Proc. Natl. Acad. Sci. USA
Vol. 78, No. 8, pp. 4694–4698, August 1981
Biochemistry

α-Fetoprotein and albumin genes are in tandem in the mouse genome

(genomic cloning/DNA sequence/unique copy DNA probes/developmental gene regulation)

ROBERT S. INGRAM, RICHARD W. SCOTT, AND SHIRLEY M. TILGHMAN

Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111

Communicated by Robert Palese Perry, March 30, 1981

ABSTRACT **The murine α-fetoprotein (AFP) and serum albumin genes most probably arose in evolution as the consequence of a duplication of a common ancestral gene. They have both been previously mapped to chromosome 5 in the mouse. We now have evidence that these genes are closely linked. By using a unique copy DNA probe derived from previously cloned AFP 5′ flanking DNA, a recombinant DNA phage has been isolated, from a bacteriophage DNA library, that contains sequences flanking the 5′ end of the AFP gene and the 3′ end of the albumin gene. Restriction endonuclease mapping and DNA sequence determination of the recombinant phage and comparison to total genomic DNA confirmed that the genes are in tandem, 13.5 kilobase pairs apart, with the albumin gene to the 5′ side of the AFP gene. Thus, they are transcribed from the same strand of DNA.**

α-Fetoprotein (AFP), which is synthesized by the embryonic liver and yolk sac, is the major serum protein of the developing mammalian fetus (1, 2). In rodents, its rate of synthesis by fetal liver begins to decline just prior to birth, at a time when the rate of synthesis of serum albumin, the major adult serum protein, is increasing (3, 4). This reciprocal expression of AFP and albumin, coupled to a number of functional and structural properties which they share (5, 6), led to the proposal that they were the products of a pair of duplicated genes. Two strong lines of evidence support this hypothesis. First, the complete mouse AFP amino acid sequence, deduced from direct determination of the nucleotide sequences of AFP cDNA and AFP chimeric cDNA plasmids (7), shows an overall 32% conservation of primary sequence with several mammalian albumins. Similar results have been obtained from more limited sequence comparisons of the human (6) and rat (8, 9) proteins. Second, structural analysis of the mouse AFP and albumin genes, by using cloned segments of genomic DNA, revealed that both genes were organized identically into 15 coding segments interrupted by 14 intervening sequences (10, 11). The corresponding coding segments in each gene were the same size, although no nucleotide cross-hybridization was detected between them.

More recently we have found that the two genes are present on chromosome 5 in the mouse (12). By using a series of hamster–mouse somatic cell hybrids in which the mouse chromosomes had differentially segregated, the presence of murine albumin and AFP DNA sequences was detected by hybridization and correlated to the presence of mouse chromosome 5 in the cell lines. In the present report, the close linkage of the AFP and albumin genes is demonstrated by the isolation and characterization of a recombinant genomic DNA clone whose sequences span the distance between them.

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

Identification of Unique Copy DNA Flanking the AFP and Albumin Genes. DNA from λAFP14 (11) was cleaved with several restriction enzymes (New England BioLabs) in both single and double digestions. The fragments were electrophoresed through 1.5% agarose gels and transferred from the agarose gels to nitrocellulose filters (Millipore) by the method of Southern (10, 13). Pretreatment, hybridization, and washing of all filters were performed as described by Wahl *et al.* (14). The DNA on the filters was hybridized to total mouse genomic DNA which had been labeled by nick-translation (15) to a specific activity of $5–10 \times 10^7$ cpm/μg. Those fragments that failed to hybridize to the labeled DNA were presumed to contain only unique copy DNA. To confirm this, the fragments of interest were isolated, labeled by nick-translation, and hybridized to an *Eco*RI digest of mouse genomic DNA on a nitrocellulose filter. The hybridization of a predicted pattern of *Eco*RI fragment(s), in the absence of any background, was taken as evidence that the isolated fragment would be a useful probe for the cloning of flanking DNA.

Isolation of Mouse Genomic DNA Clones Flanking the AFP Gene. Mouse genomic libraries, prepared from random partial *Hae* III/*Alu* I digests of BALB/c embryonic DNA or partial *Hae* III digests of the plasmacytoma MOPC41 DNA (16) and the outer *Eco*RI fragments of the vector Charon 4A (17), were plated to a density of 20,000 phage per plate and screened by the procedure of Benton and Davis (18) using a unique copy DNA fragment derived from λAFP14 that had been labeled to $4–10 \times 10^7$ cpm/μg in the presence of [^{32}P]dCTP (15).

Restriction Endonuclease Analysis of λAFP15 and Mouse Genomic DNA. Approximately 25 μg of λAFP15 DNA was digested to completion with *Sma* I, an enzyme that cuts 3.7 kilobase (kb) pairs from the *Eco*RI site in the right arm of λ Charon 4A DNA and does not cut within the insert of AFP15. Portions of this DNA were partially cleaved with other restriction enzymes whose sites were to be mapped, and the serial digests were transferred to nitrocellulose filters after electrophoresis through 0.5% agarose gels. The filters were hybridized to a labeled 3.7-kb *Eco*RI/*Sma* I fragment isolated from λ Charon 4A DNA. The sizes of the partially digested fragments that hybridized to the probe were determined and the specific sites were mapped by subtraction.

High molecular weight genomic DNA was prepared from isolated nuclei (19) from BALB/c 18-day embryos by a procedure described by Wu *et al.* (20). Portions (20-μg) were digested with restriction endonucleases; the products were fractionated by agarose gel electrophoresis and transferred to nitrocellulose filters according to Southern (13). The filters were exposed to labeled DNA fragments for hybridization, washed, and exposed to x-ray film.

Abbreviations: AFP, α-fetoprotein; kb, kilobase(s).

Determination of DNA Sequence of λAFP14 and λAFP15. A *Bam*HI/*Eco*RI fragment of λAFP14 containing the first three coding blocks of the AFP gene and the *Eco*RI B fragment of λalb 6 containing the last two coding blocks of the albumin gene were cloned into pBR322 as described (10). Selected regions of the DNA were subjected to DNA sequence analysis by using the procedure of Maxam and Gilbert (21).

RESULTS

Orientations of AFP and Albumin Genes. The identification and 5'-to-3' polarity of the cloned AFP and albumin genes had been based on a comparison of genomic fragments and several cDNA clones (10, 11). DNA sequence analysis was performed on several small genomic coding blocks to confirm these assignments. The second coding block in the AFP gene was analyzed by the Maxam and Gilbert (21) procedure starting from a *Hpa* II site at the 5' side of the block. The sequence of the coding block (Fig. 1) agrees exactly with that determined from the sequence of AFP cDNA (7) and corresponds to amino acids 10–25 of the mature protein. The data in Fig. 1, plus further sequence determinations in the two adjacent blocks (unpublished data) show that the 5' and 3' coding–intervening sequence borders occur within triplet codons. Both intervening sequence–coding borders are in good agreement with the consensus sequence generated by Lerner *et al.* (22).

When a small coding block in *Eco*RI fragment B of the albumin gene (11) was analyzed, we discovered that the orientation of the gene was opposite to that reported previously. The error was due to a misalignment of a rat albumin cDNA clone, pralb 1, onto its template (24). Instead of its 1200-base-pair insert being colinear with the mRNA, it is the product of the blunt-end ligation of two albumin double-stranded cDNAs. As a result, the insert is composed of two discontinuous segments of albumin-specific cDNA.

The correct orientation of the mouse albumin gene is now shown in Figs. 1 and 2. The sequence of the 14th coding block of the albumin gene, (lower half of Fig. 1) contains the codons for the last 13 amino acids of the protein and 29 additional base pairs of 3' untranslated sequence. The deduced amino acid sequence agrees poorly with that published for the carboxy termini of human (5 of 13) or bovine (7 of 13) albumins (25, 26), but the agreement with rat albumin (10 of 13), as determined by Sargent *et al.* (23) from the sequence of albumin cDNA clones, is good. There are only five single-base differences in the mouse and rat sequences, of which two are silent third-base substitutions. The most striking difference is an extra six base pairs in the 3' untranslated DNA of the mouse. These six base pairs are directly preceded by the same sequence, C-A-A-C-C$\frac{T}{A}$, which suggests a recent duplication of this sequence in the mouse or, less likely, a deletion of the tandem

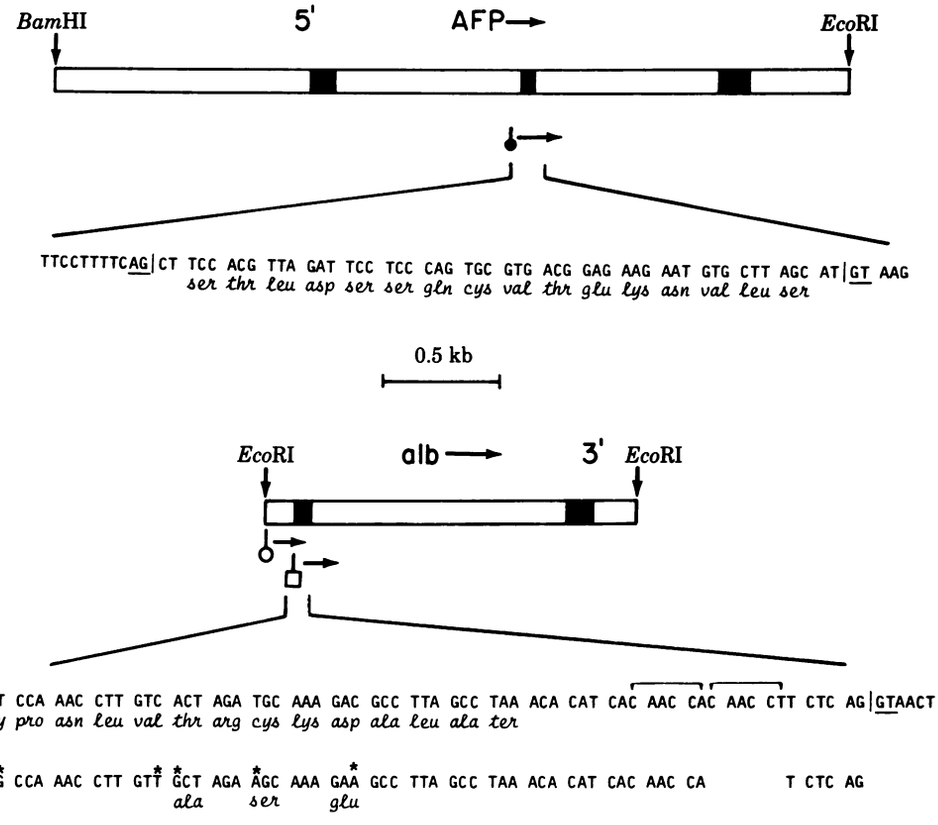

FIG. 1. DNA sequences in the AFP and albumin genes. Intervening sequence–coding borders are indicated by the vertical lines; underlined nucleotides represent the most highly conserved bases of the consensus border sequences (22). (*Upper*) Top line shows a *Bam*HI/*Eco*RI fragment of the Z *Eco*RI fragment of λAFP14 (11) (see Fig. 2) subcloned into pBR322. The dark blocks represent the positions of the first three coding blocks in the AFP gene, with the 5'-to-3' orientation as indicated. A fragment containing the second coding block was subjected to the Maxam and Gilbert (21) sequence procedure, beginning at a *Hpa* II site (dark circle) to the 5' side of the coding block. The sequence obtained is shown with the derived amino acid sequence for all complete triplet codons. (*Lower*) Top line shows the alb B *Eco*RI fragment of λalb 6 (11) (see Fig. 2, probe 3) which was subcloned into pBR322. The last two coding blocks of the albumin gene are shown as dark blocks. The sequence of the penultimate block was determined from an *Rsa* I site (open circle) at its 5' side and an *Ava* II site (open square) just within the coding sequence. The sequence obtained is shown below, with the derived amino acid sequence. Shown at the bottom is the sequence for rat albumin (23); the base substitutions are indicated by asterisks. Only those triplets that result in amino acid changes are translated below the rat nucleotide sequence. The repeated sequence in the mouse albumin 3' untranslated region is indicated by brackets.

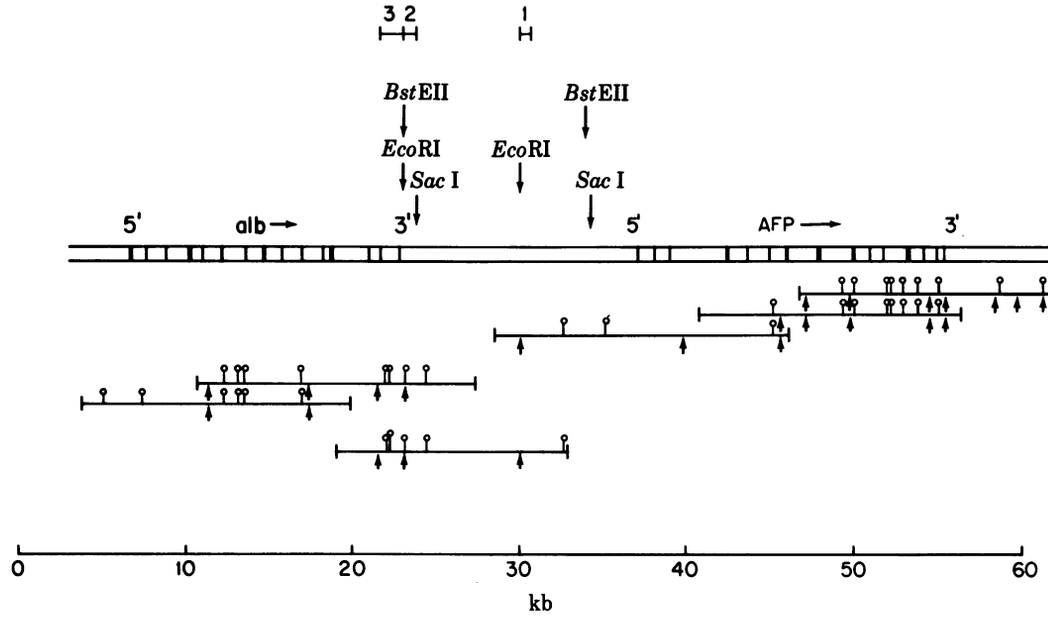

FIG. 2. Structure of the AFP–albumin gene locus in the mouse. The internal structures of the AFP and albumin genes and their positions relative to each other on the chromosome are shown in the middle. The dark blocks represent the positions and sizes of the 15 coding segments in each gene; the coding strands and direction of transcription also are indicated for each gene. Below the gene map are shown the regions of the locus contained in the overlapping λAFP and λalb phage. The arrows below each line represent internal *Eco*RI restriction sites; the circles above each line represent *Hin*dIII sites. a, λAFP8; b, λAFP7; c, λAFP14; d, λalb6; e, λalb5; f, λAFP15. Above the gene map, the DNA fragments used as probes in Figs. 2 and 3 are positioned on the chromosome. Only those *Bst*EII, *Eco*RI, and *Sac* I restriction fragments that span the region between genes and are detected in Fig. 3 are indicated.

repeat in the rat. The intervening sequence–coding borders also are in good agreement with the consensus sequences (22).

Cloning Sequences Adjacent to the AFP Gene. In order to extend the repertoire of cloned DNA sequences surrounding the AFP gene, it first was necessary to identify restriction fragments that were devoid of repetitive DNA (11) and thus could serve as hybridization probes. As a first step, λAFP14 DNA, a recombinant phage containing the 5′ end of the AFP gene (11) (see Fig. 2), was digested with various restriction endonucleases and the fragments were transferred to nitrocellulose after agarose gel electrophoresis. The fragments were hybridized to labeled total cellular mouse DNA under conditions such that only reiterated DNA would be detected. Those fragments that did not hybridize were then isolated and hybridized to an *Eco*RI digest of cellular DNA to ensure that they recognized only a single fragment. By this procedure, a 0.65-kb *Eco*RI/*Pvu* II fragment at the 5′ side of λAFP14 was identified and is referred to as probe 1 in Fig. 2.

Probe 1 was then used as a labeled hybridization probe to screen a BALB/c-derived genomic DNA library in Charon 4A by using the procedure of Benton and Davis (18). Of 5×10^5 phage screened, a single positive clone was identified and labeled "λAFP15." When λAFP15 DNA was cleaved with *Eco*RI, four internal *Eco*RI fragments were observed: 6.4, 3.5, 2.45, and 1.5 kb (Fig. 3, lane 2). The order of the *Eco*RI fragments, determined by partial digestion restriction mapping, was 5′ 2.45, 1.5, 6.4, 3.5 3′ (5′ refers to those sequences farthest upstream from the AFP gene). Of these, probe 1 hybridized to the 3′-most fragment, 3.5 kb (Fig. 3, lane 3). These results, plus extensive restriction mapping, implied that λAFP15 overlapped λAFP14 by 5.0 kb, and therefore the cloned region surrounding the AFP genes had been extended by approximately 9 kb in the 5′ direction (Fig. 2).

The cleavage sites of several other restriction enzymes (*Bam*HI, *Hin*dIII, *Sal* I, *Xho* I, *Pvu* II, *Pst* I, *Xba* I, *Sac* I) were determined in λAFP15 and a restriction map was constructed. Sixteen cleavage sites at the 5′ end coincided exactly with the map of the 3′ distal region of λalb 6 (11), a cloned segment of the 3′ half of the mouse albumin gene. In Fig. 2, only the *Hin*dIII sites have been included to illustrate the identity between λAFP15 and λalb 6; however, the same results were ob-

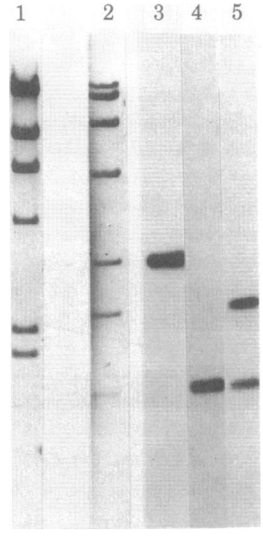

FIG. 3. Presence of both AFP and albumin flanking DNA in λAFP15. λAFP15 DNA (1 μg) was digested with *Eco*RI; the fragments were separated on a 1% agarose gel (lane 2) and transferred to multiple nitrocellulose filters. These were hybridized individually to: lane 3, an *Eco*RI/*Pvu* II fragment of λAFP14 (probe 1); lane 4, a 1.5-kb *Eco*RI fragment of λalb 6 (probe 2); and lane 5, the 0.5-kb insert of a 3′ distal albumin cDNA clone, pmalb 1 (11). Lane 1 shows a *Hin*dIII digest of λ*cI*857; the visible fragments are, from top to bottom, 22.0, 9.4, 6.6, 4.4, 2.3, and 1.9 kb.

tained with every enzyme used. Thus, it appeared that the insert in λAFP15 DNA contained sequences derived from the albumin gene.

To test this possibility, an *Eco*RI digest of λAFP15 was hybridized to a 1.5-kb *Eco*RI fragment [alb *Eco* B (11), shown in Fig. 1 and as probe 3 in Fig. 2] of λalb 6, which contains the last two coding blocks of the albumin gene and is composed of only unique copy DNA. An identically sized (1.5 kb) *Eco*RI fragment in λAFP15 was recognized (Fig. 3, lane 4). In addition, the 0.5-kb insert of an albumin cDNA clone, pmalb 1, which contains sequences complementary to the last five coding blocks of the albumin gene, hybridized to the same 1.5-kb *Eco*RI fragment plus the adjacent 2.45-kb fragment (Fig. 3, lane 5). From both the restriction mapping and hybridization data, we conclude that λAFP15 contains the 3′ end of the albumin gene as well as 5′ AFP flanking DNA (Fig. 2). This result places the two genes in tandem on the chromosome with the 3′ terminus of the albumin gene 13.5 kb upstream from the AFP gene.

It was possible that one or more of the recombinant phage shown in Fig. 2 had been generated after a deletion, *in vivo* or *in vitro*, during the construction of the library. In that case, the predicted 13.5-kb distance between genes would be incorrect. To ensure that this was not the case, BALB/c genomic DNA, prepared from 18-day embryos, was cleaved with restriction enzymes that cut within the gap between the genes. For example, if λAFP15 represents the true intergenic distance, *Sac* I digestion of genomic DNA should generate a 9.2-kb fragment that would hybridize to probe 1. Such a fragment was detected in genomic DNA (Fig. 4, lane 2). Likewise, probe 2 from λalb 6 should hybridize to the genomic fragment equivalent to the internal 6.4-kb *Eco*RI fragment of λAFP15, and it did (Fig. 4, lane 3). Finally, a 9.5-kb *Bst*EII fragment was found to hybridize to both probes. These experiments demonstrate that the organization of the murine AFP and albumin genes, as determined by the analysis of DNA sequences in λAFP14 and λAFP15, reflects their original structure in the chromosome.

DISCUSSION

Previous reports (5–11) have argued that the similarities in both the sequence and the organization of the AFP and albumin genes made it highly likely that the genes arose as the consequence of a duplication of a common ancestral gene. This argument is not changed by reversing the polarity of the albumin gene because the sizes of the 15 coding blocks in both genes form a symmetrical pattern. The present finding of the close linkage of these two genes in the mouse genome strongly reinforces the likelihood of an early duplication. The two genes have apparently been maintained in close proximity for a long period of time. Reports of a fetal-specific α-globulin, with properties and tissue distributions similar to those of mammalian AFP, in chickens (27) and sharks (28) would argue that the duplication must have occurred 300–500 million years ago. Certainly the absence of any nucleotide cross-homology, as determined by hybridization, between coding sequences supports a distant date for the duplication. Therefore, finding them in tandem raises the question of whether this close proximity is required for their reciprocal expression during fetal and neonatal development.

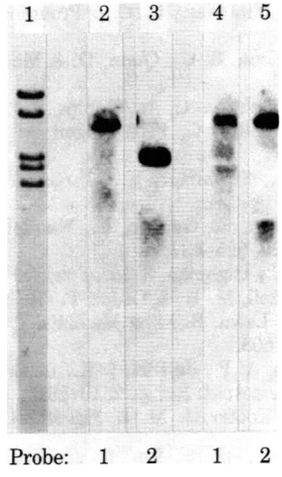

FIG. 4. Detection of intergenic region of the AFP–albumin locus in mouse genomic DNA. High molecular weight DNA from BALB/c embryos was cleaved with *Sac* I (lane 2), *Eco*RI (lane 3), or *Bst*EII (lanes 4 and 5). Lane 1 shows a *Bam*HI digest of λ*cI*857 DNA in which the fragment sizes, from top to bottom, are: 17.4, 12.3, 7.3, 6.5, and 5.54 kb.

Although various multigene families have recently been described in detail, their chromosomal locations have been determined in only a few instances. Of these, the AFP and albumin genes represent a unique example of a relatively ancient gene pair whose expression is sequential rather than coordinate during development. Thus, there is little precedent to assist one in assessing the likelihood that these genes would have remained in tandem without selective pressure. When one examines those gene families of longevity equal to that of AFP and albumin, no consistent linkage pattern is apparent. In the globin gene family, the α- and β-globin gene clusters, thought to have diverged about 500 million years ago, are unlinked in humans (29–31), mice (32), and chickens (33). Likewise, light and heavy chain immunoglobulin gene clusters arose from a common ancestor at approximately the same time and are present at separate chromosomal loci (34, 35). These gene families differ from the AFP and albumin genes in that they are examples of multimeric protein genes, in which coordinate regulation of the individual subunits is accomplished in the absence of linkage. It may be that once the α and β chains evolved to perform noninterchangeable functions in the hemoglobin molecule, the need to protect the cell from the loss of one gene cluster by unequal crossing-over may have provided the selective advantage for dispersion of the two gene families. The presence of linked α- and β-globin genes in *Xenopus laevis* (36) suggests that the dispersion occurred after the generation of amphibians and before that of chickens.

The gene clusters that have remained closely linked in higher organisms include old genes, like the histone genes (37, 38) and the A-B pair of chorion genes in *Bombyx mori* (39) and more recently established families such as the β-globin gene clusters in humans (40), rabbits (41), mice (42), and chickens (43) and the ovalbumin/X/Y gene cluster in chickens (44). These gene clusters are thought to be fixed in the genome by at least two different mechanisms. As pointed out by Tiemeier *et al.* (45), the net effect of the rapid accumulation of point mutations, deletions, and insertions in intervening and flanking sequences is to decrease the target size for recombination, thereby limiting the rate of unequal crossovers. That this process is working in the AFP and albumin gene locus is reflected in the complete divergence, in terms of both size and sequence, of the corresponding 14 intervening sequences in each gene (11). The flanking sequences as well bear no residual homology except for a repeated sequence present in many copies in the genome.

A second mechanism for maintaining gene families in tandem over a long period in evolution may be their establishment in opposite orientations relative to the direction of transcription. If the inversion were of sufficient size, this would be as effective a mechanism to prevent the loss of one gene copy as the dis-

persion of the genes or the divergence of their noncoding sequences. That divergent orientation is seen in each of the oldest gene clusters—that is, the chorion A-B gene pair (39) and the histone genes in *Drosophila* (38)—argues that gene inversion is not a rare event.

As yet, there is little evidence to implicate the structural organization of a multigene family locus in its regulation in higher organisms. In the human β-globin locus, it has been argued, on the basis of the different extents of gene deletions in individual cases of β-thalassemia and hereditary persistence of fetal hemoglobin, that the switch from the synthesis of fetal γ- to adult β-globin chains requires DNA sequences between the two gene sets which would act in *cis* (46–48). That is, the developmental regulation of the locus requires the genes to be tandemly arranged, in the order of their temporal expresson; embryonic→fetal→adult β-globin. Certainly, the demonstration that the adult-expressed albumin gene is found upstream from the fetal AFP gene argues that there is no absolute requirement for such an arrangement in all developmentally regulated loci. Although there have been 14 reported cases of hereditary analbuminemia in humans (49), in which there is a decrease or absence of serum albumin, as well as an animal model in rats (50), compensatory synthesis of AFP in the adults has not been tested. An analysis of the genetic defects in these cases may illuminate the mechanism underlying the reciprocal expression of the AFP and albumin genes in development.

We are grateful to Jon G. Seidman and Philip Leder for providing the mouse DNA libraries. This work was supported by Grant CA 28050 from the National Cancer Institute, a Basil O'Connor Starter Grant, Grants CA 06927 and RR 05539 from the National Institutes of Health awarded to The Institute for Cancer Research, and an appropriation from the Commonwealth of Pennsylvania. S.M.T. is the recipient of an American Cancer Society Junior Faculty Award; R.W.S. has a Postdoctoral Fellowship from the U.S. Public Health Service.

1. Abelev, G. I. (1974) *Transplant. Rev.* **20**, 3–37.
2. Ruoslahti, E. & Seppala, M. (1979) *Adv. Cancer Res.* **29**, 276–346.
3. Tamaoki, T., Miura, K., Lin, T. & Banks, P. (1976) in *Oncodevelopmental Gene Expression*, eds. Fishman, W. H. & Sell, S. (Academic, New York), pp. 115–122.
4. Sala-Trepat, J. M., Dever, J., Sargent, T. D., Thomas, K., Sell, S. & Bonner, J. (1979) *Biochemistry* **18**, 2167–2178.
5. Ruoslahti, E. & Engvall, E. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 4641–4644.
6. Ruoslahti, E. & Terry, W. D. (1976) *Nature (London)* **260**, 804–805.
7. Gorin, M. B., Cooper, D. L., Eiferman, F., van de Rijn, P. & Tilghman, S. M. (1981) *J. Biol. Chem.* **256**, 1954–1959.
8. Liao, W. S. L., Hamilton, R. W. & Taylor, J. M. (1980) *J. Biol. Chem.* **255**, 8046–8049.
9. Innis, M. A. & Miller, D. L. (1980) *J. Biol. Chem.* **255**, 8994–8996.
10. Gorin, M. B. & Tilghman, S. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1351–1355.
11. Kioussis, D., Eiferman, F., van de Rijn, P., Gorin, M. B., Ingram, R. S. & Tilghman, S. M. (1981) *J. Biol. Chem.* **256**, 1960–1967.
12. D'Eustachio, P., Ingram, R. S., Tilghman, S. M. & Ruddle, F. (1981) *Somat. Cell Genet.* **7**, 289–294.
13. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
14. Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3683–3687.
15. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
16. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, E. (1978) *Cell* **15**, 687–701.
17. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L. & Smithies, O. (1977) *Science* **196**, 161–169.
18. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
19. Schibler, U., Marcu, K. B. & Perry, R. P. (1978) *Cell* **15**, 1495–1509.
20. Wu, C., Wong, Y.-C. & Elgin, S. C. R. (1979) *Cell* **16**, 807–814.
21. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
22. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) *Nature (London)* **283**, 220–224.
23. Sargent, T. D., Yang, M. & Bonner, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 243–246.
24. Kioussis, D., Hamilton, R., Hanson, R. W., Tilghman, S. M. & Taylor, J. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4370–4374.
25. Behrens, P. Q., Spiekerman, A. M. & Brown, J. R. (1975) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **34**, 591.
26. Brown, J. R. (1975) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **34**, 591.
27. Lindgren, J., Vaheri, A. & Ruoslahti, E. (1974) *Differentiation* **2**, 233–236.
28. Gitlin, D., Perricelli, A. & Gitlin, J. D. (1973) *Comp. Biochem. Physiol. B* **46**, 207–215.
29. Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W. F., Ruddle, F., Lawrence, J., Creagan, R. & Kucherlapati, R. (1977) *Cell* **12**, 205–218.
30. Gusella, J., Varsanyi-Breiner, A., Kao, F.-T., Jones, C., Puck, T. T., Keys, C., Orkin, S. & Housman, D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5239–5243.
31. Jeffreys, A. J., Craig, I. W. & Francke, U. (1979) *Nature (London)* **281**, 606–608.
32. Popp, R. A. (1979) in *Inbred and Genetically Defined Strains of Laboratory Animals*, eds. Altman, P. L. & Katz, D. D. (Am. Soc. Exp. Biol., Bethesda, MD), Part 1, p. 105.
33. Hughes, S. H., Stubblefield, E., Payvar, F., Engel, J. D., Dodgson, J. B., Spector, D., Cordell, B., Schimke, R. T. & Varmus, H. E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1348–1352.
34. Swan, D., D'Eustachio, P., Leinwand, L., Seidman, J., Keithley, D. & Ruddle, F. H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2735–2739.
35. D'Eustachio, P., Pravtcheva, D., Marcu, K. & Ruddle, F. H. (1980) *J. Exp. Med.* **151**, 1545–1550.
36. Jeffreys, A. J., Wilson, V., Wood, D., Simons, J. P., Kay, R. M. & Williams, J. G. (1980) *Cell* **21**, 555–564.
37. Kedes, L. (1979) *Annu. Rev. Biochem.* **48**, 837–870.
38. Lifton, R. P., Goldberg, M. L., Karp, R. W. & Hogness, D. S. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1047–1051.
39. Jones, C. W. & Kafatos, F. C. (1980) *Cell* **22**, 855–867.
40. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., de Riel, 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.
41. Lacy, E., Hardison, R. C., Quon, D. & Maniatis, T. (1979) *Cell* **18**, 1273–1283.
42. Jahn, C. L., Hutchison, C. A., Phillips, S. J., Weaver, S., Haigwood, N. L., Voliva, C. F. & Edgell, M. H. (1980) *Cell* **21**, 159–168.
43. Dodgson, J. B., Strommer, J. & Engel, J. D. (1979) *Cell* **17**, 879–887.
44. Heilig, R., Perrin, F., Gannon, F., Mandel, J. L. & Chambon, P. (1980) *Cell* **20**, 625–637.
45. Tiemeier, D. C., Tilghman, S. M., Polsky, F. I., Seidman, J. G., Leder, A., Edgell, M. H. & Leder, P. (1978) *Cell* **14**, 237–244.
46. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1979) *Nature (London)* **279**, 598–603.
47. Tuan, D., Biro, A. P., de Riel, J. K., Lazarus, H. & Forget, B. G. (1979) *Nucleic Acids Res.* **6**, 2519–2544.
48. Bernards, R., Kooter, J. M. & Flavell, R. A. (1979) *Gene* **6**, 265–280.
49. Buehler, B. A. (1978) *Ann. Clin. Lab. Sci.* **8**, 283–286.
50. Nagase, S., Shimamune, K. & Shumiya, S. (1979) *Science* **205**, 590–591.