Fine Structure Genetic Analysis of a β-Globin Promoter

RICHARD M. MYERS, KIT TILLY, TOM MANIATIS

A novel procedure for saturation mutagenesis of cloned DNA was used to obtain more than 100 single base substitutions within the promoter of the mouse β-major globin gene. The effects of these promoter substitutions on transcription were determined by transfecting the cloned mutant genes into HeLa cells on plasmids containing an SV40 transcription enhancer, and measuring the levels of correctly initiated β-globin transcripts after 2 days. Mutations in three regions of the promoter resulted in a significant decrease in the level of transcription: (i) the CACCC box, located between −87 and −95, (ii) the CCAAT box, located between −72 and −77, and (iii) the TATA box, located between −26 and −30 relative to the start site of transcription. In contrast, two different mutations in nucleotides immediately upstream from the CCAAT box resulted in a 3- to 3.5-fold increase in transcription. With two minor exceptions, single base substitutions in all other regions of the promoter had no effect on transcription. These results precisely delineate the *cis*-acting sequences required for accurate and efficient initiation of β-globin transcription, and they establish a general approach for the fine structure genetic analysis of eukaryotic regulatory sequences.

THE ANALYSIS OF SINGLE BASE MUTATIONS IN PROKARYOTIC promoters and regulatory sequences has led to significant insights into the mechanisms of transcription initiation and gene regulation (*1–3*). Mutations that were initially identified on the basis of phenotype played an essential role in the study of in vivo and in vitro interactions of RNA polymerase and regulatory proteins with *cis*-acting regulatory sequences (*4*). Single base mutations in these regulatory sequences were obtained by in vivo mutagenesis and genetic selection. However, genetic selections for promoter and regulatory mutations in higher eukaryotes have thus far been difficult to establish. This limitation has been circumvented by introducing mutations into specific regions of cloned DNA in vitro (*5*; *6–8* for reviews), and then studying the consequences of these mutations on in vivo or in vitro transcription.

The eukaryotic promoters that have been most thoroughly characterized by this molecular genetic approach are those of the herpesvirus thymidine kinase (tk) (*5*, *9*, *10*), the SV40 T-antigen (see *11* for references), and mammalian β-globin (*12–15*) genes. Comparison of the DNA sequences immediately upstream from the messenger RNA (mRNA) initiation sites of a number of mammalian β-globin genes revealed three highly conserved regions (*16–18*), which became the focus of in vitro mutagenesis studies (*12–15*). The first region occurs at about −90 relative to the cap site and contains the sequence GCCACACCC (*14*, *17*); this sequence is a common feature of adult β-globin gene promoters from many species, but does not appear in the promoters of many other eukaryotic genes. The second region contains the sequence CCAAT and appears at about −75; the CCAAT box is a common element in many eukaryotic promoters transcribed by RNA polymerase II (*16*, *19*). The third region, called the TATA box, is located at about −30 of most eukaryotic genes and appears to participate in determining the site of transcription initiation (*20*). The functional significance of the conserved β-globin promoter elements was established initially by the analysis of a series of 5′ and internal deletions (*12*, *14*). A more precise localization of the promoter elements was subsequently established with the use of a linker scanning mutagenesis procedure (*15*). Finally, a small number of single base mutations within the conserved sequences was generated and analyzed (*14*).

A number of procedures have been used for introducing single base mutations into specific DNA sequences (*7*, *8*). We have developed a method for rapidly introducing large numbers of single base substitutions into cloned DNA fragments (*21*). This technique includes the generation of mutations by chemical treatment of single-stranded DNA in vitro, synthesis of a complementary DNA strand, and subsequent separation of mutant and wild-type duplex DNA fragments by denaturing gradient gel electrophoresis. Since this procedure does not involve a genetic selection for mutant promoters, both neutral substitutions as well as those that alter promoter function are obtained. Moreover, the distribution of point mutations is not biased toward conserved sequence elements. Analysis of large numbers of both types of substitutions allows the precise and unambiguous identification of promoter sequences required for accurate and efficient transcription. We used this saturation mutagenesis procedure to introduce 130 different random single base substitutions into the mouse β-major globin promoter region (*21*). In this article, we report the analysis of the effects of these mutations on β-globin gene transcription by means of a transient expression assay (*22*, *23*). We find that base substitutions that alter β-globin gene transcription are confined to the three conserved regions, with two minor exceptions. Moreover, two single base substitutions near the CCAAT box result in a significant increase in the level of β-globin gene transcription and are therefore identified as promoter "up mutations".

Transcription assay. The effects of single base substitutions on β-globin transcription were analyzed by a HeLa cell transient expression assay. The β-globin genes used in our study were carried on plasmids containing the SV40 72–base pair (bp) repeat sequence

Richard M. Myers and Kit Tilly were postdoctoral fellows and Tom Maniatis is a professor in the Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138. R. M. Myers is now in the Department of Physiology, University of California, San Francisco, CA 94143, and K. Tilly is in the Department of Physiological Chemistry, University of Wisconsin, Madison, WI 53706.

(*24*), since efficient β-globin transcription in HeLa cells requires an enhancer element (*23*). A diagram of the mouse β-globin plasmids used in the HeLa cell transfection experiments is shown in Fig. 1. Each of the single-base promoter mutations was analyzed in a "test" plasmid containing an 8-bp Bgl II linker at +26 of the mouse β-globin gene (Fig. 1A). The reference plasmid is identical to the test plasmid except that it contains a wild-type promoter, and it does not contain the linker at +26 (Fig. 1B). The linker DNA makes it possible to distinguish transcripts derived from the test gene from those derived from the reference gene. Both plasmids contain 106 bp of 5′ flanking DNA and the first 475 bp of the mouse β-major globin gene. This portion of the gene is fused to the remaining part of the human β-globin gene at a common Bam HI site in the second exon. The mouse-human hybrid β-globin gene was used to allow direct comparisons to previous deletion and linker scanning mutation analyses of the mouse β-globin promoter (*15*). Identical levels of transcription are observed with the intact mouse and the hybrid β-globin gene. Both the test and reference plasmids contain the SV40 transcription enhancer sequence located 800 bp downstream from the poly(A) addition site of the human β-globin gene. The plasmids also contain a 300-bp GC-rich DNA sequence and the bacteriophage M13 origin of replication immediately upstream from the promoter; these DNA sequences were used in the saturation mutagenesis procedure for obtaining the promoter mutations (*21*). The presence of these sequences does not affect the levels of transcription from the wild-type and mutant β-globin promoters.

The effects of single base substitutions in the β-globin promoter were determined by comparing the levels of correctly initiated RNA derived from the test and reference plasmids cotransfected into HeLa cells (*25*). A 1:1 mixture of supercoiled test and reference plasmids was transfected into HeLa cells as a calcium phosphate coprecipitate. Forty-eight hours after transfection, total cellular RNA was prepared, and the relative quantities of β-globin transcripts from the test and reference genes were determined by a 5′ S1 nuclease assay (*26*). A 156-nucleotide (nt) single-stranded DNA probe, labeled at its 5′ end and complementary to 80 nt of the 5′ end of the test gene mRNA transcripts and to 76 nt of the 5′ flanking region of the gene, was prepared and annealed in molar excess to the mRNA samples (Fig. 2). The probe forms an 82-bp duplex with mRNA synthesized from the test genes, but forms an imperfect duplex containing an 8-nt "loop-out" at +26 when annealed to mRNA synthesized from the reference gene. On treatment of the RNA-probe mixture with S1 nuclease, the test gene transcripts yield an 82-nt protected probe fragment, whereas transcripts from the reference genes generate a 48-nt protected probe fragment as a result of nuclease cleavage at the 8-nt loop out in the probe. These labeled protected products were fractionated by electrophoresis on denaturing polyacrylamide gels and detected by autoradiography. The ratio

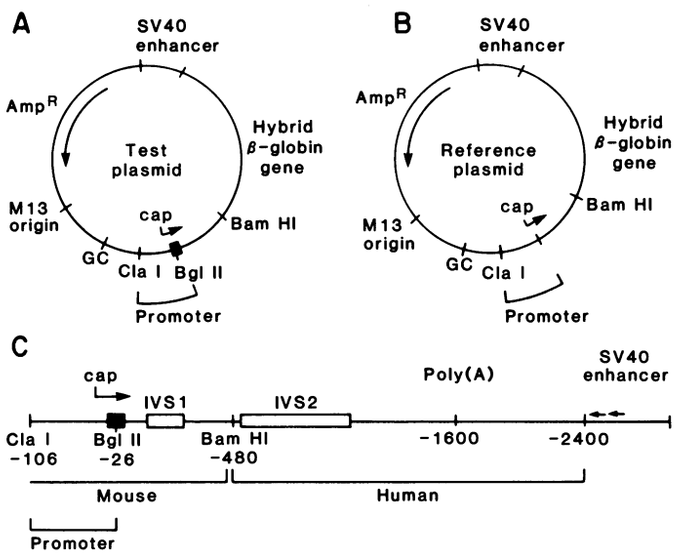

Fig. 1. Diagram of test and reference β-globin gene plasmids. (A) Test gene plasmid. Counterclockwise from the top, the plasmid contains pBR322 sequences from 4363 to 2440, including the ampicillin resistance gene; 571 bp of the bacteriophage M13 origin of replication (from 5943 to 5372); a 300-bp GC-rich sequence called the GC-clamp, which is used in the mutagenesis scheme (*21*); the mouse β-major globin gene promoter from −106 (Cla I site) to +26 (Bgl II site); the mouse β-major globin gene to a Bam HI site in the second exon at +475; the human β-globin gene from a Bam HI site at the analogous position in the second exon to +2400 in the 3′ flanking sequences of the gene; and a 300-bp SV40 DNA fragment containing the viral enhancer sequence. The unique Bgl II site was created by inserting an 8-bp synthetic oligonucleotide linker (CCCATGGG) into a Hinc II site at +26 of the mouse β-major globin gene. This linker allows mRNA molecules synthesized from this plasmid to be distinguished from β-globin transcripts synthesized from the reference gene. Mutant test genes were made by inserting the mutant promoter fragments into the Cla I–Bgl II sites of the test plasmid. (B) Reference gene plasmid. The plasmid is identical to the test gene plasmid except that it lacks the Bgl II linker at +26 in the β-globin gene. Therefore, β-globin transcripts made from this gene are eight nucleotides shorter than test gene transcripts. (C) Detailed map of the hybrid β-globin test gene. The promoter fragment extends from a Cla I linker inserted at −106 to a Bgl II linker inserted at +26. The mouse β-major globin gene extends from the Bgl II site to a Bam HI site appearing in the first intron at +475. The human β-globin gene is fused to the mouse gene at the common Bam HI site in the first intron, and extends through the poly(A) site at +1550 to a Pst I site at +2163. At this junction, the SV40 enhancer region is attached to the gene.

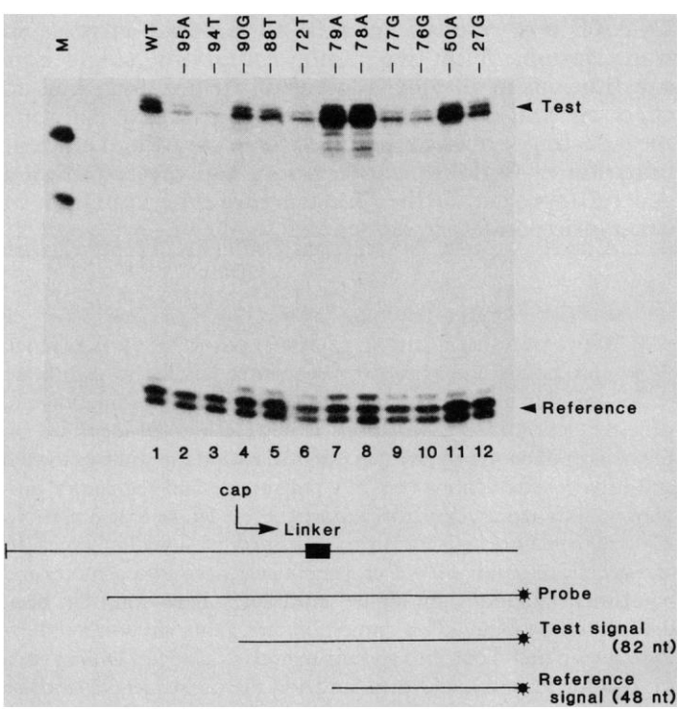

Fig. 2. A typical S1 nuclease analysis of β-globin transcripts. "Test" and "Reference" indicate the positions of the S1 nuclease–protected DNA fragments expected for the test and reference RNA molecules, respectively. (Lane 1) Wild-type test promoter. (Lanes 2 to 12) Mutant test promoters with the names indicated at the top of each lane. For example, −95A in lane 2 refers to a mutant promoter carrying a G to A transition at −95 in the test gene promoter. The lines drawn at the bottom of the figure represent the test gene containing a Bgl II linker at +26, the single-stranded DNA probe complementary to the wild-type test gene region, and the protected probe fragments generated by S1 nuclease treatment when the probe is annealed to RNA derived from the test and reference β-globin genes.

of radioactive signals corresponding to transcripts from the test and reference genes was quantitated by scanning the autoradiograms with a densitometer. These data were used to calculate the relative transcription level (RTL) of each mutant promoter (*5*) (legend to Fig. 3).

Analysis of mutants. The results of a typical transcription analysis of base substitutions in the β-globin promoter are presented in Fig. 2. An example of a base substitution that does not affect the level of transcription is a G to A transition at −50. The level of transcription observed with this template (lane 11) is indistinguishable from that of the wild-type promoter (lane 1). In contrast, mutations at several other positions within the promoter result in a significant decrease in transcription compared to the wild-type promoter (lanes 2 to 6, 9, 10, and 12). In addition, two promoter mutations result in a three- to fourfold increase in transcription (lanes 7 and 8). These up promoter mutations are located at positions −79 and −78.

The results of a transcription analysis of all of the single base mutations in the β-globin promoter analyzed are presented in Fig. 3 and Table 1. Examination of the histogram reveals that most single base substitutions have no effect on the level of β-globin transcription. As predicted from earlier studies (*12–15*), base substitutions that alter the RTL are primarily localized to the three highly conserved regions within the promoter. Single base substitutions in the CACCC box region result in a 2.5- to 10-fold decrease in the level of transcription (Fig. 3 and Table 1), while mutations in the nucleotides immediately surrounding this region have no effect on the RTL's.

The second region of the promoter that is sensitive to single base substitutions is the CCAAT box. The sequence of this region is 5′ GGCCAATCT 3′ in β-globin genes from a number of different species (*16*). The G to A transitions at positions −79 and −78, the first two bases in this region, result in a significant increase in the RTL (Figs. 2 and 3 and Table 1). In contrast, both a transition and a transversion mutation at the next nucleotide, a C at −77, result in approximately tenfold decreases in the RTL. Single base mutations in the remainder of the positions within the CCAAT box lead to a three- to fivefold decrease in the RTL. A promoter containing two base substitutions, one at −75 and the other at −74, results in a 40- to 50-fold decrease in the RTL. Single base substitutions in the region between the CCAAT and TATA boxes have no effect on the RTL, with one exception. Both C to T and C to G substitutions at −37 cause 2.5-fold decreases in the transcription level (Table 1). The nucleotides immediately surrounding −37 have no effect on the RTL.

The third region of the promoter that is sensitive to single base substitutions is the TATA box (Fig. 3). Single base substitutions in the TATA box result in a two- to fourfold decrease in the RTL (Fig. 3 and Table 1). In addition to the observed decrease in transcription levels, some of these mutations cause a slight alteration in the site of transcription initiation.

A number of promoter mutations between the TATA box and the cap site, and between the cap site and +26 were analyzed, but a difference in the RTL was observed only with the +1 A to G transition. The significance of the twofold decrease in the level of accumulated transcripts observed with this mutation is uncertain, since the base substitution changes the sequence of the mRNA. The observed decrease could therefore be the result of a decrease in the stability of globin RNA, possibly due to less efficient capping at the 5′ end of the transcript.

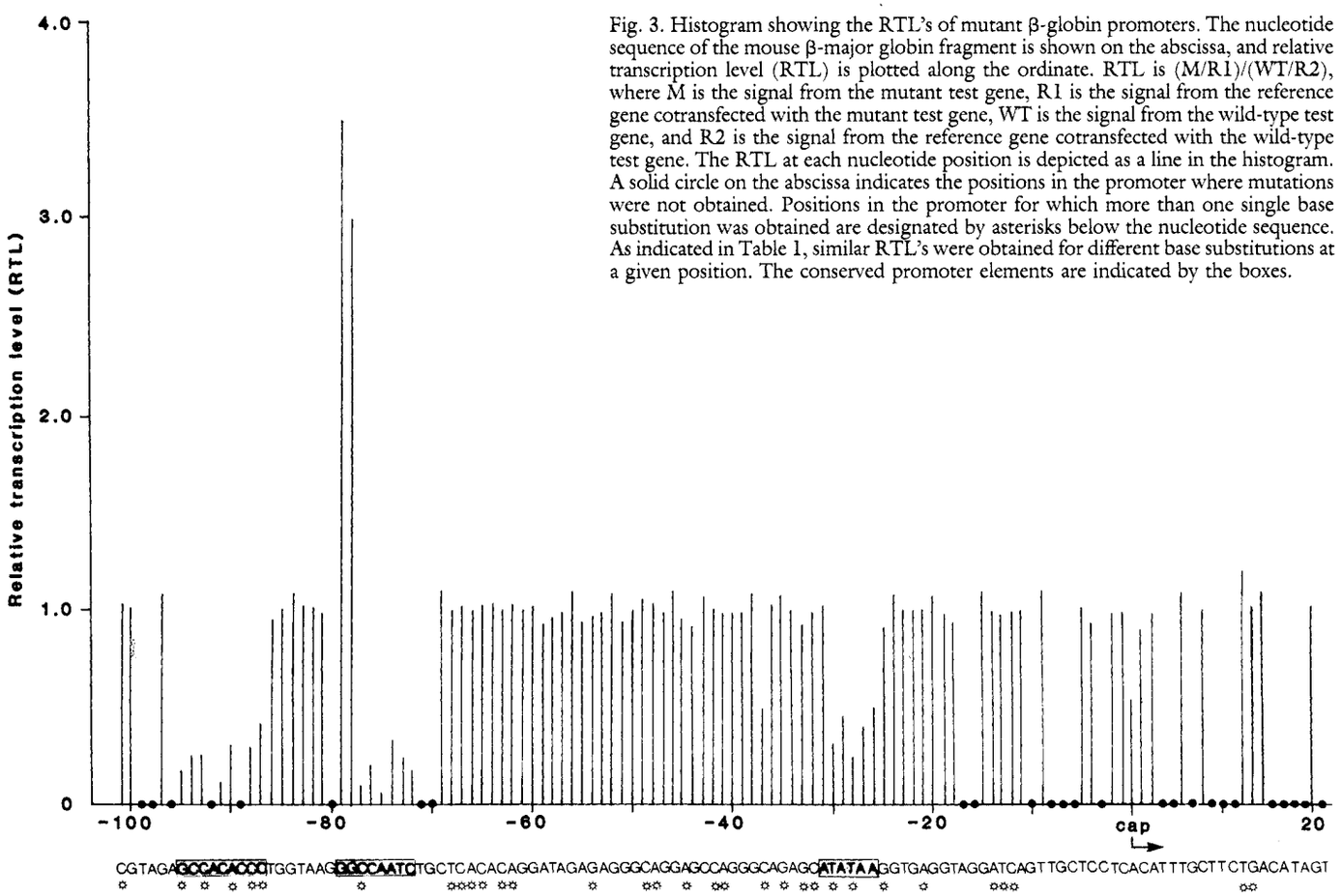

Fig. 3. Histogram showing the RTL's of mutant β-globin promoters. The nucleotide sequence of the mouse β-major globin fragment is shown on the abscissa, and relative transcription level (RTL) is plotted along the ordinate. RTL is (M/R1)/(WT/R2), where M is the signal from the mutant test gene, R1 is the signal from the reference gene cotransfected with the mutant test gene, WT is the signal from the wild-type test gene, and R2 is the signal from the reference gene cotransfected with the wild-type test gene. The RTL at each nucleotide position is depicted as a line in the histogram. A solid circle on the abscissa indicates the positions in the promoter where mutations were not obtained. Positions in the promoter for which more than one single base substitution was obtained are designated by asterisks below the nucleotide sequence. As indicated in Table 1, similar RTL's were obtained for different base substitutions at a given position. The conserved promoter elements are indicated by the boxes.

Eukaryotic promoter elements. Detailed characterization of different eukaryotic promoters has revealed a common pattern of organization (*27, 28*). Most promoters consist of a TATA box located approximately 30 bp upstream from the start site of transcription, and upstream promoter elements located between 30 and 100 bp upstream from the cap site. In vivo and in vitro transcription analyses and DNA binding studies have thus far identified five different upstream promoter elements:

I	GGGGCGG (*28, 29*)
II	CCAAT (*10, 16, 19, 22, 30*)
III	GCCACACCC (*14, 17*)
IV	GGCCACGTGACC (*31, 32*)
V	ATGCAAAT (*33, 34*)

Each of these sequences is required for wild-type levels of transcription in the promoters in which they are found. Moreover, each element binds to a specific transcription factor in vitro. Different promoters appear to have different combinations of one or more of these elements. Multiple copies of element I are present in the promoter regions of a number of viral and cellular genes (*11, 28, 29*). Two copies of this element and one copy of element II are present in the herpesvirus thymidine kinase promoter (*10, 30*). As mentioned above, β-globin promoters contain one or more copies of elements II and III. Element IV has been identified only in the adenovirus major late promoter (*31, 32*). Element V, in contrast, is present in the mouse immunoglobulin heavy chain gene promoter (*33*), and is also located 225 bp upstream from vertebrate U1 and U2 small nuclear RNA (snRNA) genes (*34*). At least two of these elements are also found within essential regions of transcription enhancer sequences; elements III (*35, 36*) and V (*33*) are found in the SV40 and heavy chain immunoglobulin gene enhancers, respectively. It seems likely that promoter strength is determined, at least in part, by the strength, number, and arrangement of these types of elements in the region immediately upstream from the TATA box. Although the five promoter elements mentioned above are the most thoroughly characterized to date, it is likely that many other such elements will be found within other promoters. In addition to these elements, which are required for constitutive promoter function, a number of other elements have been identified within inducible and tissue specific enhancers and within yeast upstream activating sequences (*27*).

The CACCC box element. Single base substitutions in the sequence GCCACACCC decrease the level of transcription from 2.5- to 10-fold. The functional 5′ boundary of this sequence is either −96 or −95, since a mutation at −96 was not obtained. This upstream promoter element is present only once within the mouse β-globin promoter, but appears to be duplicated in many other β-globin promoters (*14, 16, 18*). In fact, two copies of this sequence are required for maximal levels of transcription of the rabbit β-globin gene (*14*). The existence of a naturally occurring β-thalassemia promoter mutation in the CACCC sequence provides evidence that this element is necessary for maximal levels of β-globin gene transcription in erythroid cells (*37, 38*). In addition, major DNase I and S1 nuclease hypersensitive sites are located near the CACCC sequence in a number of mammalian β-globin genes (*39–42*).

The CACCC box may function as an upstream element in promoters other than globin, and it may play a role in enhancer function. The latter possibility is suggested by the observation that a similar sequence element is required for maximal activity of the SV40 and bovine papillomavirus enhancers (*35, 36*). As with other promoter elements, the CACCC box binds specifically to a factor present in HeLa cell nuclear extracts (*43*). Thus, the CACCC element appears to share many of the properties identified with the more thoroughly characterized upstream promoter elements mentioned above (*28*).

The CCAAT box element. In addition to the CCAAT box sequence of β-globin genes, this promoter element has also been extensively characterized in the herpesvirus tk gene and the Moloney murine sarcoma virus (MSV) LTR promoters (*10, 30*). The functional CCAAT box in the tk promoter is in the opposite orientation relative to the cap site, and the effect of a number of single base mutations within this sequence on in vivo transcription has been determined (*10*). In addition, a linker scanning mutation that

Table 1. Tabulation of the relative transcription levels of mutant mouse β-major globin promoters in HeLa cells. Each mutant gene is named according to the position of the single base substitution in the promoter relative to the start site of transcription. The first letter (A, G, C, or T) following the number indicates the nucleotide present at that position in the wild-type promoter, while the second letter indicates the substituted nucleotide. For example, −101 C-A refers to a mutant promoter in which the C at −101 is changed to an A (only the mutant bases are shown for the two double mutants at positions −75/−74, and −28/−26). RTL denotes the relative transcription level as defined in the legend to Fig. 3.

MUTANT	RTL	MUTANT	RTL	MUTANT	RTL	MUTANT	RTL	MUTANT	RTL	MUTANT	RTL	MUTANT	RTL
-101 C-A	1.05	-82 A-G	0.97	-64 A-C	1.02	-49 C-G	1.11	-35 G-A	1.11	-21 A-G	1.01	+3 A-G	0.99
-101 C-T	0.93	-81 A-G	0.98	-63 C-A	1.00	-48 A-T	1.01	-35 G-T	0.99	-21 A-C	1.00	+6 T-A	1.09
-100 G-A	1.01	-79 G-A	3.50	-63 C-T	0.99	-48 A-G	0.92	-34 A-G	0.91	-20 G-A	1.09	+8 C-T	1.00
-97 G-A	1.10	-78 G-A	2.90	-62 A-C	0.95	-47 G-A	0.99	-33 G-T	1.00	-19 G-A	0.97	+12 T-C	1.21
-95 G-A	0.17	-77 C-T	0.09	-62 A-G	1.10	-46 G-A	1.13	-33 G-A	0.93	-18 T-C	0.95	+13 G-A	0.99
-95 G-T	0.20	-77 C-G	0.10	-61 G-A	0.99	-45 A-C	1.01	-32 C-A	0.99	-15 G-A	1.12	+13 G-C	0.99
-94 C-T	0.25	-76 C-G	0.20	-60 G-A	1.02	-45 A-G	0.96	-32 C-T	0.96	-14 A-C	0.90	+14 A-C	1.10
-93 C-A	0.24	-75G-74T	0.02	-59 A-G	0.91	-44 G-A	0.92	-32 C-G	1.09	-14 A-G	0.98	+14 A-G	0.92
-93 C-T	0.30	-74 A-G	0.33	-58 T-C	0.96	-43 C-A	1.08	-31 A-G	1.01	-13 T-A	0.92	+19 G-A	0.99
-91 C-A	0.11	-73 T-A	0.24	-57 A-G	0.99	-42 C-A	1.01	-30 T-A	0.33	-13 T-G	0.91	+24 G-A	1.01
-90 A-G	0.30	-72 C-T	0.18	-56 G-A	1.12	-42 C-G	1.00	-30 T-C	0.33	-12 C-G	0.99		
-90 A-C	0.23	-69 C-A	1.12	-55 A-G	0.93	-42 C-T	0.93	-29 A-G	0.45	-12 C-T	1.04		
-88 C-T	0.34	-68 T-C	0.98	-54 G-A	0.92	-41 A-C	0.98	-28G-26G	0.25	-11 A-G	0.99		
-88 C-A	0.29	-68 T-G	1.11	-54 G-T	0.95	-41 A-G	1.09	-27 A-G	0.40	-9 T-A	1.11		
-87 C-T	0.40	-67 C-A	1.01	-53 A-G	0.99	-40 G-A	0.99	-26 A-G	0.50	-5 T-C	1.02		
-87 C-A	0.41	-67 C-T	1.05	-52 G-A	1.11	-39 G-A	0.98	-25 G-T	0.92	-4 C-T	0.94		
-86 T-C	0.92	-66 A-G	0.96	-51 G-A	0.93	-38 G-A	1.12	-25 G-A	1.02	-2 T-A	0.98		
-85 G-A	1.01	-66 A-C	0.99	-50 G-A	1.02	-37 C-T	0.56	-24 G-A	1.11	-1 C-T	0.98		
-84 G-A	1.11	-65 C-A	1.00	-49 C-T	1.00	-37 C-G	0.44	-23 T-C	1.00	+1 A-G	0.45		
-83 T-C	1.07	-65 C-T	1.10	-49 C-A	0.98	-36 A-G	1.02	-22 G-A	0.99	+2 C-T	0.90		

disrupts the tk CCAAT box leads to a significant decrease in in vitro transcription (*30*). A comparison of the effects of point mutations within the CCAAT box on transcription in vivo is presented in Table 2. Base changes at the same position of the CCAAT box in different promoters have similar effects.

All of the base substitutions that occur within the CCAAT sequence result in a decrease in transcription in all of the promoters thus far examined. However, an increase in the RTL is observed with certain base substitutions at positions adjacent to this sequence in both the tk and β-globin promoters. Two base substitutions near the tk promoter CCAAT box result in a 1.5-fold increase in the RTL (*10*) (Table 2), while a 3- to 3.5-fold increase is observed when either of the two bases immediately upstream from the β-globin CCAAT box is changed (Fig. 3). Each of the two promoter up mutations in the tk CCAAT box region produce a sequence that is closer to the consensus sequence of globin gene promoters. The nucleotide immediately downstream from the CCAAT box is usually C, while the two nucleotides that precede this sequence are usually purines (*16, 18*). The two mouse β-globin promoter up mutations change the GG dinucleotide immediately upstream from the CCAAT sequence to AG or GA. Each of these three possible dinucleotide pairs is found adjacent to the CCAAT box of other β-globin genes. Although GG occurs at this position in adult β-globin genes from different species, AG and GA are found in fetal and embryonic globin promoters respectively (*16, 18*). It is possible that these up promoter mutations act by affecting the affinity of a transcription factor for the CCAAT box, or by altering the conformation of this factor when it is bound to the promoter (*10*).

Cellular factors that specifically interact with the CCAAT box have, in fact, been identified in nuclear extracts of human HeLa cells (*30*) and rat liver cells (*10*). The HeLa cell factor has been shown to stimulate in vitro transcription from the tk promoter (*30*), while both the HeLa and rat cell factors protect a specific region within and surrounding the tk CCAAT box from deoxyribonuclease (DNase) I cleavage (*10, 30*). The DNase footprint in both cases includes 14 to 16 bases upstream from the CCAAT box and between 4 and 9 bases downstream from the sequence. It is likely that the factor identified in the two types of cell extracts is the same. The HeLa cell factor has also been shown to bind specifically to the human β-globin CCAAT box (*43*).

The TATA box element. As in the case of most other eukaryotic promoters examined (*19, 28*), linker scanning mutations (*15*) and single base mutations (Fig. 3) within the mouse β-globin TATA box result in a significant decrease in the level of accurately transcribed RNA. In addition, some of these mutations alter the start site of transcription. Naturally occurring single base mutations in the TATA box of the human β-globin gene associated with β-thalassemia result in decreases in the transcription level in erythroid cells and in the HeLa cell transcription assay (*44*). Like the upstream promoter elements, cellular factors that specifically interact with TATA boxes have also been identified (*31, 45, 46*). A linker scanning mutation (*15*) and single base mutations (Fig. 3) in the region upstream from the β-globin TATA box also result in small decreases in the RTL. Although the significance of this observation is not clear, it is possible that mutations at −37 affect the affinity of factors that interact with the TATA box.

The fine structure genetic analysis of the β-globin promoter has made it possible to precisely delineate the sequence elements necessary for accurate and efficient initiation of transcription. The validity of the enhancer-dependent HeLa cell transcription assay for analyzing β-globin promoter mutants has been demonstrated by showing that naturally occurring promoter mutations that decrease the level of transcription in human erythroid cells also decrease the level of transcription in the HeLa cell assay (*38, 44*). Moreover, linker scanning mutations and single base substitutions have been shown to result in comparable decreases in the transcription of cloned β-globin genes introduced into mouse erythroleukemia cells and HeLa cells (*15, 47*). Although single base substitutions in the regions that lie outside the conserved promoter elements have no effect on transcription in the HeLa cell assay, we note that some of these sequences are highly conserved in mammalian evolution. It is therefore possible that these sequences play an important role in a function not detected in the HeLa cell assay. Whether they may be involved in globin gene regulation in erythroid cells remains to be determined.

Table 2. Comparison of CCAAT box mutations in several different promoters. The wild-type sequence of the CCAAT box and surrounding nucleotides of four promoters is shown at the top. In the case of the herpesvirus tk promoter, the CCAAT box is in an inverted orientation relative to the cap site of the gene, but is presented in the same orientation as that of the other promoters. The sequence of the mouse and rabbit β-globin CCAAT regions are identical. Single base substitutions in each of the CCAAT boxes are indicated beneath the wild-type sequences, and the observed RTL's are shown to the right of the substitutions. Mouse β-globin, Table 1; rabbit β-globin (*14*); HSV tk and MSV LTR (*10*).

MSV LTR	A	A	C	C	A	A	T	C	
HSV tk	C	G	C	C	A	A	T	G	
Mouse/Rabbit β	G	G	C	C	A	A	T	C	RTL
Mouse β									
	A								3.50
		A							2.90
			T						0.09
			G						0.10
				G					0.20
					G	T			0.02
						G			0.33
							A		0.24
								T	0.18
Rabbit β									
	T	T							0.12
			T						0.24
				T					0.12
					G				0.12
HSV tk									
	G								1.50
		C							0.70
			G						0.18
				G					0.30
					C				0.20
						C			0.30
							G		0.50
								C	1.50
MSV LTR									
			G						0.28
				G					0.48
					C				0.20
						C			0.70
							G		0.30

REFERENCES AND NOTES

1. P. Youderian, S. Bouvier, M. M. Susskind, *Cell* **30**, 843 (1982).
2. B. J. Meyer, R. Maurer, M. Ptashne, *J. Mol. Biol.* **139**, 163 (1980).
3. R. Losick and M. Chamberlin, Eds., *RNA Polymerase* (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1976).
4. M. Ptashne, *A Genetic Switch* (*Cell*/Blackwell, Cambridge, MA), in press.
5. S. L. McKnight and R. Kingsbury, *Science* **217**, 316 (1982).
6. D. Shortle, D. DiMaio, D. Nathans, *Annu. Rev. Genet.* **15**, 265 (1981).
7. M. Smith, *ibid.* **19**, 423 (1985).
8. D. Botstein and D. Shortle, *Science* **229**, 1193 (1985).
9. S. L. McKnight, R. C. Kingsbury, A. Spence, M. Smith, *Cell* **37**, 253 (1984).
10. B. J. Graves, P. F. Johnson, S. L. McKnight, *ibid.* **44**, 565 (1986).
11. D. Gidoni *et al.*, *Science* **230**, 511 (1985).

12. G. C. Grosveld, E. de Boer, C. K. Shewmaker, R. A. Flavell, *Nature (London)* **295**, 120 (1982).
13. G. C. Grosveld, A. Rosenthal, R. A. Flavell, *Nucleic Acids Res.* **10**, 4951 (1982).
14. P. Dierks *et al.*, *Cell* **32**, 695 (1983).
15. P. Charnay, P. Mellon, T. Maniatis, *Mol. Cell. Biol.* **5**, 1498 (1985).
16. A. Efstratiadis *et al.*, *Cell* **21**, 653 (1980).
17. E. Lacy and T. Maniatis, *ibid.*, p. 545.
18. R. C. Hardison, *J. Biol. Chem.* **258**, 8739 (1983).
19. C. Benoist, K. O'Hare, R. Breathnach, P. Chambon, *Nucleic Acids Res.* **8**, 127 (1980).
20. R. Breathnach and P. Chambon, *Annu. Rev. Biochem.* **50**, 349 (1981).
21. R. M. Myers, L. S. Lerman, T. Maniatis, *Science* **229**, 242 (1985).
22. P. M. Mellon, V. Parker, Y. Gluzman, T. Maniatis, *Cell* **27**, 279 (1981).
23. J. Banerji, S. Rusconi, W. Schaffner, *ibid.*, 299.
24. Single base substitutions were generated in the promoter region of the mouse β-major globin gene [D. A. Konkel, S. M. Tilghman, P. Leder, *Cell* **15**, 1125 (1978)] as described (*21*). All plasmids were prepared by the alkaline lysis method [H. C. Birnboim and J. Doly, *Nucleic Acids Res.* **7**, 1513 (1979)], and purified by CsCl–ethidium bromide buoyant density centrifugation [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1982), pp. 93–94]. The concentration of each test DNA was estimated by measuring the ultraviolet absorption spectrum, and by digestion of the plasmid DNA with restriction enzymes followed by agarose-gel electrophoresis and ethidium bromide staining. We estimate that experimental variations due to differences in DNA concentration were less than 10 percent. More than half of the promoter mutants were tested with two separate preparations of plasmid DNA, and no significant variation in the relative transcription level was observed between different preparations of the same test plasmid.
25. Plasmid DNA was introduced into HeLa cells by the calcium phosphate coprecipitation procedure [M. Busslinger, N. Moschonas, R. A. Flavell, *Cell* **27**, 289 (1981)]. For each transfection, 20 μg each of supercoiled test and reference plasmid were mixed and coprecipitated with calcium phosphate in a final volume of 2 ml. This mixture was distributed equally to two 10-cm diameter tissue culture plates containing HeLa cells at approximately 50 percent confluency in 10 ml of medium (DME, 10 percent fetal calf serum, penicillin at 1000 U/ml, and streptomycin at 1 mg/ml). The plates were incubated for 16 hours at 37°C, at which time the medium was removed, the plates were rinsed with phosphate-buffered saline (PBS), and fresh medium was added. After an additional 32-hour incubation at 37°C, total cellular RNA was isolated as described [J. Favalaro, R. Treisman, R. Kamen, *Methods Enzymol.* **65**, 718 (1980)].
26. S1 nuclease analysis was performed as follows. Polynucleotide kinase and γ-labeled ^{32}P ATP (ICN; 3000 Ci/mmol) were used to label a Dde I site at +82 in the wild-type test gene plasmid. The labeled fragment was digested with Hae III, which cleaves at −76 in the promoter region, and the single-stranded, 5′ end-labeled probe fragment was purified by denaturing polyacrylamide gel electrophoresis. The probe, 0.02 pmol (approximately 10^5 count/min) was mixed with 20 μg of total cellular RNA in 30 μl of hybridization buffer (80 percent formamide, 40 m*M* Pipes, *p*H 6.4, 0.4*M* NaCl, and 1 m*M* EDTA). The mixture was heated at 90°C for 10 minutes, incubated at 30°C for 12 to 16 hours, and treated with S1 as described [R. F. Weaver and C. Weissmann, *Nucleic Acids Res.* **7**, 1175 (1979)]. Autoradiographic signals were quantitated as described (*15*).
27. E. Serfling, M. Jasin, W. Schaffner, *Trends Genet.* **1**, 224 (1985).
28. W. S. Dynan and R. Tjian, *Nature (London)* **316**, 774 (1985).
29. ———, *Cell* **32**, 669 (1983).
30. K. A. Jones, K. R. Yamamoto, R. Tjian, *ibid.* **42**, 559 (1985).
31. M. Sawadogo and R. G. Roeder, *ibid.* **43**, 165 (1985).
32. R. W. Carthew, L. A. Chodosh, P. A. Sharp, *ibid.*, p. 439.
33. F. G. Falkner and H. G. Zachau, *Nature (London)* **310**, 71 (1984); T. G. Parslow, D. L. Blair, W. J. Murphy, D. K. Granner, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2650 (1984); Y. Bergman, D. Rice, R. Grosschedl, D. Baltimore, *ibid.*, p. 7041; J. O. Mason, G. T. Williams, M. S. Neuberger, *Cell* **41**, 479 (1985); H. Singh, R. Sen, D. Baltimore, P. A. Sharp, *Nature (London)* **319**, 154 (1986).
34. A. Krol, E. Lund, J. E. Dahlberg, *EMBO J.* **6**, 1529 (1985); G. Ciliberto, R. Buckland, R. Cortese, L. Philipson, *ibid.*, p. 1537; M. Ares, M. Mangin, A. M. Weiner, *Mol. Cell. Biol.* **5**, 1560 (1985); I. W. Mattaj, S. Leinherd, J. Jiricny, E. M. DeRobertis, *Nature (London)* **316**, 163 (1985).
35. H. Weiher, M. König, P. Gruss, *Science* **219**, 626 (1983).
36. H. Weiher and M. R. Botchan, *Nucleic Acids Res.* **12**, 2901 (1984).
37. S. H. Orkin *et al.*, *Nature (London)* **296**, 627 (1982).
38. R. Treisman, S. H. Orkin, T. Maniatis, *ibid.* **302**, 591 (1983).
39. J. B. Margot and R. C. Hardison, *J. Mol. Biol.* **184**, 195 (1985).
40. M. Groudine, T. Kohwi-Shigematsu, R. Gelinas, G. Stamatoyannopoulos, T. Papayannopoulou, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7551 (1983).
41. M. Sheffery, P. A. Marks, R. A. Rifkind, *J. Mol. Biol.* **172**, 417 (1984).
42. E. R. Hofer, R. Hofer-Warbinek, J. E. Darnell, *Cell* **29**, 887 (1982).
43. K. Jones, J. Kadonaga, R. Tjian, personal communication.
44. S. E. Antonarakis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1154 (1984).
45. B. L. Davison, J. M. Eggly, E. R. Mulvihill, P. Chambon, *Nature (London)* **301**, 680 (1983).
46. C. Parker and J. Topol, *Cell* **36**, 357 (1984).
47. R. M. Myers, K. Tilly, T. Maniatis, unpublished results.
48. We thank S. McKnight and R. Tjian for communicating results prior to publication, and S. Goodbourn, B. Kingston, and S. McKnight for critical reading of the manuscript. Supported by grants from the NSF and NIH (T.M.), by Special Fellowship of the Leukemia Society of America, Inc. (R.M.M.), and by a postdoctoral fellowship from the Damon Runyon Walter Winchell Cancer Fund (K.T.).

17 December 1985; accepted 13 March 1986