Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 1392–1396, March 1983
Genetics

Unusual structure of the chicken embryonic α-globin gene, π'

(DNA sequence analysis/Southern blotting technique/large introns)

JAMES DOUGLAS ENGEL*, DREW J. RUSLING*, KEVIN C. MCCUNE*, AND JERRY B. DODGSON†

*Department of Biochemistry, Molecular and Cell Biology, Northwestern University, Evanston, Illinois 60201; and †Departments of Microbiology, Public Health, and Biochemistry, Michigan State University, East Lansing, Michigan 48824

Communicated by Irving Klotz, November 15, 1982

ABSTRACT We report the DNA sequence of the globin locus encoding the chicken embryonic α-globin, π'. The structure differs significantly from that of the two chicken adult α-globin genes, α^A and α^D, as well as from that of previously studied adult α-globin genes in that the introns of the π' gene are substantially larger than those in adult α-globin loci. In contrast, the π' introns are structurally similar to the only other expressed embryonic α-globin gene reported to date, the human ζ gene. While completing the sequence of the π' gene, we determined that only one chromosomal locus within the chicken genome hybridizes to a π' central exon probe. These data lead to the conclusion that if the equimolar chicken embryonic α-globin polypeptides, called π and π', are indeed independently transcribed, then that transcription occurs from alleles of the same gene; however, we favor the possibility that the π gene does not actually exist. This conclusion is drawn from the observation that the two chromosomal alleles of embryonic α-globins (represented by recombinant bacteriophage λCαG5 and λCαG7) both encode π'.

The α- and β-globin genes of the chicken make up a small multigene family that is subject to temporal regulation of gene expression during development (switching). This regulation can be defined as occurring both within a single cell type (e.g., transient β^H expression in the definitive erythroid cell) and in differentially related cell types [e.g., ρ versus β expression during the switch from embryonic globin synthesis in primitive erythroid cells to adult globin synthesis in definitive erythroid cells (1)]. To study the expression of those genes that are regulated during avian embryogenesis, we have isolated and characterized both α- and β-globin linkage groups on recombinant bacteriophage (2–5).

In this paper, we report the complete DNA sequence for the member of the globin gene family of the chicken that had not heretofore conclusively been identified in chromosomal DNA, that encoding the embryonic chicken α-globin. The gene sequence here reported encodes π', one of two major α-globins expressed exclusively in the primitive erythroid line. These two globins are reported to differ one from another by a single amino acid at position 124 of the polypeptide chain, being either alanine (π) or glutamate (π') (6).

We were unable to demonstrate linkage of the other major embryonic α-globin gene, π, to within 12 kilobase pairs (kbp) of the α-globin gene cluster (4), and we present evidence from the analysis of two α-globin genomic recombinants that only a single chromosomal locus codes for the embryonic α-globin polypeptide chains, thereby indicating that π and π' are produced by the same gene. Southern blotting data support this conclusion, in that only a single chromosomal restriction endonuclease fragment is visualized in genomic maps when they are probed with a central exon subclone of π'.

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.

MATERIALS AND METHODS

Isolation and Characterization of Embryonic α-Globin Chromosomal Recombinant Bacteriophages. Isolation of recombinants λCαG5, λCαG6, and λCαG7 from the original amplified λ Charon 4A/chicken chromosomal DNA linker library (2) has been described (4, 5). The position of the π or π' gene within these recombinants was determined by positive blot hybridization with a radioactive cDNA probe prepared from globin mRNA isolated from embryos 4.5 days after fertilization, and lack of hybridization to the same restriction fragments with an adult globin-specific cDNA probe (ref. 4; data not shown). The chromosomal α-globin domain as defined by these hybridizations is shown by Fig. 1*A*, and, in particular, the locus reported herein is shown in Fig. 1*B*. Subclones prepared from this segment of the chromosome and containing the embryonic α-globin locus are shown in Fig. 1*C*, and sequence analysis strategy for the locus is shown in Fig. 1*D*. pBR322 subclones were prepared as described (3, 4); we have also previously described conditions used for chromosomal Southern blotting (3, 7), end-labeled fragment isolation (3, 8), and DNA sequence analysis and gel electrophoresis conditions (9, 10). The only sequencing modifications used in this study were to dry the 0.4-mm gels onto Whatman 3MM paper before autoradiography and then expose the dried gel without intensifying screens. Although this necessitated increasing the exposure time 3- to 4-fold, the increase in band resolution of sequencing ladders was substantially enhanced. Recombinant DNAs were propagated according to current National Institutes of Health guidelines for recombinant DNA research.

RESULTS

DNA Sequence of the π'-Globin Gene. In our original report of the chicken α-globin linkage group (4), we alluded to the close proximity of the embryonic π or π' gene to the adult genes. Some uncertainty existed regarding the precise location of the embryonic gene, because the adult genes (α^A and α^D) are also expressed at low levels in the primitive (embryonic) erythroid lineage (1), and these adults genes are also quite closely linked, thus giving rise to a complex mapping blot hybridization pattern of embryonic globin cDNA to chromosomal recombinants λCαG5 and λCαG7. A variety of overlapping subclones in the embryonic α-globin region were prepared and fine-structure restriction maps of these subclones were derived by using the technique of Smith and Birnstiel (11). After these results had been obtained, the sequence of the locus was entirely determined; the data from these experiments are shown in Figs. 1 and 2. The DNA sequences of λCαG5 and λCαG7 in the region of the third exon (described by subclones pα5BH20 and pα7BR2; Fig. 1*C*) are identical, and they predict the same glutamate at position 124 of the polypeptide chain (Fig. 2), thereby identifying both

Abbreviation: bp, base pair(s).

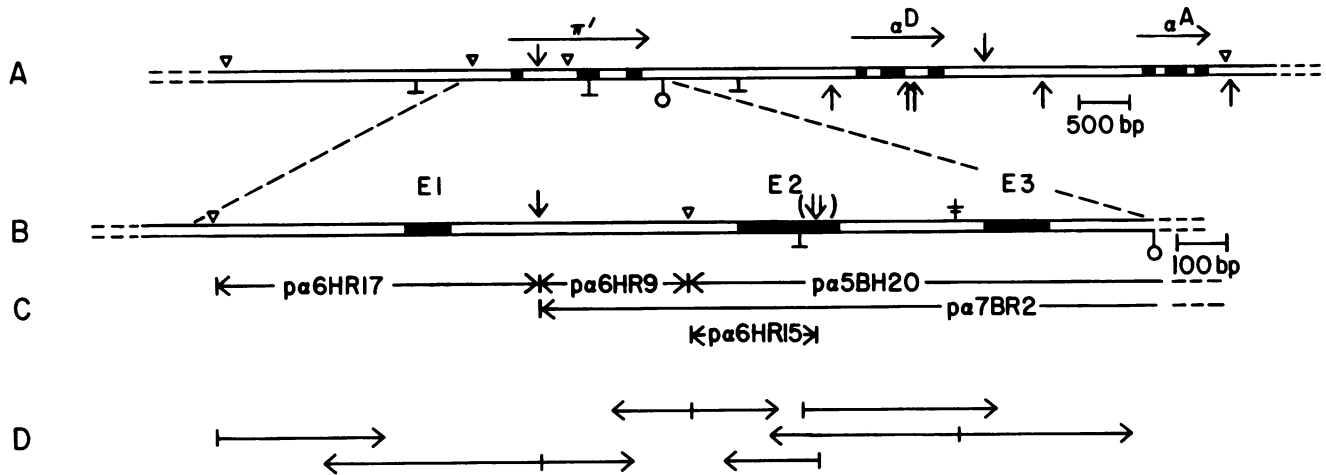

FIG. 1. Chromosomal location of and sequencing strategy for the embryonic chicken α-globin gene. (*A*) Position of the π′-globin gene relative to the adult α-globins, α^D and α^A (4, 5). The transcriptional direction (5′ to 3′) and approximate primary transcript size are shown above each gene, depicted in the chromosome as a series of black boxes (three exons for each of the three genes). Also depicted is an abbreviated restriction map of the α-globin chromosomal domain. (*B*) An expanded section of the chromosome in the region of the π′-globin gene. Restriction enzymes used in analysis of the locus are ↓, *Eco*RI; ↑, *Bam*HI; ▽, *Hin*dIII; ⊥, *Taq* I; ‡, *Hin*fI; ♁, *Sac* I (*Sst* I). As in *A*, black boxes correspond to the position of the three π′ exons. The arrow in parentheses above exon 2 (E2) shows the position of the *Eco*RI linker of λCαG6 (4). (*C*) Subclones used in the sequence analysis of the gene were generated from parental λ recombinants λCαG5, λCαG6, and λCαG7 (4). For example, pα6HR15 denotes: pBR322 subclone, λC<u>α</u>G<u>6</u>, <u>H</u>*ind*III/*Eco*<u>R</u>I, isolate <u>15</u>. (*D*) Sequencing strategy for the locus. The vertical line in each case represents the position of the 5′ end label, and the arrowhead represents the direction and maximal extent of DNA sequence determined from a single labeled restriction site end (shown in *B*). The sequence of each fragment was determined multiple times.

recombinants as encoding the π′-globin (6). Although the λCαG5 and λCαG7 subclones containing the third exon are identical in their respective coding segments, a single base difference was detected in the 3′ untranslated segment of the respective subclones (Fig. 2; data not shown). Thus, λCαG5 and λCαG7 are assumed to be representative of the two different chromosomal alleles of this gene within the linker library.

The π′ DNA sequence (Fig. 2) has both expected and unexpected structural features typical for α-globin genes. Introns occur within codon 31 and between codons 99 and 100, assuming that the normal splice-junction rule (5′ G-T . . . A-G 3′ for nucleotides that directly abutt the intron/exon boundaries) applies to this gene (15); thus intron position in all α-globins is precisely conserved (16–18). Consensus sequences implicated *in vivo* and *in vitro* for accurate and abundant transcription exist 5′ to the gene coding sequence, and these C-C-A-A-T (nucleotides 253 to 257) and A-T-A (nucleotides 300 to 302) conserved DNA sequences are in a physical location with respect to the gene consistent with their presumptive enhancer and promoter specificity function (12, 13, 19). Similarly, 125 nucleotides 3′ to the termination codon we found the A-A-T-A-A-A recognition sequence (nucleotides 1,806–1,811), the poly(A) processing signal (14, 20). Thus, this gene contains all landmark consensus sequences we have come to expect in many RNA polymerase II-transcribed genes.

Introns of the π′-Globin Gene. The major unexpected finding in the DNA sequence shown in Fig. 2 was the size of the two introns. In those α-globin genes that have been reported to date, the intron sizes are relatively highly conserved (Fig. 3). The 5′ intron usually shows most extreme size conservation, varying between 113 nucleotides (17) and 147 nucleotides (5). The 577-nucleotide 5′ intron in the π′ gene reported here is thus 4–5 times larger than that in any other expressed adult α-globin gene. The 3′ intron is also unusually large (294 nucleotides) when compared to adult α-globin loci (5, 18) with the exception of the minor adult chicken α-globin gene, α^D (281 nucleotides).

Recently, the sequences of the human embryonic α-globin gene, termed ζ, and a closely linked pseudogene, ψζ, have also been reported (21). Both the gene and the related pseudogene bear striking structural similarity to the π′ gene reported here. Both ζ and ψζ contain simple, tandemly repeated sequences in their introns, whereas the π′ gene does not exhibit these structural features. If one deletes these simple sequences from the ζ 5′ intron and then compares the sizes of the analogous introns of π′ and ζ, one finds the introns to be quite similar in size (617 bp vs. 577 bp for the 5′ introns and 239 vs. 294 bp for the 3′ introns in ζ and π′, respectively). Yet, despite this structural similarity, and as might reasonably be expected from the length of time since these genes originally diverged, π′ and ζ share only limited sequence homology. The evolutionary implications will be considered in *Discussion*.

Allelism at the π/π′ Globin Locus? One question arising during these and prior investigations (1) that neither we nor others had been able to resolve was whether π and π′ polypeptides were encoded by two separate genes, or if, in fact, they were alleles of the same gene. It has been reported that segregation of alleles had not been observed in analysis of the hemoglobins present in individual embryos, nor was allelism evident in erythroid multicell colony bursts in culture (6, 22). Furthermore, the sequence analysis of the π and π′ polypeptides demonstrated a single amino acid difference at position 124 of the polypeptides, also indicating a single change in charge between the two (6). Thus we were initially perplexed by our inability to isolate a recombinant containing another embryonic α-globin gene either linked to or physically distinct from the rest of the α-globin gene cluster.

During the early part of this work, the π/π′ gene was mapped to contiguous DNA fragments summing to greater than 1,500 nucleotides; therefore, it was assumed that this hybridizing locus contained both the π and π′ genes and that these genes were very closely linked in the genome. Later, the DNA sequence of this locus (Fig. 2) showed that this earlier assumption was in error and that the locus instead encoded a single, abnormally large, embryonic α-globin gene, π′.

Attempts were then made to determine if one or two separate embryonic α-globin genes existed in the genome by studies on chromosomal DNA rather than of isolated recombinants. Be-

```
 agcttcttagtttgctcaaaatccattgaaaggcacgtcctaatcctccatcactagaatgttccagcactatctctgctcatacgtgc
Hind III 100
ctggtcaacttttcaggccacacggccaatacgtgttcagaagcaagaagctgataacaccagcaaaccaggaggccagtgcactgaccc
             200
ttagcgaaatgggctttgtttcacaaggagataagggtctcccacctccatgggtgggcctccggagtgaCCAATgagtgtggacagatgc
                   300
caaggcccgctctcctcttcctctttATAaccggggctgcgagggcactcagtacaacctgctctgggtgttcactgaagggagcctgagc
                             400
cagcactctcctgcacaatgGCA CTG ACC CAA GCT GAG AAG GCT GCC GTG ACC ACC ATC TGG GCA AAG GTG GCT
                    Ala Leu Thr Gln Ala Glu Lys Ala Ala Val Thr Thr Ile Trp Ala Lys Val Ala
                                                                         500
ACC CAG ATT GAG TCC ATT GGG CTG GAA TCA CTG GAG AG gtaagtcacccacagcacccccccaaagggtgcccccc
Thr Gln Ile Glu Ser Ile Gly Leu Glu Ser Leu Glu Ar(g)
                                                                                   600
tgactttgctgttaggatgcatcttgtttcagtgctgtatgagtgagacccatacagtcgtgttaggactgatgagaactgcttgatgag
                                             Eco RI
ctgcatgtttttaacatgatttttcttactggagttacacccgtgctatgaattcaaaggcattcatactcggtgtcccgaaataggg
 700
ttatcgctaccaggaaatgaactcaaatagatttatcacataagtcgccatgtaatggacattaagagaatagctgtccacatttgttgt
          800
tgttcagagctatttaaatgctatttacaatcaatctaacatatcgtgatatccctgctgggactccgagtaccactcataccaatcctg
                  900                                                       Hind III
gttcattcagctcacagcagtttgaagaccttgcaaaattggcttacggtataatataggaaacgtggtgtcagtttcaagcttccaaag
                           1000
ctctattaattgctgaaggagtgtatttcaagtgcggatcatgcgtgccttttccactgctgaactttgttgtcttgttctcctccag G
                                                         1100
CTT TTT GCC AGC TAT CCT CAG ACG AAA ACC TAC TTC CCT CAC TTT GAT GTC AGC CAA GGC TCA GTT CAG
Leu Phe Ala Ser Tyr Pro Gln Thr Lys Thr Tyr Phe Pro His Phe Asp Val Ser Gln Gly Ser Val Gln
                                                                        Taq I
CTT CGT GGT CAC GGC TCC AAG GTC CTG AAT GCC ATT GGG GAA GCT GTG AAG AAC ATC GAT GAC ATT AGA
Leu Arg Gly His Gly Ser Lys Val Leu Asn Ala Ile Gly Glu Ala Val Lys Asn Ile Asp Asp Ile Arg
    1200
GGT GCT TTG GCC AAA CTC AGC GAG CTG CAT GCT TAC ATC CTC AGG GTG GAC CCA GTG AAC TTC AAG gtg
Gly Ala Leu Ala Lys Leu Ser Glu Leu His Ala Tyr Ile Leu Arg Val Asp Pro Val Asn Phe Lys
                                   1300
agtgggcacgctttcagggatgaaaactaccagtcacagaactagaggccacaggtcatttagactaatgggagcttcatccctgactgtg
                                            1400
tggtacagcctcatgttgctcctggttccccagctgattttaacctgagactggatcttctagcaaaccatctaatccagtttaagctgt
                                                    1500
tacctggctgggacatgtatttttttcattccagagcatgttgcagactgttagattggcatctcttactgattacaatggatttcatct
                                            Hinf I                               1600
ttgtttttccctttcag  CTG CTT TCC CAC TGT ATC CTG TGC TCT GTG GCT GCC CGC TAT CCC AGT GAT TTC
                   Leu Leu Ser His Cys Ile Leu Cys Ser Val Ala Ala Arg Tyr Pro Ser Asp Phe

ACC CCA GAA GTT CAT GCT GCG TGG GAC AAG TTC CTG TCC AGC ATT TCC TCT GTT CTG ACT GAG AAA TAC
Thr Pro Glu Val His Ala Ala Trp Asp Lys Phe Leu Ser Ser Ile Ser Ser Val Leu Thr Glu Lys Tyr
                        1700
AGA taa atggcttccacactgggttagggacgtgcatccacggcacacaacagctgccaagttctggggtattcttcctatgcagtccc
Arg ...
                                  1800
cccaactcccctgcgcaggggctcggccacctgcagaccacAATAAAtaattcgactgtgatctatggttc
```

FIG. 2. DNA sequence of the π'-globin locus. The DNA sequence of the π' gene determined as depicted in Fig. 1 is shown. Because no independent transcription initiation or polyadenylylation sites were determined in this study, numbering in the locus arbitrarily begins at the first (leftmost) *Hin*dIII site of Fig. 1*B*. By analogy to other mRNA cap sites (5), we presume that transcription initiation begins at nucleotide 327. Nucleotides are numbered in multiples of hundreds, the last digit placed over the appropriate nucleotide; in between hundreds, a dot is placed over the 50th nucleotide in the sequence. Consensus sequences common to other RNA polymerase II-transcribed genes, and flanking the transcribed sequence, are capitalized and underlined (5, 12, 13, 14; see text). Codons for the π'-globin are capitalized directly over the amino acids encoded in the π' polypeptide (6). Restriction enzyme recognition sites are indicated for ease of orientation relative to the map of Fig. 1*B*. The single nucleotide difference detected in pα5BH20 and pα7BR2 (see text) is an C versus T, respectively, at position 1,716.

cause the DNA sequence of the π' structural gene had already been determined (Fig. 2), it was reasoned that by using in chromosomal Southern blots a combination of restriction endonucleases that cleaved both within and outside of the π' globin locus and by using the π' central exon as probe, one should be able to deduce whether one or two different loci complementary to the π' central exon probe existed in genomic DNA. Because the π and π' polypeptides are identical within this region (amino acids 32–99) (6), reason further dictated that a π' central exon probe should detect both π and π' sequences in chromosomal DNA. If the coding sequences for these polypeptides were found at two distinct chromosomal locations by these criteria (i.e., if two bands were detected on genomic blots) they must be encoded by separate genes; if only one band was de-

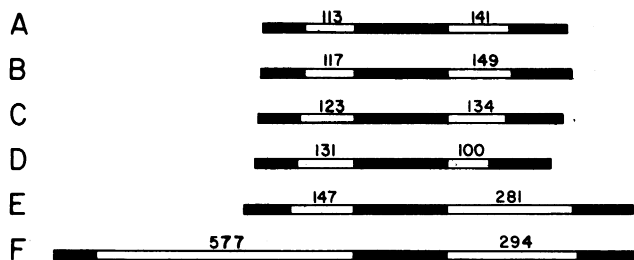

FIG. 3. Intron sizes in vertebrate α-globin genes. The following intron sizes are shown: (*A*) human adult α_2-globin (17), (*B*) human adult α_1-globin (18), (*C*) mouse adult α-globin (16), (*D*) chicken adult α^A-globin, (*E*) chicken adult α^D-globin (5), and (*F*) chicken embryonic π'-globin. Numbers above each blank correspond to the size of the 5′ (leftmost) and 3′ (rightmost) introns in each gene. In each case, the three exons (black boxes) are identical in length.

tectable, the most likely possibility was that the two proteins were encoded by alleles of the same gene.

The results of this chromosomal blotting experiment are shown in Fig. 4. Clearly, only one locus complementary to the π' central exon probe is detected on these filters, whether the DNA is digested with restriction enzymes that give large (*Eco*RI, *Bam*HI, *Hin*dIII) or small (*Taq* I) genomic DNA digestion fragments. Each of these restriction enzymes produces bands precisely predicted by the map of the π'-globin locus (Fig. 1*B*), and it was therefore tentatively concluded that embryonic chicken α-globins must be synthesized from alleles of the same gene at a common chromosomal locus or that in fact only one gene (that encoding π'-globin) exists in the chicken genome (see *Discussion*).

DISCUSSION

The DNA sequence of the 5′-most locus (relative to transcription) in the chicken α-globin gene cluster corresponds to that of

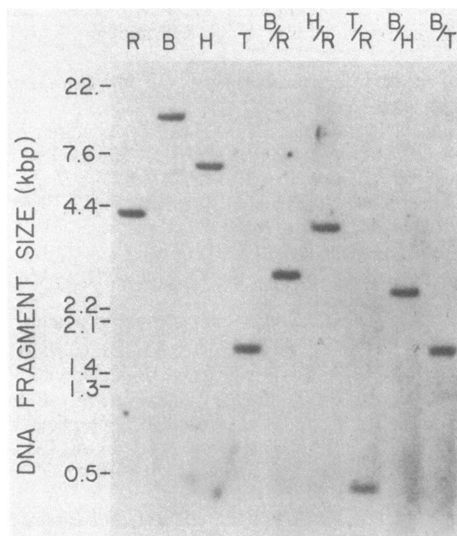

FIG. 4. Chromosomal Southern blot hybridization to the π' central exon. Three micrograms of chromosomal chicken erythrocyte DNA was digested with *Eco*RI (R), *Bam*HI (B), *Hin*dIII (H), or *Taq* I (T), as well as combinations of these, electrophoresed on a 0.8% agarose gel, and blotted, hybridized, and washed as described (3). The probe used for hybridization was nick-translated pα6HR15 (Fig. 1*C*), which contains 150 nucleotides of π' central exon coding sequence and 99 nucleotides of 5′ intron sequence. Fragment sizes were determined by hybridization to restriction enzyme-digested, internal λ and pBR322 markers run in a separate lane on the same gel (not shown).

the embryonic α-globin gene, π' (Fig. 1). The gene is typical of most other eukaryotic RNA polymerase II-transcribed genes in most respects: (*i*) consensus C-A-A-T (13) and A-T-A (12) boxes flank the 5′ end of the gene, whereas a consensus polyadenylylation sequence A-A-T-A-A-A (14) is found in the sense strand 125 nucleotides 3′ to the termination codon; (*ii*) splice junction sequences 5′ G-T . . . A-G 3′ border the exons and interrupt the coding sequence in positions typical of other α-globins (5, 16–18).

The results reported here do not conclusively resolve the question of whether embryonic α-globins π and π' are separate genes or are alleles of the same gene. As reported previously (4) and in these experiments, only a single locus hybridizing to embryonic α-globin cDNA has been found, and that locus (from the sequence of the 3′ exons of two separate recombinant subclones, one from each chromosome; Fig. 1*C*) corresponds to the coding sequence for the embryonic gene π' (Fig. 2). Chromosomal Southern blot analysis would similarly indicate the presence of only a single locus encoding this polypeptide (Fig. 4). Contrary to these data, analysis of the globin polypeptide chains derived from erythroid bursts in chicken embryo mesenchymal organ culture demonstrates the presence of both globin chains in every burst examined (with sometimes as few as three erythroid precursors per burst) (22), thereby implying the expression of two separate and nonallelic embryonic α-globin genes. We offer below two separate explanations to resolve these apparently disparate results.

The first possibility is that all, or a large portion, of the α-globin gene cluster has been duplicated in the chicken genome, and that, after this duplication, globin genes π and π' diverged by one amino acid, whereas genes α^D and α^A did not diverge. This explanation would resolve the apparent conflict of these results with those of π and π' polypeptide sequence analysis (6), but it also places other constraints on the proferred explanation. First, Southern blotting of chromosomal DNA from a wide variety of chickens or chicken cell lines showed no evidence that the adult chicken α-globin genes were present in 2-fold excess to the adult β-globin gene (7). Because the α-globin and β-globin loci are on different chromosomes (23), it seems very unlikely that both loci have been independently duplicated. Furthermore, if the α-globin locus has been duplicated, it has apparently been duplicated on the same (or a nearly identical) chromosome because α-globin cDNA appears to hybridize to a single chromosome when chicken macrochromosomes are separated on sucrose gradients (23). Second, analysis of a wide variety of chicken α-globin gene recombinants (4, 5, 24, 25) has failed to reveal significant nucleotide divergence among the clones such as might be expected if we had cloned two separate gene clusters. This suggests that, if a duplication in the α-globin locus has occurred, it has occurred (or has been corrected by gene conversion) very recently in evolution. Note that the duplicated adult α-globin genes of humans are presumed to be the result of a relatively recent gene conversion event, and even so, the genes still have considerable sequence divergence at their 3′ ends (17, 18). An explanation analogous to duplication of the α-globin gene region might involve duplication solely of the ancestral embryonic α-globin gene followed by divergence to π and π'. In this case, the duplicated π gene must then lie outside the approximately 30 kbp around the chicken α-globin gene locus that we have mapped in detail. This further requires that the genomic DNA fragments produced at the π and π' loci upon restriction digestion with those enzymes chosen for this analysis (Fig. 4; *Eco*RI, *Bam*HI, *Hin*dIII, and *Taq* I) happen to be the same size. Although the above considerations do not rule out the possibility of duplication of the α-globin locus, they nonetheless weigh heavily against it.

The second explanation to resolve this apparent ambiguity would be to assume that there indeed exists only a single embryonic α-globin gene, π', which encodes glutamic acid at position 124 of the polypeptide chain (6), and that the identification of a polypeptide π, which encodes an alanine residue at that same position, is the result of an error in protein sequence analysis. This alternative supposes that a single structural gene exists in both chromosomes encoding π', and, therefore, this single translation product is the only embryonic α-globin chain in chickens. Why then, do the embryonic α-globins resolve into equimolar species on acid/urea gels if not for this position 124 (alanine vs. glutamate) charge difference? A possible explanation resides in the work which originally reported the sequence of the two polypeptides. Globin π was found to be NH_2-terminally blocked, whereas globin π' was found to be ". . . at least partially unblocked" (6). Assuming that "π" is completely blocked and "π'" completely unblocked at the NH_2 terminus (i.e., equimolar quantities of blocked and unblocked π', in reality), this would account for a single charge difference resolving the two equimolar species of the same polypeptide on acid/urea/polyacrylamide gels. One major prediction for experimental results emerges from this explanation: we would expect newly synthesized chicken embryonic α-globin polypeptides to migrate identically on acid/urea gels in the absence of NH_2-terminal blocking. Second, although not explicitly required by the above explanation, we might expect embryonic hemoglobin P to contain one blocked and one unblocked α-globin polypeptide per tetrameric hemoglobin P molecule. Thus, final resolution of this question must await further experiments.

Finally, the fact that the sizes of the introns in this embryonic gene vary greatly from those in other expressed α-globins deserves some note; in particular these introns bear striking structural similarity to the human embryonic α-globin gene, ζ (21). In computer-assisted sequence comparisons between the 5' introns of π' and ζ, four segments of significant nucleotide sequence homology (i.e., greater than 90% identity) were found. One of these homologous regions (nucleotides 940–949, Fig. 2) has characteristics of possible structural significance, a match between a 10-nucleotide segment in the central portion of the π' intron and the splice donor junction of the ζ gene intron. One possibility for a functional role of this internal splice junction donor sequence could be as a normal donor in a multistep nuclear RNA processing mechanism whereby the intron is excised in segments, rather than in a single splicing event (26). Pertinent to this possibility is the observation that several potential splice-acceptor sequences exist between this homology unit and the canonical splice-acceptor sequence bordering the second exon (Fig. 2).

Although these two genes are quite similar to one another in size, we have detected only limited DNA sequence homology in comparisons of the respective loci between the introns of ζ and π'. Thus the relative conservation of physical size in the chicken and human embryonic α-globin genes is not reflected in a similar conservation of intron DNA sequence. It has recently been postulated that the human and chicken embryonic α-globins diverged from their common ancestral gene long before avian/mammalian divergence (21, 27). We find it intriguing then, that although no sequences for control of transcription have been found internal to RNA polymerase II-transcribed genes in surrogate genetic experiments (transformation, transfection, and microinjection), the embryonic α-globins of chickens and humans have maintained, with substantial integrity, the size of the introns of these related genes for about 400 million years. The high level of evolutionary conservation of intron size among members of developmentally similar gene types in widely divergent species further suggests that the size and structure of these introns may be of fundamental significance in control of globin gene expression (28).

We are grateful for the skilled technical assistance of Ed Diekhoff and Sheryl Niemiec, as well as for access to the Stanford MOLGEN Project and the National Institutes of Health SUMEX-AIM computer analysis system. We thank Tom Maniatis for a preprint of the structural analysis of the human ζ-globin gene and Steve Hughes for the suggestion to dry our sequencing gels. We also acknowledge support by the National Institutes of Health (Grant HL24415 to J.D.E. and Grant GM28837 to J.B.D.) and the Leukemia Research Foundation. J.D.E. gratefully acknowledges the support of an American Cancer Society Faculty Research Award.

1. Brown, J. L. & Ingram, V. M. (1974) *J. Biol. Chem.* **249**, 3960–3972.
2. Dodgson, J. B., Strommer, J. S. & Engel, J. D. (1979) *Cell* **17**, 879–887.
3. Dolan, M., Sugarman, B. J., Dodgson, J. B. & Engel, J. D. (1981) *Cell* **24**, 669–677.
4. Dodgson, J. B., McCune, K. C., Rusling, D. J., Krust, A. & Engel, J. D. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5998–6002.
5. Dodgson, J. B. & Engel, J. D. (1983) *J. Biol. Chem.* **258**, in press.
6. Chapman, B. S., Tobin, A. J. & Hood, L. E. (1980) *J. Biol. Chem.* **255**, 9051–9059.
7. Engel, J. D. & Dodgson, J. B. (1978) *J. Biol. Chem.* **253**, 8239–8246.
8. Girvitz, S. C., Bacchetti, S., Rainbow, A. J. & Graham, F. L. (1980) *Anal. Biochem.* **106**, 492–496.
9. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
10. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255–2274.
11. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3**, 2387–2399.
12. Goldberg, M. (1979) Dissertation (Stanford Univ., Stanford, CA).
13. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeReil, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Bechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 653–668.
14. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
15. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
16. Nishioka, Y. & Leder, P. (1979) *Cell* **18**, 875–882.
17. Liebhaber, S. A., Goossens, M. J. & Kan, Y. W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7054–7058.
18. Michelson, A. M. & Orkin, S. H. (1980) *Cell* **22**, 371–377.
19. McKnight, S. L., Gavis, E. R., Kingsbury, R. & Axel, R. (1981) *Cell* **25**, 385–398
20. Fitzgerald, M. & Shenk, T. (1981) *Cell* **24**, 251–260.
21. Proudfoot, N. J., Gil, A. & Maniatis, T. (1982) *Cell* **31**, 553–563.
22. Pine, K. S. & Tobin, A. J. (1976) *Dev. Biol.* **49**, 556–562.
23. Hughes, S. H., Stubblefield, E., Payvar, F., Engel, J. D., Dodgson, J. B., Spector, D., Cordell, B., Schmike, R. T. & Varmus, H. E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1348–1352.
24. Deacon, N. J., Shine, J. & Naora, H. (1980) *Nucleic Acids Res.* **8**, 1187–1199.
25. Richards, R. I. & Wells, J. R. E. (1980) *J. Biol. Chem.* **255**, 9306–9311.
26. Kinneburgh, A. J. & Ross, J. (1979) *Cell* **17**, 915–921.
27. Czeliesniak, J., Goodman, M., Hewett-Emmitt, D., Weiss, M. L., Venta, P. L. & Tashian, R. E. (1982) *Nature (London)* **298**, 297–300.
28. Engel, J. D. & Dodgson, J. B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2596–2660.