THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 256, No. 11, Issue of June 10, pp. 5524-5531, 1981
Printed in U.S.A.

Complete Amino Acid Sequence of the Major Early Embryonic β-like Globin in Chickens*

(Received for publication, July 7, 1980, and in revised form, October 13, 1980)

Barbara S. Chapman‡ and Allan J. Tobin
From the Department of Biology, University of California, Los Angeles, California 90024
Leroy E. Hood
From the Division of Biology, California Institute of Technology, Pasadena, California 91125

The ρ globin is the major β-like chain found in 5-day-old chick embryos. In association with two unique early embryonic α-like globins, it forms the two major hemoglobins of early chick development. This paper presents the complete amino acid sequence of the ρ globin. There are no amino acid differences between the ρ chain and the adult chicken β chain at known Bohr effect or organophosphate-binding positions, and there are only 19 differences altogether. The ρ globin ought to be functionally equivalent to the adult chicken β globin. Since the adult and embryonic chains are very similar in sequence, they may be products of a relatively recent gene duplication in the chicken β globin gene family. The possibility of a gene correction event is discussed.

Early embryonic hemoglobins are normally seen only during very early stages of development in association with yolk sac-derived erythroid cells (Kitchen and Brett, 1974). Electrophoretically distinct hemoglobins have been described in early embryos of several species of birds (Borgese and Bertles, 1965; Manwell *et al.*, 1966; Beaupain *et al.*, 1979) and mammals (Craig and Russel, 1964; Kleihauer and Stoffler, 1968; Pataryas and Stamatoyannopoulos, 1972; Kitchen and Brett, 1974; Jelkmann and Bauer, 1977). These hemoglobins disappear during ontogeny and are replaced by fetal and adult hemoglobins.

Early embryos of birds and mammals produce at least two types of β-like globin chains unique to the embryo (Fantoni *et al.*, 1967; Steinheider *et al.*, 1972; Brown and Ingram, 1974; Gale *et al.*, 1979). These combine with at least two types of α-like globins to form several hemoglobins. One of the α-like globins is found only in the early embryo (Melderis *et al.*, 1974). In the chicken, the four early chick embryonic hemoglobins are thought to have the following chain compositions: HbP[1]($\pi_2\rho_2$), HbP′ ($\pi'_2\rho_2$), HbE ($\alpha^A{}_2\epsilon_2$), and HbM ($\alpha^D{}_2\epsilon_2$) (Brown and Ingram, 1974). Since HbP and HbP′ account for 64% of the hemoglobin in the hemolysates, the ρ globin is the major early β-like globin and ε globin is the minor.

The genes encoding α- and β-like embryonic globins, together with late embryonic and adult globin genes, are arranged in two multigene families in birds and mammals (Dodgson *et al.*, 1979; Engel and Dodgson, 1980; Fritsch *et al.*, 1980; Ginder *et al.*, 1979; Lacy *et al.*, 1979; Lauer *et al.*, 1980; Jahn *et al.*, 1980). The early embryonic genes in these families have evolved in interesting ways with respect to the adult genes, with different patterns in the α-like and β-like globin gene families (Melderis *et al.*, 1974; Steinheider *et al.*, 1975; Chapman, 1981).

In order to investigate the structure, evolution, and perhaps the physiological function of early embryonic hemoglobins, we have determined the amino acid sequences of several α- and β-like globins from chick embryos. In this paper, we report the complete amino acid sequence of the major early embryonic β-like globin in chickens, designated ρ. This globin appears to be functionally equivalent to the adult β globin in chickens, from which it differs at 19 positions. The small number of amino acid differences suggests that the ρ and β globins are products of a relatively recent gene duplication. An unexpected distribution of substitutions in the COOH-terminal region of the ρ globin suggests a possible correction

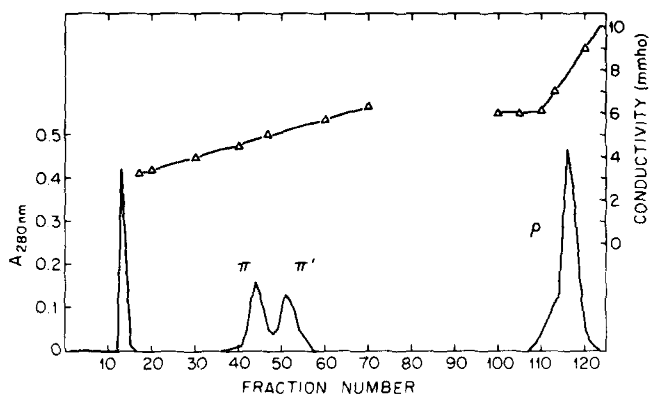


FIG. 1. **Separation of α- and β-like globin chains from HbP and HbP′.** Approximately 10 mg of reduced, carboxyamidomethylated globin was dissolved in starting buffer (freshly deionized 8 M urea brought to pH 4.3 with formic acid and containing 50 mM NaCl). This was chromatographed on Whatman CM52 equilibrated with the same buffer. After an initial wash with 50 ml of starting buffer, the α-like globins were eluted with a gradient of 50 to 150 mM NaCl in 8 M urea, pH 4.3. The β-like globin was eluted with a second gradient of 150 to 300 mM NaCl. Protein was detected by absorbance at 280 nm, and conductivity was monitored at every 10th fraction. Five-ml fractions were taken at a flow rate of 38 ml/h through a column (1.0 × 10 cm). CAM, carboxyamidomethyl.

* This research was supported by National Research Service Award F32 HL05553 (to B. S. C.), Grants PCM 76-02859 and 78-20767 from the National Science Foundation (to A. J. T.), and Grant GM-06965 from the National Institutes of Health (to L. E. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address, Department of Biochemistry, University of California, Berkeley, CA 94720. To whom correspondence should be addressed.

[1] The abbreviations used are: Hb, hemoglobin; CM, carboxymethyl; PTH, phenylthiohydantoin; HPLC, high pressure liquid chromatography.

Downloaded from www.jbc.org at CALIFORNIA INSTITUTE OF TECHNOLOGY on December 24, 2006

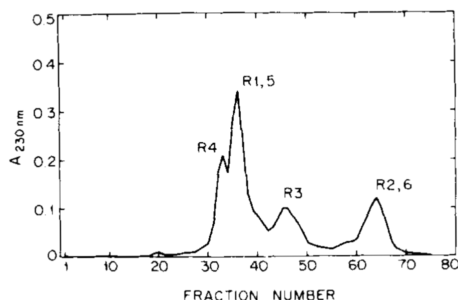


FIG. 2. **Separation of peptides resulting from cleavage at arginine residues.** Approximately 2 mg of reduced, carboxyamidomethylated ρ globin was cleaved under mild acid conditions, succinylated, and digested with trypsin. Peptides were separated on a column (1.6 × 72 cm) of Bio-Gel P-10 in 0.1 M ammonium bicarbonate, pH 7.8. Fractions of 2.0 ml were collected at a flow rate of 6 ml/h; peptides were monitored by absorbance at 230 nm and identified by sequenator analysis. The arginine fragments are numbered starting from the NH_2 terminus.

of the ρ gene against an adult β-like gene since their divergence.

EXPERIMENTAL PROCEDURES

Isolation of the Major β-like Globin from Early Embryos—Hemolysates prepared from 5-day-old White Leghorn embryos (Chapman and Tobin, 1979) were separated into their component hemoglobins by ion exchange chromatography on CM-Sephadex or CM-52 (Brown and Ingram, 1974; Cirotto and Geraci, 1975). Fractions containing HbP and HbP′ were converted to globin by acid/acetone precipitation (Rossi-Fanelli *et al.,* 1958), then reduced and alkylated with iodoacetamide (McKean *et al.,* 1973). The globin was desalted by gel filtration on Bio-Gel P-10 (Bio-Rad) and lyophilized. To separate β-like from α-like globin, the protein was dissolved in 8 M urea, pH 4.3, containing 50 mM NaCl, and applied to a column of CM-52 (Whatman). The α-like chains were first eluted with a gradient of 50 to 150 mM NaCl in 8 M urea (Moss and Hamilton, 1974; Chapman *et al.,* 1980); then the β-like chains were eluted with a second gradient of 150 to 300 mM NaCl in 8 M urea (Fig. 1).

Automated Edman Degradation and PTH Identification—The globin was sequenced with a modified Beckman Instruments 890B

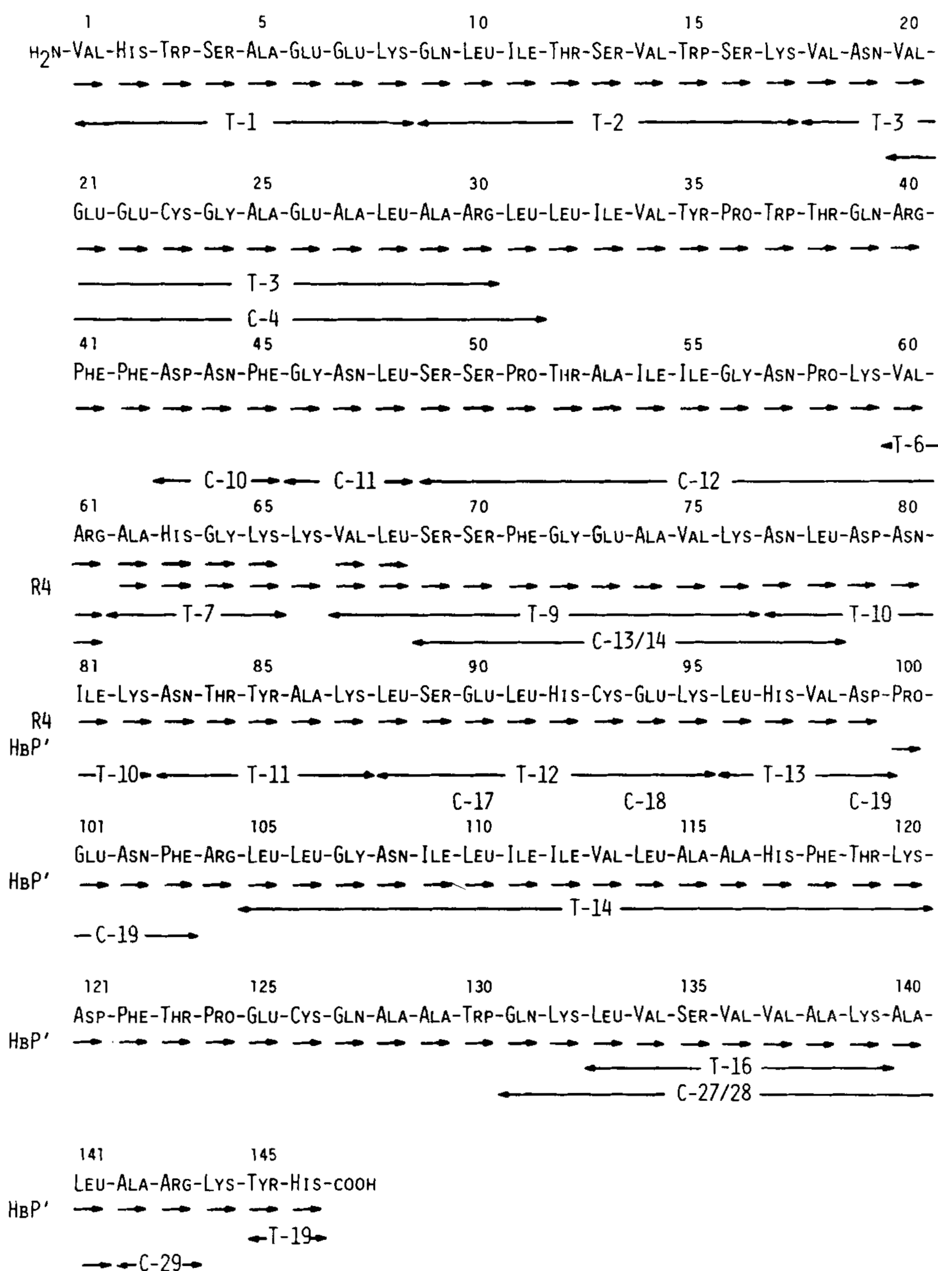


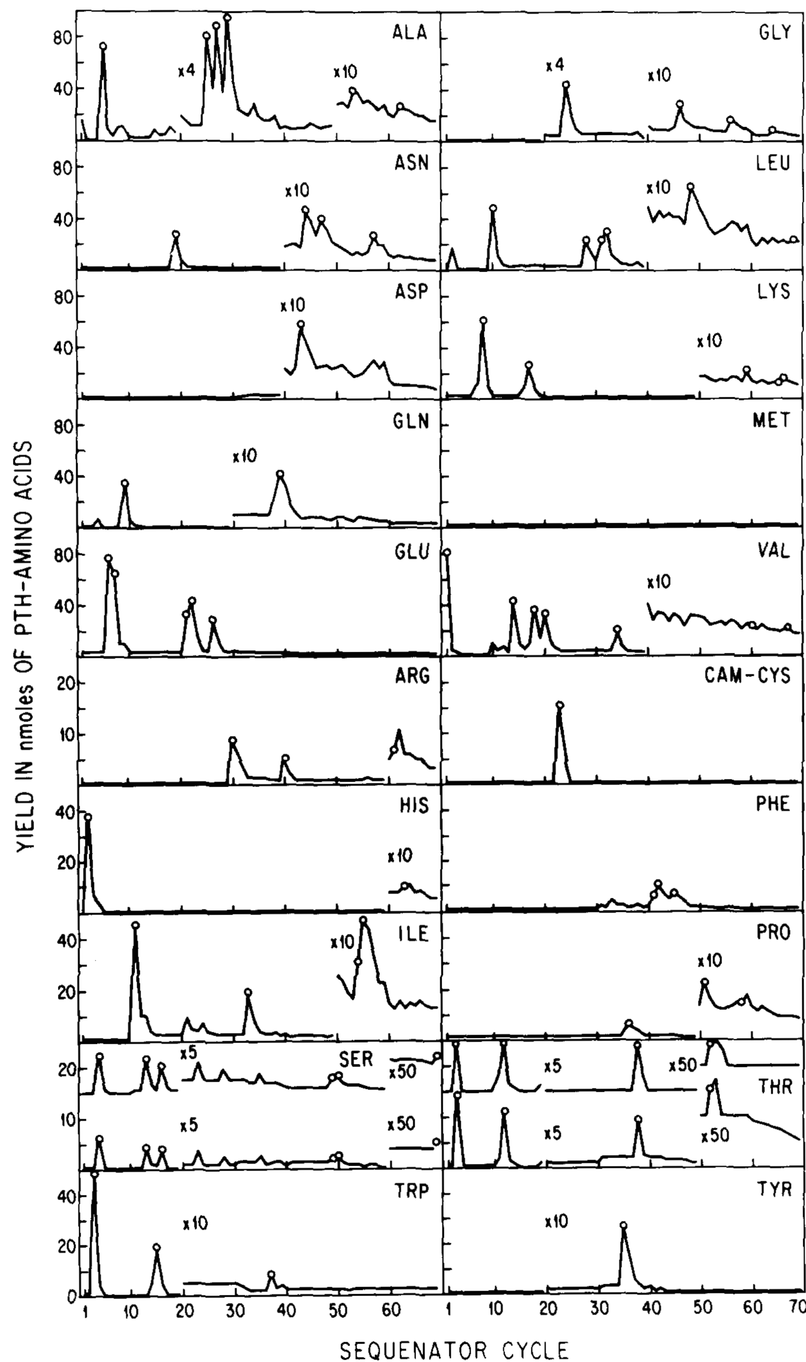
FIG. 3. **The complete amino acid sequence of the ρ chain of early chick HbP.** Composite data from Tables I and II and Figs. 4–6 are shown. →→, automated Edman degradations identified by HPLC; T-*n*, tryptic peptides numbered from the NH_2 terminus; C-*n*, chymotryptic peptides. The peptides are identified by amino acid composition. Peptides produced by incomplete cleavage with chymotrypsin are indicated with a *slash*.

Downloaded from www.jbc.org at CALIFORNIA INSTITUTE OF TECHNOLOGY on December 24, 2006

sequenator (Hunkapiller and Hood, 1978), and with a newly designed sequenator (Hunkapiller and Hood, 1980). All reagents, solvents, and procedures were as previously described (Chapman *et al.*, 1980). PTH-derivatives were identified by a reverse phase high pressure liquid chromatography system (Johnson *et al.*, 1979).

Preparation of Fragments for Sequencing—Carboxyamidomethylated globin was specifically cleaved at the aspartic acid-proline bond using mild acid treatment (Piszkiewicz *et al.*, 1970) in 6 M guanidine HCl (Chapman *et al.*, 1980). The COOH-terminal portion of the polypeptide was isolated and desalted by gel filtration on Bio-Gel P-10 in 0.5% formic acid. Arginine cleavage fragments were prepared by trypsin digestion of succinylated globin (Chapman *et al.*, 1980), and were separated by gel filtration on Bio-Gel P-10 in 0.1 M ammonium bicarbonate (Fig. 2). All fragments were lyophilized and stored at −20 °C.

Tryptic and Chymotryptic Peptide Compositions—Purified globin was digested with L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin or with chymotrypsin (Chapman *et al.*, 1980) and separated in two dimensions on thin layer sheets (Brown and Ingram, 1974). Peptides were identified by ninhydrin staining, eluted, and



1 5 10 15 20
Val-His-Trp-Ser-Ala-Glu-Glu-Lys-Gln-Leu-Ile-Thr-Ser-Val-Trp-Ser-Lys-Val-Asn-Val-
25 30 35 40
Glu-Glu-Cys-Gly-Ala-Glu-Ala-Leu-Ala-Arg-Leu-Leu-Ile-Val-Tyr-Pro-Trp-Thr-Gln-Arg-
45 50 55 60
Phe-Phe-Asp-Asn-Phe-Gly-Asn-Leu-Ser-Ser-Pro-Thr-Ala-Ile-Ile-Gly-Asn-Pro-Lys-Val-
65 70
Arg-Ala-His-Gly-Lys-Lys-Val-Leu-

Fig. 4. **Yield of PTH-derivatives from sequenator analysis of the NH_2 terminus of the ρ chain.** Approximately 100 nmol of ρ globin was degraded using a Quadrol single cleavage program (Hunkapiller and Hood, 1978). Derivatives from 70 cycles were analyzed by HPLC; 10-μl injections were normalized to 100% of the sample and the peak heights were converted to nanomoles by use of a 1.0-nmol standard. The average repetitive yield throughout was 94%. Sequenator lag was significantly increased after the proline at cycle 36 and became larger than the signal after the proline at cycle 58, but the moderate background allowed identification of the principal residues throughout. PTH-threonine and PTH-serine are shown as double plots in order to indicate levels of a characteristic threonine product in the 313 nm absorbance channel and a serine product appearing between histidine and tyrosine in the chromatogram. To facilitate plotting on a linear scale, yields of PTH-derivatives from later cycles have been multiplied by the factors indicated. ○, residues of the ρ chain. The sequence is listed at the *bottom* of the figure. CAM, carboxyamidomethyl.

Downloaded from www.jbc.org at CALIFORNIA INSTITUTE OF TECHNOLOGY on December 24, 2006

hydrolyzed (Chapman *et al.,* 1980). Quantitative analyses were done with a Durrum D-500 amino acid analyzer.

RESULTS

The complete amino acid sequence of the ρ globin is shown in Fig. 3. The entire sequence was established with three sequenator runs: an NH_2-terminal analysis of residues 1-68 (Fig. 4); a determination of all 38 residues of an internal arginine cleavage fragment ending at the mild acid cleavage site between residues 99 and 100 (Fig. 5); and a run of the final 47 amino acids beginning at residue 100 (Fig. 6). More than 85% of the sequence was confirmed by amino acid compositions, and charge and NH_2-terminal analyses of tryptic and chymotryptic peptides (Tables I and II; Fig. 3).

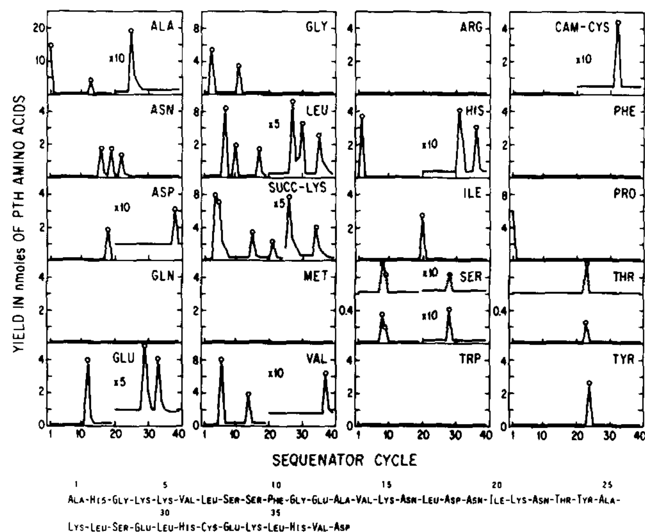


FIG. 5. **Yield of PTH-derivatives from sequenator analysis of an internal arginine fragment R4.** Twenty nmol of the R4 fragment shown in Fig. 2 were degraded and 40 cycles were analyzed by HPLC. The repetitive yield averaged 92%. The yield was probably low because the small size of the peptide allowed it to wash out during the run. The sequence of the fragment is shown.

NH_2-terminal Sequence Analysis—The NH_2 terminus of the ρ chain was unblocked, allowing an unambiguous determination of 68 residues. The valine residue at cycle 60 was represented by a clear signal over background (Fig. 4), although it does not appear as a peak on the plot shown. The recovery of all PTH-derivatives from this sample was reduced (probably an artifact of HPLC loading). Residue 65 (Lys) was obscured by developing lag in the sequence, combined with repetition of Lys at position 66. Lag is a term used when referring to the appearance of PTH-derivatives from a previous cycle in a subsequent cycle, and is the result of incomplete coupling or cleavage of a residue, so that some is cleaved or coupled in the steps following. At position 65, Lys was identified in an overlapping sequence and by the composition of tryptic peptide T-7 (Table II). Neither the tryptic nor chymotryptic peptides for the region between residues 32-42 were isolated from contaminating peptides, and were therefore not useful in confirming the sequence. However, reliable data were obtained from both sequenator and HPLC for these residues.

Sequence of Residues 62-99—Arginine fragments were prepared from the large aspartic acid-proline cleavage fragment of the ρ globin and separated by gel filtration (Fig. 2). The largest fragment contained residues 62-99, giving a 7-residue

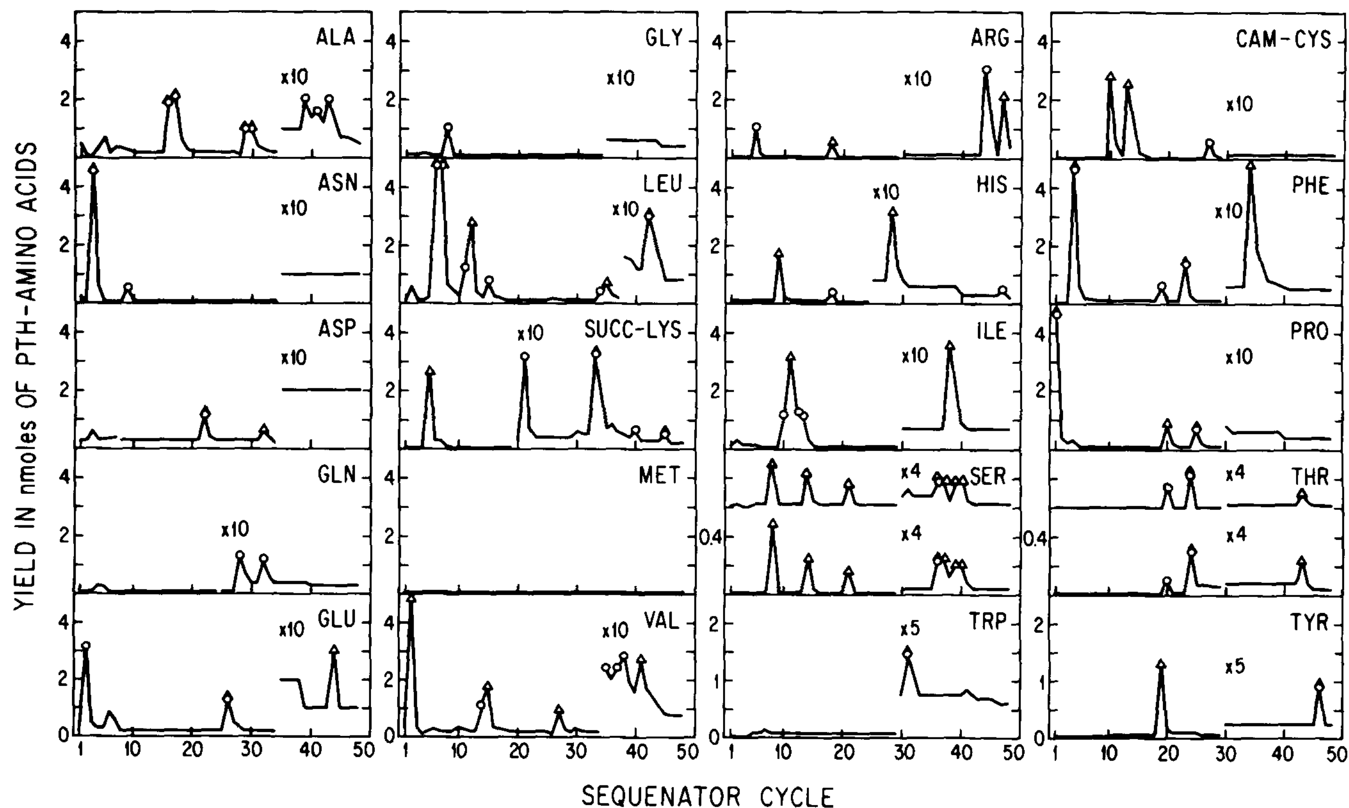


FIG. 6. **Yield of PTH-derivatives from sequenator analysis of the COOH-terminal fragment of HbP′ cleaved by mild acid treatment.** A mixture of ρ chains and π′ chains was analyzed. Only 25 pmol of the COOH-terminal histidine was recovered, but the identification of this residue was confirmed by peptide composition. ○, indicates a residue of ρ chain; Δ, indicates a residue of π′ chain. The sequences of the π′ and ρ fragments are shown in Fig. 7. CAM, carboxyamidomethyl.

Downloaded from www.jbc.org at CALIFORNIA INSTITUTE OF TECHNOLOGY on December 24, 2006

overlap to the NH_2-terminal sequence and ending with the Asp residue at position 99 (Fig. 5). Absence of background and lag in the sequenator analysis of this fragment allowed unambiguous determination of every residue. This analysis was confirmed by tryptic and chymotryptic peptide compositions (Tables I and II), which also linked this sequence to the COOH-terminal sequence (peptide C-19; Fig. 3).

Analysis of the COOH-terminal 47 Residues—Since the

TABLE I

Chymotryptic peptides

Peptides were separated by thin layer electrophoresis on polyamide sheets at pH 6.4 followed by chromatography (Brown and Ingram, 1974). Approximately 10 nmol of enzymatically hydrolyzed globin were loaded per sheet. Peptides were identified by ninhydrin staining and were hydrolyzed in 6 N HCl, and their compositions were determined. Values shown are molar ratios (greater than 0.2) of amino acid residues in each peptide. Identified peptides are numbered from the NH_2 terminus; partial cleavage products are indicated with a slash. Numbers in parentheses are the expected integer values based on sequenator analysis.

	C-4 20–31	C-10 43–45	C-11 46–48	C-12[a] 49–61	C-13/14 69–78	C-17 89–91	C-18 92–96	C-19 97–103	C-27/28[a] 131–141	C-29[a] 142–143
Aspartic acid or asparagine	0.4	1.7 (2)	1.2 (1)	1.3 (1)	1.0 (1)		0.4	1.6 (2)		
Threonine				0.5 (1)						
Serine				1.2 (2)	1.5 (2)	0.6 (1)	0.4		0.9 (1)	
Glutamic acid or glutamine	2.9 (3)				1.3 (1)	0.7 (1)	1.4 (1)	0.7 (1)	1.4 (1)	
Proline				1.6 (2)				0.8 (1)		
Glycine	1.0 (1)		0.9 (1)	1.2 (1)	1.0 (1)					
Alanine	2.5 (3)			1.1 (1)	1.1 (1)			0.3	2.4 (2)	1.0 (1)
Cysteine	0.9 (1)						0.8 (1)			
Valine	0.8 (1)			1.2 (1)	1.1 (1)			0.7 (1)	2.7 (3)	
Methionine										
Isoleucine				1.6 (2)						
Leucine	1.6 (2)		0.9 (1)		1.1 (1)	0.8 (1)	0.7 (1)	0.4	1.6 (2)	
Tyrosine										
Phenylalanine		1.2 (1)			0.9 (1)			0.9 (1)		
Histidine					0.3		0.6 (1)	1.1 (1)		
Lysine				1.2 (1)	0.7 (1)		0.6 (1)		1.6 (2)	
Arginine	0.9 (1)			1.2 (1)						1.0 (1)
Tryptophan										
Residues	12	3	3	13	10	3	5	7	11	2
Yield (nmol)	0.3	0.3	0.2	0.2	0.4	0.7	0.5	0.3	0.4	0.3
NH_2 terminus	Val	Asp	Gly	Ser	Ser	Ser	His	His	Gln	Ala
Net charge	−2	−1	0	0	0	−1	+1	−1	−2	+1

[a] These peptides were obtained by chymotrypsin cleavage of arginine fragments. The lysines are succinylated.

TABLE II

Tryptic peptide composition

Peptides were separated by thin layer electrophoresis on polyamide sheets at pH 6.4 followed by chromatography (Brown and Ingram, 1974). Approximately 10 nmol of enzymatically hydrolyzed globin were loaded per sheet. Peptides were identified by ninhydrin staining and were hydrolyzed in 6 N HCl, and their compositions were determined. Values shown are molar ratios (greater than 0.2) of amino acid residues in each peptide. Identified peptides are numbered from the NH_2 terminus; partial cleavage products are indicated with a slash. Numbers in parentheses are the expected integer values based on sequenator analysis.

	T-1 1–8	T-2 9–17	T-3 18–30	T-6 60–61	T-7 62–65	T-9 67–76	T-10 77–82	T-11 83–87	T-12 88–95	T-13[a] 96–99	T-14 105–120	T-16 133–139	T-19 145–146
Aspartic acid or asparagine			1.4 (1)			0.4	2.8 (3)	0.9 (1)	0.5	1.1 (1)	1.1 (1)		
Threonine		0.7 (1)						0.7 (1)			0.8 (1)		
Serine	0.9 (1)	1.1 (2)	0.3			1.6 (2)			0.9 (1)			0.8 (1)	
Glutamic acid or glutamine	1.8 (2)	0.8 (1)	3.1 (3)			1.1 (1)		0.3	2.4 (2)	0.4			
Proline													
Glycine	0.3		1.0 (1)		1.3 (1)	1.2 (1)		0.4			1.3 (1)		
Alanine	0.9 (1)		2.6 (3)		0.7 (1)	1.1 (1)		0.9 (1)	0.4		2.3 (2)	1.0 (1)	
Cysteine			1.0 (1)						0.8 (1)				
Valine	0.5 (1)	1.2 (1)	1.7 (2)	1.0 (1)		1.6 (2)				0.7 (1)	1.2 (1)	2.5 (3)	
Methionine													
Isoleucine		0.8 (1)					0.8 (1)				2.2 (3)		
Leucine		1.2 (1)	0.9 (1)			1.1 (1)	1.0 (1)		1.4 (2)	0.7 (1)	3.1 (4)	1.1 (1)	
Tyrosine								0.7 (1)			0.3		0.9 (1)
Phenylalanine						0.7 (1)					1.0 (1)		
Histidine	0.6 (1)				0.9 (1)				0.8 (1)	0.7 (1)	0.8 (1)		1.1 (1)
Lysine	0.5 (1)	0.8 (1)			0.9 (1)	0.6 (1)	0.5 (1)	0.7 (1)	0.7 (1)		0.8 (1)	0.9 (1)	
Arginine			0.7 (1)	1.0 (1)									
Tryptophan	ND[b] (1)	ND[b] (1)											
Residues	8	9	13	2	4	10	6	5	8	4	16	7	2
Yield (nmol)	0.2	0.2	0.4	0.3	0.2	0.6	0.4	0.4	0.1	0.5	0.2	0.1	0.2
NH_2 terminus	Val	Gln	Val	Val	Ala	Val	Asn	Asn	Leu	Leu	Leu	leu	Tyr
Charge	0	+1	−2	+1	+2	0	0	+1	0	0	+2	+1	+1

[a] This peptide was obtained by trypsin digestion of the large aspartic acid-proline cleavage fragment.
[b] ND, not determined.

Downloaded from www.jbc.org at CALIFORNIA INSTITUTE OF TECHNOLOGY on December 24, 2006

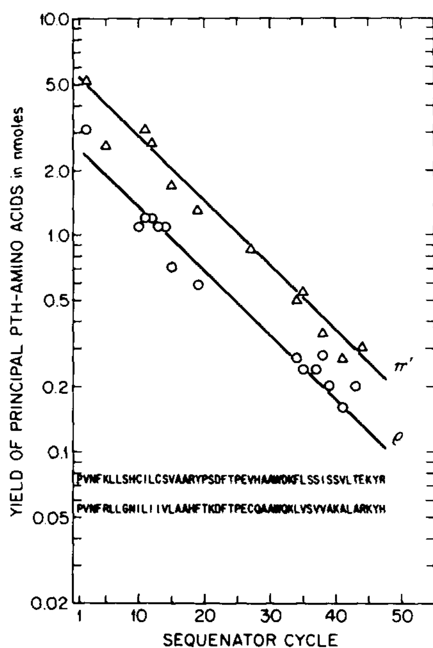


FIG. 7. **Repetitive yield of ρ and π' globin residues.** Yelds of principal PTH-derivatives of ρ and π' are plotted for each cycle of the sequenator analysis shown in Fig. 6. Residues of the π' globin represent about 25% of the PTH-derivative yield at each cycle. Repetitive yield, calculated by linear regression analysis, was 94% for each globin.

complete sequence of the π' globin was known (Chapman *et al.,* 1980), the COOH-terminal sequences of the ρ and π' globins were analyzed simultaneously. For this determination, HbP′ was separated from the other hemoglobins (Cirotto and Geraci, 1975), converted to globin, and cleaved at the Asp-Pro bonds of the α- and β-like chains. The large (NH_2-terminal) and small (COOH-terminal) fragments were separated, and the mixture of small fragments was analyzed. Fig. 6 shows both the π' and ρ chain sequences. Residue assignments were confirmed by peptide compositions except for residues 121–130. Sequenator data for positions 121–130 appear to be reliable, and Fig. 7 shows that PTH recoveries and repetitive yields are consistent with the assignments made.

DISCUSSION

Comparison of the ρ Chain with Adult Chicken β Chain—The early embryonic ρ globin differs from adult β globin at 19 of 146 positions. Eleven of these changes are conservative with respect to polarity and charge. There are no substitutions in positions identified as heme contacts, in residues implicated in the Bohr effect, or in amino acids forming the organophosphate-binding site (Ladner *et al.,* 1977). Table III shows the four differences in functional residues between the ρ and β chains. Two changes in $\alpha_1\beta_1$ contact positions interact with residues in the π and π' globins that are altered with respect to the α^A chain of adult chicken HbA (Matsuda *et al.,* 1971; Ladner *et al.,* 1977; Chapman *et al.,* 1980). The substitution of Asp in ρ for Ala in β at $\alpha_1\beta_2$ contact position 43 is probably not significant, since many mammals and marsupials have Asp at this position in their β globins (Dayhoff, 1976).

Implications of the Structural Similarity between Major Adult and Embryo β Chains—Comparison of the ρ globin sequence with the adult β globin shows that residues expected to form the organophosphate-binding site (positions 1, 2, 82, and 143) are identical (Matsuda *et al.,* 1973; Ladner *et al.,* 1977). The same is true of the predicted principal alkaline Bohr effect residue (position 146) and all 13 predicted heme contact residues. Hemoglobins containing the ρ globin and those containing the adult β globin should respond similarly to organophosphates and pH changes. Indeed, HbP and HbP′ demonstrate reduced oxygen affinity in the presence of inositol hexaphosphate, as do late embryonic and adult hemoglobins (Cirotto and Geraci, 1975; Isaacks *et al.,* 1976). Because there are functionally significant hemoglobin residues yet to be identified, functional properties of a globin chain cannot be established with certainty from an analysis of structure (Eaton, 1980). Nevertheless, in the chick embryo, as in the rabbit embryo, the unusually small Bohr effect and high oxygen affinity of isolated early embryonic hemoglobins must be properties of the α-like rather than the β-like globin chains (Jelkmann and Bauer, 1978; Chapman *et al.,* 1980).

If the ρ chain is functionally equivalent to the adult β chain, why have a ρ globin? One possibility is that the ρ globin is a separate genetic entity, capable of being coordinately regulated with the genes for the π and π' globins. Brown and Ingram (1974) have inferred from quantitative analyses of hemoglobins during development that the ρ and β chains are coordinately expressed with embryonic α-like and adult α-like chains, respectively. A common evolutionary solution to the problem of differential temporal, tissue-specific, and quantitative gene regulation is through gene duplication (Ohno, 1970; Zuckerkandl, 1978).

A second rationale for the existence of the ρ globin is that it has a specific function in forming stable hemoglobins with the unusual π and π' globins. These α-like chains differ by nearly 45% from the adult α globins of the chicken, and probably play a specialized role in the early embryo (Chapman *et al.,* 1980). The ρ globin chain has been found only in association with the π and π' globins. Perhaps the four substitutions in $\alpha\beta$ contact positions (Table III) permit ρ chains to form stable hemoglobins with the divergent π and π' chain structure.

Evolution—There are at least four β-like globin chains produced in the chicken during ontogeny (Brown and Ingram, 1974; Moss and Hamilton, 1974). These are ρ and ϵ (early embryonic), β^H (late embryonic), and β (adult). The genes

TABLE III

Comparison of chicken ρ, β, and ϵ globin amino acid sequences at variant positions

Those positions where amino acid differences occur are shown. At positions where ρ and β globins are both different from ϵ globin, the residues are in italics.

Residue No.	β	ρ	ϵ	Function
4	Thr	Ser	Ser	
13	Gly	Ser	Ser	
14	Leu	Val	Val	
16	Gly	Ser	Ser	
21	Ala	Glu	Glu	
43	Ala	Asp	Ala	$\alpha_1\beta_2$
44	Ser	Asn	Ser	
55	*Leu*	*Ile*	Met	$\alpha_1\beta_1$
59	Met	Lys	Lys	
69	Thr	Ser	Ser	
73	Asp	Glu	Glu	
85	Phe	Tyr	Tyr	
86	Ser	Ala	Ala	
87	Gln	Lys	Lys	
94	Asp	Glu	Asp	
108	Asp	Asn	Asp	$\alpha_1\beta_1$
116	*Ala*	*Ala*	Ser	$\alpha_1\beta_1$
119	*Ser*	*Thr*	Ala	$\alpha_1\beta_1$
120	*Lys*	*Lys*	Arg	
125	*Glu*	*Glu*	Ala	
128	*Ala*	*Ala*	Phe	$\alpha_1\beta_1$
135	*Arg*	*Ser*	Asn	
139	His	Lys	His	

Downloaded from www.jbc.org at CALIFORNIA INSTITUTE OF TECHNOLOGY on December 24, 2006

ASBMB
The Journal of Biological Chemistry
jbc

encoding these globins are arranged in a cluster as are the β-like globin genes of man, rabbit, and mouse (Dodgson *et al.*, 1979; Fritsch *et al.*, 1980; Lacy *et al.*, 1979; Jahn *et al.*, 1980). The amino acid sequence of the ρ chain differs from the β chain at 19 positions, suggesting that its gene diverged following a relatively recent duplication in the β-like globin multigene family (Hood *et al.*, 1975). Assuming a constant rate of evolution, the divergence time for the chick ρ and β genes is comparable to that of the human β and δ globin genes (10 differences) or the bovine β and γ globin genes (23 differences).

The early embryonic β-like globins of the chicken appear to have been derived from more recent gene duplication than the early β-like globins of man. The 2-fold lower number of amino acid differences accumulated between the most divergent pair of the chicken β-like globins compared to the differences between the human β and γ globins (40 differences) is evidence that the human β globin gene cluster may be twice as old as that of the chicken. If fixation of gene duplication occurs at an approximately constant rate, one might expect roughly twice as many human as chicken β-like genes. Dodgson *et al.* (1979) have evidence for four chicken β-like globin genes, whereas Fritch *et al.* (1980) find 7 human genes. Further tests of this hypothesis will be possible when the cluster sizes and sequences have been established for β-like globin genes in other species. Cursory examination of partial nucleotide sequences of mouse β-like globin genes suggests that the seven genes of the mouse cluster differ by at least as many amino acid replacement changes as the human globin genes (Jahn *et al.*, 1980).

While the chicken α-like globins specific to the early embryo are highly diverged from the adult α globins, the early embryonic β-like globins maintain close homology with the adult β globin. This pattern of homologies within the α- and β-like globin gene families has also been inferred for the globins of mice and rabbits from amino acid compositions of early embryonic and adult globins (Melderis *et al.*, 1974; Steinheider *et al.*, 1975). Since the early embryonic β-like globins in each species are most similar to the adult β-globin in that species, then the embryonic β-like genes must have been derived from the adult gene after divergence of the species. Sequence homology is presumably maintained by contraction and expansion cycles in the β-globin multigene family and perhaps by a gene correction mechanism (Hood *et al.*, 1975; Slightom *et al.*, 1980).

Comparison of the amino acid sequences of the chicken ρ and β globins with the chicken ε globin[2] reveals a pattern of amino acid substitutions suggesting correction of the ρ globin gene against an adult β gene, either β or β^H (Table III). Through the first 108 residues, ρ chains differ from β chains at 16 positions, while ε chains differ from β chains at 12 positions. However, in the last 38 residues, only three changes relative to β occur in ρ, while six are found in ε. If strong selection accounted for the conservation of the ρ and β COOH-terminal sequences, then the ε globin should be likewise conserved, since ε combines with adult-type α globins as does β. A gene correction mechanism seems a more likely explanation for the similarity of ρ and β COOH-terminal sequences.

It is difficult to determine from amino acid sequence data the endpoints of a genetic event. Assuming that the recombination event extended from the coding sequence for residue 105 (at the junction between the large intervening sequence and the β-globin coding sequence) to beyond the COOH-terminal coding sequence, it is necessary to explain amino acid substitutions in ρ globin for β globin residues 108, 119, 135, and 139. There are two possibilities. Either the β and ρ globin genes have diverged since the correction event, or the ρ globin gene was corrected against the β^H gene, whose sequence is unknown. A nucleotide or peptide sequence for β^H would be useful in distinguishing between these alternatives.

Recently, Eaton (1980) has suggested that globins may evolve new or better functions through replacement of whole sequences in single steps. Perhaps the family of chicken β-like globin genes is a testing ground for his hypothesis.

[2] B. S. Chapman, L. E. Hood, and A. J. Tobin, manuscript in preparation.

Acknowledgments—We owe sincere thanks to Dr. James Schilling for teaching BSC amino acid sequencing techniques. Thanks also to Margaret Kowalczyk for precision illustrations and to Lori Erdley for patient assistance in the preparation of the manuscript.

Note Added in Proof—The complete sequence of a ρ globin cDNA clone obtained by I. Roninson and V. M. Ingram (manuscript in preparation) suggests that another ρ globin (ρ′) is found in domestic chickens. Careful review of our amino acid sequence data (Fig. 6) reveals small amounts of the ρ′ globin product encoded by their nucleotide sequence. The ρ′ globin differs from the sequence shown in Fig. 3 at residues 125, 129, 139, and 143, and represents approximately 15% of the total globin sequenced (Fig. 6).

REFERENCES

Beaupain, D., Martin, C., and Dieterlen-Lievre, F. (1979) *Blood* **53,** 212–225
Borgese, T., and Bertles, J. (1965) *Science* **148,** 509–511
Brown, J., and Ingram, V. (1974) *J. Biol. Chem.* **249,** 3960–3972
Chapman, B., and Tobin, A. (1979) *Dev. Biol.* **69,** 375–387
Chapman, B., Tobin, A., and Hood, L. (1980) *J. Biol. Chem.* **255,** 9051–9059
Chapman, B. (1981) in *Hemoglobins in Development and Differentiation* (Stamatoyannopoulos, G., and Nienhuis, A., eds) Alan R. Liss, Inc., New York, in press
Cirotto, C., and Geraci, G. (1975) *Comp. Biochem. Physiol. A. Comp. Physiol.* **51,** 159–163
Craig, M., and Russel, E. (1964) *Dev. Biol.* **10,** 191–201
Dayhoff, M. (1976) *Atlas of Protein Sequence and Structure.* National Biomedical Research Foundation, Silver Spring, Md.
Dodgson, J., Strommer, J., and Engel, J. (1979) in *Eukaryotic Gene Regulation* (Axel, R., Maniatis, T., and Fox, C. F., eds) pp. 383–392, Academic Press, New York
Eaton, W. (1980) *Nature* **284,** 183–185
Engel, J., and Dodgson, J. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2596–2600
Fantoni, A., Bank, A., and Marks, P. (1967) *Science* **157,** 1327–1329
Fritsch, E., Lawn, R., and Maniatis, T. (1980) *Cell* **19,** 959–972
Gale, R., Clegg, J., and Huehns, E. (1979) *Nature* **280,** 162–164
Ginder, G., Wood, W., and Felsenfeld, G. (1979) *J. Biol. Chem.* **254,** 8099–8102
Hood, L., Campbell, J., and Elgin, S. (1975) *Annu. Rev. Genet.* **9,** 305–353
Hunkapiller, M., and Hood, L. (1978) *Biochemistry* **17,** 2124–2133
Hunkapiller, M., and Hood, L. (1980) *Science* **207,** 523–525
Isaacks, R., Harkness, D., Adler, J., and Goldman, P. (1976) *Arch. Biochem. Biophys.* **173,** 114–120
Jahn, C., Hutchison, C., Phillips, S., Weaver, S., Haigwood, N., Vivola, C., and Edgell, M. (1980) *Cell* **21,** 159–168
Jelkmann, W., and Bauer, C. (1977) *Pfluegers Arch. Eur. J. Physiol.* **372,** 149–156
Jelkmann, W., and Bauer, C. (1978) *Pfluegers Arch. Eur. J. Physiol.* **377,** 75–80
Johnson, N., Hunkapiller, M., and Hood, L. (1979) *Anal. Biochem.* **100,** 335–339
Kitchen, H., and Brett, I. (1974) *Ann. N. Y. Acad. Sci.* **241,** 653–671
Kleihauer, E., and Stoffler, G. (1968) *Mol. Gen. Genet.* **101,** 59–69
Lacy, E., Hardison, R., Quon, D., and Maniatis, T. (1979) *Cell* **18,** 1273–1283
Ladner, C., Heidner, E., and Perutz, M. (1977) *J. Mol. Biol.* **114,** 385–414
Lauer, J., Shen, C., and Maniatis, T. (1980) *Cell* **20,** 119–130
Manwell, C., Baker, C., and Betz, T. (1966) *J. Embryol. Exp. Morphol.* **16,** 65–81
Matsuda, G., Takei, H., Wu, K., and Shiozawa, T. (1971) *Int. J.*

Downloaded from www.jbc.org at CALIFORNIA INSTITUTE OF TECHNOLOGY on December 24, 2006

Protein Res. **3,** 173–174
Matsuda, G., Maita, T., Mizumo, K., and Ota, H. (1973) *Nat. New Biol.* **244,** 244
McKean, D., Potter, M., and Hood, L. (1973) *Biochemistry* **12,** 749–759
Melderis, H., Steinheider, G., and Ostertag, W. (1974) *Nature* **250,** 774–776
Moss, B., and Hamilton, E. (1974) *Biochim. Biophys. Acta* **371,** 379–391
Ohno, S. (1970) *Evolution by Gene Duplication,* Springer-Verlag, New York
Pataryas, H., and Stamatoyannopoulos, G. (1972) *Blood* **39,** 688–696
Piszkiewicz, D., Landon, M., and Smith, E. (1970) *Biochem. Biophys. Res. Commun.* **40,** 1173–1178
Rossi-Fanelli, A., Antonini, E., and Caputo, A. (1958) *Biochim. Biophys. Acta* **30,** 608–615
Slightom, J., Blechl, A., and Smithies, O. (1980) *Cell* **21,** 627–638
Steinheider, G., Melderis, H., and Ostertag, W. (1972) *Haematol. Bluttransfus.* **10,** 225–235.
Steinheider, G., Melderis, H., and Ostertag, W. (1975) *Nature* **257,** 714–716
Zuckerkandl, E. (1978) *J. Mol. Evol.* **12,** 57–89

The Journal of Biological Chemistry
Downloaded from www.jbc.org at CALIFORNIA INSTITUTE OF TECHNOLOGY on December 24, 2006