Identification of DNA sequences required for mouse APRT gene expression

Michael K.Dush⁺, Michael R.Briggs^{1§}, Melanie E.Royce, Dennis A.Schaff[¶], Sohaib A.Khan, Jay A.Tischfield² and Peter J.Stambrook*

Department of Anatomy and Cell Biology, University of Cincinnati College of Medicine, Cincinnati, OH 45267, ¹Howard Hughes Medical Institute, Department of Biochemistry, University of California—Berkeley, Berkeley, CA 94720 and ²Department of Medical Genetics, Indiana University School of Medicine, Indianapolis, IN 46223, USA

Received April 12, 1988; Revised and Accepted August 2, 1988

ABSTRACT

The mouse *aprt* promoter contains four GC boxes, which bind transcription factor Sp1 *in vitro*, and lacks both TATA and CCAAT boxes. Removal of the two most distal GC boxes of this promoter had little effect on APRT enzyme levels produced in a transient expression assay. Deletion of the distal three GC boxes resulted in a 50% reduction, and deletion of all GC boxes resulted in essentially complete loss of APRT activity. There are two predominant transcription start sites which are located within the region containing the GC boxes. The promoter behaved as a relatively strong promoter when compared to the RSV LTR promoter in a transient CAT assay, and operated in one orientation only. No upstream anti-sense transcripts were detected in either mouse CAK or liver cells, confirming that the mouse *aprt* promoter, unlike some other GC-rich promoters appears not to support bidirectional transcription.

INTRODUCTION

Accurate and efficient transcription initiation by RNA polymerase II requires several separate DNA elements, which together comprise a promoter. Initial studies of promoter elements associated with highly expressed viral and cellular genes, identified several conserved sequences whose mutation resulted in the loss of either accuracy or efficiency of transcription initiation. The first promoter element to be identified was the TATA box, located around position -30 with respect to the site of transcription initiation. A second series of elements, located 5' to the TATA box, were subsequently shown to contribute to wild-type levels of gene expression. An early model suggested that RNA polymerase II promoters consist of at least two elements, a TATA box and one or more upstream elements, all of which are conserved in sequence, position, and orientation with respect to the transcription initiation site. The TATA box appeared necessary for accurate transcription initiation (1-8) while the upstream elements contributed to efficient initiation (1,3-6,8-12).

It is generally accepted that this RNA polymerase II promoter architecture applies in many cases, but is subject to substantial variation and is not the

only possible polymerase II promoter motif. The TATA box is more variable in sequence than originally appreciated and is not always an obligatory component of the promoter. The promoter regions of "housekeeping genes" (13-18) deviate from the original promoter model in several important respects. Specifically, they are very G+C-rich and frequently lack a TATA box or TATA-like sequence. They also usually contain one or more GC boxes whose sequence, CCGCCC or its reverse complement, comprises the core of a consensus decanucleotide that interacts with the cellular transcription factor Sp1. Like the majority of genes, housekeeping genes produce transcripts that are commonly represented in low abundance within the mRNA population (19). A complete understanding of RNA polymerase II promoter function will require not only an appreciation of how TATA box-containing promoters operate, but also of how sequences within these GC-rich promoters direct accurate and efficient transcription initiation.

In this report we characterize the promoter region of the gene that encodes mouse adenine phosphoribosyltransferase (APRT). The APRT enzyme, which constitutes the pathway that converts adenine to AMP, is present at low steady-state levels in most or all cell types and appears to be subject to little if any regulation.

MATERIALS AND METHODS

Plasmid Construction

The starting plasmid for these studies, pSAM 3.1, contains a 3.1 kb fragment encompassing a functional mouse *aprt* gene (20) cloned between the EcoRI and SphI sites of pBR328. The nucleotide sequence of this fragment, as well as of two APRT cDNAs, has been previously reported (18). The plasmid pSAM 3.1 was modified to include an additional 1.3 kb of 3' flanking sequence distal to the polyadenylation site, which may be important for efficient transcription termination and 3' processing (21-24), and was designated pSAM 4.4. To facilitate subcloning, a 400 bp KpnI fragment, containing an internal HindIII site, was excised from the largest intron (intron 2) of the *aprt* gene, and the resulting 4.0 kb EcoRI - HindIII fragment containing the mouse gene and 3' flanking sequences was inserted into the polylinker region of plasmid pIBI20 and designated pIBI/4.0. Plasmid pIBI/4.0 was as efficient as pSAM 4.4 in transfecting *Aprt*⁻ cells to the *Aprt*⁺ phenotype (data not shown), and is considered wild-type for these studies. To construct the plasmid pIBI/3.7, which has the distal putative promoter deleted, pIBI/4.0 was linearized with XmnI,

followed by addition of an EcoRI linker at the XmnI site. Removal of the resultant 322 bp EcoRI fragment produced pIBI/3.7.

Deletion Mutagenesis and DNA Sequencing

Progressive deletions of the promoter region were produced in either of two ways. The first is the "band-aid" method originally described by Dale *et al.* (26) using the "Cyclone" kit purchased from International Biotechnologies Inc. (IBI). Phage particles were produced following infection with helper phage, and single stranded DNA was isolated from phage particles in the culture supernatant. Unidirectional deletion followed the basic methodology described by Dale *et al.* (25). In the second method, plasmids were linearized with EcoRI and digested for increasing times with BAL-31 nuclease (New England Biolabs). Aliquots of the BAL-31 reaction were pooled and termini with overhanging 5' nucleotides were filled in with the Klenow fragment of *E. coli* DNA polymerase I. After heat inactivation of the Klenow fragment, the samples were digested to completion with HindIII and the partially deleted fragments re-cloned into pIBI20. Deletion mutants were sequenced using the chain termination method as previously described (26).

5' Mapping by Primer Extension

A 22 residue oligonucleotide was synthesized that extended from position +11 to -11 (Figure 1A) and was complementary to the messenger RNA. The oligonucleotide was end-labeled with T4 polynucleotide kinase and [γ -³²P]ATP, hybridized with poly A⁺ RNA, and extended with reverse transcriptase and all four dNTPs as previously described (27). The extended products were fractionated on an acrylamide sequencing gel and displayed by autoradiography.

Cell Culture and DNA Transfections

Mouse Aprt⁻ LS-24b cells, were grown as previously described (28). To eliminate possible Aprt⁺ revertants, although none have been observed, the cells were passaged through 50 μ g/ml 2,6-diaminopurine (DAP). Stable transfections, using the calcium phosphate precipitation method were performed as previously described (29,30), as was selection for Aprt⁺ colonies (28) in medium containing 0.05 mM azaserine, 0.1 mM adenine and 4 μ g/ml alanosine (AAA medium).

For transient expression, 2x10⁶ cells per 100 mm plate were exposed to the DNA-calcium phosphate precipitate for 4-6 hr, followed by a 90 second glycerol shock and addition of fresh medium. After 48 hr the cells were collected and suspended in 100 μ l of 0.25 M Tris HCl, pH 7.4. The cells were lysed by two cycles of freeze-thawing and the cellular debris pelleted by centrifugation.

Aliquots of the resultant supernatant were used for either APRT or CAT assay, depending upon the transfecting DNA.

CAT and APRT Enzyme Assays

CAT assays were performed as described by Gorman *et al.* (31), and relative APRT values were determined after 0, 15, 30 and 60 minutes of incubation by a rapid filter binding assay (32) that measures the conversion of labeled adenine to AMP. In addition, a 5 μ l aliquot of the APRT assay mixture was removed at 60 minutes, mixed with 50 μ l of a solution containing 50 mM adenine, adenosine, and AMP, and chromatographed on a thin layer cellulose plate to separate reaction products and substrate using 5% (w/v) Na_2HPO_4 as the solvent. The markers were visualized by UV illumination, cut out and their radioactivity measured.

DNase Protection Experiments

DNase I protection experiments were performed as previously described (33-35), using a crude fraction of Spl that elutes from a heparin column with 0.3 M KCl as well as a highly purified Spl preparation (36). The fragment used for footprinting was prepared by ^{32}P end-labeling of a 232 bp FokI fragment that extends from position -210 (Figure 1) to +23 (not shown). This fragment was cleaved by digestion with FnuDII (position -6, Figure 1), and the resultant 205 bp fragment was gel purified and used for Spl binding and footprint analysis.

RESULTS

Organization of the Promoter Region

We have previously reported the nucleotide sequence of the BALB/c mouse APRT gene and its 5' flanking region (18). Inspection of the DNA sequence upstream of the ATG translation start codon revealed two discrete regions each containing a potential promoter (Figure 1). The distal region contains a nucleotide sequence and organization with striking similarity to known β -globin promoters (Figure 1B), and is followed by an open reading frame, distinct from the APRT structural gene, that is sufficient to encode an 86 amino acid peptide. Their relative positions are depicted in Figure 1A. The more proximal functional promoter is similar to that of other housekeeping genes. It lacks TATA and CCAAT sequences, contains multiple GC boxes, and like promoter regions of other housekeeping genes, contains GC boxes that are part of a larger G+C rich domain.

To determine whether or not upstream regions are important for aprt expression, we first constructed the plasmid pIBI/3.7, which lacks sequences

A

B

-110	-83	-26	+1	
CCACACCCT- -	GGCCAATT- -	CATTATA- -	TTTACAC	APRT upstream
-92	-77	-31	+1	
CCACACCCT- -	GGCCAATCT- -	CATAAA- -	CTTACAC	Rabbit β -globin
-94	-79	-32	+1	
CCACACCCT- -	GGCCAATCT- -	CATATA- -	CTCACAT	Mouse β -major

Figure 1. Organization of the region upstream of the APRT translation start codon. A) The numbered scale bar is in base pairs, and is numbered negatively from the first nucleotide upstream of the ATG start codon. The solid vertical rectangles represent the position of GC boxes to which SphI binds *in vitro*. The asterisks represent the predominant transcription start sites as defined by primer extension. The horizontal open rectangle depicts the position of the distal β -globin promoter-like sequence, and the diagonally shaded rectangle identifies the position of the open reading frame. The horizontal lines beneath the scale bar identify each of the deletion mutants and their 5' end points. The solid vertical arrows identify the 5' end points of the deletion mutants used for transient expression in Figure 2. The letters indicate the following restriction sites: Fn=FnuDII; Fo=FokI; R=EcoRI; Sc=ScaI; Sm=SmaI; and X=XmnI. B) Comparison of the β -globin promoter-like sequence (open rectangle) upstream of the *aprt* gene with rabbit β -globin and mouse β -major globin promoters. The numbers reflect nucleotide distances from the β -globin transcription start sites and its presumptive counterpart in the *aprt* upstream region.

upstream of the XmnI site at position -551 (Figure 1), and compared the capacity of pIBI/3.7 and of pIBI/4.0 to transfect LS-24b cells. As shown in Table 1, pIBI/3.7 and pIBI/4.0 produced approximately the same number of *Aprt*⁺ colonies and supported production of approximately equivalent levels of APRT activity. To more precisely define the sequences required for gene expression, the "band-aid" method (25) was used to produce a series of progressive deletion mutants spanning the region from the former XmnI site to the ATG start codon (Figure 1A). A subset of these mutants were compared for their ability to confer an *Aprt*⁺ phenotype on *Aprt*⁻ LS-24b cells following trans-

TABLE 1
Stable and Transient Transfections with 5' Deletion
Mutants of the Aprt Gene

Transfecting Plasmid	Stable Transfections	Transient APRT Expression
	Colonies/2x10 ⁵ Cells	APRT Activity Relative to pIBI/4.0
pIBI/4.0	116	1.0
pIBI/3.7	138	0.86
pIBI/-320	128	1.05
pIBI/-295	121	ND
pIBI/-255	110	0.90
pIBI/-245	115	ND
pIBI/-226	178	ND
pIBI/-160	149	1.01
pIBI20	0	0.03

LS-24b cells were transfected with each of a series of 5' deletion mutants of the aprt gene. The effect of the deletions upon aprt expression was measured by the efficiency of colony production in stable transfections and the relative production of APRT activity in a transient expression assay. pIBI20 represents vector sequences only and pIBI/4.0 is a plasmid containing a control aprt gene with all of the upstream sequence presented in Figure 1. pIBI3.7 contains an aprt gene deleted to the XmnI site at -550. The 5' end point of each of the other deletion mutants is indicated by its designated number. The values presented represent the average of 3 independent determinations.

fection. Deletion to position -160 had no effect on transfection efficiency or on enzyme activity levels in transfectants (Table 1), placing the 5' border of the promoter within 160 bp from the translation start codon.

Additional mutant genes, with progressive 5' deletions from position -160 to the translation start codon were tested for their capacity to produce functional transcripts as measured by APRT activity in a transient expression assay. Since it was likely that one or more of the GC boxes would be a functionally significant component of the promoter, mutants which retained integral numbers of GC boxes were analyzed. As shown in Figure 2, deletion of the two distal GC boxes (mutant -121) had little effect. Deletion of the 3 distal GC boxes (mutant -88) resulted in about a 50% reduction in APRT activity, and

Figure 2. Transient APRT activities in LS-24b cells following transfection with mutant APRT genes. APRT deficient LS-24b cells were transfected with mutant APRT genes containing progressive deletions in 5' flanking sequences as described in Methods. Forty-eight hours after transfection, cells were collected and lysates assayed for APRT activity. The data are presented as counts incorporated into AMP as a function of time. Each point represents the average of 3 determinations. Symbols: +, pIBI/-160; *, pIBI/-121; □, pIBI/-88; Δ, pIBI/-66; •, pIBI/-47. The numbers indicate deletion end-points which are shown in Figure 1.

deletion of all four GC boxes (mutant -66) resulted in an almost complete loss of APRT activity. The efficiency of each transfection was monitored by co-transfection with pRSVCAT and an aliquot of each cell extract was assayed for CAT activity, which remained relatively constant between samples (data not shown). The filter binding assay, which only measures AMP production, can potentially be biased by 5'-nucleotidase activity, which dephosphorylates AMP to adenosine. Therefore, an aliquot of each extract was chromatographed to measure AMP conversion to adenosine, which was always less than 5% of the total AMP produced. This indicates that values obtained from the filter disc assays accurately reflect relative APRT activities of the transfectants. Thus, the transfection data suggest that DNA sequence elements between posi-

Figure 3. DNase I footprint analysis of Spl binding sites in the APRT promoter region. The figure displays an autoradiographic exposure of an 8% polyacrylamide sequencing gel used to analyze the products of a partial DNase I digest of the end-labeled 205 bp *FokI*/*Fnu*DI fragment (Figure 1). Lanes 1 and 12) control reactions (no Spl); lanes 2-6) increasing amounts of a 0.3M KCl heparin fraction containing partially purified Spl; lanes 7-11) increasing amounts of highly purified Spl; lane m) marker DNA fragments whose sizes are indicated. The numbered boxes at the right of the figure represent the position of the GC boxes, with 1 being the most proximal and 4 being the most distal.

tions -121 and -66, which include the two most proximal GC boxes, are required for *aprt* gene expression in LS-24b cells.

Spl Binding

The presence of GC boxes within the proximal promoter region, and their apparent importance for expression, suggests the participation of Spl in *aprt* gene transcription. To test whether or not Spl binds one or more of the GC boxes *in vitro*, an end-labeled *FokI*/*Fnu*DI fragment extending from -210 to -6 (Figure 1), was mixed with partially purified Spl or with Spl purified to 95% homogeneity (36). Following partial DNase I digestion, the reaction products

Figure 4. 5' mapping of mouse *sprt* mRNA by primer extension. A) Sequencing ladder for size determination of primer extended RNAs. B) RNA samples were hybridized with the 5' end-labeled oligonucleotide described in the Methods. The primer was extended with reverse transcriptase, and the extension products resolved on a 6% polyacrylamide sequencing gel. Lane a) mouse CAK cell poly A⁺ RNA; b) Balb/c mouse liver poly A⁺ RNA; c) *E. coli* tRNA; d) total RNA from Balb/c liver. The arrows identify the position of primer extended products. The nucleotide distances of primer extended termini from the translation start codon are indicated. The sequencing ladder and primer extension separation were performed in the same gel, but for this illustration, the ladder was exposed to film for 18 hr and the primer extended products were exposed for 72 hr.

and protected sequences were displayed on a DNA sequencing gel. As seen in Figure 4, Spl interacts with each of the four GC boxes, but provides differing levels of protection against DNase 1 digestion that are more apparent with the partially purified Spl preparation. Regions containing GC boxes 1 and 3 are

TABLE 2
Orientation Requirement of the *aprt* Promoter
for Transient CAT Expression

Plasmid	CAT Activity		
	Experiment 1	Experiment 2	Experiment 3
pRSVCAT	100	100	100
pEMBLCATFnu ⁺	15%	10%	17%
pEMBLCATFnu ⁻	ND	1.6%	2.1%
pEMBLCAT	ND	1.6%	2.2%
pEMBLCAT ΔFnu ⁺	ND	ND	14%
pEMBLCAT ΔFnu ⁻	ND	ND	2.5%

LS-24b cells were transfected with derivatives of pEMBLCAT containing *aprt* promoter sequences in forward (+) or reverse (-) orientation. CAT determinations were performed 48 hr after transfection. The inserted fragment in the pEMBLCATFnu construct extends from -402 to -6 as shown in Figure 1. The inserted fragment in the pEMBLCAT ΔFnu construct extends from -160 to -6. In all experiments, the CAT activities were compared to those obtained with pRSVCAT, and all values represent the average of three separate transfections.

less effectively protected than regions containing GC boxes 2 and 4 (Figure 3).

Transcription Start Sites

The position of transcription initiation was assessed by primer extension as described in the Methods, and produced two major bands coincident with positions -116 and -73 with respect to the translation initiation codon (Figure 4). The position of these transcription start sites with respect to the GC boxes is depicted by asterisks in Figure 1A.

Strength and Directionality of the *aprt* Promoter

The relative strength of the mouse *aprt* promoter was assessed by inserting it in front of a promoterless bacterial CAT gene and testing its ability to produce CAT in a transient expression assay. The APRT promoter/CAT gene construct designated pEMBLCATFnu⁺ was comprised of a 397 bp *Fnu*DI fragment, extending from position -402 to position -6 (Figure 1A), inserted into the polylinker of pEMBLCAT upstream of and in the same transcriptional orientation as the CAT gene. A second, shorter promoter fragment extending from -159 to -6, containing all four GC boxes, was inserted at the same site and in the

same orientation, and was designated pEMBLCAT Δ Fnu⁺. The plasmid pEMBLCAT contains the CAT gene coding region and an SV40 polyadenylation signal, but no promoter or enhancer. Transfection of parallel LS-24b cultures with each of the promoter-containing constructs and with pRSVCAT demonstrates that the mouse aprt promoter facilitates CAT expression about 10-17% as efficiently as the RSV LTR promoter (Table 2).

Bidirectional transcription from several promoters, including SV40 (37), dihydrofolate reductase (DHFR) (38-40), and Ki-ras (41), mediated by GC boxes, suggests that it may be a common phenomenon. To test this possibility for aprt, each of the above promoter-containing fragments was inserted into pEMBLCAT at the same site, but in reverse orientation with respect to the CAT gene, and designated pEMBLCATFnu⁻ and pEMBLCAT Δ Fnu⁻, respectively. When transfected into LS-24b cells, CAT activity produced by either construct was negligible and indistinguishable from that produced by the parental plasmid pEMBLCAT (Table 2), suggesting that the mouse aprt promoter does not function bidirectionally.

To further test for bidirectional transcription, a 606 bp ScaI/SmaI fragment (Figure 1A) immediately upstream from the promoter region was used in Northern blots to probe RNA from adult mouse liver and from Ap^rt⁺ mouse CAK cells. No discrete hybridization bands were visible, even after an eight day exposure. In contrast, hybridization of the same Northern blot with an 800 bp BamHI probe containing most of the ap^rt cDNA, but labeled to a specific activity about one tenth that of the ScaI/SmaI probe, produced a discrete band within 18 hours (data not shown). This result reaffirms that in mouse liver and CAK cells, the APRT promoter does not support transcription in both directions; or if it does, sense transcripts are more abundant by at least two orders of magnitude than transcripts from the opposite strand. Since the upstream ScaI/SmaI probe is labeled on both strands, and contains the upstream open reading frame (see Figure 1), the absence of detectable RNA following an 8 day exposure also suggests that the upstream β -globin promoter-like sequence does not support transcription in mouse liver and CAK cells.

DISCUSSION

The mouse ap^rt promoter has characteristics common to those of housekeeping genes that are expressed in all tissues. Many of these promoters lack TATA and CCAAT motifs, but are contained within sequences that are G+C-rich (13-18, 42,43). Within the G+C-rich domain of the mouse ap^rt promoter, there are 4 copies of the consensus decanucleotide sequence which, in vitro, bind

the Spl transcription factor. Whether Spl binds any or all of these sequences in vivo is unknown, but is under investigation. Progressive 5' deletions have defined some of the regions important for mouse *aprt* expression. Removal of the most distal GC box alone or of the two most distal GC boxes together, permitted APRT activity approaching that of wild-type. Sequential removal of the next two GC boxes resulted in a 50% decrease and a loss of APRT activity, respectively. In this context, it is worth noting that the human *aprt* promoter region lacks the two distal GC boxes, although other GC boxes exist at different sites (44,45), and it bears little overall sequence homology to the mouse promoter (18). However it has two GC boxes that are similarly spaced and positioned as the two proximal GC boxes in the mouse promoter. Promoter and enhancer regions of other genes have cis-acting sequences, whose removal results in elevated promoter activity (46-48). No such negative regulatory elements were revealed by the deletion mutants used in these experiments.

Distal to the functional promoter is a sequence which has striking similarity to known β -globin promoters and which is followed by an open reading frame sufficient to encode 86 amino acids. This sequence contains TATA and CCAAT boxes appropriately positioned with respect to a potential transcription start site. There is also a nine nucleotide sequence which occurs at approximately the same position 5' to all β -globin genes examined to date, and which appears to be a functional component of both the rabbit (8) and mouse (49,50) β -globin promoters. This promoter-like sequence, however, appears not to function in adult liver or in the cultured cells used for these experiments. Its removal has no effect upon transfection efficiency or upon the capacity to produce APRT enzyme activity during transient expression, and it produces no detectable transcript under conditions at least 80 fold more sensitive than are required to detect the *aprt* gene transcript. Inserting this promoter-like sequence in front of the neomycin-resistance gene is insufficient to confer resistance to G418 in transfected recipient cells (M. Dush, unpublished data). However, since it closely resembles a β -globin promoter, which requires linkage to an enhancer for efficient expression in non-erythroid cells (51-53), inclusion of an appropriate enhancer may be necessary to permit its activation in cultured cells. This sequence does support α -amanitin sensitive transcription initiation in an in vitro transcription initiation assay (D. Luse, personal communication). Thus, it will be informative to determine whether, as in the case of the porphobilinogen deaminase gene (54), this potential promoter may be subject to cell type-specific or developmental regulation.

The coding regions of mouse and hamster *aprt* genes are highly conserved

(55); however, their 5' flanking sequences are divergent. Furthermore, transcription start sites in the two rodent genes appear to be different. In the hamster, transcription begins predominantly about 65 nucleotides upstream of the translation initiation site (55). In the mouse, primer extension defines two primary transcription start sites at about -116 and -73. The basis for the difference in these transcription initiation start sites is unclear, but it may reflect a difference in position and organization of potential Sp1 binding sites. Despite these differences, limited similarities between the two rodent aprt promoters have been noted (55). Immediately upstream of the translation start codon, 9 of the first 10 base pairs are conserved in the mouse (18) and hamster (55,56). There is also a 26 nucleotide sequence extending from -83 to -108 with strong homology to a sequence in the hamster promoter, and to a lesser extent, to a sequence within the presumptive human promoter (55). Significantly, in the mouse, deletion to -87, which includes most of this sequence, reduces APRT activity in transient expression by about 50% (Figure 2).

The mouse aprt promoter operates efficiently in transient expression experiments in cultured cells. Insertion of a fragment containing the four GC boxes in front of the bacterial CAT gene, affirms its promoter function. Comparison with the RSV LTR promoter confirms that the mouse aprt promoter is relatively strong, supporting CAT production with about 15% the efficiency of the RSV LTR. This observation is consistent with reports indicating that other G+C-rich regions behave as strong promoters in CAT transient expression assays (55,57,58).

Some promoters with GC boxes support bidirectional transcription. Those within each of the 21 bp repeats of the SV40 promoter region are required for early and late gene expression (37), and produce transcripts that proceed in opposite directions. Point mutations within individual GC boxes decrease the level of both early and late transcription, and suggest that individual GC boxes participate in promoting bidirectional transcription. Similarly, the promoter regions of mouse (38,39) and Chinese hamster (40) DHFR genes, and Ki-ras (41) genes support bidirectional transcription. They all possess GC boxes which, in the mouse DHFR gene (59), have been shown to bind Sp1 in vitro. These observations, plus the fact that the distal GC box of the HSV tk promoter will function in either orientation, suggest that the contribution of GC boxes to bidirectional transcription may be a general phenomenon. This property, however, does not extend to the mouse aprt promoter, since neither construct, in which the promoter was inverted with respect to the CAT gene,

supported production of CAT activity. This observation is consistent with a similar recent report for the hamster *aprt* promoter (55).

Although Sp1 binding to GC boxes appears necessary for transcription of some genes, its binding to all GC boxes in a promoter is not always required for function. There are 10 GC boxes upstream of the Ha-ras transcription start site, but only 6 bind Sp1 and only the three proximal binding sites are needed for full expression (58). In contrast, all of the five GC boxes located within a 500 bp region upstream of the HMG CoA reductase promoter appear important for expression (57). In the mouse *aprt* promoter, all four GC boxes bind Sp1 *in vitro* but confer differing degrees of protection from DNase I digestion. The distal three GC boxes conform completely to the consensus Sp1 decanucleotide binding site (36), whereas the most proximal displays homology at 9 of the 10 bases.

It is likely that the effects of Sp1 binding to DNA upon transcription are dependent upon its interaction with other factors, and are not caused by Sp1 binding alone. In this context, we note the above-mentioned sequence homologies that appear in the known mammalian *aprt* promoters, and that also may participate in protein binding functions. In addition, rodent and human *aprt* promoters comprise part of an HTF island (60), with elevated CpG dinucleotide frequencies extending through the first two exons and first intron (44). Methylation of the 5' half of the hamster *APRT* gene interferes with expression (61), suggesting that the CpG-rich domain must remain hypomethylated for function. Bird has proposed that such CpG-rich regions may interact with one or more proteins that help maintain its hypomethylated state and may interact with other proteins to influence transcription (60). Whether such additional proteins exist and whether they interact with Sp1 remains to be determined.

ACKNOWLEDGEMENTS

We thank Estrella Feliciano for expert technical assistance, Amy Bertino for critical reading of the manuscript and Susan Eder for her help in manuscript preparation. The Sp1 binding was performed in Robert Tjian's laboratory and the Howard Hughes Medical Institute. This work was supported by Public Health Service grants CA 36897 and DK 38185 from the National Institutes of Health. M.K.D. was supported by a fellowship from the Albert J. Ryan Foundation.

Present addresses: ⁺Department of Anatomy, University of California—San Francisco, San Francisco, CA 94143, USA, [§]Zoologisches Institut, Universitat Zurich, Winterthurer Strasse, 190 CH8057, FRG and [¶]Germplasm Quality and Enhancement Laboratory, B-001, Room 311, BARC-West, Beltsville, MD 20705, USA

*To whom correspondence should be addressed

REFERENCES

1. Benoist, C. and Chambon, P. (1981) *Nature* **290**, 304-310.
2. Ghosh, P.K., Lebowitz, P., Frisque, F.J. and Gluzman, Y. (1981) *Proc. Natl. Acad. Sci.* **78**, 100-104.
3. Fromm, M. and Berg, P. (1982) *J. Mol. Appl. Gen.* **1**, 457-481.
4. Fromm, M., and Berg, P. (1983) *J. Mol. Appl. Gen.* **2**, 127-135.
5. McKnight, S.L., Gavis, E.R., Kingsbury, R. and Axel, R. (1981) *Cell* **25**, 385-398.
6. McKnight, S.L. and Kingsbury, R.C. (1982) *Science* **217**, 316-324.
7. Grosschedl, R. and Birnstiel, M.L. (1980) *Proc. Natl. Acad. Sci.* **77**, 1432-1436.
8. Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissman, C. (1983) *Cell* **32**, 695-706.
9. McKnight, S.L. (1982) *Cell* **31**, 355-365.
10. McKnight, S.L., Kingsbury, R.C., Spence, A. and Smith, M. (1984) *Cell* **37**, 253-262.
11. Grosschedl, R., Waslylyk, B., Chambon, P. and Birnstiel, M.L. (1981) *Nature* **294**, 178-180.
12. Scott, R.W. and Tilghman, S.M. (1983) *Mol. Cell Biol.* **3**, 1295-1309.
13. Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S. and Goldstein, J.L. (1984) *Cell* **38**, 275-285.
14. Melton, D.W., Konecki, D.S., Brennard, J. and Caskey, C.T. (1984) *Proc. Natl. Acad. Sci.* **81**, 2147-2151.
15. McGrogan, M., Simonsen, C.C., Smouse, D.T., Farnham, P.J. and Schimke, R.T. (1985) *J. Biol. Chem.* **260**, 2307-2314.
16. Singer-Sam, J., Keith, D., Tani, K., Simmer, R.L., Shively, L., Lindsay, S., Yoshida, A. and Riggs, A.D. (1984) *Gene* **32**, 409-417.
17. Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkvens, T.M., van der Voorn, L., van Ormondt, H. and van der Eb, A.J. (1985) *EMBO J.* **4**, 437-443.
18. Dush, M.K., Sikela, J.M., Khan, S.A., Tischfield, J.A. and Stambrook, P.J. (1985) *Proc. Natl. Acad. Sci.* **82**, 2731-2735.
19. Hastie, N.D. and Bishop, J.O. (1976) *Cell* **9**, 761-774.
20. Sikela, J.M., Khan, S.A., Feliciano, E., Trill, J., Tischfield, J.A. and Stambrook, P.J. (1983) *Gene* **22**, 219-228.
21. Woychik, R.P., Lyons, R.H., Post, L. and Rottman, F.M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3944-3948.
22. Gil, A. and Proudfoot, N.J. (1984) *Nature* **312**, 473-474.
23. McLaughlan, J., Gaffney, D., Witton, J.L and Clements, J.B. (1985) *Nucleic Acids Res.* **13**, 1347-1365.
24. Falk-Pederson, E., Logan, J., Shenk, T. and Darnell, J.E. (1985) *Cell* **40**, 897-905.
25. Dale, R.M.K., McClure, B.A. and Houchins, J.P. (1985) *Plasmid* **13**, 31-40.
26. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
27. Sazer, S. and Schimke, R.T. (1986) *J. Biol. Chem.* **261**, 4685-4690.
28. Tischfield, J.A., Trill, J.J., Lee, Y.I., Coy, K. and Taylor, M.W. (1982) *Mol. Cell Biol.* **2**, 250-257.
29. Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T. and Axel, R. (1979) *Cell* **16**, 777-785.
30. Graham, F.L. and van der Eb, A.J. (1973) *Virology* **52**, 456-467.
31. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell Biol.* **2**, 1044-1051.
32. Hershey, H.V. and Taylor, M.W. (1978) *Prep. Biochem.* **8**, 453-462.
33. Dynan, W.S. and Tjian, R. (1983) *Cell* **32**, 669-680.

34. Dynan, W.S. and Tjian, R. (1983) *Cell* 35, 79-87.
35. Jones, K.A. and Tjian, R. (1985) *Nature* 317, 179-182.
36. Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R. (1986) *Science* 234, 47-52.
37. Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P. and Tjian, R. (1986) *Science* 230, 511-517.
38. Farnham, P.J., Abrams, J.M. and Schimke, R.T. (1985) *Proc. Natl. Acad. Sci.* 82, 3978-3982.
39. Crouse, G.F., Leys, E.J., McEwan, R.N., Frayne, E.G. and Kellems, R.E. (1985) *Mol. Cell Biol.* 5, 1847-1858.
40. Mitchell, P.J., Carothers, A.M., Han, J.H., Harding, J.D., Kas, E., Venolia, L. and Chasin, L.A. (1986) *Mol. Cell Biol.* 6, 425-440.
41. Hoffman, E., Trusko, S.P., Freeman, N. and George, D.L. (1987) *Molec. Cell. Biol.* 7, 2592-2596.
42. Yang, J.K., Masters, J.N. and Attardi, G. (1984) *J. Mol. Biol.* 176, 169-187.
43. Lin, P.F., Zhao, S.Y. and Ruddle, F.H. (1983) *Proc. Natl. Acad. Sci.* 80, 6528-6532.
44. Broderick, T.P., Schaff, D.A., Bertino, A.M., Dush, M.K., Tischfield, J.A. and Stambrook, P.J. (1987) *Proc. Natl. Acad. Sci. USA* 84, 3349-3353.
45. Hidaka, Y., Tarle, S.A., O'Toole, T.E., Kelley, W.N. and Palella, T.D. (1987) *Nucleic Acids Res.* 15, 9086.
46. Gorman, C.M., Rigby, P.W.J. and Lane, D.P. (1985) *Cell* 42, 517-526.
47. Laimins, L.A., Holmgren-Konig, M. and Khowry, G. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3151-3155.
48. Kuhl, D., Fuente, J., Chaturved, M., Parimoo, S., Ryals, J., Meyer, F. and Weissman, C. (1987) *Cell* 50, 1057-1069.
49. Charnay, P., Mellon, P. and Maniatis, T. (1985) *Mol. Cell Biol.* 5, 1498-1511.
50. Myers, R.M., Tilly, K. and Maniatis, T. (1986) *Science* 232, 613-618.
51. Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981) *Cell* 27, 279-288.
52. Humphries, R.K., Ley, T., Turner, P., Moulton, A.D. and Nienhuis, A.W. (1982) *Cell* 30, 173-183.
53. Treisman, R., Green, M.R. and Maniatis, T. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7428-7432.
54. Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goosens, M. and Romeo, P.H. (1988) *Proc. Natl. Acad. Sci. USA* 85, 6-10.
55. Park, J.H. and Taylor, M.W. (1988) *Mol. Cell. Biol.* 8, 2536-2544.
56. Nalbantoglu, J., Phear, G.A. and Meuth, M. (1986) *Nucleic Acids Res.* 14, 1914.
57. Osborne, T.F., Goldstein, J.L. and Brown, M.S. (1985) *Cell* 42, 203-212.
58. Ishii, S., Merlino, G.T. and Pastan, I. (1986) *Science* 230, 1378-1381.
59. Dynan, W.S., Sazer, S., Tjian, R. and Schimke, R.T. (1986) *Nature* 319, 246-248.
60. Bird, A.P. (1986) *Nature* 321, 209-213.
61. Keshet, I., Yisraeli, J. and Cedar, H. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2560-2564.