Title
Isolation and characterization of the QM promoter.

Journal
Nucleic Acids Research, 24(11)

ISSN
0305-1048

Authors
Farmer, Andrew A.
Johnsen, John I.
Loftus, Thomas M.
et al.

Publication Date
1996-06-01

2158–2165 *Nucleic Acids Research, 1996, Vol. 24, No. 11*

Isolation and characterization of the *QM* promoter

Andrew A. Farmer, John I. Johnsen, Thomas M. Loftus, Kelly P. Smith and Eric J. Stanbridge*

Department of Microbiology and Molecular Genetics, University of California, Irvine, College of Medicine, Irvine, CA 92715, USA

Received January 11, 1996; Revised and Accepted April 9, 1996

GenBank accession no. U37218

ABSTRACT

This report describes the isolation, sequencing and preliminary characterization of the first 1 kb of the 5′-regulatory region of the human *QM* gene. This region and the 5′-half of the transcribed region of the *QM* gene are enriched for C and G nucleotides with no bias against CpG dinucleotides—indicative of a CpG island. Several consensus GC boxes are present within the sequence. Most are clustered at the distal end, with one site present in the proximal 200 bp of the promoter. Electrophoretic mobility shift experiments and luciferase assays done in insect cells transfected with an Sp1 expression construct suggest that most of these sites can bind Sp1 or a closely related factor. In addition, the promoter is shown to be responsive to cAMP via a response element (CRE) in the proximal promoter. Studies with 5′-end and internal deletion mutants suggest that elements in the distal promoter exert their positive effect through interactions with a proximal element(s). Candidate proximal elements include the proximal GC box and a 43 bp region between a *Kpn*I site (at –182) and a *Sma*I site (at –139).

INTRODUCTION

The *QM* gene was first identified by subtractive hybridization as a gene for which increased expression correlated with the non-tumorigenic phenotype of a Wilms' tumor microcell hybrid (1). Southern analysis showed *QM* to be a member of a large, multigene family in mammals (1). *QM* is the only member of this family that is known to be expressed. Other family members so far isolated appear to have arisen by retrotransposition events and may be pseudogenes.

The QM protein is a highly basic, 25 kDa protein that shows no significant similarity to any other known human proteins. However, QM homologs have been isolated from a diverse array of other species, encompassing not only the animal, but also the plant and fungal kingdoms (2,3). Amongst these homologs there has been striking conservation, such that even yeast (*Saccharomyces cerevisiae*) and human QM are 70% identical at the amino acid level (2). Much of this conservation is within extensive domains of similarity that stretch over the first 170 residues of the protein (2).

The function of QM remains obscure. It has been reported that QM can bind to c-Jun *in vitro* and repress c-Jun-mediated transcriptional activation, suggesting that QM may be a novel transcription factor (4). However, as yet, no *in vivo* data, such as co-immunoprecipitation of c-Jun and QM from cell extracts, has been presented in support of the observed *in vitro* association. Moreover, subfractionation studies done in this laboratory suggest that QM is located on the membrane of the endoplasmic reticulum and not in the nucleus (T.M.Loftus and E.J.Stanbridge, unpublished results). The yeast homolog of QM has also been localized to the cytosolic compartment by immunocytochemistry (5). Thus, the apparent association of c-Jun and QM may not be physiological, although the formal possibility that QM translocates to the nucleus under certain conditions, or in quantities undetectable by the methods used, cannot be excluded.

Whatever the true role of QM, its normal function is critical to eukaryotes. This is suggested both by its extreme conservation across the animal, plant and fungal kingdoms (2) and by the finding that deletion of the yeast homolog of *QM* is lethal in yeast (5). Not surprisingly, *QM* appears to be expressed in all mammalian tissues examined (1,6). However, the level of *QM* expression shows considerable variation between different tissues, as well as within tissues at different stages of development (1,6). In particular, there appears to be an inverse correlation between the level of *QM* expression and developmental stage. For example, Northern analysis of normal mouse tissues from different stages of development revealed a decrease in *QM* expression in heart, kidney, liver and skin between embryonic and adult stages (1). Moreover, the mouse homolog of *QM* was isolated by subtractive hybridization as a gene whose expression decreases 70% upon differentiation of pre-adipocytes into mature adipocytes (7,3). A similar reduction in *QM* expression was also observed in differentiating rat adipocytes (3). In plants also, decreased expression of *QM* is associated with differentiation into adult tissues (8). Taken together, these findings suggest that *QM* may have a function in cell growth or differentiation. However, it remains to be demonstrated whether the changes in *QM* expression are a cause or consequence of the changes in cell growth/differentiation. As a first step toward a better understanding of the possible mechanisms for regulating QM expression we report here on the isolation, sequencing and preliminary characterization of the 5′-regulatory region of human *QM*.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Generation of wild-type and mutant *QM* promoter reporter constructs

A luciferase reporter construct containing ~1 kb of sequence upstream of the *QM* transcription start site was generated in two steps. First, a fragment extending from an *Xho*I site 995 bases upstream of the transcription start site to a *Dpn*I site in the second exon was cloned into pGL2basic (Promega) at the *Xho*I and blunted *Hin*dIII sites. This was done as a three-part ligation using an *Xho*I–*Aat*II fragment and an *Aat*II–*Dpn*I fragment. Ligation of the *Dpn*I half-site to the filled-in *Hin*dIII site recreates a *Hin*dIII site. Second, the construct generated in the first step (pGL2QM1) was opened with *Hin*dIII and *Aat*II, and the excised fragment further digested to generate an *Aat*II–*Fsp*I fragment. This fragment was then religated into the opened pGL2QM1 together with an oligomer encoding from the *Fsp*I site to the fifth base of the first exon of QM, followed by a *Hin*dIII site. The sequence of this oligomer is: sense strand, 5′-GCAGGCGGAGGAGCGCCTCTTA; antisense strand, 5′-AGCTTAAGAGGCGCTCCTCCGCCTGC. 5′-End and internal deletion mutants were generated from this construct by digestion with various restriction enzymes followed by reclosure of the vector. (see Fig. 3).

Sequencing

Sequencing was done by the chain termination method (9) using Sequenase 1.0® (US Biochemical). The primers used were as follows:

GL1, 5′-TGTATCTTATGGTACTGTAACTG;
GL2, 5′-CTTTATGTTTTTGGCGTCTTCCA;
QM35, 5′-GTAAGAACCATAGAGTCCTGT;
QM2-53, 5′-AGCACAGTGGAGTGGGAA;
QM2-35, 5′-TTCCCACTCCACTGTGCT;
QM3-35, 5′-CGTTAACTGTGACAGACGTA;
QM3-53, 5′-CTCTCAGAAATATACGTCTG;
QM4-35, 5′-GAGAAATCTCCACGGAGA;
QM5-53, 5′-CGGGTTGACAAAGGAACG.

Primers GL1 and GL2 are specific for the flanking vector sequences (pGL2Basic). The remaining primers are *QM* specific.

Cell culture conditions

NIH3T3 cells were maintained in DMEM with 10% bovine calf serum (Hyclone). SL2 (Schneider line 2) insect cells were cultured in Shields and Sang M3 insect medium (Sigma) supplemented with 10% inactivated bovine calf serum.

Transient transfection and luciferase assays

For transfection, 4×10^5 NIH3T3 cells or 5×10^6 SL2 cells were plated in a 60 mm diameter dish and allowed to grow overnight. For each experiment, equimolar quantities of the various promoter constructs were used and the total mass of DNA was standardized between transfections using pGL2basic DNA. For the insect cell experiments, promoter constructs were transfected either in the presence or absence of the Sp1 expression vector pPacSp1 (10). The pGL2promoter vector (Promega), which contains six Sp1 binding sites from the SV40 promoter, served as a positive control for Sp1 activity. All experiments were done in triplicate. The DNA was transfected by calcium phosphate precipitation, as previously described (11). After two days, the cells were washed with PBS and lysed in 300 µl of luciferase reporter lysis buffer (Promega). The lysate was clarified by centrifugation, and the supernatant transferred to a new tube. The nuclear pellet was extracted for DNA and used to standardize transfection efficiency. Luciferase activity of the supernatant was assayed using Promega's luciferase assay system and a Monolight 2010 luminometer (Analytical Luminescence Laboratories).

Standardization of transfection efficiencies

NIH3T3 transfection efficiencies were standardized to the quantity of luciferase DNA present in the nuclei of the transfected cells. To do this, the concentration of DNA in one sample was determined and the volume containing 5 µg DNA calculated. This volume was then taken from each sample and transferred to a nylon membrane (Nytran; Schleicher and Schuell) using a slot blot apparatus (Schleicher and Schuell). The membrane was then probed using a 2.5 kb (*Hin*dIII–*Cla*I) luciferase fragment from pGL2basic that had been random prime labeled (12). After overnight hybridization at 65°C, the blot was washed (once in 2× SSC, 0.2% SDS for 15 min at 25°C, once in 0.2× SSC, 0.2% SDS for 30 min at 25°C and once in 0.2× SSC, 0.2% SDS for 30 min at 65°C) and exposed to X-ray film (X-Omat; Kodak). The resulting autoradiogram was scanned onto a Macintosh 660AV computer using a Hewlett Packard flatbed color scanner and densitometry was done on the image using the program NIH Image 1.5f.

For the insect cell experiments, luciferase activity was standardized to the quantity of total cellular protein in the luciferase extracts.

Treatment with cyclic AMP

Dibutyryl cyclic AMP (DBcAMP; Sigma) was solvated in 100% dimethyl sulfoxide (DMSO) and stored at –20°C. The drug was added to the cells at doses of 0.1–1.0 mM 36 h post-transfection. Cells were harvested 12 h later and assayed for luciferase activity as described above. Cells treated with 0.4% DMSO (the final concentration of DMSO in the DBcAMP treatments) acted as a control.

Electrophoretic mobility shift assays (EMSAs)

EMSAs were done using the following oligomers and restriction fragments of the *QM* promoter as probes:

GC box consensus oligomer, 5′-attcgatcGGGGCGGGGCgagc (Santa Cruz);
GC box mutant oligomer, 5′-attcgatcGG<u>TT</u>CGGGGCgagc (Santa Cruz);
CRE consensus oligomer, 5′-agagattgccTGACGTCAgagagctag (Santa Cruz);
CRE mutant oligomer, 5′-agagattgccTG<u>TG</u>GTCAgagtag (Santa Cruz);
QM wild-type CRE, 5′-tatggtcaTGACGTCTgacagagc;
QM mutant CRE, 5′-tatggtcaTG<u>TG</u>GTCTgacagagc.

For these oligomers, only the sense strand is shown. Upper case nucleotides represent the binding site and underlined bases depict mutations. The *QM* promoter restriction fragments containing potential GC boxes were *Pst*I (at –925)–*Bgl*I (at –844), *Fsp*I (at –866)–*Pvu*II (at –702) and *Pvu*II (at –203)–*Sma*I (at –139). The positions of these fragments are detailed in Figure 4A. All probes were phosphorylated with [γ-^{32}P]ATP using T4 polynucleotide kinase (NEB) and purified over Sephadex G-50 columns. Restriction fragment probes were dephosphorylated using calf intestinal alkaline phosphatase prior to being labeled. For each reaction, 6.5 pmol probe was labeled using 70 µCi [γ-^{32}P]ATP (>5000 Ci/mmol) in a reaction volume of 15 µl. Binding reactions were done by mixing 5 µg HeLa nuclear extract (Promega) with ~20

000–40 000 c.p.m. probe in a mixture containing 10 mM Tris, pH 7.5, 50 mM NaCl, 1 mM DTT, 1 mM Na_2EDTA, 5% glycerol and 1 µg poly(dI·dC). The reaction volume was 20µl. Reactions were incubated at room temperature for 15 min and then electrophoresed through a 6%, 29:1 acrylamide:bis polyacrylamide gel in 1× TGE (25 mM Tris base, 190 mM glycine, 1 mM Na_2EDTA) at room temperature, with water cooling, for ~2 h at a constant 200 V. After electrophoresis, the gel was dried and autoradiographed. For competition experiments the extracts were first incubated with excess competitor DNA for 15 min. The probe was then added, and the mixture incubated a further 15 min prior to electrophoresis. For super-shift experiments with anti-Sp1 antiserum (1C6; Santa Cruz), 1 µg antiserum was mixed with the extract for 15 min at room temperature, prior to adding the probe.

Sequence analysis

Analysis of the *QM* 5′-flanking region and transcribed sequence for G/C content and dinucleotide composition was done using the program COMPOSITION (13). Sequence comparisons were done using BLAST (14). Analysis of the 5′-regulatory region for potential transcription factor binding sites was done using DNA Strider 1.2 for the Macintosh together with a database of transcription factor binding sites compiled by H.Mangalam (Department of Microbiology and Molecular Genetics, University of California, Irvine).

RESULTS

Isolation and sequencing of the *QM* 5′-regulatory region

Several groups have mapped the human *QM* gene to Xq28 (6,15,16). In particular, Bione *et al.* mapped *QM* to a single cosmid within a 450 kb contig stretching from the *G6PD* locus to the color vision genes (16). To clone the *QM* promoter from this cosmid, a probe specific for the 5′-end of genomic *QM* was generated by PCR using primers that lie within the first and fifth introns of the gene (DC1, 5′-TAGGTCTGTTCTCGTCTTG and DC2, 5′-AATGTAGAGACTCCAACTGC). The amplified fragment was 'TA'-cloned into pCRII (Invitrogen) and an ~500 bp fragment encoding from intron 1 through to the *Not*I site in the second intron was isolated by digestion with *Not*I and *Eco*RI. This fragment was used to probe restriction digests of the cosmid. In this way, a *Not*I–*Eco*RI fragment that extended ~6 kb 5′ of the *QM* transcription start site was identified and cloned into $pBSK^+$ (Stratagene). Subsequently, a 1 kb fragment encoding from nucleotides –995 to +5 relative to the transcription start site was cloned into the firefly luciferase reporter vector pGL2Basic (Promega). This fragment was sequenced in both directions, with the exception of the region 3′ of the CRE (Fig. 1A). The sequence of this segment has been published previously (15) and we found no differences between the published sequence and that described here.

Sequence analysis

A striking feature of the sequence shown in Figure 1A is its high G/C content, which averages 65% over its length (Fig. 2A). Within the gene itself the percentage G/C falls, though at 52%, it continues to remain above the 40% average G/C content for the human genome (17). In addition to the high G/C content, the usual bias against CpG dinucleotides over GpC is not seen (Fig. 2B).

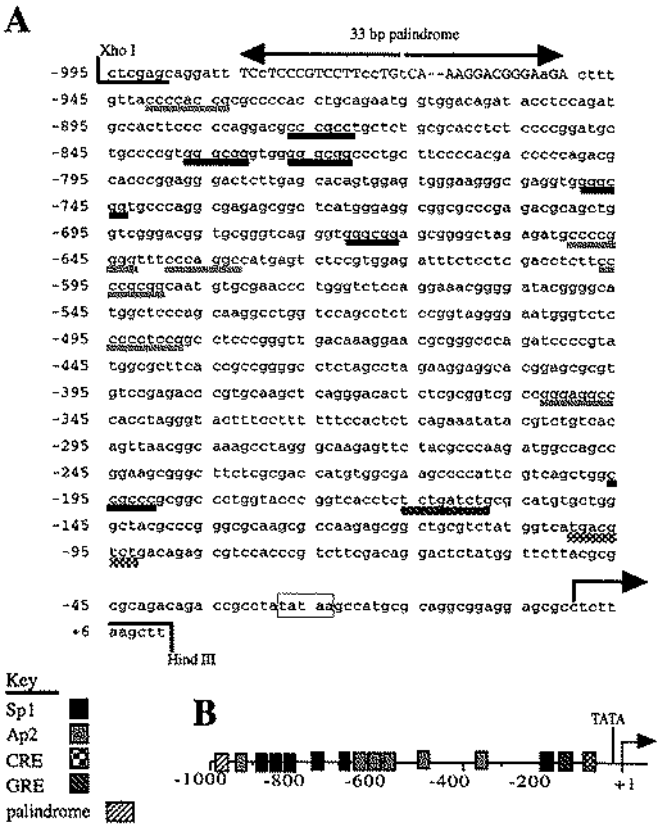

Figure 1. Sequence of the first 1000 bp of the *QM* 5′-regulatory region. (**A**) The sequence of the *QM* 5′-regulatory region from an *Xho*I site at position –995 through to an engineered *Hind*III site following the fifth base of the first exon is shown. Putative transcription factor binding sites are underlined according to a scheme given in the key. The 33 bp palindromic sequence is also noted. Upper case letters are used within this sequence to highlight bases that have direct counterparts in the palindrome. The two dashed lines were inserted within the palindrome to facilitate its alignment. The TATA sequence is boxed and the transcription start site is given by an arrow. (**B**) A schematic representation of the *QM* 5′-regulatory region is shown. Putative transcription factor binding sites are noted as in the key.

In higher vertebrates, the ratio CpG:GpC is typically ~0.25 (18). In contrast, this ratio is close to 0.8 within the QM promoter and rises further to 1.2 over exons 1–2 (Fig. 2B). Although G/C content falls after exon 2, the CpG:GpC ratio remains high (~0.8) through exon 4, where it returns abruptly to the more common figure of ~0.2. The high CpG:GpC ratio of the *QM* promoter region was not unexpected, since a cluster of rare restriction enzyme sites had been mapped to the 5′-end of the gene (15,16). Indeed, the cosmid contig that contains the *QM* sequence was developed by looking specifically for sequences on the X chromosome that contain CpG islands, as defined by both restriction mapping and the absence of methylation on CpG dinucleotides (16).

The promoter sequence was screened for potential transcription factor binding sites. This revealed several putative consensus GC boxes and AP-2 sites, expected in a G/C-rich sequence. Two putative GC boxes were also found in the G/C-rich region of the transcribed sequence. In addition, a consensus cAMP response element (CRE) and a putative glucocorticoid receptor site were identified. There is a canonical TATA box at position –28. However, no CCAAT box is present; the CCAAT box is a common regulatory element in TATA

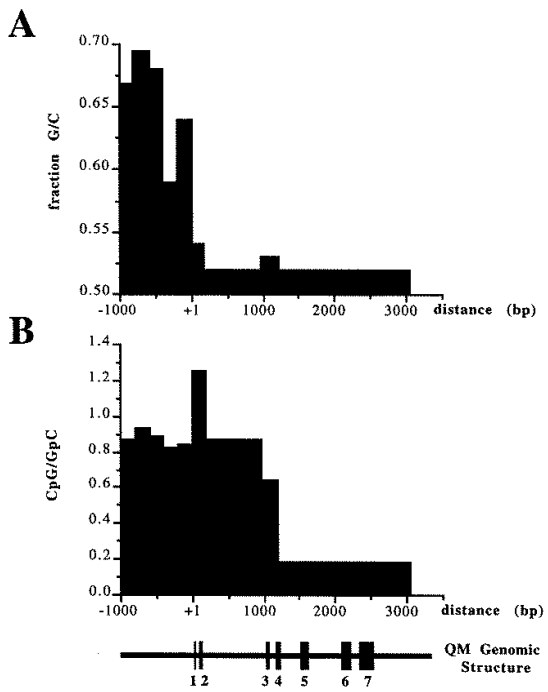

Figure 2. Variation in G/C content and CpG:GpC ratio over the 5′-regulatory and transcribed regions of the *QM* gene. (**A**) A histogram depicting the variation in G/C content across the 5′-regulatory and transcribed regions of the *QM* gene. (**B**) A histogram showing the ratio of CpG:GpC across the 5′-regulatory and transcribed regions of the *QM* gene. For orientation, a schematic representation of the *QM* genomic region is given below the histograms. In this, numbered solid bars denote exons.

box-containing promoters. Finally, a 33 bp palindromic sequence was found at the very 5′-end of the region sequenced. Neither this element, nor the promoter as a whole, share any strong similarity to other known promoter sequences, as determined by searching the combined Genbank and EMBL databases using BLAST (14) as detailed in Materials and Methods. The positions of all the putative sites are detailed in Figure 1A and B.

Serial deletion analysis of the *QM* promoter

To assess the contribution of various regions of the promoter to its overall activity, a series of 5′-end and internal deletion mutants was generated using appropriate restriction enzymes. These constructs (shown in Figure 3) were transfected into NIH3T3 cells, and promoter activity was measured as described in Materials and Methods. The results of this analysis are tabulated in Figure 3. Fully 60% of the activity seen in the 1000 bp promoter fragment is lost on deletion of the 5′-most 460 bases of the promoter up to the unique *Stu*I site (at –533). This region contains five putative consensus GC boxes, four putative AP-2 sites and has the highest G/C content (Figs 1A and 2A). The region containing the 33 bp palindrome and the most distal AP-2 site at the very 5′-end of the sequence analyzed is not responsible for this activity, since loss of this sequence alone (*Pst*I construct, Fig. 3) does not result in any loss of promoter activity. In fact, loss of this region gives a mild but significant increase in activity (>95% confidence by Student's *t*-test), suggesting that it may have a repressor function.

Promoter activity declines further (60%) to 14% activity with removal of sequences up to the *Kpn*I site at position –182. This region contains two further putative AP-2 binding sites and the remaining GC box. Activity then drops sharply to just 1–2% of maximal activity upon deletion up to the *Sma*I site (–139), just 43 bp less than the *Kpn*I fragment. This segment contains a putative glucocorticoid response element. Further reductions, ultimately down to the region surrounding the TATA box have little further effect. Interestingly, an internal deletion that removes ~200 bp between the *Avr*II site at –281 and the *Aat*II site at –99 generates a promoter with only 1% of the activity of the 1000 bp construct. While it should be borne in mind that this deletion does not retain the distal elements in the same positions relative to the TATA box as in the wild-type construct, it does suggest that the distal elements alone are not sufficient to provide promoter activity. Nonetheless, they do contribute considerably to it, as witnessed by the effect of deleting them (*Stu*I construct).

Specific binding to the GC boxes

As discussed above, there is considerable loss of promoter activity upon removal of the distal half of the promoter. This region is very G/C rich and, as noted in Figures 1 and 3, contains all but one of the putative GC boxes. GC boxes bind members of the Sp1 family of transcription factors. Thus, it seemed possible that Sp1, or a related factor, acting through these sites might be responsible for the transcriptional activity of this region. As a first step towards answering this question, we tested the ability of these sites to bind to such a factor. To do this, three restriction fragments were chosen that together covered all but one of the putative GC boxes. These were *Pst*I (at –925)–*Bgl*I (at –844), *Fsp*I (at –866)–*Pvu*II (at –702) and *Pvu*II (at –203)–*Sma*I (at –139). The positions of these fragments are outlined in Figure 4A. These fragments were gel purified and labeled with [γ-^{32}P]ATP as described in Materials and Methods and used in EMSAs to determine if Sp1 or an Sp1-like factor could specifically bind to them. Two typical autoradiograms are shown in Figure 4B and C. This shows that both the *Fsp*I–*Pvu*II and the *Sma*I–*Pvu*II fragments bind a factor that can be specifically competed by excess unlabeled consensus Sp1 oligomer, but not by an excess of a mutant Sp1 oligomer. No similar binding was found using the *Pst*I–*Bgl*I fragment (Fig. 4C). Moreover, under the conditions used, no similarly migrating species is seen using a labeled mutant Sp1 oligomer as a probe (Fig. 4B and C). These results suggest that at least some of the putative GC boxes within the *QM* promoter can bind Sp1 or an Sp1-like factor that may contribute to the activity of the *QM* promoter.

The effect of Sp1 expression on *QM* promoter activity

To further investigate the potential functional role of Sp1 in regulating activity of the *QM* promoter, the luciferase assays were repeated using SL2 insect cells, which lack endogenous Sp1. In these experiments, the various *QM* promoter constructs were transfected either in the presence or absence of an Sp1 expression vector. The results are tabulated in Figure 3, together with the results from the NIH3T3 experiments described above. In the absence of any exogenous Sp1, the activity of all the promoter constructs is minimal. In contrast, addition of Sp1 results in a pattern of promoter activity closely resembling that seen in the NIH3T3 assays. In particular, a large drop in activity is seen with removal of the distal half of the promoter up to the *Apa*I site, and both the *Sma*I fragment (lacking all GC boxes) and the *Avr*II–*Aat*II internal deletion construct (lacking the proximal GC box) have little activity either in the presence or absence of Sp1. Together

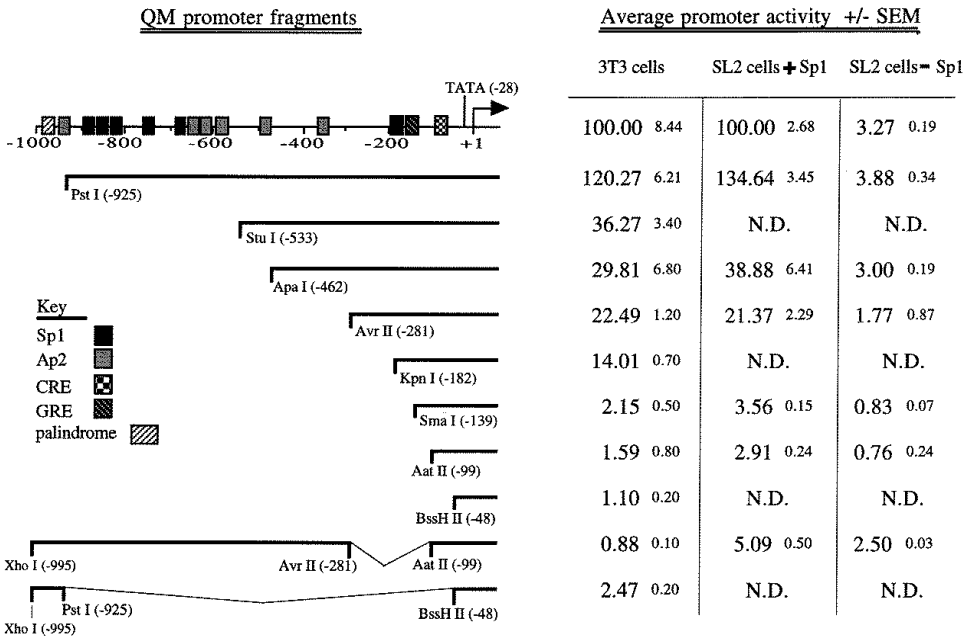

Figure 3. *QM* 5′-regulatory region deletion mutants and promoter activity in NIH3T3 and SL2 cells. The various 5′-end and internal deletion mutants generated from the 1000 bp promoter fragment are shown below a schematic of the region. Putative transcription factor binding sites and the position of the 33 bp palindrome are shown, as detailed in the key. Figures in parentheses give the distance in bases of the various restriction sites from the transcription start site. All sequences end following the fifth base of the first exon in the *Hin*dIII site of pGL2Basic, into which all the fragments were cloned. The activities of the various promoter constructs in NIH3T3 and SL2 cells, as a percentage of the luciferase activity obtained with the 1000 bp *Xho*I fragment, are tabulated to the right. For each construct, the average ± SEM activity derived from triplicate assays is shown. Assays in SL2 insect cells were either in the presence or absence of exogenous Sp1, as noted in the table. For the NIH3T3 experiments, transfection efficiencies were standardized between samples to the amount of luciferase DNA in each sample. For the SL2 experiments, transfection efficiencies were standardized to total cellular protein. N.D., not determined.

with the EMSA data, these data strongly suggest that the GC boxes identified are functional Sp1 binding sites and that synergy between the distal GC boxes and an element(s) in the proximal promoter is important for promoter function.

Interestingly, the *Pst*I fragment (lacking the palindrome sequence) still shows significantly greater activity than the *Xho*I fragment, as was seen in the NIH3T3 experiments.

QM transcription can be induced by cAMP

As mentioned above, there is a near-consensus CRE (5′-TGAC-GTCT-3′) in the proximal region of the *QM* promoter (position –99 with respect to the transcription start site). We have obtained evidence that *QM* transcription is responsive to cAMP through this site. As shown in Figure 5, the 1000 bp promoter fragment is responsive to the cAMP analog DBcAMP at doses of ≥0.5 mM. This effect is not seen with DMSO alone, used to solubilize the DBcAMP. The effect of DBcAMP (0.5 mM) was also tested on the various deletion mutants. As shown in Figure 6A and B, deletion up to the *Sma*I site (position –139) does not block the response to DBcAMP. However, deletions that remove the CRE [*Aat*II, Δ(*Avr*II–*Aat*II) and Δ*Aat*II] result in a loss of responsiveness (Fig. 6B and C). Significantly, the Δ*Aat*II construct (in which the CRE alone was destroyed by opening the promoter construct at the overlapping *Aat*II site, blunting by fill-in with Klenow fragment and reclosing of the vector) has the same basal activity as the wild-type promoter yet fails to respond to DBcAMP (Fig. 6C).

The functionality of the putative CRE was also assayed by EMSA. As shown in Figure 7, both a labeled consensus CRE oligomer and an oligomer encoding the wild-type *QM* site produce a similar band pattern with a HeLa cell nuclear extract. In both cases, the upper doublet can be competed by either excess consensus oligomer or excess wild-type *QM* oligomer. However, neither excess mutant consensus oligomer nor excess mutant *QM* oligomer can compete this binding. Moreover, the upper band of the doublet is not seen in shifts done using the mutant *QM* oligomer as the probe. Based on previous studies (19), we suspect that the upper band represents CREB/ATF dimers and the lower band CREB/ATF monomers, which bind to the CRE half-site with low affinity. The mutant *QM* CRE retains an intact half-site and thus should be able to bind CREB/ATF monomers. Together these data suggest that the *QM* CRE is able to bind members of CREB/ATF family and that activation of these via cAMP-dependent phosphorylation mediates the increase in *QM* transcription in response to cAMP.

DISCUSSION

The human *QM* gene has been mapped to a single cosmid within a 450 kb contig on Xq28 that was generated by probing an Xq28-specific cosmid library with probes for CpG islands (16). In all cases, the CpG island probes mapped to the 5′-ends of genes contained within the contig (16). For *QM*, the presence of a CpG island was inferred by the presence of a cluster of rare restriction enzyme sites 5′ of the gene (15,16). The sequencing data presented here shows that the *QM* promoter, as well as the first 1 kb of transcribed sequence, is G/C rich (averaging 65% G/C) and lacks the usual bias seen against CpG dinucleotides. This confirms the presence of a CpG island at the 5′-end of the *QM* gene.

It has been estimated that up to 50% of all human genes are associated with CpG islands (20). Most of these are housekeeping

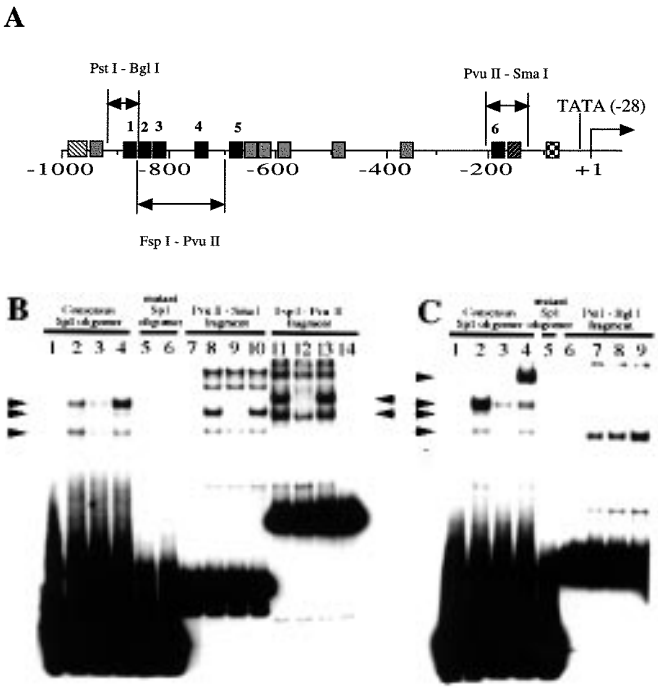

Figure 4. Binding to GC boxes in the *QM* promoter. EMS assays showing the binding of Sp1 or a related factor to regions of the *QM* promoter and to a consensus Sp1 oligomer. (**A**) A schematic of the *QM* promoter showing the positions of the GC boxes (numbered 1–6) and the three restriction fragments used in the EMSAs. (**B**) EMSA. Sp1 consensus oligomer (lanes 1–4); mutant Sp1 oligomer (lanes 5–6); *QM Pvu*II–*Sma*I fragment (lanes 7–10); *QM Fsp*I–*Pvu*II fragment (lanes 11–14). Free probe (lanes 1, 5, 7 and 14); probe + extract only (lanes 2, 6, 8 and 11); extract + 250× unlabelled consensus Sp1 oligomer (lanes 3, 9 and 12); extract + 250× unlabelled mutant Sp1 oligomer (lanes 4, 10 and 13). (**C**) EMSA. Sp1 consensus oligomer (lanes 1–4); mutant Sp1 oligomer (lane 5); *QM Pst*I–*Bgl*I fragment (lanes 6–9). Free probe (lanes 1 and 6); probe + extract only (lanes 2, 5 and 7); extract + 250× unlabelled consensus Sp1 oligomer (lanes 3 and 8); extract + 1 µg Sp1 antiserum (1C6; Santa Cruz) (lane 4); extract + 250× unlabelled mutant Sp1 oligomer (lane 9). The positions of GC box-specific shifts in (B) and (C) are denoted by the solid black arrow heads. In (C) the Sp1-specific supershift is marked by a gray arrowhead.

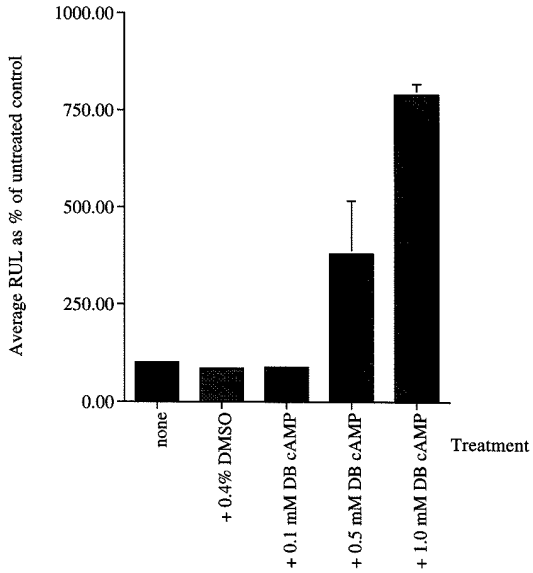

Figure 5. Dose–response characteristics of the 1000 bp 5′-regulatory fragment to DBcAMP. Histogram showing the response of the 1000 bp 5′-regulatory region to DBcAMP. Each treatment was assayed in triplicate and average activities are plotted. Error bars represent SEMs. Treatments with DBcAMP were for 12 h with 0.1, 0.5 or 1.0 mM DBcAMP.

genes, all of which have CpG islands (21). However, 40% of all sequenced tissue-restricted genes (e.g. α-globin) are also associated with CpG islands (17,22). Whereas bulk vertebrate DNA is highly methylated, CpG islands are almost never methylated at cytosines *in vivo* (except on the inactivated mammalian X chromosome and, in some instances, of imprinting; 21,23). Thus, CpG islands appear to be remnants of ancestral invertebrate DNA which, unlike vertebrate DNA, is mostly unmethylated at cytosine residues and shows no bias against CpG dinucleotides (23). It is the methylation of CpG within the vertebrate genome that accounts for the observed paucity, since 5-methylcytosine is prone to mutation via deamination to thymidine (24). Thus, methylated CpG dinucleotides become replaced over time by TpG and CpA dinucleotides. By remaining unmethylated, CpG islands escape this mutational loss and so are not biased against CpG. Why vertebrates should have methylated most of their DNA except for these short stretches that overlap the 5′-ends of many genes remains unclear. Their localization to the 5′-ends of genes is provocative. However, it is not clear whether transcription at these sites is the result or the cause of the lack of methylation. Thus, it remains to be explained why only some genes have islands and why the islands of tissue-specific genes (e.g. α-globin) remain unmethylated even in tissues where they are not expressed.

An interesting feature of many genes associated with CpG islands is the lack of TATA and CCAAT box sequences. The lack of a TATA box in these promoters results in a wide variation in the position of transcription initiation (25). The *QM* gene, however, retains a TATA box and initiates transcription from a single site (1). Likewise, the promoter for triose phosphate isomerase (TPI) is also G/C rich, but has both TATA and CCAAT boxes and initiates transcription from a single site (26). The mechanisms that enable some G/C-rich promoters to function in the absence of a TATA box remain to be defined. The importance of GC content (27) and the presence of long palindromic sequences in the region of transcript initiation (28) and polypurine/polypyrimidine tracts (29) have all been proposed. It would be interesting to determine if the TATA boxes in G/C-rich promoters such as those for *QM* and *TPI* are functionally redundant or still required for efficient transcription of these genes. If the TATA box is required, then comparison of the difference between TATA-less and TATA-containing G/C-rich promoters may yield insights into the mechanism of TATA-independent transcription in G/C-rich promoters.

Consistent with its high G/C content, there are multiple putative GC boxes within the *QM* promoter. These are binding sites for members of the Sp1 family of transcription factors. In the *QM* promoter, most of these sites are clustered in the 5′-half of the sequence, which is the most G/C rich. The clustering of distal GC boxes is quite common (30). In addition, a single GC box is often found in the proximal promoter (30), as is seen here for *QM* (Fig. 1). Synergism between adjacent Sp1 sites, as well as between distal and proximal sites, can occur through the formation of multimeric Sp1 complexes, dependent on activation domains in Sp1 (30). This may be occurring in the *QM* promoter. Deletion analysis shows that the region containing the distal GC boxes accounts for 60% of basal transcription activity. Moreover, an internal deletion

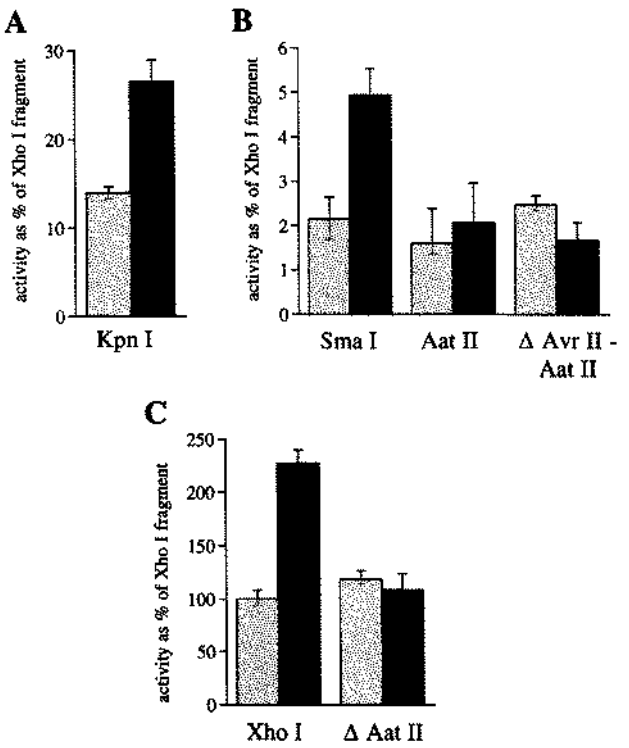

Figure 6. Effect of DBcAMP on the activity of various deletion mutants. Histogram showing the response of various deletion mutants of the *QM* 5′-regulatory region to 0.5 mM DBcAMP for 12 h. Each bar represents an average of triplicate assays. Error bars represent SEMs. Activities are given as percentage activity relative to the 1000 bp fragment in the absence of DBcAMP. (**A**) *Kpn*I fragment ± DBcAMP. (**B**) *Sma*I, *Aat*II and Δ(*Avr*II–*Aat*II) mutants ± DBcAMP. (**C**) Full-length (*Xho*I) fragment and Δ*Aat*II mutant ± DBcAMP. For each experiment, gray bars represent activity of the untreated control and black bars represent the activity following DBcAMP treatment.

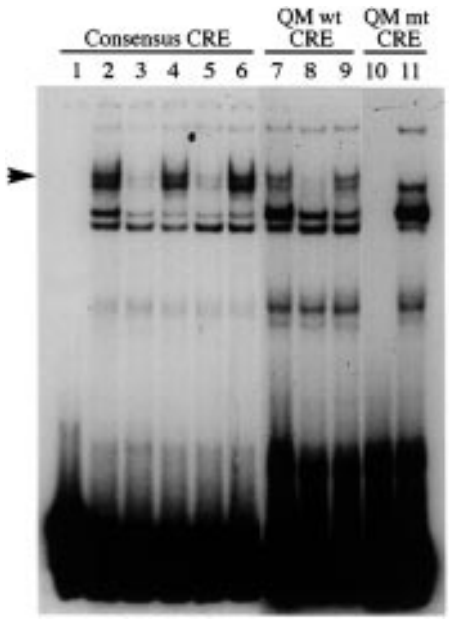

Figure 7. Binding of CREB to the CRE in the *QM* promoter. EMSA showing the binding of CREB to the CRE in the *QM* promoter and to a consensus CRE oligomer. Consensus CRE oligomer (lanes 1–6); wild-type *QM* CRE (lanes 7–9); mutant *QM* CRE (lanes 10 and 11). Free probe (lanes 1 and 10); probe + extract only (lanes 2, 7 and 11); extract + 250× unlabelled consensus CRE oligomer (lanes 3 and 8); extract + 250× unlabelled mutant CRE oligomer (lanes 4 and 9); extract + 250× unlabelled wild-type *QM* CRE oligomer (lane 5); extract + 250× unlabelled mutant *QM* CRE oligomer (lane 6) The dimeric CREB complex is denoted by a solid black arrow head. The lower most pair of bands appear to be non-specific.

[Δ(*Avr*II–*Aat*II) construct] that removes proximal sequences, including the proximal GC box, results in almost no activity from the promoter, even though the distal GC boxes remain. This was the case in both NIH3T3 and SL2 cells transiently transfected with an Sp1 expression construct. One caveat, however, is that this deletion alters the spacing between the distal elements and the TATA box. It is possible that other elements may also be involved in the apparent synergy between proximal and distal elements. In particular, the 43 bp region between the *Kpn*I site at –182 and the *Sma*I site at –139 seems to be important. This region is also lost in the Δ(*Avr*II–*Aat*II) construct. Moreover, while deletion up to the *Kpn*I site removes all GC boxes, the promoter still retains some 15% of its full activity. Yet, this is all lost when the promoter is further deleted up to the *Sma*I site (see Fig. 3). Finer deletion analysis, site-directed mutagenesis and EMSAs should further resolve the functions and interactions of these elements.

Additional support for the functional activity of at least some of the GC boxes has come from gel shift analysis using various restriction fragments of the *QM* promoter (presented in Fig. 4) and from luciferase assays done on the promoter fragments transfected into insect cells in the presence or absence of Sp1 (Fig. 3). The EMSA studies revealed GC box-specific binding both to the proximal GC box (box 6, see Fig. 4A), present on a *Pvu*II–*Sma*I fragment, and to an *Fsp*I–*Pvu*II fragment covering GC boxes 2–4. A *Pst*I–*Bgl*I fragment containing the most distal GC box (box 1, see Fig. 4A) did not show GC box-specific binding. Thus, this site may not be functional. From the present data, it is not possible to determine which GC box(es) in the *Fsp*I–*Pvu*II fragment is responsible for the binding seen with this fragment. Interestingly, all the *QM* promoter GC boxes, except the non-functional distal-most element, share the same core sequence (5′-GGGCGG-3′). This sequence, recognized by the second and third zinc fingers of Sp1 and related proteins, is the most critical determinant of binding (31). Thus, it is possible that all of these sites (2–6) are functional. Work ongoing in the laboratory will address this question. Although our gel shift competition experiments demonstrate specific binding to the *QM* GC boxes, we were unable to identify the factor binding to these sequences in the gel shifts. However, the results of the promoter activity experiments obtained using insect cells not only confirm the functional role of the distal and proximal GC boxes in regulating *QM* transcription, but also support the argument that Sp1 is able to transactivate *QM* promoter activity through these sites.

The *QM* promoter was found to be responsive to cAMP. Sequence analysis revealed the presence of a near-consensus CRE in the proximal promoter (5′-TGACGTCT-3′), as well as five putative AP-2 sites. Responses to cAMP can occur through both types of sites (32). In the case of the *QM* promoter, only the CRE appears to respond to cAMP, since deletion of this element, alone, abolishes the cAMP response. In contrast, 5′-end deletions of the promoter that remove the AP-2 sites but which retain the CRE do not affect cAMP responsiveness. Gel shift data further support the functionality of the *QM* CRE. The wild-type *QM* CRE generates an identical pattern of shifted complexes with nuclear

extract as an oligomer encoding a consensus CRE, whereas an oligomer encoding a mutant *QM* CRE does not show this pattern of binding. Moreover, the wild-type *QM* sequence can compete with the consensus CRE for binding of the CRE-specific factors, but the mutant sequence cannot. We have also found that the wild-type, but not the mutant, *QM* CRE can bind to the recombinant DNA binding/dimerization domain of CREB (Santa Cruz) (data not shown). It is interesting that the AP-2 sites do not appear to function in the cAMP response of the *QM* promoter. There are at least two other promoters that contain AP-2 sites and a CRE in which the cAMP response is found to occur only through the CRE (33,34). In contrast, there are also reports of promoters that have consensus CRE sequences that bind CREB in gel shift experiments but which do not respond to cAMP (34). Sometimes, this is cell type dependent, e.g. the rat cytochrome c promoter has a consensus CRE that functions in NIH3T3 but not COS cells (35). It is becoming clear for other sites too that context is important in determining activity. Spacing as well as the presence of other neighboring sites can be important in determining the activity of a given site (36).

In addition to previously identified binding sites, our analysis also revealed a 33 bp near-perfect palindrome ~900 bp upstream of the transcription start site. Removal of this element resulted in a mild but significant increase (>95% confidence by Student's *t*-test) in promoter activity, suggesting a potential repressor activity may bind to this site. It will be of interest to determine the nature of the factor(s) binding here. As discussed above, potentially novel positive elements may also be present in the 43 bp region between the *Kpn*I and *Sma*I sites.

In summary, this study has revealed several mechanisms for controlling activity of the *QM* promoter, including both known and novel elements. The interactions between the distal Sp1-responsive GC boxes and proximal promoter elements highlight the considerable potential for synergy between different promoter elements and, thus, support the growing understanding that transcription factor binding sites do not necessarily function independently, but can have significant functional interactions with each other, depending on their context within the promoter. Presumably, such interactions are important in increasing the subtlety of transcriptional regulation. The potential for regulation of the *QM* promoter is evident. How this potential for regulation determines the impact of this essential eukaryotic gene on cell growth and differentiation remains to be determined.

ACKNOWLEDGEMENTS

The authors thank Richard Schuppek for his excellent technical assistance and Dr H.Mangalam for the use of his transcription factor database. The cosmid containing the genomic *QM* sequence from Xq28 was a kind gift of D.Toniolo (Instituto di Genetica Biochemica ad Evoluzionistica, Pavia, Italy). This work was funded by The National Cancer Institute, National Institutes of Health grant GM 20379 and a Post-doctoral Fellowship (KPS), The Norwegian Cancer Society (JIJ), a National Science Foundation Graduate Fellowship (TML) and a University of California Cancer Research Coordinating Committee Post-doctoral Associate Award (16137; AAF/EJS).

REFERENCES

1 Dowdy,S.F., Lai,K-M., Weissman,B., Matsui,Y., Hogan,B.L.M. and Stanbridge,E.J. (1991) *Nucleic Acids Res.*, **19**, 5763–5769.
2 Farmer, A.A., Loftus,T.M., Mills,A.A., Sato,K.Y., Neill,J.D., Tron,T., Yang,M., Trumpower,B.L. and Stanbridge,E.J. (1994) *Hum. Mol. Genet.*, **3**, 723–728.
3 Eisinger,D.P., Jiang,H.P. and Serrero,G.A. (1993) *Biochem. Biophys. Res. Commun.*, **196**, 1227–1232.
4 Monteclaro,F.S. and Vogt,P.K. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6726–6730.
5 Tron,T., Yang,M., Dick,F.A., Schmitt,M.E. and Trumpower,B.L. (1995) *J. Biol. Chem.*, **270**, 9961–9970.
6 Kaneko,K., Kobayashi,H., Onodera,O., Miyatake,T. and Tsuji,S. (1992) *Hum. Mol. Genet.*, **1**, 529–533.
7 Jiang,H.P., Harris,S.E. and Serrero,G. (1992) *Cell Growth Differentiat.*, **3**, 21–30.
8 Marty,I., Brugidou,C., Chartier,Y. and Meyer,Y. (1993) *Plant J.*, **4**, 265–278.
9 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
10 Courey,A.J. and Tjian,R. (1988) *Cell*, **55**, 887–898.
11 Okayama,H. and Chen,C.A. (1991) In Murray,E.J. (ed.), *Gene Transfer and Expression Protocols.* Humana Press, pp. 15–21.
12 Feinberg,A.P. and Vogelstein,B.A. (1984) *Anal. Biochem.*, **137**, 266–267.
13 *Program Manual for the Wisconsin Package, Version 7.* (1991) Genetics Computer Group, 575 Science Drive, Madison, WI 53711, USA.
14 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
15 van den Ouweland,A.M.W., Kioschis,P., Verdijk,M., Tamanini,F., Toniolo,D., Poustka,A. and van Oost,B.A. (1992) *Hum. Mol. Genet.*, **1**, 269–273.
16 Bione,S., Tamanini,F., Maestrini,E., Tribioli,C., Poustka,A., Torri,G., Rivella,S. and Toniolo,D. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 10977–10981.
17 Bird,A.P., Taggart,M.H., Nicholls,R.D. and Higgs,D.R. (1987) *EMBO J.*, **6**, 999–1004.
18 Tykocinski,M.L. and Max,E.E. (1984) *Nucleic Acids Res.*, **12**, 4385–4396.
19 Yamamoto,K.K., Gonzalez,G.A., Biggs,W.H. and Montminy,M.R. (1988) *Nature*, **334**, 494–498.
20 Bird,A.P. (1993) *Cold Spring Harbor Symp. Quant. Biol.*, **58**, 281–285.
21 Macleod,D., Charlton,J., Mullins,J. and Bird,A.P. (1994) *Genes Dev.*, **8**, 2282–2292.
22 Larsen,F., Gundersen,G., Lopez,R. and Prydz,H. (1992) *Genomics*, **13**, 1095–1107.
23 Bird,A.P. (1987) *Trends Genet.*, **3**, 342–347.
24 Coulondre,C., Miller,J.H., Farabaugh,P.J. and Gilbert,W. (1978) *Nature*, **274**, 775–780.
25 Koller,E., Hayman,A.R. and Trueb,B. (1991) *Nucleic Acids Res.*, **19**, 485–491.
26 Brown,J.R., Daar,I.O., Krug,J.R. and Maquat,L.E. (1985) *Mol. Cell. Biol.*, **5**, 1694–1706.
27 Dynan,W.S. (1986) *Trends Genet.*, **2**, 196–197.
28 Mavrothalassitis,G.J., Watson,D.K. and Papas,T.S. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1047–1051.
29 Ackerman,S.L., Minden,A.G. and Yeung,C.Y. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 11865–11869.
30 Pascal,E. and Tjian,R. (1991) *Genes Dev.*, **5**, 1646–1656.
31 Berg,J.M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11109–11110.
32 Roesler,W.J., Vandenbark,G.R. and Hanson,R.W. (1988) *J. Biol. Chem.*, **263**, 9063–9066.
33 Shepard,A.R., Zhang,W. and Eberhardt,N.L. (1994) *J. Biol. Chem.*, **269**, 1804–1814.
34 Jungling,S., Cibelli,G., Czardybon,M., Gerdes,H.H. and Thiel,G. (1994) *Eur. J. Biochem.*, **226**, 925–935.
35 Evans,M.J. and Scarpulla,R.C. (1989) *J. Biol. Chem.*, **264**, 14361–14368.
36 Tjian,R. and Maniatis,T. (1994) *Cell*, **77**, 5–8.