MOLECULAR AND CELLULAR BIOLOGY, Sept. 1993, p. 5439–5449
0270-7306/93/095439-11$02.00/0

Vol. 13, No. 9

Nuclear Protein-Binding Sites in a Transcriptional Control Region of the Rabbit α-Globin Gene

SUSAN E. YOST, BRIAN SHEWCHUK, AND ROSS HARDISON*

Department of Molecular and Cell Biology, The Center for Gene Regulation, The Pennsylvania State University, University Park, Pennsylvania 16802

Received 28 December 1992/Returned for modification 8 April 1993/Accepted 18 May 1993

The 5′-flanking and internal regions of the rabbit α-globin gene, which constitute a CpG island, are required for enhancer-independent expression in transfected cells. In this study, electrophoretic mobility shift assays revealed that a battery of nuclear proteins from both erythroid and nonerythroid cells bind specifically to these regulatory regions. Assays based on exonuclease III digestion, methylation interference, and DNase I footprinting identified sequences bound by proteins in crude nuclear extracts and by purified transcription factor Sp1. In the 5′ flank, recognition sites for the transcription factors α-IRP (positions −53 to −44 relative to the cap site), CP1 (−73 to −69), and Sp1 (−95 to −90) are bound by proteins in K562 cell nuclear extracts, as are three extended upstream regions. Two recognition sites for Sp1 in intron 1 are also bound both by proteins in crude nuclear extracts and by purified Sp1. The sequences CCAC in intron 2 and C_5 in the 3′-untranslated region also bind proteins. A major binding site found in exon 1, TATGGCGC, matches in sequence and methylation interference pattern the binding site for nuclear protein YY1, and binding is inhibited through competition by YY1-specific oligonucleotides. The protein-binding sites flanking and internal to the rabbit α-globin gene may form an extended promoter.

The α- and β-globin gene families are regulated in a tissue-specific and stage-specific manner to produce embryonic, fetal, and adult hemoglobins in erythrocytes (16). However, the coordinate expression of these two gene families is paradoxical. In both humans and rabbits, the two gene families are located on different chromosomes (17, 67, 68, 70) in very different sequence contexts. The α-globin genes are in a very dense, GC-rich isochore (6), are embedded in CpG islands (21, 29), and are unmethylated in all tissues examined (7). In contrast, the β-globin genes are in a very light, AT-rich isochore (6), are deficient in CpG dinucleotides (29), and are heavily methylated in nonexpressing tissues (58, 62).

These differences in genomic context are reflected in striking differences in the mechanisms regulating the two types of genes. In the absence of a viral enhancer, the human and rabbit α-globin genes are efficiently expressed when transfected into either erythroid or nonerythroid cell lines (11, 12, 33, 46). In contrast, the β-globin genes require either a viral enhancer for transient expression in nonerythroid HeLa cells (2, 33, 61) or chemical induction after stable integration into erythroid MEL cells (10, 65). Despite these differences, the α- and β-globin genes are coordinately expressed in equal amounts to produce a functional hemoglobin tetramer in normal erythrocytes. In fact, both types of human globin genes are appropriately regulated when their chromosomes are introduced into murine erythroid cells by cell fusion (17, 53, 64), indicating some higher level of control that does not function when the genes are removed from their normal chromosomal context. Prominent candidates for such higher-level control elements are the locus control regions (27, 31, 49).

The *cis*-acting sequences of the human and rabbit α-globin genes required for enhancer-independent expression in transfected cells map both 5′ and internal to the genes (9, 11, 34a, 69). Although the expression of the human α-globin gene has been studied extensively, protein-DNA interactions have been most extensively characterized for the promoter of the mouse α1-globin gene. In particular, transcription factor CP1 binds to a CCAAT box, and another nuclear protein, called α-IRP, binds to a GC-rich inverted repeat in the proximal 5′ flank (3, 15, 39, 40). Although this part of the 5′ flank is quite well conserved among mammalian α-globin genes, the binding site for erythroid cell-specific transcription factor GATA-1, located 5′ to the mouse α1 CCAAT box (4, 51), is not found in other mammalian α-globin genes (29; this paper). Also, the mouse α1-globin gene does not have a CpG island and in fact requires an enhancer for expression in transfected cells (63). Proteins binding in the CpG island could be important for the regulation of human and rabbit α-globin genes (63, 69).

Since both 5′-flanking and internal regions of the rabbit and human α-globin genes are important for expression and several of these regions differ dramatically in sequence from the well-characterized mouse α1-globin promoter, it is important to establish which segments are involved in binding nuclear proteins. This study reports the results of a series of binding and footprinting assays that identify protein-binding sites in the 5′-flanking and internal regions of the rabbit α-globin gene. These binding sites map in the regions implicated in regulating the expression of the α-globin gene and coincide with a CpG island. Hence, these binding sites, in the context of a CpG island, are likely to be involved in the regulated expression of the rabbit α-globin gene.

MATERIALS AND METHODS

Plasmid constructs for binding assays. Short restriction fragments spanning the α-globin gene were subcloned into pBluescript I KS(+) to generate probes for DNA-binding proteins (Fig. 1). Plasmid pBS+136 contains the 5′ flank of the rabbit α-globin gene from positions −173 (*Hin*fI) to −36 (*Ava*I) (numbering relative to the cap site, +1). Plasmid

* Corresponding author.

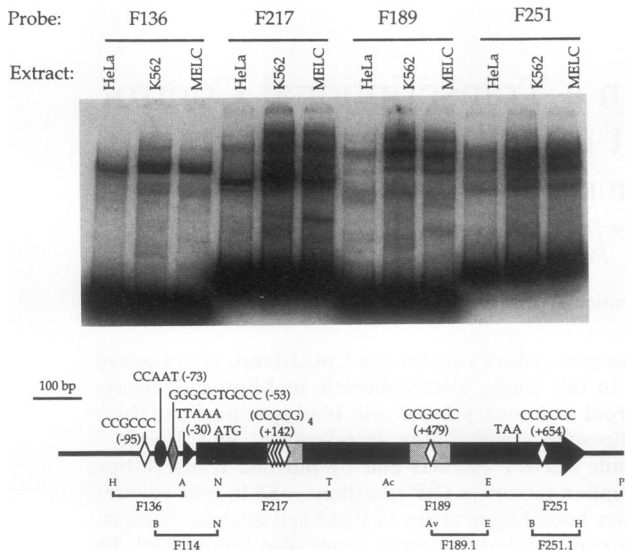

FIG. 1. EMSA with probes from the 5′-flanking and intragenic regions of the rabbit α-globin gene. Crude nuclear extracts prepared from HeLa, K562, and MEL cells, indicated above each lane, were assayed with the α-globin gene probes identified below the line drawing. The restriction sites used for cloning each fragment are indicated as follows: H, *Hin*fI; B, *Bgl*I; A, *Ava*I; N, *Nco*I; T, *Taq*I; Ac, *Acc*I; Av, *Ava*II; E, *Eco*RI; P, *Pvu*II. The recognition sites for several known transcription factors are also indicated (CCGCCC for Sp1, CCAAT for CP1, GGGCGTGCCC for αIRP, and TTAAA for the TATA-binding protein).

pBS+217 contains most of exon 1, all of intron 1, and part of intron 2 from +35 (*Nco*I) to +253 (*Taq*I). pBS+189 contains part of exon 2, all of intron 2, and part of exon 3 from +356 (*Acc*I) to +545 (*Eco*RI). pBS+251 contains most of exon 3 and the 3′ flank (to 84 bp past the polyadenylation site) from +545 (*Eco*RI) to +796 (*Pvu*II). The inserts were excised from each plasmid with *Bam*HI and *Eco*RI and labeled at the 3′ end with [α-^{32}P]dATP and Klenow DNA polymerase or labeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase. Fragment 114 extends from the *Bgl*I site at −83 in the 5′ flank to the *Nco*I site at +35 in exon 1. Fragment F189.1 (94 bp) was isolated by digesting pBS+189 with *Ava*II (+454) and *Eco*RI (+545), and fragment F251.1 (87 bp) was isolated by digesting pBS+251 with *Bgl*I (+640) and *Hin*fI (+727). Both labeled and unlabeled fragments were electroeluted from an 8% polyacrylamide gel for use as labeled probes or unlabeled specific competitors in electrophoretic mobility shift assays (EMSAs).

Preparation of crude nuclear extracts. Nuclear extracts of K562, HeLa, and MEL cells were prepared by the procedure of Dignam et al. (18) by Dounce homogenization with 0.42 M NaCl and dialysis with 20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonate (HEPES)–20% glycerol–0.2 mM EDTA–100 mM KCl. All procedures were performed at 4°C, and all buffers contained 0.5 mM dithiothreitol and the protease inhibitors 0.3 mM phenylmethylsulfonyl fluoride, 1 μM leupeptin, and 1 μM pepstatin A. The resulting extracts were adjusted to a protein concentration of 1 μg/μl. Purified Sp1 from HeLa cells was purchased from Promega.

EMSAs. Protein-DNA complexes were detected on the basis of alterations in electrophoretic mobility (23). Binding reaction mixtures (25-μl final volume) contained 1.0 ng of an end-labeled, double-stranded DNA fragment in binding buffer [10 mM HEPES (pH 7.9), 1 mM EDTA, 1 mM dithiothreitol, 80 mM KCl, 10% glycerol, 0.08 μg of poly(dI-dC) per μl, 0.3 μg of bovine serum albumin per μl]. Nuclear extract (9 μg) or purified Sp1 (1 "footprinting unit" [Promega]) was added, the mixture was incubated at room temperature for 15 min, and then samples were run on a nondenaturing 5% polyacrylamide gel with 45 mM Tris–45 mM borate–1 mM EDTA (pH 8.3) at 200 V (8 V/cm) for approximately 2.5 h. The gel was then dried and autoradiographed. For most competition experiments, the unlabeled competitor DNA and nuclear extract were added to binding buffer on ice, the labeled probe was immediately added, and the mixture was incubated at room temperature for 15 min (25, 36) prior to electrophoresis.

Exonuclease III digestion assay. The binding reaction for the exonuclease III digestion assay was the same as that for EMSAs, but with the addition of 6 mM $MgCl_2$. Blocks to exonucleolytic cleavage (66) were identified by digesting the protein-DNA complex with 3 μl of exonuclease III (80 U; New England Biolabs) for 10 min at 30°C. After the reaction was quenched with 35 μl of 10 mM EDTA–1% sodium dodecyl sulfate (SDS), proteins were removed by phenol extraction and ethanol precipitation. The exonuclease III-resistant DNA was dissolved in 8 μl of 80% formamide loading dye, denatured at 90°C for 5 min, and analyzed by electrophoresis on a denaturing 8% acrylamide–0.4% bisacrylamide–7 M urea gel at 48 V/cm for 1 h. Markers consisted of end-labeled fragments subjected to purine, G, or C cleavage reactions (45).

Methylation interference assay. Guanine nucleotides whose methylation prevents the binding of proteins (60) were identified by partial methylation of end-labeled DNA fragments with dimethyl sulfate (45), binding to nuclear extracts in a reaction mixture scaled up twofold to 50 μl, and separation of the bound and free DNA fragments by electrophoresis, as for EMSAs. The DNA fragments were electroeluted, ethanol precipitated, cleaved in 100 μl of 1 M piperidine at 90°C for 30 min, and analyzed on a denaturing polyacrylamide gel as described above for the exonuclease III digestion assay.

DNase I footprint assay. Regions of the labeled fragments protected by bound proteins from cleavage by DNase I (24) were identified by the method of Jackson et al. (34). An end-labeled, double-stranded DNA fragment (1.0 ng) was incubated at room temperature for 20 min with increasing amounts (0 to 45 μl) of a nuclear extract (or purified Sp1) in a final volume of 50 μl of binding buffer (as described above for EMSAs). The binding reaction mixture was then mixed with 50 μl of 5 mM $CaCl_2$–10 mM $MgCl_2$, and 5 μl of 10-ng/μl DNase I was added. Samples of unbound DNA (no added protein) were digested on ice for 90 s, and samples of bound DNA (with added protein) were digested at room temperature for 120 s. The reactions were quenched with 100 μl of 200 mM NaCl–20 mM EDTA–1% SDS–250 μg of yeast tRNA per μl, and the samples were extracted with phenol-chloroform, ethanol precipitated, and analyzed on a denaturing polyacrylamide gel as described above for the exonuclease III digestion assay.

Oligonucleotides used in binding assays. The complementary oligonucleotides containing two Sp1-binding sites (5′TCGATGGGCGGAGTTAGGGGCGGGACTA and 5′TCGATAGTCCCGCCCCTAACTCCGCCCA; binding sites are underlined) cover the sequence of sites 5 and 6 from the simian virus 40 early promoter (26). The αI1 oligonucleotides (5′CCGCCCCGCCCCGCCCCGCCCG and 5′CGGGCGGGGCGGGGCGGGGCGG) are from a sequence in the first

intron of the rabbit α-globin gene (13). The oligonucleotides containing one GATA-1-binding site (5'GATCCGGGCAAC TGATAAGGATTCCCA and 5'GATCTGGGAATCCTTAT CAGTTGCCCG) are from the mouse α-globin promoter (51). The oligonucleotides containing two AP-1- or NFE-2-binding sites (5'TCGACTCAAGCACAGCAATGCTGAGT CATGATGAGTCATGCTGAGGCTTA and 5'TCGATAAG CCTCAGAATGACTCATCATGACTCAGCATTGCTGT GCTTGAG) are from hypersensitive site 2 of the human α-globin gene locus control region (47). The oligonucleotides containing one CP1-binding site (5'TCGAGAAATTAAC CAATCAGCGCACTCTCACAGC and 5'TCGAGCTGTA GAGTGCGCTGATTGGTTAATTTC) are derived from the sequence of the human α-globin promoter (39). The oligonucleotides containing the YY1-binding site (5'AATTCGTTT TGCGACATTTGCGACACG and 5'AATTCGTGTCGCAA AATGTCGCAAAACG) are from the region near position −60 of the P5 promoter of adeno-associated virus (59).

RESULTS

Nuclear proteins bind specifically to sequences 5′ to and within the rabbit α-globin gene. The rabbit α-globin gene was divided into the fragments shown in the lower panel of Fig. 1 and tested for the ability to bind nuclear proteins in EMSAs (23). Fragment 136 contains upstream promoter elements that include consensus binding sites for Sp1 (37), CCAAT-binding proteins (39, 55), and α-IRP (3). Fragment 217 includes most of exon 1, all of intron 1, and part of exon 2. This fragment also includes a tandem repeat of $(CCCCG)_4$ that generates consensus binding sites for Sp1 (37). Fragment 189 contains all of intron 2 along with parts of exons 2 and 3, and fragment 251 contains most of exon 3 and the 3′ flank to 84 bp past the polyadenylation site. Both of these latter fragments contain matches to the consensus binding sites for Sp1 (Fig. 1).

Nuclear proteins from nonerythroid HeLa cells and from erythroid K562 and MEL cells bind to all four fragments of the rabbit α-globin gene (Fig. 1). The similarity in the patterns of protein-DNA complexes from all three sources indicates that few if any of the proteins binding to these gene fragments are erythroid cell specific. Competition with unlabeled DNA fragments and with oligonucleotides containing consensus binding sites (Fig. 2) indicated that most of the complexes are sequence specific. In other EMSAs (68a), the amount of the competitor was titrated from 50 to 400 ng for the unlabeled probe DNA fragments or from 25 to 200 ng for oligonucleotides. Each oligonucleotide that competed for the formation of a complex was effective at 25 ng, a quantity representing a 200-fold molar excess over the quantity of the labeled probe, and the unlabeled probe DNA fragments were effective at 50 to 100 ng (a 50- to 100-fold molar excess), as detailed below.

The probe from the 5′ flank (fragment F136) generated a prominent doublet in the EMSAs (Fig. 1 and 2A), showing that two major complexes (C1 and C2) are formed with K562 nuclear extracts. These two complexes are inhibited through competition by a 50-fold molar excess of an unlabeled fragment but not by nonspecific DNA or oligonucleotides containing binding sites for the transcription factors Sp1, AP-1, and GATA-1. Complex C1 (F136) is inhibited by 25 ng (a 200-fold molar excess) of CP1-specific oligonucleotides (Fig. 3), indicating that factor CP1 accounts for one of the major complexes formed with this probe.

The 5′ half of the α-globin gene (fragment F217) binds to proteins in K562 nuclear extracts to form several complexes

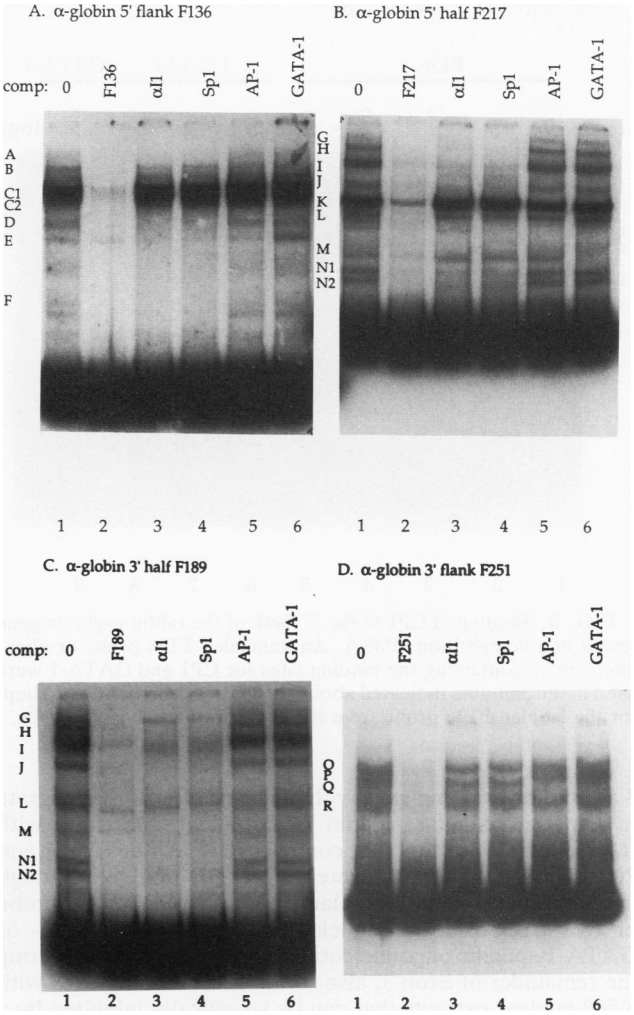

FIG. 2. Specificity of the binding of proteins in crude K562 nuclear extracts to regions of the α-globin gene tested in competition EMSAs. Competition assays were done with 400 ng of self DNA (unlabeled probe), oligonucleotides containing known binding sites for Sp1, AP-1, and GATA-1, or αI1 oligonucleotides, which are from a sequence in the first intron of the rabbit α-globin gene containing the Sp1-binding sites. The probes used for each experiment are indicated above each panel, and the source of the excess competitor (comp) DNA is listed above each lane.

(G through N; Fig. 2B), with complex K being the most abundant (Fig. 2B). These complexes are specifically inhibited through competition by a 100-fold molar excess of unlabeled probe. The more slowly moving complexes (G through J) are inhibited by Sp1-specific oligonucleotides but not by AP-1- or GATA-1-specific oligonucleotides, indicating that proteins recognizing Sp1-binding sites bind within the 5′ half of the gene. Complexes G through J are also specifically inhibited by αI1 oligonucleotides, which contain a portion of the first intron of the rabbit α-globin gene consisting of tandem repeats of the sequence CCCCG, suggesting that these potential Sp1 recognition sites are bound by proteins to form complexes G through J. The more quickly migrating complexes, M, N1, and N2, are not found in all batches of the extracts, indicating that they represent binding to degradation products.

The fragment containing intron 2 and parts of exons 2 and

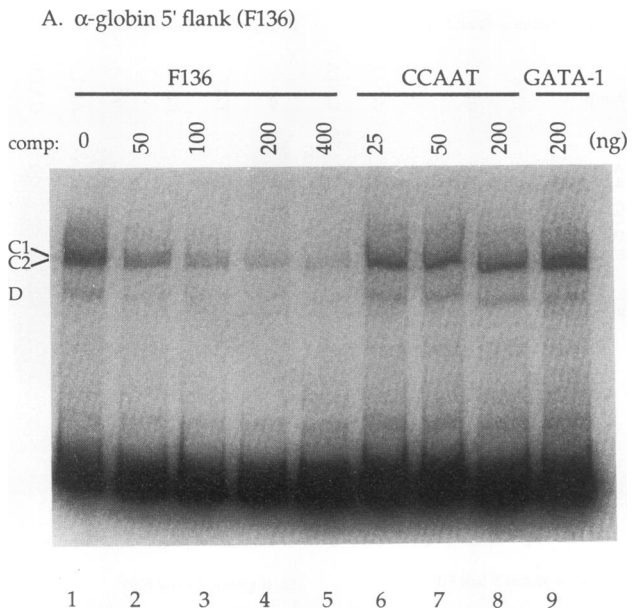

FIG. 3. Binding of CP1 to the 5′ flank of the rabbit α-globin gene tested in a competition EMSA. An unlabeled F136 probe or oligonucleotides containing the binding sites for CP1 and GATA-1 were used in the amounts indicated above the lanes as competitors (comp) for the labeled F136 probe from the 5′ flank of the α-globin gene.

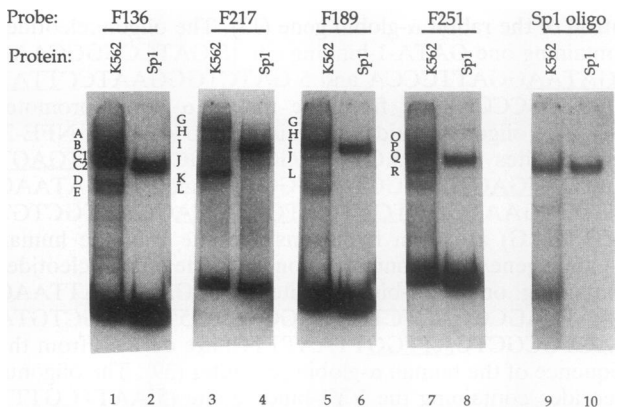

FIG. 4. Comparison by EMSAs of the binding of crude K562 nuclear extracts and purified Sp1 to the α-globin gene. Probes from the 5′-flanking and intragenic regions of the α-globin gene (Fig. 1) and Sp1-specific oligonucleotides (containing two Sp1-binding sites) were tested for binding to crude K562 nuclear extracts or pure Sp1 (indicated above the lanes).

3 (F189) forms complexes with nuclear extracts that migrate in a pattern similar to that of complexes obtained with fragment F217, except that complex K is lacking (Fig. 1 and 2C). The bound K562 proteins are all inhibited through competition by a 50-fold molar excess of an unlabeled probe or by Sp1-specific oligonucleotides but not by AP-1- or GATA-1-specific oligonucleotides. Fragment F251, covering the remainder of exon 3, also forms several complexes with K562 nuclear extracts that can be specifically inhibited by a 100-fold molar excess of an unlabeled probe. One complex (P) is also inhibited by Sp1-specific oligonucleotides (Fig. 2D).

Sp1 binds to flanking and internal sequences of the α-globin gene. Binding assays with purified Sp1 generated single complexes with mobilities similar but not identical to those of complexes from crude nuclear extracts (Fig. 4). Complex I (F217 and F189) generated by K562 extracts migrates slightly more slowly than the complex obtained with pure Sp1, and this complex is inhibited through competition by an excess of Sp1-specific oligonucleotides (Fig. 2B and C). However, other complexes obtained with F217 and F189 are also inhibited by the Sp1-specific oligonucleotides (e.g., G, H, and J), perhaps indicating that multimers of interacting proteins including Sp1 generate those complexes. The mobilities of the complexes obtained with pure Sp1 vary with different fragments; for example, the complexes obtained with F136 and F251 migrate at about the same rate as the complexes obtained with the Sp1-specific oligonucleotides, but the complexes obtained with F217 and F189 migrate more slowly, perhaps because of multimers of Sp1 binding to these regions of the gene. These binding data demonstrate that Sp1 can bind to fragments of the rabbit α-globin gene, but few if any of the complexes seen with crude nuclear extracts correspond to a simple complex between Sp1 and the probe.

Localization of protein-binding sites in the 5′ flank of the rabbit α-globin gene. The specific sequences interacting with nuclear proteins were identified by monitoring blocks to exonuclease III digestion (66), interference with binding by methylation of guanine nucleotides (60), and DNase I footprinting (24). The major blocks to exonucleolytic digestion of labeled fragment F136 (positions −129 to −122 and −115 to −110; Fig. 5A) are at the 5′ ends of upstream footprint regions UF2 and UF3, which are also protected from DNase I digestion (Fig. 5 and 6). Additional exonuclease III digestion blocks are found flanking the GC box (positions −95 to −90, matching the consensus site for Sp1) and upstream of position −140 (Fig. 5A). The blocks to exonuclease III digestion are eliminated by an excess of unlabeled competitor (Fig. 5A, lane 6), indicating that the binding is specific.

The sequence in complex C2 bound by proteins (Fig. 2A) was identified by a methylation interference assay. Fragment F114 (Fig. 1), which overlaps fragment F136, forms only one major complex in a mobility shift assay; this complex cannot be inhibited through competition by CP1-specific oligonucleotides, showing that fragment F114 does not form complex C1 but rather forms complex C2 (68a). Methylation of nucleotides from positions −44 to −55, in particular, the guanine residues (in boldface type) of the sequence **GGGCGTGCCC**, prevents the formation of complex C2 with nuclear extracts from either HeLa or K562 cells (Fig. 5B). This sequence is the one previously identified as the binding site for α-IRP (3), located between the CCAAT and TATA sequences (Fig. 6).

A DNase I footprint analysis with K562 nuclear extracts revealed several regions of protection in fragment F136 (Fig. 5C and D); the results are summarized in Fig. 6. One of the most prominent footprints on the lower strand covers the α-IRP-binding site and adjacent sequences from positions −39 to −65 (Fig. 5C, lane 2). Purified Sp1 also protects a smaller region near the α-IRP-binding site (positions −44 to −60 on the lower strand) (Fig. 5C, lane 4) and enhances cleavage 5′ to the α-IRP-binding site on the upper strand (Fig. 5D, lane 4), indicating that the protein binding to α-IRP may be Sp1 or a related protein. This result correlates with the mobility shift data (Fig. 4) showing that Sp1 forms a complex with probe F136 that moves slightly faster than complex C2, the complex involving the α-IRP-binding site

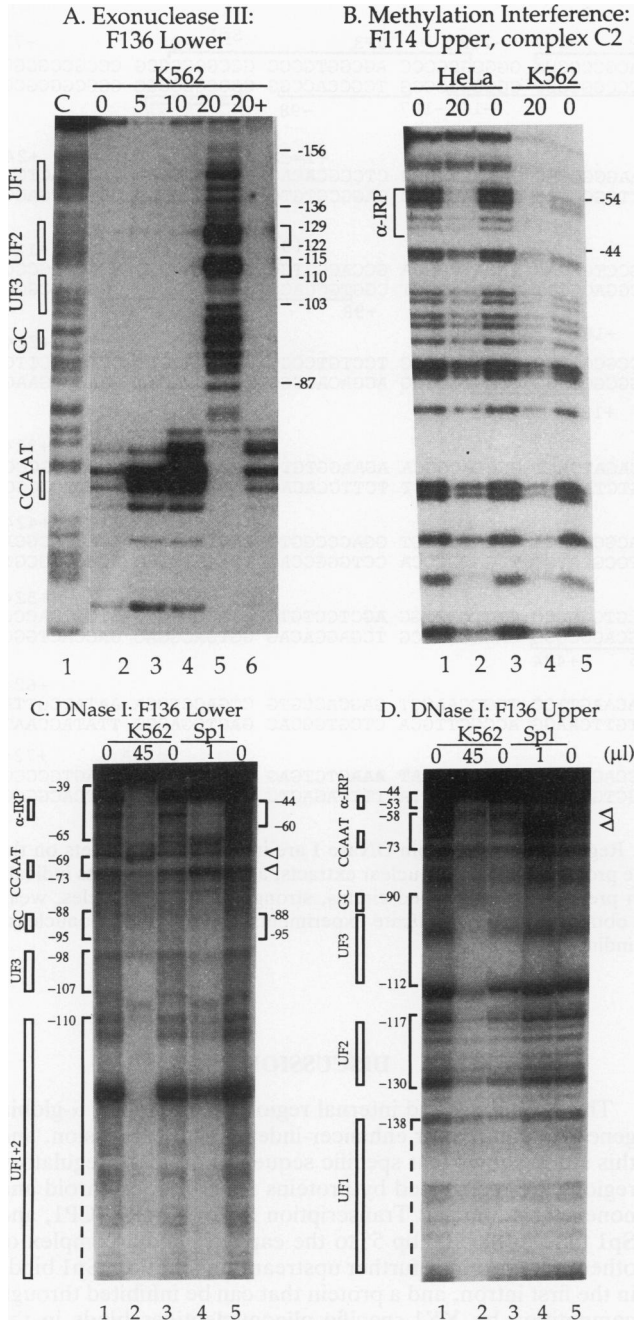

FIG. 5. Localization of binding sites in the 5′ flank of the rabbit α-globin gene. (A) Exonuclease III assay of fragment F136 (labeled on the 5′ end of the lower strand). Lanes: 1, DNA sequence reaction for cytosines (C); 2, no extract; 3 to 5, increasing amounts (5 to 20 μl) of extract; 6, 20 μl of extract inhibited through competition by 200 ng of a specific unlabeled probe. (B) Methylation interference assay of complex C2 formed with fragment F114 (labeled on the 3′ end of the upper strand). Lanes: 1, 3, and 5, unbound probe; 2 and 4, probe bound by protein from HeLa and K562 cell nuclear extracts (20 μl), respectively. (C) DNase I footprinting assay of fragment F136 (labeled on the 3′ end of the lower strand). Lanes: 1, 3, and 5, unbound probe; 2, probe incubated with 45 μl of K562 nuclear extract; 4, probe incubated with 1 footprinting unit of purified Sp1. (D) DNase I footprinting assay of fragment F136 (labeled on the 5′ end of the upper strand). Lanes are as indicated for panel C. Sequence motifs and positions of binding sites are marked to the side of the autoradiographs, on the basis of the upper-strand sequence. Protected regions are shown in brackets, and hypersensitive sites are indicated by arrowheads. The 5′ limit of UF1 could not be determined (indicated by broken brackets).

(Fig. 5B). The small difference in mobility in the EMSA and the protection of a larger region in the footprint assay indicate that additional proteins besides the Sp1-like protein bind to this region.

A match for the consensus Sp1-binding site (GC box) is found upstream of the CCAAT box at positions −90 to −95. This region is more difficult to examine by footprint analysis because the free DNA is cleaved inefficiently by DNase I, but some protection by both K562 nuclear extracts and Sp1 is apparent on the lower strand (Fig. 5C, lanes 2 and 4). Likewise, the CCAAT box (positions −69 to −73) in free DNA is cleaved inefficiently, but the addition of K562 nuclear extracts causes enhanced cleavage at position −69 on the lower strand (Fig. 5C, lane 2) and protection on the upper strand (Fig. 5D, lane 2). The footprinting and competition EMSA data (Fig. 3) strongly indicate that a protein, most likely CP1 (39), binds to the CCAAT box. Three other broad regions upstream of the GC box are also protected by proteins in the K562 nuclear extracts (Fig. 5C and D). These extended upstream footprint regions, UF1, UF2, and UF3 (Fig. 6), could result from the formation of large complexes with multiple proteins.

Localization of protein-binding sites internal to the rabbit α-globin gene. Complexes formed between fragment F217 and K562 nuclear extracts revealed several blocks to digestion with exonuclease III in the first intron (Fig. 7A), some of which (positions +156 and +163) are at the 3′ ends of potential Sp1-binding sites in the $(CCCCG)_4$ segment (Fig. 6). DNase I footprint assays show that these tandem repeats are bound by pure Sp1 (Fig. 7C and D), and with K562 proteins, the footprint in this segment extends through the splice junction (positions +128 to +167; Fig. 7D, lane 3). Enhanced cleavage is also seen adjacent to the $(CCCCG)_4$ segment when Sp1 is bound. These data, along with the competition EMSA data (Fig. 2) and data from EMSAs with Sp1 (Fig. 4) strongly indicate that Sp1 or a protein recognizing a similar binding site binds in the $(CCCCG)_4$ segment of intron 1 (Fig. 6). The $(CCCCG)_4$ sequence forms a dimer of the high-affinity decanucleotide-binding site for Sp1, CCCC GCCCCG (37). An additional footprint is seen further 3′ in the intron, from positions +181 to +190 (Fig. 6).

The other major complex in the 5′ end of the rabbit α-globin gene appears to be formed with nuclear protein YY1 (59). An additional footprint with K562 nuclear extracts covers positions +98 to +116 (Fig. 7C, lane 4, and Fig. 7D, lane 3); this region is in the 3′ portion of exon 1 (Fig. 6). Examination of complex K (Fig. 2B) by methylation interference showed that methylation of guanine residues (in boldface type) in the sequence **GTATGGCGCC** prevents the formation of this abundant complex (Fig. 7B). This sequence and methylation pattern are similar to those of the binding site for YY1 (Fig. 8A), and 10 ng (an 84-fold molar excess) of a duplex oligonucleotide containing the YY1-binding site prevents the formation of complex K between K562 nuclear extracts and α-globin gene probe F217 (Fig. 8B).

DNase I footprint analyses of intron 2 with a K562 nuclear extract showed protection of the sequence GTGG at positions +466 to +463 on the lower strand (Fig. 6 and 9A). Also, binding of this extract to the 3′-untranslated region of the gene protects a string of G residues on the lower strand (positions +664 to +660 in Fig. 6; Fig. 9B). For both of these probes, a nearby sequence containing the consensus Sp1-binding site (positions +479 to +484 in intron 2 and positions +654 to +659 in the 3′-untranslated region) is not cleaved efficiently by DNase I in the absence of the extract; hence, protein binding cannot be assessed well by footprinting.

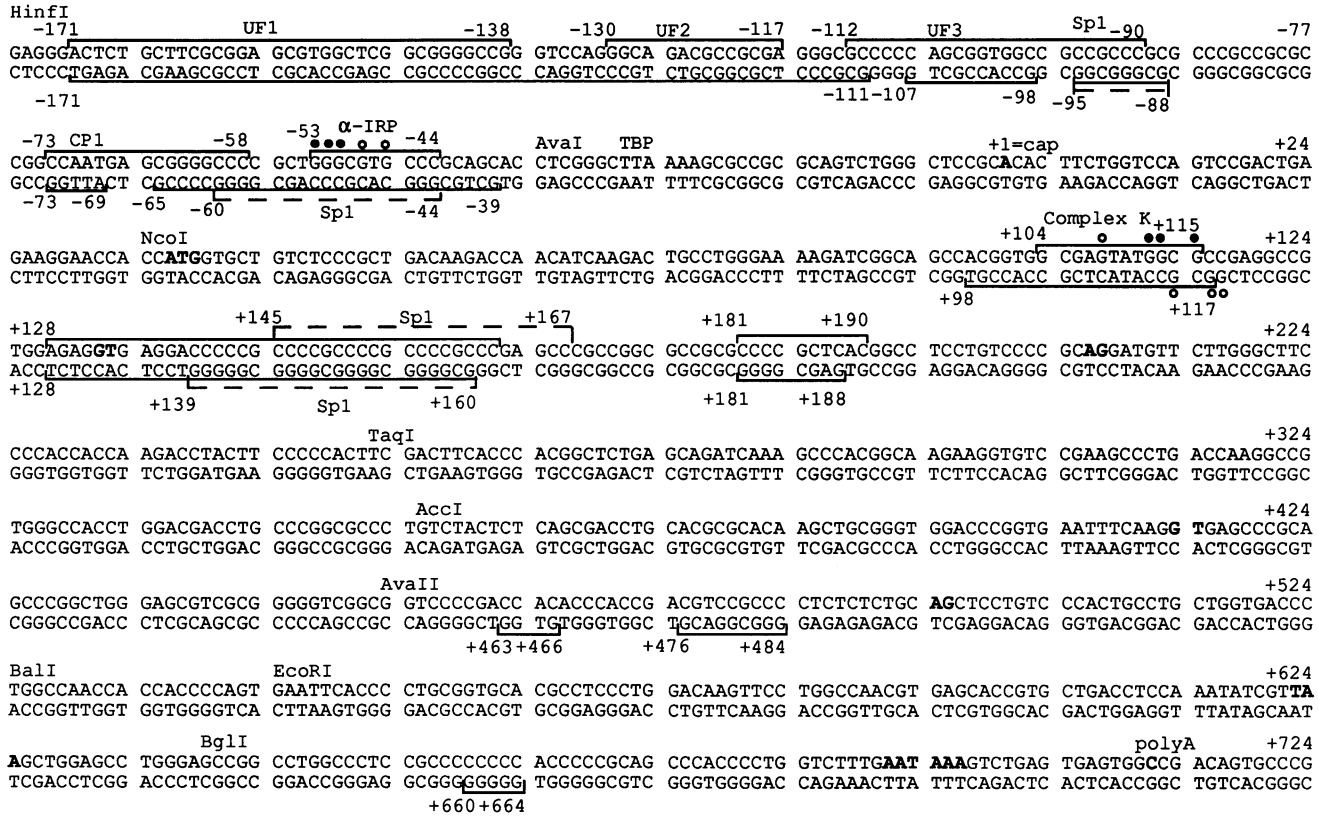

FIG. 6. Summary of protein-binding sites in the rabbit α-globin gene. Regions protected from DNase I are indicated by brackets on the upper and lower strands of the rabbit α-globin gene; solid brackets indicate protection by K562 nuclear extracts, and broken brackets indicate protection by Sp1. Circles mark guanine nucleotides whose methylation prevents binding (filled circles, strong effect; open circles, weak effect). This figure summarizes the results shown in Fig. 5, 7, and 9 and obtained in many replicate experiments. Restriction endonuclease cleavage sites used to generate the fragments are shown. TBP, TATA-binding protein.

However, the GC box in intron 2 does show a slightly reduced intensity of cleavage with K562 nuclear extracts (Fig. 9A), and purified Sp1 enhances cleavage at position +484 (68a). Also, Sp1 forms a complex with both of these regions of the gene in mobility shift assays (Fig. 4).

Patterns of conserved sequence in mammalian α-globin genes. The four-way alignment of the goat, human, rabbit, and mouse α-globin genes (Fig. 10) showed that the 5′-flanking region from the CP1-binding site to the cap site is highly conserved and thus is probably involved in a common regulatory mechanism. However, the mouse sequence is quite divergent in the further 5′ flank, instead containing several binding sites for a unique transcription factor, CP2 (41), and GATA-1 (4). The other three, more G+C-rich α-globin genes continue to be aligned through the 5′-flanking regions identified in binding assays. However, none of the hexanucleotide consensus Sp1-binding sites (CCGCCC) in the 5′ flanks of the genes are aligned. Upstream footprint region UF3 is well conserved among the human, goat, and rabbit sequences. The K site, or YY1, is quite well conserved, although this conservation is partly a result of selection for the protein sequence encoded. In intron 1, only the rabbit sequence has the adjacent pair of decanucleotide-binding sites for Sp1, but the hexanucleotide consensus Sp1-binding site is found in this same region in the goat sequence and further downstream in the human sequence.

DISCUSSION

The 5′-flanking and internal regions of the rabbit α-globin gene are required for enhancer-independent expression, and this study shows that specific sequences in these regulatory regions are recognized by proteins from both erythroid and nonerythroid nuclei. Transcription factors α-IRP, CP1, and Sp1 bind in the 100 bp 5′ to the cap site, and a complex of other proteins binds further upstream. In addition, Sp1 binds in the first intron, and a protein that can be inhibited through competition by YY1-specific oligonucleotides binds in the first exon. Finally, proteins bind to a CAC sequence in the second intron and to a string of C residues in the 3′-untranslated region. Proteins Sp1 (37) and CP1 (41) are known to be transcriptional activators, and it is likely that α-IRP is also an activator (3); hence, the binding of these proteins is probably involved in the enhancer-independent expression of the α-globin gene in transient assays. The requirement for both flanking and internal regions of the α-globin gene for efficient expression (9, 11, 34a, 69) suggests that the α-globin gene may contain an extended promoter that is part of a CpG island. Our identification of multiple protein recognition sequences flanking and internal to the gene supports this model. It is notable that no binding site for erythroid cell-specific transcriptional activator GATA-1 is found in the G+C-rich α-globin genes of humans, rabbits, and goats (Fig. 10), although a strong GATA-1-binding site is

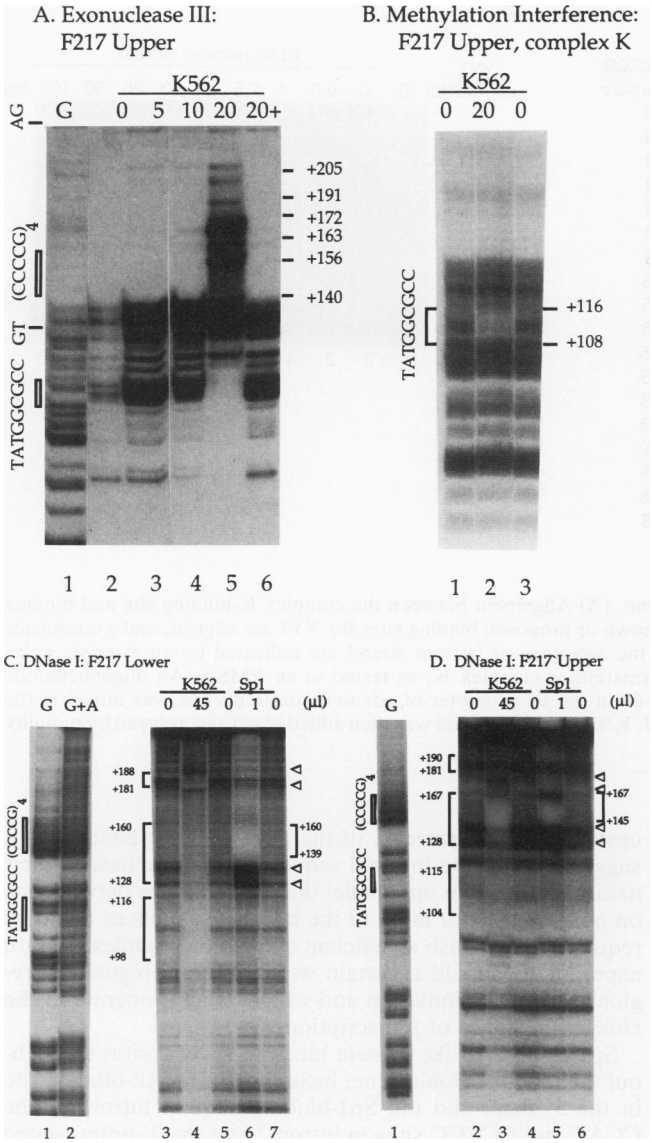

FIG. 7. Localization of binding sites in the 5′ end of the rabbit α-globin gene. (A) Exonuclease III assay of fragment F217 (labeled on the 5′ end of the upper strand). Lanes: 1, DNA sequence reaction (G); 2, no extract; 3 to 5, increasing amounts (5 to 20 μl) of K562 nuclear extract; 6, 20 μl of extract inhibited through competition by 200 ng of unlabeled probe. (B) Methylation interference assay of complex K formed with fragment F217 (labeled on the 5′ end of the upper strand). Lanes: 1 to 3, methylation-generated cleavage patterns of the unbound probe (lanes 1 and 3) or the probe after forming complex K (lane 2). (C) DNase I footprinting assay of fragment F217 (labeled on the 3′ end of the lower strand) with protein in K562 nuclear extracts or purified Sp1. Lanes: 1 and 2, sequencing reactions; 3 to 7, DNase I digestion of the probe with no added protein (lanes 3, 5, and 7), incubated with K562 nuclear extract (lane 4), or incubated with purified Sp1 (lane 6). (D) DNase I footprinting assay of fragment F217 (labeled on the 5′ end of the upper strand). Lanes: 2 to 6, DNase I digestion patterns of unbound probe (lanes 2, 4, and 6), probe incubated with K562 nuclear extract (lane 3), and probe incubated with purified Sp1 (lane 5). The sequencing reaction (lane 1) is the G reaction of the lower strand labeled on the 3′ end. Sequence motifs marked to the left of the autoradiographs are based on the sequence of the upper strand. Protected regions are shown in brackets, and hypersensitive sites are indicated by arrowheads.

present in the 5′ flank of the mouse α-globin gene (4) and in virtually all other control regions for erythroid genes (19). This fact indicates that the regulation of these highly G+C-rich genes differs in some way from that of other erythroid genes. However, GATA-1 is doubtless involved at some level, since a binding site has been found in the distal major control region of the human α-like globin gene cluster (35).

The protein α-IRP binds to a G-rich sequence that is reminiscent of the GC box recognized by Sp1 (Fig. 6). In fact, purified α-IRP binds to Sp1 recognition sites in the simian virus 40 promoter as readily as it does to the site in the α-globin gene promoter (3), but it differs from Sp1 in chromatographic behavior and in molecular weight (40). The conserved region of the α-IRP-binding site (Fig. 10) matches the consensus sequence of the Sp1-binding site, NRGG CGNRN, derived both from sequence matches and from considerations of the sequence specificity of the Zn fingers in the binding site (5). We show that purified Sp1 binds to the α-IRP-binding site but that Sp1-specific oligonucleotides do not displace the complex formed with K562 nuclear extracts, indicating that some protein in the extracts has a higher affinity than does Sp1 for this sequence in the 5′ flank of the rabbit α-globin gene. This suggestion supports the contention of Kim et al. (40) that the α-IRP protein is related to but distinct from the ubiquitous Sp1.

A sequence at the 3′ end of exon 1 is bound by an abundant protein to form major complex K. A strong candidate for the protein forming this complex is YY1, on the basis of alignments of binding-site sequences and the ability of a YY1-specific oligonucleotide to compete for the formation of the complex. Nuclear protein YY1 (also called δ, NF-E1, CF1, UCRBP, and CSBP1) was identified independently in several different systems and has a variety of effects on gene expression. YY1 (59) is involved in both positive and negative regulation of the P5 promoter of adeno-associated virus; it also binds to the initiator sequence and is required for basal levels of transcription (57). A protein called δ binds to the internal regulatory regions of the mouse genes for ribosomal proteins L30 and L32 (1), which are required for expression (14). A protein called NF-E1 binds to the downstream enhancer of the immunoglobulin κ gene; at this location it can repress expression of the gene (50). Analysis of cDNA clones for YY1, δ, and NF-E1 show that these are the same protein (30, 50, 59). Other studies have shown that this protein binds in the 5′ flank of c-*myc*, at which it can apparently activate transcription (54), in the long terminal repeat of Moloney leukemia virus, at which it inhibits transcription (22), in the internal sequences of the dihydrofolate reductase (20) and Surf-1 (42) genes, at which it appears to play a positive role, and in the 5′ flanks of the γ- and ε-globin genes, at which it may be involved in transcriptional silencing (28). YY1 also binds to the μE1 site (50, 54), initially identified in the intronic enhancer of immunoglobulin heavy-chain genes; hence, it may also bind to sequences implicated as μE1-related sites, such as the highly conserved sequences near the beginning of full-length long interspersed repetitive elements in humans and rabbits (52) or in the enhancer of the insulin gene (38).

The facts that YY1 can bind to a variety of positions within or flanking particular genes and exert such a wide range of effects indicate that either it has multiple functional domains (30) or perhaps it acts as a mediator for other proteins that have a positive or a negative function. For example, YY1 is required for the activation of the P5 promoter by adenovirus-encoded protein E1A (59). In several cases, the YY1-binding sites are internal to the gene, as

A.

Gene	Sequence	Position	Protein	Reference
α-globin	CGGCGCCATACTCG	+110*		this paper
P5 AAV	TTGCGACATTTTGC	-59	YY1	60
P5 AAV	GGTCTCCATTTTGA	+1	YY1	60
rpL30	CCCGGCCATCTTGG	+20	δ	30
rpL32	GGCTGCCATCTGTT	+34	δ	30
rpL32	GGCGGCCATCCGCC	+67	δ	30
DHFR	CGCTGCCATCATGG	+50	δ	30
κE3'	CACCTCCATCTTGT	3' enh.	NF-E1	51
IgHE	ATCGGCCATCTTGA	enh. 364*	μE	55
c-myc	CCCGACCATTTTCT	-255*	CF1	55
c-myc	ATACGCCATGTACC	-385*	CF1	55
α-actin	CGTCGCCATATTTG	-84*	CF1	55
Surf-1,2	AGCAGCCATCTTTG	-78	Su1	43
L1Oc	TTCGGCCATCTTGT	+20*		53
L1Hs	TTCGGCCATCTTGG	+17*		53
insulin	ATCCGCCATCTGCC	-108	IEF1	39
insulin	ATCCGCCATCTGGC	-235	IEF1	39
γ-globin	TTCTGACATATTGG	-1092*	CSBP1	28
ε-globin	CTCTCCCATTCTCC	-288*	CSBP1	28

Consensus NKCSGCCATCTTSN

B

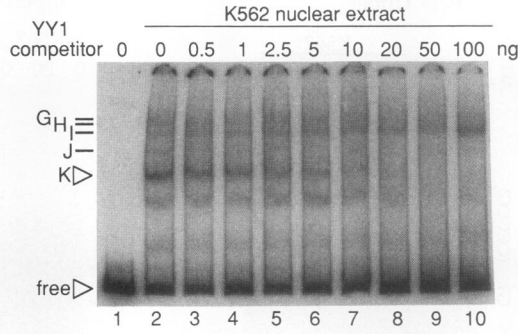

FIG. 8. Evidence that YY1 binds in exon 1 of the rabbit α-globin gene. (A) Alignment between the complex K-binding site and binding sites for protein YY1. Sequences and positions (the cap site is +1) for known or proposed binding sites for YY1 are aligned, and a consensus sequence is shown; the invariant core is underlined. Sequences from the antisense or bottom strand are indicated by an asterisk. enh., enhancer. (B) Competition of YY1-specific oligonucleotides for the formation of complex K, as tested in an EMSA. An oligonucleotide containing the sequence for the YY1-binding site located at position −60 in the P5 promoter of adeno-associated virus was mixed in the amounts indicated above the lanes with 1 ng of end-labeled fragment F217. K562 nuclear extract was then added (2 μl) and assayed by mobility shifts.

is the binding site for complex K in the rabbit α-globin gene. The intronic regulatory sequences in the ribosomal protein genes function only when placed 3′ to the cap site (14), and current data suggest this may also be the case for the human α-globin gene. Fusion of the 5′ flank plus exon 1 and intron 1 of the human α-globin gene to a reporter gene resulted in efficient expression, but placement of the internal sequences upstream or downstream of the gene was not effective (9), suggesting that the internal sequences operate best in their natural context. In one model that explains this dependence on position, one or more of the internal sequences could be required to establish an efficient elongation complex. Future experiments should ascertain which internal regulatory regions operate at initiation and which, if any, operate at the elongation stages of transcription.

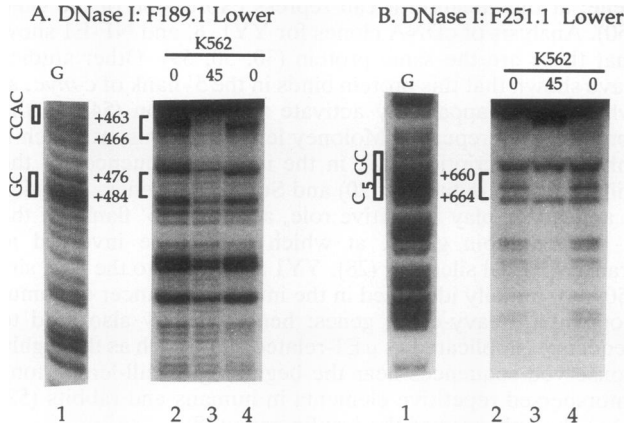

FIG. 9. Localization of binding sites in the second intron and 3′-untranslated regions of the rabbit α-globin gene. (A) DNase I footprinting assay of fragment F189.1 (Fig. 1) (labeled on the 5′ end of the lower strand). Lanes: 1, DNA sequence reaction (G); 2 and 4, DNase I cleavage pattern for the unbound probe; 3, DNase I cleavage pattern for the probe incubated with 45 μl of K562 nuclear extract. (B) DNase I footprinting assay of fragment F251.1 (labeled on the 5′ end of the lower strand). Lanes are as indicated for panel A.

Sp1 or an Sp1-like protein binds to several sites throughout the rabbit α-globin gene, including the α-IRP-binding site in the 5′ flank and the Sp1-binding sites in intron 1. The CCAC and CCCCC sites in intron 2 and the 3′-untranslated region are similar to sequences that have been shown to bind Sp1, despite their deviation from the Sp1-binding site consensus sequence (43). In no case did the mobility of complexes with pure Sp1 match that of any complex with a K562 nuclear extract (Fig. 4). This result shows that the binding factors in K562 cells either are not identical to Sp1 from HeLa cells or include other proteins in a heterotypic complex. The DNase I footprints formed with crude extracts near the Sp1-binding sites in fragments F136 and F217 were broader than those seen with purified Sp1, supporting the idea that a large complex of many proteins may form on these DNA segments. Perhaps this extended region of protection reflects the formation of a protein-DNA complex peculiar to a CpG island. For example, the three upstream region footprints in the region 100 to 170 bp upstream from the transcription initiation site show broad protection of these G+C-rich sequences. One model for transcriptional regulation (8) suggests that specific binding proteins maintain hypomethylated CpG islands by sterically excluding DNA methyltransferases. These extended protected regions in the 5′ flank of the rabbit α-globin gene may represent the binding of specific CpG island DNA-binding proteins.

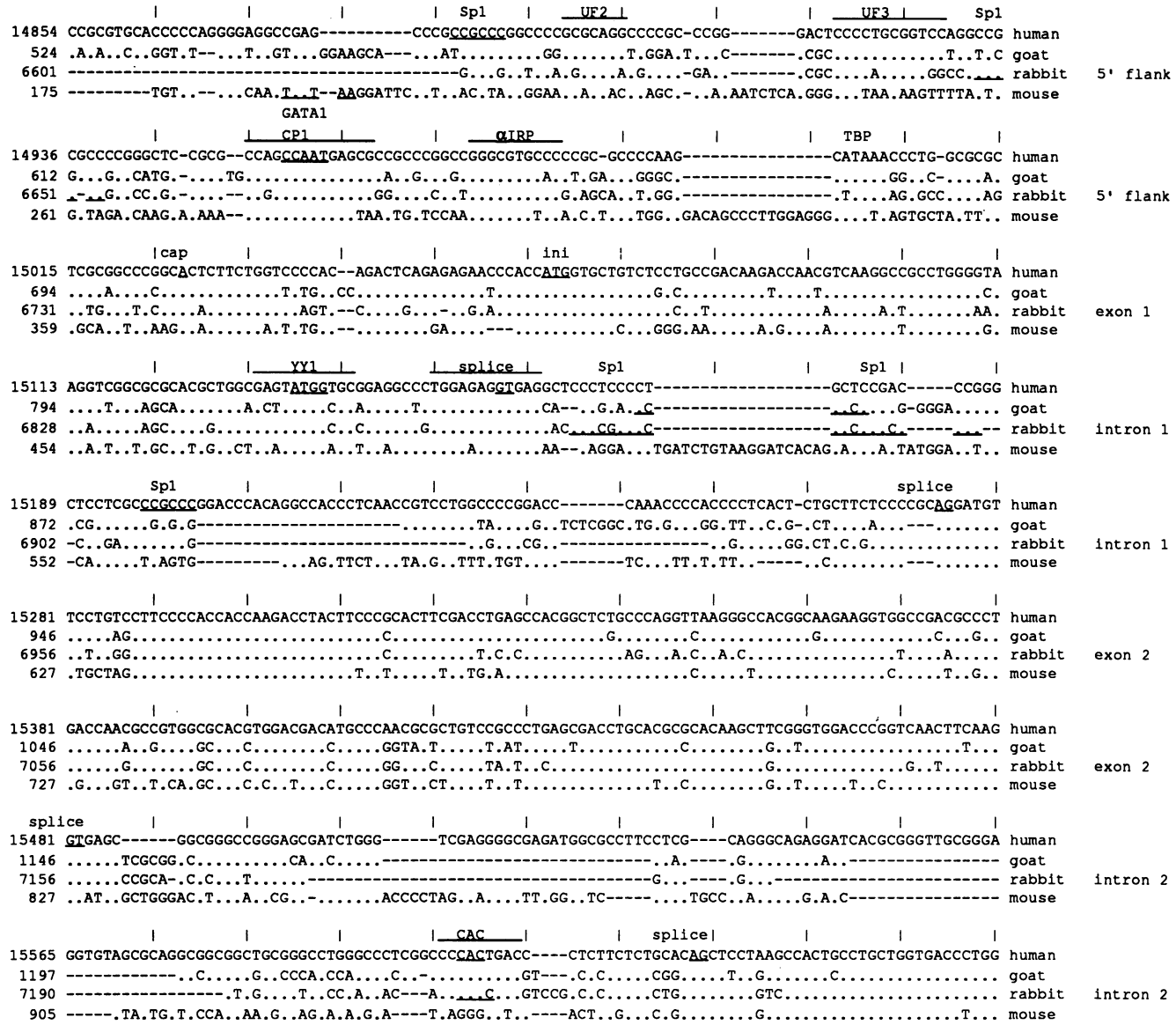

FIG. 10. Alignment of human, goat, rabbit, and mouse α-globin gene sequences. The program *yama* (10a) was used to construct the multiple alignment, starting with pairwise alignments from the program *sim* (32); the region from the 5′ flank through intron 2 is shown. Matching nucleotides are indicated by periods, and dashes represent gaps introduced to improve the alignment. Well-conserved binding sites are overlined, and core-binding sites are underlined. Consensus Sp1-binding sites, whether conserved or not, are underlined; in most cases, these are the hexanucleotide consensus, but in rabbit intron 1 they are the decanucleotide consensus. The cap site, translation initiation codon (ini), and splice junctions are indicated. Sources of the α-globin gene sequences are human (44), goat $^{II}\alpha$ (56), rabbit (13, 29), and mouse (48). TBP, TATA-binding protein.

ACKNOWLEDGMENTS

We thank Don Wojchowski for providing the oligonucleotides containing binding sites for Sp1, AP-1, CACCC, and GATA-1 and Webb Miller for generating the multiple alignment of the α-globin genes.

This work was supported by Public Health Service grants DK-27635 and HL-44491, a PHS Research Career Development Award (DK-01589) to R.H., a Hatch Grant from The Pennsylvania State University Agricultural Experiment Station, and a Grant-in-Aid of Research from Sigma Xi, the Scientific Research Society, to S.E.Y.

REFERENCES

1. **Atchison, M. L., O. Meyuhas, and R. P. Perry.** 1989. Localization of transcriptional regulatory elements and nuclear factor binding sites in mouse ribosomal protein gene rpL32. Mol. Cell. Biol. **9**:2067–2074.
2. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
3. **Barnhart, K. M., C. G. Kim, S. S. Banerji, and M. Sheffery.** 1988. Identification and characterization of multiple erythroid cell proteins that interact with the promoter of the murine α-globin gene. Mol. Cell. Biol. **8**:3215–3226.
4. **Barnhart, K. M., C. G. Kim, and M. Sheffery.** 1989. Purification and characterization of an erythroid cell-specific factor that binds the murine α- and β-globin genes. Mol. Cell. Biol. **9**:2606–2614.
5. **Berg, J. M.** 1992. Sp1 and the subfamily of zinc finger proteins

with guanine-rich binding sites. Proc. Natl. Acad. Sci. USA **89:**11109–11110.
6. **Bernardi, G., B. Olofsson, J. Filipski, M. Zerial, J. Salivas, G. Cuny, M. Meunier-Rotival, and F. Rodier.** 1985. The mosaic genome of warm-blooded vertebrates. Science **228:**953–958.
7. **Bird, A., M. Taggart, R. Nicholls, and D. Higgs.** 1987. Non-methylated CpG-rich islands at the human α-globin locus: implications for evolution of the α-globin pseudogene. EMBO J. **6:**999–1004.
8. **Bird, A. P.** 1986. CpG-rich islands and the function of DNA methylation. Nature (London) **231:**209–213.
9. **Brickner, H. E., X.-X. Zhu, and G. F. Atweh.** 1991. A novel regulatory element of the human α-globin gene responsible for its constitutive expression. J. Biol. Chem. **266:**15363–15368.
10. **Chao, M., P. Mellon, P. Charnay, T. Maniatis, and R. Axel.** 1983. The regulated expression of β-globin genes introduced into mouse erythroleukemia cells. Cell **32:**483–493.
10a. **Chao, K.-M., W. Miller, and R. Hardison.** Submitted for publication.
11. **Charnay, P., R. Treisman, P. Mellon, M. Chao, R. Axel, and T. Maniatis.** 1984. Differences in human α- and β-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell **38:**251–263.
12. **Cheng, J. F., D. E. Krane, and R. C. Hardison.** 1988. Nucleotide sequence and expression of rabbit globin genes ζ1, ζ2 and ζ3. J. Biol. Chem. **263:**9981–9993.
13. **Cheng, J. F., L. Raid, and R. C. Hardison.** 1986. Isolation and nucleotide sequence of the rabbit gene cluster ψζ-α1-ψα. J. Biol. Chem. **261:**839–848.
14. **Chung, S., and R. P. Perry.** 1989. Importance of introns for expression of mouse ribosomal protein gene rpL32. Mol. Cell. Biol. **9:**2075–2082.
15. **Cohen, R. B., M. Sheffery, and C. G. Kim.** 1986. Partial purification of a nuclear protein that binds to the CCAAT box of the mouse α_1-globin gene. Mol. Cell. Biol. **6:**821–832.
16. **Collins, F. S., and S. M. Weissman.** 1984. The molecular genetics of human hemoglobin. Prog. Nucleic Acid Res. Mol. Biol. **31:**315–462.
17. **Deisseroth, A., and D. Hendrick.** 1978. Human α-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukemia cells. Cell **15:**55–63.
18. **Dignam, J., R. L. Lebovitz, and R. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11:**1475–1489.
19. **Evans, T., G. Felsenfeld, and M. Reitman.** 1990. Control of globin gene transcription. Annu. Rev. Cell Biol. **6:**95–124.
20. **Farnham, P. J., and A. L. Means.** 1990. Sequences downstream of the transcription initiation site modulate the activity of the murine dihydrofolate reductase gene. Mol. Cell. Biol. **10:**1390–1398.
21. **Fischel-Ghodsian, N., R. D. Nicholls, and D. R. Higgs.** 1987. Unusual features of CpG-rich (HTF) islands in the human α-globin complex: association with nonfunctional pseudogenes and presence within the 3′ portion of the ζ genes. Nucleic Acids Res. **15:**9215–9225.
22. **Flanagan, J. R., K. G. Becker, D. L. Ennist, S. L. Gleason, P. H. Driggers, B.-Z. Levi, E. Appella, and K. Ozato.** 1992. Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. Mol. Cell. Biol. **12:**38–44.
23. **Fried, M., and D. M. Crothers.** 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9:**6505–6525.
24. **Galas, D., and A. Schmitz.** 1978. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5:**3157–3170.
25. **Gaub, M.-P., M. Bellard, I. Scheuer, P. Chambon, and P. Sassone-Corsi.** 1990. Activation of the ovalbumin gene by the estrogen receptor involves the fos-jun complex. Cell **63:**1267–1276.
26. **Gidoni, D., J. T. Kadonaga, H. Barrera-Saldana, K. Takahashi, P. Chambon, and R. Tjian.** 1985. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science **230:**511–517.
27. **Grosveld, F., G. B. van Assendelft, D. R. Greaves, and G. Kollias.** 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell **51:**975–985.
28. **Gumucio, D. L., H. Heilstedt-Williamson, T. A. Gray, S. A. Tarle, D. A. Shelton, D. Tagle, J. Slightom, M. Goodman, and F. S. Collins.** 1992. Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human γ- and ε-globin genes. Mol. Cell. Biol. **12:**4919–4929.
29. **Hardison, R. C., D. Krane, D. Vandenbergh, J.-F. Cheng, J. Mansberger, J. Taddie, S. Schwartz, X. Huang, and W. Miller.** 1991. Sequence and comparative analysis of the rabbit α-like globin gene cluster reveals a rapid mode of evolution in a G+C rich region of mammalian genomes. J. Mol. Biol. **222:**233–249.
30. **Hariharan, N., D. E. Kelley, and R. P. Perry.** 1991. δ, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. Proc. Natl. Acad. Sci. USA **88:**9799–9803.
31. **Higgs, D., W. Wood, A. Jarman, J. Sharpe, J. Lida, I. M. Pretorious, and H. Ayyub.** 1990. A major positive regulatory region located far upstream of the human α-globin gene locus. Genes Dev. **4:**1588–1601.
32. **Huang, X., R. Hardison, and W. Miller.** 1990. A space-efficient algorithm for local similarities. Comput. Appl. Biosci. **6:**373–381.
33. **Humphries, R. K., T. Lee, P. Turner, A. D. Moulton, and A. W. Nienhuis.** 1982. Differences in human α-, β-, and δ-globin gene expression in monkey kidney cells. Cell **30:**173–183.
34. **Jackson, P., T. Evans, J. Nickol, and G. Felsenfeld.** 1989. Developmental modulaton of protein binding to β-globin gene regulatory sites within chicken erythrocyte nuclei. Genes Dev. **3:**1860–1873.
34a. **James-Pederson, M., et al.** Submitted for publication.
35. **Jarman, A., W. Wood, J. Sharpe, G. Gourdon, H. Ayyub, and D. Higgs.** 1991. Characterization of the major regulatory element upstream of the human α-globin gene cluster. Mol. Cell. Biol. **11:**4679–4689.
36. **Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian.** 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48:**79–89.
37. **Kadonaga, J. T., K. A. Jones, and R. Tjian.** 1986. Promoter-specific activation of RNA polII transcription by Sp1. Trends Biochem. Sci. **11:**20–23.
38. **Karlsson, O., M. D. Walker, W. J. Rutter, and T. Edlund.** 1989. Individual protein-binding domains of the insulin gene enhancer positively activate β-cell-specific transcription. Mol. Cell. Biol. **9:**823–827.
39. **Kim, C., and M. Sheffery.** 1990. Physical characterization of the purified CCAAT transcription factor α-CP1. J. Biol. Chem. **265:**13362–13369.
40. **Kim, C. G., K. M. Barnhart, and M. Sheffery.** 1988. Purification of multiple erythroid cell proteins that bind the promoter of the α-globin gene. Mol. Cell. Biol. **8:**4270–4281.
41. **Kim, C. G., S. L. Swendeman, K. M. Barnhart, and M. Sheffery.** 1990. Promoter elements and erythroid cell nuclear factors that regulate α-globin gene transcription in vitro. Mol. Cell. Biol. **10:**5958–5966.
42. **Lennard, A., and M. Fried.** 1991. The bidirectional promoter of the divergently transcribed mouse *Surf-1* and *Surf-2* genes. Mol. Cell. Biol. **11:**1281–1294.
43. **Lewis, C. D., S. P. Clark, G. Felsenfeld, and H. Gould.** 1988. An erythrocyte-specific protein that binds to the poly(dG) region of the chicken β-globin gene promoter. Genes Dev. **2:**863–873.
44. **Liebhaber, S., M. Goossens, and Y. W. Kan.** 1980. Cloning and complete nucleotide sequence of human 5′-α-globin gene. Proc. Natl. Acad. Sci. USA **77:**7054–7058.
45. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
46. **Mellon, P., V. Parker, Y. Gluzman, and T. Maniatis.** 1981. Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system.

Cell **27**:279–288.
47. **Ney, P., B. Sorrentino, K. McDonagh, and A. Nienhuis.** 1990. Tandem AP-1-binding sites within the human β-globin dominant control region function as an inducible enhancer in erythroid cells. Genes Dev. **4**:993–1006.
48. **Nishioka, Y., and P. Leder.** 1979. The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. Cell **18**:875–882.
49. **Orkin, S.** 1990. Globin gene regulation and switching: circa 1990. Cell **63**:665–672.
50. **Park, K., and M. L. Atchison.** 1991. Isolation of a candidate repressor/activator, NF-E1 (YY-1, δ), that binds to the immunoglobulin κ 3′ enhancer and the immunoglobulin heavy-chain μE1 site. Proc. Natl. Acad. Sci. USA **88**:9804–9808.
51. **Plumb, M., J. Frampton, H. Wainwright, M. Walker, K. Macleod, G. Goodwin, and P. Harrison.** 1989. GATAAG: a *cis*-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. Nucleic Acids Res. **17**:73–92.
52. **Price, D. K., J. A. Ayres, D. Pasqualone, C. H. Cabell, W. Miller, and R. C. Hardison.** 1992. The 5′ ends of LINE1 repeats in rabbit DNA define subfamilies and reveal a short sequence conserved between rabbits and humans. Genomics **14**:320–331.
53. **Pyati, J., R. S. Kucherlapati, and A. Skoultchi.** 1980. Activation of human β-globin genes from non-erythroid cells by fusion with murine erythroleukemia cells. Proc. Natl. Acad. Sci. USA **79**:4337–4341.
54. **Riggs, K. J., K. T. Merrell, G. Wilson, and K. Calame.** 1991. Common factor 1 is a transcriptional activator which binds in the c-*myc* promoter, the skeletal α-actin promoter, and the immunoglobulin heavy-chain enhancer. Mol. Cell. Biol. **11**: 1765–1769.
55. **Santoro, C., N. Mermod, P. Andrews, and R. Tjian.** 1988. A family of CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature (London) **334**:218–224.
56. **Schon, E., S. Wernke, and J. Lingrel.** 1982. Gene conversion of two functional goat α-globin genes preserves only minimal flanking sequences. J. Biol. Chem. **257**:6825–6835.
57. **Seto, E., Y. Shi, and T. Shenk.** 1991. YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*. Nature (London) **354**:241–245.
58. **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. USA **77**:6634–6638.
59. **Shi, Y., E. Seto, L.-S. Chang, and T. Shenk.** 1991. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell **678**:377–388.
60. **Siebenlist, U., R. Simpson, and W. Gilbert.** 1980. *E. coli* RNA polymerase interacts homologously with two different promoters. Cell **20**:269–281.
61. **Treisman, R., M. R. Green, and T. Maniatis.** 1983. *cis* and *trans* activation of globin gene transcription in transient assays. Proc. Natl. Acad. Sci. USA **80**:7428–7432.
62. **van der Ploeg, L. H., and R. A. Flavell.** 1980. DNA methylation in the human γ-δ-β-globin locus in erythroid and nonerythroid tissues. Cell **19**:947–958.
63. **Whitelaw, E., P. Hogben, O. Hanscombe, and J. Proudfoot.** 1989. Transcriptional promiscuity of the human α-globin gene. Mol. Cell. Biol. **9**:241–251.
64. **Willing, M., A. Nienhuis, and W. Anderson.** 1979. Selective activation of a human β- but not a γ-globin gene in human fibroblast and mouse erythroleukemic cell hybrids. Nature (London) **277**:534–538.
65. **Wright, S., E. deBoer, F. G. Grosveld, and R. A. Flavell.** 1983. Regulated expression of the human β-globin gene family in murine erythroleukaemia cells. Nature (London) **305**:333–336.
66. **Wu, C.** 1985. An exonuclease protection assay reveals heat-shock element and TATA box DNA-binding proteins in crude nuclear extracts. Nature (London) **317**:84–87.
67. **Xu, J., and R. Hardison.** 1989. Localization of the β-like globin gene cluster and the genes for parathyroid hormone and c-Harvey *ras* 1 to region q14-q21 of rabbit chromosome 1 by *in situ* hybridization. Cytogenet. Cell Genet. **52**:157–161.
68. **Xu, J., and R. Hardison.** 1991. Localization of the α-like globin gene cluster to region q12 of rabbit chromosome 6 by *in situ* hybridization. Genomics **9**:362–365.
68a. **Yost, S. E.** 1991. Ph.D. thesis. The Pennsylvania State University, University Park.
69. **Yost, S. E., M. James-Pederson, J. Xu, R. Miller, T. Zeigler, and R. C. Hardison.** 1991. Intragenic sequences and proteins regulating the rabbit α-globin gene, p. 220–234. *In* G. Stamatoyannopoulos and A. W. Nienhuis (ed.), The regulation of hemoglobin switching. Johns Hopkins University Press, Baltimore.
70. **Zabel, B., H. Kronenberg, G. Bell, and T. Shows.** 1985. Chromosome mapping of genes on the short arm of human chromosome 11: parathyroid hormone gene is at 11p15 together with genes for insulin, c-Harvey-ras1, and β-hemoglobin. Cytogenet. Cell Genet. **39**:200–205.