Sandermann, H., Jr., Tisue, G. T., and Grisebach, H. (1968), *Biochim. Biophys. Acta 165*, 550.
Schabel, F. M., Jr. (1968), *Chemotherapy 13*, 321.
Schardein, J. L., and Sidwell, R. W. (1968), *Antimicrob. Ag. Chemotherap.*, 158.
Seto, H., Otake, N., and Yonehara, H. (1968), *Agr. Biol. Chem. (Tokyo) 32*, 1299.
Sidwell, R. W., Arnett, G., and Schabel, F. M., Jr. (1969), *Progr. Antimicrob. Anticancer Chemother., Proc. Int. Congr. Chemother., 6th Meeting, 2*, 44.
Sidwell, R. W., Dixon, G. J., Schabel, F. M., Jr., and Kaump, D. H. (1968), *Antimicrob. Ag. Chemother.*, 148.
Sloan, B. J., Miller, F. A., Ehrlich, J., McLean, I. W., Jr., and Machamer, H. E. (1968), *Antimicrob. Ag. Chemother.*, 161.
Suhadolnik, R. J. (1970), Nucleoside Antibiotics, New York, N. Y., John Wiley & Sons, Inc.
Suhadolnik, R. J., and Chock, S. P. (1971), 162nd National Meeting of the American Chemical Society, Washington, D. C.
Suhadolnik, R. J., Chock, S. P., and Uematsu, T. (1971), 162nd National Meeting of the American Chemical Society, Washington, D. C.
Suhadolnik, R. J., and Farmer, P. B. (1971), 162nd National Meeting of the American Chemical Society, Washington, D. C.
Suhadolnik, R. J., Weinbaum, G., and Meloche, H. P. (1964), *J. Amer. Chem. Soc. 86*, 948.
Tono, H., and Cohen, S. S. (1962), *J. Biol. Chem. 237*, 1271.
Uematsu, T., and Suhadolnik, R. J. (1972), *Biochemistry* (in press).
Volk, W. A. (1959), *J. Biol. Chem. 234*, 1931.

Phylogeny of Hemoglobins. β Chain of Frog (*Rana esculenta*) Hemoglobin*

Jean-Pierre Chauvet and Roger Acher†

ABSTRACT: The amino acid sequence of the β chain of frog hemoglobin (*Rana esculenta*) has been studied. The five fragments produced by tryptic cleavage of the trifluoroacetylated, carboxymethylated β chain were first isolated. After removal of the trifluoroacetyl groups and further cleavage with trypsin, 15 peptides were purified and sequenced. The order of the tryptic peptides in each fragment was determined through chymotryptic peptides and the alignment of the fragments was established by isolating arginine-containing peptides from a chymotryptic digest of the β chain. The frog β chain comprises 140 residues. When the 26 known mammalian β, δ, and γ chains and frog β chain are compared, 54 positions out of 146 are invariant in all the proteins. Of 47 amino acids involved in the interactions of β chain with either α chains or heme group, 30 are invariant. Two long sequences (28–40 and 96–108) seem particularly stable. The comparison of β chains from eutherians, metatherians (kangaroo), and amphibians (frog), which have diverged approximately 80, 130, and 320 million years ago, respectively, reveals that the number of amino acid substitutions is dependent but not proportional to time. These results are confronted with current concepts of evolution.

Among the proteins which were chosen for a study on molecular evolution, the hemoglobin family is one of the most attractive. Because hemoglobin is particularly abundant and easy to purify from red cells, the protein seems amenable to structural investigations in almost all vertebrates. On the other hand, the "molecule" is generally built with polypeptide chains of 140–150 residues in length and determination of the complete amino acid sequence can be performed under rather good conditions with the current techniques of protein chemistry. However, separation of α and β chains turned out to be more difficult for lower vertebrates than for mammals and probably for this reason our knowledge on hemoglobins of lower vertebrates is very limited since to the present time only the α chains of the chicken (Matsuda *et al.*, 1970) and of the carp (Hilse *et al.*, 1966) have fully been sequenced.

The choice of hemoglobin as an evolutionary tracer is not only determined by practical reasons. In contrast to most enzymes, this protein takes in account a whole physiological function, the oxygen transport, and hence is directly subjected to selective pressure; the so-called phenotypic character on which selection acts is apparently related to a few structural genes and not to many genes as in a polymolecular function (Simpson, 1964). In another point of view, the vertebrates, except fishes, used dissolved oxygen in the first part of their life and aerial oxygen in the second part and the switch might be associated with a molecular change. This change has actually

* From the Laboratory of Biological Chemistry, University of Paris VI, 96, Boulevard Raspail, Paris (6e), France. *Received June 21, 1971.* This study was supported in part by a grant from the Délégation à la Recherche Scientifique et Technique (66-00-146). Maps of chymotryptic peptides from tryptic fragments, tables of tryptic peptides subjected to Edman degradation and to chymotryptic hydrolysis, and tables of amino acid compositions and of yields of chymotryptic from tryptic fragments and from β chain will appear following these pages in the microfilm edition of this volume of the journal. Singles copies may be obtained from the Business Operations Office, Books and Journals Division, American Chemical Society, 1155 Sixteenth St., N.W., Washington, D. C. 20036, by referring to code number BIO-11-916. Remit check or money order for $3.00 for photocopy or $2.00 for microfiche.

† To whom inquiries should be directed.

been detected in mammals, birds, and amphibians and to what extent it can reflect at the biochemical level the recapitulation law of Haeckel deserves special attention.

On the phylogenetic scale, amphibians are located between aquatic and terrestrial vertebrates. Because of this position, amphibian species seem of particular interest for the search of transition molecules. The example of neurohypophysial hormones is rather suggestive. Isotocin is present in bony fishes but mestocin appears in amphibians and remains in reptiles and birds (Acher *et al.*, 1970). The comparison of amphibian proteins to those of fishes, on one hand, and those of reptiles on the other, might reveal significant changes. Furthermore in the case of hemoglobin, it is known that during the metamorphosis of tadpoles, the tadpole protein is replaced by an adult protein (Moss and Ingram, 1968) and the comparison of the first to fish hemoglobins could be very instructive.

The present work is devoted to *Rana esculenta* hemoglobin. *R. esculenta* is a current European frog which is used for food and the collection of material in large amounts is relatively easy. The purification of the major component of *R. esculenta* hemoglobin as well as the separation of α and β chain has been described elsewhere (Chauvet and Acher, 1967b, 1968, 1971). The amino acid sequence of β chain has been determined and this paper gives a detailed description of the work, the results of which have been briefly reported (Chauvet and Acher, 1970).

Material and Methods

β Chain and Derivatives. Hemoglobin from adult frog *Rana esculenta* was prepared as previously described by fractionated precipitation with ammonium sulfate and chromatography on carboxymethyl-Sephadex C-50. The chromatography resolved hemoglobin into two components which have apparently similar amino acid compositions but different electrophoretic mobilities (Chauvet and Acher, 1967b, 1971). The major component (hemoglobin II) was used for this study. Reduced globin was subjected to countercurrent distribution and α and β chains were separated with a yield of 60–70%. From 100 ml of frog blood, about 1.0–1.5 g (60–100 μmole) of each chain can be obtained. The purity of the material was checked by starch gel electrophoresis and by N-terminal analysis (Chauvet and Acher, 1968, 1971).

Carboxymethylation of the SH groups was performed according to Crestfield *et al.* (1963). β chain (100 mg) is dissolved in 10 ml of 0.5 M Tris-HCl buffer (pH 8.5) containing 8 M urea and 20 mg of EDTA. β-Mercaptoethanol (0.1 ml; Fluka) is added. After 4 hr, 268 mg of iodoacetic acid (BDH) dissolved in 1 ml of 1 N NaOH is added and the pH is maintained at 8.5 for 30 min by addition of 1 N sodium hydroxide with a Radiometer pH-Stat. Aminoethylation was carried out by adding 0.3 ml of ethylenimine (Fluka) in the place of iodoacetic acid, the pH being fixed at 8.5 for 60 mn by addition of 1 N HCl according to Raftery and Cole (1963). The pH is lowered to 3.0 with 1 N HCl and carboxymethylated or aminoethylated β chain is dialyzed for removing salts and urea, precipitated by acetone, redissolved in 10 ml of water, and lyophilized (yield for carboxymethylated β chain, 99 mg; for aminoethylated β chain, 96 mg).

Trifluoroacetylation of carboxymethylated β chain was performed according to Goldberger and Anfinsen (1962). Carboxymethylated β chain (100 mg) is dissolved in 10 ml of water and the pH is brought to 9.95 with 2 N KOH. *S*-Ethyl trifluorothioacetate (0.5 ml) is added and the pH is maintained at 9.95 for 90 min with 2 N KOH by using a Radiometer pH-Stat. The pH is then lowered to 2.5 with 12 N HCl and the derivative precipitates. The product is recovered by centrifuging, washed three times with 20 ml of 0.01 N HCl, suspended in 10 ml of water, and lyophilized (yield 103 mg). For removing the trifluoroacetyl groups, the derivatives are treated with 1 M piperidine (pH 12.4) at 5° for 2 hr (Goldberger and Anfinsen, 1962).

Electrophoresis. Starch gel electrophoresis is used for checking the homogeneity of β chain. Gel is prepared with a formate buffer (pH 1.9) (Muller, 1960) and staining is carried out with Amido Schwarz 10B (Merck). Paper electrophoresis is employed for checking the purity of large peptide fragments; Whatmann No. 3MM paper in a buffer pH 3.7 (pyridine–acetic acid–water, 1:10:298, v/v) is used and staining is carried out with 0.1% ninhydrin in alcohol.

Enzymatic Cleavages. β chain is split by trypsin or chymotrypsin in order to obtain fragments for amino acid sequence determination. Trypsin (twice crystallized Worthington, batch TRSF 6144-5) (EC 3.4.4.4) is used with an enzyme: substrate weight ratio 1:100 for 3 hr in 0.1 M ammonium bicarbonate (pH 8.0) at 37°. Two equal fractions of the enzyme are added at times 0 and 90 min. Selective cleavages at arginine residues are obtained by using trifluoroacetylated carboxymethylated β chain. Chymotrypsin (three-times crystallized Worthington, batch DCI 6078) (EC 3.4.4.5) is employed with an enzyme:substrate weight ratio 2:100 for 3 hr in 0.1 M ammonium bicarbonate (pH 8.0) at 37°.

Purification of Peptide Fragments. For fractionating the mixture of the large fragments produced by trypsin hydrolysis of the trifluoroacetylated carboxymethylated β chain, gel filtration on columns of either Sephadex G-50 (4 × 100 cm) or Sephadex G-25 (2 × 90 cm) is used with 0.1 M acetic acid. The flow (50 and 15 ml per hr) is fixed with a Milton Roy pump. Fractions of 5 ml are collected and peptides are detected by absorbance at 232 and 280 nm with a Beckman DU spectrophotometer. When two components are not resolved by gel filtration on Sephadex G-50, the separation is completed either by another gel filtration on Sephadex G-25 or by paper electrophoresis.

Isolation of tryptic or chymotryptic peptides is carried out by using the fingerprinting technique (Katz *et al.*, 1959) under the conditions described by Baglioni (1961). High-voltage electrophoresis at 35 V/cm on Whatman No. 3MM paper with a pyridine–acetate buffer (pH 6.4; pyridine–acetic acid–water 10:0.4:90, v/v) in cooled tanks and descending chromatography in an isoamyl alcohol–pyridine–water solvent (30:30:35, v/v) are performed. Staining is carried out with ninhydrin (0.1% in alcohol) or with reagents specific for arginine (Acher and Crocker, 1952), tyrosine (Acher and Crocker, 1952), histidine (Sanger and Tuppy, 1951), and tryptophan (Smith, 1953). For preparative purposes, staining is performed with dilute ninhydrin (0.01% in alcohol); the spots are cut, washed with ether, and eluted with 0.1 M acetic acid. The materials are kept frozen at −20° and used either for amino acid analysis or structural studies.

Amino Acid Analysis and Sequence Techniques. Aliquots of peptides (100–200 nmoles) are hydrolyzed in evacuated, sealed tubes with 6 N HCl at 110° for 24, 48, or 72 hr. After removal of HCl by evaporation in a desiccator with sodium hydroxide, hydrolysates are analyzed according to Spackman *et al.* (1958) on a Spinco automatic analyzer Model 120B fitted with a high-sensitivity colorimeter. N-Terminal sequences are determined by using the dinitrofluorobenzene technique of Sanger (Chauvet *et al.*, 1966) and the stepwise degradation procedure of Edman in the paper strip variant

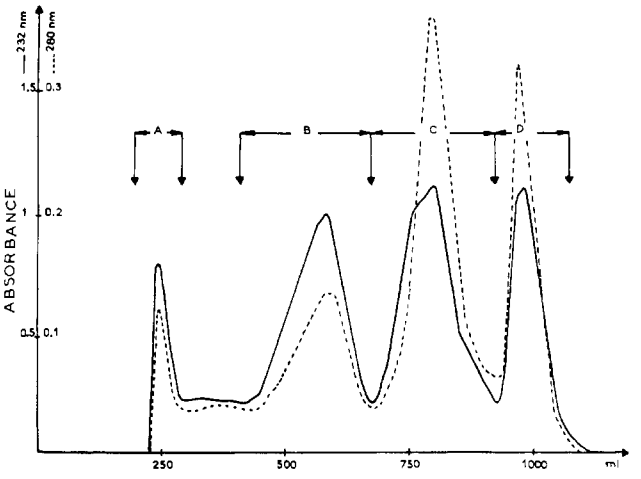

FIGURE 1: Fractionation of the tryptic hydrolysate of the trifluoroacetylated carboxymethylated β chain on a Sephadex G-50 column (after removal of trifluoroacetyl groups; details in the text).

(Fraenkel-Conrat, 1954; Schroeder *et al.*, 1963) with slight modifications. Reaction with phenyl isothiocyanate (500 nmoles of peptide treated at 40° with 0.2 ml of 20% phenyl isothiocyanate in pyridine) is performed twice, respectively, for 4 and 2 hr, the paper being washed three times with benzene between the two steps. The cleavage of phenylthiocarbamyl peptide is carried out in a desiccator by exposing the paper strip to vapors of 6 N HCl and glacial acetic acid in separate beakers for 8 hr under a pressure of 100 mm; phenylthiohydantoin amino acids are extracted with acetone. Dinitrophenyl and phenylthiohydantoin derivatives are identified by thin-layer chromatography (Brenner *et al.*, 1961; Bailey, 1967).

The C-terminal sequence of β chain is determined by hydrazinolysis (Niu and Fraenkel-Conrat, 1955) and by using carboxypeptidases A and B (Fraenkel-Conrat *et al.*, 1955).

Results

Purification of Trifluoroacetyl Fragments. Because frog β chain contains 10 residues of lysine and 4 residues of arginine-

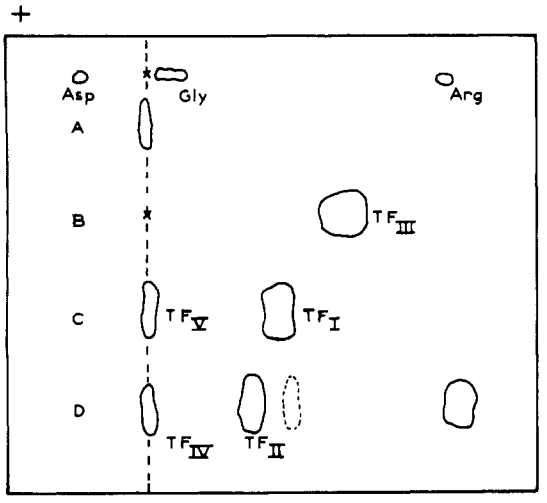

FIGURE 2: Paper electrophoresis of fractions A–D, obtained by gel filtration (lysine residues of the fragments are unblocked; details in the text).

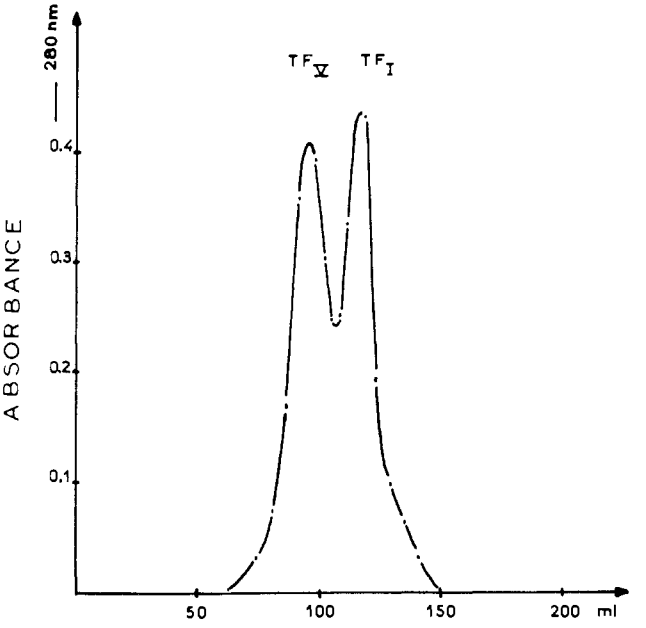

FIGURE 3: Separation of the fragments TF_I and TF_V on a Sephadex G-25 column.

tryptic hydrolysis gives 15 peptides which can be separated by the fingerprinting technique. However a "core" precipitates during the tryptic digest of carboxymethylated β chain and the separation of the tryptic units is not very satisfactory. The blocking of lysine residues by trifluoroacetyl groups reduced the number of trypsin cleavages to four and allow the obtaining of five trifluoroacetyl fragments.

β chain, carboxymethylated and trifluoroacetylated as described in Material and Methods, is subjected to trypsin hydrolysis for 3 hr. Because of the poor solubility of the protected peptides at pH 9, trifluoroacetyl groups are removed before fractionation. Hydrolyzed carboxymethylated trifluoro acetylated β chain (100 mg) is treated with 5 ml of 1 M piperi

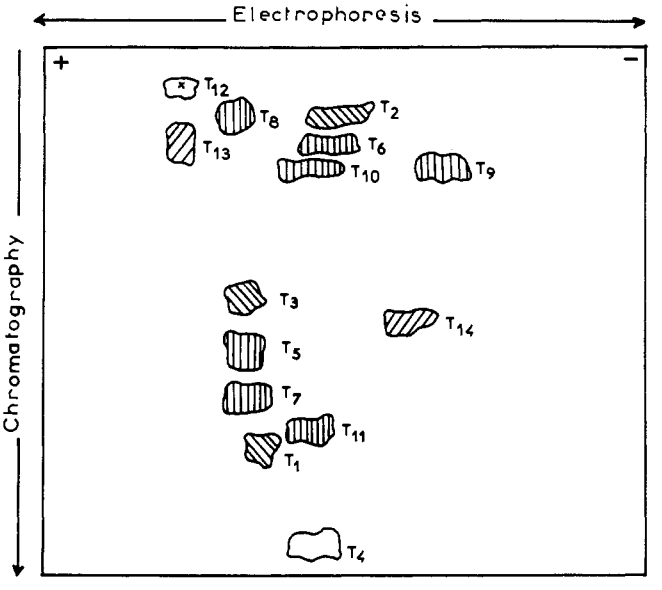

FIGURE 4: Tryptic peptide map of β chain. ▨, peptides from fragment TF_I; ▭, $TF_{II} = T_4$; ▥, peptides from fragment TF_{III}; ⊠, $TF_{IV} = T_{12}$; ▨, peptides from fragment TF_V.

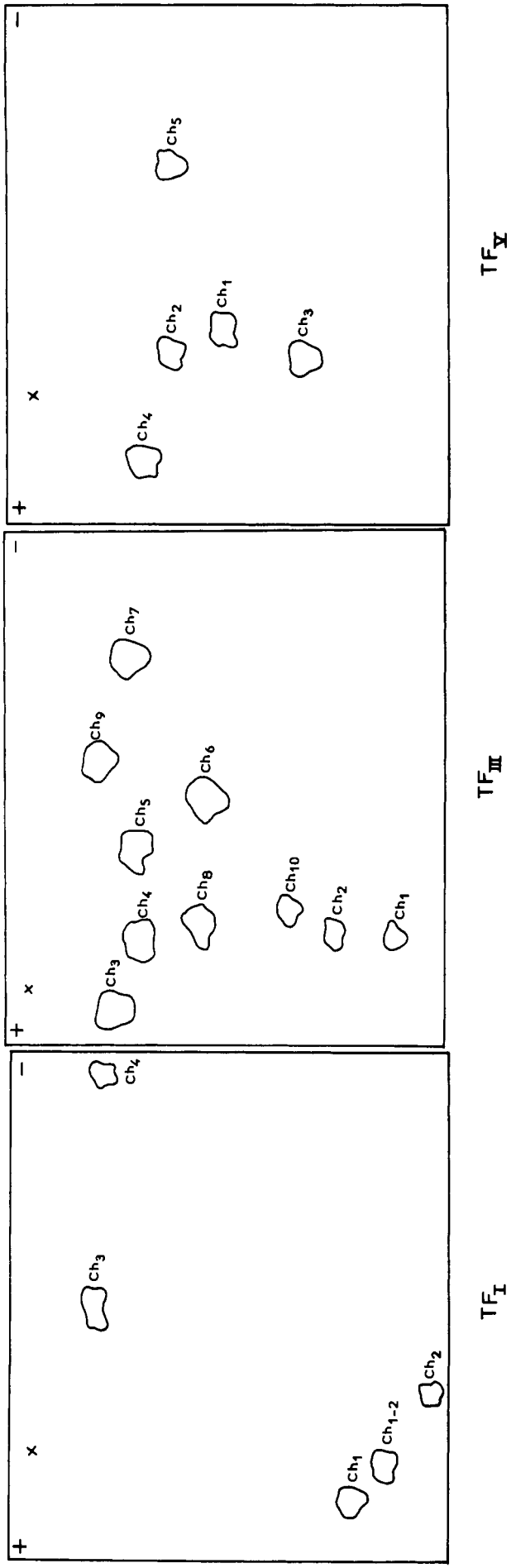

FIGURE 5: Chymotryptic peptide maps of fragments TF_I, TF_{III}, and TF_V respectively.

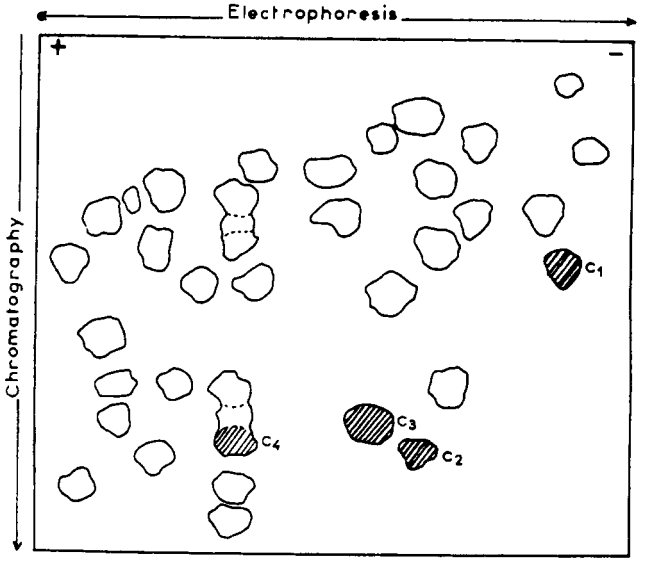

FIGURE 6: Chymotryptic peptide map of β chain. ▨, arginine-containing peptides.

dine for 2 hr at 5°. The pH is brought to 3.6 by addition of acetic acid and the mixture is filtered on a Sephadex G-50 column (4 × 100 cm) in 0.1 M acetic acid. After washing with about 1200 ml of acetic acid, four peaks are detected by absorbance at 232 and 280 nm (Figure 1). Four fractions are collected as shown by the arrows, concentrated in a rotary evaporator, and lyophilized. The weights of fractions A, B, C, D are, respectively, 13, 30, 28, and 81 mg, the fourth fraction containing probably some salts. Each fraction is examined by electrophoresis on paper (Figure 2). Fraction A is probably the intact chain because its amino acid composition is very close to that of β chain. Fraction B seems homogeneous at pH 3.7; this fragment will be called TF_{III} because of its position in the sequence of β chain. Fraction C contains two components TF_I and TF_V which are separated by another filtration on a Sephadex G-25 column (Figure 3). Fraction D also contains two components TF_{II} and TF_{IV}; they can be separated either by paper electrophoresis at pH 3.7 or by selective extraction: the freeze-dried material is suspended in 0.5 M acetic acid (20 mg/ml); TF_{IV} is insoluble and TF_{II} is further purified by paper electrophoresis at pH 3.7. The amino acid compositions of the five fragments are given in Table I.

Amino Acid Sequence of Fragment TF_I. This peptide contains 24 amino acid residues, including 2 lysine and 1 arginine residues (Table I). The presence of tryptophan is detected by specific reaction. The N-terminal residue is glycine as for the intact β chain (Chauvet and Acher, 1968) and because none of the four other trifluoroacetyl peptides have a N-terminal glycine, this fragment is obviously N terminal in the sequence. After removal of trifluoroacetyl groups, trypsin splits TF_I into three peptides T_1, T_2, and T_3 which are separated by the fingerprinting technique (Figure 4) and which contain 11, 5, and 8 residues, respectively (Table I). The Edman degradation permits determination of the complete sequence of T_2 and T_3 and the sequence of eight residues in T_1 as shown in Table II. Chymotryptic hydrolysis of T_1 yields four products which are separated by the fingerprinting technique and analyzed. T_1-Ch_3 gives the position of the ninth residue (Trp) and T_1Ch_4 the sequence of the last two residues of T_1 (Gly-Lys). For T_3 the analysis of chymotryptic peptides confirms the amino acid sequence (Table II).

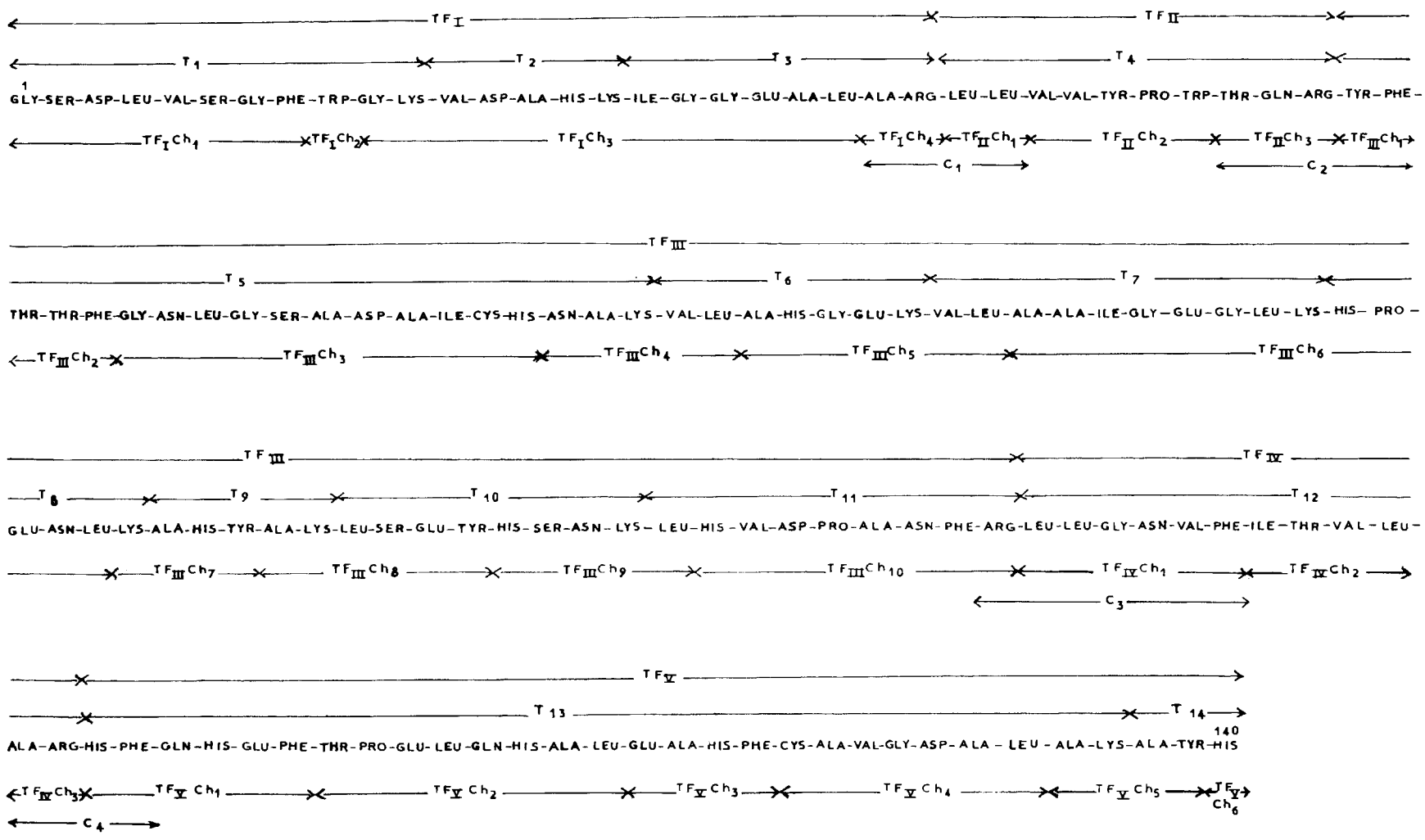

FIGURE 7: Alignment of the residues of β chain. TF, tryptic fragments from trifluoroacetylated β chain; TF Ch, chymotryptic peptides from fragments TF; T, tryptic peptides from β chain; C, chymotryptic peptides from β chain.

TABLE I: Amino Acid Compositions of the Trifluoroacetylated Fragments and of Their Constitutive Tryptic Units.[a]

	TF_I		T_1		T_2		T_3	TF_{II} T_4	TF_{III}		T_5		T_6		T_7		T_8		T_9		T_{10}		T_{11}	TF_{IV} T_{12}	TF_V		T_{13}		T_{14}
Lys	2.05	2	1.00	1	1.00	1			5.76	6	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1			1.12	1	1.00	1	
His	1.24	1			0.99	1			6.24	6	0.74	1	0.95	1			0.68	1	0.93	1	0.95	1	1.09	1	4.63	5	3.70	4	1.12	1

Arg	1.00 1			1.00 1	1.00 1	1.00 1							1.00 1	1.00 1			
Asp	2.25 2	1.00 1	0.95 1			6.86 7	2.98 3			1.13 1		0.76 1	2.30 2	1.20 1	1.33 1	1.25 1	
Thr					0.91 1	2.05 2	1.40 2							0.89 1	1.23 1	0.94 1	
Ser	1.82 2	1.72 2				3.36 3	1.10 1					1.34 2					
Glu	1.25 1			0.99 1	0.86 1	4.00 4		1.15 1	1.05 1	1.10 1		0.96 1			4.81 5	4.70 5	
Pro					0.79 1	2.05 2				0.95 1			1.18 1		1.12 1	0.78 1	
Gly	5.25 5	3.15 3		2.03 2		4.95 5	1.71 2	1.10 1	1.96 2					0.99 1	0.95 1	0.95 1	
Ala	2.98 3		0.89 1	2.01 2		8.85 9	2.40 3	1.06 1	2.12 2		1.92 2		1.32 1	0.92 1	5.65 6	4.60 5	1.00 1
Cys						0.57 1	0.53 1								0.33 1	0.35 1	
Val	1.87 2	0.82 1	0.93 1		1.52 2	2.67 3		0.79 1	1.08 1				0.96 1	1.76 2	0.90 1	0.89 1	
Met																	
Ile	0.84 1			0.98 1		1.63 2	0.69 1		0.66 1					0.78 1			
Leu	2.13 2	1.06 1		1.18 1	2.19 2	6.65 7	1.15 1	0.95 1	1.86 2	1.10 1		0.82 1	1.02 1	3.10 3	3.00 3	3.02 3	
Tyr					1.09 1	2.65 3	0.70 1				0.54 1	0.59 1			0.67 1		0.83 1
Phe	1.11 1	0.92 1				2.95 3	1.73 2						1.30 1	1.08 1	2.65 3	3.16 3	
Trp	1	1			1												
Total	24	11	5	8	10	64	19	7	10	6	5	8	9	12	30	27	3

[a] TF, tryptic fragments obtained from trifluoroacetylated β chain; T, tryptic units isolated from TF fragments and from β chain. For each peptide, the left-hand column gives experimental results for a 24 hr-hydrolysis (the values given are molar ratios, arginine or lysine being taken as reference, except for TF_V and T_{14}; values below 0.3 residue have been omitted); the right-hand column indicates the integral values. Cysteine is determined as carboxymethylated cysteine and tryptophan is detected by specific color reaction.

The alignment of T_1, T_2, and T_3 in the fragment TF_I can be deduced from the N-terminal glycine residue of T_1 and the C-terminal arginine residue of T_3. Confirmation is obtained by chymotrypsin hydrolysis of the fragment TF_I. Four products are isolated (Figure 5) and have the following amino acid compositions, TF_ICH_1: Gly, Ser, Asp, Leu, Val, Ser, Gly, Phe; TF_ICH_2: Trp; TF_ICh_3: Gly, Lys, Val, Asp, Ala, His, Lys, Ile, Gly, Gly, Glu, Ala, Leu; and TF_ICH_4: Ala, Arg. The results are summarized in Table II.

Amino Acid Sequence of Fragment TF_{II}. The peptide contains ten residues one of which is arginine and no lysine. It is identical with a peptide T_4 isolated from the digest of carboxymethylated β chain (Figure 4). It contains a tryptophan residue as shown by the specific reagent. The Edman degradation permitted the determination of the sequence of this peptide (Table II). Chymotryptic hydrolysis gives three products which are analyzed. The amino acid compositions of T_4Ch_1, T_4Ch_2, and T_4Ch_3 are in accord with the results of the Edman technique (Table II).

Amino Acid Sequence of Fragment TF_{III}. This fragment contains 64 residues, 6 of which are lysine and 1, arginine. The N-terminal amino acid is tyrosine determined by Edman procedure and the C-terminal residue is arginine because of the blocking of lysines. After unblocking, trypsin splits the polypeptide at the level of the six lysines and gives seven peptides called T_5 to T_{11} which can be separated by the fingerprinting technique (Figure 4). The amino acid compositions are shown in Table I, the tryptic peptides containing 19, 7, 10, 6, 5, 8, and 9 residues, respectively. The complete sequences of all these peptides can be determined by Edman degradation except for T_5 in which the sequence of the first seven residues is established.

T_5 is split by chymotrypsin into four peptides T_5Ch_1, T_5Ch_2, T_5Ch_3, and T_5Ch_4 containing 2, 3, 11, and 3 residues. T_5Ch_1 and T_5Ch_2 confirm the N-terminal sequence of T_5, T_5Ch_3 studied by Edman degradation gives the central sequence of 11 residues, and T_5Ch_4 gives the C-terminal sequence of 3 residues as shown in Table II. The structure of T_5 is confirmed with aminoethylated β chain. The aminoethylcysteine residue in 16th position allows an additional tryptic cleavage and at the place of T_5, two peptides T_5AE_1 and T_5AE_2 with 15 and 4 residues, respectively, are found in the digest (Table II).

The seven peptides (T_5–T_{11}) are subjected to chymotryptic hydrolysis for confirming the sequences. T_8 and T_{11} are not split. The others give fragments which are analyzed; the results, shown in Table II, are in agreement with those obtained with Edman degradation.

The alignment of the seven tryptic peptides is determined through chymotryptic peptides of TF_{III}. T_5 is obviously the N-terminal unit because it is the sole tryptic peptide having a N-terminal tyrosine as TF_{III}. T_{11} is the C-terminal unit because of its C-terminal arginine. Chymotryptic digest of TF_{III} gives ten peptides called $TF_{III}Ch_1$ to $TF_{III}Ch_{10}$ (Figure 5), the amino acid compositions of which are determined. These compositions allow the ordering of the seven tryptic units of TF_{III} (*cf.* Figure 7).

Amino Acid Sequence of Fragment TF_{IV}. This fragment has twelve residues, one of which is arginine, and no lysine. It is identical with a peptide found in the tryptic digest of carboxymethylated β chain and called T_{12} because of its position in the chain. This peptide has a poor solubility in solvents used for paper chromatography or paper electrophoresis and remains at the starting spot (Figure 4). The N-terminal sequence determined by Edman procedure is Leu-Leu-Gly-Asn-Val-Phe-Ile-Thr. Chymotryptic hydrolysis of T_{12} yields

TABLE II: Sequence of the Tryptic Units of β Chain.[a]

Peptide		Sequence	No. of Residues
		1 11	
TF_I	T_1	Gly-Ser-Asp-Leu-Val-Ser-Gly-Phe-Trp-Gly-Lys	11
		T_1Ch_1 T_1Ch_2 T_1Ch_4	
		T_1Ch_3	
		12 16	
	T_2	Val-Asp-Ala-His-Lys	5
		17 24	
	T_3	Ile-Gly-Gly-Glu-Ala-Leu-Ala-Arg	8
		T_3Ch_1 T_3Ch_2	
		25 34	
TF_{II}	T_4	Leu-Leu-Val-Val-Tyr-Pro-Trp-Thr-Gln-Arg	10
		T_4Ch_1 T_4Ch_2 T_4Ch_3	
		35 53	
TF_{III}	T_5	Tyr-Phe-Thr-Thr-Phe-Gly-Asn-Leu-Gly-Ser-Ala-Asp-Ala-Ile-Cys-His-Asn-Ala-Lys	19
		T_5Ch_1 T_5Ch_2 T_5Ch_3 T_5Ch_4	
		T_5AE_1 T_5AE_2	
		54 60	
	T_6	Val-Leu-Ala-His-Gly-Glu-Lys	7
		T_6Ch_1 T_6Ch_2	
		61 70	
	T_7	Val-Leu-Ala-Ala-Ile-Gly-Glu-Gly-Leu-Lys	10
		T_7Ch_1 T_7Ch_2	
		71 76	
	T_8	His-Pro-Glu-Asn-Leu-Lys	6
		77 81	
	T_9	Ala-His-Tyr-Ala-Lys	5
		T_9Ch_1 T_9Ch_2	
		82 89	
	T_{10}	Leu-Ser-Glu-Tyr-His-Ser-Asn-Lys	8
		$T_{10}Ch_1$ $T_{10}Ch_2$	
		90 98	
	T_{11}	Leu-His-Val-Asp-Pro-Ala-Asn-Phe-Arg	9
		99 110	
TF_{IV}	T_{12}	Leu-Leu-Gly-Asn-Val-Phe-Ile-Thr-Val-Leu-Ala-Arg	12
		$T_{12}Ch_1$ $T_{12}Ch_2$ $T_{12}Ch_3$	
		111	
TF_V	T_{13}	His-Phe-Gln-His-Glu-Phe-Thr-Pro-Glu-Leu-Gln-His-Ala-Leu-Glu-Ala-His-Phe-Cys-Ala-	27
		$T_{13}Ch_1$ $T_{13}Ch_2$ $T_{13}Ch_3$	
		$T_{13}AE_1$	
		137	
		Val-Gly-Asp-Ala-Leu-Ala-Lys	
		$T_{13}Ch_4$	
		$T_{13}AE_2$	
		138 140	
	T_{14}	Ala-Tyr-His	3

[a] TF, tryptic fragments of trifluoroacetylated β chain; T, tryptic units isolated from fragments TF and from β chain; TCh, chymotryptic peptides isolated from a tryptic unit (T_1Ch_1, T_1Ch_2, etc.). →, sequence determined by the Edman method: arrow above the sequence, degradation performed on the tryptic unit; arrow under the sequence, degradation performed on chymotryptic peptides of the tryptic unit.

three peptides, $T_{12}Ch_1$, $T_{12}Ch_2$, and $T_{12}Ch_3$. $T_{12}Ch_1$ corresponds to the first six residues. $T_{12}Ch_2$ gives the sequence Ile-Thr-Val-Leu and $T_{12}Ch_3$ gives the C-terminal sequence Ala-Arg. The structure of TF_{IV} is shown in Table II.

Amino Acid Sequence of Fragment TF_V. This fragment contains 30 residues, one of which is lysine, and no arginine. It is the C-terminal trifluoroacetylated fragment. After unblocking of lysine, trypsin splits TF_V into two peptides, T_{13} which has 27 residues and T_{14} which has 3 residues and is the C-terminal peptide because it has no lysine and no arginine. These peptides are purified by the fingerprinting technique (Figure 4).

The first ten residues of T_{13} are located by Edman degradation (Table II). Chymotryptic digestion of T_{13} gives 4 peptides, $T_{13}Ch_1$, $T_{13}Ch_2$, $T_{13}Ch_3$, and $T_{13}Ch_4$. $T_{13}Ch_1$ is the N-terminal peptide and the sequence of the three others is determined by Edman procedure (Table II). The alignment of $T_{13}Ch_1$ and $T_{13}Ch_2$ can be deduced from the N-terminal sequence of T_{13}, and C-terminal position of $T_{13}Ch_4$ is deduced from the C-terminal lysine. When tryptic digestion is performed on aminoethylated β chain, at the place of T_{13} two peptides, $T_{13}AE_1$ and $T_{13}AE_2$, with 19 and 8 residues, respectively, are found and the amino acid compositions confirm the structure of T_{13} and the position of the cysteine residue.

T_4 has 3 residues and its structure is determined by Edman degradation. The alignment of T_{13}–T_{14} can be deduced from the C-terminal residues and is confirmed by chymotryptic digestion of TF_V and isolation of a chymotryptic peptide TF_VCh_5 (Ala, Lys, Ala, Tyr). The sequence of TF_V is shown in Table II.

Alignment of the Trifluoroacetylated Fragments. The position of TF_I is deduced from its N-terminal glycine, identical to the N-terminal residue of the β chain, the other trifluoroacetylated fragments having Leu, Tyr, Leu, and His as N-terminal residues, respectively. The C-terminal position of TF_V is deduced from its C-terminal sequence Ala-Tyr-His, identical to the C-terminal sequence of β chain (Chauvet and Acher, 1971).

The alignment of the three other fragments is determined by isolating arginine peptides in chymotryptic digest of β chain. Chymotryptic hydrolysis of β chain gives many peptides (Figure 6); among them the Sakaguchi reaction reveals four peptides containing arginine. These peptides are C_1, Ala-Arg-Leu-Leu; C_2, Thr-Gln-Arg-Tyr-Phe-; C_3, Arg-Leu-Leu-Gly-Asn-Val-Phe; C_4, Ala-Arg-His-Phe. The peptide C_4 could be interpreted for an alignment TF_I–TF_V or TF_{IV}–TF_V because TF_I and TF_{IV} have the same C-terminal sequence Ala-Arg. Because TF_V is obviously the C-terminal fragment of β chain and TF_I the N-terminal fragment, it is evident that the right alignment is TF_{IV}–TF_V. C_2 gives the alignment TF_{II}–TF_{III} without ambiguity so that the positions of the five fragments can be deduced. C_1 confirms the alignment TF_I–TF_{II}. Figure 7 shows the results.

Discussion

Methodology in Sequence Determination. The reversible blocking of lysine residues and subsequent tryptic cleavages at arginine residues were the best way for obtaining a few large fragments from β chain because this protein has only four arginines and the absence of methionine excludes the classical split with cyanogen bromide. The trifluoroacetyl blocking, previously used for the determination of amino acid sequences in glyceraldehyde-3-phosphate dehydrogenase (Davidson *et al.*, 1967), pepsin (Perham and Jones, 1967), and parathyroid hormone (Brewer and Ronan, 1970), was very satisfactory for β chain, the protected fragments being soluble at alkaline pHs. However, the citraconyl blocking (Dixon and Perham, 1968), now used in the laboratory, seems somewhat more convenient.

Because β chain contains ten lysines and four arginines, tryptic cleavage gives peptides having usually around ten residues, the sequence of which is often entirely determined by Edman technique. Further cleavage of tryptic units with chymotrypsin is used for confirmation or for obtaining shorter peptides which are sequenced by Edman procedure (*cf.* T_5, T_{12}, and T_{13}). The tryptic units are ordered in trifluoroacetyl fragments by isolating chymotryptic peptides from the latter, and trifluoroacetyl fragments themselves are ordered through chymotryptic peptides from β chain. Because β chain has many residues which are suitable for chymotrypsin cleavage (2 tryptophans, 5 tyrosines, 7 phenylalanines, 17 leucines), the wanted overlapping peptides were found and no additional enzyme was necessary.

The expected chymotrypsin specificity for tryptophan, tyrosine and phenylalanine residues was observed except in

34 35 36 37

the sequence Arg-Tyr-Phe-Thr, the cleavage occurring only after the second aromatic. In the trypsin inhibitor of Kunitz, a similar sequence Arg-Tyr-Phe-Tyr-Asn exists and chymotrypsin splits after both tyrosine residues but not between Phe and Tyr (Chauvet and Acher, 1967a). In the B chain of insulin the sequence Arg-Gly-Phe-Phe-Tyr-Thr is cleaved by chymotrypsin after the second Phe and after Tyr but not between the two Phe (Sanger and Tuppy, 1951). It seems that when there are two consecutive aromatic residues in the sequence, chymotrypsin preferentially splits after the second. About leucines, 10 out of 17 leucine bonds were split but the reasons for which 7 bonds were not cleaved are not clear. The short time used (3 hr) and probably the great number of suitable residues have limited the cleavages at the expected amino acids (Neil *et al.*, 1966).

Except for the purification of the trifluoroacetyl fragments of β chain, isolation of peptides was performed with the fingerprinting technique. A fingerprint usually can supply 100–300 nmoles of peptide and material obtained from a few fingerprints is sufficient for structural studies. The paper strip variant of Edman degradation turned out to be very convenient because it needs small amounts of product and five peptides can easily be worked simultaneously.

Features of β-Chain Family. Only mammalian β chains are so far known and general characteristics of β chain family cannot yet be established. However, there are among all the known β chains some common features which can be contrasted with those of α-chain family. The mammalian β chains usually have 146 residues but shorter chains such as β chains of sheep or goat hemoglobins C with a 5-residue deletion at the N-terminal end have been found (Boyer *et al.*, 1967); Huisman *et al.*, 1967). Among abnormal human hemoglobins, the Gun Hill hemoglobin has a β chain with a 5-residue deletion in positions 91–95 (or 92–96 or 93–97), which cannot bind the heme group (Bradley *et al.*, 1967). β chains with 145 residues have been found in sheep, goat, and ox as well as in a pathological human hemoglobin (Jones *et al.*, 1966). Compared to the 146-residue mammalian β chains, the frog β chain has 6 residues deleted from the N-terminal portion (Figure 8). This deletion could be peculiar to the species and has not necessarily an evolutionary meaning. However, a similar deletion of 6 amino acids from the NH_2 terminus has been observed in the β chain of *Rana catesbeiana* hemoglobin (Baldwin and Riggs, 1971).

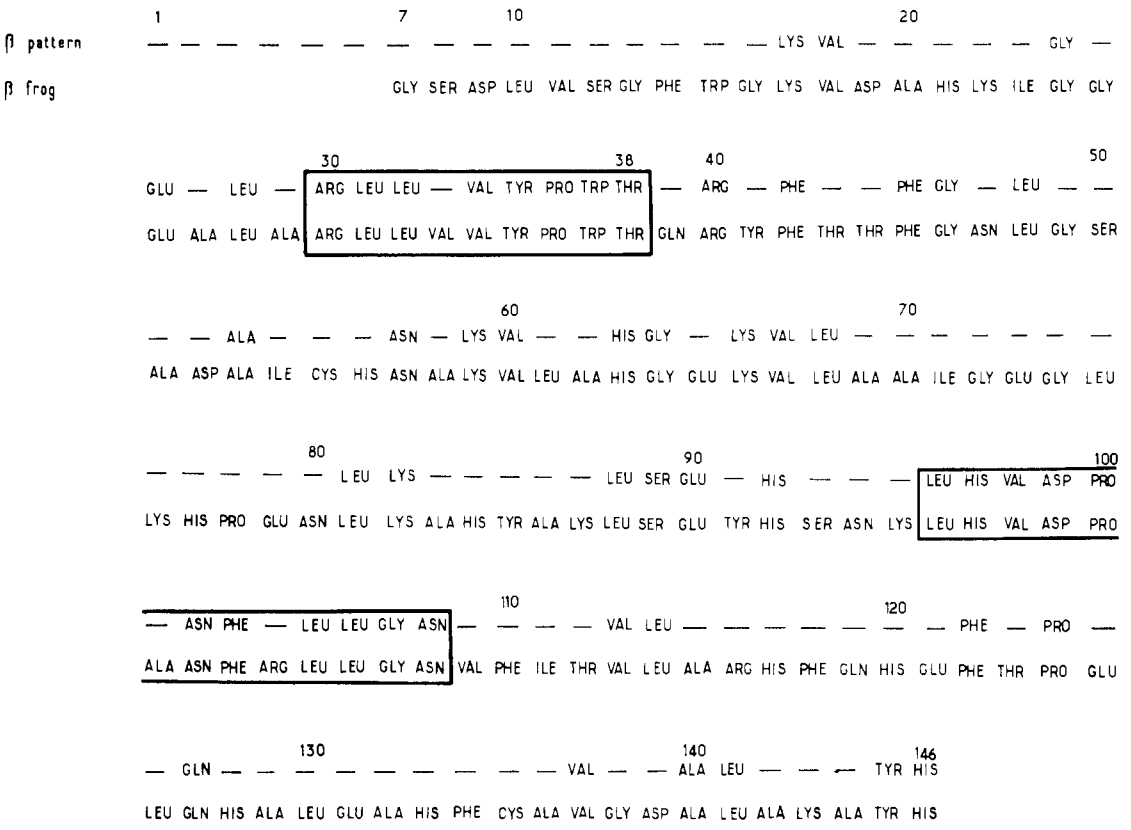

FIGURE 8: Pattern of the β-chain family. The first line shows the 54 invariant positions in all the known β, δ, and γ chains (26 mammalian and one amphibian chains) assuming a frame of 146 residues. The second line shows the primary structure of frog β chain. The two nearly invariant sequences of the β-chain family are in boxes.

It is noteworthy that in 26 β, δ, and γ chains, there is a nearly invariant sequence of 13 residues (28–40 in human β chain numbering) Leu-Gly-Arg-Leu-Leu-Val-Val-Tyr-Pro-Trp-Thr-Gln-Arg which can be regarded as characteristic of the family (Dayhoff, 1969). There are two exceptions for position 33. In the place of Val, there is Leu in rhesus monkey (Matsuda *et al.*, 1971) and Ile in grey kangaroo (Air *et al.*, 1971). There is one exception for position 39. In the place of Gln there is Arg in llama (Braunitzer *et al.*, 1964). It is of interest to find an almost identical sequence in frog β chain (with Ala in the place of Gly in position 29 in human numbering). Another virtually constant sequence of 13 residues exists between residues 96 and 108, and in frog β chain this sequence is actually found with a substitution of Ala for Glu in position 101 (human numbering). Hence, these two nearly immutable long sequences can probably be used for recognition of a β chain in lower vertebrates in which the distinction between α and β types might be difficult.

The α-chain family is more stable evolutionarily than the β-chain group and it is also possible to recognize a nearly invariant sequence of 15 residues between the positions 88 and 102 (Dayhoff, 1969), so that the two families can be distinguished by their particular sequential "features." It may be recalled that in 30 eukaryotic cytochromes *c*, an 11-residue invariant sequence was found (Nolan and Margoliash, 1968). The reason of the stability of long sequences might be that strict fits of the polypeptide chain to the prosthetic group, other chains and ligands are necessary for the biological function.

Evolution of β Chains. Twenty-six mammalian β, γ, or δ chains are so far known (Dayhoff, 1969); 70 positions out of 146 are constant, about one-third being in the 2 long sequences above mentioned. The number of invariant positions decreases to 54 when the frog β chain is included in the family. The distribution of these 54 positions along the chain increases the contrast already noted between the isolated positions and those included in the two long sequences (Figure 8). According to Perutz (Perutz *et al.*, 1968; Perutz and Lehmann, 1968) the "vital" positions are those for the $\alpha_1\beta_2$ (or $\beta_1\alpha_2$) contacts; in horse β chains there are 9 residues involved in these contacts: Pro-36, Trp-37, Gln-39, Arg-40, His-97, Val-98, Asp-99, Glu-101, Asn-102; these residues are found in frog β chain except Glu-101 which is replaced by Ala. In contrast the positions which are involved in the $\alpha_1\beta_1$ contacts seem to be less strict since out of 18 positions (Perutz *et al.*, 1968), only Arg-30, Val-33, Val-34, Tyr-35, Asn-108, Phe-122, Pro-124, Gln-127, are found in all the β chains.

The 21 positions which are occupied by residues in contact with the heme group (Perutz *et al.*, 1968) are relatively stable since 16 are invariant: Leu-31, Thr-38, Phe-42, Phe-45, His-63, Lys-66, Val-67, Leu-88, His-92 (bound to iron), Leu-96, Val-98, Asn-102, Phe-103, Leu-106, Val-137, Leu-141. So there are 30 invariant residues for 47 involved in the interactions of β chain either with α chains or with heme group. The "immutability" in this case reaches 64% in contrast to 25% for the other positions of the polypeptide chain.

The Bohr effect, which is related to the quaternary structure of hemoglobin, is supposed to involve, in β chain, a salt bridge between His-146 and Asp-94 (Perutz *et al.*, 1969; Perutz, 1970) and a thiol group of Cys-93 (Riggs, 1961). All these positions are invariant in mammalian β chains, but in frog β-chain Asn replaces Asp in position 94 and Ser replaces Cys in position 93. However, there are two new Cys residues in

TABLE III: Rate of Variation of β Chains.[a]

Order Primates			Order Artiodactyla				Vertebrates				
Hominoidea			*Suborder Ruminantia*				*Class Mammalia*				
Man	0		Sheep A	0			Eutheria				
Chimpanzee	0	20 m	Goat A	4			Primates (man)	0			
Gorilla	1		Sheep B	7	40 m		Artiodactyla				
			Bovine B	16		60 m	Ruminantia (sheep A)	27	80 m	130 m	
Ceboidea		60 m	Goat C	15			Suiformes (pig)	24			
Spider monkey	6		Sheep C	16			Tylopodes (llama)	21			
			Suborder Suiformes				Perissodactyla (horse)	25			320 m
Cercopithecoidea			Pig	26			Rodentia (mouse)	25			
Japanese monkey	7		*Suborder Tylopodes*				Metatheria				
Rhesus monkey	8		Llama	27			Marsupialia (kangaroo)	38			
							Class Amphibia				
							Anura (frog)	67			

[a] Number of amino acid substitutions (Dayhoff, 1969) and approximate time since the common ancestor (in m, million years) according to paleontological data (Simpson, 1950; Romer, 1966).

positions 55 and 135. A Bohr effect has been observed for the hemoglobins of the adult *R. catesbeiana* and not for the hemoglobins of the tadpole (Agarwal and Riggs, 1969) but the comparison between the amino acid sequences of the respective β chains remains to be made.

It is noteworthy that the nearly invariant sequence 30–40 could be explained by the necessity of the contacts $\beta_1\alpha_1$ (30, 33, 34, 35), $\beta_1\alpha_2$ (36, 37, 39, and 40), and β heme (38). In the same way, the permanent sequence 96–108 could be justified, but less strongly, by contacts between β chain and heme (96, 98) in one hand and the two α chains (97, 98, 99, 101, 102, 108) in the other. However, because some adaptative substitutions seem possible, the explanation of the stability of peculiar parts of the chain might also involve some intrinsic factors.

If the evolution of β chain was mainly dependent upon the interactions with the two α chains and the heme group, the evolution of α chain should have had approximately the same restrictions and consequently the same rate. However, the variability of α chain is clearly less great (Dayhoff, 1969). It would be of interest to compare the variabilities of α and β chains to that of myoglobin which, being limited only by the contacts between the polypeptide chain and heme, should be greater.

May the number of substitutions indicate the phylogenetic distance between two species? There is evidence that two species belonging to the same genus might have nearly identical β chain. Rhesus monkey (*Macaca mulatta*) and japanese monkey (*Macaca fuscata*) have hemoglobins with identical β chains except for a single residue (Matsuda *et al.*, 1971). The β chain of grey kangaroo (*Macropus giganteus*) differs from that of red kangaroo (*Macropus rufus*) by a single substitution (Air *et al.*, 1971). Between two species belonging to two different genera of the same family or superfamily, the number of substitutions is rather small. Among apes, the chimpanzee has a hemoglobin β chain identical with human β chain and the gorilla has only one amino acid substitution. The β chain of sheep hemoglobin A differs from its homolog of goat hemoglobin A by only four residues (Dayhoff, 1969), the two species belonging respectively to the genera Ovis and Capra. Although the distinction between genera, families, orders, etc., does not indicate precise units of phylogenetic distance, we can assume that homologous proteins are probably more akin when the species belong to the same genus the genera to the same family, the families to the same suborder and so on.

Table III indicates the number of substitutions in two orders of mammals, Primates and Artiodactyla, with man as reference for the first and sheep hemoglobin A for the second. This number increases with time but not proportionally. The rate is about 5 substitutions/10 million years for Artiodactyla but much less for Primates. Between several orders of eutherian mammals which separated about 80 million years ago, the number of substitutions is approximately 25 whatever the order taken as reference (Dayhoff, 1969) and the average rate is only 3 substitutions/10 million years. For the marsupial–eutherian divergence which probably occurred at the early cretaceous period, 130 million years ago, there are about 40 substitutions whatever the species taken as reference (Air *et al.*, 1971) and the average rate is 3 substitutions/10 million years. If the frog β chain is compared to mammalian β chains, there are 64–68 substitutions or deletions following the species chosen; the time elapsed from the common ancestor of mammals and amphibians is about 320 million years (Simpson, 1950) and the average rate is 2 substitutions/10 million years. We can theoretically admit that the number of substitutions in each line is half of the number of amino acid differences. It is noteworthy that the time scale for the evolution of organisms is derived from fossil evidence and at the time of species divergence, the populations may have been small so that the probability that fossils will be found is very low (Dayhoff, 1969).

Several authors have assumed that proteins have evolved with a rate of substitution statistically constant (Zuckerkandl and Pauling, 1965; Kimura, 1969; Air *et al.*, 1971). In the case of the β chain, the number of substitutions seems time dependent but not proportional since the greater the divergence time, the lower the average rate per million years. The fact that the number of substitutions is only time-dependent and apparently not related to each specific evolutionary line could suggest that the observed variations are not selective but neutral, and have only to be compatible with the biological function which remains apparently stationary (Zuckerkandl and Pauling, 1965; Nolan and Margoliash, 1968; Kimura, 1969). On the other hand, because the average rate of substitutions

decreases when the time increases, the explanation might be that the positions of the polypeptide chain have different degrees of tolerance, and those with high rates of random substitutions are counted for one substitution although they "mutate" several times during the long periods of divergence (Nolan and Margoliash, 1968). These free positions might be the positions on the surface of the molecule such as those of the human β chain for which substitutions were observed without clinical symptoms (Perutz and Lehmann, 1968). In fact almost all these "neutral" positions (21 out of 23) have varied during the evolution; in contrast the positions for which pathological mutations were observed are generally among the 54 invariants. It would be of interest to determine for the 92 variable positions, the limits of the variability in relation with the chemical analogy of the residues, on one hand, and the genetic code, on the other.

The comparison of the percentages of homology between α and β chains in each class of vertebrates is also of interest for checking the hypothesis of a common ancestor of both chains (Ingram, 1963). If the duplication of the ancestral monomeric hemoglobin has occurred in the very early vertebrates, and if it is assumed that sequence changes have occurred in a parallel way with morphological changes, the similarity between α and β chains might be greater in primitive fishes than in mammals. However, so far, in nonmammalian species, either α or β chain is fully known but not both so that the comparison cannot yet be done.

Because the actual problem is to relate a possible drift of protein sequences to the morphological drift of organisms, it would be appropriate to wonder how the concepts used to explain the second can be applied to the first. Anatomical variations are filtered to a given direction by natural selection so that the successive steps are explained by an increasing benefit for the organism. The observed amino acid substitutions could express either a real evolution, that means substitutions ordered by selection pressure to improve the biological property of the protein, or a neutral variability of the molecule. Because homologous proteins are usually recognized by their function which remains apparently stationary, the improvement could often be questionable. Neutral substitutions not related to the general morphological evolution are probably to be found if the biological properties are identical. In contrast proteins with evolutionary substitutions must have functions quantitatively or qualitatively different; they might be related to the phylogenetic lines built by paleontologists (Romer, 1966) because a coordination between the individual evolutions of the components of an organism seems more or less to exist. Furthermore, the comparison between proteins functionally different but with structural similarities and synthesized by cells from the same embryological origin, such as enterosecretory proteins (Adelson, 1971), could throw some light on the molecular mechanism of evolution.

Acknowledgment

The authors thank Miss Christiane Devaux for her skilled technical assistance. They are grateful to Dr. Albert Jöhl for the synthesis of *S*-ethyl thiofluoroacetate.

References

Acher, R., Chauvet, J., and Chauvet, M. T. (1970), *Eur. J. Biochem. 17*, 509.

Acher, R., and Crocker, C. (1952), *Biochim. Biophys. Acta 9*, 704.

Adelson, J. W. (1971), *Nature (London) 229*, 321.

Agarwal, S. J., and Riggs, A. (1969), *J. Biol. Chem. 244*, 2372.

Air, G. M., Thompson, E. O. P., Richardson, B. J., and Sharman, G. B. (1971), *Nature (London) 229*, 39.

Baglioni, C. (1961), *Biochim. Biophys. Acta 48*, 489.

Bailey, J. L. (1967), Techniques in Protein Chemistry, Amsterdam, Elsevier Publishing Co.

Baldwin, T. O., and Riggs, A. F. (1971), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 30*, 1295.

Boyer, S. H., Hathaway, P., Pascasio, F., Bordley, J., Orton, C., and Naughton, M. A. (1967), *J. Biol. Chem. 242*, 2211.

Bradley, B., Jr., Wohl, R. C., and Rieder, R. F. (1967), *Science 157*, 1581.

Braunitzer, G., Hilse, K., Rudloff, V., and Hilschmann, N. (1964), *Advan. Protein Chem. 19*, 34.

Brenner, M., Niederwieser, A., and Pataki, G. (1961), *Experientia 17*, 145.

Brewer, H. B., Jr., and Ronan, R. (1970), *Proc. Nat. Acad. Sci. U. S. 67*, 1862.

Chauvet, J., and Acher, R. (1967a), *Bull. Soc. Chim. Biol. 49*, 985.

Chauvet, J., Nouvel, G., and Acher, R. (1966), *Biochim. Biophys. Acta 115*, 121.

Chauvet, J. P., and Acher, R. (1967b), *C. R. Acad. Sci. Ser. A 265*, 2084.

Chauvet, J. P., and Acher, R. (1968), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 5*, 305.

Chauvet, J. P., and Acher, R. (1970), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 10*, 136.

Chauvet, J. P., and Acher, R. (1971), *Int. J. Protein Res. 3*, 261.

Crestfield, A. M., Moore, S., and Stein, W. H. (1963), *J. Biol. Chem. 238*, 622.

Davidson, B. E., Sajgo, M., Noller, H. F., and Harris, J. I. (1967), *Nature (London) 216*, 1181.

Dayhoff, M. O. (1969), Atlas of Protein Sequence and Structure, Vol. 4, Silver Spring, Md., The National Biomedical Research Foundation.

Dixon, H. B. F., and Perham, R. N. (1968), *Biochem. J. 109*, 312.

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

Fraenkel-Conrat, H., Harris, J. I., and Levy, A. L. (1955), *Methods Biochem. Anal. 2*, 359.

Goldberger, R. F., and Anfinsen, C. B. (1962), *Biochemistry 1*, 401.

Hilse, K., Sorger, V., and Braunitzer, G. (1966), *Hoppe-Seyler's Z. Physiol. Chem. 344*, 166.

Huisman, T. H. J., Adam, H. R., Dimmock, M. D., Edwards, W. E., and Wilson, J. B. (1967), *J. Biol. Chem. 242*, 2534.

Ingram, V. M. (1963), The Hemoglobins in Genetics and Evolution, New York, N. Y., Columbia University Press.

Jones, R. T., Brimhall, B., Huisman, T. H. J., Kleihauer, E., and Betke, K. (1966), *Science 154*, 1024.

Katz, A. M., Dreyer, W. J., and Anfinsen, C. B. (1959), *J. Biol. Chem. 234*, 2897.

Kimura, M. (1969), *Proc. Nat. Acad. Sci. U. S. 63*, 1181.

Matsuda, G., Maita, T., Ota, H., Tachikawa, I., Tanaka, Y., Araya, A., and Nakashima, Y. (1971), *Int. J. Protein Res. 3*, 53.

Matsuda, G., Takei, H., Wu, K. C., Minuzo, K., and Shiozawa, T. (1970), *Proc. Int. Congr. Biochem., 8th Abst. 4*.

Moss, B., and Ingram, V. M. (1968), *J. Mol. Biol. 32*, 481.

Muller, C. G. (1960), *Nature (London) 186*, 643.

Neil, G. L., Nieman, C., and Hein, G. E. (1966), *Nature (London) 210*, 903.

Niu, C. I., and Fraenkel-Conrat, H. (1955), *J. Amer. Chem. Soc. 77*, 5882.
Nolan, C., and Margoliash, E. (1968), *Annu. Rev. Biochem. 37*, 727.
Perham, R. N., and Jones, G. M. T. (1967), *Eur. J. Biochem. 2*, 84.
Perutz, M. F. (1970), *Nature (London) 228*, 734.
Perutz, M. F., and Lehmann, H. (1968), *Nature (London) 219*, 902.
Perutz, M. F., Muirhead, H., Cox, J. M., and Goaman, L. C. G. (1968), *Nature (London) 219*, 131.
Perutz, M. F., Muirhead, H., Mazzarella, L., Crowther, R. A., Greer, J., and Kilmartin, J. V. (1969), *Nature (London) 222*, 1240.
Raftery, M. A., and Cole, R. D. (1963), *Biophys. Biochem. Res. Commun. 10*, 467.
Riggs, A. (1961), *J. Biol. Chem. 236*, 1948.
Romer, A. S. (1966), Vertebrate Paleontology, Chicago, Ill., The University of Chicago Press.
Sanger, F., and Tuppy, H. (1951), *Biochem. J. 49*, 481.
Schroeder, N. A., Shelton, J. R., Shelton, J. B., and Cormick, J. (1963), *Biochemistry 2*, 1353.
Simpson, G. G. (1950), The Meaning of Evolution, London, Oxford University Press.
Simpson, G. G. (1964), *Science 146*, 1535.
Smith, E. (1953), *Nature (London) 171*, 43.
Spackman, D. H., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30*, 1190.
Zuckerkandl, E., and Pauling, L. (1965), *in* Evolving Genes and Proteins, Bryson, V., and Vogel, H. J., Ed., New York, N. Y., Academic Press, p 97.

Molecular Weight and Circular Dichroism Studies of Bovine and Ovine Pituitary Growth Hormones†

Thomas A. Bewley and Choh Hao Li*

ABSTRACT: The state of aggregation and the conformation of bovine and ovine pituitary growth hormones have been studied by means of exclusion chromatography, osmotic pressure, and circular dichroism measurements in acidic, slightly alkaline, and strongly alkaline solutions. The dimer of both proteins (mol wt 41,000–42,000) may be dissociated into the monomer form (mol wt ≈ 22,000) under either acidic (pH 3.6) or strongly alkaline (pH 11.5) conditions. In acidic solution, this dissociation is not accompanied by significant changes in the secondary structures of either protein, the α-helix contents being 45–50% in all cases. Some alterations in the tertiary structures does occur however, as evidenced by changes in the circular dichroic spectra in the region of side-chain absorption.

Sedimentation studies in the ultracentrifuge have previously indicated molecular weights of 46,000 for BGH[1] (Li and Pedersen, 1953) and 47,800 for SGH (Papkoff and Li, 1958). The possibility of a dissociation of BGH into smaller components in highly alkaline solutions was also strongly suggested by Li and Pedersen (1953). Subsequently, considerable evidence for the dissociation of BGH under various acidic, basic and denaturing conditions has appeared in the literature. These results have been summarized by Dellacha *et al.* (1968), who also present their own evidence indicating that BGH may be completely dissociated from the dimeric form (mol wt 40,000–45,000) in neutral or slightly alkaline solution, to the monomer (mol wt ≈ 21,000) in acidic solutions of low ionic strength.[2] This monomer weight for BGH is in excellent agreement with a minimum molecular weight of 20,846 calculated by Fellows and Rogol (1969) from the composition of fragments obtained by treatment of the hormone with cyanogen bromide.

The conformation of BGH has also been a subject of considerable interest. The results of a number of earlier investigations have been summarized by Aloj and Edelhoch (1970). There is some doubt as to the state of aggregation of the hormone in some of these earlier studies. With this in mind, Edelhoch and Lippoldt (1970) have presented a more recent conformational study of this protein based on circular dichroism measurements in solvents which might be expected to provide predominantly monomer or dimer forms of the hormone.

Although BGH and SGH have been shown to have a number of important chemical and immunochemical similarities (Papkoff and Li, 1958; Moudgal and Li, 1961; Hayashida, 1969), relatively little is known about the ovine hormone. In this study, we have reinvestigated the dissociation of BGH, both in strong alkali and under the acidic conditions described by Dellacha *et al.* (1968), with the goal of isolating and characterizing samples of the monomeric and dimeric forms of the protein. We have estimated the molecular weights of each form by exclusion chromatography and osmotic pressure. Circular dichroism measurements were then carried out on the purified monomer and dimer. Finally, par-

† The Hormone Research Laboratory, University of California, San Francisco, California 94122. *Received July 14, 1971.* This work was supported in part by the American Cancer Society, the Geffen Foundation and the Allen Foundation.

[1] Abbreviations used are: BGH, bovine pituitary growth hormone; SGH, ovine pituitary growth hormone; HGH, human pituitary growth hormone; LTH, ovine pituitary lactogenic hormone; HCS, human chorionic somatomammotropin; CD, circular dichroism.

[2] 0.1 M glycine hydrochloride buffer, pH 2.2–3.6, μ = 0.045–0.01.