© 1995 Oxford University Press

Nucleic Acids Research, 1995, Vol. 23, No. 17 3473–3480

TBP binding and the rate of transcription initiation from the human β-globin gene

Michael Antoniou+, Ernie de Boer§, Eugenia Spanopoulou†, Ali Imam and Frank Grosveld*,†

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

Received May 19, 1995; Revised and Accepted August 8, 1995

ABSTRACT

DNA–protein interaction studies *in vitro* revealed several factors binding over the TATA box and the region of transcription initiation (cap) site of the human β-globin promoter; TATA binding protein TBP at –30, Sp1 at –19, GATA-1 at –12 and +5, YY1 at –9 and a novel factor C1 over the site of initiation (–4 to +7). Point mutants which specifically abolish the binding of each of these proteins were tested in a β-globin locus control region (LCR) construct which allows quantitative comparisons at physiological levels of transcription. Only mutants which drastically affect the binding of TBP resulted in decreased levels of transcription. A threshold value of TBP binding of 15–30% of wild type was sufficient to give normal levels of transcription. This indicates that the association of TF IID with the TATA box is not limiting in the rate of initiation of transcription.

INTRODUCTION

The initiation of transcription by RNA polymerase II (Pol II) in TATA box containing promoters is currently thought to be dependent on the efficiency of binding of the TF IID complex to the promoter. The binding of TF IID at the promoter initiates a sequential interaction of the basic transcription factors TF IIA, B, E, F, G, H and RNA Pol II with the template to form the initiation complex (for review see ref. 1).

Although initially discovered as a large, multi-component complex (>700 kDa; 2,3) that would bind at the TATA element at –30 of many class II genes, TF IID is now known to be involved in transcription from promoters that lack a canonical TATA box (4,5). In this latter situation, other factors binding at the promoter are envisaged to act, at least in part, in a 'tethering' capacity helping to anchor TF IID to the template. From the study of such class II promoters, a second element involved in initiation of transcription was discovered, namely, the 'initiator' (Inr) which is usually located over the transcriptional start or cap site (6). In addition to the TATA box, different Inr elements have been identified for different promoters (7).

The TATA binding protein (TBP) component of TF IID has been cloned and characterized from many different species (8) and has since been shown to be a 'core' component of the initiation complex of Pol I and Pol III promoters where it forms part of the SL1 and TF IIIB/SNAP$_c$ factors, respectively (9,10 and refs therein). A different set of TBP associated factors (TAFs), allows it to function as a nucleation centre for the basic transcriptional machinery on the promoters of all classes of genes (11). TAFs which mediate both activation and repression of transcription have been described (12,13).

Studies *in vitro* suggest that the rate of association of TF IID with the promoter is slow and therefore thought to be rate limiting in the formation of the pre-initiation complex (14,15). However, it is not clear that this is the case *in vivo*. Transient transfection assays have suggested that TF IID may be rate limiting on TATA-containing but not TATA-lacking promoters (16). The main problem associated with such studies is that they do not study the initiation of transcription at the same level as normally occurs *in vivo* in a natural chromatin environment and in the context of all the regulatory sequences for a given gene. However, the discovery of the key regulatory elements in the human β-globin gene locus has made such an analysis possible.

The high level erythroid and developmental stage-specific expression of the β-like globin genes is under the primary control of the locus control region (LCR) which is located 5′ of the entire gene cluster. The β-globin LCR consists of five developmentally stable, erythroid-specific regions of DNaseI hypersensitivity (17–19; Raguz and Grosveld, in preparation) and is defined by its ability to confer physiological levels of expression of the globin genes independent of the position of integration in the host genome of transgenic mice (17) and stable transfected erythroid tissue culture cells (20,21).

In earlier studies we have shown that the βLCR can fully activate transcription from a minimal β-globin promoter consisting of a TATA box (–30), CCAAT (–75) and CACC (–90) elements (22). These results indicated that erythroid specific transcription by the βLCR was mediated through the CCAAT and CACC regions, because deletion of these elements leads to the loss of erythroid specific induction of transcription. In addition,

* To whom correspondence should be addressed

Present addresses: +Department of Experimental Pathology, UMDS, Guy's Hospital, London Bridge, London SE1 9RT, UK, §Erasmus University Rotterdam, Faculty of Medicine, Department of Cell Biology and Genetics, PO Box 1738/dr., Molewaterplein 50, 3000 DR Rotterdam, The Netherlands and †The Rockefeller Institute, 1230 York Ave, New York, NY 10021, USA

there have been reports suggesting that factors binding at the TATA box region of β-globin genes are also involved in erythroid-specific transcription. In the mouse β-maj globin promoter, a 10 bp direct repeat element just upstream of the TATA box has been defined (23,24) and for the adult chicken β-globin gene, the TATA box region also binds the erythroid factor GATA-1 (25).

Because of these results and to gain further insight into the molecular mechanisms of LCR/promoter interactions, we undertook a study to characterize the factors binding at the TATA box and cap region of the human β-globin gene and to assess their role in transcription initiation *in vivo*. To achieve this, we used the βLCR in stable transfected MEL cells to obtain quantitative comparisons between different mutant promoters. Our data clearly indicate that although experiments *in vitro* showed several factors binding over this region, only TF IID is functional at the TATA box *in vivo*. Mutants capable of strongly binding GATA-1 at this site, give rise to transcripts with aberrant start sites. The results also suggest that upstream promoter binding factors may play a more active role in positioning TF IID on a TATA-less promoter than previously thought.

MATERIALS AND METHODS

DNA constructs

As before, the hybrid human β-globin/mouse H2K gene (22) was used as the reporter system for the functional analysis of the promoter mutants in these studies. This gene consists of the β-globin promoter from the *Sna*BI site at –265 to a *Hin*dIII linkered site at +30 from the start of transcription. This promoter is joined to a *Hin*dIII linkered *Nru*I site at +9 in exon I of the H2K gene. The promoter region of this hybrid gene extending from the –265 *Sna*BI site to the *Not*I site at +85 within exon I was sub-cloned between the *Hin*cII and *Not*I sites of pBluescript (Stratagene). Unique *Sma*I and *Xba*I sites at –40 and +11, respectively, were created by point mutagenesis within this sub-clone. All other promoter mutants were constructed by the direct substitution of the sequences between –40 and +11 by a series of overlapping oligonucleotides containing the relevant mutation. All mutant promoters were confirmed by sequence analysis. The wild-type promoter of the β–H2 hybrid gene in the microlocus LCR expression vector (22), was replaced with these mutant elements as *Cla*I–*Not*I fragments. The microlocus consists of all four elements of the LCR as a 6.5 kb fragment which retains full activity (21,26). All constructs were linearised at the *Pvu*I site within the ampicillin resistance gene of the vector prior to transfection into MEL and L-cells (see below).

Electrophoretic mobility shift assays (EMSA)

Electrophoretic mobility shift assay analysis was carried out using total nuclear protein extracts from HeLa, K562 and MEL cells as described previously (27). Bacterially expressed TBP was provided by Steve Jackson (Cambridge, UK) and the TBP EMSAs were quantitated by phosphorimage analysis, using a Molecular Dynamics phosphorimager (Fig. 3). Methylation interference experiments were carried out as described previously (27).

Expression studies

The maintenance, transfection by electroporation and selection of stable transfected pools of MEL cells was exactly as previously described using 50 μg linearised plasmid DNA and 3×10^7 cells per sample (28). Murine fibroblast L-cells were grown in DMEM containing 10% (v/v) foetal calf serum. L-cells were transfected with Lipofectin (Gibco-BRL) according to the manufacturer's instructions, using 20 μg linearised plasmid DNA, 100 μg lipofectin and 3×10^6 cells seeded on a 10 cm diameter Petri dish. Stable transfected clones were selected in 500 μg/ml G418 (Gibco-BRL) and were obtained at a frequency of ~1 in 10^4.

The preparation of RNA and analysis by S1-nuclease protection assays were carried out as before using 5′ end-labelled DNA probes (22,28).

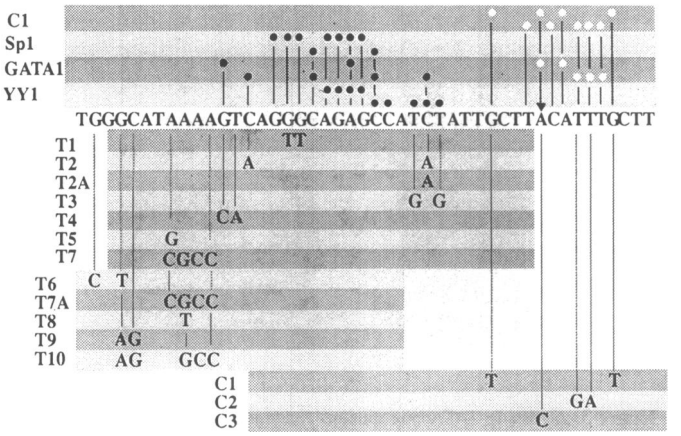
In the case of the various mutations, homologous probes were isolated for the analysis in order to measure aberrantly initiated transcripts. The results were analysed by autoradiography and quantitated by phosphorimage analysis. A 350 bp *Cla*I–*Not*I fragment from the 5′ end of the β–H2K hybrid gene was used to give a 85 nt S1-nuclease protected product (22). A 700 bp *Hin*dIII-*Nco*I fragment from the 5′ half of the mouse β-maj globin gene was used which gives a 96 nt S1-protected product from the 5′ portion of the second exon of the mRNA (27). The β-actin probe was a 320 bp *Bam*HI–*Ava*I fragment from a human β-actin cDNA clone (29) which gives a 100 nt S1-protected product from the homologous murine mRNA (30). The products were resolved on a 6% polyacrylamide gel in the presence of 8 M urea.

RESULTS

TATA box and Cap site binding factors

Initial *in vitro* DNaseI footprint experiments on the human β-globin promoter using total nuclear extracts from erythroid (murine erythroleukaemia, K562) and non-erythroid (HeLa) cells, revealed an extended area of protection from the TATA box to past the site of initiation of transcription (data not shown). To gain insight into the factors binding over this region, EMSA were employed using the same nuclear extracts and two overlapping oligonucleotides from –34 to –1 (TATA) and from –22 to +9 (Cap; Fig. 1, Table 1). Both probes showed several retarded bands, indicating that multiple proteins could potentially bind over the TATA box (T1, T2 and T3) and cap site regions (C1, C2A, C2B and C3). The bands between T2 and T3 or C2 and C3 as well as those below T3, are non-specific as they do not self-compete. One of these factors (T2/C2B) was detected only with the erythroid extracts and was bound by both the TATA and Cap probes. Another factor (T3/C3) that was not erythroid-specific, also seemed to bind to both TATA and Cap oligonucleotides. A slow migrating species was seen with both probes (T1 and C1). However, this did not appear to be the same protein, as it was not cross competable by the two probes used (Fig. 2A). Sequence comparison, migration characteristics and competition experiments provided identification of most of these factors (Fig. 2A) and the results are summarized in Table 2 (upper panel). The factor T2/C2B is GATA-1 as it was competed efficiently by a probe corresponding to the –200 GATA-1 binding site from the human β-globin promoter (27; Fig. 2A, lanes GATA) and could be supershifted with GATA-1 antibodies (31; data not shown). The residual binding activity with the Cap probe after GATA competition is the weakly binding factor C2A (Fig. 1) which has

Table 1. Methylation interference and point mutants at the human β-globin TATA box and cap site regions



The sequence of the normal (WT) promoter is shown on a white background (inverse filled triangle represents initiating nucleotide). Above the sequence are the methylation interference sites for the C1, Sp1, GATA-1 and YY1 complexes (see Figs 1 and 2). In each grey bar the protected bases of the sense strand are shown on top, the protected bases on the anti-sense strand on the bottom. The TATA region protection is shown in black dots; the cap site protection is shown in white. The cap site (+1) is at the second A from the right. Below the sequence are the different oligonucleotides with the mutations indicated in the grey bars. The length of each bar indicates the length of the oligonucleotide used in the EMSAs in Figures 1 and 2.

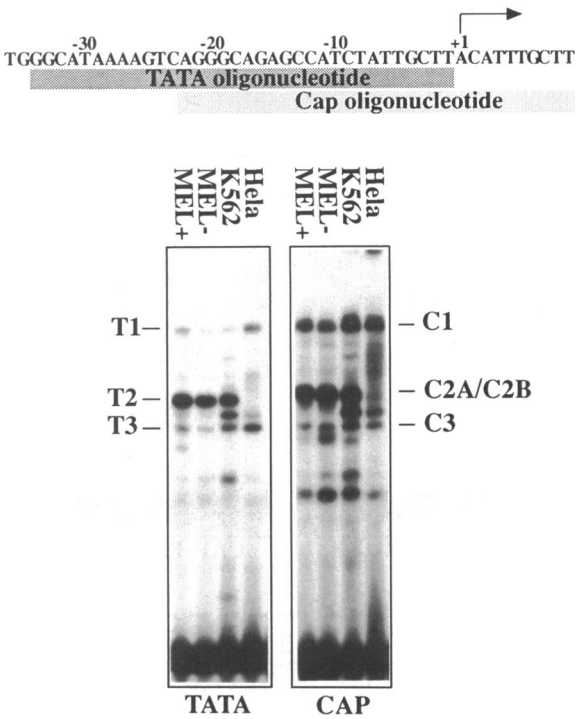


Figure 1. Protein factors binding to the TATA box and cap site regions of the human β-globin promoter. Upper panel: nucleotide sequence of the human β-globin proximal promoter from –40 to +11 relative to the cap site (+1). The extent of the two overlapping oligonucleotides used to assess protein factors binding to this region in EMSAs (lower panel) are marked; –35 to –1 (TATA probe) and –23 to +11 (Cap probe). These oligonucleotides are equivalent in length to the mutants used in subsequent experiments (Table 1). Lower panel: EMSAs employing the TATA and Cap probes described above (see Materials and Methods). The nuclear extracts used were from the following cell types: murine erythroleukaemia (MEL) both before (–) and after (+) 2 days of induced erythroid differentiation; human myelogenous leukaemia K562; HeLa. The extract used in a given assay is indicated at the top of each lane. The various complexes formed with the TATA probe (T1–T3) and Cap probe (C1–C3) are marked.

a slightly slower mobility than GATA-1 and is not affected by GATA-1 antibodies (data not shown). The factor designated T3/C3 was competed by an oligonucleotide corresponding to the YY1 binding site (32,33) from the cap region of the adeno-associated virus p5 gene (Fig. 2A, lanes YY1). The protein T1 appears to be Sp1 as it is competed by the binding site for this factor from the SV40 virus enhancer (Fig. 2A, lanes Sp1) and the complex can be supershifted with antibodies to Sp1 (data not shown). The factors C1 and C2A were not competable by any of the probes that were tried, including that for the initiator element from the TdT gene (data not shown). Methylation protection experiments (summarized in Table 1) were used to determine the position of the binding sites of each of the proteins. This showed that in addition to TF IID, the erythroid factor GATA-1 and Sp1 bind just downstream of the TATA box. The GATA binding region is a non-consensus site corresponding to a GATGGC sequence between –9 and –14 on the anti-sense strand (34). The protein which appears to be YY1, which can act both as an initiator and a transcriptional repressor (32,33) is also found at this position. (Note that the YY1 oligonucleotide has a weak GATA-1 binding activity which cross competes T2 and C2B.) Over the cap site, three proteins are seen to bind, namely, GATA-1 and two unidentified factors which are not erythroid-specific and are designated C1 and C2A (Fig. 1). This GATA-1 site again corresponds to a non-consensus TAAG element between +1 and –3 on the anti-sense strand (34). We did not detect any binding of the well characterized initiator protein TF II I (35). The binding of TF IID was only detectable in high salt conditions in crude extracts and was difficult to quantitate precisely (2,36 and data not shown). We therefore used recombinant TBP as a measure of TF IID binding at the TATA box under the low salt conditions used in all the other EMSA experiments (see below).

Information from methylation interference data was used to generate mutations for each of the proteins. These mutants are summarized in Table 1 and we used EMSA to show that the mutations specifically abolish the binding of each of the factors without creating any new activity (Fig. 2B), mT1 for Sp1, mT2 and mT2A for GATA-1 and mT3 for YY1. Of these mutants mT3 also showed a reduced binding for the TATA box factor TBP (Fig. 3) although we currently have no explanation for this observation. In addition to mT3, we generated a further set of TATA mutants which show a decreasing capacity to bind TBP (Fig. 3, Table 2), namely mT10 and mT7 which bind ~4% and <4% respectively, mT5 (13%), mT6 (30%) and mT9 (50%).

The mutants for each of the factors were subsequently built into a β-globin promoter extending to the *Sna*BI site at –265. To facilitate cloning of the mutant oligonucleotides, the promoter was first modified by point mutations at –40 and +11 to create *Sma*I and *Xba*I sites, respectively (see Materials and Methods). The introduction of these mutations does not affect the transcription of the promoter (data not shown). Each of the mutant promoters, in turn, was incorporated into the microlocus LCR

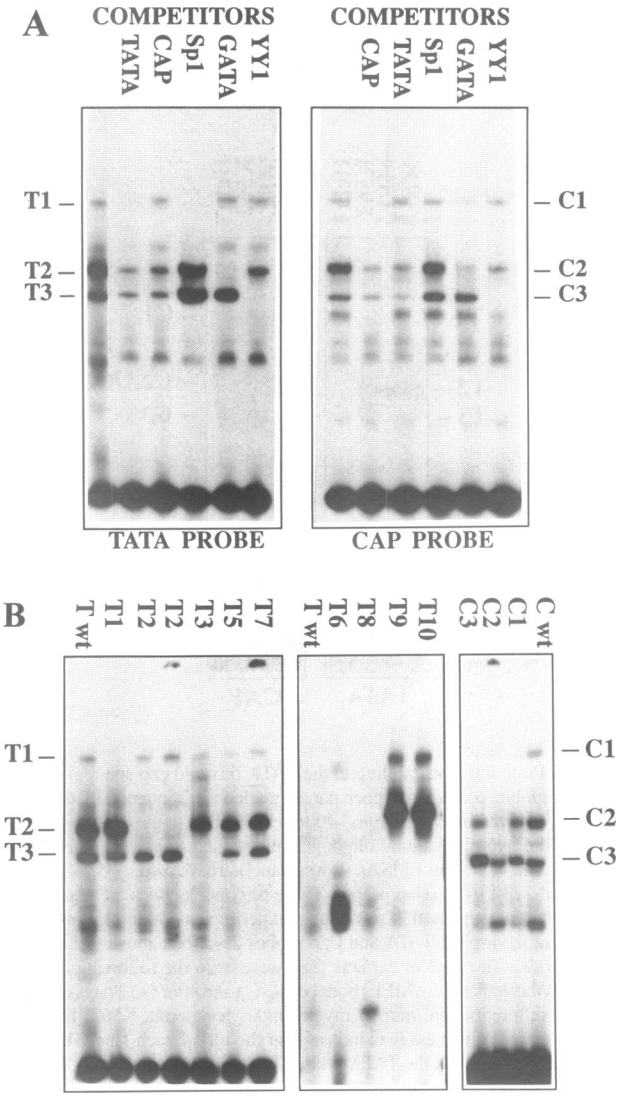


Figure 2. Identification of factors binding at the TATA box and cap site region of the human β-globin gene. (**A**) EMSAs in the presence of various competitor or mutated oligonucleotides were performed to gain insight into the identity of the factors binding to the TATA and Cap site probes. The TATA and Cap probes (0.2 ng) described in Figure 1 were used in a series of EMSA in the presence of 100-fold excess (50 ng) of unlabelled competitor oligonucleotide which specifically binds the following factors: transcription factor Sp1, erythroid factor GATA-1 and initiator protein YY1. Reactions were also self and cross competed with TATA or Cap probes. Assays in the absence of competitor are also shown. (**B**) Analysis of the mutant TATA box and cap site oligonucleotides (Table 1). Note that the Twt oligonucleotide in the middle panel is shorter (–38 to –11) than that used in (A) or the left hand panel of this figure (–34 to –1) and terminates at the same position as T6, T8, T9 and T10 (Table 1). Complexes T1, T2 and T3 are therefore absent in these assays. The nuclear extract used in all cases was from uninduced MEL cells. The complexes marked are the same as in Figure 1.

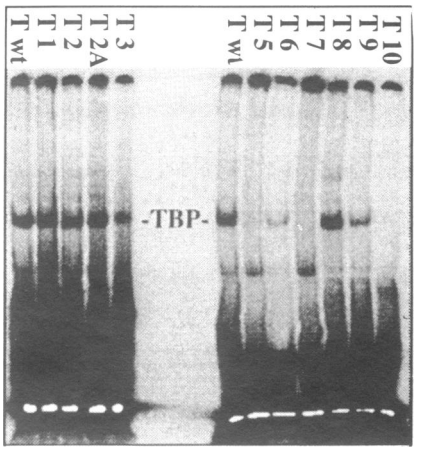


Figure 3. Assessing TATA box mutants for binding of TBP. The TATA box mutants listed in Table 1 were assessed by EMSAs for their ability to bind recombinant TATA binding protein (TBP). The oligonucleotide probes (0.2 ng) used were the same as those described in Figure 2. The wild type (Twt) sequence acted as a control. This figure shows the phosphorimage which was used to quantify the data (Table 2). The result with the wild type human β-globin TATA box (Twt) was taken as 100%.

expression cassette as a β–H2 hybrid gene (22). Table 2 shows a list of the mutants built into this expression system. These constructs were then introduced into MEL cells by electroporation and two or three large pools of stable transfected, G418 resistant clones selected (28). RNA was then isolated from cells that had been induced to undergo erythroid differentiation for 4 days in the presence of 2% DMSO. Expression of the transfected gene was assessed by an S1-nuclease protection assay using end-labelled DNA probes (Fig. 4).

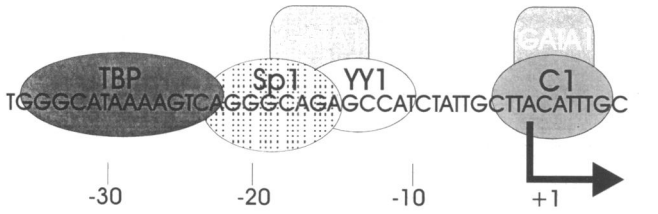
Functional analysis of TATA box mutants; TBP binding is not rate limiting in transcription initiation

Mutants that prevent the binding of Sp1 (mT1), GATA-1 (mT2) and YY1 (mT3) near the TATA box (Fig. 2B), still gave full transcriptional activation (Fig. 4A, lanes mT1, mT2 and mT3). The only mutants in this region that decreased transcription (Fig. 4B,C) were those which markedly reduced the binding of TBP, namely mT5, mT7 and mT10 (Fig. 3) of which mT5 (–29 A to C) is a natural β-thalassaemia mutation (37). Interestingly, three of the mutants that showed a reduced binding for TBP namely mT3, mT6 and mT9 (Fig. 3), did not result in decreased transcription (Fig. 4A–C). Only when the ability to bind TBP is reduced still further (mT5, mT7 and mT10), does this result in decreased expression (Table 2), in agreement with the lower levels of β-globin RNA found in patients with the mT5 mutation (37). As the β–H2K hybrid mRNA is stable, the steady state cytoplasmic levels of this sequence are directly proportional to rates of transcription. Therefore, these data indicate that maximum initiation of transcription can be achieved once a certain threshold in the efficiency of TF IID binding to the promoter is reached. This also clearly suggests that the ability of the TATA box to bind TF IID is not limiting in the rate of initiation of transcription.

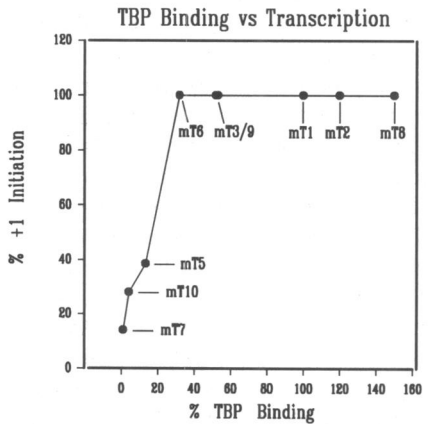
Functional analysis of Cap site mutants

Three mutants of the cap site region were also tested. Mutant mC1 affects only the binding of C1 while mC2 abolishes binding of C1 and GATA-1 (see DNA contact sites, Table 1). Neither of these caused a decrease in transcription (Fig. 4E, lanes mC1 and mC2). In addition, a natural β^+-thalassaemia mutant (mC3) which consists of a change of the initiating A to a C residue (38), was

Table 2. Efficiency of TBP binding and expression levels from the TATA box and cap site mutants of the human β-globin promoter in MEL cells



Construct	Mutation	TBP binding	Correct start	Other starts
WT		100	100	
mT1	Sp1	100	100	
mT2	GATA-1	120	100	
mT3	YY1	53	100	
mT5	TBP	13.3	38.5	
mT6	TBP	32	100	
mT7	TBP	<4	14	
mT8	Mouse TATA	150	100	
mT9	Chicken TATA	52	100	
mT10	GATA-1+/TBP-	4.2	28	8.4
mC1	C1		100	
mC2	GATA-1		100	
mC3	+1A➤C		64.5	40



The top panel shows a schematic representation of the position of the different proteins binding to the human β-globin promoter sequence. The middle panel shows the quantitation of the efficiency of TBP binding to the different mutant oligonucleotides (see Table 1 and Fig. 3) and the amount of RNA transcribed from the +1 or downstream positions of the constructs containing these mutations (Fig. 4). The wild type (WT) promoter is taken as a 100% standard for both TBP binding and expression levels of correctly initiated transcripts. The bottom panel shows a graphic illustration of the amount of RNA transcribed from these constructs versus the amount of TBP binding.

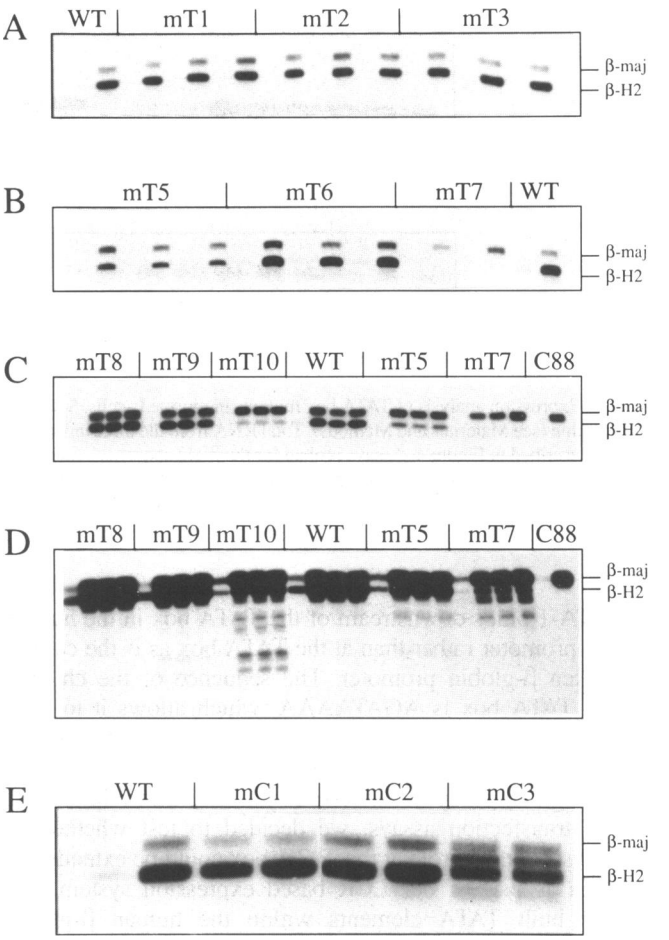


Figure 4. Expression analysis of TATA box and cap site mutants in MEL cells. The TATA box and cap site mutants for the various factors listed in Table 1 were built into the βLCR expression cassette and used to generate stable transfected, G418 resistant pools of MEL cells (see Materials and Methods). Total RNA was isolated from the large pools of clones that were produced after they had been induced to undergo 4 days of erythroid differentiation in the presence of 2% DMSO. Samples (2 μg) were simultaneously assayed for the β-H2 transgene and the endogenous β-maj globin sequences by S1-nuclease protection using end-labelled DNA probes which detect the 5′ end of the mRNAs. The S1-nuclease protected fragment for the β-H2 mRNA is 85 nucleotides and that for β-maj 96 nucleotides. Homologous probes for the TATA and cap mutants were employed. Panel D is a longer exposure of panel C and shows the minor bands corresponding to initiation events downstream of the normal Cap site. WT refers to the result obtained with the wild type promoter. mT1–mT10 refer to the TATA box mutants and mC1–mC3 to the cap mutants. The results were quantified by phosphorimage analysis correcting for variation in erythroid induction levels by using β-maj globin as an internal standard. The level obtained with the wild type promoter was taken as 100% expression (see Table 2). Different specific activity probes were used on different gels. Only a small part of the gels is shown for space reasons.

also tested. This mutation also reduces the binding of the C1, but not C2A and GATA-1 proteins. It results in a small decrease of the total amount of transcription, of which a large part is initiated aberrantly, in particular, between 5 and 8 nucleotides upstream of the wild type cap site (Fig. 4E, lanes mC3). We conclude that the reduction in the level of transcription observed with the mC3, is largely due to the fact that the initiation nucleotide A has been changed, leading to aberrant starts, rather than to the fact that C1 binding has been affected. The reduced level of RNA correlates well with the observed very mild thalassaemia which arises from this mutation in human subjects (38).

Testing specialized TATA elements

Our observation that GATA-1 binding at the human β-globin TATA box does not appear to be functionally significant, is at variance with the results obtained with this same region from the chicken gene (25). This could possibly be explained by the fact

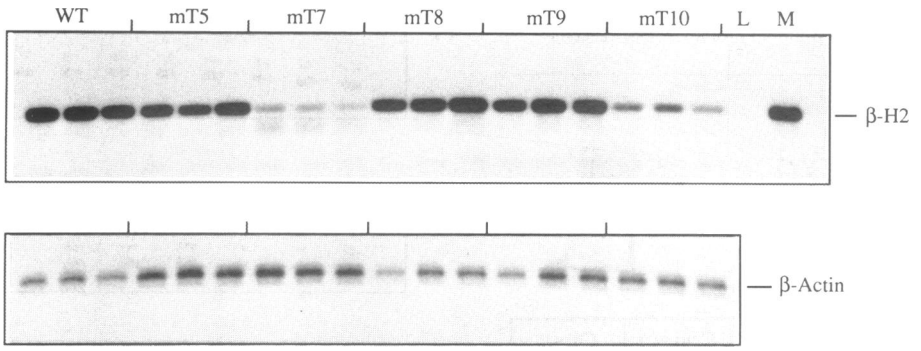


Figure 5. Expression analysis of TATA box mutants in murine L-cells. Selected TATA box mutants (see Table 1) were stably transfected into mouse fibroblast L-cells by lipofection (see Materials and Methods). Total RNA from the three independent pools of clones that were generated was then analysed by an S1-nuclease protection assay, as described in Figure 4. Assays probed for the β-H2 sequences (upper panel) contained 20 μg RNA per reaction. Samples (2 μg) were also assayed for β-actin mRNA as a loading control. The β-H2 probe gives the same β-H2 S1-protected product as in Figure 4. The β-actin probe gives a 100 nucleotide S1-protected product. RNA from untransfected L-cells (L) and β-H2 transfected MEL (M) acted as negative and positive controls, respectively.

that GATA-1 binds downstream of the TATA box in the human β-globin promoter rather than at the TATA box as is the case in the chicken β-globin promoter. The sequence of the chicken β-globin TATA box is AGATAAAA, which allows it to bind GATA-1 strongly as opposed to the corresponding human element (ACATAAAA). Since the function of GATA-1 in the chicken β-globin study was assessed by *in vitro* transcription and transient transfection assays, we decided to test whether its proposed specialized role at the TATA box could be extended to chromatin as part of our LCR based expression system. We therefore built TATA elements within the human β-globin promoter with varying affinities for TF IID and GATA-1 (Figs 2B and 3 and Table 2). The TATA box regions of the mouse (mT8), human (wt) and chicken (mT9) β-globin genes have a different affinity for TBP (Fig. 3), which are 150, 100 and 50%, respectively, (Table 1) using the human sequence as the standard. In contrast, the chicken sequence has a much higher affinity for GATA-1 as expected (Fig. 2B, Table 2). In addition, we built a mutant TATA element, mT10 (AGATAGCC), which binds GATA-1 (Fig. 2B, lane T10) as strongly as the chicken element, but which has minimal binding of TBP (Fig. 3). The severe TATA mutant mT7, acted as a control which does not bind TBP (Fig. 3, Table 2), but which retained GATA-1 binding at –12 like the other constructs in this series (Fig. 2B).

Expression of these mutants in MEL cells revealed that the wild type human β-globin TATA box (Fig. 4C, lane WT, Table 2), directed initiation of transcription with equal efficiency as that from the mouse (Fig. 4C, lanes mT8) or chicken (Fig. 4C, lanes mT9) genes. Thus, the higher affinity for GATA-1, seen with the chicken TATA box does not result in any physiological effect. The only difference that could be discerned between the mouse or human and chicken elements was the appearance of aberrant transcripts initiated downstream of the natural cap site (Fig. 4D). Some of these appear to be GATA-1 stimulated transcripts and some were the result of weak TF IID binding as shown by the different starts in mT7 and mT10 (Fig. 4D). The mT10 mutant has the same GATA-1 binding capacity as the chicken TATA box (mT9), but very low affinity for TBP (Fig. 3, Table 2). Interestingly, the mT10 (GATA-1) promoter produced 2-fold more correctly initiated transcripts than the severe TATA mutant mT7 (Fig. 4C, Table 2). In order to confirm that this difference was due to the observed higher affinity for TBP by mT10 than that by mT7 (Fig. 3), we carried out transfections in mouse fibroblast L-cells which are devoid of GATA-1. Stable transfected pools of 200–300 clones of L-cells carrying the same WT, mT7, mT8, mT9 and mT10 promoter constructs as used in MEL cells, were produced. Expression of the transgene was assessed by S1-nuclease protection as before (Fig. 5). The results show that all of the promoters give only the normal site of transcription initiation; that is, the aberrant transcription starts seen with the mT9 and mT10 promoters in MEL cells, are absent in L-cells. This supports the notion that these aberrant initiated products in MEL cells are due to GATA-1 binding at the TATA box region of these constructs. This suggests that GATA-1 and TF IID binding at the TATA box do not co-operate in the initiation of transcription, but that GATA competes with normal TATA box directed initiation at +1. However, attempts to show this directly by co-transfection of GATA expression vectors in L-cells failed because we were unable to obtain levels of GATA-1 comparable to those seen in MEL cells (data not shown).

It is also important to note that the relative differences in transcription efficiency observed in the L-cells are different to those observed in MEL cells. For example, the WT/mT7 ratio is higher in MEL cells (7-fold) than in L cells (3- to 4-fold), suggesting that conclusions on tissue specific promoters should preferably be based on results obtained in homologous systems.

Finally, relatively high levels of correctly initiated transcripts observed with the TATA box mutants mT7 (14%) and mT10 (28%) (Fig. 4C,D; Table 2) suggests that either upstream promoter factors or an Inr element (or both) are involved in placing TF IID at the TATA box region. Since we did not detect binding of the previously characterized Inr binding protein TF II-I (35) to the transcriptional start site, we decided to reassess if the factor C1 (Fig. 1) could function in this capacity. The analysis of double mutants consisting of the mT7 or mT10 TATA box and mC1 cap site combinations (Table 1), resulted in a further reduction in the level of correctly initiated transcripts relative to that obtained with the single mT7 and mT10 promoters (data not shown). This suggests that C1 protein can function as an Inr protein in the context of a weak TATA element but whose physiological significance at present remains unclear.

DISCUSSION

In this study we have exploited the human β-globin LCR expression system to study the factors involved in the initiation of transcription from the β-globin gene. This system allows quantitative functional expression assays to be performed at physiological levels *in vivo* in contrast to reports on this topic which predominantly use *in vitro* transcription assays with nuclear extracts or transient transfection experiments. We have previously demonstrated that the βLCR can fully activate erythroid specific transcription through the proximal promoter (–90, CACC and –75, CCAAT) elements of the β-globin gene (22). In this study we investigated factors binding over the TATA box and Cap site regions.

Protein–DNA binding studies *in vitro* using nuclear extracts from erythroid (MEL, K562) and non-erythroid (HeLa) cells showed that in addition to TBP, three other factors were found to bind around the TATA box of the human β-globin gene (Table 2). These were the erythroid factor GATA-1, the ubiquitous transcription factor Sp1, a transcriptional activator (39) and a YY1-like protein which can function both as a transcriptional inhibitor as well as an initiator protein (32,33). For reasons that are presently not clear, we find no evidence for the binding activity reported by Stuve and Myers (24). Expression studies in MEL cells with LCR driven promoters carrying mutations which abolish the binding of Sp1, YY1 or GATA-1, resulted in wild type levels of transcription (Fig. 4A and B; Table 2), indicating that these proteins have no significant role at the TATA box *in vivo*.

The various TATA box mutants tested in this study allowed us to address the question of TBP binding efficiency in regulating initiation of transcription. Electrophoretic mobility shift assays with recombinant TBP (Fig. 3), indicated the following order of binding efficiency for this factor by the TATA elements tested: mouse (mT8), 150%; human (wt), 100%; chicken (mT9) and mT3, 50%; mT6, 30%; mT5, 13%; mT10, 4%; mT7, <4% using human (wt) as a 100% reference. Despite the fact that the binding of TBP to the mT9, mT3 and mT6 mutants (Table 2) is between 50 and 30% of the normal human sequence, equal rates of transcription initiation are observed *in vivo* (Fig. 4A and B). Again, in the case of the T5 (–28 A to C) mutant which binds TBP at 13% of wild type, transcription is still 30% compared to wild type. Interestingly, the mT5 level of transcription is in agreement with that observed in the β-thalassaemia patients which carry this mutation (37), emphasizing that our model system mimics the *in vivo* situation. However, it is known that TF IID partially purified from HeLa cells produces a DNaseI footprint over the adenovirus major late (AdML) promoter which extends from the TATA box to past the cap site (2,40). This is in marked contrast to recombinant TBP which only footprints the TATA box element (41–43). The extended footprint given by TF IID is now known to be due to the $TAF_{II}150$ component of this complex (44). We therefore also attempted to correlate *in vitro* TF IID rather than TBP binding with transcriptional efficiency *in vivo* by isolating an enriched TF IID fraction from MEL cells and compared it to a highly enriched TF IID containing fraction from human (HeLa) cells (a generous gift from P. Verrijzer and R. Tjian). However, none of these fractions (MEL or HeLa) gave EMSAs which could be quantitated with either the wild type or mutant oligonucleotides used in the TBP binding experiments. Nevertheless, our results indicate that the efficiency of TBP (and probably TF IID) binding at the TATA box is not the rate limiting step in transcription and that once a threshold value has been achieved, maximum initiation of transcription will occur. This threshold value would appear to be quite low and of the order of 15–30% of a strong binding site like the human TATA box which is comparable to that from the AdML promoter (36).

The TATA box mutants mT7 and mT10 have only a weak TBP binding activity (Fig. 3) which is much lower than the inherent TBP binding capacity at the –30 region of some native TATA-less promoters (45). However, these two mutants still gave significant levels of correctly initiated transcripts (Fig. 4C; Table 2). This may be due to number of different reasons. First, since $TAF_{II}150$ can bind to DNA independently of TBP (44), this may reflect the residual *in vivo* binding capacity of TF IID to these TATA box mutants due to this TAF component of the complex. However, this is unlikely to completely account for the observed levels of expression given that TATA box mutants of the Ad2 E1B gene are unable to compete the binding of a purified HeLa cell TF IID fraction to the wild type element (5). This indicates that the binding of $TAF_{II}150$ to DNA is weak compared to the TF IID complex as a whole. Alternatively, the binding of TF IID to the mT7 and mT10 mutants may be mediated by an Inr function or by factors associated with upstream promoter elements such as the –200 GATA-1, –85 CACC or –75 CCAAT regions which could position TBP as part of TF IID at the natural TATA position through some tethering capacity, either by interacting directly with TF IID or through an intermediary, co-activator protein or TAF (46). However, the observation that most of the transcripts have aberrant downstream start sites (Fig. 4D) would indicate that the 'tethering' capacity of these upstream factors is generally a loose one, resulting in the positioning of TF IID and subsequently the initiation complex as a whole, over an extended region rather than a single site. This resembles the situation for transcripts from many TATA-less promoters which usually have a cluster of start sites and no demonstrated Inr function.

ACKNOWLEDGEMENTS

We are grateful to Steve Jackson for sending us the Sp1 antiserum and TBP recombinant protein, to Doug Engel for the GATA-1 antibody, Peter Verrijzer and Bob Tjian for the HeLa cell TF IID fraction and to Peter Rigby for helpful comments on the manuscript. MA was supported by Zeneka (formerly ICI) Pharmaceuticals (UK). FG is the recipient of a Howard Hughes International Research Scholarship. This work was supported by the MRC (UK) and NWO (NL).

REFERENCES

1 Zawel, L. and Reinberg, D. (1993) In Cohn,W.E and Moldave, K. (eds) *Progress in Nucleic Acid Research Molecular Biology.* Academic Press, San Diego, USA, pp. 67–108.
2 Nakajima, N., Horikoshi, M. and Roeder, R. (1988) *Mol. Cell. Biol.*, **8**, 4028–4039.
3 Dynlacht, B.D., Hoey, T. and Tjian, R. (1991) *Cell*, **66**, 563–576.
4 Pugh, B. and Tjian, R. (1991) *Genes Dev.*, **5**, 402–408.
5 Zhou, Q., Lieberman, P., Boyer, T. and Berk, A. (1992) *Genes Dev.*, **6**, 1964–1974.
6 Smale, S. and Baltimore, D. (1989) *Cell*, **57**, 103–108.
7 Weis, L. and Reinberg, D. (1992) *FASEB J.*, **6**, 3300–3309.
8 Hernandez, N. (1993) *Genes Dev.*, **7**, 1291–1308.
9 Sharp, P. A. (1992) *Cell*, **68**, 819–821.
10 Rigby, P. (1993) *Cell*, **72**, 7–10.
11 Goodrich, J.A. and Tjian, R. (1994) *Curr. Opin. Cell Biol.*, **6**, 403–409.
12 Roeder, R. (1991) *Trends Biochem. Sci.*, **16**, 402–408.
13 Gill, G. and Tjian, R. (1992) *Curr. Opin. Genet. Dev.*, **2**, 236–242.
14 Reinberg, D. and Roeder, R. (1987) *J. Biol. Chem.*, **262**, 3310–3321.

15 Schmidt, M., Zhou, Q. and Berk, A. (1989) *Mol. Cell. Biol.*, **9**, 3299–3307.
16 Colgan, J. and Manley, J. (1992) *Genes Dev.*, **6**, 304–315.
17 Grosveld, F., Blom van Assendelft, G., Greaves, D. and Kollias, G. (1987) *Cell*, **51**, 975–985.
18 Tuan, D., Solomon, W., Li, Q. and London, I. M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6384–6388.
19 Forrester, W., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. (1987) *Nucleic Acids Res.*, **15**, 10 159–10 177.
20 Blom van Assendelft, G., Hanscombe, O., Grosveld, F. and Greaves, D. R. (1989) *Cell*, **56**, 969–77.
21 Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. and Greaves, D. R. (1989) *Nature*, **338**, 352–355.
22 Antoniou, M. and Grosveld, F. (1990) *Genes Dev.*, **4**, 1007–1013.
23 Cowie, A. and Myers, R. (1988) *Mol. Cell. Biol.*, **8**, 3122–3128.
24 Stuve, L. and Myers, R. (1990) *Mol. Cell. Biol.*, **10**, 972–981.
25 Fong, T. and Emerson, B. (1992) *Genes Dev.*, **6**, 521–532.
26 Collis, P., Antoniou, M. and Grosveld, F. (1990) *EMBO J.*, **9**, 233–240.
27 deBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988) *EMBO J.*, **7**, 4203–4212.
28 Antoniou, M. (1991) In Murray,E.J. (ed.) *Methods in Molecular Biology. Vol. 7: Gene Transfer and Expression Protocols*, Humana Press Inc., Clifton, NJ, pp. 421- 434.
29 Gunning, P., Ponte, P., Okayama, H., Engel, J., Blau, H. and Kedes, L. (1983) *Mol. Cell. Biol.*, **3**, 787–795.
30 Eccles, S., Sarner, N., Vidal, M., Cox, A. and Grosveld, F. (1990) *The New Biologist*, **2**, 801–811.
31 Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J. D. and Yamamoto, M. (1993) *Nature*, **362**, 466–468.
32 Seto, E., Shi, Y. and Shenk, T. (1991) *Nature*, **354**, 241–245.
33 Shi, Y., Seto, E., Cheng, L. and Shenk, T. (1991) *Cell*, **67**, 377–388.
34 Whyatt, D.J., deBoer, E. and Grosveld, F. (1993) *EMBO J.*, **12**, 4993–5005.
35 Roy, A., Meisterernst, P., Pognonec, P. and Roeder, R. (1991) *Nature*, **354**, 245–348.
36 Spanopoulou, E. (1990) Ph.D. Thesis, University of London.
37 Huang, S., Wong, C., Antonarakis, T., Ro-Lein, T., Lo, W. and Kazazian, H. (1986) *Hum. Genet.*, **74**, 152–164.
38 Wong, C., Dowling, C., Saiki, R., Higuchi, H., Erlich, H. and Kazazian, H. (1987) *Nature*, **330**, 384–386.
39 Gidoni, D., Kadonaga, J. T., Barrera, S. H., Takahashi, K., Chambon, P. and Tjian, R. (1985) *Science*, **230**, 511–517.
40 Sawadogo, M. and Roeder, R.G. (1985) *Cell*, **43**, 165–175.
41 Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M. and Roeder, R. (1990) *Nature*, **346**, 387–390.
42 Kao, C., Lieberman, M., Schmidt, M., Zhou, Q., Pei, R. and Berk, A. (1990) *Science*, **248**, 1646–1649.
43 Peterson, M., Tanese, N., Pugh, B. and Tjian, R. (1990) *Science*, **248**, 1525–1630.
44 Verrijzer, C.P., Yokomori, K., Chen, J.-L. and Tjian, R. (1994) *Science*, **264**, 933–941.
45 Wiley, S.R., Kraus, R.J. and Mertz, J.E. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 5814–5818.
46 Chen, J.L., Attardi, L.D., Verrijzer, C.P., Yokomori, K. and Tjian, R. (1994) *Cell*, **79**, 93–105.