MOLECULAR AND CELLULAR BIOLOGY, Jan. 1985, p. 147–160
0270-7306/85/010147-14$02.00/0

Vol. 5, No. 1

Transcription Unit of the Rabbit β1 Globin Gene

MARK L. ROHRBAUGH,† JOSEPH E. JOHNSON III, MAGDALENA D. JAMES, AND ROSS C. HARDISON*

Department of Biochemistry, Microbiology, and Molecular and Cell Biology, The Pennsylvania State University, University Park, Pennsylvania 16802

Received 7 September 1984/Accepted 15 October 1984

We have hybridized pulse-labeled nuclear transcripts to cloned DNA fragments from the rabbit β-like globin genes to determine the developmental timing, extent, and asymmetry of their transcription. The fetal-adult gene β1 was transcribed in fetal liver but not embryonic nuclei, whereas genes β3 and β4, which encode embryonic globin polypeptides, were transcribed only in embryonic nuclei. This shows that the switch from embryonic to fetal-adult globin production in rabbits is accomplished primarily by differential transcription of the β-like globin genes. Gene β1 was subdivided into M13 subclones and tested for hybridization to nascent RNA. The nucleotide sequence of the 3′ flanking region of gene β1 was also determined for 2,447 base pairs past the polyadenylation [poly(A)] site. No transcripts were found 5′ to the cap site, but asymmetric transcription of gene β1 proceeded at a high level through the gene and past the poly(A) addition site for 603 nucleotides. The level of transcription declined after this, gradually dropping through the next 568 nucleotides. No polymerases were found on a fragment that begins 1,707 nucleotides past the poly(A) site; this fragment was part of a segment of repetitive DNA. These data show that the transcription unit of gene β1 begins at or near the cap nucleotide and extends at least 1,171 but no more than 1,706 nucleotides past the poly(A) addition site. The DNA segment that precedes the region of declining transcription contained an inverted repeat and encoded a short RNA transcribed by RNA polymerase II from the strand opposite the β1 transcript. These two features may function to attenuate the transcription of gene β1. An inverted repeat and a potential polymerase II transcription unit were also found in the homologous segment 3′ to the human β-globin gene. A short DNA segment close to the 3′ end of the β1 transcription unit was transcribed more actively than the surrounding DNA, and it contained sequences that match the consensus internal control region for RNA polymerase III. This DNA segment may contain a separate polymerase III transcription unit. A member of the D repeat family located 3′ to gene β1 was not transcribed in its entirety coordinately with β1.

The rabbit β-like globin gene family is a set of structurally related genes that are differentially expressed during development (22, 32). Gene β3 and β4 are expressed in embryonic life, gene β1 is expressed in fetal and adult life (22, 50), and gene ψβ2 is an unexpressed pseudogene (33). The coordinate change in the amounts of polypeptides, mature mRNAs, and pre-mRNAs from each active gene suggests that the switch in expression from embryonic to fetal-adult β-like globin genes is accomplished primarily by differential transcription of the genes (50). Transcriptional regulation of rabbit β-globin gene expression is demonstrated more directly in this paper. Experiments designed to study possible mechanisms for this regulation have shown no observable difference in the extent of methylation of CCGG sites in the vicinity of gene β1 in embryonic and adult erythropoietic tissues (56), whereas the transcriptional activation of gene β1 is accompanied by the appearance of a DNase I-hypersensitive, S1-sensitive site 5′ to the gene (J. Margot and R. Hardison, unpublished data).

In studying the regulated transcription of the rabbit β-like globin genes, it is important to identify the transcription units for the genes, that is, what segments of DNA are transcribed into the primary transcript. By studying the abundant precursor RNAs from the fetal-adult gene β1, Grosveld et al. (19) showed that the 5′ and 3′ ends of the precursor RNAs were identical to the ends of the mature mRNA, and therefore no RNA from the flanking regions was present in the stable precursors. These studies did not address possible unstable transcripts from the flanking region. Transcription of gene β1 by RNA polymerase II in cell-free extracts begins at the capped nucleotide (20). One can infer from this that RNA polymerase II does not transcribe the 5′ flanking region at a high level in vivo.

Analysis of nascent transcripts allows one to examine the process of transcription in nuclei without the complications of RNA processing or turnover. Previously initiated transcription complexes in isolated nuclei can continue to synthesize RNA under appropriate conditions. By incubating the nuclei with a pulse of radiolabeled UTP, the newly synthesized RNA will be radioactively labeled before significant processing of the RNA occurs (21, 41). Hybridization of the labeled nascent RNA to purified, cloned DNA fragments can then be used to define the transcription unit of a particular gene (28). Experiments following this procedure have shown that transcription of the mouse adult β^{maj} gene continues past the polyadenylic acid [poly(A)] addition site for at least 1,200 nucleotides (28, 29). Thus the transcription unit is larger than the region that encodes mRNA, with the excess transcription extending into the 3′ flanking region. The exact limits of transcription of mouse β^{maj} were not determined in these earlier studies, however, and an initial report of precise termination (52) has been retracted (53).

In this paper, we present a map of nascent transcripts in the region around rabbit globin gene β1, along with the nucleotide sequence extending to 2,447 nucleotides past the poly(A) site. Transcription of β1 in fetal liver nuclei extends past the poly(A) site for about 600 nucleotides, at which point the level of transcription begins to decline. A transcriptionally silent region occurs about 1,700 nucleotides past the poly(A) site within a region of repetitive DNA. An inverted

* Corresponding author.
† Present address: Department of Botany, University of Minnesota, St. Paul, MN 55108.

repeat and a short, polymerase II-transcribed RNA opposing the β1 transcript are found just before the region where β1 transcription begins to decline. These structures may be important in attenuating transcription of β-globin genes.

MATERIALS AND METHODS

Enzymes. Restriction endonucleases, T4 DNA ligase, *Escherichia coli* DNA polymerase I (large fragment), and bacterial alkaline phosphatase were purchased from New England Biolabs or Bethesda Research Laboratories. Proteinase K and pancreatic DNase I were purchased from Sigma Chemical Co.

Plasmid DNAs. The isolation of recombinant λ clones containing genes β1, β3, and β4 has been previously described (32, 37). The present study used a 5.6-kilobase pair *Pst*I subclone of gene β1 that contains a considerable amount of 3′ flanking sequences; the large intervening sequence subclones of genes β3 and β4, pEco Bam 0.88-β3 and pEco Bam 0.88-β4, respectively (32); clone pEco 1.65, which contains a member of the C repeat family located between genes β3 and ψβ2; and clone pEco 1.85, which contains a member of the D repeat family located 3′ to gene β1 (57).

Subclones of the β1 region in M13. DNA fragments within and flanking gene β1 were subcloned into the M13 bacteriophage vectors mp8 and mp9 (42–44). Either gel-purified single fragments or a mixture of restriction fragments from either pPst 5.6-β1 or pEco 1.85 were ligated with the replicative form of mp8 or mp9 DNA (cleaved with the appropriate restriction endonuclease) and transfected into competent (12) *E. coli* JM103 (42). DNA samples from recombinant phage (colorless or light blue on 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) were identified by the size of the inserts after excision from the replicative form with restriction endonucleases, by the ability of the single-stranded phage DNA to anneal to clones containing the complementary strand, and by the nucleotide sequence as determined by the dideoxynucleotide method (25, 55). All short DNA clones were sequenced through the insert into the M13 vector DNA, and only those clones containing single inserts were used in the hybridization assays. A map of the cloned DNA fragments is given in Fig. 1, and further details about constructing the clones are described by M. L. Rohrbaugh (Ph.D. thesis, The Pennsylvania State University, University Park, 1984) and in the New England Biolabs M13 manual.

Isolation of nuclei. Nuclei were prepared according to Groudine et al. (21). The uterus was dissected from a pregnant New Zealand white rabbit (anesthetized with Rompun and Ketaset) and submerged in 1× SSC solution (SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0) at 4°C. Embryos (12 days of gestation) or fetal livers (18 days of gestation) were disaggregated in 5 to 8 ml of 1× SSC solution, and cells were pelleted at 160 × *g* in an HB-4 rotor (Sorvall RC-5B centrifuge) for 5 min. Cells were lysed in ~5 ml of reticulocyte standard buffer (0.01 M Tris [pH 7.4], 0.01 M NaCl, 0.003 M $MgCl_2$) containing 0.5% Nonidet P-40, and nuclei were pelleted at 160 × *g* (HB-4 rotor), gently washed again, and suspended at a DNA concentration of 0.7 to 1.5 mg/ml in a solution of 40% glycerol, 50 mM Tris [pH 8.3], 5 mM $MgCl_2$, and 0.1 mM EDTA. This preparation was used immediately or stored at −70°C for up to 5 months.

Nuclear transcription and RNA isolation. Transcription in isolated nuclei and subsequent RNA isolation were performed with modifications of procedures described by Groudine et al. (21) and McKnight and Palmiter (41). The transcription reaction (0.2 ml) contained 0.125 ml of nucleus preparation (88 to 188 μg of DNA), 20 μl of 10× transcription buffer (1× transcription buffer is 2.5 mM dithiothreitol, 1 mM $MgCl_2$, 70 mM KCl, 0.25 mM CTP, 0.25 mM GTP, 0.5 mM ATP), 200 pmol of unlabeled UTP, and 200 pmol (~400 μCi) of [α-^{32}P]UTP (~2,000 Ci/mmol; ICN Pharmaceuticals). Nuclei were incubated at 26°C for 20 min and the reaction was quenched with DNase I (purified over a UMP-agarose column [3, 67]) in the presence of 10 mM vanadyl-ribonucleoside complex for 10 min. The reaction mixture was deproteinized at 42°C for 30 min at final concentrations of 1% sodium dodecyl sulfate–5 mM EDTA–10 mM Tris (pH 7.4)–10 mM vanadyl-ribonucleoside complex and 0.1 mg of proteinase K per ml. After several extractions with an equal volume of Tris-saturated phenol/sevag (chloroform-isoamyl alcohol, 24:1), tRNA was added to 0.1 mg/ml and the nucleic acids were precipitated in 5% trichloroacetic acid–30 mM sodium pyrophosphate at 4°C for 30 min. The pellet was collected by centrifugation in a microfuge for 5 min and washed three times with cold trichloroacetic acid–5% sodium pyrophosphate solution. RNA was purified by redissolving the pellet in 0.3 ml of DNase buffer [20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5, 5 mM $MgCl_2$, 1 mM $CaCl_2$] and treating with 20 μg of DNase I per ml–10 mM vanadyl-ribonucleoside complex at 37°C for 30 min. Further treatment with proteinase was performed at final concentrations of 15 mM EDTA, 1% sodium dodecyl sulfate, 10 mM vanadyl-ribonucleoside complex, and 25 μg of proteinase K per ml at 37°C for 30 min. After several phenol/sevag extractions, the aqueous phase was precipitated with 0.25 M NaOAc and 2.5 volumes of ethanol.

Dot blot hybridizations. Dot blots were prepared according to Kafatos et al. (31) with the modifications given by Schleicher & Schuell. Either M13 single-stranded or plasmid (4 μg) DNA was treated with 0.3 ml of 57 mM Tris (pH 7.4)–0.2 N NaOH–6.7× SSC at 80°C for 10 min. The samples were placed on ice and neutralized with 60 μl of 2 M HEPES (pH 7.5). A Schleicher & Schuell filter manifold was used to spot denatured DNA onto nitrocellulose (Schleicher & Schuell) presoaked in 10× SSC. Nitrocellulose was removed from the aparatus, washed in 6.7× SSC, air dried, and baked in vacuo at 80°C for 3 h.

Hybridization conditions are described by Wahl et al. (64) with the modifications described by Lacy et al. (32). Labeled nuclear RNA (10^6 cpm) was redissolved in 50 μl of TE buffer (10 mM Tris, pH 7.5, 1 mM EDTA), denatured at 60°C for 5 min, and added to 2 ml of preheated hybridization solution containing no dextran sulfate. After 3 to 4 days of hybridization at 42°C, the blots were washed for 20 min at 65°C twice each in successive solutions of 2× SSC, 1× SSC, and 0.3× SSC, each of which contained 0.1% sodium pyrophosphate and 0.1% sodium dodecyl sulfate. The washed blots were exposed to Kodak XAR-5 film with an intensifying screen at −70°C for 2 to 7 days. Some blots were treated with RNase A (58) after hybridization and washing; the resulting autoradiograms were identical to those obtained without RNase treatment.

To probe for repetitive DNA, 4 μg of the recombinant M13 single-stranded DNAs was spotted onto nitrocellulose, baked, hybridized in 10 ml of hybridization solution (32) with 3 ng of nick-translated (38) rabbit liver genomic DNA (specific activity, 10^8 dpm/μg) per ml for 16 h at 42°C, washed at a final 1× SSC concentration, and exposed to X-ray film for 2 days.

DNA sequence determination. Plasmid DNA fragments were end labeled and sequenced by the base-specific chemical cleavage method (39, 40). Recombinant M13 clones were sequenced by the dideoxynucleotide chain termination method (25, 55) with modifications described by Rohrbaugh (Ph.D. thesis, 1984).

Dot plot sequence comparisons. The nucleotide sequences of the 3′ flanking regions of the rabbit β1 and human β-globin genes were compared graphically, using a version of the program MATRIX (70) developed to run on an IBM XT microcomputer with an Epson FX-80 printer. Similar results were obtained by using R. Britten's program DOT run on an Apple IIE microcomputer.

RESULTS

Transcription map of the β1 gene region. To examine the extent and asymmetry of transcription in and around the β1 globin gene, DNA fragments were subcloned into the single-stranded phage vectors M13 mp8 and mp9 (44). Most fragments were cloned in both orientations so that transcription from each strand could be measured. The cloned DNA was hybridized to labeled nascent RNA from fetal liver nuclei in four separate experiments and to labeled RNA from embryo nuclei in another experiment. Figure 1 shows the results of the hybridization, along with a map of the DNA segments used in the analysis.

Transcription of gene β1 continues past the poly(A) addition site in fetal liver nuclei, but no detectable transcription occurs 5′ to the cap site (Fig. 1, experiments I to IV). Neither strand of fragment A, which ends at a *Pvu*II site 10 base pairs (bp) before the cap nucleotide, hybridizes to the labeled RNA. This result agrees with the observation that initiation in vitro occurs primarily at the cap nucleotide, but not upstream (20). Gene β1 and the 3′ flanking portion are transcribed in fetal liver nuclei but not in embryo nuclei. The mRNA-complementary (antisense) strand of fragment B (from the second intron), fragment C (3′ untranslated and 3′

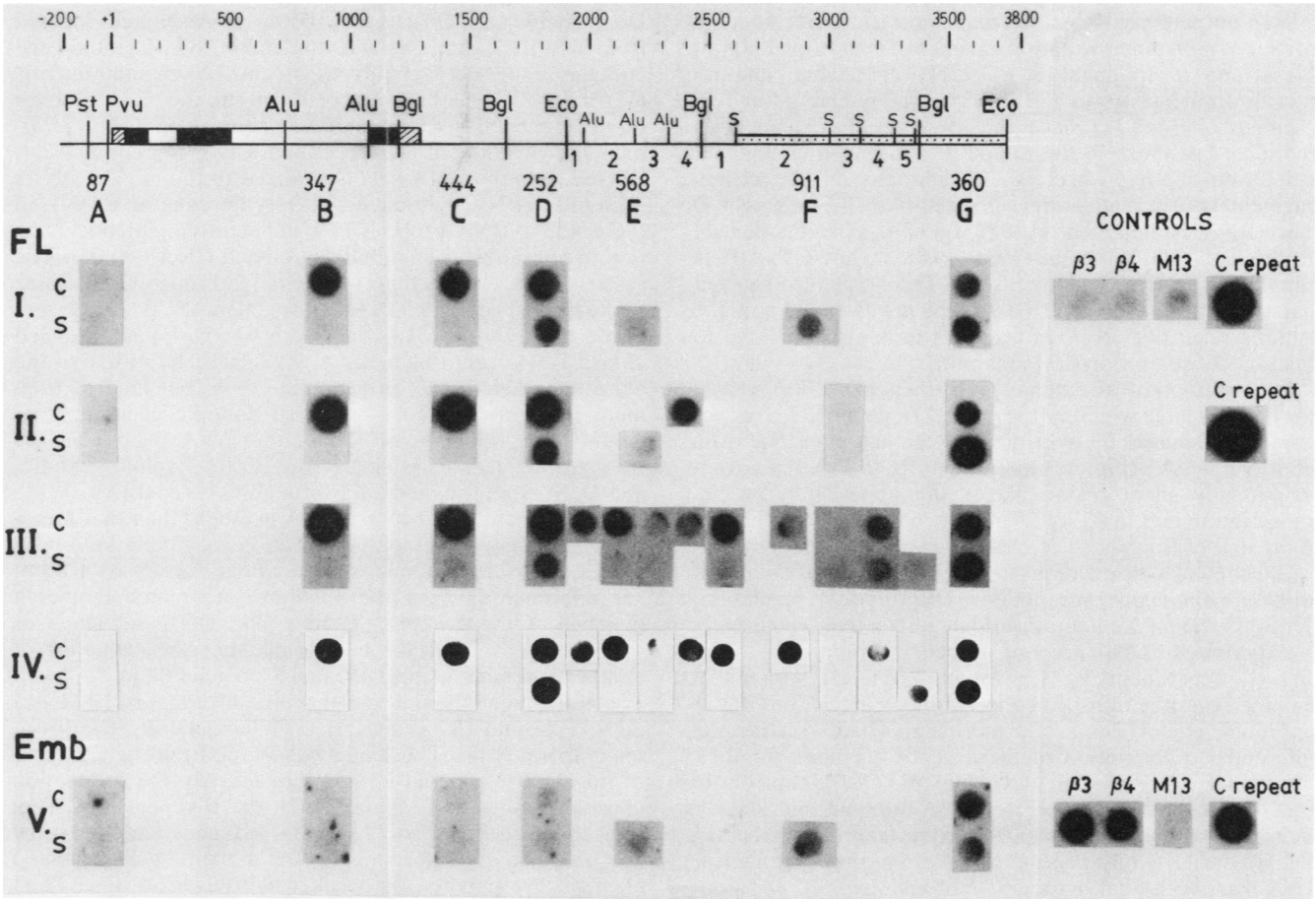


FIG. 1. Transcription analysis of the β1 gene locus. The upper portion shows a restriction map of the region around the β1 gene. The β1 gene is shown as a box with intervening sequences unshaded, polypeptide-coding regions shaded, and untranslated regions hatched. The region containing repetitive DNA is indicated by the stippled box at the right end of the map. The scale of distance in base pairs is shown above the map; the cap nucleotide is +1. Restriction site abbreviations are as follows: Pst, *Pst*I; Pvu, *Pvu*II; Alu, *Alu*I; Eco, *Eco*RI; Bgl, *Bgl*II; S, *Sau*3AI. Sizes of fragments A through G are indicated below the map. Fragments E and F were subdivided into four *Alu*I fragments and five *Sau*3AI fragments, respectively. The lower portion shows the results of hybridizing labeled nascent transcripts to the cloned DNAs. Fragments A through F were subcloned into M13 mp8 or mp9 in either orientation such that the strand complementary to (c) or synonymous with (s) the β1 mRNA is present. Single-stranded DNA was spotted onto dot blots and probed with RNA labeled in isolated nuclei. Dots containing a particular subclone are shown below the position of the fragment on the restriction map. Data from four separate experiments with labeled RNA from fetal liver (FL) nuclei (I, II, III, and IV) are shown. A duplicate of the dot blot shown in experiment I was probed with labeled RNA from embryo nuclei (V). The Es dot in experiments I, II, and V and the Fs dot in experiments I and V contained DNA from the mRNA-synonymous strand of the entire E fragment and the entire F fragment, respectively. Controls included plasmids containing the large intervening sequence of gene β3 or β4, a C family repeat (pEco 1.65), and double-stranded M13 mp8 as a negative control.

flanking region), and fragment D [3′ flanking region, 352 to 604 bp past the poly(A) site] hybridize to fetal liver nuclear transcripts at equivalent levels (Fig. 1, experiments I to IV), but they do not hybridize to embryonic nuclear transcripts (Fig. 1, experiment V). This transcription is asymmetric for fragments B and C as shown by the absence of hybridization to the mRNA-synonymous (sense) strands. The mRNA-synonymous strand of fragment D, however, does hybridize to fetal liver nuclear transcripts, but at a lower level than the mRNA complementary strand. As shown in experiments II, III, and IV (Fig. 1), the labeled nascent transcripts from fetal liver hybridize to the mRNA-complementary strand of fragments E [604 to 1,172 bp past poly(A)], part of F (F1, F2, and F4), and G. These hybridization signals are weaker than those seen for fragments B, C, and D except for F1c, which produces a strong hybridization signal. As will be shown below, fragments A through F1 are present as single copies in the rabbit genome, but fragments F2 through G contain repetitive DNA. One cannot determine the actual source of the RNA hybridizing to fragments of repetitive DNA by this assay.

RNA polymerase II synthesizes a short transcript from the mRNA-synonymous strand of fragment D. Hybridization to this strand of fragment E is barely detectable, and no hybridization occurs to the mRNA-synonymous strand of fragment C (Fig. 1). Since the sequences in fragment D hybridize but those in the 5′ and 3′ adjacent fragments (E and C) do not hybridize, the transcript is no larger than fragment D (252 nucleotides). Synthesis of the fragment D short RNA is inhibited by 0.5 μg of α-amanitin per ml (Rohrbaugh, Ph.D. thesis, 1984), which shows that it is transcribed by RNA polymerase II. This study also showed that transcription of gene β1 (fragments B to E) and the flanking repetitive DNA in fragment G is also sensitive to this low concentration of α-amanitin.

Transcription of β1 extends no further than 1,707 bp past the poly(A) site, which is the end of fragment F2. Neither strand of fragment F3 hybridizes to the nascent RNA from fetal liver nuclei (Fig. 1, experiments II to IV). This transcriptionally silent region places the upper limit on the transcription unit of β1 as 2,994 bp (from the cap nucleotide to the end of fragment F2).

The control hybridizations in Fig. 1 confirm that the rabbit β-like globin genes are under transcriptional regulation. Genes β3 and β4 are not transcribed in fetal liver nuclei (Fig. 1, experiment I) but are transcribed in embryonic nuclei (Fig. 1, experiment V). Conversely, gene β1 with its 3′ flanking region is transcribed in fetal liver nuclei but not in embryonic nuclei (Fig. 1, cf. experiments I and V). Thus, the differential transcription of these genes accounts for their differential expression in development. Other controls in Fig. 1 show that the M13 vector DNA does not hybridize to the nuclear transcripts, and DNA from a short, interspersed repeat, the C family repeat (6, 57), hybridizes to nuclear RNA from both fetal liver and embryos.

Attenuation of transcription in the 3′ flanking region. The internal and 3′ flanking portions of gene β1 are transcribed at an equivalent level through fragment D, but the level of transcription declines through fragment E. This is shown most clearly in Fig. 2, which is a plot of the hybridization results from Fig. 1 after they were quantitated by microdensitometry and adjusted to hybridization units per base pair. The mRNA-complementary strands of fragments B, C, and D show a high level of hybridization to nuclear RNA, although hybridization to fragment C was somewhat lower in two experiments. After fragment D, the level of hybridization declines steadily through fragment E and eventually is undetectable in fragment F3. These results show that virtually all of the RNA polymerases that initiate at the cap site of β1 proceed on through fragment D, but fewer of the polymerases continue on through fragment E, and essentially none of them reach fragment F3 [1,707 bp past poly(A)]. Thus, the transcription of gene β1 attenuates about 600 bp past the poly(A) addition site.

The data in Fig. 2 have two unexpected features. Although transcription of the β1 region is largely asymmetric, a short transcript confined to fragment D is synthesized from the mRNA-synonymous strand. A second unusual feature is that the 133-bp fragment F1c is transcribed at a much higher level than the surrounding DNA. This contrasts with the pattern of steadily declining transcription from E1 through F2 and suggests that fragment F1 may contain a separate transcription unit. The hybridization to the repetitive DNA in fragments F and G is analyzed below.

Repetitive DNA in the 3′ flanking region. To interpret the nascent RNA hybridization data, one must know the repetition frequency of the DNA fragments. Hoeijmakers-van Dommelen et al. (27) showed that a DNA segment located from 0.9 to 2.1 kilobases 3′ to the end of the β1 globin gene contains moderately repetitive DNA, and Shen and Maniatis (57) mapped a D family repeat within the 1.85-kilobase *Eco*RI fragment composed of fragments E, F, and G. To map the position of the repetitive DNA more precisely, labeled genomic DNA was hybridized to the cloned DNAs used in the RNA hybridization. Only the repetitive DNA in the genomic DNA probe is at a sufficiently high concentration to hybridize to immobilized, cloned DNA (16, 57). The results in Fig. 3 show that fragments F2 through G hybridize to the labeled genomic DNA, but fragments B, C, D, E2, and F1 do not hybridize. In confirmatory experiments, labeled cloned DNA from fragments A through F1 hybridize to the expected single bands in genomic DNA, but labeled fragment F2 hybridizes to a smear of multiple genomic fragments (data not shown). Thus, the DNA from gene β1 through fragment F1 is single copy in the haploid genome, and DNA from fragments F2 through G is repetitive.

This segment of repetitive DNA is longer than that measured previously from analysis of D repeat family heteroduplexes in the electron microscope (about 935 versus 420 bp; see reference 57), but the position does match that of a member of the D-repeat family. The DNA sequence of fragments F and G (see below) matches the sequence of another D-repeat member located 5′ to gene β4 in a 3.4-kbp *Eco*RI fragment and the sequence of a D repeat isolated as a cDNA plasmid (R. Printz, J. Johnson, and R. Hardison, unpublished data). The sequences match from the right end of fragment G through F3 and into F2. (By the orientation chosen by Shen and Maniatis [57], the D repeat runs from right to left in Fig. 1 and 2.) These sequence matches show that fragments F2 through G are a continuous repeated element. Presumably, the weaker hybridization signal from labeled genomic DNA to fragments F3 and F4 in Fig. 3 reflects the absence of F3 and F4 DNA from some other members of the D repeat family.

Most of the DNA fragments comprising the D family repeat are transcribed, usually from both strands (Fig. 2). Only fragment F3 reproducibly shows no detectable hybridization from either strand. (Since no subclones containing only fragment F2s or F5c were obtained, these were not included in the analysis.) The most active transcription is from fragment Gs, and this transcription declines to undetectable levels by fragment F3s. These results represent the

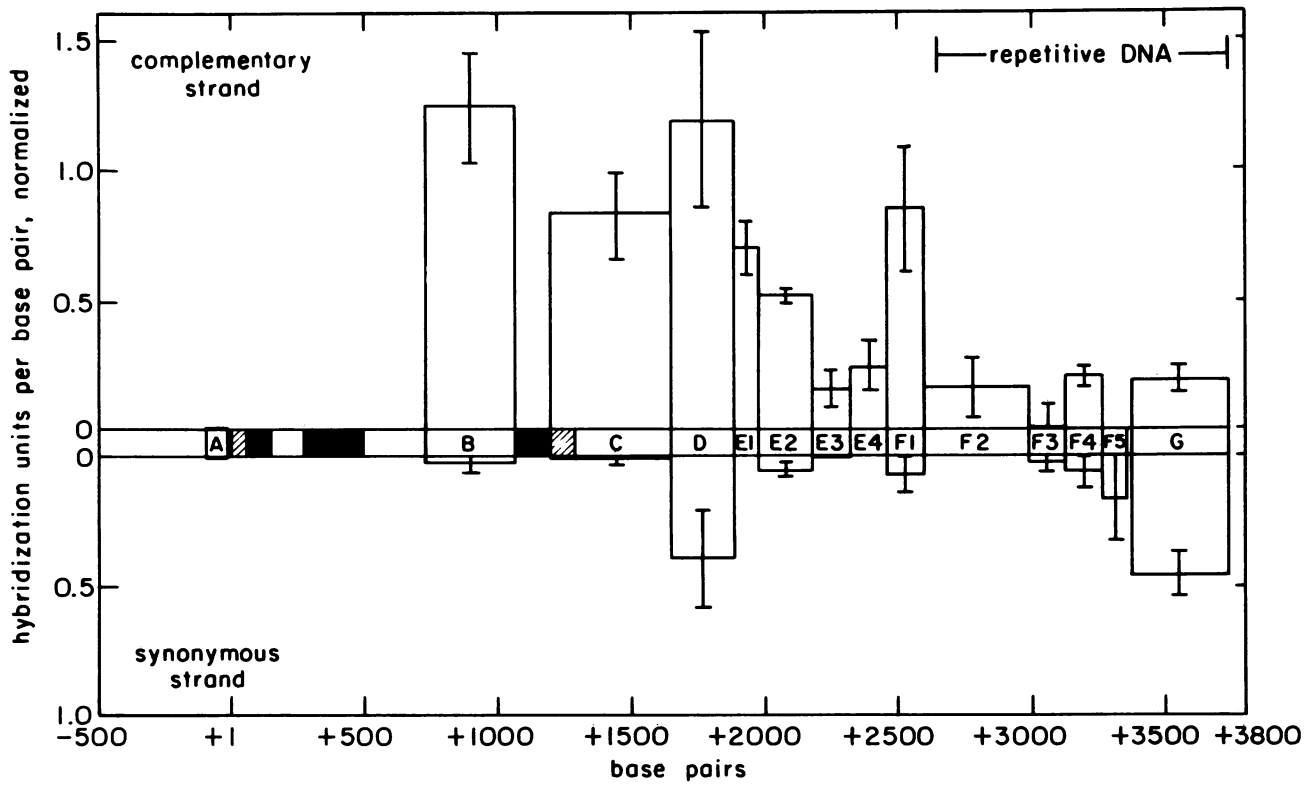


FIG. 2. Quantitative analysis of the transcripts from the β1 gene locus. The hybridization data shown in Fig. 1 were quantitated by microdensitometry, converted to hybridization units per base pair, and normalized so that the hybridization unit value per base pair for the sum of fragments B and C is 1.0. Data are plotted as bars along a map of the β1 gene region, with the width of the bar being the size of the fragment. Hybridization to the mRNA-complementary (antisense) strand is plotted above the gene map and hybridization to the mRNA-synonymous (sense) strand is plotted below the map. The error bars are ±one standard deviation for values with three determinations or the range of data for values with two determinations. The abscissa is in base pairs, with +1 being the cap nucleotide and sequences before the cap having negative numbers.

sum of transcription from all members of the D repeat family. Some of the individual repeats could be included in other transcription units in either orientation, which would account for the symmetry of transcription. One cannot determine from our assays whether any of this RNA came from the copy of the repeat 3′ to gene β1. However, the absence of hybridization to fragment F3 is informative. If any part of the repetitive DNA element located 3′ to β1 is transcribed in fetal liver nuclei (coordinately with gene β1), those transcripts do not continue throughout the entire repetitive DNA region.

Nucleotide sequence of the 3′ flanking region. We determined the nucleotide sequence of the region 3′ to β1 through fragment G to obtain the sequence of the transcribed region and to search for structures that could account for the attenuation of transcription after fragment D. Knowledge of the DNA sequence of this region was also necessary to identify the clones used in the nuclear RNA hybridization assays. The strategy for determining the sequence is presented in Fig. 4. Both the dideoxynucleotide chain termination technique for the M13 clones (25, 55) and the chemical degradation technique for the plasmid clones (39, 40) were used to determine the sequence. The sequence was determined from both strands over 69% of the region, each strand was sequenced multiple times, and many segments were determined by both sequencing techniques. By labeling the *Hpa*I sites at positions 1763 and 2452, we were able to sequence through the *Bgl*II and *Eco*RI sites that bound fragments D and E. This provided critical overlaps in the sequence.

The nucleotide sequence of the β1 globin gene region is presented in Fig. 5. This figure combines previously reported sequence data from the 5′ flank of β1 allele 1 (10) and the region from −75 to +1324 from β1 allele 2 (22) with the new sequence data (from β1 allele 2) for a total of 4,161 nucleotides. The region containing repetitive DNA lies between nucleotides 2598 and 3735.

Fragment D (nucleotides 1640 through 1892), which precedes the region where transcription declines, contains two interesting features. An inverted repeat is found between nucleotides 1664 and 1717, with the dyad axis between nucleotides 1689 and 1690. The RNA transcript in this region can be drawn as the stem-and-loop structure shown in Fig. 6 (upper strand), which has a calculated free energy of formation of −25.7 (54) or −25.8 (61) kcal (ca. −107.6 or −108 kJ).

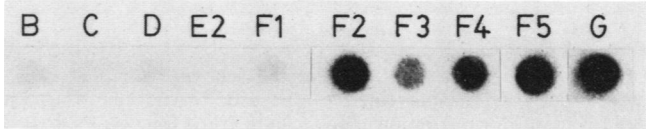


FIG. 3. Location of repetitive DNA flanking gene β1. Single-stranded DNA from M13 subclones of fragments B through G were spotted onto nitrocellulose and probed with labeled rabbit genomic DNA. Fragments containing repetitive DNA produce a signal on the autoradiogram.

Therefore, it is possible that the RNA could form this secondary structure in the primary transcript. Because of the inverted repeat in fragment D, the primary transcript for gene β1 will partially complement the mRNA-synonymous strand of fragment D (Ds) in two segments of 16 nucleotides. This complementarity, however, does not account for the hybridization of nuclear RNA to Ds. One can calculate that the approximate melting temperature is 34°C for this partial duplex (two groups of 16 nucleotides, 17% mismatch, 56% guanine plus cytosine) in the formamide hybridization conditions, and the melting temperature is 47°C in the 0.3× SSC wash (equations reviewed in reference 36). This is 8°C below the hybridization temperature and 21°C below the wash temperature, so the partial duplex should not be present on the filters. Also, if the RNA were hybridized to Ds DNA only in the inverted repeat region, most of the labeled RNA should be sensitive to RNase treatment. However, we observed no decrease in the Ds signal after treatment with RNase A (data not shown), which shows that the bulk of the labeled RNA (not just the inverted repeat) is in a hybrid with the DNA on the filter.

The second interesting feature of fragment D is a promoter consensus sequence for RNA polymerase II (see reference 2 for review). A short transcript from the strand opposite to the β1 transcript (RNA sequence synonymous with the lower strand in Fig. 5) comes from fragment D (Fig. 1). The sequence CATAAA (nucleotides 1850 to 1845 on the lower strand) and CAAT (nucleotides 1908 to 1905 on the lower strand) are at the correct positions to serve as −85 "CAAT" and −30 "ATA" promoter sequences for a transcript beginning around position 1819 and continuing to the beginning of fragment D (nucleotide 1640). The 3′ end of this transcript could possibly fold into the stem-and-loop structure shown in the lower portion of Fig. 6. The closest sequence to the AATAAA poly(A) addition signal (49) is the sequence AATAAGA located from nucleotides 1648 to 1642, but Higgs et al. (26) have shown that AATAAG in an α-thalassemia gene is not used efficiently as a poly(A) addition signal. The first ATG begins at 1744 and is followed by an open reading frame of 34 codons before the end of fragment D at the *Bgl*II site. However, we have not yet determined the stability, intracellular location, or translational capacity of this RNA. Thus the region from nucleotides 1908 to 1643 (lower strand of Fig. 5) contains most of the DNA sequences known to be required for transcription by RNA polymerase II, and the data in Fig. 1 show that this segment is transcribed into a short RNA in fetal liver nuclei.

The consensus sequence for an RNA polymerase III internal control region (62) is found in fragments E4 and F1. A match with the anterior control region for tRNA genes (box A) begins at nucleotide 2440 and is followed 76 nucleotides downstream by a match with the posterior control region (box B) at nucleotide 2526 (Fig. 7A). The distance between the match with box A and box B (76 nucleotides) exceeds the known values for the internal control regions of tRNA genes (33 nucleotides) but is close to the distance found for *Alu* repeats (61 nucleotides; reference 62). Also, the distance between the two internal control elements can be varied without decreasing significantly the efficiency of transcription (7, 30). The region just before fragment F1 matches for 23 of 33 nucleotides (with two gaps) with the internal control region for a 5S RNA gene (13) (Fig. 7B). These nucleotide sequence comparisons suggest that this segment of DNA (approximately nucleotides 2420 to 2560 in fragments E4 and F1) can be transcribed by RNA polymerase III in the same direction as the β1 globin gene. However, the sequences do not match completely with either tRNA or 5S RNA genes. A separate, short transcription unit could be proposed to explain the high level of transcription of fragment F1c relative to surrounding DNA (Fig. 1 and 2), and the sequence data suggest that this separate transcription unit could be read by RNA polymerase III. No matches with the RNA polymerase II promoter consensus sequence could be found in fragments E4 and F1.

Although an adenine-plus-thymine (A+T)-rich region might be expected to result in a less stable RNA-DNA duplex in the transcription complex, such an A+T-rich region is not sufficient by itself to terminate transcription by RNA polymerase II. Fragment C contains long tracts of polydeoxythymidylic acid-polydeoxyadenylic acid and is

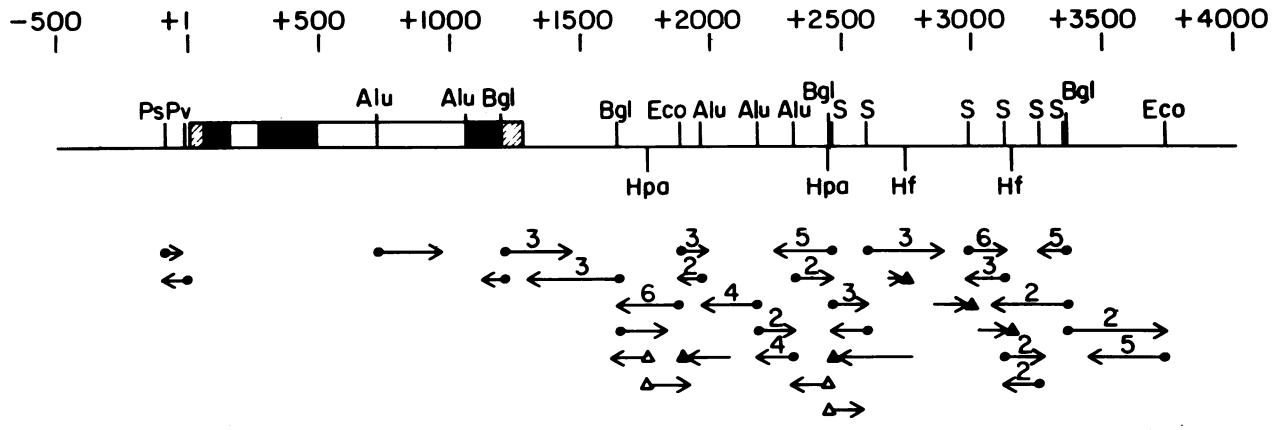


FIG. 4. Strategy for determining the sequence of the region 3′ to gene β1. Restriction sites used to generate M13 subclones for dideoxynucleotide sequencing (55) and used to end label fragments for Maxam and Gilbert (39) sequencing are shown on the top line. Abbreviations for restriction endonuclease are as follows: Ps, *Pst*I; Pv, *Pvu*II; Alu, *Alu*I; Bgl, *Bgl*II; Eco, *Eco*RI; S, *Sau*3AI; Hpa, *Hpa*I; Hf, *Hin*fI. The β1 globin gene is shown as a box with hatched untranslated regions, black protein-coding regions, and unshaded intervening sequences. The length of each arrow indicates the segment of DNA sequenced from a given site. The orientation of the arrow designates the strand sequenced: right arrows show that the mRNA-synonymous strand was read and left arrows show that the mRNA-complementary strand was read. The number of times a sequence was determined is given by the numeral over the arrows; arrows without numerals designate single determinations. Symbols: ●, M13 clones; ▲, 3′ end-labeled restriction sites; △, 5′ end-labeled restriction sites.

80% A+T from nucleotides 1507 to 1589. However, this fragment is transcribed at a level comparable to that of fragment B, and β1 transcription continues on through fragment D (Fig. 1 and 2). In conjunction with other factors, however, the A+T content of the DNA may play some role in terminating transcription. We note that fragment F2, which immediately precedes the transcriptionally silent fragment F3, is very A+T rich (87% A+T between nucleotides 2773 and 2885).

Conservation of 3′ flanking structures between rabbit and human β-globin genes. The observation that β1 transcription attenuates and eventually terminates in the region from 604 to 1,706 bp past the poly(A) addition site suggests that DNA sequences within this region may be involved in stopping transcription. If so, similar sequences (or secondary structures) could be conserved in homologous genes from related species. The recent determination of the sequence of the δ-β-globin gene region of human chromosome 11 (48) has allowed us to compare the 1,500 nucleotides past the poly(A) addition site between rabbit β1 and human β. The results are displayed as a dot plot in Fig. 8, where direct matches are shown as a diagonal with negative slope. The two sequences match for about 830 nucleotides past the poly(A) site of human β [680 nucleotides past the poly(A) site of rabbit β1], with two interruptions due to two apparent deletions in the rabbit DNA (or inserts in the human DNA) which show as offsets in the diagonal. The matches are not perfect, as shown by the gaps in the dotted diagonal, but some similarity in sequence has clearly been retained through fragment E1 in the rabbit DNA. Beyond this point, no long matches are observed between rabbit and human DNA. The rabbit repetitive DNA does not begin until fragment F2 [1,310 nucleotides past the poly(A)] and the first *Alu* repeat 3′ to the human β-globin gene begin 1,760 nucleotides past the poly(A) (11, 48). Thus, the loss of matching sequences is not a result of insertion of known repetitive DNA.

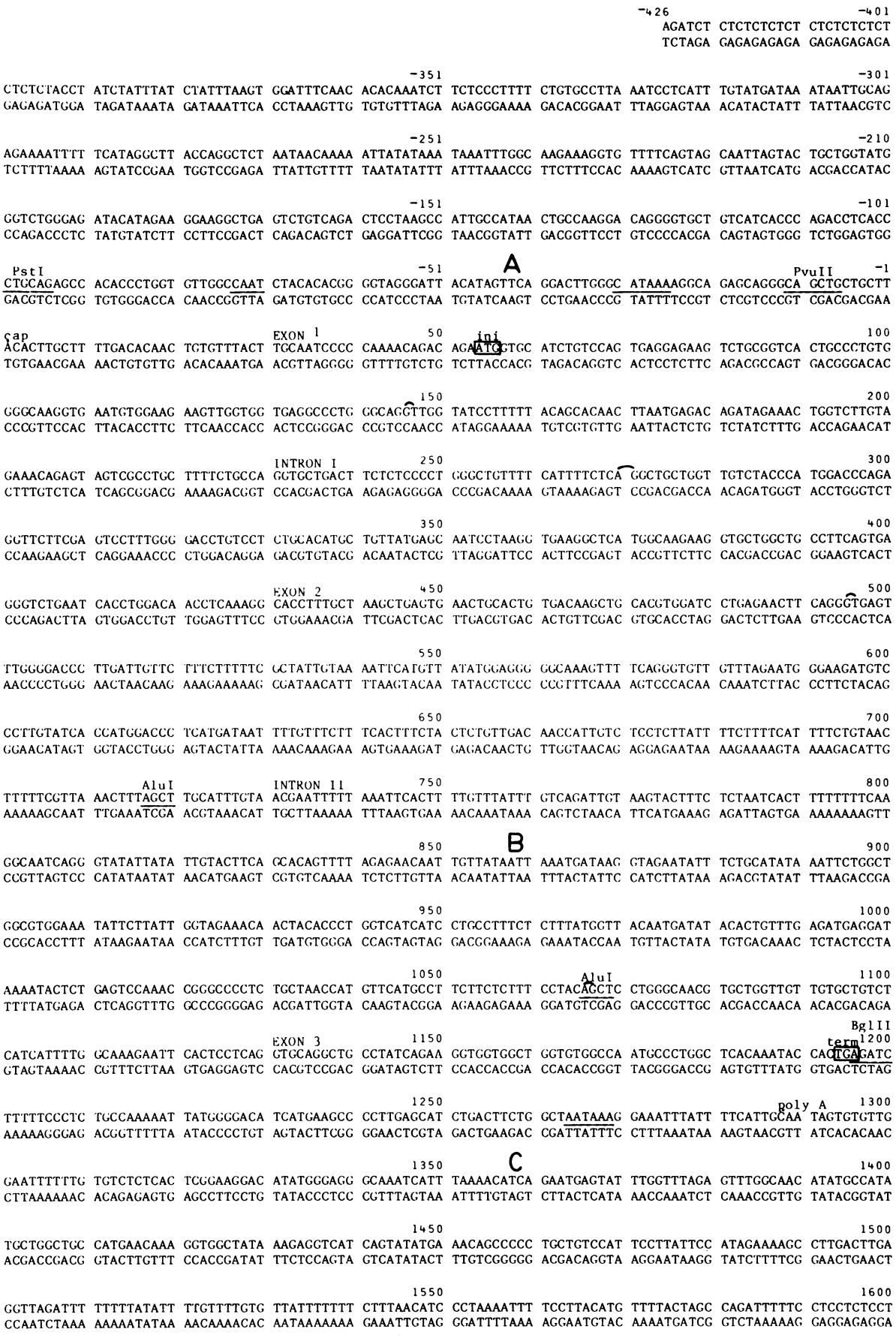
Not only is there overall conservation of sequences in the 3′ flanking region of rabbit β1 and human β, but an inverted repeat and a potential transcription unit opposed to the β-globin transcripts are also found 3′ to the human β-globin gene. These features are diagrammed in Fig. 9, which shows an alignment of the mRNA-complementary strands (lower strand in Fig. 5) of human and rabbit DNAs in the region corresponding to rabbit fragment D. The alignment was deduced by inspection, following the pattern given by the dot plot in Fig. 8. In the human sequence, the sequences CTAAT (nucleotides 13,749 to 13,745) and AATAAGA (nucleotides 13,688 to 13,682) are positioned to serve as −85 "CAAT" and −30 "ATA" sequences for a posible transcript beginning about nucleotide 13,660. Although these sequences are similar to promoter consensus sequences, they differ from the consensus in two critical nucleotides. Dierks et al. (9) showed that mutating the β-globin promoter sequences in vitro to CTAAT for the CAAT box and CATAAGA for the ATA box reduced the relative transcriptional level approximately eight-and twofold, respectively, but did not eliminate transcription entirely. These mutant sequences are very similar to the ones noted above in the 3′ flanking region of the human β-globin gene. Thus, these human sequences may have some promoter activity, albeit less than the β-globin promoter. An inverted repeat is found between nucleotides 13,530 and 13,461 with the dyad axis between nucleotides 13,498 and 13,497. Although these features are also found in the rabbit DNA, the homologous DNAs are not involved in the same structures. For example, the CAAT and ATA boxes in the rabbit DNA have diverged in the human DNA, but other CAAT and ATA boxes are found 17 to 11 nucleotides downstream. Likewise, the inverted repeats in rabbit and human DNAs are offset, with the dyad axis of the human inverted repeat located about 10 nucleotides downstream from that of the rabbit inverted repeat. This sequence comparison shows that the 3′ flanks of rabbit and human β-globin genes share a common ancestry, some nucleotide sequences have been conserved, and the potential for transcription by RNA polymerase II and for stem-and-loop formation by the RNA transcripts has also been retained. The transcriptional capacity of this region of the human β-globin 3′ flank has not yet been assayed in vivo or in vitro.

DISCUSSION

The data in this paper lead to six principal conclusions. (i) The rabbit β-like globin genes are differentially transcribed at different stages of development. (ii) The transcription unit of the rabbit β1 globin gene is very large, with transcription beginning at or near the cap site and continuing at a high level for 1,891 nucleotides [603 nucleotides past the poly(A) addition site]. Transcription continues, but a declining levels, for 2,459 nucleotides [the end of fragment E, 1,171 nucleotides past poly(A)] and goes no further than 2,994 nucleotides [the end of fragment F2, 1,706 nucleotides past poly(A)]. Since fragment F contains repetitive DNA, the source of the RNA hybridizing to it cannot be determined by this assay, and thus the endpoint β1 transcription cannot be determined exactly. (iii) Transcription of gene β1 does not terminate at a discrete site, but instead it attenuates after fragment D and continues to decline throughout fragment E. (iv) Fragments E4 and F1 appear to contain a separate transcription unit, possibly transcribed by RNA polymerase III. (v) Fragment D includes an inverted repeat and a short transcription unit in the opposite orientation to gene β1. The short RNA is transcribed by RNA polymerase II, as shown by its α-amanitin sensitivity and the presence of polymerase II promoter consensus sequences upstream. (vi) The repeated DNA located 3′ to gene β1 is not transcribed in its entirety in fetal liver nuclei.

Previous work has shown that the transcription unit for many eucaryotic polymerase II-transcribed genes is larger than the cap to poly(A) distance. The late transcription unit of adenovirus (15, 46), early regions 2 and 4 of adenovirus (45), and late transcripts of simian virus 40 (14) continue past one or more poly(A) addition sites. The 3′ terminus of each of these mRNAs is presumably generated by specific processing events, including cleavage and polyadenylation. UV transcription mapping has shown that the transcription units of mouse immunoglobulin genes (17) and *Drosophila* heat shock genes (5) are much larger than the corresponding mRNAs, but these studies did not directly address whether the flanking regions are transcribed. The most abundant pre-mRNA from the mouse β^{maj} globin gene corresponds to the cap to poly(A) region of the gene (8, 51, 65). Transcription of this gene, however, extends well past the poly(A) addition site for over 1,300 nucleotides (28, 29). Transcription mapping of chicken α- and β-globin genes also suggests that the primary transcript for each gene extends past the poly(A) addition site (34, 63, 66). Our observation of a large transcription unit for the rabbit β1 globin gene that extends well past the poly(A) addition site fits this pattern described for other polymerase II-transcribed genes.

Some β-related globin genes are also transcribed upstream from the cap site. Hofer and Darnell (28) report a small amount of transcription 5′ to the cap site of the mouse β^{maj}


-426
-401
AGATCT CTCTCTCTCT CTCTCTCTCT
TCTAGA GAGAGAGAGA GAGAGAGAGA
-351
-301
CTCTCTACCT ATCTATTTAT CTATTTAAGT GGATTTCAAC ACACAAATCT TCTCCCTTTT CTGTGCCTTA AATCCTCATT TGTATGATAA ATAATTGCAG
GAGAGATGGA TAGATAAATA GATAAATTCA CCTAAAGTTG TGTGTTTAGA AGAGGGAAAA GACACGGAAT TTAGGAGTAA ACATACTATT TATTAACGTC
-251
-210
AGAAAATTTT TCATAGGCTT ACCAGGCTCT AATAACAAAA ATTATATAAA TAAATTTGGC AAGAAAGGTG TTTTCAGTAG CAATTAGTAC TGCTGGTATG
TCTTTTAAAA AGTATCCGAA TGGTCCGAGA TTATTGTTTT TAATATATTT ATTTAAACCG TTCTTTCCAC AAAAGTCATC GTTAATCATG ACGACCATAC
-151
-101
GGTCTGGGAG ATACATAGAA GGAAGGCTGA GTCTGTCAGA CTCCTAAGCC ATTGCCATAA CTGCCAAGGA CAGGGGTGCT GTCATCACCC AGACCTCACC
CCAGACCCTC TATGTATCTT CCTTCCGACT CAGACAGTCT GAGGATTCGG TAACGGTATT GACGGTTCCT GTCCCCACGA CAGTAGTGGG TCTGGAGTGG
PstI
-51
A
PvuII
-1
CTGCAGAGCC ACACCCTGGT GTTGCCCAAT CTACACACGG GGTAGGGATT ACATAGTTCA GGACTTGGGC ATAAAAGGCA GAGCAGGGCA GCTGCTGCTT
GACGTCTCGG TGTGGGACCA CAACCGGTTA GATGTGTGCC CCATCCCTAA TGTATCAAGT CCTGAACCCG TATTTTCCGT CTCGTCCCGT CGACGACGAA
cap
EXON 1
50
ini
100
ACACTTGCTT TTGACACAAC TGTGTTTACT TGCAATCCCC CAAAACAGAC AGAATGGTGC ATCTGTCCAG TGAGGAGAAG TCTGCGGTCA CTGCCCTGTG
TGTGAACGAA AACTGTGTTG ACACAAATGA ACGTTAGGGG GTTTTGTCTG TCTTACCACG TAGACAGGTC ACTCCTCTTC AGACGCCAGT GACGGGACAC
150
200
GGGCAAGGTG AATGTGGAAG AAGTTGGTGG TGAGGCCCTG GGCAGGTTGG TATCCTTTTT ACAGCACAAC TTAATGAGAC AGATAGAAAC TGGTCTTGTA
CCCGTTCCAC TTACACCTTC TTCAACCACC ACTCCGGGAC CCGTCCAACC ATAGGAAAAA TGTCGTGTTG AATTACTCTG TCTATCTTTG ACCAGAACAT
INTRON I
250
300
GAAACAGAGT AGTCGCCTGC TTTTCTGCCA GGTGCTGACT TCTCTCCCCT GGGCTGTTTT CATTTTCTCA GGCTGCTGGT TGTCTACCCA TGGACCCAGA
CTTTGTCTCA TCAGCGGACG AAAAGACGGT CCACGACTGA AGAGAGGGGA CCCGACAAAA GTAAAAGAGT CCGACGACCA ACAGATGGGT ACCTGGGTCT
350
400
GGTTCTTCGA GTCCTTTGGG GACCTGTCCT CTGCACATGC TGTTATGAGC AATCCTAAGG TGAAGGCTCA TGGCAAGAAG GTGCTGGCTG CCTTCAGTGA
CCAAGAAGCT CAGGAAACCC CTGGACAGGA GACGTGTACG ACAATACTCG TTAGGATTCC ACTTCCGAGT ACCGTTCTTC CACGACCGAC GGAAGTCACT
EXON 2
450
500
GGGTCTGAAT CACCTGGACA ACCTCAAAGG CACCTTTGCT AAGCTGAGTG AACTGCACTG TGACAAGCTG CACGTGGATC CTGAGAACTT CAGGGTGAGT
CCCAGACTTA GTGGACCTGT TGGAGTTTCC GTGGAAACGA TTCGACTCAC TTGACGTGAC ACTGTTCGAC GTGCACCTAG GACTCTTGAA GTCCCACTCA
550
600
TTGGGGACCC TTGATTGTTC TTTCTTTTTC GCTATTGTAA AATTCATGTT ATATGGAGGG GGCAAAGTTT TCAGGGTGTT GTTTAGAATG GGAAGATGTC
AACCCCTGGG AACTAACAAG AAAGAAAAAG CGATAACATT TTAAGTACAA TATACCTCCC CCGTTTCAAA AGTCCCACAA CAAATCTTAC CCTTCTACAG
650
700
CCTTGTATCA CCATGGACCC TCATGATAAT TTTGTTTCTT TCACTTTCTA CTCTGTTGAC AACCATTGTC TCCTCTTATT TTCTTTTCAT TTTCTGTAAC
GGAACATAGT GGTACCTGGG AGTACTATTA AAACAAAGAA AGTGAAAGAT GAGACAACTG TTGGTAACAG AGGAGAATAA AAGAAAAGTA AAAGACATTG
AluI
INTRON II
750
800
TTTTTCGTTA AACTTTAGCT TGCATTTGTA ACGAATTTTT AAATTCACTT TTGTTTATTT GTCAGATTGT AAGTACTTTC TCTAATCACT TTTTTTTCAA
AAAAAGCAAT TTGAAATCGA ACGTAAACAT TGCTTAAAAA TTTAAGTGAA AACAAATAAA CAGTCTAACA TTCATGAAAG AGATTAGTGA AAAAAAAGTT
850
B
900
GGCAATCAGG GTATATTATA TTGTACTTCA GCACAGTTTT AGAGAACAAT TGTTATAATT AAATGATAAG GTAGAATATT TCTGCATATA AATTCTGGCT
CCGTTAGTCC CATATAATAT AACATGAAGT CGTGTCAAAA TCTCTTGTTA ACAATATTAA TTTACTATTC CATCTTATAA AGACGTATAT TTAAGACCGA
950
1000
GGCGTGGAAA TATTCTTATT GGTAGAAACA ACTACACCCT GGTCATCATC CTGCCTTTCT CTTTATGGTT ACAATGATAT ACACTGTTTG AGATGAGGAT
CCGCACCTTT ATAAGAATAA CCATCTTTGT TGATGTGGGA CCAGTAGTAG GACGGAAAGA GAAATACCAA TGTTACTATA TGTGACAAAC TCTACTCCTA
1050
AluI
1100
AAAATACTCT GAGTCCAAAC CGGGCCCCTC TGCTAACCAT GTTCATGCCT TCTTCTCTTT CCTACAGCTC CTGGGCAACG TGCTGGTTGT TGTGCTGTCT
TTTTATGAGA CTCAGGTTTG GCCCGGGGAG ACGATTGGTA CAAGTACGGA AGAAGAGAAA GGATGTCGAG GACCCGTTGC ACGACCAACA ACACGACAGA
BglII
EXON 3
1150
term
1200
CATCATTTTG GCAAAGAATT CACTCCTCAG GTGCAGGCTG CCTATCAGAA GGTGGTGGCT GGTGTGGCCA ATGCCCTGGC TCACAAATAC CACTGAGATC
GTAGTAAAAC CGTTTCTTAA GTGAGGAGTC CACGTCCGAC GGATAGTCTT CCACCACCGA CCACACCGGT TACGGGACCG AGTGTTTATG GTGACTCTAG
1250
poly A
1300
TTTTTCCCTC TGCCAAAAAT TATGGGGACA TCATGAAGCC CCTTGAGCAT CTGACTTCTG GCTAATAAAG GAAATTTATT TTCATTGCAA TAGTGTGTTG
AAAAAGGGAG ACGGTTTTTA ATACCCCTGT AGTACTTCGG GGAACTCGTA GACTGAAGAC CGATTATTTC CTTTAAATAA AAGTAACGTT ATCACACAAC
1350
C
1400
GAATTTTTTG TGTCTCTCAC TCGGAAGGAC ATATGGGAGG GCAAATCATT TAAAACATCA GAATGAGTAT TTGGTTTAGA GTTTGGCAAC ATATGCCATA
CTTAAAAAAC ACAGAGAGTG AGCCTTCCTG TATACCCTCC CGTTTAGTAA ATTTTGTAGT CTTACTCATA AACCAAATCT CAAACCGTTG TATACGGTAT
1450
1500
TGCTGGCTGC CATGAACAAA GGTGGCTATA AAGAGGTCAT CAGTATATGA AACAGCCCCC TGCTGTCCAT TCCTTATTCC ATAGAAAAGC CTTGACTTGA
ACGACCGACG GTACTTGTTT CCACCGATAT TTCTCCAGTA GTCATATACT TTGTCGGGGG ACGACAGGTA AGGAATAAGG TATCTTTTCG GAACTGAACT
1550
1600
GGTTAGATTT TTTTTATATT TTGTTTTGTG TTATTTTTTT CTTTAACATC CCTAAAATTT TCCTTACATG TTTTACTAGC CAGATTTTTC CTCCTCTCCT
CCAATCTAAA AAAAATATAA AACAAAACAC AATAAAAAAA GAAATTGTAG GGATTTTAAA AGGAATGTAC AAAATGATCG GTCTAAAAAG GAGGAGAGGA

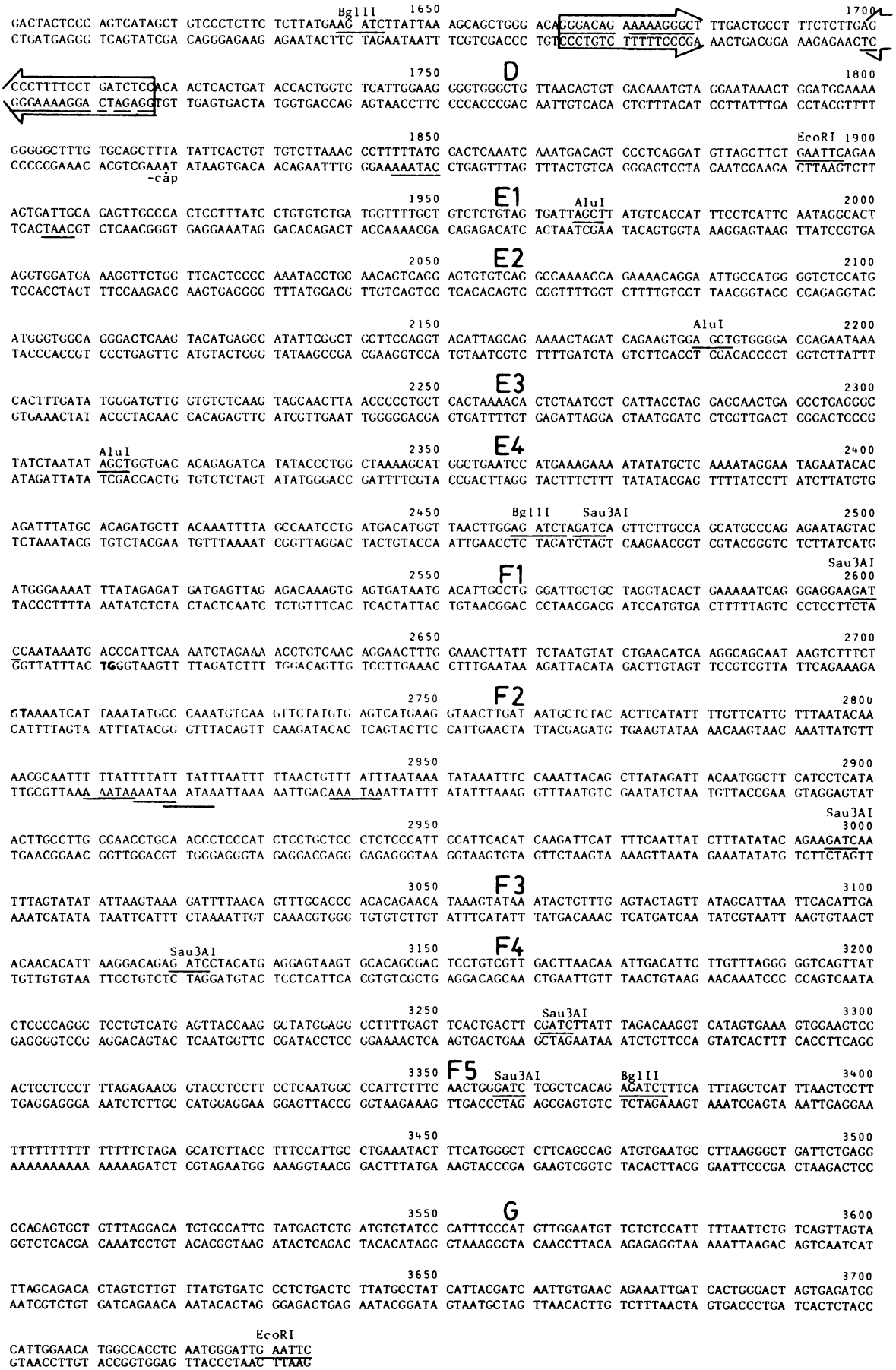


FIG. 5. Sequence of the β1 gene locus. The sequence of both strands of the β1 gene from 426 base pairs 5′ to the cap site to 2,447 bp 3′ to the poly(A) addition site is shown. The cap nucleotide is numbered as +1, and upstream sequences have negative numbers. The sequence of allele 2 of the β1 gene from 75 bp 5′ to the cap site to 36 bp 3′ to the poly(A) addition site (nucleotides −75 to +1324) was reported in Hardison et al. (22). The 5′ flanking sequence from −426 to the cap site of allele 1 was determined by Dierks et al. (10). Consensus sequences, including the CAAT and ATA boxes in the 5′ flanking region and the AATAAA poly(A) addition signal, are underlined. Restriction endonuclease cleavage sites that form the boundaries of the fragments used in the hybridization assays are indicated and the fragments are labeled. The inverted repeat in fragment D is indicated by opposing arrows. Exons and introns in gene β1 are labeled, and the ATG and TGA codons for translation initiation and termination are boxed.

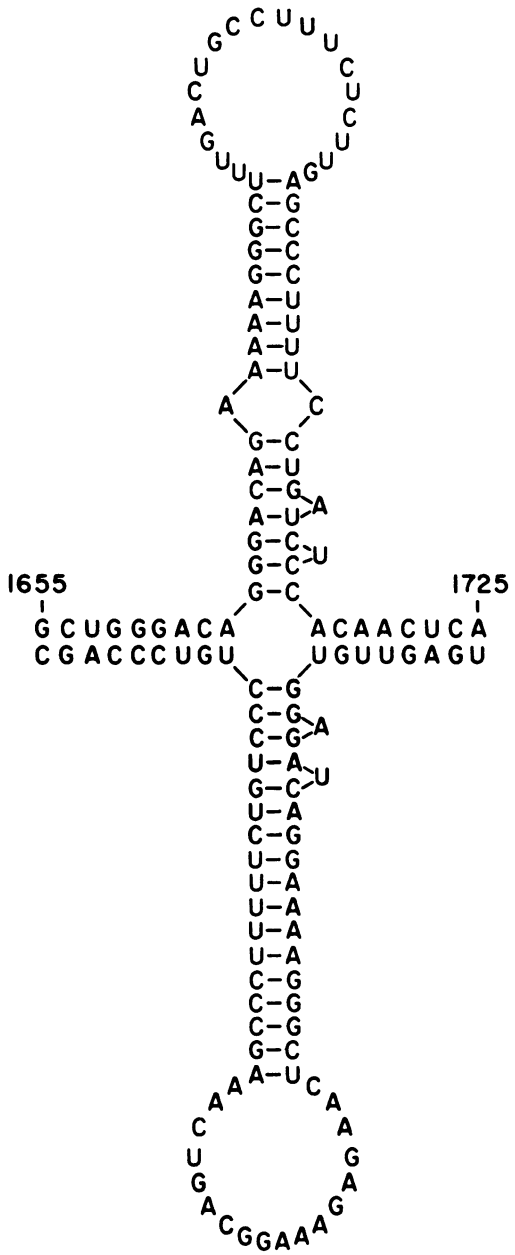


FIG. 6. Possible secondary structure of transcripts from fragment D. Transcription of the inverted repeat in fragment D DNA (nucleotides 1664 through 1717) could possibly result in the formation of a stem-and-loop structure in the RNA. One such structure is shown, with paired bases in the stem indicated by horizontal lines. A potential secondary structure in the short transcript from fragment D (sequence synonymous with the lower strand in Fig. 5) is shown in the lower portion of the figure. It differs from the upper structure in having a G-U base pair in place of the A-C interior loop found in the upper stem.

globin gene, Carlson and Ross (4) observed a polymerase III-transcribed RNA that initiated 5′ to the cap site of the human β-globin gene, and Allan et al. (1) found several upstream start points for transcription of the human ε-globin gene. However, in each of these cases, the major site of initiation is at the cap site. The only indications of upstream transcription from the rabbit β1 gene are the aberrant RNAs which are detected when heterologous cells are transformed with the β1 globin gene (68). When the gene is introduced via a viral vector, such RNAs are minor species compared with normal β1 RNAs (9, 10, 18). Some of the aberrant RNAs have 5′ ends which map to regions upstream from the cap site (9), and the most abundant aberrant RNAs have "5′ termini" which map to positions +42 to +48. Grosveld et al. (18) suggest that these 5′ termini are in fact not termini, but represent splice sites of RNAs which arise from initiation upstream from the cap site. Using nuclei from a normal erythroid tissue, we do not detect any transcription upstream from the cap site of the rabbit β1 globin gene. Apparently the aberrant β1 RNA species observed previously are an artifact of the transfection assay.

No discrete site for termination of transcription by RNA polymerase II has been identified in the systems studied to date. Previous work with the mouse β^{maj} globin gene in MEL cells (28, 29) has shown that transcription equimolar to that within the gene continues for about 580 nucleotides past the poly(A) site (the end of mouse β^{maj} fragment E). The next 710 nucleotides (mouse fragment F) were found to be transcribed at an equimolar level in one experiment (28), but more recent experiments show that this region is transcribed at a lower level than internal gene sequences (52, 53). The next 590 nucleotides (mouse fragment G) show very little transcription (29, 52). This pattern of high level of transcription for 580 nucleotides past the poly(A) site followed by a decreasing level of transcription for a total of about 1,880 nucleotides past the poly(A) site is strikingly similar to that described in this report for the rabbit β1 globin gene. In both the mouse and rabbit β-globin genes, transcription declines downstream from a region about 600 nucleotides past the poly(A) site. We refer to this region as an attenuator because the number of transcription complexes decreases beyond this point. Since this region lies well past the sequences found in the precursor RNAs, it is possible that polyadenylation of the pre-mRNA occurs independently of (and perhaps prior to) the transcription of the attenuator region. Thus this attenuator may not be directly involved in the formation of

A. tRNA gene

```
                                                    T
Consensus     RRYNNARYGG...(33-43)...GATCRANNC
              ' ' ' ' ' ' ' ' '                 ' ' ' ' ' ' '
E4 + F1       GATGACATGG...(76).....GTTAGAGAC
          2440                              2526
```

B. 5S RNA gene

```
                50                                          83
Human           AGCTAAGCAGGGTCGGGCCTGGTTAGTACTTGGA
                ' ' ' ' ' '  ' '  ' '  ' '  ' ' ' ' ' '  ' ' ' ' ' ' '
E4 + F1         AGCCAATCCTGAT--GACATGGTTA--ACTTGGA
            2430                                        2459
```

FIG. 7. Alignment of RNA polymerase III promoter consensus sequences with similar regions in fragments E4 and F1. (A) The top line shows the consensus sequence of RNA polymerase III anterior and posterior control regions (box A and box B) for tRNA genes (62). The anterior control sequence was aligned by inspection to a segment of fragment E4 (nucleotides 2440 to 2449), and the posterior control region matched a sequence found in fragment F1 (nucleotides 2526 to 2534). The distance between the control regions is given in base pairs. (B) The top line shows the internal control region (nucleotides 50 to 83) of the human 5S RNA gene (13). This region was matched by inspection to nucleotides 2430 to 2459 in fragment E4. Matching nucleotides are marked and gaps are inserted to improve the alignment.

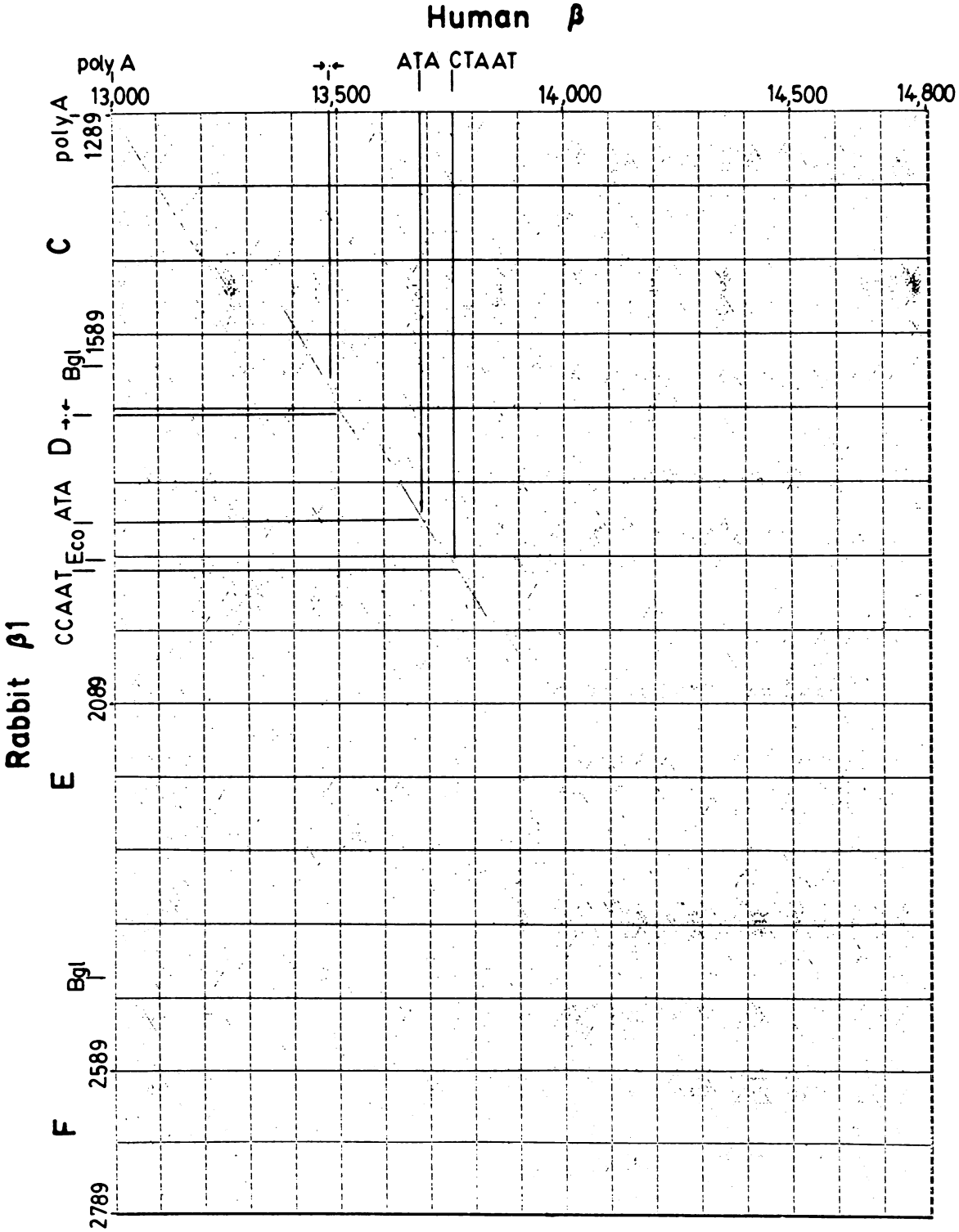


FIG. 8. Similar sequences found in the β-globin gene 3′ flanking gene regions of human and rabbit DNA. The dot plot shows similar sequences as a diagonal of dots with negative slope for direct matches and positive slope for matches with the reverse complement. The human sequence is arrayed along the horizontal axis, starting with the nucleotide after the poly(A) addition site and extending for 1,800 nucleotides. The sequence is from Poncz et al. (48) and uses their numbering system. The rabbit sequence is arrayed along the vertical axis, beginning at the nucleotide after the poly(A) addition site and extending for 1,500 nucleotides. The positions of inverted repeats (opposing arrows) and ATA and CAAT promoter consensus sequences are indicated. The rabbit DNA axis is also labeled with restriction sites and DNA fragment letters for comparison with other figures. This dot plot was obtained from the program MATRIX (70) searching for 13 matches in a 15-nucleotide window. A similar pattern is obtained at other criteria and with other programs.

globin mRNA. In this sense it is clearly different from other attenuators, such as those that regulate bacterial amino acid biosynthesis operons (reviewed in reference 69) or the attenuator recently identified in simian virus 40 (24). In these examples, the attenuator precedes the coding portion of the RNA, and modulation of transcription at these attenuators regulates the level of expression of the downstream genes. If the attenuator described for the β-globin genes has a role in regulation, it is distinctly different from that of the familiar amino acid biosynthesis operons.

We propose that two structures found in fragment D may cause transcription to attenuate. The first structure is the inverted repeat which could potentially form a stem-and-loop structure in the RNA transcript. A stem-and-loop structure is a common feature of procaryotic transcription terminators (reviewed in reference 47). Hay et al. (24) have proposed that a stem-and-loop structure is involved in the premature termination of simian virus 40 transcription, and they have shown that the stability of the RNA secondary structure affects the efficiency of pausing at the attenuator site (23). Similarly, the formation of this stem-and-loop structure in the β1 globin RNA transcript could impede the progress of the RNA polymerase. The second structure found in fragment D is the short transcription unit. Since it is on the strand opposite that of the β1 transcript, the RNA polymerases from the short transcript should interfere with the progress of the RNA polymerases from gene β1, perhaps causing some of them to dissociate from the transcription complex. The polymerases transcribing β1 would also block the progress of polymerases synthesizing the short (Ds) transcript. If the polymerases transcribing Ds pause at the stem-and-loop structure in the RNA (Fig. 6), then the large number of polymerases transcribing β1 in the opposite direction could cause dissociation of the Ds polymerases at the inverted repeat, and thus the short transcript would be confined to fragment D. Such an attenuation by collision could be an effective means of limiting the amount of transcription from a given gene. This aspect of the proposal assumes that the short transcript is synthesized in the same cells that are transcribing β1. The source of the nuclei used in our assays, the fetal liver, is about 80% erythropoietic (59), and although it is likely that both β1 and the short D fragment RNA are transcribed in the same cells, it is possible that the D fragment transcript is made in the 20% non-erythropoietic cells. One observation that favors coordinate transcription of β1 and the Ds fragment is that neither are transcribed in embryonic cells (see Fig. 1), but further experiments are required to test this point. The inverted repeat and the short transcript are found in the DNA fragment upstream from the fragment where declining transcription is observed. Thus, if our proposal is correct, the effect of these structures on transcription is not exerted immediately. Only after RNA polymerase has progressed about another 200 nucleotides does the level of transcription decline noticeably, and the transcription continues to decline as the polymerase proceeds further.

FIG. 9. Alignment of rabbit and human sequences in the region corresponding to rabbit fragment D. By using the information in the dot plot of Fig. 8, the rabbit and human sequences in the region corresponding to fragment D were aligned by inspection. The human sequence (H) is on the top line, numbered according to Poncz et al. (48). The rabbit sequence (R) is on the bottom line, numbered according to Fig. 5. In both cases, the sequence of the mRNA-complementary strand (lower strand in Fig. 5) is given. Matching nucleotides are indicated by asterisks, and gaps are inserted to improve the alignment. Sequences matching the CAAT (−85) and ATA (−30) promoter consensus sequences are overlined (human) or underlined (rabbit), and the potential start site for transcription (a purine about 30 nucleotides downstream from ATA) is labeled as ~cap. The inverted repeats are indicated by opposing arrows.

Both of these structural features, the inverted repeat and the potential for an opposing transcript, are found in the homologous 3′ flanking region of the human β-globin gene. This observation is consistent with their having some function in the expression of the β-globin genes. Some other sequences within the 3′ flanking region are also conserved between rabbits and humans, and these could play a role in attenuation as well, perhaps as factor recognition sites. The transcriptionally silent F3 region is preceded by a very A+T-rich segment in the F2 fragment. Although a high A+T content alone is not sufficient to terminate transcription (the A+T-rich fragment C is transcribed at a high level), a high A+T content in conjunction with other factors could lead to a cessation of transcription. Our proposal that the *cis*-acting inverted repeat and opposing transcript are involved in attenuation of transcription does not preclude the additional involvement of *trans*-acting, diffusable factors. In particular, high salt-soluble factors have been proposed to function in transcription termination of rRNA (35), histone mRNA (60), and simian virus 40 RNA (23). If the short transcript from fragment D is not rapidly degraded, it could act in *trans* to form a double-stranded segment of RNA in the primary transcript of gene β1. Such a novel structure could potentially act as a recognition site for a regulatory factor.

A portion of the D repeat family member located 3′ to gene β1 is not transcribed in fetal liver nuclei (fragment F3).

However, internal regions of the repeat do hybridize to nascent RNA, but these RNAs could come from any member of the repeat family. Thus, the entire D repeat element located 3′ to β1 is not transcribed coordinately with the β1 globin gene, but we cannot eliminate the possibility that internal portions of this D repeat member may be coordinately transcribed with β1. The D repeat and C repeat DNAs also hybridize to embryonic nuclear RNA, so these repeat families are not under obvious developmental control. Thus, the D and C repeat families differ from some of the repeats in the chicken β-globin gene cluster which are transcribed in a stage-specific manner (63).

ACKNOWLEDGMENTS

We thank J. E. Darnell, Jr. for information prior to publication, S. Zweig for the MATRIX computer program, C. Kappel for assistance in the sequence determination, and T. Peters for typing the manuscript.

This work was supported by Public Health Service grants AM27635 and AM31961 from the National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases.

LITERATURE CITED

1. **Allan, M., G. Lanyon, and J. Paul.** 1983. Multiple origins of transcription in the 4.5 kb upstream of the ε-globin gene. Cell **35**:187–197.
2. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
3. **Brison, O., and P. Chambon.** 1976. A simple and efficient method to remove ribonuclease contamination from pancreatic deoxyribonuclease preparations. Anal. Biochem. **75**:402–409.
4. **Carlson, D. P., and J. Ross.** 1983. Human β-globin promoter and coding sequences transcribed by RNA polymerase III. Cell **34**:857–864.
5. **Carlson, J. O., and D. E. Pettijohn.** 1979. Lengths of transcriptional units coding for the heat shock proteins in Drosophila melanogaster. J. Mol. Biol. **132**:141–161.
6. **Cheng, J.-F., R. Printz, T. Callaghan, D. Shuey, and R. Hardison.** 1984. The rabbit C family of short, interspersed repeats: nucleotide sequence determination and transcriptional analysis. J. Mol. Biol. **176**:1–20.
7. **Ciliberto, G., G. Traboni, and R. Cortese.** 1982. Relationship between the two components of the split promoter of eukaryotic tRNA genes. Proc. Natl. Acad. Sci. U.S.A. **79**:1921–1925.
8. **Curtis, P. J., and C. Weissman.** 1976. Purification of globin mRNA from DMSO induced Friend cells and detection of a putative globin mRNA precursor. J. Mol. Biol. **106**:1061–1075.
9. **Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**:695–706.
10. **Dierks, P., A. van Ooyen, N. Mantei, and C. Weissman.** 1981. DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5′ terminus. Proc. Natl. Acad. Sci. U.S.A. **78**:1411–1415.
11. **Duncan, C. H., P. Jagadeeswaran, R. C. Wang, and S. M. Weissman.** 1981. Structural analysis of templates and RNA polymerase III transcript of Alu family sequences interspersed among the human β-like globin genes. Gene **13**:185–196.
12. **Enea, V., G. Vovis, and N. Zinder.** 1975. Genetic studies with heteroduplex DNA of bacteriophage f1. Asymmetric segregation, base correction and implications for the mechanisms of genetic recombination. J. Mol. Biol. **96**:495–509.
13. **Erdmann, V., E. Huymans, A. Vandenberghe, and R. DeWachter.** 1983. Collection of published 5S and 5.8S RNA sequences. Nucleic Acids Res. **11**:r105–r133.
14. **Ford, J., and M.-T. Hsu.** 1978. Transcription pattern of in vivo labeled late simian virus 40 RNA: equimolar transcription beyond the mRNA 3′ terminus. J. Virol. **28**:795–801.
15. **Fraser, N. W., J. Nevins, E. Ziff, and J. E. Darnell., Jr.** 1979. The major late adenovirus type-2 transcription unit: termination is downstream from the last poly(A) site. J. Mol. Biol. **129**:643–656.
16. **Fritsch, E., C.-K. J. Shen, R. Lawn, and T. Maniatis.** 1980. The organization of repetitive sequences in mammalian globin gene clusters. Cold Spring Harbor Symp. Quant. Biol. **44**:761–775.
17. **Gilmore-Hebert, M., K. Hercules, M. Kamaroiny, and R. Wall.** 1978. Variable and constant regions are separated in the 10-kbase transcription unit coding for immunoglobin K light chains. Proc. Natl. Acad. Sci. U.S.A. **75**:6044–6048.
18. **Grosveld, G. C., E. de Boer, C. K. Shewmaker, and R. A. Flavell.** 1982. DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. Nature (London) **295**:120–126.
19. **Grosveld, G. C., A. Koster, and R. A. Flavell.** 1981. A transcription map for the rabbit β-globin gene. Cell **23**:573–584.
20. **Grosveld, G. C., C. K. Shewmaker, P. Jat, and R. A. Flavell.** 1981. Localization of DNA sequences necessary for transcription of the rabbit β-globin gene *in vitro*. Cell **25**:215–226.
21. **Groudine, M., M. Peretz, and H. Weintraub.** 1981. Transcriptional regulation of hemoglobin switching in chicken embryos. Mol. Cell. Biol. **1**:281–288.
22. **Hardison, R., E. Butler III, E. Lacy, T. Maniatis, N. Rosenthal, and A. Efstratiadis.** 1979. The structure and transcription of four linked rabbit β-like globin genes. Cell **18**:1285–1297.
23. **Hay, N., and Y. Aloni.** 1984. Attenuation in SV40 as a mechanism of transcription-termination by RNA polymerase B. Nucleic Acids Res. **12**:1401–1414.
24. **Hay, N., H. Skolnik-David, and Y. Aloni.** 1982. Attenuation in the control of SV40 gene expression. Cell **29**:183–193.
25. **Heidecker, G., J. Messing, and B. Gronenborn.** 1980. A versatile primer for DNA sequencing in the M13 mp2 cloning system. Gene **10**:69–73.
26. **Higgs, D. R., S. E. Y. Goodbourn, J. Lamb, J. B. Clegg, D. J. Weatherall, and N. J. Proudfoot.** 1983. Alpha thalassemia caused by a polyadenylation signal mutation. Nature (London) **306**:398–400.
27. **Hoeijmakers-van Dommelen, H. A. M., G. C. Grosveld, E. deBoer, R. A. Flavell, J. M. Varley, and A. J. Jeffreys.** 1980. Localization of repetitive and unique DNA sequences neighboring the rabbit β-globin gene. J. Mol. Biol. **140**:531–547.
28. **Hofer, E., and J. E. Darnell, Jr.** 1981. The primary transcription unit of the mouse β-major globin gene. Cell **23**:585–593.
29. **Hofer, E., R. Hofer-Warbinek, and J. E. Darnell, Jr.** 1982. Globin RNA transcription: a possible termination site and demonstration of transcriptional control correlated with altered chromatin structure. Cell **29**:887–893.
30. **Hofstetter, H., A. Kressmann, and M. Birnsteil.** 1981. A split promoter for a eukaryotic tRNA gene. Cell **24**:573–585.
31. **Kafatos, F., C. W. Jones, and A. Efstratiadis.** 1979. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res. **7**:1541–1552.
32. **Lacy, E., R. Hardison, D. Quon, and T. Maniatis.** 1979. The linkage arrangement of four rabbit β-like globin genes. Cell **18**:1273–1283.
33. **Lacy, E., and T. Maniatis.** 1980. The nucleotide sequence of a rabbit β-globin pseudogene. Cell **21**:545–553.
34. **Landes, G. M., B. Villeponteau, T. M. Pribyl, and H. G. Martinson.** 1982. Hemoglobin switching in chickens: Is the switch initiated post-transcriptionally? J. Biol. Chem. **257**:11008–11014.
35. **Leer, Y., D. Tiryaki, and O. Westergaard.** 1979. Termination of transcription in nucleoli isolated from Tetrahymena. Proc. Natl. Acad. Sci. U.S.A. **76**:5563–5566.
36. **Lewin, B.** 1980. Gene expression, vol. 2, 2nd ed., p. 503–508. John Wiley & Sons, Inc., New York.
37. **Maniatis, T., R. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15**:687–701.
38. **Maniatis, T., A. Jeffrey, and D. G. Kleid.** 1975. Nucleotide sequence of the rightward operator of phage λ. Proc. Natl.

Acad. Sci. U.S.A. **72**:1184–1188.
39. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
40. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
41. **McKnight, G. S., and R. Palmiter.** 1979. Transcriptional regulation of the ovalbumin and conalbumin genes by steroid hormones in chick oviduct. J. Biol. Chem. **254**:9050–9058.
42. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
43. **Messing, J., R. Crea, and P. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
44. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
45. **Nevins, J. R., J.-M. Blanchard, and J. E. Darnell.** 1980. Transcription units of Adenovirus type 2: termination of transcription beyond the poly(A) addition site in early regions 2 and 4. J. Mol. Biol. **144**:377–386.
46. **Nevins, J. R., and J. E. Darnell.** 1978. Steps in the processing of Ad 2 mRNA: poly(A)$^+$ nuclear species are conserved and poly(A) addition precedes splicing. Cell **15**:1477–1493.
47. **Platt, T.** 1981. Termination of transcription and its regulation in the tryptophan operon of E. coli. Cell **24**:10–23.
48. **Poncz, M., E. Schwartz, M. Ballantine, and S. Surrey.** 1983. Nucleotide sequence analysis of the δβ-globin gene region in humans. J. Biol. Chem. **258**:11599–11609.
49. **Proudfoot, N., and G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
50. **Rohrbaugh, M. L., and R. C. Hardison.** 1983. Analysis of rabbit β-like globin gene transcripts during development. J. Mol. Biol. **164**:395–417.
51. **Ross, J., and D. A. Knecht.** 1978. Precursors of α and β globin mRNA's. J. Mol. Biol. **119**:1–20.
52. **Salditt-Georgieff, M., and J. E. Darnell, Jr.** 1983. A precise termination site in the mouse β^{maj}-globin transcription unit. Proc. Natl. Acad. Sci. U.S.A. **80**:4694–4698.
53. **Salditt-Georgieff, M., and J. E. Darnell, Jr.** 1984. Retraction of "A precise termination site in the mouse β-major globin transcription unit." Proc. Natl. Acad. Sci. U.S.A. **81**:2274.
54. **Salser, W.** 1978. Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. **42**:985–1002.
55. **Sanger, F., S. Nicklen, and A. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
56. **Shen, C.-K. J., and T. Maniatis.** 1980. Tissue-specific DNA methylation in a cluster of rabbit β-like globin genes. Proc. Natl. Acad. Sci. U.S.A. **77**:6634–6638.
57. **Shen, C.-K. J., and T. Maniatis.** 1980. The organization of repetitive sequences in a cluster of rabbit β-like globin genes. Cell **19**:379–391.
58. **Soeiro, R., and J. E. Darnell.** 1969. Competition hybridization by "pre-saturation" of HeLa cell DNA. J. Mol. Biol. **44**: 551–562.
59. **Sorenson, G.** 1963. Hepatic hematocytopoiesis in the fetal rabbit: a light and electron microscopic study. Ann. N.Y. Acad. Sci. **111**:44–69.
60. **Stunnenberg, H. G., and M. L. Birnstiel.** 1982. Bioassay for components regulating eukaryotic gene expression: a chromosomal factor involved in the generation of histone mRNA 3′ termini. Proc. Natl. Acad. Sci. U.S.A. **79**:6201–6204.
61. **Tinoco, I., Jr., P. Borer, B. Dengler, M. Levine, O. Uhlenbeck, D. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
62. **Traboni, C., G. Ciliberto, and R. Cortese.** 1982. A novel method for site-directed mutagenesis: its application to a eukaryotic $tRNA^{pro}$ gene promoter. EMBO J. **1**:415–420.
63. **Villeponteau, B., G. Landes, M. Pankratz, and H. Martinson.** 1982. The chicken β-globin gene region: delineation of transcription units and developmental regulation of interspersed DNA repeats. J. Biol. Chem. **257**:11015–11023.
64. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. **76**:3683–3687.
65. **Weaver, R., and C. Weissman.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. **7**:1175–1193.
66. **Weintraub, H., A. Larsen, and M. Groudine.** 1981. α-Globin gene switching during the development of chicken embryos: expression and chromosome structure. Cell **24**:333–344.
67. **Wierenga, R., J. Haizinga, W. Gaastra, G. Welling, and J. Beintema.** 1973. Affinity chromatography of porcine pancreatic ribonuclease and reinvestigation of the N-terminal amino acid sequence. FEBS Lett. **31**:181–185.
68. **Wold, B., M. Wigler, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel.** 1979. Introduction and expression of a rabbit β-globin gene in mouse fibroblasts. Proc. Natl. Acad. Sci. U.S.A. **76**:5684–5688.
69. **Yanofsky, C.** 1981. Attenuation in the control of expression of bacterial operons. Nature (London) **289**:751–758.
70. **Zweig, S. E.** 1984. Analysis of large nucleic acid dot matrices on small computers. Nucleic Acids Res. **12**:767–776.