DNA-binding properties of MafR, a global regulator of *Enterococcus faecalis*

Sofía Ruiz-Cruz, Ana Moreno-Blanco, Manuel Espinosa and Alicia Bravo*

Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain

***Correspondence:** Dr. Alicia Bravo, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, E-28040 Madrid, Spain
Tel: +34 918373112
Fax: +34 915360432
Email: abravo@cib.csic.es

ABSTRACT

Global transcriptional regulators play key roles during bacterial adaptation to environmental fluctuations. Protein MafR from *Enterococcus faecalis* was shown to activate the transcription of many genes on a genome-wide scale. We proposed that MafR is a global regulator of the Mga/AtxA family. Here, we purified an untagged form of the MafR protein and found that it binds to linear double-stranded DNAs in a non-sequence-specific manner. Moreover, multiple MafR units (likely dimers) bind sequentially to the DNA molecule generating multimeric complexes. On DNAs that contain the promoter of the *mafR* gene, MafR recognizes a potentially curved DNA region. We discuss that a characteristic of the Mga/AtxA regulators might be their ability to recognize particular DNA shapes across the bacterial genomes.

Keywords: *Enterococcus faecalis*, global regulators, gene expression, protein-DNA interactions

Running title: Binding of MafR to DNA

Abbreviations: dNTP, deoxynucleoside triphosphate; dsDNA, double-stranded DNA; EMSA, electrophoretic mobility shift assay; IPTG, isopropyl-β-D-thiogalactopyranoside; OD, optical density; PCR, polymerase chain reaction; PEI, polyethyleneimine.

INTRODUCTION

Bacterial adaptation to a new niche usually requires global changes in gene expression. Many of these changes are coordinated by proteins that function as global transcriptional regulators. The ability of such proteins to recognize multiple DNA sites across the bacterial genome makes possible to adjust the gene expression pattern in response to environmental fluctuations. Several findings from structural and biochemical studies have shown that simple protein-DNA recognition mechanisms do not exist (1). Rohs *et al.* (2) classified two main readout mechanisms: base readout and shape readout. In the base readout mechanism, proteins recognize the unique chemical signatures of the DNA bases. In contrast, in the shape readout mechanism, proteins recognize a sequence-dependent DNA shape. Nevertheless, based on the structures of numerous protein-DNA complexes, it has been reasoned that particular proteins use likely a combination of readout mechanisms to achieve DNA binding specificity (2).

The Gram-positive bacterium *Enterococcus faecalis* is able to colonize different niches of the human host. It is generally found as a harmless commensal of the gastrointestinal tract. However, in immunocompromised hosts, *E. faecalis* can cause a variety of infections, such as urinary tract infections, endocarditis and bacteraemia (3, 4, 5). Our knowledge of the regulatory elements involved in the adaptation of *E. faecalis* to particular host niches is still very limited. Genome-wide microarray assays designed for the *E. faecalis* strain OG1RF showed that protein MafR activates, directly or indirectly, the expression of numerous genes (6). Some of them were found to be up-regulated during the growth of *E. faecalis* in blood and/or in human urine (7, 8). MafR (482 amino acids) has sequence similarity to three regulatory proteins: AtxA (40.7%; 475 amino acids) from *Bacillus anthracis*, Mga*Spn* (38.8%; 493 amino acids) from *Streptococcus pneumoniae*, and Mga (31.3%; 530 amino acids) from *S. pyogenes* (6). The three proteins are members of an emerging class of global transcriptional regulators involved in virulence (the Mga/AtxA family) (9, 10, 11). Furthermore, according to the Pfam database of protein families (12), MafR has two putative DNA-binding domains within the N-terminal region, the so-called HTH_Mga (Family PF08280, residues 11-69) and Mga (Family PF05043, residues 76-164) domains, which are

also present in Mga, AtxA and Mga*Spn* (6, 13, 14). Based on these findings, we proposed that MafR is a potential regulator of the Mga/AtxA family. However, the DNA-binding properties of MafR remain to be investigated.

In vitro protein-DNA interaction studies have been reported for Mga and Mga*Spn*, but not for AtxA. During exponential growth of *S. pyogenes*, Mga activates directly the transcription of several virulence genes. Some of them encode factors important for adherence to host tissues and for evasion of the host immune responses (15). *In vitro* experiments showed that a His-tagged Mga protein binds to regions located upstream of the target promoters. The position of the Mga binding sites with respect to the start of transcription differs among the promoters tested (13, 16). Although a consensus Mga binding sequence was initially proposed (17), subsequent sequence alignments revealed that the sites recognized by Mga exhibit a low sequence identity (13.4%) (16). Additionally, it has been shown that the His-tagged Mga protein is able to form higher-order oligomers in solution (18).

The pneumococcal Mga*Spn* protein activates directly the transcription of a four-gene operon (*spr1623-spr1626*) of unknown function. This activation requires a region located upstream of the target promoter (14). *In vitro* experiments showed that Mga*Spn* binds to linear double-stranded (ds) DNAs with little or no sequence specificity. Moreover, Mga*Spn* is able to generate multimeric complexes on linear DNAs (10). Additional results supported that Mga*Spn* recognizes structural features in its target DNA as follows: (i) Mga*Spn* binds preferentially to DNA sites that contain a potential intrinsic curvature flanked by regions of bendability, (ii) Mga*Spn* has a high affinity for a naturally occurring curved DNA, and (iii) Mga*Spn* has a preference for AT-rich DNA sites (10, 19). Because of these results, we proposed that a preference for particular DNA structures rather than for specific DNA sequences might be a general feature of the global regulators that constitute the Mga/AtxA family (19). In agreement with this hypothesis, sequence similarities in the promoter regions of the genes regulated by AtxA are not apparent, and some of those promoter regions are intrinsically curved (20).

In the present work, we purified an untagged form of the MafR protein and analysed its DNA-binding properties. By gel retardation assays, we found that MafR binds to linear dsDNAs in a non-sequence-specific manner. Multiple units

of MafR (likely dimers) bind orderly on the same DNA molecule generating multimeric complexes. Moreover, by footprinting experiments, we found that MafR binds to a potentially curved DNA region. Our results support that recognition of sequence-dependent DNA shapes might be a hallmark of the global regulators that belong to the Mga/AtxA family.

MATERIALS AND METHODS

Oligonucleotides, bacterial strains, and plasmids

Oligonucleotides used in this work are listed in Table 1. Chromosomal DNA was isolated from *E. faecalis* V583 (21), *E. faecalis* OG1RF (22), and *Streptococcus pneumoniae* R6 (23). *E. faecalis* JH2-2 (24) was used as a host for plasmids based on pDLF (6) and pDLS. The expression vector pDLS is a pDL287 (25) derivative. It has a unique restriction site for *Sph*I downstream of the pneumococcal *PsulA* promoter (26). For its construction, a 202-bp region (promoter *PsulA*) of the R6 genome was amplified by PCR using primers *pSul*F and *pSul*R. The PCR product was digested with *Cla*I, and the 181-bp restriction fragment was ligated to *Cla*I-linearized pDL287. Plasmids pDLF*mafR*$_{V583}$ and pDLS*mafR*$_{V583}$ carry the *P2493::mafR*$_{V583}$ and *PsulA::mafR*$_{V583}$ fusion genes, respectively. For their construction, a 1,546-bp region of the V583 chromosome was amplified using the *mafSph*F and *mafSph*R primers. After *Sph*I digestion, the 1,514-bp restriction fragment was inserted into the *Sph*I site of both expression vectors pDLF and pDLS.

For protein overproduction, an inducible expression system based on *Escherichia coli* BL21 (DE3) (a gift of F. W. Studier) and pET24b (Novagen) was used. To overproduce MafR$_{V583}$, a 1,502-bp region of the V583 chromosome was amplified by PCR using the Up*mafR* and Dw*mafR* primers. These primers have a single restriction site for *Nde*I and *Xho*I, respectively. The amplified product was digested with both enzymes, and the 1,470-bp digestion product was inserted into pET24b (plasmid pET24b-*mafR*$_{V583}$). To overproduce MafR$_{V583}$-His and MafR$_{OG1RF}$-His, a 1,481-bp DNA region was amplified by PCR using the Up*mafR*

and Dw*mafR*-His primers. The amplified DNA was digested with *Nde*I and *Xho*I, and the 1,448-bp digestion product was inserted into pET24b (plasmids pET24b-*mafR*$_{V583}$-His and pET24b-*mafR*$_{OG1RF}$-His). To overproduce MafR$_{OG1RF}$Δ3N-His, a 1,472-bp region of the OG1RF chromosome was amplified by PCR using the Up*mafR*-Δ3N and Dw*mafR*-His oligonucleotides. The 1,439-bp digestion product was inserted into pET24b (plasmid pET24b-*mafR*$_{OG1RF}$Δ3N-His).

DNA isolation

For small-scale preparations of plasmid DNA, the High Pure Plasmid Isolation Kit (Roche Applied Science) was used with the modifications reported for *Enterococcus* (6). Chromosomal DNA was isolated from *E. faecalis* and *S. pneumoniae* as previously described (26).

Growth and transformation of bacteria

E. faecalis was grown in Bacto™ Brain Heart Infusion (BHI) medium, which was supplemented with kanamycin (250 µg/ml) when the cells harboured a plasmid based on pDLF or pDLS. *E. coli* cells carrying a derivative of pET24b were grown in tryptone-yeast extract (TY) medium supplemented with kanamycin (30 µg/ml). Bacteria were grown at 37ºC. The protocols used to transform *E. faecalis* and *E. coli* by electroporation have been described (27, 28).

Western blot

Plasmid-carrying enterococcal cells were grown as indicated above to an optical density at 650 nm (OD_{650}) of 0.3 (exponential phase). To prepare whole-cell extracts, cells were concentrated 40-fold in buffer LBW (25 mM Tris-HCl, pH 7.6, 0.5 mM EDTA, 0.2 mg/ml lysozyme, 260 units/ml mutanolysin), and incubated at 37ºC for 10 min. Total proteins were separated by SDS-polyacrylamide (10%) gel electrophoresis. Pre-stained proteins (Invitrogen) were run in the same gel as molecular weight markers. Proteins were transferred electrophoretically to immunoblot polyvinylidene difluoride (PVDF) membranes (Bio-Rad) using a Mini Trans Blot (Bio-Rad) as reported previously (26). Membranes were probed with rabbit polyclonal antibodies against MafR$_{V583}$. Antigen-antibody complexes were detected using antirabbit horseradish peroxidase-conjugated antibodies, the Immun-Star™ HRP substrate kit (Bio-Rad), and the Luminescent Image Analyzer

LAS-3000 (Fujifilm Life Science). Rabbit polyclonal antibodies against $MafR_{V583}$ were produced in the Animal Facility of the Centro de Investigaciones Biológicas, CSIC (Madrid, Spain). For the immunizations, purified protein ($MafR_{V583}$) and traditional protocols (Freund´s complete adjuvant, Freund´s incomplete adjuvant, subcutaneous administration) were used.

Polymerase chain reaction (PCR)

The Phusion High-Fidelity DNA polymerase (Thermo Scientific) and the Phusion HF buffer were used. Reaction mixtures (50 µl) contained 5-30 ng of template DNA, 20 pmol of each primer, 200 µM each deoxynucleoside triphosphate (dNTP), and one unit of DNA polymerase. PCR conditions were reported previously (26). PCR products were purified with the QIAquick PCR purification kit (QIAGEN).

PCR-amplification of chromosomal DNA regions

By PCR, four regions of the V583 chromosome were amplified: (i) a 217-bp region (coordinates 2888932-2889148) using the 3012A and 3013A primers, (ii) a 227-bp region (2888932-2889158) using the 3012A and 3013B primers, (iii) a 260-bp region (coordinates 2888858-289117) using the 3012B and 3013C primers, and (iv) a 321-bp region (coordinates 94488-94808) using the 0091G2 and 0092A2 primers. From the *S. pneumoniae* R6 chromosome, a 322-bp region (coordinates 1598010-1598331) was amplified using the 26A and EM1 primers.

Annealing of complementary oligonucleotides

Oligonucleotides 20A, 26A, 32A, 40A and their complementary oligonucleotides (20B, 26B, 32B, 40B) (Table 1) were used to generate small dsDNA fragments. Equimolar amounts of complementary oligonucleotides were annealed in buffer TE (2 mM Tris-HCl, pH 8.0, 0.2 mM EDTA) containing 50 mM NaCl. Reaction mixtures (150 µl) were incubated at 95ºC for 10 min, cooled down slowly to 37ºC, kept at 37ºC for 10 min, and placed on ice for 10 min.

Overproduction and purification of proteins

E. coli BL21 (DE3) cells harbouring a pET24b derivative were grown as indicated above to an OD_{600} of 0.45. Then, isopropyl-β-D-thiogalactopyranoside (IPTG; 1

mM) was added. After 25 min, cells were treated with rifampicin (200 µg/ml) for 60 min. Cells were then collected by centrifugation, washed twice with buffer VL (50 mM Tris-HCl, pH 7.6, 5% glycerol, 1 mM DTT, 1 mM EDTA) containing 200 mM NaCl, and stored at -80ºC.

To purify $MafR_{V583}$, bacterial cells were concentrated 40-fold in buffer VL containing 200 mM NaCl and a protease inhibitor cocktail (Roche). To purify His-tagged proteins, an EDTA-free protease inhibitor cocktail (Roche) was used. Cells were lysed using a pre-chilled French pressure cell, and the whole-cell extract was centrifuged to remove cell debris. The clarified extract was mixed with 0.2% polyethyleneimine (PEI), kept on ice for 30 min, and centrifuged at 9,000 rpm in an Eppendorf F-34-6-38 rotor for 20 min at 4ºC. $MafR_{V583}$, as well as the His-tagged variants, were recovered in the PEI pellet, which was subsequently washed twice with buffer VL containing 200 mM NaCl. $MafR_{V583}$ and the His-tagged variants were eluted with buffer VL containing 500 mM NaCl. Proteins were precipitated with ammonium sulphate (70% saturation; 60 min on ice), followed by a centrifugation step (9,000 rpm for 20 min at 4ºC).

In the case of $MafR_{V583}$, the ammonium sulphate precipitate was dissolved in buffer VL containing 200 mM NaCl, and dialyzed against the same buffer at 4ºC. The protein preparation was loaded onto a heparin affinity column (HiPrep Heparin, GE Healthcare). To elute $MafR_{V583}$, a linear gradient of NaCl (200-600 mM) in buffer VL was used. Fractions containing $MafR_{V583}$ were identified by Coomassie-stained SDS-polyacrylamide (12%) gels, pooled, and dialyzed against buffer VL containing 200 mM NaCl. Then, the protein preparation was concentrated (Vivaspin-20, GE Healthcare), applied to a HiLoad Superdex 200 gel-filtration column, and subjected to fast-pressure liquid chromatography (Biologic Duoflow, Bio-Rad) at 4ºC. The running buffer contained 200 mM NaCl. Protein fractions containing $MafR_{V583}$ were pooled, concentrated and stored at -80ºC.

In the case of His-tagged proteins, the ammonium sulphate precipitate was dissolved in buffer S-His (10 mM Tris-HCl, pH 7.6, 5% glycerol, 300 mM NaCl, 1 mM DTT), and dialyzed against the same buffer at 4ºC. Imidazole (10 mM) was added to the protein preparation, which was then loaded onto a nickel affinity column (HisTrap HP column, GE Healthcare). To elute the His-tagged protein, a linear gradient of imidazole (10 to 250 mM) in buffer S-His was used. Fractions

containing the His-tagged protein were pooled, dialyzed against buffer VL containing 200 mM NaCl, concentrated, and stored at -80ºC.

Protein concentration was determined using a NanoDrop ND-2000 Spectrophotometer (Bio-Rad). $MafR_{V583}$, $MafR_{OG1RF}$-His and $MafR_{OG1RF}\Delta 3N$-His were subjected to amino terminal sequencing by Edman degradation using a Procise 494 Sequencer (Perkin Elmer).

Gel filtration chromatography

Protein $MafR_{V583}$ was injected into a HiLoad Superdex 200 gel-filtration column using a Biologic Duoflow system (Bio-Rad). Buffer VL containing 200 mM NaCl was used to equilibrate the column and as running buffer. The column was calibrated with various proteins of known Stokes radius: alcohol dehydrogenase (ADH; 45 Å), albumin (A; 35.5 Å), ovalbumin (O; 30.5 Å) and carbonic anhydrase (CA; 20.1 Å). Elution positions were monitored at 280 nm. The K_{av} value was calculated as $(V_e-V_o)/(V_t-V_o)$, where V_e is the elution volume, V_o is the void volume (determined by elution of blue dextran), and V_t is the total volume of the packed bed. Data were plotted according to Siegel and Monty (29).

Analytical ultracentrifugation

Experiments were performed at 12ºC in an Optima XL-I analytical ultracentrifuge (Beckman-Coulter) equipped with an UV-visible optical detection system, using an An50Ti rotor and Epon-charcoal standard double sector centerpieces (12 mm optical path). Sedimentation velocity assays were carried out at 48,000 rpm. $MafR_{V583}$ (5 and 10 µM; 350 µl) was equilibrated in buffer AU (50 mM Tris-HCl, pH 7.6, 1 mM EDTA, 0.1 mM DTT; 3% glycerol and 200 mM NaCl). The $MafR_{V583}$ sedimentation coefficient was estimated applying a direct linear least squares boundary modelling of the sedimentation velocity data using the SEDFIT program (version 12.0) (30). The sedimentation coefficient was corrected to standard conditions to obtain the corresponding $S_{20,W}$ value using the SEDNTERP program (31). The translational frictional coefficient (*f*) of $MafR_{V583}$ was determined from the molecular mass and sedimentation coefficient values (32), whereas the frictional coefficient of the equivalent hydrated sphere (f_0) was estimated using a hydration of 0.37 g H_20/g protein (33). With these parameters the translational frictional ratio (f/f_0) was calculated ($MafR_{V583}$ hydrodynamic shape).

Sedimentation equilibrium assays were performed at two protein concentrations (5 and 10 μM). Samples (90 μl) were centrifuged at two successive speeds (8,000 and 10,000 rpm) and absorbance readings were done after the sedimentation equilibrium was reached. The absorbance scans were taken at 280 and 291 nm, depending on $MafR_{V583}$ concentration. In all cases, the baseline signals were measured after high-speed centrifugation (40,000 rpm). Apparent average molecular masses of $MafR_{V583}$ were determined using the HETEROANALYSIS program (34). The partial specific volume of $MafR_{V583}$ was 0.742 ml/g, calculated from the amino acid composition with the SEDNTERP program (31).

Radioactive labelling of DNA fragments

Oligonucleotides were radioactively labelled at the 5' end using [γ-^{32}P]-ATP (3,000 Ci/mmol; Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs) as described (10). The 5' labelled oligonucleotides were used for PCR amplification (labelling at either the coding or the non-coding strand).

Electrophoretic mobility shift assays

In general, binding reactions (10-20 μl) contained either non-labelled DNA (10 nM) or ^{32}P-labelled DNA (1-2 nM) and different amounts of the purified protein. The binding buffer contained 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 0.2 mM EDTA, 1% glycerol, 50 mM NaCl, and 0.5 mg/ml bovine serum albumin (BSA). When indicated, non-labelled competitor calf thymus DNA and ^{32}P-labelled DNA were added simultaneously to the binding reaction. Reaction mixtures were incubated at room temperature for 20 min. Free and bound DNA forms were separated by electrophoresis on native polyacrylamide (6% or 8%) gels (Mini-PROTEAN system, Bio-Rad). Gels were pre-electrophoresed (20 min) and run at 100 V and room temperature. Labelled DNA was visualized using a Fujifilm Image Analyzer FLA-3000.

DNase I footprinting assays

Binding reactions (50 μl) contained 2 nM of ^{32}P-labelled DNA, 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 0.2 mM EDTA, 1% glycerol, 50 mM NaCl, 0.5 mg/ml BSA, 1 mM $CaCl_2$, 10 mM $MgCl_2$, and different concentrations of $MafR_{V583}$. Reaction

mixtures were incubated at room temperature for 20 min. Then, DNA was digested using 0.04 units of DNase I (Roche Applied Science). After 5 min at room temperature, reactions were stopped by adding 25 µl of Stop DNase I buffer (2 M ammonium acetate; 0.8 mM sodium acetate, 0.15 M EDTA). DNA was precipitated with ethanol, dried and dissolved in 5 µl of loading buffer (80% formamide, 10 mM NaOH, 0.1% bromophenol blue, 0.1% xylene cyanol, 1 mM EDTA). After heating at 95°C for 5 min, samples were loaded onto 8 M urea-6% polyacrylamide gels. Dideoxy-mediated chain termination sequencing reactions were run in the same gel. Labelled products were visualized using a Fujifilm Image Analyzer FLA-3000. The intensity of the bands was quantified using the Quantity One software (Bio-Rad).

In silico prediction of intrinsic curvature

The bendability/curvature propensity plots were calculated with the bend.it server (35) (http://hydra.icgeb.trieste.it/dna/bend_it.html) as described previously (10).

RESULTS

Purification of an untagged form of the MafR$_{V583}$ protein

To overproduce and then purify an untagged form of the MafR$_{V583}$ protein, we used a heterologous system based on the *E. coli* strain BL21 (DE3) and the inducible expression vector pET24b. The procedure used to purify MafR$_{V583}$ involved basically four steps: (i) precipitation of DNA and MafR$_{V583}$ with PEI using a low ionic strength buffer; (ii) elution of MafR$_{V583}$ from the PEI pellet using a higher ionic strength buffer; (iii) fractionation of proteins by heparin chromatography; and (iv) fractionation of proteins by gel filtration chromatography (Figure 1A). Purified MafR$_{V583}$ was analysed by SDS-polyacrylamide (12%) gel electrophoresis. It migrated between standard proteins of 45 and 66 kDa, which is consistent with the theoretical molecular weight of the MafR$_{V583}$ monomer (56,247 Da). Further determination of the N-terminal amino acid sequence (eight

residues) of MafR$_{V583}$ by Edman degradation showed that the first Met residue was not removed.

Next, we obtained polyclonal antibodies against MafR$_{V583}$ and demonstrated that they are suitable for detection of MafR$_{V583}$ in enterococcal whole-cell extracts by Western blotting (Figure 1B). Specifically, we inserted the *mafR*$_{V583}$ gene into the expression vectors pDLF (6) and pDLS (this work), which are based on the enterococcal *P2493* promoter and the pneumococcal *PsulA* promoter, respectively (26). Each recombinant plasmid (pDLF*mafR*$_{V583}$ and pDLS*mafR*$_{V583}$) was introduced into the enterococcal JH2-2 strain, which is a plasmid-free strain. As internal control, plasmid pDLF ('empty' vector) was introduced into JH2-2. Compared to cells harbouring plasmid pDLS*mafR*$_{V583}$, the amount of MafR$_{V583}$ was ~3-fold higher in cells harbouring plasmid pDLF*mafR*$_{V583}$ (Figure 1B). This result is consistent with our previous results, which showed that the activity of the *P2493* promoter is ~3.8-fold higher than the activity of the *PsulA* promoter in *E. faecalis* JH2-2 cells (26).

Hydrodynamic behaviour of MafR$_{V583}$

Gel filtration chromatography allowed us to determine the molecular size (Stokes radius) of the untagged MafR$_{V583}$ protein. MafR$_{V583}$ eluted from the column as a single peak (Figure 2A). The elution volume was used to calculate the K_{av} value as indicated in Materials and Methods. A calibration curve was obtained by loading onto the column several standard proteins of known Stokes radius (Figure 2A). The Stokes radius of MafR$_{V583}$ determined from the calibration curve was 43 Å, slightly lower than the value of the alcohol dehydrogenase standard protein (45 Å, 150 kDa). This result indicated that MafR$_{V583}$ behaves as a dimer in solution. This conclusion was further confirmed by analytical ultracentrifugation experiments (sedimentation velocity and sedimentation equilibrium) (Figure 2B). At 5 and 10 µM of MafR$_{V583}$, the sedimentation velocity profiles showed a major peak (92-98%) with an $S_{20,w}$ value of 6 S. A minor peak (4-6%) corresponding to a molecular species of higher sedimentation coefficient ($S_{20,w}$ = 6.4 S) was also observed. Sedimentation equilibrium assays showed that, at 5 µM of MafR$_{V583}$, the experimental data are best fit to an average molecular mass ($M_{w,a}$) of 118,000 ± 1,000 Da, a value that corresponds with the theoretical mass of a

$MafR_{V583}$ dimer (112,495 Da). A similar average molecular mass was determined at 10 μM (121,000 ± 1,000 Da). Thus, under the conditions tested, the $MafR_{V583}$ dimer is the major molecular species in the protein preparation. The frictional ratio (f/f_0) calculated from the analytical ultracentrifugation assays was 1.34, indicating that the hydrodynamic behaviour of the $MafR_{V583}$ dimer deviates from the behaviour corresponding to a rigid spherical particle (f/f_0 = 1.0). From these results we conclude that the shape of the $MafR_{V583}$ dimer is an ellipsoid.

$MafR_{V583}$ generates multimeric complexes on linear dsDNA

The pneumococcal Mga*Spn* regulator, a member of the Mga/AtxA family, was shown to generate multimeric complexes on linear dsDNAs (10). In this work, we analysed whether $MafR_{V583}$ has a similar ability. Firstly, we performed electrophoretic mobility shift assays (EMSAs) using a 217-bp DNA fragment (coordinates 2888932-2889148 of the V583 chromosome) (Figure 3A), which contains the promoter of the *mafR* gene (promoter *Pma*) (6). The ^{32}P-labelled DNA fragment (2 nM) was incubated with increasing concentrations of $MafR_{V583}$ in the presence of non-labelled competitor calf thymus DNA (5 μg/ml) (Figure 4A). At 20 nM of $MafR_{V583}$, free DNA and a protein-DNA complex (C1) were observed. However, as the concentration of $MafR_{V583}$ was increased, complexes of lower electrophoretic mobility appeared sequentially and faster-moving complexes disappeared gradually. Secondly, we performed dissociation experiments (Figure 4B). The ^{32}P-labelled DNA fragment (2 nM) was incubated with $MafR_{V583}$ (260 nM) in the absence of competitor DNA to generate higher-order protein-DNA complexes (Figure 4B, lane 1). Then, different amounts of calf thymus DNA were added to the reaction mixtures. As the concentration of competitor DNA was increased, protein-DNA complexes moving faster appeared gradually, as well as free DNA molecules. Taken together, these results indicated that multiple units of $MafR_{V583}$ bind orderly on the same linear DNA molecule generating multimeric complexes. Moreover, the $MafR_{V583}$ units are able to dissociate orderly from the higher-order complexes.

Further EMSAs using different linear dsDNAs indicated that $MafR_{V583}$, like the Mga*Spn* regulator (10), binds to DNA with little or no sequence specificity. Specifically, we used a 321-bp DNA fragment from the enterococcal V583 chromosome (coordinates 94488-94808) and a 322-bp DNA fragment from the

pneumococcal R6 chromosome (coordinates 1598010-1598331). Although both DNAs have a similar A+T content (72.3% and 71.1%, respectively), sequence similarities between them are not apparent. As shown in Figure S1 (Supplementary material), $MafR_{V583}$ was able to form multimeric complexes on both DNA fragments. Moreover, the pattern of complexes was similar to that generated by $MafR_{V583}$ on the 217-bp DNA fragment (Figure 4A).

By EMSA, we also analysed the ability of a His-tagged $MafR_{V583}$ protein ($MafR_{V583}$-His) to interact with linear dsDNAs. This variant of $MafR_{V583}$ carries the Leu-Glu-6xHis peptide (His-tag) fused to its C-terminus. As shown in Figure S2 (Supplementary material), $MafR_{V583}$-His was able to generate multiple complexes on a 260-bp DNA fragment (coordinates 2888858-2889117 of the V583 chromosome) that contains the *Pma* promoter (see Figure 3A). Thus, the presence of the His-tag at the C-terminal end of $MafR_{V583}$ does not affect its ability to generate multimeric complexes.

$MafR_{V583}$ binds to a potentially curved DNA region

Regulators of the Mga/AtxA family appear to bind DNA with low sequence specificity. It has been shown that all established Mga-binding sites exhibit only 13.4% identity (16). In the case of AtxA, sequence similarities in its target promoters are not apparent, and *in silico* and *in vitro* analyses revealed that the promoter regions of several target genes are intrinsically curved (20). Moreover, *in vitro* studies indicated that Mga*Spn* recognizes structural features in its DNA targets rather than specific nucleotide sequences (10, 19). To further investigate the DNA-binding properties of $MafR_{V583}$, we performed DNase I footprinting experiments. We used a 227-bp DNA fragment (coordinates 2888932-2889158) that contains the *Pma* promoter (see Figure 3A). Figure 3B shows the bendability/curvature propensity plot of this DNA fragment according to the bend.it program (35). The highest magnitude of curvature propensity (~14 degrees per helical turn) is located at position -104 relative to the transcription start site of the *mafR*$_{V583}$ gene. Thus, the region upstream of the *Pma* promoter is potentially curved. Moreover, such a predicted curvature is located at a region of conspicuous bendability.

The 227-bp DNA fragment (2 nM) was ^{32}P-labelled either at the 5´-end of the coding strand or at the 5´-end of the non-coding strand (Figure 5). On the coding

strand and at 80 nM of MafR$_{V583}$, two regions protected against DNase I digestion were observed, from position -69 to -80 and from position -88 to -104. On the non-coding strand and at 120 nM of MafR$_{V583}$, diminished cleavages were observed from -73 to -80 and from -88 to -101. These results indicated that, on the 227-bp DNA, MafR$_{V583}$ binds preferentially to a region located upstream of the *Pma* promoter (between positions -69 and -104) (see Figure 3A). Such a site is adjacent to the peak of the potential curvature (Figure 3B). On both DNA strands and at 200 nM of MafR$_{V583}$, regions protected against DNase I digestion were observed along the DNA fragment, which is consistent with the pattern of protein-DNA complexes observed by EMSA (Figure 4A).

MafR$_{V583}$ binds to a 32-bp DNA but not to a 26-bp DNA

To define the minimum size of DNA needed for binding of MafR$_{V583}$, we used dsDNAs of 20-, 26-, 32- and 40-bp, which were obtained by annealing of complementary oligonucleotides. The sequence of such oligonucleotides (Table 1) was based on the pneumococcal R6 genome, and identical sequences were not found in the enterococcal V583 genome. The non-labelled dsDNAs were incubated with increasing concentrations of MafR$_{V583}$. Reaction mixtures were then analysed by electrophoresis on native polyacrylamide gels. In the case of the 20-bp DNA, a faint band was observed at 0.3 µM of MafR$_{V583}$ (Figure 6A). However, its intensity did not change significantly as the protein concentration was increased. Most of the DNA moved as free DNA even at high protein concentrations (6-9 µM). In the case of the 26-bp DNA, most of the DNA moved as free DNA at 4 µM of MafR$_{V583}$ (Figure 6B). Different results were obtained with the 32-bp DNA (Figure 6B) and the 40-bp DNA (not shown). In both cases, a protein-DNA complex was detected at 0.2 µM of MafR$_{V583}$, and its amount increased as the amount of free DNA decreased. From these results we conclude that the minimum size of DNA required for MafR$_{V583}$ binding is between 26-bp and 32-bp. In the case of the pneumococcal Mga*Spn* regulator, the minimum DNA size for binding was reported to be between 20-bp and 26-bp (10).

Protein MafR$_{OG1RF}$-His, but not MafR$_{OG1RF}$Δ3N-His, binds to DNA

The genome sequence of the enterococcal OG1RF strain was published in 2008 (22). Compared to strain V583 (21), the OG1RF genome contains 227 unique

open reading frames but has fewer mobile genetic elements. Despite this difference between both genomes, the nucleotide sequence of the region that spans coordinates 2888932 to 2889103 in V583 is identical in OG1RF (see Figure 3A). Such a region contains the *Pma* promoter (6) and the primary binding site of $MafR_{V583}$ (this work). However, in contrast to $MafR_{V583}$, MafR encoded by the OG1RF genome ($MafR_{OG1RF}$) has five amino acid changes (Ala37Thr, Gln131Leu, Met145Thr, Ser193Asn, Ile388Ser), and three of them (Ala37Thr, Gln131Leu, Met145Thr) are located within the predicted DNA-binding domain (residues 11 to 164). To analyse whether these changes affect the formation of protein-DNA complexes, we purified a His-tagged $MafR_{OG1RF}$ protein ($MafR_{OG1RF}$-His). $MafR_{OG1RF}$-His carries the His-tag fused to its C-terminus. EMSA experiments showed that $MafR_{OG1RF}$-His is able to form multimeric complexes on the 217-bp DNA fragment that contains the *Pma* promoter (Figure 7). The pattern of complexes was similar to that generated by $MafR_{V583}$ (Figure 4A). Thus, the amino acid substitutions of $MafR_{OG1RF}$ do not affect its interaction with DNA.

We identified previously the transcription initiation site of the $mafR_{V583}$ gene at coordinate 2889071, and proposed that the first ATG codon is likely the translation start site (6) (Figure 3A). Translation from this ATG generates a protein of 482 amino acid residues ($MafR_{V583}$ and $MafR_{OG1RF}$). However, there is a second ATG codon that might function as a translation start site. Translation from this second ATG would result in a variant that lacks the first three amino acid residues (here named $MafR_{OG1RF}\Delta3N$). To analyse whether the lack of these residues (Met-Tyr-Ser) impairs the formation of protein-DNA complexes, we purified a His-tagged $MafR_{OG1RF}\Delta3N$ protein ($MafR_{OG1RF}\Delta3N$-His). The sequence of its N-terminal end (eight residues) was confirmed by Edman degradation. EMSA experiments showed that $MafR_{OG1RF}\Delta3N$-His has lost the capacity to interact with DNA (Figure 7). From these results we conclude that (i) the first ATG codon is the translation start site, and (ii) the first three amino acid residues of $MafR_{OG1RF}$ are crucial for its structure and/or function.

DISCUSSION

Bacteria have evolved complex regulatory networks to rapidly adapt to environmental fluctuations. Global transcriptional regulators are key elements in such networks due to their ability to activate and/or repress the expression of multiple genes through a variety of mechanisms. In *E. faecalis*, various transcriptome analyses support the notion that bacterial adaptation is associated with global changes in gene expression (7, 8, 36). Our previous work identified MafR as a protein involved in global regulation of gene expression. Moreover, we proposed that MafR is a member of the Mga/AtxA family of global regulators. This proposal was based on amino acid sequence similarities and predictions of functional domains (6). However, the interaction of MafR with DNA has not previously been investigated.

In this work, we purified an untagged form of the $MafR_{V583}$ protein. Using $MafR_{V583}$ and various linear dsDNAs, we have found that the DNA-binding behaviour of $MafR_{V583}$ is very similar to the one described for the pneumococcal Mga*Spn* regulator (10). First, gel retardation assays indicated that $MafR_{V583}$ is able to bind DNA with low sequence specificity. Second, on a 227-bp DNA fragment that contains the *Pma* promoter, footprinting experiments showed that $MafR_{V583}$ binds preferentially to a site that is adjacent to the peak of a potential curvature. Such a binding site is located upstream of the *Pma* promoter (positions -69 to -104). In the case of Mga*Spn*, two primary binding sites were identified by footprinting experiments: the *P1623B* promoter site (positions -60 to -99) and the *Pmga* promoter site (positions -23 to +21) (10). Both binding sites have a low sequence identity but contain an intrinsic curvature flanked by regions of bendability. Regarding the Mga and AtxA regulators, a low sequence identity has been found in the target promoters (13, 16, 20). Moreover, the promoter regions of some AtxA target genes are intrinsically curved (20).

Another interesting finding of our current study is that $MafR_{V583}$, like Mga*Spn* (10), is able to generate multimeric complexes on linear dsDNAs. In both cases, the pattern of complexes observed by EMSA is compatible with a sequential binding of multiple protein units (likely dimers) to the same DNA molecule. Furthermore, in both cases, the protein units dissociate sequentially from the higher-order complexes in the presence of increasing amounts of competitor

DNA. Formation of multimeric protein-DNA complexes has not been reported for the Mga regulator. Nevertheless, a correlation between its ability to oligomerize in solution (without DNA) and its ability to activate transcription has been shown (18). Regarding AtxA, protein-DNA interaction studies have not been reported, but protein-protein interaction analyses revealed that AtxA exists in a homo-oligomeric state (37).

The genome sequence of the *E. faecalis* V583 clinical isolate was published in 2003 (21). Five years later the genome sequence of the OG1RF strain was published (22). This strain is a derivative of the OG1 human isolate. At present, 557 *E. faecalis* genomes have been totally or partially sequenced (NCBI, Genome Assembly and Annotation Report, 22/12/2017). Among them, we have found that 146 *E. faecalis* genomes encode a protein that is identical to $MafR_{V583}$ (Table S1). Furthermore, we have found that 50 *E. faecalis* genomes encode a protein that is identical to $MafR_{OG1RF}$ (Table S2). Compared to $MafR_{V583}$, $MafR_{OG1RF}$ has five amino acid substitutions, and three of them (Ala37Thr, Gln131Leu, Met145Thr) are located within the putative DNA-binding domain (Table 2). Despite this difference between $MafR_{V583}$ and $MafR_{OG1RF}$, we have shown that both proteins are able to generate multimeric complexes on linear dsDNAs, which could be an indication of functional conservation. Moreover, we have shown that removal of the first three amino acids in $MafR_{OG1RF}$, and likely in $MafR_{V583}$, results in a protein unable to interact with DNA. Similar DNA binding properties are expected for MafR from the strains that belong to the groups GA2, KS19, MTmid8, B653, X98, AZ19 and Com1 (Table S3). Compared to $MafR_{V583}$, MafR from such strains has one to four amino acid substitutions, and all of them are present in $MafR_{OG1RF}$ (Table 2).

In conclusion, recognition of specific sites across the genome by transcriptional regulators is essential for controlling gene expression. Different mechanisms for protein recognition of specific DNA sites have been characterized. In some cases, DNA binding proteins recognize intrinsic DNA structural characteristics rather than particular nucleotide sequences (2, 38). Examples where both DNA base sequence and shape recognition are required for protein binding have been also reported (39, 40, 41). Our study on the DNA-binding properties of MafR reinforces that recognition of particular DNA structures might be a general feature of the global regulators that constitute the Mga/AtxA family.

ACKNOWLEDGMENTS

We thank Dr. Virtu Solano-Collado for helpful discussions. This work was supported by grants BIO2016-76412-C2-2-R (AEI/FEDER, UE) and BIO2015-69085-REDC from the Spanish Ministry of Economy and Competitiveness.

REFERENCES

1. Siggers, T. and Gordân, R. (2014) Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res*. **42**, 2099-2111

2. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B. and Mann, R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233-269

3. Ramsey, M., Hartke, A. and Huycke, M. (2014) The physiology and metabolism of Enterococci. *In: Gilmore MS, Clewell DB, Ike Y, et al., editors. Enterococci: From Commensals to Leading Causes of Drug Resistant Infection (Internet).* Boston: Massachusetts Eye and Ear Infirmary, 1-43

4. Hollenbeck, B.L. and Rice, L.B. (2012) Intrinsic and acquired resistance mechanisms in enterococcus. *Virulence* **3**, 421-569

5. Fisher, K. and Phillips, C. (2009) The ecology, epidemiology and virulence of *Enterococcus*. *Microbiology* **155**, 1749-1757

6. Ruiz-Cruz, S., Espinosa, M., Goldmann, O. and Bravo, A. (2016) Global regulation of gene expression by the MafR protein of *Enterococcus faecalis*. *Front. Microbiol.* **6**:1521

7. Vebø, H.C., Snipen, L., Nes, I.F. and Brede, D.A. (2009) The transcriptome of the nosocomial pathogen *Enterococcus faecalis* V583 reveals adaptive responses to growth in blood. *PLoS ONE* **4**, e7660

8. Vebø, H.C., Solheim, M., Snipen, L., Nes, I.F. and Brede, D.A. (2010) Comparative genomic analysis of pathogenic and probiotic *Enterococcus faecalis* isolates, and their transcriptional responses to growth in human urine. *PLoS ONE* **5**, e12489

9. Hondorp, E.R., Hou, S.C., Hause, L.L., Gera, K., Lee, C.-E. and McIver, K.S. (2013) PTS phosphorylation of Mga modulates regulon expression and virulence in the group A streptococcus. *Mol. Microbiol.* **88**, 1176-1193

10. Solano-Collado, V., Lurz, R., Espinosa, M. and Bravo, A. (2013) The pneumococcal Mga*Spn* virulence transcriptional regulator generates multimeric complexes on linear double-stranded DNA. *Nucleic Acids Res.* **41**, 6975-6991

11. Hammerstrom, T.G., Horton, L.B., Swick, M.C., Joachimiak, A., Osipiuk, J. and Koehler, T.M. (2015) Crystal structure of *Bacillus anthracis* virulence regulator AtxA and effects of phosphorylated histidines on multimerization and activity. *Mol. Microbiol.* **95**, 426-441

12. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-D285

13. Hondorp, E.R. and McIver, K.S. (2007) The Mga virulence regulon: infection where the grass is greener. *Mol. Microbiol.* **66**, 1056-1065

14. Solano-Collado, V., Espinosa, M. and Bravo, A. (2012) Activator role of the pneumococcal Mga-like virulence transcriptional regulator. *J. Bacteriol.* **194**, 4197-4207

15. Ribardo, D.A. and McIver, K.S. (2006) Defining the Mga regulon: comparative transcriptome analysis reveals both direct and indirect regulation by Mga in the group A streptococcus. *Mol. Microbiol.* **62**, 491-508

16. Hause, L.L. and McIver, K.S. (2012) Nucleotides critical for the interaction of the *Streptococcus pyogenes* Mga virulence regulator with Mga-regulated promoter sequences. *J. Bacteriol.* **194**, 4904-4919

17. McIver, K., Heath, A., Green, B. and Scott, J. (1995) Specific binding of the activator Mga to promoter sequences of the *emm* and *scpA* genes in the group A streptococcus. *J. Bacteriol.* **177**, 6619-6624

18. Hondorp, E.R., Hou, S.C., Hempstead, A.D., Hause, L.L., Beckett, D.M. and McIver, K.S. (2012) Characterization of the group A streptococcus Mga virulence regulator reveals a role for the C-terminal region in oligomerization and transcriptional activation. *Mol. Microbiol*. **83**, 953-967

19. Solano-Collado, V., Hüttener, M., Espinosa, M., Juárez, A. and Bravo, A. (2016) Mga*Spn* and H-NS: Two unrelated global regulators with similar DNA-binding properties. *Front. Mol. Biosci.* **3**, 60

20. Hadjifrangiskou, M. and Koehler, T.M. (2008) Intrinsic curvature associated with the coordinately regulated anthrax toxin gene promoters. *Microbiology* **154**, 2501-2512

21. Paulsen, I.T., Banerjei, L., Myers, G.S.A., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F. *et al.* (2003) Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science* **299**, 2071-2074

22. Bourgogne, A., Garsin, D.A., Qin, X., Singh, K.V., Sillanpaa, J., Yerrapragada, S., Ding, Y., Dugan-Rocha, S., Buhay, C., Shen, H. *et al.* (2008) Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol.* **9**, R110-R110

23. Hoskins, J., Alborn, W.E., Jr., Arnold, J., Blaszczak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.-J., Fuller, W. *et al.* (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* **183**, 5709-5717

24. Jacob, A.E. and Hobbs, S.J. (1974) Conjugal transfer of plasmid-borne multiple antibiotic resistance in *Streptococcus faecalis* var. *zymogenes*. *J. Bacteriol.* **117**, 360-372

25. LeBlanc, D.J., Chen, Y.Y. and Lee, L.N. (1993) Identification and characterization of a mobilization gene in the streptococcal plasmid, pVA380-1. *Plasmid* **30**, 296-302

26. Ruiz-Cruz, S., Solano-Collado, V., Espinosa, M. and Bravo, A. (2010) Novel plasmid-based genetic tools for the study of promoters and terminators in *Streptococcus pneumoniae* and *Enterococcus faecalis*. *J. Microbiol. Methods* **83**, 156-163

27. Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127-6145

28. Shepard, B.D. and Gilmore, M.S. (1995) Electroporation and efficient transformation of *Enterococcus faecalis* grown in high concentrations of glycine. *Methods Mol. Biol*. **47**, 217-226

29. Siegel, L.M. and Monty, K.J. (1966) Determination of molecular weights and frictional ratios of proteins in impure systems by the use of gel filtration and density gradient centrifugation. Applications to crude preparations of sulfite and hydroxylamine reductases. *Biochim. Biophys. Acta* **112**, 346-362

30. Schuck, P. and Rossmanith, P. (2000) Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers*, **54**, 328–341

31. Laue, T.M., Shah, B.D., Ridgeway, T.M. and Pelletier, S.L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. *In: Harding S. E., Rowe, A., Horton, J. C. editors. Analytical Ultracentrifugation in Biochemistry and Polymer Sciences. Cambridge: Royal Society of Chemistry,* 90-125

32. van Holde, K.E. (1985) Physical Biochemistry. *Englewood Cliffs: Prentice-Hall*.

33. Pessen, H. and Kumosinski, T.F. (1985) Measurements of protein hydration by various techniques. *Methods Enzymol.* **117**, 219-255

34. Cole, J.L. (2004) Analysis of heterogeneous interactions. *Methods Enzymol*. **384**, 212-232

35. Vlahovicek, K., Kaján, L. and Pongor, S. (2003) DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res.* **31**, 3686-3687

36. Lindenstrauss, A.G., Ehrmann, M.A., Behr, J., Landstorfer, R., Haller, D., Sartor, R.B. and Vogel, R. (2014) Transcriptome analysis of *Enterococcus faecalis* toward its adaption to surviving in the mouse intestinal tract. *Arch. Microbiol*. **196**, 423-433

37. Hammerstrom, T.G., Roh, J.H., Nikonowicz, E.P. and Koehler, T.M. (2011) *Bacillus anthracis* virulence regulator AtxA: oligomeric state, function and CO_2 -signalling. *Mol. Microbiol.* **82**, 634-647

38. Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H.J., Rohs, R. and Mann, R.S. (2015) Deconvolving the recognition of DNA shape from sequence. *Cell* **161**, 307-318

39. Deng, Z., Wang, Q., Liu, Z., Zhang, M., Machado, A.C.D., Chiu, T.P., Feng, C., Zhang, Q., Yu, L., Qi, L. *et al.* (2015) Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. *Nature Communications* **6**, 7642

40. Ding, P., McFarland, K.A., Jin, S., Tong, G., Duan, B., Yang, A., Hughes, T.R., Liu, J., Dove, S.L., Navarre, W.W. *et al.* (2015) A novel AT-rich DNA recognition mechanism for bacterial xenogeneic silencer MvaT. *PLOS Pathogens* **11**, e1004967

41. Al-Zyoud, W.A., Hynson, R.M.G., Ganuelas, L.A., Coster, A.C.F., Duff, A.P., Baker, M.A.B., Stewart, A.G., Giannoulatou, E., Ho, J.W.K., Gaus, K. *et al.* (2016) Binding of transcription factor GabR to DNA requires recognition of DNA shape at a location distinct from its cognate binding site. *Nucleic Acids Res.* **44**, 1411-1420

FIGURE LEGENDS

Figure 1. (A) Purification of MafR$_{V583}$. Proteins were analysed by SDS-polyacrylamide (12%) gel electrophoresis. Gels were stained with Coomassie Blue. Lane M: molecular weight standards (in kDa) were run in the same gel (LMW Marker, GE Healthcare). Lane 1: total proteins from cells non-treated with IPTG. Lane 2: total proteins from cells treated with IPTG for 25 min. Lane 3: total proteins from cells treated with IPTG (25 min) and then with rifampicin (60 min). Purification steps (lanes 4-10): cleared cell lysate (lane 4); supernatant after DNA precipitation with PEI in the presence of 200 mM NaCl (lane 5); proteins eluted from the PEI pellet using a buffer that contains 200 mM NaCl (lanes 6, 7); proteins eluted from the PEI pellet using a buffer that contains 500 mM NaCl (lane 8); protein preparation after heparin affinity chromatography (lane 9); protein preparation after gel filtration chromatography (lane 10). **(B) Detection of MafR$_{V583}$ in *E. faecalis* cell extracts by Western blotting.** Rabbit polyclonal antibodies against MafR$_{V583}$ were used. Total proteins from JH2-2 cells harbouring plasmid pDLF (lane 2), plasmid pDLF*mafR*$_{V583}$ (lane 3), and plasmid pDLS*mafR*$_{V583}$ (lane 4) were separated by SDS-polyacrylamide (10%) gel electrophoresis. Purified MafR$_{V583}$ (lane 1) was run in the same gel.

Figure 2. (A) Stokes radius of MafR$_{V583}$ determined by gel filtration chromatography. Upper part: elution profile of MafR$_{V583}$ on a HiLoad Superdex 200 gel-filtration column. Lower part: four standard proteins of known Stokes radius were loaded on the same column. CA: carbonic anhydrase. O: ovalbumin. A: albumin. ADH: alcohol dehydrogenase. **(B) Analytical ultracentrifugation analysis of MafR$_{V583}$.** The sedimentation equilibrium profile of MafR$_{V583}$ (5 µM) was taken at 280 nm. The lower part shows the experimental data (circles) and the best fit (continuous line) to a single species with a $M_{w,a}$ = 118,000 Da. The residuals to the fit are shown in the upper part. Inset: sedimentation velocity profile of the same MafR$_{V583}$ sample.

Figure 3. (A) Relevant features of the region located upstream of the *mafR* gene. The nucleotide sequence of the region that spans coordinates 2888932 to

2889103 of the *E. faecalis* V583 chromosome is shown. The stop codon (TAG) of EF3012, the main sequence elements of the *Pma* promoter (-35 box and -10 box), and the start codon (ATG) of *mafR* are indicated in boldface letters. The transcriptional terminator (TT) of EF3012, the transcription start site (+1 position) of *mafR*, and the putative Shine-Dalgarno sequence (SD) of *mafR* (6) are shown. The brackets denote the primary $MafR_{V583}$-binding site defined by DNase I footprinting assays in this work. **(B) The region upstream of the *Pma* promoter is potentially curved.** The bendability/curvature propensity plot of the 227-bp DNA fragment (coordinates 2888932-2889158 of the V583 chromosome) was calculated using the bend.it program (35). The transcription start site (+1) of the *mafR*$_{V583}$ gene is indicated with an arrow. The location of the main sequence elements of the *Pma* promoter (-35 and -10 elements) is shown (grey box). The site recognized preferentially by $MafR_{V583}$ on the 227-bp DNA fragment (this work, see Figure 5) is indicated (black box).

Figure 4. (A) Formation of $MafR_{V583}$-DNA complexes. The ^{32}P-labelled 217-bp DNA fragment (2 nM) was incubated with the indicated concentration of $MafR_{V583}$ in the presence of non-labelled calf thymus DNA (5 µg/ml). Reaction mixtures were loaded onto a native gel (6% polyacrylamide). All the lanes came from the same gel. Bands corresponding to free DNA (F) and some $MafR_{V583}$-DNA complexes (C1, C2, C3, and C4) are indicated. **(B) Dissociation of $MafR_{V583}$-DNA complexes.** The indicated concentration of non-labelled competitor calf thymus DNA was added to pre-formed $MafR_{V583}$-DNA complexes (260 nM $MafR_{V583}$, 2 nM ^{32}P-labelled 217-bp DNA).

Figure 5. DNase I footprints of complexes formed by $MafR_{V583}$ on the 227-bp DNA fragment. Either the coding or the non-coding strand relative to the *Pma* promoter was ^{32}P-labelled at the 5´-end. All the lanes displayed came from the same gel. Dideoxy-mediated chain termination sequencing reactions were run in the same gel (lanes A, C, G, T). Densitometer scans corresponding to DNA without protein (black line) and DNA with the indicated concentration of protein (grey line) are shown. The regions protected against DNase I digestion are indicated with brackets. The indicated positions are relative to the transcription start site of the *mafR*$_{V583}$ gene.

Figure 6. EMSAs using MafR$_{V583}$ and small DNA fragments. DNA fragments were obtained by annealing of complementary oligonucleotides. **(A)** Oligonucleotides 20A/20B were used to generate the 20-bp DNA. **(B)** Oligonucleotides 26A/26B and 32A/32B were used to generate the 26- and 32-bp DNAs, respectively. The indicated concentration of MafR$_{V583}$ was mixed with 300 nM (20- and 26-bp) or 200 nM (32-bp) of DNA. Binding reactions were analysed by native polyacrylamide (8%) gel electrophoresis. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad).

Figure 7. EMSAs with MafR$_{OG1RF}$-His and MafR$_{OG1RF}$Δ3N-His The 217-bp DNA fragment (10 nM) was incubated with the indicated concentration of protein. Reaction mixtures were loaded onto a native gel (6% polyacrylamide). Bands corresponding to free DNA (F) and protein-DNA complexes (C1, C2, C3, and C4) are indicated. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad).

Table 1. Oligonucleotides used in this work

Name	Sequence (5' to 3')[(a)]
*pSul*F	TGTTAATGGG**ATCGAT**TTCTGTTTG
*pSul*R	GACAT**ATCGAT**CACTCCC**GCATGC**ATTTTCATC
*mafSph*F	TTTTTATCCGTATTC**GCATGC**AAAAGGAGG
*mafSph*R	AACCAAACGAT**GCATGC**CGAAAGAAAGC
Up*mafR*	GCAAAAGGAGGTTTT**CATATG**TACTCCATG
Dw*mafR*	AGCCAAAAAA**CTCGAG**AATGTCCTCGCTAG
Dw*mafR*-His	CCTCGCTAGTT**CTCGAG**AAAATAAGAATGA
Up*mafR*-Δ3N	GGTTTTGCCATGTAC**CATATG**TTAAAACGT
3012A	AGGAATGGCTGTTGTAACCA
3012B	AGTGCGGCTCCTGTCGGTAA
3013A	AACAAACGAATTTGCCGAAGC
3013B	CAACTGTTCCAACAAACG
3013C	CCGTTATCACACGTTTTAACA
0091G2	GGCTATTTTGATGCACATATCTG
0092A2	CCCGCCTTCCTTCCCTTGCTC
EM1	AGTTGAATGTTTAAAGAAATGATGG
26A	TTCTTTGTGGTATAATTGCAAGAGGT
26B	ACCTCTTGCAATTATACCACAAAGAA
20A	TATATTGTCTCCGTAGTGTT
20B	AACACTACGGAGACAATATA
40A	TATATCATGCTATACCTATTCTTTGTGGTATAATTGCAAG
40B	CTTGCAATTATACCACAAAGAATAGGTATAGCATGATATA
32A	TTCTTTGTGGTATAATTGCAAGAGGTTTAATC
32B	GATTAAACCTCTTGCAATTATACCACAAAGAA

(a) Restriction sites are in bold, and the base changes that generate restriction sites are underlined.

Table 2. Amino acid substitutions in the indicated MafR proteins compared to $MafR_{V583}$

Strain [(a)]	Amino acid residue					Identical [(b)]
	37	131	145	193	388	
V583	A	Q	M	S	I	146
GA2	T					2
KS19	T		T			6
MTmid8			T		S	2
B653	T	L	T			1
X98	T		T		S	15
AZ19	T	L	T	N		8
Com1	T	L	T		S	7
OG1RF	T	L	T	N	S	50

Data from the National Center for Biotechnology Information (Identical Protein Groups) (22/12/2017)
(a) Name of the *E. faecalis* strain that represents the group
(b) Number of strains that belong to the group (identical MafR) (see Tables S1, S2 and S3)

Figure 1

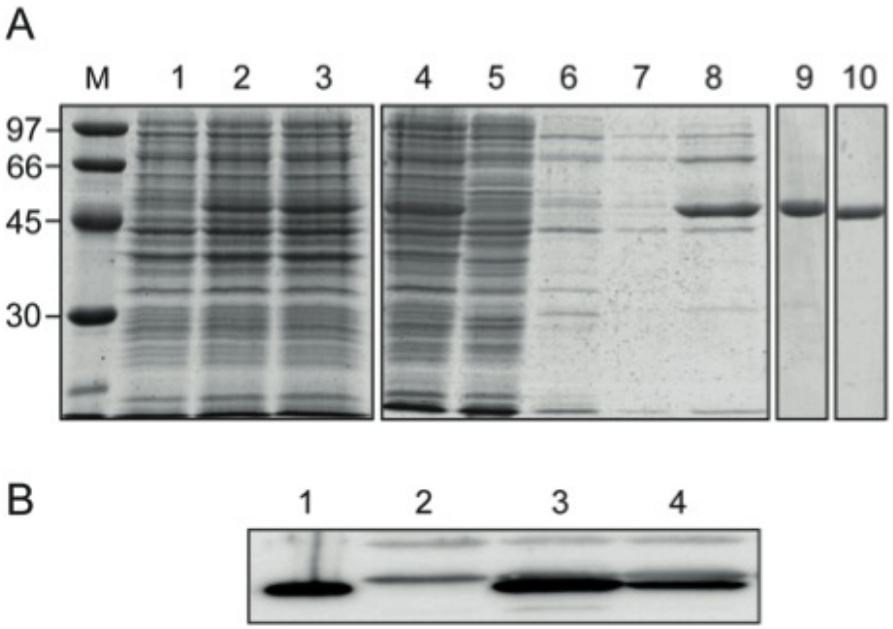


Figure 2

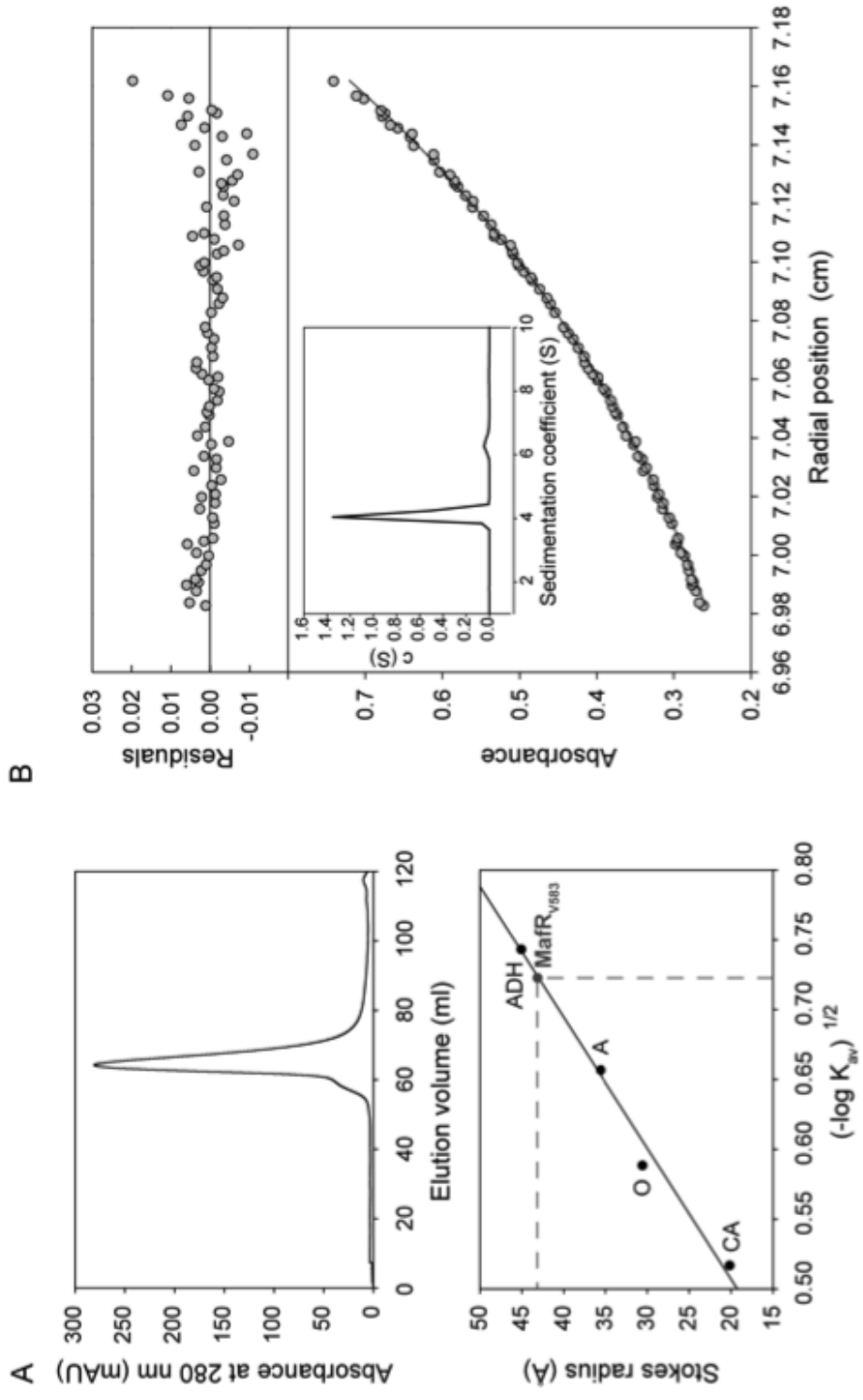

A
Absorbance at 280 nm (mAU)
Elution volume (ml)
Stokes radius (Å)
(-log K_{av})$^{1/2}$
ADH
$MafR_{V583}$
A
O
CA
B
Residuals
Absorbance
Radial position (cm)
Sedimentation coefficient (S)
c (S)

Figure 3

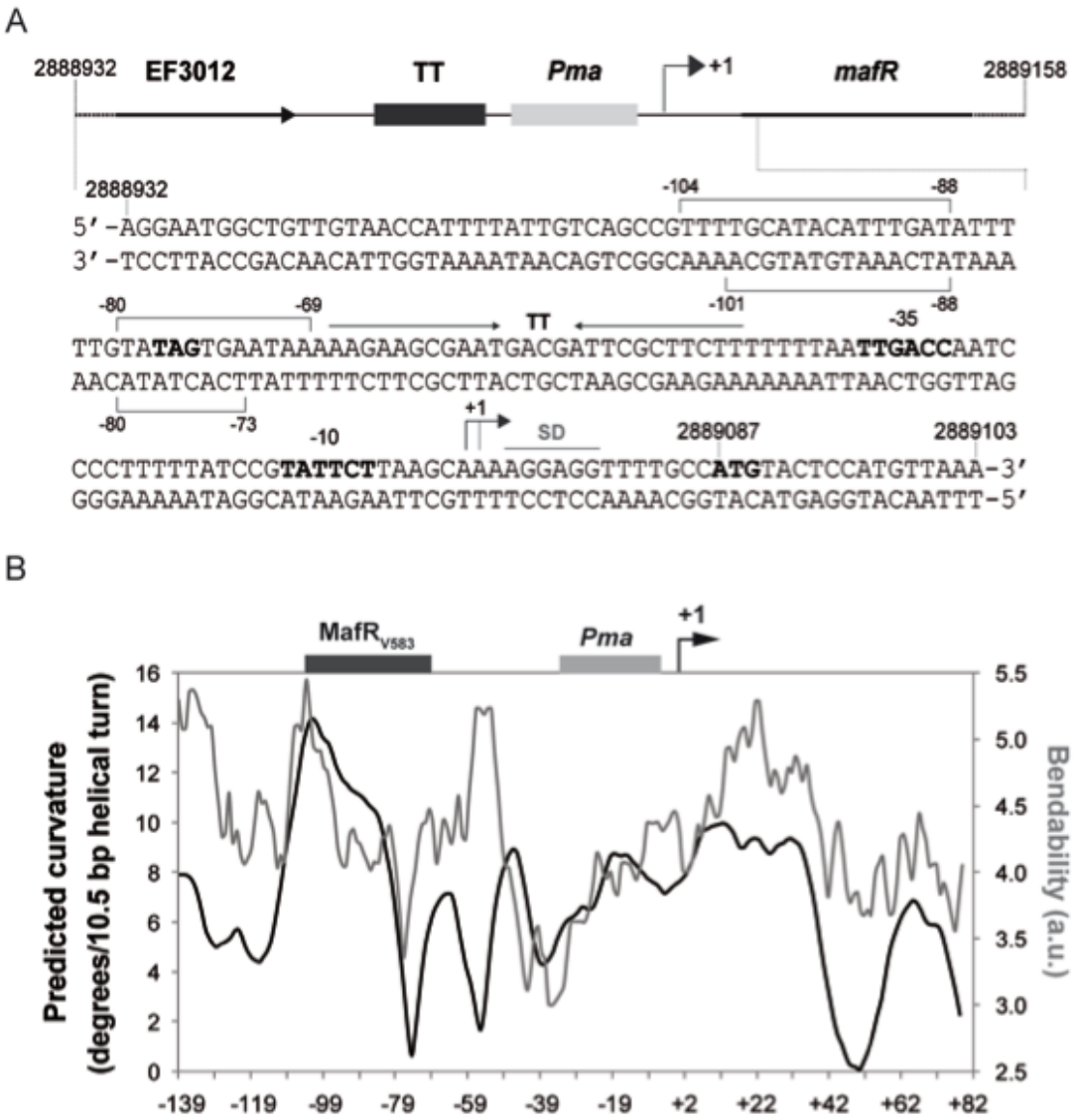

A
2888932
EF3012
TT
Pma
+1
mafR
2889158
2888932
-104
-88
5′-AGGAATGGCTGTTGTAACCATTTTATTGTCAGCCGTTTTGCATACATTTGATATTT
3′-TCCTTACCGACAACATTGGTAAAATAACAGTCGGCAAAACGTATGTAAACTATAAA
-80
-69
-101
-88
TT
-35
TTGTATAGTGAATAAAAAGAAGCGAATGACGATTCGCTTCTTTTTTTAATTGACCAATC
AACATATCACTTATTTTTCTTCGCTTACTGCTAAGCGAAGAAAAAAATTAACTGGTTAG
-80
-73
+1
-10
SD
2889087
2889103
CCCTTTTTATCCGTATTCTTAAGCAAAAGGAGGTTTTGCCATGTACTCCATGTTAAA-3′
GGGAAAAATAGGCATAAGAATTCGTTTTCCTCCAAAACGGTACATGAGGTACAATTT-5′
B
MafR V583
Pma
+1
Predicted curvature
(degrees/10.5 bp helical turn)
Bendability (a.u.)
16
14
12
10
8
6
4
2
0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
-139
-119
-99
-79
-59
-39
-19
+2
+22
+42
+62
+82

Figure 4

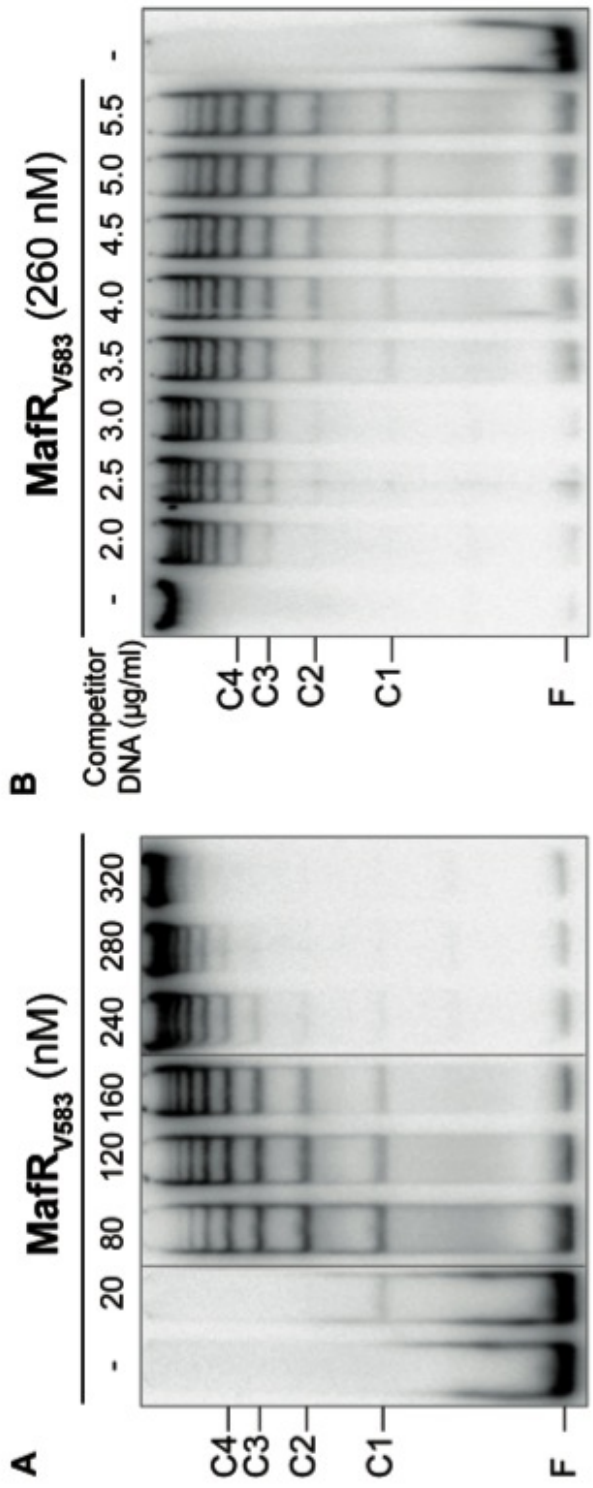

A
MafR$_{V583}$ (nM)
- 20 80 120 160 240 280 320
C4
C3
C2
C1
F
B
MafR$_{V583}$ (260 nM)
Competitor DNA (µg/ml)
- 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 -
C4
C3
C2
C1
F

Figure 5

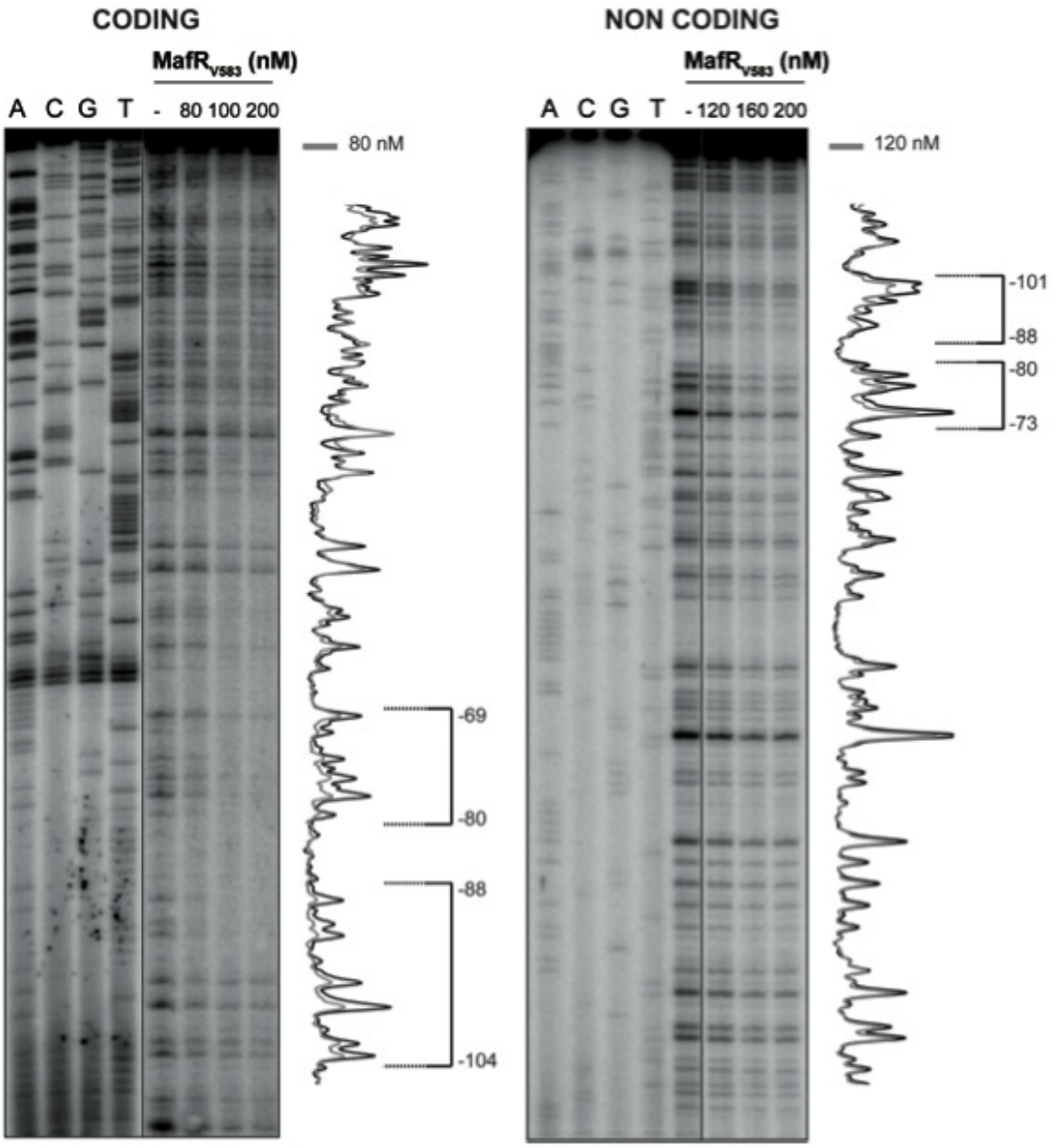

CODING
MafR$_{V583}$ (nM)
A C G T - 80 100 200
80 nM
-69
-80
-88
-104
NON CODING
MafR$_{V583}$ (nM)
A C G T - 120 160 200
120 nM
-101
-88
-80
-73

Figure 6

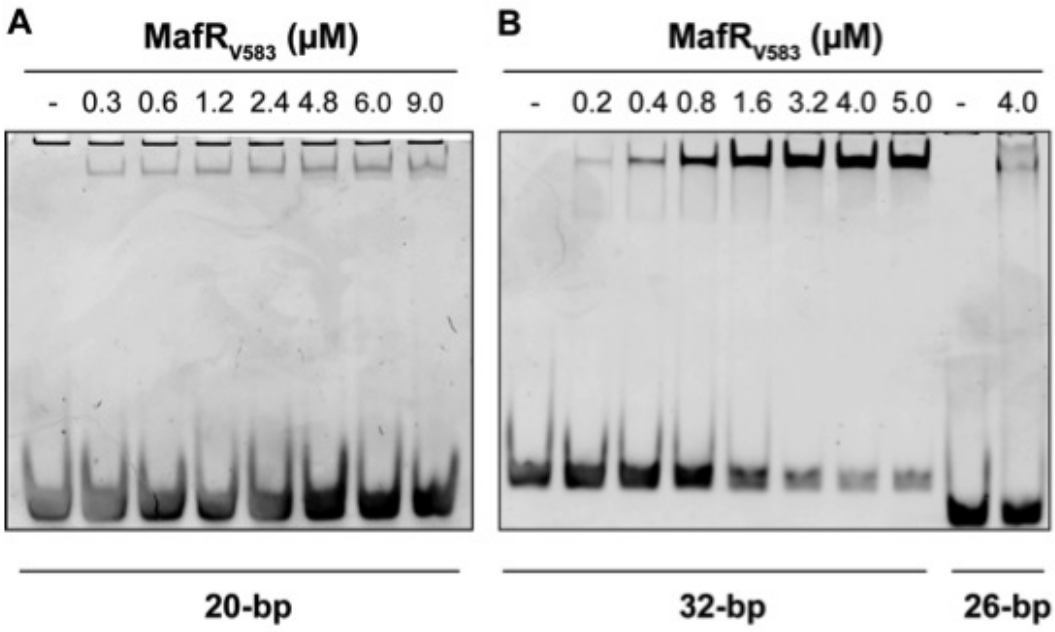
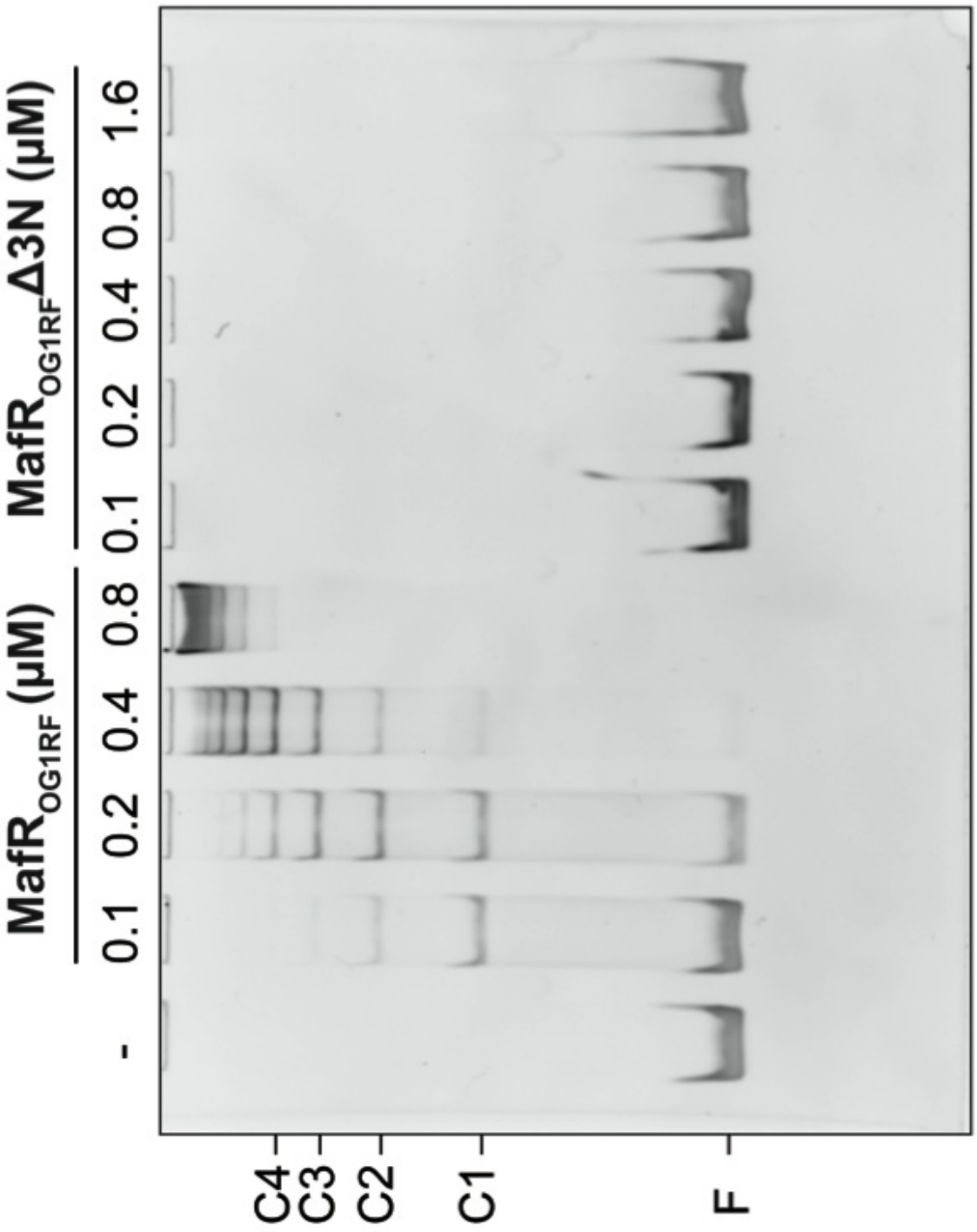

A
MafR$_{V583}$ (μM)
- 0.3 0.6 1.2 2.4 4.8 6.0 9.0
20-bp
B
MafR$_{V583}$ (μM)
- 0.2 0.4 0.8 1.6 3.2 4.0 5.0 - 4.0
32-bp
26-bp

Figure 7


MafR$_{OG1RF}$ (μM)
-
0.1
0.2
0.4
0.8
MafR$_{OG1RF}$Δ3N (μM)
0.1
0.2
0.4
0.8
1.6
C4
C3
C2
C1
F