Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 7452–7456, August 1992
Biochemistry

MAZ, a zinc finger protein, binds to c-*MYC* and *C2* gene sequences regulating transcriptional initiation and termination

STEVEN A. BOSSONE*, CLAUDE ASSELIN†, AMANDA J. PATEL‡, AND KENNETH B. MARCU*‡§¶

Departments of *Pathology, ‡Biochemistry and Cell Biology, and §Microbiology, State University of New York, Stony Brook, NY 11794; and †Department of Anatomy and Cell Biology, University of Sherbrooke, Sherbrooke, Quebec, Canada J1H 5N4

Communicated by William J. Lennarz, May 11, 1992 (received for review March 17, 1992)

ABSTRACT **ME1a1, a 16-base-pair nuclear factor binding site residing between the c-*MYC* P_1 and P_2 transcription initiation sites, is required for P_2 activity. A cDNA encoding a 477-amino acid zinc finger protein designated MAZ (*MYC*-associated zinc finger protein) was cloned from a HeLa λgt11 library by screening with a concatamerized ME1a1 binding site probe. In addition to six potential zinc fingers of the Cys_2His_2 type, MAZ contains an amino-terminal proline-rich domain and several polyalanine tracts. Its mRNA was present in all human tissues tested except for kidney, as a doublet of approximately 2.5 and 2.7 kilobases, along with differentially expressed minor species. MAZ bound specifically to the wild-type ME1a1 sequence but not to a ME1a1 mutant that also failed to yield P_2 activity. When expressed as a fusion protein in a pMAL-c vector, MAZ binds with specificity to a GA box sequence (GGGAGGG) found in the c-*MYC* P_2 promoter, to the P_2 attenuator region within the gene's first exon, and to a related sequence involved in the transcriptional termination of the *C2* gene. *MAZ* may encode a transcription factor with dual roles in transcription initiation and termination.**

The c-*MYC* protooncogene encodes a nuclear phosphoprotein with a lengthy history in the genesis of a variety of malignancies (reviewed in ref. 1). Remarkably, direct evidence for the physiologic role of the c-*MYC* gene product as a DNA sequence-specific transcription factor has only recently accumulated, through the discovery of its DNA binding site (2) and a related polypeptide, MAX, which dimerizes with MYC to optimize DNA binding (3). A conditional c-*MYC* allele was reported to induce the expression of the α-prothymosin gene (4), implicating the c-MYC polypeptide as a transcription factor. One could therefore envision a model whereby a series of genetic insults could lead to inappropriate expression of the c-*MYC* gene and in turn lead to the abnormal expression of other cellular genes that are directly or indirectly controlled by c-MYC, thus leading to neoplasia. The definition of normal c-MYC control pathways should reveal initiating events leading to normal and abnormal cellular growth.

Numerous positive and negative regulatory elements have been reported for the c-*MYC* gene which regulate transcription from two initiation sites specified by the P_1 and P_2 promoters (1). Transcripts initiating from P_1 (5) and P_2 (6, 7) are also regulated by a block to transcriptional elongation. ME1a1, a nuclear factor site at +97 to +118 with respect to P_1, was identified by deletional mutagenesis as being necessary for P_2 initiation (8). In addition to a 180-nucleotide sequence within the gene's first exon, ME1a1 was required for transcriptional blockage (9). Attenuation of P_1-initiated transcripts was shown to be dependent upon sequences residing between P_1 and P_2, inclusive of the ME1a1 site (5). A comparison of the sequences required for P_1 and P_2 blockage revealed a region in the P_2 attenuator region homologous to the ME1a1 site. In addition, another ME1a1-like sequence has been reported in the termination region of the human complement component 2 (*C2*) gene (10).

We have extended our earlier studies on the requirement of the ME1a1 site for P_2 usage (8) by employing site-directed mutagenesis, and we also report the cloning and partial characterization of a zinc finger protein that specifically binds to the ME1a1 and ME1a1-like sites.‖

MATERIALS AND METHODS

Plasmids. pMBgCAT is a *MYC* promoter–chloramphenicol acetyltransferase (*CAT*) fusion construct (8); pMAZ-MBP consists of the 2.2-kilobase (kb) *Eco*RI insert of the ME1a1-binding λgt11 19-2-3-1 cDNA clone fused to the maltose-binding protein (MBP) gene in pMAL-c (New England Biolabs); pMAZ is the 2.422-kb *Eco*RI insert of a MAZ (for *MYC*-associated zinc finger protein) λ ZAP II cDNA (Stratagene) inserted into pBluescript II KS (−) (Stratagene).

Oligonucleotide-Directed Mutagenesis and RNase Protection. Specific mutations were introduced into pMBgCAT at the nuclear factor binding sites indicated in Fig. 1 *Middle* by using oligonucleotide-directed mutagenesis (ref. 11, pp. 8.1.1–8.1.5). *MYC–CAT* constructs were cotransfected into CV-1 cells along with the pSV_2neo selectable marker vector by the method of Chen and Okayama (12). Stable transfectants (more than 100 Geneticin-resistant colonies) were obtained and total RNA was prepared as described previously (8). RNase protection was performed essentially as described (ref. 11, pp. 4.7.1–4.7.6) with some modifications. An RNA probe spanning the c-*MYC* promoter (−7 to +337) was prepared in pBluescribe (Stratagene) and 5×10^5 cpm was incubated with 30 μg of total RNA at 54°C for 12 hr in buffer containing 80% (vol/vol) formamide, digested with RNase, and electrophoresed on a 6% polyacrylamide sequencing gel, which was dried and autoradiographed. Bands were quantitated and ratios of P_1 to P_2 were determined by laser densitometry.

Library Screening and Filter Hybridization. A HeLa cell λgt11 library (gift of Paula Henthorn and Tom Kadesch, University of Pennsylvania) was screened by the modified method of Vinson *et al.* (13), using a DNA probe consisting of concatamerized ME1a1 sites. Filter binding assays of purified phage stock were performed essentially as above.

Fusion Protein Expression. *MAZ* was expressed as a fusion protein by cloning in the *Eco*RI site of pMAL-c and using conditions specified by the supplier (New England Biolabs).

Abbreviations: MBP, maltose-binding protein; MAZ, *MYC*-associated zinc finger protein; CAT, chloramphenicol acetyltransferase; GMSA, gel mobility shift assay.

¶To whom reprint requests should be addressed at: Department of Biochemistry and Cell Biology, Life Sciences Building, State University of New York, Stony Brook, NY 11794-5215.

‖The sequence reported in this paper has been deposited in the GenBank data base (accession no. M94046).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MAZ-MBP fusion protein was partially purified by passage over heparin-Sepharose in lysis buffer (ref. 11, p. 16.5.4) and eluted in lysis buffer containing 0.7 M KCl.

***In Vitro* Transcription and Translation.** pMAZ was digested with *Bam*HI or *Hin*dIII and transcribed with T3 or T7 RNA polymerase, respectively, using a Transprobe T kit (Pharmacia) to produce a capped transcript. Approximately 100 ng of mRNA was used in an *in vitro* translation reaction using a rabbit reticulocyte lysate (Promega). Brome mosaic virus mRNA was used as a control. Ten microliters was analyzed by electrophoresis in an SDS/10% polyacrylamide gel, dried, and autoradiographed.

Production of mRNA and Northern Analysis. mRNA was prepared from tissue culture cells by using the FastTrack kit (Invitrogen, San Diego) and analyzed by Northern hybridization using standard techniques (ref. 11, pp. 4.9.1–4.9.7). Human multiple tissue Northern blots were obtained from Clontech.

Gel Mobility Shift Assay (GMSA). GMSA was done as previously described (8) with minor modifications. Crude HeLa nuclear extract [prepared according to Dignam *et al.* (14)] was incubated with an end-labeled oligonucleotide probe with 2 μg of the nonspecific competitor poly(dI-dC)·poly(dI-dC) (Pharmacia). Binding to oligonucleotide probes by partially purified MAZ-MBP was assayed in the presence of 100 ng of nonspecific competitor.

RESULTS

Requirement of the ME1a1 Site for Appropriate P_1/P_2 Promoter Usage. The CV-1 monkey fibroblast line was stably transfected by $Ca_3(PO_4)_2$ precipitation with a murine c-*MYC* promoter–*CAT* fusion construct, pMBgCAT, containing either the wild type or site-directed mutations engineered into the ME1a1 or ME1a2 nuclear factor binding sites of the P_2 promoter (Fig. 1). Total cellular RNAs from large populations of stable transfectants were prepared and subjected to RNase protection analysis using a uniformly labeled RNA probe of the c-*MYC* promoter region and 3′ flanking exon 1 sequences (−7 to +337). Densitometric analysis revealed a normal P_1-to-P_2 ratio of 1:5 for the wild-type construct (Fig. 1, lane 4). When the P_2 promoter was mutated, converting a GGGG sequence in the core of the ME1a1 nuclear factor binding site to ATTC, the P_1-to-P_2 ratio shifted to 10:1, with P_2 activity virtually abolished (Fig. 1, lane 6). When the ME1a1 site was replaced by an octamer binding site (ME1a1 as oct), the normal P_1-to-P_2 ratio was not restored (Fig. 1, lane 3). When the ME1a2 site was mutated and tested as above, the P_1-to-P_2 ratio shifted to 1:2 (Fig. 1, lane 2), indicating that the ME1a2 site contributed to P_2 usage but was not as critical as ME1a1 in these fibroblastic cell lines.

cDNA Encoding a Polypeptide Exhibiting Sequence-Specific Binding to the ME1a1 Site. In light of the above data demonstrating that the ME1a1 site is required for P_2 usage, we sought to obtain a cDNA clone encoding a polypeptide that bound specifically to ME1a1. A HeLa cell λgt11 expression library was screened by using a DNA probe consisting of tandem repeats of the ME1a1 sequence (13). One million bacteriophage plaques were screened and a single positive clone (19-2-3-1) that survived multiple rounds of purification was obtained (Fig. 2*A*). When a mutant ME1a1 probe (containing the mutation shown in Fig. 1 *Middle* to abolish P_2 activity) was used in the filter binding assay, no signal was observed in comparison with the wild-type probe (Fig. 2*B*). Having shown that the Sp1 transcription factor binds to the ME1a1 site with low affinity (8), we sought to characterize the binding specificity of 19-2-3-1. We utilized a series of concatamerized oligonucleotide probes containing various mutations in the ME1a1 binding region in a filter binding assay, along with a probe containing high-affinity Sp1 binding

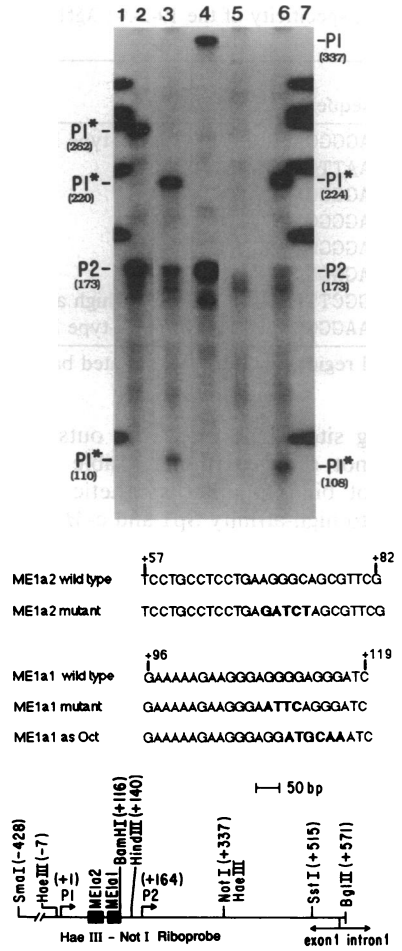

FIG. 1. RNase protection assay. Oligonucleotide-directed mutations were introduced into pMBgCAT to specifically abrogate binding to the ME1a1 or ME1a2 protein binding sites (see boldface type in the sequences in *Middle*). CV-1 cells were stably transfected, 30 μg of total RNA was incubated with a c-*MYC* RNA antisense probe (−7 to +337), and the resultant RNase digestion products were analyzed on a denaturing 6% polyacrylamide gel (*Top*). The asterisk indicates P_1-initiated RNAs, which were fragmented because a wild-type probe was used. The sizes of the P_1, P_1*, and P_2 bands are indicated in nucleotides. Lanes 1 and 7, molecular weight markers of 310, 281, 271, 234, 194, and 118 base pairs (bp) in descending order (*Hae* III-digested DNA of phage φX174); lane 2, ME1a2 mutant; lane 3, ME1a1 as oct; lane 4, pMBgCAT (wild type); lane 5, CV-1 nontransfected control; and lane 6, ME1a1 mutant. A map of murine c-*MYC* is shown in *Bottom*.

sites (GGGGCGGGGC) (15). Table 1 summarizes these results. An 8-bp sequence, GGGAGGGG, was revealed as the

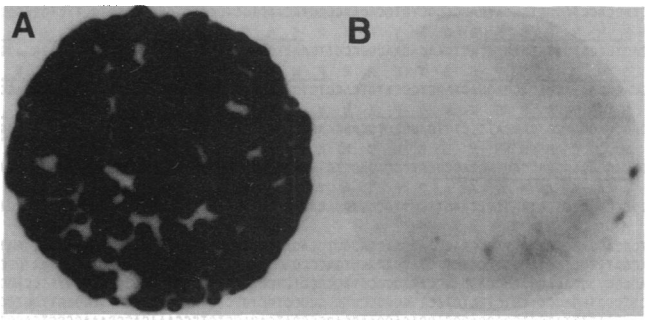

FIG. 2. Filter binding assay. λgt11 cDNA clone 19-2-3-1 probed with a concatamerized oligonucleotide probe. (*A*) Filter probed with wild-type ME1a1 binding sites. (*B*) Filter probed with the mutant ME1a1 binding site shown in Fig. 1.

Table 1. Binding specificity of the 19-2-3-1 λgt11 MAZ cDNA clone

Probe sequence		19-2-3-1 binding
GAAAAAGAAGGGAGGGGAGGGATC	(wild-type ME1a1)	+
GAAAAAGAAGGGAATTCAGGGATC		−
GAAAAACTTGGGAGGGGAGGGATC		+
GAAAATCTAGGGAGGGGAGGGATC		+
GAAATTCAAGGGAGGGGAGGGATC		+
GAATTTGAAGGGAGGGGAGGGATC		+
TCGACGGGGCGGGGCTTACTGC	(Sp1 high affinity)	−
TCCTGCCTCCTGAAGGGCAGCGTTCG	(wild-type ME1a2)	−

The underlined regions indicate the mutated bases in the ME1a1 oligonucleotide.

factor's binding site, since mutations outside of this core sequence did not reduce binding (Table 1). In addition, 19-2-3-1 did not bind to other synthetic oligonucleotides corresponding to high-affinity Sp1 and c-*MYC* ME1a2 binding sites (Table 1). The GGGAGGGG sequence is shorter than the large DNase I footprint previously published for the ME1a1 site, but it represents the core of the ME1a1 methylation interference pattern obtained with crude nuclear extract (8). We have additional evidence that a nuclear factor encoded by another gene binds just upstream of this core sequence and to a related motif in the ME1a2 site (unpublished observation).

The ME1a1 Binding Clone Encodes a Zinc Finger Protein. The 19-2-3-1 cDNA insert was sequenced (16, 17). A λ Zap II HeLa library (Stratagene) was screened with the 19-2-3-1 cDNA as a probe to obtain a larger clone, 2422 bp (Fig. 3*A*). The latter clone contained an ATG that was in a reasonable context to serve as an initiation codon according to Kozak (18) and a polyadenylylation signal (AATAAA, indicated in Fig. 3*A*). An open reading frame of 477 amino acids was identified (Fig. 3*A*) potentially encoding a polypeptide with a predicted molecular mass of approximately 52.5 kDa. *In vitro* transcription and translation revealed a polypeptide product with an apparent molecular mass of 58.5 kDa (Fig. 4) that, in a GMSA, generated a complex with an ME1a1 probe that migrated closely with the band obtained with crude HeLa nuclear extract (data not shown).

Upon analysis of the amino acid sequence, six potential zinc fingers were found of the Cys_2His_2 type (Fig. 3), three of which (fingers 1, 4, and 5 in Fig. 3*B*) conformed to the consensus $F/YXCX_{2-4}CX_3FX_5LX_2HX_{3-4}H$ (19), while the remaining three contained a substitution in one of the conserved hydrophobic residues thought to stabilize the formation of the zinc finger (fingers 2, 3, and 6 in Fig. 3*B*). Zinc-finger-containing proteins have been shown to bind DNA in a sequence-specific manner (20), and examples have been found among transcription factors (21, 22), developmental control genes (23), and at least one gene associated with neoplastic disease (24). A proline-rich region resides at the amino-terminal portion of the polypeptide and consists of 22% proline residues in the first 150 amino acids. Proline-rich regions can function as transcriptional activation domains as exemplified by AP-2 (25) and CCAAT transcription factor (CTF) (26). Another striking feature of the primary amino acid sequence is the presence of several polyalanine tracts (Fig. 3). These domains are thought to form α-helices and have been found in a number of other proteins involved in regulating *Drosophila* development, including runt (27), engrailed (28), and even-skipped (29). Searches of the GenBank and European Molecular Biology Laboratory data bases in Dec. 1992 did not reveal any similar published sequences. We designate this protein MAZ for *MYC*-associated zinc finger protein.

A

```
CGGCTCAGCGGGGGCCGAGGCCATGTTCCCGGTGTTTCCTTGCACGCTGCTGGCCCCCCCCTTCCCCGTGCTGGGCCTGGACTCCCGGGG 90
                  M  F  P  V  F  P  C  T  L  L  A  P  P  F  P  V  L  G  L  D  S  R  G
GGTGGGCGGCCTCATGAACTCCTTCCCGCCACCTCAGGGTCACGCCCAGAACCCCCTGCAGGTCGGGGCTGAGCTCCAGTCCCGCTTCTT 180
  V  G  G  L  M  N  S  F  P  P  P  Q  G  H  A  Q  N  P  L  Q  V  G  A  E  L  Q  S  R  F  F
TGCCTCCCAGGGCTGCGCCCAGAGTCCATTCCAGGCCGCGCCGGCGCCCCCGCCCACGCCCCAGGCCCCGGCGGCCGAGCCCCTCCAGGT 270
  A  S  Q  G  C  A  Q  S  P  F  Q  A  A  P  A  P  P  P  T  P  Q  A  P  A  A  E  P  L  Q  V
GGACTTGCTCCCGGTGCTCGCCGCCGCCCAGGAGTCCGCCGCGGCTGCTGCGGCCGCTGCCGCCGCTGCTGCCGCCGTCGCTGCCGCGCC 360
  D  L  L  P  V  L  A  A  A  Q  E  S  A  A  A  A  A  A  A  A  A  A  A  A  A  V  A  A  A  P
CCCGGCCCCTGCCGCCGCCTCTACGGTGGACACAGCGGCCCTGAAGCAGCCTCCGGCGCCCCCTCCGCCAACCCCGCCAGTGTCGGCGCC 450
  P  A  P  A  A  A  S  T  V  D  T  A  A  L  K  Q  P  P  A  P  P  P  P  P  P  P  V  S  A  P
CGCGGCCGAGGCCGCGCCCCCCGCCTCCGCCGCCACTATCGCCGCGGCGGCGGCCACCGCCGTCGTAGCCCCAACCTCGACGGTCGCCGT 540
  A  A  E  A  A  P  P  A  S  A  A  T  I  A  A  A  A  A  T  A  V  V  A  P  T  S  T  V  A  V
GGCCCCGGTCGCGTCTGCCTTGGAGAAGAAGACAAAGAGCAAGGGGCCCTACATCTGCGCTCTGTGCGCCAAGGAGTTCAAGAACGGCTA 630
  A  P  V  A  S  A  L  E  K  K  T  K  S  K  G  P  Y  I  C  A  L  C  A  K  E  F  K  N  G  Y
CAATCTCCGGAGGCACGAAGCCATCCACACGGGAGCCAAGGCCGGCCGGGTCCCCTCGGGTGCTATGAAGATGCCGACCATGGTGCCCCT 720
  N  L  R  R  H  E  A  I  H  T  G  A  K  A  G  R  V  P  S  G  A  M  K  M  P  T  M  V  P  L
GAGCCTCCTGAGCGTGCCCCAGCTGAGCGGAGCCGGCGGGGGAGGGGGAGAGGCGGGTGCCGGCGGCGGCGCTGCCGCAGTGGCCGCCGG 810
  S  L  L  S  V  P  Q  L  S  G  A  G  G  G  G  G  E  A  G  A  G  G  G  A  A  A  A  V  A  A  G
TGGCGTGGTGACCACGACCGCCTCGGGGAAGCGCATCCGGAAGAACCATGCCTGCGAGATGTGTGGCAAGGCCTTCCGCGACGTCTACCA 900
  G  V  V  T  T  T  A  S  G  K  R  I  R  K  N  H  A  C  E  M  C  G  K  A  F  R  D  V  Y  H
CCTGAACCGACACAAGCTGTCGCACTCGGACGAGAAGCCCTACCAGTGCCCGGTGTGCCAGCAGCGCTTCAAGCGCAAGGACCGCATGAG 990
  L  N  R  H  K  L  S  H  S  D  E  K  P  Y  Q  C  P  V  C  Q  Q  R  F  K  R  K  D  R  M  S
CTACCACGTGCGCTCACATGACGGCGCTGTGCACAAGCCCTACAACTGCTCCCACTGTGGCAAGAGCTTCTCCCGGCCGGATCACCTCAA 1080
  Y  H  V  R  S  H  D  G  A  V  H  K  P  Y  N  C  S  H  C  G  K  S  F  S  R  P  D  H  L  N
CAGTCACGTCAGACAAGTGCACTCAACAGAACGGCCCTTCAAATGTGAGAAATGTGAGGCAGCTTTCGCCACGAAGGATCGGCTGCGGGC 1170
  S  H  V  R  Q  V  H  S  T  E  R  P  F  K  C  E  K  C  E  A  A  F  A  T  K  D  R  L  R  A
GCACACAGTACGACACGAGGAGAAAGTGCCATGTCACGTGTGTGGCAAGATGCTGAGCTCGGCTTATATTTCGGACCACATGAAGGTGCA 1260
  H  T  V  R  H  E  E  K  V  P  C  H  V  C  G  K  M  L  S  S  A  Y  I  S  D  H  M  K  V  H
CAGCCAGGGTCCTCACCATGTCTGTGAGCTCTGCAACAAAGGTACTGGTGAGGTTTGTCCAATGGCGGCGGCAGCGGCAGCGGCGGCAGC 1350
  S  Q  G  P  H  H  V  C  E  L  C  N  K  G  T  G  E  V  C  P  M  A  A  A  A  A  A  A  A  A
GGCAGCAGCGGCAGCAGTAGCAGCCCCTCCCACAGCTGTGGGCTCCCTCTCGGGGGCGGAGGGGGTGCCTGTGAGCTCTCAGCCACTTCC 1440
  A  A  A  A  V  A  A  P  P  T  A  V  G  S  L  S  G  A  E  G  V  P  V  S  S  Q  P  L  P
CTCCCAACCCTGGTGAGCTCCAAGTTGGTTGCGGGGGAGAGGGGAGAATGGAGTAGAGTCCCTTGGTACAAGCTCCTCTCCCCCCTCTTT 1530
  S  Q  P  W●
TCCCACCAACTCCTATTTCCCTACCAACCAAGGAGCCTCCAGAAGGAAAGGAGGAAGAAATGTTTTCTTAGGGGAATTCGCTAGGTTTTA 1620
ACGATTTGCTTCTCCTGCTCCTCTTCTATCAGACCTGACCCCACACAAACCTGTCCCCTCGGTTGTGTTGAAGTCCCCTGGACAGTGGGC 1710
AGGGGTGGCAGAGGACACGAGCAGCCACTGCCCGTACCCCCTCTCCTCTCTGTAAGCCCATGCCCTGTCTTCCCAGGGACTTGTGAGCCT 1800
CTTCCCTCGACGGTCCTCTTCTCTCCTTCCAGTCCTCTCCCCCTGCTGTCTGCAGCCCCTCCCCGGGGAGTTGGTGCTTTCTTTTCCTTT 1890
TTTTTTTTTTTCCAGGGGGAGGGAGGAGAGGAAGGAGGGGGATCAGAGCTGTCCCAAAGAGGGAAAGCGGTGAGGTTTGAGGAGGGGCA 1980
GAAGCAGGGCCGGCAAAGGTTGTACCTTCATAAGGTGGTATCGGGGGGTTGGGGTCAGGCCCTGAACATCGTCCTACTTGAGAATCTGTC 2070
AGGGGAAAAAGTCAAGGGGAGCAGGAGGAAGAGCCAGGAGGGCCAGAGGCAGAGAAGAGATGGAGTCTTAGGGGCCAGGGTGAGCCAGGG 2160
GTCCAGGGCCTAGAGGTGCTTCTGGGGGGGGGGAATGCAGCCAGTGTCCCCCTCCCCTCTTCCACCCCAGCTCCAGCCCTGGTCTTGTC 2250
TTTTCATCCCTCTTCCCCACGACAGAAGAAGTTGTGGCCCTGGCATGTCATCGTGTTCCTGTGTCCCCTGCATGTACCCCACCCTCCACC 2340
CCTTCCTTTTGCGCGGACCCCATTACAATAAATTTTAAATAAAATCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2422
```

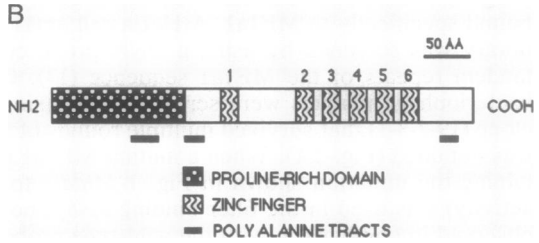

FIG. 3. (*A*) Nucleic acid and predicted amino acid sequence of cDNA encoding MAZ. Zinc fingers are underlined, polyalanine tracts are indicated by brackets, and the polyadenylylation signal is double-underlined. (*B*) Schematic of the protein product of *MAZ*, with scale bar indicating 50 amino acids.

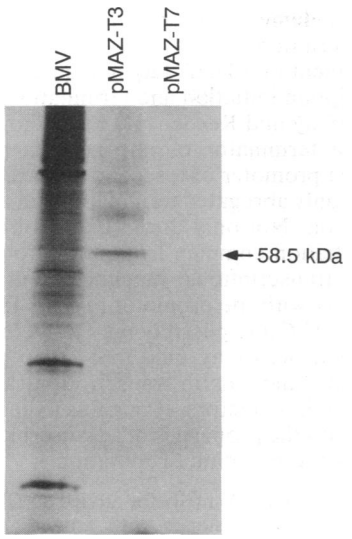

FIG. 4. SDS/PAGE of *in vitro* translated product of *MAZ*. BMV, brome mosaic virus mRNA; pMAZ-T3, mRNA transcript with T3 RNA polymerase; pMAZ-T7, mRNA transcript with T7 RNA polymerase.

***MAZ* Is Differentially Expressed in Human Tissues.** Northern analysis revealed that *MAZ* is predominately expressed as a broad band of approximately 2.4–3.0 kb (Fig. 5*A*), which was resolved into two species, 2.5 and 2.7 kb, in a shorter exposure (Fig. 5*B*). The 2.5- and 2.7-kb transcripts were present at comparable levels in all tissues tested except for kidney, where they were nearly absent (Fig. 5*A*). Prolonged exposure also revealed three minor species, one of 1.9 kb expressed mainly in brain, placenta, and lung, and to a lesser extent in heart, a second of 2.1 kb in pancreas, and a third of approximately 500 bp in brain and pancreas (Fig. 5*A*). The same Northern blot was reprobed with the ubiquitously expressed glyceraldehyde-3-phosphate dehydrogenase gene as an internal RNA quality control (Fig. 5*C*).

MAZ-MBP Protein Chimera Binds to Two Sites in the c-*MYC* Gene and to the Terminator of the *C2* Gene. Recent reports have identified sequences involved in transcriptional termination that are homologous to the ME1a1 site (5, 10). The region that contains the ME1a1 site was found to participate in premature termination of *MYC* P_1-initiated transcripts (5). A second ME1a1-like sequence, located toward the end of the c-*MYC* first exon, plays a role in the premature termination of P_2-initiated transcripts (c-*MYC* att, Fig. 6*B*) (5). A third sequence homologous to ME1a1 was found in the *C2* gene's terminator region, and similar, if not identical, proteins were found to be present in HeLa cell nuclear extracts that bound to the *C2* termination and ME1a1 sites (10). To further characterize the DNA-binding properties of MAZ, MAZ-MBP was prepared and analyzed by GMSA. A specific comigrating band appeared with all three probes (ME1a1, *C2*, and c-*MYC* att sites) and disappeared upon the addition of a 100-fold molar excess of a specific oligonucleotide competitor, but it was unaffected by the presence of a 100-fold molar excess of an unrelated competitor (see Fig. 6*A*). Each of the MAZ-MBP·DNA complexes was more retarded than the native complexes obtained with the HeLa crude nuclear extract (Fig. 6*A*). The MBP contributes 42 kDa to the recombinant MAZ-MBP fusion protein. Therefore, the migration of the MAZ-MBP·DNA complex could be more retarded than the DNA·protein complex obtained with crude nuclear extract. No binding was detected with MBP alone (data not shown). The three binding sites contain a GGGAGGG core motif (Fig. 6*B*), which agrees with

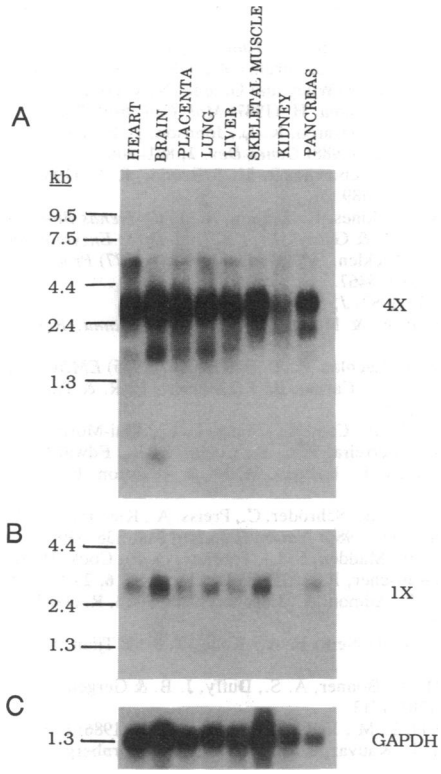

FIG. 5. Northern analysis of *MAZ* with various human tissue mRNAs. (*A*) Northern blot showing tissue distribution of MAZ mRNA, 16-hr exposure. (*B*) Same Northern blot as in *A* with a 4-hr exposure. (*C*) Control with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) probe.

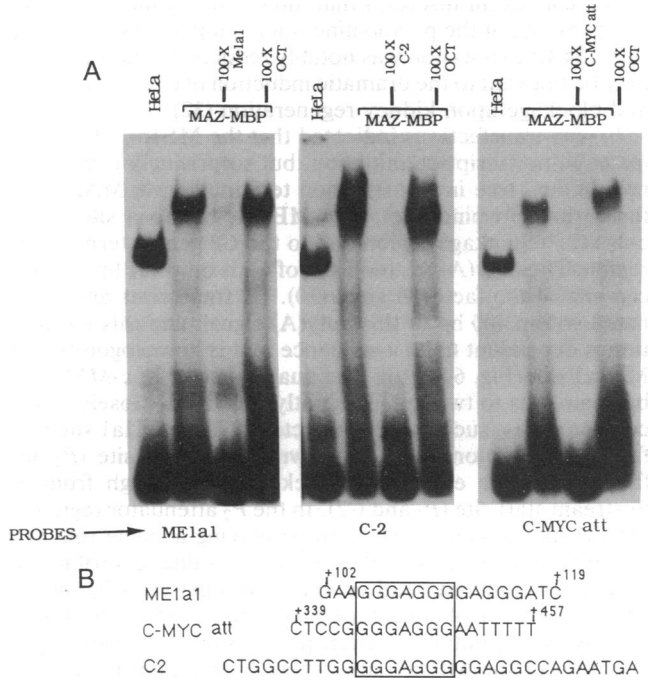

FIG. 6. GMSA of MAZ-MBP with various oligonucleotide probes. (*A*) MAZ-MBP incubated under conditions described in *Materials and Methods*, electrophoresed on a nondenaturing 4% polyacrylamide gel. Each probe was incubated with either 5 μg of HeLa nuclear extract or MAZ-MBP in the presence and absence of a 100-fold molar excess of a homologous competitor or an unrelated competitor containing an octamer (Oct) binding site as indicated at the top. (*B*) Sequence of the oligonucleotide probes used, with the consensus binding site boxed. Numbers given are relative to P_1.

the methylation interference pattern obtained with MAZ-MBP (data not shown). The differences in the intensities of the bound complexes could reflect the contributions of flanking sequences.

DISCUSSION

ME1a1 and ME1a2 define major nuclear factor binding sites in the c-*MYC* P_1 and P_2 promoters (8). Site-directed mutagenesis experiments revealed that initiation of c-*MYC* transcription at the gene's major promoter, P_2, requires a wild-type ME1a1 nuclear factor binding site. We also presented the structure and described some of the properties of a human cDNA clone encoding a zinc finger protein that specifically binds to a portion of the ME1a1 sequence. MAZ possesses the architecture of a transcription factor in that it has six potential zinc fingers and a proline-rich region, structures which are known to function as DNA-binding and transcriptional-activation domains, respectively (20–23, 25, 26). Another intriguing feature of the primary amino acid sequence is the existence of polyalanine tracts. These structures have the potential to form α-helices, and related sequences at the amino terminus of the kruppel gene product line up with polyalanine tracts present in other *Drosophila* gene products such as even-skipped and engrailed (30). The polyalanine region of kruppel was shown to facilitate the repression of a reporter gene with multiple kruppel binding sites in mammalian cells (30), while the polyalanine tract and adjacent sequences functioned at low concentration as a transcriptional activator in *Drosophila* cells (31).

Northern analysis revealed that *MAZ* is expressed in human heart, brain, placenta, lung, liver, skeletal muscle, and pancreas (Fig. 5*A*). Its mRNA exists predominately as two major species, approximately 2.5 and 2.7 kb. Minor species were seen in heart, brain, placenta, lung, and pancreas. Indeed, we have obtained a second cDNA of *MAZ*, which carries an insertion that alters the coding region and removes one of the polyalanine tracts (unpublished observation). *MAZ* expression was notably reduced in kidney, which may be relevant to the dramatic induction of c-*MYC* initiation and blockage upon kidney regeneration (32).

In vivo transfections indicated that the ME1a1 site participates in transcription initiation, but surprisingly it may also have a dual role in transcription termination. A MAZ-MBP fusion protein binds not only ME1a1, but to a site in the c-*MYC* P_2 blockage region and to the *C2* gene's termination region. The poly(A) addition site of *C2* is only 421 bp 5′ of the cap site of the factor B gene (10). *C2* transcripts are terminated within 400 bp of this poly(A) signal, and this termination is dependent upon a sequence that is homologous to the ME1a1 site (Fig. 6*B*) (10). The dual promoter of c-*MYC* may be analogous to two independently regulated, closely linked cellular genes, such as *C2* and factor B. The ME1a1 site may facilitate initiation from the downstream start site (P_2 and factor B) while effectively blocking readthrough from an upstream start site (P_1 and *C2*). In the P_2 attenuator region an ME1a1-like sequence seems to be playing a single role as a termination factor, as contrasted with its dual contributions to initiation and termination when it lies between P_1 and P_2. Close inspection of our RNase protection experiment (Fig. 1) reveals that P_1-initiated transcripts are slightly elevated upon mutation of the ME1a1 site, while P_2 initiation is virtually abrogated. This observation suggests that the ME1a1 mutation might have increased P_1 activity by abrogating its transcriptional block, thereby allowing for enhanced polymerase readthrough. Different sequences flanking the MAZ binding site may help to dictate MAZ properties by facilitating its interactions with different accessory factors. The binding of MAZ may also cause a perturbation in the local region of DNA, possibly generating a bend that could bring other cis-acting elements into position to exert a regulatory effect on transcription.

The involvement of a DNA target sequence in the regulation of transcription initiation and termination is not without precedent. McStay and Reeder (33) have shown that in two rRNA genes a terminator of one gene is present 60 bp upstream of the promoter of a second gene. Deletion of this terminator not only abrogated termination but also negatively affected initiation. Not only does binding of the terminator protect the initiation complex from readthrough by the upstream gene's transcriptional machinery but also this site directly interacts with the promoter region. This situation is analogous to c-*MYC* and possibly the *C2*–factor B scenarios. ME1a1-like sequences may bind more than one protein to accomplish dual functions in transcription, and MAZ may represent one of these factors. It remains to be tested directly whether MAZ has the properties of a transcription factor that can effect either termination or initiation.

We thank Ms. Margery Connelly for advice on DNA sequencing, Ms. Margo Reyes for assistance with artwork and manuscript preparation, and Ms. Rebecca Ashfield and Dr. Nick Proudfoot for supplying information prior to publication. C.A. was supported by a grant from the Medical Research Council of Canada. This work was supported by National Institutes of Health Grant CA36246 awarded to K.B.M.

1. Marcu, K. B., Bossone, S. A. & Patel, A. P. (1992) *Annu. Rev. Biochem.* **61**, 809–860.
2. Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N. & Weintraub, H. (1990) *Science* **250**, 1149–1151.
3. Blackwood, E. M. & Eisenman, R. N. (1991) *Science* **251**, 1211–1217.
4. Eilers, M., Schirm, S. & Bishop, J. M. (1991) *EMBO J.* **10**, 133–141.
5. Wright, S., Mirels, L. F., Calayag, M. C. B. & Bishop, J. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 11383–11387.
6. Nepveu, A. & Marcu, K. B. (1986) *EMBO J.* **5**, 2859–2866.
7. Bentley, D. L. & Groudine, M. (1986) *Nature (London)* **321**, 702–706.
8. Asselin, C., Nepveu, A. & Marcu, K. B. (1989) *Oncogene* **4**, 549–558.
9. Miller, H., Asselin, C., Dufort, D., Yang, J.-Q., Gupta, K., Marcu, K. B. & Nepveu, A. (1989) *Mol. Cell. Biol.* **9**, 5340–5349.
10. Ashfield, R., Enriquez-Harris, P. & Proudfoot, N. J. (1991) *EMBO J.* **10**, 4197–4207.
11. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1987) *Current Protocols in Molecular Biology* (Wiley and Greene, New York).
12. Chen, C. & Okayama, H. (1987) *Mol. Cell. Biol.* **7**, 2745–2752.
13. Vinson, C. R., LaMarco, K. L., Johnson, P. F., Landschulz, W. H. & McKnight, S. L. (1988) *Genes Dev.* **2**, 801–806.
14. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
15. Kadonaga, J., Jones, K. & Tjian, R. (1986) *Trends Biochem.* **11**, 20–23.
16. Maxam, A. M. & Gilbert, W. (1977) *Methods Enzymol.* **65**, 499–559.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Kozak, M. (1987) *J. Mol. Biol.* **196**, 947–950.
19. Johnson, P. F. & McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58**, 799–839.
20. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609–1614.
21. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. (1987) *Cell* **51**, 1079–1090.
22. Sukhatme, V. P., Cao, X., Chang, L. C., Tsai-Morris, C. H., Stamenkovich, D., Ferreira, P. C. P., Cohen, D. R., Edwards, S. A., Shows, T. B., Curran, T., LeBeau, M. M. & Adamson, E. D. (1988) *Cell* **53**, 37–43.
23. Rosenberg, U. B., Schröder, C., Preiss, A., Kienlin, A., Côté, S., Riede, I. & Jäckle, H. (1986) *Nature (London)* **319**, 336–339.
24. Morris, J. F., Madden, S. L., Tournay, O. E., Cook, D. M., Sukhatme, V. P. & Rauscher, F. J., III (1991) *Oncogene* **6**, 2339–2348.
25. Williams, T., Admon, A., Luscher, B. & Tjian, R. (1988) *Genes Dev.* **2**, 1557–1569.
26. Mermod, N., O'Neill, E. A., Kelly, T. J. & Tjian, R. (1989) *Cell* **58**, 741–753.
27. Kania, M. A., Bonner, A. S., Duffy, J. B. & Gergen, J. P. (1990) *Genes Dev.* **4**, 1701–1713.
28. Macdonald, P. M., Ingham, P. & Struhl, G. (1986) *Cell* **47**, 721–734.
29. Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. (1985) *Cell* **40**, 37–43.
30. Licht, J. D., Grossel, M. J., Figge, J. & Hansen, U. M. (1990) *Nature (London)* **346**, 76–79.
31. Sauer, F. & Jäckle, H. (1991) *Nature (London)* **353**, 563–566.
32. Asselin, C. & Marcu, K. B. (1989) *Oncogene Res.* **5**, 67–72.
33. McStay, B. & Reeder, R. H. (1990) *Mol. Cell. Biol.* **10**, 2793–2800.