Structural Basis of Alternative DNA Recognition by Maf Transcription Factors[▽]

Hirofumi Kurokawa,[1,2]* Hozumi Motohashi,[2] Shinji Sueno,[3] Momoko Kimura,[2] Hiroaki Takagawa,[3] Yousuke Kanno,[3] Masayuki Yamamoto,[2]† and Toshiyuki Tanaka[3]†*

Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Sendai 980-8577,[1] Tohoku University Graduate School of Medicine, Sendai 980-8575,[2] and Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba 305-8572,[3] Japan

Maf transcription factors constitute a family of the basic region-leucine zipper (bZip) factors and recognize unusually long DNA motifs (13 or 14 bp), termed the Maf recognition element (MARE). The MARE harbors extended GC sequences on each side of its core motif, which is similar to TRE or CRE (7 or 8 bp) recognized by the AP1 and CREB/ATF families, respectively. To ascertain the structural basis governing the acquirement of such unique DNA recognition, we determined the crystal structure of the MafG-DNA complex. Each MafG monomer consists of three helices in which the carboxyl-terminal long helix organizes one DNA-contacting element and one coiled-coil dimer formation element. To our surprise, two well-conserved residues, Arg57 and Asn61 in the basic region, play critical roles in Maf-specific DNA recognition. These two residues show unique side-chain orientations and interact directly with the extended GC bases. Maf-specific residues in the amino-terminal and basic regions appear to indirectly stabilize MARE recognition through DNA backbone phosphate interactions. This study revealed an alternative DNA recognition mechanism of the bZip factors that bestows specific target gene profiles upon Maf homodimers or Maf-containing heterodimers.

Since the discovery of the avian v-Maf oncogene (31), a number of v-Maf-related cellular proteins have been identified. The Maf family of transcription factors has been found to harbor a unique and highly conserved basic region-leucine zipper (bZip) structure (27). The dimeric Maf factors recognize a palindromic sequence referred to as the Maf recognition element (MARE), of which there are two types: a 13-bp T-MARE (TGCTGA^G/$_C$TCAGCA) and a 14-bp C-MARE (TGCTGA^{GC}/$_{CG}$TCAGCA) (16, 17). A 7-bp TPA (12-o-tetradeca-noylphorbol 13-acetate)-responsive element (TRE) (TGA^G/$_C$TCA) and an 8-bp cyclic AMP-responsive element (CRE) (TGA^{GC}/$_{CG}$TCA) constitute the core sequence of the T-MARE and the C-MARE, respectively. It has been shown that the TRE is recognized by the transcription factor AP1 (Jun-Fos heterodimer), while the CRE is engaged by CRE binding protein (CREB/ATF). These cores of the MARE are flanked on each side by extended GC elements (underlined), presenting the MARE as a 5′-TGC-(TRE or CRE)-GCA-3′ structure. This long recognition sequence is unique to the Maf protein family, and it is this DNA binding specificity that contributes to the important functions of the Maf proteins (28). Indeed, target genes activated by Maf homodimers or Maf-containing heterodimers are significantly different from those activated by AP1 or CREB proteins.

The Maf proteins have been subdivided into two groups: the large Maf proteins and the small Maf proteins. Large Maf proteins include MafA, MafB, c-Maf, and Nrl, which possess a transactivation domain and activate transcription by forming a homodimer. On the contrary, small Maf proteins lack a canonical transactivation domain and include MafG, MafK, MafF, and MafT (42). While small Maf proteins form homodimers that repress transcription (26, 29), they also form heterodimers with either the cap'n'collar (CNC) protein family (NF-E2 p45, Nrf1, Nrf2, and Nrf3) or the Bach protein family (Bach1 and Bach2) (12). The heterodimers activate or repress transcription depending on the functional domains of the CNC and Bach proteins.

The critical contributions of Maf proteins in vivo have been elucidated by gene disruption experiments in mouse (27). Target genes of Maf proteins have also been identified (14, 32, 37); for instance, c-Maf activates crystalline genes during lens development, and MafA activates insulin gene expression in pancreatic β-cells. An important observation is that Maf homodimers have a strict requirement for the extended GC sequence flanking each side of the TRE or CRE core of the MARE. In contrast, the NF-E2 binding site (TGCTGA^G/$_C$TCAT) and the antioxidant response element (ARE) (TGA^G/$_C$NNNGC) are composed of half a TRE and half a MARE. The NF-E2 binding site is an important *cis*-regulatory element for the expression of erythroid genes (30), while the ARE is important for the expression of antioxidant and detoxification enzyme genes (39). The critical contribution of the flanking GC sequence to the activities of these *cis*-acting elements revealed that small Maf proteins heterodimerize with NF-E2 p45 and Nrf2 to interact with the NF-E2 binding site and the ARE,

---

* Corresponding author. Mailing address for Hirofumi Kurokawa: Tohoku University Graduate School of Medicine, 2-1 Seiryo-cho, Aoba-ku, Sendai 980-8575, Japan. Phone: 81-22-717-8085. Fax: 81-22-717-8090. E-mail: hkuro@m.tains.tohoku.ac.jp. Mailing address for Toshiyuki Tanaka: Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8572, Japan. Phone: 81-29-853-6706. Fax: 81-29-853-6706. E-mail: ttanaka @tara.tsukuba.ac.jp.
† M.Y. and T.T. contributed equally to this study.
▽ Published ahead of print on 21 September 2009.

respectively (11, 13). Small Maf proteins bind to the MARE half, and CNC and Bach proteins are positioned on the TRE half, indicating that CNC and Bach proteins possess similar binding specificities for AP1-type recognition. Thus, the importance of the flanking sequence of the MARE has been recognized, but the structural basis for the specific recognition of the MARE by Maf proteins has yet to be attained.

In this regard, at least two structural features of Maf proteins are known to be involved in their recognition of the MARE. One is the deviation of critical amino acid residues in the basic region from those highly conserved in other bZip proteins (4). In particular, a tyrosine in the basic region of Maf is a unique substitution of the critical alanine residue found in all the other bZip proteins that is essential for DNA recognition (5, 7, 18, 19). The other is the extended homology region (EHR) on the N-terminal side of the basic region (41). The EHR is highly conserved among Maf proteins and plays an important role in stabilizing the DNA binding of MafG (17). However, the role of the EHR for target DNA sequence specificity has not been elucidated yet because of the lack of Maf-DNA complex structure.

To explore the structural basis by which Maf proteins recognize the unique DNA sequence of the MARE, we performed X-ray crystallography of a MafG homodimer in complex with its cognate DNA. The complex structure including MafG polypeptides (covering the EHR, basic region, and leucine zipper region) and the MARE oligonucleotide duplex enabled us to estimate its unique DNA binding mode directly. We also compared the crystal structure of the MafG homodimer with the crystal structures of Fos-Jun and Skn1, both of which do not demand the GC flanking region for DNA binding, and found that a single α-helix of one monomer grips the half-site of the DNA in the major groove in all three cases. To our surprise, Maf-specific structural elements, including Tyr64 and Thr58 residues in the basic region and EHR en bloc, were not involved in the base recognition of the extended region. However, in sharp contrast to the case of Fos-Jun, invariant Asn61 in the basic region of MafG has a distinct side-chain orientation enabling the recognition of the extended MARE sequence, and another invariant residue, Arg57, contacts with the flanking guanine base. Thus, the Maf-specific elements appear to stabilize the unique side-chain orientations of conserved Asn61 and Arg57 toward the extended GC bases through indirect interactions involving DNA backbone phosphates.

MATERIALS AND METHODS

Plasmid construction, bacterial expression, and protein purification for crystallization. The cDNA fragments encoding mouse MafG(1-123), MafG(21-123), and MafG(45-123) were amplified and digested to generate NdeI-BamHI cDNA fragments that were subcloned into the pET15b (Novagen) vector. Likewise, the expression vectors for the truncated mutants MafG(21-110) and MafG(21-117) were prepared. The resultant plasmids containing particular MafG fragments were transformed into BL21-CodonPlus(DE3)-RIL (Stratagene) cells for overexpression. Crude bacterial lysates were purified by Ni-nitrilotriacetic acid Superflow affinity chromatography (Qiagen). The partially purified proteins were cleaved with thrombin (Calbiochem) and further purified sequentially using Q-Sepharose Fast Flow (GE Healthcare) ion-exchange chromatography, Ni-nitrilotriacetic acid Superflow affinity chromatography, S2 (Bio-Rad) ion-exchange chromatography, HiLoad 16/60 Superdex 75-pg (GE Healthcare) size exclusion chromatography, and Benzamidine Sepharose 6B affinity chromatography. Approximately 25 mg of purified protein was obtained from 1 liter of

culture. To express the selenomethionine (Se-Met)-substituted MafG protein, *Escherichia coli* methionine auxotrophic strain B834(DE3) (Novagen) was used.

Crystallization and X-ray data collection. The MafG-DNA complex was crystallized by the hanging-drop vapor diffusion method. The MafG-DNA complex (0.3 mM in 20 mM HEPES [pH 7.0], 0.15 M sodium chloride, and 5 mM dithiothreitol) was crystallized by mixing with an equal volume of buffer containing 0.1 M sodium acetate (pH 5.0), 40 mM magnesium chloride hexahydrate, 8% 2-methyl-2,4-pentandiol, and 4 mM Tris(2-carboxyethyl)phosphine at 25°C. The crystals were soaked in cryoprotecting mother liquor and flash-frozen in a nylon CryoLoop in a cold nitrogen stream before data collection. Diffraction data were collected with a microfocus X-ray source on a beam-line BL-17A apparatus at Photon Factory (KEK, Tsukuba, Japan). A crystal was picked up with a nylon loop and immediately frozen with cryogenic nitrogen gas.

Structure determination. The structure of the MafG-DNA complex was solved using the multiwavelength anomalous diffraction method with a Se-Met MafG (Table 1). Diffraction data were processed with HKL2000 (34). Se-Met sites were located using SHELXD (40). Phases from these sites were calculated and refined by solvent flattening, as implemented in SHELXE (40). The ensuing electron density was of good quality, enabling the main chain to be clearly traced and most of the side chains of the proteins to be well fitted. Refinement against the 2.8-Å data set was completed by iterative cycles of model building with Xfit (25) and crystallographic refinement using torsion angle molecular dynamics with CNS, version 1.2 (2). The electron density was of high quality throughout the majority of the map. Figures were drawn by use of Molscript (20) and PyMol (3).

EMSA. Electrophoretic mobility shift assays (EMSAs) were carried out as reported previously (22, 44). Briefly, purified proteins were mixed with probes and incubated in gel shift buffer [20 mM HEPES (pH 7.9), 20 mM KCl, 5 mM dithiothreitol, 4 mM MgCl₂, 1 mM EDTA, 100 μg/ml bovine serum albumin, and 400 μg/ml poly(dI-dC)] at 37°C for 30 min. The resulting mixture was subjected to polyacrylamide gel electrophoresis and visualized by autoradiography. The homodimer binding activity was examined using 2 to 250 nM of the MafG(1-123), MafG(21-123), and MafG(45-123) proteins. The consensus T-MARE probe was generated by annealing the two oligonucleotides, 5′-TCGAGCTCGGAAT<u>TGC TGACTCAGCA</u>TTACTC-3′ and 5′-CGAGAGTAA<u>TGCTGAGTCAGCA</u>AT TCCGAGCT-3′ (underlined bases comprise 13-bp MARE or MARE-related sequences [BglHS4: GAGTGACTCAGCC], which are flanked by arbitrary sequences).

The missing-phosphate assay was performed according to a protocol described previously (43). MafG(1-123) and p45ΔN were used at 0.3 μM and 0.1 to 3.0 μM, respectively. The parental double-stranded DNA probe was generated by annealing the sense strand 5′-ACGGAAT<u>GAGTGACTCAGCC</u>TTACTCCAT ACGTTG-3′ (BglHS4-S) and the antisense strand 5′-TCAACGTATGGAGTA A<u>GGCTGAGTCACTC</u>ATTCCG-3′ (BglHS4-AS). To generate the "nicked MARE 1" probe, BglHS4-S and the 5′ portion (5′-TCAACGTATGGAGTAA <u>G</u>-3′) of the complementary strand were annealed, and each 5′ end was labeled with [γ-³²P]ATP. The 3′ portion (5′-<u>GCTGAGTCACTC</u>ATTCCG-3′) of the complementary strand was then annealed. The "nicked MARE 2" probe was generated by annealing BglHS4-AS and the 5′ portion (5′-ACGGAAT<u>GAGTG ACTCAGC</u>-3′) of the complementary strand. Each 5′ end was labeled with [γ-³²P]ATP before annealing the 3′ portion (5′-<u>C</u>TTACTCCATACGTTG-3′) of the complementary strand.

Protein structure accession number. The atomic coordinates (accession number 3A5T) from this study have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org).

RESULTS

Structural features of the MARE and Maf proteins. Of the various bZip transcription factors, only the Maf family members recognize the exceptionally long *cis*-acting DNA sequence of the MARE. A TRE or CRE core flanked on each side by GC sequences constitutes the MARE (Fig. 1A). No other member of the bZip family requires the GC flanking sequence adjoining the core (Fig. 1A). Previous reports (4, 17) suggested that at least two structural elements are responsible for the unique DNA recognition property of the Maf proteins. One is a tyrosine residue in the basic region (shown in pink in Fig. 1B and D), and the other is the EHR domain localized on the N-terminal side of the basic region (Fig. 1C and D), both of

TABLE 1. Data collection and refinement statistics

Parameter	Value		
	Infection	Peak	High-remote
Data collection			
Cells a, b, c (Å)	40.36, 185.24, 230.26	40.37, 185.28, 229.98	40.38, 185.22, 229.93
Space group		$C222_1$	
Wavelength (Å)	0.97939	0.97898	0.96416
Resolution (Å)	40–2.80 (2.90–2.80)[a]	40–2.80 (2.90–2.80)	40–2.80 (2.90–2.80)
Measured no. of reflections	284,443	255,595	271,269
No. of unique reflections	22,101	22,427	22,005
Completeness (%)	98.9 (92.6)	97.8 (88.8)	97.9 (88.1)
Redundancy (%)	7.0 (5.5)	6.3 (4.4)	6.8 (5.1)
R_{sym}[b]	0.068 (0.182)	0.075 (0.220)	0.079 (0.221)
Mean $I/\sigma(I)$	35.4 (7.5)	34.3 (7.1)	42.4 (11.9)
Refinement			
R_{work}/R_{free}[c]	0.261/0.294		
No. of atoms	2,129		
No. of protein atoms	1,438		
No. of nucleic acid atoms	609		
Water and ion	84		
B factors (Å^2)			
Protein (subunit A, subunit B)	74.5, 97.0		
Nucleic acid (strand C, strand D)	49.5, 47.1		
Water and ion	46.4		
Root mean square deviations			
Bond length (Å)	0.013		
Bond angle (°)	1.4		

[a] The highest-resolution shell is shown in parentheses.
[b] $R_{sym} = (\Sigma|I - <I>|/\Sigma I)$, where I is the observed intensity and $<I>$ is the average intensity of symmetry-related reflections.
[c] $R_{work} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} is the R value calculated for a subset of 10% of randomly selected reflections not used in the refinement.

which are highly conserved in Maf proteins. Members of the CNC family contain a unique CNC domain in the position corresponding to the EHR, but the CNC domain is apparently diverged from the EHR domain.

Crystallization and structure determination of MafG. Since full-length MafG was prone to degradation during purification, it was expressed in bacteria and purified as the three MafG truncations, MafG(1-123), MafG(21-123), and MafG(45-123), as shown in Fig. 1C. EMSAs with a MARE probe revealed that the deletion of the EHR [i.e., MafG(45-123)] reduced the DNA binding ability significantly. The deletion of 20 amino acids from the N terminus [MafG(21-123)] slightly weakened the DNA binding ability. Although the N terminus of MafG was considered to make an auxiliary contribution to DNA binding, MafG(21-123) retained substantial DNA binding ability. Therefore, MafG(21-123) containing the EHR and bZip domains was used for cocrystallization. A 15-bp MARE sequence termed MARE25 (GTGCTGACTCATCAG) (Fig. 2A) was chosen as a target DNA (16). The MARE25 sequence contains the NF-E2 binding site, which is critical for erythroid-specific gene regulation, and was previously shown to bind the Maf homodimer tightly (44). As shown in Fig. 2A, the central nucleotide of MARE25 was set as 0, and the other nucleotides were numbered accordingly. The rationale behind this choice was that since MARE25 contains one nucleotide deviation from the GC-(TRE core)-GC consensus sequence, we could assess the singular DNA recognition of each MafG monomer.

The orthorhombic crystals diffracted anisotropically up to

2.6 Å using synchrotron X-ray radiations. The asymmetric unit included two MafG fragments and one double-stranded 15-bp DNA. The structure was solved by Se-Met multiple-wavelength anomalous diffraction and was refined at a resolution of 2.8 Å (Table 1). In the electron density map, residues 113 to 123 lacked electron density, suggesting a disordered structure. To improve the quality of the crystal, we also tried crystallization with C-terminally truncated MafG(21-117) and MafG(21-110), but those attempts were unsuccessful.

Overall structure of the MafG-DNA complex. Since MARE25 is asymmetric, each MafG monomer in the dimeric MafG-DNA complex recognized a slightly different DNA sequence but displayed a similar overall structure (Fig. 2A and B). The asymmetry of the DNA sequence used in the crystallization allowed a distinction to be made between the two MafG monomers. Subunit A was in contact with the GA flanking region, and subunit B was in contact with the GC flanking region. The structural differences between subunits A and B of MafG appear to be due simply to the different crystal packing. Our analysis revealed the presence of three α-helices, H1 (residues 26 to 30), H2 (residues 35 to 39), and H3 (residues 46 to 99), in MafG. The Maf-specific structural element EHR consists of H1, H2, and the N-terminal region of H3. This observation agrees well with the previously reported solution structure of monomeric MafG(1-76) (21). H3 (residues 46 to 99) corresponds to the bZip structure. H1 and H2 within the EHR interact with the basic region (residues 46 to 62 in H3) through both hydrophobic and hydrophilic interactions. Like GCN4
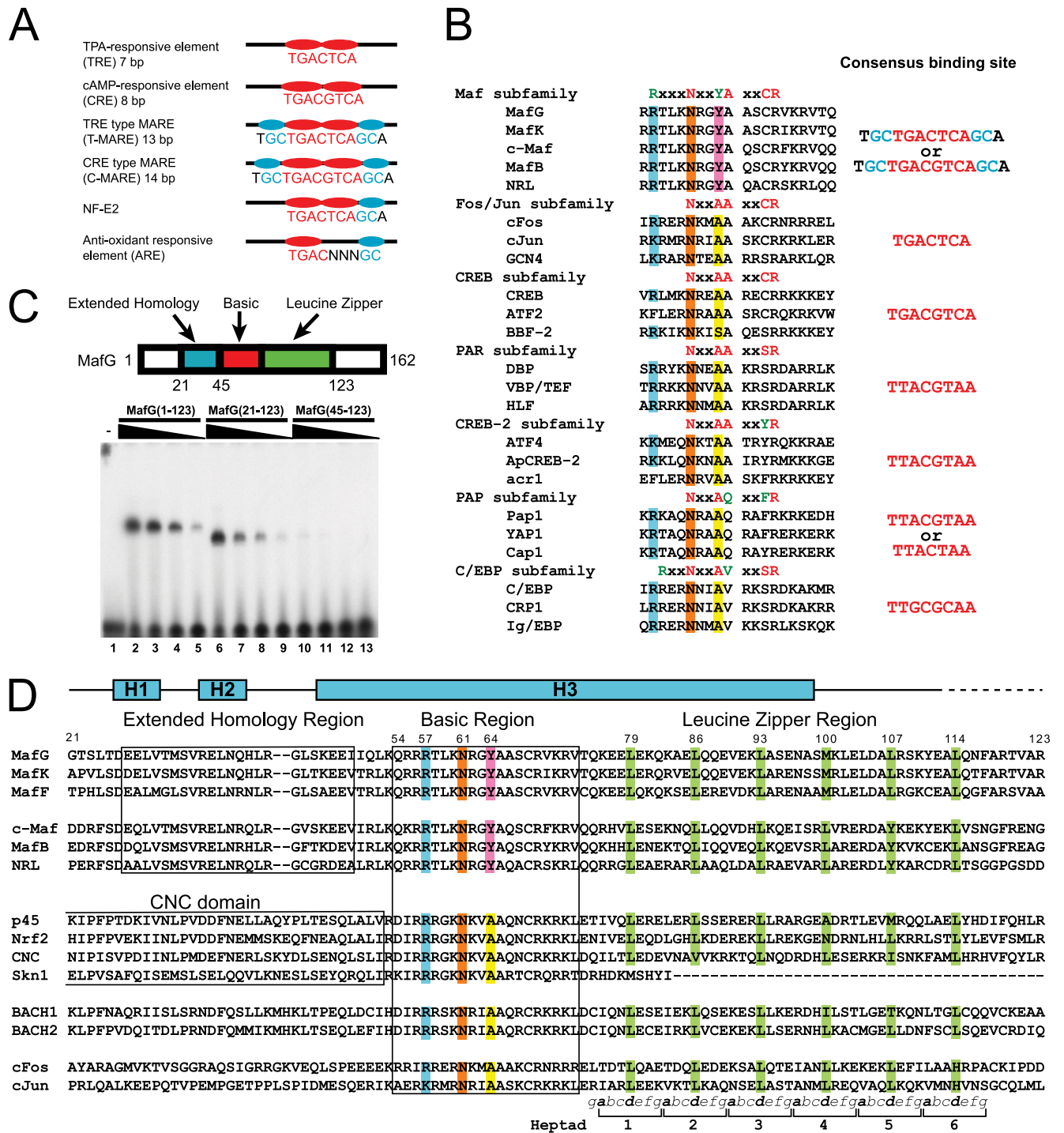
FIG. 1. Maf recognition elements and sequence conservation in Maf-type, CNC-type, and AP1-type bZip factors. (A) Consensus DNA sequences recognized by bZip transcription factors. The TRE and the CRE lie within the T-MARE and the C-MARE, respectively, and act as the established core sequences of these MAREs. Three bases on each side of the core form the flanking sequence, which is critical for the recognition of the MARE by the Maf protein family. The NF-E2 binding site and ARE are asymmetrical sequences composed of half of the TRE site and half of the MARE site with one flanking sequence. The core and flanking regions are in red and blue, respectively. cAMP, cyclic AMP. (B) Sequence alignment of various subfamilies of dimerizing bZip transcription factors and their consensus sequences. The conserved arginine (Arg57 of MafG) and asparagine (Asn61 of MafG) are highlighted in blue and orange, respectively. Maf-specific tyrosine and the corresponding alanine (serine for BBF-2) are in pink and yellow, respectively. (C) EMSA of MafG mutant molecules with a probe containing the consensus MARE. MafG(1-123), MafG(21-123), and MafG(45-123) were incubated with the MARE probe. The concentrations of each protein are 250 nM (lanes 2, 6, and 10), 50 nM (lanes 3, 7, and 11), 10 nM (lanes 4, 8, and 12), and 2 nM (lanes 5, 9, and 13). (D) Sequence alignment of small Maf, large Maf, CNC, Bach, and AP1-type transcription factors. The EHR of Maf proteins, the basic regions, and the partial CNC domains of the CNC family are boxed. H1, H2, and H3 along the top indicate the α-helical regions. The amino acid sequences of the EHR and CNC domains are aligned to each other according to structural correspondence. The amino acid positions in the leucine heptad repeats are indicated at the bottom, and amino acids at position "d" are highlighted in green.
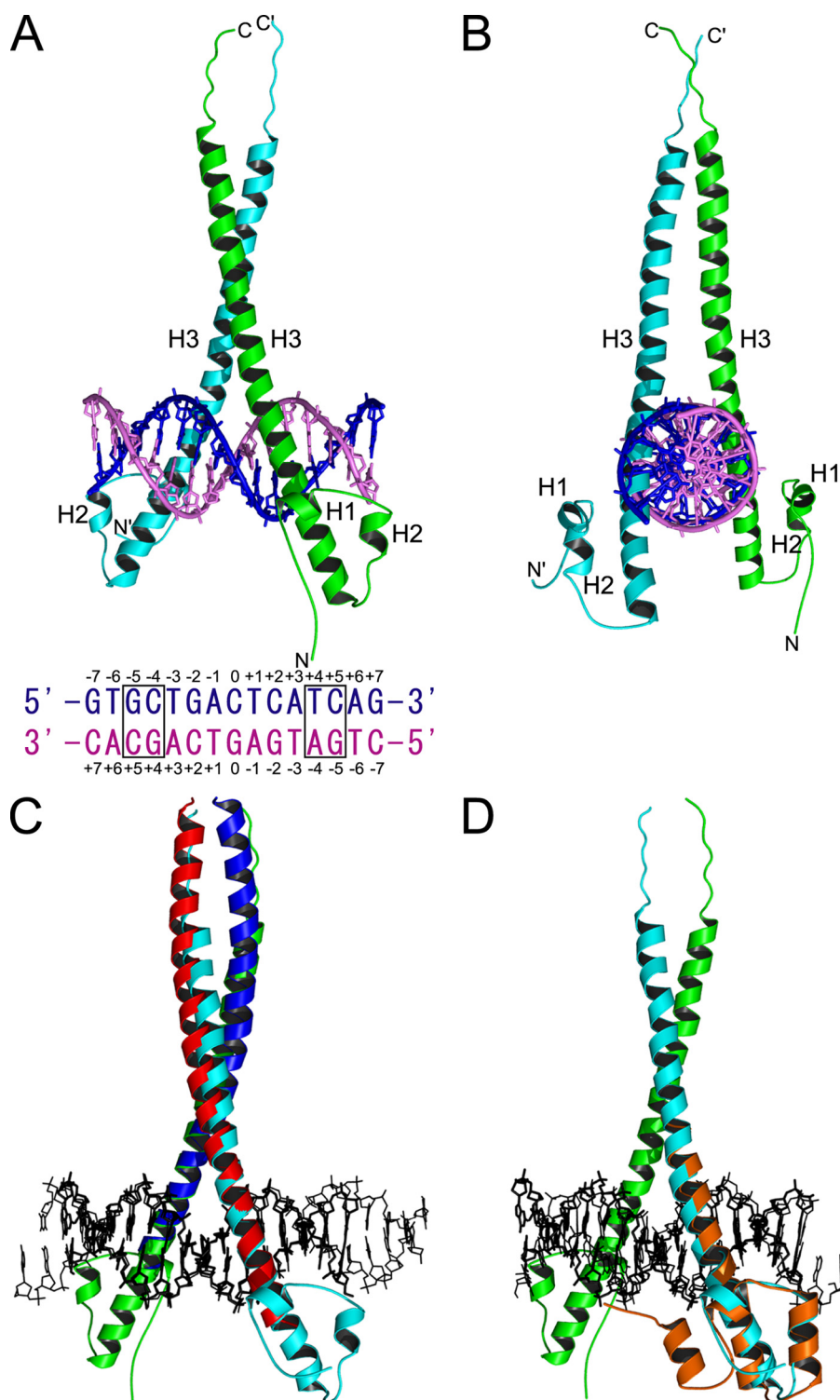
FIG. 2. Main-chain structure of MafG in complex with MARE DNA. (A and B) Orthogonal views of MafG in complex with its cognate DNA. MafG consists of three helices: H1 and H2 in the EHR and H3 in both the basic region that recognizes DNA and the leucine zipper region that forms a dimer with the other subunit. The sequence of the duplex oligonucleotide MARE25 that was used for cocrystallization is shown (A). Subunit A (in green) is in contact with the GA flanking region, and subunit B (in cyan) is in contact with the GC flanking region. (C) Comparison with Fos (in red)-Jun (in purple) in complex with DNA. The AP1 cores are superimposed. Structural alignment yields root mean square deviations for 106 superimposed $C\alpha$ atoms ($rmsd_{106C\alpha}$) of 2.5 Å when the AP1 cores are superimposed and 1.7 Å when the protein regions are superimposed. (D) Comparison of MafG with Skn1 (in orange). The protein regions are used for superimposition. Structural alignment yields $rmsd_{57C\alpha}$ of 2.4 Å when the AP1 half-cores are superimposed and 1.8 Å when only the protein regions are superimposed.
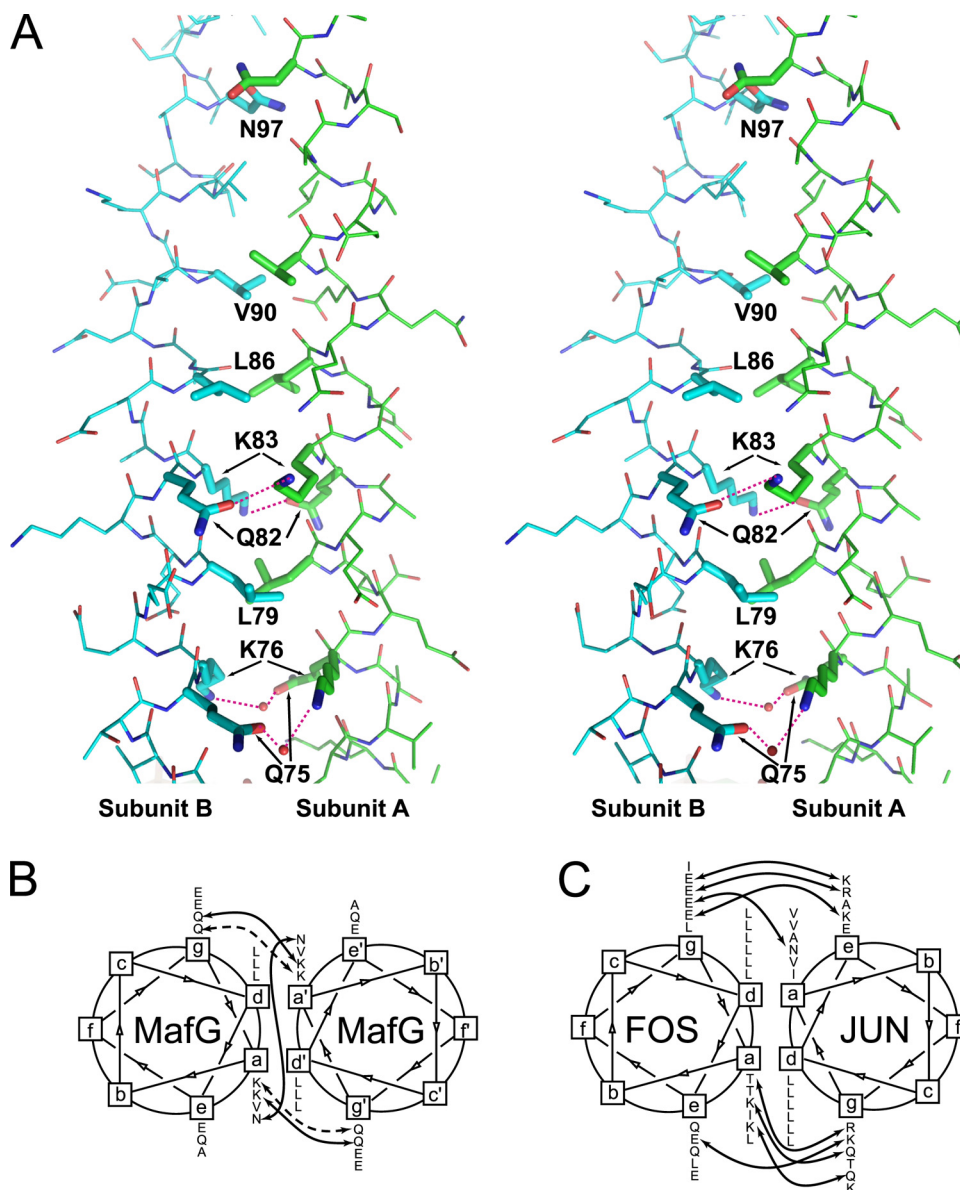
FIG. 3. Interhelical interactions. (A) Detailed view of the leucine zipper region. Amino acids constituting the heptad repeat region (subunit A, in green) and those forming interactions as a counterpart (subunit B, in blue) are indicated. Dotted lines indicate hydrogen bonds. Intersubunit interactions between each Gln82 and Lys83 and between each Gln75 and Lys76 via a water molecule are indicated. (B and C) Helical-wheel diagram of the MafG homodimer (B) and the Fos-Jun heterodimer (C). The coiled-coil sequence is read from the N terminus to the C terminus outwards from the wheel. Arrows with solid lines and broken lines indicate direct and indirect (via a water molecule) electrostatic interactions, respectively.

and Fos-Jun, the MafG homodimer embraces the major groove of the DNA via its two long H3 helices akin to a pair of chopsticks (Fig. 2A and B). The DNA in the complex is in a straight B form and has an average helical rise and twist of 3.2 Å and 34.6°, respectively.

We compared three structures: the MafG-DNA complex, the AP1-type bZip transcription factor (heterodimeric Fos-Jun [7]) in complex with the TRE, and the CNC-type bZip transcription factor (monomeric Skn1 [38]) in complex with the TRE half-site. This comparison revealed that the basic region of MafG shows practically the same structure as the basic region of the Fos-Jun dimer (Fig. 2C). While Skn1 has an

additional H0 helix on the N-terminal side, its H1 and H2 are essentially similar to H1 and H2 of the MafG EHR (Fig. 2D). Thus, in spite of the marked difference in the recognition sequence, MafG displays a main-chain structure similar to those of the factors possessing AP1-type and CNC-type binding specificities.

More closely examining the leucine zipper region of MafG, the interhelical contacts are mediated by leucine residues in the repeated heptad motif, as was described for other bZip proteins. The homodimer is stabilized by a large hydrophobic interface involving Lue79, Leu86, and Val90 (Fig. 3A). Figure 3B shows a helical-wheel representation of the MafG homo-

dimer and indicates interhelical electrostatic interactions. The MafG coiled coil reveals a few unusual residues (Lys76, Lys83, and Asn97) at the "a" positions in the heptad repeat (Fig. 1D). The side chain of each Lys76 hydrogen bonds with Gln75 at the "g" position of the other monomer via a water molecule (Fig. 3A and B). Lys83 at the "a" position shows intersubunit electrostatic interactions with Gln82 at the "g" position, while the Asn97 residues at the "a" position also link the subunits by hydrogen bonding (Fig. 3B). The two pairs of electrostatic interactions (Gln82-Lys83 and Gln75-Lys76) appear to be conserved or replaced by similar interactions (Gln-Arg, Glu-Lys, or Glu-Arg) in all Maf protein family members (Fig. 1D). The intersubunit Asn97 interaction is specific to small Maf proteins, since the corresponding residue is replaced with Ile or Val for large Maf proteins. This may, at least in part, account for the fact that small Maf proteins cannot form heterodimers with large Maf proteins (15). Interestingly, large Maf can form a heterodimer with Fos or Jun (17). In contrast, small Maf can interact with Fos but not with Jun (15). Interhelical electrostatic contacts of the Fos-Jun heterodimer (Fig. 3C) appear to be distinct from those of the MafG homodimer (Fig. 3B), and conserved interhelical interactions were not observed. Comparisons between the MafG homodimer (Fig. 3B) and the Fos-Jun heterodimer (Fig. 3C) suggest that additional interhelical contacts that determine the dimerization specificity should be formed upon the heterodimerization of MafG with Fos.

Critical contribution of Arg57 and Asn61 to MafG DNA recognition. The recognition of the MARE by Maf consists of a broad specificity against the TRE/CRE core element but a strict requirement of the flanking region (17, 44). Each monomer of the dimeric MafG-DNA complex interacts with the major groove containing a TRE half-site (Fig. 4A and B) and the flanking regions (Fig. 4A and B). The invariant Asn61 is located deep within the major groove of the recognition site (Fig. 4A and B), and its electron density in the simulated-annealing F_o-F_c omit map for each subunit was unambiguously visible (Fig. 4C and D). We found that the orientation of the Asn61 side chain of MafG is distinct from the side-chain orientation of other bZip factors such as Fos and Skn1 (Fig. 4E and F). While the side chains of the invariant Asn residues in Fos and Skn1 orientate toward the TRE core, the side chain of MafG Asn61 bears toward the flanking regions. Arg57 is another residue that is well conserved in the bZip factors, but irrespective of its conservation among bZip factors, Arg57 is involved in DNA recognition only in Maf proteins. Our results indicate that Arg57 plays a critical role in the recognition of the extended DNA sequence. Indeed, the simulated-annealing F_o-F_c omit map revealed an unambiguous side-chain orientation of Arg57 in each subunit (Fig. 4C and D).

We summarized all contacts between the MafG homodimer and the MARE25 DNA duplex in Fig. 5A. Showing very good agreement with the case for AP1-type bZip factors, conserved Ala65 residues of MafG create van der Waals contacts with the T base at positions +1 and −3. Arg69 of subunit B forms a hydrogen bond with the G base at position 0 through a water molecule. A series of salt bridges is formed between Arg62 and the phosphate backbone of thymine at position +1, between Arg69 and the phosphate backbone of adenine at position −1, and between Lys71 and the phosphate oxygen of thymine at

position −3. Arg56 of subunit B interacts with the phosphate backbone of thymine at position −7.

The DNA base recognition modes of the MafG monomers and Skn1 are depicted in Fig. 5B to F, concomitant with those of Pap1 and C/EBP. The conserved amino acid side chains of the protein motifs that make direct contacts with the DNA bases of the consensus DNA sequence as well as hydrogen bonds and van der Waals contacts are shown.

Asn61 of MafG subunit A forms a direct hydrogen bond with adenine (position −4) in the GA flanking sequence (Fig. 5B) and an indirect hydrogen bond with cytosine (position +2) of the opposite sequence (Fig. 5A) in the core region of MARE25. Asn61 of MafG subunit B creates van der Waals forces with cytosine (position −4) in the GC flanking region of MARE25 (Fig. 5C). Thus, MafG Asn61 recognizes mainly the flanking region, although it forms weak indirect hydrogen bonding with the TRE core region. In stark contrast to MafG, the invariant Asn511 of Skn1 interacted only with cytosine (position +2) in the TRE region (Fig. 5D). AP1-type transcription factors such as Fos, Jun, and GCN4 show basically the same DNA recognition mode as Skn1 (7, 9, 39). In the *Schizosaccharomyces pombe* AP1-type factor Pap1, the corresponding Asn residue also recognizes the TRE core region (Fig. 5E) (6).

Arg57 in MafG subunit A forms critical hydrogen bonds with the O6 of guanine at position −5 and the O4 of thymine at position +4 (Fig. 5A and B), while Arg57 in MafG subunit B forms hydrogen bonds to the O6 of guanine at position −5 and the O6 of guanine at position +4 (Fig. 5A and C). It should be noted that Arg57 of MafG participates in DNA base recognition by changing its side-chain orientation; Arg57 in MafG shows a distinct side-chain orientation compared to the corresponding arginine residues of AP1-type and CNC-type factors. In fact, Arg507 and Arg143 residues are the residues corresponding to Arg57 of MafG in Skn1 and Fos, respectively, but the side chain of Arg507 in Skn1 interacts with a backbone phosphate of DNA (see Fig. 7B), and that of Arg143 in Fos is exposed to solvent (7), and these side chains are not involved in DNA base recognition.

Utilization of Arg57 for DNA recognition makes the Maf subfamily possess a unique and exceptionally long consensus protein sequence for DNA base recognition (13 residues, Rxxx NxxYAxxCR) (Fig. 5B and C). In contrast, those of AP1-type/CNC-type factors (NxxAAxx$^C/_S$R) and the yeast AP1-type factor Pap1 (NxxAQxxFR) contain a 9-residue sequence, and the arginine residue corresponding to Arg57 of MafG is not included in the interaction with DNA (Fig. 5D and E). Another characteristic protein sequence is RxxNxxAVxxSR for the C/EBP subfamily, where the arginine residue involving DNA recognition is 1 amino acid closer to the asparagine residue than in MafG. The arginine residue of C/EBP at the position corresponding to Arg57 of MafG in the consensus sequence does not participate in DNA base recognition (Fig. 5F).

Thus, we surmise that the distinct side-chain orientation of Arg57 and Asn61 explains why MafG requires the flanking region for binding and recognizes the TRE core with specificity relatively broader than that of AP1-type and CNC-type transcription factors.

Roles of other Maf-specific elements in the basic region for MARE recognition. Tyr64 in the basic region has been shown to be one of the key determinants of the DNA recognition
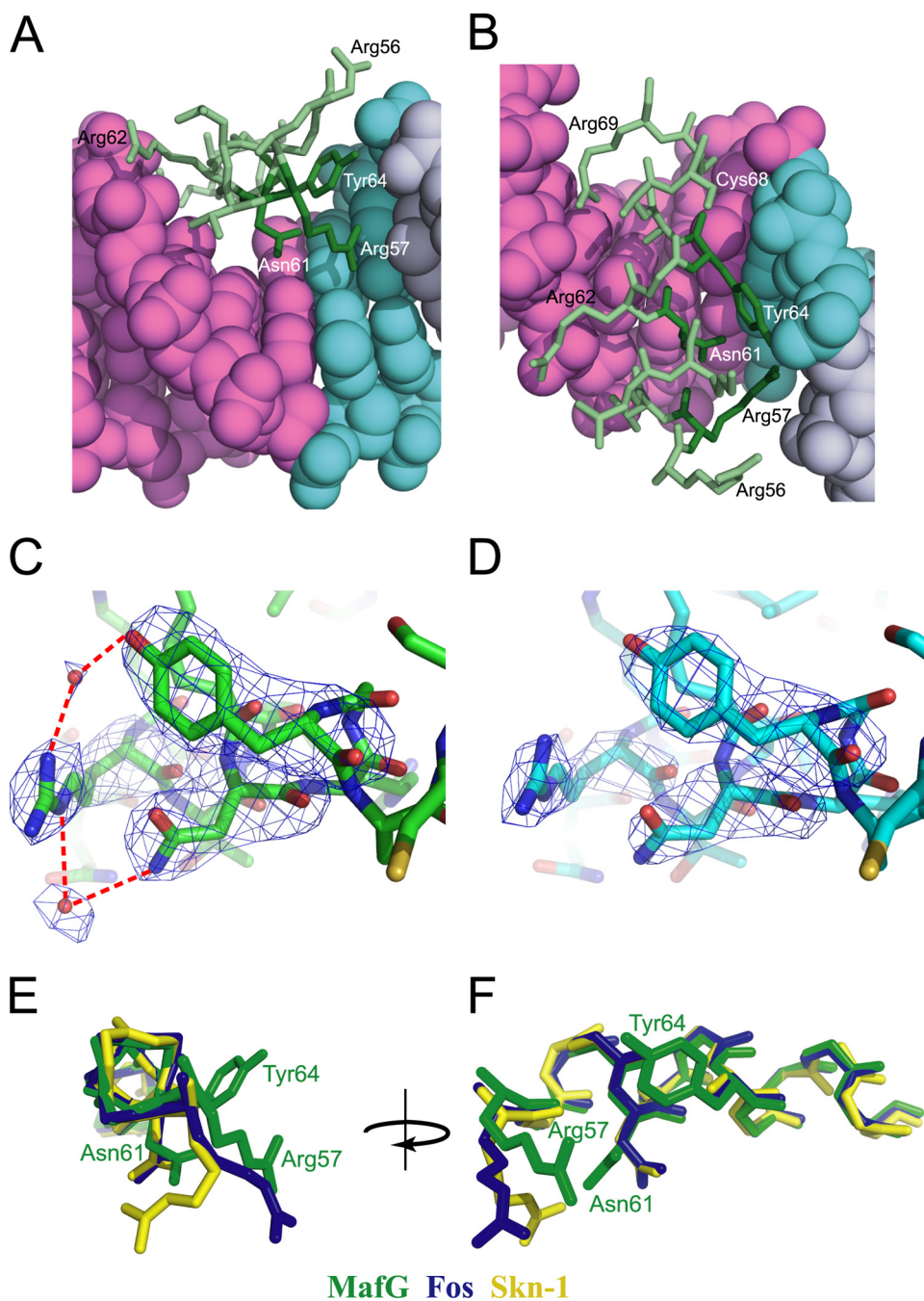
MafG **Fos** **Skn-1**

FIG. 4. Conserved Asn61 and Arg57 in the basic region of MafG contact with the MARE sequence. (A and B) Orthogonal views of the basic region of subunit A. Some of the amino acids making critical contacts are indicated. The AP1 core and extended sequence elements (Fig. 1A) are in pink and blue, respectively. (C and D) The simulated-annealing Fo-Fc omit map is contoured at 5 σ. Arg57, Asn61, and Tyr64 were omitted from the calculation. Note that the electron density of these three residues is unambiguously visible. (C) In subunit A, side chains of Tyr64 and Arg57 contact via a water molecule. An indirect interaction between Arg57 and Asn61 was also observed. Hydrogen bonds are shown as broken lines. (D) In subunit B, these interactions via a water molecule were not observed. (E and F) Comparison of the side chains of conserved residues within MafG (green), c-Fos (blue), and Skn1 (yellow). The image in E was rotated by 90° and is shown in F.

specificity of the Maf protein family (4, 18). Whereas Tyr64 is conserved in all Maf proteins, its matching residue in the other bZip transcription factors is alanine. The electron density in the simulated-annealing Fo-Fc omit map shows the side-chain orientation of Tyr64 in each subunit clearly (Fig. 4C and D), and against our expectation that this residue may be involved

in the recognition of unique bases in the MARE, Tyr64 does not form any direct contacts with DNA bases, but instead, the hydroxyl group of Tyr64 interacts with the phosphate backbone of guanine at position −5 (Fig. 5A).

To examine the input of Tyr64 to the reinforcement of MafG-DNA complex formation, we performed a missing-
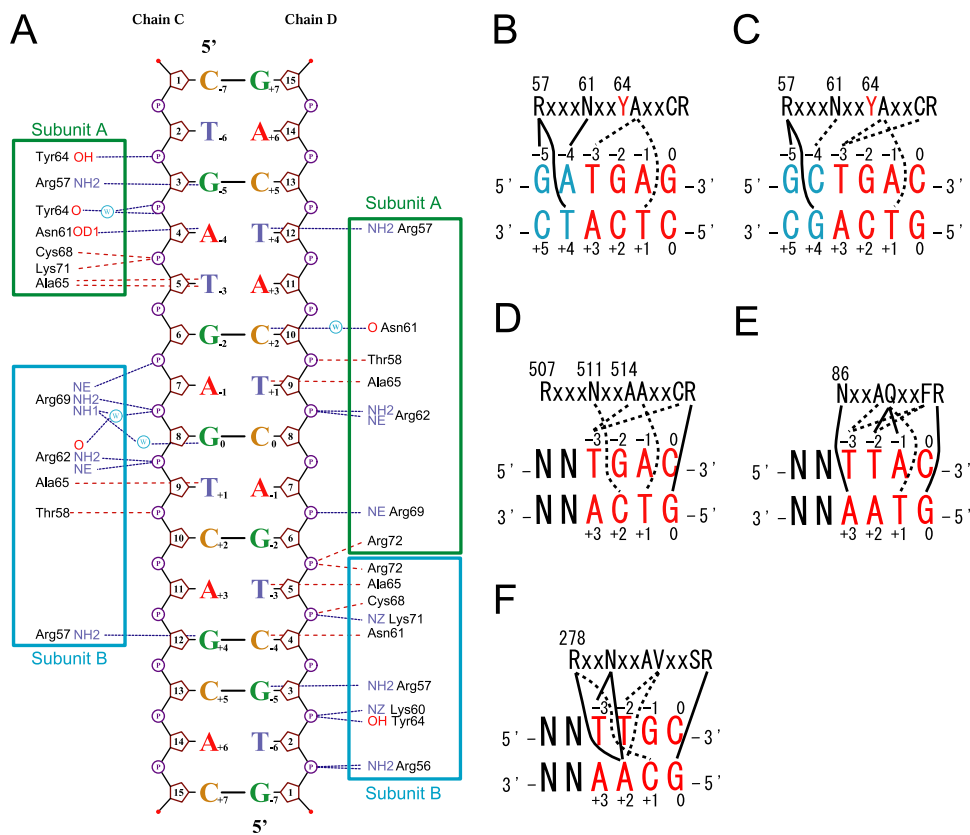
FIG. 5. MARE recognition by the basic region of MafG and comparison with AP1-type bZip factors. (A) The phosphate groups of MARE25 engaged by the MafG protein are shown. Chain C and chain D are the bottom and top strands of MARE25, respectively (Fig. 2A). The figure was drawn by NUCPLOT (23). (B to F) The conserved amino acid side chains of the motifs that make direct contacts with the DNA bases of the consensus sequence are shown for MafG subunit A (B), MafG subunit B (C), Skn1 (D), Pap1 (E), and C/EBP (F). Hydrogen bonds and van der Waals contacts are shown as solid and broken lines, respectively. PDB accession numbers 1SKN for Skn1 (D), 1GD2 for Pap1 (E), and 1GU4 for C/EBP (F) were used to estimate the DNA base recognition mode.

phosphate assay, which is a modified EMSA using a nicked probe lacking one of the backbone phosphates. To allow the unidirectional binding of MafG to DNA, we adopted an asymmetric MARE variant (BglHS4-MARE) possessing a flanking GC sequence on one side as a probe (44) and a p45-MafG heterodimer as a *trans*-acting factor requiring a single GC in the MARE flanking region (Fig. 6A). The p45-MafG heterodimer efficiently bound to the control probe without a nick (Fig. 6B). In contrast, the removal of the 5′-phosphate of guanine at position −5 (nicked MARE-1) dramatically reduced the binding of the heterodimer, while the removal of the 3′-phosphate of cytosine at position +5 (nicked MARE-2) did not affect the binding (Fig. 6B). This testifies that the interaction between Tyr64 and the backbone phosphate of guanine at position −5 is critical for the high-affinity binding of MafG to the MARE.

The crystal structure shows that Tyr64 of subunit A forms an indirect hydrogen bond with the side chain of Arg57 via a water molecule in subunit A (Fig. 4C). Tyr64 appears to stabilize the side-chain orientation of Arg57 that is optimal for the recognition of the flanking guanine. The hydrogen-bonding network extends from Tyr64 to Asn61 through the indirect hydrogen bond formation between Arg57 and Asn61, and the side chain of Asn61 in turn flips out and interacts with the extended DNA

base. Therefore, Tyr64 is vital for maintaining the side chains of both Arg57 and Asn61 in a spatial arrangement ideal for extended GC recognition and consequently stabilizes Maf-type DNA recognition by forming a hydrogen-bonding network with these side chains and extended DNA bases. Corresponding water molecules were not observed in subunit B in the current crystal structure (Fig. 4D), perhaps due to the higher B factor (Table 1) of the corresponding region.

Thr58 is a Maf-specific residue that is well conserved in all Maf proteins. Intriguingly, Thr58 is involved in the interaction that appears to be important for establishing Maf-type DNA recognition. Thr58 develops a van der Waals contact with the backbone phosphate of cytosine at position +2 and forms a concurrent hydrogen bond with Gln54 (Fig. 7A). Gln54 is also well conserved in all Maf proteins and appears to stabilize the Thr58-phosphate interaction by fixing the side-chain orientation of Thr58. Arg507, which corresponds to Arg57 of MafG in the Skn1-DNA complex, forms a hydrogen bond to the analogous backbone phosphate of cytosine at position +2 (38). Arg507 also makes a hydrogen-bonding network with the invariant Asn511 via a water molecule and a DNA backbone phosphate (Fig. 7B). This hydrogen-bonding network seems to lock the side chain of invariant Asn511 into a position that is most favorable for
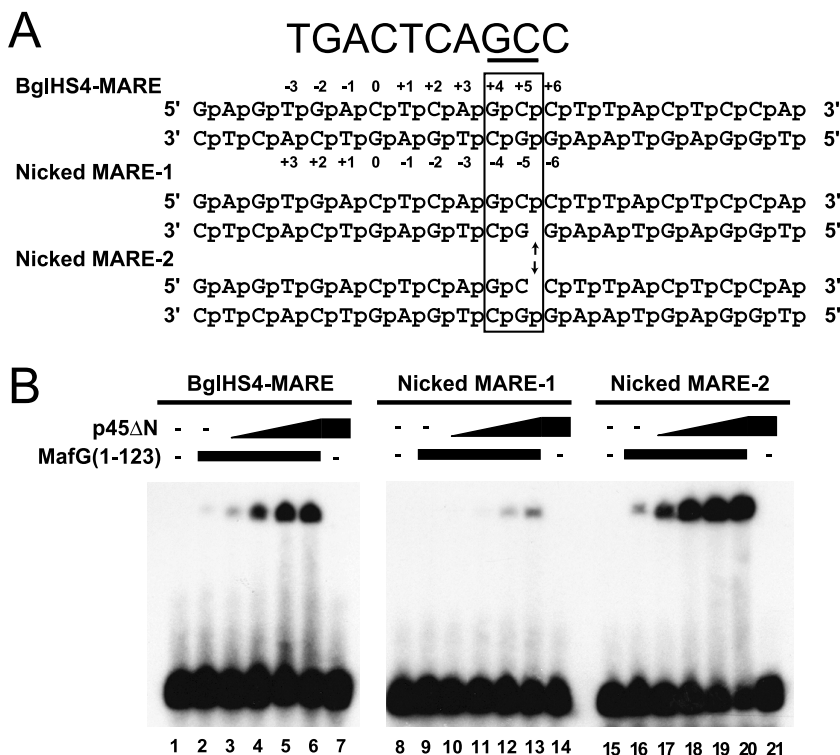
FIG. 6. A missing-phosphate assay demonstrates the critical contribution of the backbone phosphate in the flanking region of MARE. (A) Three probe types were used for EMSAs. BglHS4-MARE is a parent probe of intact double-stranded oligonucleotides containing the conserved core region and one side of the flanking region, which allows the oriented binding of the p45-MafG heterodimer but not of the MafG homodimer. The nicked MARE-1 and MARE-2 probes lack the 5′ phosphate of the flanking "GC" bases and the corresponding phosphate on the other strand, respectively. (B) EMSA of p45ΔN-MafG(1-123) with intact double-stranded parent probe (lanes 1 to 7) and the nicked probes (lanes 8 to 21). p45ΔN is an N-terminally truncated p45 that contains CNC and bZip motifs. The concentrations of p45ΔN are 0.1 μM (lanes 3, 10, and 17), 0.3 μM (lanes 4, 11, and 18), 0.9 μM (lanes 5, 12, and 19), and 3 μM (lanes 6, 7, 13, 14, 20, and 21), with 0.3 μM of MafG added in lanes 2 to 6, 9 to 13, and 16 to 20.

CNC-type DNA recognition (Fig. 4E and F). It is remarkable that the well-conserved arginine and asparagine residues of MafG (Arg57 and Asn61) and Skn1 (Arg507 and Asn511) form a hydrogen-bonding network in a distinct

manner by using a different DNA backbone phosphate as a scaffold in the case of MafG guanine at position −5 and in the case of Skn1 cytosine at position +2 (Fig. 5A and 7B). The Thr58-phosphate interaction in the MafG-DNA com-
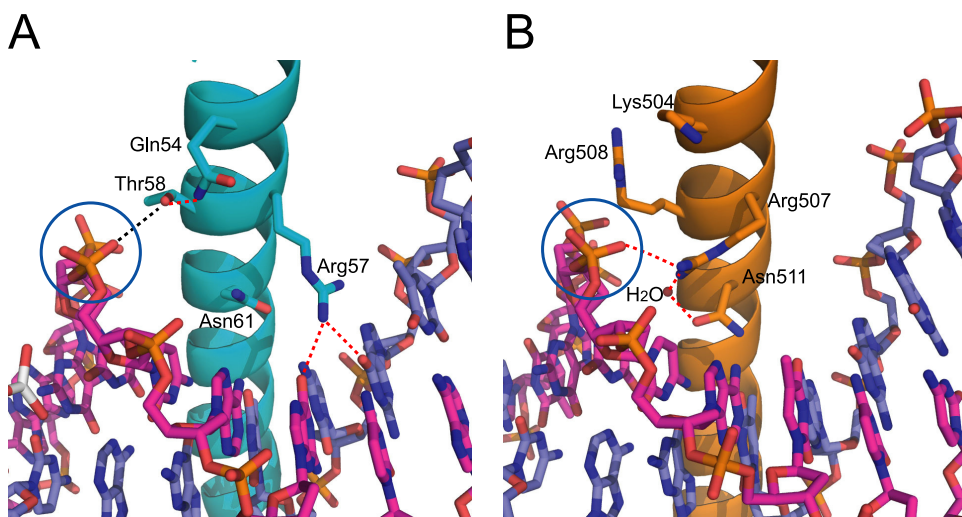

FIG. 7. Interactions of the DNA binding domain of MafG with DNA. (A) van der Waals forces between Maf-specific Thr58 and the DNA backbone phosphate. The hydrogen bond between Gln54 and Thr58 fixes the Thr58 side chain in an optimal orientation against the backbone phosphate. (B) The DNA backbone phosphate referred to above (A) interacts with Arg507 of Skn1. Arg507 corresponds to Arg57 of MafG and forms indirect contacts with the invariant Asn511 via a water molecule, thereby stabilizing the orientation of the Asn511 side chain for recognition of the core TRE half-region.
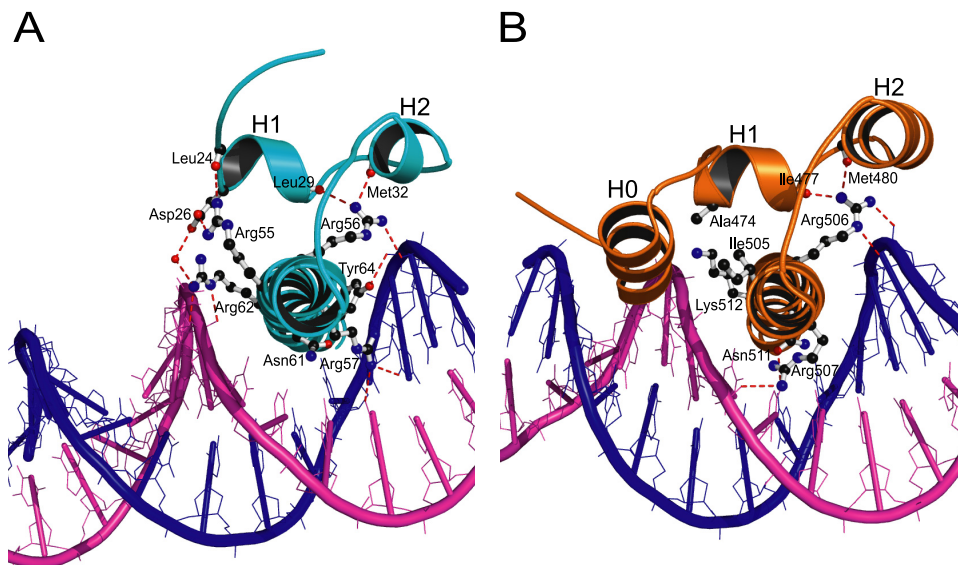
FIG. 8. Comparison of the interactions between basic regions and juxtaposed ancillary domains. (A) Schematic representation of the interactions between the EHR and the basic region of MafG (subunit A). (B) Graphic depiction of the contacts made between the CNC domain and the basic region of Skn1.

plex appears to prevent Arg57 from forming a hydrogen bond with the backbone phosphate of cytosine at position +2, which would otherwise be optimal for CNC-type DNA recognition. Thus, Thr58 stabilizes the alternative side-chain orientation of Arg57, being well suited for extended GC base recognition.

Roles of the EHR in MARE recognition. The Maf-specific EHR element is located on the N-terminal side of the basic region. As shown in Fig. 8A, H1 and H2 of the EHR interact with the basic region in H3. Main-chain oxygen atoms of Leu29 and Met32 interact with the side chain of Arg56. These interactions appear to stabilize the interaction of Arg56 with the DNA backbone phosphate. The side chain of Asp26 in the EHR domain forms electrostatic relations with Arg55 and with Arg62 via a water molecule, which secures the hydrogen bonds between Arg62 and the DNA backbone phosphate.

In a primary amino acid sequence alignment, the EHR is found to be a Maf-specific element. However, where the tertiary structure is concerned, a similarity between the EHR and the CNC domain of Skn1 was previously suggested (21). Our present study revealed similar structural features between the MafG EHR and Skn1 CNC domain. Comparison of the EHR with the Skn1 CNC domain is depicted in Fig. 2D. We also compared the interaction between the EHR and the basic region of MafG with the interaction between the CNC domain and the basic region of Skn1 (Fig. 8A and B, respectively). Ile477 and Met480 in the Skn1 CNC domain interact with Arg506 (Fig. 8B), which resembles the interaction found between the EHR (Leu29 and Met32) and Arg56 in MafG (Fig. 8A). The electrostatic interactions of Asp26 with Arg55 and Arg62 in MafG correspond to the hydrophobic contacts between Ala474 in the CNC domain and Ile505 in the basic region of Skn1. Although the overall structures and part of the interactions with the basic region resemble each other, the

amino acids involved and each interaction among them are substantially different.

The truncation of the EHR significantly reduced the DNA binding capability of MafG in an EMSA experiment (Fig. 1C). Therefore, DNA binding is rendered stable by means of indirect contact between the EHR and the DNA backbone phosphate. Considering the similarity of the interactions mediated by the MafG EHR and the Skn1 CNC domain, the latter also stabilizes DNA binding via interactions with the basic region. Therefore, these results suggest that the roles that the EHR and CNC play may be more oriented toward stabilizing DNA binding rather than conferring specificity to DNA recognition.

DISCUSSION

The current crystal structure analysis of a MafG homodimer in complex with a *cis*-acting MARE revealed a new mode of bZip protein-DNA recognition. This study stemmed from the observation that Maf proteins recognize an unusually long *cis*-acting sequence compared with the sequences bound by other bZip transcription factors, such as the AP1 and CREB/ATF families (27). Examination of the crystal structure revealed that Maf proteins recognize mainly the extended GC (or GA) sequence that flanks the TRE core of the MARE owing to the unique side-chain orientations of Maf Arg57 and Asn61.

It was previously reported that the replacement of tyrosine residue Y287 in c-Maf with alanine does not affect c-Maf binding to the MARE with base substitutions in the extended GC elements, but the mutation inhibits c-Maf binding to the MARE with base substitutions within the TRE core region (4). We recently verified that the substitution of the conserved alanine of the CNC family factor Nrf2 with tyrosine converts the binding specificity of Nrf2 to that of Maf proteins (18). We originally interpreted this observation to be direct contacts being made by Tyr64 of MafG

(corresponding to Y287 of c-Maf) with the DNA bases in the extended binding region such that an alanine-to-tyrosine substitution might create alternative base contacts deep within the major groove. However, the present study has revealed that Arg57 and Asn61 play critical roles in the recognition of the MARE. This was a surprising revelation, as these two residues are not specific to Maf family proteins.

Current data also revealed the indispensable roles that Tyr64 plays for MafG to recognize the MARE. Tyr64 of subunit A makes contact with the backbone phosphate of guanine at position −5 and concomitantly interacts with the side chain of Arg57 via a water molecule. The resulting side-chain orientation of Arg57 enables engagement with the extended GC and GA bases as well as indirect hydrogen bonding with the invariant Asn61, which consequently recognizes the extended region of the MARE. Thus, Maf proteins recognize and bind to the extended DNA sequence in the MARE by virtue of the unique side-chain orientations of these two residues.

We also found that the Maf-specific Thr58 residue contributes to Maf-type DNA recognition by associating with the DNA backbone phosphate in a manner distinct from that of Tyr64. The Tyr64-phosphate interaction with guanine at position −5 stabilizes the hydrogen bond network that is optimal for MARE recognition. On the other hand, the Thr58-phosphate interaction with cytosine at position +2 stabilizes MARE recognition by destabilizing the hydrogen-bonding network that is preferable for CNC-type recognition. Consequently, the Maf-specific structural elements in the basic region, i.e., Tyr64 and Thr58, stabilize the side-chain orientations that are optimal for MARE recognition, and this indicate that changes in the structural elements that do not interact directly with DNA bases but specify the side-chain orientations of amino acids that do interact directly with the DNA bases generate diversity in the binding sequence of bZip factors by altering the fine structure of the basic region.

A previously reported Pap1 structural analysis revealed a somewhat related strategy for generating diversity in the *cis*-acting DNA sequence (6). Pap1 target sequences are different from those recognized by AP1 or CREB/ATF factors. The glutamine and phenylalanine residues unique to the basic region of Pap1 (*N*xx*AQ*xx*F*R) interact directly with the unique bases of the target DNA. The invariant Asn86 (in italics) of Pap1, corresponding to Asn61 of MafG, contributes to the distinct specificity of Pap1 through its alternatively positioned side chain. While Pap1 appears to have adopted the MafG-related strategy, Asn86 of Pap1 recognizes adenine at position +3 in the core region but not the extended DNA sequence (Fig. 5E). We surmise that the use of the same conserved amino acids but in a different manner by taking advantage of distinct side-chain orientations may represent a common strategy for transcription factors to create diversity in their target sequences during molecular evolution.

Our previous nuclear magnetic resonance study suggested structural similarity between the MafG EHR and CNC domain (21), which is consistent with the observations made in this study. The binding of transcription factors to their cognate DNA sequences is stabilized by interactions between the DNA and key residues in the basic regions and the EHR or the CNC domain without those residues directly contacting with the DNA. These results support the contention that although direct base contacts may be restricted to the basic region, other

domains may determine how the basic region is exploited to accept the specific DNA sequence. This phenomenon is also the case for VBP, a member of the PAR subfamily of bZip factors (8).

Biological processes mediated by Maf homodimers or Maf-containing heterodimers depend heavily on the presence of the GC (or GA) sequences flanking the core TRE/CRE sequence. Consistent with the critical role of the extended sequence in *cis*, Maf-type recognition specificity in *trans* was previously shown to define a unique array of target genes (18). As described above, the Nrf2(A502Y) mutant, in which Ala502 (corresponding to MafG Tyr64) is substituted with tyrosine, acquires the Maf-type binding specificity. Since most of the cytoprotective genes are activated by Nrf2-MafG, but not by Nrf2(A502Y)-MafG, the difference between the Maf-type and CNC-type recognition specificities appears to be critical for gene regulation in vivo. Our present study has cast a light on the molecular mechanisms underlying such differential and biologically significant DNA recognitions.

Several missense mutations in the *cMAF* gene in humans and mice have been identified and cause congenital cataracts (9, 24). Mutations reported previously for the basic region of human c-Maf include R299S, which corresponds to Arg62 of MafG (9). The mutation would make c-Maf defective in DNA binding since the corresponding MafG Arg62 is involved in a stabilizing network of electrostatic interactions among Asp26 and Arg55. Similarly, R291Q (Arg291 corresponds to Arg57 of MafG) is a mutation in the basic region of mouse c-Maf. The R291Q mutation in mouse c-Maf was previously reported to alter the DNA binding specificity (24). The mutation resulted in a selective alteration in the DNA binding affinity for target oligonucleotides containing variations of the core CRE and TRE, suggesting that the c-Maf R291Q mutant might acquire DNA binding specificity similarly to AP1-type DNA recognition. With a glutamine residue in place of Arg57, the side chain of Asn61 may fail to flip out to contact with the flanking region of the MARE.

It is interesting that the hyperactivation of Maf-containing dimers often relates to pathological states (1, 36), which are due to the enhanced recruitment of p300. A quantitative enhancement of c-Maf transcription was also observed for roughly half of the patients with multiple myeloma, and *c-MAF* gene translocation has been found in 5 to 10% of the cases (10). A increased accumulation of the Nrf2-MafG heterodimer turned out to be quite common in lung cancer and confers drug resistance to tumor cells (33, 35). An elucidation of the Maf-type DNA recognition mechanisms may facilitate the development of small molecules that can disrupt the Maf-DNA interaction, thereby alleviating pathological states related to the hyperactivation of Maf-containing dimers.

In summary, the delineation of the crystal structure of the MafG homodimer in complex with a MARE oligonucleotide unequivocally demonstrated that two residues, Arg57 and Asn61, in the basic region of MafG, which are well-conserved among bZip transcription factors, play critical roles for MafG recognition of the MARE through direct contacts with the extended GC bases in the motif. Whereas the Maf-specific structural elements EHR and Tyr64 do not interact directly with the extended GC bases, these elements stabilize the side-chain orientation of Arg57 and Asn61 residues indirectly and

contribute to the specific binding of MafG to the unusually long MARE motifs.

REFERENCES

1. **Bessant, D. A., A. M. Payne, K. P. Mitton, Q. L. Wang, P. K. Swain, C. Plant, A. C. Bird, D. J. Zack, A. Swaroop, and S. S. Bhattacharya.** 1999. A mutation in NRL is associated with autosomal dominant retinitis pigmentosa. Nat. Genet. **21:**355–356.
2. **Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. **54:**905–921.
3. **DeLano, W. L.** 2002. The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA.
4. **Dlakic, M., A. V. Grinberg, D. A. Leonard, and T. K. Kerppola.** 2001. DNA sequence-dependent folding determines the divergence in binding specificities between Maf and other bZIP proteins. EMBO J. **20:**828–840.
5. **Ellenberger, T. E., C. J. Brandl, K. Struhl, and S. C. Harrison.** 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. Cell **71:**1223–1237.
6. **Fujii, Y., T. Shimizu, T. Toda, M. Yanagida, and T. Hakoshima.** 2000. Structural basis for the diversity of DNA recognition by bZIP transcription factors. Nat. Struct. Biol. **7:**889–893.
7. **Glover, J. N., and S. C. Harrison.** 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. Nature **373:**257–261.
8. **Haas, N. B., C. A. Cantwell, P. F. Johnson, and J. B. Burch.** 1995. DNA-binding specificity of the PAR basic leucine zipper protein VBP partially overlaps those of the C/EBP and CREB/ATF families and is influenced by domains that flank the core basic region. Mol. Cell. Biol. **15:**1923–1932.
9. **Hansen, L., H. Eiberg, and T. Rosenberg.** 2007. Novel MAF mutation in a family with congenital cataract-microcornea syndrome. Mol. Vis. **13:**2019–2022.
10. **Hurt, E. M., A. Wiestner, A. Rosenwald, A. L. Shaffer, E. Campo, T. Grogan, P. L. Bergsagel, W. M. Kuehl, and L. M. Staudt.** 2004. Overexpression of c-maf is a frequent oncogenic event in multiple myeloma that promotes proliferation and pathological interactions with bone marrow stroma. Cancer Cell **5:**191–199.
11. **Igarashi, K., K. Kataoka, K. Itoh, N. Hayashi, M. Nishizawa, and M. Yamamoto.** 1994. Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. Nature **367:**568–572.
12. **Igarashi, K., and J. Sun.** 2006. The heme-Bach1 pathway in the regulation of oxidative stress response and erythroid differentiation. Antioxid. Redox Signal. **8:**107–118.
13. **Itoh, K., T. Chiba, S. Takahashi, T. Ishii, K. Igarashi, Y. Katoh, T. Oyake, N. Hayashi, K. Satoh, I. Hatayama, M. Yamamoto, and Y. Nabeshima.** 1997. An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. Biochem. Biophys. Res. Commun. **236:**313–322.
14. **Kataoka, K., S. I. Han, S. Shioda, M. Hirai, M. Nishizawa, and H. Handa.** 2002. MafA is a glucose-regulated and pancreatic beta-cell-specific transcriptional activator for the insulin gene. J. Biol. Chem. **277:**49903–49910.
15. **Kataoka, K., K. Igarashi, K. Itoh, K. T. Fujiwara, M. Noda, M. Yamamoto, and M. Nishizawa.** 1995. Small Maf proteins heterodimerize with Fos and may act as competitive repressors of the NF-E2 transcription factor. Mol. Cell. Biol. **15:**2180–2190.
16. **Kataoka, K., M. Noda, and M. Nishizawa.** 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.
17. **Kerppola, T. K., and T. Curran.** 1994. A conserved region adjacent to the basic domain is required for recognition of an extended DNA binding site by Maf/Nrl family proteins. Oncogene **9:**3149–3158.
18. **Kimura, M., T. Yamamoto, J. Zhang, K. Itoh, M. Kyo, T. Kamiya, H. Aburatani, F. Katsuoka, H. Kurokawa, T. Tanaka, H. Motohashi, and M.** Yamamoto. 2007. Molecular basis distinguishing the DNA binding profile of Nrf2-Maf heterodimer from that of Maf homodimer. J. Biol. Chem. **282:**33681–33690.
19. **Konig, P., and T. J. Richmond.** 1993. The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. J. Mol. Biol. **233:**139–154.
20. **Kraulis, P. J.** 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24:**946–950.
21. **Kusunoki, H., H. Motohashi, F. Katsuoka, A. Morohashi, M. Yamamoto, and T. Tanaka.** 2002. Solution structure of the DNA-binding domain of MafG. Nat. Struct. Biol. **9:**252–256.
22. **Kyo, M., T. Yamamoto, H. Motohashi, T. Kamiya, T. Kuroita, T. Tanaka, J. D. Engel, B. Kawakami, and M. Yamamoto.** 2004. Evaluation of MafG interaction with Maf recognition element arrays by surface plasmon resonance imaging technique. Genes Cells **9:**153–164.
23. **Luscombe, N. M., R. A. Laskowski, and J. M. Thornton.** 1997. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. Nucleic Acids Res. **25:**4940–4945.
24. **Lyon, M. F., R. V. Jamieson, R. Perveen, P. H. Glenister, R. Griffiths, Y. Boyd, L. H. Glimcher, J. Favor, F. L. Munier, and G. C. Black.** 2003. A dominant mutation within the DNA-binding domain of the bZIP transcription factor Maf causes murine cataract and results in selective alteration in DNA binding. Hum. Mol. Genet. **12:**585–594.
25. **McRee, D. E.** 1999. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. **125:**156–165.
26. **Motohashi, H., F. Katsuoka, J. A. Shavit, J. D. Engel, and M. Yamamoto.** 2000. Positive or negative MARE-dependent transcriptional regulation is determined by the abundance of small Maf proteins. Cell **103:**865–875.
27. **Motohashi, H., T. O'Connor, F. Katsuoka, J. D. Engel, and M. Yamamoto.** 2002. Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. Gene **294:**1–12.
28. **Motohashi, H., and M. Yamamoto.** 2004. Nrf2-Keap1 defines a physiologically important stress response mechanism. Trends Mol. Med. **10:**549–557.
29. **Nagai, T., K. Igarashi, J. Akasaka, K. Furuyama, H. Fujita, N. Hayashi, M. Yamamoto, and S. Sassa.** 1998. Regulation of NF-E2 activity in erythroleukemia cell differentiation. J. Biol. Chem. **273:**5358–5365.
30. **Ney, P. A., B. P. Sorrentino, C. H. Lowrey, and A. W. Nienhuis.** 1990. Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. Nucleic Acids Res. **18:**6011–6017.
31. **Nishizawa, M., K. Kataoka, N. Goto, K. T. Fujiwara, and S. Kawai.** 1989. v-maf, a viral oncogene that encodes a "leucine zipper" motif. Proc. Natl. Acad. Sci. USA **86:**7711–7715.
32. **Ogino, H., and K. Yasuda.** 1998. Induction of lens differentiation by activation of a bZIP transcription factor, L-Maf. Science **280:**115–118.
33. **Ohta, T., K. Iijima, M. Miyamoto, I. Nakahara, H. Tanaka, M. Ohtsuji, T. Suzuki, A. Kobayashi, J. Yokota, T. Sakiyama, T. Shibata, M. Yamamoto, and S. Hirohashi.** 2008. Loss of Keap1 function activates Nrf2 and provides advantages for lung cancer cell growth. Cancer Res. **68:**1303–1309.
34. **Otwinowski, A., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276:**307–325.
35. **Padmanabhan, B., K. I. Tong, T. Ohta, Y. Nakamura, M. Scharlock, M. Ohtsuji, M. I. Kang, A. Kobayashi, S. Yokoyama, and M. Yamamoto.** 2006. Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. Mol. Cell **21:**689–700.
36. **Perveen, R., J. Favor, R. V. Jamieson, D. W. Ray, and G. C. Black.** 2007. A heterozygous c-Maf transactivation domain mutation causes congenital cataract and enhances target gene activation. Hum. Mol. Genet. **16:**1030–1038.
37. **Rehemtulla, A., R. Warwar, R. Kumar, X. Ji, D. J. Zack, and A. Swaroop.** 1996. The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. Proc. Natl. Acad. Sci. USA **93:**191–195.
38. **Rupert, P. B., G. W. Daughdrill, B. Bowerman, and B. W. Matthews.** 1998. A new DNA-binding motif in the Skn-1 binding domain-DNA complex. Nat. Struct. Biol. **5:**484–491.
39. **Rushmore, T. H., M. R. Morton, and C. B. Pickett.** 1991. The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. J. Biol. Chem. **266:**11632–11639.
40. **Schneider, T. R., and G. M. Sheldrick.** 2002. Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. **58:**1772–1779.
41. **Swaroop, A., J. Z. Xu, H. Pawar, A. Jackson, C. Skolnick, and N. Agarwal.** 1992. A conserved retina-specific gene encodes a basic motif/leucine zipper domain. Proc. Natl. Acad. Sci. USA **89:**266–270.
42. **Takagi, Y., M. Kobayashi, L. Li, T. Suzuki, K. Nishikawa, and M. Yamamoto.** 2004. MafT, a new member of the small Maf protein family in zebrafish. Biochem. Biophys. Res. Commun. **320:**62–69.
43. **Wang, H., L. P. McIntosh, and B. J. Graves.** 2002. Inhibitory module of Ets-1 allosterically regulates DNA binding through a dipole-facilitated phosphate contact. J. Biol. Chem. **277:**2225–2233.
44. **Yamamoto, T., M. Kyo, T. Kamiya, T. Tanaka, J. D. Engel, H. Motohashi, and M. Yamamoto.** 2006. Predictive base substitution rules that determine the binding and transcriptional specificity of Maf recognition elements. Genes Cells **11:**575–591.