Predictive base substitution rules that determine the binding and transcriptional specificity of Maf recognition elements

Tae Yamamoto[1], Motoki Kyo[2], Terue Kamiya[2], Toshiyuki Tanaka[3], James Douglas Engel[4], Hozumi Motohashi[1,*] and Masayuki Yamamoto[1,5,*]

[1]*Graduate School of Comprehensive Human Sciences and Center for Tsukuba Advanced Research Alliance, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba 305-8577, Japan*

[2]*TOYOBO Co. Ltd. Biotechnology Frontier Project, 10-24 Toyo-Cho, Tsuruga, Fukui 914-0047, Japan*

[3]*Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba 305-8572, Japan*

[4]*Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109-0616, USA*

[5]*ERATO Environmental Response Project, Japan Science and Technology Corporation, 1-1-1 Tennoudai, Tsukuba 305-8577, Japan*

Small Maf transcription factors possess a basic region-leucine zipper motif through which they form homodimers or heterodimers with CNC and Bach proteins. Different combinations of small Maf and CNC/Bach protein dimers bind to *cis*-acting DNA elements, collectively referred to as Maf-recognition elements (MAREs), to either activate or repress transcription. As MAREs defined by function are often divergent from the consensus sequence, we speculated that sequence variations in the MAREs form the basis for selective Maf:Maf or Maf:CNC dimer binding. To test this hypothesis, we analyzed the binding of Maf-containing dimers to variant sequences of the MARE using bacterially expressed MafG and Nrf2 proteins and a surface plasmon resonance-microarray imaging technique. We found that base substitutions in the MAREs actually determined their binding preference for different dimers. In fact, we were able to categorize MAREs into five groups: MafG homodimer-orientd MAREs (Groups I and II), ambivalent MAREs (Group III), MafG:Nrf2 heterodimer-orientd MAREs (Group IV), and silent MAREs (Group V). This study thus manifests that a clear set of rules pertaining to the *cis*-acting element determine whether a given MARE preferentially associates with MafG homodimer or with MafG:Nrf2 heterodimer.

Introduction

Specific interactions between transcription factors and *cis*-acting DNA sequence motifs generate the molecular basis of transcriptional regulation in which diversity is created by the particular pool of transcription factor dimers available for binding. To elaborate, the basic region-leucine zipper (bZip) superfamily of transcription factors bind *cis* DNA motifs through the formation of heterodimers or homodimers comprised of various combinations of different family members (Newman & Keating 2003). In this manner, the bZip superfamily factors establish a functional network, with different combinations of dimeric transcription factors provoking a wide range of biological responses (Jochum *et al*. 2001; Motohashi *et al*. 2002). This suggests that the relative abundance of different partner molecules is critical for the regulation of specific subsets of target genes.

To elucidate how such diversity is created in a transcriptional regulatory system, the DNA binding specificity of various transcription factor dimers has been examined. In most prior investigations, transcription factors (and various mutant molecules) have been studied with one or a few fixed *cis*-regulatory elements derived from target genes in electrophoretic mobility shift assays (EMSA) (Chapman-Smith *et al*. 2004) or surface plasmon resonance (SPR) assays (Grinberg & Kerppola 2003). Consensus DNA sequences for the preferred binding of each transcription factor dimer have also been determined using PCR-EMSA amplification and purification methods. However, this 'binding site selection' method collects a relatively diverse group of DNA sequences that share the ability to interact with specific transcription factors and usually only those

Communicated by: Shunsuke Ishii

**Correspondence*: E-mail: masi@tara.tsukuba.ac.jp, hozumin@md.tsukuba.ac.jp

DOI: 10.1111/j.1365-2443.2006.00965.x

sequences having the highest affinity are recovered (Kataoka *et al.* 1994; Kerppola & Curran 1994; Johnsen *et al.* 1998). Whereas extensive efforts have been made to extract a feature common to the diverse *cis*-regulatory elements, very few attempts have been made to examine whether the diversity in those elements contributes to the selection of transcription factors that interact with them. Thus, our current view is biased in this respect and the genuine relationship between sequence changes in *cis*-acting elements and gene regulation *in vivo* remains obscure.

Members of the Maf family belong to a group of transcription factors that share a unique bZip structure (reviewed in Motohashi *et al.* 2002). This family consists of four large Maf proteins, c-Maf, MafB, NRL and L-Maf/A-Maf, that possess N-terminal transactivation domains, and three small Maf proteins, MafF/MafT, MafG and MafK, that lack activation domains. Both large and small Maf proteins can form homodimers. In addition, the small Maf proteins can heterodimerize with Cap'n'collar (CNC) family bZip proteins (p45 NF-E2, Nrf1/LCR-F1, Nrf2/ECH and Nrf3) and also with the closely related Bach family proteins (Bach1 and Bach2).

Recognition sequences for small Maf:CNC and small Maf:Bach heterodimers are present in the enhancer and promoter regions of cellular target genes. These critical *cis*-regulatory elements were identified as the NF-E2 binding motif and the anti-oxidant/electrophile responsive element (ARE/EpRE) (Friling *et al.* 1990; Ney *et al.* 1990; Rushmore *et al.* 1991) through the analyses of erythroid-specific genes and cellular defense genes, respectively. The consensus sequence for NF-E2 was reported to be TGCTGAGTCAT (Ney *et al.* 1990), while that for the ARE is TGACnnnGC (Rushmore *et al.* 1991). NF-E2 (small Maf:p45) was shown to interact with the NF-E2 binding motif, while small Maf:Nrf2 heterodimers and small Maf homodimers interact with the ARE/EpRE (Motohashi *et al.* 2002).

A consensus DNA sequence for the binding of Maf homodimers has been determined by binding site selection (Kataoka *et al.* 1994; Kerppola & Curran 1994). The conclusions stated that large Maf homodimers recognize a relatively long palindromic DNA sequence designated the Maf recognition element (MARE), TGCTGA$^{G}/_{C}$TCAGCA and TGCTGA$^{GC}/_{CG}$TCAGCA. These two sequences contain either a TPA-responsive element (TRE; TGA$^{G}/_{C}$TCA) or a cAMP-responsive element (CRE; TGA$^{GC}/_{CG}$TCA) at their core. In addition, these two MARE motifs also contain extended elements on both sides of the core sequence (flanking region; 5′-TGC-core-GCA-3′). MARE motifs containing a TRE or CRE are referred to as T-MARE and C-MARE, respectively. The requirement for a flanking region in the MARE distinguishes the Maf family proteins from members of the AP-1 or CREB/ATF family (Swaroop *et al.* 1992; Kerppola & Curran 1994; Kusunoki *et al.* 2002). Because the NF-E2 binding motif and ARE share a strong similarity to the MARE consensus sequence and because small Maf:CNC and small Maf:Bach heterodimers interact with the MARE, the collective term for elements recognizing Maf-containing homodimers and heterodimers is MAREs (Motohashi *et al.* 2002).

Recent studies suggested that differences in sequence among the MAREs may be functionally important. For example, gene disruption of various MARE-interacting transcription factors in mouse affects the transcription of different subsets of MARE-dependent genes, resulting in distinct phenotypes (Motohashi *et al.* 2002). One reason for the diverse phenotypes encountered in these mutants may be that each transcription factor has a unique expression profile, while another may be that coexisting bZip heterodimers differentially interact with different MAREs. We envisaged that this differential interaction, if genuine, might be attained by either the cooperative binding of bZip factors to the MAREs (Nioi *et al.* 2003; Ramirez-Carrozzi & Kerppola 2003) or sequence variations within the MAREs. Since the latter possibility was most immediately tractable, we examined how sequence variation in the MAREs might affect MARE-dependent transcriptional activity.

We recently described and validated an SPR-microarray imaging technique that detects interactions between transcription factors and double-stranded DNA efficiently, quantitatively and systematically (Kyo *et al.* 2004). SPR is a label-free, real time, optical detection method for interactions between a soluble ligate and an immobilized ligand. We combined the SPR imaging system with the microarray technique. DNA microarray is assembled by immobilizing double-stranded DNAs at up to 96 spots on the gold surface and, with a protein in solution injected across the surface, the protein-DNA interactions at all spots are detected simultaneously. Utilizing this technique, we examined the binding affinity of homodimeric MafG, one of the small Maf family members, to several MARE-related DNA sequences. Since the binding affinity of the MafG homodimer to various MAREs correlated well with those determined by EMSA, we applied the SPR-microarray imaging method to binding analyses of Maf-containing homodimers and heterodimers. We discovered that each MARE sequence specifies the particular dimeric transcription factor complex with which it effectively interacts. Furthermore, we found that the MAREs can be classified into five distinct groups based on the kinetic values of binding to the MafG homodimer or MafG:Nrf2 heterodimer.

This categorization revealed a straightforward set of base alteration rules for MARE elements that actively determine the binding specificity for dimeric transcription factor complexes and govern the diversity of responses mediated by MAREs in the regulation of gene expression.

Results

SPR-microarray analysis of MafG homodimer binding to consensus and mutant MAREs

We first examined the binding profile of a MafG homodimer to a library of 40 sequence variants of MAREs using SPR-microarray imaging. The DNA microarrays were assembled using double-stranded oligonucleotides containing MARE variants generated by systematic base alterations (Fig. 1A and Table 1). DNA preparations with symmetric base alterations in the core and flanking regions were designated CO-S (core-symmetric mutation) and FL-S (flanking-symmetric mutation), respectively. DNAs with base alterations on one side of the MARE were called CO-A (core-asymmetric mutation) and FL-A (flanking-asymmetric mutation). DNAs with a central base alteration were referred to as CEN-G, CEN-A and CEN-T. CEN-C and CEN-G correspond to the MARE consensus sequence. MafG protein was expressed in bacteria and purified, then examined directly since MafG is capable of binding to DNA as a homodimer.

Representative SPR signals and the kinetic values calculated from them are shown in Fig. 1 and Table 2, respectively. An SPR signal value indicates the amount of MafG protein binding to the DNA. Lines tangent to the points of association and dissociation of an SPR signal curve reflect the association rate and dissociation rate, respectively. Although the binding affinity to the variant MARE motifs differs greatly, the curves obtained with the MafG homodimer preparation were essentially monotonous in shape (Fig. 1B,C). We therefore divided the variants into three groups: MAREs with high binding affinity (Group H; this group was subsequently subdivided into three, based on the affinities to the MafG:Nrf2 heterodimer (below)), low binding affinity (Group L) and without demonstrable binding (Group N). The highest affinities were obtained with the complete

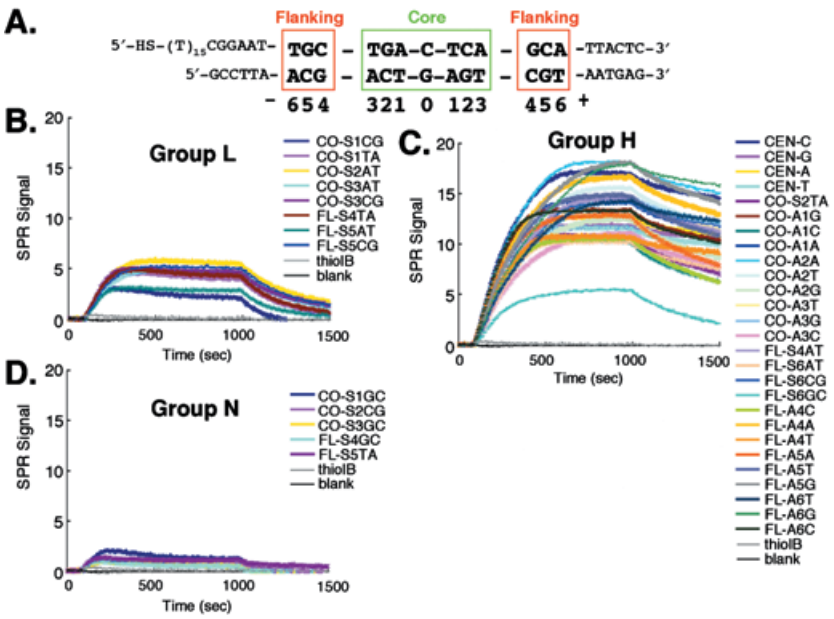

Figure 1 The design of oligonucleotides for SPR-microarray imaging and SPR binding curves generated by association with the MafG homodimer. (A) The consensus sequence of the T-MARE is shown in bold letters and the flanking and core region elements are boxed in red and green, respectively. The positions of the nucleotides are indicated by the numbers beneath the MARE. (B–D) The SPR binding curves of the MafG homodimer. Buffer containing MafG started to flow at 0 s and the flow was terminated at 1000 s. The buffer was replaced with one without MafG and flow continued for a further 500 s. Representative results obtained from more than three independent experiments are shown. The MARE variants with different binding affinities were categorized into three groups: Group L, low affinity (B), Group H, high affinity (C) and Group N, no affinity (D). Designations of the oligonucleotides are shown on the right of each graph and their sequences and calculated kinetic data are described in Tables 1 and 2, respectively. Blank and thiolB indicate SPR signals obtained from a spot without test DNAs or with nonspecific oligonucleotides, respectively.

Table 1 MARE variants with systematic base alterations in the core and flanking regions

	Name	MARE-related sequence with base alterations in the core region	Name	MARE-related sequence with base alterations in the flanking region
MARE consensus	CEN-C	TGCTGACTCAGCA		
	CEN-G	TGCTGAGTCAGCA		
Center variations	CEN-A	TGCTGAaTCAGCA		
	CEN-T	TGCTGAtTCAGCA		
Symmetric variations	CO-S1CG	TGCTGcCgCAGCA	FL-S4GC	TGgTGACTCAcCA
	CO-S1GC	TGCTGgCcCAGCA	FL-S4TA	TGtTGACTCAaCA
	CO-S1TA	TGCTGtCaCAGCA	FL-S4AT	TGaTGACTCAtCA
	CO-S2TA	TGCTtACTaAGCA	FL-S5TA	TtCTGACTCAGaA
	CO-S2AT	TGCTaACTtAGCA	FL-S5AT	TaCTGACTCAGtA
	CO-S2CG	TGCTcACTgAGCA	FL-S5CG	TcCTGACTCAGgA
	CO-S3AT	TGCaGACTCtGCA	FL-S6AT	aGCTGACTCAGCt
	CO-S3CG	TGCcGACTCgGCA	FL-S6CG	cGCTGACTCAGCg
	CO-S3GC	TGCgGACTCcGCA	FL-S6GC	gGCTGACTCAGCc
Asymmetric variations	CO-A1G	TGCTGACgCAGCA	FL-A4C	TGCTGACTCAcCA
	CO-A1C	TGCTGACcCAGCA	FL-A4A	TGCTGACTCAaCA
	CO-A1A	TGCTGACaCAGCA	FL-A4T	TGCTGACTCAtCA
	CO-A2A	TGCTGACTaAGCA	FL-A5A	TGCTGACTCAGaA
	CO-A2T	TGCTGACTtAGCA	FL-A5T	TGCTGACTCAGtA
	CO-A2G	TGCTGACTgAGCA	FL-A5G	TGCTGACTCAGgA
	CO-A3T	TGCTGACTCtGCA	FL-A6T	TGCTGACTCAGCt
	CO-A3G	TGCTGACTCgGCA	FL-A6G	TGCTGACTCAGCg
	CO-A3C	TGCTGACTCcGCA	FL-A6C	TGCTGACTCAGCc

Bases altering from the consensus MARE sequence are shown in lower-case letters. CEN-C and CEN-G indicate the consensus MARE sequences and CEN-A and CEN-T are central base variations. DNAs with symmetric base alterations in the core and flanking regions were designated CO-S and FL-S, respectively. DNAs with base alterations on one side of the MARE were called CO-A (core-asymmetric mutation) and FL-A (flanking-asymmetric mutation).

consensus sequences CEN-C and CEN-G. We found that, in general, symmetric base alterations weakened or abolished binding of the MafG homodimer and thus these mutations were placed into the low-affinity group (Group L) or non-binding group (Group N), respectively. Exceptions were mutants with alterations at position −6/+6 (FL-S6AT, FL-S6CG and FL-S6GC), CO-S2TA and FL-S4AT. In contrast, asymmetric base alterations either in the core region or in the flanking region did not inhibit binding (Group H), even though the corresponding symmetric mutations were completely incompatible with MafG association. Thus, the asymmetric mutants CO-A1C, CO-A2G, CO-A3C, FL-A4C and FL-A5A were categorized into Group H, while their corresponding symmetric mutants CO-S1GC, CO-S2CG, CO-S3GC, FL-S4GC and FL-S5TA were categorized into Group (N), indicating that the conservation of a 'one half site' is sufficient to stabilize MafG homodimer binding.

Mixtures of Nrf2 and MafG generate complex and heterogeneous SPR patterns

To prepare heterodimeric molecules containing MafG, we chose Nrf2 as the representative CNC partner molecule of MafG. A fragment of Nrf2 containing the CNC and bZip domains was bacterially expressed and purified. Equimolar Nrf2 was mixed with MafG and applied to the same DNA microarrays as examined in the MafG homodimer analysis. Surprisingly, the binding curves obtained with this preparation were quite heterogeneous in shape (left panels of Fig. 2) compared to those recovered in examining the MafG homodimer (Fig. 1). Three curve patterns were initially categorized. The first pattern was one of slow association, slow saturation to a plateau and slow dissociation (parabolic curve; Fig. 2A, left panel). The second pattern displayed rapid association and rapid dissociation (square curve; Fig. 2D, left panel). The third pattern showed the composite characteristics

Table 2 SPR kinetic values of the MafG homodimer and classification of MARE-related sequences based on binding profiles

Name	MARE-related sequence with base alterations −6 −4 −2 0 +2 +4 +6	SPR kinetics of homodimer k_a [10^4/M/s]	k_d [10^{-4}/s]	K_D [10^{-8} M]	Group
CEN-C	TGCTGACTCAGCA	7.71 ± 3.25	3.83 ± 2.29	0.46 ± 1.28	
CEN-G	TGCTGAGTCAGCA	7.24 ± 0.69	4.80 ± 0.67	0.61 ± 0.65	
CO-S1CG	TGCTGcCgCAGCA	12.64 ± 3.47	43.95 ± 10.45	3.44 ± 12.3	
CO-S1TA	TGCTGtCaCAGCA	12.09 ± 2.81	33.84 ± 7.51	2.70 ± 7.62	
CO-S2AT	TGCTaACTtAGCA	10.36 ± 0.4	23.01 ± 3.37	2.07 ± 3.44	
CO-S3AT	TGCaGACTCtGCA	10.05 ± 1.14	30.66 ± 8.31	2.80 ± 5.99	L
CO-S3CG	TGCcGACTCgGCA	10.15 ± 1.18	25.43 ± 3.4	2.36 ± 3.89	
FL-S4TA	TGtTGACTCAaCA	12.57 ± 1.63	50.54 ± 27.17	4.06 ± 27.54	
FL-S5AT	TaCTGACTCAGtA	11.27 ± 3.38	72.95 ± 37.38	7.54 ± 53.2	
FL-S5CG	TcCTGACTCAGgA	12.70 ± 2.03	44.22 ± 22.46	3.15 ± 12.18	
CO-S2TA	TGCTtACTaAGCA	9.98 ± 0.57	10.52 ± 1.98	0.98 ± 1.62	
CO-A1G	TGCTGACgCAGCA	10.09 ± 2.21	9.12 ± 1.75	0.86 ± 2.05	
CO-A1C	TGCTGACcCAGCA	10.16 ± 2.36	14.84 ± 2.71	1.44 ± 4.42	
CO-A1A	TGCTGACaCAGCA	9.51 ± 2.53	7.84 ± 0.64	0.83 ± 2.52	
CO-A2T	TGCTGACTtAGCA	11.03 ± 1.79	10.01 ± 2.72	0.88 ± 3.14	
CO-A2G	TGCTGACTgAGCA	10.24 ± 2.75	14.22 ± 2.38	1.38 ± 4.35	
CO-A3T	TGCTGACTCtGCA	9.75 ± 2.18	9.75 ± 2.03	0.97 ± 2.68	
CO-A3G	TGCTGACTCgGCA	9.65 ± 1.59	9.04 ± 1.91	0.89 ± 2.3	
CO-A3C	TGCTGACTCcGCA	9.73 ± 1.75	13.61 ± 3.49	1.33 ± 4.22	
FL-S6AT	aGCTGACTCAGCt	9.03 ± 1.52	9.43 ± 3.22	0.96 ± 2.42	
FL-S6GC	gGCTGACTCAGCc	10.15 ± 1.63	29.91 ± 11.27	2.81 ± 9.73	H
CEN-A	TGCTGAaTCAGCA	8.19 ± 0.95	5.16 ± 0.42	0.59 ± 0.59	
CEN-T	TGCTGAtTCAGCA	8.23 ± 1.51	5.87 ± 0.38	0.68 ± 1.12	
CO-A2A	TGCTGACTaAGCA	9.28 ± 2.71	6.11 ± 0.89	0.67 ± 2.13	
FL-S6CG	cGCTGACTCAGCg	8.56 ± 1.36	6.37 ± 2.46	0.69 ± 2.32	
FL-A4A	TGCTGACTCAaCA	9.41 ± 0.4	9.97 ± 0.55	0.99 ± 0.92	
FL-A5A	TGCTGACTCAGaA	9.24 ± 0.17	13.63 ± 0.28	1.37 ± 0.51	
FL-A5T	TGCTGACTCAGtA	7.46 ± 0.14	10.09 ± 0.36	1.25 ± 0.45	
FL-A5G	TGCTGACTCAGgA	6.37 ± 0.12	7.92 ± 0.11	1.15 ± 0.33	
FL-A6T	TGCTGACTCAGCt	6.00 ± 0.16	5.75 ± 0.65	0.89 ± 1.15	
FL-A6G	TGCTGACTCAGCg	5.93 ± 0.15	5.12 ± 0.52	0.80 ± 0.94	
FL-A6C	TGCTGACTCAGCc	7.47 ± 0.33	9.01 ± 0.19	1.12 ± 0.3	
FL-S4AT	TGaTGACTCAtCA	11.46 ± 1.19	12.27 ± 4.27	1.02 ± 3.88	
FL-A4C	TGCTGACTCAcCA	11.58 ± 1.13	18.17 ± 1.2	1.48 ± 2.45	
FL-A4T	TGCTGACTCAtCA	7.80 ± 0.14	5.84 ± 0.47	0.69 ± 0.64	
mG2cryst	TGCcaACaCAGCA	9.95 ± 2.35	27.45 ± 10.27	2.74 ± 12.9	L
hOPSIN	TGCTGAtTCAGCc	10.02 ± 1.77	16.60 ± 7.91	1.50 ± 4.89	H
hNQO1	TGCTGAGTCActg	12.31 ± 4.11	41.69 ± 10.84	3.68 ± 22.55	H

Altered bases are shown in lower-case letters. Group N sequences were also examined, but excluded from here because of low SPR signals. Abbreviations: k_a, association rate constant; k_d, dissociation rate constant; K_D (= k_d/k_a), dissociation constant.

rapid association, slow saturation to a plateau and relatively slow dissociation (Fig. 2B,C, left panels).

We suspected that a mixture of MafG homodimers and MafG:Nrf2 heterodimers in the preparation might be responsible for these heterogeneous binding patterns. This possibility was addressed experimentally by altering the relative molar ratio of Nrf2 to MafG (Fig. 3). MARE-related sequences with base changes in the core regions (Fig. 3A) and the flanking regions (Fig. 3B) were used. In the absence of Nrf2, all of the binding curves were parabolic. By increasing the relative abundance of Nrf2, square curves began to dominate and all the observed curves were square when a ten-fold molar excess of Nrf2 was mixed with MafG (Fig. 3A,B and see right panels

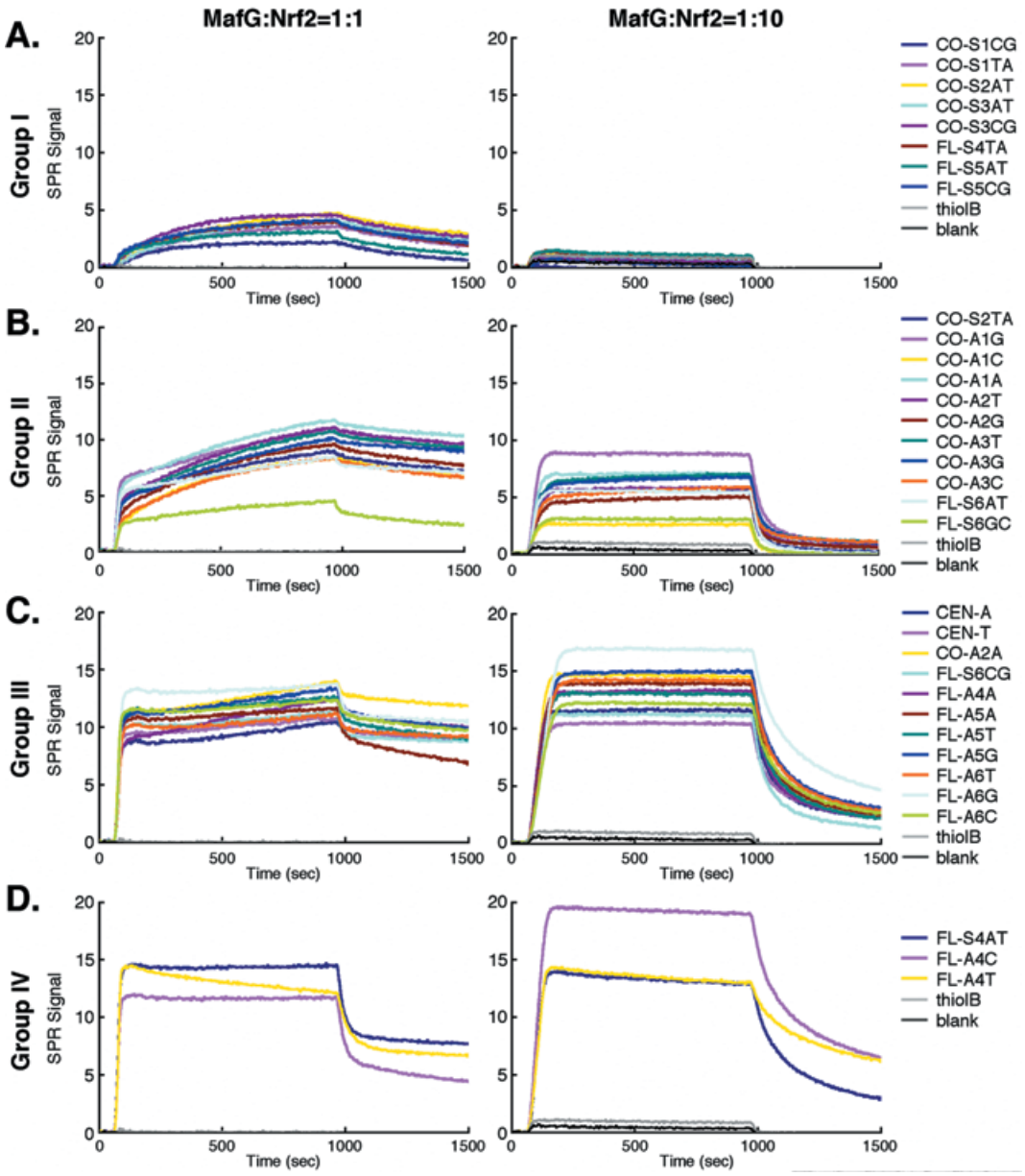

Figure 2 Heterogeneous SPR curves of MafG:Nrf2 heterodimer binding and classification of the signal patterns. Buffer containing a 1 : 1 molar ratio (left panels) or a 1 : 10 molar ratio (right panels) of MafG to Nrf2 started to flow at 0 s and the flow was terminated at 1000 s. Buffer without MafG:Nrf2 was then allowed to flow for a following 500 s. Representative results obtained from more than three independent experiments are shown. Signal curves defining four groups, (A) Group I, (B) Group II, (C) Group III and (D) Group IV, are shown. The sequence designations of each group are shown on the right hand side of the corresponding graphs. The kinetic data calculated from SPR signals obtained from using a 1 : 10 ratio mixture are described in Table 3.

of Fig. 2). When only Nrf2 was applied, no binding was observed to any of the MARE variants (data not shown). From these data, we provisionally concluded that the parabolic curves represent the interaction of MafG homodimers with the MAREs, whereas MafG:Nrf2 heterodimers account for the square curves.

MAREs can be categorized into five distinct groups

When we compared the SPR curves obtained from using equimolar mixtures of MafG and Nrf2 to those obtained from using a 1 : 10 molar ratio of MafG to Nrf2, we noticed that the MAREs fell into five groups (Fig. 2

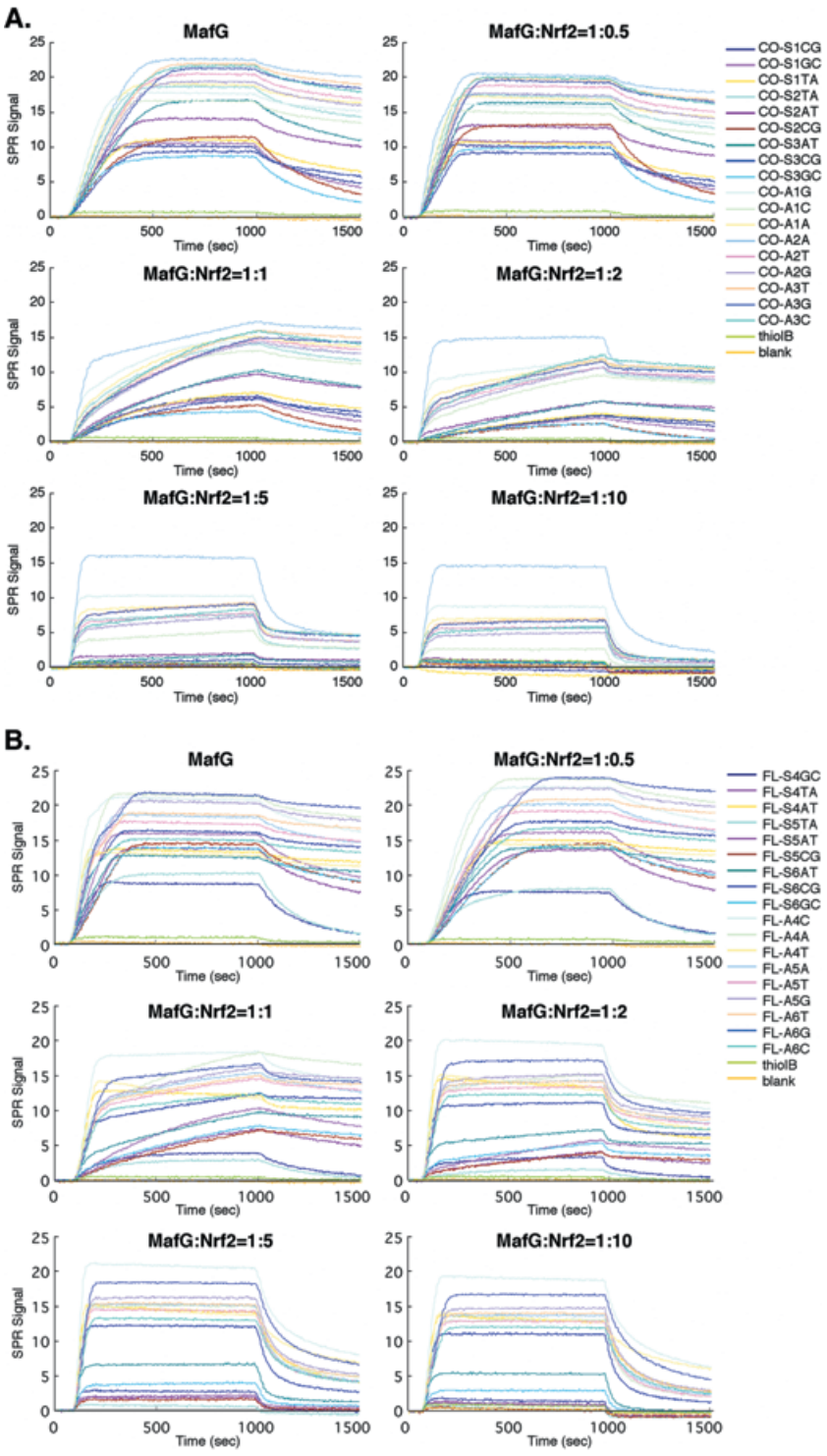

Figure 3 Titration study of SPR signal curves of 40 MARE variants with systematic base alteration. MARE-related sequences with base changes in the core and flanking regions were used in (A) and (B), respectively. Six protein preparations containing different ratios of Nrf2 to MafG were used. The flow of buffer containing MafG or various molar ratios of Nrf2 and MafG commenced at 0 s and was stopped at 1000 s. This was replaced with buffer containing no MafG:Nrf2 and the flow continued for a following 500 s. In the absence of Nrf2, all of the binding curves were parabolic. By increasing the relative abundance of Nrf2, square curves began to dominate and all the observed curves displayed this squared shape feature when a ten-fold molar excess of Nrf2 was mixed with MafG. Representative results obtained from more than three independent experiments are shown. Designations of the oligonucleotides are shown on the right hand side of the graphs. Blank and thiolB indicate SPR signals obtained from a spot without test DNAs or with nonspecific oligonucleotides, respectively.

and data not shown). These groupings correlated well with those determined for MafG homodimer binding (Fig. 1). Group I displayed parabolic curves when MafG and Nrf2 were equimolar and no substantial interaction occurred when Nrf2 was ten-fold in excess of MafG. This was interpreted as exclusively homodimeric binding by these sequences. The kinetic values of Group I sequences for homodimer binding were relatively low (Table 2) and this group corresponds to Group L of MafG homodimer binding. Group II displayed curves with composite features;

close to parabolic binding at ratios of 1 : 1 and displaying square curves with low plateaus at 1 : 10 ratios. Group III had curves with composite features that were close to being square-shaped under 1 : 1 conditions and square with high plateau levels under 1 : 10 conditions. We surmise that Group II and III sequences interact with both the MafG homodimer and the MafG:Nrf2 heterodimer depending on their relative concentrations. Group IV displayed square interaction profiles even at 1 : 1 molar ratios. This suggests that the MafG:Nrf2 heterodimer interacts with these sequences almost exclusively by preventing MafG homodimer binding. Thus, Group IV sequences strongly prefer the heterodimer. MAREs categorized into Groups II–IV correspond to Group H of MafG homodimer binding (Table 2). Group V displayed no protein binding under either condition and the binding sequences in this group correspond to Group N (Table 2).

Kinetic analysis of MafG:Nrf2 heterodimer binding

Kinetic values for heterodimer binding were calculated using protein preparations containing MafG and Nrf2 at a 1 : 10 ratio, since all MafG molecules heterodimerize with Nrf2 under this condition (Table 3). Base alterations categorized into Groups I and V (i.e. Groups L and N in Table 2), which either diminish or abolish MafG

Table 3 SPR kinetic values of the MafG:Nrf2 heterodimer and classification of MARE-related sequences based on binding profiles

	MARE-related sequence with base alterations	SPR kinetics of heterodimer			
Name	−6 −4 −2 0 +2 +4 +6	k_a [10^4/M/s]	k_d [10^{-4}/s]	K_D [10^{-8} M]	Group
CEN-C	TGCTGACTCAGCA	2.43 ± 0.84	7.05 ± 0.57	3.23 ± 1.04	
CEN-G	TGCTGAGTCAGCA	2.60 ± 0.32	5.04 ± 0.24	2.21 ± 0.35	
CO-S2TA	TGCTtACTaAGCA	2.98 ± 0.54	14.16 ± 2.36	5.39 ± 0.85	
CO-A1G	TGCTGACgCAGCA	2.68 ± 0.84	10.82 ± 1.14	4.93 ± 1.48	
CO-A1C	TGCTGACcCAGCA	2.31 ± 1.37	11.29 ± 2.89	6.69 ± 2.65	
CO-A1A	TGCTGACaCAGCA	2.45 ± 0.74	11.00 ± 1.88	5.34 ± 1.29	
CO-A2T	TGCTGACTtAGCA	2.04 ± 0.63	10.51 ± 2.41	6.11 ± 1.40	
CO-A2G	TGCTGACTgAGCA	1.74 ± 0.52	10.60 ± 2.75	7.10 ± 1.49	II
CO-A3T	TGCTGACTCtGCA	1.73 ± 0.55	10.49 ± 2.22	7.32 ± 1.92	
CO-A3G	TGCTGACTCgGCA	1.59 ± 0.52	9.52 ± 2.35	7.39 ± 2.53	
CO-A3C	TGCTGACTCcGCA	1.75 ± 0.99	11.78 ± 2.53	9.61 ± 4.53	
FL-S6AT	aGCTGACTCAGCt	2.63 ± 0.61	15.73 ± 3.70	7.12 ± 2.59	
FL-S6GC	gGCTGACTCAGCc	2.09 ± 1.06	19.25 ± 6.12	13.63 ± 8.80	
CEN-A	TGCTGAaTCAGCA	2.78 ± 0.62	9.59 ± 0.59	4.02 ± 0.74	
CEN-T	TGCTGAtTCAGCA	2.23 ± 0.28	7.25 ± 0.70	3.74 ± 0.85	
CO-A2A	TGCTGACTaAGCA	2.26 ± 0.33	8.12 ± 0.67	4.10 ± 0.66	
FL-S6CG	cGCTGACTCAGCg	2.40 ± 0.27	9.39 ± 1.65	4.49 ± 1.18	
FL-A4A	TGCTGACTCAaCA	2.14 ± 0.18	8.21 ± 1.58	4.37 ± 1.15	
FL-A5A	TGCTGACTCAGaA	2.01 ± 0.09	7.43 ± 1.11	4.16 ± 0.80	III
FL-A5T	TGCTGACTCAGtA	1.91 ± 0.08	6.65 ± 0.80	3.92 ± 0.64	
FL-A5G	TGCTGACTCAGgA	1.62 ± 0.12	6.12 ± 0.71	4.26 ± 0.82	
FL-A6T	TGCTGACTCAGCt	1.58 ± 0.13	6.44 ± 0.75	4.62 ± 0.95	
FL-A6G	TGCTGACTCAGCg	1.52 ± 0.15	4.93 ± 0.39	3.70 ± 0.69	
FL-A6C	TGCTGACTCAGCc	1.80 ± 0.19	6.26 ± 0.78	3.98 ± 0.93	
FL-S4AT	TGaTGACTCAtCA	2.84 ± 0.07	6.24 ± 0.56	2.46 ± 0.28	
FL-A4C	TGCTGACTCAcCA	2.56 ± 0.32	5.47 ± 0.51	2.43 ± 0.36	IV
FL-A4T	TGCTGACTCAtCA	3.24 ± 0.12	3.00 ± 0.19	1.03 ± 0.06	
MARE23	caaTGACTCAttg	2.42 ± 0.65	21.84 ± 7.43	11.02 ± 5.66	IV
hOPSIN	TGCTGAtTCAGCc	2.46 ± 0.51	13.90 ± 1.74	6.61 ± 1.67	III
hNQO1	TGCTGAGTCActg	2.63 ± 0.23	9.56 ± 0.61	4.10 ± 0.49	IV
hBglHS4	gGCTGACTCActc	2.23 ± 0.37	15.28 ± 1.84	7.97 ± 1.99	IV
mGSTY	TGCTttGTCAcCA	2.77 ± 0.97	21.98 ± 10.60	8.88 ± 3.61	IV

Altered bases are shown in lower-case letters. Base changes conferring features of homodimer preference are underlined and those conferring heterodimer preference are double-underlined. Type I and V MARE sequences were also examined, but the results were excluded from this table because of low SPR signals. Abbreviations: k_a, association rate constant; k_d, dissociation rate constant; K_D (= k_d/k_a), dissociation constant.

homodimer binding, also failed to bind to MafG:Nrf2 heterodimers. Substantial heterodimer binding was detected for the sequences bearing high affinity for the MafG homodimer (i.e. Groups II–IV in Table 3 that correspond to Group H in Table 2). In general, both the ka (association time constant) and kd (dissociation time constant) were larger for MafG:Nrf2 heterodimers than for the MafG homodimer (compare Tables 2 and 3). It should be noted that the differences in kinetic values for the homodimer and heterodimer also reflect the differential patterns of the observed SPR signals: larger ka and kd values indicate rapid association and dissociation, which generate the square curves, while smaller ka and kd values indicate slow association and dissociation, which generate the parabolic curves.

In general, MARE sequences within Groups II and IV had relatively larger and smaller K_D values, respectively, while Group III K_D values were intermediate. These results are consistent with the patterns of SPR curves obtained with equimolar mixtures of Nrf2 and MafG (left panels of Fig. 2B–D). At the 1 : 1 ratio, the square signal curves of the Group IV sequences suggested the exclusive binding of these sequences to heterodimers, since their affinities for the MafG:Nrf2 heterodimer were high. The signal curves of Group II sequences were close to parabolic, suggesting the preferential interaction of homodimers to those MAREs, since their affinities for the MafG:Nrf2 heterodimer were low.

The SPR-based classification of MAREs is compatible with that of EMSA

To validate the accuracy of the SPR-based MARE classifications, we performed EMSA using three different protein preparations and representative sequences chosen from each group of classification as probes. We first examined a protein mixture consisting of a 1 : 1 molar ratio of MafG to Nrf2 that generated the strongest heterodimer binding to CEN-C, the MARE consensus sequence (Fig. 4, top panel, lanes 1–26). Group I MAREs showed weak interactions (lanes

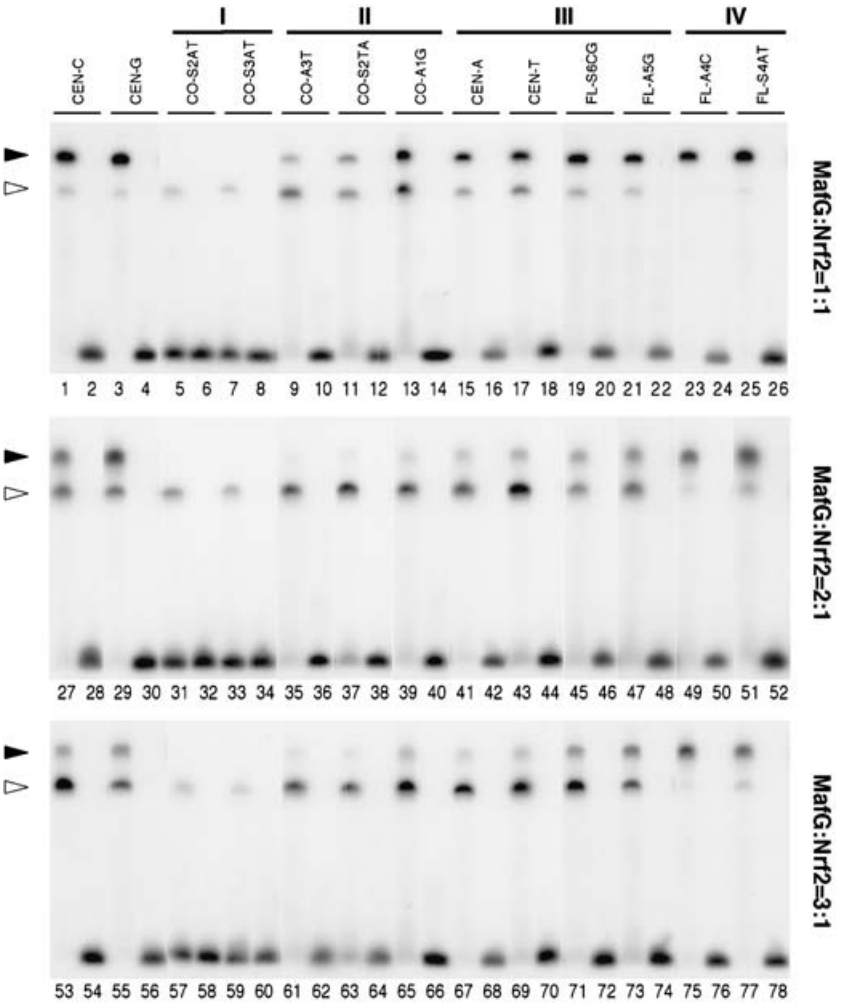

Figure 4 EMSA of recombinant proteins with representative symmetric and asymmetric 13 MARE-related sequences as probes. Three protein preparations containing different ratios of Nrf2 to MafG were used. The homodimer and heterodimer binding activities to each probe were examined under heterodimer-enriched conditions (top panel, lanes 1–26), homodimer-enriched conditions (bottom panel, lanes 53–78) and intermediate conditions (middle panel, lanes 27–52). Black and white arrowheads indicate heterodimer and homodimer binding, respectively. The designations of oligonucleotides used as probes are indicated above the lanes of the top panel. The group number of each MARE variant is indicated above the sequence name. The odd-numbered lanes contain recombinant protein mixture and the even-numbered lanes contain probes only.

5–8) and association exclusively with the homodimer (white arrowheads), but no binding by the heterodimer (black arrowheads). Group II and III sequences interacted with both homodimers and heterodimers (lanes 9–22), but Group II showed a preference for the homodimer while Group III favored the heterodimer. Group IV MAREs interacted solely with MafG:Nrf2 heterodimers (lanes 23–26).

We next used a protein mixture at a 2 : 1 molar ratio of MafG to Nrf2 that generated comparable CEN-C binding activities between the homodimer and heterodimer (Fig. 4, middle panel, lanes 27–52). Only homodimer binding was observed for Group I and II sequences (lanes 31–40). Homodimer binding increased with Group III MAREs (lanes 41–48), whereas heterodimer binding was still dominant in the case of Group IV MAREs (lanes 49–52). The molar ratio of MafG to Nrf2 in the protein mixture was adjusted to 3 : 1 and generated more abundant homodimer binding to CEN-C (Fig. 4, bottom panel, lanes 53–78). Using the 3 : 1 ratio, homodimer binding further increased with Group III MAREs (lanes 67–74). While the binding observed for Groups I and II was predominantly homodimeric (lanes 57–66), heterodimers dominated in Group IV binding (lanes 75–78). In perfect accordance with the SPR-based classification of the MAREs, these EMSA results demonstrated that, regardless of changes in the relative abundance of homodimers and heterodimers, Group I and Group II MAREs interact preferentially with homodimers, whereas Group IV MAREs interact more favorably with heterodimers. Group III MAREs appear permissive for roughly equivalent binding to both types of dimer.

Single base alterations determine the binding preference of MAREs

The base substitutions that led to the categorization of individual MAREs into five distinct groups are summarized in Fig. 5. This figure lists the mutations that were identified and categorized based on the results of SPR analyses and EMSA. All Group I sequences are symmetric mutants, while Group II includes most of the core asymmetric mutants and three symmetric mutants as exceptions (CO-S2TA, FL-S6AT and FL-S6GC). Group III includes the central base mutants and most of the flanking region asymmetric mutants. Group IV includes one symmetric mutant with a C/G to A/T change at position −4/+4 (FL-S4AT) and two asymmetric mutations at position +4, namely G to C (FL-A4C) and G to T (FL-A4T). Although Group IV sequences were mainly bound by the heterodimer (Figs 2D and 4), the homodimer also strongly interacted with these sequences

		Center	Core region			Flanking region		
Nucleotide position		0	+1	+2	+3	+4	+5	+6
Consensus of MARE		C/G	T	C	A	G	C	A
Group I	Symmetric		g/a	t	t/g	a	t/g	
	Asymmetric							
Group II	Symmetric			a				t/c
	Asymmetric		g/c/a	t/g	t/g/c			
Group III	Symmetric	a/t						g
	Asymmetric			a		a	a/t/g	t/g/c
Group IV	Symmetric					t		
	Asymmetric					c/t		
Group V	Symmetric		c	g	c	c	a	
	Asymmetric							

Figure 5 Summary of base changes that determine the binding preference for the MafG homodimer and MafG:Nrf2 heterodimer. Altered bases are listed according to the nucleotide positions and the group of MARE variants possessing the altered base. Symmetric 13-bp MARE variants are shaded in gray. Base changes conferring features of homodimer preference are in red and those of heterodimer preference are in blue. Base changes maintaining a preference for both homodimer and heterodimer are shown in black, whereas base changes that completely inhibit the binding of both types of dimer are in green.

in the absence of the heterodimer (Table 2 and EMSA data not shown). All the MAREs incapable of interacting with either the MafG homodimer or MafG:Nrf2 heterodimer and classified into Group V correspond to symmetric mutants.

In summary, the features of the MARE variants can be summarized as follows: homodimer-orientd MAREs (Groups I and II), heterodimer-orientd MAREs (Group IV), ambivalent MAREs (Group III) and silent/functionless MAREs (Group V). Thus, these results unequivocally demonstrate that single base changes can modulate the MafG homodimer vs. MafG:Nrf2 heterodimer binding preference of MAREs.

Binding preferences of endogenous MAREs

To verify the binding preferences of the MAREs proposed here, we examined a group of genomic MARE-related sequences to determine whether or not they obeyed the rules described above. To this end, we selected five well-characterized MARE sequences present in endogenous gene regulatory regions (genes abbreviated in bold in Table 4) and examined their properties by both SPR imaging and EMSA. MAREs were selected from the mouse γF crystallin (mG2cryst; Rajaram & Kerppola 2004), human rhodopsin (hOPSIN; Kumar *et al.* 1996), human NQO-1 (hNQO1; Xie *et al.* 1995), human β-globin (hBglHS4; Stamatoyannopoulos *et al.* 1995) and mouse GST-Ya genes (mGSTY; Rushmore *et al.* 1991; Xie *et al.* 1995). The two artificial MAREs, CEN-C and

Table 4 Functional MARE-related sequences in regulatory regions of endogenous target genes

Possible transacting dimers	Gene name MARE consensus (CEN-C)	MAREs TGCTGACTCAGCA	Group*	References
Small Maf/p45	Human β-globin LCR HS-4 (hBglHS4)	GGCTGACTCActc	IV†	Stamatoyannopoulos *et al.* (1995)
	Human β-globin LCR HS-2	TGCTGAGTCAtgA	IV	Ney *et al.* (1990)
	Mouse β-globin LCR HS-2	TGCTGAGTCAtgc	IV	Moon & Ley (1990)
	Human porphobilinogen deaminase	TGCTGAGTCActg	IV	Mignotte *et al.* (1989)
	Human thromboxane synthase	TGCTGAtTCAttc	IV	Deveaux *et al.* (1997)
Small Maf/Nrf2	Human NQO1 (hNQO1)	TGCTGAGTCActg	IV†	Xie *et al.* (1995)
	Rat GST-P	TGCTGAaTCAtag	IV	Sakai *et al.* (1988)
	Mouse A170	TGCTGAGTCAtag	IV	Okazaki *et al.* (1999)
	Mouse heme oxygenase-1	TGCTGtGTCAttg	IV	Alam *et al.* (2000)
	Human GCLC	cGCTGAGTCAcgg	IV	Mulcahy *et al.* (1997)
	Human GCLM	TGCTtAGTCAttg	IV	Moinova & Mulcahy (1998)
	Rat, Mouse GST-Ya (mGSTY)	TGCTttGTCAcCA	IV†	Rushmore *et al.* (1991); Xie *et al.* (1995)
Large Maf homodimer	Mouse crystallin γF (mG2cryst)	TGCcaACaCAGCA	I†	Rajaram & Kerppola 2004
	Rat crystallin γD	TGCcaACgCAGCA	I	Klok *et al.* (1998)
	Chicken crystallin αA	TGCTGACcacGtt	II	Sharon-Friling *et al.* (1998)
	Chicken δ1 crystallin	TGCTGAtcCtGCA	II	Ogino & Yasuda (1998)
	Mouse type II collagen	gGCTctGTatGCg	I	(Huang *et al.* 2002)
	Mouse insulin	gGCTGAagCtGCA	II	Kataoka *et al.* (2002)
	Human rhodopsin gene (hOPSIN)	TGCTGAtTCAGCc	III†	Kumar *et al.* (1996)

Altered bases are shown in the same way as described in the footnote of Table 3. *Classification of each sequence was predicted from its alterations (see Fig. 5). †Adequacy of the classification was confirmed by EMSA, shown in Fig. 6D.

MARE23, were additionally examined as controls (Kyo *et al.* 2004). The rationale for these experiments was as follows. Since the mG2cryst MARE bears asymmetric mutations at positions −3, −2 and +1, we predicted that asymmetric core mutations on both sides might be equivalent to symmetric core mutations and that mG2cryst might display Group I properties, i.e. preferential association with MafG homodimers. We also predicted that the hOPSIN sequence might be permissive for binding to both dimers, since it possesses the central base alteration and asymmetric change at position +6, both of which are characteristic of Group III. Since hNQO1, hBglHS4, mGSTY and MARE23 possess features of Group IV (namely a G to C change at position +4 for the first three genes and a symmetric mutation from C/G to A/T at position −4/+4 in MARE23), we expected that these four MAREs might show preferential affinity for the MafG:Nrf2 heterodimer in spite of the additional mutations.

SPR responses were recorded and kinetic values were determined for three different protein preparations: MafG alone, a 1 : 1 mixture of MafG and Nrf2 and a 1 : 10 mixture of MafG and Nrf2 (Fig. 6A–C and Tables 2 and 3). In accordance with our expectations, mG2cryst showed a typical Group I sequence association, while hOPSIN showed a curve similar to typical Group III binding. Typical Group IV bindings were observed for hNQO1, hBglHS4, mGSTY and MARE23 when 1 : 1 and 1 : 10 mixtures of MafG and Nrf2 were examined (Fig. 6B,C). It was noted, however, that hBglHS4 and MARE23 displayed atypical binding properties for Group IV sequences since they did not show any significant interactions with the MafG homodimer (Fig. 6A).

When EMSA was performed using the protein mixture of MafG and Nrf2 that gave rise to strong heterodimer binding for CEN-C (Fig. 6D, lane 2), homodimer binding was dominant for mG2cryst (lane 8), as was the case for Group I MAREs, while heterodimer binding dominated associations with MARE23, hNQO1, hBglHS4 and mGSTY (lanes 5, 14, 17 and 19), as was the case for Group IV MAREs. Like Group III MAREs, both homodimer and heterodimer binding were observed for hOPSIN (lanes 11). When MafG was applied, homodimer binding was hardly observed for MARE23 and hBglHS4 (lanes 4 and 16), as expected from the SPR data. This suggested that the additional symmetric

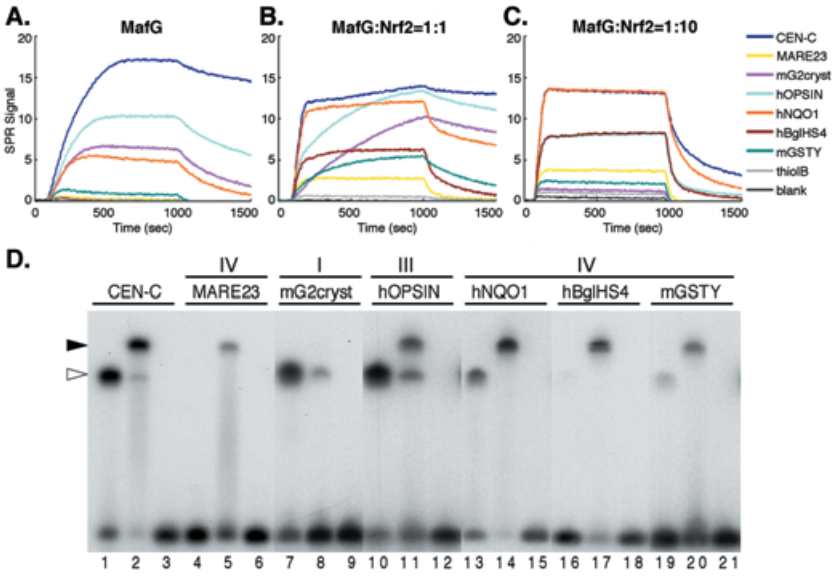

Figure 6 SPR signal curves and EMSA of MARE-related sequences from the regulatory regions of endogenous genes. SPR signals were monitored for the five MARE-related sequences together with CEN-C and MARE23, and the protein binding to these sequences was also examined in EMSA. Representative curves obtained from more than three independent experiments are shown: SPR signals of MafG homodimer (A); 1 : 1 mixture of MafG and Nrf2 (B); 1 : 10 mixture of MafG and Nrf2 (C). Sequence designations are shown on the right hand side of panel C and their sequences and calculated kinetic data are described in Tables 2 and 3. (D) EMSA with MafG homodimer and MafG:Nrf2 heterodimer. Lanes 1, 4, 7, 10, 13, 16 and 19 contain 0.36 μM MafG; lanes 2, 5, 8, 11, 14, 17 and 20 contain 0.6 μM each of Nrf2 and MafG; and lanes 3, 6, 9, 12, 15, 18 and 21 contain probes only. Black and white arrowheads indicate heterodimer and homodimer binding, respectively. The sequence designations of the probes are indicated above the lanes. The group number of each MARE variant is indicated above the sequence name.

alterations at positions –6/+6 and –5/+5 found in hBglHS4 and MARE23, respectively, significantly reduce the affinity for the MafG homodimer. With further analysis on MAREs with multiple base alterations, we found that heterodimer-orientd Group IV sequences can actually be divided into two subgroups, both of which possess the characteristic "G to T" or "G to C" substitution at position +4: MAREs that bind the MafG homodimer and those that don't (data not shown).

These results demonstrate that the relationship between the base alterations and the binding preferences summarized in Fig. 5 are applicable as general rules for predicting preferential MafG homodimer or MafG:Nrf2 heterodimer binding to genomic MARE-related sequences.

Variations in MAREs elicit distinct *trans*-acting potentials to Maf-containing homo- and heterodimers

We finally examined the functional significance of sequence variations within MAREs in regulating the transcriptional activation/repression process. We constructed four different luciferase reporter genes, containing MAREs from either the hNQO1, hBglHS4, hOPSIN or mG2cryst genes (Fig. 7A), and examined the responses of these genes to MafG homodimer and MafG:Nrf2 heterodimer co-transfection. Since MafG does not possess a canonical transcriptional activation domain, MafG homodimer interactions should result in transcriptional repression, while MafG:Nrf2 heterodimer interactions should be reflected by transcriptional activation.

When increasing amounts of MafG plasmid was transiently expressed with a constant amount of reporter gene, the luciferase activity was repressed, with the exception of hBglHS4 (left panels of Fig. 7B–E, lanes 1–4). When increasing amounts of Nrf2 were co-transfected, the luciferase activity increased except for the reporter driven by mG2cryst (right panels of Fig. 7B–E, lanes 5–9). Since Nrf2 cannot bind to DNA as a monomer or homodimer, the transfected Nrf2 must dimerize with an endogenous small Maf. Hence, these results correlate well with the binding properties obtained from the SPR imaging and EMSA; hNQO1 and hOPSIN interact with both the homodimer and the heterodimer, whereas hBglHS4 selectively interacts with the heterodimer and mG2cryst interacts exclusively with the homodimer. We

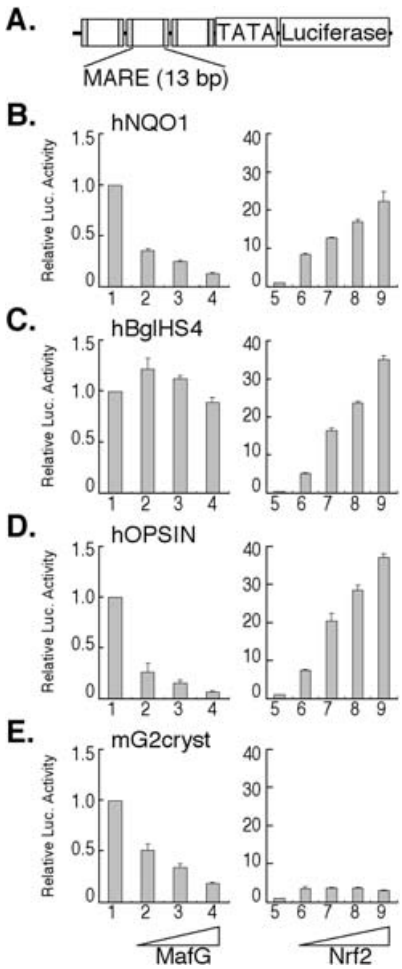

Figure 7 MAREs differentially transduce activation by MafG:Nrf2 heterodimer or repression by MafG homodimer according to their Group. (A) A reporter gene was constructed by linking three sets of a 13 bp-MARE (indicated by an open box) flanked by 6 bp-oligonucleotides (indicated by shaded boxes) in front of the rabbit β-globin TATA box and cDNA encoding luciferase. 293T cells were co-transfected with MafG or Nrf2 expression vectors and a reporter gene containing (B) hNQO1, (C) hBglHS4, (D) hOPSIN or (E) mG2cryst MARE. Lanes 1 and 5 indicate the relative luciferase activity generated by reporter gene in the absence of Nrf2 or MafG expression vectors. Relative luciferase activities in the presence of MafG expression vector (B–E: lane 2, 0125 μg; lane 3, 0.25 μg; lane 4, 0.5 μg) or Nrf2 expression vector (B, E: lane 6, 0.1 μg; lane 7, 0.25 μg; lane 8, 0.5 μg; lane 9, 1 μg. C, D: lane 6, 0.01 μg; lane 7, 0.05 μg; lane 8, 0.1 μg; lane 9, 0.25 μg) are shown. The relative luciferase activities were calculated against the activity generated by the reporter gene alone. The average values of three independent experiments are indicated and the error bars indicate standard deviations.

conclude from these data that the sequence variations within these MAREs alter their preferential interaction with transcription factors to elicit distinct responses, which in turn directly reflects the diversity in MARE-dependent gene activation or repression.

Discussion

In this study, we examined how the MafG homodimer and MafG:Nrf2 heterodimer interact with MAREs using SPR-microarray imaging. We found that MAREs can be categorized into five reasonably well-defined groups that bind to the MafG homodimer and/or the MafG:Nrf2 heterodimer based on the kinetic values derived from SPR-microarray imaging. This classification was verified by the analysis of previously identified MAREs that reside within the regulatory regions of endogenous target genes. The results demonstrate that various MAREs possess distinct binding specificities for MafG homodimers and MafG:Nrf2 heterodimers, which generates diversity in the regulation of gene expression. Thus, single base changes in the MARE sequence appear to be primary determinants for the differential binding of dimeric bZip transcription factors that coexist in cells.

SPR has not been commonly used to analyze the competitive binding of multiple proteins, as the contribution of each protein to the DNA binding becomes rather cloudy when multiple competitive proteins are coordinately applied in SPR analysis. Therefore prior to this work, the only studies on molecular interactions that could be analyzed by SPR examined one protein in solution interacting with single DNA motifs immobilized on the solid surface. We wanted to determine whether subtle variations within the *cis*-acting motifs contribute to the selection of specific *trans*-acting dimers from among multiple, related dimer populations. We unexpectedly challenged the first SPR analysis of the competitive binding of multiple transcription factors to a single DNA motif.

The two types of dimer that form in solution as a result of mixing Nrf2 and MafG are the MafG homodimer and MafG:Nrf2 heterodimer. We therefore examined the competitive binding of the MafG:Nrf2 heterodimer and MafG homodimer to genuine and synthetic MAREs. The coexistence of the MafG homodimer with the MafG:Nrf2 heterodimer in solution, both of which could interact with the MAREs, gave rise to strikingly different SPR signals. Titration studies using different molar ratios of Nrf2 and MafG provided an important clue for interpreting these heterogeneous patterns and finally revealed that the different patterns correspond to homodimer and

heterodimer binding. Thus, a unique feature of this study is that MARE variants were categorized primarily based on the SPR imaging of mixed homo- and hetero-dimer populations. This knowledge enabled us to distinguish MAREs that preferentially bind to the MafG homodimer, the MafG:Nrf2 heterodimer, or exhibit no preference under conditions in which both dimers coexist. It seems reasonable to propose that this condition may mimic physiological conditions in the cell. The calculated kinetic values for each dimer from the patterns recorded under the two most extreme conditions, when either only MafG homodimer or MafG:Nrf2 heterodimer was available, were found to be quite consistent with the classification based on the SPR curves.

The affinity for the MafG homodimer was generally reduced by symmetric mutations either in the core region or in the flanking region. Symmetric mutations at positions −5/+5 and −4/+4 strongly inhibited homodimer interactions (FL-S4GC, FL-S4TA, FL-S5TA, FL-S5AT and FL-S5CG). This is consistent with previous studies showing that a "GC" sequence in the flanking region (−5/+5 and −4/+4) is critical for stabilizing Maf protein binding (Kerppola & Curran 1994). Another feature of the homodimer binding profiles was that complete conservation of a half MARE was sufficient for the formation of a high-affinity binding site for the homodimer. When a half site of a Group V MARE, to which homodimers do not bind, is connected to another completely conserved consensus half MARE, the MafG homodimer can bind with high affinity (data not shown). This result indicates that not only the interaction between the dimer and the DNA, but also that between each dimer subunit, contributes to the stability of the bZip dimer/DNA complex. This conclusion is consistent with the analyses of another bZip family protein, GCN4 (Vinson *et al.* 2002), as well as studies examining bHLH/PAS family proteins (Chapman-Smith *et al.* 2004) and bHLH/Zip family proteins (Park *et al.* 2004).

MAREs that are well characterized *in vivo* are aligned in Table 4. Many erythroid- or megakaryocyte-specific genes possess critical MAREs in their regulatory regions (see References in Table 4), which were originally identified as NF-E2 binding sites (Mignotte *et al.* 1989). Small Maf:p45 heterodimers are the prime candidates that may interact with such MAREs (Andrews *et al.* 1993; Igarashi *et al.* 1994). Alternatively, MAREs are also known as the ARE (Rushmore *et al.* 1991) or the EpRE (Friling *et al.* 1990), which reside in the regulatory regions of phase II detoxifying enzyme genes and oxidative stress response genes (see References in Table 4). Small Maf:Nrf2 heterodimers, and on specific occasions small Maf:Nrf1 or small Maf:Nrf3 heterodimers, interact with these MAREs and, together with small Maf homodimers and small Maf:Bach1 heterodimers, either activate or repress transcription (Motohashi *et al.* 2002; Leung *et al.* 2003). All of these MAREs possess unique base alterations that are characteristic of Group IV sequences, namely the change of "G" to "C" or "T" at position +4. This observation coincides precisely with previous reports, as the consensus sequence for NF-E2 binding was reported to be "TGCTGA$^{G}/_{C}$TCAT" (Ney *et al.* 1990), while the ARE/EpRE consensus was reported to be *A*/GTGACNNNGC (Friling *et al.* 1990; Rushmore *et al.* 1991). Furthermore, *in vitro* binding site selections employing the MafG:Nrf1 heterodimer revealed that it binds preferentially to "TGCTGAGTCAT" (Johnsen *et al.* 1998).

We have proposed that the CNC-small Maf heterodimer (a transcriptional activator) and the small Maf homodimer (a transcriptional repressor) compete for the same binding sequence, and that the quantitative balance between the two finally determines the transcriptional output in megakaryocytes (Nagai *et al.* 1998; Motohashi *et al.* 2000). Based on these results, genes possessing Group III MARE sequences would be highly responsive to both small Maf:CNC heterodimers and small Maf homodimers. In contrast, genes possessing Group IV MARE sequences would be easily activated by low concentrations of small Maf:CNC heterodimers, whereas higher concentrations of small Maf homodimers would be required for binding to the same MARE sites to bring about transcriptional repression.

Another important cluster of MAREs has been identified in the genes regulated by large Maf proteins (see References in Table 4). MAREs with characteristics belonging to Groups I and II are found among these MAREs. GC/GC at positions −5–4/+4 +5, which are essential for DNA recognition by the Maf proteins, are well conserved in these MAREs. We observed that c-Maf, one of the large Mafs, can activate the reporter gene driven by mG2cryst, a typical Group I sequence (data not shown). Thus, genes possessing Group I/II MAREs seemed to be regulated specifically by the large Maf family of transcription factors. One exception for this prediction is the hOPSIN MARE motif in the rhodopsin gene. Since the hOPSIN MARE can respond to both Maf homodimers and Maf:CNC heterodimers, the rhodopsin gene may be sensitive to the relative abundance of these two kinds of dimer. Nrl, one of the large Maf proteins, is known to be a key activator of rhodopsin gene expression in rod cells of the retina (Kumar *et al.* 1996). In addition to the Nrl homodimer, CNC factors expressed in the rod cells, if any, could be involved in rhodopsin gene regulation through the MARE as a heterodimer with small Maf.

Central base alterations from C/G to A/T can be found in several genes regulated by large Maf factors, including the human rhodopsin gene (Kumar *et al.* 1996), the chicken δ1 crystalline gene (Ogino & Yasuda 1998) and the mouse insulin gene (Kataoka *et al.* 2002). An intriguing feature of the central base mutants is their high affinities for Maf containing dimers, including the MafG homodimer, MafG:Nrf2 heterodimer, large Maf homodimers and large Maf:Fos heterodimers, and their low affinities for the Jun homodimer and Jun:Fos heterodimer (AP-1) (see Tables 2 and 3) (Kataoka *et al.* 1994, 1995; Kumar *et al.* 1996). Hence, the central base alterations cited above appear to generate MAREs that are more specific for Maf-containing dimers, both homodimers and heterodimers, than for other bZip factors and exclude interference from the many other bZip transcription factors that may exist in a given type of differentiated cell.

In summary, based on these SPR-microarray studies, we conclude that MAREs create a huge diversity in the regulation of gene expression through single base alterations. Such subtle changes in the MARE sequences are sufficient to alter rather drastically the binding specificity for any given transcription factor dimer and thereby impart enormous diversity to gene regulatory networks formed by the Maf/CNC families of transcription factors.

Experimental procedures

Oligonucleotide DNAs

The oligonucleotides used for SPR-microarray imaging were synthesized by Hokkaido System Science and Qiagen as previously described (Kyo *et al.* 2004). The oligonucleotides with the thiol group protected were designed as 5′-HS-$(T)_{15}$-CGGAAT$(N)_{13}$ TTACTC-3′, with the 15-base thymine stretch on the 5′-end, and the complementary oligonucleotides were synthesized against the variable region with fixed 6-base regions on both sides. The variable 13-base regions were altered systematically and 39 mutations were generated in the MARE consensus sequence 5′-TGCTGA-C-TCAGCA-3′. There were nine types of symmetric variations in the core region (5′-TGC$(N)_3$-C-$(N)_3$GCA-3′) and nine types of symmetric variations in the flanking region (5′-$(N)_3$TGA-C-TCA$(N)_3$-3′). The rest were 18 asymmetric single base pair mutations (5′-$(N)_6$-C-TCAGCA-3′) and 3 central base mutations (5′-TGCTGA-N-TCAGCA-3′) (Fig. 1A and Table 1). In addition to these systematic variants, five oligonucleotides were prepared based on the functional MARE-related sequences found in the regulatory regions of several endogenous genes (Tables 2 and 3). One artificial sequence, MARE23, was used as a negative control for MafG homodimer binding (Kyo *et al.* 2004). The double-stranded portions of oligonucleotides used in the SPR detection, composed of a 13 bp variable sequence flanked by 6 bp fixed regions on both sides, were synthesized separately and used for probes in EMSA. ThiolB (5′-GCCAGCTTATTCAACTAG-3′) was used as an indifferent sequence as the negative control.

DNA array assembly and SPR-microarray imaging analysis

Assembly of the double-stranded DNA array and SPR-microarray imaging were performed as previously described (Kyo *et al.* 2004). The recombinant proteins were applied to the array surface at 100 μL/min in SPR buffer (20 mM HEPES, 250 or 300 mM NaCl, 4 mM $MgCl_2$, 1 mM EDTA, 1 mM Tris (2-carboxyethyl) phosphine and 0.005% Tween20). One hundred nanometers MafG protein solution was used for the SPR detection of MafG homodimer and the concentration of Nrf2 was changed from 0 to 1 μM for heterodimer analysis. Three hundred nanometers NaCl was used in the homodimer analysis, while 250 mM was used in the heterodimer analysis, because the higher salt concentration lowers the signal levels. The buffer pH was set at 7.9 as the optimal condition common to both dimers. The orientation of asymmetric MARE variants was also examined and it was found that it did not substantially affect the kinetic values, but did affect the maximal SPR signal levels when the heterodimer was applied. SPR signals whose levels of saturation were less than 10% of that of CEN-C were not considered to be significant. The kinetic values were calculated with a program based on the simple reversible reaction model (George *et al.* 1995). For the calculation of kinetic values, the concentrations of the dimers were used, which were 50 nM for MafG homodimer and 100 nM for MafG:Nrf2 heterodimer.

Nrf2 and MafG protein preparation

MafG protein was prepared as previously described (Kyo *et al.* 2004). For the preparation of His_6-tagged Nrf2, the *Hin*dIII/*Not*I fragment from pKI-256 (Itoh *et al.* 1997), which encodes the bZip domain of Nrf2, was cloned into the *Xho*I site of pET-15b (Novagen). The crude bacterial lysate containing His_6-tagged Nrf2 was purified with ProBond resin (Invitrogen).

EMSA

EMSA was performed as previously described (Kyo *et al.* 2004). The binding activity of the homodimer was examined using 0.36 μM of MafG. For the heterodimer binding analysis, mixtures of MafG:Nrf2 were prepared, with a set concentration of MafG of 0.6 μM and three different concentrations of Nrf2. Stronger heterodimer binding to CEN-C was generated by adding 0.6 μM Nrf2. A half and one third of the concentrations of Nrf2 were added for generating comparable binding activities between the homodimer and heterodimer and for generating stronger homodimer binding to CEN-C, respectively.

Reporter constructs and expression plasmids

Four kinds of luciferase reporter plasmids were constructed by replacing the *Mlu*I-*Nhe*I fragment of pRBGP2 (Igarashi *et al.* 1994) with one of the triplicate MARE-related sequences

(5′-CGGAAT$(N)_{13}$TTACTC-3′). Mouse MafG was expressed under the control of the *mafK* IM promoter (Motohashi *et al.* 1996) and mouse Nrf2 was expressed under the control of the EF promoter (Itoh *et al.* 1999).

Cell culture and transfection

293T cells were maintained in DMEM (Sigma) with 10% fetal bovine serum and 1% penicillin-streptomycin. The luciferase reporter plasmid and MafG or Nrf2 expression vector were transiently introduced into 293T cells using FuGENE 6 Transfection Reagent (Roche). pRL-TK (Promega) was used for normalizing the transfection efficiency. Cells were harvested with 1 × reporter lysis buffer (Promega) 36 h after transfection. Luciferase activity was measured using the Dual-Luciferase reporter assay system kit (Promega) and a luminometer (Berthold).

Acknowledgements

We are grateful to Drs Robert M. Corn (University of California-Irvine) for critical discussion and Fumiki Katsuoka (University of Tsukuba) for discussion and advice. This work was supported by grants from JST-ERATO (MY), the Japanese Ministry of Education, Culture, Sports, Science and Technology (HM and MY), Ministry of Health, Labor and Welfare (MY), Atherosclerosis Foundation (MY), the Yamanouchi Foundation for Research on Metabolic Disorders (HM) and Uehara Memorial Foundation (HM).

References

Alam, J., Wicks, C., Stewart, D., *et al.* (2000) Mechanism of heme oxygenase-1 gene activation by cadmium in MCF-7 mammary epithelial cells. Role of p38 kinase and Nrf2 transcription factor. *J. Biol. Chem.* **275**, 27694–27702.

Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. & Orkin, S.H. (1993) Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* **362**, 722–728.

Chapman-Smith, A., Lutwyche, J.K. & Whitelaw, M.L. (2004) Contribution of the Per/Arnt/Sim (PAS) domains to DNA binding by the basic helix-loop-helix PAS transcriptional regulators. *J. Biol. Chem.* **279**, 5353–5362.

Deveaux, S., Cohen-Kaminsky, S., Shivdasani, R.A., *et al.* (1997) p45 NF-E2 regulates expression of thromboxane synthase in megakaryocytes. *EMBO J.* **16**, 5654–5661.

Friling, R.S., Bensimon, A., Tichauer, Y. & Daniel, V. (1990) Xenobiotic-inducible expression of murine glutathione S-transferase Ya subunit gene is controlled by an electrophile-responsive element. *Proc. Natl. Acad. Sci. USA* **87**, 6258–6262.

George, A.J., French, R.R. & Glennie, M.J. (1995) Measurement of kinetic binding constants of a panel of anti-saporin antibodies using a resonant mirror biosensor. *J. Immunol. Methods* **183**, 51–63.

Grinberg, A.V. & Kerppola, T. (2003) Both Max and TFE3 cooperate with Smad proteins to bind the plasminogen activator inhibitor-1 promoter, but they have opposite effects on transcriptional activity. *J. Biol. Chem.* **278**, 11227–11236.

Huang, W., Lu, N., Eberspaecher, H. & De Crombrugghe, B. (2002) A new long form of c-Maf cooperates with Sox9 to activate the type II collagen gene. *J. Biol. Chem.* **277**, 50668–50675.

Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M. & Yamamoto, M. (1994) Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature* **367**, 568–572.

Itoh, K., Chiba, T., Takahashi, S., *et al.* (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.

Itoh, K., Wakabayashi, N., Katoh, Y., *et al.* (1999) Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev.* **13**, 76–86.

Jochum, W., Passegue, E. & Wagner, E.F. (2001) AP-1 in mouse development and tumorigenesis. *Oncogene* **20**, 2401–2412.

Johnsen, O., Murphy, P., Prydz, H. & Kolsto, A.B. (1998) Interaction of the CNC-bZIP factor TCF11/LCR-F1/Nrf1 with MafG: binding-site selection and regulation of transcription. *Nucleic Acids Res.* **26**, 512–520.

Kataoka, K., Han, S.I., Shioda, S., Hirai, M., Nishizawa, M. & Handa, H. (2002) MafA is a glucose-regulated and pancreatic beta-cell-specific transcriptional activator for the insulin gene. *J. Biol. Chem.* **277**, 49903–49910.

Kataoka, K., Igarashi, K., Itoh, K., *et al.* (1995) Small Maf proteins heterodimerize with Fos and potentially act as competitive repressors of NF-E2 transcription factor. *Mol. Cell. Biol.* **152**, 180–2190.

Kataoka, K., Noda, M. & 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.

Kerppola, T.K. & Curran, T. (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.

Klok, E.J., van Genesen, S.T., Civil, A., Schoenmakers, J.G. & Lubsen, N.H. (1998) Regulation of expression within a gene family. The case of the rat gammaB- and gammaD-crystallin promoters. *J. Biol. Chem.* **273**, 17206–17215.

Kumar, R., Chen, S., Scheurer, D., *et al.* (1996) The bZIP transcription factor Nrl stimulates rhodopsin promoter activity in primary retinal cell cultures. *J. Biol. Chem.* **271**, 29612–29618.

Kusunoki, H., Motohashi, H., Katsuoka, F., Morohashi, A., Yamamoto, M. & Tanaka, T. (2002) Solution structure of the DNA-binding domain of MafG. *Nat. Struct. Biol.* **9**, 252–256.

Kyo, M., Yamamoto, T., Motohashi, H., *et al.* (2004) Evaluation of MafG interaction with Maf recognition element arrays by surface plasmon resonance imaging technique. *Genes Cells* **9**, 153–164.

Leung, L., Kwong, M., Hou, S., Lee, C. & Chan, J.Y. (2003) Deficiency of the Nrf1 and Nrf2 transcription factors results in

early embryonic lethality and severe oxidative stress. *J. Biol. Chem.* **278**, 48021–48029.

Mignotte, V., Wall, L., deBoer, E., Grosveld, F. & Romeo, P.H. (1989) Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. *Nucleic Acids Res.* **17**, 37–54.

Moinova, H.R. & Mulcahy, R.T. (1998) An electrophile responsive element (EpRE) regulates beta-naphthoflavone induction of the human gamma-glutamylcysteine synthetase regulatory subunit gene. Constitutive expression is mediated by an adjacent AP-1 site. *J. Biol. Chem.* **273**, 14683–14689.

Moon, A.M. & Ley, T.J. (1990) Conservation of the primary structure, organization, and function of the human and mouse β-globin locus-activating regions. *Proc. Natl. Acad. Sci. USA* **87**, 7693–7697.

Motohashi, H., Igarashi, K., Onodera, K., *et al.* (1996) Mesodermal- vs. neuronal-specific expression of MafK is elicited by different promoters. *Genes Cells* **1**, 223–238.

Motohashi, H., Katsuoka, F., Shavit, J., Engel, J.D. & Yamamoto, M. (2000) Positive or negative MARE-dependent transcriptional regulation is determined by the abundance of small Maf proteins. *Cell* **103**, 65–875.

Motohashi, H., O'Connor, T., Katsuoka, F., Engel, D.J. & Yamamoto, M. (2002) Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. *Gene* **294**, 1–12.

Mulcahy, R.T., Wartman, M.A., Bailey, H.H. & Gipp, J.J. (1997) Constitutive and beta-naphthoflavone-induced expression of the human gamma-glutamylcysteine synthetase heavy subunit gene is regulated by a distal antioxidant response element/TRE sequence. *J. Biol. Chem.* **272**, 7445–7454.

Nagai, T., Igarashi, K., Akasaka, J., *et al.* (1998) Regulation of NF-E2 activity in erythroleukemia cell differentiation. *J. Biol. Chem.* **273**, 5358–5365.

Newman, J.R. & Keating, A.E. (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* **300**, 2097–2101.

Ney, P.A., Sorrentino, B.P., Lowrey, C.H. & Nienhuis, A.W. (1990) Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucleic Acids Res.* **18**, 6011–6017.

Nioi, P., McMahon, M., Itoh, K., Yamamoto, M. & Hayes, J.D. (2003) Identification of a novel Nrf2-regulated antioxidant response element (ARE) in the mouse NAD(P)H:quinone oxidoreductase 1 gene: reassessment of the ARE consensus sequence. *Biochem. J.* **374**, 337–348.

Ogino, H. & Yasuda, K. (1998) Induction of lens differentiation by activation of a bZIP transcription factor, L-Maf. *Science* **280**, 115–118.

Okazaki, M., Ito, S., Kawakita, K., *et al.* (1999) Cloning, expression profile, and genomic organization of the mouse STAP/A170 gene. *Genomics* **60**, 87–95.

Park, S., Chung, S., Kim, K.M., *et al.* (2004) Determination of binding constant of transcription factor myc-max/max-max and E-box DNA: the effect of inhibitors on the binding. *Biochim. Biophys. Acta* **1670**, 217–228.

Rajaram, N. & Kerppola, T.K. (2004) Synergistic transcription activation by Maf and Sox and their subnuclear localization are disrupted by a mutation in Maf that causes cataract. *Mol. Cell. Biol.* **24**, 5694–5709.

Ramirez-Carrozzi, V. & Kerppola, T. (2003) Asymmetric recognition of nonconsensus AP-1 sites by Fos-Jun and Jun-Jun influences transcriptional cooperativity with NFAT1. *Mol. Cell. Biol.* **23**, 1737–1749.

Rushmore, T.H., Morton, M.R. & Pickett, C.B. (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.

Sakai, M., Okuda, A. & Muramatsu, M. (1988) Multiple regulatory elements and phorbol 12-O-tetradecanoate 13-acetate responsiveness of the rat placental glutathione transferase gene. *Proc. Natl. Acad. Sci. USA* **85**, 9456–9460.

Sharon-Friling, R., Richardson, J., Sperbeck, S., *et al.* (1998) Lens-specific gene recruitment of zeta-crystallin through Pax6, Nrl-Maf, and brain suppressor sites. *Mol. Cell. Biol.* **18**, 2067–2676.

Stamatoyannopoulos, J.A., Goodwin, A., Joyce, T. & Lowrey, C.H. (1995) NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *EMBO J.* **14**, 106–116.

Swaroop, A., Xu, J., Pawer, H., Jackson, A., Skolnick, C. & Agarwal, N. (1992) A conserved retina-specific gene encodes a basic motif/leucine zipper protein. *Proc. Natl. Acad. Sci. USA* **89**, 266–270.

Vinson, C., Myakishev, M., Acharya, A., Mir, A.A., Moll, J.R. & Bonovich, M. (2002) Classification of human B-ZIP proteins based on dimerization properties. *Mol. Cell. Biol.* **22**, 6321–6335.

Xie, T., Belinsky, M., Xu, Y. & Jaiswal, A.K. (1995) ARE- and TRE-mediated regulation of gene expression. Response to xenobiotics and antioxidants. *J. Biol. Chem.* **270**, 6894–6900.

Received: 13 December 2005
Accepted: 12 March 2006