

# The 5'-AT-rich half-site of Maf recognition element: a functional target for bZIP transcription factor Maf

Tomonori Yoshida<sup>1,2,\*</sup>, Tsuyoshi Ohkumo<sup>1,3</sup>, Shoko Ishibashi<sup>1,4</sup> and Kunio Yasuda<sup>1</sup>

<sup>1</sup>Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama 8916-5, Ikoma, Nara, 630-0101, Japan, <sup>2</sup>Regulatory Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA, <sup>3</sup>Graduate School of Frontier Biosciences, Osaka University, Suita, Osaka, 565-0871, Japan and <sup>4</sup>Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK

Received as resubmission March 15, 2005; Accepted May 23, 2005

## ABSTRACT

**The Maf family of proteins are a subgroup of basic region-leucine zipper (bZIP) transcription factors, which recognize a long palindromic DNA sequence [TGCTGAC(G)TCAGCA] known as the Maf recognition element (MARE). Interestingly, the functional target enhancer sequences present in the  $\alpha$ -A-crystallin gene contain a well-conserved half-site of MARE rather than the entire palindromic sequence. To resolve how Maf proteins bind to target sequences containing only MARE half-sites, we examined their binding activities using electrophoretic gel mobility shift assays as well as *in vitro* and *in vivo* reporter assays. Our results indicate that the 5'-flanking region of the MARE half-site is required for Maf proteins to bind both *in vitro* and *in vivo*. The critical 5'-flanking sequences for c-Maf were determined by a selection and amplification binding assay and show a preference for AT-rich nucleotides. Furthermore, sequence analysis of the regulatory regions of several target genes also suggests that AT-rich sequences are important. We conclude that Maf can bind to at least two types of target sequences, the classical MARE (palindrome type) and a 5'-AT-rich MARE half-site (half-site type). Our results provide important new insights into the DNA binding and site selection by bZIP transcription factors.**

## INTRODUCTION

Many basic region-leucine zipper (bZIP) type transcription factors have been identified to date. They can be divided into a variety of subgroups within the bZIP superfamily [reviewed in (1)]. Their common feature is the bZIP domain

that contains a region rich in basic and hydrophobic residues with a leucine every seven amino acids. In order to bind to specific target DNA sequences, bZIP transcription factors form homo- or heterodimers through their leucine zipper region. The leucine zipper forms a coiled-coil structure, consisting of two parallel  $\alpha$ -helices, that mediates dimerization. The basic residues located on the N-terminal side of the leucine zipper recognize specific DNA sequences and contact the DNA directly. Because these transcription factors form dimer-DNA complexes in which each monomer interacts with adjacent half-sites, target sequences represent a symmetric or nearly symmetric DNA sequence. For example, the bZIP transcription factors GCN4, CREB and AP-1 (Fos/Jun heterodimer) recognize TGA(C/G)TCA, TGACGTCA and TGACTCA, respectively (2–8).

The *v-maf* oncogene was initially identified as a transforming gene of the avian retrovirus AS42 (9,10). Its cellular homolog, the *c-maf* proto-oncogene, encodes a bZIP transcriptional factor with homology to *c-fos* and *c-jun*. It also contains the Maf extended homology region (EHR), forming a distinct subgroup in the bZIP superfamily (9,11). Members of the Maf family are usually subdivided into two classes: the large Maf proteins—c-Maf (10), MafB (12), NRL (13) and L-Maf (14); and the small Maf proteins—MafG (15), MafF and MafK (16). Large Maf subfamily members have a transactivation domain, whereas small Maf family members do not. Small Maf proteins have been shown to be a sub-unit of the NF-E2 transcription factor that regulates erythroid-specific gene expression (15). Furthermore, small *maf* genes are also involved in the oxidative stress response and central nervous system development [reviewed in (17)].

The target DNA sequences of v-Maf, identified by selection and amplification binding (SAAB) assay, are relatively long palindromic sequences (TGCTGACTCAGCA and TGCTGACGTCAGCA), which were termed Maf recognition elements (MAREs) (10). Known MAREs include the phorbol 12-O-tetradecanoate-13-acetate-responsive element

\*To whom correspondence should be addressed. Tel: +1 858 453 4100, ext. 1851; Fax: +1 858 535 8194; Email: yoshida@salk.edu

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org

(TRE; TGACTCA) and the cyclic AMP-responsive element (CRE; TGACGTCA), to which the AP-1 and CREB/ATF family of bZIP transcription factors bind. Importantly, Maf proteins can form heterodimers among family members, as well as with other bZIP proteins, to bind to MAREs and MARE-like sequences (18). For example, the AP-1 transcription factor complex is composed of homo- and heterodimers of the Jun and Fos family (19). Maf can form a heterodimer with Jun and Fos and bind to MAREs (18). *c-maf*, *mafB* and *L-maf* can transform chicken embryonic fibroblast cells when over expressed, showing their oncogenic activities (10,12). A recent study showed that Maf-induced transformation requires transactivation through MARE and that Maf and Jun share downstream target genes involved in cell transformation (20). Therefore, it is likely that Maf proteins have a function for promoting cell proliferation through MAREs.

We have previously identified two lens-specific enhancer elements,  $\alpha$ CE1 and  $\alpha$ CE2, in the promoter region of the chicken  $\alpha$ A-crystallin gene (21). L-Maf was identified as a factor that binds to the  $\alpha$ CE2 element and activates the  $\alpha$ A-crystallin gene (14). Interestingly, the  $\alpha$ CE2 element contains only a half-site of MARE [CTCCGCATTCTGCTGACCAC] (21)] raising the question of how L-Maf functions through the element.

In this study, we find that Maf proteins bind to the half-site and that the 5'-flanking region of the half-site is critical for this activity. Furthermore, we show that the preferred 5'-flanking sequences are rich in A or T. Based on our results, we conclude that Maf proteins can bind to two target sequences: the well-described palindromic MARE and a 5'-AT-rich MARE half-site. These results shed new light onto DNA binding and site selection by bZIP transcription factors and, as such, advanced our understanding of this important class of regulatory factors.

## MATERIALS AND METHODS

### Plasmid construction

For effector plasmids used in the reporter assays, the full-length open reading frames for chicken *L-maf*, *mafB* and *c-maf* were cloned into pEFX3-flag. To generate VP16-LdN166, VP16-LdN175 and VP16-LdN199, the VP16-activation domain, derived from pCMX-VP16 (25), was inserted into pEFX3-flag, which carried a mutation in one of HindIII sites (N-terminal side of FLAG sequence). L-Maf derivatives (LdN166, LdN175 and LdN199) were amplified by PCR and cloned into the Asp718 (Roche)/BamHI sites of pEFX3-fg-VP16. The pEFX3-LdN166 used in this paper is identical to pEFX3-fg-L-maf (bZIP) (26). For pEFX3-fg-VP16-MafG, pET15-MafG was kindly provided by Dr Yamamoto. The mouse MafG-coding sequence was amplified by PCR and inserted as a Asp718/BamHI fragment into pEFX3-fg-VP16. Reporter plasmids, p $\alpha$ A-1 to 6 and p $\alpha$ A-1(pal)X6- $\beta$ Luc, were generated as follows. Oligonucleotides were annealed and inserted into the Asp718/BamHI site of pSP73 (Promega). Each fragment generated by BamHI/HindIII digest was cloned into the BamHI/HindIII site of pSP73- $\alpha$ A-#. This process was repeated to generate six copies of  $\alpha$ A-# in pSP73 (pSP73- $\alpha$ A-#X6). Finally, the six copies of  $\alpha$ A-# fragment were inserted into pGVB- $\beta$ Luc. pGVB (TOYO INKI) was modified to make pGVB- $\beta$ Luc,

which has the multicloning sites followed by the chicken  $\beta$ -actin basal promoter (-55 to +53) inserted upstream of the luciferase gene.  $\alpha$ A-3 has been described as pmt- $\alpha$ A-3- $\beta$ Luc in a previous report (14). To make  $\alpha$ A-#X6-GFP, the *luciferase* gene was replaced with green fluorescent protein (GFP) (27) in  $\alpha$ A-3- $\beta$ Luc. For recombinant proteins, the full-length cDNAs for chicken *L-maf* and *mt-L-maf* were cloned into pGEX-6P-1 (Pharmacia) in-frame. mt-L-maf was constructed by the quick mutagenesis method (Stratagene), following the manufacturer's instructions.

### Cell culture and transfection

Primary cell cultures were performed as described previously (14,28). Transient transfections were carried out by calcium phosphate co-precipitation as described previously (28). For the reporter assay, each transcription mixture contained 300 ng of the luciferase reporter plasmid, 250 ng of pUC119 plasmid as carrier, 160 ng of pEFX3 empty plasmid, 60 ng of expression plasmid (pEFX3-fg-*maf* plasmids) and 250 ng of the  $\beta$ -galactosidase expression plasmid (pEFX- $\beta$ -gal) (14) as an internal control. Cells were cultured in 12-well plates (Coaster). Forty-eight hours post-transfection, the cells were harvested, and luciferase and  $\beta$ -galactosidase activities were measured. The  $\beta$ -galactosidase activities were used for normalization of the transfection efficiency. Each transfection was performed in duplicate and data were averaged from a minimum of three independent experiments.

### Protein expression and purification

To obtain recombinant Maf proteins (L-Maf, c-Maf and mt-L-Maf), glutathione S-transferase (GST)-fusion protein expression was induced by the addition of isopropyl- $\beta$ -D-thiogalactopyranoside (final 0.1 mM) and incubation at 37°C for 3 h. After induction, cell pellets were lysed by sonication and cleared by centrifugation. Cleared lysates were incubated for 2 h at 4°C with glutathione sepharose 4B beads (Amersham Pharmacia). GST-Maf proteins were released off the beads by PreScission Protease (Amersham Pharmacia), which cleaved between the GST moiety and Maf protein. Purified recombinant Maf proteins were used for electrophoretic mobility shift assay (EMSA) (Figures 1B, 2A, 2C, 7A, 7B, 8C and 8D) and DNase I footprint assay (Figure 7C).

### SAAB assay

For the SAAB assay, GST-c-Maf was eluted by reduced Glutathione (GSH) from glutathione sepharose 4B beads. GSH was removed via dialysis. The purified GST-c-Maf protein was re-bound to glutathione sepharose 4B beads and then mixed with double-stranded random oligonucleotides. The oligonucleotide sequence was of the form 5'-TGGGCACTATTTA-TATCAAC-N25-AATGTCGTTGGTGGCCC-3' containing a T7 promoter sequence. The oligonucleotides were annealed with primer-1 (5'-CGCGGATCCTAATACGACTCACTA-TAGGGGCCACCAACGACATT-3') and the DNA was made double stranded with Klenow fragment. The binding reactions were performed at 4°C for 30 min. Reactions were washed three times with NET-150 (20 mM Tris-HCl, pH 7.4, 1 mM EDTA and 150 mM NaCl). DNA bound to c-Maf was then purified by phenol-chloroform extraction.

The recovered DNA was subjected to a PCR using primer-1 and primer-2 (5'-CCCGACACCCGCGGATCCATGGGCA-CTATTATATCAC-3'). The DNA was amplified 18 times using the following cycles: 30 s at 94°C, 30 s at 50°C and 30 s

at 72°C. After three selection, the purified fragments were digested with BamHI cloned into pBluescript and sequenced.

### Electrophoretic mobility shift assays

Complementary oligonucleotides were annealed in annealing buffer (10 mM Tris-HCl, pH 8.0, 100 mM NaCl) at 70°C and cooled slowly to room temperature. Oligonucleotides (10 pmol) were end-labeled with [ $\alpha$ -<sup>32</sup>P]dATP by filling in the GATC overhang using Klenow fragment or with [ $\gamma$ -<sup>32</sup>P]ATP using T4 polynucleotide kinase.  $\alpha$ A series oligonucleotides contained cohesive ends while 40 series oligonucleotides carried blunt ends. Recombinant Maf proteins (Figures 1B, 2A, 2C, 7A, 7B, 8C and 8D) or *in vitro* translated proteins (1–5  $\mu$ l) produced by the TNT Coupled Reticulocyte Lysate System (Promega) (Figures 3B, 4B and 8B) were incubated with labeled probe (7.5–10 fmol) in 10  $\mu$ l of Reaction Buffer [10 mM HEPES-KOH, pH 7.9, 0.1 mM EDTA, 2 mM DTT, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 0.5  $\mu$ g/ $\mu$ l BSA, 20% glycerol and 30–100 ng/ $\mu$ l poly(dI-dC)-poly(dI-dC)] for 30 min at 30°C. DNA-protein complexes were resolved by electrophoresis through a 4% polyacrylamide gel (acrylamide: bisacrylamide ratio of 39:1) in running buffer (20 mM Tris, 196 mM glycine and 1 mM EDTA) for 2–2.5 h at 150 V.

### DNase I footprint assays

Binding reactions were performed as described for EMSA. Labeled probes were prepared by PCR against a plasmid carrying 1 copy of  $\alpha$ A or its variants and a <sup>32</sup>P-labeled primer. After the binding reactions, DNase I (12.5 ng) was added, and the samples were incubated for 10 min at room temperature. After ethanol precipitation, denatured samples were resolved by electrophoresis through a 8% polyacrylamide sequencing gel (SequaGel, National Diagnostics) in running buffer (0.5 $\times$  TBE).

### Transgenesis of *Xenopus laevis*

Transgenesis of *X.laevis* was performed essentially as described by Kroll and Amaya (29). We used XmnI to linearize the reporter and control plasmids for the REMI reactions.

## RESULTS

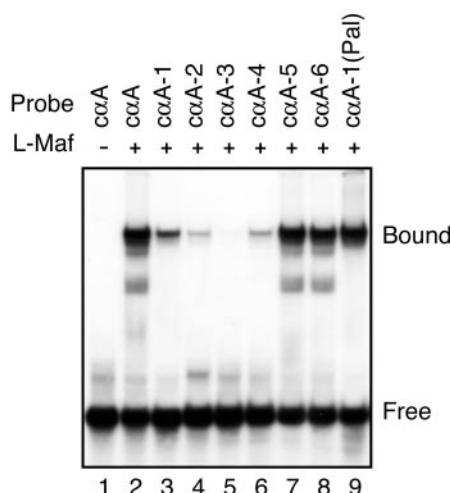
### L-Maf requires 5'-flanking sequence to bind to the MARE half-site

*In vitro* experiments have shown that Maf homodimers recognize palindromic sequences, termed MAREs (30).

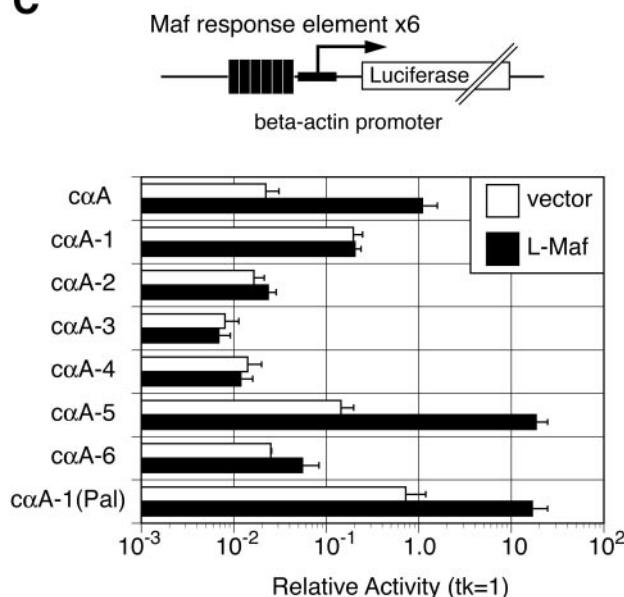
A

Probe	Sequence
$\alpha$ A	CATTTCT <u><b>TGCTGACC</b></u> ACAGTTGCCCTTC
$\alpha$ A-1	Cgcct <u><b>TCTGCTGACC</b></u> ACAGTTGCCCTTC
$\alpha$ A-2	CATTgag <u><b>GCTGACC</b></u> ACAGTTGCCCTTC
$\alpha$ A-3	CATTTC <u><b>TcgAGACC</b></u> ACAGTTGCCCTTC
$\alpha$ A-4	CATTTC <u><b>TGCTtca</b></u> ACAGTTGCCCTTC
$\alpha$ A-5	CATTTC <u><b>TGCTGAC</b></u> aggGTTGCCCTTC
$\alpha$ A-6	CATTTC <u><b>TGCTGACC</b></u> AccggGCCCTTC
$\alpha$ A-1(Pal)	Cgcct <u><b>TGCTGACTCAGCA</b></u> GCCTTC
$\alpha$ (Pal)	CATTTC <u><b>TGCTGACTCAGCA</b></u> GCCTTC
$\alpha$ (PalAT)	CATTTC <u><b>TGCTGACTCAGCA</b></u> ACTTTA

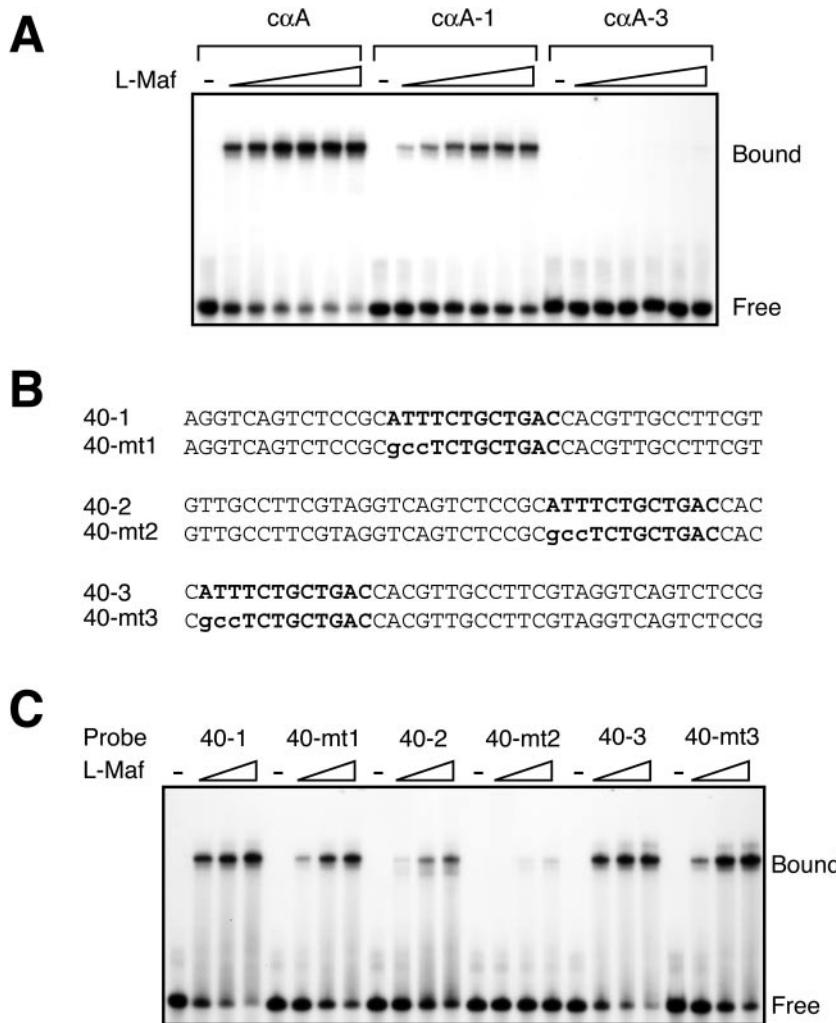
B



C



**Figure 1.** The 5'-flanking region of the Maf recognition element is important for transcriptional activity of L-Maf. (A) The sequences of the oligonucleotide probes used in this study. Substituted nucleotides are shown in lowercase letters. MARE is shown in bold capital letters. (B) EMSA using recombinant L-Maf ( $2.5 \times 10^{-7}$  nM) was analyzed for its ability to bind to the oligonucleotide probes shown in (A). (C) The reporter construct is schematically represented at the top of the figure. Closed rectangles represent the six copies of  $\alpha$ A, its mutants ( $\alpha$ A-1–6), or palindromic MARE [ $\alpha$ A-1(Pal)], and the bold line represents the chicken  $\beta$ -actin minimal promoter. Neural retina cultures were transiently co-transfected with an *L-maf* expression plasmid and each reporter construct. Luciferase activities are represented, relative to that of the luciferase gene driven by the HSV tk promoter (tk-Luc). Average values and error bars represent duplicate assays of at least three individual experiments.



**Figure 2.** The effect of the 5'-flanking region on Maf binding (A) Various amounts of L-Maf ( $0$ ,  $5.0 \times 10^1$ ,  $1.3 \times 10^2$ ,  $2.5 \times 10^2$ ,  $3.8 \times 10^2$ ,  $5.0 \times 10^2$  and  $1.0 \times 10^3$  nM) were subjected to EMSA with c $\alpha$ A, c $\alpha$ A-1 and c $\alpha$ A-3 oligonucleotide probes. (B) The sequences of probes used in (C). The 5'-flanking region and MARE half-site are indicated in bold. Mutated sequences are indicated by lower case letters. The oligonucleotides are as shown and do not contain any additional sequences on either end. (C) Increasing amounts of L-Maf ( $0$ ,  $2.5 \times 10^2$ ,  $5.0 \times 10^2$  and  $1.0 \times 10^3$  nM) were subjected to EMSA with the oligonucleotide probes shown in (B).

*mabB*, *c-maf* and *L-maf* are expressed in the lens and regulate lens-specific expression of crystallin genes (14,26,31–34). However, lens-specific enhancer elements contain only a half-site of MARE (28). To investigate the mechanisms by which L-Maf binds to the MARE half-site, we performed EMASAs using a series of mutant oligonucleotide probes, including a Maf-responsive element derived from the  $\alpha$ A-crystallin promoter (c $\alpha$ A), its mutant derivatives (c $\alpha$ A-1–6) and a palindromic MARE [c $\alpha$ A-1(Pal)] (Figure 1A). Recombinant L-Maf protein efficiently bound the c $\alpha$ A MARE half-site derived from the chicken  $\alpha$ A-crystallin enhancer,  $\alpha$ CE2, consistent with previous reports (26) (Figure 1B). Mutations in the core sequences of the half-site (c $\alpha$ A-2, c $\alpha$ A-3 and c $\alpha$ A-4) dramatically reduced the DNA-binding activity (Figure 1B, lanes 4, 5 and 6). Mutations outside of the core sequences at the 3' end (c $\alpha$ A-5 and c $\alpha$ A-6) did not affect the binding activity of L-Maf (Figure 1B, lanes 7 and 8). Surprisingly, we found that mutations in the 5'-flanking region of the MARE half-site (c $\alpha$ A-1) caused a significant

reduction of the DNA-binding activity of L-Maf (Figure 1B, lane 3). Furthermore, the introduction of an adjacent half-site to form a palindromic MARE sequence rescued the binding reduction caused by the mutation in 5'-flanking region [c $\alpha$ A-1(Pal); Figure 1B, lane 9].

To address the functional consequences of these mutations on transcriptional activation by L-Maf, reporter plasmids carrying six copies of the c $\alpha$ A, its mutant derivatives or palindromic MARE were co-transfected into primary retina cultures with an *L-maf* expression vector. As shown in Figure 1C, L-Maf could activate reporter gene expression through c $\alpha$ A, c $\alpha$ A-5, c $\alpha$ A-6 or c $\alpha$ A-1(Pal) (Figure 1C), but not via c $\alpha$ A-1, c $\alpha$ A-2, c $\alpha$ A-3 or c $\alpha$ A-4. Interestingly, we found that there was poor activation through c $\alpha$ A-6 even though L-Maf binds well. We speculate that the mutation created a novel binding sequence and that an unknown factor may repress activation through c $\alpha$ A-6. Taken together, the binding data from the EMSA and the expression data from the reporter indicate that L-Maf requires the 5'-flanking region

to bind to the MARE half-site and that this interaction is critical for gene activation *in vivo*.

In order to know how effective the 5'-flanking region is for Maf proteins to bind  $\text{c}\alpha\text{A}$ , we compared the binding specificity among  $\text{c}\alpha\text{A}$ ,  $\text{c}\alpha\text{A-1}$  and  $\text{c}\alpha\text{A-3}$  by EMSA (Figure 2). We used various amount of L-Maf proteins, and K constants estimated for  $\text{c}\alpha\text{A}$  and  $\text{c}\alpha\text{A-1}$  are  $\sim 2.5 \times 10^2$  and  $5.0 \times 10^2$  nM, respectively. Furthermore, we examined the effect of 5'-flanking region in the various location of oligonucleotides (Figure 2B). Interestingly, Maf still can bind to 40-2 that lacks the 3'-region next MARE half-site although the affinity is lower than that for 40-1. In this situation, it is clear that 5'-flanking region is important for binding. These results show that 5'-flanking region is important to bind MARE half-site.

### Helix H1 in the extended homology region is involved in the binding of the 5'-flanking sequence

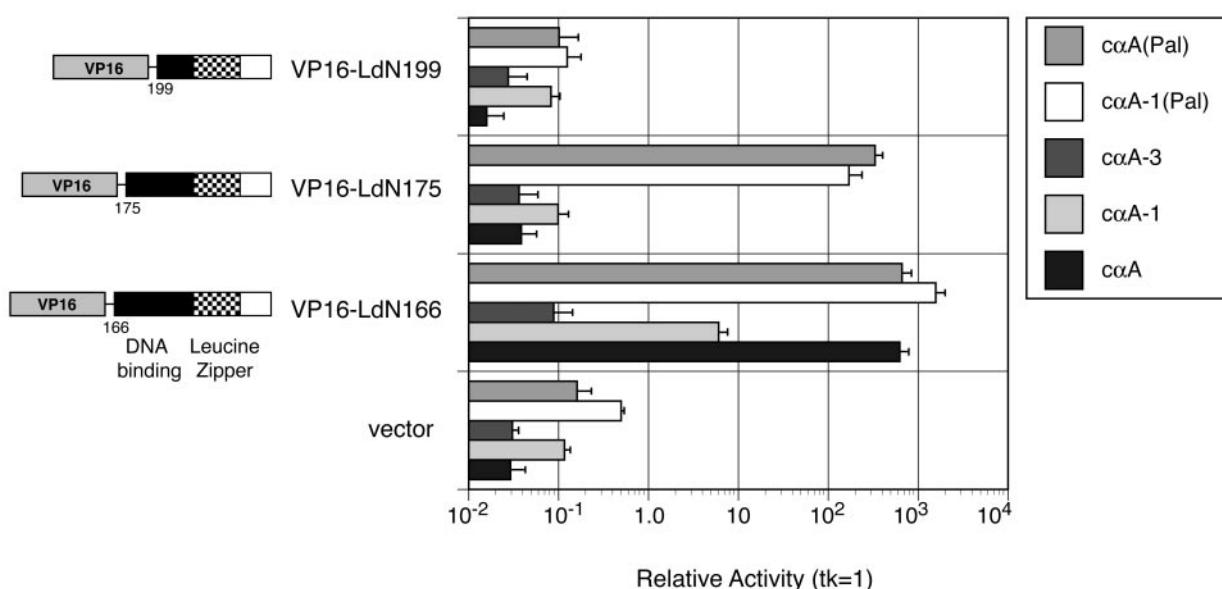
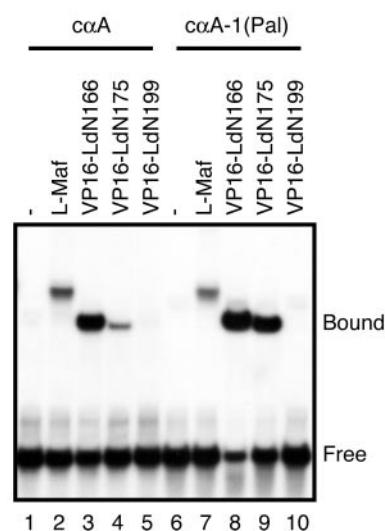
Our data show that the 5'-flanking region of the MARE half-site is required for interaction with L-Maf. To define the region(s) of Maf proteins responsible for the recognition of the 5'-flanking region, we introduced deletions to the DNA-binding domain of L-Maf (Figure 4A). An *L-maf* expression vector, containing the VP16 activation domain instead of its own, was co-transfected into primary retina cultures with a luciferase reporter gene driven by  $\text{c}\alpha\text{A}$  or its mutant derivatives (Figure 3A). As expected, VP16-LdN166, containing the whole DNA-binding domain of L-Maf, activated luciferase activity via six copies of the  $\text{c}\alpha\text{A}$  ( $\text{c}\alpha\text{AX6-}\beta\text{Luc}$ ) (Figure 3A). However, we note that VP16 activation domain shows higher activity than that of L-Maf. The structures of the mutant forms of the VP16/L-Maf chimeric proteins used as effectors are shown schematically in Figure 3A. VP16-LdN166 retained the ability to activate both reporter genes  $\text{c}\alpha\text{A(Pal)}$  and  $\text{c}\alpha\text{A-1(Pal)}$ . VP16-LdN199, containing only the basic region, failed to activate either reporter. VP16-LdN175 promoted the luciferase activities driven by the palindromic MAREs [ $\text{c}\alpha\text{A(Pal)}$  and  $\text{c}\alpha\text{A-1(Pal)}$ ], but failed to activate the reporter gene through the  $\text{c}\alpha\text{A}$ , MARE half-site. These observations indicate that the nine amino acids between 166 and 174 are necessary for the ability of L-Maf to activate transcription through the MARE half-site but not the palindromic MARE.

We next tested the DNA-binding activities of these chimeric proteins on  $\text{c}\alpha\text{A}$  or the palindromic MARE [ $\text{c}\alpha\text{A-1(Pal)}$ ] by EMSA (Figure 3B). As shown in Figure 3B, L-Maf and VP16-LdN166 bound to both target sequences efficiently (Figure 3B, lanes 2, 3, 7 and 8). Although VP16-LdN175 bound to the palindromic MARE, its binding activity for  $\text{c}\alpha\text{A}$  was dramatically reduced (Figure 3B, lanes 4 and 9). As expected, VP16-LdN199 did not bind to either oligonucleotide probe (Figure 3B, lanes 5 and 10). These results show that the regions between amino acids 166 and 174 are necessary for the binding of L-Maf to the MARE half-site. The DNA-binding domain of Maf proteins is characterized by the Maf EHR, located on the N-terminal side of the basic region (11). An NMR study for murine MafG revealed that the DNA-binding region consists of three individual helices (35). The nine amino acids identified in this study form part of the first helix, H1 (Figure 4A). Our binding and transcription data therefore suggest that H1 is involved in recognition of the 5'-flanking sequences of the MARE half-site.

Highly conserved amino acid sequences in the DNA-binding domain of Maf proteins suggest a similar structure and, as a result, DNA-binding specificity. Therefore, we reasoned that other Maf family members might also recognize the 5'-flanking region of MARE half-site. To examine this possibility, the binding activities of MafB, c-Maf and MafG proteins were tested by EMSA using the MARE half-site probes (Figure 4B). Because MafG does not contain an activation domain, a VP16 activation domain was fused to its coding sequence. Expression plasmids for MafB, c-Maf and VP16-MafG were transcribed and translated *in vitro*. MafB, c-Maf and VP16-MafG bound to the  $\text{c}\alpha\text{A}$  and  $\text{c}\alpha\text{A-1(Pal)}$ , but not to the  $\text{c}\alpha\text{A-1}$  and  $\text{c}\alpha\text{A-3}$  (Figure 4B). To test whether the DNA-binding activities correlated with transcriptional activity, we performed reporter assays using the luciferase reporter gene driven by  $\text{c}\alpha\text{A}$  ( $\text{c}\alpha\text{AX6-}\beta\text{Luc}$ ), its mutants ( $\text{c}\alpha\text{A-1X6-}\beta\text{Luc}$  and  $\text{c}\alpha\text{A-3X6-}\beta\text{Luc}$ ) or palindromic MARE ( $\text{c}\alpha\text{A-1(Pal)}\text{X6-}\beta\text{Luc}$ ) (Figure 4C). Each *maf* expression plasmid was co-transfected into primary retina cultures along with the reporter plasmid. All Maf proteins showed high transcriptional activities through the  $\text{c}\alpha\text{A}$  as well as the palindromic MARE. However, VP16-MafG, MafB and c-Maf failed to transactivate the reporter gene through the  $\text{c}\alpha\text{A-1}$ . As expected, mutations in the core sequences ( $\text{c}\alpha\text{A-3}$ ) abolished the transcriptional activation by all Maf proteins. These results suggest that the decrease in transcriptional activity of each of the Maf proteins on the mutant enhancers was caused by the reduction of their DNA-binding activity. Furthermore, our data raise the possibility that all Maf family members recognize the 5'-flanking region of MARE half-site.

### The 5'-flanking region is important for Maf proteins to function *in vivo*

Our data show that Maf proteins recognize the 5'-flanking region of MARE half-sites, and that this interaction is crucial for their transcriptional activities. However, it is possible that excess Maf proteins might result in artificial effects in our cell culture experiments. Therefore, we generated transgenic frogs to test independently whether the 5'-flanking region functions as a *cis*-element for gene regulation *in vivo*. We constructed a reporter plasmid with GFP as the reporter gene (Figure 5A). We generated transgenic embryos containing each reporter plasmid using the restriction-enzyme-mediated integration (REMI) method (29). An RFP gene driven by the cytomegalovirus (CMV) promoter was co-integrated to detect transgenic embryos. It has been well established that genes co-introduced into sperm nuclei are also co-integrated into the genome of embryos (36). At stage 37/38, RFP and GFP signals were examined under the fluorescent microscope (Figure 5B). Transgenic embryos injected with  $\text{c}\alpha\text{AX6-GFP}$  expressed GFP in lens, rhombomeres, neural crest and pronephros (Figure 5C). *XmafB* has been shown to be expressed in all of these tissues in *Xenopus* (33). *XL-maf* is also expressed in lens. The overlapping expression profiles suggest that we could detect the transcriptional activities of XMafB and XL-Maf *in vivo*. Interestingly, we could not detect GFP expression in transgenic embryos containing the  $\text{c}\alpha\text{A-3X6-GFP}$  that contained mutations in the core sequence of MARE, even though they expressed RFP (Figure 5B). Importantly, we also could not detect GFP signals in transgenic

**A****B**

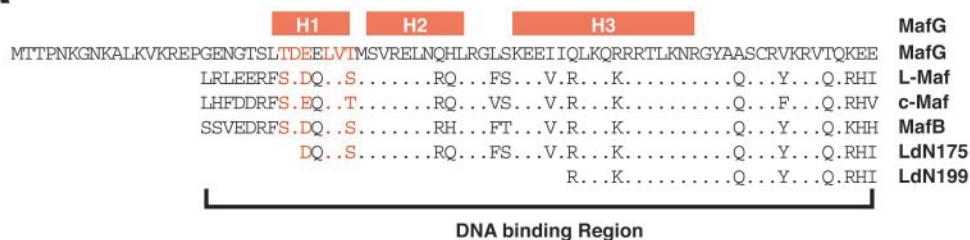
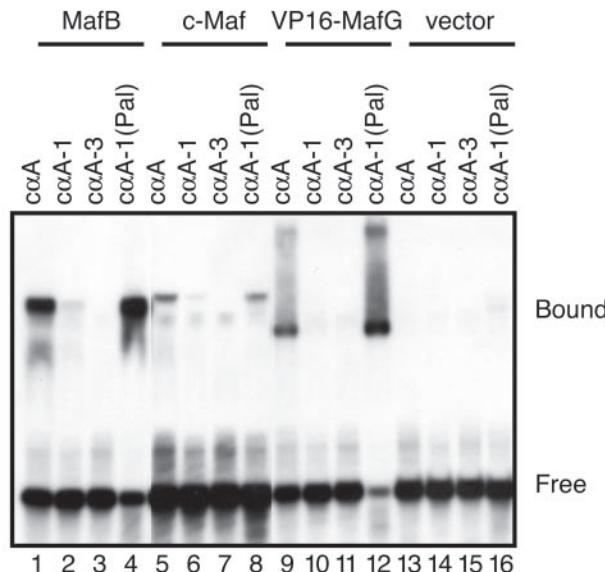
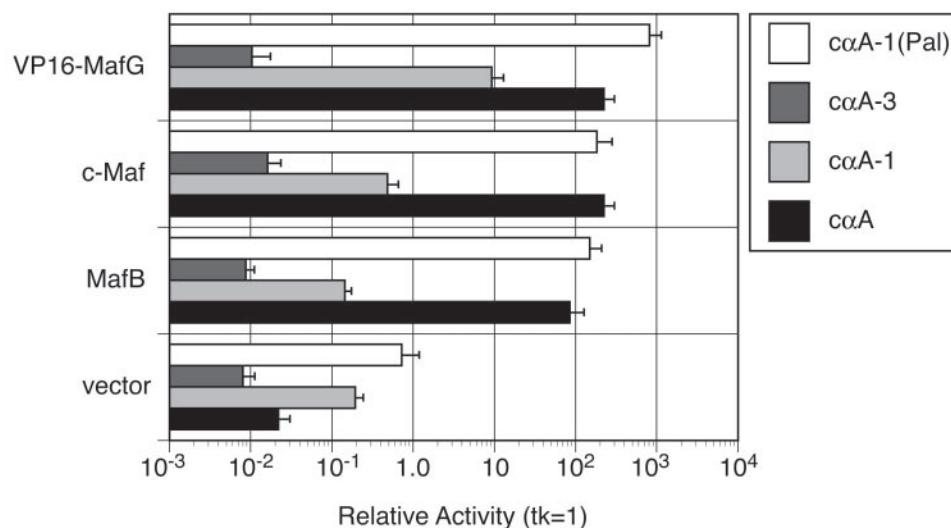
**Figure 3.** The N-terminus of the DNA-binding domain is responsible for the binding of L-Maf to the 5'-flanking sequence. (A) Effects of deletions of the DNA-binding domain of L-Maf on transcriptional activity. Structures of the VP16-L-Maf chimeric proteins are represented schematically in the left panel. Neural retina cultures were transiently co-transfected with each chimeric VP16-L-Maf expression plasmid and the reporter construct. Luciferase activities are shown, relative to that of the luciferase gene driven by the HSV tk promoter (tk-Luc). Average values and error bars represent duplicate assays of at least three individual experiments. (B) DNA-binding activities of the VP16-L-Maf chimeric proteins were analyzed by EMSA. All proteins were prepared *in vitro* using a reticulocyte lysate.

embryos containing the coαA-1X6-GFP that contained mutations in the 5'-flanking region of the MARE half-site, although they expressed CMV-driving RFP (Figure 5B). These data suggest that the 5'-flanking region is an important *cis*-element *in vivo*, and that it contributes to the regulation of specific expression in the lens, neural crest, rhombomeres and pronephros. The coαA-1(Pal)X6-GFP reporter gene containing mutations in the 5'-flanking region of palindromic MARE was expressed in the same region as coαA-X6-GFP, but we detected weak, ubiquitous GFP signals throughout the transgenic embryos (Figure 5B). These results are consistent with the results obtained from reporter assays in cultured cells,

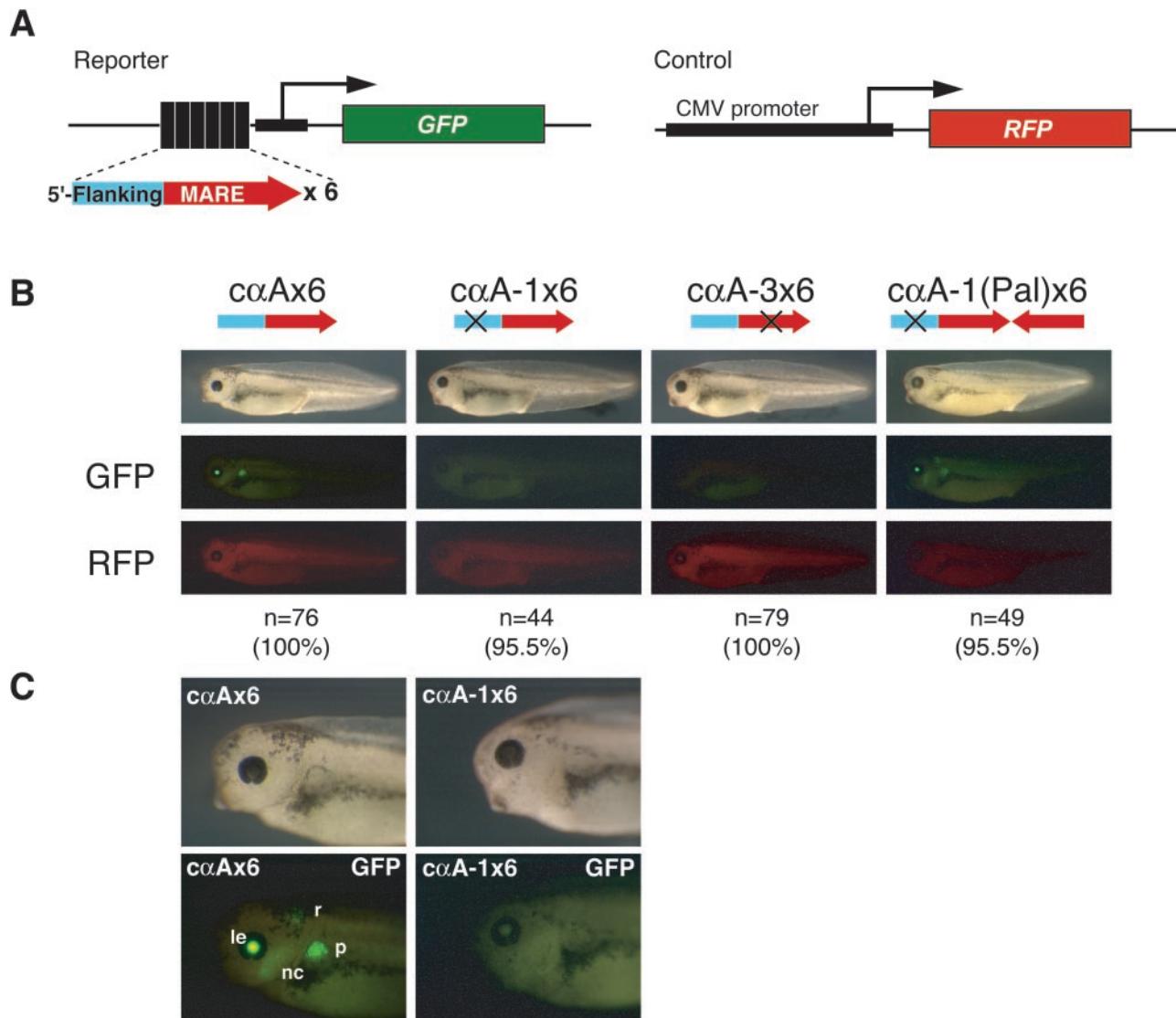
suggesting that the 5'-flanking region is important for Maf proteins to regulate target genes *in vivo*.

#### Maf proteins require AT-rich sequences in the 5'-flanking region to bind to the MARE half-site

We found that Maf proteins recognize the 5'-flanking region in order to bind efficiently to the MARE half-site. To identify specific DNA sequences within the 5'-flanking region, we utilized the SAAB assay and bacterially expressed c-Maf fused to GST. After three rounds of selection, most of the sequences obtained from the SAAB assay contained the

**A****B****C**

**Figure 4.** Maf proteins recognize the 5'-flanking region to bind to the MARE half-site. (A) Sequence alignment of the Maf DNA-binding domains, including the L-Maf deletion mutants, used in this study. Secondary structure elements of MafG are shown on the top for reference (35). The DNA-binding domain of MafG consists of three alpha-helices (H1–H3). High homology in the DNA-binding region among Maf proteins suggests that they share similar structure with MafG. (B) DNA-binding activities of MafB, c-Maf and VP16-MafG. Each Maf protein was prepared *in vitro* using reticulocyte lysates and analyzed by EMSA using coA, mutants or palindromic MARE oligonucleotides as probes. (C) MafB, c-Maf and VP16-MafG show high transcriptional activity through coA depending on the 5'-flanking region. Each Maf expression plasmid was co-transfected into primary retina cultures with each reporter construct. Luciferase activities are represented, relative to that of the luciferase gene driven by the HSV tk promoter (tk-Luc). Average values and error bars represent duplicate assays of at least three individual experiments.



**Figure 5.** The 5'-flanking region is important for Maf proteins to function *in vivo*. Transgenic embryos containing a GFP reporter gene were generated by REMI. (A) The structure of reporter and control genes used. (B) Transgenic embryos were confirmed by analysis of RFP signal driven by the CMV promoter. GFP driven by the c $\alpha$ A was expressed in lens (le), rhombomeres (r), neural crest (nc) and pronephros (p), where *XmafB* and *XL-maf* are expressed. GFP signals driven by c $\alpha$ A-1 and c $\alpha$ A-3 were not detected in transgenic embryos, although they expressed RFP. Number of embryos tested (RFP positive) and ratio of indicated phenotypes to RFP positive embryos are shown by percentage. (C) Magnification of the embryos described above.

MARE half-site (Figure 6A). Analysis of the cloned sequences revealed that the 5'-flanking region of the selected half-sites had a sequence rich in either A or T. As shown in Figure 6B, the positions from -4 to -2 contained 76, 70 and 73% of either A or T. These results suggest that natural MARE half-sites should have AT-rich sequences in their 5'-flanking region *in vivo*.

To test this hypothesis, we aligned the sequences of a well-characterized lens-specific enhancer (21,37–40) or targets of Maf proteins (Figure 6C). *IL-4*, *Hoxa3* and *Hoxb3* are target genes of c-Maf and MafB, respectively (22–24). As shown in Figure 6C, well-conserved half-sites were observed. Importantly, the 5'-flanking regions of the half-site exhibited AT-rich sequences. This fact is consistent with the sequences obtained

from the SAAB assay. Furthermore, we calculated the ratio (%) of AT to GC in the 5'-flanking region of sequences obtained from SAAB or targets of Maf (Figure 6D). The results are similar and show that positions -2 to -4 are rich in AT. Therefore, we conclude that Maf proteins can bind to both the palindromic MARE and half-sites containing AT-rich sequences in their 5'-flanking region.

#### Maf directly interacts with the AT-rich sequence

We have shown that the 5'-AT-rich sequence is important for Mafs to bind to the MARE half-site. However, it is unclear that the AT-rich sequences are important for palindromic MARE binding. We were also interested in how the AT-rich sequence

**A**

atatacAACGGGCTGAATTTTGCTGACCGTCatgtcggtgg  
tataggggccaccaacgatt**GCTGAC**GCTCTTAGATAGTGTGAT  
caacgacatt**TGGTGAGACATTGCTGAC**ACAGCCGgttgatataat  
atcaacGCAGGCCCTAAACT**TGCTGAC**GCaatgtcggtgg  
tttatatacAACGTCCATTAGTT**GCTGAC**ACTAGTCatgtcggtgg  
ttatatacAACTAACTTGGCTTGCTGACACAGCTCaatgtcggtgg  
acgacatt**CGACAATTGCGAGTGCTGACT**TGGCgttgatataatag  
tatacAACAGGTTCTGAAAT**GCTGAT**GCCTCAatgtcggtgg  
ccaccaacgacatt**TGCTGTTGCTGATTGGTATA**ACATgttgatat  
caacttatacAACGCTATT**GCTGAGTCGTTGTTGAA**atgtc  
atcaacAAGTAA**TGCTGATCTTGCTGAA**AGAatgtcggtgg  
ggggccaccaacgacatt**ACCTGCTGATTAAACACCATG**tttgatata  
ttCGGTACAGCATTAAAT**GCTGAG**tgtatataatagtgc  
ggggccaccaacgacatt**TAGTGCTGAGTCGATGGACATGGCA**gtt  
attATATGTGATT**CGTGGAAATGCCGAC**gttgatataatagtgc  
acgacatt**GGAGTGTAGAAAATGCAGACGTACG**gttgatataatag  
agggggccaccaacgacatt**TGTGATGACAGGTCTAAACACGT**ttt  
ttAAGTTGGCATAACGTAATAAT**GACG**tttgatataatagtgc  
tagggggccaccaacgacatt**CTGATGACATGCGATGTA**AAAACgtt  
tttatatacAACGGCAGGATGCT**GACTCGTGA**Atatgtcggtgg  
tatacAACCTCTAACATTT**TACTGACCTAG**Caatgtcggtgg  
aacgacatt**GAAGAAAGAGTAAGCTGACGCGT**Cgttgatataatag  
acgacatt**CGACCAAGGAAGTAGCTGACGCAT**Cgttgatataatag  
ataggggccaccaacgacatt**TAGTGACGCTTCACTTGATAGGT**ctg  
attGTACCAGAGAATCTATTGCG**GCTGAC**gttgatataatagtgc  
tagggggccaccaacgacatt**TGCTGTTGCTGAGCAATGCCCCTG**  
tttatatacACGTAAGCGTCT**GCTGTTGGT**Caatgtcg  
gacatt**GAAAACCTCGCTGAATGCTGATAC**gttgatataatagtgc  
TCCTTCCTTATGGTCGACGTT**GCTA**atgtcggtggccccctata  
gggacttattatatacACG**TGCTGATAC**CGACGGCAATAAGCCTata  
tatacAACCTTGCCGTT**TGCGCACTACAG**Caatgtcggtgg  
attTTTGCGCCGAGATTGCGAAC**Gt**tgatataatagtgc  
tttatatacACCAT**TGCTGCGCCGCA**TGCAatgtcggtgg  
atcaac**TCGGAACGCTCC**CACTG**CGACGGCG**Caatgtcggtgg  
ttGGTATGCGGGTAATT**TGTTG**GAgttgatataatagtgc  
tttatatacACG**TGCTGTTTCTC**CT**GAGTT**CGGGGaatgtcggtgg  
gcactatttatatacACG**TGCTGATTGGT**ATCGACATT**Caatgt**

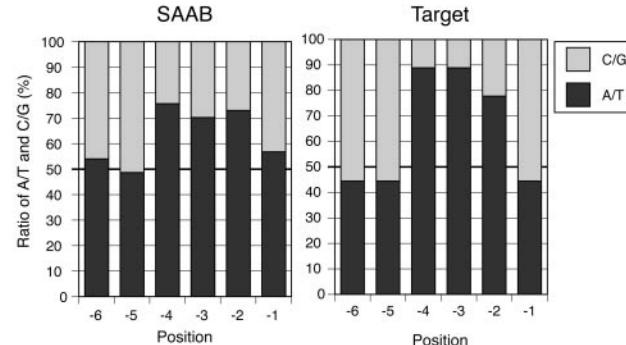
**Consensus** **TGCTGAC****B**

c-Maf (N=37)

	-5	-4	-3	-2	-1	T	G	C	T	G	A	C	1	2	3	4	5	6	7
A	8	15	11	12	6	4	1	3	3	2	32	2	6	4	10	10	9	7	14
C	5	4	5	7	4	2	2	<b>33</b>	2	1	0	<b>25</b>	2	15	5	3	14	5	7
G	14	5	6	3	12	0	<b>34</b>	0	2	<b>33</b>	2	5	16	7	12	14	6	10	9
T	10	13	15	15	32	0	1	<b>30</b>	1	3	7	13	11	10	10	8	15	7	
							<b>32</b>	<b>34</b>	<b>33</b>	<b>30</b>	<b>33</b>	<b>32</b>	<b>25</b>						
A/T	18	28	26	27	21														
(%)	49	76	70	73	57														

**C**

gene	sequence	reference
Mouse $\alpha$ -crystallin	<b>TGTCTATGCTCAGCTCGGGGAGA</b>	Nakamura et al. 1990
Mouse $\gamma$ -F-crystallin	<b>CTTTTGTGCTGTTCTGCACACACA</b>	Liu et al. 1991
Chicken $\alpha$ -crystallin	<b>CATTTCTGCTGACCACGTTGCCTTC</b>	Matsuo and Yasuda 1992
Chicken $\beta$ 1-crystallin	<b>AGGATGTCATGACTGGCCGCCGCA</b>	Roth et al. 1991
Chicken $\delta$ 1-crystallin	<b>TGTTGTTGCTCACCTACATGGACA</b>	Funahashi et al. 1991
Mouse <i>Hoxa-3</i>	<b>CCAAACTGCTGACGCGACTCTACC</b>	Manzanares et al. 1999
Mouse <i>Hoxb-3</i>	<b>CGTAACCTGCTGACTTAGGGGATGA</b>	Manzanares et al. 1997
Mouse <i>Hoxb-3</i>	<b>CAAATTGCGACACCTACATTCTT</b>	Manzanares et al. 1997
Mouse <i>IL-4</i>	<b>TAAAGTTGCTGAAACCAAGGGAAA</b>	Ho et al. 1996
consensus	<b>AAA</b> <b>TTT</b> <b>TGCTGAC</b>	

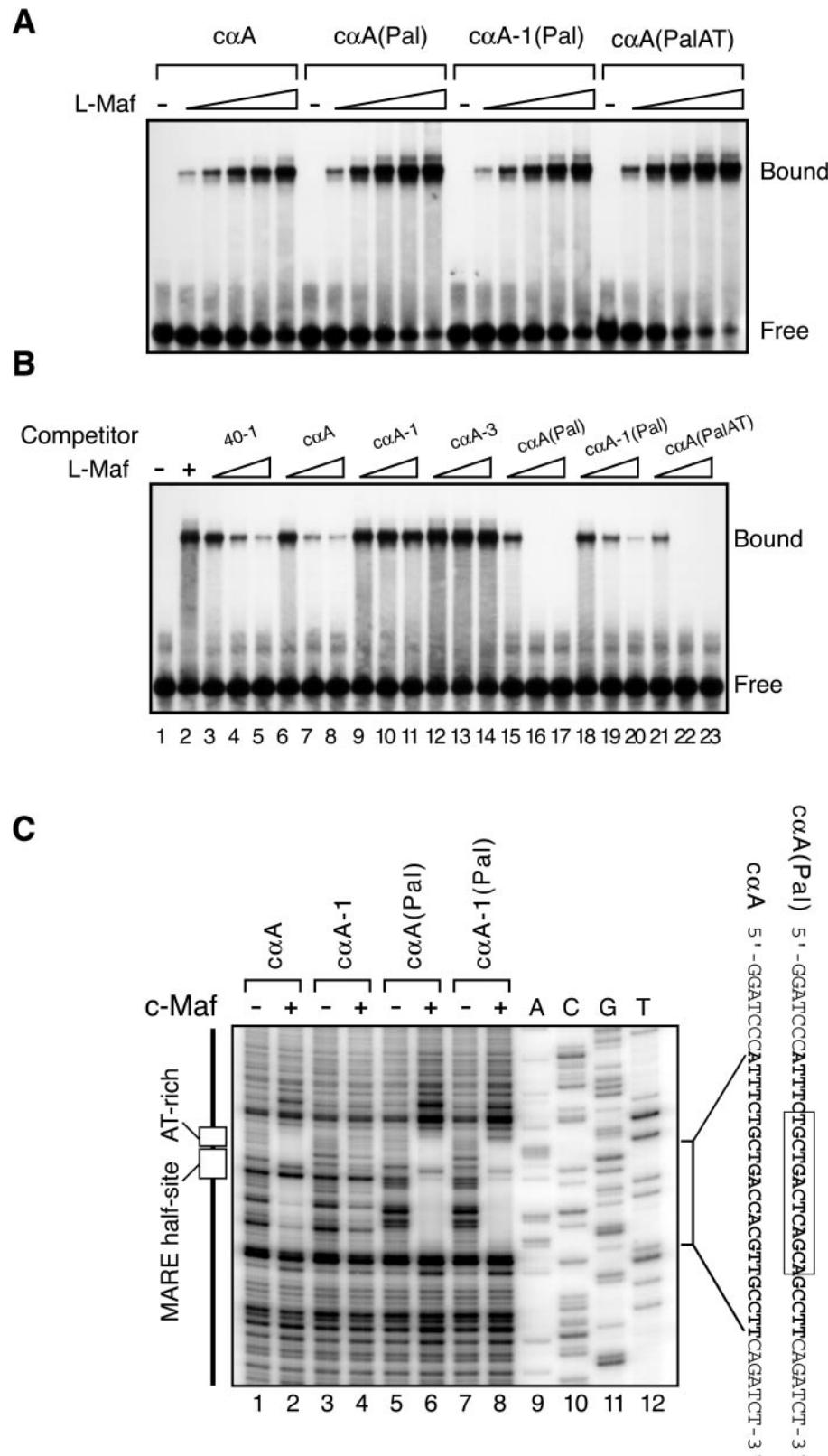
**D**

**Figure 6.** The MARE half-site flanked by AT-rich sequences in the 5' region is a target sequence for Maf proteins. (A) The sequences selected by affinity for c-Maf (SAAB assay). Sequences of the MARE half-site are shown in bold capital letters, while the non-random sequences are shown in lowercase letters. (B) The table shows the number and percentage of frequencies for the 4 nt in the selected sequences shown in (A). (C) Comparison of the coA sequences with the regulatory regions of several crystallin genes, *Hoxa3*, *Hoxb3* and *IL-4*. The conserved MARE half-site is boxed, and consensus sequences are shown in bold capital letters. (D) The ratio of AT to GC in the 5'-flanking regions of MARE selected *in vitro* by SAAB assay and those observed *in vivo*.

contributes to the binding of Maf to the MAREs. First, we compared the binding activity of L-Maf among palindromic MAREs that have AT-rich sequences on one side [coA(Pal)], both sides [coA(PalAT)] or neither side [coA-1(Pal)] (Figure 7A) using various amounts of L-Maf protein; the estimated K values for coA(Pal), coA-1(Pal) and coA(PalAT) were  $\sim 2.5 \times 10^1$ ,  $1.0 \times 10^2$  and  $1.3 \times 10^1$  nM, respectively. In order to determine the sequences that have a strong affinity for sequence-specific binding, L-Maf was incubated with labeled 40-1 probe (0.75 nM) and the indicated unlabeled competitors (100 $\times$ , 500 $\times$  and 1000 $\times$ , Figure 7B). 40-1 and coA showed the same response, consistent with the fact that they have the same sequence differing only in oligonucleotide length (lanes 3–8). The bound complex was not competed with unlabeled coA-1 or coA-3 (lanes 9–14). coA-1 (Pal) showed the same activity as coA, which was

weaker than coA (Pal) (lane 15–20). Strikingly, coA(PalAT) A(PalAT) showed the strongest competition activity (lanes 21–23). Therefore, AT-rich sequences appear to contribute to Mafs binding to palindromic MAREs.

To analyze directly how Maf interacts with AT-rich sequences, we performed a DNase I footprinting analysis (Figure 7C). We examined protection from DNase I on four different sequences, coA, coA-1, coA(Pal) and coA-1(Pal) by recombinant c-Maf protein. AT-rich and MARE half-site regions were well protected, while the 3' region was weakly protected in coA (lane 2). We detected a strong footprint on the palindromic MARE in coA-1(Pal) (lane 8). In coA(Pal), a more 5' region, including the AT-rich sequence, was protected. These observations are consistent with our other data, suggesting that the affinity of Maf binding varies depending on the combination of AT-rich sequences and half-site MAREs.



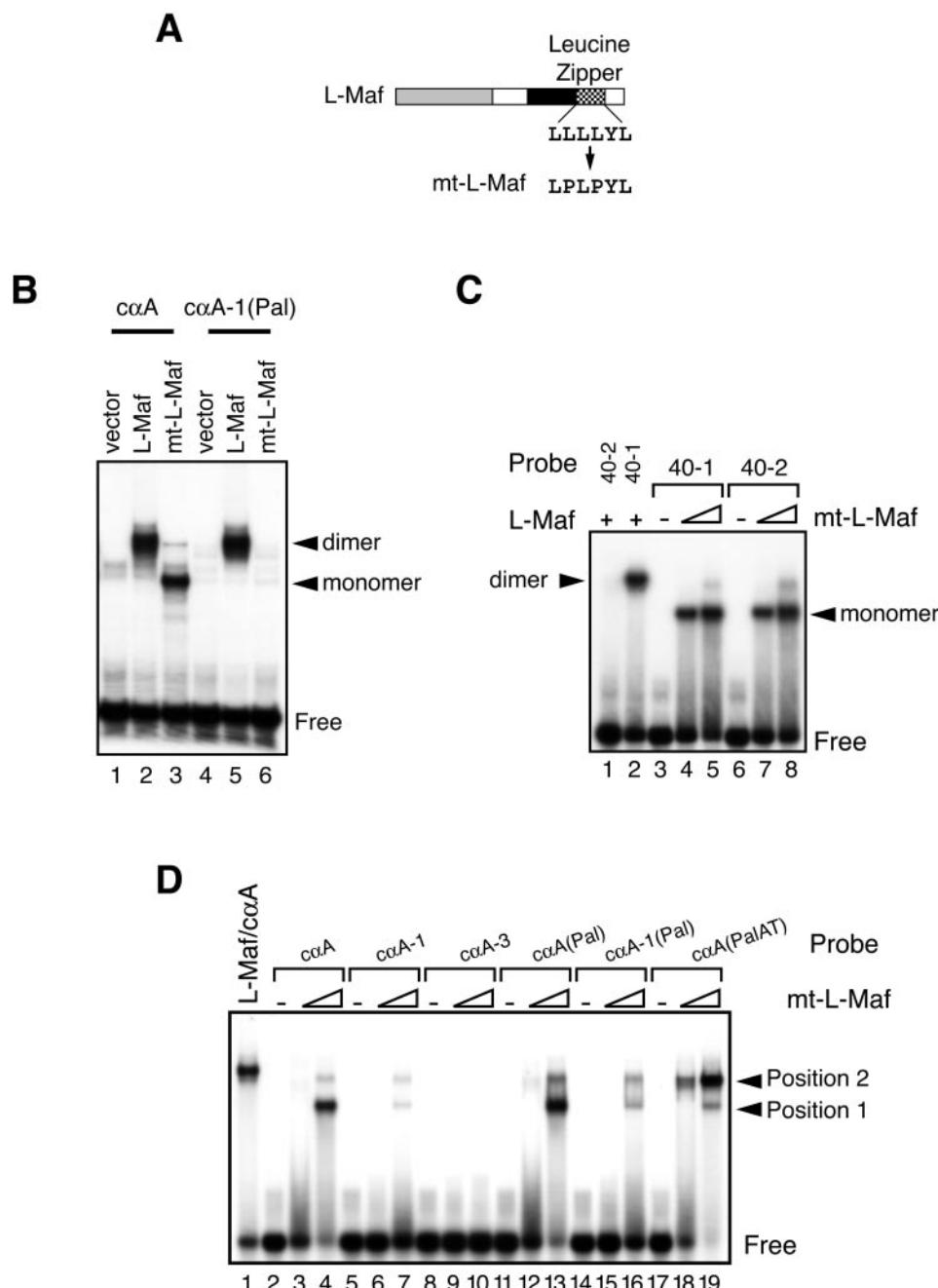
**Figure 7.** The 5'-AT-rich sequence increases the binding activity for palindromic MARE. (A) Increasing amounts of L-Maf (0, 5,  $1.3 \times 10^1$ ,  $2.5 \times 10^1$ ,  $5 \times 10^1$  and  $2.5 \times 10^2$  nM) were subjected to EMSA with the c $\alpha$ A, c $\alpha$ A(Pal), c $\alpha$ A-1(Pal) and c $\alpha$ A(PalAT) oligonucleotide probes. (B) Comparison of binding activities by competition analysis. Recombinant L-Maf was incubated with labeled 40-1 probe (0.75 nM) and indicated competitors (100 $\times$ , 500 $\times$  and 1000 $\times$ ). Lane 1 shows only 40-1 probe and lane 2 shows the complex of L-Maf and 40-1 probe. (C) Recombinant c-Maf and four different probes [c $\alpha$ A, c $\alpha$ A-1, c $\alpha$ A(Pal) and c $\alpha$ A-1(Pal)] were subjected to DNase I footprinting assay. Results of the assay are shown (lanes 1–8), along side sequencing ladders for reference (lanes 9–12). The sequence protected by c-Maf is shown to the right in bold. The palindromic MARE is boxed. The 5'-AT rich region and MARE half-site are shown to the left in rectangles.

The results indicate that Maf interacts with AT-rich sequences directly and that AT-rich half-site MAREs constitute a subset of targets of Maf proteins.

#### AT-rich half-site MARE is recognized by one Maf protein molecule

bZIP transcription factors are known to dimerize and bind to their palindromic target sequences. In this study, Maf proteins

were shown, not only to bind to the palindrome MARE, but also to the MARE half-site. To investigate how L-Maf binds to target sequences that contain only a half-site of MARE, the binding activities of L-Maf and its mutant (mt-L-Maf) were analyzed by EMSA. mt-L-Maf is not able to dimerize because of a substitution of the 2nd and 4th leucine residues to proline in the leucine zipper region (Figure 8A) (30). *In vitro* translated L-Maf bound efficiently to the c $\alpha$ A and c $\alpha$ A-1(Pal) (Figure 8B, lanes 2 and 5). Interestingly, *in vitro* translated



**Figure 8.** Maf binds to the 5'-AT-rich half-site of MARE as a dimer. (A) Schematic representation of mt-L-Maf carrying mutation in the leucine zipper region. (B) *In vitro* translated L-Maf and mt-L-Maf proteins were subjected to the EMSA with c $\alpha$ A and c $\alpha$ A-1(Pal) oligonucleotide probes. (C) Recombinant mt-L-Maf recognizes only the half-site MARE. L-Maf (2.5  $\times$  10<sup>2</sup> nM) and mt-L-Maf (7.5  $\times$  10<sup>2</sup> and 1.9  $\times$  10<sup>3</sup> nM) were incubated with 40-1 and 40-2. Lanes 1 and 2 show complexes of recombinant L-Maf (2.5  $\times$  10<sup>2</sup> nM) and 40-2 or 40-1, respectively. (D) The binding activities of recombinant mt-L-Maf. mt-L-Maf (7.5  $\times$  10<sup>2</sup> and 1.9  $\times$  10<sup>3</sup> nM) was incubated with the indicated probes. Lane 1 shows the complex of L-Maf (2.5  $\times$  10<sup>2</sup> nM) and c $\alpha$ A.

mt-L-Maf bound c $\alpha$ A but not c $\alpha$ A-1 (Figure 8B, lanes 3 and 6). The shifted complex migrated more quickly than that of the wild-type L-Maf complex. L-Maf appeared to bind to the c $\alpha$ A-1(Pal) as a homodimer, because mt-L-Maf failed to bind to this probe (Figure 8B, lane 6). To exclude the possibility that the lysate contains Maf-like binding activity, we purified recombinant mt-L-Maf protein from *Escherichia coli*, and examined the binding activities. Although the binding activity of L-Maf for 40-2 is lower than that for 40-1, mt-L-Maf did not discriminate between the two probes (Figure 2B and C and Figure 8C). This result suggests that mt-L-Maf recognizes only the half-site of MARE. In order to extend this analysis, we used the EMSA with mt-L-Maf protein and various targets [c $\alpha$ A, c $\alpha$ A-1, c $\alpha$ A-3, c $\alpha$ A(pal), c $\alpha$ A-1(pal) and c $\alpha$ A(PalAT)]. As expected, recombinant mt-L-Maf bound to c $\alpha$ A but not to c $\alpha$ A-1 and c $\alpha$ A-3 (Figure 8D, lanes 2–10). Because mt-L-Maf cannot dimerize, it failed to bind efficiently to c $\alpha$ A-1(pal) (lanes 14–16). In contrast, mt-L-Maf bound efficiently to both c $\alpha$ A(pal) and c $\alpha$ A(PalAT), although there were two different complexes. The predominant complex formed on c $\alpha$ A(pal) ran with the same mobility as that formed by mt-L-Maf and c $\alpha$ A, indicating monomer binding (lanes 11–13, Position 1). The major complex formed on c $\alpha$ A(PalAT) showed more retardation than that of Position 1 (lanes 17–19, Position 2). This band represented a similar mobility with that of wild-type L-Maf and c $\alpha$ A (lane 1), suggesting that the complex contains homodimers. Because mt-L-Maf cannot dimerize, each mt-Maf protein appeared to bind to each AT-rich MARE half-site of the c $\alpha$ A(PalAT) probe without dimerization. These results suggest that Maf proteins recognize the AT-rich half-site MARE using a single subunit of the Maf-dimer.

## DISCUSSION

We have shown that the 5'-flanking regions of MARE half-sites are important features in determining how Maf proteins recognize and bind their targets. The SAAB assay and EMSA revealed that AT-rich sequences in the 5'-flanking region are crucial for Maf protein binding as is the half-site itself. Maf proteins are unique among bZIP transcription factors as they recognize a relatively long palindromic DNA sequence. This specific feature is derived from the structure of the DNA-binding domain containing the Maf EHR (35,41). The binding activities uncovered here are also likely a result of this unique structure of the Maf proteins. Our findings are important for helping define novel target genes of the Maf family. They also provide detailed mechanistic information regarding Maf protein site selection.

### Maf proteins require AT-rich sequences in the 5'-flanking region to bind to the MARE half-site

We found that AT-rich sequences in the 5'-flanking region are important for Maf proteins to bind to the MARE half-site. Consistent with this, MARE half-sites flanked by AT-rich sequences were found in many lens-specific promoters whose sequences have been shown to affect lens-specific transcriptional activation. In our previous study, a mutation in the 5'-flanking region of the  $\alpha$ CE2 core sequence was shown to reduce its activity as a lens-specific enhancer

(21). Furthermore, the *IL-4* gene, a target of c-Maf, contains an AT-rich region, and mutation of this region (AAAGTGCTGAA to cccGTTGCTGAA) leads to a decrease in the transcriptional activity of c-Maf *in vivo* (24). Kr-1, Kr-2 and KrA are rhombomere-specific enhancers and known targets of MafB (22,23). We found that these sequences also resemble 5'-AT-rich MARE half-sites (Figure 6C). Furthermore, using c-Maf and the SAAB assay, a preference for 5'-AT-rich half-sites was uncovered. Our sequence analysis indicated that A or T is present 70% of the three at positions –2 to –4 (Figure 6B and D). These results suggest that positions –2 to –4 are recognized by Maf and show that Maf requires 5'-AT-rich sequence to bind efficiently to the MARE half-site.

Previously, the palindromic MARE sequence was shown to be a recognition site of c-Maf using the SAAB assay (30). However, we obtained predominantly MARE half-sites rather than palindromic MAREs in our analysis. Our experiments utilized GST fusions of Maf protein, while the previous report used maltose binding protein fusions. While it is possible that the fusion partner has an effect on dimerization and that this is responsible for the observed discrepancy, we believe that our results reflect the true sequence specificity of c-Maf for the following reasons. The dimerization mutant, mt-L-Maf, bound the 5'-AT-rich half-site (c $\alpha$ A) but not to the half-site of MARE lacking AT-rich sequences in its 5'-flanking region [c $\alpha$ A-1(Pal), Figure 8]. Furthermore, mt-L-Maf showed the same sequence preference for binding as L-Maf. Therefore, Maf proteins seem to require the AT-rich sequences to bind to the half-site of MARE, regardless of whether they are homodimers or monomers.

### AT-rich MARE half-sites can function *in vivo*

We have shown that AT-rich sequences are functional *in vivo* using transgenic *Xenopus*. The reporter genes were introduced into all cells of transgenic embryos; therefore, it should have been able to respond to the activity of all large Maf proteins. We found that the reporter genes, both c $\alpha$ A and c $\alpha$ A-1(Pal), are activated specifically in lens, pronephros and rhombomere. These GFP positive regions overlap with known *maf* gene expression regions obtained from *in situ* hybridizations (33). Furthermore, we found GFP signal in neural crest cells. It is known that *mafB* is expressed in some neural crest cells (42,43) and that MafB activities are detected in the migrating neural crest cells derived from rhombomere 5 (44). Our findings are consistent with these previous reports and demonstrate that endogenous Maf proteins can activate transcription through the MARE half-sites. Importantly, our results also indicate that the 5'-AT-rich regions are essential for this activity.

### Residues in helix H1 of the Maf DNA-binding domain are necessary for the recognition of the AT-rich sequences

The DNA-binding region of MafG has been shown to consist of three helices and the amino acid sequence of this region is highly conserved among L-Maf, MafB, c-Maf and MafG (35) (Figure 4A). Our results show that they also have a comparable dependency on the 5'-AT-rich sequences to bind the MARE half-site. This suggests that this highly conserved region of Maf proteins is important to the recognition of AT-rich sequences. VP16-LdN175, lacking part of helix H1, failed

to bind to c $\alpha$ A efficiently, while it bound to the palindromic MARE. Although VP16-LdN175 did not recognize the AT-rich sequence, it may bind to the palindromic MARE as a dimer more stably than to the half-site. It has been shown that Helix 2 recognizes TGC of TGCTGAC (35). Our results suggest that the H1 region recognizes AT-rich sequences.

#### Maf proteins bind to the 5'-AT-rich MARE half-site as a dimer

mt-L-Maf, which lacked the ability to dimerize, bound to the 5'-AT-rich half-site efficiently as a monomer depending on 5'-AT rich sequence. L-Maf appears to bind to the half-site as a dimer, even though the 5'-AT-rich MARE half-site has only one recognition site.

These results raise the question of which target sequence is the preferred one for Maf protein binding. We compared the affinity of several types of Maf response element and concluded that the order of preference is the following: half-site < AT-half-site < Palindrome < AT-Palindrome < AT-Palindrome-AT. Our results strongly suggest that while Maf proteins indeed bind to palindromic MARE sequences with high affinity, they also bind to targets lacking such structures. As demonstrated here, Maf proteins bind to targets

containing only MARE half-site. Binding to these targets is dependent upon AT-rich sequences immediately upstream of the half-site. Interestingly, while the AT-rich sequences and MARE half-site sequences are important for Maf recognition, 3' sequence of the half-site is also required for Maf binding. As we have shown, Maf proteins bind to their targets as homodimers. Consistent with this result, we observe DNase I protection of 3' sequence of the half-site, indicating that Maf proteins recognize sequences other than the MARE consensus. We reasoned that if Maf proteins recognize targets containing 5'-AT-rich MAREs as dimers and that one of the dimers interacts with 3' sequence of the half-site, then there may be some preference as to which bases are found in these 3' positions. Indeed, positions +2, +4 and +5 score highly in C, G and C, respectively (Figure 6B). This observation may reflect the sequence requirement placed on the binding site by the Maf proteins themselves.

Our data are consistent with a model by which Maf proteins bind as homo-dimers to MAREs containing either palindromes or 5'-AT-rich MARE half-sites (Figure 9). The DNA-binding domain of one subunit of the dimer interacts with the 5' half-site, while the remaining binding domain binds specifically to sequences immediately 3' to the 5' half-site. We propose that Maf targets require only a single MARE half-site and that the 3' sequences can be of varying composition. The palindromic MARE represents one example of the sequence flexibility of Maf target sequences. Our data raise the possibility that Maf proteins are capable of binding to a much larger repertoire of sequences than previously thought. It follows that the number of Maf target genes is also likely to be much greater than previously appreciated.

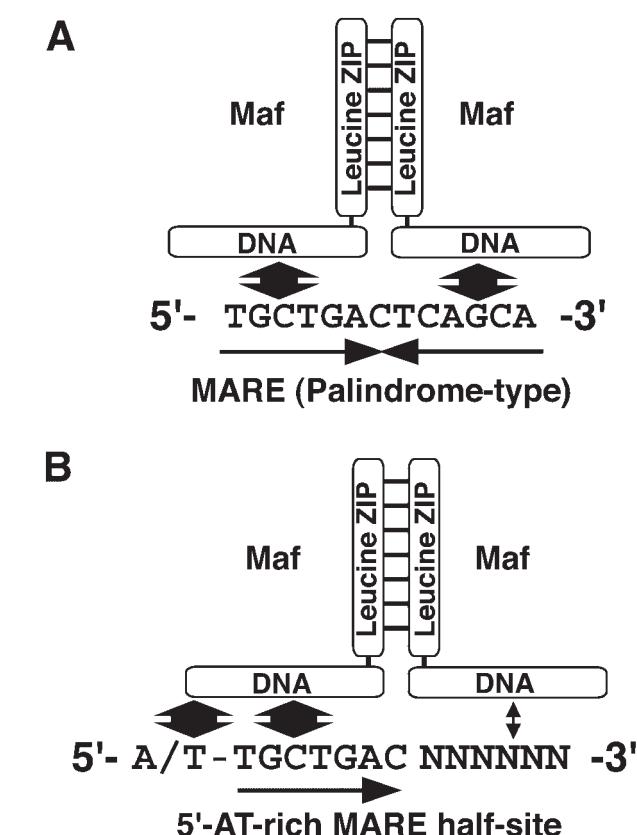
#### ACKNOWLEDGEMENTS

The authors thank Dr Masayuki Yamamoto for MafG DNA and Dr Kazuhiko Umesono for VP16 and GFP expression plasmids. The authors thank Drs Reiko Landry and Christopher Murawsky for critical reading of the manuscript. The authors also thank Drs Kazuhiko Umesono, Hajime Ogino and colleagues in Yasuda's laboratory for technical advice and helpful comments. The authors thank Drs Enrique Amaya and Katherine A. Jones for their help. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan and The Mitsubishi Foundation and Foundation for Nara Institute of Science and Technology. Funding to pay the Open Access publication charges for this article was provided by Nara Institute of Science and Technology (NAIST).

*Conflict of interest statement.* None declared.

#### REFERENCES

- Vinson,C., Myakishev,M., Acharya,A., Mir,A.A., Moll,J.R. and Bonovich,M. (2002) Classification of human b-Zip proteins based on dimerization properties. *Mol. Cell. Biol.*, **22**, 6321–6335.
- Lee,W., Mitchell,P. and Tjian,R. (1987) Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell*, **49**, 741–752.
- Montminy,M.R., Sevarino,K.A., Wagner,J.A., Mandel,G. and Goodman,R.H. (1986) Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. *Proc. Natl Acad. Sci. USA*, **83**, 6682–6686.



**Figure 9.** Proposed model for Maf protein binding to the 5'-AT rich MARE half-site. Maf proteins have two types of target sequences, a palindrome-type (**A**) and half-site type (**B**). (**A**) Maf binds to the MARE palindrome as homodimer and there are specific interactions between each DNA-binding region and a half-site of MARE (TGCTGAC). (**B**) Our results show that Maf proteins can also bind as homodimers on MARE half-sites that contain an AT rich 5' region. This region is critical for Maf protein binding and subsequent transcriptional activation.

4. Montminy,M.R. and Bilezikjian,L.M. (1987) Binding of a nuclear protein to the cyclic-AMP response element of the *somatostatin* gene. *Nature*, **328**, 175–178.
5. Angel,P., Imagawa,M., Chiu,R., Stein,B., Imbra,R.J., Rahmsdorf,H.J., Jonat,C., Herrlich,P. and Karin,M. (1987) Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. *Cell*, **49**, 729–739.
6. Hai,T. and Curran,T. (1991) Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. *Proc. Natl Acad. Sci. USA*, **88**, 3720–3724.
7. Hoeffler,J.P., Meyer,T.E., Yun,Y., Jameson,J.L. and Habener,J.F. (1988) Cyclic AMP-responsive DNA-binding protein: structure based on a cloned placental cDNA. *Science*, **242**, 1430–1433.
8. Olliphant,A.R., Brandl,C.J. and Struhl,K. (1989) Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mol. Cell. Biol.*, **9**, 2944–2949.
9. Nishizawa,M., Kataoka,K., Goto,N., Fujiwara,K.T. and Kawai,S. (1989) *v-maf*, a viral oncogene that encodes a ‘leucine zipper’ motif. *Proc. Natl Acad. Sci. USA*, **86**, 7711–7715.
10. Kataoka,K., Nishizawa,M. and Kawai,S. (1993) Structure–function analysis of the *maf* oncogene product, a member of the b-Zip protein family. *J. Virol.*, **67**, 2133–2141.
11. Kerppola,T.K. and Curran,T. (1994) Maf and Nrl can bind to AP-1 sites and form heterodimers with Fos and Jun. *Oncogene*, **9**, 675–684.
12. Kataoka,K., Fujiwara,K.T., Noda,M. and Nishizawa,M. (1994) MafB, a new Maf family transcription activator that can associate with Maf and Fos but not with Jun. *Mol. Cell. Biol.*, **14**, 7581–7591.
13. Swaroop,A., Xu,J.Z., Pawar,H., Jackson,A., Skolnick,C. and Agarwal,N. (1992) A conserved retina-specific gene encodes a basic motif/leucine zipper domain. *Proc. Natl Acad. Sci. USA*, **89**, 266–270.
14. Ogino,H. and Yasuda,K. (1998) Induction of lens differentiation by activation of a b-Zip transcription factor, L-Maf. *Science*, **280**, 115–118.
15. Igarashi,K., Kataoka,K., Itoh,K., Hayashi,N., Nishizawa,M. and Yamamoto,M. (1994) Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature*, **367**, 568–572.
16. Fujiwara,K.T., Kataoka,K. and Nishizawa,M. (1993) Two new members of the *maf* oncogene family, *mafK* and *mafF*, encode nuclear b-Zip proteins lacking putative trans-activator domain. *Oncogene*, **8**, 2371–2380.
17. Motohashi,H., O'Connor,T., Katsuoka,F., Engel,J.D. and Yamamoto,M. (2002) Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. *Gene*, **294**, 1–12.
18. Blank,V. and Andrews,N.C. (1997) The Maf transcription factors: regulators of differentiation. *Trends Biochem. Sci.*, **22**, 437–441.
19. Curran,T. and Franzia,B.R.,Jr (1988) Fos and Jun: the AP-1 connection. *Cell*, **55**, 395–397.
20. Kataoka,K., Shioda,S., Yoshitomo-Nakagawa,K., Handa,H. and Nishizawa,M. (2001) Maf and Jun nuclear oncoproteins share downstream target genes for inducing cell transformation. *J. Biol. Chem.*, **276**, 36849–36856.
21. Matsuo,I. and Yasuda,K. (1992) The cooperative interaction between two motifs of an enhancer element of the chicken  $\alpha$ A-crystallin gene,  $\alpha$ CE1 and  $\alpha$ CE2, confers lens-specific expression. *Nucleic Acids Res.*, **20**, 3701–3712.
22. Manzanares,M., Cordes,S., Kwan,C.T., Sham,M.H., Barsh,G.S. and Krumlauf,R. (1997) Segmental regulation of *Hoxb-3* by kreasler. *Nature*, **387**, 191–195.
23. Manzanares,M., Cordes,S., Ariza-McNaughton,L., Sadl,V., Maruthainar,K., Barsh,G. and Krumlauf,R. (1999) Conserved and distinct roles of kreasler in regulation of the paralogous *Hoxa3* and *Hoxb3* genes. *Development*, **126**, 759–769.
24. Ho,I.C., Hodge,M.R., Rooney,J.W. and Glimcher,L.H. (1996) The proto-oncogene *c-maf* is responsible for tissue-specific expression of interleukin-4. *Cell*, **85**, 973–983.
25. Perlmann,T., Rangarajan,P.N., Umesono,K. and Evans,R.M. (1993) Determinants for selective RAR and TR recognition of direct repeat HREs. *Genes Dev.*, **7**, 1411–1422.
26. Yoshida,T. and Yasuda,K. (2002) Characterization of the chicken L-Maf, MafB and c-Maf in crystallin gene regulation and lens differentiation. *Genes Cells*, **7**, 693–706.
27. Ogawa,H., Inouye,S., Tsuji,F.I., Yasuda,K. and Umesono,K. (1995) Localization, trafficking, and temperature-dependence of the *Aequorea* green fluorescent protein in cultured vertebrate cells. *Proc. Natl Acad. Sci. USA*, **92**, 11899–11903.
28. Matsuo,I., Kitamura,M., Okazaki,K. and Yasuda,K. (1991) Binding of a factor to an enhancer element responsible for the tissue-specific expression of the chicken  $\alpha$ A-crystallin gene. *Development*, **113**, 539–550.
29. Kroll,K.L. and Amaya,E. (1996) Transgenic *Xenopus* embryos from sperm nuclear transplants reveal FGF signaling requirements during gastrulation. *Development*, **122**, 3173–3183.
30. Kataoka,K., Noda,M. and Nishizawa,M. (1994) Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun. *Mol. Cell. Biol.*, **14**, 700–712.
31. Sakai,M., Imaki,J., Yoshida,K., Ogata,A., Matsushima-Hibaya,Y., Kuboki,Y., Nishizawa,M. and Nishi,S. (1997) Rat *maf* related genes: specific expression in chondrocytes, lens and spinal cord. *Oncogene*, **14**, 745–750.
32. Yoshida,K., Imaki,J., Koyama,Y., Harada,T., Shinmei,Y., Oishi,C., Matsushima-Hibya,Y., Matsuda,A., Nishi,S., Matsuda,H. and Sakai,M. (1997) Differential expression of *maf-1* and *maf-2* genes in the developing rat lens. *Invest. Ophthalmol. Vis. Sci.*, **38**, 2679–2683.
33. Ishibashi,S. and Yasuda,K. (2001) Distinct roles of *maf* genes during *Xenopus* lens development. *Mech. Dev.*, **101**, 155–166.
34. Kawauchi,S., Takahashi,S., Nakajima,O., Ogino,H., Morita,M., Nishizawa,M., Yasuda,K. and Yamamoto,M. (1999) Regulation of lens fiber cell differentiation by transcription factor c-Maf. *J. Biol. Chem.*, **274**, 19254–19260.
35. Kusunoki,H., Motohashi,H., Katsuoka,F., Morohashi,A., Yamamoto,M. and Tanaka,T. (2002) Solution structure of the DNA-binding domain of MafG. *Nature Struct. Biol.*, **9**, 252–256.
36. Hartley,K.O., Hardcastle,Z., Friday,R.V., Amaya,E. and Papalopulu,N. (2001) Transgenic *Xenopus* embryos reveal that anterior neural development requires continued suppression of BMP signaling after gastrulation. *Dev. Biol.*, **238**, 168–184.
37. Nakamura,T., Donovan,D.M., Hamada,K., Sax,C.M., Norman,B., Flanagan,J.R., Ozato,K., Westphal,H. and Piatigorsky,J. (1990) Regulation of the mouse  $\alpha$ A-crystallin gene: isolation of a cDNA encoding a protein that binds to a *cis* sequence motif shared with the major histocompatibility complex class I gene and other genes. *Mol. Cell. Biol.*, **10**, 3700–3708.
38. Funahashi,J., Kamachi,Y., Goto,K. and Kondoh,H. (1991) Identification of nuclear factor delta EF1 and its binding site essential for lens-specific activity of the delta 1-crystallin enhancer. *Nucleic Acids Res.*, **19**, 3543–3547.
39. Liu,Q.R., Tini,M., Tsui,L.C. and Breitman,M.L. (1991) Interaction of a lens cell transcription factor with the proximal domain of the mouse gamma F-crystallin promoter. *Mol. Cell. Biol.*, **11**, 1531–1537.
40. Roth,H.J., Das,G.C. and Piatigorsky,J. (1991) Chicken beta B1-crystallin gene expression: presence of conserved functional polyomavirus enhancer-like and octamer binding-like promoter elements found in non-lens genes. *Mol. Cell. Biol.*, **11**, 1488–1499.
41. Dlakic,M., Grinberg,A.V., Leonard,D.A. and Kerppola,T.K. (2001) DNA sequence-dependent folding determines the divergence in binding specificities between Maf and other b-ZIP proteins. *EMBO J.*, **20**, 828–840.
42. Moens,C.B., Cordes,S.P., Giorgianni,M.W., Barsh,G.S. and Kimmel,C.B. (1998) Equivalence in the genetic control of hindbrain segmentation in fish and mouse. *Development*, **125**, 381–391.
43. Cordes,S.P. and Barsh,G.S. (1994) The mouse segmentation gene *kr* encodes a novel basic domain-leucine zipper transcription factor. *Cell*, **79**, 1025–1034.
44. Theil,T., Ariza-McNaughton,L., Manzanares,M., Brodie,J., Krumlauf,R. and Wilkinson,D.G. (2002) Requirement for downregulation of kreasler during late patterning of the hindbrain. *Development*, **129**, 1477–1485.