Structure of the Genes That Do Not Rearrange

Structure and in vitro Transcription of Human Globin Genes

Nicholas J. Proudfoot, Monica H. M. Shander, Jim L. Manley
Malcolm L. Gefter, Tom Maniatis

Summary. The α-like and β-like subunits of human hemoglobin are encoded by a small family of genes that are differentially expressed during development. Through the use of molecular cloning procedures, each member of this gene family has been isolated and extensively characterized. Although the α-like and β-like globin genes are located on different chromosomes, both sets of genes are arranged in closely linked clusters. In both clusters, each of the genes is transcribed from the same DNA strand, and the genes are arranged in the order of their expressions during development. Structural comparisons of immediately adjacent genes within each cluster have provided evidence for the occurrence of gene duplication and correction during evolution and have led to the discovery of pseudogenes, genes that have acquired numerous mutations that prevent their normal expression. Recently, in vivo and in vitro systems for studying the expression of cloned eukaryotic genes have been developed as a means of identifying DNA sequences that are necessary for normal gene function. This article describes the application of an in vitro transcription procedure to the study of human globin gene expression.

The recent application of molecular cloning procedures to the human globin gene family has led to significant advances in the understanding of the structure, chromosomal arrangement, and evolution of these genes [see references in (*1-4*)]. Individual members of this relatively small gene family are expressed at different times during development, and mutations have been described that affect the expression of adult α-globin or β-globin genes (α- and β-thalassemia) or the switch from fetal to adult globin gene expression. Thus, human globin genes provide a model system for studying the molecular genetics of eukaryotic gene regulation and the molecular basis of human genetic disease.

Nicholas J. Proudfoot is a senior research fellow, Monica H. M. Shander is a postdoctoral fellow, and Tom Maniatis is a professor in the Division of Biology, California Institute of Technology, Pasadena 91125. Jim L. Manley is a postdoctoral fellow and Malcolm L. Gefter is a professor in the Department of Biology at the Massachusetts Institute of Technology, Cambridge 02139.

Chromosomal Arrangement and Structure of Globin Genes

The known α-like and β-like globins are listed in Table 1. The earliest embryonic hemoglobin tetramer, Hb(Gower 1), consists of ε (β-like) and ζ (α-like) polypeptide chains. Beginning at approximately 8 weeks of gestation, the embryonic chains are gradually replaced by the adult α-globin chain and two different fetal β-like chains, designated $^{G}\gamma$ and $^{A}\gamma$. The γ chains differ only in the presence of glycine or alanine, respectively, at position 136. During the transition period between embryonic and fetal development, Hb(Gower 2) ($\alpha_2\epsilon_2$) and Hb(Portland) ($\zeta_2\gamma_2$) are detected. HbF ($\alpha_2\gamma_2$) eventually becomes the predominant Hb tetramer throughout the remainder of fetal life. Beginning just before birth, the γ-globin chains are gradually replaced by the adult β-globin and δ-globin polypeptides. At 6 months after birth, 97 to 98 percent of the hemoglobin is HbA ($\alpha_2\beta_2$), while HbA_2 ($\alpha_2\delta_2$) accounts for approximately 2 percent. Small amounts of HbF (1 percent) are also found in adult peripheral blood. The site of erythropoiesis changes from the yolk sac in the early embryo, to the developing liver, spleen, and bone marrow in the fetus, and finally to the bone marrow in adults [see references in (*3, 4*)].

In summary, the α-like and β-like globin gene families have coordinated programs for differential gene expression. The primary difference between the two gene families is that two switches in gene expression (embryonic to fetal to adult) are observed for the β-like genes, while a single switch results in the turnoff of embryonic ζ-globin production early in fetal life.

Chromosomal Arrangement and Structure of Globin Genes

The entire β-globin (*5*) and α-globin (*6*) gene clusters have been isolated in sets of overlapping bacteriophage recombinants, which were obtained from libraries of random, high-molecular-weight, human DNA (*7, 8*). The linkage arrangement of the human α-like and β-like globin gene clusters, which was established by genomic blotting (*9, 10*) and molecular cloning (*5, 6, 8, 11, 12*) experiments, is shown in Fig. 1. Although the size of the two gene clusters differs by almost a factor of 2, the genes in both clusters are arranged on the chromosome in the order of their expression during development. A similar pattern of gene organization has been found in the rabbit (*13, 14*)

0036-8075/80/0919-1329$02.00/0 Copyright © 1980 AAAS

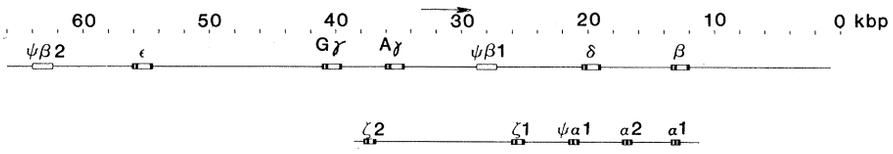

Fig. 1. The linkage arrangement of human β-like and α-like globin genes. The top line shows the relative locations of the five functional β-like globin genes and the two β-like sequences, which do not correspond to any known β-globin polypeptide chain. The bottom line shows the map of the four functional α-like globin genes and the α-globin pseudogene $\psi\alpha 1$. The mRNA coding and intervening sequences are designated by filled and open rectangles, respectively. The direction of transcription of all the globin genes is indicated by the arrow.

and mouse (*15*) β-like globin gene clusters. All of the known human globin polypeptides can be accounted for by the genes shown in Fig. 1. However, both gene clusters contain additional sequences that are detected by globin gene hybridization probes but cannot be identified with any of the known globin polypeptides. Three genes, designated $\psi\beta 1$, $\psi\beta 2$ (*5*), and $\psi\alpha 1$ (*6, 16*), fall into this category. Structural analysis of the $\psi\alpha 1$ gene indicates that it is a pseudogene (a gene that displays significant homology to a functional gene but has mutations that prevent its expression) (*16*). Similar pseudogenes have been identified at corresponding positions in the rabbit (*13, 17*) and mouse (*15*) β-like globin gene clusters.

Globin gene fine structure. Since the discovery of an intervening sequence (intron) in the rabbit β-globin (*18*) and the mouse α-globin (*19*) and β-globin (*20*) genes, two introns have been identified in all of the functional globin genes thus far studied [see references in (*2*, *21*)]. In particular, the five expressed human β-like globin genes are interrupted by two introns at identical locations: the first, 122 to 130 base pairs (bp) in length, is located between codons 30 and 31; and the second, 850 to 900 bp, is between codons 104 and 105 (*2*) (Fig. 2). Similarly, the locations of introns in the human (*22*) and mouse (*23*) α-globin genes are identical and are analogous to the positions of introns in β-like globin genes (Fig. 2). In the case of the mouse (*24*), rabbit (*13, 25*), and human (*26, 27*) β-globin genes, both the messenger RNA (mRNA) coding sequences (exons) and intron sequences are transcribed to produce a detectable nuclear mRNA precursor which is processed or spliced to give a mature globin mRNA. Furthermore, the 5′ ends of the mouse and rabbit β-globin nuclear precursors are coterminal with mature mRNA (*25, 28*). These nuclear precursors could be the primary globin mRNA transcript. In this case, the sequence encoding the 5′ end of the mature mRNA would correspond to the transcriptional initiation site. Alternatively, these nuclear transcripts may represent intermediates in the mRNA maturation. In this case, the transcriptional initiation site would be located proximal to the 5′ end of the mRNA sequence.

The complete nucleotide sequences of the five human β-like globin genes have been determined (*29, 30*), and a detailed comparison of these and other mammalian globin gene sequences has been presented (*2*). This sequence comparison revealed interesting sequence homologies in regions that are potentially involved in globin gene transcription and splicing. In particular, alignment of the sequences on the 5′ sides of the human β-like globin genes revealed two blocks of sequence homology, which are present in analogous positions adjacent to most eukaryotic genes (*31*). The first homology block is an AT-rich (A, adenine; T, thymine) sequence originally identified in the *Drosophila* histone gene cluster called the Hogness box (*32*). A comparison of a number of different β-like globin genes has revealed that the AT-rich sequence CATAAA (C, cytosine) is found 31 ± 1 bp on the 5′ side of the mRNA capping sites, but that the sequence shared by all of the β-like genes is ATA. This sequence was therefore designated the ATA box (*2*). The second homology block (designated the CCAAT box) is located 77 ± 10 bp on the 5′ side of each gene. A possible role of these sequences in transcriptional initiation, RNA processing, or both, has been discussed (*2*).

Previous comparisons of noncoding sequence in human, mouse, and rabbit β-like globin genes indicated that these regions diverged by deletion and addition as well as by simple base substitution (*2, 33, 34*). Examination of the nucleotide sequences surrounding putative deletion sites suggests that short (two to eight nucleotide sequences) and direct repeats are involved in the generation of deletions. This pattern is remarkably similar to that observed for preferred deletion sites (hot spots) in the lac *i* gene of *Escherichia coli* (*35*). A model for the involvement of short, direct repeat sequences in the generation of deletions in the noncoding regions of β-like globin genes during evolution has been proposed (*2*).

Table 1. Globin polypeptides and hemoglobin tetramers detected at different stages of human development.

Globin polypeptides	Embryo	Fetus	Adult
α-like	ζ	α	α
β-like	ϵ	$^{G}\gamma$, $^{A}\gamma$	δ, β
Tetramers	$\zeta_2\epsilon_2$ (Gower 1) $\alpha_2\epsilon_2$ (Gower 2) $\zeta_2\gamma_2$ (Portland)	$\alpha_2\gamma_2$ HbF	$\alpha_2\delta_2$ HbA$_2$ $\alpha_2\beta_2$ HbA

Globin Gene Duplication

A common feature of globin gene clusters is the occurrence of two immediately adjacent genes, which are coordinately expressed during a given developmental stage. Examples of this are the human δ-β, $^{G}\gamma$-$^{A}\gamma$, $\alpha 1$-$\alpha 2$, and $\zeta 1$-$\zeta 2$ globin gene pairs. The δ and β genes are highly homologous in the coding regions, but the noncoding sequences within and surrounding the two genes have diverged considerably (*2*). Extensive divergence of noncoding regions has also been observed in some other closely linked, coordinately expressed globin gene pairs (*13, 15, 33*). In contrast, the two members of the $^{G}\gamma$-$^{A}\gamma$ gene pair are virtually identical to one another throughout their coding, intervening, and flanking sequences (*30*). Although the nucleotide sequences of linked human α-globin genes have not yet been determined, restriction mapping and heteroduplex analysis of the $\alpha 1$ and $\alpha 2$ genes indicate that the sequences within and flanking these two genes are virtually identical. Each α-globin gene is located within an approximately 4-kbp (kilobase pair) region of homology interrupted by two small regions of nonhomology (*6*).

The extensive sequence homology within and flanking the $^{G}\gamma$-$^{A}\gamma$ and $\alpha 1$-$\alpha 2$ gene pairs appears to be the product of a mechanism for gene matching during evolution (*6, 30, 36*). Based on the nearly

identical distribution of restriction sites surrounding the α-globin genes in a number of primate species, it has been suggested that the α-globin gene duplication occurred before the time of primate divergence (*36*). Differences between the α-globin amino acid sequences of various primate species are consistent with sequence drift following primate divergence (*37*). However, intraspecies comparisons show much less divergence, indicating that the α-globin genes within a species have been corrected against one another. Maintenance of homology among a family of evolving genes within a species has been termed "concerted" evolution (*36*). Gene conversion and expansion-contraction of gene number by homologous but unequal crossing-over have been proposed as mechanisms for concerted evolution (*38*).

The precise end points of the α-globin gene duplication unit have been located by nucleotide sequence analysis (*16*). The left end point of the duplication is located immediately adjacent to the putative poly(A) (polyadenylate) addition site of ψα1, while the right end point is found 15 bp on the 3′ side of the poly(A) addition site of α1. The 15-bp sequence on the 3′ side of the poly(A) addition sites of ψα1, α2, and α1 consists of a repeated pentanucleotide (GCCTG) (G, guanine), separated by TGTGT. The occurrence of this sequence in all three genes and its location with respect to the end points of the α-globin gene duplication suggest that this sequence might be associated with the mechanism by which the genes were duplicated or corrected (*16*). Zimmer *et al.* (*36*) and Lauer *et al.* (*6*) have proposed a model for α-gene correction that involves interchromosomal, unequal crossing-over events. Proudfoot and Maniatis (*16*) have suggested that the pentanucleotide repeat acts as a boundary or terminator for the recombination event. Evidence that α-globin gene sequence matching could occur by expansion and contraction of gene number by unequal crossing-over is provided by the frequent occurrence of chromosomes containing one (*39*) or three (*40*) adult α-globin genes in some human populations (*6, 36*). The chromosome containing only one α-globin gene is associated with the common form of α-thalassemia, designated α-thalassemia 2. Comparison of the end points of the deletion associated with this disorder, with the location of blocks of homologous sequence within the α1-α2 gene duplication, strongly suggests that the deletion results from unequal crossing-over between homologous sequences (*6*). Interestingly, deletions that are indistinguishable from those found in α-thalassemia 2 occur in the cloned gene cluster during propagation in *E. coli* (*6*).

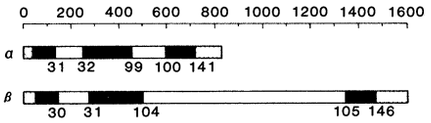

Fig. 2. The fine structure of α- and β-globin genes. The canonical structures for the human α-like and β-like globin genes are drawn to approximate scale. Solid and open boxes represent coding (exon) and noncoding (intron) sequences, respectively. The α-like globin genes contain introns of approximately 95 and 125 bp, located between codons 31 and 32, and 99 and 100, respectively. The β-like globin genes contain introns of approximately 122 to 130 and 850 to 900 bp, located between codons 30 and 31, and 104 and 105, respectively.

Analysis of the complete nucleotide sequences of cloned $^{G}\gamma$ and $^{A}\gamma$ genes has led to formulation of a specific intrachromosomal gene conversion model to explain sequence matching between linked genes (*30*). The $^{G}\gamma$ and $^{A}\gamma$ genes on one chromosome are identical in the region on the 5′ side of the center of the large intron, yet show greater divergence on the 3′ side of that position. Examination of the boundary between the conserved and divergent regions revealed a block of "simple sequence" DNA poly-(TG) (*30*). Slightom *et al.* (*30*) have proposed that this simple sequence is a preferred site for initiation of recombination events that lead to unidirectional gene conversion.

The α- and β-Globin Pseudogenes

As was mentioned above, globin gene sequences that cannot be identified with known globin polypeptide chains have been detected in several mammalian species (*5, 6, 13, 15–17, 41*). Nucleotide sequence analysis of a rabbit β pseudogene (β2) (*17*), a human α pseudogene (ψα1, Fig. 1) (*16*), a mouse β pseudogene (*15*), and a mouse α pseudogene (*41*) has demonstrated a variety of structural differences between each gene and its functional counterpart. The human β-like sequences (ψβ1 and ψβ2, Fig. 1) (*5*) have not yet been extensively characterized.

Each pseudogene that has been analyzed exhibits 75 to 80 percent sequence homology when compared with its corresponding normal gene (excluding the 5′ side of the mouse β pseudogene, which is not homologous to the adult mouse β gene) (*15*). None of these pseudogenes can encode a functional globin polypeptide due to the presence of small deletions or insertions that result in alterations of the translational reading frame. In many cases, these frameshifts lead to the presence of in-phase termination codons. In addition, one or more of the intron-exon junctions of rabbit ψβ2, mouse βH3, and ψα1 are different from the sequence common to splicing junctions in globin genes and all expressed genes studied to date (*42*). Thus, even if these pseudogenes are transcribed, it is unlikely that they would normally produce an mRNA.

It is interesting to note that in all of the mammalian globin gene clusters thus far characterized, a pseudogene is found between the embryonic (or fetal) genes and the adult genes. It is possible that pseudogenes have some as yet unidentified function in globin gene clusters. Alternatively, pseudogenes may be the products of gene duplication and subsequent sequence divergence (*16, 17*). The variation in human α-globin gene number observed in present-day populations and the location of ψα1 within the α-like globin gene cluster are consistent with the latter possibility. As shown in Fig. 2, ψα1, α2, and α1 are separated from each other by approximately 4 kbp, which is the size of the α1-α2 duplication unit noted above. The nucleotide sequence of ψα1 indicates that it is α-like rather than ζ-like (*16*). It therefore seems possible that ψα1 was once part of a set of three functional α-globin genes.

A novel mouse α-globin pseudogene has recently been described (*41*). As in the case of the pseudogenes described above, the mouse α-globin pseudogene has frameshift mutations that would result in premature translational termination. However, unlike the other pseudogenes, the mouse gene is missing both introns. The mechanism by which this pseudogene arose and its location with respect to the normal α-globin gene cluster is unknown.

Repetitive Sequences in Globin Gene Clusters

Cross-hybridization experiments between the intragenic sequences of the human β-gene clusters (*5*) and α-gene clusters (*43, 47*) revealed a nonglobin repeat sequence that is interspersed within the globin gene clusters and also repeated many times in the human genome. Nucleotide sequence analysis of the repetitive sequences within the β-globin gene cluster (*44*) indicates that they are members of a particular repeat sequence family, the Alu family, which is reiterated approximately 300,000 times in the human genome (*45*). In addition, the re-

peats are transcribed in vitro by RNA polymerase III (*46, 47*). They show sequence homology with an abundant class of small nuclear RNA's (*44, 48*) and with double-stranded, heterogeneous, nuclear RNA (*5, 44, 47, 49*). Finally, the repeats contain a sequence that is homologous to a sequence found near the replication origin of SV40, polyoma, and BK DNA tumor viruses (*44*). A similar set of repetitive sequences has been identified with a cluster of rabbit β-like globin genes (*50*). At present, there is little information regarding the expression or function of these interesting repetitive elements in vivo.

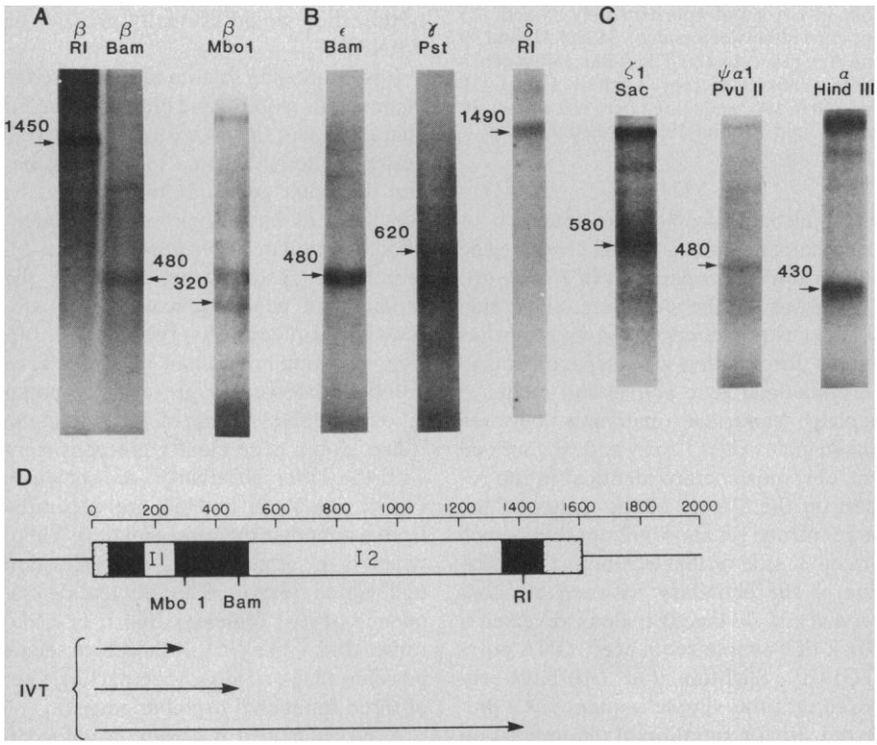

Fig. 3. In vitro transcription of human β-like and α-like globin genes. The human globin genes were originally isolated as lambda recombinants and subsequently were subcloned into pBR322 as follows: β, Pst 4.4 kbp; δ, Pst 2.3 kbp; $^{G}\gamma$, Pst 4.0 kbp; ε, Bam 0.7 kbp; α1, Sst 4.3 kbp; ψα1, Bam-Hind III 7.3 kbp; ζ1, RI (linker)-Bam 4.7 kbp (*5–7, 12*). Subcloned DNA's were digested with various restriction enzymes chosen to cleave the gene sequence at a specific position. In some cases, these digests were purified by phenol extraction and ethanol precipitation and then added directly to in vitro transcription reactions. Alternatively, the globin genes containing DNA fragments were isolated by horizontal agarose gel electrophoresis followed by electroelution into 10 m*M* tris-Cl, *p*H 8.0, 0.1*M* NaCl. The eluant was passed over a 0.5-ml DEAE-Sephadex column (equilibrated with 0.1*M* tris-Cl, *p*H 8.0, 0.1*M* KCl). After extensive washing with equilibration buffer, the column was eluted with 0.1*M* KCl (0.3 ml), then 0.4*M* KCl (0.3 ml), and finally 0.6*M* KCl (0.6 ml). The 0.6*M* KCl wash was collected in 50-μl amounts which were assayed for DNA content by agarose gel electrophoresis. The DNA obtained was precipitated twice with ethanol and washed with 70 percent ethanol. Whole cell in vitro transcription extracts were prepared according to Manley *et al.* (*54*). In vitro transcription reactions were as described (*54*) with some modifications. Plasmid DNA digests (50 μg) were used in reactions but only 1 to 5 μg of purified DNA fragments. A ribonuclease inhibitor, ribonucleoside-vanadyl complex (*73*), was added at 10 m*M* to the deoxyribonuclease step after the in vitro transcription reaction. The RNA purification was simplified to two extractions with phenol and chloroform and one extraction with chloroform followed by one ethanol precipitation with carrier transfer RNA (tRNA) and sodium acetate (0.25*M*). Finally, the RNA was dissolved in aqueous 5 m*M* methyl mercury and run on 2 percent agarose, 5 m*M* methyl mercury gels (*74*) with an Alu restriction enzyme digest of pBR322 (*75*) as size markers. After electrophoresis, the gel was soaked in 0.5*M* ammonium acetate to inactivate the methyl mercury, stained with ethidium bromide, and photographed under ultraviolet light. The presence of a discrete 18*S* RNA band was indicative of a fully recovered, undegraded RNA sample. Finally, the gel was dried and autoradiographed in a cassette with a preflashed film and intensifying screen at −70°C (*76*). Exposures of 6 hours were normally sufficient. (A) In vitro transcripts at the β-globin gene truncated as indicated in (D). (B) In vitro transcripts of the β-like globin genes ε, $^{G}\gamma$, and δ truncated as indicated. (C) In vitro transcripts of the α-like globin genes ζ1, ψα1, and α2 truncated as indicated. The numbers adjacent to the arrows indicate the size of marked band which, in each case, agreed with the predicted transcript sizes. However, in the case of ζ1, the 5′ end of the gene has not been identified by sequence analysis. Since in many cases the different lanes were derived from different experiments, the intensities of the various globin gene transcripts cannot be directly compared. However, the βRI, βBam, and δRI came from the same fractionation and exposure as did ψα1 and α. These particular band intensities are therefore comparable.

In vitro Transcription of Human Globin Genes

Although the structural studies described above have provided much useful information, the identification of regulatory sequences, such as transcriptional initiation sites, binding sites for regulatory proteins, and RNA processing sites, require the use of in vivo or in vitro assays for gene expression. In vivo assays include the use of the DNA-mediated gene transfer procedure (*51*) and SV40 vector systems (*52*). The recent development of cell-free extracts for RNA polymerase II-dependent transcription of cloned eukaryotic genes provides an in vitro approach to the study of globin gene expression.

Two in vitro transcription systems have been described. One system consists of a cytoplasmic extract, which requires the addition of purified RNA polymerase II for activity (*53*), while the other system consists of a concentrated whole cell extract with endogenous RNA polymerase II activity (*54, 55*). In both systems, specific transcription of adenovirus genes was demonstrated by the fact that the capped 5′ terminus of the in vitro transcript is indistinguishable from that found in vivo (*53, 54*). The general applicability of these in vitro transcription systems was recently demonstrated by specific transcription of the chicken conalbumin and ovalbumin genes (*56*) and the mouse β-globin gene (*57*). We report the results of an in vitro transcription study of human globin genes for which a whole cell extract procedure was used (*54*).

Analysis of embryonic, fetal, and adult globin gene transcripts. We used a truncated template assay (*53*) to determine whether individual globin genes can function as templates for specific in vitro transcription. The principle of this assay is illustrated in Fig. 3D. A DNA fragment containing the human β-globin gene is digested with a restriction enzyme that recognizes one or more sites within the gene. For example, if the human β-globin gene is digested with Eco RI, Bam HI, or Mbo 1, transcripts of approximately 1450, 480, and 320 nucleotides, respectively, should be detected in vitro if transcription begins near the mRNA capping site of the gene. Such transcripts are in fact made when Eco RI-, Bam HI-, or Mbo 1-truncated β-globin gene fragments are added to in vitro transcription extracts (Fig. 3A). Similarly, in vitro transcripts of the expected size are observed when the human globin genes ε, $^{G}\gamma$, and δ, are analyzed (Fig. 3B). The efficiency of in vitro

transcription of the ε- and γ-globin genes is approximately the same as that of the β-globin gene. In contrast, it appears that the δ-globin gene is somewhat less efficiently transcribed. A quantitative analysis of this consistently observed difference is in progress. An analysis of the in vitro transcription of the human α-like globin genes is shown in Fig. 3C. The embryonic ζ1 and the adult α2 globin genes are transcribed with an efficiency comparable to that of the β-like genes. The pseudogene, ψα1, is also accurately transcribed, although at a lower efficiency.

We can make two conclusions on the basis of these results. (i) All of the human globin genes assayed appear to comprise individual transcription units. (ii) With the exception of the δ-globin gene, the embryonic, fetal, and adult genes are transcribed with roughly equal efficiencies in an extract prepared from HeLa cells, which do not ordinarily express globin genes. Thus, the interaction of different globin promoters with RNA polymerase II in vitro is approximately the same, and the mechanisms that mediate tissue-specific transcription do not operate in vitro.

Analysis of the 5′ end of β-globin RNA transcribed in vitro. The results of the experiments shown in Fig. 3, A to C, suggest that the 5′ ends of in vitro globin gene transcripts are near to their respective mRNA capping sites. To define these 5′ ends precisely, we used two different procedures. First, DNA fragments of 50 to 100 bp were isolated from the first exon of β-, ε-, and α-globin genes (Fig. 4B). These fragments were end-labeled, strand-separated, and used in primer extension experiments (see legend to Fig. 4). As shown in Fig. 4B, extension of the β-, ε-, and α-globin gene exon 1 primers with reverse transcriptase should produce DNA fragments of 115, 130, and 75 nucleotides, respectively, when an mRNA template is used. Fragments of exactly these sizes are observed when either mRNA or in vitro transcripts of the three globin genes are used as templates (Fig. 3A). We conclude, therefore, that the 5′ ends of the in vitro β-, ε-, and α-globin transcripts are coterminal with their mRNA capping sites. This conclusion was confirmed and extended by structural analysis of the 5′ end of the β-globin in vitro transcript.

The β-globin RNA synthesized in vitro contains the same 5′ cap structure as authentic β-globin mRNA. This was shown by analyzing the ribonuclease T1 oligonucleotides that bound to dihydroxylboryl cellulose (an affinity column that binds 3′ hydroxyl groups in RNA, in-

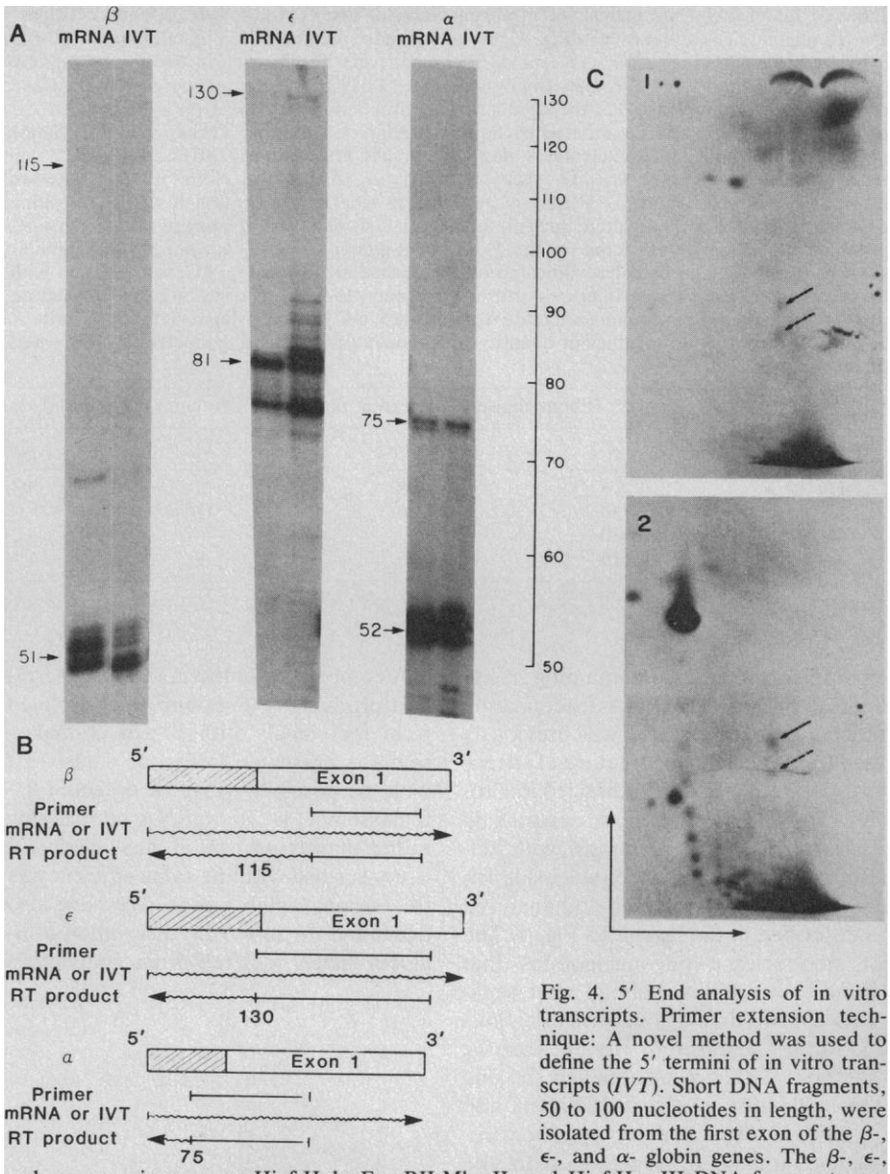

Fig. 4. 5′ End analysis of in vitro transcripts. Primer extension technique: A novel method was used to define the 5′ termini of in vitro transcripts (*IVT*). Short DNA fragments, 50 to 100 nucleotides in length, were isolated from the first exon of the β-, ε-, and α- globin genes. The β-, ε-, and α-gene primers were Hinf-Hph, Eco RII-Mbo II, and Hinf-Hae III DNA fragments, respectively. In each case the primer was end-labeled by partially filling the 5′ sticky end; the *E. coli* DNA polymerase I A fragment (*77*) and high specific activity (α^{32}P)-labeled nucleotide triphosphate were used. The end-labeled fragment strands were readily separated on 12 percent polyacrylamide 7*M* urea gels after being denatured by boiling in formamide since the two strands are of different lengths. The labeled antisense single strand DNA fragment was then annealed in 0.1*M* NaCl to either mRNA or unlabeled IVT RNA (100°C, 5 minutes; 60°C, 1 hour). The annealed mixture was then added directly to a reverse transcriptase (*RT*) reaction with all four unlabeled deoxyribonucleoside triphosphates (*78*). An equal volume of formamide was added to each reaction and, after denaturation by boiling, direct fractionation on polyacrylamide 7*M* urea gels was carried out. Maxam and Gilbert sequence ladders (*79*) were run on the same gel to determine precisely the sizes of primer extension products. (A) The primer extension data on the β-, ε-, and α-globin genes. The α and β mRNA's were obtained from purified human globin mRNA while ε mRNA was obtained from total poly(A) RNA of hemin-induced K562 cells (*80*). The colinear mRNA and IVT extension products are denoted by their sizes (deduced from the sequence ladder, which is not shown) and by arrows. The complex set of bands close to the original primer position (denoted by arrows and size) result from modification of excess unannealed primer by reverse transcriptase. The strong intermediate band in the β primer mRNA slot probably derives from premature termination due to secondary or tertiary structure in the mRNA template. (B) Line diagrams illustrating the different components involved in the β, ε, and α primer extension experiments. (C) Dihydroxylboryl cellulose-selected T1 oligonucleotides synthesized in vitro. The 100-μl reaction mixtures contain pBR-β, Pst 4.4 kbp DNA which had been digested with Bam HI as template, and 600 μCi of either [α-^{32}P]GTP (C1) or ATP (C2). After a 1-hour incubation (as above), RNA was purified, digested with ribnuclease T1, and selected on columns of dihydroxylboryl cellulose. Oligonucleotides were then purified and fractionated in two dimensions (*81*). The arrows at the bottom of the figure indicate the direction of electrophoresis (horizontal) and homochromatography (vertical). The solid arrow in each panel indicates the oligonucleotide that nearest neighbor analysis indicates has a sequence identical to the capped T1 oligonucleotide found in authentic human β-globin mRNA. The dashed arrow indicates an oligonucleotide that appears to contain the same sequence but may differ in its methylation pattern (see legend to Table 2).

Table 2. Nearest neighbor analysis of T1 oligonucleotides that bind to dihydroxylboryl cellulose. The T1 oligonucleotides shown in Fig. 4C as well as those obtained from reaction mixtures that contained [α-^{32}P]UTP or CTP as a labeled precursor were fractionated in two dimensions by electrophoresis and homochromatography (*81*). The T1 oligonucleotides were recovered and digested with ribonuclease T2; the products were fractionated by electrophoresis on DEAE paper at *p*H 3.5 (*81*) and quantitated by liquid scintillation counting. Three lines of evidence suggest that bound T1 oligonucleotides (denoted by solid arrows in Fig. 4C) contained a 5′ cap structure: (i) the presence of a T2-resistant moiety, X; (ii) 2 moles of [α-^{32}P]phosphate are incorporated into X when [α-^{32}P]ATP is used as the label, an indication that the A residue adjacent to the cap is methylated; and (iii) further analysis of X with the use of ribonuclease P1 provided the sequence expected for the 5′ end of β-globin mRNA (data not shown). Similar analysis of the oligonucleotides, indicated by the dashed arrows in Fig. 4C, is consistent with their containing the same primary structure. It appears, though, to contain a cap 2 structure (that is, the base of the second nucleotide from the cap, a C residue, is also methylated at its 2′ position). However, an insufficient quantity of radioactivity in this oligonucleotide prevented further analysis.

Labeled nucleotide	Ribonuclease T2 digestion products (^{32}P counts per minute)			
	X	A + G	C	U
U	5	48		98
C	56	60	12	
G	50			42
A	115	15	9	

cluding cap structures containing 3′ hydroxyl) by standard RNA fractionation procedures. This method was previously used to analyze the 5′ structure of adenovirus 2-specific RNA synthesized in vitro (*54*). RNA, labeled in four separate in vitro transcription reactions with different (α-^{32}P)-labeled ribonucleoside triphosphates, was extracted and analyzed as described in the legend to Fig. 4. The T1 ribonuclease oligonucleotides that bound to the affinity column were separated by two-dimensional fractionations and further analyzed by ribonuclease T2 digestion. The results obtained for one oligonucleotide are shown in Fig. 4C and in Table 2. The nearest neighbor frequencies obtained, together with the identification of a T2-resistant moiety, are consistent with this oligonucleotide having the same sequence and structure as the in vivo T1 oligonucleotide found at the 5′ terminus of human β-globin mRNA (*58*), GpppA$_{(m)}$CATTTG(C). Further evidence that this oligonucleotide contains a cap structure is described in the legend to Table 2.

In vitro transcription of abnormal globin genes. Low levels of mRNA have been associated with the mutant β-globin genes of patients with β-thalassemia. The most frequently occurring mutations in β-globin gene expression are called β^0- and β^+-thalassemia. The β^+-thalassemia is characterized by reduced levels of β-globin polypeptide production, while in β^0-thalassemia there is no detectable β-globin synthesis (*3*). Recent experiments indicate that in the cases studied thus far, the molecular defect in β^+-thalassemia is abnormal processing of the β-globin mRNA precursor (*26, 27*). In contrast, the molecular defects in β^0-thalassemia are quite heterogeneous (*59*).

A comparison of the in vitro transcription products of β-globin genes isolated from individuals with β^+- or β^0-thalassemia is presented in Fig. 5. Consistent with the possibility that the defect in β^+-thalassemia is in mRNA processing rather than transcription, a β^+ gene (*60*) is transcribed with the same efficiency as the normal β-globin gene. We have also examined the in vitro transcription of β-globin genes isolated from individuals with a type of β^0-thalassemia in which trace amounts of β-globin mRNA are produced. As shown in Fig. 5B, the efficiency of in vitro transcription of a recently isolated β^0-thalassemic gene of this type (*61*) appears to be the same as that of the normal gene. Thus, as in the case of the β^+-thalassemia gene, the in vitro assay does not reveal an obvious defect in transcription.

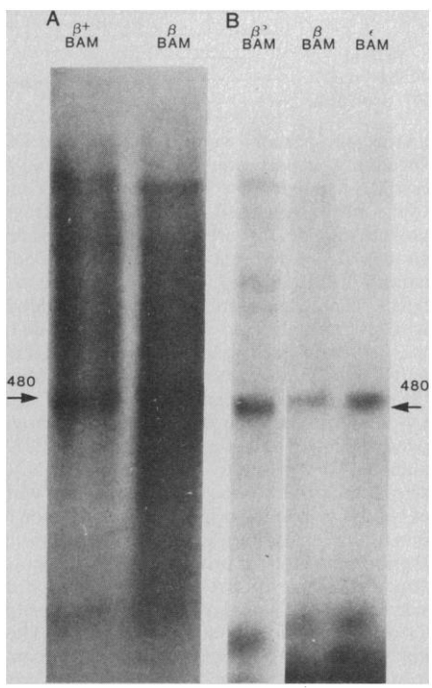

Fig. 5. In vitro transcription of β-thalassemic globin genes. In vitro transcripts (IVT) obtained from the β-globin genes of two patients, one (A) with β^+-thalassemia (*60*) and the other (B) with β^0-thalassemia (*61*). The β^+ IVT was run alongside a control β IVT. Similarly, the β^0 IVT was run alongside β and ϵ IVT controls.

Analysis of RNA Polymerase II Promoter Sites

In vitro transcription systems have been used in conjunction with molecular cloning and in vitro mutagenesis procedures as a means of defining RNA polymerase II initiation (promoter) sites. The most thoroughly studied example is the late promoter of adenovirus 2, the promoter for genes that are expressed late in adenovirus lytic growth. The DNA sequence corresponding to the 5′ end of the in vivo late gene transcript and the structure of the capped 5′ end of this transcript has been determined (*62*). In vitro transcription of a cloned DNA fragment containing the late promoter yields a transcript whose capped 5′ end is identical to that of the in vivo transcript (*53, 54*). An α-amanitin inhibition experiment demonstrated that this specific in vitro transcript is RNA polymerase II-dependent (*53, 54*).

The sequences necessary for specific in vitro transcription of late gene promoter fragments have been identified by in vitro mutagenesis experiments. As in the case of globin genes, the sequence ATA is located 31 nucleotides on the 5′ side of the mRNA capping site of the adeno 2 late promoter. This sequence was shown to be necessary and sufficient for synthesis of normal levels of specific transcripts by construction and in vitro analysis of deletions encompassing the region from −52 to +33 nucleotides from the mRNA capping site (*63*). More extensive deletion analysis demonstrated that the ATA box is necessary but not sufficient to support optimal levels of in vitro transcription (*64*). For example, deletion of the 5′ region to nucleotide −47 leads to a threefold reduction in promoter activity. Deletion of all sequences on the 3′ side of nucleotide +2 leads to a twofold stimulation of in vitro transcription. However, further deletions to nucleotide −2 result in an approximately threefold reduction in RNA synthesis. Thus, although the mRNA capping site is required for optimal levels of in vitro transcription, it does not appear to be necessary for the production of a normal transcript.

Corden *et al.* (*65*) report similar results on the basis of the analysis of the transcription products of deleted adenovirus late promoter DNA fragments; they used the in vitro transcription system described by Weil *et al.* (*53*). A similar analysis of the chicken conalbumin gene has shown that the ATA box is required for specific in vitro transcription (*56*, *65*).

In contrast to the examples cited above, the ATA box is not required for specific in vitro transcription of the SV40 late region. This region lacks the ATA sequence but is efficiently transcribed in vivo as well as in vitro (*63*). Specific transcripts are also detected when the SV40 early region, which does contain an ATA box, is examined in vitro.

Not all promoters that are known to be active in vivo are also active in vitro. For example, the adenovirus early region II promoter does not function in vitro even though transcripts originating from this site are detected in vivo. It is interesting to note that an ATA box is not located on the immediate 5′ side of the apparent transcription start point (*66*).

An ATA sequence does not appear to be necessary for the accurate transcription of cloned eukaryotic genes in the *Xenopus* oocyte injection system. Grosschedl and Birnsteil (*67*) reported the faithful synthesis of sea urchin histone mRNA in this system. The deletion of 5′ flanking sequence including the ATA box does not prevent mRNA synthesis. Similarly, Wickens *et al.* (*68*) have reported the synthesis of chicken ovalbumin protein in oocytes following the injection of either intact genes or a deleted gene fragment that lacks the 5′ flanking region and a part of the 5′ noncoding region and first intron.

Experiments to identify the nucleotide sequence required for accurate in vitro transcription of the human β-globin gene are in progress (*69*). Our preliminary results indicate that the CCAAT box is not necessary for normal β-globin gene in vitro transcription while a region of sequence including the ATA box appears to be essential. As shown in Fig. 3, we have reproducibly found that the δ-globin gene and ψα1-globin pseudogene are transcribed at lower efficiencies when compared to the β- and α-globin genes, respectively. Figure 6 compares the 5′ end regions of these four genes and illustrates the sequence differences found in δ and ψα1 at normally conserved positions (*2*). Clearly, any one of these differences could account for lower efficiency of in vitro transcription. Work is in progress to define which sequences in δ and ψα1 are responsible for the low levels of transcription.

In summary, the analysis of the structure and function of a number of normal and mutant ribonuclease II polymerase promoters have provided the first indications of the sequences essential for eukaryotic gene transcription. It is not surprising that the initial series of experiments revealed that different promoters behave differently when assayed by in vitro transcription. A more detailed analysis of mutagenized DNA templates as well as protein factors affecting transcription will be necessary to achieve a better understanding of eukaryotic promoters. In addition, it is essential to compare the behavior of normal and mutated promoter sequences in a variety of in vitro and in vivo assay systems.

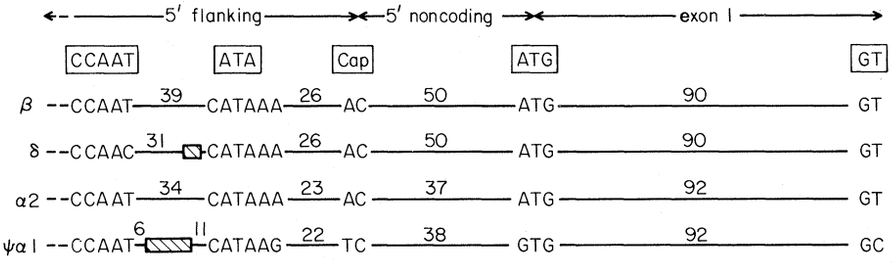

Fig. 6. Line diagrams comparing the 5′ ends of the β, δ, α2, and ψα1 genes. Open boxes indicate sequences found in all the expressed globin genes. The number of base pairs between homologous sequences is indicated. In order to align the 5′ flanking sequences of the δ and ψα1 genes with other globin genes, it was necessary to introduce the deletions in the region between the CCAAT and ATA box. These putative deletions are indicated by the crosshatched boxes.

Conclusion

Advances in the molecular biology of human globin genes in conjunction with the information provided by clinical investigations of inherited disorders in globin gene expression make it possible to study the molecular genetics of globin gene regulation. For example, cloned hybridization probes, containing specific regions of the α- or β-globin gene clusters have been used to map deletions that are associated with certain types of α- or β-thalassemias [see references in (*1*)]. The most interesting conclusion derived from these studies is that deletions, which alter the normal pattern of globin gene expression during development, map many kilobase pairs away from the genes that are affected. One interpretation of this observation is that globin gene clusters consist of one or more functional domains and that deletions within these domains alter the chromosome structure and thereby affect the normal pattern of differential globin gene expression (*1*, *10*, *70*). A correlation between gene activity and alterations in chromosome structure is suggested by the observation that actively transcribed genes are more sensitive to deoxyribonuclease I digestion than nontranscribed genes (*71*). Recently Stalder *et al.* (*72*) have shown that deoxyribonuclease I sensitivity extends for many kilobase pairs on either side of active chicken globin genes. This suggests that globin gene activation is associated with a structural alteration of a large region of chromosomal DNA.

Further development of in vitro and in vivo systems for studying globin gene expression, the isolation and structural characterization of abnormal globin genes, and the application of site-directed mutagenesis procedures to normal globin genes should make it possible to identify sequences involved in the regulation of human globin gene expression.

References and Notes

1. T. Maniatis, E. F. Fritsch, J. Lauer, R. M. Lawn, *Annu. Rev. Genet.*, in press; T. Maniatis, in *Cell Biology, A Comprehensive Treatise*, L. Goldstein and D. M. Prescott, Eds. (Academic Press, New York, 1980), vol. 3, p. 564.
2. A. Efstratiadis *et al.*, *Cell* **21**, 653 (1980).
3. D. J. Weatherall and J. B. Clegg, *ibid.* **16**, 467 (1979).
4. H. F. Bunn, B. G. Forget, H. M. Ranney, *Human Hemoglobins* (Saunders, Philadelphia, 1977).
5. E. F. Fritsch, R. M. Lawn, T. Maniatis, *Cell* **19**, 959 (1980).
6. J. Lauer, C. K. J. Shen, T. Maniatis, *ibid.* **20**, 119 (1980).
7. T. Maniatis, R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, A. Efstratiadis, *ibid.* **15**, 687 (1978).
8. R. M. Lawn, E. F. Fritsch, R. C. Parker, G. Blake, T. Maniatis, *ibid.*, p. 1157.
9. R. A. Flavell, J. M. Kooter, E. De Boer, P. F. R. Little, R. Williamson, *ibid.*, p. 25; J. G. Mears, F. Ramirez, D. Leibowitz, A. Bank, *ibid.*, p. 15; S. H. Orkin, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5950 (1978); P. F. R. Little, R. A. Flavell, J. M. Kooter, G. Annison, R. Williamson, *Nature (London)* **278**, 227 (1979); R. Bernards, P. F. R. Little, G. Annison, R. Williamson, R. A. Flavell, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4827 (1979); D. Tuan, P. A. Biro, J. K. DeRiel, H. Lazarus, B. G. Forget, *Nucleic Acids Res.* **6**, 2519 (1979).
10. E. F. Fritsch, R. M. Lawn, T. Maniatis, *Nature (London)* **279**, 598 (1979).
11. O. Smithies *et al.*, *Science* **202**, 1284 (1978); F. Ramirez, A. L. Burns, J. G. Mears, S. Spence, D. Starkman, A. Bank, *Nucleic Acids Res.* **7**, 1147 (1979).
12. N. J. Proudfoot and F. E. Baralle, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5435 (1979).
13. R. C. Hardison, E. T. Butler, E. Lacy, T. Maniatis, N. Rosenthal, A. Efstratiadis, *Cell* **18**, 1285 (1979).
14. E. Lacy, R. C. Hardison, D. Quon, T. Maniatis, *ibid.*, p. 1273.
15. C. L. Jahn, C. A. Hutchinson III, S. J. Phillips, S. Weaver, N. L. Haigwood, F. C. Voliva, M. H. Edgell, *ibid.* **21**, 159 (1980).
16. N. J. Proudfoot and T. Maniatis, *ibid.*, p. 537.
17. E. Lacy and T. Maniatis, *ibid.*, p. 545.

18. A. J. Jeffreys and R. A. Flavell, *ibid.* **12**, 1097 (1977).
19. A. Leder, H. I. Miller, D. H. Hamer, J. G. Seidman, B. Norman, M. Sullivan, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 6187 (1978).
20. S. M. Tilghman, D. C. Tiemeier, J. G. Seidman, B. M. Peterlin, M. Sullivan, J. Maizel, P. Leder, *ibid.*, p. 725.
21. J. Abelson, *Annu. Rev. Biochem.* **48**, 1035 (1979).
22. S. A. Liebhaber, M. Goosens, R. Poon, Y. W. Kan, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
23. Y. Nishioka and P. Leder, *Cell* **18**, 875 (1979).
24. S. M. Tilghman, P. J. Curtis, D. C. Tiemeier, P. Leder, C. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1309 (1978); A. J. Kinniburgh, J. E. Mertz, J. Ross, *Cell* **14**, 681 (1978); A. J. Kinniburgh and J. Ross, *ibid.* **17**, 915 (1979).
25. R. A. Flavell, R. Bernards, G. C. Grosveld, H. A. M. Hooijmakers-VanDommelen, J. M. Kooter, E. De Boer, in *Eukaryotic Gene Regulation*, R. Axel, T. Maniatis, C. F. Fox, Eds. (Academic Press, New York, 1980), pp. 335–354.
26. L. E. Maquat, A. J. Kinniburgh, L. R. Beach, G. R. Honig, J. Lazerson, W. B. Ershler, J. Ross, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4287 (1980).
27. J. A. Kantor, P. H. Turner, A. W. Nienhuis, *Cell* **21**, 149 (1980).
28. R. F. Weaver and C. Weissman, *Nucleic Acids Res.* **7**, 1175 (1979).
29. R. M. Lawn, A. Efstratiadis, C. O'Connell, T. Maniatis, *Cell* **21**, 647 (1980); R. Spritz, J. DeRiel, B. Forget, S. Weissman, *ibid.*, p. 639; F. E. Baralle, C. Shoulders, N. J. Proudfoot, *ibid.*, p. 621.
30. J. Slightom, A. Blechl, O. Smithies, *ibid.*, p. 627.
31. C. Benoist, K. O'Hare, R. Breathnach, P. Chambon, *Nucleic Acids Res.* **8**, 127 (1980).
32. M. Goldberg, thesis, Stanford University (1979).
33. D. A. Konkel, J. V. Maizel, P. Leder, *Cell* **18**, 865 (1979).
34. A. van Ooyen, J. van Den Berg, N. Mantei, C. Weissman, *Science* **206**, 337 (1979).
35. P. J. Farabaugh and J. H. Miller, *J. Mol. Biol.* **126**, 847 (1978).
36. E. A. Zimmer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2158 (1980).
37. M. O. Dayhoff, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1972).
38. L. Hood, J. H. Campbell, S. C. R. Elgin, *Annu. Rev. Genet.* **9**, 305 (1975).
39. S. Orkin, J. Old, H. Lazarus, C. Altay, A. Gurgey, D. Weatherall, D. Nathan, *Cell* **17**, 33 (1979); S. Embury, R. Lebo, A. Dozy, Y. W. Kan, *J. Clin. Invest.* **63**, 1307 (1979).
40. M. Goossens, A. Dozy, S. Embury, Z. Zachariades, M. Hadjiminas, G. Stamatoyannopoulos, Y. W. Kan, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 518 (1980); D. R. Higgs, J. M. Old, L. Pressley, J. B. Clegg, D. J. Weatherall, *Nature (London)* **284**, 632 (1980).
41. Y. Nishioka, A. Leder, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2806 (1980); E. F Vanin, G. I. Goldberg, P. W. Tucker, O. Smithies, *Nature (London)*, in press.
42. R. Breathnach, C. Benoist, K. O'Hare, F. Gannon, P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4853 (1978); M. Lerner, J. Boyle, S. Mount, S. Wolin, J. Steitz, *Nature (London)* **283**, 220 (1980).
43. C.-K. J. Shen, unpublished observations.
44. W. G. Jelinek *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1398 (1980); F. E. Baralle and N. J. Proudfoot, unpublished observations; O. Smithies, unpublished observations.
45. C. M. Houck, F. P. Rinehart, C. W. Schmid, *J. Mol. Biol.* **132**, 289 (1979).
46. C. Duncan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5095 (1979).
47. E. F. Fritsch, C. K. J. Shen, R. M. Lawn, T. Maniatis, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
48. W. G. Jelinek and L. Leinwood, *Cell* **15**, 205 (1978).
49. H. D. Robertson, E. Dickson, W. Jelinek, *J. Mol. Biol.* **115**, 571 (1977).
50. C.-K. J. Shen and T. Maniatis, *Cell* **19**, 379 (1980).
51. A. Pellicer *et al.*, *Science* **209**, 1414 (1980); N. Mantee, W. Boll, C. Weissmann, *Nature (London)* **281**, 40 (1979); M. Wigler *et al.*, *Cell* **16**, 777 (1979); B. Wold, M. Wigler, E. Lacy, T. Maniatis, S. Silverstein, B. Axel, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5684 (1979).
52. D. H. Hamer and P. Leder, *Nature (London)* **281**, 35 (1979); H. Mulligan and P. Berg, *Science* **209**, 1422 (1980).
53. P. A. Weil, D. S. Luse, J. Segall, R. G. Roeder, *Cell* **18**, 469 (1979).
54. J. L. Manley, A. Fire, A. Cano, P. A. Sharp, M. L. Gefter, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3855 (1980).
55. C. Parker, personal communication.
56. B. Waslyk, C. Kedinger, J. Corden, O. Brison, P. Chambon, *Nature (London)* **285**, 367 (1980).
57. D. S. Luse and R. G. Roeder, *Cell* **20**, 691 (1980).
58. F. E. Baralle, *ibid.* **12**, 1085 (1977).
59. J. M. Old, N. J. Proudfoot, W. G. Wood, J. I. Longley, J. B. Clegg, D. J. Weatherall, *ibid.* **14**, 289 (1978).
60. L. Burns and A. Bank, unpublished observations.
61. M. H. M. Shander, unpublished observations.
62. E. B. Ziff and R. M. Evans, *Cell* **15**, 1436 (1978).
63. J. L. Manley *et al.*, in preparation.
64. S. L. Hu and J. L. Manley, in preparation.
65. J. Corden *et al.*, *Science* **209**, 1406 (1980).
66. C. C. Baker, J. Herisse, G. Courtois, F. Galibert, E. B. Ziff, *Cell* **18**, 569 (1979).
67. R. Grosschedl and M. L. Birnsteil, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1432 (1980).
68. M. Wickens, S. Woo, B. O'Malley, J. Gurdon, *Nature (London)* **285**, 628 (1980).
69. V. Parker, N. J. Proudfoot, M. H. M. Shander, T. Maniatis, unpublished observations.
70. R. Bernards, J. M. Kooter, R. A. Flavell, *Gene* **6**, 265 (1979); L. H. T. van der Ploeg, A. Konings, M. Oort, D. Roos, L. Bernini, R. Flavell, *Nature (London)* **283**, 637 (1980).
71. H. Weintraub and M. Groudine, *Science* **193**, 848 (1976).
72. J. Stalder, A. Larsen, J. D. Engel, M. Dolan, M. Groudine, H. Weintraub, *Cell* **20**, 451 (1980).
73. S. L. Berger and C. S. Birkenmeier, *Biochemistry* **18**, 5143 (1979).
74. J. M. Bailey and N. Davidson, *Anal. Biochem.* **70**, 75 (1976).
75. J. G. Sutcliffe, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77 (1979).
76. R. Laskey, in *Methods Enzymol.*, **65**, 363 (1980).
77. H. Klenow and K. Overgaard-Hansen, *FEBS Lett.* **6**, 25 (1970).
78. N. J. Proudfoot, *Cell* **10**, 559 (1977).
79. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
80. T. R. Rutherford, J. B. Clegg, D. J. Weatherall, *Nature (London)* **280**, 164 (1979).
81. G. G. Brownlee, *Determination of Sequences in RNA: Laboratory Techniques in Biochemistry and Molecular Biology*, T. S. Work and E. Work, Eds. (North-Holland/American Elsevier, Amsterdam, 1972).
82. The authors are grateful to C. O'Connell, who was involved in the preliminary stages of this work, and to G. Attardi for providing HeLa cells. Supported by NIH grants.

21 July 1980

Mouse Globin System: A Functional and Evolutionary Analysis

Philip Leder, J. Norman Hansen, David Konkel, Aya Leder, Yutaka Nishioka, Carol Talkington

The globin genes have a special place among the systems first examined by means of recombinant technology (*1*). This is so because more than 50 years of intense study had created an array of questions that could now be dealt with at the molecular genetic level and also because nature had conveniently arranged for the development of the red blood cell to occur in such a way as to make globin messenger RNA (mRNA) abundantly available. The combination of interesting genetic phenomena and the availability of probes (globin mRNA) made the globins an early and promising target for gene cloning. Proudfoot *et al.* (*2*) have reviewed the progress made in understanding human globin genes. We describe what has been learned from studying the mouse.

The Mouse Globin Gene System

Globin gene expression in the mouse begins during intrauterine development with the appearance of a primitive population of nucleated red blood cells in the embryonic yolk sac. In contrast to adult mouse erythrocytes, these cells produce three embryonic globins, one α-like (X) and two β-like (Y and Z), whose expression is limited to this special cell population and this specific period of development (*3*). Adult α-globin also appears in yolk-sac red cells, but continues to be produced in nonnucleated adult erythrocytes, where it is accompanied by the appearance of adult β-globin. The two adult globins are produced in relatively equivalent amounts from the third and final week of gestation throughout the lifetime of the organism (*4*).

Genetic and structural studies had indicated that the BALB/c mouse ex-

The authors are investigators at the Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20205.

0036-8075/80/0919-1336$01.50/0 Copyright © 1980 AAAS