Human HMG box transcription factor HBP1: a role in hCD2 LCR function

Talgat Zhuma, Richard Tyrrell[1], Belaïd Sekkali, George Skavdis, Alexander Saveliev[2], Mauro Tolaini, Kathleen Roderick, Trisha Norton, Steve Smerdon[1], Steve Sedgwick[3], Richard Festenstein[2] and Dimitris Kioussis[4]

Division of Molecular Immunology, [1]Division of Protein Structure and [3]Division of Yeast Genetics, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

[2]Present address: Department of Medicine, Imperial College of Science, Technology and Medicine, MRC Clinical Sciences Centre, Hammersmith Hospital, Du Cane Road, London WC12 0NN, UK

[4]Corresponding author

The locus control region (LCR) of the human CD2 gene (hCD2) confers T cell-specific, copy-dependent and position-independent gene expression in transgenic mice. This LCR consists of a strong T cell-specific enhancer and an element without enhancer activity (designated *HSS3*), which is required for prevention of position effect variegation (PEV) in transgenic mice. Here, we identified the HMG box containing protein-1 (HBP1) as a factor binding to *HSS3* of the hCD2 LCR. Within the LCR, HBP1 binds to a novel TTCATT-CATTCA sequence that is higher in affinity than other recently reported HBP1-binding sites. Mice transgenic for a hCD2 LCR construct carrying a deletion of the HBP1-binding sequences show a propensity for PEV if the transgene integrates in a heterochromatic region of the chromosome such as the centromere or telomere. We propose that HBP1 plays an important role in chromatin opening and remodelling activities by binding to and bending the DNA, thus allowing DNA–protein and/or protein–protein interactions, which increase the probability of establishing an active locus.
Keywords: chromatin/HBP1/HMG box/locus control region/position effect variegation

Introduction

Locus control regions (LCRs) direct expression of linked heterologous and homologous transgenes in a tissue-specific, position-independent, transgene copy number-dependent manner (Grosveld *et al*., 1987). Studies on the human CD2 (hCD2) and β-globin LCRs have demonstrated that the LCR achieves the position-independent expression in T cells and erythroid cells of transgenic mice, respectively, by overcoming heterochromatin-mediated position effect variegation (PEV; Festenstein *et al*., 1996; Milot *et al*., 1996a). The hCD2 LCR consists of at least two distinct elements: a transcriptional enhancer and a region without enhancer activity (designated *HSS3*) necessary for preventing PEV (Lake *et al*., 1990; Festenstein *et al*., 1996). The functional differences between the LCR and enhancer were evident from the early experiments on the human β-globin LCR using transgenic and cell transfection approaches. It has been shown that whereas the *HS3* and *HS4* regions of the β-globin LCR hardly increase the level of expression of the linked β-globin gene in cell transfection assays, they are able to render expression of the transgene copy number-dependent and site integration-independent (Talbot *et al*., 1990; Pruzina *et al*., 1991). Modular structures of LCRs have also been described for the human ADA (Aronow *et al*., 1995) and the T-cell receptor α (αTCR; Ortiz *et al*., 1997) gene loci. It appears, therefore, that LCRs consist of multiple elements such as enhancers and sequences necessary for establishing an open chromatin configuration, thereby preventing mosaic expression of transgenes when they are integrated into heterochromatic areas such as telomeres or the centromeres (Kioussis and Festenstein, 1997).

PEV, or the phenomenon of mosaic expression within a cell lineage of genes placed adjacent to a euchromatin–heterochromatin junction, has been well characterized in yeast and *Drosophila* (Allshire *et al*., 1994; Weiler and Wakimoto, 1995; Grunstein, 1997; Wallrath, 1998). Although PEV in mammals has not been studied in detail, certain characteristics of PEV described in mice bear a strong resemblance to those seen in other organisms. For instance, it appears that the decision whether to express the hCD2 transgene carrying the hCD2 LCR with *HSS3* deleted is a stochastic one and, once made, is maintained through subsequent cell divisions (Kioussis and Festenstein, 1997, 1998). Similar observations were made in experiments on human β-globin (Milot *et al*., 1996a) and β-lactoglobulin transgenic mice (Dobie *et al*., 1996). Mosaic expression patterns of transgenes have also been described in other systems (Pravtcheva *et al*., 1994; Robertson *et al*., 1995; Guy *et al*., 1996). Although most of the studies on PEV in mammals have been carried out using transgenic technology (Garrick *et al*., 1996; Dobie *et al*., 1997; Kioussis and Festenstein, 1997; Grosveld *et al*., 1998), a growing body of evidence has emerged implicating position effects in several human diseases caused by chromosomal translocations (Milot *et al*., 1996b; Kleinjan and van Heyningen, 1998). To date, experiments indicate that LCRs have the ability to overcome PEV effects. However, in spite of many exciting breakthroughs, the mechanism of action of the LCRs remains a subject of hot debate (Higgs, 1998) and very little is known about proteins that bind to LCRs and mediate the chromatin-opening and PEV-preventing function.

In this study, we identified HBP1 as a protein that binds to *HSS3* and contributes to the hCD2 LCR function of

preventing PEV. HBP1 belongs to the HMG family of proteins, with LEF1 being the closest homologue (Travis *et al.*, 1991). HBP1 was cloned originally as a protein able to complement a potassium channel defect in yeast (Lesage *et al.*, 1994). The function of this protein remains the subject of current investigation, and recently several groups have reported the ability of HBP1 to interact with proteins of the retinoblastoma family, to induce morphological transformation of cells in culture and to act as a transcriptional repressor of the cyclin D1, p21 and N-*myc* genes (Lavender *et al.*, 1997; Tevosian *et al.*, 1997; Gartel *et al.*, 1998; Shih *et al.*, 1998; Yee *et al.*, 1998). Here we show that HBP1 binds to the *HSS3* sequences within the hCD2 LCR and appears to contribute to the prevention of PEV. Our data suggest that the role of HBP1 in the cell cycle and differentiation may be mediated via chromatin remodelling activity and the establishment of open chromatin structures.

Results

Yeast one-hybrid assay identifies human HBP1 as binding to HSS3 of the hCD2 LCR

To identify *trans*-acting factors that bind to the *HSS3* region of the hCD2 LCR and, potentially, determine the LCR function, an *in vitro* DNase I footprint analysis of the *HSS3* region was performed using nuclear protein extracts. The DNase I footprinting of the 116 bp *Nde*I–*Bsr*GI fragment from *HSS3* resulted in identification of the FT1 and FT2 regions protected with both thymus and liver extracts (Figure 1). Analysis of the footprinted sequences revealed that the FT1 region contains a TTCA motif, which is repeated five times within the region and is also present in the FT2 region (Figure 1). To identify proteins binding to the FT1 and FT2 sequences, the Matchmaker (Clontech) Jurkat T-cell cDNA library was screened using the one-hybrid assay in yeast. Using the FT1 sequence as a bait, we isolated a library plasmid carrying a 1.3 kb insert homologous to the rat HBP1 cDNA (Lesage *et al.*, 1994). Sequence comparison has shown that the isolated 1.3 kb fragment consists of a 0.35 kb region encoding the C-terminal part of the protein including the HMG box DNA-binding domain plus 0.95 kb of the 3′-untranslated region (3′-UTR). A full-length open reading frame (ORF) cDNA of the human HBP1 (Figure 2A) was subsequently isolated from the Jurkat λ-ZAP cDNA library (Stratagene) screened with the 1.3 kb hHBP1 cDNA probe isolated in the yeast one-hybrid screen. Sequence comparison of the isolated ORF of the human HBP1 protein with that of the rat HBP1 revealed 91% homology in the region coding for the HMG box domain and 89% in the rest of the sequence. At the protein level, homology was 100% in the HMG domain and 91% in the N-terminal part of the protein.

hHBP1 HMG box binds to a novel TTCATTCATTCA motif within the hCD2 LCR

To identify the precise HBP1-binding site within the LCR, a recombinant HMG box domain (residue 433–513) protein was overexpressed in *Escherichia coli* using the pET-22b expression system (Novagen). The recombinant protein was used to footprint a 116 bp LCR fragment containing the FT1 and FT2 protected regions identified
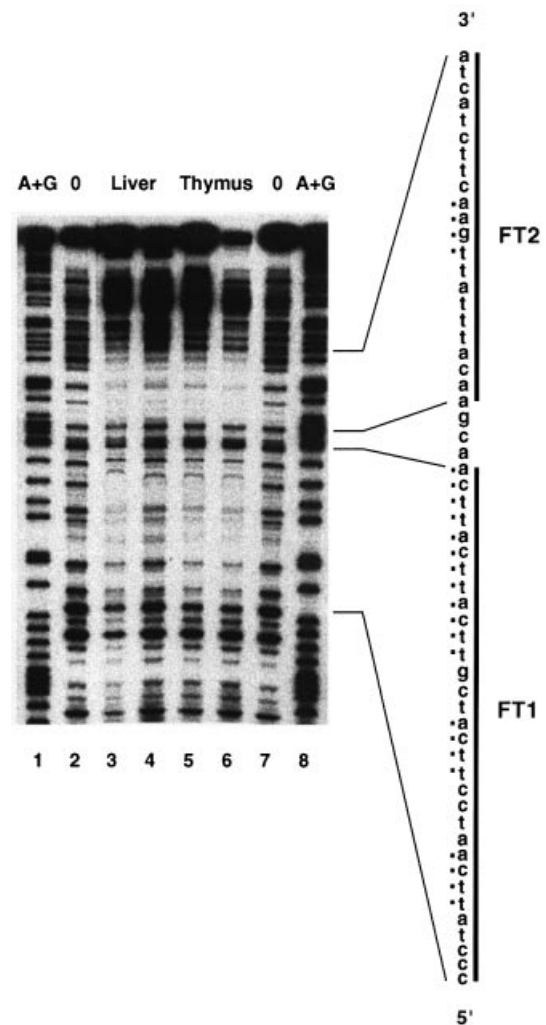

Fig. 1. *In vitro* DNase I footprint assay. A 116 bp *Nde*I–*Bsr*GI fragment (encompassing residues 1382–1497 of the 2 kb *Hin*dIII–*Hin*dIII LCR) labelled at the 5′ *Nde*I end was used in an *in vitro* DNase I footprint assay. Lanes 1 and 8, A + G DNA ladder; lanes 2 and 7, DNA control without extract; lanes 3 and 4, DNA incubated with 30 µg of liver nuclear protein extracts per lane in the presence of 1 and 2 µg of DNase I, respectively; lane 5 and 6, using thymus extract. Two protected areas marked by the solid lines designated FT1 and FT2 are clearly visible in lanes containing the thymus and liver extracts, and the boundaries of the footprints are 1402–1434 and 1438–1459 [with the poly(A) signal coinciding with the *Hin*dIII site being zero], respectively. The FT1 is more prominent with thymus extract than with liver. This is consistent with the band shift experiments (not shown) revealing that the FT1-binding protein(s) are relatively more abundant in the thymus than in the liver nuclear extracts.

with thymus and liver nuclear extracts. The FT1 region contains two HBP1-binding sites with apparently different affinities (Figure 2B). Thus, the ATCGTTCATTCATT-CAACG (FT1.S) region was visible using low concentrations of the recombinant box (25 ng), whereas protection of the other region, TACACCCTATTCAATCCTT (FT1.W), became apparent with higher concentrations (100 ng). To fine map and determine the boundaries of the HBP1-binding sites, 15 double-stranded overlapping oligonucleotides, 16 bp each, spanning a 48 bp region containing the FT1 footprint, were used in gel retardation assays (Figure 3). Such analysis showed that three oligonucleotides ATCGTTCATTCATTCA (FT1.J), CGTTCATTC
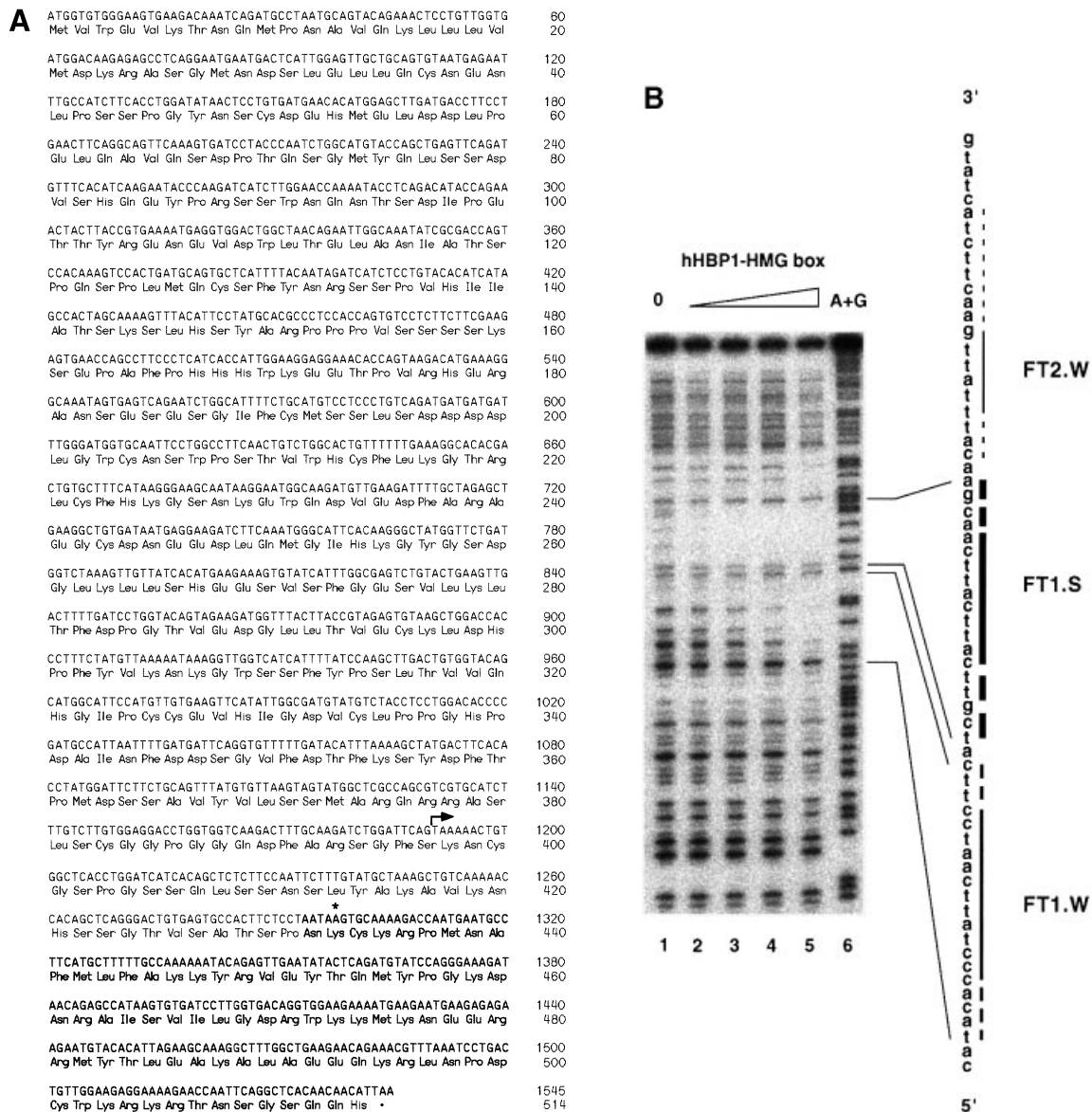
A

```
ATGGTGTGGGAAGTGAAGACAAATCAGATGCCTAATGCAGTACAGAAACTCCTGTTGGTG        60
Met Val Trp Glu Val Lys Thr Asn Gln Met Pro Asn Ala Val Gln Lys Leu Leu Leu Val      20

ATGGACAAGAGAGCCTCAGGAATGAATGACTCATTGGAGTTGCTGCAGTGTAATGAGAAT       120
Met Asp Lys Arg Ala Ser Gly Met Asn Asp Ser Leu Glu Leu Leu Gln Cys Asn Glu Asn      40

TTGCCATCTTCACCTGGATATAACTCCTGTGATGAACACATGGAGCTTGATGACCTTCCT       180
Leu Pro Ser Ser Pro Gly Tyr Asn Ser Cys Asp Glu His Met Glu Leu Asp Asp Leu Pro      60

GAACTTCAGGCAGTTCAAAGTGATCCTACCCAATCTGGCATGTACCAGCTGAGTTCAGAT       240
Glu Leu Gln Ala Val Gln Ser Asp Pro Thr Gln Ser Gly Met Tyr Gln Leu Ser Ser Asp      80

GTTTCACATCAAGAATACCCAAGATCATCTTGGAACCAAAATACCTCAGACATACCAGAA       300
Val Ser His Gln Glu Tyr Pro Arg Ser Ser Trp Asn Gln Asn Thr Ser Asp Ile Pro Glu     100

ACTACTTACCGTGAAAATGAGGTGGACTGGCTAACAGAATTGGCAAATATCGCGACCAGT       360
Thr Thr Tyr Arg Glu Asn Glu Val Asp Trp Leu Thr Glu Leu Ala Asn Ile Ala Thr Ser     120

CCACAAAGTCCACTGATGCAGTGCTCATTTTACAATAGATCATCTCCTGTACACATCATA       420
Pro Gln Ser Pro Leu Met Gln Cys Ser Phe Tyr Asn Arg Ser Ser Pro Val His Ile Ile     140

GCCACTAGCAAAAGTTTACATTCCTATGCACGCCCTCCACCAGTGTCCTCTTCTTCGAAG       480
Ala Thr Ser Lys Ser Leu His Ser Tyr Ala Arg Pro Pro Pro Val Ser Ser Ser Ser Lys     160

AGTGAACCAGCCTTCCCTCATCACCATTGGAAGGAGGAAACACCAGTAAGACATGAAAGG       540
Ser Glu Pro Ala Phe Pro His His His Trp Lys Glu Glu Thr Pro Val Arg His Glu Arg     180

GCAAATAGTGAGTCAGAATCTGGCATTTTCTGCATGTCCTCCCTGTCAGATGATGATGAT       600
Ala Asn Ser Glu Ser Glu Ser Gly Ile Phe Cys Met Ser Ser Leu Ser Asp Asp Asp Asp     200

TTGGGATGGTGCAATTCCTGGCCTTCAACTGTCTGGCACTGTTTTTTGAAAGGCACACGA       660
Leu Gly Trp Cys Asn Ser Trp Pro Ser Thr Val Trp His Cys Phe Glu Lys Gly Thr Arg     220

CTGTGCTTTCATAAGGGAAGCAATAAGGAATGGCAAGATGTTGAAGATTTTTGCTAGAGCT       720
Leu Cys Phe His Lys Gly Ser Asn Lys Glu Trp Gln Asp Val Glu Asp Phe Ala Arg Ala     240

GAAGGCTGTGATAATGAGGAAGATCTTCAAATGGGCATTCACAAGGGCTATGGTTCTGAT       780
Glu Gly Cys Asp Asn Glu Glu Asp Leu Gln Met Gly Ile His Lys Gly Tyr Gly Ser Asp     260

GGTCTAAAGTTGTTATCACATGAAGAAAGTGTATCATTTGGCGAGTCTGTACTGAAGTTG       840
Gly Leu Lys Leu Leu Ser His Glu Glu Ser Val Ser Phe Gly Glu Ser Val Leu Lys Leu     280

ACTTTTGATCCTGGTACAGTAGAAGATGGTTTACTTACCGTAGAGTGTAAGCTGGACCAC       900
Thr Phe Asp Pro Gly Thr Val Glu Asp Gly Leu Leu Thr Val Glu Cys Lys Leu Asp His     300

CCTTTCTATGTTAAAAATAAAGGTTGGTCATCATTTTATCCAAGCTTGACTGTGGTACAG       960
Pro Phe Tyr Val Lys Asn Lys Gly Trp Ser Ser Phe Tyr Pro Ser Leu Thr Val Val Gln     320

CATGGCATTCCATGTGTGTGAAGTTCATATTGGCGATGTATGTCTACCTCCTGGACACCCC      1020
His Gly Ile Pro Cys Cys Glu Val His Ile Gly Asp Val Cys Leu Pro Pro Gly His Pro     340

GATGCCATTAATTTTGATGATTCAGGTGTTTTTGATACATTTAAAAGCTATGACTTCACA      1080
Asp Ala Ile Asn Phe Asp Asp Ser Gly Val Phe Asp Thr Phe Lys Ser Tyr Asp Phe Thr     360

CCTATGGATTCTTCTGCAGTTTATGTGTTAAGTAGTATGGCTCGCCAGCGTCGTGCATCT      1140
Pro Met Asp Ser Ser Ala Val Tyr Val Leu Ser Ser Met Ala Arg Gln Arg Arg Ala Ser     380

TTGTCTTGTGGAGGACCTGGTGGTCAAGACTTTGCAAGATCTGGATTCAGTAAAAACTGT      1200
Leu Ser Cys Gly Gly Pro Gly Gly Gln Asp Phe Ala Arg Ser Gly Phe Ser Lys Asn Cys     400

GGCTCACCTGGATCATCACAGCTCTCTTCCAATTCTTTGTATGCTAAAGCTGTCAAAAAC      1260
Gly Ser Pro Gly Ser Ser Gln Leu Ser Ser Asn Ser Leu Tyr Ala Lys Ala Val Lys Asn     420

CACAGCTCAGGGACTGTGAGTGCCACTTCTCCTAATAAGTGCAAAAGACCAATGAATGCC      1320
His Ser Ser Gly Thr Val Ser Ala Thr Ser Pro **Asn Lys Cys Lys Arg Pro Met Asn Ala**   440

TTCATGCTTTTTGCCAAAAAATACAGAGTTGAATATACTCAGATGTATCCAGGGAAAGAT      1380
**Phe Met Leu Phe Ala Lys Lys Tyr Arg Val Glu Tyr Thr Gln Met Tyr Pro Gly Lys Asp**   460

AACAGAGCCATAAGTGTGATCCTTGGTGACAGGTGGAAGAAAATGAAGAATGAAGAGAGA      1440
**Asn Arg Ala Ile Ser Val Ile Leu Gly Asp Arg Trp Lys Lys Met Lys Asn Glu Glu Arg**   480

AGAATGTACACATTAGAAGCAAAGGCTTTGGCTGAAGAACAGAAACGTTTAAATCCTGAC      1500
**Arg Met Tyr Thr Leu Glu Ala Lys Ala Leu Ala Glu Glu Gln Lys Arg Leu Asn Pro Asp**   500

TGTTGGAAGAGGAAAAGAACCAATTCAGGCTCACAACAACATTAA                   1545
**Cys Trp Lys Arg Lys Arg Thr Asn Ser Gly Ser Gln Gln His  ·**                        514
```

B

Fig. 2. hHBP1 and footprint assay using the recombinant protein. (**A**) DNA and amino acid sequences. The figure shows the 1545 bp sequence of the isolated cDNA coding for the HBP1 protein of 514 amino acids. The beginning of the 345 bp fragment identified by one-hybrid yeast assay coding for the C-terminal part of the protein including the HMG box is marked by an arrow. Sequences corresponding to the HMG domain are highlighted in bold. The asterisk marks the first amino acid of the recombinant HMG box used in the footprint and gel retardation assays below. (**B**) Footprint with the HBP1 HMG box. A 116 bp *Nde*I–*Bsr*GI fragment (Figure 1) labelled at the 5′ *Nde*I end was used in an *in vitro* DNase I footprint assay. Lane 1, DNA control without recombinant HMG box; lanes 2–5, DNA incubated with 25, 50, 100 and 200 ng of the recombinant box, respectively; lane 6, A + G DNA ladder. Two protected areas, FT1.S and FT1.W, are clearly visible within the FT1 region. The protected sequences are marked by the solid lines, the thick and thin lines representing the stronger (FT1.S) and weaker (FT1.W) protected areas, respectively. At the highest amount of 200 ng, a weakly protected area (FT2.W) becomes visible within the FT2 region. The dashed lines correspond to potentially protected nucleotides, as these areas were not digested by DNase I even in lane 1, without the recombinant protein.

ATTCAAC (FT1.K) and TTCATTCATTCAACGA (FT1.L), sharing a common TTCATTCATTCA motif, bind the recombinant HBP1 HMG box better than the other oligonucleotides. The TTCATTCATTCA motif is located within the FT1.S region corresponding to the higher affinity footprint. Thus, it appears that the TTCATTCATTCA sequence within the FT1 region of the hCD2 LCR contains the minimum binding consensus motif for the hHBP1 HMG domain.

The TTCATTCATTCA motif is a single site. The TTCATTCATTCA motif identified is partly homologous to the (T/A)(T/A)CAA(A/T)G sequence which corresponds to the binding consensus for several of the HMG proteins including LEF1, TCF1, ROX1, SOX4, SRY and STE11 (Landsman and Bustin, 1993; Oosterwegel *et al.*, 1993; Grosschedl *et al.*, 1994). Because of this resemblance, we speculated that the identified TTCATTCATTCA motif may, in fact, consist of two overlapping TTCATTCA binding sites. To determine whether the TTCATTCATTCA site contains redundant binding elements, we used three 16 bp oligonucleotides designated FT1.J.mC, FT1.J.mD and FT1.J.mE (Figure 4), corresponding to mutants of FT1.J in which one of the TTCA repeats was mutated into GGCA (the mutated nucleotides are underlined). We reasoned that if the TTCATTCA motif was a minimal
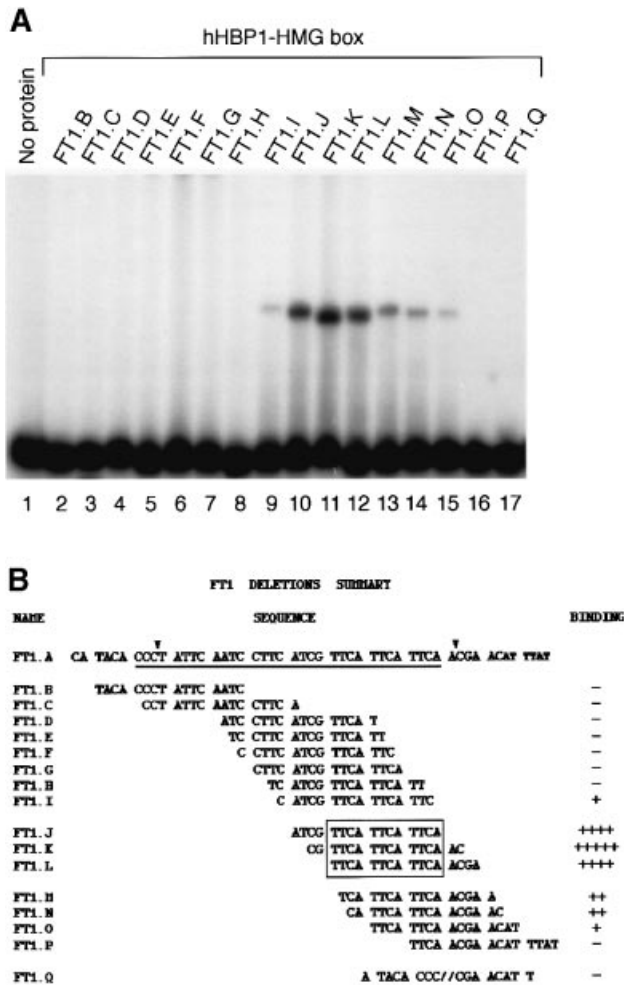
A

hHBP1-HMG box

B

FT1 DELETIONS SUMMARY

NAME	SEQUENCE	BINDING
FT1.A	CA TACA CCCT ATTC AATC CTTC ATCG TTCA TTCA TTCA ACGA ACAT TTAT	
FT1.B	TACA CCCT ATTC AATC	−
FT1.C	CCT ATTC AATC CTTC A	−
FT1.D	ATC CTTC ATCG TTCA T	−
FT1.E	TC CTTC ATCG TTCA TT	−
FT1.F	C CTTC ATCG TTCA TTC	−
FT1.G	CTTC ATCG TTCA TTCA	−
FT1.H	TC ATCG TTCA TTCA TT	+
FT1.I	C ATCG TTCA TTCA TTC	+
FT1.J	ATCG TTCA TTCA TTCA	++++
FT1.K	CG TTCA TTCA TTCA AC	+++++
FT1.L	TTCA TTCA TTCA ACGA	++++
FT1.M	TCA TTCA TTCA ACGA A	++
FT1.N	CA TTCA TTCA ACGA AC	++
FT1.O	TTCA TTCA ACGA ACAT	+
FT1.P	TTCA ACGA ACAT TTAT	−
FT1.Q	A TACA CCC//CGA ACAT T	−

Fig. 3. Fine mapping of the HMG box-binding motif. (**A**) Gel retardation assay using the recombinant hHBP1 HMG domain and overlapping FT.1B–FT1.Q double-stranded oligos covering the FT1 region. The recombinant protein was incubated with the 5′ end-labelled oligonucleotides in the presence of excess poly(dI–dC). The intensity of the bands was assessed using a PhosphorImager. Retarded complexes are visible using the FT1.I–FT.O oligos, with maximum intensity corresponding to the FT1.J–FT.1L probes. (**B**) Deletion summary diagram representing the alignment of the FT1.A sequence and the FT.1B–FT1.Q overlapping oligos covering the FT1.A area (the sense strand sequences are shown). The FT1.A sequence (underlined) comprises nucleotides from 1396 to 1446 of the hCD2 LCR containing the 32 bp FT1 footprint (Figure 1) identified with the thymus and liver nuclear extract. The TTCATTCATTCA motif common for the three oligos FT1.J–FT1.L, which bind the HMG box better than the others, is highlighted by a rectangle. The two arrows mark the 30 bp region deleted from the hCD2 LCR in subsequent experiments with transgenic mice. The FT1.Q oligo represents the site generated by the junction of the two 8 bp regions on either side of the deleted 30 bp sequence. In a gel retardation experiment, no binding by the HMG box was observed. The region within FT1.A used as a bait to clone hHBP1 in the one-hybrid assay is highlighted by larger letters.

hHBP1 box-binding site then the GGCATTCATTCA (FT1.J.mC) and TTCATTCAGGCA (FT1.J.mE) mutants would still be able to bind to the HMG box through the remaining intact TTCATTCA motif. If, however, TTCATTCATTCA was a single binding site, then these mutant oligonucleotides would not be able to bind the protein. Similar mutations had been reported to abolish the TCF1- and SOX4-binding sites (Oosterwegel *et al.*, 1991; van de Wetering *et al.*, 1993). The gel retardation

A

hHBP1-HMG box

B

MUTANTS SUMMARY

NAME	SEQUENCE	BINDING
LEF1/TCF1	GAGC TTCA AAGG GTGC	−
SOX4	TGAG AACA AAGC GCTC	−
SRY	ACAG AACA ATGG GTGC	−
MYC1	TGCT GAGA ATGG GAAG	−
MYC2	GACC TTCA ATGG GGGG	+
MYC3	AGTG CAAA ATGG GAGG	−
FT1.J	ATCG TTCA TTCA TTCA	++++
FT1.mA	ATCG TTGA TTGA TTGA	−
FT1.mB	ATCG GGCA GGCA GGCA	−
FT1.mC	ATCG GGCA TTCA TTCA	−
FT1.mD	ATCG TTCA GGCA TTCA	−
FT1.mE	ATCG TTCA TTCA GGCA	+/−
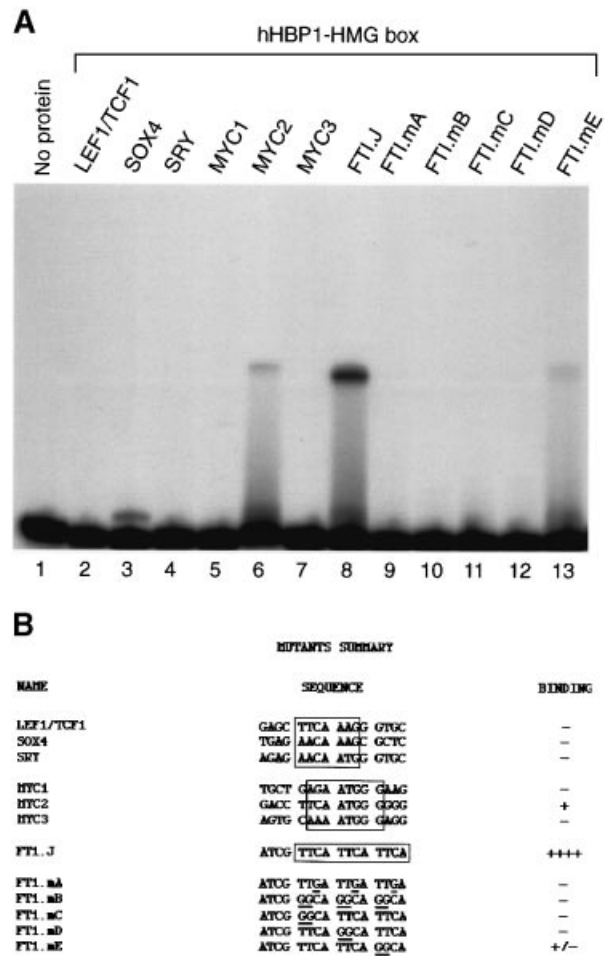
Fig. 4. Band shift assay with mutant oligos. (**A**) Gel retardation assay using the recombinant hHBP1 HMG domain and various mutants of the FT1.J oligo. Lane 1, FT1.J oligo with no protein; lanes 2–13 correspond to the following oligos incubated with the HMG box: lanes 2–4, LEF1/TCF1, SOX4 and SRY, respectively; lanes 5–7, MYC1, MYC2 and MYC3, respectively; lane 8, FT1.J; lanes 9–13, FT1.mA–FT1.mE mutants, respectively. Formation of the retarded complexes was observed for the FT1.J, MYC2 and FT1.mE oligos, with the strongest signal corresponding to FT1.J. (**B**) The mutant summary diagram represents the alignment of the FT1.J sequence with the FT1.mA–FT1.mE mutants, and with the LEF1/TCF1, SOX4, SRY and MYC1–MYC3 oligos. The consensus motifs are highlighted by the rectangles.

experiment showed that the oligonucleotides corresponding to the above mutants indeed were unable to bind the hHBP1 HMG box. Thus, TTCATTCATTCA appears to represent a single binding site (Figure 4).

The TTCATTCATTCA motif has a higher affinity than other HBP1-binding sites reported. While this work was in progress, it was reported that the mouse HBP1 HMG domain is able to bind the AGAATGGG, TCAATGGG and AAAATGGG motifs located within the N-*myc* promoter. Three TCAATGGG motifs were also reported within the cytomegalovirus (CMV) promoter and thought to be implicated in promoter silencing induced by the mHBP1 protein (Tevosian *et al.*, 1997; Yee *et al.*, 1998). Because the TTCATTCATTCA motif identified differed from the sequences reported in the N-*myc* promoter, it was decided to compare the HBP1 HMG box-binding ability with the sequences found in the LCR and N-*myc*1
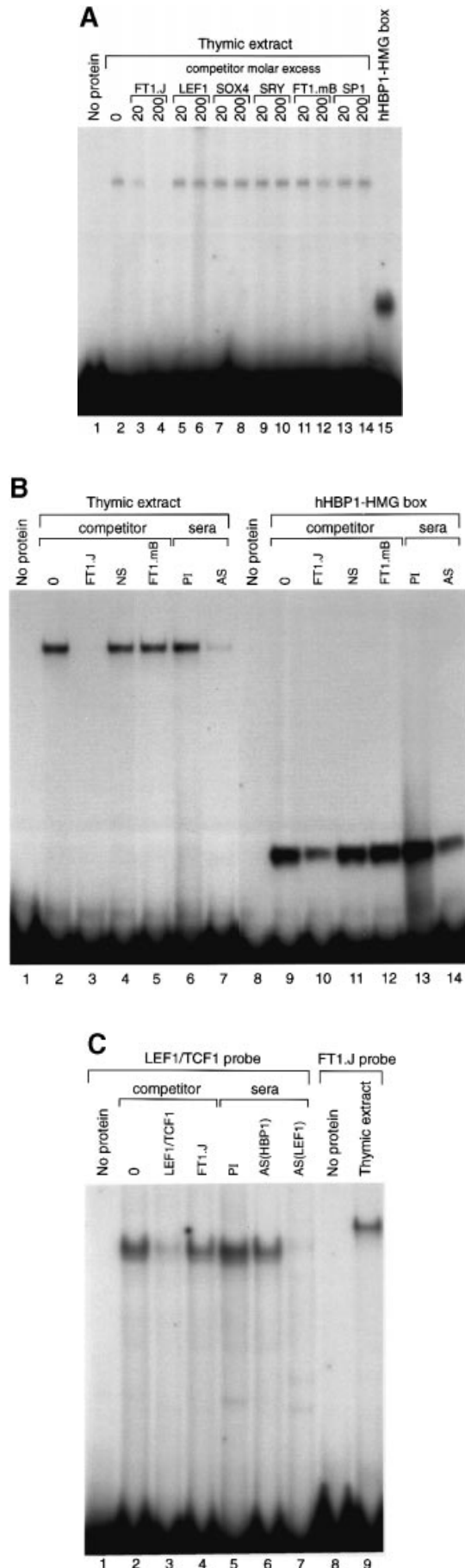
promoter. A band shift experiment (Figure 4) showed that the ATCGTTCATTCATTCA (FT1.J) oligo binds the recombinant HBP1 HMG box better than the TGCTGAGAATGGGAAG (MYC1), GACCTTCAATGGG-GGG (MYC2) and AGTGCAAAATGGGAGG (MYC3) oligos corresponding to the 466–581, 492–507 and 317–332 sequences within the N-*myc* promoter (as in DDBJ/EMBL/GenBank accession No. X632811; the core motifs are highlighted by larger letters). Thus it appears that TTCATTCATTCA is the highest affinity HBP1-binding motif described to date.

The hHBP1 HMG box binding is sequence specific. To assess the DNA-binding specificity of hHBP1, we tested whether the recombinant HMG box would recognize the binding sites of other HMG box-containing proteins. The hHBP1 HMG domain does not bind to the oligonucleotides containing consensus binding sequences of the LEF1/TCF1, SOX4 and SRY proteins (Figure 4) (Giese *et al.*, 1991; van de Wetering *et al.*, 1991, 1993). Thus, it appears that the recombinant hHBP1 HMG domain binds to the novel TTCATTCATTCA motif of the LCR in a sequence-specific manner.

Recombinant HBP1 HMG domain has DNA-binding characteristics identical to the binding activity found within a T-cell nuclear extract

To demonstrate that the protein present in the thymus nuclear extract which binds to the FT1 sequences is HBP1, a band shift experiment was performed using the thymus nuclear extract with the FT1.J oligonucleotide containing the hHBP1 HMG box-binding motif, TTCATTCATTCA. The experiment revealed a single retarded band that could not be competed by either LEF1/TCF1- or SOX4-binding sites or other non-specific oligonucleotides (Figure 5A), demonstrating that the observed binding is sequence

Fig. 5. Recombinant HBP1 HMG box has DNA-binding characteristics identical to the binding activity found in the T-cell extract. (**A**) Gel retardation assays using the thymus nuclear extract, the FT1.J probe and various cold oligonucleotide competitors. The nuclear extract was incubated with 5′ end-labelled FT1.J probe in the presence of poly(dI–dC) and the cold competitors where appropriate. The complexes and the free probe were separated on a non-denaturing polyacrylamide gel and visualized using autoradiography. All lanes contain the FT1.J probe with: lane 1, no extract; lane 2, extract, but no cold competitor; lanes 3–14, extract plus cold competitors as follows: lanes 3 and 4, FT1.J; lanes 5 and 6, LEF1/TCF1; lanes 7 and 8, SOX4; lanes 9 and 10, SRY; lanes 11 and 12, FT1.mB; lanes 13 and 14, SP1 (in 20 and 200 M excess, respectively); lane 15, probe incubated with the recombinant HBP1 HMG box. The sequences of oligonucleotides are shown in Figure 4B. (**B**) Band shift experiments with thymic nuclear extract (lanes 2–7) and recombinant HMG box (lanes 9–14) with either oligonucleotide competitors (lanes 3–6 and 10–12) or with rabbit antisera raised against the recombinant hHBP1 HMG box. All lanes contain the FT1.J probe. Lanes 1 and 8, no extract/protein; lanes 2 and 9, without oligonucleotide competitors; lanes 3 and 5 and 10–12 contained a 200 M excess of the following cold competitors: lanes 3 and 10, FT1.J; lanes 4 and 11, non-specific; lanes 6 and 12, FT1.mB. Lanes 6 and 13 contain pre-immune sera and lanes 7 and 14 contain antisera raised against the HBP1 HMG box. (**C**) Gel retardation assays using the LEF1/TCF1 probe: lane 1, no thymic extract; lane 2, extract without oligonucleotide competitors; lanes 3 and 4, extract plus a 200 M excess of the LEF1/TCF1 or FT1.J cold competitor probes, respectively; lanes 5–7, extract plus pre-immune sera, HBP1 antisera and LEF1 antisera, respectively. In lanes 8 and 9, FT1.J probe was used with and without extract, respectively.
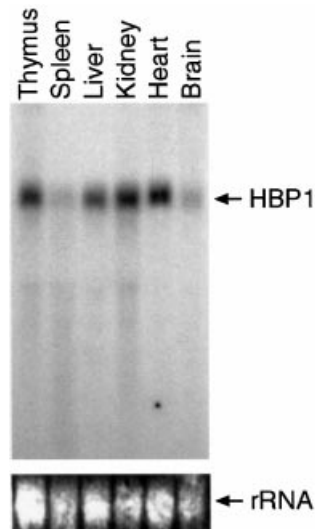
Fig. 6. HBP1 mRNA present as a single species. RNA for the Northern blot was isolated from adult mouse tissues and hybridized with the 3′-UTR of the mHBP1 cDNA probe. The lower panel shows ethidium bromide staining of the gel for 28S rRNA as a loading control.

specific and is not caused by LEF1, TCF1 or SOX4 proteins. The fact that one retarded band is detected in the band shift experiments indicates that there is one species of binding protein in the thymus extract. This is supported by Northern blot analysis with mouse HBP1 3′-UTR cDNA, which showed that HBP1 mRNA is expressed as a single transcript of 2.5 kb in various tissues (Figure 6).

Furthermore, Figure 5B shows that the binding activity in the thymus nuclear extract could be blocked by a polyclonal antibody raised against the HMG box. These experiments demonstrate that the recombinant HBP1 HMG domain and the binding activity observed in a thymocyte nuclear extract have identical DNA-binding characteristics. This strongly suggests that the activity detected in the nuclear protein extract is due to HBP1.

The conclusion is supported by additional experiments demonstrating a high specificity of the HBP1 HMG box antibody. It was found in band shift experiments (Figure 5C) with an oligonucleotide probe containing the LEF1/TCF1-binding site and thymus nuclear protein extract that the HBP1 HMG box antibody did not affect the LEF1/TCF1-specific binding, whereas a control LEF1 antibody blocked the binding. As the HMG box domains of LEF1/TCF1 (Oosterwegel *et al.*, 1991; Travis *et al.*, 1991) are thought to be amongst the highest in homology to that of HBP1 among the known HMG proteins, this result strongly indicated a low cross-reactivity of the HBP1 HMG box antibody with other HMG box proteins. Figure 5C shows that the complex formed using the LEF1/TCF1-specific oligonucleotide probe is qualitatively different from the one formed using the FT1 oligonucleotide (of the same size), thus providing further evidence that the FT1-binding activity in thymic extract is unlikely to be due to binding of other HMG box proteins, such as LEF1/TCF1. The experiments have also shown that the LEF1/TCF1-specific binding could not be competed by an FT1 oligonucleotide, suggesting inefficient binding

of the LEF1/TCF1 proteins to the HBP1-binding motif identified within FT1.

Deletion of the FT1 region from the LCR results in PEV

It has been shown previously in our laboratory that deleting the *HSS3* region from the LCR results in PEV in transgenic mice carrying a ΔHSS3-LCR hCD2 minigene (Festenstein *et al.*, 1996). To determine *in vivo* the contribution of the FT1 region and, potentially, the role of P1 in the regulation of LCR function, a 30 bp sequence including the FT1 footprint was deleted from the LCR. The pattern of expression of the hCD2 minigene with such an LCR (designated ΔFT1-LCR) was analysed in transgenic mice (Figure 7A). Seven transgenic mouse lines, further referred to as FT1.1–7, were generated and the expression of the transgenic hCD2 on thymocytes was analysed by flow cytometry, which allows measurement of the hCD2 protein in individual cells. The pattern of transgene expression in the FT1.1–7 transgenic lines was compared with that in the Mg4 line carrying a hCD2 minigene integrated in the centromere. The hCD2 minigene in the Mg4 line is integrated at the centromere but contains the full LCR and, therefore, expression of the transgene does not variegate (Festenstein *et al.*, 1996). To quantitate the degree of variegation, we estimated the percentage of thymocytes falling outside the $1.3\times$ SD (standard deviation) interval measured from the mean of hCD2-positive thymocytes. Such calculations for the Mg4 line resulted in an average K_{PEV} of 9%, indicating that the transgenic hCD2 minigene is expressed unimodally on virtually all thymocytes, so that the peak's shape is close to the normal distribution, which corresponds to $K_{PEV} = 10\%$. Analysis of the ΔFT1 transgene expression on thymocytes has shown that in the FT1.1 transgenic line, the K_{PEV} value was 29%, indicating that 20% more cells were found outside the normal distribution peak than in the non-variegating Mg4 line. In addition, the FT1.2 line also had noticeably more hCD2-negative cells ($K_{PEV} = 16\%$) than the Mg4 background control. Fluorescence *in situ* hybridization (FISH) experiments showed that the ΔFT1-LCR hCD2 transgene in the FT1.1 and FT1.2 lines was integrated in the centromere and centromere border, respectively (Figure 7B).

Interestingly, the FT1.7 line gave rise to litters that contained littermates with two different transgene expression pattern phenotypes within the same litter. One group of littermates, named FT1.7B, had noticeably more hCD2-negative thymocytes ($K_{PEV} = 28\%$), whereas the other group, FT1.7A ($K_{PEV} = 13\%$), had very few hCD2-negative thymocytes. The frequency of the two phenotypes is roughly equivalent, with half the transgenic littermates showing the variegating pattern. That littermates in these two groups do belong to the same line was confirmed by Southern hybridization (not shown). FISH analysis revealed that the transgene in the FT1.7 line is located close to the telomere (Figure 7B). As mammalian telomere-induced PEV has not been described in detail, it is not clear whether the observed variable phenotype is due to the telomeric location of transgene.

The rest of the transgenic lines, FT1.3–6, expressed the transgenic hCD2 on thymocytes in a fairly unimodal manner, resulting in an average K_{PEV} within the 8–13%
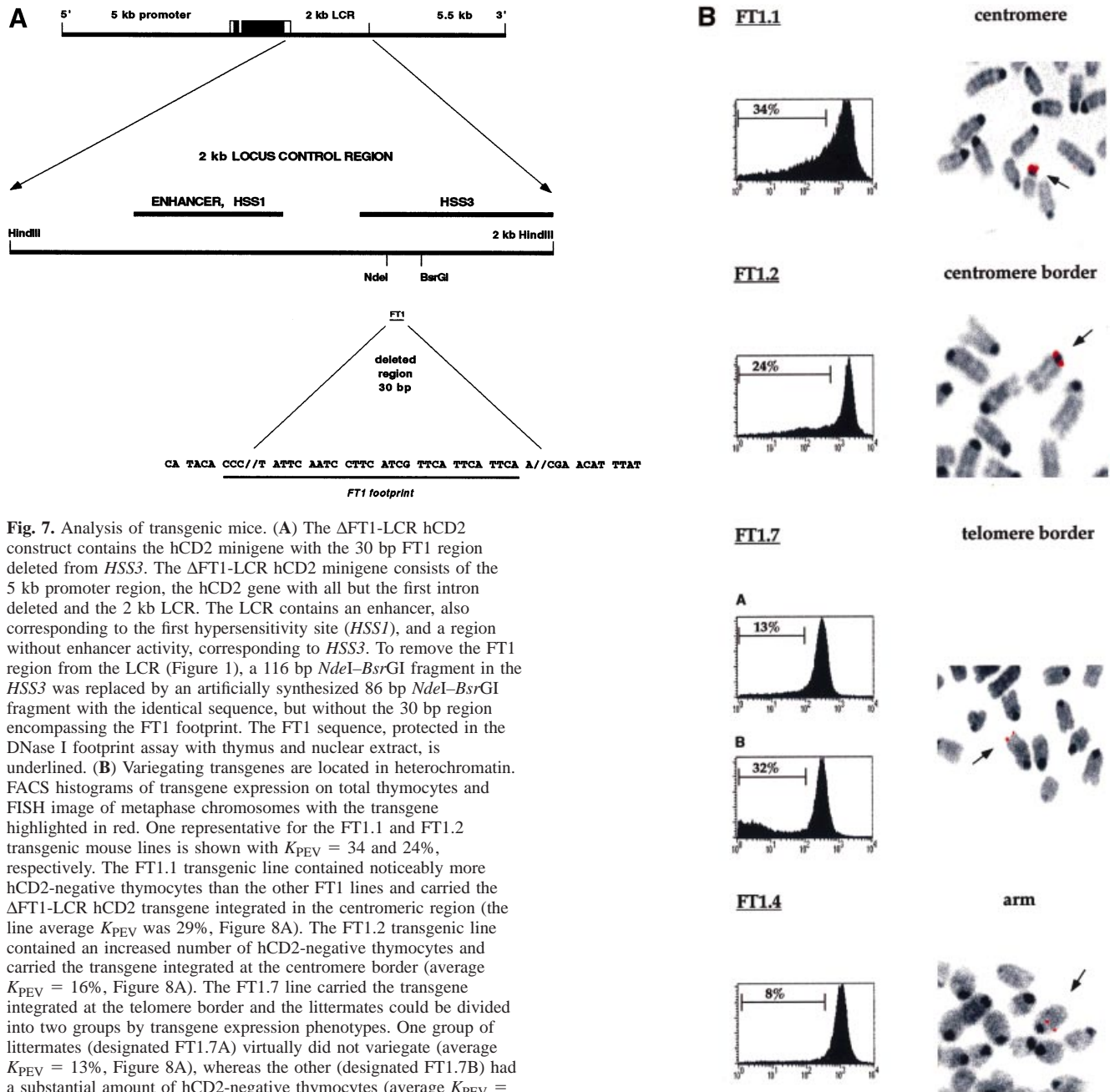
Fig. 7. Analysis of transgenic mice. (**A**) The ΔFT1-LCR hCD2 construct contains the hCD2 minigene with the 30 bp FT1 region deleted from *HSS3*. The ΔFT1-LCR hCD2 minigene consists of the 5 kb promoter region, the hCD2 gene with all but the first intron deleted and the 2 kb LCR. The LCR contains an enhancer, also corresponding to the first hypersensitivity site (*HSS1*), and a region without enhancer activity, corresponding to *HSS3*. To remove the FT1 region from the LCR (Figure 1), a 116 bp *Nde*I–*Bsr*GI fragment in the *HSS3* was replaced by an artificially synthesized 86 bp *Nde*I–*Bsr*GI fragment with the identical sequence, but without the 30 bp region encompassing the FT1 footprint. The FT1 sequence, protected in the DNase I footprint assay with thymus and nuclear extract, is underlined. (**B**) Variegating transgenes are located in heterochromatin. FACS histograms of transgene expression on total thymocytes and FISH image of metaphase chromosomes with the transgene highlighted in red. One representative for the FT1.1 and FT1.2 transgenic mouse lines is shown with K_{PEV} = 34 and 24%, respectively. The FT1.1 transgenic line contained noticeably more hCD2-negative thymocytes than the other FT1 lines and carried the ΔFT1-LCR hCD2 transgene integrated in the centromeric region (the line average K_{PEV} was 29%, Figure 8A). The FT1.2 transgenic line contained an increased number of hCD2-negative thymocytes and carried the transgene integrated at the centromere border (average K_{PEV} = 16%, Figure 8A). The FT1.7 line carried the transgene integrated at the telomere border and the littermates could be divided into two groups by transgene expression phenotypes. One group of littermates (designated FT1.7A) virtually did not variegate (average K_{PEV} = 13%, Figure 8A), whereas the other (designated FT1.7B) had a substantial amount of hCD2-negative thymocytes (average K_{PEV} = 28%, Figure 8A). Representatives of FT1.7A and FT1.7B are shown with K_{PEV}s of 13 and 32%, respectively. The FT1.4 line expressed the ΔFT1-LCR hCD2 transgene on virtually all thymocytes with K_{PEV} = 8% (corresponding to the average K_{PEV}s in these lines, Figure 8A) and carried the transgene integrated in the long arm of the chromosome. As a control, we used the Mg4 line with an average K_{PEV} of 9% and carrying the hCD2 minigene with an intact LCR integrated in the centromere.

range. FISH analysis has shown that the ΔFT1-LCR hCD2 transgene in these lines was integrated in the long arm of the chromosomes (Figure 7B; data not shown). A summary of the degree of variegation of transgene expression for all seven transgenic ΔFT1-LCR hCD2 lines is presented in Figure 8A. We found that the extent of variegation varied particularly within FT1.1, FT1.2 and FT1.7 variegating transgenic lines, resulting in higher standard deviations. For example, among the 18 mice analysed for FT1.1, line variegation ranged from a K_{PEV} of 13 to 38%, with an

average K_{PEV} of 29%. Such a dispersion within the same line could be due to effects of genetic background on PEV, so that the same transgene variegates differently on the CBA and C57Bl/10 backgrounds (R.Festenstein and D.Kioussis, in preparation).

The bar chart in Figure 8A demonstrates that the FT1.1 line carrying a transgene integrated in the centromere has the highest K_{PEV}. The FT1.2 and FT1.7 mice carried the transgene integrated at the centromere and telomere borders, respectively, and had a relatively high number of
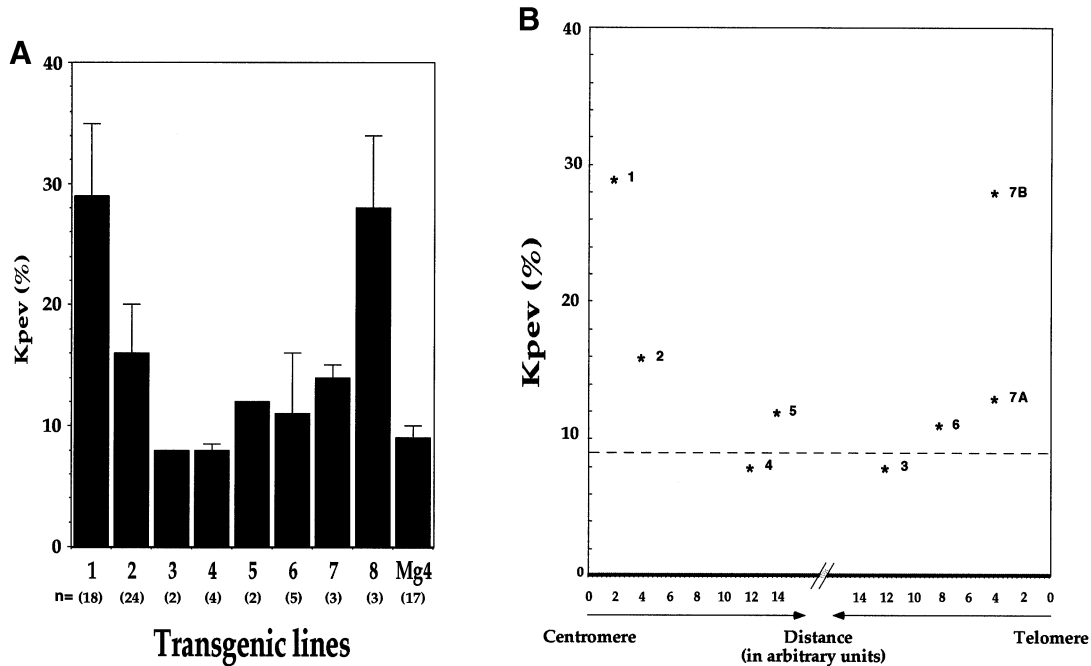
Fig. 8. Summary of the degree of variegation. (**A**) The histogram summarizes average K_{PEV} values for the FT1.1–7 and Mg4 transgenic lines. Each column represents the average K_{PEV} for a single line calculated as an arithmetic mean of the K_{PEV} of n mice analysed from the same line. The number of mice analysed is indicated under the column for each line (n). Columns 1–6 correspond to the FT1.1–6 lines, whereas columns 7 and 8 correspond to the FT1.7A and FT1.7B mice, respectively. The histogram demonstrates that the FT1.1 and FT1.7B mice have a significantly higher K_{PEV} than the others. The SD is represented by the bars above the respective columns; the representatives of the FT1.3 and FT1.5 lines had virtually identical K_{PEV}s and their SDs are not shown. Each of the FT1.1–7 lines carried 20, 22, 23, 8, 25, 6 and 5 copies of the transgene, respectively, with Mg4 carrying 7 copies (SDs were within a 20% range). No correlation between transgene copy number and degree of variegation was observed. (**B**) K_{PEV} depends on the distance to the heterochromatin. The graph shows an inverse correlation between K_{PEV} and the distance from the ΔFT1-LCR hCD2 transgene to a heterochromatic region such as the centromere or telomere. Each of the asterisks represents one of the FT1.1–7 transgenic lines. For each line, two mice were analysed and the arithmetic mean of distances measured from the transgene to the centromere or telomere on two daughter chromosomes was calculated. (FT1.7A and FT1.B mice were analysed as distinct lines and are represented by separate asterisks.) The distance was measured using the NIH Image Software (NIH) and is presented on the graph in arbitrary units. In all lines, the SD was within a 15% range. The broken line correspond to $K_{PEV} = 9\%$ of the non-variegating control Mg4 line.

CD2-negative thymocytes. The other lines had a K_{PEV} within the 8–13% range and carried the transgene integrated in the long arm of the chromosome. It appears, therefore, that deleting the FT1 sequence from the LCR renders the ΔFT1-LCR hCD2 minigene subject to PEV when the transgene is integrated into heterochromatic regions such as the centromere or telomere or their borders. Crude as the estimates are using FISH, our data also suggest an inverse correlation between the distance from the transgene to the telomere or centromere and the degree of PEV: the closer the transgene is placed to the centromeric or telomeric heterochromatin, the higher the degree of variegation (Figure 8B). Although it was reported that transgene silencing in mammals may be dependent on the number of copies of the transgene (Garrick *et al.*, 1998), in our study we did not see any correlation between transgene copy number and degree of variegation. Indeed, each of the FT1.1–7 lines carried 20, 22, 23, 8, 25, 6 and 5 copies of the transgene, respectively, with Mg4 carrying 7 copies.

Discussion

To understand the mechanism underlying LCR function, we aimed to identify proteins that interact with the hCD2 LCR. Using DNase I footprint analysis, the binding sites of proteins interacting with the LCR and, potentially, regulating its function were mapped. These footprints

were used as baits to screen a T cell-specific library in a one-hybrid assay that resulted in the isolation of the human HBP1. Screening of the one-hybrid and λ-ZAP T cell-specific cDNA libraries did not identify any HBP1 homologues, suggesting that HBP1 is a single gene. In addition, Northern blot analysis has identified only one mRNA band, supporting the hypothesis that the HBP1 protein probably exists in a single form.

Using yeast one-hybrid and various *in vitro* DNA-binding assays, we have shown that HBP1 interacts with the FT1 region of the LCR in a sequence-specific manner. The identified HBP1-binding motif TTCATTCATTCA is different from the canonical (T/A)(T/A)CAA(A/T)G consensus sequence of other HMG box proteins (Landsman and Bustin, 1993; Oosterwegel *et al.*, 1993; Grosschedl *et al.*, 1994) and represents the highest affinity among the HBP1-binding sites identified to date (Tevosian *et al.*, 1997; Yee *et al.*, 1998). Although the sites within the N-*myc* promoter were mapped originally using the mouse HBP1 HMG domain, it is reasonable to expect that both mouse and human HBP1 HMG domains recognize identical or highly similar sequences, since the human and mouse HMG boxes are identical at the amino acid level. Interestingly, the CCCATTCT (MYC1), CCCATTGA (MYC2) and CCCATTTT (MYC3) motifs within the N-*myc* promoter resemble the CCTATTCA motif within the weak hHBP1-binding site, FT1.W (Figure 2B). The presence of two HBP1 sites (FT1.S and

FT1.W) may account for the larger footprint caused by the nuclear extract as compared with the footprint mapped with the purified HBP1 HMG box. Alternatively, it is possible that the larger region of protection observed with nuclear extracts is due to a larger protein complex comprising HBP1 as the DNA-binding core.

The FT1 footprint is located within the *HSS3* region, which does not have any enhancer activity, but is necessary for the prevention of PEV in transgenic mice (Festenstein *et al*., 1996). Here, we show that the deletion of FT1 results in the variegation of transgene expression when the transgene is integrated close to heterochromatic areas, such as centromeres or telomeres. Our results show that the FT1 region is required for the LCR function and suggest that HBP1 plays a role in the prevention of PEV. Therefore, our data implicate a chromatin-opening activity and suggest the possibility that HBP1 interacts with the chromatin remodelling machinery, such as SWI–SNF and/ or deacetylase–acetyltransferase complexes. Involvement of DNA-binding proteins in chromatin remodelling processes has been demonstrated previously. For instance, it has been shown that the DNA-binding GAGA factor is able to act as a PEV modifier in the context of the NURF nucleosome remodelling complex, a member of the SWI–SNF family (Farkas *et al*., 1994; Tsukiyama *et al*., 1994, 1995a,b; Granok *et al*., 1995).

How HBP1 binding to the hCD2 LCR can mediate the prevention of PEV remains to be investigated. However, our results are consistent with other work on HBP1. Thus, it has been shown that HBP1 binds to RB family proteins (Lavender *et al*., 1997; Tevosian *et al*., 1997), and the latter are capable of interaction with the BRG1 and BRM components of the SWI–SNF complex as well as with the HDAC1 histone deacetylase (Dunaief *et al*., 1994; Strober *et al*., 1996; Brehm *et al*., 1998; Maghaghi-Jaulin *et al*., 1998). These data support our hypothesis that HBP1 could interact with the chromatin remodelling complexes by recruiting mediator molecules such as the RB family proteins. It is tempting to speculate that HBP1, which contains the same RB-binding LXCXE motif as HDAC1, is capable of competitively dissociating HDAC1 from chromatin complexes, thus preventing deacetylation of local histones and impeding the establishment of a repressive chromatin state.

It is also possible that HBP1 plays an architectural role in the LCR function, causing DNA bending and/or other structural changes and thus facilitating interaction between proteins that establish and maintain an active chromatin configuration. Such a role was suggested for the LEF1 HMG protein in the regulation of the ADA LCR (Haynes *et al*., 1996). The mutation of the LEF1-binding site within the ADA enhancer/LCR resulted in loss of site-independent transgene expression. The ability of HMG proteins to recognize altered DNA structures and bend DNA supports the notion that they play a central role in formation and functioning of enhanceosomes. Thus, it has been shown that the assembly and function of the αTCR enhancer complex is dependent on LEF1-induced bending (Giese *et al*., 1995; Love *et al*., 1995). Furthermore, it was found that the protein–protein interaction between components of the β-interferon enhanceosome are facilitated by HMG-I(Y) protein altering the enhancer DNA structure (Thanos and Maniatis, 1992; Du *et al*., 1993; Thanos *et al*., 1993; Du and Maniatis, 1994; Grosschedl, 1995).

While designing the experiment for assessing the effect of the FT1 deletion on hCD2 expression in transgenic mice, we considered a theoretical possibility of perturbations of nucleosomes and LCR-binding proteins by the deletion *per se*. To minimize such a possibility, we deleted 30 bp of nucleotides exactly, corresponding to a whole number (three) of full turns of the DNA double helix, and thus minimizing local perturbations. We have found that in thymocytes of the FT1.1 and FT1.2 variegating mice, the *HSS3* site remains hypersensitive (data not shown), indicating that the FT1 deletion *per se* causes minimal disturbances to the surrounding chromatin and does not prevent the other proteins from binding to the LCR and establishing an open DNase I-hypersensitive chromatin configuration.

Furthermore, the effect of the deletion is not as dramatic as that caused by the deletion of the whole *HSS3* region, suggesting that, in addition to HBP1, other *HSS3*-binding proteins (such as the FT2-binding factors and others) also contribute to the prevention of PEV. Potential involvement of other proteins in the regulation of the LCR function was demonstrated in the experiment with a partial deletion of the *HSS3* region which leaves FT1 and FT2 intact, but removes other footprints. Mice carrying such transgenes were subject to PEV (Festenstein *et al*., 1996), indicating that additional *HSS3*-binding proteins are also involved in prevention of PEV. The idea that several LCR-binding proteins are responsible for prevention of PEV is consistent with the mass action model of PEV (Locke *et al*., 1988), which implicates many elements in the establishment of euchromatin or heterochromatin structures (Kioussis and Festenstein, 1997).

The notion that heterochromatin spreads along the chromosomes from the centromere or telomere to the long arms is in line with our results demonstrating the inverse correlation between the degree of PEV and the distance of the transgene from the heterochromatin centres: the closer the transgene is placed to the centromere or telomere, the higher the degree of variegation (Renauld *et al*., 1993; Csink and Henikoff, 1996). Factors that determine how far the heterochromatin spreads in each cell are unknown and, according to the mass action model (Locke *et al*., 1988), this is a stochastic process depending on the local concentration of various chromatin components. This model is consistent with recent findings (Festenstein *et al*., 1999) demonstrating that the overexpression of centromere–heterochromatin-associated M31 protein in variegating ΔHSS3-LCR hCD2 mice results in an enhancement of variegation. It is worth noting that a similar overexpression of M31 in the ΔFT1-LCR hCD2 variegating mice did not result in enhancement of variegation (data not shown), perhaps indicating that the ΔFT1-LCR transgene is less sensitive to the repressive influence of heterochromatin than ΔHSS3-LCR hCD2. Another factor that may affect transgene silencing in mammals is the number of copies of the transgene (Garrick *et al*., 1998). However, in our study, we did not see any correlation between transgene copy number and degree of variegation (Figure 8A).

LCRs are powerful elements that are able to regulate expression of genes in a developmental- and tissue-specific manner by opening chromatin, thus rendering the gene

locus transcriptionally active (Kioussis and Festenstein, 1998). We show here that HBP1 contributes to the regulation of LCR function. This, taken together with the known role of HBP1 as a regulator of differentiation (Lesage *et al*., 1994; Lavender *et al*., 1997; Tevosian *et al*., 1997; Gartel *et al*., 1998; Shih *et al*., 1998; Yee *et al*., 1998), suggests a connection between chromatin opening and lineage commitment, indicating that the role of HBP1 in the cell cycle and differentiation may be mediated via chromatin remodelling.

Materials and methods

Isolation of human HBP1 cDNA

Screening of the Matchmaker one-hybrid and λ-ZAP Jurkat T cell cDNA libraries was performed according to the manufacturer's instructions (Clontech and Stratagene, respectively). DNA was sequenced using an ABI PRISM Dye-Terminator Cycle Kit (Perkin Elmer) and Automated Sequencer 377 (Applied Biosystems) as per the manufacturer's instructions. The data were analysed using ABI Sequence Analysis, Factura, AutoAssembler (ABI–Perkin Elmer) and LaserGene (DNA*) softwares.

Overexpression of recombinant HMG box in E.coli and generation of rabbit antisera

To overexpress a recombinant HBP1 HMG box, the DNA encoding the HMG box was amplified from isolated hHBP1 cDNA using the primers 5′-GGGGAATTCCATATGAAGTGCAAAAGACCAATGAAT-3′ and 5′-GGGAAGCTTCTACTCGAGTGAGCCTGAATTGGTTCTTTT-3′. The PCR product was digested with *Nde*I and *Hin*dIII, cloned into pET-22b plasmid and transformed into *E.coli* BL21 (DE3) (Novagen). Induction and purification of recombinant were performed following the manufacturer's recommendations and the protein was stored at –20°C.

The recombinant protein was injected into rabbits and immune antisera were raised. The LEF1 antisera was a generous gift from Dr R.Grosschedl.

Preparation of nuclear protein extract. DNAse I footprint and band shift assays

Nuclear extracts from fresh mouse tissue (thymus and liver) were prepared as described (Dignam *et al*., 1983). Probes were labelled either at the sense or the antisense strands by T4 kinase with [γ-^{32}P]ATP. For a typical binding reaction, nuclear protein extract (30 µg) or recombinant HMG (50 ng), probe (3000 c.p.m.) and poly(dI–dC) (1 µg) were incubated in 20 µl containing 20 mM HEPES pH 7.9, 1 mM MgCl$_2$, 60 mM KCl, 8% glycerol. After treatment with DNase I (Boehringer Mannheim), the probe was phenol/chloroform purified and analysed on an 8% polyacrylamide–8 M urea sequencing gel. Maxam–Gilbert A + G sequencing reactions were performed using a kit (Biotechnology System NEN RP) according to the manufacturer's instructions.

For band shift assays, double-stranded oligonucleotides were purified on a non-denaturing polyacrylamide gel and labelled by T4 kinase with [γ-^{32}P]ATP. For a typical binding reaction, nuclear extract (0.2 µg) or recombinant HMG (20 ng) was incubated with probe (20 000 c.p.m., equalling 0.1 ng) in the presence of poly(dI–dC) (20 ng). The mixture was loaded onto a 5% acrylamide gel and electrophoresed in 0.5× TBE at 200 V up to 4 h. In competition experiments, non-labelled competitor DNA was added with the poly(dI–dC).

Mice

Mice [CBA/Ca and C57Bl/10 and (CBA/Ca×C57Bl/10)F$_1$] were bred at the National Institute for Medical Research.

Evaluation of hCD2 expression and calculation of K$_{PEV}$

A total of 10^6 thymocytes were incubated for 30 min at 4°C with CD4 RED (Boehringer Mannheim), CD8 PE (Catlag Laboratories) and fluorescein isothiocyanate (FITC)-conjugated or biotinylated anti-hCD2 (OKT11) (Festenstein *et al*., 1996) antibodies and subsequently analysed using a Beckton Dickinson FACS sorter and CellQuest programme.

K_{PEV} was calculated as the percentage of thymocytes falling outside of the 1.3× SD interval measured from the mean of hCD2-positive thymocytes. The K_{PEV} for normal distribution is 10%. K_{PEV} for the variegating lines was calculated as the percentage of thymocytes falling outside of the 1.3× SD interval measured from the mean of peak of hCD2-positive thymocytes using the SD of the non-variegating control

Mg4 line (SD$_{Mg4}$). For the lines with a transgene expression level significantly different from that of Mg4, the SD used was adjusted according to the formula SD = SD$_{Mg4}$×P/P$_{Mg4}$, where P and P$_{Mg4}$ are the means of the peaks of hCD2-positive thymocytes in the considered line and Mg4 control, respectively.

Fluorescence in situ hybridization (FISH)

Metaphase spreads were obtained from transgenic mouse spleens cultured for 2 days after lipopolysaccharide stimulation (Sigma, final concentration 20 mg/ml). The hCD2 DNA probe was labelled and hybridized with the metaphase spreads following procedures previously described (Festenstein *et al*., 1996). Slides were then mounted in antifade (Vector) and counterstained in 4′,6-diamidino-2-phenolindole (DAPI). They were examined using a Zeiss Axiophot fluorescence microscope and the images collected by cooled CCD camera (Photometrics) using capture software (Digital Scientific).

Acknowledgements

References

Allshire,R.C., Javerzat,J.P., Redhead,N.J. and Cranston,G. (1994) Position effect variegation at fission yeast centromeres. *Cell*, **76**, 157–169.

Aronow,B.J., Ebert,C.A., Valerius,M.T., Potter,S.S., Wiginton,D.A., Witte,D.P. and Hutton,J.J. (1995) Dissecting a locus control region: facilitation of enhancer function by extended enhancer-flanking sequences. *Mol. Cell. Biol.*, **15**, 1123–1135.

Brehm,A., Miska,E.A., McCance,D.J., Reid,J.L., Bannister,A.J. and Kouzarides,T. (1998) Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature*, **391**, 597–601.

Csink,A.K. and Henikoff,S. (1996) Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature*, **381**, 529–531.

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) Accurate transcription initiation by RNA polymerase II in soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.

Dobie,K.W., Lee,M., Fantes,J.A., Graham,E., Clark,A.J., Springbett,A., Lathe,R. and McClenaghan,M. (1996) Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. *Proc. Natl Acad. Sci. USA*, **93**, 6659–6664.

Dobie,K., Mehtali,M., McClenaghan,M. and Lathe,R. (1997) Variegated gene expression in mice. *Trends Genet.*, **13**, 127–130.

Du,W. and Maniatis,T. (1994) The high mobility group protein HMG I (Y) can stimulate or inhibit DNA binding of distinct transcription factor ATF-2 isoforms. *Proc. Natl Acad. Sci. USA*, **91**, 11318–11322.

Du,W., Thanos,D. and Maniatis,T. (1993) Mechanisms of transcriptional synergism between distinct virus-inducible enhancer elements. *Cell*, **74**, 887–898.

Dunaief,J.L., Strober,B.E., Guha,S., Khavari,P.A., Alin,K., Luban,J., Begemann,M., Crabtree,G.R. and Goff,S.P. (1994) The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. *Cell*, **79**, 119–130.

Farkas,G., Gausz,J., Galloni,M., Reuter,G., Gyurkovics,H. and Karch,F. (1994) The Trithorax-like gene encodes the *Drosophila* GAGA factor. *Nature*, **371**, 806–808.

Festenstein,R., Tolaini,M., Corbella,P., Mamalaki,C., Parrington,J., Fox,M., Miliou,A., Jones,M. and Kioussis,D. (1996) Locus control region and heterochromatin-induced position effect variegation. *Science*, **271**, 1123–1125.

Festenstein,R., Sharghi-Namini,S., Fox,M., Roderick,K., Tolaini,M., Norton,T., Savelieu,A., Kioussis,D. and Singh,P. (1999) Heterochromatin protein 1 (M31) modifies mammalian PEV in a dose and chromosomal context dependent manner. *Nature Genet.*, in press.

Garrick,D., Sutherland,H., Robertson,G. and Whitlow,E. (1996) Variegated expression of a globin transgene correlates with chromatin accessibility but not methylation status. *Nucleic Acids Res.*, **24**, 4902–4909.

Garrick,D., Fiering,S., Martin,D.I. and Whitlow,E. (1998) Repeat induced gene silencing in mammals. *Nature Genet.*, **18**, 56–59.

Gartel,A.L., Goufman,E., Tevosian,S.G., Shih,H., Yee,A.S. and

Tyner,A.L. (1998) Activation and repression of p21$^{WAF1/CIP1}$ transcription by RB binding proteins. *Oncogene*, **17**, 3463–3469.

Giese,K., Amsterdam,A. and Grosschedl,R. (1991) DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev.*, **5**, 2567–2578.

Giese,K., Kingsley,C., Kirshner,J.R. and Grosschedl,R. (1995) Assembly and function of a TCRα enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein–protein interactions. *Genes Dev.*, **9**, 995–1008.

Granok,H., Leibovitch,B.A., Shaffer,C.D. and Elgin,S.C. (1995) Chromatin. Ga-ga over GAGA factor. *Curr. Biol.*, **5**, 238–241.

Grosschedl,R. (1995) Higher-order nucleoprotein complexes in transcription: analogies with site-specific recombination. *Curr. Opin. Cell Biol.*, **7**, 362–370.

Grosschedl,R., Giese,K. and Pagel,J. (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet.*, **10**, 94–100.

Grosveld,F., van Assendelft,G.B., Greaves,R.D. and Kollias,G. (1987) Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell*, **51**, 975–985.

Grosveld,F., de Boer,E., Dillon,N., Gribnau,J., McMorrow,T., Milot,E., Trimborn,T., Wijgerde,M. and Fraser,P. (1998) The dynamics of globin gene expression and position effects. *Novartis Found. Symp.*, **N214**, 67–86.

Grunstein,M. (1997) Molecular model for telomeric heterochromatin in yeast. *Curr. Opin. Cell Biol.*, **9**, 383–387.

Guy,L.G., Kothary,R., DeRepentigny,Y., Delvoye,N., Ellis,J. and Wall,L. (1996) The β-globin locus control region enhances transcription of but does not confer position-independent expression onto the *lacZ* gene in transgenic mice. *EMBO J.*, **5**, 3713–3721.

Haynes,T.L., Thomas,M.B., Dusing,M.R., Valerius,M.T., Potter,S.S. and Wiginton,D.A. (1996) An enhancer LEF-1/TCF-1 site is essential for insertion site-independent transgene expression in thymus. *Nucleic Acids Res.*, **24**, 5034–5044.

Higgs,D.R. (1998) Do LCRs open chromatin domains? *Cell*, **95**, 299–302.

Kioussis,D. and Festenstein,R. (1997) Locus control regions: overcoming heterochromatin-induced gene inactivation in mammals. *Curr. Opin. Genet. Dev.*, **7**, 614–619.

Kioussis,D. and Festenstein,R. (1998) Chromatin structure and lineage determination. In Monroe,J.G. and Rothenberg,E.V. (eds), *Molecular Biology of B-Cell and T-Cell Development*. Humana Press, Totowa, NJ, pp. 127–146.

Kleinjan,D.J. and van Heyningen,V. (1998) Position effect in human genetic disease. *Hum. Mol. Genet.*, **7**, 1611–1618.

Lake,R., Wotton,D. and Owen,M. (1990) A 3′ transcriptional enhancer regulates tissue-specific expression of the human CD2 gene. *EMBO J.*, **9**, 3129–3136.

Landsman,D. and Bustin,M. (1993) A signature for the HMG-1 box DNA-binding proteins. *BioEssays*, **15**, 539–546.

Lavender,P., Vandel,L., Bannister,A.J. and Kouzarides,T. (1997) The HMG-box transcription factor HBP1 is targeted by the pocket proteins and E1A. *Oncogene*, **14**, 2721–2728.

Lesage,F., Hugnot,J.-P., Amri,E.-Z., Grimald,P., Barhnin,J. and Lazdunski,M. (1994) Expression cloning in K$^+$ transport defective yeast and distribution of HBP1, a new putative HMG transcriptional regulator. *Nucleic Acids Res.*, **22**, 3685–3688.

Locke,J., Kotarski,M.A. and Tartof,K.D. (1988) Dosage-dependent modifiers of positional effect variegation in *Drosophila* and a mass action model that explains their effect. *Genetics*, **120**, 181–198.

Love,J.J., Li,X., Case,D.A., Giese,K., Grosschedl,R. and Wright,P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791–795.

Magnaghi-Jaulin,L., Groisman,R., Naguibneva,I., Robin,P., Lorain,S., Le Villain,J.P., Troalen,F., Trouche,D. and Harel-Bellan,A. (1998) Retinoblastoma protein represses transcription by recruiting a histone deacetylase. *Nature*, **391**, 601–605.

Milot,E. *et al.* (1996a) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell*, **87**, 105–114.

Milot,E., Fraser,P. and Grosveld,F. (1996b) Position effects and genetic disease. *Trends Genet.*, **12**, 123–126.

Oosterwegel,M., van de Wetering,M., Dooijes,D., Klomp,L., Winoto,A., Georgopoulos,K., Meijlink,F. and Clevers,H. (1991) Cloning of murine TCF-1, a T-cell specific transcription factor interacting with functional motifs in the CD3- and T-cell receptor enhancers. *J. Exp. Med.*, **173**, 1133–1142.

Oosterwegel,M., van de Wetering,M. and Clevers,H. (1993) HMG box proteins in early T-cell differentiation. *Thymus*, **22**, 67–81.

Ortiz,B.D., Cado,D., Chen,V., Diaz,P.W. and Winoto,A. (1997) Adjacent DNA elements dominantly restrict the ubiquitous activity of a novel chromatin-opening region to specific tissues. *EMBO J.*, **16**, 5037–5045.

Pravtcheva,D.D., Wise,T.L., Ensor,N.J. and Ruddle,F.H. (1994) Mosaic expression of an Hprt transgene integrated in a region of Y heterochromatin. *J. Exp. Zool.*, **268**, 452–468.

Pruzina,S., Hanscombe,O., Whyatt,D., Grosveld,F. and Philipsen,S. (1991) Hypersensitive site 4 of the human β globin locus control region. *Nucleic Acids Res.*, **19**, 1413–1419.

Renauld,H., Aparicio,O.M., Zierath,P.D., Billington,B.L., Chhablani,S.K. and Gottschling,D.E. (1993) Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength and by SIR3 dosage. *Genes Dev.*, **7A**, 1133–1145.

Robertson,G., Garrick,D., Wu,W., Kearns,M., Martin,D. and Whitelaw,E. (1995) Position-dependent variegation of globin transgene expression in mice. *Proc. Natl Acad. Sci. USA*, **92**, 5371–5372.

Shih,H.H., Tevosian,S.G. and Yee,A.S. (1998) Regulation of differentiation by HBP1, a target of the retinoblastoma protein. *Mol. Cell. Biol.*, **18**, 4732–4473.

Strober,B.E., Dunaief,J.L., Guha,S. and Goff,S.P. (1996) Functional interactions between the hBRM/hBRG1 transcriptional activators and the pRB family of proteins. *Mol. Cell. Biol.*, **16**, 1576–1583.

Talbot,D., Philipsen,S., Fraser,P. and Grosveld,F. (1990) Detailed analysis of the site 3 region of the human β-globin dominant control region. *EMBO J.*, **9**, 2169–2177.

Tevosian,S., Shih,H., Mendelson,K., Sheppard,K.-A., Paulson,K. and Yee,A. (1997) HBP1: a HMG box transcriptional repressor that is targeted by the retinoblastoma family. *Genes Dev.*, **11**, 383–396.

Thanos,D. and Maniatis,T. (1992) The high mobility group protein HMG I (Y) is required for NF-κB-dependent virus induction of the human IFN-β gene. *Cell*, **71**, 777–789.

Thanos,D., Du,W. and Maniatis,T. (1993) The high mobility group protein HMG I (Y) is an essential structural component of a virus-inducible enhancer complex. *Cold Spring Harbor Symp. Quant. Biol.*, **58**, 73–81.

Travis,A., Amsterdam,A., Belanger,C. and Grosschedl,R. (1991) *LEF-1*, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function. *Genes Dev.*, **5**, 880–894.

Tsukiyama,T., Becker,P.B. and Wu,C. (1994) ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature*, **367**, 525–532.

Tsukiyama,T. and Wu,C. (1995a) Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell*, **83**, 1011–1020.

Tsukiyama,T., Daniel,C., Tamkun,J. and Wu,C. (1995b) ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell*, **83**, 1021–1026.

van de Wetering,M., Oosterwegel,M., Dooijes,D. and Clevers,H. (1991) Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J.*, **10**, 123–132.

van de Wetering,M., Oosterwegel,M., van Norren,K. and Clevers,H. (1993) Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J.*, **12**, 3847–3854.

Wallrath,L.L. (1998) Unfolding the mysteries of heterochromatin. *Curr. Opin. Genet. Dev.*, **8**, 147–153.

Weiler,K.S. and Wakimoto,B.T. (1995) Heterochromatin and gene expression in *Drosophila*. *Annu. Rev. Genet.*, **29**, 577–605.

Yee,A.S., Shih,H.H. and Tevosian,S.G. (1998) New perspectives on retinoblastoma family functions in differentiation. *Front. Biosci.*, **3**, D532–D547.