



HAL
open science

Characterization of the first tetrameric transcription factor of the GntR superfamily with allosteric regulation from the bacterial pathogen *Agrobacterium fabrum*

Armelle Vigouroux, Thibault Meyer, Anaïs Naretto, Pierre Legrand, Magali Aumont-Nicaise, Aurélie Di cicco, Sébastien Renoud, Jeanne Doré, Daniel Lévy, Ludovic Vial, et al.

► To cite this version:

Armelle Vigouroux, Thibault Meyer, Anaïs Naretto, Pierre Legrand, Magali Aumont-Nicaise, et al.. Characterization of the first tetrameric transcription factor of the GntR superfamily with allosteric regulation from the bacterial pathogen *Agrobacterium fabrum*. *Nucleic Acids Research*, 2021, 49 (1), pp.529-546. 10.1093/nar/gkaa1181 . hal-03367389

HAL Id: hal-03367389

<https://hal.science/hal-03367389>

Submitted on 6 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Characterization of the first tetrameric transcription factor of the**
2 **GntR superfamily with allosteric regulation from the bacterial**
3 **pathogen *Agrobacterium fabrum***

4 Armelle Vigouroux^{1†}, Thibault Meyer^{2†}, Anaïs Naretto^{1#}, Pierre Legrand³, Magali Aumont-Nicaise¹,
5 Aurélie Di Cicco⁴, Sébastien Renoud², Jeanne Doré², Daniel Lévy⁴, Ludovic Vial², Céline Lavire^{2*},
6 Solange Moréra^{1*}

7 1 Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-
8 sur-Yvette, France

9 2 Université Claude Bernard Lyon 1, CNRS, INRAE, VetAgro Sup, UMR Ecologie Microbienne, 69622
10 Villeurbanne, France

11 3 Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin, 91192, Gif-sur-Yvette, France

12 4 Laboratoire Physico Chimie Curie, Institut Curie, PSL Research University, CNRS UMR 168, Paris
13 75005, France

14 #Present Address: Anaïs Naretto, Department of Biochemistry, Vanderbilt University, Nashville,
15 Tennessee, 37235, USA

16 † These authors contributed equally to this work

17 * To whom correspondence should be addressed. Tel: +33169824213; E-mail:
18 solange.morera@i2bc.paris-saclay.fr; Tel: +33426237126 celine.lavire@univ-lyon1.fr

19
20 **ABSTRACT**

21 A species-specific region, denoted SpG8-1b allowing hydroxycinnamic acids (HCAs) degradation is
22 important for the transition between the two lifestyles (rhizospheric versus pathogenic) of the plant
23 pathogen *Agrobacterium fabrum*. Indeed, HCAs can be either use as trophic resources and/or as
24 induced-virulence molecules. The SpG8-1b region is regulated by two transcriptional regulators, namely,
25 HcaR (Atu1422) and Atu1419. In contrast to HcaR, Atu1419 remains so far uncharacterized. The high-
26 resolution crystal structures of two fortuitous citrate complexes, two DNA complexes and the apoform
27 revealed that the tetrameric Atu1419 transcriptional regulator belongs to the VanR group of Pfam
28 PF07729 subfamily of the large GntR superfamily. Until now, GntR regulators were dimeric. Here, we
29 showed that Atu1419 represses three genes of the HCAs catabolic pathway. We characterized both
30 the effector and DNA binding sites and identified key nucleotides in the target palindrome. From
31 promoter activity measurement using defective gene mutants, structural analysis and gel-shift assays,
32 we propose N⁵,N¹⁰-methylene tetrahydrofolate as the effector molecule, which is not a direct
33 product/substrate of the HCA degradation pathway. The Zn²⁺ ion present in the effector domain has
34 both a structural and regulatory role. Overall, our work shed light on the allosteric mechanism of
35 transcription employed by this GntR repressor.

37

38

39 INTRODUCTION

40 *Agrobacterium fabrum* has two lifestyles: it can interact with a large variety of plants as a rhizosphere
41 inhabitant or as a pathogen when it harbors a tumor-inducing plasmid (which is a virulence plasmid),
42 and transfers a portion of this to the plant cells upon infection, resulting in the crown-gall disease (1-3).
43 *A. fabrum* possesses a species-specific region, denoted SpG8-1b (a region present in strains of this
44 species but absent from other *Agrobacterium* species), located in the circular chromosome and
45 responsible for hydroxycinnamic acids (HCAs) degradation such as ferulic acid, caffeic acid and *p*-
46 coumaric acid (4,5) (Figure 1). These latter compounds are common plant secondary metabolites being
47 precursors of lignin incorporated into plant cell walls. They are abundantly released in soil during the
48 decay of root cells and are significant environmental molecules for soil-and plant-interacting bacteria
49 (6). Although HCAs are generally a strong bacterial repellent, they appear to be chemoattractants in
50 the case of rhizobia and agrobacteria for which they can be used as trophic resources and/or induced-
51 virulence molecules (7-11). We have previously showed that HCAs degradation *via* the SpG8-1b region
52 interferes with virulence genes expression suggesting that this metabolic pathway is important for the
53 transition between the two lifestyles (rhizospheric versus pathogenic) of *Agrobacterium* (11). Such a
54 transition requires a fine-tune regulation of gene expression to express the appropriate genes at the
55 right time (12-14).

56

57 The SpG8-1b genomic region encodes six enzymes (Figure 1A). Five of them operate in this sequential
58 order for the degradation of ferulic acid leading to different intermediates (5): Atu1416 (a feruloyl-CoA
59 synthase), Atu1417 (an enoyl-CoA hydratase), Atu1415 (a phenylhydroxypropionyl-CoA
60 dehydrogenase), Atu1421 (a 4-hydroxy-3-methoxyphenyl- β -ketopropionyl-CoA β -keto-thiolase) and
61 Atu1420 (a tetrahydrofolate-dependent vanillate O-demethylase). Atu1420 shares 56% sequence
62 identity with *Sphingomonas paucimobilis* and *Sphingobium* sp. SYK-6 homologues of known structures
63 (PDB 5TL4 (15) and PDB 5X1I (16), respectively). This enzyme degrades vanillic acid into
64 protocatechuic acid using the tetrahydrofolate (H4F) cofactor and producing the N5-methyl-
65 tetrahydrofolate compound (MH4F) (5). Protocatechuic acid is then processed by *pca* genes to produce
66 intermediates of the tricarboxylic acid (TCA) cycle (17). Based on sequence identity of 29% with the
67 N5,N10-methylene tetrahydrofolate reductase MetF of *Sphingobium* sp. SYK-6, Atu1418 the sixth
68 enzyme of the SpG8-1b genomic region, is likely a 5,10-methylene tetrahydrofolate reductase which
69 transforms MH4F into N5,N10-methylenetetrahydrofolate (MEF), which would, in turn, allow the
70 regeneration of H4F (4,18) (Figure 1A).

71 Although the SpG8-1b genomic region degrades both ferulic and *p*-coumaric acids, this latter induces
72 the gene expression of the first part of the SpG8-1b region while ferulic acid induces that of the whole
73 pathway (11). Therefore, two expression units are differently regulated suggesting the existence of two
74 regulators. These regulators are Atu1419 (predicted to be a GntR member) and Atu1422 denoted HcaR

75 (for hydroxycinnamic acid catabolic repressor belonging to the MarR family), respectively (4,11) (Figure
76 1B). HcaR is the repressor of its own transcription and that of the first *atu1416* and *atu1417* genes of
77 HCA degradation pathway (11). However, because *atu1415* and *atu1416* belong to the same
78 transcription unit as well as *atu1421* and *hcaR* do, HcaR also regulates *atu1415* and *atu1421* genes
79 expression, totalizing the regulation of five genes (Figure 1B). In contrast to HcaR, the second
80 regulatory protein Atu1419 likely involved in the second part of the pathway corresponding to the
81 vanillate degradation remains uncharacterized so far.

82

83 Herein, we investigated the molecular role and structural aspects of Atu1419 combining *in vitro* and *in*
84 *vivo* approaches. We first proved that Atu1419 was a transcription repressor of three genes of the
85 second part of the HCA degradation pathway. None of HCA degradation intermediates could release
86 Atu1419 from DNA binding. Nonetheless, structural analysis of five high-resolution crystal structures of
87 Atu1419 in apoform, in complex with a fortuitous citrate molecule bound to the effector domain
88 originated from the crystallization condition (2 structures), in complex with DNA and in complex with
89 both DNA and citrate, helped us infer a possible effector molecule, which was confirmed by gel shift
90 assays and microcalorimetry. The structures revealed Atu1419 to be a member of the VanR group of
91 the FadR C-terminal domain (FCD; Pfam PF07729) subfamily of the large GntR superfamily of
92 transcriptional factors (>93135 members in Pfam database) (19-22). The FCD subfamily encompasses
93 two groups of regulators namely FadR and VanR. So far, the dimeric FadR from *Escherichia coli* was
94 the best characterized of the FadR group of FCD-GntR regulators shown to be regulated by the acyl-
95 CoA effector and able to bind specific palindromic DNA through a winged Helix-Turn-Helix (wHTH) motif
96 (23). Like FadR, Atu1419 possesses a characteristic molecular architecture, composed of a conserved
97 N-terminal DNA binding domain containing the wHTH motif and a C-terminal effector
98 binding/oligomerization domain. Unlike FadR, this latter domain of Atu1419 displays six helices instead
99 of seven observed in members of FadR group. Atu1419 is the first example of a transcriptional regulator
100 of the whole GntR superfamily to be tetrameric. Our work brings new insights into mechanistic aspects
101 of such repressor, which uses an induced-allosteric mechanism for DNA release upon effector binding.

102

103 MATERIAL AND METHODS

104 Bacterial strains and growth conditions

105 The bacteria and plasmids used for this study are listed in Supplementary Table S1. *Escherichia coli*
106 were grown routinely, with shaking (150 rpm), at 37°C in LB medium. Growth media were supplemented
107 with appropriate antibiotics (tetracycline, 10 µg/ml; gentamicin, 15 µg/ml; ampicillin, 100 µg/ml) when
108 necessary. The *A. fabrum* strains were grown with shaking (160 rpm), at 28°C in YPG (Yeast Peptone
109 Glucose)-rich medium or in AT minimal medium supplemented with 10 mM succinate and 10 mM
110 ammonium sulfate. AT minimal medium was supplemented with 750 µM of ferulic acid or citrate and
111 with the appropriate antibiotic (gentamicin, 10 µg/ml). Ferulic acid and MH4F were obtained from Sigma
112 Aldrich (St. Louis, USA) and MEF from Merck Company (Switzerland).

113

114 Construction of the deletion mutant C58 Δ *atu1419* and transcriptional fusions

115 The C58 Δ *atu1419* and C58 Δ *atu1420* strains were constructed according to a strategy as described
116 (4,11). Vectors containing the recombinant region (amplified by PCR with specific primers listed in
117 Supplementary Table S2), flanking downstream and upstream of the *atu1419* or *atu1420* genes, were
118 introduced into *A. fabrum* C58 by electroporation. Single-crossover integration was selected by
119 gentamycin resistance on YPG medium plates. Gentamycin-resistant colonies were spread on YPG
120 plates containing 5% of sucrose to obtain plasmid excision and double-crossover events leading to
121 nonpolar mutants. *atu1419* and *atu1420* deletions were confirmed by PCR analysis and DNA
122 sequencing (GenoScreen, Lille, France).

123 pOT1e (24) transcriptional fusions of the promoter regions of SpG8-1b genes, namely *Patu1419* and
124 *Patu1420* were generated as described (11) (Supplementary Table S2 for specific primers). Reporter
125 constructions were introduced into *A. fabrum* C58 wild-type and derivatives by electroporation, and
126 gentamycin-resistant colonies were selected.

127 Measurement of promoter activity

128 Genes expression was measured after 24 hours in *A. fabrum* with a pOT1e plasmid harboring an eGFP
129 transcriptional fusion (24) as described (11). Results were normalized by dividing the fluorescence level
130 by the optical density at 600 nm values. At least five technical replicates and two biological replicates
131 were performed for each condition. Differences between conditions were determined with Tukey test
132 (P-value= 0.05).

133 Oligonucleotides and DNA preparation

134 The synthetic palindromic oligodeoxyribonucleotide 5'-ATGTATACAT-3' was purchased from Sigma-
135 Aldrich (Darmstadt, Germany). Oligonucleotide solution in sterile water at 2 mM was hybridized by

136 heating to 90°C for 5 min and cooling in the crystallization room at 18°C overnight to produce a 10-mer
137 DNA.

138 The *atu1420* promoter regions were amplified with specific primer pairs listed in Supplementary Table
139 S2. The *atu1416-1417*, *atu1418-1419*, *virB* and *hcaR* promoter regions were prepared as previously
140 described (11) and purified with a PCR Clean-up kit (Macherey-Nagel, Düren, Germany).

141 **Cloning, expression and purification of Atu1419 and Atu1419-H3A mutant**

142 Coding sequences for Atu1419 was amplified by PCR adding a C-terminal 6-Histidine tag and using
143 *atu1419F* and *atu1419R* primers and was inserted into the NdeI/XhoI restriction sites of the pET-20b
144 vector (Novagen, Merck Biosciences, France). The nucleotide sequence was confirmed by DNA-
145 sequence analysis (GATC Biotech, Mulhouse, France). *E. coli* BL21 competent cells transformed with
146 pET-20b-Atu1419 were grown in LB media until OD₆₀₀ of 0.8 and protein production was induced by
147 0.5 mM isopropyl β-D-thio-galactopyranoside (IPTG) for 3h at 37°C. Cells were centrifuged at 4000 g
148 for 15 min at 4°C, resuspended in a buffer of 50 mM Tris-HCl pH 8, 300 mM NaCl and 20 mM imidazole
149 and disrupted by sonication. After centrifugation at 25000 g for 30 min at 4°C, the filtrated supernatant
150 was loaded onto a 5 ml His-Trap HP column (GE Healthcare, Chicago, Illinois, USA). After a washing
151 step of 6% with 50 mM Tris-HCl pH 8, 300 mM NaCl and 300 mM imidazole (Buffer B), protein elution
152 was performed with Buffer B. Protein fractions were loaded onto a gel filtration column (HiLoad 26/60
153 Superdex 200 prep grade, GE Healthcare) equilibrated with 50 mM Tris-HCl pH 8 and 150 mM NaCl.
154 The protein fractions were pooled, concentrated and stored at -80°C.

155 The synthetic gene (Genscript, Piscataway, New Jersey, USA) coding for Atu1419 mutant (H192A,
156 H141A and H214A) namely Atu1419-H3A was inserted into pET-20b. The Atu1419-H3A mutant was
157 expressed and purified as the wild-type protein described above.

158 **Size exclusion chromatography (SEC) and SEC-MALS.** For size exclusion chromatography (SEC)
159 and SEC coupled to multi-angle light scattering (SEC-MALS) analyses, apo Atu1419 was prepared at
160 25 μM and Atu1419 in complex with the palindromic DNA was prepared with a ratio protein:DNA of 1:2
161 in a buffer containing 50 mM Tris pH 8 and 150 mM NaCl. The same buffer was used as the mobile
162 phase for SEC using a Superdex 200 10/300 GL column on an AKTA FPLC system (GE Healthcare)
163 and on a Shimadzu HPLC. Multiangle light scattering was detected with a MiniDAWN TREOS light
164 scattering module and a refractometer Optilab T-rEX (Wyatt Technology).

165 **Electrophoretic mobility shift assay (EMSA)**

166 Atu1419 or Atu1419-H3A mutant were mixed with different promoter regions (*Patu1416-1417*,
167 *Patu1418-1419*, *PhcaR*, *Patu1420* and *PvirB*) and with variants of the *atu1418-1419* and *atu1420*
168 regions. The intergenic regions of *virB* was used as a nonspecific control probe. Gel-shift assays were
169 performed in 10 μl reaction mixture containing 30 nM of DNA probe without and with Atu1419 at different
170 concentrations in 50 mM Tris-HCl pH 8 and 150 mM NaCl. 50 to 300 μM of MEF, MH4F, H4F or citrate

171 were added for testing their influence on Atu1419-DNA complex formation. After incubation for 30 min
172 at room temperature, the samples were separated by electrophoresis in TBE buffer (45 mM Tris-HCl
173 pH 8, 45 mM boric acid and 1 mM EDTA) on non-denaturing 6% or 12% polyacrylamide gels at 150 V
174 and 4°C for 2 h. Gels were then stained with either SYBR® Green EMSA nucleic acid gel stain
175 (Invitrogen, Carlsbad, CA, USA) or ethidium bromide for 20 min. DNA was visualized under UV light
176 (Fisher Bioblock Scientific, Illkirch, France or UVP BioDoc-it2 Imager, Analytic Jena, Germany).

177

178 **Crystallization and structure determination of Atu1419**

179 Crystallization conditions are summarized in Table 1. For all protein samples (140 µM tetrameric protein
180 alone or in complex with 700 µM palindromic DNA), conditions were screened using QIAGEN kits
181 (Valencia, CA, USA) with a Mosquito nanodrop robot (TTP Labtech, Melbourn, Great Britain) and were
182 manually optimized at 20°C in hanging drop by mixing equal volumes of the protein or protein-DNA
183 solution with precipitant solution. Crystals were transferred to a cryoprotectant solution (paraffin oil or
184 mother liquor supplemented with 20% PEG 400) and flash-frozen in liquid nitrogen. Diffraction data
185 were collected at 100 K on PROXIMA 1 and PROXIMA 2 beamlines at synchrotron SOLEIL (Saint-
186 Aubin, France). Intensities were integrated using the XDS package (25) (Table 1).

187 The first structure of Atu1419 in $P2_12_12_1$ space group was determined at 2.35 Å resolution by single-
188 wavelength anomalous dispersion (SAD) method at the peak absorption energy of Zinc (Table 1). The
189 presence of a zinc metal in crystals was found thanks to an X-ray fluorescence emission scan on the
190 beamline. The steps of zinc ion substructure determination, phases calculation and density modification
191 were performed using CRANK from CCP4 (Collaborative Computational Project, Number 4) and a
192 partial model was built using BUCCANEER (CCP4). This model was then used to calculate the phases
193 of a higher resolution dataset at 2 Å resolution leading to the complete polypeptide chain model. A
194 tetramer is present in the asymmetric unit. Other structure determinations were performed by molecular
195 replacement with PHASER (26) using the first refined structure of Atu1419 (monomer, dimer or
196 tetramer). Because of the anisotropy of the diffraction of Atu1419-DNA complex in $P6_422$ crystals and
197 apo Atu1419, the DEBYE and STARANISO programs developed by Global phasing Ltd were applied
198 to the data scaled with XDS using the STARANISO server (<http://staraniso.globalphasing.org/>). These
199 programs perform an anisotropic cut-off of merge intensity data on the basis of an analysis of local $I/\sigma(I)$,
200 compute Bayesian estimates of structures amplitudes, taking into account their anisotropic fall-off, and
201 apply an anisotropic correction to the data. The corrected anisotropic amplitudes were used for further
202 refinement. Refinement of each structure was performed with BUSTER-2.10 (27) employing TLS
203 groups and NCS restraints. Inspection of the density maps and manual rebuilding were performed using
204 COOT (28). Refinement details of each structure are shown in Table 1. Molecular graphics images were
205 generated using PyMOL (<http://www.pymol.org>).

206 **Circular dichroism experiments (CD)**

207 Circular dichroism in the far-UV region was performed using a spectropolarimeter (Jasco J-810, Jasco,
208 Lisses, France) equipped with a water-cooled Peltier unit (Jasco circular dichroism spectrometer model
209 J810). Apo Atu1419 and Atu1419-H3A mutant were both concentrated at 25 μM in 50 mM Tris pH 8
210 and 150 mM NaCl. Spectra were recorded in a cell width of 0.1 mm path length (121.QS, Hellma, Hellma
211 Analytics, Müllheim, Baden Württemberg, Germany) from 190 to 240 nm at 20°C. Five consecutive
212 scans from each sample were merged to produce an averaged spectrum; the spectra were corrected
213 using buffer baselines measured under the same conditions. Data were recorded in mdeg and
214 converted using the mean residues ellipticity method ($\text{deg}\cdot\text{cm}^2\cdot\text{mol}^{-1}$). Secondary structure estimates
215 were derived from the normalized spectra using the CDSSTR, SELCON3, CONTIN of the DICHROWEB
216 server (29,30).

217 **Differential scanning calorimetry (auto PEAQ DSC)**

218 Thermal stability of 10 μM apo Atu1419, 10 μM Atu1419 in the presence of 100 μM MEF and 20 μM
219 Atu1419-H3A mutant was performed by DSC on an auto PEAQ DSC (Malvern, France) in a standard
220 buffer. Each measurement was preceded by a baseline scan with the standard buffer. Scans were
221 performed at 1 $\text{K}\cdot\text{min}^{-1}$ between 20 and 90°C. The heat capacity of the buffer was subtracted from that
222 of the protein sample before analysis. Thermodynamic parameters were determined by fitting the data
223 to the following equation:

$$224 \quad \Delta C_p(T) = \frac{K_d(T) \Delta H_{cal} \Delta H_{vH}}{[1 + K_d(T)]^2 RT^2}$$

225 where K_d is the equilibrium constant for a two-state process, ΔH_{vH} is the enthalpy calculated on the
226 basis of a two-state process and ΔH_{cal} is the measured enthalpy.

227 **Isothermal titration microcalorimetry measurements (ITC)**

228 Isothermal titration microcalorimetry experiments were performed with an ITC200 isothermal titration
229 calorimeter from MicroCal (Malvern, Orsay, France). The experiments were carried out at 20°C. Protein
230 concentration in the microcalorimeter cell (0.2 ml) was 100 μM . Nineteen injections of 2 μl MEF solution
231 at 1.2 mM were performed at intervals of 180 s while stirring at 500 rpm. The experimental data were
232 fitted to theoretical titration curves with software supplied by MicroCal (ORIGIN®).

233 **Electron Microscopy**

234 Atu1419 (150 nM) was incubated with equimolar *Patu1418-1419* region (370 bp containing two
235 palindromes separated by 190 bp) for 10 min at room temperature in a buffer containing 50 mM Tris pH
236 8 and 150 mM NaCl. Samples were deposited on a glow-discharged carbon coated grid and stained
237 with 2% uranyl acetate. Images were recorded with a Lab6 Tecnai Spirit operating at 80kV and a
238 Quemesa Olympus CCD camera. Nominal magnification was X 45 000 corresponding to 3.17 Å/pixel.

239

240 RESULTS AND DISCUSSION

241 The transcriptional regulator *Atu1419* is the second repressor of the HCAs degradation pathway

242 To study the role of *Atu1419* in transcriptional regulation of the HCAs degradation genes, we
243 constructed reported fusions with pOT1e plasmids containing each promoter region of the SpG8-1b
244 region (Figure 1) cloned upstream the *egfp* gene. This allowed reporting the transcription of each gene
245 by measuring the eGFP fluorescence. The *atu1419* and *atu1420* plasmid reporter gene fusions were
246 constructed in this study. Those corresponding to *atu1416* and *atu1418* were already available (11).
247 The plasmid reporter fusions were introduced into the wild-type *A. fabrum* C58 strain to evaluate gene
248 expression in response to ferulic acid. Compared to the fluorescence level measured in the absence of
249 ferulic acid, the fluorescence level measured in the presence of ferulic acid was higher for the four
250 reporter fusions (Table 2). Hence, *atu1416*, *atu1418*, *atu1419* and *atu1420* genes were induced in the
251 presence of ferulic acid. The four plasmid reporter fusions were also introduced into the *C58Δatu1419*
252 mutant strain to compare genes expression with that of the wild-type C58 strain in the absence of ferulic
253 acid. The fluorescence level measured after 24 hours for *atu1418*, *atu1419* and *atu1420* reporter
254 fusions were respectively 3.67, 3.92 and 1.58 times higher in the *C58Δatu1419* mutant strain than in
255 the wild-type strain (Table 2). Thus, in the strain lacking *Atu1419* regulatory protein (*C58Δatu1419*), the
256 reporter fusions for *atu1418*, *atu1419* and *atu1420* genes were constitutively expressed. In contrast,
257 the fold change between *C58Δatu1419* and the wild type strains was 0.63 and 0.97 for the *atu1416*
258 gene fusion and the empty pOT1e vector, respectively indicating that *atu1416* gene expression was not
259 repressed by *Atu1419* (Table 2). These results showed that *Atu1419* was a transcriptional repressor
260 that regulates *atu1418*, *atu1420*, and its own transcription.

261 DNA-binding targets

262 To characterize the target DNA sequence of *Atu1419*, we purified *Atu1419* (theoretical molecular weight
263 of ~26 542 Da for a monomer), which appeared as a tetramer in solution according to gel filtration/SEC-
264 MALS chromatography (Supplementary Figure S1). Using the intergenic regions (between 300 and
265 400-base pair (bp)) of *atu1416-1417*, *atu1418-1419*, *atu1420*, *hcaR* and *virB* (negative control gene
266 outside the SpG8-1b region) as DNA probes for binding assays (Figure 2A), we showed that *Atu1419*
267 bound to the *atu1418-1419* and *atu1420* promoter regions only, in agreement with the gene expression
268 data described above.

269 In silico comparison of the promoter regions of *atu1418-1419* and *atu1420* using BPROM program (31)
270 revealed three 10-mer identical palindromic sequences (5'-ATGTATACAT-3'), two in *atu1418-1419*
271 promoter region and one in *atu1420* promoter region (Figure 2B). A palindromic sequence overlaps the
272 -10 box regulatory element of *atu1418* and *atu1420*, whereas in the *atu1418-1419* promoter sequence,
273 an additional palindromic sequence is found between the -10 and the -35 regulatory elements of the
274 *atu1419* (Figure 2B). Using electrophoretic mobility shift assay (EMSA), we showed that *Atu1419* was
275 able to bind each palindrome of the *atu1418-1419* promoter sequence without the presence of the other
276 palindrome meaning that only one palindrome site was required for binding (Figure 2C). Nonetheless,

277 the intensity of the shifted bands was greater for the DNA fragment containing both palindromes (Figure
278 2C). We also analyzed the quaternary structure of Atu1419 in complex with the 10-bp palindrome using
279 gel filtration/SEC-MALS measurements (Supplementary Figure S1). Atu1419 was also tetrameric upon
280 DNA binding.

281 **Crystal structures of Atu1419 in complex with a fortuitous ligand**

282 We solved the first structure of Atu1419 from crystals grown in sodium citrate buffer using SAD method
283 at the peak absorption energy of Zinc at 2 Å resolution in the space group $P2_12_12_1$ (Table 1). The zinc
284 ion comes from *E. coli* protein expression because no metal was added during protein purification and
285 crystallization. Four molecules (A, B, C and D) are present in the asymmetric unit and form a tetramer
286 (a dimer of dimer) (Figure 3A), which is consistent with the observation that Atu1419 is a tetramer in
287 solution (Supplementary Figure S1). Each monomer consists of an N-terminal DNA binding domain
288 (residues 1-71) and a C-terminal all α -helical effector binding domain (residues 76-244) (Figure 3B).
289 The DNA binding domain is composed of three helices ($\alpha 1$ - $\alpha 3$) with helix $\alpha 2$ and helix $\alpha 3$ forming the
290 helix-turn-helix motif and two anti-parallel β -strands ($\beta 1$ - $\beta 2$) connected by a small loop designated as
291 the wing motif. The secondary element with six helices ($\alpha 4$ - $\alpha 9$) and topology of the C-terminal domain
292 place Atu1419 in the VanR group of the FCD subfamily of GntR transcriptional regulators.
293 Approximately 2776 Å² of accessible surface area is buried upon dimerization for dimers AB or CD
294 corresponding to an average of 11.7% of the total surface area of each monomer and 41 amino acids
295 per monomer. The interface between dimers AB and CD involves side chains of both domains: helix $\alpha 3$
296 and strand $\beta 1$ (residues 49-67) of the DNA binding domain and the three helices $\alpha 4$ (residues 78-100),
297 $\alpha 7$ (residues 150-163) and $\alpha 9$ (residues 208 and 212) of the effector binding domain (Figure 3B). Arg72
298 in the linker region participates to the dimer interface. Two salt bridges with Glu57-Arg63' (the prime
299 refers to the second monomer in the dimer), and two polar interactions Glu57-Ser53' and Asn62-Arg155'
300 are located in the N-terminal domain whereas fourteen H-bonds/salt-bridges including Asp81, Glu84,
301 Arg95, Arg100, Ser150, Arg155, Glu156, Glu208 and Arg212 belong to the C-terminal domain. The
302 contacts of dimer AD through eleven polar interactions/salt bridges bury 686 Å² per subunit and concern
303 23 amino acids, which are located in the loop between helices $\alpha 5$ - $\alpha 6$ (mainly residues 123-125), helices
304 $\alpha 8$ (residues 183-194) and $\alpha 9$ (residues 203-216) (Figure 3B). The dimer BC displays a smaller
305 interface of 557 Å² per subunit comprising 21 amino acids and the same structural elements as for
306 dimer AD. The total surface contact area between two neighbouring subunits within the tetramer is
307 characteristic of biological interactions (32).

308 Remarkably, the four subunits within the tetramer are not identical: molecule A shows large
309 conformational changes as indicated by an average root mean square deviation (RMSD) of 1.28 Å for
310 all C α atoms compared with the subunits B/C/D, which are more similar (average RMSD of 0.7 Å for all
311 C α atoms). This is due to the presence of a citrate originated from the crystallization condition, which
312 binds molecule A in the effector binding site in a different orientation (orientation A; Figures 3C and 3D)
313 than that for citrates in subunits B/C/D (orientation B; Figures 3C and 3E), which bind similarly (Figure

314 3C). Four loop regions (shown in red in Figure 3C), comprised of residues 64-67 of the wing between
315 β 1 and β 2 of the wHTH motif, residues 119-130 between helices α 5 and α 6, residues 169-184 between
316 helices α 7 and α 8 and residues 225-244 corresponding to the end of the effector binding domain, can
317 move between 3 and 10 Å. The loop region 169-184 can drastically rearrange upon ligand binding. In
318 molecule A, the C α atoms of Ser169, Lys175 and Arg183 are respectively, 0.92 Å, 6.84 Å and 7.34 Å
319 away from those in molecules B/C/D, allowing the side chains of Lys175, Arg183 and the NH main
320 chain of Ser169 to interact with the citrate molecule (Figures 3C and 3D). These latter protein-citrate
321 interactions cannot exist in subunits B/C/D. Nonetheless, all bound citrates share interactions with
322 Arg86, Tyr133, Asn137, His141, His214, Asn221 and a Zn²⁺ ion that co-purified with Atu1419 (Figures
323 3D and 3E). Thirteen polar interactions are observed between monomer A and the citrate molecule
324 (orientation A), which is buried within the monomer, leaving only 26.6 Å² or 8.4% of the molecule surface
325 exposed to solvent. In contrast, there are only ten polar interactions between monomers B/C/D and the
326 citrate with a similar buried surface area.

327 The orientation B of the citrate was also observed in another structure of Atu1419 solved at a higher
328 resolution of 1.75 Å but in a different space group (*P*2₁2₁2). Here, two similar monomers (RMSD of 0.98
329 Å for all C α atoms) are in the asymmetric unit and form a dimer, which in turn form a tetramer by the
330 crystal symmetry with globally unchanged dimer interfaces (1553 Å² and 701 Å² per subunit for dimers
331 AB and AD, respectively) compared with those from the *P*2₁2₁2₁ structure. They resemble subunits
332 B/C/D of the *P*2₁2₁2₁ structure and bind similarly a citrate molecule. Nonetheless, their citrates are
333 slightly shifted by 0.6 Å towards Arg86 losing the interaction with Asn221 compared with those in the
334 *P*2₁2₁2₁ structure (Supplementary Figure S2).

335 **Crystal structures of Atu1419 in complex with the palindromic DNA**

336 The structures of two palindromic DNA-Atu1419 complexes solved at 2.79 Å resolution (*P*6₄22 space
337 group) and 2.05 Å resolution (*C*222₁ space group) present a distinct neighboring crystal packing with
338 an asymmetric unit containing a monomer bound to a single DNA strand and a dimer bound to the 10-
339 mer palindromic DNA, respectively (Table 1, Supplementary Figure S3). Nonetheless, a tetramer,
340 where dimers AB and CD bind each to a DNA palindrome, was reconstituted by crystal symmetries
341 (Figures 4A and 4B) in line with the conservation of Atu1419 tetramer in solution upon DNA binding.
342 The three monomers of the asymmetric units (one in *P*6₄22 and two in *C*222₁) are similar with an
343 average RMSD value of 1 Å for all C α atoms making almost identical interactions with DNA, although
344 the positions of the DNA binding domains within the dimers between the two crystal structures do not
345 completely overlap (Figure 4C). This observation explains the different crystal packing resulting of
346 flexibility from both the DNA ligand and the DNA binding domains of Atu1419. Each monomer
347 recognizes a half-site DNA with helices α 2 and α 3 of the HTH motif making nine polar interactions with
348 both strands in the major groove of the DNA (Figures 4D and 4E). Helix α 3 via Ser44, Thr46, Arg45
349 and Arg49 side chains is responsible for six hydrogen bonds with both strands, half with oxygens of
350 phosphate groups 5' of guanine (G3) at position 3 on one strand and of adenine (A5) at position 5 on

351 the complementary strand. The other half is made by the guanidinium group of Arg45, which provides
352 specific contacts with the N7 and O6 atoms of G3 on one strand and by OG1 of Thr46 with the N7 of
353 A5 on the other strand (Figure 4C). The remaining H-bonds come from helix $\alpha 2$ via the Arg31 side chain
354 and the main chain amino group of Glu34, and an additional interaction is present with the main chain
355 NH of His9 in helix $\alpha 1$. They consist of phosphate contacts with A1 and T2 on one DNA strand and T4
356 on the other.

357 All Atu1419-DNA subunits contain a bound Zn^{2+} ion at the same position as that observed in both
358 Atu1419-citrate complexes. In contrast to the protein-DNA complex in $C222_1$ space group, which
359 crystallized in MES buffer, the $P6_422$ protein-DNA complex reveals a bound citrate in the effector
360 binding pocket. This citrate adopts the orientation B described in both Atu1419-citrate complexes,
361 except for molecule A (orientation A) in the $P2_12_12_1$ structure. Because both tetrameric Atu1419-DNA
362 structures are similar, the presence of the bound citrate in orientation B in all subunits of one DNA
363 complex has clearly no effect on Atu1419 for DNA binding and DNA release, in agreement with EMSA
364 assay (Figure 4G) and promoter activity (Table S3).

365 We modified the specifically recognized guanine base G3 by an adenine base and its partner base C8
366 by a thymine on each strand within the unique palindrome of the *atu1420* promoter region. Gel-shift
367 assays showed that Atu1419 was no longer able to bind *Patu1420s* AT mutant (Figure 4F). Therefore,
368 the guanine at position 3 has an essential role in the DNA palindrome binding by Atu1419 meaning that
369 Arg45 has a key role in recognizing this guanine base.

370 **Crystal structure of apo Atu1419 and structural comparison**

371 We solved the structure of Atu1419 in the apoform at 2.7 Å resolution using Tris-HCl buffer instead of
372 citrate buffer in the crystallization condition to avoid any bound citrate (Table 1). The asymmetric unit
373 contains a tetramer with a bound Zn^{2+} ion in each effector binding domain. Asn137 and three histidine
374 (His141, His192 and His214) side chains form the metal binding site of Atu1419 and chelate the zinc
375 ion with a classical tetrahedral coordination geometry and average distances around 2.2 Å (Figure 5A).
376 We mutated the three histidine (His141, His192 and His214) into alanine to make the triple point mutant
377 Atu1419-H3A and alter the metal binding site. The zinc ion holds helices $\alpha 6$, $\alpha 8$ and $\alpha 9$ involved in the
378 tetrameric interface far from the DNA binding domain, and likely interacts with the effector as shown by
379 the structures solved with a citrate taking the place of the effector. Atu1419-H3A is correctly folded as
380 checked by using circular dichroism (Supplementary Figure S4) and shares the same secondary
381 structure content (about 58% α -helices and 1% β -sheets) as the wild-type protein in line with what is
382 observed in all crystal structures of Atu1419. Similar to the wild-type protein, this mutant is tetrameric
383 and capable of binding *atu1418-1419* and *atu1420* promoter regions.

384 The dimeric (dimers AB/CD) and tetrameric interfaces (dimers AD/BC) covering around 1400 Å² and
385 516 Å² per subunit, respectively, are comparable to those of Atu1419-citrate complexes and Atu1419-
386 DNA complexes. Subunits A and D overlap well (RMSD of 0.53 Å), as well as subunits B and C (RMSD

387 of 0.83 Å), whereas subunits A and B or C and D show a RMSD over 2 Å, mainly due to differences in
388 the position of their HTH motif and the end of their C-terminal domain, evidencing large flexibility
389 (Supplementary Figure S5A). This is confirmed by the structural comparison of the DNA binding
390 domains of the apo dimer AB with those of the Atu1419-DNA complex (Figure 5B). While subunits A
391 are rather similar, helix α 3 of the apoform subunit B clashes into the DNA sugar-phosphate backbone.
392 Indeed, this helix is far from its optimized position for DNA binding, which is perpendicular to the helical
393 axis of the DNA. A displacement up to 7 Å would be required to correctly places its wHTH motif. A
394 similar structural analysis between the DNA binding domains of the Atu1419-citrate complexes and
395 Atu1419-DNA complexes led to similar conclusions (Supplementary Figures S5B-D).

396 It is noteworthy that the conformation of the effector binding pocket in all subunits bound or not to DNA
397 and containing a citrate in orientation B is similar to that of the empty pocket observed in the apo form
398 and one DNA complex. This suggests that this citrate may mimic a non-productive-like binding in the
399 effector binding site.

400

401 **Comparison with other FCD members**

402 Atu1419 belongs to the VanR subgroup of the FCD subfamily of the large GntR superfamily due to the
403 number of helices in its effector binding domain. The FCD subfamily is divided into two groups; VanR
404 and FadR. Members that are structurally characterized in FadR subgroup possess an additional helix
405 after the linker region at the beginning of the effector binding domain (23,33-35). This major difference
406 leads to a domain-swapped quaternary structure, in which the DNA binding domain of FadR subunit A
407 crosses the dimerization interface to be in contact with the effector binding domain of subunit B (35-37).
408 This is not observed in the VanR subgroup where both domains from the same subunit are associated.

409 Atu1419 is now the seventh VanR member to be structurally characterized. The six others are
410 *Thermotoga maritima* TM0439 (PDBs 3SXY/3FMS (38)), *Ralstonia eutropha* JMP134 YP_298823.1
411 (PDB 3IHU no related publication), *Pseudomonas syringae* PS5454 (PDB 3C7J, no related publication),
412 *Rhodococcus* sp. *RHA1* (PDB 2HS5, no related publication), *Escherichia coli* McbR/YncC (PDB 4P9F
413 (37)) and *Listeria monocytogenes* MouR (PDB 6EP3 (39)). Structural comparison using SSM-EBI
414 (<http://www.ebi.ac.uk/msd-srv/ssm>) reports RMSD of 3.19 Å over 190 C α atoms/3.46 Å over 184 C α
415 atoms between the full-length of Atu1419 and MouR (PDB 6EP3)/TM0439 (PDB 3SXY), respectively.
416 The same search using the C-terminal domain of Atu1419 only improves the RMSD values to 2.59 Å
417 and 2.71 Å over 127 residues with the effector binding domains of TM0439 and MouR corresponding
418 to 15% and 13% sequence identity, respectively (Supplementary Figure S6A). The best match concerns
419 the effector binding domain of McbR/YncC (PDB 4P9F) with a RMSD of 2.23 Å over 132 residues and
420 18% sequence identity, meaning that there are major significant differences between Atu1419 and the
421 six other VanR members. While only the C-terminal domain is responsible for dimerization for most
422 VanR regulators (37), Atu1419 is similar to McbR/YncC in using both the N- and C-terminal domains

423 as the dimeric interface. However, the dimer of Atu1419 does not structurally resemble that of
424 McbR/YncC (Supplementary Figure S6B).

425 Like Atu1419, two other VanR members, which are TM0439 and PS5454 display a metal binding site
426 with three histidine residues and an asparagine or an aspartate residue involved in the metal
427 coordination (37,38). These histidine residues are conserved and structurally close (Supplementary
428 Figure S6C). LldR (PDB 2DI3) from the FadR group binds a zinc ion chelated by these three conserved
429 histidines and an aspartate (35). It was proposed that an additional subgroup within the FCD subfamily
430 could be created for regulators from both FadR and VanR members, which are capable of binding to
431 metal ions (35,38). Using thermal denaturation experiments by differential scanning calorimetry (DSC),
432 T_m of 52.6°C and 34°C were measured with the wild-type Atu1419 and Atu1419-H3A mutant
433 respectively (Supplementary Figure S7). The drastic difference of almost 20°C indicates that the metal
434 assists in stabilizing the structure affording an explanation as to why Atu1419-H3A could not be
435 concentrated over 50 μ M without precipitating, in contrast to the wild-type protein (140 μ M). All attempts
436 to crystallize Atu1419-H3A were unsuccessful.

437 Atu1419 is the first VanR member to be structurally characterized in complex with DNA whereas several
438 FadR-DNA complexes are available (23,33,34). Each dimer of Atu1419 within the tetramer can bind an
439 identical palindrome. Similarly to regulators with a WHTH motif such as MarR regulators (40,41), FCD
440 regulators bind to one half-site of the palindromic DNA, with the dimerization interface helping to
441 establish the spacing between the two half-sites. The DNA binding domains of Atu1419, which are
442 involved in the dimerization interface, allow Atu1419 to bind a short palindrome of 10 base pairs. As in
443 *E. coli* FadR (PDB 1H9T (23) and 1HW2 (33)) and in *Vibrio cholerae* FadR (PDB 4P9U(34)), the DNA-
444 binding site of each Atu1419 monomer is recognized by conserved residues such as Glu34 in helix α 2,
445 Arg45 (specific DNA contacts), Thr46 and Arg49 (phosphate backbone contacts) in helix α 3 that interact
446 with the major groove and Gly66 in the tip of the wing (residues 64-68) between β 1 and β 2. Nonetheless,
447 the wing is not crucial for DNA interaction of Atu1419 unlike FadR, which recognizes a pseudo-
448 palindrome of 17 base pairs. Indeed, His65, which specifically recognizes a DNA base in FadR is a
449 glycine in Atu1419. DNA binding domains superposition of the Atu1419-DNA complex with those of *E.*
450 *coli* and *V. cholerae* FadR-DNA complexes shows that although the positions of subunits A are rather
451 similar, the second subunit of Atu1419 makes steric clashes with the DNA bound to FadR and that of
452 FadR is too far from the DNA bound to Atu1419 (Supplementary Figure S8). Atu1419 displays only one
453 specific interaction shared in FadR with the guanine present within the major groove of DNA *via* their
454 conserved Arg45, a typical feature of the GntR family (22). The other specific DNA contacts of FadR
455 (*via* Arg35, Thr44, Thr46 and His65) are not comparable with those of Atu1419. The recognized DNA
456 sequence of Atu1419 which is 5'-ATGTATACAT-3' is in agreement with the predicted DNA signature
457 for the GntR family: 5'-(N)_yGT(N)_xAC(N)_y-3' where the number x and y vary (20). Later, this DNA binding
458 signature was suggested to be modified as 5'-TNG(N)_nCNA-3' on the basis of base-specific interactions
459 (22) or as 5'-TGGTN_xACCA-3' for FadR subgroup (35) but both sequences are not appropriate for
460 Atu1419.

461 The DNA binding of Atu1419 requires conformational changes of the DNA binding domains. Indeed,
462 the distance between the two DNA recognition helices (Thr46 of helix α_3) narrows from 12 Å in the
463 apoform to 9.3 Å in the Atu1419-DNA complex, for a productive interaction with DNA (Figure 5C). A
464 similar observation was reported for *Vibrio cholerae* FadR (PDB 4P9U (34)). The structures of FadR
465 repressors in complex with their effector show significant conformational changes transmitted from the
466 effector binding domain to the DNA binding domain, leading to a conformational state that is no longer
467 favorable for interaction with DNA (22,23,34). The fundamental process wherein the binding of a ligand
468 or effector molecule alters the activity of the protein at a distant site is defined as an allosteric
469 mechanism. In the case of Atu1419, the citrate in orientation A bound in the effector binding pocket
470 induces large conformational changes from the C-terminal domain to the N-terminal domain. It affects
471 four different protein regions: the wing of the DNA binding domain and three regions of the effector
472 binding domain including the C-terminus protein which is in contact with the DNA binding domain
473 (Figure 3C). The presence of this citrate clearly reveals the plasticity of the effector binding pocket and
474 the large potential rearrangement of the repressor. It also shows that within the tetramer, one subunit
475 can adopt a different conformation around the effector binding pocket and further away (Figure 3C)
476 compared to the three others subunits, highlighting an allosteric mechanism. The orientation A of the
477 citrate may mimic an active-like position of the physiological effector part bound to the metal ion.

478 **Atu1419 effector is the N5,N10-methylenetetrahydrofolate (MEF)**

479 The effector molecule bound to the members of the GntR superfamily are often related to catabolic
480 substrates or intermediates of the pathway controlled by the transcription factor (19,20), and several
481 molecules are produced during the ferulic acid degradation pathways (5). Because the citrate molecule
482 was able to bind to the effector binding site of Atu1419, we searched for molecules resembling citrate
483 among those described in the HCA pathway (Figure 1A), and we asked whether the substrate-cofactor
484 H4F or the product-cofactor M4HF of the O-demethylase Atu1420 enzyme could be a potential effector.
485 Indeed, citrate compound with three carboxylate groups mimics two carboxylate groups located at one
486 end of the H4F or MH4F molecules. Alternatively, MEF, the putative product of Atu1418 enzyme, which
487 possesses the same two carboxylate groups could also be a potential effector of Atu1419.

488 To verify this, *Patu1418*, *Patu1419* and *Patu1420* plasmid reporter gene fusions were introduced into
489 the *C58 Δ atu1418* and *C58 Δ atu1420* mutant strains. For each strain, the fluorescence level was
490 monitored in the presence or absence of ferulic acid and compared to the wild-type *A. fabrum* C58
491 (Figures 6A-C). For the *C58 Δ atu1418* strain, the overproduction of fluorescence observed in the
492 presence of ferulic acid is significantly lower than that in the wild-type C58 and *C58 Δ atu1420* strains.
493 These results indicate that the presence of *atu1418* gene is important for the full induction of *atu1418*,
494 *atu1419* and *atu1420* genes expression in the presence of ferulic acid. The accumulation of MH4F and
495 the lack of MEF production in the defective *C58 Δ atu1418* mutant suggests that MEF could be the
496 effector of Atu1419. MEF is not stable in aqueous *in vitro* solutions as it enters a rapid reversible
497 equilibrium into H4F (42,43). Nonetheless, immediately using freshly-made MEF solution for gel-shift
498 assays, we were able to show that MEF could relieve repression of Atu1419 (Figure 6D) whereas H4F

499 and MH4F had no effect (Figure 6E). In the C58 Δ *atu1418* mutant, induction of *Patu1418*, *Patu1419* and
500 *Patu1420* was not totally abolished, probably due to functional redundancy of N5,N10-methylene
501 tetrahydrofolate reductase activity in *A. fabrum* genome, which allows to keep a basal pool of folate and
502 intermediates essential for the cell (44).

503 The interaction between MEF and Atu1419 was confirmed by DSC with a T_m of 63.5°C for Atu1419 in
504 complex with MEF compared to the T_m of 52.6°C for the apo Atu1419 (Figure 6F). This was also
505 confirmed by isothermal titration microcalorimetry (ITC, Figure 6G). Because MEF instability did not
506 allow a return to the baseline between two injections, we could not rigorously determine a dissociation
507 constant K_D value, which is around 20 μ M. Injection of MEF towards buffer solution was a control of its
508 instability effect (Figure 6H). No interaction could be observed between MH4F and Atu1419 by DSC
509 and ITC, in agreement with the results from EMSA assays.

510 **Atu1419 tetramer is the biologically active form**

511 We analyzed the quaternary structure of Atu1419 in complex with MEF using gel filtration measurement
512 (Supplementary Figure S9). Atu1419 remains tetrameric when bound to MEF. Thus, Atu1419 is a
513 tetramer in its apo form, upon DNA binding or effector binding suggesting that this quaternary structure
514 is the biologically active form. Both crystal structures of Atu1419 in complex with DNA revealed that
515 each dimer of Atu1419 within the tetramer can bind a palindromic site of 10 bp, separated by
516 approximately 10 nM (Figure 3A). This separation distance is compatible with DNA loop formation
517 allowing the simultaneous binding of two distant palindromes (190 bp apart) within the intergenic region
518 of *atu1418-1419*. Thus, Atu1419 tetramer could repress gene expression *via* DNA looping, such as
519 several negative regulators in prokaryotes (47) with the example of the intensively studied tetrameric
520 *lac* operon repressor (48).

521 We searched for a second potential DNA binding site of Atu1419 in *atu1420* region and found a
522 degenerated palindrome (8 bp conserved over the 10 bp palindrome; Figure 7A). The degenerated
523 palindrome is separated by 109 bp from the palindrome and is located downstream within the open
524 reading frame of *atu1420* gene (Figure 7A). We then performed EMSA assays with three
525 oligonucleotides containing either the palindrome, the degenerated palindrome, or both and identified
526 in the gel a retarded band for each oligonucleotide proving the formation of a stable non-covalent
527 protein-DNA complex (Figure 7B). Thus, the repressor can recognize a palindrome and a degenerated
528 palindrome providing it the opportunity to bind simultaneously to two DNA sites in each region (*atu1418-*
529 *1419* and *atu1420*), in agreement with its oligomeric state. Similarly, the structure of *Cupriavidus*
530 *necator* CbnR transcriptional regulator (a tetramer composed of dimer of dimer) from LysR family (PDBs
531 1IXC and 1IZL (45)) is compatible with an interaction with two DNA binding sites on a bended DNA
532 fragment, where the authors suggested that the quaternary structure of the CbnR tetramer should be
533 changed to relax the bent DNA that binds to the CbnR tetramer (45). In contrast to the quaternary
534 organization of Lac repressor (a tetramer comprising a pair of dimers loosely associated with each other)
535 which suggests that the structural change of each dimer seems to be able to occur independently,

536 conformational changes of one subunit of Atu1419 and CbnR can be easily propagated to the other
537 parts of the tetramer through inter-subunit interactions. Noteworthy, single molecule experiments
538 revealed that a repressor securing a DNA loop is a stronger transcriptional roadblock for RNA
539 polymerases than one bound to a single DNA site (49).

540 Finally, we investigated the binding between the tetrameric Atu1419 and the Patu1418-1419 region of
541 370 bp containing two palindromes (190 bp apart) using negative staining electron microscopy. We
542 measured the size of the protein (~10nM) and the length of the DNA fragment (~125 nM), which are in
543 good agreement with the expected values (10 nM and 129 nM, respectively Figure 8). We were able to
544 observe three categories of binding: a tetramer bound to one DNA fragment, two tetramers bound to
545 one DNA fragment and a tetramer forming a DNA loop (Figure 8). The loops were measured at ~60-65
546 nM compatible with the length of 190 bp, which separates two palindromes.

547 CONCLUSION

548 Hydroxycinnamic acids are involved in the initiation of the plasmid-encoded pathogenicity program of
549 *A. fabrum* and can be degraded through the expression of the species-specific chromosomal gene
550 cluster composed of eight genes, which are regulated by two repressors HcaR/Atu1422 and Atu1419.
551 We have previously shown that HcaR, a dimeric protein from the MarR family, is the transcriptional
552 repressor of its own transcription, *atu1421* gene and that of the first three genes (*atu1415*, *atu1416* and
553 *atu1417*) of HCA degradation pathway. Here, we characterized Atu1419 as a transcriptional repressor
554 of its own repression in addition to two other genes (*atu1418* and *atu1420*). We identified the short
555 palindromic region composed of 10 base pairs and a degenerated palindrome bound to Atu1419 and
556 the N5,N10-methylenetetrahydrofolate (MEF) as the effector molecule disrupting Atu1419-DNA
557 interaction, allowing the expression of the second part of the HCA degradation pathway. Remarkably,
558 Atu1419 repressor is not regulated by the direct catabolic substrates/intermediates of the pathway. An
559 unexpected outcome was that Atu1419 is a tetrameric regulator whereas all GntR regulators so far
560 studied have been shown to be dimeric (20-22). This quaternary structure allows the repressor to bind
561 two distant DNA sites simultaneously making DNA loop repression (Figure 8) and providing a fine-tune
562 mechanism in the transcriptional regulation of the appropriate genes at the right time. Our structural
563 and biophysical study of Atu1419 revealed different structural rearrangements around the fortuitous
564 citrate bound in the effector binding domain within the tetramer and a double role of the Zn²⁺ ion in
565 protein stability and effector binding. This suggests an induced-allosteric mechanism by MEF, involving
566 conformational changes “in cascade” which enhance interactions between the DNA binding domain and
567 the end of the effector binding domain to constrain the flexibility of the HTH motif to prevent it from DNA
568 binding.

569 ACCESSION NUMBERS

570 The atomic coordinates and structure factors have been deposited at the Protein Data bank under PDB
571 ID 6Z74 and 6ZA0 for Atu1419 in complex with a fortuitous citrate in *P2₁2₁2₁* and *P2₁2₁2₂*, respectively,

572 PDB ID 6ZAB and 6ZA3 for Atu1419 in complex with DNA in *P6₄22* and *C222₁*, respectively and PDB
573 ID 6ZA7 for apo Atu1419.

574 **SUPPLEMENTARY DATA**

575 Supplementary Data are available at NAR online: Supplementary Figures S1 to S9 and Tables S1 to
576 S3.

577 **ACKNOWLEDGEMENT**

578 We acknowledge SOLEIL for provision of synchrotron radiation facilities (proposals ID 20130869,
579 20140774, 20150780 and 20170872) in using Proxima beamlines. We are grateful to Nadine Assrir
580 (ICSN, Gif sur Yvette) and Pascal Retailleau (ICSN, Gif sur Yvette) for access to the CD equipment and
581 help in data processing. We are grateful to Christophe Velours for performing SEC-MALS experiment
582 on the Protein-protein interactions platform. We thank the Cell and Tissue Imaging (PICT IBiSA), Institut
583 Curie, member of the French National Research Infrastructure France-BioImaging (ANR10-INBS-04).
584 We thank Andrew Saurin for a critical reading of the manuscript.

585 **FUNDING**

586 AV, SM, TM, LV, CL was supported by CNRS (Mission pour l'interdisciplinarité, Agromics 2014-2016).
587 This work was submitted to fulfill the requirements for a doctorate of biology at ED341-E2M2 from
588 Université de Lyon, granted from the French Ministère de l'Education Nationale, de l'Enseignement
589 Supérieur et de la Recherche (T.M). This work has benefited from the I2BC crystallization and Protein-
590 protein interactions platforms supported by FRISBI ANR-10-INSB-05-01.

591

592 **CONFLICT OF INTEREST**

593 The authors declare that they have no competing interests.

594

595

596 **FIGURES LEGENDS**

597 **Figure 1. (A)** Ferulic acid degradation pathway: a coenzyme A is added to ferulic acid by Atu1416, a
598 feruloyl-CoA synthase. Feruloyl-CoA is then converted by Atu1417, an enoyl-CoA hydratase, into 4-
599 hydroxy-3-methoxyphenyl- β -hydroxypropionyl (HMPHP)-CoA, which is in turn transformed into vanillic
600 acid by Atu1415, a phenylhydroxypropionyl-CoA dehydrogenase, and then Atu1421, a 4-hydroxy-3-
601 methoxyphenyl- β -ketopropionyl-CoA (HMPKP)-CoA β -keto-thiolase. The O-demethylase Atu1420
602 degrades vanillic acid into protocatechuic acid using tetrahydrofolate (H4F) as a cofactor and produces
603 N5-methyl-tetrahydrofolate (MH4F). Protocatechuic acid enters the cycle of Krebs. Atu1418 enzyme
604 was proposed to be involved in the recycling of H4F from transformation of MH4F into MEF, a compound
605 that can be enzymatically or spontaneously converted to H4F (18,50). Dashed arrows indicated putative
606 reactions based on sequence similarities and references (4,18,44,50). **(B)** Organization of the SpG8-
607 1b region for which genes expression is regulated by the two transcriptional factors HcaR and Atu1419
608 (their genes are framed in red). Promoters are shown with small arrows in black for HcaR dependent-
609 transcription and in blue for putative Atu1419 dependent-transcription indicating the direction of gene
610 transcription.

611 **Figure 2.** Atu1419 regulation mechanism. **(A)** Gel shift analysis of 30 nM of each promoter region
612 *Patu1416-1417*, *Patu1418-1419*, *Patu1420*, *PhcaR (atu1422)* incubated without and with 250 nM
613 tetrameric Atu1419 (ratio protein:DNA of 8.3). *PvirB* was used as a promoter region control. **(B)** In silico
614 analysis of the *atu1418-1419* and *atu1420* promoter regions performed with the BPROM program (31):
615 -10 and -35 boxes are underlined, and the palindromic sequences are shown in red. The transcription
616 initiation sites are shown with an arrow and the translation start sites are indicated in bold. Palindromic
617 regions are separated by 190 bp in *atu1418-1419* region. **(C)** Gel mobility shift assay analysis of 30 nM
618 of three parts of the promoter region of *atu1418-1419* containing each one palindrome (P1 or P2) or
619 none (P3) incubated without and with 250 nM tetrameric Atu1419 (ratio protein:DNA of 8.3).

620 **Figure 3. (A)** Wire representation of Atu1419 tetramer in complex with a fortuitous citrate molecule and
621 a co-purified Zn²⁺ ion, shown as sticks and ball, respectively, bound to the effector binding domain in
622 the structure at 2 Å resolution (molecule A is in cyan whereas molecules B/C/D in gray). **(B)** Left: ribbon
623 representation of dimer AB within the tetramer in **A**. The secondary structural elements are indicated in
624 subunit A and the orange elements form the dimeric interface. The linker joining the N-terminal DNA
625 binding domain and the C-terminal effector binding domain is shown in black. Right: ribbon
626 representation of dimer AD within the tetramer in **A**. The secondary structural elements are indicated in
627 subunit A and the orange elements form the other dimeric interface **(C)** Superposition of the subunits A
628 in cyan and B in gray of the tetramer in **A** (the subunits B, C, D adopt the same fold). The red parts
629 indicate major conformational changes between both subunits. A close-up view of the citrate and ion
630 binding sites in the effector/metal binding domain shown in blue and orange for subunits A and B,
631 respectively. **(D)** Interactions between the bound citrate in orientation A and in blue/Zn²⁺ (blue ball) and
632 subunit A in cyan. Hydrogen bonds are shown as dashed lines in black (distance below 3.2 Å) and
633 metal contacts are shown as dashed lines in red. Residues involved in the interactions are labeled and

634 shown as sticks. Citrate is shown in its Fo-Fc omit map contoured at 4σ in subunit A. (E) Interactions
635 between the bound citrate in orientation B and in orange/ Zn^{2+} (orange ball) and subunit B in gray.
636 Hydrogen bonds are shown as dashed lines in black (distance below 3.2 Å) and metal contacts are
637 shown as dashed lines in red. Residues involved in the interactions are labeled and shown as sticks.
638 Citrate is shown in its Fo-Fc omit map contoured at 4σ in subunit A.

639 **Figure 4.** Cartoon representation of the tetrameric Atu1419-DNA complex (A) in the $P6_422$ structure
640 with the monomer of the asymmetric unit shown in purple. (B) in the $P2_12_12_1$ structure with the dimer of
641 the asymmetric unit shown in magenta. (C) Superposition of the DNA binding domains of the dimers
642 AB of the $P6_422$ structure in purple and gray for subunits A and B, respectively and of the $P2_12_12_1$
643 structure in magenta. The palindromic 10-mer DNA is in green and orange in the $P6_422$ and $P2_12_12_1$
644 structures, respectively. (D) View of the DNA binding domain of the dimer AB of the $P6_422$ structure in
645 purple and gray for subunits A and B, respectively bound to the palindromic 10-mer DNA shown in
646 green. The secondary elements are indicated in subunit A. Arg45 in helix $\alpha 3$ is the key residue, which
647 interacts with the essential guanine base at position 3. This residue is shown as sticks. (E) View showing
648 a close-up of the Atu1419-DNA interface within a monomer and schematic diagram of Atu1419-DNA
649 contacts. Nucleotide bases that interact with Atu1419 are shown in green. (F) Gel mobility shift assay
650 analysis of *Patu1420* incubated without and with Atu1419, and with Atu1419 and 100 μM citrate (ratio
651 tetrameric protein:DNA of 8.3). (G) Gel mobility shift assay analysis of *Patu1420s* containing a unique
652 palindrome or *Patu1420s* AT, in which the palindrome was mutated with the recognized guanine at
653 position 3 replaced by an adenine and its cytosine partner replaced by a thymine, incubated without
654 and with Atu1419 (ratio tetrameric protein:DNA of 4.6).

655 **Figure 5.** (A) Cartoon representation of the C-terminal domain except for helix $\alpha 5$ shown in ribbon of
656 Atu1419 with a co-purified Zn^{2+} ion and a close-up view of the Zn^{2+} binding site. The four amino acid
657 residues bound to the metal ion (orange ball) are shown as sticks. (B) Superposition of dimers AB of
658 apo Atu1419 (in orange) and Atu1419-DNA complex (in purple and gray). The DNA is in green. (C)
659 Close-up view showing the superposition of their DNA binding domains colored as in B with a view at
660 90° .

661 **Figure 6.** (A) (B) (C) *Patu1418*, *Patu1419* or *Patu1420* expression in *A. fabrum* C58 wild-type strain
662 compared with that in $C58\Delta atu1420$ and $C58\Delta atu1418$ mutants in the presence of ferulic acid. Values
663 correspond to normalized fluorescence intensity in the presence of ferulic acid corrected by subtracting
664 values without ferulic acid. Different letters indicate statistical differences between conditions (one-way
665 ANOVA and Tukey test; P-value=0.05). (D) Gel mobility shift assay analysis of *Patu1420* DNA region
666 incubated without and with Atu1419 (ratio tetrameric protein:DNA of 17) and N5,N10-
667 methylenetetrahydrofolate at different concentrations (0 to 300 μM). (E) Gel mobility shift assay analysis
668 of *Patu1420* DNA region incubated without and with Atu1419, and with Atu1419 and H4F, MH4F or
669 MEF at 100 μM . (F) Differential scanning calorimetry thermograms of apo Atu1419 (black) and Atu1419
670 in the presence of MEF (magenta). The table below indicates the T_m . DSC experiments were performed

671 twice. (G) Isothermal titration microcalorimetry (ITC) experiments of Atu1419 towards MEF. The top
672 panel shows heat differences upon injection of MEF and low panel shows integrated heats of injection
673 with the best fit (solid line) to a single binding model using Microcal ORIGIN. (H) Control of MEF injection
674 in the buffer solution by ITC.

675 **Figure 7.** (A) *atu1420* promoter region: -10 and -35 boxes are underlined, the palindromic sequence
676 and the degenerated palindromic sequence (109 bp apart) are in red. The transcription initiation site is
677 shown with an arrow and the translation start site is indicated in bold. (B) Gel mobility shift assay
678 analysis of four parts of the promoter region of *atu1420* containing each one palindrome or one
679 degenerated palindrome (P1 or P2), or both (P1+P2) or none (P3) incubated without and with 250 nM
680 or 500 nM tetrameric Atu1419 (ratio protein:DNA of 8.3 or 16.6).

681 **Figure 8.** Interactions between the tetrameric Atu1419 repressor and *Patu1418-1419* region of 370 bp
682 containing two palindromes separated by 190 bp visualized by electron microscopy and corresponding
683 model. (A) Image of Atu1419 tetramer as a control, (B) Image of *Patu1418-1419* region as a control,
684 (C) Image of one tetramer bound to one palindrome only, (D) Image of two tetramers bound each to
685 one palindrome, (E) Image of one tetramer and a DNA loop.

686

687 REFERENCES

- 688 1. Nester, E.W. (2014) Agrobacterium: nature's genetic engineer. *Front Plant Sci*, **5**, 730.
689 2. Dessaux, Y. and Faure, D. (2018) Niche Construction and Exploitation by Agrobacterium: How
690 to Survive and Face Competition in Soil and Plant Habitats. *Curr Top Microbiol Immunol*, **418**,
691 55-86.
692 3. Meyer, T., Thiour-Mauprivez, C., Wisniewski-Dye, F., Kerzaon, I., Comte, G., Vial, L. and Lavire,
693 C. (2019) Ecological Conditions and Molecular Determinants Involved in Agrobacterium
694 Lifestyle in Tumors. *Front Plant Sci*, **10**, 978.
695 4. Lassalle, F., Campillo, T., Vial, L., Baude, J., Costechareyre, D., Chapulliot, D., Shams, M.,
696 Abrouk, D., Lavire, C., Oger-Desfeux, C. *et al.* (2011) Genomic species are ecological species
697 as revealed by comparative genomics in Agrobacterium tumefaciens. *Genome Biol Evol*, **3**,
698 762-781.
699 5. Campillo, T., Renoud, S., Kerzaon, I., Vial, L., Baude, J., Gaillard, V., Bellvert, F., Chamignon,
700 C., Comte, G., Nesme, X. *et al.* (2014) Analysis of hydroxycinnamic acid degradation in
701 Agrobacterium fabrum reveals a coenzyme A-dependent, beta-oxidative deacetylation pathway.
702 *Appl Environ Microbiol*, **80**, 3341-3349.
703 6. Bhattacharya, A., Sood, P. and Citovsky, V. (2010) The roles of plant phenolics in defence and
704 communication during Agrobacterium and Rhizobium infection. *Mol Plant Pathol*, **11**, 705-719.
705 7. Guo, M., Huang, Z. and Yang, J. (2017) Is there any crosstalk between the chemotaxis and
706 virulence induction signaling in Agrobacterium tumefaciens? *Biotechnol Adv*, **35**, 505-511.
707 8. Parke, D., Ornston, L.N. and Nester, E.W. (1987) Chemotaxis to plant phenolic inducers of
708 virulence genes is constitutively expressed in the absence of the Ti plasmid in Agrobacterium
709 tumefaciens. *J Bacteriol*, **169**, 5336-5338.
710 9. Kape, R., Parniske, M. and Werner, D. (1991) Chemotaxis and nod Gene Activity of
711 Bradyrhizobium japonicum in Response to Hydroxycinnamic Acids and Isoflavonoids. *Appl*
712 *Environ Microbiol*, **57**, 316-319.
713 10. Kalogeraki, V.S., Zhu, J., Eberhard, A., Madsen, E.L. and Winans, S.C. (1999) The phenolic vir
714 gene inducer ferulic acid is O-demethylated by the VirH2 protein of an Agrobacterium
715 tumefaciens Ti plasmid. *Mol Microbiol*, **34**, 512-522.

- 716 11. Meyer, T., Renoud, S., Vigouroux, A., Miomandre, A., Gaillard, V., Kerzaon, I., Prigent-
717 Combaret, C., Comte, G., Morera, S., Vial, L. *et al.* (2018) Regulation of hydroxycinnamic acid
718 degradation drives *Agrobacterium fabrum* lifestyles. *Mol Plant Microbe Interact.*
- 719 12. Duprey, A., Reverchon, S. and Nasser, W. (2014) Bacterial virulence and Fis: adapting
720 regulatory networks to the host environment. *Trends Microbiol*, **22**, 92-99.
- 721 13. Valentini, M., Gonzalez, D., Mavridou, D.A. and Filloux, A. (2018) Lifestyle transitions and
722 adaptive pathogenesis of *Pseudomonas aeruginosa*. *Curr Opin Microbiol*, **41**, 15-20.
- 723 14. Barton, I.S., Fuqua, C. and Platt, T.G. (2018) Ecological and evolutionary dynamics of a model
724 facultative pathogen: *Agrobacterium* and crown gall disease of plants. *Environ Microbiol*, **20**,
725 16-29.
- 726 15. Kohler, A.C., Mills, M.J.L., Adams, P.D., Simmons, B.A. and Sale, K.L. (2017) Structure of aryl
727 O-demethylase offers molecular insight into a catalytic tyrosine-dependent mechanism. *Proc*
728 *Natl Acad Sci U S A*, **114**, E3205-E3214.
- 729 16. Harada, A., Kamimura, N., Takeuchi, K., Yu, H.Y., Masai, E. and Senda, T. (2017) The crystal
730 structure of a new O-demethylase from *Sphingobium* sp. strain SYK-6. *FEBS J*, **284**, 1855-
731 1867.
- 732 17. Parke, D. (1995) Supraoperonic clustering of *pca* genes for catabolism of the phenolic
733 compound protocatechuate in *Agrobacterium tumefaciens*. *J Bacteriol*, **177**, 3808-3817.
- 734 18. Abe, T., Masai, E., Miyauchi, K., Katayama, Y. and Fukuda, M. (2005) A tetrahydrofolate-
735 dependent O-demethylase, LigM, is crucial for catabolism of vanillate and syringate in
736 *Sphingomonas paucimobilis* SYK-6. *J Bacteriol*, **187**, 2030-2037.
- 737 19. Hoskisson, P.A. and Rigali, S. (2009) Chapter 1: Variation in form and function the helix-turn-
738 helix regulators of the GntR superfamily. *Adv Appl Microbiol*, **69**, 1-22.
- 739 20. Rigali, S., Derouaux, A., Giannotta, F. and Dusart, J. (2002) Subdivision of the helix-turn-helix
740 GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem*,
741 **277**, 12507-12515.
- 742 21. Suvorova, I.A., Korostelev, Y.D. and Gelfand, M.S. (2015) GntR Family of Bacterial
743 Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution.
744 *PLoS One*, **10**, e0132618.
- 745 22. Jain, D. (2015) Allosteric control of transcription in GntR family of transcription regulators: A
746 structural overview. *IUBMB Life*, **67**, 556-563.
- 747 23. van Aalten, D.M., DiRusso, C.C. and Knudsen, J. (2001) The structural basis of acyl coenzyme
748 A-dependent regulation of the transcription factor FadR. *EMBO J*, **20**, 2041-2050.
- 749 24. Allaway, D., Schofield, N.A., Leonard, M.E., Gilardoni, L., Finan, T.M. and Poole, P.S. (2001)
750 Use of differential fluorescence induction and optical trapping to isolate environmentally
751 induced genes. *Environ Microbiol*, **3**, 397-406.
- 752 25. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr*, **66**, 125-132.
- 753 26. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J.
754 (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.
- 755 27. Blanc, E., Roversi, P., Vornrhein, C., Flensburg, C., Lea, S.M. and Bricogne, G. (2004)
756 Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta*
757 *Crystallogr D Biol Crystallogr*, **60**, 2210-2221.
- 758 28. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta*
759 *Crystallogr. D Biol. Crystallogr.*, **60**, 2126-2132.
- 760 29. Sreerama, N. and Woody, R.W. (2000) Estimation of protein secondary structure from circular
761 dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an
762 expanded reference set. *Anal Biochem*, **287**, 252-260.
- 763 30. Whitmore, L. and Wallace, B.A. (2004) DICHROWEB, an online server for protein secondary
764 structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*, **32**, W668-
765 673.
- 766 31. Solovyev, V. and Salamov, A. (2011) Automatic Annotation of Microbial Genomes and
767 Metagenomic Sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine*
768 *and Environmental Studies. (Ed. R. W. Li) Nova Science Publishers*, 17.
- 769 32. Janin, J., Rodier, F., Chakrabarti, P. and Bahadur, R.P. (2007) Macromolecular recognition in
770 the Protein Data Bank. *Acta Crystallogr D Biol Crystallogr*, **63**, 1-8.
- 771 33. Xu, Y., Heath, R.J., Li, Z., Rock, C.O. and White, S.W. (2001) The FadR.DNA complex.
772 Transcriptional control of fatty acid metabolism in *Escherichia coli*. *J Biol Chem*, **276**, 17373-
773 17379.

- 774 34. Shi, W., Kovacicova, G., Lin, W., Taylor, R.K., Skorupski, K. and Kull, F.J. (2015) The 40-
775 residue insertion in *Vibrio cholerae* FadR facilitates binding of an additional fatty acyl-CoA
776 ligand. *Nat Commun*, **6**, 6032.
- 777 35. Gao, Y.G., Suzuki, H., Itou, H., Zhou, Y., Tanaka, Y., Wachi, M., Watanabe, N., Tanaka, I. and
778 Yao, M. (2008) Structural and functional characterization of the LldR from *Corynebacterium*
779 *glutamicum*: a transcriptional repressor involved in L-lactate and sugar utilization. *Nucleic Acids*
780 *Res*, **36**, 7110-7123.
- 781 36. van Aalten, D.M., DiRusso, C.C., Knudsen, J. and Wierenga, R.K. (2000) Crystal structure of
782 FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold.
783 *EMBO J*, **19**, 5167-5177.
- 784 37. Lord, D.M., Uzgoren Baran, A., Soo, V.W., Wood, T.K., Peti, W. and Page, R. (2014)
785 McbR/YncC: implications for the mechanism of ligand and DNA binding by a bacterial GntR
786 transcriptional regulator involved in biofilm formation. *Biochemistry*, **53**, 7223-7231.
- 787 38. Zheng, M., Cooper, D.R., Grosseohme, N.E., Yu, M., Hung, L.W., Cieslik, M., Derewenda, U.,
788 Lesley, S.A., Wilson, I.A., Giedroc, D.P. *et al.* (2009) Structure of *Thermotoga maritima* TM0439:
789 implications for the mechanism of bacterial GntR transcription regulators with Zn²⁺-binding
790 FCD domains. *Acta Crystallogr D Biol Crystallogr*, **65**, 356-365.
- 791 39. Pinheiro, J., Lisboa, J., Pombinho, R., Carvalho, F., Carreaux, A., Brito, C., Pontinen, A.,
792 Korkeala, H., Dos Santos, N.M.S., Morais-Cabral, J.H. *et al.* (2018) MouR controls the
793 expression of the *Listeria monocytogenes* Agr system and mediates virulence. *Nucleic Acids*
794 *Res*, **46**, 9338-9352.
- 795 40. Otani, H., Stogios, P.J., Xu, X., Nocek, B., Li, S.N., Savchenko, A. and Eltis, L.D. (2016) The
796 activity of CouR, a MarR family transcriptional regulator, is modulated through a novel
797 molecular mechanism. *Nucleic Acids Res*, **44**, 595-607.
- 798 41. Cogan, D.P., Baraquet, C., Harwood, C.S. and Nair, S.K. (2018) Structural basis of
799 transcriptional regulation by CouR, a repressor of coumarate catabolism, in
800 *Rhodopseudomonas palustris*. *J Biol Chem*, **293**, 11727-11735.
- 801 42. Danenberg, P.V., Gustavsson, B., Johnston, P., Lindberg, P., Moser, R., Odin, E., Peters, G.J.
802 and Petrelli, N. (2016) Folates as adjuvants to anticancer agents: Chemical rationale and
803 mechanism of action. *Crit Rev Oncol Hematol*, **106**, 118-131.
- 804 43. Jagerstad, M. and Jastrebova, J. (2014) 5,10-Methylene-tetrahydrofolate dissociates into
805 tetrahydrofolate and formaldehyde at physiological pH and acidic pH, typical conditions used
806 during sample extraction and LC-MS/MS analysis of biological samples. *Biomed Chromatogr*,
807 **28**, 1041-1042.
- 808 44. Feirer, N., Xu, J., Allen, K.D., Koestler, B.J., Bruger, E.L., Waters, C.M., White, R.H. and Fuqua,
809 C. (2015) A Pterin-Dependent Signaling Pathway Regulates a Dual-Function Diguanylate
810 Cyclase-Phosphodiesterase Controlling Surface Attachment in *Agrobacterium tumefaciens*.
811 *mBio*, **6**, e00156.
- 812 45. Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K. and Senda, T. (2003) Crystal
813 structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two
814 subunit forms and molecular bases for causing and changing DNA bend. *J Mol Biol*, **328**, 555-
815 566.
- 816 46. Amouyal, M., Mortensen, L., Buc, H. and Hammer, K. (1989) Single and double loop formation
817 when deoR repressor binds to its natural operator sites. *Cell*, **58**, 545-551.
- 818 47. Bylino, O.V., Ibragimov, A.N. and Shidlovskii, Y.V. (2020) Evolution of Regulated Transcription.
819 *Cells*, **9**.
- 820 48. Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan,
821 R.G. and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with
822 DNA and inducer. *Science*, **271**, 1247-1254.
- 823 49. Voros, Z., Yan, Y., Kovari, D.T., Finzi, L. and Dunlap, D. (2017) Proteins mediating DNA loops
824 effectively block transcription. *Protein Sci*, **26**, 1427-1438.
- 825 50. De Brouwer, V., Zhang, G.F., Storozhenko, S., Straeten, D.V. and Lambert, W.E. (2007) pH
826 stability of individual folates during critical sample preparation steps in prevision of the analysis
827 of plant folates. *Phytochem Anal*, **18**, 496-508.

828

Table 1. Crystallographic data and refinement parameters

	Zn ²⁺ SAD	Atu1419 Citrate	Atu1419 Citrate	Atu1419-DNA#	Atu1419-DNA§	Apo Atu1419†
PDB code		6Z74	6ZA0	6ZAB	6ZA3	6ZA7
Wavelength (Å)	1.282290	0.97541	0.978570	1	1	0.97934
Crystallization conditions	5% PEG 4K, 0.2 M AS, 0.1 M Na-Citrate pH 5.6	5% PEG 4K, 0.2 M AS, 0.1 M Na-Citrate pH 5.6	5% PEG 4K, 0.2 M AS, 0.1 M Na-Citrate pH 5.6	10% Terbutanol, 0.1 M NaCitrate pH 5.6, 2 mM MgCl ₂	25% PEG 400, 0.1 M MES pH 6.7, 0.2 M Na-Acetate	12% PEG 4K, 0.2 Ammonium Sulfate, 0.1 M Tris/HCl pH 8.5
Za	4	4	2	1	2	4
Space group	<i>P2₁2₁2₁</i>	<i>P2₁2₁2₁</i>	<i>P2₁2₁2</i>	<i>P6₄22</i>	<i>C222₁</i>	<i>P2₁</i>
Cell parameters (Å, °)	a= 77.4 b= 114.1 c= 130.5	a= 77.2 b= 113.7 c= 130.5	a= 72.2 b= 145.6 c= 41.2	a= 179.5 b= 179.2 c= 96.8	a= 62.4 b= 112.4 c= 179.6	a= 52.1 b= 79.0 c= 140.6 β= 95.1
Resolution (Å)	49.8-2.35 (2.42-2.35)	45.6-2 (2.12-2)	24.0-1.65 (1.69-1.65)	45.61-2.8 (2.87-2.8)	47.64-2.0 (2.05-2.0)	43.38-2.34 (2.41-2.34)
No. of observed reflections	518098 (29677)	467551 (72154)	239555 (17376)	444041 (30119)	376669 (23170)	504810 (31752)
No. of unique reflections	48018 (3262)	78344 (12280)	51917 (3674)	23089 (1642)	43067 (3095)	47943 (3255)
Completeness (%)	99.4 (92.1)	99.6 (98.1)	97.9 (95.8)	99.7 (88.4)	99.9 (99.1)	99.4 (92.3)
Completeness Staraniso (%)				68.7 (11.9)	97.7 (79.3)	71.5 (30.6)
<i>R</i> _{sym} (%)	8.8 (78.6)	12 (155.5)	8.1 (127.3)	11.6 (236.4)	8.6 (190.3)	10.9 (128.8)
<i>R</i> _{pim} (%)	2.8 (26.6)	5.4 (78.1)	4.2 (64.4)	2.7 (57.7)	3.1 (79.3)	3.5 (41.5)
<i>I</i> / <i>σ</i> (<i>I</i>)	18 (2.3)	10 (1.12)	10.7 (1.1)	18.8 (1.1)	13.7 (0.9)	13.0 (1.4)
CC _{1/2}	0.999 (0.809)	0.998 (0.582)	0.999 (0.461)	0.999 (0.469)	0.999 (0.421)	0.999 (0.696)
<i>R</i> _{cryst} (%)		19.1	16.9	18.8	19.2	19.4
<i>R</i> _{free} (%)		21.4	18.9	19.9	22.8	22.2
rms bond deviation (Å)		0.01	0.001	0.009	0.01	0.009
rms angle deviation (°)		0.99	0.95	1	1	1.02
Average B (Å ²)						
Protein A/B/C/D		42/53.1/60/57.5	27.9/29.2	86.4	50.7/50.1	66.2/74.3/73.2/80.8
Zn ²⁺		35/41.4/39.7/45.2	18.3/18.6	106.9	43.3/46.2	46.2/63.2/69.3/66.6
citrate		44.4/45.2/44.5/45.7	23.1/22	127.9		
DNA				65.3	60.2	
solvent		50.5	36.7	56.3	48.	50
^a Clashscore		4.56	1.6	3.2	2.11	2.88
MolProbity score		1.51	0.93	2.07	1.24	1.38
^a Ramachandran plot (%)						
Favoured		99.36	98.73	94.94	98.73	98.2
Outliers		0	0	0	0	0.21

Values for the highest resolution shell are in parentheses. CC_{1/2} = percentage of correlation between intensities from random half-dataset

^aCalculated with MolProbity

Numbers in italic account for statistical values after ellipsoidal mask application by Staraniso.

A dataset collected from a crystal, which diffracted anisotropically to 3.4 Å along *b** and 0.894*a** - 0.447*b** and 2.7 Å along *c**

† Two datasets collected from a crystal, which diffracted anisotropically to 3.2 Å along 0.81*a** - 0.58*c**, 2.3 Å along *b** and 2.2 Å along *c**

§ One dataset collected from a crystal, which diffracted anisotropically to 2.02 Å along *a**, 1.99 Å along *b** and 1.95 Å along *c**

Table 2. SpG8 1-b genes expression in the wild-type (WT) and C58 Δ *atu1419* strains. Values refer to fold change at 24 hours obtained by comparison of genes expression in the WT C58 strain with and without 750 μ M of ferulic acid and correspond to the mean of three biological replicates with at least six technical replicates. Fold change at 24 hours without ferulic acid was obtained by comparison of genes expression in the C58 Δ *atu1419* and the WT strains and corresponds to the mean of at least two biological replicates and six technical replicates. Empty pOT1e corresponds to plasmid without any promoting region before the *egfp* gene in order to measure basal expression of the system.

Genes	Fold change C58 WT in the presence/in the absence of ferulic acid	Fold change C58Δ<i>atu1419</i>/WT in the absence of ferulic acid
<i>Patu1416</i>	2.53	0.63
<i>Patu1418</i>	1.44	3.67
<i>Patu1419</i>	1.81	3.92
<i>Patu1420</i>	3.31	1.58
Empty pOT1e	1.12	0.97

Figure 1.

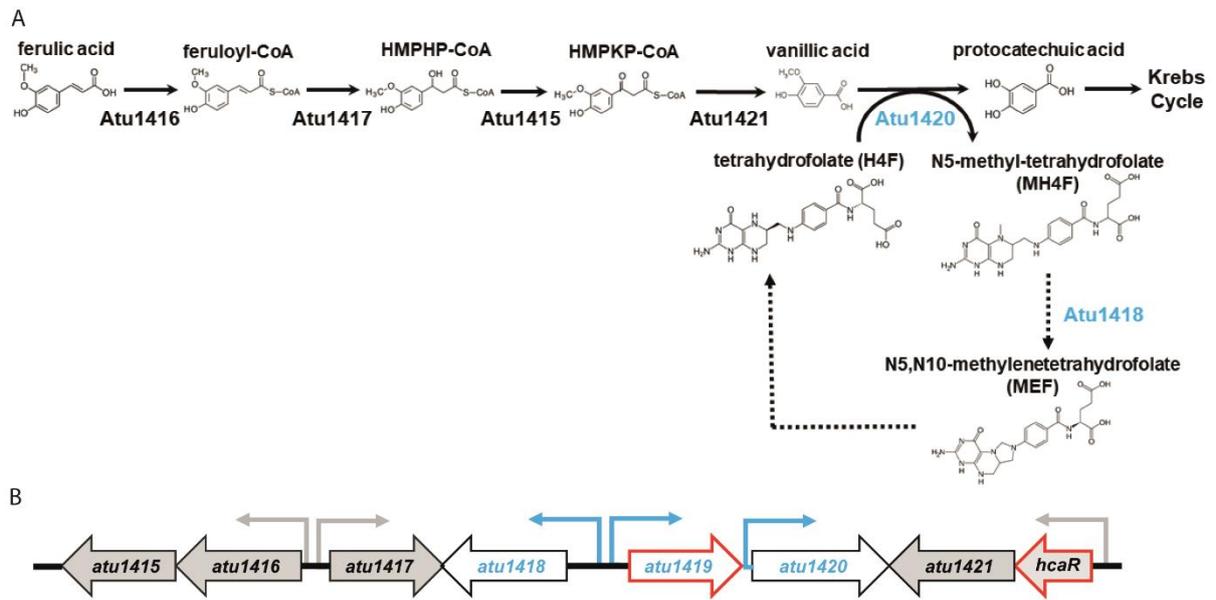


Figure 2.

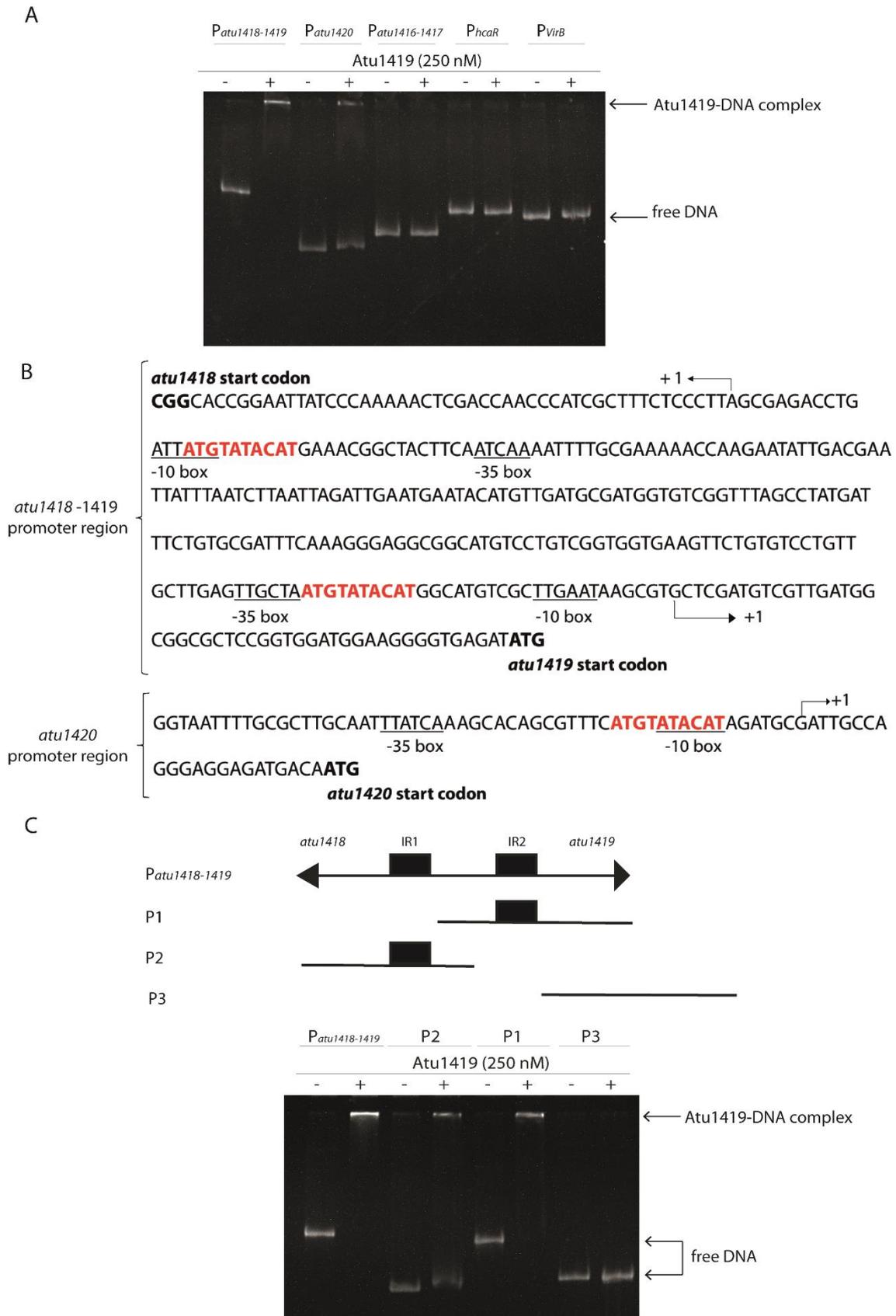


Figure 3.

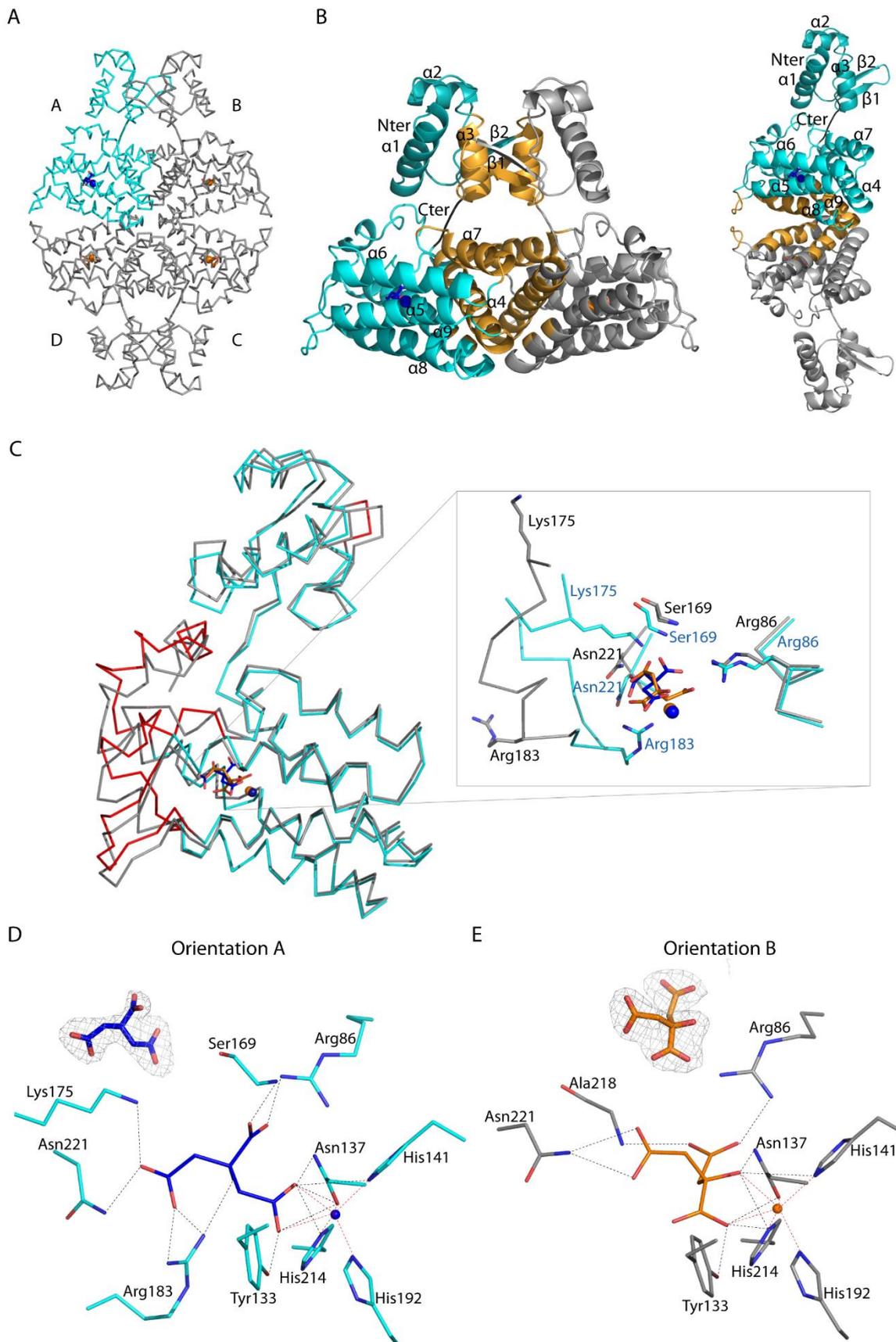


Figure 4.

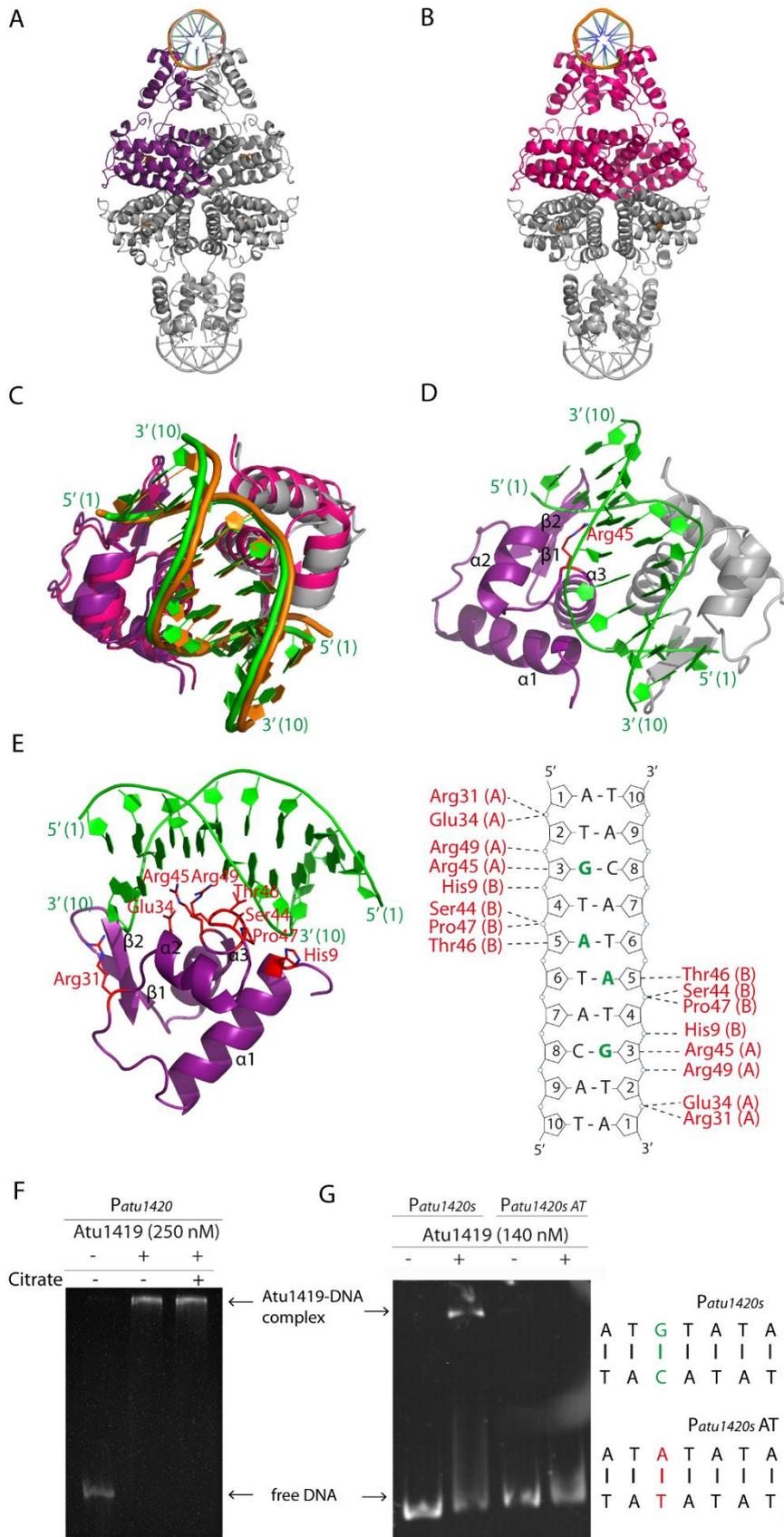
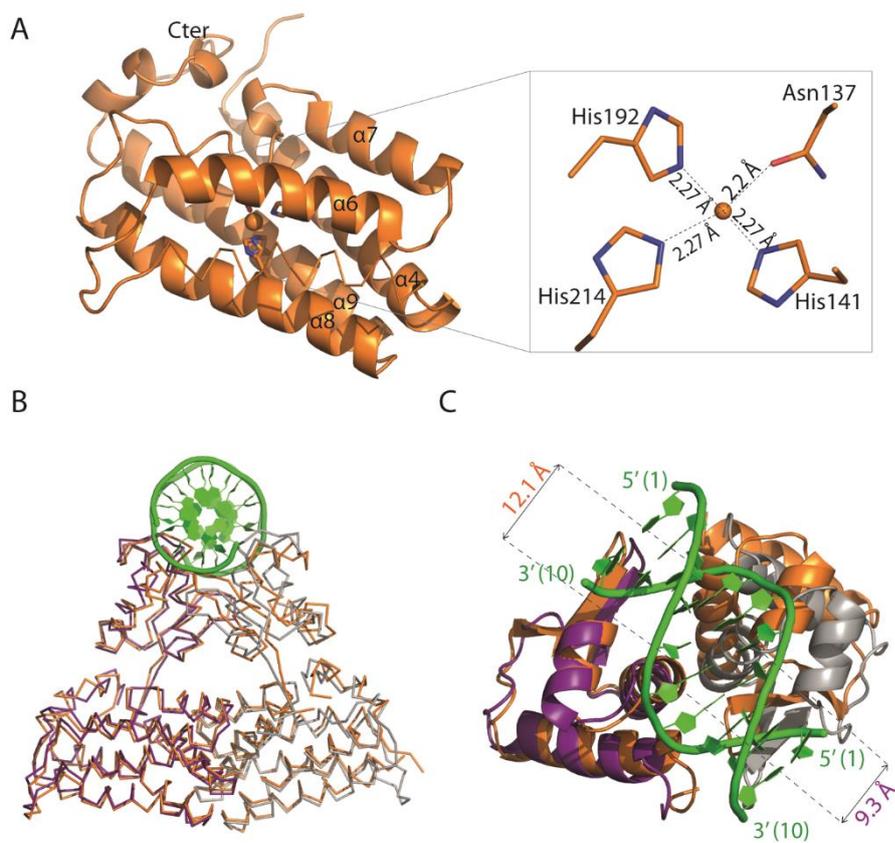
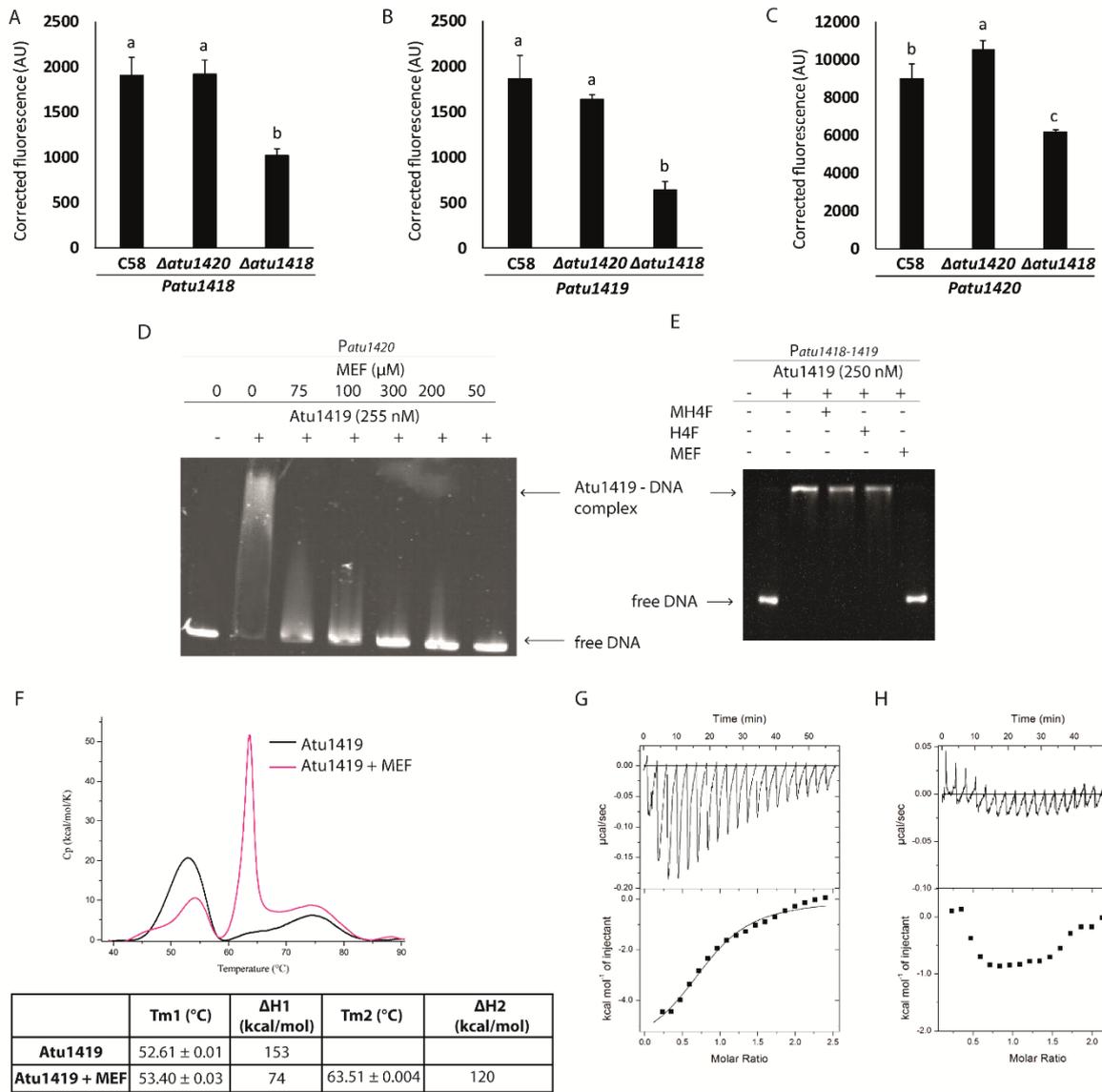


Figure 5.



829

830 **Figure 6.**



831

832

833

834 **Figure 7**

