

A novel bipartite antitermination system widespread in conjugative elements of Gram-positive bacteria

Andrés Miguel-Arribas¹, Jorge Val-Calvo¹, César Gago-Córdoba¹, José M. Izquierdo¹, David Abia², Ling Juan Wu³, Jeff Errington³ and Wilfried J.J. Meijer^{1,*}

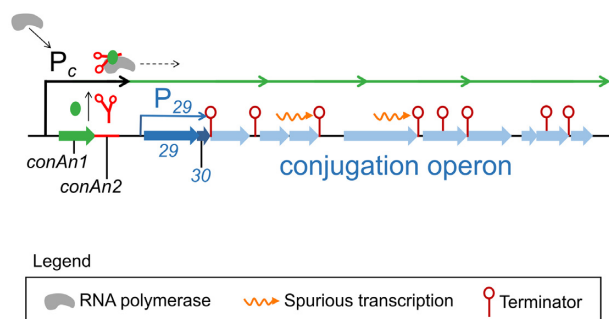
¹Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), C. Nicolás Cabrera 1, Universidad Autónoma de Madrid, Canto Blanco, 28049 Madrid, Spain, ²Bioinformatics Facility, Centro de Biología Molecular “Severo Ochoa”, (CSIC-UAM), C. Nicolás Cabrera 1, Universidad Autónoma de Madrid, Canto Blanco, 28049 Madrid, Spain and ³Centre for Bacterial Cell Biology, Biosciences Institute, Newcastle University, Richardson Road, Newcastle Upon Tyne, NE2 4AX, UK

Received February 22, 2021; Revised April 09, 2021; Editorial Decision April 13, 2021; Accepted April 23, 2021

ABSTRACT

Transcriptional regulation allows adaptive and coordinated gene expression, and is essential for life. Processive antitermination systems alter the transcription elongation complex to allow the RNA polymerase to read through multiple terminators in an operon. Here, we describe the discovery of a novel bipartite antitermination system that is widespread among conjugative elements from Gram-positive bacteria, which we named *conAn*. This system is composed of a large RNA element that exerts antitermination, and a protein that functions as a processivity factor. Besides allowing coordinated expression of very long operons, we show that these systems allow differential expression of genes within an operon, and probably contribute to strict regulation of the conjugation genes by minimizing the effects of spurious transcription. Mechanistic features of the *conAn* system are likely to decisively influence its host range, with important implications for the spread of antibiotic resistance and virulence genes.

GRAPHICAL ABSTRACT



INTRODUCTION

Transcriptional regulation contributes importantly to adaptive and coordinated gene expression, which are crucial for the survival of all biological entities in order to respond adequately to changing environmental conditions. The transcription process can be divided into three stages, -initiation, elongation and termination (1,2). Bacteria have two types of termination signals: Rho-dependent and intrinsic (aka Rho-independent) terminators (2–7). Rho-dependent terminators require the RNA-dependent helicase Rho, which, after binding to specific sites, translocates along the mRNA resulting eventually in dissociation of the transcription elongation complex (TEC) (3,8,9). Intrinsic terminators, though, do not require any additional factor to terminate transcription; they are typically characterized by a GC-rich inverted repeat followed by a stretch of Ts in the non-template DNA strand. When transcribed into RNA, this region forms a stem-loop structure followed by a tract of Us, which is sufficient to terminate transcription, although termination may be enhanced by the NusA transcription factor (10,11). Thus, a transcription terminator marks the end of a transcriptional unit that can consist of one or several genes (6). However, many terminators have evolved to terminate transcription only under certain circumstances, allowing conditional expression of the gene(s) located downstream. Such terminators, which include riboswitches, act via the so-called attenuation mechanism in which the RNA region can fold in a terminator or antiterminator structure depending on the presence or absence of a specific signal (12, and references therein). These attenuation systems that control a single terminator (13–15) are fundamentally different from the processive antitermination (P-AT) mechanisms, which do not depend on differences in the configuration of the terminator. Instead, P-AT mechanisms require the association of an antitermination (AT) factor with the TEC. Once associated, the altered TEC can read through multiple intrinsic terminators and

*To whom correspondence should be addressed. Tel: +34 91 196 4539; Email: wmeijer@cbm.csic.es

act over great distances. So far, only a limited number of P-AT systems have been described; several of them achieve P-AT when an antiterminator protein engages with a transcribing RNAP. While the primary sequences of some of these, like the phage λ AT proteins N and Q, do not share similarity with the transcription elongation factors, others, like RfaH, ActX, TaA, UpxY and Loap, share similarity with the elongation factor NusG. In a few other cases, the AT factor is an RNA element that achieves P-AT in largely unknown ways (16–19). The characteristics and mode of action of the relatively few known P-AT systems have recently been reviewed by Goodson and Winkler (20).

Here, we describe the discovery of a novel P-AT system which we call *conAn* (conjugation antitermination), located at the start of the conjugation operon present on plasmid pLS20 of the Gram-positive (Gram+) bacterium *Bacillus subtilis*. In contrast to all known P-AT systems, this is a bipartite system composed of a large RNA element responsible for AT and a protein component that is required for P-AT. The *conAn* system permits transcription to read through multiple intrinsic terminators present in the very long conjugation operon, and probably contributes to proper expression of the conjugation genes by minimizing the effects of spurious transcription. The system serves at least one other purpose: differential expression of one or more genes inside the conjugation operon during conjugation and non-conjugation conditions. In addition, we show that bipartite *conAn* type P-AT systems are most likely present on hundreds of conjugative plasmids of Gram+ bacteria. Finally, we found that *conAn* systems differ in their functionality in heterologous hosts, which likely depends on the compatibility of the *conAn* system with the TEC of the host. This could have important consequences for the ability of conjugative elements to spread antibiotic resistance.

MATERIALS AND METHODS

Bacterial strains, plasmids, media and oligonucleotides

Escherichia coli, *Bacillus subtilis* and *Enterococcus faecalis* strains were grown in Luria-Bertani (LB) medium (21) and *Listeria innocua* strains were grown in Brain Heart Infusion (BHI) medium (22). *Lactobacillus casei* strains were grown in de Man, Rogosa and Sharpe (MRS) medium (23). All bacteria were grown in liquid media with shaking or on 1.5% LB agar plates at 37°C. When appropriate, media were supplemented with the following antibiotics: ampicillin (Amp), 100 μ g/ml; erythromycin (Em), 1.5 μ g/ml (*B. subtilis*, *L. innocua*, *L. casei* and *E. faecalis*) or 150 μ g/ml (*E. coli*), chloramphenicol (Cm), 5 μ g/ml; spectinomycin (Spec), 100 μ g/ml; and kanamycin (Kan), 10 μ g/ml (*B. subtilis*) or 40 μ g/ml (*E. coli*). *B. subtilis* strains used were isogenic with *B. subtilis* strain 168. Bacterial strains are listed in Supplementary Table S7. Plasmids and oligonucleotides used are listed in Supplementary Table S8 and S9, respectively. All oligonucleotides were purchased from Isogen Life Science, The Netherlands. Synthetic DNA constructs were obtained from Shinegene (Shanghai, China). These constructs were provided in the vector pUC57Km. The inserts corresponding to the synthetic DNA constructs were amplified by PCR using oligonucleotides M13Fw and

M13Rev. Next, the fragments were digested with appropriate enzymes and cloned in the final vector. Alternatively, inserts were provided directly in the final vector.

Transformation

Escherichia coli competent cells were prepared and transformed using standard methods (21). Generation of competent *B. subtilis* cells and transformations were prepared as described (24). *E. faecalis*, *L. casei* and *L. innocua* cells were transformed as described (25–27). Transformants were selected on LB, BHI and MRS plates with appropriate antibiotics.

Construction of plasmids and strains

DNA techniques were performed using standard molecular methods (21). Plasmids were isolated from *E. coli* strains using ‘Wizard Plus SV Minipreps DNA Purification Systems’ (Promega). PCR fragments were purified using ‘Wizard SV Gel and PCR Clean-Up System’ (Promega). All enzymes used were purchased from New England Biolabs, USA. Cloned PCR fragments were checked for their correctness by sequence analysis. Total DNA extracted from pLS20cat harbouring strain PKS11 was used as template to amplify pLS20cat regions by PCR. In the same way, total DNA extracted from the plasmid p576 and pAW63 harbouring strains *Bacillus pumilus* NRS576 and *B. thuringiensis* subsp. *kurstaki* HD73, respectively, were used as template to amplify *conAn* regions of these plasmids. Specific information of the construction of each plasmid, including the oligonucleotides and restriction enzymes used to amplify and clone the fragments, is presented in Supplementary Table S8. General information of the use and characteristics of the different parental vectors is given below.

Marker less in-frame deletions of gene 28 and its downstream regions on pLS20cat were constructed making use of the pMiniMAD2 vector (28). The method is based on single cross-over integration of a temperature-sensitive replicon via homologous recombination at restrictive temperature followed by permitting replication of the integrated replicon by growth at the permissive temperature to provoke deletion of replicon thereby generating the desired deletion (29). The deletion strategy is described in (30). In short, in a first step, around 600 bp of the interested ‘UP’ and ‘Down’ regions were amplified by PCR, and then joined in a second step by overlapping PCR. The resulting fused PCR product was purified and digested with NdeI and BamHI and cloned into the pMiniMAD2 vector digested with the same enzymes. Using this strategy, the following derivatives of pMiniMAD2 were constructed: pAND86, pAND98, pAND99 and pAND100. These plasmids were used for the construction of derivatives of pLS20cat. pAND86 was used to construct strain AND81 which contains pLS20cat derivative lacking most of gene 28 (pLS20 Δ 28). Similarly, pAND98, pAND99 and pAND100 were used to generate strains AND98 (pLS20 Δ ds-I), AND99 (pLS20 Δ P₂₉) and AND100 (pLS20 Δ ds-II) (see Supplementary Table S7). More details regarding the construction of plasmids is described in Supplementary Tables S8 and S9.

Plasmid pAX01 is a *B. subtilis* *lacA* integration vector designed for the construction of DNA regions fused to the

xylose-inducible promoter P_{xyl} , and the cassette is associated with an erythromycin resistance gene for selection in *B. subtilis*. This vector was used for controlled expression of gene(s) of interest in *B. subtilis* and it can be used in combination with another *B. subtilis* integration vector such as pDR110, which is designed for placing constructs at the *amyE* locus having genes fused to the IPTG-inducible promoter P_{spank} . We used pAX01 to construct plasmids pPKS17 and pAND178 and the strains AND84, AND178P and CG14 (see Supplementary Table S1, S2 respectively). Double-crossover integration into the chromosome was checked by PCR using the primer sets [pAX1-Up/pAX1-Dn] and [pAX2-Up/pAX2-Dn] (Supplementary Table S3). When used to study the effect of ectopic expression of a given gene placed under the control of the inducible P_{xyl} promoter, an overnight grown culture was diluted in pre-warmed LB supplemented with 1% of xylose.

The vector pKSsfGFP was previously constructed in our lab based on the vector pJS104 (30,31). The principal characteristics of this plasmid are: (i) high copy number and Amp^R in *E. coli*, (ii) *amyE* gene integrative cassette by double cross-over and Cm^R in *B. subtilis* and (iii) promoter-less superfolder *gfp* (*gfp*) gene that is preceded by the unique restriction sites BamHI, HindIII, EcoRI, NheI and SpeI. The *gfp* reporter gene used in these studies corresponds to a version that contains several modifications to enhance the stability and other features of the encoded green fluorescent protein (32). For simplicity, we refer to this recombinant gene as *gfp*. DNA fragments were cloned upstream the *gfp* gene on pKSsfGFP and the resulting derivatives were used to construct *B. subtilis* strains that contain at the *amyE* locus a single copy of the *gfp* gene transcriptionally fused to the cloned region. pKSsfGFP derivatives were isolated from *E. coli* strains and used to transform competent *B. subtilis* 168 cells. Transformants were selected on LB plates containing Cm. Double-crossover integration into the chromosome was checked by the loss of amylase activity.

Plasmid pDR110 is a *B. subtilis amyE* integration vector that contains the IPTG-inducible P_{spank} promoter and is associated with a spectinomycin resistance gene for selection in *B. subtilis*. This vector was used to construct pAND101 that is designed for studying terminator and/or antiterminator activities *in vivo*. Plasmid pAND101 was constructed as follows. The superfolder *gfp* gene (*gfp*) was amplified by PCR using oligos oA98 and oA88 and pKSsfGFP vector as template. This PCR fragment was then digested with NheI and SpeI and subsequently cloned into vector pDR110 digested with the same enzymes. Vector pAND101 can be used for two objectives: (i) test the terminator activity and (ii) test the antitermination activity. For these purposes, first a PCR fragment containing a putative terminator is cloned at the multiple cloning site (MCS) located in between the P_{spank} promoter and the *gfp* gene on plasmid pAND101. The resulting plasmid can be used to construct a *B. subtilis* strain containing at its *amyE* locus the construct P_{spank} -[insert]-*gfp*, which can be used to study terminator or promoter activity of the cloned region. Second, a DNA fragment containing an antiterminator system *conAn_x* can subsequently be placed upstream of the first cloned fragment containing the putative terminator. The resulting derivative of pAND101 can then be used to construct a *B. sub-*

tilis strain containing at its *amyE* locus the construct P_{spank} -[insert-II (AT system)]-[insert-I (Ter)]-*gfp*, which can be used to study the functionality of the cloned *conAn* region, or to study whether the terminator is resistant or not to the antitermination system.

Plasmid pAT18 shuttle vector replicates as a high-copy number plasmid in *E. coli* and as a low-copy number plasmid in Gram+ bacteria. The plasmid contains the replication functions of plasmid pAM β 1, allowing it to replicate in a broad host-range of Gram+ bacteria (33). It is associated with an erythromycin resistance gene for selection in Gram+ bacteria. This plasmid was used to study the *conAn* antitermination systems of different Gram+ plasmids in their own genus. Derivatives of pAT18 were constructed as follows. First, the cassette containing P_{spank} -[possible insert]-*gfp-lacI* was isolated from the pAND101, or derivative of interest, by digestion with *Pst*I and *Bam*HI. This fragment was ligated to pAT18 digested with the same enzymes. The ligation mixture was then used to transform competent *E. coli* DH5 α cells. Positive clones were detected by colony PCR using primers oM13RevExt and oSeqpKS-GFP-Dn and checked by digestion with *Pst*I and *Bam*HI.

Conjugation assays

Conjugation was carried out in liquid medium as described previously (34). The effect of ectopic expression of a given gene placed under the control of the inducible P_{spank} and/or P_{xyl} promoter on conjugation was studied by adding the inducer (1 mM IPTG or 1% xylose) to prewarmed LB medium used to dilute overnight cultures of the donor cells.

RNA isolation and RNA sequencing

Total RNA was isolated from exponentially growing cells ($OD_{600} = 0.8-1$) by using RNA protect solution and RNeasy Mini Kit from Qiagen, following the manufacturer's protocol. RNA sequencing and bioinformatical analysis of RNAseq data was done as described previously (35).

Flow cytometry

Fluorescence quantification using Flow cytometry was performed as described before (30). All different bacteria were grown in liquid media with shaking until the culture reached an $OD_{600} = 0.8-1$, with the exception of *E. faecalis* strains that were recollected at $OD_{600} = 0.5$. Fluorescence levels were expressed as the mean value of the Geomean values of 100 000 cells obtained in three independent experiments.

Identification of putative termination sequences

Conjugative plasmid sequences were screened for the presence of putative Rho-independent transcriptional terminators by (i) the 'ARNold' Web server (rna.igmors.u-psud.fr/toolbox) that uses two algorithms: Erpin and RNAmotif (36,37) and by (ii) the TransTermHP Web server (transterm.cbcbg.umd.edu) which uses an algorithm that is distinct from the Erpin and RNAmotif algorithms (38).

Identification of genes encoding proteins showing significant similarity to ConAn1_{pLS20} homologs

Each of the ConAn1 sequences was used as query in multiple rounds of Psi-blastp searches against the NCBI nr database (version 2.5.0+, November 2019) (39,40). This search resulted in the detection of 662 sequences sharing significant similarity with the nine ConAn1 sequences, with *e*-value less than $1e-7$. The program 'USEARCH' (version 8.0.1517_i86linux32) was then used to identify and remove redundant sequences showing 100% identity (41), resulting in 531 unique hits. MMSEQ2 program was used to create clusters of proteins with more than 95% of sequence identity and coverage over 80%, resulting in 141 clusters. One representative sequence was selected from each cluster. The evolutionary history was inferred using the Neighbor-Joining method (42). The optimal tree with the sum of branch length = 48.74942208 is shown. The evolutionary distances were computed using the JTT matrix-based method (43) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1.97). All ambiguous positions were removed for each sequence pair (pairwise deletion option). The final dataset contained a total of 343 positions. Evolutionary analyses were conducted in MEGA X (44). The tree image was generated using Dendroscope program (45). Colouring of different branches were added manually.

RESULTS

Deletion of pLS20cat gene 28 greatly reduces conjugation efficiency

The conjugation genes of pLS20cat are located in a single large operon spanning genes 28 to 74 (35, for review see, 46). Intriguingly, the putative 172-residue protein encoded by the first gene of the conjugation operon, gene 28, does not show significant similarity to any protein of known function. To gain insight into its function we constructed a derivative of pLS20cat, pLS20Δ28, containing a markerless in-frame deletion of gene 28 and used strain AND81 harbouring pLS20Δ28 to determine its conjugation efficiency. In agreement with previous results, the conjugation efficiency obtained for wild type pLS20 (strain PKS11) was in the range of $2.0E-3$ transconjugants per donor cells (34,35). However, deletion of gene 28 had a dramatic effect on conjugation, causing the efficiency to drop more than 10 000-fold (Figure 1B). This demonstrates that gene 28 plays a key role in the conjugation process. Gene 28 complementation assays are described below.

Protein p28 acts at the level of transcription but not as a transcriptional activator

The lack of similarity between protein p28 with proteins of known function complicated designing an experimental approach to unravel its function in the conjugation process. We reasoned, however, that protein p28 should either play a role at the level of transcription or at a later stage, and that RNAseq could be used to discriminate between these two possibilities. We therefore compared the expression lev-

els of the conjugation genes by RNAseq analysis of the wild type plasmid pLS20cat and its derivative pLS20Δ28. The results presented in Supplementary Figure S1 show that the absence of gene 28 severely affected the expression of most of the downstream conjugation genes but, strikingly, not that of genes 29 or 30, demonstrating that the absence of gene 28 does not affect the activity of the main conjugation promoter P_c located upstream of gene 28 (47). The strong decrease in expression downstream of gene 30 for pLS20Δ28 might be explained by the presence of a protein p28-activated promoter located upstream of gene 31. To test this, we constructed strain CG14 with two chromosomal cassettes: one containing the upstream region of gene 31 fused to a *gfp* reporter gene, and another containing a copy of gene 28 controlled by a xylose-inducible promoter. This strain did not show significant fluorescence regardless of whether grown in the presence or absence of xylose (supplementary Table S1), indicating that the upstream region of gene 31 does not contain a protein p28-inducible promoter.

pLS20cat gene 30 is followed by a functional and efficient transcriptional terminator, Ter₃₀

The results described above raised the possibility that gene 28 encodes a transcriptional antiterminator that acts on a transcriptional terminator downstream of gene 30. In line with this possibility, *in silico* analysis suggested the presence of a putative intrinsic transcriptional terminator near the end of gene 30, which we tentatively named Ter₃₀ (Figure 2A). To test whether Ter₃₀ was a genuine intrinsic terminator we constructed a series of strains in which DNA fragments with potential transcriptional terminators were inserted between an IPTG-inducible P_{spank} promoter and a *gfp* reporter gene (Figure 2B). A functional transcriptional terminator would result in decreased fluorescence levels. Only low fluorescence levels were obtained when the cultures were grown in the absence of IPTG, showing that the P_{spank} promoter was well repressed under these conditions (Figure 2C). As expected, relatively high fluorescence levels were obtained for the strain with no terminator (P_{spank} -*gfp*, strain AND101) when grown in the presence of 1 mM IPTG. Similar high fluorescence levels were obtained for control strain AND202 with a >800 bp fragment with no known terminator (Figure 2C), demonstrating that increasing the distance between the promoter and the *gfp* gene by more than 800 bp did not significantly affect the expression of *gfp*. Importantly, only background levels of fluorescence were observed for strain AND127 with the putative Ter₃₀ terminator, establishing that Ter₃₀ comprises a functional transcriptional terminator.

The main conjugation promoter P_c , located upstream of gene 28, is considerably stronger than the fully induced P_{spank} promoter (30). To test if Ter₃₀ could terminate transcription from the P_c promoter, we constructed strains in which *gfp* (strain AND148) or the [Ter₃₀-*gfp*] construct (strain AND357) was downstream from the P_c promoter (Figure 2B). The much higher fluorescence levels obtained for the strain AND148 in which *gfp* was fused to P_c (Figure 2C), confirmed that P_c is stronger than the P_{spank} promoter. Importantly, as observed for P_{spank} , the presence of

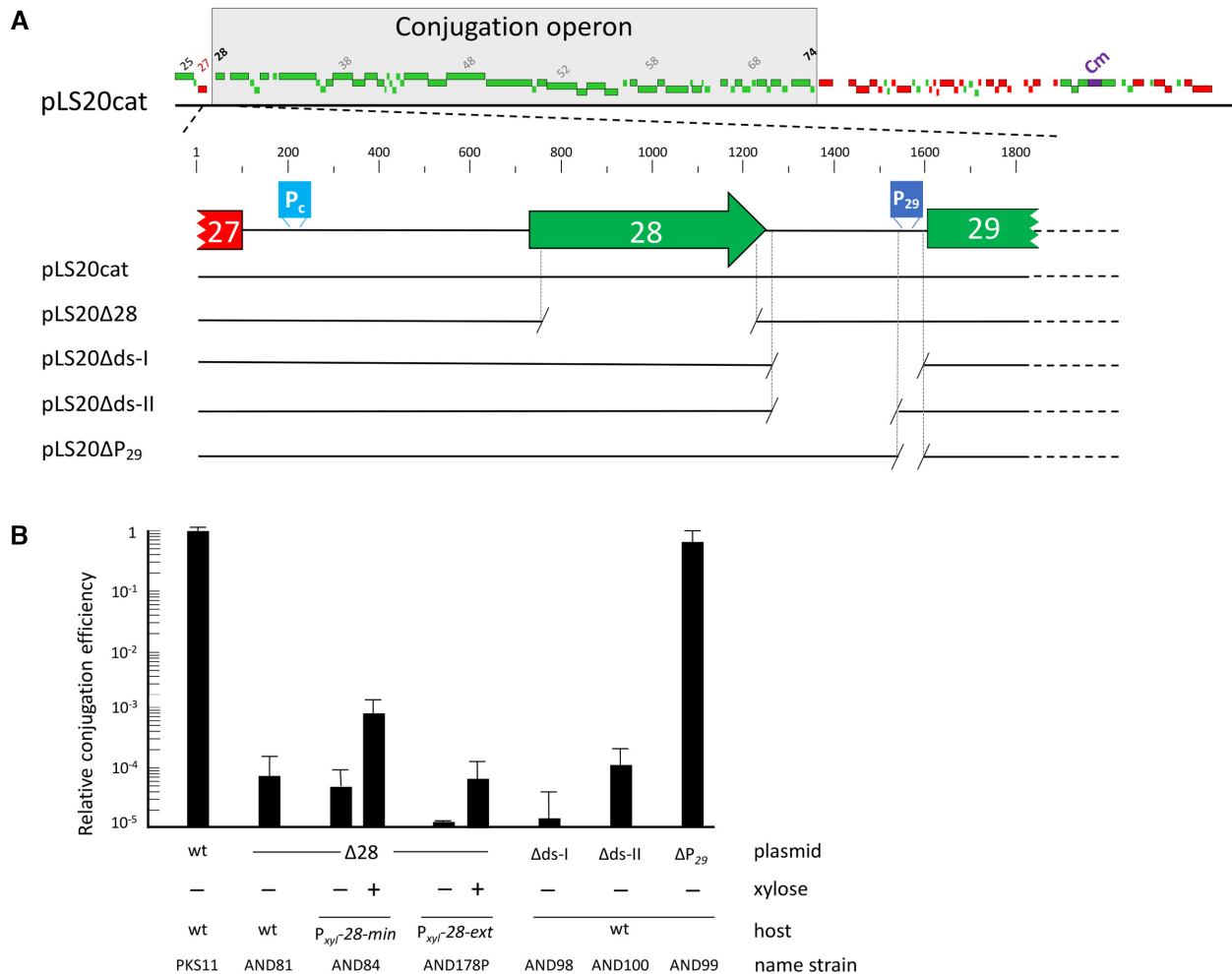


Figure 1. pLS20cat gene 28 is required for efficient conjugation. (A) Genetic map of pLS20cat. Upper panel shows a linear map of pLS20cat. Rightward and leftward oriented genes, and the chloramphenicol resistance gene (Cm) are indicated with green, red and blue rectangles, respectively. The conjugation operon spanning genes 28–74 is indicated. A blow up of gene 28 and its downstream region, as well as the regions deleted in the various derivatives of pLS20cat studied here are shown in the lower panel. (B) Relative conjugation efficiencies of donor strains harbouring pLS20cat (wt) or the derivative pLS20Δ28 (Δ28), pLS20Δds-I (Δds-I), pLS20Δds-II (Δds-II) or pLS20ΔP₂₉ (ΔP₂₉). Strains AND84 and AND178P contain an ectopic copy of gene 28 without (AND84, P_{xyI} -28-min) or with the 315 bp region downstream of gene 28 (AND178P, P_{xyI} -28-ext) placed under the control of the xylose-inducible P_{xyI} promoter. Conjugation efficiencies were calculated as the number of transconjugants per donor, and their efficiencies are expressed relative to the wild type plasmid pLS20cat. At least three independent experiments were performed for each donor. Error bars represent standard deviation.

Ter₃₀ downstream of the P_c promoter in strain AND357 resulted in a ~50-fold decrease in fluorescence level (Figure 2C). These results agree with previous observations that terminators can terminate transcription with different intrinsic efficiencies but that the efficiency is not or only moderately influenced by the strength of the upstream promoter (48–52). Importantly, these results showed that Ter₃₀ is an efficient transcriptional terminator.

pLS20 contains a *cis*-acting bipartite antitermination system composed of gene 28 and its downstream region

pLS20cat gene 29 is preceded by a constitutive weak promoter, P₂₉, that drives expression of genes 29 and 30, independent of the P_c promoter (30). Besides promoter P₂₉, the rather large (383 bp) intergenic gene 28–29 region contains some conspicuous features (see below). To test for possible AT activity the experiments described below were based on

constructs containing gene 28 with or without 315 bp of the downstream DNA, which we refer to as *extended* (ext) or *minimal* (min) constructs, respectively.

First we tested whether ectopic expression of gene 28-*min* or 28-*ext* could restore efficient conjugation of plasmid pLS20Δ28. Strains harbouring pLS20Δ28, and containing an ectopic copy of 28-*min* or 28-*ext* under the control of the xylose inducible P_{xyI} promoter at the chromosomal *lacA* locus, were used as donor strains in conjugation experiments. Ectopic expression of 28-*min* (strain AND84) or 28-*ext* (strain AND178P) did not restore efficient conjugation (Figure 1B), nor did it elicit AT at the Ter₃₀ terminator located elsewhere on the chromosome (*amyE*), as measured by fluorescence from a P_{spank} -[Ter₃₀]-*gfp* construct (Supplementary Table S2).

To analyse whether gene 28-*min*/ext could antiterminate Ter₃₀ in a *cis* configuration we engineered strains in which copies of gene 28-*min* or 28-*ext* were located up-

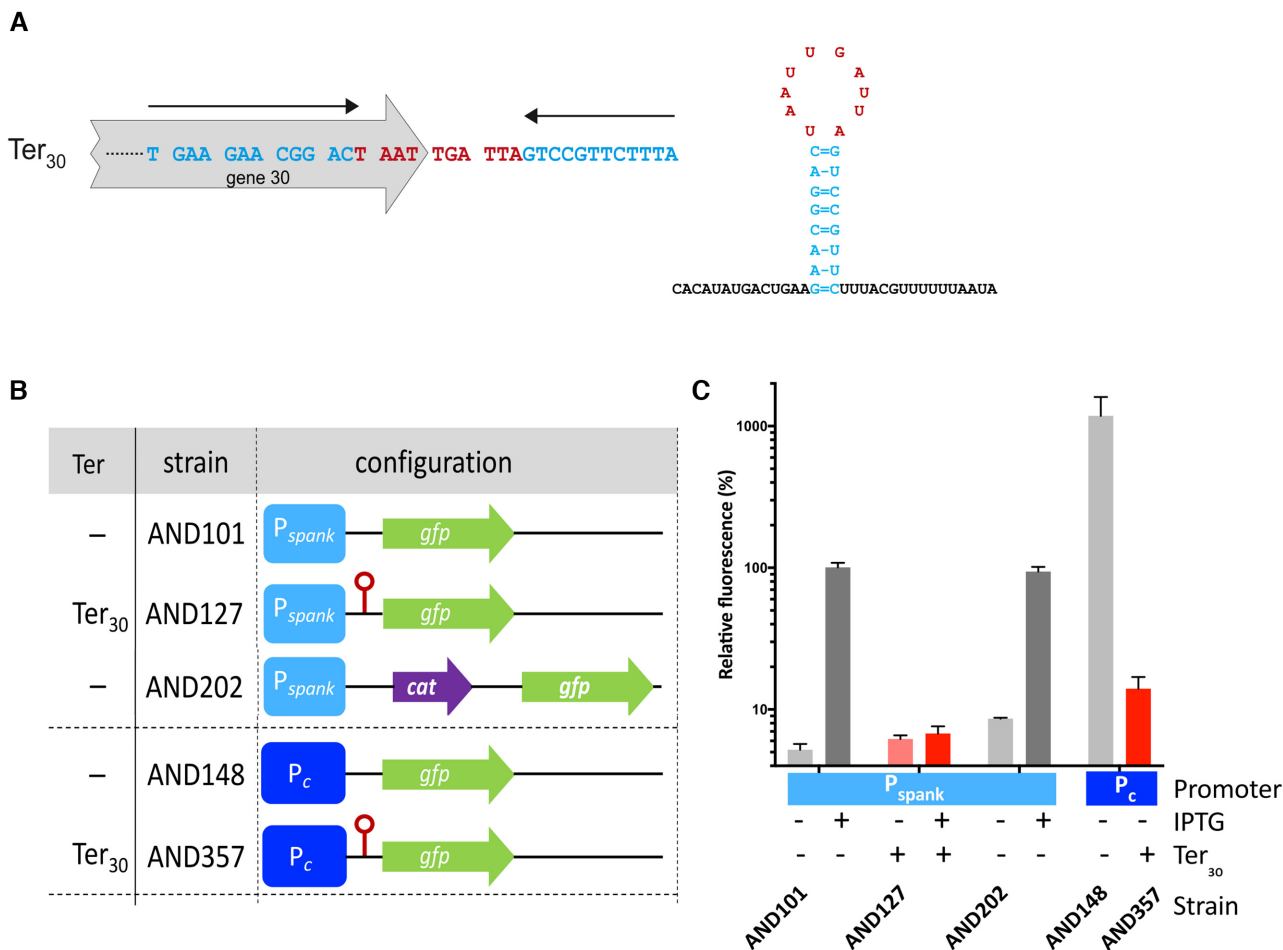


Figure 2. pLS20cat gene 30 is followed by a functional transcriptional terminator. (A) Features of terminator Ter₃₀ located near the end of pLS20cat gene 30. pLS20cat was screened for putative intrinsic terminators by (i) the ARNold Web server (rna.igmors.u-psud.fr/toolbox) that uses two algorithms: Erpin and RNAmotif (36,37), and by (ii) the TransTermHP Web server (transterm.cbcb.umd.edu) that uses a distinct algorithm (38). The pLS20cat-region containing an inverted repeated sequence that may form a stem loop structure when transcribed into RNA is shown; nucleotides predicted to form part of the stem are shown in blue and indicated with black arrows; and those predicted to form the loop are shown in red. Ter₃₀ overlaps with the C-terminal region of gene 30 (grey arrow). The right panel presents the predicted stem–loop structure including the twelve downstream nucleotides. The same colour code is used as in the left panel. (B) Schematic view of the relevant genetic features of the different strains used. The P_{spank} and P_c promoters are indicated with light and dark blue boxes, respectively. The *gfp* reporter (*gfp*) and the chloramphenicol resistance gene (*cat*) are indicated with green and purple arrows, respectively. The transcriptional terminator Ter₃₀ is indicated with a red lollipop symbol. (C) Fluorescence level determined by FACS analyses of cells grown in the absence or presence of 1 mM IPTG relative to the fluorescence levels obtained for strain AND101 grown in the presence of IPTG. Samples were withdrawn from late exponentially grown cultures (OD₆₀₀ between 0.8 and 1). Color code bars: grey and red, strains having a construct lacking or containing Ter₃₀, respectively; light and dark, strains grown in the absence or presence of IPTG, respectively. Fluorescence levels are expressed as the mean value of the Geomean values of 100 000 cells obtained in three independent experiments. Error bars represent standard deviation.

stream of Ter₃₀ in the same transcriptional unit (i.e. P_{spank}-[28-min/ext]-[Ter₃₀]-*gfp*), together with control strains lacking Ter₃₀ (see Figure 3A). As expected, when grown in the presence of IPTG, high fluorescence levels were observed for the control strains AND101, AND115 and AND118 lacking Ter₃₀ (Figure 3B). Importantly, similarly high fluorescence levels were also obtained for the strain in which gene 28-ext (strain AND210), but not 28-min (strain AND181), preceded Ter₃₀ (Figure 3B). These results indicate that: (i) gene 28, together with (part of) the downstream region, function as a transcriptional antiterminator that efficiently eliminates transcriptional termination at Ter₃₀ and (ii) the AT system is only functional when located in *cis*, probably to limit its activity to the conjugation operon. To support the conclusion that the gene 28 downstream region

is essential for AT, we constructed plasmid pLS20Δds-I, a derivative of pLS20cat in which the 356 bp intergenic region after the stop codon of gene 28 up to the RBS of gene 29 (downstream region I) was deleted (Figure 1A). The absence of this region had similar negative effects on the efficiency of conjugation as did deletion of gene 28 (Figure 1B, strain AND98). Based on these results we refer to gene 28 as *conAn1* (conjugation Antitermination element 1) and the downstream region as *conAn2* (conjugation Antitermination element 2).

The *conAn2* region does not include promoter P₂₉

The antiterminator protein Q of bacteriophage λ binds between the -35 and -10 boxes of the late operon pro-

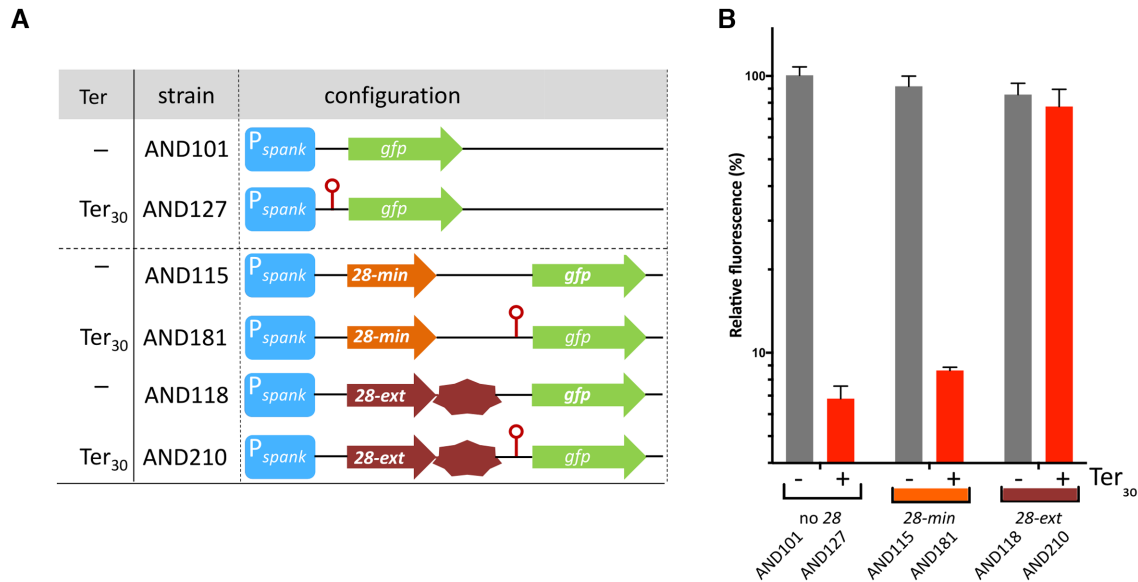


Figure 3. pLS20cat gene 28 and its downstream region inhibit the termination activity of Ter₃₀ when they precede the terminator on the same transcriptional unit. (A) Relevant genetic features of the different strains used. Light and dark brown arrows indicate pLS20cat gene 28 without (28-min) and with its 315 bp downstream region (28-ext), respectively. Other symbols and color codes are the same as in Figure 2B. (B) FACS-determined fluorescence levels of cells of different strains normalized against the fluorescence level obtained for AND101 cells. Grey and red bars, strains having a construct lacking or containing Ter₃₀, respectively. Relevant genotypes of the strains are given in panel A. Conditions and symbols are identical as described in legend of Figure 2C. Error bars represent standard deviation.

motor P_R , which allows Q to associate with the RNA polymerase (RNAP) after it binds to P_R , generating an RNAP-Q complex that can antiterminate the downstream t_R terminator (53). To test whether ConAnI used promoter P_{29} to associate with the RNA polymerase, we constructed two derivatives of pLS20cat lacking either the 80 bp region containing promoter P_{29} (30), or containing P_{29} but lacking the remaining intergenic gene 28–29 region (Figure 1A). The results presented in Figure 1B show that the absence of the P_{29} promoter (pLS20 ΔP_{29} , strain AND99) had almost no effect on the conjugation efficiency compared with pLS20cat, whereas deletion of the intergenic region (pLS20 Δds -II, strain AND100) resulted in much lower conjugation efficiencies, similar to those for pLS20 $\Delta 28$ and pLS20 Δds -I (Figure 1B). These results demonstrate that all or part of the 283 bp DNA region located between *conAnI* and the P_{29} promoter is required for conjugation, but P_{29} is not.

conAn can antiterminate other transcriptional terminators located inside or outside the conjugation operon

In silico analyses predicted that the conjugation operon of pLS20 contains 22 additional potential intrinsic terminators besides Ter₃₀, that might terminate transcripts originating from the conjugation promoter P_c , and thus be regulated by the *conAn* system (Supplementary Figure S2A and Table S3). Five of these, all predicted to constitute an intrinsic terminator by at least two of the three algorithms used, were chosen at random for further analysis. These putative transcriptional terminators, located near the end or downstream of genes 33, 35, 47, 53 and 62, are referred to as Ter₃₃, Ter₃₅, Ter₄₇, Ter₅₃ and Ter₆₂, respec-

tively. As a control, we also analysed a transcriptional terminator taken from downstream of the chloramphenicol resistance gene from an unrelated genetic element, plasmid pC194 (54), which we named Ter_{cat}. The predicted features of these putative terminators are given in Supplementary Figure S3. The same strategy as used above for Ter₃₀ was employed to determine whether any of these regions contain a functional terminator, and if so whether the *conAn* system was able to overcome transcriptional termination at that site. Supplementary Figure S2C shows that the presence of each of the six predicted terminators preceding the *gfp* gene resulted in a clear decrease in the level of fluorescence compared to the control strain AND101 lacking a terminator. This shows that each region indeed contains a functional transcriptional terminator, most likely corresponding to the sequences given in Supplementary Figure S3. Different reductions in fluorescence were obtained, indicating that transcription was terminated with different efficiencies.

To test whether the *conAn* system was able to antiterminate these terminators, we engineered a set of strains in which the *conAn* system was placed in front of each terminator, resulting in the configuration (P_{spank} -[*conAnI*+2]-[Ter_x]-*gfp*). When grown in the presence of IPTG, significant increases in the level of fluorescence were obtained for each of these strains compared to the corresponding strain lacking the *conAnI*+2 system (Supplementary Figure S2C). Such a [*conAnI*+2]-dependent increase in fluorescence was also observed for the strain containing Ter_{cat} from plasmid pC194. Therefore, the *conAn* system of pLS20 acted as an antiterminator on all of the transcriptional terminators analysed, regardless of whether they form part of the conjugation operon.

ConAn2 exerts antitermination activity and ConAn1 is required for processive antitermination

We anticipated that both *conAn1* and *conAn2* would be required for the observed anti-termination effects. To our surprise though, we found that strain AND131 in which Ter₃₀ was preceded by only *conAn2* (i.e. P_{spank}-[*conAn2*]-[Ter₃₀]-*gfp*) displayed similarly high fluorescence levels compared to the isogenic strain AND210 containing both *conAn1* and *conAn2* (Figure 4), demonstrating that the *conAn2* region is sufficient for antitermination at the Ter₃₀ terminator. Results presented in Supplementary Figure S2C show that *conAn2* can antiterminate at least five other terminators.

These results suggest that *conAn2* is mechanistically responsible for AT, which contradicts the results in the earlier section that AT on pLS20cat required both *conAn1* and *conAn2* (Figure 1B). One potentially important difference between the native setting on pLS20cat and the AT screening system is that in the former case Ter₃₀ (the first terminator of the conjugation operon) is located 1,475 bp downstream of *conAn2*, whereas this distance is only 285 bp in the AT screening system. To examine if the distance between the *conAn2* and the terminator affected the efficiency of *conAn1*-independent AT we constructed derivatives of the AT screening strain with insertions of 824 bp or 1475 bp between *conAn2* and Ter₃₀ (i.e. P_{spank}-[*conAn2*]-[spacer]-[Ter₃₀]-*gfp*) (see Figure 4). Indeed, increasing the distance between *conAn2* and the terminator by 824 bp (strain AND223) or 1475 bp (strain AND262) caused fluorescence levels to drop by 73 and 84%, respectively (Figure 4). Importantly, high level fluorescence (i.e. efficient antitermination) was regained when *conAn1* was added to these constructs (i.e. *conAn1*+2, Figure 4 strains AND222 and AND261). These results show that *conAn1* is required for *conAn2* to act at long distances (i.e. P-AT).

A short GC-rich region and an inverted repeated structure are essential for *conAn2* function

The sequence of the gene 28–29 intergenic region is shown in Figure 5A. The region absent in pLS20Δds-II, which had an extremely low conjugation efficiency, has four striking features. First, a 20 bp region located near the 5' end is duplicated in inverse orientation near the 3' end of the region. In other words, a 238 bp long central region is flanked by a pair of 20 bp inverted repeat sequences. Second, a 16 bp GC-rich region (75% versus 42% for the entire *conAn2* region) is located 50 bp from the 5' end of the fragment. Third and fourth, inverted repeated structures that could form stem-loop structures in RNA with calculated free energies of −10.1 and −6.4 kCal/mol are located around positions 75 ('SL1') and 225 ('SL2') from the 5' end, respectively. To test if these features were important for antitermination we used our standard antitermination assay (effect on expression of *gfp* downstream of Ter₃₀) to compare mutant and wild type versions of ConAn2 in a ConAn1+2 context. A schematic presentation of the positions that were subjected to mutagenesis is given in Figure 5B. Details of the introduced mutations in each fragment are given in supplementary Figure S4.

The strains were grown in the presence of IPTG and subjected to FACS analyses. As shown in Figure 5B, al-

terations in the 20 bp inverted sequences located at 5' end (strain AND154), or at the 3' end (strain AND156) or at both ends (strain AND157) (constructs *conAn2*_{b-d}) did not significantly affect the level of antitermination. However, mutations in the 16 bp GC rich region (*conAn2*_e, strain AND159) and/or the region SL1 (*conAn2*_{f-g}, strains AND168, AND162) greatly reduced the level of fluorescence, suggesting that both of these features are required for efficient antitermination. Based on these results, it was not surprising that efficient antitermination was also abolished with *conAn2*_h (strain AND170), which contained the combined mutations of *conAn2*_{e-g}. Finally, antitermination was also affected in *conAn2*_{i-j}, which contained mutations in region SL2 (strains AND165 and AND172), although the effects of these mutations were less dramatic than those in the GC rich region, or SL1. In summary, the GC-rich and SL1 regions are crucial for efficient antitermination, whereas region SL2 has a less important role.

The antitermination processivity function of ConAn1 is confined to the N-terminal half of the protein

To see whether *conAn*-like systems were present on other genomes or plasmids, we performed *in silico* analyses and found that numerous conjugative plasmids of Gram+ bacteria encode a protein that shares similarity with ConAn1 of pLS20. To distinguish between putative homologs, we have added the name of the plasmid (subscript) to the *conAn* element name; i.e. the *conAn1* and *conAn2* elements of pLS20 are referred to as *conAn1*_{pLS20} and *conAn2*_{pLS20}, respectively. Remarkably, many of the identified putative ConAn1 homologs were smaller in size and corresponded to approximately the N-terminal half of ConAn1_{pLS20}. To test whether the N- or C-terminal halves of ConAn1_{pLS20} were functional *in vivo* we constructed strains JV40 and JV41 that contained the AT screening cassette with the configuration (P_{spank}-[X]-[*conAn2*_{pLS20}]-[824 bp spacer]-[Ter₃₀]-*gfp*), in which 'X' corresponds to *conAn1*_{Nter} or *conAn1*_{Cter}, respectively. As shown in Figure 4, high and low fluorescence levels were obtained for cells of strains JV40 and JV41, respectively, showing that the function required for processive AT is contained in the N-terminal half of ConAn1_{pLS20}.

***conAn* antitermination systems are present on other conjugative plasmids of Gram-positive bacteria, including pathogens**

Selection of conjugative plasmids from Gram-positive bacteria to analyze if they contain a bipartite conAn-type antitermination system. We selected eight (putative) conjugative plasmids of different Gram+ bacteria, including several pathogens, to test if they possess a bipartite *conAn*-like system. Although little is known about conjugative plasmids from Gram+ bacteria in general, there are a few exceptions. The *Bacillus thuringiensis* plasmid pAW63 has been studied for >20 years, and the enterococcal plasmids pAD1 and pCF10 for over 40 years (55–57). This was a main reason to select these three plasmids. Another reason to include pAD1 and pCF10 was that they contain virulence determinants (55,56). Moreover, pAD1 and pCF10 are harboured by *E. faecalis* strains which allowed comparison of *conAn*-type systems present on different plasmids harboured by

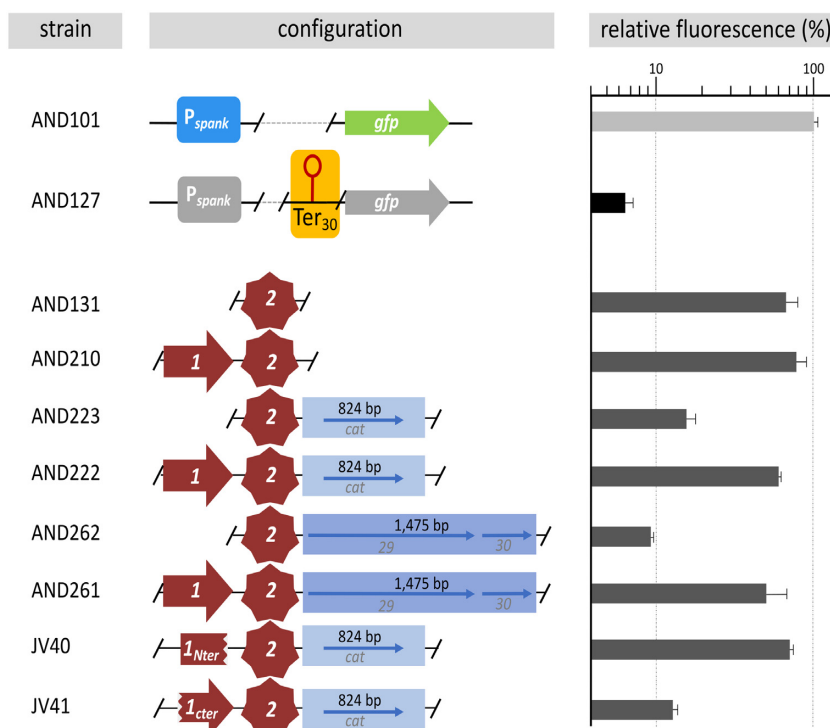


Figure 4. *conAn2* exerts antitermination activity and *conAn1* is required for processive antitermination. Cells of late exponential cultures grown in the absence or presence of 1 mM IPTG were subjected to FACS analyses to determine their fluorescence level. Cells corresponded to isogenic strains containing at their *amyE* locus a cassette in which no or different elements (see also text) are located in between the IPTG-inducible P_{spank} promoter and *gfp* as schematically shown. Conditions and symbols are identical as described in legend of Figure 2C. The regions of *conAn1* present in strains JV40 and JV41 correspond to codons 1–103 and 82–172, respectively. Error bars represent standard deviation.

the same bacterial species. This was also a major reason to select pFR55, which, like pAW63, is harboured by *B. thuringiensis*. Plasmid p576 was chosen for a similar reason. p576 is harboured by a *Bacillus pumilus* strain, which is closely related to *B. subtilis*, and we have shown before that p576 is closely related to pLS20 (58,59). Plasmid pN2 was chosen because it is present in *Lactobacillus paracasei* and hence, like pLS20, may play a role in horizontal gene transfer in the guts of humans and animals. Finally, plasmids p57330 and pLM5578 were selected because they are harboured by the pathogens *Clostridium sporogenes* and *Listeria monocytogenes*, respectively. *In silico* analyses suggested that each of the deduced conjugation operons of these plasmids contains multiple intrinsic transcriptional terminators (not shown), suggesting that they also require a P-AT system to overcome the termination sites for proper expression of the conjugation genes. Properties of these plasmids are given in Supplementary Table S4.

In silico identification of poorly conserved putative bipartite *conAn* systems at the beginning of the conjugation operon on each of the eight plasmids. *In silico* analyses identified a putative *conAn* system near the beginning of the (putative) conjugation operon on each of these plasmids, that would encode either a large (p576, pLM5578) or a small (pAW63, pFR55, pAD1, pCF10, pN2 and p57330) ConAn1_{pLS20} ortholog. These putative ConAn1 orthologs share only limited similarity with ConAn1_{pLS20} (ranging from 13 to 33%

identity, Supplementary Table S4); and even ConAn1 orthologs encoded by plasmids present in different strains of the same genus share low identity levels between 16 and 23%. A similar pattern was found for the putative *conAn2* regions: their predicted sizes range from about 260 to 600 bp, and the levels of sequence identity with *conAn2*_{pLS20} was between 30 and 40%, except for the putative *conAn2* region of *Bacillus pumilus* plasmid p576 which shared an identity of 70% with *conAn2*_{pLS20} (Supplementary Table S5).

All eight plasmids contain a bipartite *conAn* antitermination system that is functional in their native host. To study whether these regions constitute functional *conAn* systems we tested possible AT activity in their native genus, except for the region of *C. sporogenes* plasmid p57330, which was tested in both *Enterococcus faecalis* and *B. subtilis*. For this purpose, cassettes with the configuration (P_{spank}-[*conAn*_X]-[1.5 kb spacer]-[Ter₃₀]-gfp) were constructed in which *conAn*_X corresponds to one of the putative *conAn* systems. When *B. subtilis* strain 168 was used as a host, the cassettes containing the putative *conAn* systems were placed at the chromosomal *amyE* locus. For tests using other hosts, the cassettes containing the putative *conAn* systems were cloned onto the shuttle vector pAT18, which contains the replication functions of the low-copy broad host-range plasmid pAMβ1 (33), and then introduced into the corresponding host bacterium. Control experiments showed that terminator Ter₃₀ was functional in *E. faecalis* JH2–2, *L. in-*

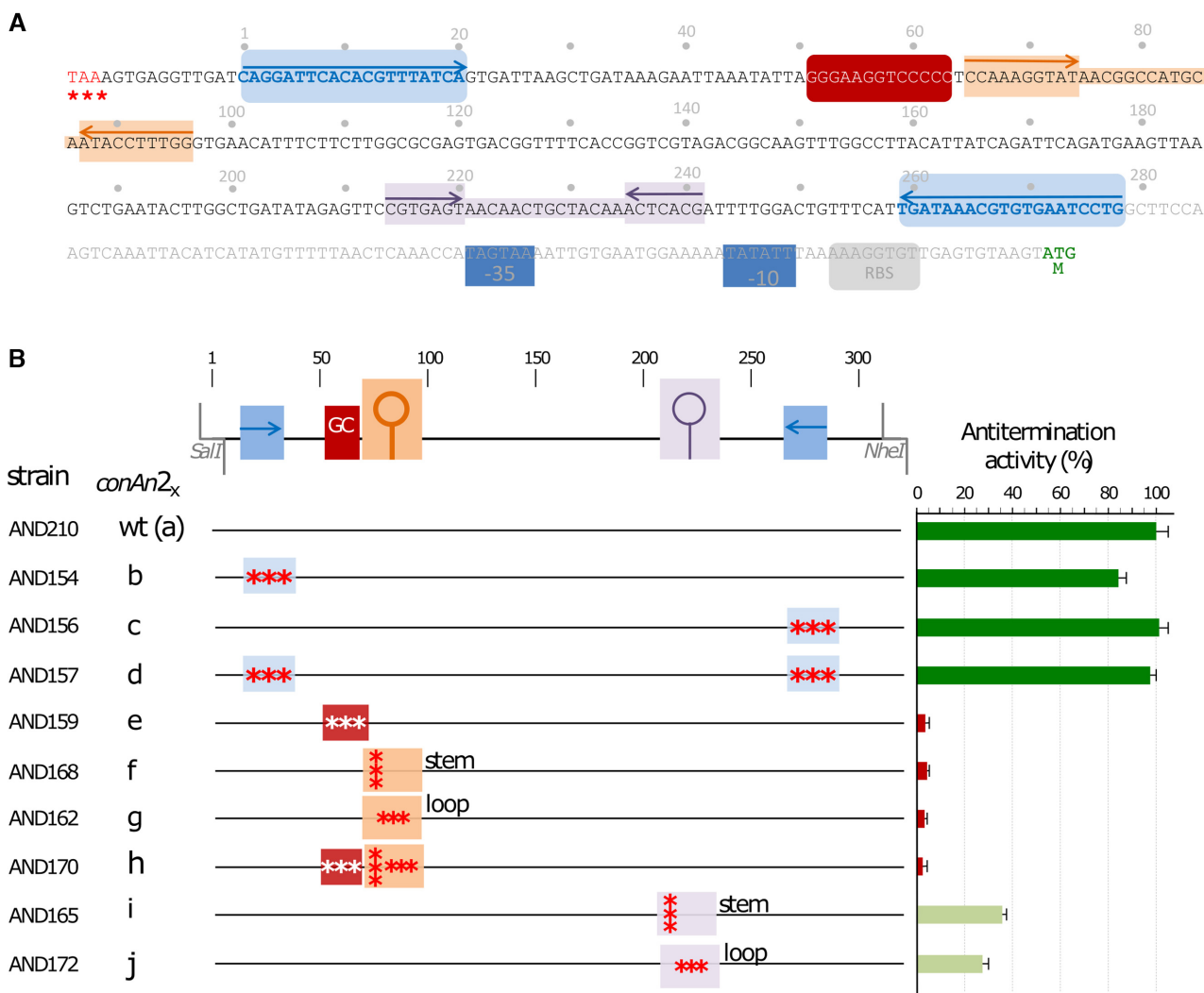


Figure 5. Features within *conAn2* crucial for antitermination. (A) Sequence of the 383 bp intergenic region between genes 28–29, together with the upstream-located TAA stop codon of gene 28 (written in red and highlighted with three red asterisks) and the downstream-located ATG start codon of gene 29 (the last 3 bp written in green and with a green letter ‘M’ reflecting the starting methionine residue). Numbering starts at the 5’ positioned 20 bp inverted repeat, which together with the 3’ positioned repeated sequence are indicated in blue and over lined with blue arrows. The sequence of the GC-rich region is shown in white on a red background. Regions highlighted in orange and purple correspond to regions predicted to form stem-loop structures in RNA with calculated free energies of –10.1 and –6.5 kcal/mol, respectively. The –35 and –10 boxes of promoter P₂₉ are indicated with blue boxes, and the likely Ribosome Binding Site (RBS) of gene 29 is indicated with a grey box. (B) Schematic representation of the *conAn2*-region and derivatives containing mutations in one or more features together with their effect on antitermination activity. Positions of the conspicuous features within the *conAn2* region are given in the top line using the same colour code as in ‘A’. Lines below the top line represent wild type (wt) and derivatives containing mutations in one or more the features. For simplicity the fragments containing different mutations are numbered as indicated. *conAn2*_{b,c,d} contain mutations in the 20 bp inverted repeats located at the 5’, 3’ end or at both ends, respectively. The mutations in the 5’ and 3’-located 20 bp in *conAn2*_b and *conAn2*_c are complementary such that these sequences constitute an inverted repeat of 20 bp in *conAn2*_d. The mutations in *conAn2*_e lower the GC content in this region from 75 to 38%. The mutations in *conAn2*_f were designed to disrupt the formation of a stem-loop when this region is transcribed into RNA. The mutations in *conAn2*_g also correspond to the stem-loop region but alter only positions of the predicted loop when this region is transcribed into RNA. *conAn2*_h contains the combined mutations present in *conAn2*_{e-g}. Finally, similar to *conAn2*_f and *conAn2*_g, the mutations in *conAn2*_i and *conAn2*_j were designed to disrupt the formation of a stem-loop when this region is transcribed into RNA, and to mutate positions in the loop, respectively. Left column gives the name of the strains containing the cassette with the configuration (P_{spank}–[*conAn1*]–[*conAn2*_x]–[Ter₃₀]–gfp), in which *conAn2*_x represents the wild type or a derivative containing the mutations indicated. The second column refers to the fragments used for cloning. Third column indicates feature(s) containing mutations, which are indicated with red asterisks. Fourth column presents the level of antitermination activity of each derivative relative to strain AND210 containing the wild type *conAn2*-region. Conditions and symbols are identical as described in legend of Figure 2C. Error bars represent standard deviation.

nocua CLIP 11262 and *L. casei* BL23, and that *P_{spank}* was induced by IPTG in these bacteria.

As shown in Table 1 and Supplementary Figures S5A and B, with the exception of the putative *conAn_{p576}* system of *B. pumilus* plasmid p576 (strain AND264), all the other regions tested displayed AT activity. These results demonstrate that: (i) in general, functional *conAn* AT systems can have very different sequences; (ii) some *conAn* systems from other bacteria can function in *B. subtilis*, even with low sequence conservation.

The conAn2 component of all eight conAn systems exerts antitermination. We also constructed strains to test if, like *conAn2_{pLS20}*, the putative *conAn2* regions from the various plasmids conferred AT activity when located immediately upstream of *Ter₃₀*. Analysis of these strains showed that all these *conAn2* regions displayed moderate or (very) efficient AT activities on *Ter₃₀* (Table 1 and Supplementary Figure S6A and B), with the exception of *conAn2_{pAD1}* which only showed a low level of AT activity. Therefore, except perhaps for *conAn2_{pAD1}*, the *conAn2* regions located downstream of their respective *conAn1* genes are also mechanistically responsible for AT in the other plasmids tested. Interestingly, although the *conAn2_{p576}* region alone resulted in good levels of AT when located just upstream of *Ter₃₀* (Supplementary Figure S6A, strain AND422), no AT activity was observed for the complete bipartite *conAn_{p576}* system when tested at a large distance from *Ter₃₀* (Supplementary Figure S5A, strain AND264), suggesting that *ConAn1_{p576}* is not functional in *B. subtilis*.

conAn systems display variable antitermination activities in heterologous hosts. Having shown that the *conAn* system of the *C. sporogenes* plasmid p57330 was functional in the heterologous hosts *B. subtilis* 168 (Supplementary Figure S5A) and *E. faecalis* JH2-2 (Supplementary Figure S5B), we next tested whether one or more of the *conAn2* regions alone and/or the [*conAn1*+2] systems of the listerial, enterococcal, lactobacillus and clostridial plasmids were functional in *B. subtilis*. As shown in Table 1 and Supplementary Figures S5A and S6A, the *conAn2* elements and the bipartite [*conAn1*+2] systems of the listerial, lactobacillus and clostridial plasmids all displayed AT activities in *B. subtilis* with different efficiencies that ranged from moderate to good. Significant AT activities were also observed for the *conAn2* regions of the enterococcal plasmids pAD1 and pCF10, although the AT levels were clearly lower than those observed for the other plasmids (Supplementary Figure S6A). Finally, while only a low level of AT was observed for the [*conAn1*+2] system of pAD1 (strain AND268), no AT activity was observed for the [*conAn1*+2] system of pCF10 (Supplementary Figure S5A, strain AND274), suggesting that, like *ConAn1_{p576}*, *ConAn1_{pCF10}* is not functional in *B. subtilis*.

In summary, the results showed: (i) that different *conAn* systems that have little shared similarity are functional in their native hosts; (ii) that some but not all *conAn* systems can be functional in a heterologous host; and (iii) proper functioning of a P-AT system depends on functionality of both the *ConAn1* and *conAn2* components in a given host.

Table 1. Functionality of *conAn* systems of conjugative plasmids of different Gram-positive bacteria

Host plasmid	<i>conAn</i> system (plasmid)	Functionality bipartite <i>conAn</i> system at 1.5 kb distance from <i>Ter₃₀</i>				Functionality <i>conAn2</i> at short distance from <i>Ter₃₀</i>			
		<i>B. subtilis</i> 168	<i>L. innocua</i> CLIP 11262	<i>E. faecalis</i> JH2-2	<i>L. casei</i> BL23	<i>B. subtilis</i> 168	<i>L. innocua</i> CLIP 11262	<i>E. faecalis</i> JH2-2	<i>L. casei</i> BL23
<i>B. subtilis</i>	pLS20	+++	x	x	x	+++	x	x	x
<i>B. pumilus</i>	p576	-	x	x	x	++	x	x	x
<i>B. thuringiensis</i>	pAW63	++	x	x	x	++	x	x	x
	pFR55	+++	x	x	x	++	x	x	x
<i>L. monocytogenes</i>	pLM5578	+	++	x	x	++	++	x	x
<i>E. faecalis</i>	pAD1	+	x	+++	x	+	x	-	x
	pCF10	-	x	+++	x	+	x	++	x
<i>L. paracasei</i>	pN2	++	x	+++	+++	++	x	+	+
<i>C. sporogenes</i>	P57330	+++	x	+++	x	++	x	++	x

The level of antitermination activity was studied by antitermination of *Ter₃₀*. Antitermination activity is expressed as the level of fluorescence observed for the different strains containing a (putative) *conAn* system with respect to the control strain lacking a *conAn* system, and considering 100% the level of fluorescence obtained by the *conAn* system of pLS20. ++, ≥70% fluorescence recovery; +, (25–70%) fluorescence recovery; -, ≤10% fluorescence recovery; x, not tested.

The *conAn* system of pLS20 is the founding member of a large family of bipartite antitermination systems present on conjugative plasmids of Gram-positive bacteria

Next, we searched databases to identify genes encoding proteins sharing similarity with these ConAn1 proteins. Each of the ConAn1 sequences was used as a query in multiple rounds of Psi-blastp searches of the NCBI nr database. These nine searches resulted in a total of 662 hits having an *E*-value of $\leq 1e-7$. The distribution of these identified hits per genus and bacterial species is shown in Supplementary Table S6. After pruning identical sequences, the list was reduced to 531 proteins. Except one hit, which has possibly been annotated incorrectly (see Supplementary Table S6), all other hits corresponded to proteins encoded by Gram+ bacteria belonging to the phylum Firmicutes (see Supplementary Table S6). 45 and 80 proteins had been annotated as PrgR or TraE, respectively, due to their similarities to the PrgR or TraE protein encoded by the enterococcus plasmids pCF10 and pAD1, respectively. All others had been annotated as 'hypothetical protein'.

The phylogenetic tree presented in Figure 6 shows that, except for the ConAn1 proteins encoded by *Bacillus* plasmids pLS20 and p576, all other ConAn1 protein tested are located in a separate branch. This indicates that the different ConAn1 proteins are evolutionary distantly related, which corroborates the low level of sequence similarity between them. It also suggests that the ConAn1 proteins identified here very likely do not represent all ConAn1 proteins. ConAn1_{pLS20} belongs to the group of 'large' ConAn1 proteins. All large ConAn1 proteins that contain an additional C-terminal domain form part of the two branches shown on a grey background in Figure 6, which shows that they form a minority. It should be mentioned that the eight putative ConAn1 members encoded by plasmids from *Exiguobacterium* and *Planococcus* species (plasmid pFR55 branch) also contained an additional domain. In these cases, the processivity domain (Pfam domain pXO2-34, PF17362) appeared to be fused to an N-terminal domain predicted to contain a helix-turn-helix DNA-binding motif (Pfam domain HTH_3, PF01381).

DISCUSSION

Here, we demonstrate the presence of a novel bipartite AT system (*conAn*) in several plasmid conjugation operons, and which is probably widespread on conjugative plasmids of Gram+ bacteria. The system is located near the beginning of the conjugation operon; the *conAn2* RNA component is essential and mechanistically responsible for antitermination, and the ConAn1 protein is required for processivity. ConAn is a genuine P-AT system because its activity is not limited to a single terminator, instead it acts on multiple terminators over a long range. The relative few processive antitermination systems described so far are constituted by either a protein or an RNA component (for review see, 20). Antiterminator proteins can act in *trans* and are recruited to transcription elongation complexes (TECs) in different ways, while the systems using RNA-based antitermination factors function only in *cis* (for review see, 20). The RNA component (*conAn2*) and the protein component (ConAn1) of the *conAn* system only function efficiently in *cis*, as ectopic expression of *conAn1* or *conAn1*+*conAn2* did

not restore efficient conjugation of pLS20Δ28 (Figure 1B). Moreover, in the *gfp*-based setup in which *conAn2* is located 1.5 kb upstream of Ter₃₀, ectopic expression of *conAn1* also did not result in efficient *conAn2*-mediated antitermination (our unpublished results). The *cis*-dependent activity of the *conAn* system would contribute to limiting antitermination strictly to the conjugation operon.

Presently, only two other RNA-based antiterminators are known: *put* elements of the lambdoid phage HK022 and the EAR element present within or upstream of an operon encoding biofilm or capsular polysaccharide proteins in *Bacilli* (16,18,19,60). Both are very different from the *conAn2* regions described here. First and most importantly, whereas *conAn2* requires ConAn1 to act at large distance, the *put* and EAR elements appear to act alone to achieve P-AT, although in the latter case the requirement of a cellular cofactor was not excluded (19). In addition, the *conAn2* region is much larger than the *put* and EAR elements.

Our *in silico* analysis suggests that *conAn*-type P-AT systems are widespread on conjugative plasmids of Gram+ bacteria, with over 500 *conAn*-like systems identified. Interestingly, most of the identified (putative) ConAn1 homologs are small monodomain proteins, corresponding to the N-terminal part of the ConAn1_{pLS20} protein that we showed is responsible for the processivity. In a minority of the ConAn1 proteins, including ConAn1_{pLS20}, this processivity domain appears to be fused to another domain of unknown function. In other putative ConAn1 members, such as those encoded by *Exiguobacterium* and *Planococcus* species, the processivity domain appears to be fused to an N-terminal domain predicted to contain a helix-turn-helix DNA-binding motif. Future studies will be required to determine the function of these additional domains.

At this moment, the exact mechanism by which *conAn2* exerts antitermination is unknown, but it may do so according to at least one of the following three models that are in line with the obtained results. In the first model, *conAn2* would exert antitermination in an indirect way, by binding to the core components of the RNAP complex or/and one or more transcription elongation factors. These interactions would alter the conformation and/or orientation of the RNA exit channel, or affect the configuration of a transcription elongation factor like NusA, such that stem-loop formation is affected when the terminator region is transcribed, or affect pausing and/or backtracking. In the second and third models, *conAn2* would enable antitermination in a direct manner. In the second model, *conAn2* would also bind directly to one or more protein components in the transcription elongation complex. This binding would place the *conAn2* structure in a specific configuration near the RNA exit channel such that the stem-loop with calculated free energy of -10 kCal/mol and/or the GC-rich region would be able to transiently hybridize with *de novo* synthesized 5' stem sequences of a terminator, thereby interfering with proper formation of a functional terminator. The third model is similar to the second one, but in this case, the *conAn2* region would have no intrinsic affinity for component(s) of the transcription elongation complex, and fully depend on ConAn1 for proper configuration.

The gut microbiome has high concentrations of bacteria and harbours a diverse reservoir of antibiotic resistance (AR) genes (61), so it would be expected to be a niche that

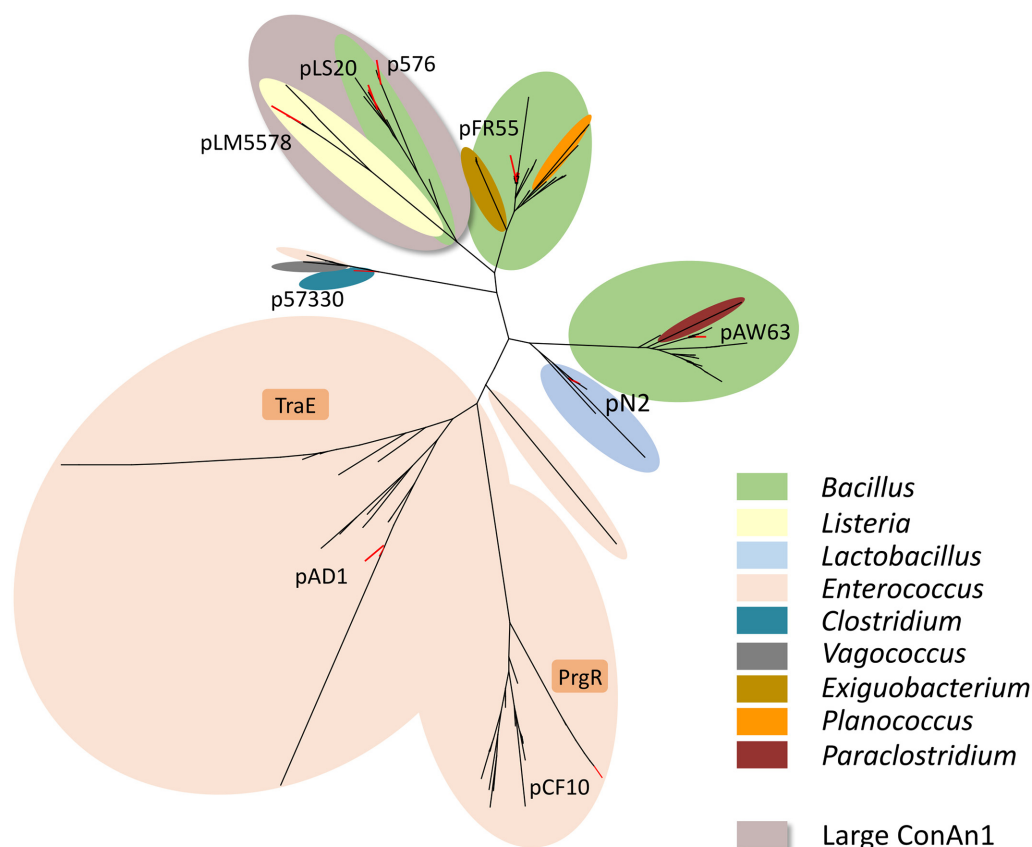


Figure 6. Phylogenetic tree of ConAn1 proteins. The phylogenetic tree was constructed using the program 'MEGA-X' and phylogeny was built using the Neighbour-joining algorithm (see Materials and methods). The tree was built using a set of 141 representative proteins having less than 95% sequence identity (obtained by MMSEQ2). The 141 ConAn1 members were clustered into nine clearly distinct branches. ConAn1 members having an additional C-terminal domain of unknown function are restricted to two branches (shown on a grey background). The nine ConAn1 members for which antitermination activity has been demonstrated (see text) are indicated. The majority of ConAn1 proteins clustered within each branch belong to the same bacterial genus, which are indicated with different colours.

is particularly apt for the transfer of AR genes between commensals and pathogenic bacteria. However, the rates of transfer of intrinsic antibiotic resistance genes between species or genera in the gut are much lower than expected (62). Our finding that several of the *conAn2* elements and ConAn1 processivity factors tested were less or hardly functional in heterologous hosts may provide an explanation for the low frequency of transfer in the gut. Our results indicate that both the ConAn1 and *conAn2* components must make functional interactions with the transcription machinery of the host to achieve efficient P-AT. The lower efficiency or non-functionality of *conAn* systems in heterologous hosts may not be surprising taking into account the low level of sequence conservation between them.

We have previously shown that the pLS20 conjugation operon is under the control of the strong P_c promoter, whose activity is strictly regulated at different levels including DNA looping. Our new results suggest that the principal function of the P_c promoter is not simply to control directly the expression of all the conjugation genes, but more to regulate the expression of the surface exclusion gene *ses_{pLS20}*, gene 30 of unknown function, and the *conAn* system. The *conAn* system in combination with multiple terminators most likely serves several purposes. First, it allows the co-ordinated expression of many genes located within a long

operon. Second, in combination with additional promoters it allows differential expression of one or more genes within the conjugation operon. An example of this is the surface exclusion gene *ses_{pLS20}*. When conjugation is not activated, the weak constitutive promoter P_{29} results in low levels of *Ses_{pLS20}*, sufficient to inhibit the transfer of pLS20 between donor cells 10-fold (30). Because transcripts starting at P_{29} will terminate at Ter_{30} located downstream of gene 30, gene 29 and 30 are expressed at low levels in all donor cells. However, in cells having activated the conjugation genes, *ses_{pLS20}* are also expressed from the much stronger P_c promoter and the elevated levels of *Ses_{pLS20}* inhibit >1,000-fold the transfer of pLS20 between two conjugation primed donor cells (30). In addition, these transcripts read through Ter_{30} , switching on the conjugation process. The *conAn* system also differentially expresses a toxin and antitoxin gene in conjugation primed and non-primed donor cells (our unpublished results, A.M., J.V., L.W., J.E. and W.M.). These examples highlight a novel function for a P-AT system: allowing differential expression of genes or cassettes embedded within the operon.

However, the pLS20 conjugation operon contains many more terminators than Ter_{30} . We consider it likely that these terminators contribute to the strict regulation of expression of the conjugation genes; i.e. ensuring that transcripts en-

coding the conjugation proteins are only generated when environmental signals are received, processed and lead to the activation of the P_c promoter. ‘Unintended’ transcription from cryptic promoters, named spurious transcription, is known to occur on a rather large scale (63–65). Without terminator sites inside the large (37 kb) conjugation operon, spurious transcription with the same directionality as the conjugation genes would generate transcripts until the end of the conjugation operon and hence undermine the strict regulation of the P_c promoter. The presence of multiple terminators inside the conjugation operon therefore will limit the likely negative effects of spurious transcription. The bypass mechanism to allow ‘intended’ transcripts initiated at P_c promoter to traverse the operon is achieved by the *conAn* antitermination system. This view is supported by our RNAseq data of a pLS20cat-harboring strain in which *rco*-, the repressor of the P_c promoter-, was ectopically overexpressed from a chromosomal locus. Although little transcription was observed for the plasmid region corresponding to the conjugation operon, low levels of transcripts were observed along the conjugation operon with some spikes, which disappeared at positions coinciding with the (putative) terminators. In summary, we propose that, the *conAn* system contributes to proper expression of the conjugation genes in three important ways: by promoting processivity across the whole 37 kb operon; by enabling differential regulation of genes within the operon; and by limiting the effects of spurious transcription.

DATA AVAILABILITY

RNAseq data have been deposited to the NCBI SRA repository (Project PRJNA717725). Flow Cytometry data is available from the corresponding author upon request.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr. Willem van Schaik for providing *Enterococcus faecalis* strain JH2–2 and plasmid pAT18, Dr Francisco García del Portillo and Dr Graciela Pucciarelli for providing *Listeria innocua* strain CLIP11262 as well as electrocompetent cells and help with *L. innocua* CLIP11262 transformation, Dr. Gloria del Solar for help with transforming *Lactobacillus casei* BL23 cells and Dr Praveen K. Singh for the construction of plasmid pKS17. We also thank Berta Raposo Ponce and other members of the flow cytometry unit and José ‘Pepe’ Belio for help preparing the figures. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

FUNDING

Ministry of Science and Innovation of the Spanish Government [PID2019_108778GB_C21 (AEI/FEDER, EU) to W.J.J.M., RTI2018-098517-B100 to J.M.I.]; Wellcome Investigator grant [209500 to J.E., L.J.W.]; EMBO Short-term Fellowship [7849]; FEMS research and training grant [FEMS-GO-2018–2019 to A.M.A.]; institutional grants

from the ‘Fundación Ramón Areces’ and ‘Banco de Santander’ to the Centro de Biología Molecular ‘Severo Ochoa’; publication fee by the CSIC Open Access Publication Initiative through its Unit of Information Resources for Research (URICI). Funding for open access charge: Ministry of Economy and Competitiveness of the Spanish Government [PID2019_108778GB_C21 (AEI/FEDER, EU) to W.J.J.M.]; URICI of the CSIC (Spanish Research Council). *Conflict of interest statement.* None declared.

REFERENCES

- Browning, D.F. and Busby, S.J. (2016) Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.*, **14**, 638–650.
- Washburn, R.S. and Gottesman, M.E. (2015) Regulation of transcription elongation and termination. *Biomolecules*, **5**, 1063–1078.
- Peters, J.M., Vangeloff, A.D. and Landick, R. (2011) Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.*, **412**, 793–813.
- Porrua, O. and Libri, D. (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.*, **16**, 190–202.
- Porrua, O., Boudvillain, M. and Libri, D. (2016) Transcription termination: variations on common themes. *Trends Genet.*, **32**, 508–522.
- Ray-Soni, A., Bellecourt, M.J. and Landick, R. (2016) Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.*, **85**, 319–347.
- Mitra, P., Ghosh, G., Hafeezunnisa, M. and Sen, R. (2017) Rho protein: roles and mechanisms. *Annu. Rev. Microbiol.*, **71**, 687–709.
- Roberts, J.W. (1969) Termination factor for RNA synthesis. *Nature*, **224**, 1168–1174.
- Roberts, J.W. (2019) Mechanisms of bacterial transcription termination. *J. Mol. Biol.*, **431**, 4030–4039.
- Greenblatt, J., McLimont, M. and Hanly, S. (1981) Termination of transcription by nusA gene protein of *Escherichia coli*. *Nature*, **292**, 215–220.
- Mondal, S., Yakhnin, A.V., Sebastian, A., Albert, I. and Babitzke, P. (2016) NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat. Microbiol.*, **1**, 15007.
- Santangelo, T.J. and Artsimovitch, I. (2011) Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.*, **9**, 319–329.
- Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867–879.
- Mars, R.A., Nicolas, P., Denham, E.L. and van Dijk, J.M. (2016) Regulatory RNAs in *Bacillus subtilis*: a Gram-positive perspective on bacterial RNA-mediated regulation of gene expression. *Microbiol. Mol. Biol. Rev.*, **80**, 1029–1057.
- Sherwood, A.V. and Henkin, T.M. (2016) Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.*, **70**, 361–374.
- King, R.A., Banik-Maiti, S., Jin, D.J. and Weisberg, R.A. (1996) Transcripts that increase the processivity and elongation rate of RNA polymerase. *Cell*, **87**, 893–903.
- Sen, R., King, R.A. and Weisberg, R.A. (2001) Modification of the properties of elongating RNA polymerase by persistent association with nascent antiterminator RNA. *Mol. Cell*, **7**, 993–1001.
- Komissarova, N., Velikodvorsky, T., Sen, R., King, R.A., Banik-Maiti, S. and Weisberg, R.A. (2008) Inhibition of a transcriptional pause by RNA anchoring to RNA polymerase. *Mol. Cell*, **31**, 683–694.
- Irnov, I. and Winkler, W.C. (2010) A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. *Mol. Microbiol.*, **76**, 559–575.
- Goodson, J.R. and Winkler, W.C. (2018) Processive antitermination. *Microbiol. Spectr.*, **6**, 1–15.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) In: *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
- Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T. et al.

- (2001) Comparative genomics of *Listeria* species. *Science*, **294**, 849–852.
23. de Man, J.C., Rogosa, M. and Sharpe, M.E. (1960) A medium for the cultivation of lactobacilli. *J. appl. Bact.*, **23**, 130–135.
 24. Bron, S., Meijer, W.J.J., Holsappel, S. and Haima, P. (1991) Plasmid instability and molecular cloning in *Bacillus subtilis*. *Res. Microbiol.*, **142**, 875–883.
 25. Shepard, B.D. and Gilmore, M.S. (1995) Electroporation and efficient transformation of *Enterococcus faecalis* grown in high concentrations of glycine. *Methods Mol. Biol.*, **47**, 217–226.
 26. Berthier, F., Zagorec, M., Champomier-Verges, M., Ehrlich, S.D. and Morel-Deville, F. (1996) Efficient transformation of *Lactobacillus sake* by electroporation. *Microbiology*, **142**, 1273–1279.
 27. Park, S.F. and Stewart, G.S. (1990) High-efficiency transformation of *Listeria monocytogenes* by electroporation of penicillin-treated cells. *Gene*, **94**, 129–132.
 28. Patrick, J.E. and Kearns, D.B. (2008) MinJ (YvjD) is a topological determinant of cell division in *Bacillus subtilis*. *Mol. Microbiol.*, **70**, 1166–1179.
 29. Arnaud, M., Chastanet, A. and Debarbouille, M. (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl. Environ. Microbiol.*, **70**, 6887–6891.
 30. Gago-Cordoba, C., Val-Calvo, J., Miguel-Arribas, A., Serrano, E., Singh, P.K., Abia, D., Wu, L.J. and Meijer, W.J.J. (2019) Surface exclusion revisited: function related to differential expression of the surface exclusion system of *Bacillus subtilis* plasmid pLS20. *Front. Microbiol.*, **10**, 1502.
 31. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.*, **24**, 79–88.
 32. Rudge, T.J., Federici, F., Steiner, P.J., Kan, A. and Haseloff, J. (2013) Cell polarity-driven instability generates self-organized, fractal patterning of cell layers. *ACS Synth. Biol.*, **2**, 705–714.
 33. Trieu-Cuot, P., Carlier, C., Poyart-Salmeron, C. and Courvalin, P. (1991) Shuttle vectors containing a multiple cloning site and a lacZ alpha gene for conjugal transfer of DNA from *Escherichia coli* to gram-positive bacteria. *Gene*, **102**, 99–104.
 34. 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.
 35. 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.
 36. Gautheret, D. and Lambert, A. (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.*, **313**, 1003–1011.
 37. Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A. and Sampath, R. (2001) RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic. Acids. Res.*, **29**, 4724–4735.
 38. Kingsford, C.L., Ayanbule, K. and Salzberg, S.L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.*, **8**, R22.
 39. Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schaffer, A.A. and Yu, Y.K. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J.*, **272**, 5101–5109.
 40. Schaffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V. and Altschul, S.F. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic. Acids. Res.*, **29**, 2994–3005.
 41. Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
 42. Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
 43. Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.*, **8**, 275–282.
 44. Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, **35**, 1547–1549.
 45. Huson, D.H. and Scornavacca, C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.*, **61**, 1061–1067.
 46. Meijer, W.J.J., Boer, D.R., Ares, S., Alfonso, C., Rojo, F., Luque-Ortega, J.R. and Wu, L.J. (2021) Multiple layered control of the conjugation process of the *Bacillus subtilis* plasmid pLS20. *Front. Mol. Biosci.*, **8**, 648468.
 47. Ramachandran, G., Singh, P.K., Luque-Ortega, J.R., Yuste, L., Alfonso, C., Rojo, F., Wu, L.J. and Meijer, W.J. (2014) A Complex genetic switch involving overlapping divergent promoters and DNA looping regulates expression of conjugation genes of a Gram-positive plasmid. *PLoS Genet.*, **10**, e1004733.
 48. Nojima, T., Lin, A.C., Fujii, T. and Endo, I. (2005) Determination of the termination efficiency of the transcription terminator using different fluorescent profiles in green fluorescent protein mutants. *Anal. Sci.*, **21**, 1479–1481.
 49. Cambray, G., Guimaraes, J.C., Mutalik, V.K., Lam, C., Mai, Q.A., Thimmaiah, T., Carothers, J.M., Arkin, A.P. and Endy, D. (2016) Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.*, **44**, 7006.
 50. Gusarov, I. and Nudler, E. (1999) The mechanism of intrinsic transcription termination. *Mol. Cell*, **3**, 495–504.
 51. Gusarov, I. and Nudler, E. (2001) Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell*, **107**, 437–449.
 52. Chen, Y.J., Liu, P., Nielsen, A.A., Brophy, J.A., Clancy, K., Peterson, T. and Voigt, C.A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods*, **10**, 659–664.
 53. Deighan, P. and Hochschild, A. (2007) The bacteriophage lambdaQ anti-terminator protein regulates late gene expression as a stable component of the transcription elongation complex. *Mol. Microbiol.*, **63**, 911–920.
 54. Horinuchi, S. and Weisblum, B. (1982) Nucleotide sequence and functional map of plasmid pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.*, **150**, 815–825.
 55. Clewell, D.B. (2011) Tales of conjugation and sex pheromones: a plasmid and enterococcal odyssey. *Mob. Genet. Elements*, **1**, 38–54.
 56. Dunny, G.M. (2013) Enterococcal sex pheromones: signaling, social behavior, and evolution. *Annu. Rev. Genet.*, **47**, 457–482.
 57. Wilks, A., Jayaswal, N., Lereclus, D. and Andrup, L. (1998) Characterization of plasmid pAW63, a second self-transmissible plasmid in *Bacillus thuringiensis* subsp. kurstaki HD73. *Microbiology*, **144** (Pt 5), 1263–1270.
 58. Singh, P.K., Ballesterio-Beltrán, 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.
 59. Val-Calvo, J., Miguel-Arribas, A., Gago-Cordoba, C., Lopez-Perez, A., Ramachandran, G., Singh, P.K., Ramos-Ruiz, R. and Meijer, W.J.J. (2019) Draft genome sequences of sporulation-impaired *Bacillus pumilus* strain NRS576 and its native plasmid p576. *Microbiol. Resour. Announc.*, **8**, e00089-19.
 60. Banik-Maiti, S., King, R.A. and Weisberg, R.A. (1997) The antiterminator RNA of phage HK022. *J. Mol. Biol.*, **272**, 677–687.
 61. Sommer, M.O., Dantas, G. and Church, G.M. (2009) Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*, **325**, 1128–1131.
 62. Ruppe, E., Ghoulane, A., Tap, J., Pons, N., Alvarez, A.S., Maziers, N., Cuesta, T., Hernando-Amado, S., Clares, I., Martinez, J.L. et al. (2019) Prediction of the intestinal resistome by a three-dimensional structure-based method. *Nat. Microbiol.*, **4**, 112–123.
 63. Lybecker, M., Bilusic, I. and Raghavan, R. (2014) Pervasive transcription: detecting functional RNAs in bacteria. *Transcription*, **5**, e944039.
 64. Wade, J.T. and Grainger, D.C. (2014) Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat. Rev. Microbiol.*, **12**, 647–653.
 65. Wade, J.T. and Grainger, D.C. (2018) Spurious transcription and its impact on cell function. *Transcription*, **9**, 182–189.