Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 5970–5974, October 1983
Genetics

Molecular evolution of the human adult α-globin-like gene region: Insertion and deletion of *Alu* family repeats and non-*Alu* DNA sequences

(gene duplication/repetitive sequences/concerted evolution/gene conversion)

JOHN F. HESS*, MICHAEL FOX†, CARL SCHMID†, AND CHE-KUN JAMES SHEN*‡

*Department of Genetics and †Department of Chemistry, University of California, Davis, CA 95616

Communicated by M. M. Green, June 30, 1983

ABSTRACT **Previous heteroduplex studies have revealed extensive sequence homology between the two human adult α-globin-like genes (α2 and α1) and their flanking regions. These homologous regions, which are interrupted by two blocks of nonhomology, each span approximately 4 kilobases [Lauer, J., Shen, C.-K. J. & Maniatis, T. (1980) *Cell* 20, 119–130]. We have determined 3 kilobases of DNA sequences within and flanking the nonhomologous blocks of these two tandem duplication units. A total of three *Alu* family repeats has been identified. Two of them are approximately 300 base pairs long and define the 3′ ends of the first homology blocks. The third *Alu* family member is a 600-base-pair-long sequence consisting of two monomeric *Alu* members arranged in a head-to-tail fashion. It is located in the 3′ portion of the first block of nonhomology in α2-gene-containing unit. We present direct evidence that this dimeric *Alu* sequence was inserted at a staggered break. The second nonhomology block is the result of insertion or deletion of a 224-base-pair sequence. From these data and the calculation of sequence divergence, we propose a history for the evolution of the human adult α-globin-like gene region. We also suggest that DNA insertion elements may disrupt gene correction processes in the two duplication units containing α2- and α1-globin genes.**

The human α-globin-like and β-globin-like gene clusters are excellent systems to study the mechanisms of eukaryotic gene expression, molecular evolution, and the molecular basis of genetic diseases (reviewed in refs. 1 and 2). Recent molecular cloning studies have established the physical maps of the β-globin-like gene cluster and the α-globin-like gene cluster (for references, see ref. 2). In both cases, the genes are arranged in order of their expression during development, are transcribed from the same strand, and are clustered in a contiguous stretch of DNA: 50 kilobases (kb) for the β-globin-like cluster and 30 kb for the α-globin-like cluster. The nucleotide sequences of all the globin genes and their immediate flanking regions have been determined (for references, see refs. 2 and 3), and the evolutionary relationship of the α-globin-like and β-globin-like genes has been established (4, 5).

The two human adult α-globin-like genes (α2 and α1) have almost identical nucleotide sequences (for references, see refs. 2 and 3). Comparison of restriction sites within the regions flanking the 5′ side of the α2- and α1-globin genes as well as electron microscope heteroduplex studies demonstrate that the sequence homology between the α2- and α1-globin genes extends beyond the coding regions and into the flanking intergenic regions (6). This extensive homology suggests the existence of mechanism(s) for sequence matching of the two adult α-globin genes and 5′ flanking regions (6, 7). The otherwise excellent sequence homology between the regions flanking the human α2- and α1-globin genes is interrupted by a 1-kb region of nonhomology 5′ to α2-globin gene, and by 0.5-kb and 0.2-kb regions of nonhomology located 5′ to the α1-globin gene (ref. 6; this study). Several *Alu* family members map within or near these regions of nonhomology, and they can be transcribed *in vitro* by RNA polymerase III (8). The *Alu* family is the major family of interspersed repeats in human DNA (9).

The consequence of this tandem arrangement of homologous α-globin gene sequences is the occurrence of two types of DNA deletions in *Escherichia coli* during propagation of recombinant phages (6) and deletions of one of the adult α-globin genes in patients with α-thalassemia 2 (10, 11). Thus, a detailed analysis of the sequence organization of these tandem duplication units should provide information of both evolutionary and genetic importance. A complete determination of the DNA sequences contained within and flanking the nonhomology blocks should also reveal whether they contain sequences characteristic of the prokaryotic and eukaryotic transposable DNA elements (for a review, see ref. 12).

MATERIALS AND METHODS

DNA Sequence Determination of 5′ α1-Globin Gene Region. The sequence 5′ to the α1-globin gene was determined entirely by the partial chemical cleavage method of Maxam and Gilbert (13). Plasmid DNA originated from the clone pRHα2 (6). DNA fragments analyzed were labeled at their 3′ ends by the fill-in or exchange reaction using appropriate deoxyribonucleotide[^{32}P]triphosphates (Amersham) and Klenow fragment of DNA polymerase I (Biotec, Madison, WI).

DNA Sequence Determination of 5′ α2-Globin Gene Region. The base sequence of the region 5′ to the α2-globin gene was determined primarily by shotgun cloning of *Hae* III, *Alu* I, and *Sau* 3A restriction digests of the 1.4 kb and 4.5 kb *Sma* I fragments (6) into appropriate restriction sites of M13 (mp7, mp8, and mp9 strains) phage DNAs (14). The resulting recombinants were analyzed by the dideoxy method of Sanger *et al.* (15) as adapted by Messing *et al.* (16) for use with M13. A total of 30 overlapping clones has been analyzed in this way.

RESULTS

Sequence Determination. Three homologous regions and three nonhomology blocks between the two adult α-globin gene-containing duplication units have been identified by heteroduplex analysis (6). Proudfoot and Maniatis (17) have tentatively assigned the boundaries of the two duplication units by sequence comparison, and they termed these three homology

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); bp, base pair(s).

‡To whom reprint requests should be addressed.

blocks detected by electron microscopy X, Y, and Z, respectively. In this report, the three nonhomology blocks will be referred to as nonhomology I, II, and III, respectively (Fig. 1).

As shown schematically in Fig. 1, we have determined the nucleotide sequences of the two bracketed regions, which contain the 3′ portions of the two X blocks, the nonhomology regions I and II, the two Y blocks, the nonhomology block III, and 5′ portions of the two Z blocks. These sequenced regions cover from −2,424 to −721 of the α2-globin gene-containing duplication unit and from −1,978 to −721 of the α1-globin gene-containing unit, assuming the 5′-capped nucleotide of each globin mRNA to be position +1. Approximately 170 base pairs (bp) of our DNA sequences overlap with those obtained by Michelson and Orkin (3) who have determined the sequence of the entire Z blocks including the two α-globin genes. All the DNA sequences are shown in Figs. 2 and 3. The interesting features of these DNA sequences are described below.

***Alu* Family Repeat Contained Within the X Homology Blocks.** Although the sequences of the 3′ portions of the two X blocks are highly homologous, they contain a significant number of base substitutions and small DNA insertions or deletions. The percentage of sequence divergence in the region sequenced is approximately 12.3. Both X blocks end in a copy of an *Alu* family repeat. We designate these two sequences as *Alu* 1 and *Alu* 2 (Figs. 1 and 2). The *Alu* 1 is flanked by a short direct repeat sequence (5′-A-A-A-T-A-A-A-C-T-A-A-A-A-T-C-3′ in this case) as are most other *Alu* family members (Fig. 2). This same short sequence is located on the 5′ side of *Alu* 2 but is surprisingly absent on its 3′ side. Presumably, *Alu* 1 (or *Alu* 2) existed at this same site in the ancestral adult α-globin gene region prior to the duplication event that produced the present day adult α2- and α1-globin gene-containing units. Following this duplication, we postulate that a recombination event (deletion or unequal crossing-over) eliminated the direct repeat that would otherwise be flanking the 3′ end of *Alu* 2. The 3′ terminus of this proposed recombination is undefined and could have continued to result in part or all of the region identified as nonhomology II.

A Fused Dimeric *Alu* Family Repeat Inserted Within the α2-Globin Gene-Containing Duplication Unit. The sequences immediately downstream from the two *Alu* family sequences

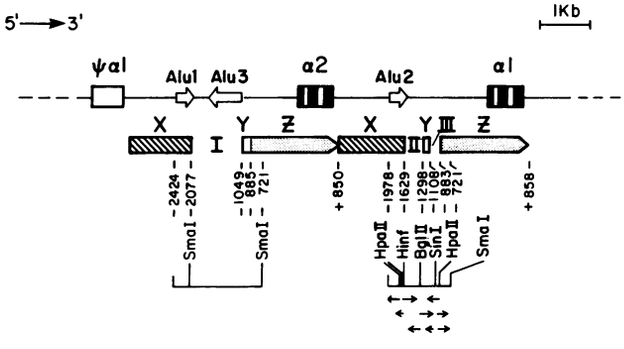

FIG. 1. Strategy for sequence determination of human adult α-globin-like gene regions. The linkage map of human ψα1-α2-α1 is derived from ref. 6. The black boxes are protein-coding sequences; the blank boxes in between them are introns. The three homology blocks X, Y, and Z and the nonhomology blocks I, II, and III are shown below the linkage map. The numbers on the boundary of each block show their positions relative to the 5′ end of α2 or α1 mRNA. A minus sign indicates the position is upstream, and a plus sign indicates it is downstream from the capping site of the mRNA. The 5′ ends of the X blocks and 3′ ends of the Z blocks contain the triply repeated sequence 5′-G-C-C-(T-G)$_4$-C-C-T-G-3′ (17) and define the boundaries of the two duplication units. Arrows below the X, Y, and Z blocks show the regions and directions of the Maxam–Gilbert DNA sequence analysis.

(*Alu* 1 and *Alu* 2) are entirely unrelated. This nonhomology appears in previous heteroduplex analysis as a 1.03-kb loop 5′ to α2-globin gene and a 0.43-kb loop 5′ to α1-globin gene (Fig. 1). Sequence analysis reveals that the exact lengths of these nonhomologous sequences, I and II, are 1,026 and 320 bp, respectively.

In the 5′ portion of the nonhomology I, there is a long stretch of the simple sequence (C-A)$_{15}$. Long runs of dinucleotide (C-A) are widely distributed in the eukaryotic genomes and can adopt a Z-DNA conformation under superhelical strain (for references, see ref. 18).

The 3′ end of the nonhomology I is occupied by a fused dimer of two *Alu* family members, the *Alu* 3 (Figs. 2 and 3), which occurs in an inverted orientation with respect to *Alu* 1 and *Alu* 2 (Fig. 1). The entire dimeric structure is flanked by a short imperfect direct repeat 5′-A-A-A-$\frac{T}{C}$-C-A-T-$\frac{G}{C}$-A-C-T-T-T-3′ (Fig. 2). This sequence is present as the unoccupied insertion site in the 5′ flanking sequences of α1-globin gene (Fig. 2). The insertion site marks the beginning of the ensuing homology between the two Y blocks. However, immediately to the left of the unoccupied insertion site flanking α1-globin gene is the short sequence 5′-A-T-C-C-C-C-C-C-A-A-A-3′, which also occurs immediately to the left of the short direct repeat on the 5′ side of *Alu* 3. The duplication of this short sequence to form the flanking direct repeats of the fused dimer is direct evidence that short dispersed repeats are inserted at staggered breaks in DNA, as has been proposed by Van Arsdell *et al.* (19) and Jagadeeswaran *et al.* (20). It has also been observed that short direct repeats form from a preexisting sequence upon insertion of repetitive sequences into the monkey α-satellite DNA (21) and the rat α-tubulin pseudogene (22).

A Non-*Alu* Deletion Sequence Within the Second Homology Y Blocks. As described above, homology corresponding to the two Y blocks resumes with the short direct repeat that flanks the dimeric *Alu* family member *Alu* 3. The homology between these two sequences is imperfect and includes many single-base substitutions. Of particular interest is a 20-bp sequence in the α1-globin unit that is not present in the α2-globin unit. This is most likely the result of deletion of the sequence 5′-G-C-A-G-G-A-G-C-T-G-G-C-C-A-G-C-C-T-C-A-3′ from the α2-globin gene-containing unit by "slipped mispairings" during DNA replication (5), because the sequence T-C-A-C-C-C-C found in α2-globin gene-containing unit is present on both sides of this 20-bp element (Fig. 2).

The Nonhomology III. Located on the 5′ side of the α1-globin gene-containing Z block is a 224-bp-long sequence that does not flank the Z block containing the α2-globin gene. This sequence, nonhomology III, is not flanked by short direct repeats and hence could result either from a simple deletion of 224 bp from the ancestral α2-globin gene-containing unit or from an insertion mechanism that does not involve the formation of short direct repeats.

Z Homology Region. We define the 5′ ends of the Z homology regions to be base −885 5′ to the α2-globin gene and base −883 5′ to the α1-globin gene. In both cases, the homology extends into and through the α-globin genes up to the 3′ end of the α-globin duplication unit (4). Our sequence analyses cover bases −885 to −721 5′ to the α2-globin gene and bases −883 to −721 5′ to the α1-globin gene (Fig. 2). There are a total of five base differences between the approximately 900-bp-long 5′ flanking regions of α2- and α1-globin genes. Four of these sequence divergences are located close to the nonhomology III (Fig. 2).

Sequence Divergence Between the Two Duplication Units. Sequence divergence resulting from point mutations has accumulated throughout the two duplication units (Table 1).

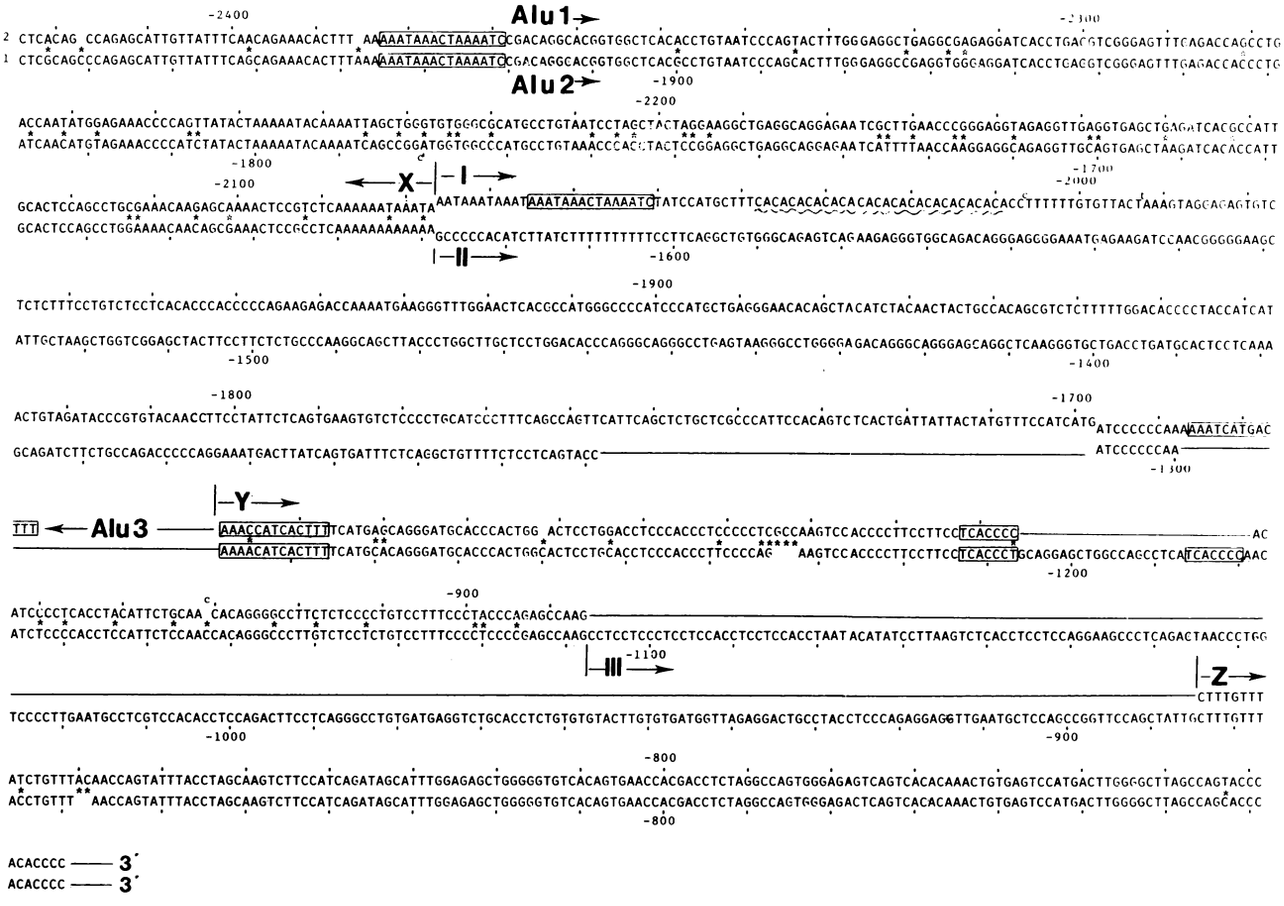

FIG. 2. Comparison of nucleotide sequences in the adult α-globin gene region as determined in this study. The regions shown cover bases $-2,424$ to -721 of the $\alpha 2$-globin gene-containing unit and bases $-1,978$ to -721 of the $\alpha 1$-globin gene-containing unit. The * denotes positions of nonhomologous bases in the X, Y, or Z blocks. The boxed sequences are the short direct repeats flanking *Alu* 1, *Alu* 2, *Alu* 3, and the 20-bp extra sequence in the $\alpha 1$-globin gene-containing Y block. Note that the direct repeat on the 3′ side of *Alu* 2 is deleted. The wavy line in the nonhomology block I denotes the stretch of poly(C-A) sequence. The fused dimeric *Alu* sequence, *Alu* 3, is located in the 3′ portion of the nonhomology block I. Its complete nucleotide sequence is shown in Fig. 3. At position -887 ($\alpha 2$), Michelson and Orkin (3) found a guanine instead of an adenine residue. For sequences beyond position -721, see ref. 3.

However, much higher sequence divergence is found in regions either near the ends of duplication units or near the nonhomology blocks. This point is discussed in more detail below.

DISCUSSION

In summary, we have determined a total of 1,704 and 1,258 bp of DNA sequences flanking the 5′ regions of the $\alpha 2$- and $\alpha 1$-globin genes, respectively. The identification of homologous and nonhomologous sequences at the level of DNA bases agrees extremely well with previous electron microscopic study (6). The base-sequence results schematically summarized in Fig. 4 show clearly that the positions of *Alu* family members (*Alu* 1, *Alu* 2, and *Alu* 3) correlate with a breakdown of homology and resumption of homology between the sequences flanking the $\alpha 2$- and $\alpha 1$-globin genes.

We propose a history of evolution of the human adult α-glo-

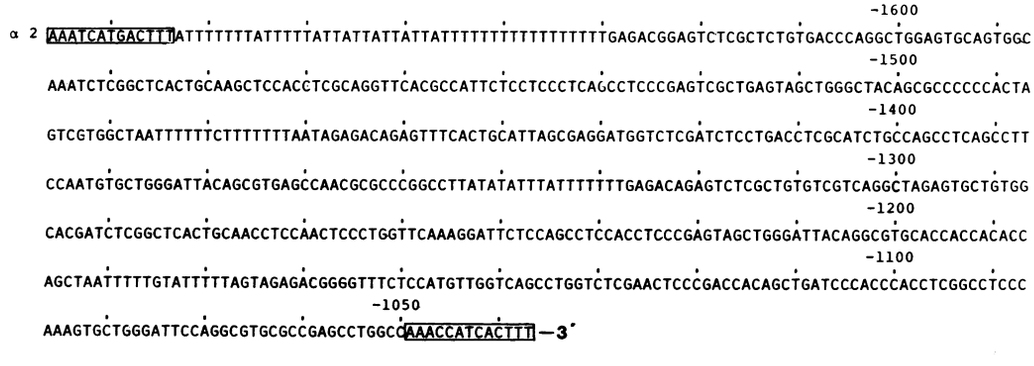

FIG. 3. Complete nucleotide sequences of the fused dimeric *Alu* family sequence *Alu* 3. The sequences of the dimeric *Alu* repeat, *Alu* 3 in Fig. 2, is shown here. Because this *Alu* sequence is arranged on the chromosome in a direction opposite that of the *Alu* 1 (or *Alu* 2), a poly(T) sequence instead of the poly(A) tail i shown immediately downstream from the 5′ short direct repeat.

Correction. In the article "Molecular evolution of the human adult α-globin-like gene region: Insertion and deletion of *Alu* family repeats and non-*Alu* DNA sequences" by John F. Hess, Michael Fox, Carl Schmid, and Che-Kun James Shen, which appeared in number 19, October 1983, of *Proc. Natl. Acad. Sci. USA* (**80**, 5970–5974), the authors request that the following changes be noted on Fig. 2. (*i*) The sequences from −730 to −721 of both the α2- and α1-strands should be C-C-A-C-C-A-C-C-C-C instead of C-C-A-C-A-C-C-C-C. (*ii*) The nucleotide at position −777 of the α1-strand should be G instead of C. (*iii*) The sequences from −890 to −881 of α2-strand should be C-C-A-A-G-T-T-T-G-T instead of C-C-A-A-G-C-T-T-T-G-T. (*iv*) The nucleotide at position −1408 of the α1-strand should be G instead of C. (All of these changes are due to typing errors.)

Table 1. Sequence divergence between tandem duplication units containing human adult α-globin-like genes

Block	Corresponding regions	Sequence divergence, bp
X	5′ portion of X blocks	1/124 (0.8%)
	Central portion of X blocks (≈500 bp)	ND
	3′ portion of X blocks	44/350 (12.3%)
Y	Y blocks	23/164 (14.0%)
Z*	5′ portion of Z blocks	4/165 (2.4%)
	Central portion of Z blocks (−1 to −720)	1/720 (0.1%)
	5′ untranslated regions	0/40
	First exon	0/93
	First intron	0/117
	Second exon	0/204
	Second intron	2/142 (1.4%)
	Third exon	0/126
	3′ untranslated regions	19/113 (17.0%)

The sequence divergence between different pairs of counterparts of the two duplication units containing human α2- and α1-globin genes is calculated and listed. The number of base differences used in the third column includes base substitutions and small (<6) DNA insertions and deletions. For example, the 7-bp extra sequence in the second intron of α1-globin gene is not considered during the calculation. Except for the 5′ portions, all the data of the Z blocks are derived from ref. 3. ND, not determined.

* From data of ref. 3.

bin-like gene region. The model consists of five major genetic events.

(i) Insertion of ancestral α-globin gene: The first stage is the formation of the ancestral duplication unit containing the ancestral α-globin gene. This may have been accomplished by a DNA transposition event of an ancestral α-globin gene before mammalian divergence, as has been proposed for the goat α-globin gene (23).

(ii) Insertion of an *Alu* family repeat: Before or after transposition of the ancestral α-globin gene, an *Alu* family repeat sequence was transposed into a site approximately 2 kb 5′ to the globin gene. This DNA insertion event must have occurred prior to the duplication of the α-globin gene because Alu family repeats having identical 5′ flanking sequences (5′-A-A-A-T-A-A-A-C-T-A-A-A-A-T-C-3′) are found in similar positions on the two tandem duplication units (*Alu* 1 and *Alu* 2; Fig. 2).

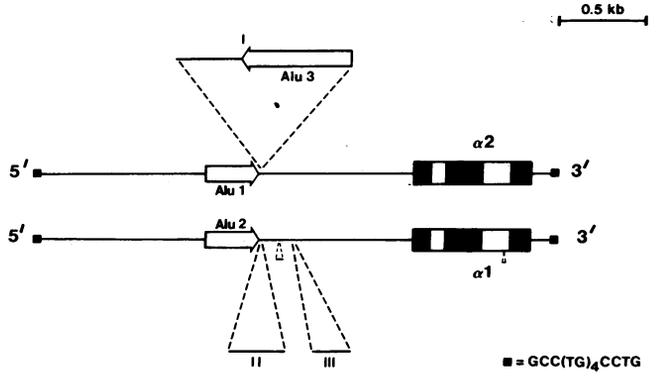

FIG. 4. Summary of the nucleotide sequence analyses in a heteroduplex representation. The main features of our sequence studies, which include the two monomeric *Alu* repeats (*Alu* 1 and *Alu* 2), the dimeric *Alu* sequence inserted in the nonhomology I, and the small deletion–insertion in the Y blocks, are pictured in this heteroduplex molecule. The small black triangle is the 7-bp extra sequence within the second intron of α1-globin gene.

(iii) Tandem duplication of 4-kb fragment containing ancestral α-globin gene and the *Alu* family repeat: The next stage of human α-globin gene evolution would be the tandem duplication of an approximately 4-kb-long DNA unit containing the ancestral adult α-globin gene and the *Alu* family repeat. This process could have been accomplished by an unequal crossing-over in a fashion similar to that proposed for the human $^{G}\gamma$-$^{A}\gamma$-globin gene pair by Shen *et al.* (24). The age of this tandem duplication event is difficult to estimate. As previously noted, *Alu* family members in primates consist of a 300-nucleotide-long inexact dimer structure, whereas rodent *Alu* family members are 150-nucleotide-long monomeric structures (9). The insertion of a 300-nucleotide-long *Alu* family member into the ancestral adult α-globin gene unit prior to its tandem duplication suggests that the duplication occurred in an ancestral primate after the divergence of primates and rodents.

(iv) Concerted evolution of the duplication units: The conservation of a high degree of homology between duplicated functional α-globin genes within each primate species has been suggested by Lauer *et al.* (6) and Zimmer *et al.* (7) to be the result of concerted evolution (7, 25). Concerted evolution could be imposed by either natural selection or a mechanism(s) of gene correction. However, there does not seem to be a strong selection for precise homology between tandemly duplicated α-globin genes. For example, the adult chicken α^{D}-α^{A}-globin gene pair (26) are highly diverged.

Two mechanisms have been proposed for gene correction: gene conversion and homologous but unequal crossing-over (25, 27, 28). Both processes seem to be operating on the human globin gene families. Although the homologous but unequal crossing-over phenomenon is fairly common for the adult α-globin genes in human populations (7, 29, 30), Slightom *et al.* (31) and Michelson and Orkin (3) have presented evidence for gene conversion of the $^{G}\gamma$-$^{A}\gamma$-globin gene pair and in the adult human α-globin gene region, respectively.

(v) Interruption of gene correction: As shown in Table 1, many point mutations have accumulated in the two duplication units. It is conceivable that at early stages after the tandem duplication event, DNA sequence differences arising from point mutations were "erased" by the processes of gene correction. However, assuming the process of either gene conversion or homologous but unequal crossing-over involves pairing of homologous DNA and recombination (25), one could imagine that mutagenic events that result in nonhomologous regions might inhibit or block gene correction. Introduction of a relatively long nonhomologous DNA segment into one of the duplication units (or gene correction units) would decrease the frequency or rate of homologous DNA pairing as well as inhibit branch migration, which is required for DNA recombination (6). The dimeric *Alu* sequence inserted in the α2-globin gene-containing unit and the 224-bp sequence (nonhomology block III) in the α1-globin gene-containing unit may play such a role. Michelson and Orkin (3) have independently proposed that the 7-bp extra sequence within the second intron of α1-globin gene blocked gene conversion during evolution and caused the high divergence of the 3′ untranslated regions. One then expects more sequence divergence to be found between two counterparts of the duplication units in regions adjacent to nonhomology blocks or the ends of the duplication units. This is indeed the case for most areas of the adult α-globin gene regions (Table 1).

Interestingly, Proudfoot and Maniatis (17) have found that the 5′ ends of the two duplication units have almost identical sequences, with only one nucleotide difference out of 124 bp determined. One possible reason for this observed difference in sequence divergence of the two ends is that the 5′-end se-

quences serve as hot spots for initiation of gene correction, which proceeds downstream towards the nonhomology I and II. The existence of polarity in gene conversion has been observed in distinct genes in several organisms including fungus and yeast (for references, see refs. 32 and 33).

Gene correction may be affected by nonhomologous sequences to different extents for different genes in different species. Although gene conversion of certain genes of yeast (34–37) and the V regions of immunoglobulin genes (38) could proceed within areas flanked by totally nonhomologous sequences, it has been demonstrated that the frequency of recombination or gene conversion is decreased by regions of nonhomology in yeast (for references, see ref. 33) and in bacteriophage λ (39).

It is well established that several families of interspersed repeats, including the *Alu* family, are dispersed throughout the genome by insertion into preexisting unoccupied DNA sequences (19, 20). If our interpretation of the sequences flanking the human adult α-globin genes is correct, then one consequence of the insertion of repetitive sequences is inhibition of the gene conversion process and induction of sequence diversity between duplicate genes and their flanking regions.

We thank Drs. Tom Maniatis, Argis Efstratiadis, Michael Turelli, and R. Holliday for many helpful suggestions. We also thank Alan Michelson and Stuart Orkin for stimulating discussions and for communicating their results to us before publication. We appreciate Kathryn Graehl's technical assistance and Candy Miller's effort in typing the manuscript. This research is supported by National Institutes of Health Grants AM 29800 (C.-K.J.S.) and GM21346 (C.W.S.).

1. Weatherall, D. J. & Clegg, J. B. (1979) *Cell* **16,** 467–479.
2. Maniatis, T., Fritsch, E. F., Lauer, J. & Lawn, R. M. (1980) *Annu. Rev. Genet.* **14,** 145–178.
3. Michelson, A. M. & Orkin, S. H. (1983) *J. Biol. Chem.*, in press.
4. Proudfoot, N. J., Gil, A. & Maniatis, T. (1982) *Cell* **31,** 553–563.
5. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21,** 653–668.
6. Lauer, J., Shen, C.-K. J. & Maniatis, T. (1980) *Cell* **20,** 119–130.
7. Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W. & Wilson, A. C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2158–2162.
8. Shen, C.-K. J. & Maniatis, T. (1982) *J. Mol. Appl. Genet.* **1,** 343–360.
9. Schmid, C. W. & Jelinek, W. R. (1982) *Science* **216,** 1065–1070.
10. Orkin, S. H., Old, J., Lazarus, H., Altay, C., Gurgey, A., Weatherall, D. J. & Nathan, D. G. (1979) *Cell* **17,** 33–42.
11. Embury, S., Lebo, R., Dozy, A. & Kan, Y. W. (1979) *J. Clin. Invest.* **63,** 1307–1310.
12. Calos, M. P. & Miller, J. H. (1980) *Cell* **20,** 579–595.
13. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
14. Messing, J. & Gronenborn, B. (1978) in *Single-Stranded DNA Phages,* eds. Denhardt, D. T., Dressler, D. H. & Ray, D. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 449–453.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
16. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–321.
17. Proudfoot, N. J. & Maniatis, T. (1980) *Cell* **21,** 537–544.
18. Nordheim, A. & Rich, A. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1821–1825.
19. Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T. & Gesteland, R. F. (1981) *Cell* **26,** 11–17.
20. Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. (1981) *Cell* **26,** 141–142.
21. Grimaldi, G. & Singer, M. F. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1497–1500.
22. Lemischka, I. & Sharp, P. (1982) *Nature (London)* **300,** 330–335.
23. Schon, E. A., Wernke, S. M. & Lingrel, J. B. (1982) *J. Biol. Chem.* **257,** 6825–6835.
24. Shen, S. H., Slightom, J. L. & Smithies, O. (1981) *Cell* **26,** 191–203.
25. Hood, L., Campbell, J. H. & Elgin, S. C. R. (1975) *Annu. Rev. Genet.* **9,** 305–353.
26. Dodgson, J. B., McCune, K. C., Rusling, D. J., Krust, A. & Engel, J. D. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5998–6002.
27. Smith, G. P. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38,** 507–514.
28. Tartoff, K. D. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38,** 491–500.
29. Goosens, M., Dozy, A., Embury, S. H., Zacharides, Z., Hadjiminas, M. G., Stamatoyannopoulos, G. & Kan, Y. W. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 518–521.
30. Higgs, D. R., Old, J. M., Pressley, L., Clegg, J. B. & Weatherall, D. J. (1980) *Nature (London)* **284,** 632–635.
31. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21,** 627–638.
32. Rossignol, J.-L., Paquette, N. & Nicolas, A. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 1343–1352.
33. Fogel, S., Mortimer, R., Lusnak, K. & Tavares, F. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 1325–1341.
34. Scherer, S. & Davis, R. W. (1980) *Science* **209,** 1380–1384.
35. Haber, J. E., Rogers, D. T. & McCuster, J. M. (1980) *Cell* **22,** 277–289.
36. Klar, A. J. S., McIndoo, J., Strathern, J. N. & Hicks, J. B. (1980) *Cell* **22,** 291–298.
37. Jackson, J. A. & Fink, G. R. (1981) *Nature (London)* **292,** 306–311.
38. Baltimore, D. (1981) *Cell* **24,** 592–594.
39. Sodergren, E. J. & Fox, M. S. (1979) *J. Mol. Biol.* **130,** 357–377.