11910–11926 *Nucleic Acids Research, 2018, Vol. 46, No. 22* *Published online 31 October 2018*
doi: 10.1093/nar/gky996

Novel regulatory mechanism of establishment genes of conjugative plasmids

Jorge Val-Calvo [1], Juan R. Luque-Ortega[2], Isidro Crespo [3], Andrés Miguel-Arribas[1], David Abia[4], Dione L. Sánchez-Hevia[5], Ester Serrano[1], César Gago-Córdoba[1], Saúl Ares[5,6], Carlos Alfonso[2], Fernando Rojo [5], Ling J. Wu [7], D. Roeland Boer [3,*] and Wilfried J.J. Meijer [1,*]

[1]Department of Virology and Microbiology, Centro de Biología Molecular “Severo Ochoa" (CSIC-UAM), Instituto de Biología Molecular “Eladio Viñuela" (CSIC), C. Nicolás Cabrera 1, Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain, [2]Centro de Investigaciones Biológicas (CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain, [3]ALBA Synchrotron Light Source, Carrer de la Llum 2-26, Cerdanyola del Vallès, Barcelona 08290, Spain, [4]Bioinformatics Facility, Centro de Biología Molecular "Severo Ochoa", [5]Centro Nacional de Biotecnología (CSIC), Darwin 3, 28049 Madrid, Spain, [6]Grupo Interdisciplinar de Sistemas Complejos (GISC) and Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganes, Spain and [7]Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Richardson Road, Newcastle Upon Tyne, NE4AX, UK

Received July 31, 2018; Revised October 05, 2018; Editorial Decision October 08, 2018; Accepted October 10, 2018

ABSTRACT

The principal route for dissemination of antibiotic resistance genes is conjugation by which a conjugative DNA element is transferred from a donor to a recipient cell. Conjugative elements contain genes that are important for their establishment in the new host, for instance by counteracting the host defense mechanisms acting against incoming foreign DNA. Little is known about these establishment genes and how they are regulated. Here, we deciphered the regulation mechanism of possible establishment genes of plasmid p576 from the Gram-positive bacterium *Bacillus pumilus.* Unlike the ssDNA promoters described for some conjugative plasmids, the four promoters of these p576 genes are repressed by a repressor protein, which we named Reg_{576}. Reg_{576} also regulates its own expression. After transfer of the DNA, these genes are de-repressed for a period of time until sufficient Reg_{576} is synthesized to repress the promoters again. Complementary *in vivo* and *in vitro* analyses showed that different operator configurations in the promoter regions of these genes lead to different responses to Reg_{576}. Each operator is bound with extreme cooperativity by two Reg_{576}-dimers. The X-ray structure revealed that Reg_{576} has a Ribbon-Helix-Helix core and provided important insights into the high cooperativity of DNA recognition.

INTRODUCTION

Horizontal gene transfer (HGT) is the process by which (large) DNA regions are transferred between bacterial cells (1–3). A single HGT event can therefore significantly alter the genetic content of a cell and, hence HGT plays an important role in the evolution of prokaryotes. In addition, HGT is also responsible for the rapid spread of antibiotic resistance genes and virulence factors in bacteria. HGT involves various routes. Of these, conjugation is majorly responsible for the spread of antibiotic resistance genes. Conjugation is the process by which a DNA element is transferred from a donor to a recipient cell via a connecting channel. A conjugative element can be located either on an autonomously replicating plasmid, which is named conjugative plasmid, or on a bacterial genome, which is then named integrated conjugative element (ICE). Conjugative elements are widespread in both Gram-negative (G-) and Gram-positive (G+) bacteria and the basic principles of the conjugation process are conserved. The initial steps of the conjugation process involve the formation of a mating pair in which a donor cell recognizes and interacts with a suitable recipient cell. Probably, this triggers a signal for processing the DNA of the conjugative element to generate the ssDNA, named T-strand, which is subsequently transferred

*To whom correspondence should be addressed. Tel: +34 91 196 4539; Fax: +34 91 196 4420; Email: wmeijer@cbm.csic.es
Correspondence may also be addressed to D. Roeland Boer. Tel: +34 93 592 4333; Fax: +34 93 592 4333; Email: rboer@cells.es
Present address: Ester Serrano, Centro Nacional de Biotecnología (CSIC), Darwin 3, 28049 Madrid, Spain. Dione L. Sánchez-Hevia, Centro de Biología Molecular ”Severo Ochoa” (CSIC-UAM), Instituto de Biología Molecular ”Eladio Viñuela” (CSIC), C. Nicolás Cabrera 1, Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain.

into the recipient cell through the connecting channel. It has to be noted that although in most conjugation systems a single DNA strand is transferred into the recipient cell, double-stranded DNA is transferred in G+ mycelial *Streptomyces* bacteria (4,5). Transfer of the DNA into the recipient cell does not imply automatically a successful conjugation event. For instance, the transferred ssDNA must be circularized and converted into dsDNA, and, in the case of ICEs, become integrated into the bacterial chromosome. In addition, bacteria possess mechanisms that protect them against incoming foreign DNA. First-line defense mechanisms possessed by most bacteria recognize and inactivate foreign DNA by restriction-modification (RM) systems that encode a restriction endonuclease (REase) and a methyltransferase (MTase). The MTase methylates specific short DNA sequences and thereby protects these sequences from being digested by the cognate REase of the RM system. Foreign DNA entering a cell that is not properly methylated (including conjugative DNA) will be digested by the REase (for review see, 6,7). At least some conjugative elements encode anti-restriction proteins, named Ard (alleviation of restriction of DNA) proteins, which impair the REase activity. Once the entered DNA is properly methylated it is no longer recognized as foreign and inactivation of the REase can be relieved. In fact, expression of anti-restriction genes must be limited to a short period of time after the conjugative element has entered the cell because prolonged expression will make the cell vulnerable to the entry of other foreign DNA, like phage DNA. Thus, anti-restriction genes are important for stable establishment of the conjugative element in the new host, and hence they are referred to as establishment genes. Another establishment gene is *psiB*, present on at least several conjugative plasmids of G- origin. As explained above, during conjugation only a single DNA strand enters the recipient cell. ssDNA is a potent trigger of the host's SOS response. The PsiB protein suppresses activation of the SOS response by binding the RecA protein (8). Like the anti-restriction genes, it is in the interest of the host that the SOS response is suppressed only for a limited time after entry of the conjugative DNA into the cell. In some conjugative elements of G- bacteria, like the F and ColIbP-9 plasmids, the anti-restriction gene *ardA* and the *psiB* gene are under the control of a special single-stranded promoter (ssDNA promoter) which ensures that they are only expressed for a short period of time after entry of the ssDNA into the recipient cell (9–12). The colIbP-9 plasmid contains three ssDNA promoters that control the expression of at least 10 genes (13). The function of most of these genes is unknown but because they are controlled by ssDNA promoters it is likely that they are functionally related, all playing a role in enhancing establishment of the conjugative element in the new host.

As a possible approach to combat the spreading of antibiotic resistance, drugs have been developed that affect the activity of some relaxase proteins, which are crucial for initiating the DNA processing step to generate the T-strand (14). Similarly, compromising the activity of establishment proteins and/or deregulating their expression may also be a fruitful strategy to impair conjugation. So far though, this approach has not attained much attention, and little is known about establishment genes and their regulation of conjugative elements present in G+ bacteria. In fact, only few conjugative plasmids of G+ bacteria have been studied so far. In our laboratory we study the related conjugative plasmids p576 from *Bacilllus pumilus* and pLS20 from *Bacillus subtilis* (15–17). In this study we focused on the transcriptional regulation of the establishment genes of plasmid p576, which contains a putative *ard* gene and six other genes that may have a role in establishment of the plasmid in the recipient cell after transfer. We have identified the repressor of their promoters, named Reg_{576}, and we show that expression of the p576 establishment genes is regulated in a fundamentally different way than the ssDNA promoters described for some conjugative plasmids of G- bacteria. The number and position of the operator sites is not the same for each promoter, and the different configurations control the sensitivity of a promoter to Reg_{576}-mediated repression. Structure determination of the apo form revealed that Reg_{576} belongs to the Ribbon-Helix-Helix (RHH) family of DNA binding proteins. Reg_{576} binds DNA as a dimer-of-dimers, providing insights into the extremely high cooperative binding. The implications of these studies for regulation of the possible establishment genes on p576 and possibly other conjugative plasmids are discussed.

MATERIALS AND METHODS

Bacterial strains, plasmids and oligonucleotides

Bacterial strains were grown in Lysogeny Broth (LB) medium or on 1.5% agar plates of LB or Spizizen minimal medium (18). When appropriate, the following antibiotics were added: ampicillin (100 μg/ml) or kanamycin (30 μg/ml) for *Escherichia coli*; erythromycin (2 μg/ml) or spectinomycin (100 μg/ml) for *B. subtilis*. When indicated, LB agar plates were supplemented with isopropyl β-D-1-thiogalactopyranoside (IPTG) or 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal) to the final concentrations specified. Supplementary Table S1 lists the strains used. Plasmids and oligonucleotides used are listed in Supplementary Tables S2 and 3, respectively. All oligonucleotides were purchased from Isogen Life Science, The Netherlands.

Transformation

Escherichia coli cells were transformed by standard methods (19). Competent *B. subtilis* cells were prepared as described before (18). Transformants were selected on LB or Spizizen minimal agar medium plates supplemented with appropriate antibiotics and in the case of minimal agar medium plates complemented with tryptophan and/or threonine (20 mg/ml).

Construction of plasmids and strains

Detailed description of the construction of plasmids and strains is available in Supplementary Data.

β-Galactosidase activity assays

β-galactosidase activities were essentially determined as described previously (20). In short, overnight grown cultures

were diluted to an $OD_{600} \approx 0.025$ in fresh 37°C LB. Growth was followed by measuring OD_{600} at regular intervals. Samples were taken every 30 min from shaking cultures until $OD_{600} > 1$ and stored at −80°C for later processing. Processing of the samples involved lysis of the cells, incubation at 28°C of the lysate with the chromogenic substrate o-nitrophenyl-beta-D-galactopyranoside (ONPG) and determination of the OD_{420} value of centrifuged sample after stopping the enzymatic reaction at the desired time.

RNA isolation

Total RNA was isolated from late exponentially growing cells using the RNeasy Mini Kit (Qiagen, Germany) according to the manufacturer's protocol. RNA protect solution (Qiagen, Germany) was used to ensure integrity of RNA during isolation and also to stop transcription at given time points.

Protein purification

Escherichia coli BL21 (DE3) cells carrying plasmid pJV9 (containing $his_{(6)}reg_{576}$) were used to inoculate 10 l of fresh LB medium supplemented with kanamycin and grown at 37°C with shaking. At an OD_{600} of 0.4, expression of $His_{(6)}$27c was induced by adding IPTG to a final concentration of 1 mM and growth was continued for 2 h. After harvesting by centrifugation, the cells were lysed and processed as described before to purify the his-tagged Reg_{576} protein by affinity purification (21). Purified protein (>95% pure, 275 μM) was desalted with PD10 columns, equilibrated in buffer B (50 mM Tris (pH7.5), 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 7 mM β-mercaptoethanol, 50% Glicerol) or buffer C (20 mM Tris (pH8), 500 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 0.1 mM β-mercaptoethanol, 1% glycerol) and stored in aliquots at −80°C. In the case of fractions used for crystallization, the target buffer used for desalting was 20 mM Tris–HCl (pH8.0), 250 mM NaCl. Protein concentration was determined by OD_{280} using an $\varepsilon_{0.1\%} = 0.411$ L g^{-1} cm^{-1}.

EMSA

Electrophoretic mobility shift assay (EMSA) assays were essentially performed as described before (22). In short, 120 ng of the polymerase chain reaction fragment was incubated for 20 min at RT in Binding Buffer (20 mM Tris HCl [pH 8.0], 1 mM EDTA, 5 mM $MgCl_2$, 100 mM KCl, 10% (v/v) glycerol) without and with increasing amounts of purified $His_{(6)}Reg_{576}$ in a total volume of 20 μl. Next, samples were incubated on ice for 10 min and were subsequently loaded onto a 2% agarose gel in 0.5·TBE. Electrophoresis was carried out in 0.5·TBE at 60 V at 4°C. After electrophoresis, the gel was stained with ethidium bromide, de-stained in 0.5·TBE and photographed under UV illumination.

Primer extension experiments (5′ RACE)

5′ RACE analysis was carried out using the First Choice RLM-RACE kit (ThermoFisher Scientific). Briefly, total RNA was purified using the RNeasy Mini Kit (Qiagen, Germany), as indicated above, and treated with RNase-free DNase I (Turbo DNA-free kit, ThermoFisher Scientific) to remove any residual DNA. A total of 4 μg of this RNA were treated with terminator 5′-phosphate-dependent exonuclease (TEX). Reactions were terminated by extraction with acid phenol followed by ethanol precipitation. Samples were treated with tobacco acid pyrophosphatase, and the RNA oligonucleotide 5′ RACE Adapter was ligated to the RNA 5′-ends using the T4 RNA ligase. After obtaining cDNA using random primers and the Super Script III reverse transcriptase (Life Technologies), a specific amplification step was performed using the 5′ RACE Outer primer in combination with either the specific reverse primer 5′ RACE lacZ-rev-Outer (for promoter P_{23c}) or with primer 5′ RACE lacZ-rev-Inner (for promoters P_{20c} and P_{27c}). The resulting products were cloned into vector pGEM-T easy (Promega) and transformed into *E. coli* NEB-5α competent cells (Biolabs). The 5′-end of the transcripts (corresponding to the point of ligation) was determined by DNA sequencing of several independent transformants.

Ultracentrifugation

Sedimentation velocity assays (SV). Samples of protein Reg_{576} alone and mixed with the different DNA variants, in 20 mM Tris, 500 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 0.1 mM β-mercaptoethanol and 1% glycerol, pH 7.4, were loaded (320 μl) into 12 mm epon-charcoal standard double-sector centerpieces. The assays were performed at 48 000 rpm in a XL-I analytical ultracentrifuge (Beckman-Coulter Inc.) equipped with both UV-VIS absorbance and Raleigh interference detection systems, using an An-50Ti rotor. Sedimentation profiles were recorded simultaneously by Raleigh interference and absorbance at 230 and 260 nm. Differential sedimentation coefficient distributions were calculated by least-squares boundary modeling of sedimentation velocity (SV) data using the continuous distribution *c(s)* Lamm equation model as implemented by SEDFIT (23). These experimental *s*-values were corrected to standard conditions using the program SEDNTERP (24) to obtain the corresponding standard *s*-values ($s_{20,w}$). Multi-signal sedimentation velocity (MSSV) data were globally analyzed by SEDPHAT software (25) using the 'multi-wavelength discrete/continuous distribution analysis' model, to determine the spectral and diffusion-deconvoluted sedimentation coefficient distributions, $c_k(s)$, from which the number and stoichiometry of protein versus DNA molecules can be derived (26). Prediction of extinction coefficients for the different DNA fragments considering duplex hypochromism at 260 nm was done by means of the Microsoft Excel® application developed by A. Tautorov (27).

Sedimentation equilibrium assays (SE). Short columns (95 μl) SE experiments of Reg_{576} alone and mixed with the different DNA fragments were carried out at speeds ranging from 5500 to 22 000 rpm and at three different wavelengths (230, 260 and 280 nm), using the same experimental conditions and instrument as in the SV experiments. A last high-speed run (48 000 rpm) was done to deplete protein and DNA from the meniscus region to obtain the cor-

responding baseline offsets. Weight-average buoyant molecular weights of Reg_{576}, DNA fragments and Reg_{576}–DNA complexes were obtained by fitting a single-species model to the experimental data using the HeteroAnalysis program (28), once corrected for temperature and solvent composition with the program SEDNTERP (24). The equilibrium binding isotherms of Reg_{576} with the different DNA variants were built using a fixed DNA concentration of 0.1 μM titrated with increasing Reg_{576} concentrations (from 0.1 to 3 μM). The amount of Reg_{576} bound to the different DNA fragments was determined from the experimental apparent buoyant mass increments, using 0.7366 as partial specific volume for Reg_{576}, calculated from its amino acid composition by SEDNTERP. The data were modeled with a three parameter Hill function, as implemented in SigmaPlot 11.0 software.

Dynamic light scattering assays (DLS). Dynamic light scattering (DLS) experiments were carried out in the same experimental conditions used in SV and SE, in a Protein Solutions DynaPro MS/X instrument (Protein Solutions, Piscataway, NJ, USA) using a 90° light scattering cuvette. Previous to measurements, samples were centrifuged during 10 min at 12 000 x *g*. Data were collected and analyzed with Dynamics V6 Software.

Estimate of molar mass of Reg_{576} from hydrodynamic measurements. The apparent molar mass of a single sedimenting solute species (M) was also calculated using measured values of the sedimentation coefficient *s* and the diffusion coefficient D according to the Svedberg equation (29),

$$M = \frac{RTs}{(1 - \bar{\nu}\rho)D} \quad (1)$$

where T, R, $\bar{\nu}$ and ρ stand for the absolute temperature, the universal gas constant, the partial specific volume of the protein, and the density of the solution, respectively.

Mathematical modeling

Basal gene expression was neglected and linear degradations were considered for all species. Repression was modeled through effective Hill functions using the Hill constants given by the number of monomers of repressor binding the promoters: in our model Reg_{576} binds promoter P_{20c}, which has two operator regions, with a Hill exponent of $h_e = 8$ and Reg_{576} represses its own expression with a Hill exponent of $h_r = 4$. Following experimental results, we consider the equilibrium constants $K_r = 0.7$ for the repressor and $K_e = 0.8$ for the establishment genes. Denoting time by t we obtain the following system of ordinary differential equations:

$$\frac{dm_r(t)}{dt} = \frac{a}{1 + \left(\frac{p_r(t)}{K_r}\right)^{h_r}} - d_m m_r(t) \quad (2)$$

$$\frac{dp_r(t)}{dt} = bm_r(t) - d_p p_r(t) \quad (3)$$

$$\frac{dm_e(t)}{dt} = \frac{3a}{1 + \left(\frac{p_r(t)}{K_e}\right)^{h_e}} - d_m m_e(t) \quad (4)$$

$$\frac{dp_e(t)}{dt} = bm_e(t) - d_p p_e(t) \quad (5)$$

where m_r is the concentration of repressor mRNA, p_r is the concentration of repressor protein, m_e is the concentration of establishment gene mRNA, p_e is the concentration of establishment gene protein, a is the repressor's maximum transcription rate, d_m its mRNA degradation rate, b is the repressor's translation rate and d_p its protein degradation rate. For simplicity we consider the same degradation and translation rates for the establishment gene; the qualitative nature of our results does not depend on this choice or on the specific values of the parameters we use. According to our results, we use a maximum transcription rate for the establishment genes that is three times higher than that of the reg_{576}. We have chosen the following dimensionless parameter values: $a = 1$, $d_m = 0.1$, $b = 0.3$, $d_p = 0.03$ and solved the equations using the ode45 solver from Matlab (30). We have no dynamical gene expression data to compare with the model predictions. Therefore, we have left units undefined, and use the model as a tool of qualitative exploration of the implications of the regulatory logic found in our experiments. Note however, that we used as equilibrium constants the results of our measurements in μM, and that the parameters which we have chosen for degradation rates are consistent with biological realistic rates in 1/min. A trivial scaling and fitting of parameter values allows us to fit the model to potential future experimental data to obtain the desired time scale for the pulses of expression observed in the model.

Crystallization and structure determination

Purified protein was concentrated to 25 mg/ml in [20 mM Tris (pH8.0), 250 mM NaCl]. Crystals of Reg_{576} were obtained by the sitting-drop vapor-diffusion method at 18°C, by equilibration of drops of 1 μl protein + 1 μl crystallization buffer (130 mM potassium bromide, 25% PEG 2000 monomethylether) against 50 μl of the crystallization buffer. Needle-shaped crystals appeared in less than a week. Cryo-cooling in liquid nitrogen was carried out using a cryo-protecting solution containing reservoir solution supplemented with 5% glycerol. Data collection was performed to the indicated resolutions at ALBA synchrotron Light Source on the BL13-Xaloc beamline (31). The crystals belonged to space group $P4_32_12$, with two protein molecules in the asymmetric unit. Data were processed with AutoPROC from Global Phasing (32). Data-collection statistics are listed in Supplementary Table S4.

Structure refinement

In silico phasing was performed using Arcimboldo Borges (33) Crystallographic refinement using refmac_5.8.0222 (34) was interspersed with manual building in Coot (35), using the 2Fo−Fc and Fo−Fc electron-density maps from refinement. Molprobity (36) was used to validate and improve the final model. Refinement statistics are presented in Supplementary Table S4. Figures were prepared using Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) and PDBSum (37, Database issue D355-D359).

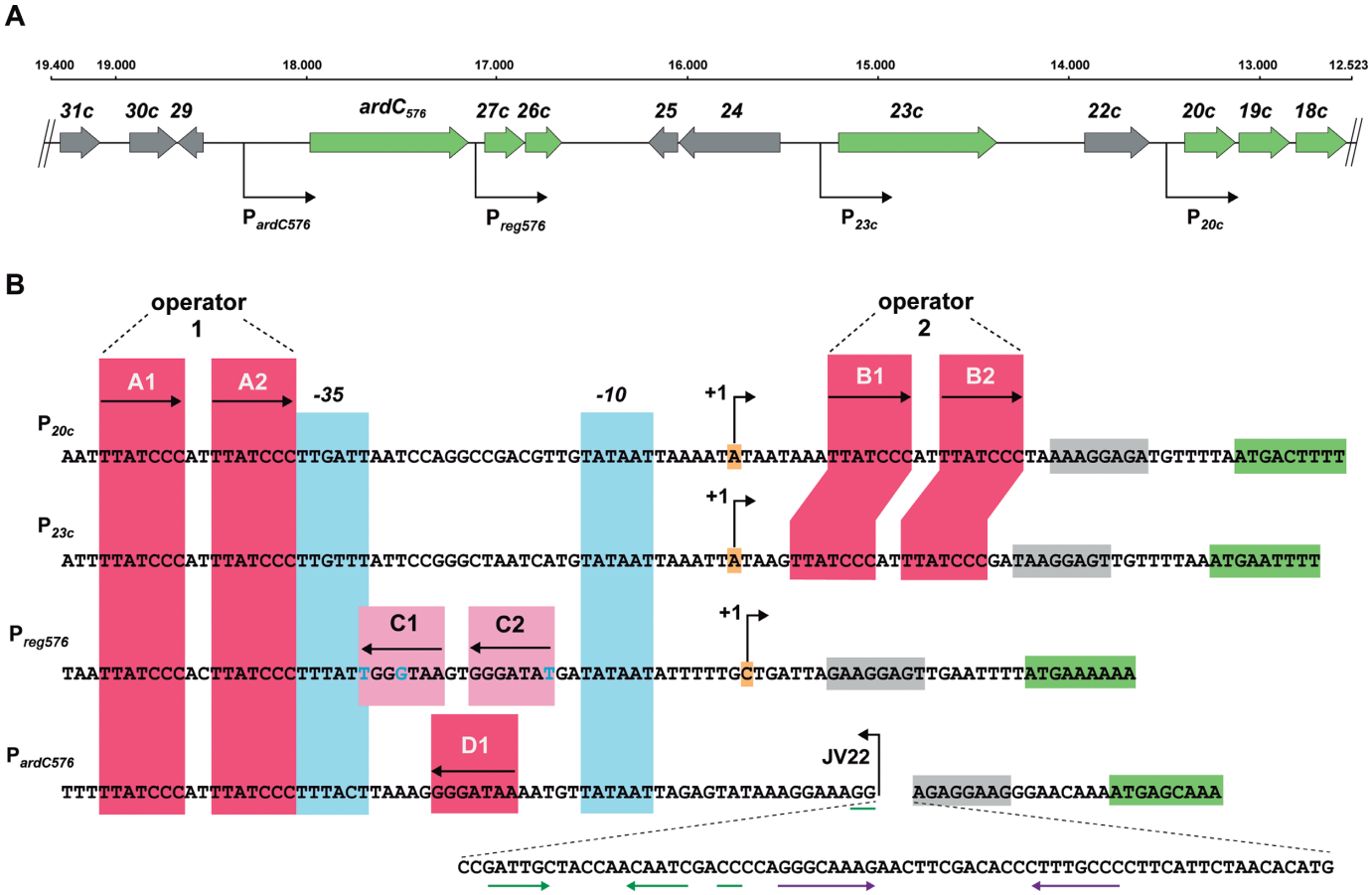


Figure 1. Heptamer sequences 5′-TTATCCC-3′ are located near the p576 promoters P_{20c}, P_{23c}, P_{reg576} (P_{27c}) and $P_{ardC576}$. (**A**) Genetic organization of p576 region encompassing genes *18–31* corresponding to plasmid positions 12 500–19 400 (given at top). Locations of promoters are indicated with bent arrows. Genes are indicated with wide arrows. Genes controlled by putative promoters P_{20c}, P_{23c}, P_{27c} (P_{reg576}) and $P_{ardC576}$ are shown in green, others in gray. (**B**) Sequences and features of promoters P_{20c}, P_{23c}, P_{27c} (P_{reg576}) and $P_{ardC576}$. Predicted RBSs are shown on a gray background. Coding sequences are highlighted in green. Hexamer sequences showing similarity to the −35 and −10 boxes of σ^A-type promoter are given in blue. Heptamer sequences 5′-TTATCCC-3′ and similar sequences containing up to two mismatches are highlighted in dark and light red, respectively. Mismatches with respect to the consensus heptamer sequence are shown in green. The inverted repeated sequences downstream of $P_{ardC576}$ predicted to form stem-loop structures when transcribed into RNA are indicated with convergent green and purple arrows. The backward directed bent arrow marked JV22 indicates the deletion end point present in $P_{ardC576\Delta75}$ lacking the inverted repeated sequences. Determined transcription start sites (see text) are highlighted in orange, indicated with bent arrows and labeled '+1'.

RESULTS

Identification of an anti-restriction gene and other putative establishment genes of p576, and detection of conserved features in their upstream regions

Plasmids p576 (43.4 kb) of *B. pumilus* NRS576 and pLS20 (64.8 kb) of *B. subtilis* natto UM3335 are related conjugative plasmids (16,38). Their replication regions, which share 72% identity, are flanked by conserved genes/operons: a bicistronic operon encoding a plasmid partitioning system is located upstream of each origin (16,39), and three regulatory genes followed by a large conjugation operon are located downstream of each origin (see Supplementary Figure S1, 16,40). However, the proteins encoded by the genes present on the remaining regions of both plasmids do not share significant similarity with proteins of known function, with one exception: gene *28c* of p576 gene and gene *82c* of pLS20cat each encodes a putative protein that shares 48 and 51% similarity, respectively, with the type C anti-restriction protein encoded by the *Shigella flexneri* broad host range plasmid pSa (see Supplementary Figure S2) (41). Accordingly, we named gene *28c* of p576 $ardC_{576}$, and gene *82c* of pLS20cat $ardC_{LS20}$. Anti-restriction genes are typical establishment genes that play important roles in the establishment of the plasmid in the recipient after its conjugative transfer. As explained in the introduction, establishment genes must be regulated such that they are expressed rapidly, but transiently, upon entry of the plasmid into the recipient. We analyzed the upstream sequences of $ardC_{576}$ for clues of possible regulatory mechanisms. This analysis revealed a putative σ^A-dependent promoter (5′-TTtACt-n14-TGnTATAAT-3′) that we named $P_{ardC576}$, which is located 105 bp upstream of the start codon of $ardC_{575}$ (see Figure 1B). Thus, (i) the 5′-TTtACt-3′ and the 5′-TATAAT-3′ hexamers are similar and identical to the consensus −35 (5′-TTGACA-3′) and −10 (5′-TATAAT-3′) boxes of σ^A type promoters; (ii) these putative boxes are separated by the optimal spacer length of 17 bp; and (iii) it contains a so-

called extended −10 motif (5′-TGn-3′) located upstream of the −10 box. Interestingly, three identical heptamer motifs (5′-TTATCCC-3′) are located near this putative promoter (see Figure 1B). Two motifs, named 'A1' and 'A2', are located immediately upstream of the putative −35 box. These two motifs have the same orientation and are separated by 2 bps. The third copy of this motif, which is named 'D1' and has the opposite orientation with respect to the other two, is located 11 bps downstream of motif 'A2' and hence is located in the spacer region of the putative promoter (see Figure 1B). Intriguingly, further analysis revealed that p576 contains in total 18 copies of this heptameric sequence. Five of these are located inside the coding region of five different (putative) genes, while the remaining 13 copies are located upstream of gene *20c* (four motifs), *23c* (four motifs), *27c* (two motifs) and $ardC_{576}$ (three motifs). These four genes are clustered in a 7.9 kb region of p576 and are all transcribed in the opposite direction to that of the conjugation genes (see Figure 1A). Similar to that observed for $P_{ardC576}$, the motifs are located near putative σ^A-dependent promoters, which we named P_{20c}, P_{23c} and P_{27c}. Whereas the putative promoters $P_{ardC576c}$ and P_{23c} are predicted to drive expression of one gene, putative promoters P_{27c} and P_{20c} are predicted to control the expression of two and three genes, respectively (Figure 1A). A schematic overview of the distribution of the conserved motifs with respect to their putative promoter is shown in Figure 1B. Strikingly, two of the mentioned heptameric motifs are located in a directed orientation immediately upstream of the −35 box of each of these four putative promoters. We will refer to such a pair of direct-repeated motifs as 'dual motif' (DM). In the case of putative promoters P_{20c} and P_{23c} another DM is located just upstream of their predicted ribosomal binding site (RBS). The putative P_{27c} promoter also contains an additional DM. However, in this case the motifs deviate slightly from the consensus sequence (5′-aTATCCC-3′ and TTAcCCt-3′, indicated as 'C1' and 'C2', respectively, in Figure 1B). In addition, this DM is located within the spacer region separating the −35 and −10 hexamers of the putative promoter and their orientation is opposite to that of the DM located upstream of the predicted −35 box. The various DMs are all separated by the dimer sequence 5′-AT-3′, except for the consensus motifs near putative P_{27c}, which are separated by 5′-AC-3′.

Working model explaining transient expression of $ardC_{576}$ and other putative establishment genes after transfer of plasmid p576 into a new host

As explained above, $ardC_{576}$ has the features of a typical establishment gene. The direct repeated motifs near the putative promoters may be the binding sites for a regulatory protein. If correct, this indicates that the same transcriptional regulator controls the activities of promoters P_{20c}, P_{23c}, P_{27c} and P_{ard576}. This, in turn, suggests that the genes controlled by these four promoters are members of the same regulatory network and that the encoded proteins would all play a role in plasmid establishment. The presence of identical or similar motifs near the putative promoters could hold the clue in achieving the temporal expression of these genes. Based on these assumptions we devised the following working model. Plasmid p576 encodes a regulatory protein that has affinity for the 5′-TTATCCC-3′ motif, and upon binding the regulatory protein would interfere with transcription from the nearby promoters and hence impair expression of the downstream-located establishment genes in cells harboring plasmid p576. During conjugation an ssDNA strand of the plasmid, but not the presumed regulatory protein, will be transferred into the new host. Once the ssDNA is converted into double-stranded plasmid DNA, the four promoters will be active resulting in expression of the establishment genes. Simultaneously, the gene encoding the repression protein will be expressed. Once sufficient level of the repressor protein has been synthesized the promoters will be shut off again. Consequently, the promoters will be expressed during a short window of time soon after entry of the DNA into the recipient. Experiments were designed to test this model and their results are presented below.

The upstream regions of p576 genes *20c*, *23c*, *27c* and $ardC_{576}$ encompassing the heptamer sequences contain active promoters

To study whether the p576 genes *20c*, *23c*, *27c* and $ardC_{576}$ are preceded by a promoter their upstream regions were each cloned in front of the *lacZ* reporter gene, and a single copy of each resulting fusion was placed at the *B. subtilis* chromosomal *thrC* locus. As a control, strain PKS7 was constructed, which contains a copy of a promoter-less *lacZ* at *thrC* (see 'Materials and Methods' section). As a first qualitative approach, colonies of the constructed strains were tooth picked onto LB agar plates supplemented with the chromogenic substrate Xgal. After overnight growth, all the colonies were blue except that of the negative control strain PKS7 (Figure 2A), demonstrating that each of the cloned fragments contains a functional promoter. The intensity of the blue color was similar for promoters P_{20c}, P_{23c} and $P_{ardC576}$, while that of promoter P_{27c} was slightly lower.

To localize the positions of the promoters we determined the transcription start sites of promoters P_{20c}, P_{23c} and P_{27c} by RNA ligase-mediated rapid amplification of 5′ cDNA ends (5′-RACE, see 'Materials and Methods' section). The obtained transcription start sites are indicated in Figure 1B and confirmed that the promoter positions predicted above are correct.

We next determined the apparent strength of these four promoters by determining the β-galactosidase activity (βGA) of late exponentially growing cultures of the strains (Figure 2B). As expected, very low levels of βGA were obtained for the negative control strain PKS7. Similar levels of βGA, -in the range of 270–325 Miller units-, were obtained for strains containing *lacZ* fused to promoters P_{20c}, P_{23c} and $P_{ardC576}$. The βGA of strain DG127c containing the P_{27c}-*lacZ* fusion was about three times lower, though. The sequences of promoters P_{20c} and P_{23c} are alike; their −10 boxes are identical, while their −35 boxes are very similar to the consensus σ^A-type promoter boxes, (see Figure 1B). Hence, it was not surprising that they displayed a similar high promoter activity. However, the relatively high activity of promoter $P_{ardC576}$ was surprising considering that its −35

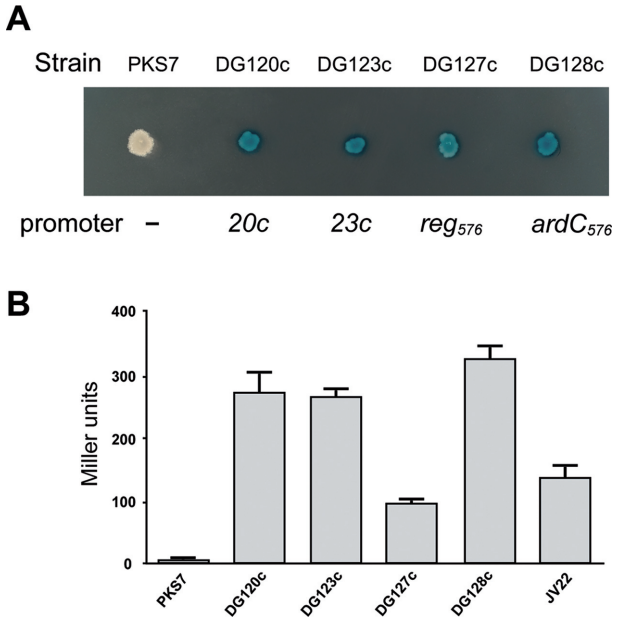


Figure 2. Promoters are located upstream of p576 genes *20c*, *23c*, *27c* (reg_{576}) and $ardC_{576}$ (*28c*). (**A**) Cells of *Bacillus subtilis* strains containing a chromosomal cassette harboring a copy of *lacZ* without being preceded by p576 sequences (PKS7), or preceded by upstream sequences of p576 gene *20c* (DG120c), *23c* (DG123c), *27c* (reg_{576}) (DG127c) or $ardC_{576}$ (DG128c) were spotted onto an LB-agar plate supplemented with Xgal and incubated overnight at 37°C. (**B**) The strength of promoters P_{20c}, P_{23c}, P_{27c} (P_{reg576}), $P_{ardC576}$, was determined by β-galactosidase assays of cell samples taken at the end of the exponential growth phase ($OD_{600} = 1$) of strains containing the corresponding transcriptional *lacZ* fusions. Strain PKS7 was included as a negative control. The data, expressed as Miller Units, correspond to the mean value of three independent experiments. T-symbols represent the standard deviation. PKS7, negative control strain; DG120, DG123, DG127, DG128 are 168-derivatives containing at their *thrC* locus copy of *lacZ* fused to the upstream sequences of p576 gene *20c*, *23c*, *27c* (reg_{576}) and *28c* ($ardC_{576}$), respectively. JV22 is a derivative of strain DG128c in which the 75 bp region located immediately upstream of the start codon of gene $ardC_{576}$ encompassing three inverted repeated sequences (see Figure 1B) is deleted.

box, like that of promoter P_{27c}, is more deviant from the consensus sequence. The apparent high activity of this promoter could be due to the generation of transcripts having a rather long untranslated region that might form secondary structures (note that the distance of the predicted −10 box to the RBS is about 70 bp longer for promoter $P_{ardC576}$ than the other three promoters, see Figure 1B) (42,43). To study this hypothesis, we constructed a derivative of strain DG128c lacking the 75 bp DNA located immediately upstream of the start codon of gene $ardC_{576}$ encompassing the inverted repeated sequences. We referred to this modified promoter as $P_{ardC576\Delta75}$. Colonies of the strain harboring cassette $P_{ardC576\Delta75}$-*LacZ* (strain JV22) were still blue when grown on LB agar plates containing Xgal, demonstrating that promoter $P_{ardC576}$ is located upstream of the region deleted in JV22 (not shown). However, βGAs assays showed that the strength of promoter $P_{ardC576\Delta75}$ was about 2-fold lower than that of $P_{ardC576}$ and similar to that of promoter P_{27c} (Figure 2B). These results confirm that the 75 bp upstream of gene $ardC_{576}$ significantly enhances expression of the downstream gene.

p576 gene *27c* encodes a repressor of promoters P_{20c}, P_{23c} and $P_{ardC576}$ and regulates its own expression

To study if p576 encodes a repressor of promoters P_{20c}, P_{23c}, P_{27c} and $P_{ardC576}$ we aimed to introduce the plasmid into the above *lacZ*-reporter strains for these promoters. However, p576 does not contain an antibiotic resistance or other markers for its selection in *B. subtilis*. Despite several attempts and using different approaches we did not succeed in labeling the p576 plasmid with an antibiotic resistance marker. We did obtain, though, a derivative of p576 in which the largest of the three EcoRI fragments, corresponding to approximately the 3′ half of the conjugation operon, was replaced by a chloramphenicol resistance gene. We introduced this derivative, which we named pBCM1, into the *B. subtilis* strain harboring the P_{20c}-*lacZ* fusion, resulting in strain BCM1. Colonies of strain BCM1 (P_{20c}-*lacZ*, pBCM1) were white when grown on LB plates containing Xgal (not shown). This result strongly indicates that p576 encodes a repressor of promoter P_{20c}.

The conjugation operons and the surrounding regions of p576 and pLS20 are similarly organized. The conjugation operon of pLS20 is repressed by default, and only under appropriate conditions the conjugation genes are activated during a rather short window of time (40). Based on this, we reasoned that it was unlikely that the repressor gene would be located within the conjugation operon of p576. We also discarded genes located on the large EcoRI fragment that is absent in pBCM1, and genes to which a function not involving DNA binding could be attributed based on similarity with known genes. The deduced proteins of the remaining genes were subjected to *in silico* analyses to identify the presence of putative DNA binding domains. Using this approach genes *20c*, *27c*, *29* and *30c* were identified as candidate repressor genes. To test whether any of these genes encodes the repressor of promoters P_{20c}, P_{23c}, P_{27c} and $P_{ardC576}$ we placed each of these four genes at the chromosomal *amyE* locus under the control of the IPTG-inducible P_{spank} promoter. Gene *27c* was an exception in the sense that this gene was cloned behind the P_{spank} promoter together with its transcriptionally coupled downstream gene *26c*. We then introduced the P_{20c}-*lacZ* reporter fusion into each of these four strains. In parallel, we also introduced into each of the four strains a cassette in which the *lacZ* gene was fused to the upstream region of p576 gene *16*, predicted to contain a promoter but without a 5′-TTATCCC-3′ motif and so would serve as a repressor-insensitive control. As shown in Figure 3A, colonies of all the eight strains, including those carrying the *lacZ* reporter for the upstream region of gene *16*, were blue when grown on agar plates supplemented with only XGal. In the presence of IPTG colonies were also blue except for strain JV5 harboring the P_{spank}-*27c-26c* cassette. These results demonstrate that gene *27c* and/or gene *26c* encodes a regulator that represses P_{20c} but not promoter P_{16}. We next tested whether gene *27c* or *26c* alone was sufficient to repress promoter P_{20c} using cassettes containing either P_{spank}-*27c* or P_{spank}-*26c*. Results presented in Figure 3B show that induction of gene *27c*, but not *26c*, resulted in repression of the P_{20c} promoter. To study whether expression of *27c* also repressed the activities of promoters P_{23c}, P_{27c} and $P_{ardC576}$, we introduced the respective tran-

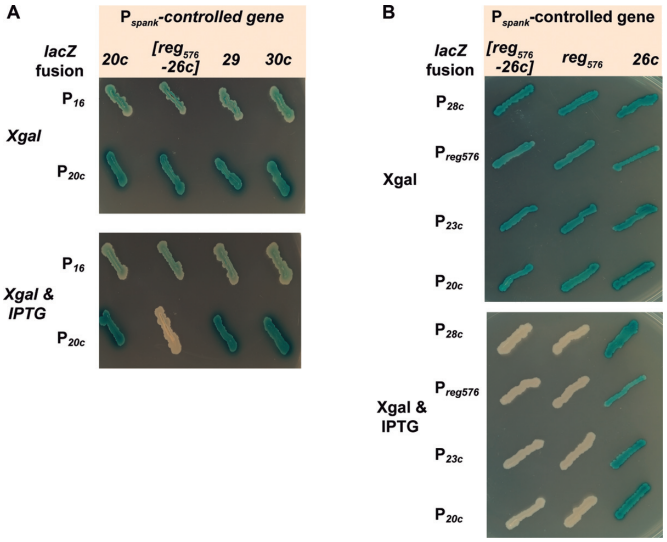


Figure 3. Gene *27c* (*reg*$_{576}$) encodes a regulator that represses promoters P$_{20c}$, P$_{23c}$ and P$_{ardC576}$, as well as its own promoter P$_{27c}$(P$_{reg576}$). (**A**) Strains containing a copy of *lacZ* fused to either promoter P$_{16}$ or P$_{20c}$ in combination with another cassette containing one of the candidate repressor genes *20c*, *27c-26c*, *29* and *30c* under the control of the IPTG-inducible promoter P$_{spank}$ (see Supplementary Table S1) were grown overnight at 37°C on plates supplemented with X-gal, or with Xgal and IPTG. (**B**) Strains containing a copy of *lacZ* fused to either promoter P$_{20c}$, P$_{23c}$ P$_{27c}$ (P$_{reg576}$), P$_{ardC576}$ in combination with another cassette containing both genes *27c* and *26c*, or only gene *27c* (*reg*$_{576}$) or gene *26c* under the control of the inducible promoter *P*$_{spank}$ were grown overnight at 37°C on plates supplemented with X-gal, or with Xgal and IPTG.

scriptional *lacZ* fusion into the strain harboring P$_{spank}$-*27c-26c*, P$_{spank}$-*27c* or P$_{spank}$-*26c* and grew colonies of the constructed strains on Xgal-containing plates supplemented or not with IPTG. Figure 3B shows that induction of gene *27c* repressed, besides P$_{20c}$, also the promoters P$_{23c}$, P$_{27c}$ and P$_{ardC576}$.

Altogether, these results show that p576 gene *27c* encodes a regulatory protein that represses the promoters P$_{20c}$, P$_{23c}$ and P$_{ardC576}$ as well as its own promoter P$_{27c}$, but not promoter P$_{16}$. Based on these results we denominated p576 gene *27c reg*$_{576}$ (**r**epressor of **e**stablishment **g**enes).

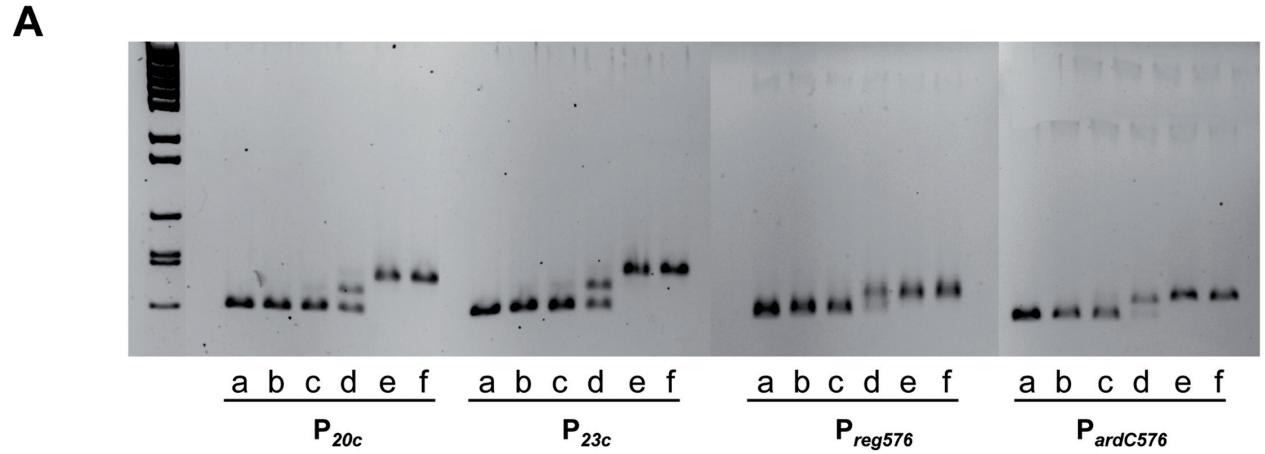
The operator site of Reg$_{576}$ is constituted by a dual 5′-TTATCCC-3′ motif

In vitro and *in vivo* approaches were used to establish the operator site of Reg$_{576}$. To facilitate the purification of Reg$_{576}$ we fused *reg*$_{576}$ in frame to a region encoding a 6- histidine tag at either the C- or N-terminus of *reg*$_{576}$ (see 'Materials and Methods' section). Both recombinant genes, *reg*$_{576}$*his*$_{(6)}$ and *his*$_{(6)}$*reg*$_{576}$, were functional as, when induced, they were able to repress the expression of the P$_{20c}$-*lacZ* fusion in *B. subtilis* (not shown). Recombinant gene *his*$_{(6)}$*reg*$_{576}$ was chosen for further studies and the purified protein, which we referred to as Reg$_{576}$, was used in EMSAs to identify the operator site. As shown in Figure 4A, Reg$_{576}$ clearly bound to each of the four promoter-containing DNA fragments at concentration as low as 42 nM, and all the DNA fragments were shifted at 380 nM, independent of the promoter analyzed. However, there was a fundamental difference between the retardation patterns of promoters P$_{20c}$/P$_{23c}$ and P$_{27c}$/P$_{arcC576}$. Only a single shift was observed for the latter two promoters whereas, depending on the protein concentration applied, up to two shifts were observed for promoters P$_{20c}$ and P$_{23c}$. These results indicate that promoters P$_{20c}$/P$_{23}$ and P$_{27c}$/P$_{ardC576}$ contain two and one Reg$_{576}$ operator sites, respectively.

We then selected promoters P$_{20c}$ and P$_{27c}$ to study the binding of Reg$_{576}$ in more detail by analyzing the effects of mutations in one or more of the repeated heptamer motifs. A summary of the DNA fragments used with the specific mutations and the results of EMSAs is given in Figure 4B; representative results of the corresponding EMSAs are shown in Supplementary Figure S3. As shown above, the wild-type P$_{20c}$ promoter (fragment F-20c) gave up to two retardation species (see also Figure 4A), but the fragments containing mutations in motifs ['A1' and 'A2' (F-mut1A)] or ['B1' and 'B2' (F-mut1B)] gave only a single shift. Furthermore, no retarded species was observed for fragment F-mut1C having mutations in all four motifs 'A1', 'A2', 'B1' and 'B2'. These results provide firm evidence that (i) Reg$_{576}$ binds specifically to the 5′-TTATCCC-3′ motifs and (ii) that motifs ['A1' and 'A2'] and ['B1' and 'B2'] each constitute an operator site. Based on these results we named the DM located upstream of the −35 box operator 1, and the DM upstream of the RBS operator 2. We next analyzed whether mutations in only one of the two motifs of an operator site affected Reg$_{576}$ binding. Only a single retarded species was observed for DNA fragments F-mut1D, F-mut1E, F-mut1F and F-mut1G demonstrating that a functional Reg$_{576}$ operator requires a dual 5′-TTATCCC-3′ motif.

Similar results were obtained for promoter P$_{27c}$. Only one retarded species was observed for fragment F-27c containing the wild-type P$_{27c}$ promoter, but no retardation was observed for fragment F-mut2A containing mutations in motifs 'A1' and 'A2'. These results showed (i) that motifs ['A1' and 'A2'] constitute a functional Reg$_{576}$ operator of promoter P$_{27c}$, and (ii) that the motifs ['C1' and 'C2'], which contain two and one deviations with respect to the consensus sequence, respectively, do not constitute a functional operator. The observation that fragments F-mut2B and F-mut2C gave one and no retarded species, respectively, further substantiates these conclusions.

We next studied how the mutations in the 5′-TTATCCC-3′ boxes analyzed by EMSA affected Reg$_{576}$-mediated repression of the P$_{20c}$ and P$_{27c}$ promoters *in vivo*. For this, *lacZ* fused to each of the wild-type or the mutated DNA regions were introduced into strain JV23 containing a copy of *reg*$_{576}$ under the control of the inducible P$_{spank}$ promoter, and the resulting strains were used to determine the strength of the P$_{27c}$ and P$_{20c}$ promoter without or with different levels of *reg*$_{576}$ induction. An overview of the transcriptional fusions and a graphical representation of the results obtained are given in Figure 5. As shown above as well as below, the motifs 'C1' and 'C2' of P$_{27c}$ do not constitute a functional operator site. Mutations in these motifs, which are located in between the −35 and −10 boxes of promoter P$_{27c}$, drastically affected its promoter activity and hence the response of these mutant derivatives to *reg*$_{576}$ induction could not be tested. Mutations in boxes 'A1', 'A2', 'B1' and/or 'B2' mod-



B

Name	Sequence	Boxes Mutated	shifts
F-20c	TATTTAATTTATCCCATTTATCCCTTGATTAATCCAGGCCGACGTTGTATAATTAAAATATAATAAATTATCCCATTTATCCCTA	–	2
F-mut1A	TATTTAATAATTGCTATTAAAGGCTTGATTAATCCAGGCCGACGTTGTATAATTAAAATATAATAAATTATCCCATTTATCCCTA	A1, A2	1
F-mut1B	TATTTAATTTATCCCATTTATCCCTTGATTAATCCAGGCCGACGTTGTATAATTAAAATATAATAAAATTAGGCATAATTGCATA	B1, B2	1
F-mut1C	TATTTAATAATTGCTATTAAAGGCTTGATTAATCCAGGCCGACGTTGTATAATTAAAATATAATAAAATTAGGCATAATTGCATA	A1, A2, B1, B2	0
F-mut1D	TATTTAATAATTGCTATTTATCCCTTGATTAATCCAGGCCGACGTTGTATAATTAAAATATAATAAATTATCCCATTTATCCCTA	A1	1
F-mut1E	TATTTAATTTATCCCATTAAAGGCTTGATTAATCCAGGCCGACGTTGTATAATTAAAATATAATAAATTATCCCATTTATCCCTA	A2	1
F-mut1F	TATTTAATTTATCCCATTTATCCCTTGATTAATCCAGGCCGACGTTGTATAATTAAAATATAATAAAATTAGGCATTTATCCCTA	B1	1
F-mut1G	TATTTAATTTATCCCATTTATCCCTTGATTAATCCAGGCCGACGTTGTATAATTAAAATATAATAAATTATCCCATAATTGCATA	B1	1
F-27c	AAAATTAATTATCCCACTTATCCCTTTATTGGGTAAGTGGGATATGATATAATATTTTTGCTGATTAG	–	1
F-mut2A	AAAATTAAAATTGCTACTAAAGGCTTTATTGGGTAAGTGGGATATGATATAATATTTTTGCTGATTAG	A1, A2	0
F-mut2B	AAAATTAATTATCCCACTTATCCCTTTATTGCCATTGTTGCAATTGATATAATATTTTTGCTGATTAG	C1, C2	1
F-mut2C	AAAATTAAAATTGCTACTAAAGGCTTTATTGCCATTGTTGCAATTGATATAATATTTTTGCTGATTAG	A1, A2, C1,C2	0

Figure 4. Determination of the operator site of Reg_{576} by EMSA. (**A**) Indications that promoters P_{20c}/P_{23c} and $P_{reg576}/P_{ardC576}$ contain two and one operator, respectively. EMSA results obtained using ∼400 bp DNA fragments (150 ng) encompassing promoters P_{20c} (F-20c), P_{23c} (F-23c), P27c (P_{reg576}, F-27c) or $P_{ardC576}$ (F-28c). –, loaded without protein (lanes 'a'). Increasing concentrations of Reg_{576} were prepared using a 3-fold dilution method, and ranged from 4.7 (lanes 'b') to 1150 nM (lanes 'f'). (**B**) Schematic representation of the DNA fragments of variants of promoter P_{20c} and P_{27c} (P_{reg576}) used as probes and summary of the EMSA results. −35 and −10 promoter boxes are highlighted in blue. 5′-TTATCCC-3′ motifs [A1, A2, B1, B2] and [C1, C2] are highlighted against a red and pink background, respectively. Deviations with respect to the consensus sequence are given in green. Introduced mutations are shown in purple.

erately improved the promoter strength (up to a maximum of 1.8- and 1.5-fold for promoter P_{27c} and P_{20c}, respectively, see Supplementary Figure S4). In agreement with the results presented in Figure 3, in the presence of IPTG the activities of the wild-type promoters P_{20c} and P_{27c} were affected although they responded differentially to the increasing levels of reg_{576} induction. When grown in the presence of 4 μM IPTG, a clear reduction in promoter activity was observed for promoter P_{20c} but not P_{27c}. In the presence of 8 μM IPTG the activities of promoter P_{20c} and P_{27c} dropped about 5- and 2-fold, respectively; and at 32 μM of IPTG the activities were reduced by about 20-fold for P_{20c} and 7-fold for promoter P_{27c}. This indicates that P_{20c} is repressed more strictly and at lower Reg_{576} concentrations than the P_{27c} promoter. P_{20c} and P_{27c} derivatives containing mutations in operator 1 (mut1A and mut2A, respectively) also responded differently to reg_{576} induction. Whereas the activity of promoter P_{27c} derivative mut2A was not affected upon reg_{576} induction, the activity of P_{20c} derivative mut1A was lower when reg_{576} was induced. Interestingly, promoter

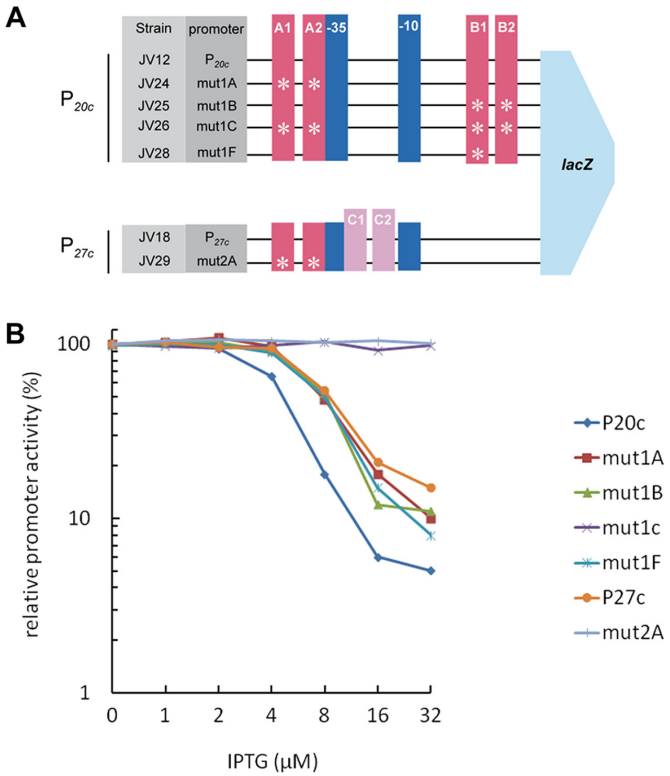


Figure 5. Effects of operator mutations on Reg_{576}-mediated repression of the p576 promoters P_{20c} and P_{27c} (P_{reg576}) *in vivo*. Strains containing transcriptional *lacZ* fusions with the wild-type sequences encompassing promoters P_{20c} or P_{reg576} (P_{27c}) or derivatives having mutations in one or more of the 5′-TTATCCC-3′ boxes were subjected to βGAs to determine their relative promoter strength. Cell samples were withdrawn from late exponential growing cultures ($OD_{600} = 1$). (**A**) Schematic overview of the features of the strains used, and cartoons showing which heptamer boxes are mutated. Mutations are identical to those shown in Figure 4B. βGAs, calculated in Miller Units, were determined for cells growing in the absence or presence of 1, 2, 4, 8, 16 or 32 μM IPTG and are plotted in (**B**) as the relative promoter activity of each construct with respect to the promoter activity measured in the absence of IPTG. Obtained Miller Units are given in Supplementary Figure S4.

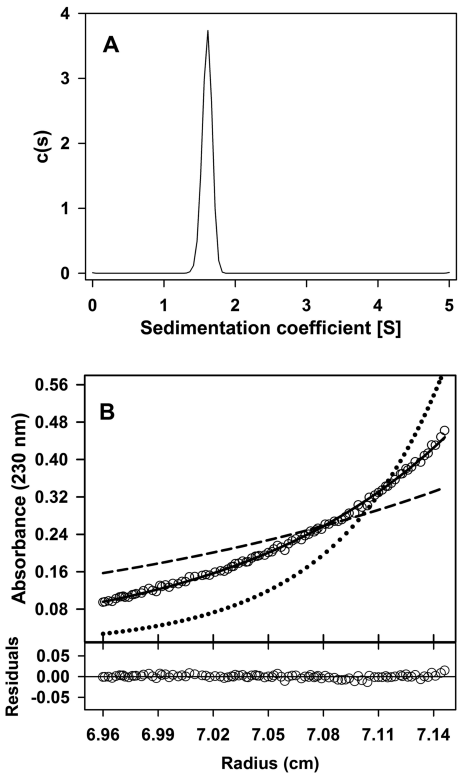


Figure 6. Reg_{576} forms dimers *in vitro*. Oligomerization state analysis of purified Reg_{576} at 3 μM in solution by analytical ultracentrifugation. (**A**) Sedimentation coefficient distribution profile obtained by SV showing a single sedimenting species. (**B**) Concentration gradient obtained by SE. Data (empty circles) are shown together with best-fit analysis assuming a protein dimer (solid line), monomer (dashed line) and tetramer (dotted line). The lower plot shows the difference between experimental data and estimated values for the protein dimer model (residuals).

P_{20c} derivative mut1A responded similarly to different reg_{576} induction levels as the native P_{27c} promoter. Moreover, a similar partial loss in Reg_{576} sensitivity was also observed for P_{20c} derivative mut1B in which motifs 'B1' and 'B2' are mutated. Full loss of responsiveness to reg_{576} induction was observed for promoter P_{20c} derivative mut1C, in which both motifs of operator 1 as well as those of operator 2 were mutated. Together these results show that the motifs 'C1' and 'C2', of promoter P_{27c}, which contain two and one deviations, respectively, with respect to the 5′-TTATCCC-3′ consensus sequence, do not constitute a functional Reg_{576} operator; and the presence of only one functional operator, constituted by either motifs 'A1-A2' or 'B1-B2', makes that the nearby promoter P_{20c} is less sensitive to Reg_{576}-mediated repression than the native promoter containing two functional operators. These results fully agreed with the EMSA results presented in Figure 4 and demonstrate that binding of Reg_{576} to the operator sequences is required for *in vivo* functionality.

At last, promoter P_{20c} derivative mut1F behaved similarly to the derivatives mut1A and mut1B. These results again are in line with the EMSA results and confirm that mutating only one of the two juxtaposed motifs destroys the functionality of the operator.

Protein Reg_{576} is a dimer in solution

To determine its oligomerization state *in vitro* we subjected purified Reg_{576} to analytical ultracentrifugation sedimentation velocity (SV, Figure 6A) and dynamic light scattering (DLS) analyses in parallel. In the SV assays, Reg_{576} at concentrations ranging from 3 to 124 μM was observed always as a single species with an experimental sedimentation coefficient of 1.6 S. This value, corrected to standard conditions ($s_{20,w} = 1.8$ S), was compatible with the theoretical mass of the nearly globular Reg_{576} dimer (Figure 6A). DLS analysis of Reg_{576} yielded a translational diffusion coefficient (D) of $7.16 \times 10^{-7} \pm 0.04$ cm^2/s, which introduced with the obtained *s*-value of 1.6 S into the Svedberg equation, resulted in an apparent molar mass of 22 093 Da, very close to the molecular mass of Reg_{576} dimers (21 779 Da). Finally, sedimentation equilibrium (SE) assays with Reg_{576} at concentrations ranging from 3 to 30 μM showed a buoyant mass of 5398 Da, corresponding to a molar mass of 21 840 ±

180 Da, matching the molecular weight of the protein dimer (Figure 6B). Together, these data show that Reg_{576} forms dimers in solution.

Two Reg_{576} dimers bind to one operator

Results presented above show that (i) a functional operator *in vivo* is constituted by a dual 5′-TTATCCC-3′ motif; (ii) Reg_{576} is unable to bind *in vitro* to an operator containing mutations in one of the two motifs and (iii) Reg_{576} is a dimer in solution. Based on these results it seemed likely that one protomer of a Reg_{576} dimer would bind to one 5′-TTATCCC-3′ motif, and the other protomer to the flanking motif. To test this prediction we performed multi-signal sedimentation velocity (MSSV) experiments using samples containing a DNA fragment without Reg_{576} or with a >100-fold excess of Reg_{576} (0.1 versus 11 μM). The fragments used were the 196 bp DNA F-P20c encompassing promoter P_{20c} and its two flanking Reg_{576} operators, or derivatives in which 5′-TTATCCC-3′ motifs have been mutated (['A1' and 'A2' (mut1A)], ['A1', 'A2', 'B1', 'B2' (mut1C)] or ['A1' (mut1D)], see Figure 7A for a schematic view of the DNA fragments used). Single absorbance plots for the DNA fragments without or with Reg_{576} (OD_{260} nM) are shown in Figure 7B. As expected, the different DNA fragments in the absence of Reg_{576} showed the same *s*-value of 5.0 S. The presence of Reg_{576} did not increase the *s*-value of fragment F-mut1C, which has mutated all the four motifs, but it did result in a major increment (from 5.0 S to 6.3 S) for F-P_{20c} (having both operators intact) and moderate (from 5.0 S to 5.6 S) increments for fragments F-mut1A (mutations in 'A1' and 'A2') and F-mut1D (mutations in only 'A1'). These results show that Reg_{576} did not bind to fragment mut1C, but that it bound to the other three DNA fragments. Based on the increase in *s*-value it seemed that fragments F-mut1A and F-mut1D bound the same or similar amounts of Reg_{576}, and that fragment F-P_{20c} bound more Reg_{576} dimers than mut1A and mut1D. However, care has to be taken with this interpretation since the *s*-value is not only determined by the molecular weight of the complex but also by its size and shape. To fully extract the maximum information enclosed in the SV data, besides the hydrodynamic separation of the complexes, we took advantage of the simultaneous absorbance data acquisition at 230 and 260 nm and globally analyzed them through SEDPHAT to get the diffusion-deconvoluted sedimentation coefficient distributions with spectral deconvolution of the absorbance signals, $c_k(s)$. Further improvement of the molar ratio resolution was achieved by using both, mass conservation constraint and multi-segmented model restriction, using our prior knowledge that only free Reg_{576} sediments in the low-*s* region and no free DNA can be found from 1 S to 3 S. The MSSV analysis of Reg_{576}–F-P_{20c} complex indicated that the areas under the peaks corresponded to a stoichiometry of 7.9 moles of Reg_{576} bound per mol of F-P_{20c}. The ratios of Reg_{576} moles with respect to DNA fragments F-mut1A and F-mut1D were 4.3 and 3.7, respectively (see Figure 7C). These results are in line with those obtained in the EMSAs using the maximum amounts of Reg_{576}. Thus, they confirm that (i) Reg_{576} binding is abolished when all four motifs are mutated; (ii) operator 1 and 2 are both binding sites for Reg_{576}; and (iii) mutation of one motif of an operator abolishes binding of Reg_{576} to that operator even at a >100-fold excess of Reg_{576} over DNA.

Contrary to our expectations though, these results indicated that one 5′-TTATCCC—3′ motif is bound by two Reg_{576} monomers (i.e. one 5′-TTATCCC—3′ motif is bound by a Reg_{576} dimer). To assure that the calculated stoichiometries determined by MSSV were correct we performed SE analytical ultracentrifugation experiments. This technique permits determining the exact molecular weight of the complexes and, since the molecular weight of the DNA and protein components are known, the obtained values can be used directly to calculate the number of protein molecules bound to a DNA molecule. We selected DNA fragments F-P_{20c}, F-mut1A and F-mut1D for this analysis. The sedimentation behavior of the different DNA fragments alone fitted well with a single sedimenting species model with apparent buoyant mass of 57 000 ± 380 Da, in good agreement with the expected molecular weight for these DNA fragments (data not shown). As explained in material and methods, a fixed DNA concentration of 0.1 μM was titrated with increasing Reg_{576} concentrations (from 0.1 to 3 μM). Figure 7D shows the binding isotherms built from the experimental buoyant mass increments obtained at low speed and 260 nm, through an empirical three parameters Hill plot (Equation 6):

$$y = \frac{a x^b}{K_d^b + x^b} \quad (6)$$

Where *y* stands for the number of proteins bound per DNA, *a* denotes the maximum number of proteins bound at saturation, *x* is the total concentration of protein, K_d is the concentration of half-maximal binding and *b* is an empirical cooperativity parameter.

Taking into account the complex stoichiometry experimentally determined for Reg_{576}–F-P_{20c}, an apparent cooperative model ($b = 7.8$) can explain the experimental binding isotherm, with a macroscopic K_d of 0.8 ± 0.1 μM. Analogously, for Reg_{576}–F-Mut1A and Reg_{576}–F-Mut1D complexes, an apparent cooperative model (b = 4.1 and 3.2, respectively) can account for the binding isotherm with a macroscopic K_d of 0.7 ± 0.1 μM. These results, together with those obtained by MSSV technique and taking into account that Reg_{576} is a dimer in solution, demonstrate unequivocally that two Reg_{576} dimers bind to one functional operator. In addition, these results confirm that mutation of only one of the two motifs of an operator abolishes binding of both Reg_{576} dimers to the operator. Altogether, the results obtained confirmed that Reg_{576} binds its operator as a dimer of dimers.

Mathematical modeling shows that high cooperativity and differential repression are key for tight temporal regulation of establishment genes

To explore the qualitative dynamics of gene expression that our experimental findings imply, we designed a minimal mathematical model for the regulation of the putative p576 establishment genes based on modeling the dynamics of the mRNA and protein of the repressor Reg_{576}, and those of

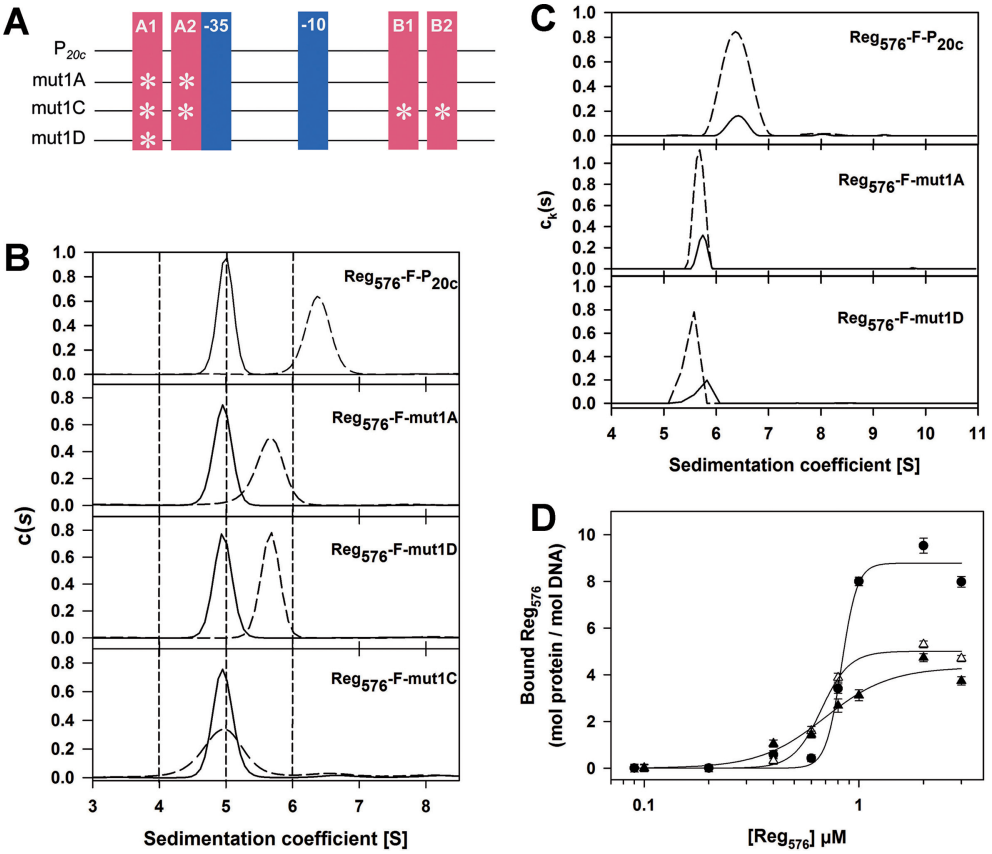


Figure 7. Evidence that Reg_{576} binds its operator as a dimer of dimers. (**A**) Cartoons illustrating with asterisks which of the 5′-TTATCCC-3′motifs are mutated in the 196 bp DNA fragments used. Mutations in the motifs are identical as those shown in Figure 4B. (**B**) Sedimentation coefficient distributions, c(*s*), obtained from SV assays at 260 nm with the different DNA fragments in the presence or absence of Reg_{576}, showing the shift in the *s*-value of Reg_{576}–DNA complexes (dashed trace) relative to the corresponding DNA alone (solid trace). To facilitate comparison of the peak positions only the *s*-range encompassing the DNA and Reg_{576}–DNA complexes is shown. (**C**) Global multi-wavelength analysis of Reg_{576}–DNA complexes and decomposition into component sedimentation coefficient distributions, $c_k(s)$, for Reg_{576} (dashed trace) and the different DNA fragments (solid trace). For clarity, when comparing the areas under the peaks ascribed to the complexes the low-*s* range where only Reg_{576} sediments is not shown. (**D**) Binding isotherms for the interaction of Reg_{576} with F-P_{20c} (black circles), F-mut1A (white triangles) and F-mut1D (black triangles). The solid curves represent the best fit of the three-parameters Hill equation to the SE experimental data.

the putative establishment gene *20c*. Since our data indicate that gene *23c* is similarly regulated, we only selected gene *20c* for modeling.

Our model is constructed with a repressor that represses itself, and also represses its targets with a tighter repression than itself (see 'Materials and Methods' section for details). Experimentally, these conditions on the regulation are justified by the high cooperativity with which Reg_{576} binds the wild-type P_{20c} promoter that is flanked by two functional operators, and which is higher than the cooperativity obtained for mutated derivatives, like mut1A, containing only one functional operator. The promoter region of reg_{576} also contains only one functional operator (see above), and results presented in Figure 5 show that promoters P_{27c} and the P_{20c} derivative mut1a respond similarly to Reg_{576}. We therefore assume that Reg_{576} binds its own promoter with a cooperativity that is similar to that determined for the P_{20c} derivative containing one functional operator. Under these conditions, in our model establishment genes have a transient time interval of expression where they are strongly expressed before being shut down by the repressor (see Figure 8). The long time repression of establishment genes is not efficient if there is not enough cooperativity in the binding of the repressor to its targets (Supplementary Figure S5) or if the cooperativity with which the establishment genes are repressed is not significantly higher than the repression of the repressor itself (Supplementary Figure S6), showing that these conditions, consistent with our experimental findings, are key properties of the mechanism regulating the temporal expression of establishment genes. Our model not only shows that the regulatory logic found in our experiments is consistent with transient pulse-like expression of target genes, but more generically shows that an auto-inhibitory motif can play this role. Negative feedback loops have been associated with keeping homeostasis, to speed up response times, taming fluctuations or the emergence of oscillations (44). Multiple binding sites for transcriptional repressors can produce regular bursting and enhance noise suppression (44), but here we show that under the appropriate conditions of high cooperativity and differential repression of targets, they can also produce pulses of expression. Pulses have usually been associated to more complicated regulatory motifs like feed-forward loops (45).

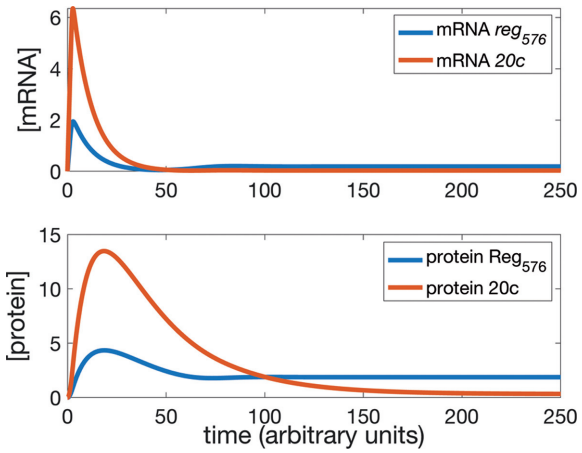


Figure 8. Transient expression of an establishment gene controlled by tight repression from a repressor that represses itself in a looser manner. Time evolution of the mRNA and protein concentrations given by our model for the repressor gene *reg*$_{576}$ and the putative establishment gene *20c*. Parameters are as described in 'Materials and Methods' section. Units are arbitrary. Time = 0 corresponds to the point where double-stranded DNA is reformed and transcription of *reg*$_{576}$ and *20c* starts. The initial values of Reg$_{576}$, protein p20 and mRNA are assumed to be zero.

The crystal structure of Reg$_{576}$ reveals that it is a dimeric Ribbon Helix Helix protein and provides a model of the tetramer bound to a double DNA heptamer

To gain insights into the structural basis of DNA recognition of Reg$_{576}$, we embarked on structure determination of the protein. The overall structure of Reg$_{576}$, determined at 2 Å, revealed that the N-terminal R7-L47 residues or Reg$_{576}$ form a Ribbon-Helix-Helix (RHH, see Figure 9). As is the case for other RHH proteins, the RHH core consists of a dimeric structure in which a double stranded anti-parallel β-sheet is formed by the two β-strands from the monomers (46). Beyond this core structure, Reg$_{576}$ contains a third α-helix (H3, residues 49–65). An additional C-terminal helix is also present in the RHH proteins Mnt (1MNT) and SSV-RH (4AAI) for which structural data are available. However, the placement of this third helix with respect to the core is quite different in Reg$_{576}$ compared to those in Mnt and SSV-RH. In Reg$_{576}$ helix H3 contacts the outward flank of α-helix 1 (H1) of the opposite monomer. In Mnt, the third helix folds back on the H2′, but on the opposite side compared to Reg$_{576}$, interacting with H1′. In the case of the SSV-RH protein, H3 folds back over the outward flank of H2′. The presence of an additional α-helix and its particular position with respect to the RHH core distinguishes Reg$_{576}$ from other RHH proteins. Indeed, we have not found this geometry, with the exception of the structure of a protein from *Nitrosomonas europaea* of unknown function (PDB code: 1ZX3).

To gain insight into the recognition of DNA by Reg$_{576}$, we compared the binding site length and separation of the Reg$_{576}$ recognition sites shown in Figure 1 with RHH proteins for which structures of the protein–DNA complex are available. The length of the Reg$_{576}$ DNA sequence is 7 bps, with a 9 bp separation between the start of the two repeats. This binding mode resembles that of the RHH proteins

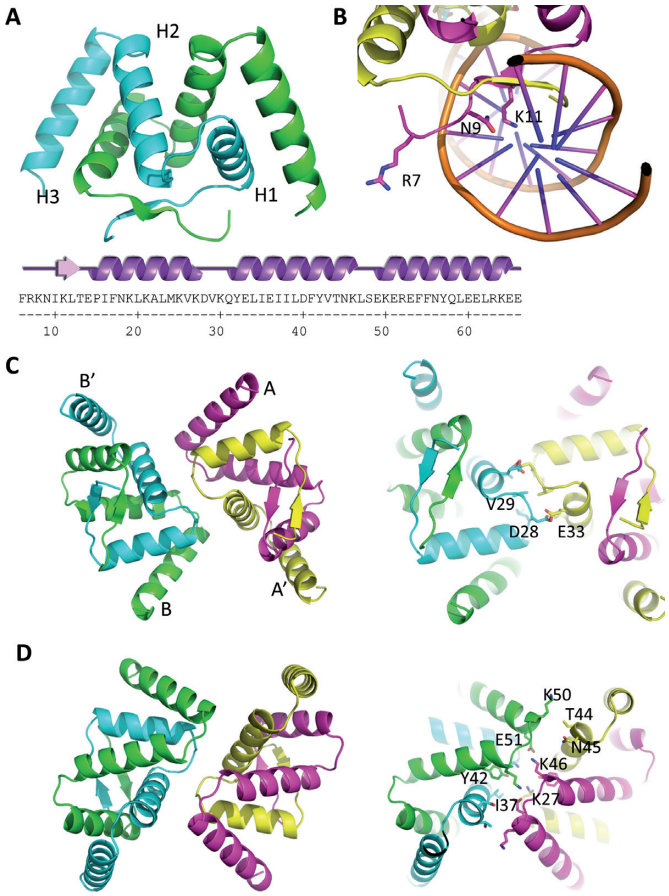


Figure 9. Representations of the structure of Reg$_{576}$. (**A**) Cartoon representation of the Reg$_{576}$ dimer, viewed on the side of H2. The lower panel shows the Reg$_{576}$ sequence and the secondary structure. (**B**) Cartoon representation of the tetrameric model, including the DNA of the Arc repressor as present in the PDB structure '1PAR'. The residues of Reg$_{576}$ involved in base interactions are shown as stick representations. (**C** and **D**) Cartoon representations of the tetramer models of Reg$_{576}$ based on the DNA-bound Arc structure (see text) highlighting the amino acid interactions between dimers (the DNA is not shown for clarity). The left panels show the overall structures from two different views (top view on the β-sheet, lower view on H2). The right panels show the residues that are involved in the dimer-to-dimer interface of the tetramer.

Arc (e.g. 1PAR), CopG (e.g. 1EA4) and AmrZ (3QOQ), which all bind to DNA repeats that are separated by approximately one turn of the DNA double helix. When we superpose two dimers of Reg$_{576}$ on the protein chains of the structure of the Arc repressor complexed to DNA (PDB code: 1PAR), the interactions formed between the Reg$_{576}$ dimers are feasible and chemically sensible, with no strong clashes between residues (Figure 9). To identify the different chains, we follow here the naming scheme as shown in Figure 9, where chains denoted by A and B are from one dimer, and A′ and B′ are the chains from the other dimer, where the prime indicates that the chains are related by C2 symmetry between the dimers of the tetramer, A and A′ being the chains with the β-strands closest to each other in the tetramer. The putative interactions between consecutive Reg$_{576}$ dimers are formed between the H1-H2 loops of monomers from the two different dimers. More precisely, residues V29 and V29′ and residues E33 and D28′, all from

chain B and B′, interact. Residues K27 from chain B interacts with Y42 from A′ and with I37 from B′. In addition, residues from the H2-H3 loop also form an interaction in the tetrameric model. Residues K46 from chain A and E51 from A′ form a salt bridge, and K46 interacts with the backbone residues from K46-S48 from chain A′ through hydrophobic interactions. Additionally, N45 from chain A comes close to the N-terminal residues of helix H3 (S48, E49 and K50). Residues from K46 to E49 constitute the loop that connects helices H2 and H3 of Reg_{576}. Together the model indicates that two Reg_{576} dimers involve more interactions than two Arc dimers. This is also reflected by an increase in the buried accessible surface area, as calculated by the program PISA (47). Whereas the buried surface between two Arc dimers (1PAR structure) is 297.8 Å^2, that of two Reg_{576} dimers is 547.6 Å^2 according our model.

DISCUSSION

Establishment genes contribute to the stable establishment of a conjugative element in the new host. Deregulating expression of these genes or inactivating the activity of the encoded proteins are potential strategies to inhibit conjugation and thereby restraining conjugation-mediated spread of antibiotic resistance. Our knowledge on establishment genes and how they are regulated is scarce, especially for those present on conjugative elements of G+ bacteria. The few establishment genes for which their function has been established include anti-restriction (*ard*) genes, and genes encoding proteins that bind ssDNA (SSB) or suppress the SOS response (PsiB). While the SSB protein is required to coat the single-stranded form of the transferred conjugative element and are important for DNA replication, the anti-restriction (Ard) and anti-SOS (PsiB) proteins improve the success rate of establishment in the recipient cell by inhibiting the restriction enzyme and SOS response, respectively. For proper functioning, these genes must be expressed rapidly upon entry of the conjugative element into the new host on the one hand, and on the other hand their expression must be silenced after a relatively short time. This is because prolonged expression may be harmful for the cell: lengthy inactivation of the restriction proteins makes the cell vulnerable to the entry of other foreign DNA like phage DNA, and extended suppression of the SOS system prevents the cell to respond appropriately to for example DNA damage. Hence, particular regulatory mechanisms must ensure that the genes are expressed rapidly but transiently after transfer of the conjugative element into the recipient cell. Most knowledge on the expression of establishment genes is based on studies of a few conjugative *E. coli* plasmids, especially F and ColIB-P9. The *ssb*, *ard* and *psiB* genes, which are present on enterobacterial plasmids of different incompatibility groups (48,49), are indeed expressed rapidly and transiently upon entry of the plasmid into the recipient cell (9,50,51). Special promoters that are only active when the DNA is in its single-stranded form (11) control these three genes, as well as other genes of unknown function. Consequently, these genes are expressed when the ssDNA strand enters the recipient but their expression is turned off when the immigrating ssDNA has been converted into double stranded DNA.

Here, we found that the p576 promoters P_{20c}, P_{23c}, P_{27c} and $P_{ardC576}$ contain in their vicinity (dual) boxes of the heptamer sequence 5′-TTATCCC-3′. We show that these promoters are similarly regulated due to binding of the Reg_{576} protein to the dual heptamer sequences. These and other results support the view that the Reg_{576}-controlled promoters are derepressed upon transfer of the plasmid into the recipient until sufficient Reg_{576} is produced to silence them again. One of the genes regulated in this way is $ardC_{576}$, whose encoded protein shows about 50% similarity with type C anti-restriction proteins encoded by conjugative plasmids of Gram-negative bacteria that play a role in establishment of plasmids in the recipient cell after conjugative transfer (9,13,16). Based on this, we propose that $ardC_{576}$ encodes an anti-restriction protein that may enhance the establishment of p576 in a recipient cell. It seems reasonable to assume that the other Reg_{576}-controlled genes encode proteins that are functionally related and that they play a role very soon after transfer of the plasmid in the recipient cell, which may include enhancing establishment of the plasmid in the new host.

The mechanism in which a promoter is derepressed until sufficient repressor is synthesized to silence its activity is very similar to zygotic induction, whose term was first used to describe the phenomenon that conjugation-mediated transfer of the *E. coli* prophage λ, but not the repressor of the lytic genes, into a recipient cell lacking the λ prophage resulted in immediate induction of the prophage and subsequent lysis of the 'zygote' (52,53). Establishment genes of the G- broad host-range plasmid pKM101 may be controlled by a zygotic induction-like system. The pKM101 establishment genes are located in seven transcriptional units that are all preceded by >400 bp long conserved DNA regions that may be binding sites for two regulatory proteins, ArdR and ArdK (54). However, the regulatory function of these repeated sequences and the putative regulatory proteins have not been studied in detail. As far as we know, this is the first detailed study of a mechanism that ensures transient expression of establishment genes without using ssDNA promoters. Intriguingly, it appears that the putative establishment genes of the related plasmid pLS20 are regulated in a fundamentally different way. This assumption is based on the observation that sequences highly similar to those that precede its putative anti-restriction gene are present upstream of several other pLS20 genes/operons, suggesting that they are co-ordinately regulated. However, these sequences do not share similarity to any of the upstream sequences of the p576 promoters studied here. It will be interesting to unravel the way these pLS20 genes are regulated.

Members of the superfamily of RHH transcription factors are involved in the regulation of diverse bacterial processes such as amino-acid biosynthesis, cell division, control of plasmid copy number and lytic cycle of bacteriophages (for review see, 46). Recently, we have shown that many auxiliary relaxosome proteins encoded by conjugative plasmids of G+ bacteria are also RHH type proteins (22). Here we show that Reg_{576} too is a RHH type protein, thereby expanding even further the range of processes in which this family of proteins is involved. To bind DNA in a sequence-specific manner, RHH proteins use a relatively short con-

served 3D structural domain (generally 36 residues) that is formed by intertwining of the RHH motif of two monomers referred to as RHH_2. Hence, most RHH proteins are dimers having a 2-fold symmetry. Reg_{576} is no exception. In many cases the operators to which RHH proteins bind are composed of two or multiple binding sites arranged as inverted or tandem repeats to which higher order RHH oligomers bind. We show that this is also the case for Reg_{576}: its operator is composed of two subsites that are arranged in a directed repeated orientation spaced by 2 bps and that this operator is bound by two Reg_{576} dimers. However, compared to other RHH proteins, Reg_{576} stands out with its extremely high level of cooperativity for binding its operator. For comparison, the well-studied 53 residue RHH Arc protein of bacteriophage P22 binds its two subsite operator (total of 21 bp) in a highly cooperative manner with a calculated Hill constant of 3.5 (55). Nevertheless, binding of individual Arc dimers to a DNA fragment containing only one subsite was observed by EMSA (56). Using EMSA and analytical ultracentrifugation techniques we were unable to detect binding of Reg_{576} to DNA fragments containing only one intact subsite, indicating that it binds the operator with extreme high cooperativity. This is supported by the high calculated Hill coefficient of 7.8. Based on the crystal structure presented here and the DNA-bound structure of the Arc repressor (PDB code: 1PAR), we constructed a model of two Reg_{576} dimers bound to two consecutive DNA repeats. The results suggest that the additional Reg_{576} helix H3 and the preceding H2-H3 loop, both at the C-terminus of the RHH core, are able to form contacts between the two dimers of the tetramer. These contacts are not observed in the Arc structure or structures of other RHH proteins bound to two DNA repeats. These additional interactions of Reg_{576} may explain the high cooperativity of the Reg_{576} tetramer binding to its operator DNA site.

Another interesting finding of this work is that Reg_{576} regulates expression of its own relatively weak promoter P_{27c}. Interestingly, the P_{27c} promoter has a different configuration compared to the strong P_{20c} and P_{23c} promoters: the latter two promoters both contain two Reg_{576} operators, whereas P_{27c} contains only one. Consequently, Reg_{576} represses P_{27c} less strictly than P_{20c} and P_{23c}. Modeling suggests that this organization results in a regulatory system in which the P_{20c} and P_{23c} promoters will be efficiently repressed by the relatively low levels of Reg_{576}.

The $P_{ardC576}$ promoter driving the expression of the putative anti-restriction gene $ardC_{576}$ appeared to be a strong promoter with a strength that is similar to that of promoters P_{20c} and P_{23c}. However, we showed that the strength of $P_{ardc576}$ is similar to that of P_{27c} and that the apparent high promoter activity is due to the presence of inverted repeated sequences in the long untranslated region upstream of the $ardC_{576}$ gene that are predicted to form secondary structures when transcribed into RNA and which probably increase the stability of the transcripts. Promoter $P_{ardc576}$ contains only one Reg_{576} operator and therefore it is expected that it responds less sensitive to Reg_{576} than promoters P_{20c} and P_{23c}. Whether this serves a biological purpose requires future research.

In summary, in this work we have provided evidence that the expression of several possible establishment genes of the conjugative plasmid p576, located in different transcriptional units, are regulated by the dimeric Reg_{576} RHH-type repressor protein by a zygotic induction mechanism, which ensures that these genes are transiently expressed when the plasmid enters the recipient cell. Two Reg_{576} dimers bind their operator that is composed of a dual 5′-TTATCCC-3′ motif with extremely high cooperativity, and the crystal structure of the Reg_{576} dimer provides an explanation for this high cooperative binding. These studies have determined for the first time in detail how establishment genes are regulated in a manner different from that of the ssDNA promoters. It would be very interesting to understand the functions of the putative p576 establishment genes other than $ardC_{576}$.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data bank under accession number 6GYG.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Germán Rivas for valuable discussions related with the ultracentrifugation experiments, Massimo Sammito and Isabel Usón for their help with the structure solution using ARCIMBOLDO, David Rudner for plasmid pDR110, José Belio for help with preparing the figures, and Margarita Salas and Jeff Errington for their support on our work. We also want to acknowledge helpful discussion with other lab members.

FUNDING

Ministry of Economy and Competitiveness of the Spanish Government [BFU2016-75471-C2-1-P (AEI/FEDER, EU) to C.A., BIO2013-41489-P (AEI/FEDER, EU) and BIO2016-77883-C2-1-P (AEI/FEDER, EU) to W.M., BIO2016-77883-C2-2-P (AEI/FEDER, EU) to R.B., BIO-2015-66203-P (AEI/FEDER, EU) to F.R., FIS2016-78313-P (AEI/FEDER, EU) to S.A.]; BIO2013-41489-P (AEI/FEDER, EU) and BIO2016-77883-C2-1-P (AEI/FEDER, EU) also supported J.V., A.M., and C.G.; Wellcome Investigator Award [209500 to Jeff Errington] supported L.W; 'Ramón y Cajal' Contract Supported S.A.; 'Agencia Estatal de Investigación' (AEI); 'Fondo Europeo de Desarrollo Regional (FEDER); European Union (EU). Funding for open access charge Ministry of Economy and Competitiveness of the Spanish Government [BIO2016-77883-C2-2-P].
Conflict of interest statement. None declared.

REFERENCES

1. Ochman,H., Lawrence,J.G. and Groisman,E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*, **405**, 299–304.
2. Frost,L.S., Leplae,R., Summers,A.O. and Toussaint,A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Micobiol.*, **3**, 722–732.

3. Thomas,C.M. and Nielsen,K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.*, **3**, 711–721.
4. Thoma,L. and Muth,G. (2016) Conjugative DNA-transfer in Streptomyces, a mycelial organism. *Plasmid*, **87–88**, 1–9.
5. Goessweiner-Mohr,N., Arends,K., Keller,W. and Grohmann,E. (2013) Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid*, **70**, 289–302.
6. Vasu,K. and Nagaraja,V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.*, **77**, 53–72.
7. Ershova,A.S., Rusinov,I.S., Spirin,S.A., Karyagina,A.S. and Alexeevski,A.V. (2015) Role of restriction-modification systems in prokaryotic evolution and ecology. *Biochemistry (Mosc)*, **80**, 1373–1386.
8. Petrova,V., Chitteni-Pattu,S., Drees,J.C., Inman,R.B. and Cox,M.M. (2009) An SOS inhibitor that binds to free RecA protein: the PsiB protein. *Mol. Cell*, **36**, 121–130.
9. Althorpe,N.J., Chilley,P.M., Thomas,A.T., Brammar,W.J. and Wilkins,B.M. (1999) Transient transcriptional activation of the IncI1 plasmid anti-restriction gene (ardA) and SOS inhibition gene (psiB) early in conjugating recipient bacteria. *Mol. Microbiol.*, **31**, 133–142.
10. Bates,S., Roscoe,R.A., Althorpe,N.J., Brammar,W.J. and Wilkins,B.M. (1999) Expression of leading region genes on IncI1 plasmid ColIb-P9: genetic evidence for single-stranded DNA transcription. *Microbiology*, **145**, 2655–2662.
11. Masai,H. and Arai,K. (1997) Frpo: a novel single-stranded DNA promoter for transcription and for primer RNA synthesis of DNA replication. *Cell*, **89**, 897–907.
12. Frost,L.S., Ippen-Ihler,K. and Skurray,R.A. (1994) Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.*, **58**, 162–210.
13. Nasim,M.T., Eperon,I.C., Wilkins,B.M. and Brammar,W.J. (2004) The activity of a single-stranded promoter of plasmid ColIb-P9 depends on its secondary structure. *Mol. Microbiol.*, **53**, 405–417.
14. Lujan,S.A., Guogas,L.M., Ragonese,H., Matson,S.W. and Redinbo,M.R. (2007) Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 12282–12287.
15. Lovett,P.S. (1973) Plasmid in *Bacillus pumilus* and the enhanced sporulation of plasmid-negative variants. *J. Bacteriol.*, **115**, 291–298.
16. Singh,P.K., Ballestero-Beltran,S., Ramachandran,G. and Meijer,W.J. (2010) Complete nucleotide sequence and determination of the replication region of the sporulation inhibiting plasmid p576 from *Bacillus pumilus* NRS576. *Res. Microbiol.*, **161**, 772–782.
17. Singh,P.K. and Meijer,W.J. (2014) Diverse regulatory circuits for transfer of conjugative elements. *FEMS Microbiol. Lett.*, **358**, 119–128.
18. Anagnostopoulos,C. and Spizizen,J. (1961) Requirements for transformation in *Bacillus subtilis*. *J. Bacteriol.*, **81**, 741–746.
19. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
20. Miller,J.H. (1982) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, NY.
21. Singh,P.K., Ramachandran,G., Duran-Alcalde,L., Alonso,C., Wu,L.J. and Meijer,W.J. (2012) Inhibition of *Bacillus subtilis* natural competence by a native, conjugative plasmid-encoded comK repressor protein. *Environ. Microbiol.*, **14**, 2812–2825.
22. Miguel-Arribas,A., Hao,J.A., Luque-Ortega,J.R., Ramachandran,G., Val-Calvo,J., Gago-Cordoba,C., Gonzalez-Alvarez,D., Abia,D., Alfonso,C., Wu,L.J. *et al.* (2017) The *Bacillus subtilis* conjugative plasmid pLS20 encodes two Ribbon-Helix-Helix type auxiliary relaxosome proteins that are essential for conjugation. *Front. Microbiol.*, **8**, 2138.
23. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1609.
24. Laue,T.M., Shah,B.D., Ridgeway,T.M. and Pelletier,S.L. (1992) Interpretation of analytical sedimentation data for proteins. In: Harding,SE, Rowe,AJ and Horton,JC (eds). *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Royal Society of Chemistry, Cambridge, pp. 90–125.
25. Schuck,P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.*, **320**, 104–124.
26. Balbo,A., Minor,K.H., Velikovsky,C.A., Mariuzza,R.A., Peterson,C.B. and Schuck,P. (2005) Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 81–86.
27. Tataurov,A.V., You,Y. and Owczarzy,R. (2008) Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. *Biophys. Chem.*, **133**, 66–70.
28. Cole,J.L. (2004) Analysis of heterogeneous interactions. *Methods Enzymol.*, **384**, 212–232.
29. Svedberg,T. and Pedersen,K.O. (1940) *The Ultracentrifuge*. Clarendon Press, Oxford.
30. Ingalls,B.P. (2013) *Mathematical Modeling in Systems Biology: an Introduction*. MIT Press, Cambridge.
31. Juanhuix,J., Gil-Ortiz,F., Cuni,G., Colldelram,C., Nicolas,J., Lidon,J., Boter,E., Ruget,C., Ferrer,S. and Benach,J. (2014) Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrotron. Radiat.*, **21**, 679–689.
32. Vonrhein,C., Flensburg,C., Keller,P., Sharff,A., Smart,O., Paciorek,W., Womack,T. and Bricogne,G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 293–302.
33. Rodriguez,D.D., Grosse,C., Himmel,S., Gonzalez,C., de Ilarduya,I.M., Becker,S., Sheldrick,G.M. and Uson,I. (2009) Crystallographic ab initio protein structure solution below atomic resolution. *Nat. Methods*, **6**, 651–653.
34. Murshudov,G.N., Skubak,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
35. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
36. Chen,V.B., Arendall,W.B. III, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
37. Laskowski,R.A. (2009) PDBsum new things. *Nucleic Acids Res.*, **37**, D355–D359.
38. Meijer,W.J.J., de Boer,A., van Tongeren,S., Venema,G. and Bron,S. (1995) Characterization of the replication region of the *Bacillus subtilis* plasmid pLS20: a novel type of replicon. *Nucleic Acids Res.*, **23**, 3214–3223.
39. Derman,A.I., Becker,E.C., Truong,B.D., Fujioka,A., Tucey,T.M., Erb,M.L., Patterson,P.C. and Pogliano,J. (2009) Phylogenetic analysis identifies many uncharacterized actin-like proteins (Alps) in bacteria: regulated polymerization, dynamic instability and treadmilling in Alp7A. *Mol. Microbiol.*, **73**, 534–552.
40. Singh,P.K., Ramachandran,G., Ramos-Ruiz,R., Peiro-Pastor,R., Abia,D., Wu,L.J. and Meijer,W.J. (2013) Mobility of the native *Bacillus subtilis* conjugative plasmid pLS20 is regulated by intercellular signaling. *PLoS Genet.*, **9**, e1003892.
41. Belogurov,A.A., Delver,E.P., Agafonova,O.V., Belogurova,N.G., Lee,L.Y. and Kado,C.I. (2000) Antirestriction protein Ard (Type C) encoded by IncW plasmid pSa has a high similarity to the "protein transport" domain of TraC1 primase of promiscuous plasmid RP4. *J. Mol. Biol.*, **296**, 969–977.
42. Sharp,J.S. and Bechhofer,D.H. (2005) Effect of 5′-proximal elements on decay of a model mRNA in *Bacillus subtilis*. *Mol. Microbiol.*, **57**, 484–495.
43. Durand,S., Tomasini,A., Braun,F., Condon,C. and Romby,P. (2015) sRNA and mRNA turnover in Gram-positive bacteria. *FEMS Microbiol. Rev.*, **39**, 316–330.
44. Lengyel,I.M. and Morelli,L.G. (2017) Multiple binding sites for transcriptional repressors can produce regular bursting and enhance noise suppression. *Phys. Rev. E*, **95**, 042412.
45. Alon,U. (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits*. CRC Press, Boca Raton.

46. Schreiter,E.R. and Drennan,C.L. (2007) Ribbon-helix-helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.*, **5**, 710–720.
47. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
48. Golub,E., Bailone,A. and Devoret,R. (1988) A gene encoding an SOS inhibitor is present in different conjugative plasmids. *J. Bacteriol.*, **170**, 4392–4394.
49. Chilley,P.M. and Wilkins,B.M. (1995) Distribution of the *ardA* family of antirestriction genes on conjugative plasmids. *Microbiology*, **141**, 2157–2164.
50. Bagdasarian,M., Bailone,A., Angulo,J.F., Scholz,P., Bagdasarian,M. and Devoret,R. (1992) PsiB, and anti-SOS protein, is transiently expressed by the F sex factor during its transmission to an *Escherichia coli* K-12 recipient. *Mol. Microbiol.*, **6**, 885–893.
51. Jones,A.L., Barth,P.T. and Wilkins,B.M. (1992) Zygotic induction of plasmid *ssb* and *psiB* genes following conjugative transfer of Incl1 plasmid Collb-P9. *Mol. Microbiol.*, **6**, 605–613.
52. Alfoldi,L., Jacob,F. and Wollman,E.L. (1957) Lethal zygosis in crossing between colicinogenic and non-colicinogenic strains of *Escherichia coli*. *C R Hebd. Seances Acad. Sci.*, **244**, 2974–2977.
53. Wollman,E.L. and Jacob,F. (1957) Processes of conjugation and recombination in *Escherichia coli*. II. Chromosomal location of phage lambda and genetic results of zygotic induction. *Ann. Inst. Pasteur. (Paris)*, **93**, 323–339.
54. Delver,E.P. and Belogurov,A.A. (1997) Organization of the leading region of IncN plasmid pKM101 (R46): a regulation controlled by CUP sequence elements. *J. Mol. Biol.*, **271**, 13–30.
55. Brown,B.M., Bowie,J.U. and Sauer,R.T. (1990) Arc repressor is tetrameric when bound to operator DNA. *Biochemistry*, **29**, 11189–11195.
56. Brown,B.M. and Sauer,R.T. (1993) Assembly of the Arc repressor-operator complex: cooperative interactions between DNA-bound dimers. *Biochemistry*, **32**, 1354–1363.