Nucleic Acids Research, 1997, Vol. 25, No. 20 **3995–4003**

Efficient transcription of an immunoglobulin κ promoter requires specific sequence elements overlapping with and downstream of the transcriptional start site

Marc R. Pelletier[+], Eunice N. Hatada, Gabriele Scholz and Claus Scheidereit*

Max-Delbrück-Center for Molecular Medicine MDC, Robert-Rössle-Str. 10, 13122 Berlin, Germany

Received July 18, 1997; Revised and Accepted September 4, 1997

ABSTRACT

The expression of immunoglobulin (Ig) genes depends on tissue-specific elements in the promoter and enhancer regions of light chain and heavy chain genes. In contrast to the complex modular character of Ig enhancers, the promoters appear to be simple, depending primarily on a conserved TATA box and octamer elements. We have analyzed the role of proximal sequences for Igκ promoter function. Igκ promoter transcription critically depends on initiator-like sequences and on a downstream element located at +24 to +39 relative to the start site. Replacement of these sequences resulted in strong reduction of promoter activity. *In vitro*, these elements were found to be more effective in extracts of lymphoid than of non-lymphoid origin. Deletion of the downstream and initiation site regions had a comparable effect on promoter activity to obliteration of the TATA box or octamer element. The downstream sequence was bound by two nuclear proteins, identical to the previously identified Ig-specific C5 and C6 complexes. Whereas C5 is found in HeLa cells and in lymphoid cells, C6 is lymphoid specific. Thus, further specific sequences in addition to the previously characterized elements, the octamer and the TATA box, are required for efficient κ promoter expression in B lymphocytes.

INTRODUCTION

The lymphoid cell restricted expression of immunoglobulin genes is dependent on specific *cis*-regulatory elements in the promoters and enhancers (see ref. 1 for review). The most critical regulatory element is the octamer motif, which is conserved in all immunoglobulin promoters and enhancers. The critical role of this motif for efficient transcription of immunoglobulin (Ig) genes in B lymphocytes has been substantiated by a number of reports (2–4, ref. 1 for review). Heavy chain promoters contain an additional heptamer sequence 5′ to the octamer. Both octamer and heptamer elements are bound by the ubiquitous Oct-1 and the B cell specific Oct-2 POU family transcription factors (5–7). The POU factors can interact with a B cell specific co-factor named, alternatively, Bob1, OBF-1 or OCA-B, to enhance their transactivation potential in a promoter-specific fashion (8–10). Inactivation of the genes encoding either Oct-2 or OCA-B by homologous recombination, however, showed that neither of these activators is essential for basal Ig expression, possibly due to functional redundancy with related factors (11–13). Since functional octamer elements also occur in genes with ubiquitous expression, further sequence information, apart from the octamer, should be required to render an octamer-containing promoter B cell specific. Except for the octamer, the only other generally conserved sequence elements in the promoters of light chain and heavy chain genes are the TATA box and, in light chain promoters, a sequence upstream of the octamer, the pentadecanucleotide element (2,14). Light chain and heavy chain promoters interact with a common upstream factor (C1) that binds upstream or downstream of the octamer site, respectively (15). C1 is the human cut-homeobox protein and represses κ gene transcription (G.S. and C.S., in preparation). Some immunoglobulin promoters have been shown to carry binding sites for further upstream factors, such as Ig/EBP, NFκY, NTF or EBF (16–19).

Whereas gene transcription by RNA polymerase II is assumed to generally require the basal transcription factors IIB, IID, IIE, IIF, IIH and IIJ, TFIIE is not required for transcription from an Ig heavy chain promoter, suggesting a putative promoter specificity of basal factor requirement (20). As another example, in contrast to E4 or IL-2 core promoters, transcriptional repression of Ig light or heavy chain promoters by a 90 kDa factor was not abolished by TFIIA (21).

Interestingly, the heavy chain promoter is bound by several factors, distinct from TFIID, over the TATA box (15,22). The contacts of one of these factors extended downstream of the transcription start site to position +28, but the functional importance of this interaction remains to be established (22). Regulatory elements which overlap with the transcription start site or reside in a downstream position have been found in several other cellular and viral genes (23,24 for review). Human TFIID

* To whom correspondence should be addressed. Tel: + 49 30 9406 3816; Fax: +49 30 9406 3866; Email: scheidereit@mdc-berlin.de

[+]Present address: Mitotix, Inc., 1 Kendall Square, Building 600, Cambridge, MA 02139, USA

interacts with downstream sequences on the adenovirus E4 promoter in the presence of the activator ATF bound to its upstream sites (25). In the *Drosophila hsp*70, *hsp*26 and histone H3 promoters, binding of TFIID depends on the TATA sequence and on specific downstream elements (from +10 to +33). The downstream interaction in the *Drosophila* genes is presumably mediated through TFIID-associated TAFs (26,27). Recently it has been demonstrated that a TBP-TAF250–TAF150 complex binds to the initiation site and to downstream sequences in several promoters (28). A conserved initiator sequence found at or around the transcription start site of many genes functionally interacts with a universally conserved protein complex, presumably TFIID (29, and references therein). *Drosophila* TFIID binds to a basal promoter element ~30 nt downstream of the RNA start site which is conserved in many TATA-less promoters (30). In some cases regulatory downstream regions have been shown to be bound by gene-specific factors. The HIV-1 LTR contains stimulatory sequences from +4 to +18 relative to the start site, as demonstrated in a reconstituted HeLa system and in transfected cells (31). A sequence element overlapping this region binds the factor LBP-1 which stimulates transcription (32). In a similar fashion, the major late promoter of SV40 and the HTLV-1 promoter are dependent on sequences downstream of the start site which are bound by the factors DRE-1 and IBP, respectively (33,34).

In this study we have analyzed the properties of the basal promoter of an Igκ light chain gene. We have identified novel basal promoter elements located at the initiation site and at a downstream position. The initiation site and downstream elements were found to be as crucial for promoter function as the conserved octamer upstream element both *in vitro* and *in vivo* and exhibited cell type specificity *in vitro*. A promoter fragment containing these sequences was bound by two nuclear proteins, one of which is lymphoid-specific. Initiator-like sequences are found in a number of human and murine immunoglobulin gene promoters and may contribute to the selective activation of these genes in B lymphocytes.

MATERIALS AND METHODS

Plasmid constructs

The wild-type T1κ(+39) promoter, the point mutants and deletion constructs, TATA-, octa- and +6 constructs are derived from pGT131 (35), which contains the murine T1 κ light chain promoter (36). tk-β-gal used for B cell transfection and the RSV-β-gal construct used for transfection of non-B cells were kindly donated by K. Mölling, -50-MLP-C2AT is described in reference 37 and pCATenh was purchased from Promega. The +39 promoter was constructed with an *Ava*I–*Pst*I fragment of pGT131 (containing the T1 κ sequence from –131 to +39).The promoter was cloned into the *Hin*dIII–*Pst*I sites of pCATenh, after filling in the *Hin*dIII site from the vector and the *Ava*I site from the fragment with Klenow polymerase. The resulting sequence of the +39/CAT construct downstream of the TATA box is, TCATATACCC GTCACACATG TACGGTACCA TTGTCATTGC AGCCAGGACT CAGCATGGAC ATGAGGACGA CCGGTCGACT CTAGAGGATC TGAGCTTGGC GAGATTTTCA GGAGCTAAGG AAGCTAAAATG. Underlined are the TATA box, the mapped initiation sites, the κ translation start site and the chloramphenicol acetyltransferase (CAT) translation start, respectively. The last 63 nt are polylinker and CAT sequences. The +6 promoter was constructed by PCR with the T7 sense primer and an oligonucleotide encoding the anti-sense sequence from –12 to +6 followed by a *Pst*I site and an extra six bases to facilitate restriction digestion of the PCR product. pGT 131 was used as template. The PCR product was digested with *Xba*1–*Pst*I, while pCATenh was cut with *Hin*dIII–*Pst*I and filled-in at the *Hin*d site with Klenow prior to ligation. For construct –2, pGT 131 was cut upstream of the promoter at the *Xba*I site and at the *Asp*718 site (at –2) and filled in with Klenow enzyme. This fragment was then ligated into pCATenh. +39/octa- and +6/octa- were constructed using PCR with a T7 primer as the upstream primer and an oligonucleotide containing T1 sequences from –46 to –75 with base substitutions changing the wild-type octamer sequence from ATTTGCAT to CGGGGCAT, with pGT 131 as a template. pCB1 (which contains the same T1 sequence as pGT but with pGEM4 as background vector to eliminate the *Sph*I site in the vector) was used as an intermediate vector and both this and the PCR product were digested with *Sph*I–*Eco*RI and ligated, resulting in +39/octa-. The *Bam*HI–*Asp*718 fragment was then isolated from this plasmid and ligated into either the +39 or the +6 vector cut with the same enzymes. +39/TATA- and +6/TATA- were constructed using PCR with an upstream primer encompassing the κ sequence from –30 to –5 but containing base substitutions transforming the wild-type sequence TCATATACC to GCGGCCGCT, creating a *Not*I site. An oligonucleotide encoding sequences downstream of the polylinker in pCATenh (CATTTTAGCTT-CCTTAGCTCCTG, the same as for primer extension) was used as an antisense primer, with the +39 construct as template. pGT 131 was employed as an intermediate vector and both this and the PCR-generated insert were cut with *Sty*1–*Asp*718 and ligated. The *Bam*HI–*Asp*718 fragment of this construct was then ligated into the +39 and +6 constructs. The run off templates were constructed by inserting a 350 bp fragment of oct-2 cDNA into the *Pst*I site in the polylinker of pGT. The plasmid was linearized with *Hin*dIII, resulting in a run off transcript of 390 nt. For the –2 run off construct, the +39 run off construct was cleaved with *Asp*718 and *Pst*I, taking out the promoter sequence from –2 to +39 and religated. This construct was linearized with *Hin*dIII, giving a run off signal of 200 nt. mINR-A, -B and -AB were derived from +39/CAT (see above) by PCR using wild-type sequence 5′- and mutant sequence 3′-primers, containing *Hin*dIII and *Pst*I sites at their ends, respectively. After restriction, products were cloned into pCAT enhancer. Random cassette mutants rdA, rdB, rdC and rdD were derived from the +39/CAT construct (see above) by PCR. The 5′-primer contained T1κ sequences from –131 to –112 and a *Hin*dIII site at the 5′-end. The 3′-primers contained the indicated random nucleotides, a *Pst*I site at the 5′-end and 20 nt complementary to the promoter at the 3′-end. PCR products were cleaved with *Hin*dIII and *Pst*I and cloned into pCAT enhancer. Individual clones were isolated and sequenced. PCR reactions were carried out with Taq DNA polymerase from Promega with magnesium concentrations from 1.5 to 4.0 mM, dNTP concentrations of 200 μM and primer concentrations of 0.5 μM. DNA fragments, PCR products and oligonucleotides were purified with NuSieve agarose (FMC) and Qiaex (Qiagen). Plasmids were purified using Qiagen Tips from Qiagen. All constructs were sequenced using the T7 sequencing kit (Pharmacia). The dTTP ladder in Figure 1 was generated with the same sequencing kit

using only ddATP to block elongation of the anti-sense strand of both the +39 and +6 constructs.

Cell culture and nuclear extracts

HeLa cells were maintained in suspension in S-MEM medium with L-glutamine (Gibco/BRL) supplemented with 2.2 g/l sodium hydrogen carbonate, 1% penicillin/streptomycin, 1% non-essential amino acids, 10 mM HEPES (pH 7.3), 5% newborn calf serum. Namalwa cells were grown in suspension in RPMI 1640 medium supplemented with 2% L-glutamine, 1% penicillin/streptomycin, 10 mM HEPES (pH 7.3) hand 7.5% fetal calf serum. COS7 and 293 cells were grown on 16 mm plates in D-MEM with 10% fetal calf serum. HeLa and Namalwa cell nuclear extracts were prepared essentially as described (38) with HEPES replaced by Tris.

In vitro transcription

Primers for primer extension were 5′-end labelled with T4 polynucleotide kinase (from NEB) according to standard protocols. The oligonucleotide used for detecting transcription from all κ promoter constructs had the sequence CATTTTAGCTTCCTTAGCTCCTG, complementary to the translation start region of the CAT gene in pCATenh. The primer complementary to the -50-MLP/C2AT construct had the sequence GGGGTGAGAGTGAATGATGATAG and produced a signal of 130 bases. The κ promoter construct (500 ng) and the -50-MLP internal control construct (150 ng) were transcribed with ~50 µg of nuclear extract (10 µl in Buffer D) in a 25 µl reaction containing 10 mM HEPES (pH 8.4), 3 mM $MgCl_2$ and 600 µM NTPs. The reactions were incubated at 30°C for 60 min and then stopped with 10 mM EDTA, 0.5% SDS and 0.5 M NaOAc (pH 5.2) in a final volume of 200 µl. After a phenol/chloroform and subsequent chloroform extraction, the RNA was precipitated with 30 µg/ml glycogen and 3 vol. EtOH. Primer extension reactions were carried out essentially according to standard methods using AMV reverse transcriptase from Promega. Run-off transcription was performed as described in reference 3. *In vitro* transcription gels were quantitated using an image scanner from Molecular Dynamics.

Transfection

Electroporation of Namalwa cells was performed using a culture in log phase with a concentration of ~700 000 cells/ml. The cells were gently pelleted and resuspended in RPMI medium with 10% FCS to a concentration of 33×10^6 cells/ml. An aliquot containing 10^6 cells (0.3 ml) was then briefly mixed with 15 µg of the TK-β-gal construct (as internal control) and 15 µg of the T1/CAT test construct in 20 µl. The reaction was then pipetted into an electroporation cuvette with a gap of 0.4 cm and pulsed with 230 V and 960 µF with a BioRad Gene Pulser. The culture was then transferred immediately to a T flask containing 10 ml RPMI with 10% FCS and incubated at 37°C in 5% CO_2 for 48 h. The cells were harvested and the lysates were prepared for CAT–ELISA as described in the kit from Boehringer. HeLa suspension cells were transfected as above except that they were pulsed with 450 V and 500 µF in S-MEM medium. Forty micrograms of each T1/CAT construct were used. Adherent cultures of 293 cells were co-transfected with 40 µg of each reporter plasmid by calcium phosphate precipitation following standard protocols. The cultures were then harvested after 48 h and extracts were prepared according to the freeze–thaw protocol described in the CAT-ELISA kit. COS7 cells were transfected with 10 µl lipofectamine (BRL) using 2 µg reporter construct. After 48 h, the cultures were harvested as described in the CAT-ELISA kit with the supplied 5× lysis Buffer. CAT-ELISA and β-gal assays were performed as described in the Boehringer manual.

DNA-binding assay

EMSA-conditions were essentially as described (15) but using HEPES (pH 7.9) and a 25 mM Tris, 250 mM glycine (pH 8.3) gel system.

RESULTS

The sequence from –3 to +6 relative to the transcription start site (ACCATTGTC) of an Igκ light chain promoter revealed a similarity to the consensus initiator sequence (YYCAYYYYY) (39). To address a possible function of the initiation site sequence a 3′ deletion was generated extending to position –2 in the κ promoter. It was compared to a construct with wild-type downstream sequences extending to position +39 in an *in vitro* transcription assay using Namalwa nuclear extract (Fig. 1). The adenovirus major late promoter (ML) was used as an internal control. A reproducible reduction of transcription of at least 3-fold was observed (Fig. 1A, lanes 1 and 2). To determine whether this was due to the sequence at the initiation site, a construct with downstream sequence to +6 was generated (Fig. 1). Surprisingly, however, restoration of these further downstream bases containing the putative initiator sequence could not counteract the reduced transcription rate of the truncated promoter (Fig. 1A, lane 3). Hence, bases between +7 and +39 in the κ promoter appear to be important for full transcriptional activity. The contribution of the well characterized octamer element on the T1 κ promoter (3,35) was measured for comparison. Point mutation of the octamer site (from –66 to –59) resulted in at least a 5-fold reduction of activity in our assay (Fig. 1A, lane 4). Thus, the downstream sequences are as critical for promoter activity as the octamer element.

To rule out that the initiation site was shifted in the deletion construct, the transcripts produced from both the +39 and the +6 constructs (Fig. 1B, lanes 1 and 2) were run alongside sequencing reactions (Fig. 1B, lanes 3 and 4). The same nucleotides were used in both cases as transcriptional initiation sites *in vitro*, proving that deletion of downstream sequences in the promoter affected the rate but not the accuracy of transcription initiation. These start site residues map to the same region as the initiation site proposed earlier for this gene (2,3).

When the linearized promoter constructs were tested in a run-off transcription assay using Namalwa nuclear extract (Fig. 1C), the same requirement of the downstream region was observed as with the supercoiled templates used in the primer extension assay. Transcription initiated from the –2 construct (lane 3) was reduced to a greater extent than that from the template with mutated octamer (lane 2). Consequently, the topology of the template did not play a role in the activating potential of the downstream element.

Ig promoters are transcriptionally active only in mature B lymphocytes, and while this cell type-specific activity has been attributed primarily to regulatory protein–DNA interactions via

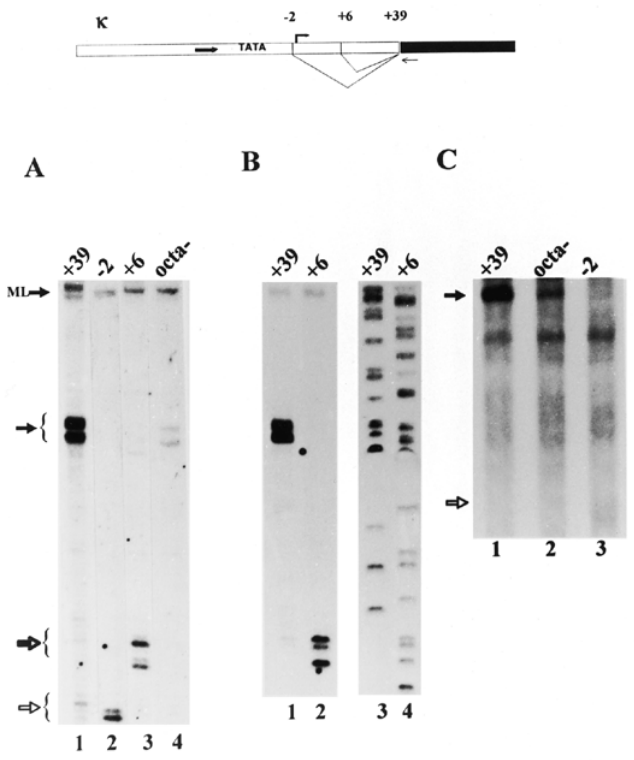

Figure 1. Effect of a downstream sequence element on κ promoter transcription in Namalwa nuclear extracts. Top, Schematic presentation of the light chain promoter constructs. The filled-in lines denote the pCATenh vector, the filled-in arrow the octamer site and the thin arrow the CAT primer. (**A**) Primer extension analysis of transcripts initiated from κ promoter constructs with sequences from –131 to either +39 (lane 1), –2 (lane 2), +6 (lane 3) or +39 with a mutated octamer sequence (lane 4). ML indicates the signal produced by the internal control construct, the second arrow denotes the +39 and octa- transcripts and the following two arrows show the +6 and –2 transcripts, respectively. (**B**) Primer extension of *in vitro* transcripts from the +39 and +6 κ constructs (lanes 1 and 2). The same plasmids and labeled primer were used to generate a T ladder representing the sense strand of the promoter constructs (lanes 3 and 4). The reactions were run simultaneously on the same gel. (**C**) Run-off transcripts from the +39 (lane 1), the octamer point mutant (lane 2) and the –2 construct (lane 3) in Namalwa extract. The filled-in arrow denotes the +39 and octa- transcripts and the open arrow the –2 transcript.

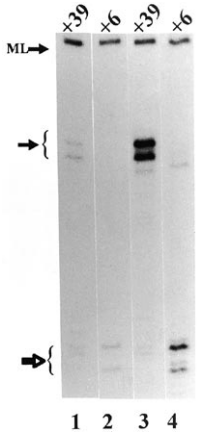

Figure 2. Cell specific effect of a κ sequence element from +7 to +39 in an *in vitro* transcription assay. The +39 (lanes 1 and 3) and +6 (lanes 2 and 4) constructs were transcribed both in HeLa (lanes 1 and 2) and Namalwa (lanes 3 and 4) nuclear extracts. The internal control signal (ML) and the κ promoter transcripts are denoted by arrows.

the octamer element (8,9 and references therein), it is possible that further elements contribute to tissue-specific promoter function. Therefore, we asked whether the identified downstream element could play a role in contributing to the cell type-specific expression of this gene. κ promoter constructs containing the natural sequences from –131 to either +6 or +39 were transcribed in either HeLa cell or Namalwa B cell nuclear extracts (Fig. 2, lanes 1 and 2 versus 3 and 4, respectively). Both extracts transcribed the ML to the same extent (Fig. 2), providing evidence that they contained the same activity of pol-II basal factors. Both promoter constructs were more active in B cell extracts than in HeLa cell extracts, as expected (3). However, the 9-fold cell type-specific difference in transcription observed with the +39 construct (Fig. 2, lanes 1 versus 3) was markedly reduced to a 3-fold difference when the downstream sequence was deleted (lanes 2 and 4). Thus, the lymphoid-specific activity of the κ promoter determined in cellular extracts appeared to be dependent to a large extent on the sequence close to and downstream of the initiation site.

We next examined how this region, which apparently contributes to cell type specificity *in vitro*, would influence transcription *in vivo*. Namalwa cells were transfected with the κ promoter constructs and examined for levels of CAT expression (Fig. 3). TK-β-gal was co-transfected and CAT levels were normalized for β-gal activity. The reduction in promoter activity upon deletion of the downstream region was even more pronounced in intact cells. The +6 construct showed an ~5-fold decrease in transcription compared to the +39 construct, confirming the observation made *in vitro*. However, when transfecting the κ promoter constructs into non-B cells, an equivalent dependence on the downstream sequence was observed. The +39 and +6 constructs were transfected into several non-B cell lines (HeLa suspension and adherent cell lines, COS7 and 293 cells). In each of the cell lines transfected and with each transfection protocol used, the +6 construct showed at least a 5-fold reduction in CAT expression as compared to the +39 construct (data not shown). The downstream portion of the κ promoter was thus required for full activity also in non-B cells. This discrepancy in cell type dependency between the *in vivo* and *in vitro* data could be due to differences in concentrations of activating factors in nuclear extracts versus intact cells or to the influence of the vector in transfection experiments.

We next analyzed to what extent the sequences from +7 to +39 would functionally interact *in vivo* with the two known control elements of the κ promoter, the octamer and TATA box. The effect of replacing the promoter sequences downstream of position +6 was in fact equivalent to obliterating the octamer element, which also reduced transcription by 5-fold in the +39 context in Namalwa cells (Fig. 3), similar to the observation made *in vitro* (Fig. 1, lanes 2 and 4). Replacement of the downstream region in the presence of the mutated octamer element resulted in a further 5-fold reduction of transcription from this weakened promoter (Fig. 3). This double mutation rendered the promoter as ineffective as the promoter-less background vector. Thus, the downstream portion of the promoter apparently functions independently of the octamer element and its associated factors.

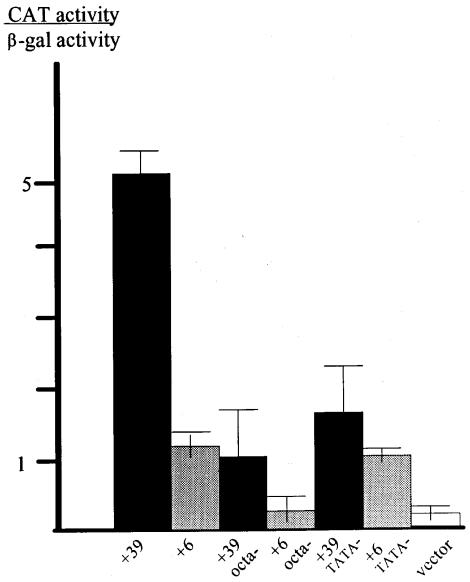

Figure 3. Activity of κ light chain promoter mutants in Namalwa cells. Namalwa cells were transfected with the indicated CAT constructs and with a thymidine kinase promoter β-gal construct used to control for transfection efficiency. The values are normalized for β-gal activity and were determined in triplicate experiments.

Further constructs were transfected which contained point mutations to convert the TATA box into a GC rich *Not*I site. The mutation of the TATA box completely inactivated the promoter *in vitro* (data not shown), suggesting that the initiation site and downstream regions strongly synergize with the TATA box region *in vitro*. In transfected B cells the TATA-less promoter exhibited the expected decrease in activity (Fig. 3). Truncation of the downstream region in the context of a mutated TATA box (+39/TATA- versus +6/TATA-) did not markedly reduce transcription. This may indicate a functional dependency between the TATA box and downstream elements.

To further delineate the sequence requirements we introduced point mutations into both the initiation site region and sequences further downstream. We first focused on a sequence with homology to an initiator element between positions –2 and +5, that reoccurs in a slightly modified version between positions +5 to +11 (Fig. 4). Either one of the T residues at +3 or at +9 or both were mutated to G, resulting in mutants mINR-A, mINR-B and mINR-AB, which otherwise retained the complete κ promoter sequences from –131 to +39 (Fig. 4). These bases are 2 nt downstream of a CA dinucleotide in the two repeated YCATTGY motifs, respectively. A T residue at the corresponding position in a consensus initiator element (YCANTYY) is most crucial for initiator function (29). When tested in an *in vitro* transcription reaction with B cell nuclear extract, mINR-A and mINR-B showed impaired activity (3- and 5-fold, respectively) compared to the wild-type promoter (Fig. 4A, lanes 1 versus 2 and 3). The double mutation drastically affected transcription, completely eliminating detectable signals (lanes 1–3 versus 4). Very similar results were obtained when the same mutants were transfected into Namalwa cells. mINR-A and mINR-B were inactivated by 40% and 70%, respectively and the double mutation caused a 5-fold inactivation compared to the wild-type promoter (Fig. 4B).

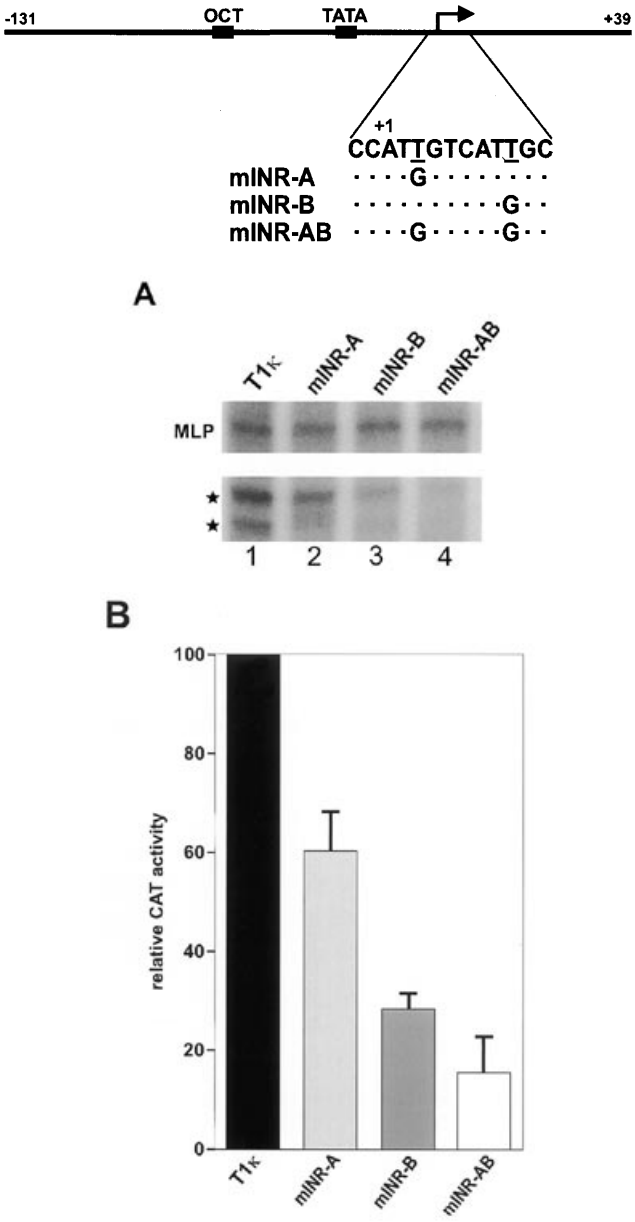

Figure 4. Analysis of initiator-like elements by point mutagenesis. Top, diagram of the κ light chain promoter showing the tandem initiator-like sequences and the point mutations introduced. (**A**) *In vitro* transcription in B cell nuclear extracts of the wild-type Igκ promoter (from –131 to +39) (lane 1) and point mutants with T to G replacements at +3, +9 or both (mINR-A, mINR-B, mINR-AB, lanes 2–4, respectively). MLP, adenovirus major late promoter internal control. (**B**) CAT assay. The wild-type Igκ promoter or mutants mINR-A, mINR-B or mINR-AB were transfected into Namalwa cells together with a thymidine kinase promoter-regulated β-gal construct. The average CAT values of three independent experiments, corrected for transfection efficiencies, are shown.

To further investigate the contribution of the initiator-like elements as well as of nucleotides further downstream, sequences were blockwise replaced by random sequences (Fig. 5). For each region several clones containing randomized sequences were tested by *in vitro* transcription in Namalwa cell nuclear extract.

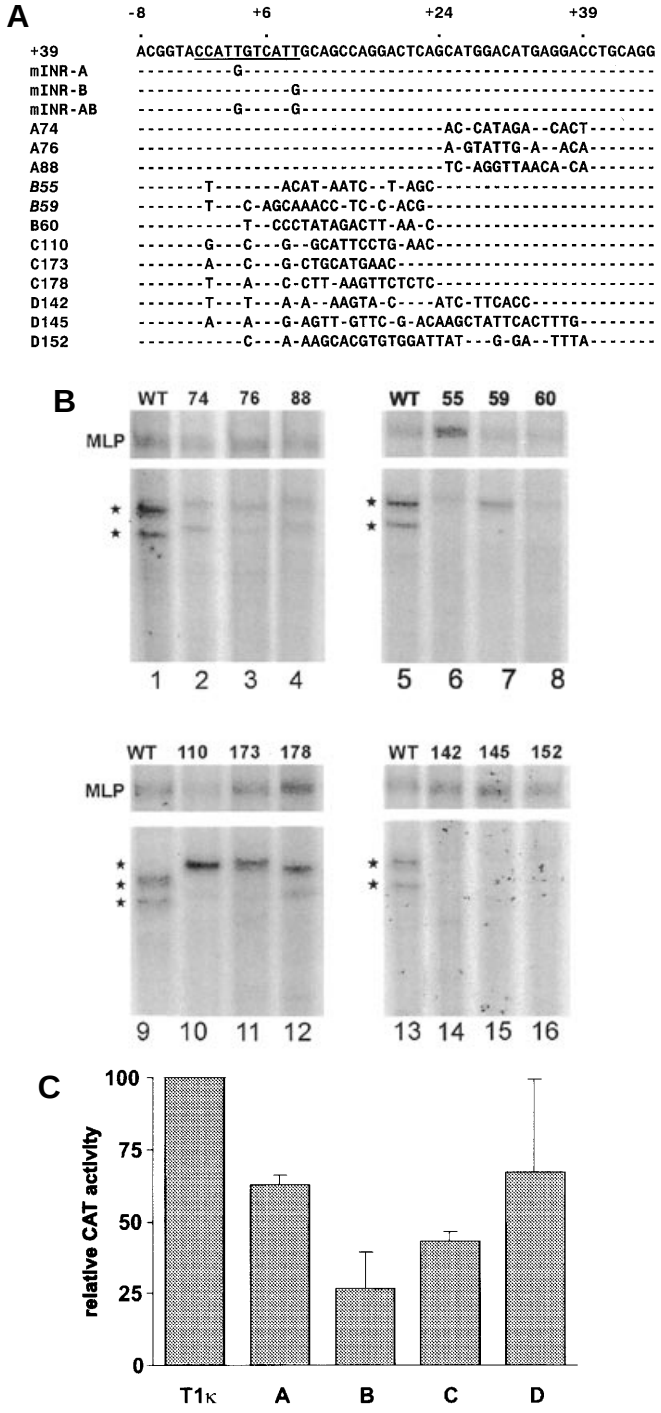

Figure 5. Analysis of initiator-like elements and further downstream sequences of the T1κ promoter by random mutagenesis. (**A**) Sequence alignment of random cassette mutants A, B, C and D with initiation region point mutants (mINR-A, mINR-B, mINR-AB), as indicated. The initiator-like sequence is underlined. Mutagenized residues are indicated. (**B**) Primer extension assay. *In vitro* transcription activities in Namalwa extract of the wild-type T1κ promoter (WT) (lanes 1, 5, 9 and 13) compared to random cassette mutants A (lanes 2–4), B (lanes 6–8), C (lanes 10–12) and D (lanes 14–16). T1κ promoter transcripts are indicated by asterisks, the internal control adenovirus major late transcript by MLP. Numbers on top of the lanes indicate the individual clones. (**C**) Random cassette mutants of the A, B, C and D series were transfected into Namalwa cells. The average CAT activity of three different clones for each cassette, as used in (B), is indicated. Each individual transfection was performed in triplicate and corrected for transfection efficiency by a co-transfected β-gal expression vector.

When the sequences from –2 to +23 were randomized (B55, B59 and B60, Fig 5A), only keeping bases within the first initiator-like motif, but eliminating the second motif, all single mutants were at least 5- to 10-fold inactivated (Fig. 5B, compare lane 5 with lanes 6–8). Thus, the sequence context in the –2 to +23 region contributes to promoter function at a high level. In a further set of mutants the second initiator-like sequence was essentially maintained in the randomized region (C110, C173 and C178, Fig 5A). These constructs showed a marked increase of *in vitro* transcription activity compared to the B series (Fig. 5B, lanes 10–12 versus 6–8), consistent with the importance of the second motif defined in Figure 4. Of note, however, is a slight shift in the selection of the start site (Fig. 5B, lanes 9 compared to 10–12). Another set of mutants (D142, D145 and D152, Fig. 5A) in which the same residues in the tandem repeat were maintained as in the C mutants, but in which further bases were randomized between nt +24 and +39 showed even more strongly reduced transcription *in vitro* (Fig. 5B, lanes 13 versus 14–16). Although individual base exchanges in the initiation site proximal region may have accounted for this effect, another possibility is a functional role of mutated sequences from +24 to +39. To assess the latter possibility, we generated constructs with mutated sequences positioned exclusively between residues +24 and +39 (A74, A76 and A88, Fig. 5A). Surprisingly, all of these were >5-fold less active than the wild-type promoter (Fig. 5B, lanes 1–4), despite the unchanged tandem array of initiator-like sequences. Consequently, the κ gene contains an extended core promoter structure with critical sequences around the transcriptional start site(s) between positions –2 and +11 and further downstream between +24 and +39.

The same mutants were also tested by transfection into Namalwa cells (Fig. 5C). The B mutants displayed a similar reduction as observed for mINR-B (Fig. 4A). The increase in activity of the C mutants relative to the B series was visible, but not as pronounced as *in vitro* (Fig. 5C). Similarly, transcription of the A and D mutants was reduced, but not as severely as *in vitro*.

Our data could be explained by two possibilities. The first is that TFIID binds to the TATA box, but also interacts in a sequence-specific manner with regions surrounding the initiation site and further downstream. In other promoters it has been observed that DNase I footprints of the TATA binding complex extended over the initiation site and into the transcribed portion of the gene (25). In *Drosophila* TATA-deficient promoters TFIID binds a conserved downstream element (30). One could also consider the possibility of a separate protein with its own DNA binding domain specific for the downstream region and interacting with the TATA complex through protein–protein interaction, resulting in a more stable protein–DNA complex. Therefore, it was interesting to see if DNA binding activity was associated with this functionally important region of the light chain promoter..

Several attempts to detect sequence specific binding in HeLa or Namalwa nuclear extracts with an oligonucleotide probe containing the sequences from +7 to +39 failed (data not shown). A possible explanation was that sequences further upstream of +7 were needed to form a complete binding site. In fact, a probe containing the sequence from –9 to +39, when incubated with Namalwa nuclear extracts, generated in EMSA two complexes, the faster migrating complex being much stronger than the slower one (Fig. 6A, lane 1). When the two κ promoter constructs containing sequences either to +39 or to +6 were used as competitors, the +39 construct competed much more efficiently

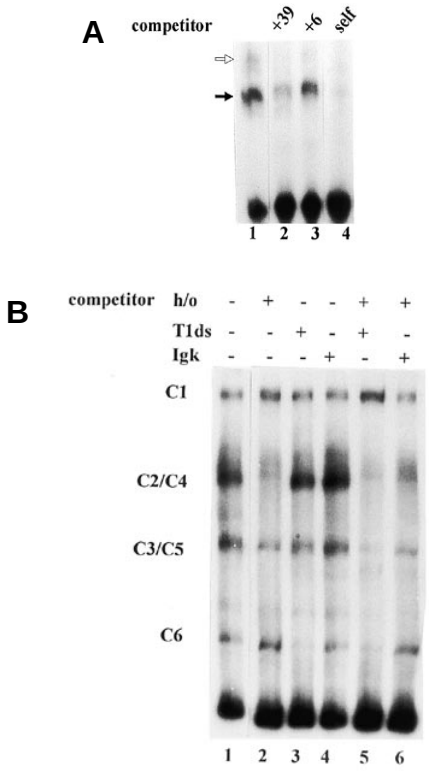

Figure 6. (**A**) Gel shift assay using the promoter sequence from –9 to +39 as the labeled probe and Namalwa nuclear extract. Promoter constructs extending to +39 (lane 2) or +6 (lane 3) were used as cold competitors (140-fold molar excess) and the self competition (lane 4) was performed at a 66-fold excess. (**B**) Gel mobility shift assay using the promoter from –131 to +39 as the labeled probe in a binding reaction with Namalwa nuclear extract. Unlabeled oligonucleotides (heptamer-octamer, h/o, κ promoter –9 to +39, T1ds and the light chain enhancer NF-κB site, Igκ) were used in a 50-fold molar excess.

for the binding of both complexes (lanes 2 and 3). Consequently, the sequence from +6 to +39 was necessary, although not sufficient, for the observed binding activity. Since the probe did not contain the TATA box, the complexes are most likely formed by proteins with their own DNA binding domains, independent of TBP.

Binding assays performed previously with a κ light chain promoter fragment extending from –131 to +39 and Namalwa cell nuclear proteins yielded six different protein–DNA complexes, C1 through to C6 (28; Fig. 6B). Two of these complexes (C2 and C3) correspond to Oct-1 and Oct-2, respectively. The complexes C4 and C5 migrate in similar positions as C2 and C3 respectively, and are visible only after competition of the octamer factors with an octamer containing oligonucleotide (15; Fig. 6B, lane 2). Complexes C1 and C4 bind to regions overlapping with and upstream of the octamer sequence, respectively (15). When an oligonucleotide containing the initiation site and the downstream sequence (from –9 to +39) was used as a competitor, the C6 complex disappeared whereas the other complexes remained essentially unchanged (lane 3). A non-specific oligonucleotide showed no effect on complex formation (lane 4). In the presence of the octamer oligonucleotide to mask Oct-1 and Oct-2, the downstream competitor clearly affected C6 and possibly C5 (compare lanes 2 and 5). The non-specific oligonucleotide added in conjunction with the octamer oligonucleotide to the binding reaction showed no additional effects (compare lanes 2 and 6). It can be concluded that the previously identified complexes C6, and most likely also C5, bind to sequences downstream of –9. Similar results were obtained when probes extending from –131 to +39 and from –131 to –2 were compared in gel retardation assays (not shown).

DISCUSSION

Immunoglobulin genes have been studied intensively as a model system for tissue-specific gene regulation in lymphoid cells. Several control elements have been identified in the promoter and enhancer regions of heavy chain and light chain genes which can confer cell type-specific activity (ref. 1, for review). In contrast to the complex modular character of the Ig enhancers, the promoters show a simple pattern of conserved elements, consisting of the TATA box and the octamer element. Despite the lack of further obvious conserved elements, both heavy chain and light chain promoters interact with other common factors, in addition to the Oct-1/Oct-2 proteins which bind to the octamer (15). In this study, we have investigated the influence of promoter sequences overlapping with and extending downstream of the transcriptional start site of an Igκ light chain gene. In both intact cells and an *in vitro* transcription system the sequence between +7 and +39 relative to the CAP site was required for full activity and its replacement resulted in an inactivation comparable to mutation of the octamer element. Whereas the downstream sequence supported transcription equally in both transfected B lymphoid and non-lymphoid cells, it had a distinct cell type specific effect *in vitro*. The efficient transcription observed in B cell nuclear extracts was much more dependent on the downstream sequence than was the weaker transcription in HeLa cell nuclear extracts. These observations suggest possible functional interactions of lymphoid specific regulators with the downstream sequences.

Several observations confirm that the downstream sequence exerts its effect via transcription activation and rule out possible influences of nucleic acid structure. First, the effect was dependent on the template:protein ratio. At higher template concentrations, the activation was titrated out (data not shown), typical for a factor dependent effect. The transcription activation also showed a dependency on extract preparation, i.e. different extract preparations exhibited *cis*-activation caused by the downstream sequence to different extents. The analysis in run-off assays also revealed a strong reduction of transcription upon replacement of the sequences from +7 to +39 (Fig. 1C), ruling out possible differences in transcript mapping efficiencies for the 3′ deletion constructs. Therefore, we expect that the downstream region acts through protein–DNA interactions on the transcriptional machinery.

Using a number of mutated templates we have identified as functional elements a tandem-array of initiator-like sequences between positions –2 and +11, and a sequence located further downstream between +24 and +39 relative to the start site of transcription. Both initiator-like sequences reveal homology to classical initiators and in each case mutation of a T residue three bases downstream of a CA dinucleotide caused strong inactivation. The same position was identified as functionally most critical in the TdT type initiator (29). A double mutation almost completely eliminated transcription, both *in vitro* and in intact cells.

A

```
T1κ          CTTCATATACCCGTCACACATGTACGGTACCATTGTCATTGCAGCCAGGACTCAGCATGGA      (22, 28)
M22009       ATTAATAGGCTGGACGACTCATGCAGGAATCAGTCCCACTCAGGACACAGCATGGA           (30)
X57828       AAGAGGATAAGACAGGAGCTCATAGTGTCCAGGTTCCACTGGGCAGTCTCGAATAGAGCTCTTGGA (25)
M64442       GAGATTTAATAATCTGATCATACACACTTCAACAGTCATTCTTGGTCAGGAGACGTTGTAGAAATG (24)
Z72382       TGGCCTGAGATAAAACCTCAAGTGTCCTCTTGCCTCCACTGATCACTCTCCTATGTTTATTTCCTC (22, 29)
L23151       TGATTTATAGGACATCCAATCTGGCTTCTGGAGTCCCACCTCGCTTCGGTGGCAGTGGGTCTGGGA (28)
M18756       GTTATAAACCAGGCCTTTGCTGTAAGAACAGAAATACATCAGGCAGGCAAAGGCATCAAGATGAA  (29)
```

B

```
T1κ      TTCATATACCCGTCACACATGTAC-----------GGTACCATTGTCATTGCAGCCAGGACTCAGCATGGACATGAGGAC
X59315   TTTTTTATGGGCTGGTCGCACCCTGTGCA------GGAGTCAGTCTCAGTC-----AGGACACAGCATGGACATGAGGGT
X12686   ATATCAATGCCTGGGTCAGAGCTCTGGAGAA-------GAGCTGCTCAGTT-----AGGACCCAGAGGGAACCATGGAAA
X59314   GATAAAAAGCTCAGCTCTATCCTTGCCTTGACTGATCAGGACTCCTCAGTTCACCTTCTCACCATGAGGCTCCCTGCTCA
```

Figure 7. Conservation of initiator-like sequences and downstream sequences in human and murine Ig genes. (**A**) Homology of initiator-like sequences between murine T1κ and other light chain promoters. Translation initiation codons are underlined, TATA boxes and potential initiator-like elements are shown in bold. GenBank accession numbers are shown on the left, distances between TATA boxes and potential initiator regions on the right. M22009, human Igκ chain variable region from γ-H chain disease; X57828, human germline immunoglobulin λ light chain gene; M64442, mouse germline IgMκ light chain gene; Z72382, murine Igκ-Vκ2(70/1) gene; L23151 and M18756, murine Igκ genes. (**B**) Alignment of sequences downstream of the TATA elements of murine T1κ (murine class V) with representative human subgroup I (X59315), subgroup II (X59314) and subgroup III (X12686) κ light chain promoters (see ref. 48 for further members of these subgroups). Gaps were introduced to show homologies. TATA elements, potential initiator like elements and conserved leader sequences are shown in bold. Translation initiation codons are underlined.

We have also investigated whether the initiator-like sequence (–2 to +11) would direct initiation without a TATA box or whether it would enhance transcription when fused to a heterologous TATA box. To this end we have compared the κ promoter intiation site tandem motif with the TdT initiator, either alone or fused to the adenovirus major late promoter TATA box. The κ sequence could not direct initiation by itself and could not synergize with the MLP TATA box in the heterologous context *in vitro* (data not shown). This may indicate that a specific TATA box sequence or additional sequences are required for its proper function. The observation that a TATA box mutation completely inactivated the κ promoter (M.R.P., unpublished data) in fact suggests that the initiator-like sequence, although as critical for transcription *in vitro* as the TATA box, is not able to direct initiation alone. The sequence from +24 to +39 had a pronounced stimulatory effect on transcription *in vitro*, when both initiator-like sequences were intact and its replacement by random sequences enhanced the effect of mutations in the initiation site region. It has been shown previously that TFIID binding to pol-II transcribed promoters leads to footprints extending to +35 (25,40). It is also known that sequences downstream of the initiation site are bound by TAFs in various *Drosophila* promoters, leading to more stable TFIID interaction (26,27,41). The Igκ promoter downstream element resides in a position similar to that of an element (DPE) found 25–30 nt downstream of initiator sequence in *Drosphila* TATA box deficient genes (30), which is bound by TFIID. Although no sequences homologous to a DPE are found in Igκ promoter downstream sequences, a similar functionally important DNA-contact of TFIID components is possible.

Using combined octamer, TATA box and downstream mutations we determined if these elements do cooperate. In transfected B cells the promoter could be inactivated almost to an equivalent extent by a mutation of either the TATA box or of the octamer or by deletion of the downstream sequence (Fig. 3). Without the original downstream sequences, a mutation of the TATA box had no effect on transcription (+6 versus +6/TATA-). Similarly, when the TATA box was mutated, replacement of the downstream sequence resulted in only a very weak reduction of promoter activity (+39/TATA- versus +6/TATA-). In contrast, without a functional octamer, the intact basal promoter was still fully responsive to the downstream sequences (+39/octa- compared to +6/octa-) and deletion of the octamer in the +6 background greatly reduced transcription (+6 versus +6/octa-). These observations suggest that the TATA sequence and the downstream sequence functionally depend on each other and are bound by the same or interacting component(s). The TFIID complex may contact both sites, or it may interact with a separate factor binding to the downstream sequence. Another interesting observation was that the octamer-containing promoter without TATA box and downstream sequence (+6/TATA-) was still as active as the intact basal promoter lacking the octamer (+39/octa-). The factors which bind to the octamer are presumably able to restore transcription by tethering TFIID to the promoter, even when the TATA box and/or downstream elements are eliminated. This view is supported by recent reports which demonstrate functional interactions between TFIID and Oct-2 and direct protein interactions between TBP and Oct-2 (42,43).

Two nuclear factors were found to bind specifically to a short probe containing both the initiation site and the downstream element, when its sequence was extended further upstream until position –9 (Fig. 5A and B). Competition experiments with different κ promoter constructs showed that sequences from +7 to +39 were required for complex-formation. Using the whole κ promoter as a probe, we showed that the +7 to +39 region is bound by the previously identified Ig promoter specific complexes C5 and C6. C5 should correspond to the major complex detected with the short probe, whereas C6 was only detectable with the long probe. In contrast to the C5 activity, which was detected in both HeLa cells and in B cells, C6 was restricted to B cells (35) and possibly contributes to the cell-type specific influence of the downstream sequence observed in extracts of lymphoid cells *in vitro* (Fig. 2). The sequence bound

by C5 and C6 also interacts with recombinant C/EBPβ (E.N.H., unpublished observation) but does not contain further obvious sites for other known transcription factors. The elucidation of the identity and of the functional relevance of C5 or C6 in κ promoter regulation has to await the purification and further analysis of these factors.

Several reports have described transcriptionally active sequences downstream of the start sites in various genes (31,32,44,45). Regulatory elements downstream of or overlapping with the transcriptional initiation site (ref. 23 for review) have also been found in genes whose expression is restricted to lymphoid cells, such as the terminal deoxynucleotidyl transferase gene and the Vγ3 receptor (39,46).

Although in several instances specific proteins have been shown to interact with Inr sequences or other downstream elements, no correlation between protein binding and Inr function has yet been demonstrated. Possibly these proteins modulate recognition of the Inr by general trancription factors, such as the TFIID complex (see ref. 29 for discussion). Similarly, TBP-associated factors may interact with both initiator region and downstream element of the κ promoter as has been shown for TATA-less promoters (30,47), and gene-specific factors, such as C5 and C6, may have modulator functions.

What may be the implications of a more complex regulation of immunoglobulin promoters? Ig genes generally contain weak non-consensus TATA boxes, consistent with the presence of initiator-like and downstream elements described here. A more complex composition of Ig core promoters may serve to provide alternative targets for transcriptional activators and repressors of these genes. The search for sequence conservation in the leader sequences and transcriptional initiation site sequences of various κ gene variable regions revealed a striking conservation of initiator-like sequences in human and murine κ light chain promoters (Fig. 7). Moreover, motifs similar to YYCANTY initiator motifs are frequently found around transcriptional start sites of a larger number of vertebrate Ig light chain and even heavy chain genes. Furthermore, the κ gene analysed here (murine class V), which is highly homologous to a number of murine κ genes, reveals extended downstream homology to human subgroup I and III Igκ variable genes (Fig. 7). The occurrence of initiator-like and downstream regulatory elements in a κ light chain promoter may provide an additional level by which immunoglobulin gene transcription can be affected during B cell development.

ACKNOWLEDGEMENTS

We thank Dr S.T. Smale for discussion. This work was funded in part by a grant from the Deutsche Forschungsgemeinschaft (SFB 344) to C.S.

REFERENCES

1 Staudt, L.M. and Lenardo, M.J. (1991) *Annu. Rev. Immunol*., **9**, 373–398.
2 Falkner, F.G. and Zachau, H.G. (1984) *Nature*, **310**, 71–74.
3 Mitzushima-Sugano, J. and Roeder, R.G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8511–8515.
4 Jenuwein, T. and Grosschedl, R. (1991) *Genes Dev.*, **5**, 932–943.
5 Clerc, R.G., Corcoran, L.M., LeBowitz, J.H, Baltimore, D. and Sharp, P.A. (1988) *Genes Dev.*, **2**, 1570–1581.
6 Scheidereit, C., Cromlish, J.A., Gerster,T., Kawakami, K., Balmaceda, C.-G., Currie, R.A. and Roeder, R.G.(1988) *Nature*, **336,** 551–557.
7 Sturm, R.A., Das, G. and Herr, W. (1988) *Genes Dev.*, **2,** 1582–1599.
8 Gstaiger, W., Knoepfel, L., Georgiev, O., Schaffner, W. and Hovens, C.M. (1995) *Nature*, **373**, 360–362.
9 Strubin, M., Newell, J.W. and Matthias, P. (1995) *Cell*, **80**, 497–506.
10 Luo, Y. and Roeder, R.G. (1995) *Mol. Cell. Biol.*, **15**, 4115–4124.
11 Corcoran, L.M., Karvelas, M., Nossal, G.J., Ye, Z.S., Jacks, T. and Baltimore, D. (1993) *Genes Dev.*, **7**, 570–82.
12 Kim, U., Qin, X.F., Gong, S., Stevens, S., Luo, Y., Nussenzweig, M. and Roeder, R.G. (1996) *Nature*, **383**, 542–547.
13 Schubart, D.B., Rolink, A., Kosco-Vilbois, M.H., Botteri, F. and Matthias, P. (1996) *Nature*, **383**, 538–542.
14 Parslow T.G., Blair, D.L., Murphy, W.J. and Granner, D.K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2650–2654.
15 Franke, S., Scholz, G. and Scheidereit, C. (1994) *J. Biol. Chem*., **269**, 20075–20082.
16 Cooper, C., Johnson, D., Roman, C., Avitahl, N., Tucker, P. and Calame, K. (1992) *J. Immunol*., **149**, 3225–3231.
17 Atchison, M.L., Delmas, V. and Perry, R.P. (1990) *EMBO J*., **9**, 3109–3117.
18 Yoza, B.K. and Roeder, R.G. (1990) *Mol. Cell. Biol*., **10**, 2145–2153.
19 Sigvardsson, M., Akerblad, P. and Leanderson, T. (1996) *J. Immunol*., **156**, 3788–3796.
20 Parvin, J.D., Timmers, H.Th.M. and Sharp, P.A. (1992) *Cell*, **68**, 1135–1144.
21 Kim, J., Parvin, J.D., Shykind, B.M. and Sharp, P.A. (1996) *J. Biol. Chem*., **271**, 18405–18412.
22 Parvin, J.D. and Sharp, P.A. (1991) *J. Biol. Chem*., **266,** 22878–22886.
23 Weis, L. and Reinberg, D. (1992) *FASEB J*., **6**, 3300–3309.
24 Gill, G. (1994) *Curr. Biol*., **4**, 374–375.
25 Horikoshi, M., Hai, T., Lin, Y.-S., Greene, M.R. and Roeder, R.G. (1988) *Cell*, **54**, 1033–1042.
26 Emanuel, P.A. and Gilmore, D.S. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 8449–8453.
27 Purnell, B.A., Emanuel, P.A. and Gilmour, D.S. (1994) *Genes Dev*., **8**, 830–842.
28 Verrijzer, C.P., Chen, J.-L., Yokomori, K. and Tjian, R. (1995) *Cell*, **81**, 1115–1125.
29 Javahery R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S.T. (1994) *Mol. Cell. Biol*., **14**, 116–127.
30 Burke, T.W. and Kadonaga, J.T. (1996) *Genes Dev*., **10**, 711–724.
31 Jones, K.A., Luciw, P.A. and Duchange, N. (1988) *Genes Dev*., **2**, 1101–1114.
32 Kato, H., Horikoshi, M. and Roeder, R.G. (1991) *Science*, **251**, 1476–1479.
33 Wiley, S.R., Kraus, R.J., Zuo, F., Murray, E.E., Loritz, K. and Mertz, J.E. (1993) *Genes Dev*., **7**, 2206–2219.
34 Kashanchi, F., Duvall, J.F., Dittmer, J., Mireskandari, A., Reid, R.L., Gitlin, S.D. and Brady, J.N. (1994) *J. Virol*., **68**, 561–565.
35 Scheidereit, C., Heguy, A. and Roeder, R.G. (1987) *Cell*, **51**, 783–793.
36 Altenburger, W., Steinmetz, M. and Zachau, H.G. (1980) *Nature*, **287**, 603–607.
37 Sawadogo, M. and Roeder, R.G. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4394–4398.
38 Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acids Res*., **11**, 1475–1489.
39 Smale, S.T. and Baltimore, D. (1989) *Cell*, **57**, 103–113.
40 Nakatani, Y., Horikoshi, M., Brenner, M., Yamamoto, T., Besnard, F., Roeder, R.G. and Freese, E. (1990) *Nature*, **348**, 86–88.
41 Verrijzer, C.P., Yokomori, K., Chen, J.-L. and Tjian, R. (1994) *Science*, **264**, 933–941.
42 Arnosti, D.N., Merino, A., Reinberg, D. and Schaffner, W. (1993) *EMBO J*., **12**, 157–166.
43 Zwillig, S., Annweiler, A. and Wirth-T. (1994) *Nucleic Acids Res*., **22**, 1655–1662.
44 Ayer, D.E. and Dynan, W.S. (1988) *Mol. Cell. Biol*., **8**, 2021–2033.
45 Jarrell, K.A. and Meselson, M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 102–104.
46 Clausell, A. and Tucker, P.W. (1994) *Mol. Cell. Biol*., **14**, 803–814.
47 Martinez, E., Chiang, C.M., Ge, H. and Roeder, R.G. (1994) *EMBO J*., **13**, 3115–3126.
48 Schable, K.F. and Zachau, H.G. (1993) *Biol. Chem. Hoppe Seyler*, **11**, 1001–1022.