized γ chains	α-Chymotrypsin[b]	40°	8	1.5
γ chains	Pepsin[c]	Room temperature	2	2
γ chain "core"	Pepsin[c]	Room temperature	2	2

[a] Worthington Biochemical Corp., twice crystallized, salt-free, lot TRSF-717. [b] Worthington Biochemical Corp., three times crystallized, lot 6003. [c] Calif. Corp. Biochem. Research, three times crystallized, B grade, lot 108814.

60 minutes. No additional enzyme was added during the peptic hydrolyses. The *p*H of the tryptic and chymotryptic hydrolyses was maintained with a *p*H-stat (Radiometer, Copenhagen, Denmark) in an atmosphere of nitrogen. In no instance did the uptake of base during tryptic hydrolysis cease completely although it occurred at a much decreased and essentially constant rate during at least half of the period of even the shorter hydrolyses. The chymotryptic hydrolyses were purposely of short duration in order to obtain a greater array of peptides. Insoluble material remained after all tryptic hydrolyses but not after chymotryptic hydrolysis; the peptic hydrolyses never contained insoluble material. At the completion of the hydrolysis, the tryptic and chymotryptic hydrolysates were taken to *p*H 6.5, and any insoluble portion was removed by centrifugation. After addition of buffer reagents to give the proper normality and after adjustment of *p*H, all three types of hydrolysate were ready for chromatographic separation of the peptides.

Hydrolysis with trypsin, chymotrypsin, or pepsin was applied only infrequently to peptides. When it was applied, in general, several μmoles of peptide were dissolved in 400 μl of a buffer of appropriate *p*H, 0.5 mg of the enzyme was added, and the hydrolysis was allowed to proceed for 17 hours at 40°. If pepsin was used, the duration of hydrolysis was 2 hours at room temperature. After proper adjustment of *p*H the hydrolysate was chromatographed.

Leucine aminopeptidase (LAP)[3] was applied in two ways to determine the structure of a peptide. Hydrolysis with LAP permitted a satisfactory determination of amino acid composition when the peptide contained tryptophyl, glutaminyl, or asparaginyl residues or when bonds resistant to acid hydrolysis, such as those between valyl residues, were present. It was useful when the peptide contained proline for, in that case, all residues N-terminal to proline except one are removed and the remaining peptide after separation from the amino acids offers a new starting point for sequential degradation. The conditions of hydrolysis were the same in both applications. A 0.2–5-μmole sample of peptide was dissolved in 200 μl of water and to this was added 20 μl of 0.025 M magnesium chloride and 20 μl of 0.5 M tris-hydroxymethylaminomethane buffer at *p*H 8.5. The *p*H was checked with narrow-range *p*H paper and adjusted to *p*H 8.5 if necessary. Finally, 20 μl of LAP (18.2 mg/ml; c_1 = 30) in solution in 0.005 M Tris buffer and 0.005 M magnesium chloride at *p*H 8.5 were added and the mixture was maintained at 40° for 15 hours. At the completion of the reaction the solvent was removed in a stream of air at 40°. The residue was then taken up in solvent appropriate for the next step.

Hydrolysis with Acetic Acid.—Partridge and Davis (1950) observed the preferential release of aspartic acid upon refluxing with dilute acid. The following conditions have been applied to peptides. A solution of several μmoles of peptide in 10 ml of 0.25 M acetic acid was refluxed for 8 hours. After evaporation of the solvent in a rotary evaporator, the residue was ready for further examination.

Complete Hydrolysis with Hydrochloric Acid.—For the determination of amino acid composition the substance was hydrolyzed completely in hydrochloric acid. When hemoglobin F or γ chains were to be analyzed, the sample (5–10 mg) in solution or as a solid was transferred to a test tube and dried to constant weight at 110°. A peptide (usually 0.3–1 μmole) in solution was likewise transferred to a test tube and dried in a stream of air at 40°. To each sample was added 2 ml of doubly glass-distilled 6 N hydrochloric acid. Each tube was evacuated at room temperature with a water aspirator and sealed. All peptide samples were heated at 110° ± 1° for 24 hours in an oven with a fan to circulate the air. Samples of hemoglobin F or γ chains were heated at 110° ± 1° for 22 and 70 hours so that the extent of the destruction of sensitive amino acids could be determined. At the completion of hydrolysis, the tubes were opened and the acid was removed in a stream of filtered air at 40°. The residues were stored at −20° until analysis was made.

Determination of Amino Acid Composition.—The amino acid composition was determined with a Model 120 amino acid analyzer (Beckman Instruments, Inc., Spinco Division, Palo Alto, California) according to the procedure of Spackman *et al.* (1958). A precipitate that came from the heme group was always present after hydrolysates of hemoglobin F had been dissolved in the buffer in which samples were applied to the columns. Prior to the analysis such a precipitate was centrifuged off and only the supernatant solution was chromatographed.

Isolation of Peptides.—Throughout this investigation the initial separation of peptides in enzymatic hydrolysates was achieved by chromatographing on Dowex-50 × 2 with volatile pyridine–acetic acid developers. Most of the zones thus isolated were still mixtures that required further separation. Initially, this further separation was carried out by paper electrophoresis or paper chromatography. These methods were soon superseded by rechromatography on Dowex-1 × 2.

The chromatographic procedures on Dowex-50 and Dowex-1 have been described in detail by Schroeder *et al.* (1962a).

In only a few instances do the data to be presented derive from peptides that were isolated by paper methods. For these few, the following conditions apply. Electrophoresis on Whatman 3MM paper was made at *p*H 6.4 with pyridine–acetic acid buffer that was prepared according to Ingram (1958) and diluted with an equal volume of water. External cooling was used during electrophoresis at 50 v/cm for 45 minutes. The peptides were detected in the usual way by guide strips that were treated with ninhydrin. Water was the eluent.

Whatman 3MM paper was likewise used for ascending paper chromatography with 7:7:6 pyridine–*iso*-amyl alcohol–water (by volume) as the developer. Detection and elution of the peptides were as described above.

Edman PTH Procedure.—The sequence of amino acids in peptides was determined primarily by means of the Edman PTH procedure. This procedure, which is based on a modification of a method of Fraenkel-Conrat (1954; Fraenkel-Conrat *et al.*, 1955), has been described briefly (Schroeder *et al.*, 1961b) and is described in detail below. Continued experience with this procedure has led to the solution of some problems in its application to peptides but most of all to an appreciation of the degree to which atmospheric pollution (smog) of the type that exists in the Los Angeles area can interfere with its successful use. It is now apparent that the high oxidizing content of smog in conjunction with high humidity can destroy a PTH-amino acid. For this reason, certain phases of the procedure more recently have been routinely carried out in a smog-free chamber that was maintained at

[3] We are indebted to Dr. Robert L. Hill for generous gifts of LAP and for assistance in its application to various problems. Abbreviations used in this paper: LAP, leucine aminopeptidase; PTH, phenylthiohydantoin; CAM or CM, carbamidomethyl or carboxymethyl derivative of cysteine.

about 25% relative humidity by passing compressed air through two filters (Commercial Filters Corp., Melrose, Mass, Model 25a and Koby Corp., Boston, Mass., Model S) before dehumidification with a Lectrodryer (Model B-6-A, Pittsburgh Lectrodryer Division, McGraw-Edison Company, Pittsburgh, Pa.). This equipment is more than adequate to maintain the desired conditions in a volume of 25 cu ft.

All reagents that were used in this procedure were of reagent grade quality and, with the exception of acetone and dioxane, were used without further purification. Both acetone and dioxane were distilled without fractionation. Dioxane was stored frozen in 20-ml portions and was used within 2 weeks of distillation.

It was convenient to use 0.4–1.0 μmole of peptide which was distributed from aqueous solution over a sufficient number of 1 × 7-cm strips of Whatman No. 1 filter paper so that each strip carried about 0.2 μmole. A small hole near one end of each strip permitted it to be suspended from glass hooks on racks of glass rod during various stages of the procedure.[4] After the sample had been applied, the strips were dried in the smog-free chamber for 0.5 hour. (Its use at this point is convenient but not essential.) Each strip was then wetted with about 0.2 ml of a 20% solution of phenylisothiocyanate in dioxane. Two strips on a rack were placed in an 8-oz screw-capped jar to which had been added 15 ml of a 1:1:1 mixture (v/v/v) of pyridine-dioxane-water. (The jar was covered with aluminum foil before the cap was applied.) The 3-hour heating period at 40° in an oven was conveniently begun at 11 A.M.

When the strips were removed from the reaction atmosphere after heating, they were somewhat translucent. They were dried only to the point at which they lost their translucency and were then placed into 13 × 100-mm test tubes which contained sufficient benzene to cover the strip. If the strips were placed in the benzene without drying, some PTC-peptide dissolved. If they were dried too thoroughly, diphenylthiourea from a side reaction was incompletely extracted by the benzene. Diphenylthiourea cannot be distinguished from PTH-methionine in the chromatographic identification and, thus, will lead to incorrect conclusions unless it is removed at this step. The extraction with benzene was continued for 1.5 hours at which time the benzene was poured off, discarded, and replaced for a second extraction of 1.5 hours. A third extraction was then made overnight. (The timing of these extractions was convenient but does not appear to be critical.)

After the third extraction with benzene, the strips were aerated in the smog-free chamber for approximately 1 hour and then transferred to a desiccator (Corning No. 3120, 200 mm) that contained 15 ml each of glacial acetic acid and 6 N hydrochloric acid in separate beakers. The pressure within the desiccator was reduced to 100 mm of mercury and the degradation was allowed to proceed for 7 hours at room temperature. The degradation was ended at about 4 P.M. and the strips were aerated overnight in the smog-free chamber. (If one is not available, it is advisable to place the strips in a desiccator over Drierite.)

The dry strips were then extracted twice with acetone for one-hour periods to remove the PTH-amino acid. After the acetone had evaporated from the strips, they were ready for a repetition of the degradation procedure.

[4] Better exposure to reagents was achieved in this way than if the strips were placed in beakers or laid in petri dishes (Fraenkel-Conrat, 1954; Fraenkel-Conrat *et al.*, 1955).

The acetone was removed from the extract under reduced pressure at 40°. The residue was then taken up in acetone and applied to paper for subsequent chromatography. To each chromatogram, the material from about 0.1 μmole of peptide was applied.

Paper chromatography was used to identify the PTH-amino acid at each step of the degradation of the peptide. The procedures were those of Sjöquist (1953) and Edman and Sjöquist (1956) with minor modifications described by Shelton and Schroeder (1960). Starched paper was used throughout and the PTH-amino acids were located by heavy spraying with the iodine-azide reagent of Sjöquist (1953). Usually, one aliquot portion was chromatographed first with solvent A of Sjöquist for tentative identification and then another portion was chromatographed with another solvent to complete the identification.

Humidity has the most influence on chromatograms with Sjöquist's solvent A. When the humidity is very low, the PTH-amino acids move very rapidly and separate poorly. This trouble may be remedied in one of two ways. Before the sample is applied, the starched paper is dipped in a 10% solution of water in acetone. As the acetone evaporates, the sample is spotted, and then the chromatogram is begun. Obviously, some experimentation is necessary here, until one learns how to judge when the paper is in the correct state for chromatography. Alternatively, the sample may be applied to the dry paper which is then equilibrated at 45–50% relative humidity for 3 hours. A relative humidity of 45–50% is readily achieved in a closed chamber with a saturated solution of potassium carbonate.

When the relative humidity is above 60%, solvent A forms multiple fronts on the paper. Again, the paper may be equilibrated at 45–50% relative humidity after the sample has been applied or the sample may be applied under a heat lamp to dry the paper. The latter method clearly requires practice.

A new developer permits PTH-leucine and PTH-isoleucine to be distinguished. The starched sheet was dipped in a 20% solution of formamide in acetone. After the acetone had evaporated, the sample was applied, and the chromatogram was developed with 3:2 benzene-heptane (v/v) by descending chromatography in which the solvent front was allowed to move at least 40 cm. The R_F values of PTH-leucine and PTH-isoleucine were 0.71 and 0.66, respectively. Because the difference is small, several standards for each PTH-amino acid should be chromatographed on the same sheet.

Special treatment must be accorded to peptides that contain histidine, cysteic acid, or non-C-terminal arginine. These PTH-amino acids are not soluble in acetone and hence are not extracted after the degradation. They may be extracted with a 5% solution of water in acetone, but this solvent also extracts the remaining peptide and so prevents further degradation. However, the material on one-half strip is sufficient for identification. Therefore, one-half strip was sacrificed and the degradation was continued on the remainder without hindrance. Accordingly, when the peptide contained one or more of these amino acids (unless it was a tryptic peptide in which C-terminal arginine was expected), the schedule of degradation was altered. After the PTH-amino acid had been extracted with acetone, the next degradation was not begun until the acetone-soluble PTH-amino acid had been identified. If no acetone-soluble derivative could be detected, one-half strip was extracted with a 5% solution of water in acetone and the material, after evaporation of the solvent, was chromatographed with

solvent C. Alternatively, electrophoresis was used for identification. A sheet of starched Whatman 3MM paper was dipped in pyridine–acetic acid buffer at *p*H 6.4 (Ingram, 1958) that had been diluted with an equal volume of water. One-half strip from the PTH procedure was placed directly on the wet paper. After equilibration for 15 minutes, electrophoresis was carried out at 15–20 v/cm for 2 hours. The PTH-amino acids were easily detected with the usual reagent. This electrophoretic procedure is similar to that which Light *et al.* (1960) have used with PTH-cysteic acid and PTH-arginine.

The DNP Procedure.—The DNP procedure has been used only infrequently in this investigation. When it was applied, the methods for the formation of the DNP-derivative, the chromatography of DNP-peptides, and the identification of the N-terminal amino acid were those that have been described in detail elsewhere (Green and Kay, 1952; Schroeder and LeGette, 1953; Schroeder and Honnen, 1953; Schroeder *et al.*, 1954).

Results

Isolation of Hemoglobin F

The heterogeneity of umbilical cord blood hemoglobin made it imperative that hemoglobin F be separated from other components before investigation of the sequence was begun. That the hemoglobin F so isolated has the marks of a homogeneous substance was borne out by prior investigations (Allen *et al.*, 1958; Matsuda *et al.*, 1960). The isolation of hemoglobin F may be accomplished without undue difficulty.

Separation of the α and γ Chains

Column chromatography of globin F has given a separation of α and γ chains that is essentially that shown by Hunt (1959). In a few instances, the interrupted gradient of Chernoff (1961) was used without appreciably improving the quality of the products. All chromatographic separations were accompanied by a small rapidly moving peak much like that which Wilson and Smith (1959) observed with horse globin. The material in this peak probably is of relatively small molecular weight because nothing could be isolated after the urea and reagents in the developer had been dialyzed away. At the very end of the chromatogram, several small peaks were present when the gradient was continued for a long time. A plateau has been observed between the α and γ peaks when the interrupted gradient of Chernoff was used; after the absorbance of the effluent reached a minimum following the emergence of the α peak, it increased, remained constant for a period, and then increased again as the γ peak emerged.

In assessing the quality of the material in the peaks, the quantity of isoleucine has been an important indicator because of its presence only in the γ chains. The α chains were free of γ chains when isolated by chromatography. The γ chains have varied somewhat in different preparations but, in general, contained 15–20% of α chains as estimated not only by the isoleucine content but also from the N-terminal residues as determined by the DNP method. The apparent separation of chains that is observed on the chromatogram is sufficiently good (indeed, with the interrupted gradient, it is very good) that so great a contamination of the γ chains with α chains was unanticipated. It has been found that the "plateau" between the peaks contains α and γ chains in about equal proportion and that the small peaks at the end of the chromatogram contain a greater percentage of α chain than the main γ peak. It may be that the contamination of the γ peak is caused by some type of αγ aggregate (perhaps, even one in which disulfide bonds had formed during prior operations) rather than by incomplete separation from the α chain.

When the precipitation method of Hayashi (1961) was used to separate the α and γ chains, the γ chains contained about 15–20% of α chain and were about equivalent in quality to those from the chromatographic separation. The α chains contained about 20% of γ chain. This method has been successful only if the precipitation has been made in dilute solution as described above and if the resulting precipitate has been washed with large volumes of the precipitating solvent.

Insolubility of globin F precludes the use of countercurrent distribution to separate the chains (W. Konigsberg, personal communication).

Amino Acid Composition of the γ Chain

Inability to obtain pure γ chains by the above methods has prevented the direct determination of their amino acid composition. The composition, therefore, had to be derived by difference from the amino acid composition of hemoglobin F. From these data were subtracted the amino acid composition of the α^A chain which was under concurrent investigation and which was assumed to be identical with the α^F chain. The latter assumption is now known to be correct. In addition, the amino acid composition of hemoglobin Bart's or γ_4 was determined; there is good evidence that the four γ chains in this molecule are normal. These data have already been presented in some detail (Schroeder *et al.*, 1961a). If they are now corrected on the basis of our present knowledge of the α chain (Braunitzer *et al.*, 1961; Konigsberg and Hill, 1962), the amino acid composition of the γ chain from these data is ala_{11} arg_3 asp_{13} $(cys/2)_1$ glu_{12} gly_{12} his_7 $ileu_4$ leu_{17} lys_{11-12} met_2 phe_8 pro_{4-5} ser_{10} thr_{10} try_4 tyr_2 val_{13}. The amino acid composition from the determined sequence requires only five minor alterations: gly_{13} lys_{12} pro_4 ser_{11} try_3.

Separation of Peptides

Peptides were isolated from a variety of enzymatic hydrolysates that were prepared under several conditions and that were applied not only to the γ chain but to derivatives. Figures that depict the chromatograms of some of these hydrolysates have been published (Schroeder *et al.*, 1961a; Schroeder *et al.*, 1962a). It is impractical to present all the chromatograms (some 101) that are pertinent to this investigation. Therefore, in Figure 1, a composite of the data for all tryptic peptides is presented in which is indicated the effluent volume at which a given peptide has been found to emerge. The bars in the upper figure show the volume of emergence of tryptic peptides from a 3.5 × 100–cm column of Dowex-50 × 2 under the conditions that have been described (Schroeder *et al.*, 1962a). The lower part shows their behavior on 1 × 100–cm columns of Dowex-1 × 2. That peptides show identical or similar behavior in this composite is not necessarily indicative of difficulty in separation: though of identical elution volume on one ion exchange resin, they may be very different on the other. Note that 1, 12, and 16, though of identical behavior on Dowex-50, are easily separated by Dowex-1. Only simple mixtures from Dowex-50 were chromatographed on Dowex-1.

It should be emphasized that in this and other figures and in the remainder of this report we have dispensed with the temporary nomenclature by which we knew these tryptic peptides in the laboratory before they were ordered along the polypeptide chain. Instead we have numbered them consecutively from the N-terminus. This method of identification is applied

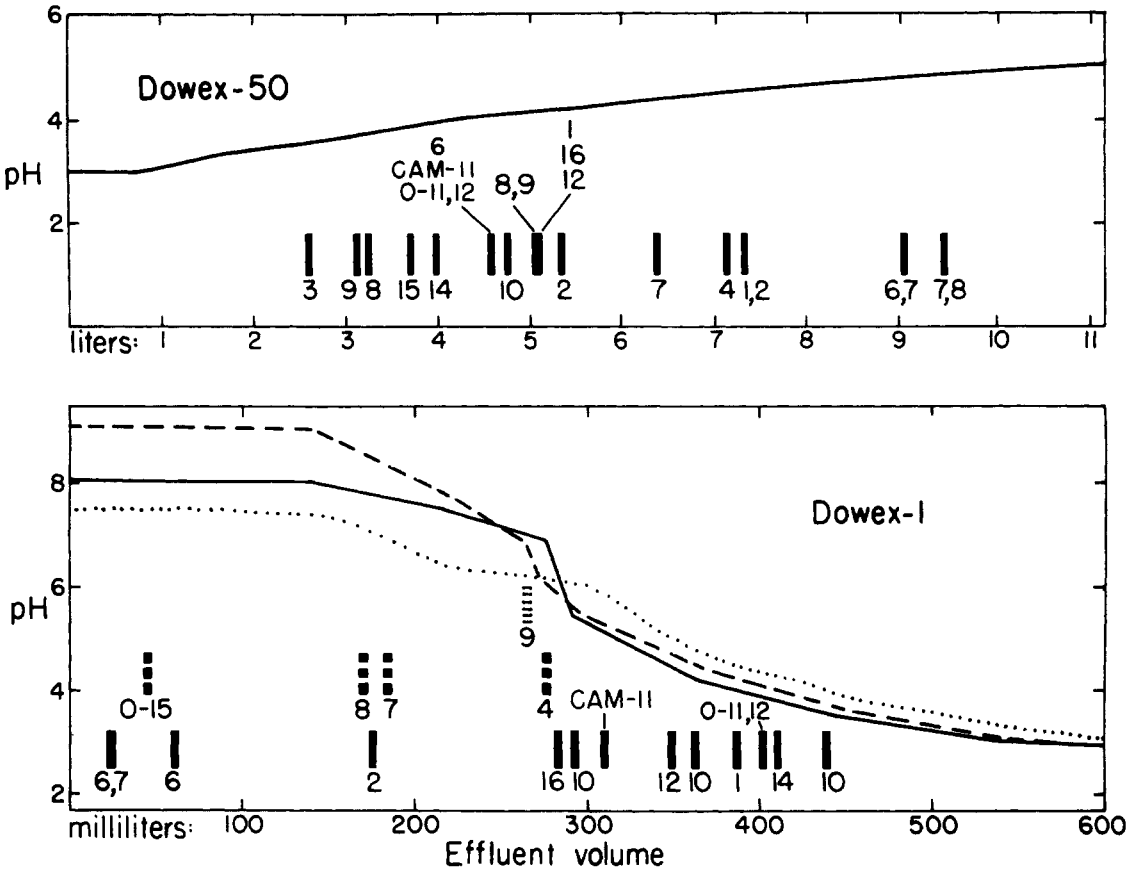

FIG. 1.—A composite of the chromatographic properties of the various tryptic peptides of the γ chain on Dowex-50 and Dowex-1. The initial developers on Dowex-1 are as follows: *p*H 8.0 buffer . . . .; *p*H 8.3 buffer——; *p*H 9.3 buffer– – – –. The representation of the bars thus also indicates the initial buffer that was used for the chromatogram. See text for further details.

to tryptic, chymotryptic, and peptic peptides which are distinguished by a preceding letter thus, T-3, P-7, etc. In a few instances, in which the methionine and/or cysteine has been oxidized, the letter "O" is prefixed to the designation, or if the carbamidomethyl derivative of cysteine has been formed the letters "CAM" are used.

It has not seemed worthwhile to present in detail the chromatographic behaviors of the thirty-five chymotryptic and the nineteen peptic peptides that were used in the derivation of the sequence of the γ chain. Instead, in the description of the derivation of the sequence, the *p*H at which a given peptide emerges both from Dowex-50 and Dowex-1 will be noted in order to give an indication of the chromatographic behavior. By reference to Figure 1 the approximate volume of emergence may be deduced.

Amino Acid Composition of Tryptic Peptides

In common with most present-day determinations of amino acid sequence, emphasis in the initial experiments was placed upon the isolation and characterization of the peptides in tryptic hydrolysates. The amino acid composition of the tryptic peptides from the γ chain is presented in Table II. These peptides were isolated from short and long hydrolyses of hemoglobin F or γ chains and of oxidized or carbamidomethylated γ chains. These data are a summary of the entire results of this investigation and not of any given hydrolysate. The tryptic peptides are listed in the order in which they follow each other in the γ chain and use the nomenclature that was described above.

The analyses of peptides T-2, O-T-15, and T-16 were obtained from LAP hydrolysates and the others from acidic hydrolysates. Peptides T-5 and T-13 have never been isolated in pure form and their amino acid compositions are derived from data that will be discussed in detail below. Tryptophan was easily determined in LAP hydrolysates as in the case of T-2 if LAP hydrolyzed the peptide completely; in the instances of T-4 and T-14 the proline blocked the action of LAP before tryptophan was released. Although proline prevented the determination of the complete amino acid composition of T-4 by LAP hydrolysis, it was nevertheless of use. This peptide contains a val-val sequence which can be hydrolyzed only by long hydrolysis which in turn destroys tyrosine to a large extent. LAP hydrolysis showed that the equal numbers of leucyl and valyl residues were present and substantiated the conclusion that two valyl residues were in the peptide. The peptides T-7 and T-7,8 are similar to peptides from the α chain which have the composition (gly_2 his lys) and (gly_2 his lys_2), respectively. These related peptides of the α and γ chains cannot be separated by the chromatographic procedures that were used. The analyses of T-7 and T-7,8 in Table II give a measure, therefore, of the degree of contamination of the γ chains by α chains in the particular sample from which these peptides were isolated and analyzed. The analysis of T-8,9 is from a hydrolysate of a DNP-peptide. After the hydrolysis of the DNP-peptide, the N-terminal residue was extracted and identified as di-DNP-lysine. The amino acids in the extracted hydrolysate were determined in the usual way except that a 10% solution of *n*-propanol in the usual *p*H 5.25 buffer was used on the short column of the amino acid analyzer in order to determine ε-DNP-lysine.

Amino Acid Sequence of the γ Chain

The amino acid sequence in the γ chain of hemoglobin

TABLE II. AMINO ACID COMPOSITION[a] OF TRYPTIC PEPTIDES OF THE γ CHAIN

	T-1	T-2	T-1,2[b]	T-3	T-4	T-5	T-6	T-7	T-6,7	T-8	T-7,8	T-9	T-8,9	T-10	CAM-T-11	T-12	O-T-11,12	T-13	T-14	O-T-15	T-16
try		0.95	Present[c]		Present[c]									1.10	1.01	1.01	2.03	(1)	0.24[d]		
lys	0.99	1.04	1.98			(1)	1.01	1.02	2.06	1.00	2.00	1.00		0.90	0.93	0.99	1.92	(1)	1.00		1.07
his	0.97		1.04					0.97	0.83		1.00									1.05	
arg				1.04	0.94				0.08								0.94				
cySO$_3$H															0.63						
CMcys															1.00	1.95	2.86	(1)			
asp	1.04		1.08	1.98	0.07	(3)			0.10			1.05	1.10	2.07						0.97	
metSO$_2$												0.81	1.11	0.08	0.95		0.87	(1)	0.94	0.99	
thr	0.94	1.95	2.60	0.89	1.00		0.06		0.12			0.86	1.16	0.12	0.98	0.05	0.86		0.95	3.00	
ser		0.98	0.94		0.21	(4)			0.27				0.31	0.10	1.98	1.01	2.86		4.00		
glu	2.05		2.02	2.01	1.14				0.17							1.04	1.04		1.08		
pro			0.13		1.18	(1)												(2)	0.09	0.98	0.05
gly	0.99	1.02	1.88	2.94	0.24	(2)		1.17	1.17		1.21	0.96	1.34	0.11	1.05		1.01	(1)	1.00	1.98	
ala		1.05	1.07	1.00	0.08	(2)		0.82	0.95		0.85	1.03	1.04	0.19	1.00		1.05	(3)	0.94	2.01	
val			0.15	2.00	1.58		0.98		0.90			1.03	0.76			1.00	0.94				
met						(1)						0.93	0.80					(1)			
ileu		1.00	0.89			(1)			0.05			1.98	1.95	1.97	1.96	0.97	2.91	(4)		0.97	
leu		0.97	1.13	1.01	1.95	(1)			0.15												0.94
tyr					0.83																
ohe	0.97		1.01			(3)							0.27		0.88	0.93	1.79	(1)	0.96		
ε-DNP-lys													0.71								
di-DNP-lys													0.53								

[a] The data are in terms of residues per peptide. Amino acids present to less than 0.05 residue are omitted. [b] A designation such as T-1,2 refers to a peptide in which the bond between the component tryptic peptides is still intact. [c] Detected by color reactions on paper. [d] Detected also by color reactions but a portion survived acidic hydrolysis.

F is depicted in Figure 2. In the succeeding paragraphs the basis for this formulation will be given. This presentation will take the following form and include the following information and conventions. The designations applied to peptides will be those that have been given above and will be applied only to tryptic, chymotryptic, and peptic peptides from hydrolysates of the γ chain itself. Their location in the γ chain may be learned from inspection of Figure 2. Peptides that were produced by further breakdown of any peptide will be given a suitable designation to facilitate reference. Discussion of the sequence will begin at the N-terminus of the γ chain and will proceed toward the C-terminus. Each tryptic peptide will be presented successively. After its sequence has been given, it will be followed by the data from which the sequence is derived. As peptides other than tryptic peptides are described, several items of information will be given in abbreviated form: the amino acid composition in residues, the results of the application of the Edman degradation by means of half arrows under the abbreviations of the residues, the extent of contamination (in per cent) by amino acids that are present in far less than molar quantity (if this contamination exceeds 10%), and the *p*H at which the peptide emerged from Dowex-50 and Dowex-1. Initial buffers of *p*H 8.0, 8.3, and 9.3 were used in the chromatography on Dowex-1 (Schroeder *et al.*, 1962a); these are indicated by the Roman numerals I, II, and III, respectively, that are connected to the data for Dowex-1. As an example peptide C-5 may be used:

C-5 $\overset{0.77}{\underset{\rightarrow}{\text{val}}}\text{–}\overset{0.77}{\underset{\rightarrow}{\text{val}}}\text{–}\overset{0.99}{\underset{\rightarrow}{\text{tyr}}}\text{–}\overset{1.00\ \text{pos.}}{\underset{\rightarrow}{\text{pro}}}\text{–try}$ (12%)

Dowex-50-*p*H 4.2; Dowex-1-(II)-*p*H 2.9

The values above the abbreviations for the amino acids are the amino acid composition in residues, the half arrows under the abbreviations depict the results of the successful applications of the PTH procedure, the percentage to the right in parentheses is the maximum degree of contamination, and the *p*H of emergence from the columns is shown below.

When the data that established the sequence of a tryptic peptide have been presented, they will be followed by the information that links two adjacent tryptic peptides.

T-1,2: gly–his–phe–thr–glu–glu–asp–lys–ala–thr–ileu–thr–ser–leu–try–gly–lys.—This double tryptic peptide is hydrolyzed into its components to a slight extent only in 90 minutes of tryptic digestion; in 24 hours, hydrolysis is virtually complete. The amino acid composition of T-1,2 was firmly established by its analysis and by those of T-1 and T-2 that were isolated from longer hydrolysates (Table II).

The N-terminus of T-1,2 was identified as glycine by the DNP procedure, but the PTH procedure was never applied to T-1,2 as such. When 3 μmoles of T-1,2 were digested with trypsin for 24 hours and the digest was chromatographed on Dowex-50, three peptides were isolated. One of these, t-1, clearly is from the N-terminal position and is equivalent to T-1.

t-1 $\overset{1.11}{\underset{\rightarrow}{\text{gly}}}\text{–}\overset{0.92}{\underset{\rightarrow}{\text{his}}}[5]\text{–}\overset{0.93}{\underset{\rightarrow}{\text{phe}}}\text{–}\overset{0.93}{\underset{\rightarrow}{\text{thr}}}\text{–}\overset{0.91}{\underset{\rightarrow}{\text{glu}}}\text{–}\overset{0.91}{\underset{\rightarrow}{\text{glu}}}\text{–}\overset{0.98}{\underset{\rightarrow}{\text{(asp, lys)}}}$ $\overset{1.04}{}$ (26%)

t-2 $\overset{0.99}{\underset{\rightarrow}{\text{ala}}}\text{–}\overset{0.87}{\underset{\rightarrow}{\text{thr}}}\text{–}\overset{0.98}{\underset{\rightarrow}{\text{ileu}}}[5]\text{–}\overset{0.87}{\underset{\rightarrow}{\text{thr}}}\text{–}\overset{1.03}{\underset{\rightarrow}{\text{ser}}}\text{–}\overset{1.05\ \text{pos.}}{\underset{\rightarrow}{\text{leu}}}[5]\text{–try}$ (27%)

t-3 $\overset{1.04}{\underset{\rightarrow}{\text{(gly}}}\ \overset{0.96}{\underset{\rightarrow}{\text{lys)}}}$ (34%)

[5] No PTH-amino acid was detected at the second degradation of t-1 because of our inability at the time to handle PTH-histidine properly. Likewise, isoleucine and leucine are placed correctly (from other data) although the PTH procedure as then used did not distinguish them at this point.

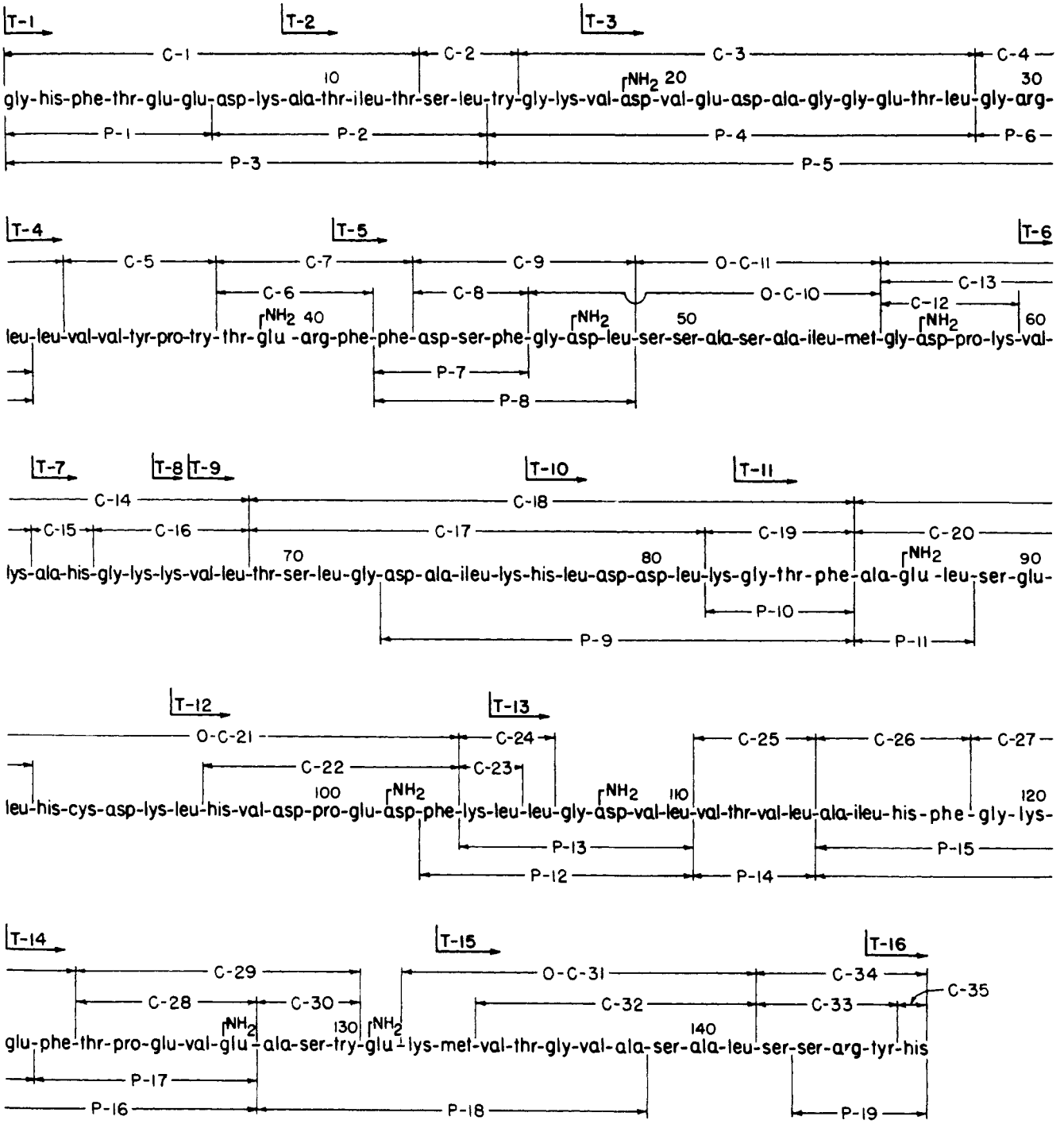

FIG. 2.—The amino acid sequence of the γ chain of human hemoglobin F. The positions and designations of the peptides that are discussed in the text are given.

These sequences were substantiated by a variety of peptides and experiments. For example, LAP hydrolysis of t-1 gave an initially surprising but useful result as follows:

1.03 0.92 1.01 1.04 1.90 0.46 0.46
(gly, his, phe, thr, glu, asp, lys) (11%)

This result is taken to mean that the sequence asp-lys is C-terminal. This inability of LAP to hydrolyze aspartic acid completely is useful because those residues that are equimolar in quantity to aspartic acid may be deduced to be C-terminal to it in the peptide.

The positioning of isoleucine and leucine in t-2 and hence T-2 came from chymotryptic peptide C-2:

C-2 0.85 1.00 pos.
ser–leu–try (16%)
Dowex-50-*p*H 4.3; Dowex-1-(II)-*p*H 3.0

Peptides T-2, T-4, and T-14 contain tryptophan but only T-2 contains serine, leucine, and tryptophan. Therefore, isoleucine and leucine must be placed as shown.

Other peptides that substantiate the sequence and composition of T-1,2 are as follows:

P-1 1.00 1.03 0.98 0.98 1.00 1.00
gly–his–phe–thr–glu–glu
Dowex-50-*p*H 4.15; Dowex-1-(II)-*p*H 3.65

P-2 0.99 0.99 1.03 0.85 0.97 0.85 0.87 1.19
asp–lys–ala–thr–ileu–thr–ser–leu
Dowex-50-*p*H 3.7; Dowex-1-(II)-*p*H 6.1

P-3 1.03 0.97 1.00 2.75 1.94 1.01 1.04 1.00 1.00 0.87 0.93
(gly, his, phe, thr, glu, asp, lys, ala, ileu, ser, leu)
Dowex-50-*p*H 4.15; Dowex-1-(II)-*p*H 3.9

C-1 1.09 0.91 0.95 2.55 1.92 1.25 1.07 1.04 0.91
(gly, his, phe, thr, glu, asp, lys, ala, ileu) (49%)
Dowex-50-*p*H 4.1; Dowex-1-(II)-*p*H 3.8

The γ chains have been shown to have the N-terminal sequence gly-his-phe- by PTH degradation (Shelton and Schroeder, 1960). The only tryptic peptide that fulfills the requirement is T-1 (or T-1,2) which accordingly must be in the N-terminal position of the γ chain.

T-3: val–aspNH$_2$–val–glu–asp–ala–gly–gly–glu–thr–leu–gly–arg.—This peptide emerged most rapidly from a Dowex-50 column and required no further purification. Edman degradation failed after five steps but gave the sequence val–aspNH$_2$–val–glu–asp–(ala,gly$_3$,glu,thr,leu, arg). When 7 μmoles of peptide T-3 was hydrolyzed with acetic acid for 6 hours, four compounds were isolated by paper electrophoresis:

AA-1 1.89 1.12 1.18 (val, asp, glu)

AA-2 2.06 2.03 1.23 (val, asp, glu)

AA-3 aspartic acid

AA-4 0.97 1.02 1.02 1.26 0.90 1.00 1.02 0.94 ala–gly–gly–glu–thr–leu–gly–arg (14%)

Peptides AA-1 and AA-2 contained obvious contamination but are of little import in the determination of the structure of T-3 because AA-4 permits the entire sequence to be written.

The link between T-2 and T-3 was supplied by peptic peptide P-4 and chymotryptic peptide C-3:

P-4 0.50[6] 1.00 0.99 0.99 1.00 0.99 0.98 1.00 1.00 1.00 1.00 try–gly–lys–val–aspNH$_2$–val–glu–(asp, ala, gly, gly, 0.98 0.95 1.05 glu, thr, leu)
Dowex-50-*p*H 4.2; Dowex-1-(II)-*p*H 2.95

C-3 0.97 1.06 0.99 1.01 0.99 1.08 1.01 1.06 0.97 0.97 1.08 gly–lys–val–aspNH$_2$–val–glu–asp–(ala, gly, gly, glu, 0.94 1.01 thr, leu) (18%)
Dowex-50-*p*H 3.85; Dowex-1-(II)-*p*H 3.2

The sequence try-gly-lys is unique in T-2. The partial sequence and composition of the remainders of P-4 and C-3 definitely identify them as stemming from T-3. Therefore T-3 is C-terminal to T-2.

T-4: leu–leu–val–val–tyr–pro–try–thr–gluNH$_2$–arg.—The problems that were associated with the determination of the amino acid composition of this peptide have already been discussed.

Six Edman degradations gave the sequence leu-leu-val-val-tyr-pro(try, thr, gluNH$_2$, arg). The following three peptides from a chymotryptic digest of γ chains are pertinent to the sequence of this peptide:

C-5 0.77 0.77 0.99 1.00 pos. val–val–tyr–pro–try (12%)
Dowex-50-*p*H 4.2; Dowex-1-(II)-*p*H 2.9

C-6 0.67 1.00 1.08 0.99 thr–gluNH$_2$–arg–phe (39%)
Dowex-50-*p*H 4.45; Dowex-1-(II)-*p*H 8.0[7]

C-7 0.81 1.03 1.00 0.96 0.96 thr–(gluNH$_2$, arg, phe, phe) (14%)
Dowex-50-*p*H 4.65; Dowex-1-(II)-*p*H 8.0[7]

Peptide T-4 is the only tryptic peptide that contains tyrosine, proline, and tryptophan or the sequence -val-val-, so C-5 must derive from it. Thus, only the three C-terminal residues of T-4 are unplaced. Arginine is present in the sequences -gluNH$_2$-arg, -gly-arg, and -ser-arg. Obviously, only the first and therefore C-6 and C-7 could have originated in T-4. These three chymotryptic peptides substantiate and complete the sequence of T-4. Also isolated in low yield from a 24-hour tryptic digest was the peptide

0.89 1.01 1.00 thr–gluNH$_2$–arg
Dowex-50-*p*H 4.15; Dowex-1-(II)-*p*H 7.85

This is the second example of the breaking of a peptide bond C-terminal to tryptophan by long tryptic hydrolysis. Peptide T-1,2 behaved identically. Such a hydrolysis has been observed by Bromer *et al.* (1957).

Chymotryptic peptide C-4 has the following composition and sequence:

C-4 1.00 0.91 1.00 1.00 gly–arg–leu–leu (13%)
Dowex-50-*p*H 4.45; Dowex-1-(II)-*p*H 8.0[7]

Actually, the sequence leu-leu is N-terminal in both T-4 and T-13. However, C-4 links T-4 to T-3 because, as will be shown below, T-13 is C-terminal to a lysyl residue.

T-5: phe–phe–asp–ser–phe–gly–aspNH$_2$–leu–ser–ser–ala–ser–ala–ileu–met–gly–aspNH$_2$–pro–lys.—Peptide T-5 comprises part of the insoluble material that was present at the end of any tryptic digest. Numerous attempts to isolate this peptide in pure form have failed. They will be discussed below. As a result, the entire composition and sequence of T-5 has been elucidated from a variety of peptides that were isolated from chymotryptic and peptic digests of the γ chain and may be listed as follows:

C-6 thr–gluNH$_2$–arg–phe } already described above
C-7 thr–gluNH$_2$–arg–phe–phe }

P-7 0.95 1.08 0.94 0.95 phe–asp–ser–phe (19%)
Dowex-50-*p*H 3.75; Dowex-1-(II)-*p*H 2.4

C-8 1.04 0.86 0.98 asp–ser–phe
Dowex-50-*p*H 3.5; Dowex-1-(II)-*p*H 2.6

P-8 1.00 0.99 0.95 1.00 1.03 0.99 1.00 phe–asp–ser–phe–gly–aspNH$_2$–leu
Dowex-50-*p*H 3.75; Dowex-1-(II)-*p*H 2.3

C-9 1.00 0.88 1.00 1.04 1.00 1.00 asp–ser–phe–gly–aspNH$_2$–leu
Dowex-50-*p*H 3.5; Dowex-1-(II)-*p*H 2.65

O-C-10 0.99 0.98 0.99 0.92 0.92 1.09 0.92 1.09 1.02 0.96 gly–aspNH$_2$–leu–ser–ser–ala–ser–ala–ileu–metSO$_2$
Dowex-50-*p*H 3.4; Dowex-1-(II)-*p*H 6.3

O-C-11 0.91 0.91 1.06 0.91 1.06 1.00 1.00 ser–ser–ala–ser–ala–ileu–metSO$_2$ (12%)
Dowex-50-*p*H 3.4; Dowex-1-(II)-*p*H 5.6

C-12 1.00 1.04 0.81 1.14 gly–aspNH$_2$–pro–lys (27%)
Dowex-50-*p*H 4.0; Dowex-1-(II)-*p*H 7.8

[6] Tryptophan was also detected by color reaction but in this instance half survived the acidic hydrolysis.

[7] A *p*H so close to that of the starting developer is indicative only of the fact that the peptide emerged in the first 150 ml of development before the *p*H change due to the gradient became effective. See Schroeder *et al.*, 1962a for chromatographic details.

In this case, let us consider first the linkage of T-5 to T-4. The sequence -thr-gluNH$_2$-arg is C-terminal in T-4. The peptides C-6 and C-7 which have already been discussed above provide the information that the sequence phe-phe must be at the N-terminus of T-5. The equating of the C-terminal phenylalanine of C-7 with the N-terminal phenylalanine of P-7 would be tenuous were it not for the fact that other phenylalanyl residues in the chain may be placed without question. The overlapping of the other peptides is good with the exception of C-11 and C-12 for which the evidence of linkage is less definitive than one would like.

The peptide C-12 has been isolated from two sources: chymotryptic digests of the γ chains themselves and of the "core." Its isolation from a chymotryptic hydrolysate of the γ chains, of course, requires an unusual type of splitting by chymotrypsin. On the other hand, our knowledge of the "core" would not permit C-12 to be associated with any tryptic peptide but T-5. From a peptic hydrolysate (2 hours) of the oxidized "core" was isolated a peptide with the amino acid composition

1.43 1.40 0.89 0.72 1.20 0.82 1.18 0.99
(ser, ala, ileu, metSO$_2$, gly, asp, pro, lys)
Dowex-50-*p*H 4.05; Dowex-1-(III)-*p*H 8.5

The material clearly was not homogeneous but is believed to be a mixture of equal parts of ser–ala–ileu–met–gly–aspNH$_2$–pro–lys and ser–ser–ala–ser–ala–ileu–met–gly–aspNH$_2$–pro–lys. The evidence, though less than desirable, suggests that C-12 follows C-11 in the sequence. For the present, this will be assumed to be correct and will be discussed again below.

T-6: val–lys.—Both the DNP and PTH procedures show the sequence to be val–lys.

A chymotryptic peptide, C-13, established the link between T-5 and T-6 if we assume that C-12 is at the C-terminus of T-5.

0.98 1.00 1.01 0.97 0.99 0.97
C-13 gly–aspNH$_2$–pro–lys–val–lys (13%)
Dowex-50-*p*H 4.3; Dowex-1-(III)-*p*H 9.0

Chymotrypsin, in this instance, has hydrolyzed a bond that is commonly resistant to its action.

T-7: ala–his–gly–lys.—The purity of T-7 depends primarily on the purity of the γ chain as has already been commented on. The DNP procedure showed that alanine was N-terminal, and the Edman degradation gave

ala–his–gly–lys

Peptide C-15 or ala–his (though of no import in establishing the total sequence) must have originated in T-7 inasmuch as this sequence occurs at no other place in the γ chain.

The bond between T-6 and T-7 is provided by T-6,7 which was isolated in small amount from a 1.5-hour tryptic hydrolysate. The pertinent data are these:

Analysis in Table II
T-6,7 val–lys–ala–his–gly–lys
Dowex-50-*p*H 4.75; Dowex-1-(II)-*p*H 7.75

In this peptide, nothing was detected in the second step of the degradation. This sometimes erratic behavior of lysine in the PTH procedure has not yet been controlled. The third and fourth steps proceeded without trouble.

T-7,8: ala–his–gly–lys–lys.—As in the case of T-7, a good analysis is dependent on the purity of the γ chains. Alanine was shown to be N-terminal by the DNP procedure. Because lysine is not the N-terminal amino acid, the peptide must be related to T-7 through the addition of lysine at the C-terminus. The sequence of chymotryptic peptide C-16 which will be discussed below is also relevant to T-7,8.

T-7,8 was present in reasonable yield in 1.5-hour hydrolysates whereas only T-7 and T-8 could be isolated from long hydrolysates.

T-8: lysine.—That lysine was present rather than some polymer was shown by the analysis of hydrolyzed and unhydrolyzed aliquot portions of the material: the quantity of lysine was equal in both samples. Electrophoresis and paper chromatography substantiated this conclusion.

T-9: val–leu–thr–ser–leu–gly–asp–ala–ileu–lys.—The Edman degradation gave

val–leu–thr–ser–leu–gly–asp–ala–(ileu, lys)

but did not distinguish between leucine and isoleucine. The usefulness of LAP in the study of aspartyl peptides is shown by its application here which yielded the analysis

1.00 0.97 0.97 1.00 0.97 0.97 0.51 0.54 0.56 0.54
(val, leu, thr, ser, leu, gly, asp, ala, ileu, lys)

and thus placed the isoleucyl residue. Further substantiation was given by acetic acid hydrolysis of T-9 and the isolation by electrophoresis of peptide AA-5.

0.93 1.00 1.04
AA-5 ala–ileu–lys (11%)

The link joining T-9 to T-7,8 is provided by chymotryptic peptide C-16:

0.96 1.04 1.04 1.11 0.84
C-16 gly (lys, lys, val, leu) (28%)
Dowex-50-*p*H 4.50; Dowex-1-(III)-*p*H 8.7

The sequence gly–lys–lys of C-16 is peculiar only to T-7,8 and the sequence val–leu is N-terminal only in T-9. Additional evidence of the link between T-8 and T-9 is provided by T-8, 9.

T-8,9: lys–val–leu–thr–ser–leu–gly–asp–ala–ileu–lys.—This peptide was present in low yield in a 1.5-hour tryptic digest. The DNP procedure showed an N-terminal lysine, and the analysis of amino acid composition of the remainder of the peptide was identical in composition to T-9 (see Table II).

A chymotryptic peptide C-14 was isolated in small amount and in rather impure state.

1.14 0.80 0.93 1.14 1.10 1.14 0.85 0.73 1.14 1.14
C-14 (gly, aspNH$_2$, pro, lys, val, lys, ala, his, gly, lys,
1.14 1.10 1.52
lys, val, leu)
Dowex-50-*p*H 4.5; Dowex-1-(III)-*p*H 8.65

It supplies a somewhat uncertain link from T-5 to T-9.

T-10: his–leu–asp–asp–leu–lys.—Although there was some difficulty in obtaining this peptide in high purity, the composition and sequence followed without problem. The Edman degradation gave

his–leu–asp–asp–leu–lys

Two chymotryptic peptides C-17 and C-18 firmly placed T-10 C-terminal to T-9 and N-terminal to T-11 thus:

0.78 0.96 0.99 1.20 0.99 1.02 0.98 1.00 1.00 0.99 0.99
C-17 thr (ser, leu, gly, asp, ala, ileu, lys, his, leu, asp,
0.99 0.99
asp, leu) (16%)
Dowex-50-*p*H 4.0; Dowex-1-(II)-*p*H 3.6

0.94 0.85 1.00 1.01 1.00 1.02 0.98 0.98 1.00 1.00 1.00 1.00
C-18 thr–ser–leu–gly–asp–ala (ileu, lys, his, leu, asp, asp,
1.00 0.98 1.01 0.94 0.98
leu, lys, gly, thr, phe)
Dowex-50-*p*H 4.2; Dowex-1-(II)-*p*H 4.5

T-11: gly–thr–phe–ala–gluNH$_2$–leu–ser–glu–leu–his–cys–asp–lys.—Peptide T-11 normally is part of the "core." However, if hemoglobin was reacted with iodoacetamide and the CAM-γ-chain that was isolated from the reaction product was digested with trypsin for 24 hours, CAM-T-11 could be isolated in appreciable yield.

The Edman degradation gave

gly–thr–phe–ala–gluNH$_2$–leu–(ser, glu, leu, his, CAM-cys, asp, lys).

Actually, the sequence of T-11 was almost completely known from examination of chymotryptic peptides before peptide CAM-T-11 was isolated, thus:

C-19 1.16 1.00 0.90 0.96 (lys, gly, thr, phe) (18%)
Dowex-50-*p*H 4.3; Dowex-1-(II)-*p*H 7.85

C-20 0.94 1.00 1.00 0.90 1.00 1.00 ala–gluNH$_2$–leu–ser–glu–leu (16%)
Dowex-50-*p*H 3.65; Dowex-1-(II)-*p*H 3.55

Peptides C-19 and C-20 must follow in sequence from the results of the Edman degradation of CAM-T-11.

Related to T-11 because of the single cysteinyl residue in the γ chain is O-C-21 which may be written as:

O-C-21 0.87 1.06 1.11 0.77 1.06 1.11 0.93 0.91 0.98 ala–gluNH$_2$–leu–ser–glu–leu–his–(cySO$_3$H, asp, 1.12 1.11 0.93 1.05 0.98 1.09 1.06 0.98 0.53 lys, leu, his, val, asp, pro, glu, aspNH$_2$, phe)
Dowex-50-*p*H 3.9; Dowex-1-(II)-*p*H 3.3

Thus, the single alanyl residue in O-C-21 was shown by one step of the Edman degradation to be N-terminal. When the remainder of the peptide was eluted from the strips, hydrolyzed with LAP, and analyzed, the results of the analysis were:

1.02 1.00 0.79 1.04 1.00 0.92
(gluNH$_2$, leu, ser, glu, leu, his)

Again the action of LAP was halted by some residue because other amino acids were present to the extent of less than 0.4 residue. This information in combination with C-20 fixes histidine after leucine in the sequence of O-C-21. The data from T-11, C-20, and O-C-21 place all residues of T-11 except cysteine, aspartic acid, and lysine.

The sequence of T-11 was completed with the help of acetic acid hydrolysis of CAM-T-11. From this hydrolysate, lysine was isolated in 93% yield and aspartic acid in 61% yield. Peptide AA-8 (32% yield) had the following composition:

AA-8 0.95 0.92 0.92 1.00 2.05 1.96 0.92 1.00 0.44 0.30 (gly, thr, phe, ala, glu, leu, ser, his, CMcys, cys/2?)
Dowex-1-(II)-*p*H 4.85

Some random hydrolysis apparently occurred, because serine and glycine were isolated in 26 and 17% yield, respectively. Peptide AA-8 contains the amino acids of CAM-T-11 except for aspartic acid and lysine. Their presence in excellent yield in the hydrolysate leads to the conclusion that the C-terminal sequence of T-11 must be –cys–asp–lys.

Peptide C-18 (see above) forms an obvious link between T-10 and T-11.

T-12: leu–his–val–asp–pro–glu–aspNH$_2$–phe–lys.—Peptide T-12 normally formed a part of the "core." It, like T-11, was isolated from CAM-γ chain after tryptic digestion for 24 hours.

Although the Edman degradation gave the entire sequence

leu–his–val–asp–pro–glu–aspNH$_2$–phe–lys,

it was already known from chymotryptic and peptic peptides, thus:

C-22 0.97 0.97 1.03 1.02 1.07 1.03 1.00 his–val–asp–pro–glu–aspNH$_2$–phe
Dowex-50-*p*H 4.0; Dowex-1-(II)-*p*H 3.2

P-12 0.76 1.00 1.04 1.04 1.00 1.05 0.97 1.04 phe–lys–leu–leu–gly–aspNH$_2$–(val, leu)
Dowex-50-*p*H 4.45; Dowex-1-(III)-*p*H 7.0

The C-terminal phenylalanine of C-22 cannot be equated unequivocally with the N-terminal phenylalanine of P-12 from these two peptides alone but it follows definitely from the sequence of T-12.

Peptide O-C-21 forms the link between T-11 and T-12 as does also O-T-11,12 (Table II) which gave the anticipated amino acids in the four degradations that were made.

T-13: leu–leu–gly–aspNH$_2$–val–leu–val–thr–val–leu–ala–ileu–his–phe–gly–lys.—Peptide T-13 is part of the "core" and like T-5 has never been isolated in the pure state. The above composition and sequence are derived from chymotryptic and peptic peptides. Thus, peptide P-12 (above) in conjunction with peptide P-13

P-13 1.20 1.00 1.00 1.05 1.02 0.98 1.00 lys–leu–leu–gly–aspNH$_2$–val–leu (17%)
Dowex-50-*p*H 4.3; Dowex-1-(II)-*p*H 8.1[7]

links six residues to peptide T-12. Although peptides C-23 and C-24

C-23 1.02 0.99 (lys, leu) (22%)
Dowex-50-*p*H 4.3; Dowex-1-(II)-*p*H 7.9

C-24 0.93 1.00 1.00 (lys, leu, leu) (26%)
Dowex-50-*p*H 4.35

probably derive from this section of the γ chain, they are of little use in defining the sequence.

There is no direct evidence that peptide P-14 (identical with C-25) should be placed next in the γ chain; the basis for this placement will be discussed below when the validity of the entire sequence is considered. Peptides P-14 and C-25 are as follows:

P-14 val–thr–val–leu
Dowex-50-*p*H 3.6; Dowex-1-(II)-*p*H 6.2

C-25 1.00 0.95 0.98 1.03 val–(thr, val, leu) (23%) (N-terminal by DNP method)
Dowex-50-*p*H 3.7; Dowex-1-(II)-*p*H 6.2

Peptides P-14 and C-25 were difficult to purify and, indeed, complete separation from contaminating peptides was attained only after converting to the DNP derivative and chromatographing in this form. Valine clearly was N-terminal and the composition of the peptide was easily determined. The PTH-degradation of P-14 was carried out with a mixture which contained some P-11 and ser-ala-leu that derived from peptides T-11 and T-15, respectively, but the results of the degradation showed the sequence that is given. Peptides P-14 and C-25 were isolated in greater yield than any other peptide in the peptic or chymotryptic hydrolysates that were studied.

The following peptides complete the sequence of peptide T-13 and, indeed, show the linkage to the succeeding tryptic peptide T-14.

C-26 0.97 0.89 1.00 1.05 ala–ileu–his–phe (23%)
Dowex-50-*p*H 4.3; Dowex-1-(II)-*p*H 6.5

C-27 0.98 1.08 1.00 0.93 gly–lys–glu–phe (25%)
Dowex-50-*p*H 4.2

1.06 0.73 0.76 0.85 1.10 1.02 1.04
P-15 (ala, ileu, his, phe, gly, lys, glu) (15%)
Dowex-50-*p*H 4.55; Dowex-1-(III)-*p*H 6.0

0.94 0.76 0.88 0.96 1.15 1.06 1.02 0.96 0.93 1.05 1.02
P-16 ala–ileu–his–phe–gly–lys–glu–(phe, thr, pro, glu,
1.05 1.02
val, gluNH$_2$) (30%)
Dowex-50-*p*H 4.45; Dowex-1-(III)-*p*H 4.5

T-14: glu–phe–thr–pro–glu–val–gluNH$_2$–ala–ser–try–gluNH$_2$–lys.—This peptide in most cases remained in the "core." Evidence suggests that its presence there was not primarily due to the resistance to hydrolysis of the lys-glu bond between T-13 and T-14 but rather to its own insolubility. Peptide T-14 was isolated in 10% yield from a 24-hour tryptic digest of the γ chain.

The Edman degradation gave

glu–phe–thr–pro–glu–val–(gluNH$_2$, ala, ser, try, gluNH$_2$, lys)

and to this degree substantiated the sequence that had been deduced from the following chymotryptic and peptic peptides:

Peptides P-15, P-16, and C-27 have already been described above.

0.96 0.76 1.14 1.00 1.00 1.00
P-17 phe–thr–pro–glu–val–gluNH$_2$
Dowex-50-*p*H 3.5; Dowex-1-(II)-*p*H 3.55

0.87 1.00 1.00 1.01 1.00
C-28 thr–pro–glu–val–gluNH$_2$
Dowex-50-*p*H 3.5; Dowex-1-(II)-*p*H 3.65

0.79 0.87 0.99 1.16 0.99 1.00 0.79 pos.
C-29 thr–pro–glu–val–gluNH$_2$–ala–ser–try (45%)[8]
Dowex-50-*p*H 3.85; Dowex-1-(II)-*p*H 2.65

1.00 0.86 0.69
C-30 ala–ser–try
Dowex-50-*p*H 4.15

0.92 0.84 0.94 1.07 0.97 0.94 0.91 1.06 0.94 0.92
P-18 ala–ser–try–gluNH$_2$–lys–met–val–thr–gly–(val, ala)
Dowex-50-*p*H 4.4; Dowex-1-(III)-*p*H 7.4

Peptide P-16 provides the link between T-13 and T-14.

T-15: met–val–thr–gly–val–ala–ser–ala–leu–ser–ser–arg.—The Edman degradation of the peptide gave

met–val–thr–gly–val–(ala, ser, ala, leu, ser, ser, arg).

On the basis of peptide P-18, an alanyl residue must follow this partial sequence of T-15. The sequence near the C-terminus of T-15 may be deduced from the following:

1.01 1.01 0.95 1.04
C-33 ser–ser–arg–tyr (22%)
Dowex-50-*p*H 4.45; Dowex-1-(II)-*p*H 7.9[7]

0.94 1.07 0.81 0.88
P-19 (ser, arg, tyr, his) (18%)
Dowex-50-*p*H 4.7; Dowex-1-(III)-*p*H 7.1

0.78 0.78 1.00 0.88 1.06
C-34 ser–ser–arg–tyr–his (13%)
Dowex-50-*p*H 5.1; Dowex-1-(III)-*p*H 6.6

The other two arginyl residues in the γ chain are associated with very different residues in T-3 and T-4. Therefore, the above three peptides must have originated in part from T-15. The remaining sequence of T-15 was elucidated as follows:

The above data place all but the seventh to ninth residues of T-15 in sequence. A chymotryptic peptide, O-C-31, was used to elucidate the sequence of these residues.

[8] The large percentage of contaminant is misleading. The contaminant could be identified as peptide C-3. The three glycyl residues of C-3, therefore, make the degree of contamination appear large according to the convention that we use to indicate maximum contamination. Actually, C-3 was present to the extent of about 15%.

1.11 0.87 2.00 0.92 1.22 2.03 1.04 1.04
O-C-31 (lys, metSO$_2$, val, thr, gly, ala, ser, leu) (16%)
Dowex-50-*p*H 4.1; Dowex-1-(II)-*p*H 8.0[7]

The composition of O-C-31 is such that it could have originated only from this portion of the γ chain. O-C-31 was digested with pepsin, and the following peptide was isolated and studied:

0.97 1.05 1.00
ser–ala–leu (17%)
Dowex-1-(II)-*p*H 5.4

The sequence of T-15, thus, is complete, and the evidence for its linkage to T-14 is provided by P-18.

T-16: tyr–his.—The N-terminal amino acid by both the DNP and PTH procedures was tyrosine.

T-16 was the only tryptic peptide that was not C-terminal in lysine or arginine. Therefore it is the logical choice for the C-terminus of γ-chain. Likewise, free histidine (C-35) was isolated from a chymotryptic hydrolysate in rather good yield. Its most logical source is the C-terminus; because of the occasional hydrolysis of bonds C-terminal to histidine, it is possible that part of the free histidine could have originated in residues 92 and 97. The link joining T-16 and T-15 is supplied by peptides C-33 and C-34 that have already been described.

The above data provide the evidence for the sequence of the 146 residues of the γ chain.

The Insoluble Portion or the "Core"

The initial tryptic hydrolysis of the γ chain was made for 1.5 hours at *p*H 8 and 40°. At the end of this time insoluble material was still present which may have increased or decreased in quantity when the *p*H was reduced to 6.5. The insoluble material accounted for almost 50% of the starting material. From the soluble portion, all of the tryptic peptides except T-5, T-11, T-12, T-13, and T-14 were isolated and partially characterized. At this time, of course, it was not known either that these were the undetected peptides or that they, indeed, constitute very close to 50% of the γ chain (71 of 146 residues). Because the insoluble portion or "core" constituted so much of the chain, it was subjected to chymotryptic hydrolysis. The peptides that were isolated clearly did not derive from the portion of the molecule that was soluble and so gave an insight into the nature of the insoluble portion. Further information was gained from chymotryptic hydrolysis of the entire γ chain. Indeed, it was possible to determine the sequence of the insoluble tryptic peptides from the chymotryptic peptides as indicated above.

This knowledge was used to devise experiments that hopefully would lead to the isolation of the peptides and the substantiation of the conclusions already drawn. Thus, when a long tryptic hydrolysis of CAM-γ chains was made, peptides T-11, T-12, and T-14 could be isolated from the soluble portion although the insoluble portion still amounted to 40%. When the oxidized γ chain was hydrolyzed for a long time with trypsin, the insoluble portion was less than 20%. The soluble portion contained, for example, peptide O-T-11,12, and probably O-T-11 and T-12 could have been isolated. Oxidation destroyed tryptophan so that peptide T-14 was not apparent. However, no evidence was found for the presence of peptides T-5 and T-13 in the soluble portion.

There are features in the amino acid composition of peptides T-5 and T-13 that are unique among the tryptic peptides of the γ chain. Both peptides contain isoleucine and peptide T-5 also has methionine. Furthermore, peptide T-5 is rich in phenylalanine and aspartic acid, and peptide T-13 in leucine. If, then,

the "core" is analyzed, the relative amounts of these amino acids furnish a clue as to the content of these peptides. From the amino acid composition of the "core" of the oxidized γ chain, it could reasonably be concluded that a considerable portion consisted of peptide T-5 but that less than a molar ratio of peptide T-13 was present because of the ratio of isoleucine to methionine. However, as noted above, neither peptide could be isolated from the soluble portion of the hydrolysate.

Numerous other attempts to isolate either peptide T-5 or T-13 in pure form have failed. The insoluble material is soluble at *p*H 3. When the *p*H was gradually raised, a precipitate formed at *p*H 5.0 and a second at *p*H 6.2. Amino acid composition suggested that the *p*H 5.0 precipitate was the richer in peptide T-5. However, repeated solution and precipitation were unsuccessful in producing a purer peptide T-5.

Because the "core" is soluble at lower *p*H, it is possible to chromatograph it. However, chromatography on Dowex-50 under the conditions that were used for other peptides was unsuccessful as was the procedure of Goldstein *et al.* (1961) for the core from the β chain. Passage through Sephadex likewise achieved no separation.

Peptide T-5 is analogous to a peptide from the β chain which is easily isolated. The two peptides differ in six of nineteen residues; the differences have thus conferred rather different characteristics upon peptide T-5.

Yield of Peptides from Hydrolysates

During this investigation, special stress has not been laid upon the determination of the yield of peptides that were isolated from the various hydrolysates. Because of the problems that were encountered with the core, the isolation of peptides received most attention throughout the study and less attention was devoted to the determination of the yield. As a result, few experiments were performed in duplicate and some isolations were necessary solely to provide material in order that the sequence of a few peptides might be completed. In all instances, chromatography on Dowex-50 was used for the initial isolation. Only rarely have the peptides from chromatography on Dowex-50 been adequately pure. Usually simple mixtures had to be separated. In our hands, paper chromatographic or electrophoretic methods were unsatisfactory for the separation of these mixtures and led to their separation by chromatography on Dowex-1. In recent experiments with another protein in which more quantitative data were desired, the yield of peptides has been unsatisfactory and has prompted an investigation of recovery from the chromatograms. Apparently, quantitative elution of peptides from Dowex-50 may be anticipated under the conditions that have been used. However, preliminary results suggest appreciable losses during chromatography on Dowex-1; further investigation is now under way.

If we consider all of the data from the various tryptic hydrolysates, the yield of peptides in the best instance for each ranged from about 40–95% of the anticipated amount. Certain peptides such as T-4, T-14, and T-15 were isolated in a yield of only 10–15%. Only a few per cent of such peptides as T-6,7 or T-8,9 were isolated, as might be expected. Peptide T-3 was always present in excellent yield; in one instance, it was 74% and in another 96%. It is noteworthy that peptide T-3 emerged first from the Dowex-50 chromatograms and needed no further purification. It is probable, to a degree at least, that the lower yield of many of the other peptides is a result of subsequent purification and that losses occurred to an extent that remains to be determined.

Discussion

Methodology of Sequence Determination

Use of Enzymes.—It is superfluous to review the experiences of others on the specificity of the enzymes that were employed in this investigation but the results in general will be mentioned.

Trypsin has shown the specificity that has come to be anticipated for it. Clearly the rate of hydrolysis of various bonds is different. The presence of aspartic acid in a position N-terminal to lysine markedly decreases the rate of hydrolysis. This sequence occurs in peptides T-1 and T-11. The isolation of T-1,2 and T-11,12 from short hydrolysates clearly indicates the resistance of the bond. Although most bonds that trypsin hydrolyzes are hydrolyzed to a large degree within 1.5 hours, such a short hydrolysis is useful because some bonds are incompletely hydrolyzed and the isolation of the corresponding peptides gives an insight into their arrangement. Longer hydrolysis splits the more resistant bonds but it has also resulted in hydrolysis of a bond C-terminal to tryptophan, an effect that was noted also by Bromer *et al.* (1957). Whether this is an inherent property of trypsin or due to contamination is not important because the extent of the hydrolysis is minor.

Chymotrypsin, of course, was far less specific in its action. In addition to the expected hydrolysis of bonds that are associated with tyrosine, phenylalanine, tryptophan, leucine, and methionine, the hydrolysis of bonds that are C-terminal to glutamine, threonine, histidine, and lysine has been observed. Although the hydrolysis of lysyl bonds is unusual, it has been observed in investigations of the α^A chain (Hill and Konigsberg, 1962) and of the α^F chain (Schroeder *et al.*, 1963) at a point where the sequence is similar to that of the γ chain; certain sequences thus may predispose to this unexpected behavior.

Pepsin has been useful mainly because of its ability to hydrolyze bonds on the N-terminal side of the aromatic amino acids and thereby to substantiate conclusions from chymotryptic hydrolyses. Pepsin's action on alanyl bonds that are untouched by chymotrypsin can be of much use.

In this investigation, the above three enzymes have been applied mainly to the γ chain itself rather than to peptides.

Leucine aminopeptidase (LAP) has proved to be of much assistance. By complete hydrolysis of the peptide, glutamine, asparagine, or tryptophan may be determined. On the other hand, the presence of aspartic acid in the peptide, in general, alters the course of action, but worthwhile information about the sequence may result nevertheless. The experience with peptide t-1 and T-9 as given above suggests that aspartic acid and all residues C-terminal to it in the peptide will be found in about half the expected amount in a LAP hydrolysate. Probably this is caused by the α- to β-rearrangement of aspartic acid that has been discussed in some detail by Naughton, Sanger, Hartley, and Shaw (1960). In the β form, aspartic acid would not be expected to be susceptible to the action of LAP. A third useful feature of LAP is its inability to hydrolyze prolyl peptides completely so that action ceases at the bond penultimate to proline. If the remaining peptide is then isolated from the reaction mixture, it offers a new starting point for sequential degradation.

Hydrolysis with Acetic Acid.—Although chemicals that react specifically with amino acid residues such as methionine or tryptophan are available (see the review by Witkop, 1961), their application to proteins or peptides is of somewhat limited usefulness because few residues of these amino acids are normally present in a protein. On the other hand, aspartic acid and/or asparagine make up an appreciable fraction of most proteins. The specific removal of aspartic acid by hydrolysis with dilute acid has been of much help in determining the sequence of peptides. From experience with γ-chain peptides as well as others, it appears that hydrolysis for 8 hours in refluxing 0.25 M acetic acid is optimal. Experience with peptide T-3 suggests that the carboxyl group of aspartic acid is freed first. Thees conditions have little or no effect on asparginyl, glutaminyl, or glutamyl residues. The method is most effective if the peptide contains a single aspartyl residue but it has also been successfully used when several are present.

Amino Acid Analyses.—It is difficult to generalize about the quality of the analysis that may be expected from any given peptide. The results may occasionally be unanticipated as in the case of peptide P-4 in which 50% of the tryptophan survived acidic hydrolysis. The partialde struction of serine, threonine, and tyrosine has commonly been observed: that of serine usually is between 5 and 15%, and of threonine between 5 and 10%, but that of tyrosine is more variable although the low content of tyrosine in the γ chains has given only a few observations. Considerable destruction of methionine may occur during the hydrolysis of a peptide but methionine does not appear to be destroyed appreciably when a protein is hydrolyzed. In a well-purified peptide, the molar ratios of the amino acids usually will not differ from an integral number by much more than 5% with the exceptions already mentioned. It is only rarely that a annalysis has not shown other amino acids in trace amounts. The source of this contamination is not always obvious although in many instances it is reasonable to conclude that a small amount of a specific peptide is present. Usually neither the slight variations from an integral number of residues nor the presence of minor contamination is sufficient to influence the conclusion as to the amino acid composition of the main constituent. This conclusion must of course be supported or invalidated by further data an by releated peptides from other hydrolysates.

Separation of Peptides.—The separation of peptides has been discussed in some detail in a previous publication (Schroeder *et al.*, 1962a). The unsatisfactory quality of peptides that had been purified by paper electrophoresis or chromatography as well as the high percentage that could not be recovered promptedt he use of successive chromatograms on columns. In the few instances of peptides that were isolated by paper methods, the degree of contamination is relatively high and representative of our experience. Some analyses of tryptic peptides in Table II are of rather recent determination on samples that had been specially isolated by column chromatography alone. They replace earlier and less satisfactory analyses of samples in whose isolation paper methods had been used. In no case was a previous conclusion about the amino acid composition altered by the second analysis, but the composition was always evident without equivocation.

Although each chromatogram by which a given peptide was isolated has not been depicted, the data are sufficient that the general chromatographic behavior can be deduced. Continued experience with the chromatographic systems has made it apparent that on Dowex-1 chromatograms a histidyl residue behaves as a neutral residue and that, like phenylalanyl and tyrosyl residues, it appreciably retards the emergence of a peptide.

The Edman Degradation.—The Edman degradation as used in this investigation is an approach that differs from most other applications. The present method permits the positive identification of the residue that has just been removed rather than its identification by the difference between the amino acid composition of the original and degraded peptide. The latter procedure has resulted in some incorrect conclusions (Smyth *et al.*, 1963). The positive approach, for example, permits the identification of asparagine and glutamine as such and thereby removes reliance on the over-all charge of the peptide as a means of identification. Asparagine presents no problem regardless of its position in the peptide. However, if glutamine is N-terminal and has been converted to pyrrolidone carboxylic acid, it is inaccessible to the degradation. If glutamine is not N-terminal, no problem has arisen. Threonine and serine appear to form and maintain a normal PTH derivative. In the initial experience with the method, the PTH-amino acids that did not extract with acetone presented a problem that has been overcome as described in the experimental part of this paper. Lysine still exhibits somewhat erratic behavior for which the cause has not been determined. Sometimes it appears in the acetone extract and is easily identified. At other times, it is not detectable. Though a gap may thus appear at this step in the determination of the sequence because no residue can be identified, yet the degradation itself apparently is normal because succeeding degradations usually proceed without difficulty.

As applied here, the Edman degradation is essentially a qualitative procedure. It can be made semiquantitative if serial dilutions of the acetone extract are chromatographed and compared with known amounts. When this was done, it was concluded that the yield per degradation is usually of the order of 80% or better. The experienced worker by inspection of the size and intensity of the spots on the chromatograms in conjunction with the amount of peptide that has been degraded has no difficulty in determining whether the yield at each degradation is normal or has decreased markedly.

The procedure as now employed with the schedule that has been indicated in the experimental part is such that six peptides may easily be under simultaneous study and on the average three residues may be identified daily. Indeed, unless several peptides are studied simultaneously, the procedure is relatively inefficient because, for example, chromatographic identification can be made as readily with several as with one sample.

Figure 3 is a montage of the final chromatograms that identified the PTH-derivative at each degradation of peptide P-8. The results are generally indicative of what may be expected from the application of the method. In this instance, six of the seven residues were determined, and the seventh, of course, followed from the amino acid composition. Both aspartic acid and asparagine were present and readily identified. The behavior of serine and aspartic acid is very similar in the solvent shown although they may still be distinguished. Actually, they are better differentiated in solvent A which preceded the use of solvent E. If the average yield at each degradation is about 80% as indicated above, one might expect to detect the PTH-derivative from the preceding degradation in each succeeding step. In Figure 3, the first three

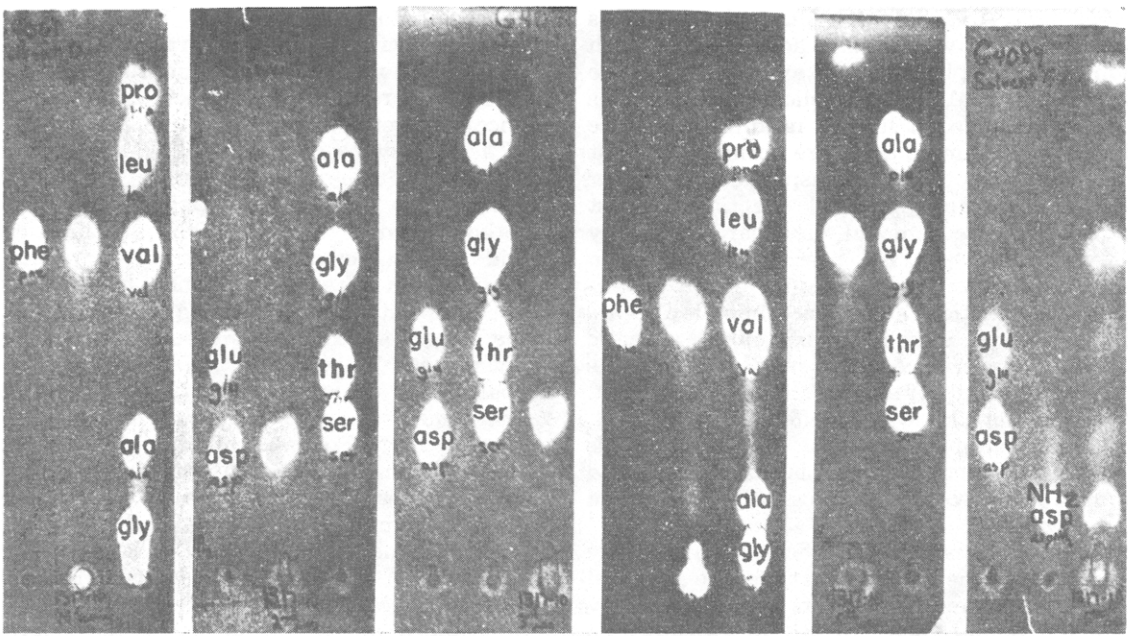

FIG. 3.—The results of the Edman degradation of peptide P-8. The final chromatogram that verified the PTH-amino acid from each degradation is shown. The spots of the reference PTH-amino acids are labeled but the spot from each sample is unmarked.

degradations show no evidence of the PTH-derivative of the preceding step. In the last three, some of the derivative from the preceding step is apparent but it by no means interferes with the interpretation of the results.[9] The application of the procedure to other peptides may be more or less successful than this example. Peptides AA-4, P-4, O-C-10, T-9, and P-18 may be cited as examples in which a greater number of degradations succeeded. On the other hand, in peptide T-3 only five steps were successful. It has been observed in this and other investigations in this laboratory that the procedure sometimes ceases to be successful without apparent cause. Why, in peptide T-3, alanine should fail to be detectable in the sixth degradation is unknown, although a contributory factor may be the preceding residues of glutamic acid and aspartic acid which tend to degrade somewhat less quantitatively than normal. If the sequence contains two or more successive identical residues, successful identification becomes difficult if the degradation has already proceeded for four or five steps. If a peptide should be contaminated with 10–20% of another peptide, this need not hinder the successful application of the method. Heavy spraying of the chromatograms will reveal the relative amounts of the two spots, and, indeed, the sequence of both may often be followed. In general, we have come to anticipate the successful determination of six to eight residues.

The Validity of the Sequence

Amino Acid Composition of the γ Chain.—Because pure γ chain was unavailable, its amino acid composition had to be determined indirectly. These data are, therefore, to a degree uncertain and in general only rather than in exact detail ascertain what residues must be accounted for in any sequence. Nevertheless, the amino acid compositions as determined and as calculated from the sequence differ in only minor detail. Indeed, the composition from the sequence contains a few more residues than were anticipated from the analyses.

Of the sixteen peptides that are to be expected from tryptic hydrolysis, fourteen have been isolated and analyzed and have had their sequence determined. Peptides T-5 and T-13 have not been isolated as such despite a variety of attempts that have been discussed in some detail. They therefore constitute a gap in the knowledge of the γ chain, and the sum of all residues in the tryptic peptides cannot be compared with the expected sum from the amino acid composition. Nevertheless, it is probable that the sequence of peptides T-5 and T-13, as well as their position in the chain, has been correctly deduced. The evidence that supports this statement requires a more general consideration of the over-all sequence and will be discussed below.

The Asparaginyl and Glutaminyl Residues.—The γ chain contains eight aspartyl, five asparaginyl, eight glutamyl, and four glutaminyl residues. Although the definite placing of amide groups is still somewhat of a problem in protein chemistry, their placement in the γ chain is based on positive evidence of several kinds.

In the first place, where applicable, LAP hydrolysis has given a direct answer as to the presence or absence of asparagine or aspartic acid, glutamine or glutamic acid. LAP hydrolysis has been useful in the study of peptides t-1, T-2, T-4, T-9, O-T-15, etc. though in different ways as has been indicated.

Furthermore, hydrolysis with acetic acid releases aspartic acid but not asparagine as is evident from results not only with the γ chain but also with other proteins. Its use in the study of the γ chain has been mainly to provide peptides for further investigation but the results substantiate other data.

Most importantly, the Edman degradation has permitted the identification of asparagine and glutamine because it identifies positively the residues that have been removed and does not rely on a subtractive or difference procedure. As might be expected, aspartic

[9] One can best judge what spots are present in only minor amount by deliberately spraying the paper heavily with the reagent. The spots are apparent only because of the bleaching of the iodine-starch complex. If the paper is only lightly sprayed, a small spot may bleach almost as completely as a large one. If the paper is heavily sprayed, the small spot cannot bleach completely and its contrast with the large spot is readily apparent.

acid and glutamic acid are not altered by the method, but neither are asparagine and glutamine, so that they appear as their PTH derivatives.

The Sequence as a Whole.—The variety of peptides by which the sequence of the γ chain has been determined in many instances has led to repeated determination of some sequences. However, as has been indicated, two parts of the sequence are subject to some uncertainty. Both these involve the two sections of the chain for which it has not been possible to isolate the tryptic peptide as such. Let us consider first the more definitely linked portions of the sequence. From the N-terminus to the 55th residue, the tryptic peptides are clearly linked together. This sequence includes a portion of peptide T-5. Likewise, residues 56 to 110 are united by many overlapping peptides. The tetrapeptide sequence val-thr-val-leu has been placed in residues 111 to 114. Finally, from residue 115 to the C-terminus an excellent series of peptides has been found.

The main questions about the correctness of the sequence must then be these: Is methionine correctly linked to glycine to form the bond between the 55th and the 56th residues? And is the sequence val-thr-val-leu correctly placed at the 111th to 114th residues? These questions could be answered unequivocally by an amino acid analysis of peptides T-5 or T-13. In the absence of such an analysis, the answers must be somewhat equivocal.

The linkage between methionine and glycine (residues 55 and 56) in T-5 is based first upon the amino acid composition (as noted above under the description of peptide T-5) of an apparent mixture of two peptides which presumably have the sequences ser-ala-ileu-met-gly-aspNH$_2$-pro-lys and ser-ser-ala-ser-ala-ileu-met-gly-aspNH$_2$-pro-lys. Other evidence is indirect and uses a comparison of the γ chain with the β chain. A peptide equivalent to T-5 is readily isolated from the β chain and is equal in length to that which is assumed for T-5. These two peptides have the following sequences:

γ: phe–phe–*asp*–ser–phe–gly–*aspNH$_2$*–leu–ser–*ser*–*ala*–*ser*–ala–*ileu*–met–gly–aspNH$_2$–pro–lys

β: phe–phe–glu–ser–phe–gly–asp–leu–ser–thr–pro–asp–ala–val–met–gly–aspNH$_2$–pro–lys

If the deduced sequence of the γ chain is correct, there are many identities here.

It should be mentioned again that the peptide val-thr-val-leu is in excellent yield in chymotryptic and peptic hydrolysates. It can, of course, be placed as it is now or inserted between residues 55 and 56, that is, between methionine and glycine in T-5. Again, we may compare T-13 with the corresponding peptide from the β chain thus:

γ: leu–leu–gly–aspNH$_2$–val–leu–val–*thr*–val–leu–ala–*ileu*–his–phe–gly–lys

β: leu–leu–gly–aspNH$_2$–val–leu–val–cys–val–leu–ala–his–his–phe–gly–lys

The identities are even more complete than between T-5 and the peptide from the β chain. The portion val-thr-val-leu differs from val-cys-val-leu only in the substitution of threonine for cysteine. It seems, therefore, most reasonable to place val-thr-val-leu as it has been placed. To place it between residues 55 and 56 would require that the β and γ chains be altogether different in sequence between residues 56 and 114. This would seem quite foreign to the type of variation in the β and γ chains that is observed in residues 1 to 55 and 115 to 146 where the sequence of the γ chain is established with some certainty.

Although yields of many tryptic peptides are considerably less than quantitative, this fact does not lessen appreciably the reliability of the sequence. The peptides could be reproducibly isolated from different hydrolysates and different samples. A nearly theoretical yield of peptides is, of course, an indicator of the homogeneity of the protein. In the case of hemoglobin with its many abnormalities, it also is an indicator of the homozygosity of the individual from whom it stemmed. This investigation of the γ chain has of necessity used hemoglobin F from a number of individuals because of the transience of hemoglobin F in the infant. Each sample was used separately and was not pooled with others. The results, therefore, stem from several individual samples from none of which was the complete sequence derived. Nevertheless, throughout this investigation, no evidence of difference in sequence from sample to sample was ever observed.

One must, of course, express the realization that any sequence by degradative methods is only a working hypothesis that cannot be proved short of synthesis. In the case of the sequence of the α, β, and γ chains of hemoglobins A and F, such equivocation has less justification because of the independent investigations of several groups. The probability of significant error in the established sequences of the α and β chains is remote. Agreement in the results of the several groups has, for the most part, been excellent. Some initial differences in sequence have been reported but these have led to reinvestigation and general accord. The sequence of the γ chain, however, is due solely to the present investigation and is more subject to error than the sequence of the α and β chains. Some sequences of the γ chain have been achieved from such a variety of data that error would seem unlikely. Other sequences need to be more firmly established. The proposed sequence of the γ chain is to a degree more certain because of homology between the β and the γ chains. The identities, of course, are far more numerous than the differences. In certain instances in which the differences seemed unusual, they were substantiated by careful reinvestigation. Peptide T-3 is an example of one difference in amino acid composition and four in sequence between equivalent peptides in the β and γ chains. It contains the sequence -glu-asp- in the fourth and fifth residues from the N-terminus of the γ peptide but -asp-glu- in the β peptide. These sequences were readily substantiated. We believe that the proposed sequence of the γ chain is basically correct.

A Comparison of the β and γ Chains and Hemoglobins A and F

Both the β and γ chains contain 146 residues. Because the α^A and α^F chains are identical, hemoglobins A and F contain equal numbers of residues although they differ slightly in molecular weight. It is in the differences in the β and γ chains that the characteristic properties of the two hemoglobins must be sought.

On the basis of the proposed sequences of the γ chain and the β chain (Braunitzer *et al.*, 1961 and personal comm.; Hill and Konigsberg, personal comm.), the seventeen differences in amino acid composition express themselves in thirty-nine differences in sequence. These differences for the most part have been noted earlier (Schroeder *et al.*, 1962b). They need correction only in that residue 80 of the β chain is asparagine and residue 94 of both chains is aspartic acid. The β and γ chains differ in the number of acidic and basic amino acid residues but the over-all balance in each is such that the γ chain in essence has two additional histidyl residues. Thus, hemoglo-

bins A and F differ in charged amino acid residues by four histidyl residues. Hence, at neutral *p*H, they would be essentially identical in charge. Probably the difficulty that is usually experienced in the electrophoretic separation of hemoglobins A and F stems from their virtually identical charge. The ease of chromatographic separation of hemoglobins A and F arises probably from the marked difference in the sequences.

Although an earlier discussion of the differences between the β and γ chain (Schroeder *et al.*, 1962b) could be enlarged upon, to do so would add little to an understanding of the properties of hemoglobins A and F. It is true, for example, that the tryptophan peak of hemoglobin F can now be readily explained. It is probably as much a result of a decrease in the tyrosine content of the γ chain as it is an increase in tryptophan content. As Jones (1961) has pointed out, the spectra of hemoglobins A and H are identical in the region of the tryptophan absorption despite the fact that the tryptophan content of hemoglobin H is greater. Actually, hemoglobins F and H are identical in tryptophan residues. The very apparent absorption by tryptophan in the spectrum of hemoglobin F probably is due to the decreased absorption by tyrosine which tryptophan replaces. Thus, the absorption by tryptophan on the edge of the tyrosine absorption is more prominent.

Through the X-ray studies of Kendrew *et al.* (1961) with sperm whale myoglobin and of Cullis *et al.* (1962) with horse hemoglobin, much has been learned about the three-dimensional structure of these molecules. One can with reasonable assurance conclude that this information is applicable also to human hemoglobins A and F. As has already been pointed out (Schroeder *et al.*, 1962b), most of the differences in sequence between the β and γ chains probably occur in the helical portions. It is to be hoped that as more is learned about hemoglobins A and F it will be possible to correlate the differences in sequence with such properties as alkali denaturation, oxygen affinity, Bohr effect, and the like.

Acknowledgments

We wish to express our thanks to Dr. D. Armstrong, Dr. M. Lazarus, and Dr. T. Perry through whom some samples of cord blood were obtained. Mr. J. T. Cua and Mr. W. D. Fenninger rendered capable assistance in some of the experiments.

References

Allen, D. W., Schroeder, W. A., and Balog, J. (1958), *J. Am. Chem. Soc. 80*, 1628.

Anson, M. L., and Mirsky, A. E. (1930), *J. Gen. Physiol. 13*, 469.

Braunitzer, G., Gehring-Müller, R., Hilschmann, N., Hilse, K., Hobom, G., Rudloff, V., and Wittmann-Liebold, B. (1961), *Z. Physiol. Chem. 325*, 283.

Bromer, W. W., Staub, A., Sinn, L. G., and Behrens, O. K. (1957), *J. Am. Chem. Soc. 79*, 2801.

Chernoff, A. I. (1961), *J. Chromatog. 6*, 252.

Clegg, M. D., and Schroeder, W. A. (1959), *J. Am. Chem. Soc. 81*, 6065.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G., and North, A. C. T. (1962), *Proc. Roy. Soc. (London), Ser. A: 265*, 161.

Edman, P., and Sjöquist, J. (1956), *Acta Chem. Scand. 10*, 1507.

Fraenkel-Conrat, H. (1954), *J. Am. Chem. Soc. 76*, 3606.

Fraenkel-Conrat, H., Harris, J. I., and Levy, A. L. (1955), *Methods Biochem. Analy. 2*, 393.

Goldstein, J., Guidotti, G., Konigsberg, W., and Hill, R. J. (1961), *J. Biol. Chem. 236*, P677.

Green, F. C., and Kay, L. M. (1952), *Anal. Chem. 24*, 726.

Hayashi, H. (1961), *J. Biochem. (Tokyo) 50*, 70.

Hill, R. J. and Konigsberg, W. (1962), *J. Biol. Chem. 237*, 3151.

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

Hunt, J. A. (1959), *Nature 183*, 1373.

Ingram, V. M. (1958), *Biochim. Biophys. Acta 28*, 539.

Jones, R. T. (1961), Ph.D. thesis, California Institute of Technology.

Jones, R. T., and Schroeder, W. A. (1963), *J. Chromatog. 10*, 421.

Jones, R. T., Schroeder, W. A., and Vinograd, J. R. (1959), *J. Am. Chem. Soc. 81*, 4749.

Kendrew, J. C., Watson, H. C., Strandberg, B. E., Dickerson, R. E., Phillips, D. C., and Shore, V. C. (1961), *Nature 190*, 666.

Konigsberg, W., and Hill, R. J. (1962), *J. Biol. Chem. 237*, 3157.

Light, A., Glazer, A. N., and Smith, E. L. (1960), *J. Biol. Chem. 235*, 3159.

Matsuda, G., Schroeder, W. A., Jones, R. T., and Weliky, N. (1960), *Blood 16*, 984.

Naughton, M. A., Sanger, F., Hartley, B. S., and Shaw, D. C. (1960), *Biochem. J. 77*, 149.

Partridge, S. M., and Davis, H. F. (1950), *Nature 165*, 62.

Rhinesmith, H. S., Schroeder, W. A., and Martin, N. (1958), *J. Am. Chem. Soc. 80*, 3358.

Rhinesmith, H. S., Schroeder, W. A., and Pauling, L. (1957), *J. Am. Chem. Soc. 79*, 4682.

Schroeder, W. A., and Honnen, L. R. (1953), *J. Am. Chem. Soc. 75*, 4615.

Schroeder, W. A., Jones, R. T., Cormick, J., and McCalla, K. (1962a), *Anal. Chem. 34*, 1570.

Schroeder, W. A., Jones, R. T., Shelton, J. R., Shelton, J. B., Cormick, J., and McCalla, K. (1961a), *Proc. Nat. Acad. Sci. U. S. 47*, 811.

Schroeder, W. A., Kay, L. M., LeGette, J., Honnen, L., and Green, F. C. (1954), *J. Am. Chem. Soc. 76*, 3556.

Schroeder, W. A., and LeGette, J. (1953), *J. Am. Chem. Soc. 75*, 4612.

Schroeder, W. A., and Matsuda, G. (1958), *J. Am. Chem. Soc. 80*, 1521.

Schroeder, W. A., Shelton, J. R., and Shelton, J. B. (1961b), *Anal. Biochem. 2*, 87.

Schroeder, W. A., Shelton, J. R., Shelton, J. B., and Cormick, J. (1962b), *Proc. Nat. Acad. Sci. U. S. 48*, 284.

Schroeder, W. A., Shelton, J. R., Shelton, J. B., and Cormick, J. (1963), *Biochemistry 2* (in press).

Shelton, J. R., and Schroeder, W. A. (1960), *J. Am. Chem. Soc. 82*, 3342.

Sjöquist, J. (1953), *Acta Chem. Scand. 7*, 447.

Smyth, D. G., Stein, W. H., and Moore, S. (1963), *J. Biol. Chem. 238*, 227.

Spackman, D. H., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30*, 1190.

Vinograd, J., and Hutchinson, W. D. (1960), *Nature 187*, 216.

Wilson, S., and Smith, D. B. (1959), *Can. J. Biochem. Physiol. 37*, 405.

Witkop, B. (1961), *Advan. Protein Chem. 16*, 221.