Volume 10 Number 16 1982 Nucleic Acids Research

Sequence requirements for the transcription of the rabbit β-globin gene *in vivo*: the −80 region

G.C.Grosveld*, A.Rosenthal+ and R.A.Flavell

Laboratory of Gene Structure and Expression, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Received 17 May 1982; Revised and Accepted 21 July 1982

ABSTRACT

We have performed a detailed analysis of DNA sequences in the −80 region of the rabbit β-globin gene that are required for transcription. A variety of rabbit β-globin gene templates deleted at various sites in the −80 region were linked to an SV40-plasmid recombinant and introduced into HeLa cells; under these conditions the rabbit β-globin gene is expressed from its own promotor. The results show that the conserved GGCCAATCT sequence is required for efficient transcription in vivo and further identify another sequence in the region from about −81 to −96 which is also required for transcription in vivo.

INTRODUCTION

The comparison of DNA sequences in the 5' flanking regions of a large number of eukaryotic genes has pointed to two conserved DNA regions. The first is located about 30 nucleotides upstream from the cap site and has the canonical sequence $\overset{T}{C}ATA\overset{T}{A}AG$ (1, 2, 3). This DNA sequence is required for specific transcription both in in vitro systems (4, 5, 6, and P. Dierks, personal communication) and in vivo. Its absence in vivo causes the loss of specificity in the choice of the site of initiation of RNA synthesis (7, 8, 9, 10); loss of this sequence is commonly (7, 10) but not always (8, 9) a strong down mutation.

The second conserved region, called the CAAT box, is localized about 80 nucleotides upstream from the site of initiation and has the canonical DNA sequence $GG\overset{T}{C}CAATCT$ (3). The evidence for the involvement of this DNA sequence in transcription is controversial. Deletion mutants which lack segments of increasing length at the 5' side of a number of eukaryotic genes have been analysed for their ability to generate specific transcripts in vivo or in vitro. Removal of the CAAT box region has little, if any, effect on transcriptional efficiency in vitro (4, 5, 6). In vivo the results are equivocal - most systems studied show that the removal of DNA sequences in this region causes a strong down mutation (11, 8, 9, 12, 10), but in at

© IRL Press Limited, Oxford, England.

0305-1048/82/1016-4951$2.00/0

least one case the deletion of these DNA sequences caused a moderate up mutation (7). In addition, in the case of the Herpes Simplex (HSV) thymidine kinase (tk) gene the removal of sequences in the CAAT box region caused only a moderate down mutation; instead, the deletion of DNA sequences somewhat upstream from this region had a strong down mutational effect (9). It is not clear whether the effect seen for the HSV tk gene is specific to that gene, or of more general relevance. In the other genes studied, the deletions analyzed do not distinguish between a CAAT box effect or an effect of the absence of those DNA sequences surrounding this DNA sequence. The interpretation of these results is also compounded by the fact that different genes have been analyzed in different transcriptional assay systems.

The results described above were, for the most part, based on experiments where the 5' flanking DNA sequences of a given gene were deleted and the deleted DNA recloned in bacteria; this results in a mutant DNA template where the truncated 5' extragenic DNA sequences abuts plasmid DNA sequences. Clearly the presence of the plasmid DNA at a given position relative to a promotor sequence might affect the results. In this article we have generated a series of specific short deletions of the DNA sequences in and/or around the CAAT box region of the rabbit β-globin gene. The mutated DNA, present on an SV40 vector, is then introduced into HeLA cells and its ability to direct the synthesis of β-globin mRNA assayed as described previously (10). The results show that removal of part or all of the GGCCAATCT sequence strongly decreases the transcriptional efficiency of the deleted template in vivo. The results also identify DNA sequences to the 5' side of this structure which are also required for transcription.

MATERIALS AND METHODS

The DNA manipulations carried out to generate the mutant DNAs are only described in outline here to save space (on request by the editors). Details of constructions can be obtained on request.

Construction of mutants in the -80 area of the rabbit globin gene

(i) Construction of mutants with deletions in and around the CAAT box (position -75).

To construct deletions in the CAAT box area of the rabbit β-globin gene we used the transient expression vector pRβ 3'SVBglII; this plasmid has been described previously (10) and contains the BamHI-EcoRI fragments of SV40 and pBR328 respectively, together with a 2070bp rabbit β-globin BglII partial fragment, as is shown in Fig. 1. This plasmid contains two BalI sites; one

at the 5' end of the CAAT-box at position -75 (where +1 is the cap nucleotide of the mRNA), and one at position +1164 in the third exon of the rabbit β-globin gene.

To optimize the yield of CAAT-box mutants, we eliminated the BalI site at +1164 by inserting a PstI-linker at that site.

20μg of this DNA, pRβ 3'SV BglII-3'Pst (see Fig. 1) was linearized with BalI at position -75. Deletion mutants around this site were produced following a procedure described previously (6). This involves digestion with Klenow-PolI in the presence of either dATP or dTTP followed by S_1 nuclease digestion and repair synthesis with reverse transcriptase. The DNA was circularized using T4 ligase transfected into E.coli HB101 and the deletion mutants screened in mini preps by HaeIII digestion. DNA sequence analysis (13) showed four different mutants; deletion mutant -76 to -77, -74 to -81, -60 to -83 and -61 to -84.

Deletion mutant -72 to -79 was constructed using the mutants -76 to -77 and -74 to -81. Both plasmid DNA's (50μg) were linearized with KpnI (Fig.1), denatured and reannealed overnight in 100μl 3 x SSC 50% formamide at 43°C.

The mixture of homo- and heteroduplex molecules was taken up in 300μl S_1-nuclease buffer (14) and digested with 3000 units S_1-nuclease at 37°C. Under these conditions, S_1-nuclease will not only cleave the DNA looping out from the heteroduplex, but will also degrade AT-rich sequences at the termini of a DNA duplex (10). Since the DNA sequence upstream from position -73 is GC-rich (see Table I), we reasoned that S_1-nuclease would specifically remove the AAT segment of the remainder of the CAAT-box. After S_1-nuclease treatment, the termini of the digested fragments were made blunt by repair synthesis using reverse transcriptase and the DNA was digested with XmaIII and BamHI (see Fig. 1). Both the resected BalI-XmaIII (0.995kb) and the resected BalI-BamHI (0.55kb) fragment were isolated from a 1% LMA-gel and ligated into the 9.2kb XmaIII-BamHI fragment of pRβ 5'SV BglII-3'Pst (see Fig. 1). Mini preps were screened and of five mutants sequenced, one provided the mutant -72 to -79.

(ii) Construction of Pst and HindIII-linker insertion mutants of pRβ 3'SV BglII-3'Pst at position -75 and derivatives of these.

PstI linker and HindIII linker insertion mutants at position -75 were constructed by ligation of phosphorylated PstI or HindIII linkers into the BalI linearized plasmid pRβ 3'SV BglII-3'Pst as described above. During the Pst-cloning experiment, two clones were picked up with a shortened PstI-PvuII promotor fragment that did not contain a PstI-linker at position -75.

Sequence analysis showed that these two clones contained the deletions -73 to -69 and -64 to -123. The two deletion mutants were therefore picked up by accident, probably due to contamination of the BalI enzyme preparation with aspecific nucleases. To construct deletion mutant -93 to -76, the PstI-linker clone was digested with PstI + SalI + BamHI. Both the 605bp, SalI-PstI (3' end at -96) and the 552bp (5' end at -76) PstI-BamHI fragments were isolated from a 1.5% LMA-gel. These two fragments were ligated into the 8.6kb SalI-BamHI fragment of pRβ 3'SV BglII 3'Pst. The -93 to -76 mutant was identified by the electrophoretic analysis of PvuII + PstI and RsaI + PvuII double digests from DNA mini preparations, which in the case of both digests produces a shortened rabbit β-globin promotor fragment. The precise co-ordinates of the deletion were verified by DNA sequence analysis.

(iii) Construction of deletion mutants round the 5' extragenic PstI site (-96) of the rabbit β-globin gene.

pRβ 3'SV BglII-3' PstI was digested with PstI and deletions generated at this site by digestion with Klenow PolI essentially as described (6).

Double deletion mutants were prepared by Klenow PolI exonuclease digestion of PstI cut mutant -69 to -73 as above. This gave mutants -104 to -94 and -112 to -97, both in the mutant background -69 to -73.

(iv) Construction of mutants containing a 20bp histone gene fragment in the PstI site at -96 in the deletion mutant -60 to -83.

Deletion mutant -60 to -83 was cut with PstI and the 20bp PstI fragment of Psamaechinus miliaris histone H1 gene was inserted at this site. Sequence analysis of the clones obtained showed four, six and seven histone inserts.

Transient expression of pRβ 3'SV BglII-mutants in HeLa cells

Transient expression was performed as described elsewhere (10) with the exception that each 90mm dish of HeLa cells was transformed with a misture of 7.5µg pRβ 3'SV BglII-mutant DNA and 7.5µg of pHβ5'SV BglII, an SV40-pBR328 vector containing the 4.7kb BglII human β-globin gene fragment derived from a normal individual (see 15 and 16). Since the latter DNA is also transcribed well in HeLa cells, it was added as an internal standard for the quantitation of the rabbit β-globin mRNA.

S_1-mapping procedure

The S_1-mapping of the rabbit β-globin mRNA in HeLa cell RNA was done as described (10). However, in the experiments described in this article, the amount of mRNA from the mutated rabbit β-globin gene was compared to the

amount of β-globin mRNA from the human β-globin gene added as an internal standard; both mRNAs were detected in a simultaneous S_1-nuclease assay. The rabbit mRNA was assayed with a 5' labelled 221bp BstNI probe (10) and the human mRNA was assayed using a 3' end labelled 148bp HinfI fragment derived from the first exon-first intron boundary of the human β-globin gene (17). The S_1-nuclease digestion products from both mRNA were separated on a 5% polyacrylamide urea gel (13); the lengths of the products were 139NT (rabbit) and 78NT (human) respectively. Both signals on the autoradiogram were scanned using a Joyce Loebl densitometer and since the human signal is a constant, the rabbit β-globin signal from each mutant can be compared via this human β-globin mRNA internal standard. Control experiments established that both probes (5ng each) were present in excess during hybridization. The hybridization signals obtained were approximately 1/100 of the plateau value (not shown).

RESULTS

General Strategy

In the experiments described here we have determined the effects of specific mutations in the CAAT-box region, situated 75bp upstream of the 5' terminus of the rabbit β-globin gene, on the synthesis of β-globin mRNA *in vivo*. To assay for this template activity, we constructed those mutants in a transient-expression vector (18, 16, 10) which contains the entire rabbit β-globin gene, flanked by 425bp of 5' extragenic sequences and 345bp of 3' extragenic sequences, linked to most of the SV40 viral genome and the plasmid pBR328. In particular, this vector contains the 72 base pair enhancer sequence which is required for the *in vivo* expression of genes present *in cis* to this sequence element (see e.g. 18). As a result of the strategy followed, all DNA molecules used for transient expression studies in these experiments are identical except for the small deletions introduced in the CCAAT-box area of the rabbit β-globin gene. Although the 5' flanking sequences of the globin gene in this vector, pRβ 3'SV BglII (see Fig. 1) are relatively short, it gives the same globin mRNA levels as are found for a similar plasmid DNA that contains 1.5kb of 5' flanking sequences, both in transient expression (10) and in stable L-cell transformants (11) which contain the corresponding β-globin gene fragments in a (presumably) integrated form. Fig. 2 shows a simple map of the rabbit gene and its promotor area in which the ATA-box, the capping-box and the CAAT-box are indicated. It also shows the BalI site at position -75 and the PstI site at -96 which we have used to introduce small deletions

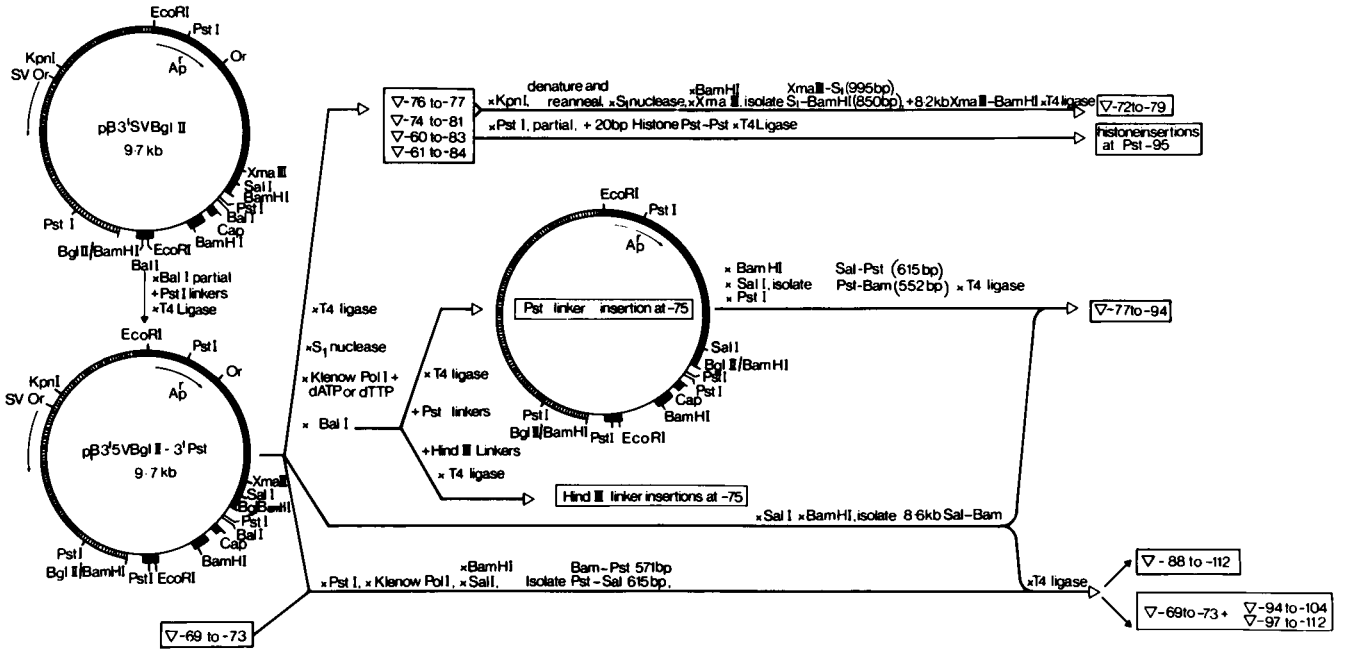


Fig. 1 Cloning strategy used to generate deletion mutants. The mutants were prepared as detailed in Materials and Methods and according to this scheme. Or refers to a replication origin and cap to the 5' end of the β-globin gene.

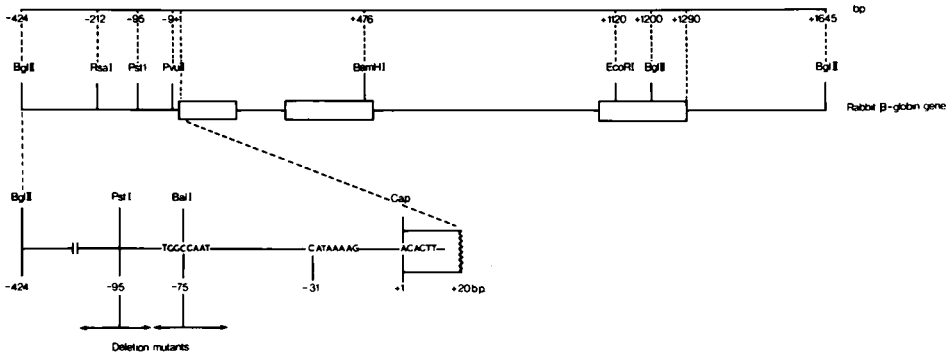
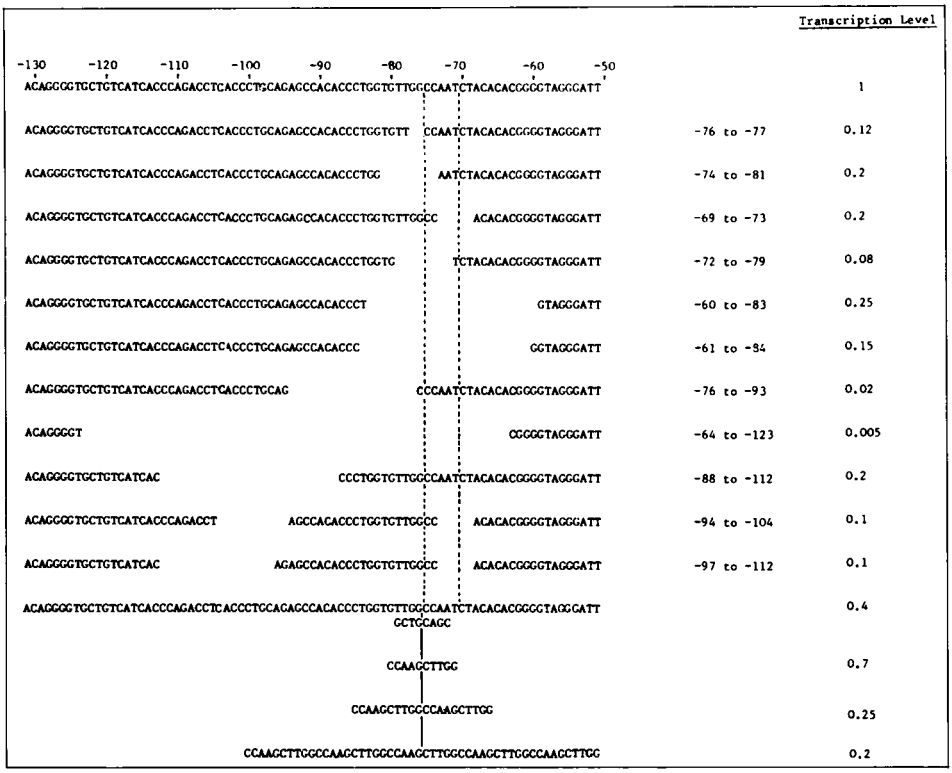


Fig. 2. General structure of the rabbit β-globin gene and the strategy used to generate deletion mutants. The rabbit β-globin gene is shown together with a number of relevant restriction enzyme cleavage sites. The bottom half of the figure shows an expanded view of the promotor region. The position of the CAAT box and ATA box sequences are shown. Deletions were introduced at the BalI site, within the CAAT box or at the PstI site upstream from this site as detailed in Fig. 1 and Materials and Methods.

in this area by exonuclease digestion of the DNA around the two sites. The fine deletion mutants were cloned and their structure established by restriction enzyme mapping and DNA sequencing (see Table I). The details of their construction are described in Materials and Methods.

Each of the mutant DNAs was introduced into HeLa cells by the Ca-phosphate method (19) and cellular RNA was recovered after 36 hours as described previously (10). Although transcription of the rabbit β-globin gene is directed by the β-globin gene promotor, the high levels of expression in this system seem to be dependent on the presence of the SV40 72bp repeat *in cis* on the molecule (18,10). After extraction of the RNA from the HeLa cells, rabbit β-globin mRNA was detected by S_1-mapping, as in reference 20. Since the aim of this study is to relate the effect of each promotor mutation to the transcription efficiency of the gene, we developed a method by which the mRNA level directed by the mutated gene could be compared with mRNA directed by a control β-globin gene present as an internal standard. To this aim we co-transformed the HeLa cells with an equimolar mixture of the respective mutant rabbit DNA and pHβ 5'SV BglII a human β-globin plasmid. This DNA has the same SV40-pBR328 vector as pβ3'SV BglII, but it contains the 4.7kb BglII human β-globin fragment interposed between the SV40 and pBR328 plasmid sequences at the BamHI site (15,16). The rabbit and human β-globin mRNA was detected by simultaneous S_1-mapping as shown in Materials and Methods and



The transcription efficiencies of deletion and insertion mutants in the -80 region of the rabbit β-globin gene. The sequence in the -80 region of each mutant template is presented, and the gaps show the DNA residues deleted. Transcription level was determined by scanning the autoradiographs of the S_1 mapping experiments and correcting the signal for the deletion mutants by reference to the internal human β-globin gene control.

Fig. 3. From control S_1-mapping experiments using in vivo rabbit and human globin mRNA, we concluded that these two probes only detect homologous mRNA (not shown); the signals obtained in the S_1-mapping experiments are directly dependent on the amount of the respective β-globin mRNA present under conditions, used in this paper, where the DNA probes are in excess during hybridization (see Materials and Methods). After autoradiography of the polyacrylamide-urea gel, the S_1-signals of both human and rabbit β-globin RNA are scanned by a densitometer. Since the human signal is used as an internal standard, all rabbit mRNA signals can be compared with each other via this human mRNA control.

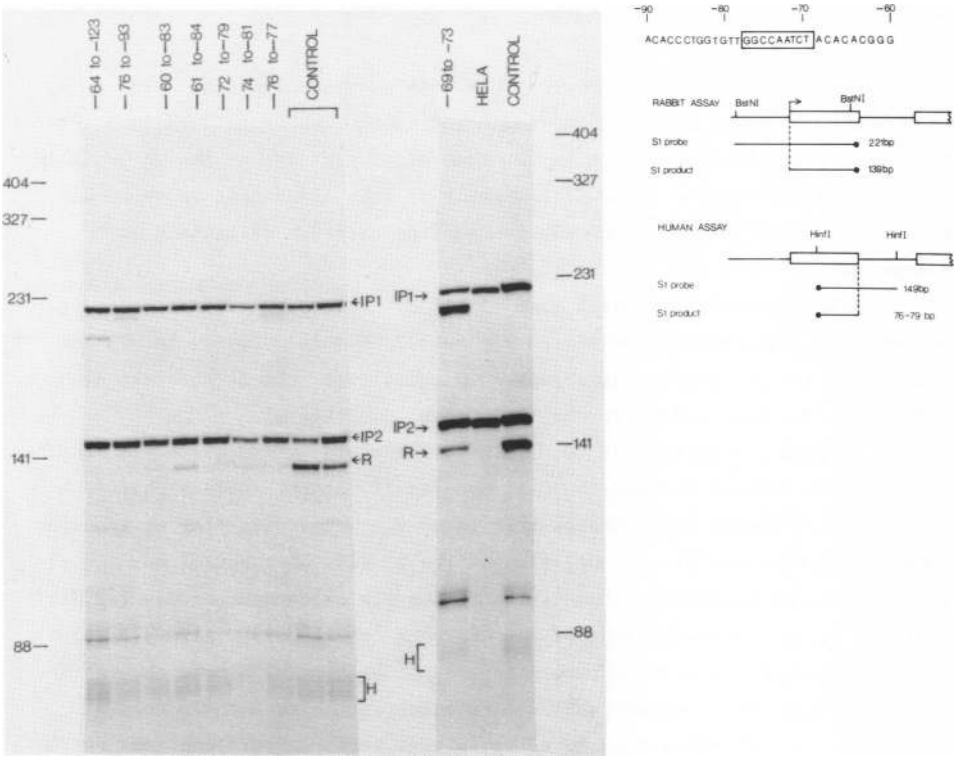


Fig. 3. Transcription in HeLa cells of the rabbit β-globin gene mutated in the -80 region. HeLa cells were transformed with the deletion mutants indicated in the presence of an equal amount of the same SV plasmid carrying the human β-globin gene. Transcripts were assayed by S_1-nuclease mapping using the probes indicated in the scheme below the figure. Transcripts initiated at the cap site of the rabbit β-globin gene give an S_1 nuclease resistant fragment of 139NT (labelled R in the figure); the SV40 rabbit β-globin plasmid also generates RNAs which are probably caused by splicing of an upstream transcript to nucleotide +42 to +48 of the rabbit β-globin gene (10). These transcripts give the cluster of fragments of about 88 nucleotides and have been described previously (11, 10).

The human β-globin gene transcripts are assayed by a HinfI fragment which generates a cluster of S_1 nuclease resistant fragments of 76-79NT (labelled H in the figure). The multiplicity of fragments is due to the redundancy of the DNA sequence at the 5' end of the first intron and the 5' end of the second exon and has been discussed extensively elsewhere (e.g. 16). A residual amount of the input fragments used is still visible (IPI at 221NT and IP2 at 149NT). As controls mock transformed HeLa cells (HELA) and cells transformed with the normal rabbit and human β-globin gene plasmids (CONTROL). The numbers at the side of the figures refer to the sizes of the labelled DNA markers.

The DNA sequence of the CAAT box region is shown; the conserved DNA sequence is boxed.

Sequence requirements of the rabbit β-globin CAAT-box area for transcription in vivo

We first analyzed a number of CAAT-box deletion mutants. The extent of deletion in these DNAs varies from deleting the two G-residues immediately to the 5' side of the CAAT-box to the removal of part of, or the whole CAAT-box and its surrounding area. The result of this experiment is shown in Fig. 3 and the effect of each deletion on transcription efficiency is summarized in Table I.

Deletion mutants lacking part of all of the conserved CAAT box sequence (3, 21), that is deletions -76 to -77, -74 to -81, -69 to -73 and -72 to -79 are all relatively strong down mutations; the effects are eight, five, five and twelve fold respectively (Fig. 3, Table I).

In deletion mutant -76 to -77 the two G residues immediately 5' to the CAAT-box are removed leaving the CAAT-box itself intact. Since this is a strong down mutation it indicates that these two G-residues play an important role in the CAAT-box area. Deletion of the 5' side of the CAAT box (in -74 to -81; the CC residues of the CAAT box plus six extra nucleotides further to the 5' side) or the 3' side of the CAAT-box (-69 to -73; AATCT) results in down mutations of intermediate strength. Deletion of almost the entire CAAT-box (CCAA plus four extra 5' nucleotides) in the mutant -72 to -79 results in the strongest down mutation in this series. This confirms the sequence requirements of both the CAAT-box and the two G-residues for transcription of the rabbit β-globin gene in vivo.

In contrast to this observation, deletion mutants -60 to -83 and -61 to -84 (see Fig. 3) are weaker mutations than -72 to -79 (five to seven-fold, see Table I), although the entire CAAT-box and surrounding nucleotides are removed. This paradoxical result might suggest that upstream sequences in these mutants now located at the position of the CAAT-box could partly take over its function. This idea was tested by insertion of multiple copies of a 20bp sequence (that is, about the same length as the deletion) into the 5' extragenic PstI site, now at position -72 relative to the cap site in these mutants. The sequence inserted is a PstI fragment from the HI histone gene of a cloned Psamaechinus miliaris histone DNA cluster (22) which is non homologous to the CAAT box area sequence (see Fig. 4). If the above arguments are correct, the insertion would displace the putative sequence, which would compensate for the lack of the CAAT region in those mutants, further to the 5' side and result in an additional down mutation; the final level of transcription in these deletion/insertion mutants should be comparable to

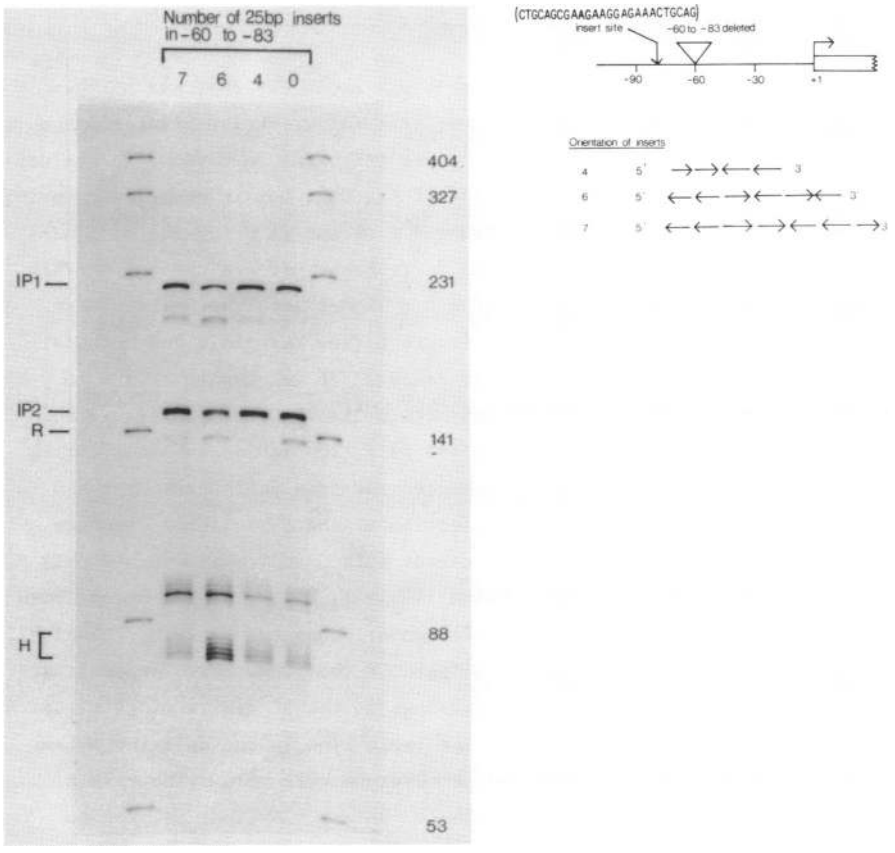


Fig. 4. The effect of insertions of a 20bp DNA fragment at position -72 of mutant template -60 to -83 on the transcription efficiency in HeLa cells. The 20bp PstI fragment with the DNA sequence indicated was introduced into this template as described in Materials and Methods and the orientation and number of inserts present at each site is indicated. The assay of transcription was performed as described in Materials and Methods and the terminology is the same as in Fig. 3 ; the rabbit fragments (R) and the human fragments (H) are indicated. The numbers on the side are the fragment sizes of the labelled DNA markers used.

deletion mutant -72 to -79. Fig. 4 shows that insertion of four, six and seven copies of the histone sequence indeed results in an extra three to ten fold down mutation (a total effect of 12 to 40 fold) which brings transcription efficiency of this mutant template to a level comparable with that of deletion mutant -72 to -79.

Requirement of sequences to the 5' side of the CAAT-box for transcription in vivo

McKnight et al. (9) have shown that DNA sequences upstream of the CAAT-box, that is at -95 to -105, are required for effective transcription of the Herpes Simplex virus thymidine kinase (tk) gene in vivo. To evaluate the role of DNA sequences to the 5' side of the CAAT box in the transcription of the rabbit β-globin gene, we constructed a mutant in which the sequences in between the PstI site at -96 and the CAAT-box at -75 are removed. This construction involved the insertion of a Pst linker into the BalI site at -75 followed by deleting the 26bp Pst fragment (see Materials and Methods). The result of this manipulation is that residue -76 is changed from G to C in addition to the deletion of 18bp (see Table I). As shown in Fig. 3 this deletion causes a strong down mutation (50 fold, see Table I). This strong down mutational effect can only be accounted for in a small part by the G to C conversion at position -76, since insertion of the PstI-linker sequence into the BalI site at -75, which also causes this G to C mutation, is only a moderate down mutation as is shown below (Table I, Fig. 7). The strong down mutation must, therefore, be due to deletion of sequences between -96 and -75; this region must in turn, therefore, be part of the promotor. In contrast to this result, the deletion of DNA sequences at the 5' limits of this region causes a more moderate down mutation (see Table I). In the deletion mutant -88 to -112 a five fold down mutation is observed (Fig. 5). The result suggests that the loss of sequences between -88 and -76 (that is, the non-overlapping regions of these two deletions) accounts for the strong down mutation in the mutant -94 to -76.

Because the deletion in mutant -88 to -112 extends to the 5' side beyond the PstI site at -96, we have investigated the contribution of those 5' sequences to the transcriptional efficiency of this mutant DNA template. Analysis of mutants in this area could map the 5' boundary of the upstream segment of the rabbit β-globin promotor region. We therefore generated double deletion mutants in the background of the CAAT-box mutant -69 to -83 (missing AATCT, see Table I); the second deletion was introduced at the PstI site at -96 and extends for various distances further to the 5' side. As is shown in Fig. 6, the transcription efficiency of the double deletion mutant templates -94 to -104 and -97 to -112 is 40% and 50% of the single deletion. We conclude that deletion of DNA in the region from -94 to -112 has relatively little effect on the efficiency of transcription.

The observation that both the CAAT box and the sequences to its 5'

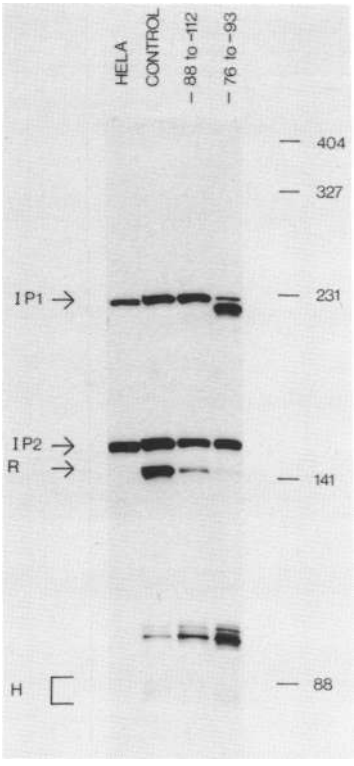


Fig. 5. Transcription efficiency of the rabbit β-globin gene mutated in the region upstream from the CAAT box. The co-transformations of the rabbit and human β-globin genes and the transcription assays were performed as described in Materials and Methods and Fig. 3. The mock transformed HeLa cells (HELA) and normal rabbit β-globin gene (CONTROL) were assayed in parallel to the mutant templates indicated.

side are actually part of the promotor is confirmed by transcription of a mutant that lacks the entire region under study. As shown in Fig. 3 deletion mutant -64 to -123 is the strongest down mutation analysed in this study (about 200 fold; see Table I) and transcription products from this template are essentially undetectable.

The effect of DNA insertions in the CAAT-box region on transcription efficiency in vivo

In addition to deletion mutants we have also examined a few mutants that contain DNA linker insertions in the CAAT box region, that is in the BalI site at position -74. Obviously, the effect of such insertions is to displace part of the promotor sequences further to the 5' side and therefore to increase the distance between the 5' and 3' parts of the upstream region of the promotor. A potential disadvantage of this approach is that the DNA linker might have a sequence-specific effect on promotor function.

We have inserted one 8bp PstI-linker, and one, two and five 10bp HindIII linkers into the BalI site of pβ3'SV BglII-3'Pst. The PstI linker

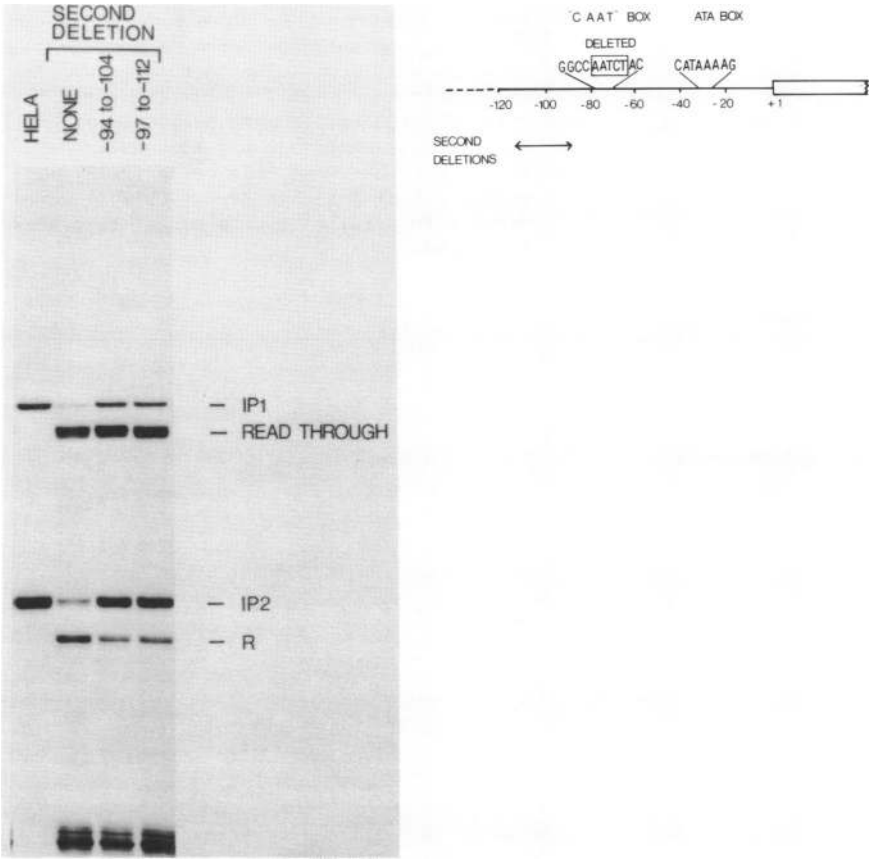


Fig. 6. The transcription efficiency of double mutant rabbit β-globin gene templates with lesions in the CAAT box and upstream from that site. The structures of the deletion mutants are indicated in the sequence below the figure. The second deletions were introduced at the PstI site and those analysed are indicated at the top of the figure. The signal obtained with the human β-globin transcripts was weak in this case and was quantitated in an over exposure of the autoradiogram. The correct rabbit transcripts were also measured relative to read-through transcripts which emanate from an upstream promotor. This approach has also been used previously (10) and gave the same results as the cotransformation approach. The terminology used is described in Fig. 3. A mock transformed HeLa cell control is included.

both changes the GG-dinucleotide at the 5' side of the CAAT box into a GC-dinucleotide (as discussed above) and the relative positions of the two separated promotor areas by 0.8 turn of the DNA helix. The HindIII linker does not have these effects because the HindIII linker sequences restores the GG-dinucleotide (see Table I) and extends the promotor area by one whole turn of the helix.

As shown in Fig. 7 and Table I insertion of one PstI linker causes a somewhat stronger down mutation (about two fold) than insertion of one HindIII linker (1.4 fold). When two or five HindIII linkers are inserted, the transcription efficiency only goes down moderately (1.6 and five fold).

DISCUSSION

We have previously described how the SV40 plasmid vector, in conjunction with HeLa cells (18) can be used to assay RNA processing (16) and transcriptional events (10). It should be recalled that the normal rabbit (18, 10) or human (15, 16) β-globin genes are transcribed specifically and their transcripts spliced correctly (16) in this sytem.

Dierks et al., (11) and Grosveld et al., (10) showed that the DNA sequences in the CAAT box region of the rabbit β-globin gene were required for efficient transcription in vivo in mouse L cells and HeLa cells respectively. In this article, we have been able to pinpoint some of the DNA sequences in this region that affect transcription.

Before considering the results in detail, a number of general points should be made. We have used (mostly) small internal deletions to detect the DNA sequences required for transcription. In the discussion that follows we shall, for the most part, assume that the effect of a given deletion on transcription is the result of the loss of the DNA segment that has been deleted. Though this is generally the most likely explanation, others are also possible. First, the elimination of a given internal DNA sequence causes the juxtaposition of two new DNA sequences. It is possible that presenting a new DNA sequence next to, say, a CAAT box could influence the way a protein interacts with that sequence. Second, the removal of any number of base pairs other than 10 at a given site causes an alteration in the relative position on any particular plane of the DNA helix. Most dramatic, obviously, are deletions of 5, 15, 25 bp and so on, where DNA sequences originally present on the same "side" of the helix now find themselves on opposite sides. This effect is likely to be most serious when a single protein, such as an RNA polymerase interacts with two regions which flank, on either side, the deletions generated. It is not clear how RNA polymerase II interacts with the ATA and the CAAT sequences (it is, indeed, not known whether RNA polymerase II interacts with the CAAT sequence,although it is likely that an interaction occurs with the ATA region). It should be noted that the distance from the CCAAT sequence to the ATA sequence is conserved at about 45bp.

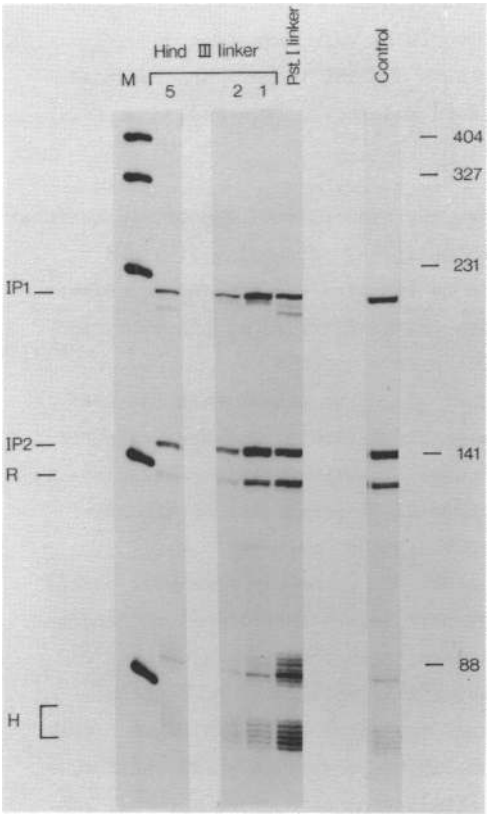


Fig. 7. Transcription efficiency of rabbit β-globin gene templates mutated by the introduction of PstI or HindIII linkers at the CAAT box. The templates generated contain either one PstI linker (GCTGCAGC) or the number of HindIII linkers (CCAAGCTTGG) as indicated. Transcription was assayed as described in Fig. 3. Control indicates transcripts from the normal rabbit β-globin gene, R indicates the rabbit transcripts and H the human transcripts. M indicates the ØX 174 x Taq YI marker fragment.

A second general criticism applies to the use of the SV40-HeLa cell system. First, the expression of a given gene requires the presence of the SV40 72bp enhancer sequence *in cis* (18). Since this is not the normal state of the rabbit β-globin gene, it is possible that certain artefacts could occur. This may not be a serious problem, since our results obtained with the SV40 system (10) agree well with those of Dierks *et al.* (11) in the L cell system in the cases where theese data overlap. In the latter system SV40 enhancers are not present. Second, it should be borne in mind that the rabbit β-globin gene present on the rabbit chromosome might be transcribed differently when present on the SV40 vector or co-transferred into L cells. No pertinent data are available on this point.

The effect of deletions in the "CAAT" box on transcriptional efficiency

A number of the internal deletions that we have studied specifically lack part or all of the 'CAAT' box. All of these deletions cause strong

down mutations which differ somewhat in their quantitative effects:

(i) -76 to -77 (deletion underlined): the loss of these two G residues causes an approximately eight fold down mutation. As in all cases, this deletion causes the juxtaposition of new sequences; in this case, the product is TTCCAATCT.

(ii) -69 to -73, GGCCAATCT: this deletion causes an approximately five fold down mutation; the deletion generates the new sequence GGCCACACA (* indicates mismatch with the wild type CAAT box) where two nucleotides of the newly juxtaposed sequence are by chance the same as that of the original CAAT box.

(iii) -74 to -81 GGTGTTGGCCAATCT; this is a five fold down mutation, in fact, it is a weaker mutation than -76 to -77, despite the fact that the DNA deleted in -76 to -77 is also deleted in this case. It is possible that the juxtaposition of the two upstream G residues to the downstream AAT sequence, partially compensates for the deletion of a segment of the 'CAAT' box; the sequence generated is CTGGAATCT.

(iv) -72 to -79 TGTTGGCCAATCT; this deletion removes the entire CAAT region with the exception of the TC residues at the 3' end of the CAAT sequence. The deletion causes the generation of a new sequence CCTGGTGTCT which bears little resemblance to the CAAT box. The effect of this deletion is also the strongest down mutation of this group of mutants, that is about twelve fold.

(v) -60 to -83 and -61 to -84 are similar large deletions which remove the CAAT region entirely. Yet the effect of these mutations, about four to six fold down, is smaller than all of the above mutants (i) to (iv). We believe that this is caused by the juxtaposition of compensatory upstream DNA sequences at the correct position with respect to the ATA box. Thus, -60 to -83 generates a new sequence which shows a reasonable match with the CAAT box region when displaced by about seven nucleotides:

```
-80          -70         -60
 '            '           '
  GACCCTGCAG  AGCCACACCCG
  *       ** ***     ** ***
 ACACCCTGGTGTTGGCCAATCTAC
 '           '           '
-90         -80         -70
```

The fact that insertion of multiples of 20 nucleotides between residues -71 and -72 has an additional down mutational effect (Fig. 4) suggests that

sequences upstream from -71 in these mutants (note that this numbering refers to the mutant sequence and not the wild type sequence) may indeed compensate for the functional DNA sequences lost in the original deletion. (We cannot, of course, exclude the alternative explanation that the inserted histone DNA segment has a sequence-specific negative effect rather than a spacing effect as discussed above).

In summary then, the CAAT box is required for transcription in vivo. In retrospect it is unfortunate that this name has been chosen since our data suggest that the GG residues that precede the CCAAT sequence are at least as important as the rest of the sequence; sequence comparisons already suggested that this sequence is as conserved as the rest of the CAAT sequence (3).

Sequence requirements to the 5' side of the CAAT region

McKnight et al. (9) have showed that DNA sequences from about -95 to -105 are required for the expression of the Herpes TK gene and that this deletion is a pronounced down mutation. We have analysed the transcription in vivo of a number of templates deleted in this general area.

(i) -76 to -93 is a deletion which removes the 18 nucleotides immediately flanking the CAAT box on the 5' side. This deletion generates the new sequence TGCAGCCCAATCT at the CAAT box. It is obvious that this is a very good CAAT sequence. Nonetheless, this mutation is a 50 fold down mutation, that is, stronger an effect than any of the CAAT box deletions. It is highly unlikely that this results from the GG -> GC change at the 5' side of the CAAT box. We have inserted a PstI linker at -75 to generate the identical sequence TGCAGCCCAATCT. This 8bp insertion has only a slight (about two fold) down mutational effect. More likely, therefore, some of the deleted DNA sequences between -75 to -95 are required for transcription. It is possible that this mutation results from the rotation of the DNA helix by about 1.7 turns, caused by the 18bp deletion. We do not consider this likely because of the very small effect on transcriptional efficiency seen by the insertion of an 8bp PstI linker which causes a rotation of 0.76 turns at position -75.

(ii) Mutant -88 to -112 is an approximately five fold down mutation. This deletion partially overlaps with -76 to -93 (which is a much stronger down mutation) and might therefore indicate that the effect seen in the mutant -88 to -112 is mainly due to the loss of the DNA sequences from -88 to -93, that is, the DNA which overlaps in these two deletions. Equally, since -76 to -93 is a much stronger down mutation than -88 to -112, it is likely that the loss of the sequences from -76 to -88 in the former mutant is of greater consequence than the loss of -88 to -93.

(iii) To investigate the -100 region for a role in transcription, we analyzed a series of deletion mutants in this region. We generated these mutants in the CAAT box mutant DNA -69 to -73 which lacks the AATCT sequence of the CAAT box - this mutation had an approximately five fold down effect. Double mutants lacking, in addition to the above, -94 to -104, or -97 to -112, appear to be transcribed at 40% and 50% of the level seen for the single mutant control which is within the error of our measurements. The elimination of DNA sequences in the region -94 to -112 has therefore only a small effect on transcription efficiency.

(iv) As stated above, deletion of the DNA sequences from -74 to -81 has only a five fold effect on transcription while the deletion from -76 to -93 has a 50 fold effect. It is unlikely therefore that the sequences from -78 to -81 (TGTT) play an important role in the upstream region. This, together with the argument raised in (ii) point to the DNA sequences from -81 to -88 (ACCCTGG) as the major contribution to the upstream part of the promotor.

Considering these data together, the mutations can be grouped in the order (going from weak to strong) -97 to -112 = -94 to -104 <-88 to -112 <<-76 to -93. This suggests that the most important DNA sequences upstream from the CAAT box region that influence transcription are localized from -94 to -76. The fact that mutations from -88 to -112 to -76 to -93 have an effect suggests that a relatively long region (or regions) of at least 30 nucleotides to the 5' side of the "CAAT" box is involved in the transcriptional process. More data will be required to analyse this region in detail.

Insertions at the CAAT box have little effect on transcription

We have attempted to ask whether there is an obligatory topological constraint on the relationship between the CAAT box and the upstream DNA sequences by inserting DNA linkers at the precise boundary of the CAAT box. Although this insertion separates the two G residues at the 5' side of the CAAT box from the CCAAT segment, the HindIII linker has two G residues at its 3' end which reconstitutes this sequence. Similarly the PstI linker insertion generates a new sequence GCCCAAT. The effect of the insertion of a single HindIII or PstI linker is very small, causing only an approximately two fold reduction in transcription efficiency, and the insertion of five HindIII linkers, or 50bp has only a five fold effect. The HindIII linker sequence does bear a strong resemblance to a CAAT box with only a single nucleotide substitution (C->G) CCAAGCT. It is therefore possible that this is the reason for the small effect. Nonetheless, the insertion of five

HindIII linkers at -76 does displace the upstream promotor region with respect to the rest of the promotor and the cap site by 50bp. It is unlikely, therefore, that a precise positioning of these upstream sequences with respect to the ATA box and cap site is required. A similar argument follows from the 8bp displacement caused by the PstI linker insertion. While this sequence does not resemble a CAAT box, it does however show some similarity with the DNA region (-83 to -76) to the 5' side of the CAAT box which it displaces.

GGTGTTGG
GGTGCAGC
(** marks the two mismatched positions)

The upstream region of the rabbit β-globin gene promotor

Our data suggest that an extended upstream sequence element around -80 is required to obtain efficient transcription in vivo. At least one part of this element is the "CAAT" box and of this 'box', the first two residues are of great importance. Deletion of the entire box (as in -72 to -79) is a stronger down mutation than deletion of only part of the conserved sequence.

There seems to be a region, to the 5' side of the 'CAAT' box, of equal importance for transcriptional efficiency, although we have not defined this region precisely. Deletions like -76 to -93, which leave an intact CAAT box, are very strong down mutations. This result is reminiscent of the data of McKnight et al. (9) who observed a strong down mutation on the deletion of -95 to -105 in the HSVI TK gene. It should be noted, however, that deletion of -94 to -104 has little effect in the case of the rabbit β-globin gene. It is of course possible that the homologous segment of the TK gene is displaced to the 5' side with respect to the rabbit sequence. In this regard it is noteworthy that the CAAT box of the TK gene is also displaced to the 5' side; it is found at -80 rather than at -75 as in the rabbit β-globin genes. There is, in fact, some sequence homology between the upstream regions of the TK and rabbit β-globin genes (-94 to -104 TK; -83 to -94 in β-globin;

A CCCCGCCCAG)
 * * *
AGCCACACCCTG)

although it is difficult as yet to assess its significance. A more precise localization of the important DNA sequences is required to elucidate the DNA sequences required for transcription in this region.

ACKNOWLEDGEMENTS

We are grateful to Dr. Meinrad Busslinger for a gift of the cloned sea urchin histone gene cluster. This work was supported by the British Medical Research Council and, in part, by a travel grant from NATO.

*Present address: Erasmus University, Rotterdam, Postbox 1738, 3000 DR Rotterdam, The Netherlands.

+Permanent address: Department of Molecular Virology, Hadassah Medical School, P.O. Box 1172, 91 010 Jerusalem, Israel.

REFERENCES

1. Goldberg, M. (1979). Sequence analysis of Drosophila histone genes. Ph.D. thesis, Stanford University, Stanford, California.
2. Proudfoot, N. J. (1979). Nature 279, 376.
3. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142.
4. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, L. and Chambon, P. (1980) Science 209, 1406-1414.
5. Hu, S. L. and Manley, J. L. (1981) Proc. Natl. Acad. Sci. U.S.A., 78, 820-824.
6. Grosveld, G. C., Shewmaker, C., Jat, P. and Flavell, R. A. (1981) Cell 23, 215-226.
7. Grosschedl, R. and Birnstiel, M. L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1432-1436.
8. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310.
9. McKnight, S. L., Gavis, E. R. and Kingsbury, R. (1981) Cell 25, 385-398.
10. Grosveld, G. C., deBoer, E., Shewmaker, C. K. and Flavell, R. A. (1982) Nature 295, 120-126.
11. Dierks, P., van Ooyen, A., Mantei, M. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 1411-1415.
12. Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981). Cell 27, 279-288.
13. Maxam, A. and Gilbert, W. (1980) Methods in Enzymology 65, 499-560.
14. Favoloro, J., Treisman, T. and Kamen, R. I. (1980) Methods in Enzymology 65, 718-749.
15. Moschonas, N., deBoer, E., Grosveld, F. G., Dahl, H. H. M., Wright, S., Shewmaker, C. K. S. and Flavell, R. A. (1981). Nucl. Acids Res. 9, 4391-4401.
16. Busslinger, M., Moschonas, N. and Flavell, R. A. (1981). Cell 27, 289-298.
17. Lawn, R. M., Efstratiadis, A., O'Connell, C. and Maniatis, T. (1980). Cell 21, 647-651.
18. Banerji, J., Rusconi, S. and Schaffner, W. (1981) Cell 27, 299-300.
19. Wigler, M., Silverstein, S., Lee, L., Pellicer, A., Cheng, Y. and Axel, R. (1977). Cell 11, 223-232.
20. Weaver, R. F. and Weissmann, C. (1979) Nucl. Acids Res., 6, 1175-1193.
21. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F.E., Shoulders, C. C. and Proudfoot, N. J. (1980) Cell 21, 653-668.
22. Busslinger, M., Portmann, R., Irminger, J. C. and Birnstiel, M. L. (1980). Nucl. Acids Res. 8, 957-977.