The housekeeping promoter from the mouse CpG island HTF9 contains multiple protein-binding elements that are functionally redundant

Maria Patrizia Somma[1,2], Claudio Pisano[2] and Patrizia Lavia[2,*]
[1]Dipartimento di Genetica e Biologia Molecolare and [2]Centro di Genetica Evoluzionistica del CNR, Universita' 'La Sapienza', Rome 00185, Italy

Received April 10, 1991; Revised and Accepted May 10, 1991

ABSTRACT

The mouse CpG-rich island HTF9 harbours the divergent RNA initiation sites shared by two genes that are both expressed in a housekeeping fashion. In this work we have analyzed the architecture of the HTF9 promoter. Gel shift assays were first employed to locate nuclear factor-binding sites within HTF9. Multiple protein-binding sites were identified across a 500 bp-long region, two of which appear to interact with novel factors. Deletion analysis was used to determine the requirements for the different sites in transient expression of a CAT reporter gene. Although multiple elements contributed to the overall promoter strength in each orientation, extensive deletions failed to affect the basal level of transcription from HTF9 in either direction. Thus, only a subset of elements is necessary to activate transcription from HTF9. Functional redundancy may be a general feature of housekeeping CpG-rich promoters.

INTRODUCTION

Eukaryotic promoters contain multiple control elements or modules, which usually identify transcription factor binding sites (rev. 1–3). Promoters of genes encoding basic 'housekeeping' functions may well represent the commonest class of promoters (4), yet much of our understanding of promoter recognition and activation derives from the study of tissue-specific genes (rev. 1, 3), while little is known on the functional organisation of housekeeping promoters. In mammals, housekeeping promoters are often described as 'atypical' due to the lack of recognizable control signals such as the TATA and CCAAT boxes (5, 6). One characteristic feature of this class of promoters is their association with domains of CpG-rich, unmethylated DNA (CpG islands), whose extent (1–2 kb) well exceeds the average promoter size. The distinctive structural properties of CpG islands have been extensively characterized (rev. in 7) and their implication in 'marking' expressed portions of the genome has been discussed (8). CpG island chromatin is distinctively accessible *in vivo* (9–11 and refs therein). These observations have lead to the suggestion that the whole CpG-rich domain might be implicated in factor binding, thereby participating in promoter activity.

Most studies of DNA:protein interactions at housekeeping promoters have focussed on the role of individual transcription factors in activating specific promoter sequences. Examples of this type of study include analysing the effect of Sp1 on expression from the mouse dihydrofolate-reductase (DHFR) (12) and adenine phosphoribosyl-transferase (APRT) (13) genes; of AP1 and AP2 on the metallothionein-2 (mtt-2) promoter (14–16); of E2F (17) and of the HIP1 protein (18) on DHFR transcription. Only a few CpG islands have been extensively dissected to produce a complete overview of the promoter architecture. In all cases in which such a dissection was achieved, the results have pointed out a complex, possibly redundant architecture, involving multiple protein-binding elements; for example, several factor-binding sites corresponding to control regions were identified in the promoter of three mouse ribosomal protein genes (19–22) and in the human phosphoglycerate-kinase (PGK) promoter both *in vitro* (23) and *in vivo* (24).

The CpG island HTF9 was originally isolated from the mouse genome during the characterization of the CpG-rich DNA genomic fraction (25). Transcription studies showed that two genes arranged head-to-head are transcribed with divergent polarity from opposite DNA strands of HTF9 (26). The lower-strand gene (HTF9-A) encodes a protein with characteristics typical of certain DNA-binding proteins and particularly resembling HMG1 (27), while the upper-strand gene (HTF9-C) encodes a structurally unrelated product of unknown function. Both divergent genes are expressed in all cell lines and tissues tested and in embryos (26). One feature of HTF9 which is unique among bidirectional promoters is that the divergent RNAs are originated at coincident sites on complementary DNA strands: thus, the upstream sequences of each gene fall within the transcribed portion of the opposite-strand gene. No TATA box is apparent on either strand, which is consistent with the heterogeneous initiation of RNA transcription seen with both

* To whom correspondence should be addressed

genes. A CCAAT sequence maps upstream of one gene only. The initiation region also includes two putative Sp1 recognition sites.

The present work was undertaken to disentangle the arrangement of the control elements within the HTF9 island. Multiple protein-binding elements were mapped over a 500bp-long portion of HTF9. Deletion analysis revealed that distinct sets of elements contribute to the efficiency of transcription on either strand, however extensive deletions were tolerated without substantial loss in the promoter activity: in fact, an 85 bp-long region retaining only an Sp1 site and a site (termed site G) for a novel factor was sufficient for transcription in both directions. These results suggest that the multiple binding sites in the CpG-rich island HTF9 are functionally redundant.

MATERIALS AND METHODS

Nuclear extracts

Crude extracts from cultured BalbC 3T3 or NIH 3T6 were prepared as in ref. 28 and precipitated overnight with 0.33 g/ml ammonium sulfate. The protein concentration was measured using the Biorad protein kit. Aliquots of the extract preparations were run on 10−12% test SDS/polyacrylamide gels.

Probes and competitors

HTF9 (accession code X05830) was cleaved with restriction enzymes to obtain fragments ranging from 90 to 250 bp in size (Fig. 1a). Overlapping fragments were also generated with different enzymes so as to leave no cleavage site uncovered. Fragments were gel-purified and eluted with standard methods. The following sequence and its complementary strand were used as a test Sp1 binding site: 5′-GATACGCGTATCGGGGCGGA-GAAACACCGT-3′. Single DNA strands were synthesized with a Coder 300 (Dupont). The purified strands were a gift from A.Felsani (Istituto di Tecnologie Biomediche CNR, Roma). For duplex annealing, both DNA strands were incubated at in 100 mM NaCl 10′ at 70°C and then shifted to room temperature. The polyoma enhancer region included between the 5020 BclI and the 5265 PvuII sites was a gift from M. Caruso (Ist. Biologia Cellulare CNR, Roma). The E2F-binding probe was a RsaI-TaqI 72 bp-long fragment from the mouse DHFR promoter (the TaqI site is at position +1 according to the numbering used in most DHFR gene reports, as it overlaps the translation initiation codon, see ref. 29). The DHFR competitor containing four Sp1 binding sites was a SmaI-RsaI (position −411 to −71) fragment.

Gel shift assays

Binding reactions contained either 100 pg of gel-purified probe which was end-labeled using the Klenow fragment of DNA polymerase I and 50 μCi of the appropriate α-^{32}P-dNTP, or 20 pg of double-stranded oligonucleotide end-labeled using T4 polynucleotide kinase and γ-^{32}P-ATP. Routinely, 3−4 μg of ammonium sulfate precipitated proteins were preincubated with 1 μg of poly (dI.dC) for 10′ on ice in a 20 μl reaction containing 25 mM Hepes (pH 7.6), 60 mM KCl, 8.7% glycerol, 1 mM DTT and 0.5 mM EDTA. In the competition experiments, prebinding reactions included a 10−100 fold molar excess of cold specific competitor DNA. Incubation was continued for a further 20′ following addition of the labeled probe. Reactions were run on 4% acrylamide gels (29:1 cross-linking ratio) in 1 mM EDTA, 3 mM Na acetate, 6.7 mM Tris (pH 7.5) or in O.5×TBE buffer at 4°C.

DNaseI footprinting

A 265 bp-long SmaI-TaqI fragment was excised from the pTHF9 A-CAT subclone and end-labeled at the TaqI site on the lower strand of HTF9. The complementary strand of HTF9 was labeled after purification of a 418 bp-long Asp718-PvuII fragment from the pTS-C subclone. Details of the subclones are given below and in Fig. 7. Footprinting reactions were essentially set up as above, except that 30−60 μg of extract and larger restriction fragments—usually 70.000 cpms of probe labeled at one end—were used. After 20′ incubation on ice, the binding reaction was diluted to 50 μl in 10 mM $MgCl_2$ and subjected to DNaseI digestion for 3′ on ice. Reactions were stopped by adding 1% SDS, 25 mM EDTA. The mixtures were phenol-extracted, ethanol precipitated, redissolved in 95% formamide, 0.1% xylene cyanol, 0.1% bromophenol blue dye and loaded onto a 7 M urea −6% acrylamide gel. At least two DNase I concentrations were used for each binding reaction.

Southwestern blot assays

40 μg of nuclear extract were run on 12% SDS-PAGE gels. Prestained markers were from Biorad. Gels were electroblotted in 20% methanol, 0.3% Tris, 1.44% glycine on NTC sheets. Filter strips corresponding to individual slots were renatured for 2−6 hrs in buffer 1 (10 mM Tris-HCl pH 8.0, 15 mM Mg acetate, 7 mM KCl, 10 mM β-MeSH, 0.1 mM EDTA and 1×Denhardt's solution), and hybridized using 300.000 cpms of probe and 10 μg of poly (dI.dC) in buffer 2 (50 mM Tris-HCl pH 8.O, 2 mM β-MeSH, 25 mM NaCl, 1 mM EDTA) for 1−3 hrs at room temperature. The EDTA concentration was raised to 5 mM in experiments designed to minimise Sp1 binding to DNA. Filters were washed several times in the same buffer, exposed and autoradiographed.

Expression constructs and transfection experiments

A 558 bp-long fragment extending from the HpaII site at position 444 to the SmaI site at position 1002 in HTF9 was gel-purified, blunt-ended using the Klenow fragment of DNA polymerase I and cold dNTPs, and cloned in both orientations in front of the CAT gene after filling in the unique HindIII site of the pSVO vector. This gave the original subclones pHTF9 A-CAT and pHTF9 C-CAT respectively. The initial 558 bp fragment was then digested with suitable restriction enzymes (see Fig.7) to give eight more opposite-orientation subclones in which the promoter region was deleted to various extents. Transfections were carried out using the calcium-phosphate precipitation method. Typically 8 millions BalbC 3T3 cells were transfected with 5 μg of construct DNA. The plasmid pSp1 used for in vivo competition experiments was constructed by self-ligating 8 copies of the Sp1 binding 30-mer and cloning into the SmaI site of pUC 19. Cotransfections were carried out using 12.5 and 25 μg of competitor pSp1 construct versus 5 μg of tester pTH-A construct, corresponding to a molar ratio of 1:30 and 1:60 Sp1 sites respectively. The total DNA amount was equalized in cotransfections and in control transfections by adding pUC19 DNA to 30 μg. Plasmid uptakes were controlled by Southern blotting of the trasfected cell DNA and probing with a gel-purified pUC ampicillin resistance gene. Each transfection experiment was repeated three to six times and was carried out each time on duplicate sets of cultures. Promoter strengths were quantitated by scintillation counting of the ^{14}C-modified and unmodified cloramphenicol after thin-layer chromatography. Controls included pSVO, pRSV (30) and pA10-CAT2 (31)

RESULTS

Identification of multiple protein-binding regions in HTF9

To locate the sequences that might be relevant in the HTF9 promoter, we firstly characterised the regions of protein interaction. The most CpG-rich region of HTF9 (1194 bp-long) was 'scanned' in gel shift assays to identify the target sites for nuclear factors. The map in Fig. 1 shows the extent of the CpG-rich region, the location of the exons and intervening sequences of both genes and the probes used (labeled 1 to 8). Each fragment (referred to according to the numbering in Fig. 1) was individually incubated with nuclear proteins and assayed in a gel retardation experiment. Results are shown in the panel in Fig. 1. Three fragments, identified by probes 1, 2 and 4 gave discrete mobility shifts after incubation with nuclear extracts. The regions encompassed by these probes map respectively to the left of (probes 1 and 2) and around the divergent RNA initiation region (probe 4), and henceforth will be referred to as the distal and proximal protein-binding regions respectively. The stability of the complexes was tested using increasing amounts (0.1 to 2.5 μg) of competitor mouse genomic DNA, of poly (dI.dC) and of E.coli genomic DNA. All complexes remained stable under all tested conditions (data not shown) suggesting that critical DNA-binding elements map there.

Distal protein binding region

The distal protein-binding region identified by probes 1 and 2 extends 400 to 160 bp upstream of the upper-strand transcript (HTF9-C) and falls in the first intron of the lower-strand gene (HTF9-A). We noticed several putative targets for known factors (Fig. 2), which include several potential Sp1 sites and a sequence identical to the TC-II site recognized by AP2 in the SV4O and polyoma enhancers (16). Site-specific competitor sequences were used to ascertain whether either Spl or AP2 interacted with the distal region. Fig. 2 shows that neither a high-affinity Spl binding oligonucleotide (sequence in Materials and Methods), nor an Spl site contained in HTF9-region 4 and characterized for its binding ability (see below) were capable of competitively inhibiting the original binding of a factor to probe 2 (Fig. 2a, lanes 6 and 8). The complex was also unaffected in competition experiments using either the AP2 site from the polyoma enhancer, or HTF9-region 6, which is identical to the AP2 site in the mtt-2 promoter (Fig. 2a, lanes 5 and 10). On the other hand, probes 1 and 2 were effective mutual competitors and must therefore be bound by the same factor (Fig. 2b, lanes 4 and 6). One similar sequence in both probes involved a consensus with the MLTF site in the adenovirus major late promoter region (32), i.e. the AGGtCACGcG homology in probe 1 and the gcCCACGTGgCC homology in probe 2. The sequence in probe 2 overlaps a BglI site, which enabled us to generate a truncated probe—indicated as probe 2*—cleaved within the MLTF-like site. No complex was formed when probe 2* was used (Fig. 2b, lanes 9 and 10). Moreover, the MLTF-disrupted probe did not competitively inhibit the binding to either probe 1 or 2 (Fig. 2b, lanes 3 and 7). Thus, the factor binding both distal elements in HTF9 has a specificity related to that of the MLTF protein.

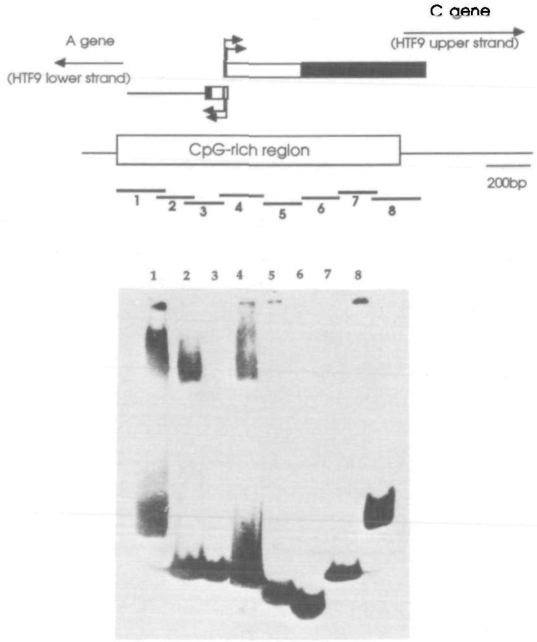

Figure 1. A. The HTF9 locus. The top part of the figure shows the exon/intron structure at the 5'end of the A and C genes (leader sequences are represented by open boxes, coding sequences by filled boxes and introns by lines). Two major divergent RNA starts are arrowed. The probes used for gel shift assays are labeled 1 to 8. **B.** Gel shift assays of probes 1−8 with 3T3 extracts (5 μg/lane).

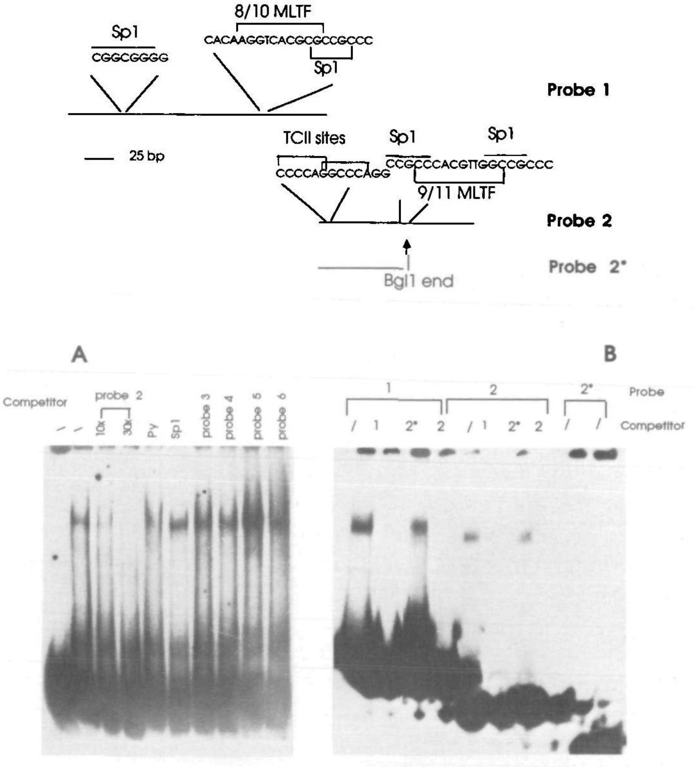

Figure 2. Top: scheme of probes 1, 2 and 2*. Sites for known factors are indicated. Probe 2* was generated by BglI digestion of probe 2 and is disrupted in the MLTF-like site. **A.** Gel shift assays with probe 2. Lane 1: no protein; 2: 5 μg of 3T3 extract and no specific competitor; 3 to 10: specific competitors (50×molar excess unless otherwise stated) are indicated above each lane. Competitors 4 (lane 8) and 6 (lane 10) contain Spl and AP2 binding sites respectively. **B.** Competition between probes 1, 2 and 2*. Lanes 1−4: binding reaction with probe 1 and 5 μg of 3T3 extract. Competitions included a 50×excess of unlabeled fragment 1 (lane 2), 2* (lane3) and 2 (lane4). Lanes 5−8: binding to probe 2 in the presence of a 50×excess of fragment 1 (lane 6), 2* (lane7) and 2 itself (lane 8). Lanes 9−10: probe 2* incubated with 5 μg of 3T3 (9) or with no extract (10).

Spl binding in the proximal region of HTF9

Probe 4 encompasses the RNA origins of both genes in HTF9. Gel shift assays had revealed multiple shifts in that region (see Fig.1). Two Spl recognition sites are apparent in probe 4: one (position 822) is homologous to the Spl box (VI) in the SV4O enhancer, which has a medium binding affinity, while the second one (position 842) is homologous to the DHFR-Spl boxes II and IV, and can be expected to represent a high-affinity binding site (rev. in 33). The latter falls between the two main divergent RNA start sites (see map in Fig. 3). To assess the actual binding ability of this Spl site, we tested the ability of probe 4 to competitively inhibit the formation of a complex on a high-affinity Spl-binding oligonucleotide. Fig. 3a shows that the Spl box from HTF9 is as effective a binding site as the oligomer itself (compare lanes 3 to 6, 4 to 7, and 5 to 8). We then used HTF9-fragment 4 as the probe to visualise the interaction of Spl with HTF9 (Fig.3). Three distinct shifts were resolved in these experiments. The most slowly migrating one (Fig. 3b, lane 5) was inhibited by a molar excess of the Spl oligomer (Fig. 3b, lanes 3 and 4), and by an excess of a DHFR promoter fragment (lane 2) which is known to bind Spl at multiple sites (12). Thus, the slowest band represents a genuine Spl complex. Two more complexes were apparent, whose intensity or mobility were not affected by a 70-fold molar excess of Spl oligomer, indicating that neither interaction was affected by variations in the binding of Spl to a neighboring site. These results suggest that the proximal region of HTF9 includes one or more protein-binding elements other than the Spl site. The complexes were designated Ubl and Ub2 respectively for their ubiquitous distribution in a variety of tested extracts (M.P.S. et al., in preparation).

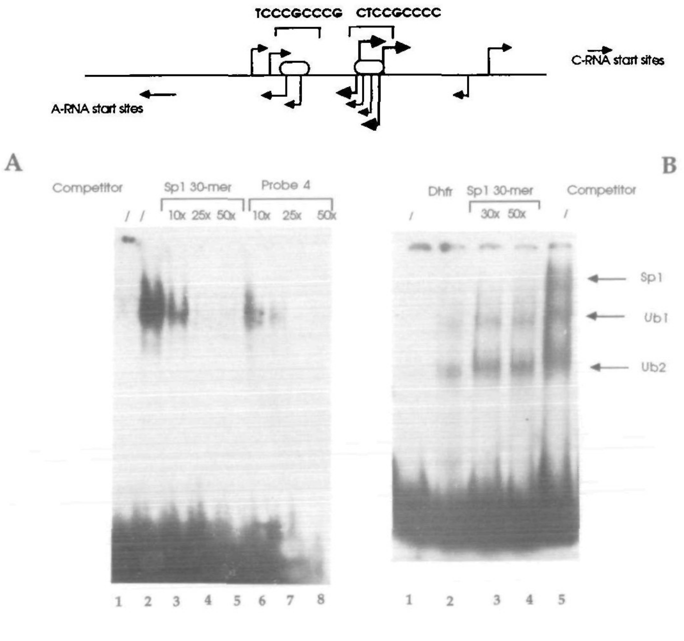

Figure 3. Top: scheme of probe 4 showing the start sites of the HTF9-A (above the line) and HTF9-C (below) transcrips. Thick arrows mark the major start sites. Two putative Spl sites are indicated. **A.** Competition between the HTF9-fragment 4 and a synthetic high-affinity Spl-binding probe. 1: no extract; 2: 3.5 μg of 3T3 extract and no specific competitor; 3–5: unlabeled Spl 30-mer added; 6–8: unlabeled probe 4 added. Molar excesses are indicated above each lane. **B.** Gel shifts of probe 4 upon incubation with 5 μg of 3T3 extract. 1: no extract; 2: 30×excess of a DHFR fragment (−441 to −71) containing four Spl binding sites; 3: 20×excess of Spl binding 30-mer; 4: 50×excess of Spl binding 30-mer; 5: no specific competitor.

Novel factor-binding elements in the proximal region

The Ubl and Ub2 complexes in Fig. 3b might have reflected either the individual binding of different proteins to their cognate sequences, or the di/multimerization of a factor bound to a single sequence. These possibilities can be distinguished by Southwestern binding assay. Nuclear proteins were fractionated on a polyacrylamide-SDS gel and their binding to labeled DNA probes was assayed. We firstly used the entire probe 4 (a 150 bp-long fragment, see map in Fig. 4). To minimise the possibility that probe 4 was sequestrated on the filter by Spl, the binding was carried out either with nuclear extracts of different origin and whose Spl content was low (M.P.S. et al., in preparation) or in high EDTA conditions which disfavour Spl binding. Three bands were detected of apparent Mr 30, 33 and 53 KDa (Fig. 4, lane 3). An additional 39 KDa band was thought to represent a non specific binding event, since it was bound by several probes (for example, lane 5 shows a control experiment with probe 3, which does not participate in any protein complex as seen by the gel shift assay in Fig. 1). We then shortened the probe to 90 bp encompassing the RNA initiation region (probe 4* in the map in Fig. 4). The latter fragment only bound the 53 KDa protein (lane 4). Thus the 30−33 KDa factors are likely to bind to sites on the left of the initiation region.

DNaseI protection experiments were carried out to characterise the target sites for such factors. A strong protection was observed at three locations. The most stable footprint extends from position 858 to position 879 (protection 3 in Fig. 5), a position which fits well with the location of the 53 KDa factor. The protected sequence (5′-CCCTGACCCCTGACCCC-3′) consists of a

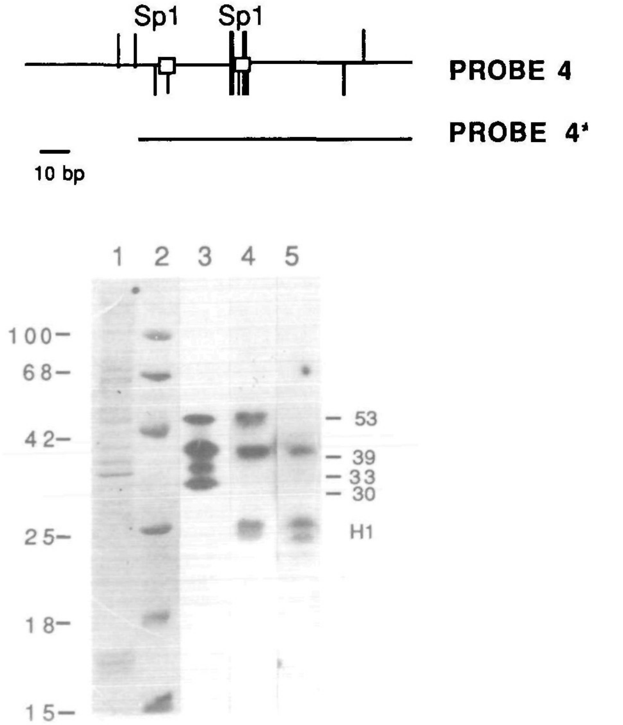

Figure 4. Top: Scheme of probes 4 and 4*. Below: Southwestern blotting of nuclear extracts. **1**: Comassie blue staining of nuclear extracts (40 μg); **2**: molecular weight markers; **3**: binding to probe 4; specific bands are indicated; **4**: binding to probe 4*; **5**: probe 3 was used as a control of non specific binding events: only a 39KDa protein and H1 are bound.

tandem duplication of the motif CCCCTGA, which resembles certain types of AP2 sites found in the growth hormone gene (TGCCCCTG), in the mtt-2 gene (CGCCTG) and in the polyoma/SV40 enhancer (GTCCCCAG) (16). However, competition with the polyoma enhancer did not affect the footprint (not shown). The efficiency of protection from DNase I cleavage suggests that the factor is either very abundant or that it binds with a high affinity to its target sequence.

As anticipated from the gel shift results, the footprint extended into protection of the Sp1 sequence (protection 2 in Fig. 5, position 838–852). The most distal CG box (i.e., the SV40 VI-like sequence) was left unprotected in these experiments.

Another protein-binding element (protection 1 in Fig. 5) was mapped further left (position 796 to 812) of the RNA start site cluster. That position falls 25–45 bp upstream of the C gene and 25–45 bp within the untranslated leader region of the A gene. That location suggests that the 30–33 KDa protein(s) generate the DNase I protection. It is possible that electrophoretic variations of one factor were visualized in the Southwestern assay—such as those of H1, which also migrates as a doublet in our experiments—or else the factor might exist in two forms, perhaps due to post-translational modifications. Both strands of the protected element, 5′-CTTTCCTCCGCGTCTG-GCGCCGG-3′, show a high homology (7/8 bp) with the E2F site in various genes (17). Since a well characterised E2F- site is contained in the DHFR promoter, whose mutation is detrimental to DHFR transcription (17), we were intrigued by the possibility that E2F did also bind to HTF9. However, the HTF9-complex was not competitively inhibited by an excess of the E2F site from the DHFR promoter (Fig. 6a). Conversely, the E2F-specific complexes formed at the DHFR site were not affected by an excess of HTF9-derived probe (Fig. 6b), indicating that these complexes are generated by unrelated factors.

Together our data suggest that two novel factors interact with the proximal region of HTF9. One is a 53 KDa protein, which binds to a unique site in the initiation region and contains a direct repeat of the motif CCCCTGA. We currently refer to the 53 KDa protein-binding site as the G element. The other is a protein of 30–33 KDa which protects a sequence similar to characterised E2F sites, but is distinguishable from E2F by site-specific competition assays. We call the target site for this protein the E element. In addition, a strong binding was observed over the Sp1 site. All three sites are closely spaced in the region containing the major start sites of divergent transcription.

Deletion mapping analysis

The experiments described above revealed multiple activities whose binding sites are distributed over 500 bp of HTF9. We wished to assess the contribution of the various protein-binding regions to transcription from HTF9 in either orientation. A 558 bp-long insert containing all positive regions in the gel-shift assay (i.e., fragments 1, 2 and 4 in Fig.1) was initially cloned in opposite orientations into a promoterless pSVO-CATvector. Deletions removing one or more factor-binding sites were

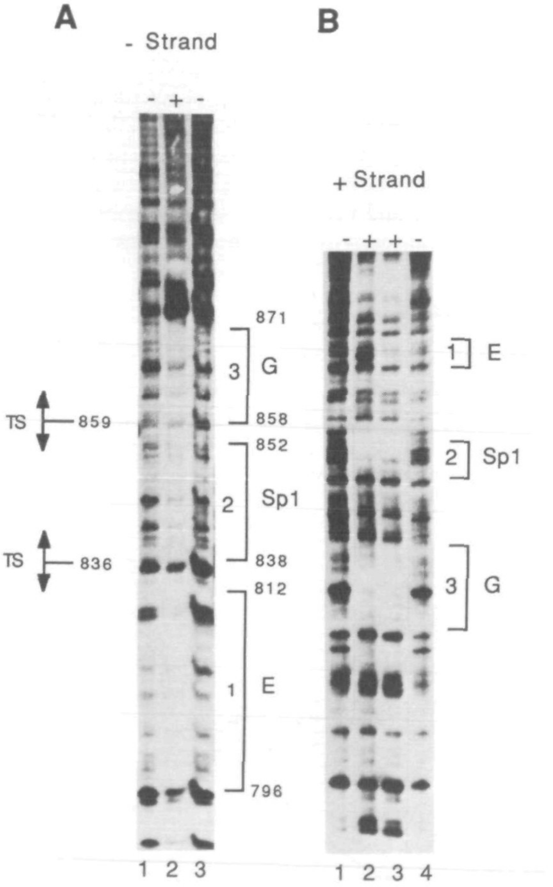

Figure 5. DNase I footprinting of the proximal region from HTF9. Protected regions are indicated. **A.** A 265-bp fragment labeled on the lower strand of HTF9 was incubated with 40 μg of 3T3 extracts. 1and 3: DNase I (20ng) cleavage pattern of the unbound probe. 2: 40 μg of 3T3 extract, followed by digestion with 200 ng DNAse I. The double-headed arrows mark the major divergent transcription starts (TS). **B.** Labeled complementary strand. The probe was incubated with 40 μg of NIH 3T6 extract (2 and 3) and digested with 150 (2) or 300 (3) ng of DNaseI. A non reproducible protection was occasionally generated by 3T6 extracts below region 3. Lanes 1 and 4: unbound probe digested with 20 (1) or 10 (4) ng of DNaseI.

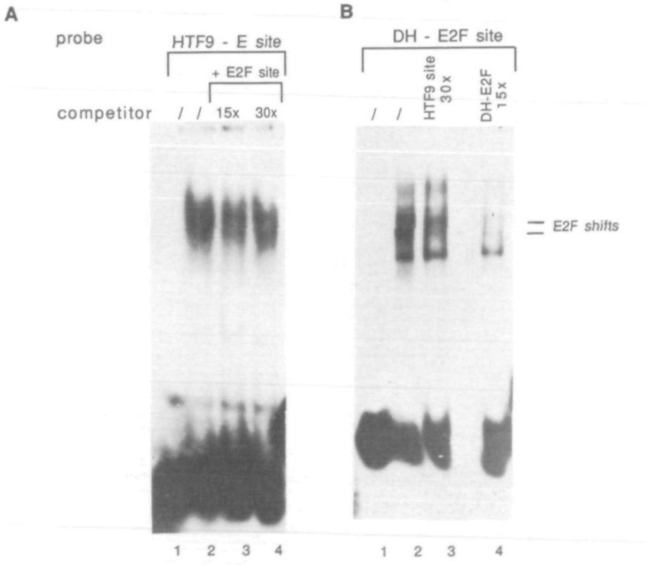

Figure 6. Competition between the E2F-like sequence of HTF9 and the E2F site from the DHFR promoter. **A.** A 70-bp fragment from the left portion of probe 4 was generated (see map in Fig.4) and incubated with 3.5 μg of extract. 1: no protein; 2: no specific competitor; 3 and 4: 15×and 30×excess of DHFR-E2F site respectively. **B.** The E2F-binding DHFR fragment was incubated as above. 1: no protein, 2: no specific competitor; 3: 30×excess of HTF9-E site; 4.15×excess of unlabeled DHFR-E2F probe. Two specific complexes formed by E2F are indicated. The band below represents a non specific complex (see 17) and is not prevented by either competitor.

progressively generated, which gave rise to ten opposite-orientation constructs (Fig. 7).

We initially measured the promoter efficiency of the original constructs carrying both the distal (probes 1 and 2) and the proximal (probe 4) regions in both orientations (subclones pHTF9 C-CAT and pHTF9 A-CAT respectively, lines 1 and 6 in Fig. 7). The entire region supported transcription of the CAT gene in both directions. The overall promoter activity was comparable in both orientations, and ranged between 40 and 45% of the transcriptional efficiency of the LTR in the pRSV construct.

To determine the contribution of the distal elements, the region corresponding to fragments 1 and 2 was deleted in both reverse-orientation constructs. The resulting subclones (pTS series, lines 2 and 7) retained 265 bp from the proximal region and had lost both MLTF-related sites. The C orientation of the deleted pTS subclone maintained 67% of the full-length promoter activity, suggesting that the distal sites together contributed some 40% of the C promoter strength. This was confirmed by assaying the promoter ability of the distal region alone (construct pM2-C, line 3). On the other hand, removal of the distal region in the A orientation increased the promoter efficiency, which now reached 80% of the pRSV transcriptional level (compare lines 6 and 7). That result may indicate that some intragenic element quenched the transcriptional activity of the original 558 bp insert in pHTF9-A. However, assay of the pM2-A construct, carrying the distal region alone in the A orientation (line 8), still showed some basal CAT expression, comparable to that driven by the minimal promoter in the pA10 control construct (14% of pRSV). This result rules out that sequences in the distal region inhibited transcription in the HTF9-A direction. The decreased activity of the full-length pHTF9-A compared to the proximal pTS-A subclone (see lines 6 and 7), may rather result from the presence of a 5′ donor without a 3′ acceptor splicing site upstream of the CAT gene in the pHTF9-A construct. Unpaired splicing sites have been reported to affect the stability and/or correct processing of the CAT-mRNA in unrelated constructs (21). Therefore, the figure obtained for the pTS-A subclone is taken to indicate the wild-type promoter efficiency of the A gene, which is underestimated in the pHTF9-A construct.

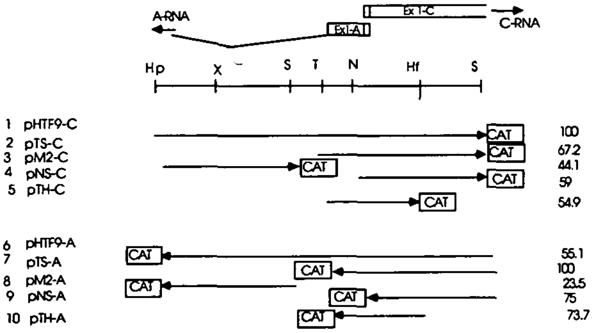

Figure 7. Deletion mapping of the HTF9 island. Top: map of the region and restriction sites used for generating the deletions. Hp=Hpa II; X=XhoI; S=SmaI; T+Taq I; N=Nar I; Hf=Hinf I. Below, the lines represent portions of HTF9 that were inserted upstream of the CAT gene in either orientation. Promoter efficiencies were quantitated as the ratio of ^{14}C-modified to unmodified CAT activity. Each figure represents the mean value from three assays at least. The strongest promoter in each orientation was given a value of 100. Absolute values were 3.03% and 5.6% of CAT conversion for pHTF9-C and pTS-A respectively, corresponding to 43% and 78% of the pRSV efficiency.

These experiments established that most of the C and all of the A promoter activity were due to elements contained in the pTS constructs. We attempted to delimit these elements further. A deletion to the left of the initiation region and removing the E site (constructs pNS-C and pNS-A, lines 5 and 10) mildly affected transcription in both directions, its effect being somewhat more noticeable in the A rather than the C direction. A deletion on the right-hand side removed a region including a CCAAT box in the 'A' direction. Although no factor binding had been detected in either the gel shift (Fig. 1) or the footprinting (Fig. 5) experiments under the conditions reported here, we wished to determine whether that sequence contributed at all to the overall promoter activity. Removal of the CCAAT box (constructs pTH-A and pTH-C, lines 4 and 9) resulted in a moderate drop of the A promoter efficiency, which however retained 73% of the whole promoter strength. Virtually no effect was observed in the C orientation. No further deletions could be generated within the proximal region without disrupting some of the multiple RNA start sites.

These data show that practically all assayed deletions around the initiation region were tolerated with no major effect on transcription from HTF9. The deletion analysis formally defines two separate promoters on each strand (summarised in Fig. 8), within which different sets of elements contribute to the overall efficiency, yet are not absolutely required for promoter activity. The upper limit of the region sufficient for transcription in both directions is defined by the 5′ boundary of the pNS subclones and the 3′ boundary of the pTH subclones. The overlap between such deletions leaves 85 bp around the initiation region, only retaining the Sp1 and the G sites, and contributing the bulk of the transcriptional activity in both directions.

DISCUSSION

The CpG-rich island HTF9 contains the origins of two genes that are arranged head-to-head and are bidirectionally transcribed from opposite DNA strands. The expression of the HTF9-associated transcripts is that of typical housekeeping genes. This work represents an initial step towards the identification of the elements

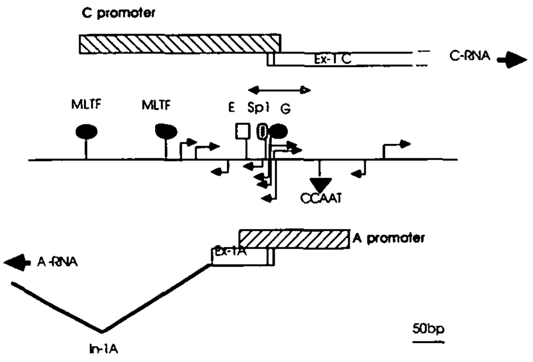

Figure 8. Summary of the binding activities over HTF9. E and G indicate the E2F-like site and the CCCTGA repeat respectively. The extent of the C and A promoters as defined by deletion mapping is represented by dashed boxes above and below the central line respectively. The open boxes indicate the exons of the A and C genes, the thick line represents the first intron in the A gene; the direction of transcription is indicated by thick arrowheads. Vertical arrows mark the RNA start sites. The horizontal double-headed arrow depicts the extent of the region sufficient for transcription in either direction.

that are part of the HTF9 promoter. Regulatory elements are often the target sites of trans-activating factors. Therefore, HTF9 was initially 'scanned' by gel shift assays to locate the regions of protein interaction. These experiments enabled us to construct a map of putative factor-binding elements across the island.

Firstly, two protein-binding sites were identified distally to the RNA initiation region, which both interacted with a factor related in specificity to the MLTF protein. Experiments carried out in an independent study of HTF9 (C.Tyndall, F.Watt, P.Molloy and M.Frommer, in preparation) have confirmed that both sites are indeed bound by MLTF. Together the MLTF-related sites contribute about 40% of the efficiency of the C promoter, while having no effect on the A promoter. Interestingly, MLTF also binds to the surfeit bidirectional promoter, and mutation of its binding site (site Su'y') decreases the transcriptional efficiency of the Surf-2 gene, possibly by altering the usage of the major Surf-2 RNA start site, while having a negligible effect on the divergently transcribed Surf-1 gene (34). It is possible that in both TATA-less promoters MLTF affects start sites usage.

Another region of protein interaction contained closely spaced protein-binding sites encompassing the initiation region. Firstly, a high-affinity Sp1 site was identified which overlaps one major divergent RNA start site. Sp1 was earlier suggested to be *per se* bidirectionally active, since Sp1 recognition sequences are found in either orientation upstream of similarly expressed genes (rev. in 5) and Sp1 binding occurs close to bidirectionally transcribed sequences (rev. in 33), in particular in the core of the 'major' bidirectional promoter of the DHFR gene (12). However, the lack of Sp1 sites in the divergent promoter of the Surf-1 and Surf-2 genes (34, 35) indicates that Sp1 binding is not an absolute requirement for bidirectional transcription to occur. Secondly, a novel element, that we termed E, was highly homologous to sites recognized by the HeLa E2F factor. Despite the sequence similarity though, the E-binding protein does not represent the mouse homologue of the E2F factor as shown by a specific competition experiment with the E2F site from the DHFR promoter. Moreover, E2F is mainly induced upon adenovirus infection of HeLa cells, while its steady-state abundance is low in non-infected cells. E2F footprinting over the DHFR promoter (17) was detected using very large amounts of nuclear extracts (200 μg, i.e. over four times the amount required for Sp1 footprinting), whereas in our experiments the E site was fully protected by ordinary amounts (40 μg) of crude extract. Thus the E site appears to interact with a novel abundant protein whose binding specificity is similar to that of E2F but distinct from it. Southwestern binding assay showed a 30–33 KDa protein doublet. Deletion of the E site had a mild effect on transcription in both orientations. Finally, one more site (G site) was identified, which also coincided with one major divergent RNA origin and consisted of a tandem duplication of the CCCTGA motif. No obvious similarity was found between that protection and known cis-active sequences, with the possible exception of certain AP2 sites. However, the G site did not compete with AP2-binding sites. In addition, AP2 cannot be renatured or visualised in Southwestern assays (15) whereas our experiments showed a 53 KDa protein binding to a probe containing the CCCTGA repeat. Thus the G site appears to interact with a previously unidentified binding activity. We believe that this novel activity is relevant for transcription from HTF9, since a region containing only the G site and a neighbouring Sp1 site is sufficient for transcription in both directions. Most Sp1-responsive promoters do not rely on the activating effect of Sp1 only, but rather on a combination of at least two factors, usually Sp1 and TFIID in most TATA box-containing genes, or Sp1 and the positioning HIP1 protein in the TATA-less DHFR gene (18). It has not been possible as yet to directly assess the role of the 53 KDa protein by generating further deletions around the G site, since that would have removed one or more RNA origins (see summary in Fig. 8). However, cotransfection experiments in which the intracellular Sp1was bound by a synthetic competitor construct *in vivo* suggest that the G site indeed plays a major role in HTF9 transcription (data not shown). It may be of interest that the motif CCCTGA is also included in one protected region (footprint 6) of the CG-rich HMGCoA reductase promoter, and is part of an element which is indispensable for HMGCoA reductase transcription (36). Finally, the proximal region of HTF9 also includes a CCAAT box contributing less than 25% of the A promoter efficiency, a weak effect compared to the role played by the CCAAT box upstream of several genes (3 and refs. therein). The moderate effect of the deletion is consistent with the lack of protein binding with the extracts used here. A detailed study of the CCAAT box and of its binding properties will be reported elsewhere (M.P.S. et al., in preparation).

The results of the protein-binding and deletion mapping experiments are summarized in Fig. 8. A short (85 bp) region around the RNA origins contained the sequences sufficient for transcription in both directions. Flanking elements contributed to, but were not essential for, the overall promoter efficiency in either orientation. It is intriguing that the protein-binding experiments identified a multiplicity of elements over 500 bp of HTF9, whereas the basal bidirectional promoter was associated with a region retaining only an Sp1 site and the G site. Transient assays may fail to identify elements that are required *in vivo*, for example for controlling the unfolding of the expressing gene (37). However, Antequera et al. (10) analysed the chromatin accessibility over 14 kb surrounding the HTF9 island. The region identified as accessible to nucleases *in vivo* corresponds to the region that we have scanned for protein binding and assayed in expression constructs.

From our data, and from the data available on other housekeeping promoters, a pattern has begun to emerge. Dissection of the bidirectional DHFR promoter revealed redundant regulatory elements, organised in a minor (distal) and a major (proximal) promoter, each capable of functioning in mutual independence (38). However, the DHFR region sufficient for minimal promoter activity is limited to a 80 bp-long fragment retaining only one out of four Sp1 sites and one site for the positioning HIP1 factor (18). The hamster HMGCoA reductase promoter contains several elements distributed over 500 bp (39), yet only two sites (footprints 4 and 6) included in an 80 bp promoter fragment are required for functioning (36). Similarly, the human hprt promoter activity is confined to a 40 bp fragment and can tolerate extensive deletions removing most of the surrounding CG-rich sequences (40). Finally, three short elements are sufficient for accurate transcription of the Surf-1 and Surf-2 genes from a common CpG-island (34). Together these data suggest that despite the large size of CpG-islands, and despite the presence of multiple potential factor-binding sites, housekeeping promoters may have a simple functional organization. Only a small subset of elements appear to be required for expression in any given cell type. The redundancy of elements may allow the flexibility for the promoter to adapt to different cellular backgrounds.

ACKNOWLEDGEMENTS

We thank A.Felsani (Istituto di Tecnologie Biomediche del CNR, Roma) for the gift of the Sp1 binding oligonucleotide, M.Caruso (Istituto di Biologia Cellulare del CNR, Roma) for the gift of the polyoma enhancer, and C.Tyndall, F.Watt, P.Molloy and M.Frommer (CSIRO Division of Biotechnology, Sydney) for communicating their unpublished data. We also thank A.Bird (Insitute of Moelcular and Cell Biology, University of Edinburgh), S.Fiering and R.Lathe (AFRC Centre for Genome Research, Edinburgh) for their critical comments on this manuscript. This work was supported by Consiglio Nazionale delle Ricerche, by Progetto Finalizzato Ingegneria Genetica and by Fondazione Cenci Bolognetti.

REFERENCES

1. Maniatis, T., Goodbourn, S. and Fischer, J. (1987) Science, **236**, 1237–1244.
2. Dynan, W.S. (1989) Cell, **58**, 1–4.
3. Mitchell, P.J. and Tjian, R. (1989) Science, **245**, 371–378
4. Hastie N.D. and Bishop J.O. (1976) Cell, **9**, 761–774
5. Dynan, W.S. (1986) Trends Genet., **6**, 196–197.
6. Melton, D. (1987) J. Cell. Sci., **88**, 267–270.
7. Bird A.P. (1986) Nature, **321**, 209–213
8. Bird, A.P. (1987) Trends Genet., **3**, 342–347.
9. Keshet, I., Lieman-Hurwitz, J. and Cedar, H. (1986) Cell, **44**, 535–543.
10. Antequera, F., Macleod, D. and Bird, A.P. (1989) Cell, **58**, 509–517.
11. Tazi J. and Bird A. (1989) Cell,
12. Dynan, W.S., Sazer, S., Tijan, R. and Schimke, R.T. (1986) Nature, **319**, 246–248.
13. Dush, M.K., Briggs, M.R., Royce, M.E., Scaff, D.A., Kahn, S.A., Tischfield, J.A. and Stambrook P.J. (1988) Nucl. Acids Res., **16**, 8509–8524.
14. Mitchell, P.J., Wang, C. and Tjian, R. (1987) Cell, **50**, 847–861.
15. Lee, W., Haslinger, A., Karin, M. and Tijan, R. (1987) Nature, **325**, 368–372.
16. Imagawa, M., Chiu, R. and Karin, M. (1987) Cell, **51**, 251–260.
17. Blake, M.C. and Azizkhan, J.C. (1989) Mol. Cell. Biol., **9**, 4994–5002.
18. Means A.L. and Farnham P.J. (1990) Mol.Cell.Biol., **10**, 653–661
19. Atchinson, M.L., Meyuhas, O. and Perry R.P. (1989) Mol. Cell. Biol., **9**, 2067–2074.
20. Chung, S. and Perry, R. (1989) Mol. Cell. Biol., **9**, 2075–2082.
21. Hariharan U., Kelley D.E. and Perry R. (1989) Genes Dev., **3**, 1789–1800
22. Hariharan U. and Perry R. (1989) Nucl. Acids Res., **17**, 5323–5337.
23. Yang, T.P., Singer-Sam, J.C., Flores, J. and Riggs, A.D. (1988) Som. Cell. Mol. Genet., **14**, 461–472.
24. Pfeifer G., Tanguay R.L., Steigerwald S., and Riggs A.D. (1990) Genes Dev., **4**, 1277–1287
25. Bird, A.P., Taggart, M., Frommer,M., Miller, O.J. and Macleod, D. (1985) Cell, **40**, 91–99.
26. Lavia, P., Macleod, D. and Bird, A. (1987) EMBO J., **6**, 2773- 2779.
27. Bressan A., Somma M. P., Lewis J., Santolamazza C., Copeland N., Gilbert D., Jenkins N. and Lavia P. (1991) Gene, in press.
28. Dignam, J.D., Lebovitz, R.M. and Roeder,R.G. (1983) Nucl. Acids Res., **11**, 1475–1489.
29. Mc Grogan M., Simonsen C., Smouse D., Farnham P. and Schimke R. (1985) J. Biol. Chem., **260**, 2307–2314.
30. Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H. (1982) Proc. Natl. Acad. Sci. USA, **79**, 6777–6781.
31. Laimins, L.A., Khoury, G., Gorman, C., Howard B. and Gruss, P. (1982) Proc. Natl.Acad. Sci. USA, **79**, 6453- 6457.
32. Miyamoto, N.G., Moncollin, V., Egly, J.M. and Chambon, P. (1985) EMBO J., **4**, 3563- 3570.
33. Kadonaga, J.T., Jones, K.A. and Tijan R. (1986) Trends Biochem., **11**, 20–23.
34. Lennard A. and Fried M. (1991) Mol. Cell. Biol., **11**, 1281–1294
35. Williams T.J. and Fried M. (1986) Mol. Cell Biol., **6**, 4558–4569.
36. Osborne T., Gil G., Brown M.S., Kowal R. and Goldstein J. (1987) Proc. Natl. Acad. Sci. USA, **84**, 3614–3618
37. Aronov B., Lattier D., Silbiger R., Dusing M., Hutton J., Jones G., Stock J., McNeish J., Potter S., Witte D. and Wiginton D. (1989) Genes Dev., **3**, 1384–1400
38. Linton, J.P., Yen, J., Selby, E., Chen, Z., Chinsky, J.M., Lin, K., Kellems, R.E. and Crouse, G.F. (1989) Mol. Cell. Biol., **9**, 3058–3072.
39. Osborne T., Goldstein J. and Brown M.S. (1985) Cell, **42**, 203–212
40. Melton D.W., McEwan C., McKie A. and Reid A.M. (1986) Cell, **44**, 319–328