Fetal Globin Expression in New World Monkeys*

(Received for publication, January 18, 1996, and in revised form, April 9, 1996)

Robert M. Johnson*, Steven Buck†, Chi-hua Chiu†, Horacio Schneider‡, Iraclilda Sampaio§, Douglas A. Gage**, Tun-Li Shen**, Maria Paula C. Schneider**, José A. Muniz††, Deborah L. Gumucio‡‡, and Morris Goodman¶

From the Departments of Biochemistry and Anatomy and Cell Biology, Wayne State Medical School, Detroit, Michigan 48201, **Mass Spectroscopy Facility, Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824, ‡Deptartment of Anatomy and Cell Biology, University of Michigan, Ann Arbor, Michigan 48109, |Departamento de Genética, Universidade Federal do Pará, Belem, Pará, Brazil, and ¶¶Centro Nacional de Primatas, Ananindeua, Pará, Brazil

Reverse phase chromatography of the globin chains of adult, newborn, and fetal erythrocytes from three species of New World monkeys (Cebus apella, Aotus azarae, and Callithrix jacchus) representing three of the seven plathyrrhine clades showed that γ-globin expression was fetal in these animals. The globins were identified by a combination of chemical sequencing and mass spectral analysis. Since γ-globin expression is fetal in the other major simian branch, the catarrhines, but embryonic in prosimian primates and nonprimate placental mammals, the evolution of fetal recruitment can now be assigned to the period between the simian-prosimian divergence (55 million years ago) and the plathyrrine-catarrhine divergence (35 million years ago). The γ-globin gene underwent tandem duplication during the same evolutionary epoch, in accord with a model that suggests that the downstream duplicated γ-gene (γ2) was free to acquire the mutations necessary for fetal recruitment. Mass spectrometric analysis of tryptic digests of the γ-globins verified the amino acid sequences deduced from genomic sequencing. Detailed analysis of high performance liquid chromatography and matrix-assisted laser desorption/ionization mass spectrometry data showed that γ2-globin in Cebus was expressed to a far greater extent than γ1-globin, supporting inferences drawn from a study of the promoter sequences. A "pre-γ"-globin was observed in C. apella and shown to be primarily the glutathionyl adduct. The other species, A. azarae and C. jacchus, also express only one γ-globin polypeptide. This work provides biochemical evidence of an evolutionary trend in the plathyrrhines to alter the duplicated γ-globin gene locus so that only one γ-globin polypeptide is expressed.

All vertebrate hemoglobins are tetramers composed of two α-type and two β-type globin chains. In humans, a cluster of related genes on chromosome 11 encodes the β-type globins: ε, γ, γ, δ, ε, δ, and β. These genes are expressed at defined times during ontogeny. In humans (1), the ε-globin gene is embryonic, expressed during the first 20% of gestation, after which it is down-regulated. The γ-genes are up-regulated at this time and are expressed for the remaining 80% of prenatal development. After birth, γ-gene expression declines as the adult δ- and β-globins become the predominant forms. The mechanism of the developmental switch from one globin gene to the next is of fundamental importance for an understanding of stage-specific gene expression. There is a possible clinical benefit as well, since an ability to maintain or revive expression of the γ-globin gene can be a route to treatment of the severe clinical conditions associated with abnormal adult β-globins, such as β-thalassemia and sickle cell anemia.

An evolutionary approach has been productive in identifying important promoter control regions. Alignment of promoter regions from globin genes with similar developmental timing has revealed conserved sequences important for transcriptional control (2-4). More recently, a detailed comparison of differences near these conserved regions in genes that are expressed either embryonically or fetally has identified base changes associated with alternative developmental timing (5). Examination of globin gene promoter structure and expression in the descendants of the primate ancestor will be particularly informative, since fetal recruitment of γ-globin occurred in this lineage (Fig. 1). The galago, a prosimian primate, has a β-globin gene cluster that differs in two significant ways from the human (2). First, the galago possesses only one β-globin gene. Second, expression of the single γ-globin gene is embryonic, rising and falling simultaneously with ε-globin expression during the first 30% of gestation. The galago δ- and β-globins are the functional globins for the remaining two-thirds of prenatal development. Thus, γ-gene expression was delayed to fetal life at some point after the divergence of the prosimian and simian lineages, about 55 million years ago.

Studies of humans and other catarrhine primates have shown that γ-globin expression is fetal in these species, but little data are available for the other major simian group, the plathyrrhines (New World monkeys). A single study has reported the presence of a globin in a newborn marmoset with an electrophoretic mobility different from the adult form, although its identification as a γ-globin was based only on the amino acid analysis of one cyanogen bromide fragment (6). We have now examined globin expression during fetal life in a group of plathyrrhine primates. Together with the earlier studies of catarrhine globin gene expression, these results place the emergence of fetal recruitment in the same 55-35 million years ago time period as the duplication of the γ-globin gene (7). In

*This work was supported by National Institutes of Health Grant HL-33940 and National Science Foundation Grant DEB9116098. It was also supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brazil) Grant 20153093-7 and Financiador de Estudos e Projetos (Brazil) Grant 66.94.0034.00. The Michigan State University-National Institutes of Health Mass Spectrometry Facility is supported in part by Grant RR-00480 from the National Institutes of Health, National Center for Research Resources, Biotechnology Area. 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.

† To whom correspondence should be addressed: Dept. of Biochemistry, Wayne State Medical School, 540 E. Canfield, Detroit, MI 48201. Tel.: 313-377-1816; Fax: 313-377-2703.
addition, evidence is presented that in the platyrhine lineage, there is a trend toward a reduction of the number of function-
ifying genes from two to one. This is associated with divergent promoter regions (8), with large scale deletion (9), and crossing over events (10).

MATERIALS AND METHODS

Blood samples were obtained from a total of 15 animals of the following species: Cebus apella (brown capped capuchin), mother 1 with fetus, mother 2 with fetus, mother 3 with a newborn and two fetuses; Callithrix jaccus (common marmoset), one mother and fetus, two other fetuses from unrelated mothers; Aotus azarae (owl monkey), one mother, together with one fetus and one newborn. The fetuses were obtained from spontaneous miscarriages occurring after one-half to two-thirds of gestational time.

Chromatography—Reverse phase HPLC1 was carried out essentially as described by Shelton et al. (11) using a Vydac C4 column on a Varian ISO 9001 apparatus. Quantitation of the amounts of the various globins was obtained by integration of the peak areas monitored at 214 nm. Globin polypeptides were purified for chemical analysis by pooling peaks from four or five runs and rechromatographing. The purified samples were lyophilized.

Hemoglobin A2 was isolated on DE52 cellulose as described by Huisman (12). Peak fractions were dialyzed against ammonium bicarbonate, lyophilized, and dissolved in 0.1% trifluoroacetic acid for HPLC analysis.

Amino acid sequencing was done with an Applied Biosystems gas phase sequencing apparatus. When required, the globins were cleaved with cyanogen bromide directly on the glass filter support used in the sequencing and sequencing was performed after washing the filter.

Mass Spectrometric Analysis—Analysis by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) were performed with a PerSeptive Biosystems Voyager Elite mass spectrometer equipped with a nitrogen laser (λ = 337 nm). Data were acquired in the linear mode of time-of-flight operation with a 50-MHz transient recorder. All experiments were performed using α-cyano-4-hydroxycinnamic acid (Aldrich) as the matrix. Saturated matrix solutions were prepared in a 1:1 solution of acetonitrile/aqueous 0.1% trifluoroacetic acid. To prepare the sample, 1 μl of the globin or tryptic digest was added to 1 μl of the matrix solution and applied to a stainless-steel sample plate. The mixture was allowed to air dry before being introduced into the mass spectrometer.

Tryptic Fragment Analysis—Tryptic digestions were carried out in 400 μl Tris-Cl, pH 8.0, 6 mM guanidine, 40 mM CaCl2, heated to 90 °C for 15 min. Sequencing grade trypsin was then added at a protein/enzyme ratio of approximately 20 (w/w). The mixture was then incubated at 37 °C in a water bath for 24 h. Digestion mixtures were analyzed by MALDI-MS without further fractionation or purification.

TCEP Reduction—Protein samples were mixed with 50 volumes of 10 mM tris[2-carboxyethyl]phosphine hydrochloride (TCEP, Pierce) and incubated at 37 °C for 30 min.

Genomic Sequencing—The γ-globin region of C. jaccus was obtained by Long polymerase chain reaction as described elsewhere (10).

RESULTS

The results of chromatographic separation of the globin chains are shown in Fig. 2 with the peak identifications. The amino acid sequences for C. apella α- and β-globins are known (13), as well as genomic sequences for the γ1- and γ2-globin regions (10), and this species was selected for detailed analysis and verification of the peak assignments. The C. apella γ-globin amino acid sequences are identical to the previously published Cebus albifrons sequences (8). For convenience, the major HPLC peaks are labeled A-E on the fetal C. apella sample, which contains the relevant globin chains.

α- and β-Globin—Peak A was identified as α-globin because of its constant presence in all blood samples. Moreover, the N-terminal sequence of peak A from C. apella was VHS-PAAEK, identical with the known α-globin sequence (13). Peak C is β-globin, with an N-terminal sequence of VHLTAAEK, in agreement with the published β-globin sequence (13). It was found in adult samples at an approximate 1:1 ratio with α-globin.

Peak B eluted at the same position as human "pre-β"-globin, which has been described as a mixture of post-translationally modified β-globin polypeptides (14), sometimes including an adduct with GSH. Edman degradation of C. apella peak B yielded the β-globin N-terminal sequence, and MALDI-MS found two major polypeptides with masses of 16,013 Da (the predicted β-globin molecular mass) and 16,318 Da, the mass expected for a β-globin-GSH adduct. Phosphine reduction of the pre-β-globin to remove the glutathionyl residue eliminated the 16,318 peak, which was shifted downward to be coincident with the 16,013 mass peak (data not shown). Thus, the HPLC peak B was a C. apella pre-β-globin, consisting largely of glutathionyl-β-globin.

Δ-Globin—A late eluting minor peak, seen only in the adult and newborn samples, was identified as δ-globin. Hemoglobin A2 was prepared by DEAE chromatography of un-denatured hemoglobin. Upon reverse phase HPLC, the late-eluting putative δ-globin was recovered, together with α-globin. In addition, Edman degradation of the putative δ-chain yielded the N-terminal sequence VHLTGEKAAVA. Although the δ-globin sequence for C. apella is not known, inspection of the known sequences of δ-globins from New World monkeys (Fig. 3) indicates that all have alanine at position 12, which distinguishes them from the otherwise nearly identical β-chains.

γ-Globins—Peaks D and E are candidate γ-globins because of their restriction to fetal samples. N-terminal sequencing of the purified Peaks D and E was unsuccessful. This was not unexpected, since the deduced amino acid sequences of the C. apella γ-globins (10) (identical to C. albifrons (8)) predicts an N-terminal serine. Such serines are normally acetylated in globins (for review, see Ref. 1).

The Major γ-Globin (Peak E)—For positive identification, advantage was taken of the two methionines of the γ sequence. Cyanogen bromide cleavage was performed on the larger peak E polypeptide, still bound to the glass filter used as a support for N-terminal sequencing. Thirteen cycles of Edman degradation yielded two amino acids at each cycle as follows: GV, NA, PG, KV, VA, KS, AA, HL, GG, AS, KR, VY, LH. These are precisely those expected from the γ-globin sequences (Fig. 4). These cyanogen bromide fragment sequences do not distinguish which of the two C. apella γ-chains was present. C. apella has two γ-globin genes (10), but γ1 and γ2 differ in only one amino acid residue (Fig. 4), position 116 (V in γ1, I in γ2). There

1 The abbreviations used are: HPLC, high performance liquid chromatography; MALDI-MS, matrix-assisted laser desorption/ionization mass spectrometry; TCEP, tris[2-carboxyethyl]phosphine hydrochloride.
Fig. 2. HPLC chromatography of blood samples. a, C. apella. Top, adult female; center, fetus; bottom, newborn. On the trace for the fetal sample, the major HPLC peaks discussed in the text are denoted A–E. b, C. jacchus mother and fetus. c, A. azarae mother and fetus. The assignment of the ß-globin in A. azarae is tentative.
is no convenient cleavage site in the vicinity of this residue, precluding identification by Edman sequencing. Therefore, MALDI-MS was used to determine the molecular weight of the intact globin (Fig. 5). Within experimental error (±0.1%), the measured mass [M + H]⁺ at m/z 15,852 was the same as the predicted value for the γ2 gene product, which is 15,805 plus a mass of 42 for N-terminal acetylation. The identification was confirmed by mass spectrometric analysis of a tryptic digest of C. apella HPLC fraction E (Fig. 5), which showed peaks corresponding to the expected [M + H]⁺ ions of peptides for the γ2 sequence (Table 1). Nearly the entire amino acid sequence predicted by the genomic analysis (10) was verified by this mass analysis, with the exception of residues 61–83, which produce a number of tryptic fragments too small to be detected in the MALDI spectra (i.e. below mass 500, where matrix-related peaks predominate). Most importantly, the fragment containing residues 106–121 with the single amino acid difference between γ1 and γ2 at position 116 has a predicted molecular mass of 1684 for γ2 and 1670 for γ1. Only the 1684 mass peptide was found, showing that the larger γ-globin peak E was γ2-globin.

The Minor γ-globin (Peak D)—The smaller HPLC peak D was largely the GSH adduct of γ2-globin. MALDI spectra of HPLC peak D always found two species with molecular masses at 16,162 and approximately 15,840 (Fig. 6). The larger molecular ion is close to the predicted mass for glutathionyl-γ2-globin, with [M + H]⁺ at m/z = 16,157 Da (15,852 for γ2-globin plus 305 for glutathionyl). Phosphorylation reduces, which cleaves the disulfide linkage to glutathione, shifted the 16,162 mass by +16 Da (16,178 Da), verifying the assignment of the disulfide linkage to glutathione, shifted the 16,162 mass by +16 Da (16,178 Da), verifying the assignment. MALDI-MS of a tryptic digest of the minor γ-globin peak D that survives reduction is uncertain. Incomplete reduction by TCEP was ruled out by MALDI-MS of the tryptic digest of reduced peak D, in which no fragments characteristic of the glutathionyl adduct were seen. There are at least two other possibilities. Schroeder et al. (14) found that human pre-β contained a post-translationally modified β-globin of uncertain chemical structure. Thus, the non-reducible component of HPLC peak D may represent a similar modification of γ2-globin. Alternatively, the residual material may represent a low level of expression of γ1-globin, although no peptide fragment characteristic of γ1-globin was observed in the MALDI-MS of a tryptic digest of HPLC peak D. An upper limit for the amount of γ1 expression can be estimated in the following way. Integration of the HPLC peaks (Fig. 2a) showed that the minor γ-globin peak D is 15% of the total γ-globins, and, in turn, MALDI-MS of D found that at least 75% was the glutathionyl derivative of γ2-globin. Thus, γ1 can contribute no more than 4% of the expressed γ-globin. C. apella γ2-globin is expressed at levels that are at least 20 times greater than γ1.

Other Species—The chromatographic separation of globins from A. azarae and C. jacchus is shown in Fig. 2, b and c. The globins were identified by comparison with C. apella and by the timing of their appearance in fetal or adult life. MALDI-MS analysis of tryptic digests of purified γ-globin from A. azarae (data not shown) verified the amino acid sequence deduced from genomic DNA sequencing (10). Note the presence of two fetal globin HPLC peaks in A. azarae, which has only one γ-globin gene (10). In this instance, the smaller peak can only be a “pre-γ-γ” resulting from a post-translational modification. Insufficient sample prevented direct chemical proof of this point, however.

C. jacchus, the common marmoset, has two γ-globin genes. The genomic coding sequence of the two genes and the deduced amino acid sequence are shown in Fig. 7. Both γ-gene codes for identical polypeptides. There is a single base difference between the genes in exon 2, but the two codons are synonymous and the amino acid is Phe in both chains. MALDI-MS analysis of the intact HPLC-purified γ-chain found a single polypeptide, and a tryptic digest yielded fragments consistent with the deduced amino acid sequence (data not shown). A “pre-γ-γ”-globin is also present in C. jacchus.

Timing of Globin Expression—All of the fetuses examined were post-embryonic. No embryonic chain or ε-globin chains were observed with HPLC, although elution conditions were varied normally, intraerythrocytic human β-globin has no GSH adduct (14, 15), but significant amounts can form during storage (16, 17).

Quantitation of γ1-globin Polypeptide—After treatment of the HPLC peak D with TCEP and reapplication to reverse phase HPLC, about 75% of the material was found to elute at the E peak position, indicating that most of the D peak was the GSH adduct of γ2-globin. The precise identity of the 25% of peak D that survives reduction is uncertain. Incomplete reduction by TCEP was ruled out by MALDI-MS of the tryptic digest of reduced peak D, in which no fragments characteristic of the glutathionyl adduct were seen. There are at least two other possibilities. Schroeder et al. (14) found that human pre-β contained a post-translationally modified β-globin of uncertain chemical structure. Thus, the non-reducible component of HPLC peak D may represent a similar modification of γ2-globin. Alternatively, the residual material may represent a low level of expression of γ1-globin, although no peptide fragment characteristic of γ1-globin was observed in the MALDI-MS of a tryptic digest of HPLC peak D. An upper limit for the amount of γ1 expression can be estimated in the following way. Integration of the HPLC peaks (Fig. 2a) showed that the minor γ-globin peak D is 15% of the total γ-globins, and, in turn, MALDI-MS of D found that at least 75% was the glutathionyl derivative of γ2-globin. Thus, γ1 can contribute no more than 4% of the expressed γ-globin. C. apella γ2-globin is expressed at levels that are at least 20 times greater than γ1.

Other Species—The chromatographic separation of globins from A. azarae and C. jacchus is shown in Fig. 2, b and c. The globins were identified by comparison with C. apella and by the timing of their appearance in fetal or adult life. MALDI-MS analysis of tryptic digests of purified γ-globin from A. azarae (data not shown) verified the amino acid sequence deduced from genomic DNA sequencing (10). Note the presence of two fetal globin HPLC peaks in A. azarae, which has only one γ-globin gene (10). In this instance, the smaller peak can only be a “pre-γ-γ” resulting from a post-translational modification. Insufficient sample prevented direct chemical proof of this point, however.

C. jacchus, the common marmoset, has two γ-globin genes. The genomic coding sequence of the two genes and the deduced amino acid sequence are shown in Fig. 7. Both γ-gene codes for identical polypeptides. There is a single base difference between the genes in exon 2, but the two codons are synonymous and the amino acid is Phe in both chains. MALDI-MS analysis of the intact HPLC-purified γ-chain found a single polypeptide, and a tryptic digest yielded fragments consistent with the deduced amino acid sequence (data not shown). A “pre-γ-γ”-globin is also present in C. jacchus.

Timing of Globin Expression—All of the fetuses examined were post-embryonic. No embryonic chain or ε-globin chains were observed with HPLC, although elution conditions were varied normally, intraerythrocytic human β-globin has no GSH adduct (14, 15), but significant amounts can form during storage (16, 17).

Quantitation of γ1-globin Polypeptide—After treatment of the HPLC peak D with TCEP and reapplication to reverse phase HPLC, about 75% of the material was found to elute at the E peak position, indicating that most of the D peak was the GSH adduct of γ2-globin. The precise identity of the 25% of peak D that survives reduction is uncertain. Incomplete reduction by TCEP was ruled out by MALDI-MS of the tryptic digest of reduced peak D, in which no fragments characteristic of the glutathionyl adduct were seen. There are at least two other possibilities. Schroeder et al. (14) found that human pre-β contained a post-translationally modified β-globin of uncertain chemical structure. Thus, the non-reducible component of HPLC peak D may represent a similar modification of γ2-globin. Alternatively, the residual material may represent a low level of expression of γ1-globin, although no peptide fragment characteristic of γ1-globin was observed in the MALDI-MS of a tryptic digest of HPLC peak D. An upper limit for the amount of γ1 expression can be estimated in the following way. Integration of the HPLC peaks (Fig. 2a) showed that the minor γ-globin peak D is 15% of the total γ-globins, and, in turn, MALDI-MS of D found that at least 75% was the glutathionyl derivative of γ2-globin. Thus, γ1 can contribute no more than 4% of the expressed γ-globin. C. apella γ2-globin is expressed at levels that are at least 20 times greater than γ1.

Other Species—The chromatographic separation of globins from A. azarae and C. jacchus is shown in Fig. 2, b and c. The globins were identified by comparison with C. apella and by the timing of their appearance in fetal or adult life. MALDI-MS analysis of tryptic digests of purified γ-globin from A. azarae (data not shown) verified the amino acid sequence deduced from genomic DNA sequencing (10). Note the presence of two fetal globin HPLC peaks in A. azarae, which has only one γ-globin gene (10). In this instance, the smaller peak can only be a “pre-γ-γ” resulting from a post-translational modification. Insufficient sample prevented direct chemical proof of this point, however.

C. jacchus, the common marmoset, has two γ-globin genes. The genomic coding sequence of the two genes and the deduced amino acid sequence are shown in Fig. 7. Both γ-gene codes for identical polypeptides. There is a single base difference between the genes in exon 2, but the two codons are synonymous and the amino acid is Phe in both chains. MALDI-MS analysis of the intact HPLC-purified γ-chain found a single polypeptide, and a tryptic digest yielded fragments consistent with the deduced amino acid sequence (data not shown). A “pre-γ-γ”-globin is also present in C. jacchus.

Timing of Globin Expression—All of the fetuses examined were post-embryonic. No embryonic chain or ε-globin chains were observed with HPLC, although elution conditions were varied normally, intraerythrocytic human β-globin has no GSH adduct (14, 15), but significant amounts can form during storage (16, 17).
extensively in an attempt to detect additional polypeptides. More importantly, in all samples, the amount of adult $\alpha$-chain was equivalent to the sum of the $\beta$- and $\gamma$-chains, indicating that embryonic $\gamma$-globin was not present. It is very significant, therefore, that in all of the fetal erythrocytes examined, substantial amounts of $\gamma$-globin polypeptides were seen (Table II), demonstrating that $\gamma$-globin expression is fetal in these platyrhine species, as it is in human and other catarrhines (6, 18). Thus, the evolutionary events that resulted in $\gamma$-gene expression during fetal life predate the catarrhine-platyrhine divergence.

There were significant amounts of $\beta$-globin in all the fetal

---

**Fig. 5. MALDI spectra of the major $\gamma$-globin of C. apella (HPLC peak E).** Top, intact polypeptide. Bottom, the unfractionated tryptic digest, which contains a peak at m/z 1684 corresponding to fragment 106-124 of $\gamma 2$. This excludes the possibility that CaE represents $\gamma 1$, since the Val-116 substitution would shift the mass of this peptide to m/z 1670.
The expected tryptic peptides and their masses were deduced from the DNA sequences of the Cebus \( \gamma \)-globins (10). All of the predicted peptides are identical for \( \gamma 1 \) and \( \gamma 2 \), except for the 105-120 peptide, which contains the single amino acid difference. CaE yielded only the 1685 mass fragment expected for \( \gamma 2 \).

| Fragment | Predicted peptides from \( \gamma \) sequence | \( \text{MH}^+ \) (mass) predicted | \( \text{MH}^+ \) (mass) observed in CaE |
|----------|---------------------------------------------|----------------------------------|----------------------------------|
| 1–18     | Acetyl-SNFTAEDKAITSLWAK                     | 1896.98                          | 1894                             |
| 18–30    | NVVEDAGGETLGR                               | 1316.64                          | 1318                             |
| 31–40    | LLVVYPWTQR                                  | 1274.73                          | 1276                             |
| 41–59    | FFDSFGSLSPSAIAGNPK                          | 1988.94                          | 1992                             |
| 83–104   | GTFGQLSEIHDCKLHVDPNFR                      | 2542.21                          | 2545                             |
| 105–120(\( \gamma 2 \)) | LLGNVLTVLAHHGK                         | 1684.03                          | 1684.6                           |
| 105–120(\( \gamma 1 \)) | LLGNVLTVLHMGK                          | 1669.60                          |                                   |
| 121–132  | EFTPEQASWWQK                                | 1449.70                          | 1451                             |
| 133–146  | MVAVGASALGSRYH                              | 1418.72                          | 1420                             |

**DISCUSSION**

The data presented here, combining HPLC and mass spectroscopic analysis, establish that \( \gamma \)-globin expression occurs in fetal life in platyrrhines, the New World monkeys. Since it is known that that humans (1) and other catarrhines (6, 18) exhibit fetal expression of \( \gamma \)-globin, the evolutionary emergence of fetal expression can now be placed earlier than the catarrhine-platyrrhine divergence (35 million years ago). Because \( \gamma \)-globin expression is embryonic in prosimians (2), the appearance of fetal expression occurred after the prosimian-simian divergence (55 million years ago). Thus, the sequence changes in the \( \beta \)-globin cluster that led to fetal recruitment of \( \gamma \)-globin occurred between 55 and 35 million years ago. Fitch et al. (7) have shown that the duplication of the stem simian \( \gamma \)-gene took place in the same time period. They suggested that the duplication of the \( \gamma \)-gene supplied a redundant \( \gamma \)-gene copy which was then free to collect the regulatory and structural sequence changes required for fetal expression and function. The finding that \( \gamma \)-gene duplication and fetal recruitment are closely linked in evolutionary time is in accord with this hypothesis.

![MALDI spectra of the minor \( \gamma \)-globin of C. apella (HPLC peak D).](http://www.jbc.org/)

There is a prominent peak at \( \text{m/z} \) 16,162 for the protonated molecule, with a corresponding doubly charged peak at \( \text{m/z} \) 8081. A shoulder on the low mass side of the \([M + H]^+\) peak has an \( \text{m/z} \) of 15,850, consistent with an unmodified \( \gamma \)-globin. Reduction of the sample with TCEP (inset) produced a shift in the mass of the larger \([M + H]^+\) from \( \text{m/z} \) 16,162 to 15,842, as expected for the reductive cleavage of a glutathione adduct.

**FIG. 6.** MALDI spectra of the minor \( \gamma \)-globin of C. apella (HPLC peak D). There is a prominent peak at \( \text{m/z} \) 16,162 for the protonated molecule, with a corresponding doubly charged peak at \( \text{m/z} \) 8081. A shoulder on the low mass side of the \([M + H]^+\) peak has an \( \text{m/z} \) of 15,850, consistent with an unmodified \( \gamma \)-globin. Reduction of the sample with TCEP (inset) produced a shift in the mass of the larger \([M + H]^+\) from \( \text{m/z} \) 16,162 to 15,842, as expected for the reductive cleavage of a glutathione adduct.
have assigned a prominent role to the proximal CCAAT box (19, 20), and it was predicted that the presence of a non-canonical CCAAT sequence (CCAAC) would markedly down-regulate the \( g_1 \) gene (5, 21). The measurements of the expressed globin polypeptides presented here verify this hypothesis, since the ratio of \( g_2 \) to \( g_1 \) expression was at least 20:1 in \( C. \) apella.

The very low expression of \( g_1 \) in \( Cebus \) species is a manifestation of a striking evolutionary trend in platyrhines toward reduction of two functional \( g \)-genes to one (8–10, 22). The platyrhine genera can be grouped in seven major clades (23), which can in turn be grouped into the families of Atelidae, Pitheciidae, and Cebidae (24). In the four genera of the Atelidae clade, a large deletion in \( g_1 \) renders it nonfunctional (9). The Pitheciidae family contains two clades. In one, represented by Callithrix species, the proximal CCAAT box of the \( g_1 \) promoter has the same disabling base change (CCAAt \[3\] CCAAc) noted in the \( C. \) apella \( g_1 \) promoter (10). In Pitheciini, the other major clade of this family, the base change CCAAT to CCgAT is found in the proximal CCAAT box of the \( g_1 \) promoter (10). The remaining four clades comprise the Cebidae family. In Cebus species, as we show here, only \( g_2 \)-globin is expressed at observable levels. In each of the two clades represented by \( A. \) azarae and Saimiri species, only a single \( g \)-globin gene exists, but in contrast with Atelidae, the mechanism was an unequal crossover event that has led to the formation of a single hybrid \( g \)-gene from the original duplicated pair (10). Only \( C. \) jacchus, representing the seventh clade, has two \( g \)-genes which are free of disabling promoter mutations (10), but code for identical polypeptides. It is unclear why significant expression from both \( g \)-globin genes has been preserved in the catarrhines while only one \( g \)-globin polypeptide is expressed in the New World monkeys.

The single hybrid \( g \)-globin gene in \( A. \) azarae has the promoter and 5'-flanking regions of a \( g_1 \) gene. The fact that this hybrid gene is expressed at much higher levels than \( g_1 \) of \( C. \) apella (but less than \( g_2 \); Table II) may be related to the fact that the proximal CCAAT box of \( A. \) azarae, although non-canonical (CcaaAAT \[8\] CCAAAC), is not the same variant as the \( C. \) apella \( g_1 \) promoter (CCAAT \[10\] CCAAAC).

The occurrence of gene duplication and fetal recruitment in the same evolutionary epoch suggests a number of possible scenarios. 1) Fetal recruitment might have preceded duplication. 2) If gene duplication occurred first, producing a redundant gene for fetal recruitment, subsequent gene conversions conferred fetal characteristics on the the second \( g \)-gene. Conversions between the \( g \)-globin genes have been frequent (25-
27), but the first explanation cannot be ruled out at the present time. The second model makes the interesting prediction that at the early stages of fetal recruitment, one of the duplicated \( \gamma \)-genes would have remained embryonic while the other was evolving toward fetal expression. It is possible that in some platyrhine species, the \( \gamma \)-globin gene may be expressed in embryonic life and thus represent a survival of this situation. A test of this suggestion will require analysis of hematopoietic tissues from early embryos of these species.

A method known as phylogenetic footprinting (2–4) identifies promoter sequence regions that have been conserved throughout mammalian evolution and are therefore likely to be functionally significant. Recently, "differential" phylogenetic footprinting was introduced, which pinpoints lineage-specific base changes near the conserved promoter elements. This analysis (5) suggests that at least one of the changes associated with fetal recruitment is an alteration in sequences immediately upstream from the proximal CCAAT box of the \( \gamma \)-gene. The changes, which occurred in the stem anthropoid ancestor, seriously impair the binding of several trans factors (G1, G2, G3, G4) which bind avidly to prosimian \( \gamma \) promoter sequences and to all \( \epsilon \)-gene promoters, i.e. to all embryonically expressed genes, but not to \( \gamma \)-genes that are fatally expressed. It remains to be determined whether these changes are necessary and sufficient to explain fetal expression of the \( \gamma \)-globin genes. Promoter sequence changes are not the only possible mechanism which might have facilitated fetal recruitment. The \( \epsilon \)-to \( \gamma \)-distance is 6 kb in the prosimian galago, but the simian \( \gamma \) duplication increased the distance between \( \epsilon \) and the downstream \( \gamma \)-2 to 11 kb. It is possible that this increased distance prevents the LCR from interacting with the \( \epsilon \)-promoter and the \( \gamma \)-2 promoter at the same time. This has been proposed as one of the events that allowed the duplicated \( \gamma \)-gene to escape the requirement for expression during the embryonic period (22).

The demonstration that \( \gamma \)-globin expression is fetal in the New World monkeys will allow the inclusion of the known promoter sequences (8, 10, 21) from these species in the sequence comparisons and phylogenetic tests that aim to define the elements associated with fetal recruitment more precisely. Another valuable comparison will be the alignment of platyrhine \( \gamma \)-gene promoters with catarrhine sequences, with the aim of detecting those differences in the promoters that are associated with the delay of \( \gamma \)-gene inactivation into postnatal life in humans.

Acknowledgments—We thank J une Snow for chemical sequencing, Mark Stout for the use of equipment, and J erry Slightom for critical reading of the manuscript.

REFERENCES
1. Bunn, H., and Forget, B. (1986) Hemoglobin: Molecular, Genetic and Clinical Aspects, W. B. Saunders Co., Philadelphia, PA
2. Tagle, D. A., Koop, B. F., Goodman, M., Slightom, J. L., Hess, D. L., and Jones, R. T. (1988) J. Mol. Biol. 203, 439–455
3. Gumucio, D. L., Hellstedt-Williamson, H., Gray, T. A., Tarle, S. A., Shelton, D. A., Tagle, D. A., Slightom, J. L., Goodman, M., and Collins, F. S. (1992) Mol. Cell. Biol. 12, 4910–4929
4. Gumucio, D. L., Shelton, D. A., Bailey, W. J., Slightom, J. L., and Goodman, M. (1993) Proc. Natl. Acad. Sci. U.S.A. 90, 6018–6022
5. Gumucio, D. L., Shelton, D. A., Blanchard-McQuate, K., Gray, T., Tarle, S., Hellstedt-Williamson, H., Slightom, J. L., Collins, F., and Goodman, M. (1994) J. Biol. Chem. 269, 15371–15380
6. Huisman, T. H., Schroeder, W. A., Keeling, M. E., Gengoian, N., Miller, A., Brodie, A. R., Shelton, J. R., et al. (1973) Biochem. Genet. 10, 309–318
7. Fitch, D. H., Bailey, W. J., Tagle, D. A., Goodman, M., Siew, L., and Slightom, J. L. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 7396–7400
8. Hayasaka, K., Skinner, C. G., Goodman, M., and Slightom, J. L. (1993) Genomics 18, 20–28
9. Meireles, C. M., Schneider, M. P., Sampaio, M. I., Schneider, H., Slightom, J. L., Chu, C., Neiswanger, K., Gumucio, D. L., Czelsniak, J., and Goodman, M. (1995) Proc. Natl. Acad. Sci. U.S.A. 92, 2607–2611
10. Chu, C.-H., Schneider, H., Schneider, M., Sampaio, I., Meireles, C. M., Slightom, J., Gumucio, D., and Goodman, M. (1996) Proc. Natl. Acad. Sci. U.S.A., in press
11. Shelton, J., Shelton, J., and Schroeder, W. A. (1984) J. Liq. Chromatogr. 7, 1969–1977
12. Huisman, T. (1986) in The Hemoglobinopathies (Huisman, T., ed) pp. 1–31, Churchill-Livingstone, Edinburgh, UK
13. Matsuda, G., Maia, T., Watanabe, B., Araya, W., Morokuma, K., Ota, Y., Goodman, M., Barnabas, J., and Prychodko, W. (1973) Hoppes-Seyler’s Z. Physiol. Chem. 354, 1513–1516
14. Schroeder, W. A., Shelton, J. B., Huynh, V., and Shelton, J. R. (1986) Hemo-
15. Birchs, W., Tushmaid, P. E., and Winterhalter, K. H. (1973) Biochemistry 12, 3667–3672
16. Allen, D., and Jandl, J. (1961) J. Clin. Invest. 40, 454–474
17. Huisman, T. H., Dozy, A. M., Horton, B. F., and Neshland, C. M. (1966) J. Lab. Clin. Med. 67, 305–373
18. De Simone, J., and Mueller, A. (1979) Blood 53, 19–27
19. Myers, R. M., Tilly, K., and Maniatis, T. (1986) Science 232, 613–618
20. Grosveld, F., Antoniou, M., Berry, M., de Boer, E., Dillon, N., Ellis, J., Fraser, P., Hurst, J., Imam, A., Majer, D., Philpisen, S., Pruzina, S., Troubolds, J., and Whyatt, D. (1991) Cold Spring Harbor Symp. Quant. Biol. 56, 7–13
21. Hayasaka, K., Fitch, D. H., Slightom, J. L., and Goodman, M. (1992) J. Mol. Biol. 224, 875–881
22. Goodman, M., Slightom, J., and Gumucio, D. (1996) in Gene Families: Structure, Function, Genetics, and Evolution (Holmes, R., and Lim, H., eds) World Scientific Publishing, Singapore, in press
23. Schneider, H., Schneider, M. P., Sampaio, I., Harada, M. L., Stanhope, M., Czelsniak, J., and Goodman, M. (1995) Mol. Phylogenet. Evol. 4, 331–349
24. Harada, M., Schneider, H., Schneider, M., Sampaio, I., Czelsniak, J., and Goodman, M. (1995) Mol. Phylogenet. Evol. 2, 207–225
25. Slightom, J. L., Theisen, T. W., Koop, B. F., and Goodman, M. (1987) J. Biol. Chem. 262, 7474–7483
26. Slightom, J. L., Koop, B. F., Xu, P., and Goodman, M. (1988) J. Biol. Chem. 263, 12427–12438
27. Fitch, D. H., Mainone, C., Goodman, M., and Slightom, J. L. (1990) J. Mol. Biol. 225, 781–793
28. Nute, P. E., and Sullivan, B. (1971) Comp. Biochem. Physiol. B Comp. Bio-
29. Tanokka, Y., Araya, A., Maia, T., and Matsuda, G. (1986) Biol. Chem. Hoppe-
30. Boyer, S. H., Crosby, E. F., Fuller, G. L., Noyes, A. N., and Adams, J. G. (1969) Ann. N. Y. Acad. Sci. 165, 360–377
31. Boyer, S. H., Crosby, E. F., Noyes, A. N., Fuller, G. F., Leslie, S. E., Donaldson, L. J., Vrablik, G. R., Schaefer, E. Jr., and Thurmon, T. F. (1970) Biochem. Genet. 5, 405–448
32. Spritz, R. A., and Giebel, L. B. (1988) Mol. Biol. Evol. 5, 21–29

Downloaded from http://www.jbc.org/ by guest on April 26, 2019
Additions and Corrections

Vol. 271 (1996) 14684–14691

Fetal globin expression in New World monkeys.

Robert M. Johnson, Steven Buck, Chi-hua Chiu, Horacio Schneider, Iracilda Sampaio, Douglas A. Gage, Tun-Li Shen, Maria Paula C. Schneider, José A. Muniz, Deborah L. Gumucio, and Morris Goodman

Page 14689, Table I: The letter A was omitted in the seventh sequence of the list. The correct table is shown below:

| Fragment | Predicted peptides from γ sequence | MH⁺ (mass) predicted | MH⁺ (mass) observed in CaE |
|----------|-----------------------------------|----------------------|---------------------------|
| 1–18     | Acetyl-SNFTAEDKAAITSLWAK          | 1896.98              | 1894                      |
| 18–30    | VNVDEGGETLR                       | 1316.64              | 1318                      |
| 31–40    | LLVVPWTR                          | 1274.73              | 1276                      |
| 41–59    | FFDSFSSLSSPSAIMNK                 | 1988.94              | 1992                      |
| 83–104   | GTPQGSELHCDKLHDVFENFR             | 2542.21              | 2545                      |
| 105–120(γ2) | LLGNVLTVLAIHGK                     | 1684.03              | 1684.6                    |
| 105–120(γ1) | LLGNVLTVLAIHGK                     | 1669.60              |                           |
| 121–132  | EFTEPEQASWQK                      | 1449.70              | 1451                      |
| 133–146  | MVAGVASALGSRHY                    | 1418.72              | 1420                      |

We suggest that subscribers photocopy these corrections and insert the photocopies at the appropriate places where the article to be corrected originally appeared. Authors are urged to introduce these corrections into any reprints they distribute. Secondary (abstract) services are urged to carry notice of these corrections as prominently as they carried the original abstracts.