Volume 9 Number 17 1981 Nucleic Acids Research

Transcription of the β-like globin genes and pseudogenes of the goat in a cell-free system

D.S.Luse[1], J.R.Haynes[2], D.VanLeeuwen[2], E.A.Schon[2], M.L.Cleary[2], S.G.Shapiro[2], J.B.Lingrel[2] and R.G.Roeder[1]

[1]Department of Biological Chemistry, Washington University School of Medicine, St. Louis, MO 63110, and [2]Department of Biological Chemistry, University of Cincinnati College of Medicine, 231 Bethesda Avenue, Cincinnati, OH 45267, USA

Received 30 April 1981

ABSTRACT

We have examined the transcription in a cell-free system of all the members of a highly regulated gene family, the β-like hemoglobin genes of the goat. These five genes, which code for embryonic, fetal, juvenile, and adult β-globin proteins, are all transcribed to roughly the same extent in the in vitro system. In all cases initiation of transcription is accurate. However, two goat β-globin pseudogenes, as well as several artificially constructed deletion mutants, are not transcribed in vitro. A common feature of the transcriptionally inactive genes is the lack of an AT-rich consensus sequence just upstream of the presumptive initiation site.

INTRODUCTION

An understanding of the mechanisms which control transcription will be fundamental in explaining the overall process of gene regulation. One promising approach to this problem involves the reconstitution of biologically meaningful transcription events using purified components. We have recently described a transcription system, consisting of a cultured cell extract (S100) and purified RNA polymerase II, which will initiate transcription on cloned DNA templates at the same sites used for initiation in vivo (1,2). Another in vitro transcription system has subsequently been described, consisting of whole cell extracts independent of exogenous RNA polymerase II (3). Both viral (1,3-6) and chromosomal (2,4,7-9) genes are accurately transcribed by these methods.

Given the availability of soluble factors which will mediate accurate transcription initiation in vitro, two kinds of questions may be immediately posed: first, do the transcription factors in the extracts discriminate among promoters in a way that reflects in vivo regulatory events, and second, what are the specific DNA sequences near the site of transcription initiation which are required for promoter recognition? In order to begin to address these questions, we have studied the transcription of the entire family of β-like globin genes of the goat in the S100 system. This gene family has been described in detail previously (10-13); it consists five developmentally regulated sequences: ε^{I} and ε^{II} (presumptive embryonic), γ

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

(fetal), β^C (juvenile), and β^A (adult). There are also two β-homologous sequences, $\psi\beta^X$ and $\psi\beta^Z$, which cannot code for globin-like proteins and are therefore termed pseudogenes. We report here that all five of the β-like goat globin genes are transcribed accurately and to approximately the same extent in the S100 system. However, the two pseudogenes, which lack the AT-rich consensus sequence (the 'TATA box'; see ref. 7) upstream of the expected initiation site, are not transcribed in vitro. To further investigate the importance of the consensus sequence, we have constructed and transcribed several deletion and rearrangement mutants of the goat globin genes; only those mutants which retain the TATA box are transcribed in vitro. During the course of this research, other laboratories reported the accurate in vitro transcription of human α and β globin genes (8) and mouse α globin genes (9); in these studies as well, template activity in vitro was dependent on the presence of the TATA sequence.

MATERIALS AND METHODS

I. Templates

The goat DNA fragments in Figure 1 were isolated from the previously described large genomic fragments cloned in Charon 4A (10,12,13). Each DNA was then subcloned by insertion into pBR322 which had been previously cleaved with the same pair of restriction enzymes used to excise that fragment. (The $\psi\beta^Z$ fragment subcloned is 2.7 kb long and is bounded by Eco R1 and Pst I sites; only 2/3 of this fragment is shown in Figure 1.)

The mutant DNAs were constructed as follows. For β^A del #1, the β^A fragment in Figure 1 was cleaved with Pvu II at the site at position -10 (see Figure 4); Hind III linkers were added and the resulting Hind III/Bam H1 fragment bearing the β^A coding region was recloned into Hind III/Bam H1-cleaved pBR322. For γ del #1, the γ fragment in Figure 1 was cleaved with Hae III at the site at position -13 (see Figure 4; note that this was a partial digest, since the fragment contains several Hae III sites); Eco R1 linkers were added and the resulting Eco R1/Bam H1 fragment bearing the γ coding region was recloned into Eco R1/Bam H1-cleaved pBR322. γ del #2 was produced from the γ fragment in Figure 1 in the same way as γ del #1, except that the site of Hae III cleavage was at position -83 instead of position -13 (see Figure 4). The recombinant plasmid, $\psi\beta^X/\beta^A$, was prepared from the $\psi\beta^X$ and β^A fragments in Figure 1. Each fragment was cut by partial Hae III digestion at a site present in both fragments at position -13 (see Figure 4); the $\psi\beta^X$ fragment containing sequences upstream of the site and the β^A fragment containing sequences downstream of the site were joined by blunt-end ligation and recloned into Eco R1/Bam H1-cleaved pBR322.

II. Preparation of cultured cell extracts

Almost all of the in vitro transcriptions reported here were done with S100 extracts prepared exactly as previously described (1,2). In recent experiments we have changed the extraction method slightly, which results in more reproducible extraction of transcription factors (J.D. Dignam, R. Lebovitz, and R.G. Roeder, manuscript in preparation). Briefly, the new method involves hypotonic lysis of the cultured cells as before, followed by low speed centrifugation to yield a crude nuclear pellet. The nuclei are then extracted with 0.35 M NaCl and the resulting extract is centrifuged at 100,000 x g and concentrated by ammonium sulfate precipitation. The only results involving the use of these extracts presented in this paper are lanes 7-11 of Figure 5. Neither transcriptional fidelity nor relative promoter strength is changed by the use of these extracts.

III. In vitro transcription

RNA was synthesized in S100 extracts as described previously (1,2), except that calf thymus RNA polymerase II was used in place of human RNA polymerase II. The calf polymerase was prepared basically by the method of Hodo and Blatti (14); an elution from heparin-agarose was substituted for the final agarose gel filtration step. The RNA polymerase was applied to heparin-agarose at 0.17 M $(NH_4)_2SO_4$ and eluted at 0.6 M $(NH_4)_2SO_4$, in the chromatography buffer D of Hodo and Blatti (14).

IV. Primer extension analysis

This is based generally on the procedure of Proudfoot et al. (8). Unlabeled run-off RNA was synthesized in vitro in a final volume of 250-500 µl (5-10 times the standard reaction size). The RNAs were purified as described (1,2) and electrophoresed on polyacrylamide gels. The unlabeled run-off RNAs were located by running identical radioactive RNAs in adjacent lanes; the unlabeled transcripts were excised from the gel and purified by the methods used previously to prepare RNAs for fingerprint analysis (1,2). DNA primer fragments (see Figure 3) were isolated by electroelution from polyacrylamide gels, treated with bacterial alkaline phosphatase, and 5' end labeled with γ-^{32}P ATP and T4 polynucleotide kinase. The RNA to be analyzed (either the in vitro transcript from 5-10 standard 50 µl reactions, or 0.1 µg of goat 9S globin mRNA) and the appropriate primer DNA fragment (10-20,000 cpm) were dissolved in 10 µl of 0.1 M NaCl, 20 mM Tris, pH 7.9, 0.1 mM EDTA, and sealed in glass capillaries. The mixtures were denatured at 100 °C for 2 min, incubated at 60 °C for 5-10 h, and then diluted into an equal volume of 80 mM Tris, pH 7.9, 10 mM $MgCl_2$, 4 mM dithiothreitol, 400 µM each dATP, dCTP, dGTP, and dTTP, and 5 units of AMV reverse transcriptase. After 5 min on ice, the reverse transcriptions were incubated at 37 °C for 30 min and then

extracted with phenol and chloroform.

RESULTS

A schematic representation of the β-like goat globin templates is given in Figure 1. Prior to in vitro transcription, each DNA is cleaved by a restriction endonuclease at a site downstream of the expected initiation site. The lengths of the run-off RNAs which would be produced as the result of accurate initiation on each of the globin templates are indicated in the figure. The cleaved templates were transcribed by RNA polymerase II in the presence of human cultured cell extracts essentially as described previously (2; see Materials and Methods). The RNAs synthesized in vitro were purified by extraction with phenol and chloroform, denatured with glyoxal and electrophoresed on polyacrylamide gels.

An autoradiogram of the transcripts produced in vitro from the five globin genes is shown in Figure 2. Lanes 1-3 contain in vitro-synthesized RNAs whose

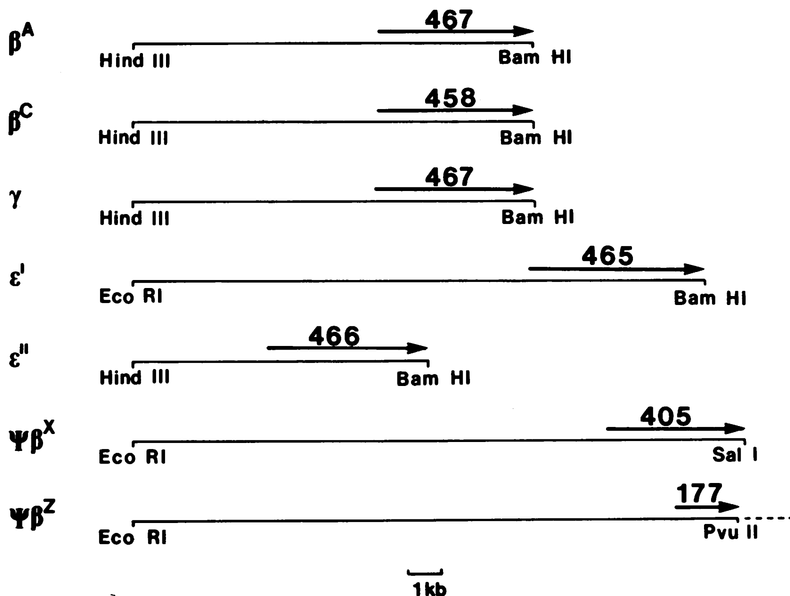


Figure 1. For each goat globin gene or psuedogene, the goat DNA framgment diagrammed above was subcloned into pBR322 as indicated in Materials and Methods. (For $\psi\beta^Z$, which was subcloned as a 2.7 kb Eco R1-Pst I fragment, the entire stretch of goat DNA is not shown.) Before transcription all of the recombinant plasmids were cleaved downstream of the expected initiation sites with the indicated restriction enzymes. The location of the anticipated initiation site and the length of the run-off RNA produced by initiation at that site are indicated for each template.

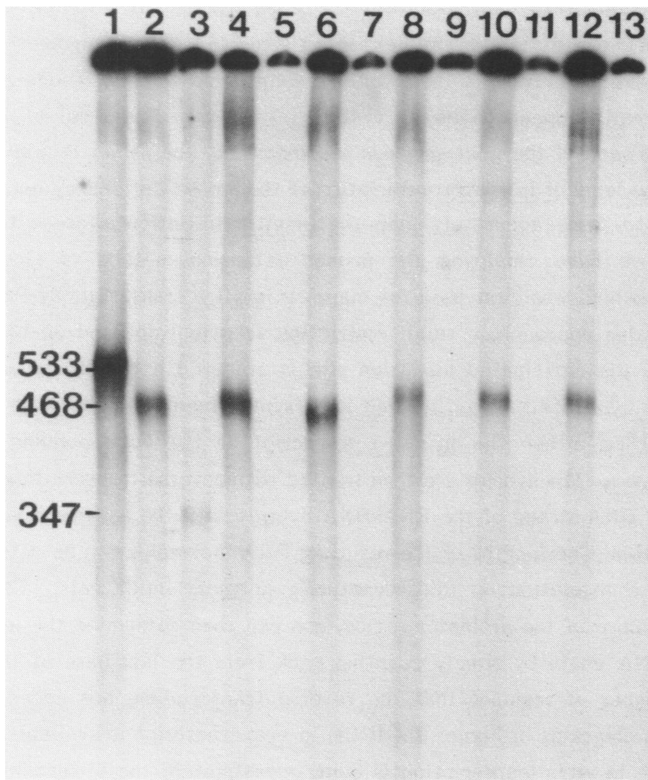


Figure 2. Gel electrophoresis of in vitro transcripts of the goat globin genes. Lanes 1-3 contain size markers, consisting of RNAs of known length synthesized in vitro from three different promoters: lane 1, Ad2 major late, 533 bases (reference 1); lane 2, mouse β-globin major, 468 bases (reference 2); lane 3, Ad2 early region Ib, 347 bases (D. Lee and R.G. Roeder, in press). Lanes 4-13 contain in vitro transcripts of the five goat β-like globin gene subclones diagrammed in Figure 1; each DNA was cleaved with Bam H1 and was expected to yield a run-off transcript of about 465 bases (see Figure 1). The templates used were: lanes 4 and 5, β^A; lanes 6 and 7, β^C; lanes 8 and 9, γ; lanes 10 and 11, ε^I; lanes 12 and 13, ε^{II}. In the right-hand lane of each pair (lanes 3, 5, 7, 9, and 11), in vitro transcription was performed in the presence of 0.5 μg/ml of α-amanitin.

sizes are known exactly: lane 1, adenovirus 2 major late transcript, 533 bases; lane 2, mouse β-globin major transcript, 468 bases; lane 3, adenovirus 2 early region 1b transcript, 347 bases. It can be seen that all five of the goat genes are transcribed in the in vitro system (lanes 4, 6, 8, 10, 12); these transcripts are the product of RNA polymerase II, since they are not produced in the presence of 0.5 μg/ml α-amanitin (lanes 5, 7, 9, 11, 13). The four transcripts (β^A, γ, ε^I, and ε^{II}) whose

sizes were expected to be approximately 465 bases co-migrate with the 468 base marker RNA and the β^C transcript (expected size, 458 bases) migrates slightly faster than the marker. Thus, to the level of resolution possible with such gels, the in vitro transcripts appear to be accurately initiated. The level of in vitro transcription of each of the goat genes is approximately the same; this level is in turn equal to the level of in vitro transcription of the mouse β-globin gene (lane 2).

In order to more accurately map the in vitro initiation sites on the goat globin genes, we have employed the primer extension method of Ghosh and colleagues (15) which was first used for mapping in vitro transcripts by Proudfoot et al. (8). In this approach, a small restriction fragment located 30-100 bases downstream of the anticipated initiation site is prepared and end-labeled. The labeled DNA is hybridized to non-labeled RNA synthesized from the same region; this RNA may be either the in vitro transcript or the corresponding mRNA synthesized in vivo. The hybrids are then treated with reverse transcriptase, which will extend the DNA strand of the RNA-DNA hybrids until the enzyme reaches the 5' end of the RNA. The lengths of the resulting DNA molecules can be determined with single base resolution on thin, denaturing polyacrylamide gels. Using the nucleotide sequence of the promoter region, one can then determine the initiating base of the RNA chain by simply counting back from the last base of the DNA primer the number of residues that the reverse transcriptase has added to the primer (see the diagrams in Figure 3). If the in vivo transcript is available for the promoter whose in vitro transcription is being investigated, the 5' termini of the two RNAs can thus be directly compared. (In this paper we make the assumption that the cap site and the initiation site are equivalent, and we use the terms interchangeably. In those cases where the question has been examined, there is no evidence of transcription initiating upstream of the cap site, either in vivo [16,17] or in a reconstituted cell-free system [18] . However, the formal possibility of extremely rapid degradation and/or transfer of upstream nucleotides has not been ruled out.)

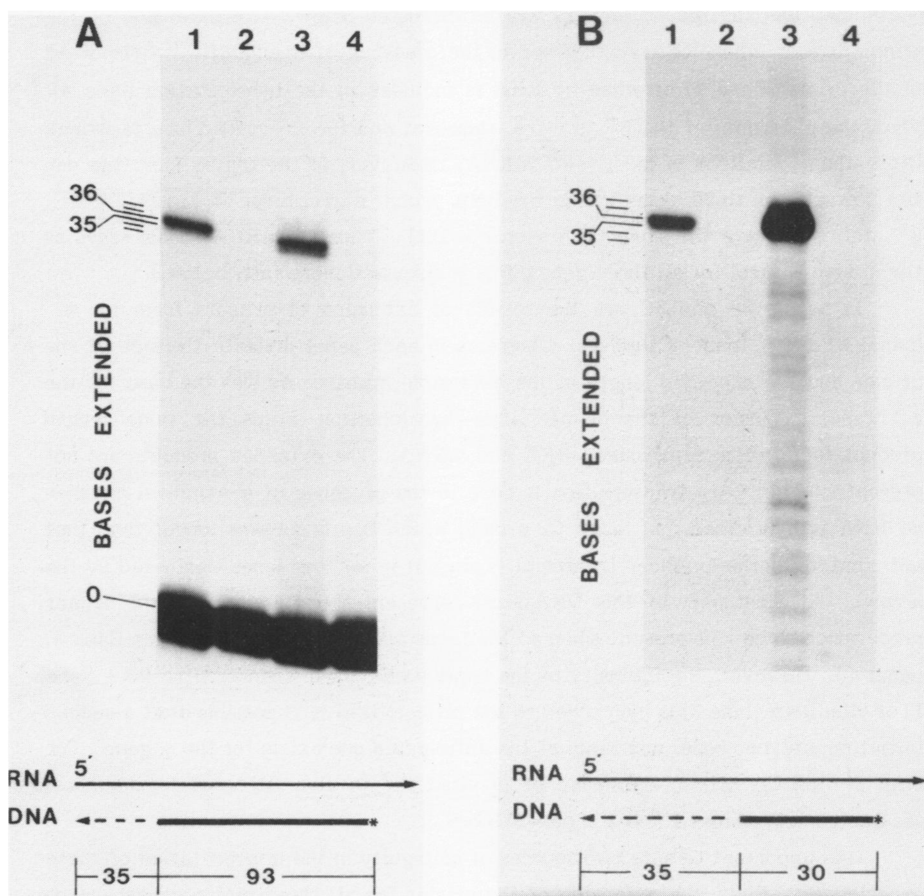
The results of such primer extension experiments on transcripts of the β^A, β^C, and γ globin genes are shown in Figure 3. As outlined in the diagram in panel A, a 93 base pair end-labeled DNA primer was hybridized to either the in vitro transcript of the β^A gene or to 9S globin mRNA and then extended with reverse transcriptase. Hybridization to the 9S mRNA should allow extension of the primer by 35 bases. The hybrid of primer and in vitro transcript will be extended by the same amount if the two RNAs initiate at the same position. A comparison of lanes 1 and 3 of panel A shows that the two extended primers have the identical length. (Extension of both primers actually results in two products differing in length by

one nucleotide; the reason for this will be discussed below.) No extension of the primer is seen when RNA synthesized in the presence of α-amanitin is hybridized to the primer (lane 2) or when no RNA is included in the hybridization (lane 4). Thus, the 5' termini of the β^A in vitro transcript and the β^A mRNA are identical. (Note that 9S mRNA is mostly β^A mRNA. However, in the region from the cap site downstream to 20 bases within the DNA primer diagrammed in Figure 3A, β^A, β^C, and γ all have the same DNA sequence [11]. Thus, 9S mRNA should serve as the in vivo control for all three genes; this is discussed more fully below.)

In panels 3B and 3C are the results of extension of primers from the β^C (panel B) and γ (panel C) genes. Diagrams in each panel indicate the size of the primer and the expected length of the extension product. As was the case for the β^A gene, extension of the primer after hybridization yields the same length product for in vitro synthesized RNA and mRNA. The extended products are not present when in vitro transcription is done in the presence of α-amanitin or when no RNA is hybridized. In panel C, a band which is a few bases longer than that anticipated for the extended transcript is present in all four lanes (indicated by the arrow). At least part of this DNA must represent a contaminant in the primer preparation since it is present when no RNA was added to the hybridization (lane 4, panel C). However, the intensity of the band is increased when in vitro RNA (lane 1) or 9S mRNA (lane 3) is hybridized to the primer; thus it is possible that a second initiation site two bases upstream of the anticipated one exists for the γ gene. For both β^C and γ, primer extension of in vitro and in vivo RNAs yields identical results (compare lanes 1 and 3 in panels B or C).

It is important to note two sources of ambiguity in the interpretation of these experiments. First, the extension of the primer for all three globin genes always results in a pair of extension products differing in length by one base. This is true for both in vitro and in vivo RNAs (see lanes 1 and 3 in panels A, B, and C). This result has been consistently observed in our hands, not only for the globin genes but also for primer extension experiments with the adenovirus 2 major late promoter (D. Lee and R.G. Roeder, in press); it is known from analysis of 5' terminal oligonucleotides that both in vivo (16) and in vitro (1) transcripts from this promoter have single 5' ends. Thus, we suspect that the doublets are artifacts of the reverse transcription reaction. However, the possibility that initiation on the goat globin genes actually occurs at two positions cannot be excluded. This point does not affect our main conclusion concerning the fidelity of in vitro transcription, since the in vitro and in vivo 5' ends are clearly identical.

It should also be noted that we have not directly mapped the 5' termini of the β^C and γ mRNAs but have only inferred the location of these termini from



sequence homology with the β^A gene. The extension of the β^A mRNA in panel A of Figure 3 places the 5' end of this RNA within one base of the position noted in Figure 4. It has previously been determined (11) that the $\beta^C$and γ genes have the same nucleotide sequence as the β^A gene in a region extending from the TATA box downstream to a point 20 bases within the primer diagrammed in Figure 3. It thus seemed reasonable to assume that initiation would occur at the same site in these genes as in β^A, although we could not prove this directly since we could not obtain pure β^C or γ mRNA. Given this assumption, β^A mRNA could be used as the in vivo control for the β^C and γ primer extensions, since it will hybridize to the β^C and γ primers in exactly the same region as the homologous mRNAs would. The results in panels B and C of Figure 3 demonstrate that in vitro transcripts of the β^C and γ genes do have the expected 5' ends. It remains a formal possibility that

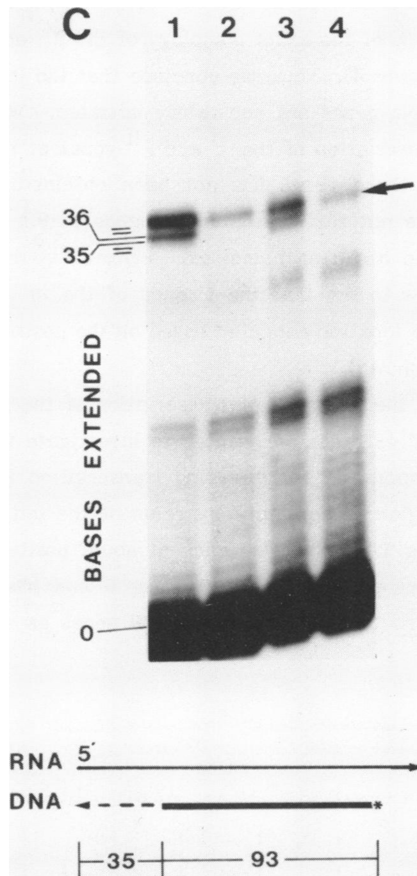


Figure 3. Unlabeled in vitro transcripts from the β^A (panel A), β^C (panel B), or γ (panel C) globin genes were hybridized to short, 5'-end labeled DNA fragments and extended by reverse transcriptase as described in the text and in Materials and Methods. In the diagram in the lower part of each panel, the heavy line indicates the DNA primer (the asterisk indicates the ^{32}P-label) and the dashed line indicates the bases added to the primer by reverse transcriptase; the lengths of the primer and of the extension (assuming accurate intiation of the RNA) are shown below the diagram. The upper part of each panel displays the results of gel electrophoresis of primers extended after hybridization to: lane 1, RNA synthesized in vitro; lane 2, RNA synthesized in vitro in the presence of 0.5 μg/ml α-amanitin; lane 3, 9S goat globin mRNA; lane 4, no RNA. In lanes adjacent to those displayed, size standards (consisting of end-labeled DNA fragments subjected to partial, base-specific cleavage by the Maxam and Gilbert technique) were run, in order to determine the number of bases that the primers were extended. The number of bases added is indicated in the left margin of each panel.

transcription in vivo may initiate at another site on these genes, but this is extremely unlikely in view of the exact homology of the β^C and γ promoters with the β^A promoter (see Figure 4). Thus, we conclude that the in vitro transcripts of the β^A, β^C, and γ globin genes are accurately initiated. We cannot assess the accuracy of in vitro transcription of the ε^I and ε^{II} genes at the single nucleotide level since mRNA for these genes has not been obtained and the nucleotide sequence of the genes is not sufficiently homologous to β^A to permit an exact assignment of the 5' end based on homology. Primer extension experiments on these genes do permit us to say that the 5' ends of the in vitro transcripts are within three bases of the location expected based on the position of the TATA box in these genes (data not shown).

Having established the accuracy of transcription of the goat globin genes in the presence of the S100 extracts, we wanted to investigate which regions of the DNA sequence were responsible for directing transcription initiation. We have concentrated on the region immediately upstream of the initiation site, since as previously mentioned the TATA box sequence at about position -25 is present in almost all genes transcribed by RNA polymerase II and has been found to be important for the in vitro transcription of class II genes as well (7-9,26-28). Of

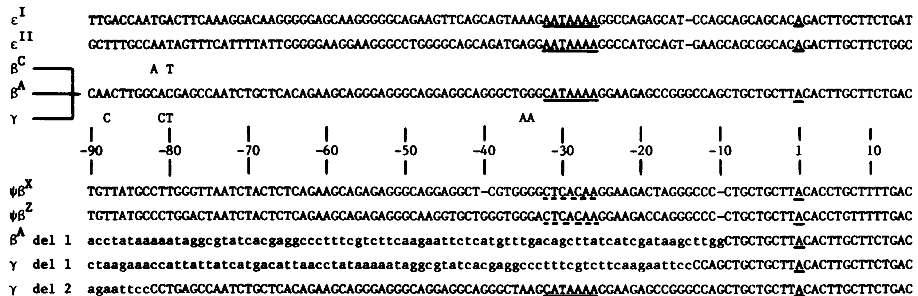


Figure 4. The DNA sequence of the promoter region is shown for the genes used in this study. Since β^A, β^C, and γ have almost the identical sequence in this region, only the β^A sequence is given in full, with the changes in β^C and γ noted above and below the β^A line. The TATA box region is underlined in those genes which have this feature; in the pseudogenes the altered TATA sequence is indicated by dashed underlining. The initiating nucleotide is known only for the $\beta^A/\beta^C/\gamma$ triplet; in the other genes it is assigned by homology with β^A. In the deletion mutants, lower case letters indicate vector (pBR322) DNA. The $\psi\beta^X/\beta^A$ recombinant is not shown; it consists simply of the $\psi\beta^X$ sequence upstream of the Hae III site (GGCC) at position -13 and of the β^A sequence downstream of this site. Sequences for β^A, β^C, γ, and $\psi\beta^X$ have been published (11,12). Sequences for the deletion mutants were not directly determined but inferred from the method of construction and the pBR322 sequence of Sutcliffe (19); other sequence information is from our unpublished observations.

particular interest are the two β-homologous pseudogenes, $\psi\beta^X$ and $\psi\beta^Z$. As noted in Figure 4, both $\psi\beta^X$ and $\psi\beta^Z$ retain a great deal of homology with β^A, β^C, and γ in the vicinity of the initiation site. If β^A and $\psi\beta^X$ are compared, 60 out of the 80 residues preceding the initiation site in β^A are present in $\psi\beta^X$ (see Figure 4). Three of the base changes are in the TATA sequence, leaving $\psi\beta^X$ without a recognizable TATA box. $\psi\beta^Z$ shares a similar overall homology with β^A and is similarly mutated in the TATA box (Figure 4). The general homology of the pseudogene "promoters" with the promoters of transcribed β-like genes, combined with the loss of the TATA box, makes transcription of the psuedogenes an interesting test of the importance of the consensus sequence.

Recombinants containing $\psi\beta^X$ or $\psi\beta^Z$ were cleaved with restriction enzymes as indicated in Figure 1. When these cleaved DNAs were incubated with RNA polymerase II in S100 extracts, no transcripts of the size expected for initiation at the promoter-homologous region of either gene were detected. Parallel incubations with other globin genes indicated that the extracts and RNA polymerase used were active. The results of a typical transcription of $\psi\beta^X$ and $\psi\beta^Z$ are shown in Figure 5, lanes 4 and 5. The expected positions for correctly initiated run-off RNAs may be judged from the size markes in lane 1 (mouse β-globin, 468 bases) and 6 (Hind III-cleaved Ad2 major late, 197 bases).

The inactivity of the $\psi\beta^X$ and $\psi\beta^Z$ templates is consistent with the lack of a TATA sequence. However, since there are numerous other changes in these genes relative to the genes known to be transcriptionally active, we cannot attribute loss of activity directly to the loss of the consensus sequence. We have therefore constructed several deletion and rearrangement mutants, based on the transcriptionally active β^A, β^C, and γ genes in order to localize the DNA region important for transcription. (Sequences of the mutants in the promoter region are given in Figure 4; details of mutant construction are in Materials and Methods.) β^A del #1 and γ del #1 were produced by cutting the genes between the TATA box and the initiation site, discarding the upstream goat DNA, and ligating vector DNA in its place. These mutants thus lack the TATA sequence but retain the native cap sites and coding regions. A recombinant, $\psi\beta^X/\beta^A$, was created between $\psi\beta^X$ and β^A by use of a common restriction site at position -13; $\psi\beta^X/\beta^A$ has the $\psi\beta^X$ sequence from position -13 upstream and the β^A sequence from position -13 downstream. It therefore also lacks the TATA sequence but retains overall homology (to about position -80) with the transcribed gene; it retains the native cap site and coding region. The final recombinant, γ del #2, is derived from the γ gene by cleavage at a restriction site 50 bases upstream of the TATA box. The goat DNA upstream of this site is discarded and vector DNA is ligated in its place. Transcription of γ del

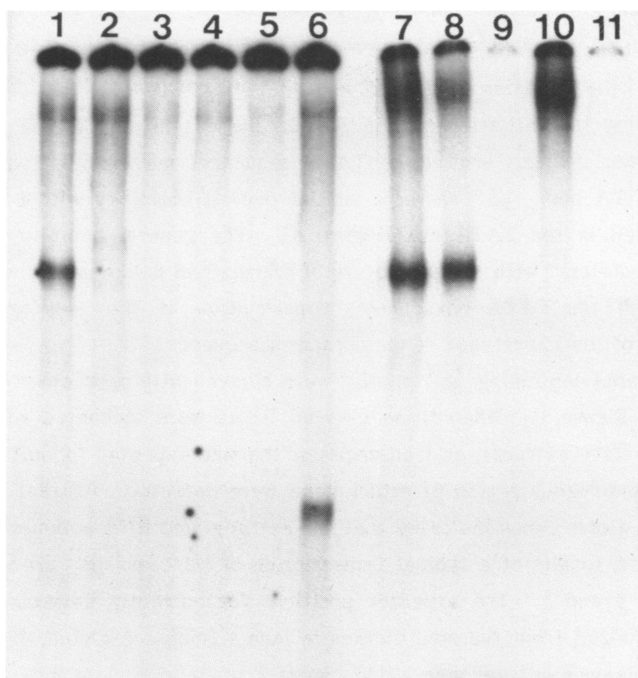


Figure 5. In vitro transcription of pseudogenes and artificially mutated genes was performed as indicated in Materials and Methods; lanes 1-6 and 7-11 were separate experiments using different batches of cultured cell extract. RNAs were synthesized in vitro from the following promoters for use as size markers: lane 1, mouse β-globin major, 468 bases (reference 2); lane 6, adenovirus 2 major late, 197 bases (reference 1); lane 7, goat γ-globin, 467 bases (see Figure 2). Templates for the other lanes, with the expected size for a correctly initiated "run-off transcript" (see text and Figure 1), are as follows: lane 2, $\psi\beta^X/\beta^A$, cleaved with Bam H1, 467 bases; lane 3, β^A del #1, cleaved with Bam H1, 467 bases; lane 4, $\psi\beta^X$ cleaved with Sal I, 405 bases; lane 5, $\psi\beta^Z$, cleaved with Pvu II, 177 bases; lane 8, γ del #2, cleaved with Bam H1, 467 bases; lane 10, γ del #1, cleaved with Bam H1, 467 bases. Lane 9 is the same reaction as lane 8, and lane 11 is the same reaction as lane 10, except that the reactions for lanes 9 and 11 contained 0.5 μg/ml α-amanitin.

#2 will thus test whether the functional in vitro promoter extends a significant distance upstream of the TATA box.

All of the mutants retain the Bam H1 restriction site 467 bases downstream of the putative initiation site; each DNA was cleaved at this site and incubated with RNA polymerase II and cultured cell extracts as usual. As can be seen in lanes 2, 3, and 8-11 of Figure 5, only γ del #2, which retains the TATA sequence, is transcribed. The amount of transcript synthesized from γ del #2 is apparently the

same as the amount produced from the parental gene, γ (compare lanes 7 and 8). The mutants which have lost the TATA box, β^A del #1 and γ del #1, and the mutant with the defective TATA box, $\psi\beta^X/\beta^A$, are not transcribed in the in vitro system.

DISCUSSION

In this report, we extend our studies of in vitro transcription of hemoglobin genes (2) to the entire family of β-like globin genes of the goat and make the following points: (i) Transcription initiation on the five goat globin genes is accurate; on the β^A, β^C, and γ genes, where the in vivo initiation sites are known or can be inferred, initiation occurs at the same base in vivo and in vitro. (ii) The extent of transcription of all five of the goat β-globin genes in the soluble system is approximately the same. (iii) The presence of a functional promoter on these globin genes is correlated with the presence of an intact TATA sequence approximately 25 bases upstream of the initiation site, as assayed by the use of both naturally occurring pseudogenes and artificially created deletion mutants.

The goat β-like genes are normally expressed at four different developmental stages (ϵ^I and ϵ^{II} in embryos, γ in the fetus, β^C in early postnatal development, and β^A in the adult), and in different tissues (ϵ^I and ϵ^{II} in blood islands, γ in the fetal liver, and β^A and β^C in the bone marrow; see ref. 13). It is therefore significant that five members of a highly regulated, differentially expressed gene family are all transcribed to about the same extent in cultured cell extracts. Other laboratories have also reported the faithful transcription of regulated cellular genes in soluble cell-free transcription systems analogous to the one employed here. These genes include: the mouse α globin genes (9), human embryonic, fetal and adult α and β globin genes (8), and the steroid-responsive ovalbumin and conalbumin genes of the chicken (4,7). In no case has initiation of transcription in vitro on a chromosomal gene been dependent on tissue or stage-specific factors. These observations confirm our previous suggestion (2) that the components present in the S100s (and related extracts) are general transcription factors, required for initiation at all class II promoters. Since the templates used for in vitro transcription are purified DNAs and not the native nucleoprotein structures, we cannot address the question of what role chromosomal components might play in modulating transcription. However, one model consistent with both our results and recent data on the conformation of active genes (20-22) would involve a general repression of transcription in chromatin. Gene activation would then be dependent on a modification of chromatin structure in the region of the gene to be transcribed. These changes would be necessary for access to the

promoter by the RNA polymerase and general transcription factors. Since we have eliminated potential differences in promoter accessibility by our use of deproteinized DNA templates, it is not surprising that we see little evidence of transcriptional regulation using the *in vitro* systems.

In view of the accuracy and broad applicability of the S100 system in transcribing class II genes, it is important to ask what features of the DNA sequence are common to the recognized promoter regions. It was considered very likely that the TATA box sequence would be a part of class II promoters since this sequence is found about 25 bases upstream of the initiation site of almost all genes transcribed by RNA polymerase II (see 7 and references therein). The five β-like goat globin genes all possess this sequence (Figure 4) and are all transcribed *in vitro* (Figure 2). However, as shown in Figure 5, we are unable to obtain transcription in the S100 system of genes which lack an intact TATA box. The presence of a native initiation site and coding region did not prove sufficient to yield an active promoter, in contrast to the case of genes transcribed *in vivo* by RNA polymerase III (23-25). Deletion of goat DNA upstream of position -80 did not reduce promoter activity in the one mutant of this type that we tested. We have also constructed mutants of β^A, β^C, and γ which retain the TATA sequence but lack the eukaryotic cap site and coding region. All of these mutants are transcribed *in vitro*, although most of them are much less efficient as templates than the parental genes (our unpublished observations). In all of our experiments, the presence of the TATA sequence is absolutely correlated with template activity; this aruges that the TATA box is, in fact, required for transcription in the S100 system.

Our results are thus in accord with more systematic studies of *in vitro* promoter requirements done in other laboratories (7,26-29). The most complete analysis of this problem has been performed by Chambon and coworkers. This group demonstrated that run-off transcripts were greatly reduced or abolished for both deletion mutants which lacked the TATA box (7) and for a mutant with a single base change in the TATA sequence (26). Among the regions which could be deleted without eliminating the *in vitro* promoter are sequences upstream of the TATA box, including the CAAT consensus sequence which appears to about position -75 in many eukaryotic class II transcription units (30). This is important to note because all five of the goat β-like globin genes, but none of the non-transcribed mutants, have a CAAT sequence somewhere between position -75 and -85 (Figure 4).

Recent results have emphasized that the functional class II promoter, both *in vivo* and *in vitro*, is more complex than the simple TATA sequence. In the case of

either SV40 early genes (31,32) or the sea urchin H2A gene microinjected into Xenopus oocytes (33,34), deletion of the TATA sequence does not abolish transcription, although new 5' termini do appear. In both of these systems (32,34), sequences upstream of the TATA box seem to be required for promoter function; upstream sequences are also implicated in promoter function in rabbit β-globin genes introduced by transformation into mouse L cells (35). Further, Mathis and Chambon (29) have shown that in vitro transcription from the SV40 early promoter occurs at a level about equal to that observed in wild type in mutants from which the TATA box has been deleted; however, no run-off transcripts are detected for these mutants because initiation occurs at many different sites. Thus, the general function of the TATA sequence, both in vivo and in vitro, appears to be to fix the site of initiation relative to itself. At least in the case of SV40, alternative sequences can substitute for the TATA box and allow in vitro and in vivo transcription to occur.

The results discussed above indicate that in vitro transcription systems cannot yet duplicate all of the features of transcription in vivo. It should be emphasized, however, that initiation of transcription in vitro occurs with extreme fidelity. Further, as we have shown here, an entire family of highly regulated genes may be accurately transcribed in S100 extracts. This system now provides an excellent starting point for the assay of other components and structures presumed to be involved in transcription. As noted above, the use of nucleoprotein templates will probably be needed in order to more faithfully reconstitute natrual transcription events. It will also be necessary to prepare extracts from tissues actually expressing the genes of interest in order to search for specific regulatory components. In this regard, it is encouraging to note that Tjian and coworkers (6) have recently succeeded in repressing the in vitro transcription of the SV40 early promoter through the binding of T antigen to template DNA, as is presumed to occur in vivo. Although this system is not an exact model for the induction of transcription by the binding of a positive effector, it does illustrate the practicality of using in vitro transcription systems to study the molecular basis of gene regulation.

REFERENCES

1. Weil, P.A., Luse, D.S., Segall, J., and Roeder, R.G. (1979) Cell **18**, 469-484
2. Luse, D.S. and Roeder, R.G. (1980) Cell **20**, 691-699
3. Manley, J.L., Fire, A., Cano, A., Sharp, P.A., and Gefter, M.L. (1980) Proc. Natl. Acad. Sci. USA **77**, 3855-3859
4. Wasylyk, B., Kedinger, C., Corden, J., Brison, O., and Chambon, P. (1980) Nature **285**, 367-373
5. Green, M.R. and Roeder, R.G. (1980) Cell **22**, 231-242

6. Rio, D., Robbins, A., Myers, R., and Tjian, R. (1980) Proc. Natl. Acad. Sci. USA **77**, 5706-5710
7. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980) Science **209**, 1406-1414
8. Proudfoot, N.J., Shander, M.H.M., Manley, J.L., Gefter, M.L., and Maniatis, T. (1980) Science **209**, 1329-1336
9. Talkington, C.A., Nishioka, Y., and Leder, P. (1980) Proc Natl. Acad. Sci. USA **77**, 7132-7136
10. Haynes, J.R., Rosteck, P., Jr., Schon, E.A., Gallagher, P.M., Burks, D.J., Smith, K. and Lingrel, J.B. (1980) J. Biol. Chem. **255**, 6355-6367
11. Haynes, J.R., Rosteck, P., and Lingrel, J.B. (1980) Proc. Natl. Acad. Sci. USA **77**, 7127-7131
12. Cleary, M.L., Haynes, J.R., Schon, E.A., and Lingrel, J.B. (1980) Nucl. Acids Res. **8**, 4791-4802
13. Lingrel, J.B., Haynes, J.R., Schon, E.A., Cleary, M.L., and Gallagher, P.M. (1981) in Organization and Expression of Globin Genes, G. Stamatoyannopoulous and A. Nienhuis, eds. (Alan R. Liss, New York), 45-57.
14. Hodo, H.G. and Blatti, S.P. (1977) Biochemistry **16**, 2334-2345
15. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P., and Weissman, S.M. (1978) J. Mol Biol. **126**, 813-846
16. Ziff, E.B. and Evans, R.M. (1978) Cell **15**, 1463-1475.
17. Hofer, E. and Darnell, J.E. (1981) Cell **23**, 585-593
18. Hagenbuchle, O. and Schibler, U. (1981) Proc. Natl. Acad. Sci. USA **78**, 2283-2286
19. Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol. **43**, 77-90
20. Stalder, J., Larsen, A., Engel, J.D., Dolan, M., Groudine, M., and Weintraub, H. (1980) Cell **20**, 451-460
21. Wu, C. (1980) Nature **286**, 854-860
22. Samal, B., Worcel, A., Louis, C., and Schedl, P. (1981) Cell **23**, 401-410
23. Kressman, A., Hofstetter, H., Di Capua, E., Grosschedl, R., and Birnstiel, M.L. (1979) Nuc. Acids Res. **7**, 1749-1763
24. Sakonju, S., Bogenhagen, D.F., and Brown, D.D. (1980) Cell **19**, 13-25
25. Bogenhagen, D.F., Sakonju, S., and Brown, D.D. (1980) Cell **19**, 27-35
26. Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule, R., and Chambon, P. (1980) Proc. Natl. Acad. Sci. USA **77**, 7024-7028
27. Hu, S-L. and Manley, J.L. (1981) Proc. Natl. Acad. Sci. USA **78**, 820-824
28. Tsai, S.Y., Tsai, M-J., and O'Malley, B.W. (1981) Proc. Natl. Acad. Sci. USA **78**, 879-883
29. Mathis, D.J. and Chambon, P. (1981) Nature **290**, 310-315
30. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., De Riel, J.K., Forget, B.G., Weissman, S.M., Slighton, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., and Proudfoot, N.J. (1980) Cell **21**, 653-668
31. Ghosh, P.K., Lebowitz, D., Frisque, R.J., and Gluzman, Y. (1981) Proc. Natl. Acad. Sci. USA **78**, 100-104
32. Benoist, C. and Chambon, P. (1981) Nature **290**, 304-310
33. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA **77**, 1432-1436.
34. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA **77**, 7102-7106
35. Dierks, P., van Ooyen, A., Mantei, N., and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA **78**, 1411-1415