Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Identification of RNAs bound by Hfq reveals widespread RNA partners and a sporulation regulator in the human pathogen *Clostridioides difficile*

Pierre Boudry[a,b], Emma Piattelli[a], Emilie Drouineau[a], Johann Peltier [a,b], Anaïs Boutserin[a], Maxence Lejars[c], Eliane Hajnsdorf [c], Marc Monot [b,d], Bruno Dupuy [b], Isabelle Martin-Verstraete [b,e], Daniel Gautheret [a], Claire Toffano-Nioche [a], and Olga Soutourina [a,b,e]

aUniversité Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France; bLaboratoire Pathogenèses des Bactéries Anaérobies, Institut Pasteur, Université de Paris, Paris, France; cUMR8261, CNRS, Université de Paris, Institut de Biologie Physico-Chimique, Paris, France; dBiomics Platform, C2RT, Institut Pasteur, Paris, France; eInstitut Universitaire de France (IUF), Paris, France

ABSTRACT

Noncoding RNAs (ncRNA) have emerged as important components of regulatory networks governing bacterial physiology and virulence. Previous deep-sequencing analysis identified a large diversity of ncRNAs in the human enteropathogen *Clostridioides (Clostridium) difficile*. Some of them are *trans*-encoded RNAs that could require the RNA chaperone protein Hfq for their action. Recent analysis suggested a pleiotropic role of Hfq in *C. difficile* with the most pronounced effect on sporulation, a key process during the infectious cycle of this pathogen. However, a global view of RNAs interacting with *C. difficile* Hfq is missing. In the present study, we performed RNA immunoprecipitation high-throughput sequencing (RIP-Seq) to identify Hfq-associated RNAs in *C. difficile*. Our work revealed a large set of Hfq-interacting mRNAs and ncRNAs, including mRNA leaders and coding regions, known and potential new ncRNAs. In addition to *trans*-encoded RNAs, new categories of Hfq ligands were found including *cis*-antisense RNAs, riboswitches and CRISPR RNAs. ncRNA-mRNA and ncRNA-ncRNA pairings were postulated through computational predictions. Investigation of one of the Hfq-associated ncRNAs, RCd1, suggests that this RNA contributes to the control of late stages of sporulation in *C. difficile*. Altogether, these data provide essential molecular basis for further studies of post-transcriptional regulatory network in this enteropathogen.

Introduction

Clostridioides difficile (formerly *Clostridium difficile*) has emerged as an important human enteropathogen. This Gram-positive anaerobic spore-forming bacterium is considered as the main cause of health-care-associated diarrhoea worldwide [1,2]. Antibiotics exposure, age over 65 years and immune deficiency have been identified as major risk factors of *C. difficile* infection (CDI). With an increase in incidence rate, severity of infection forms, occurrence of community-acquired cases and emergence of hypervirulent strains, *C. difficile* infections are becoming a key public health concern accentuated by the general ageing of population in industrial countries. Normal gut microbiota protects healthy adults from CDI, since antibiotic treatments induce a colonic dysbiosis state that subsequently facilitates the development of this pathogen from acquired or pre-existing spores in healthy carriers [2,3]. In addition to two major toxins, TcdA and TcdB, causing alterations in the actin cytoskeleton of intestinal epithelial cells, several virulence and colonization factors have been identified to contribute to CDI [4–7]. However, many aspects controlling the infection cycle of this emerging pathogen still remain poorly understood. Among the most challenging traits of CDI are high recurrence rates that could

be associated at least in part with the ability to form highly resistant spores [6,8]. Additional mechanisms contributing to the successful adaptation of this pathogen inside the host remain to be explored. A better understanding of regulatory processes governing CDI cycle and pathogenesis is urgently needed to put forward new ideas for alternative therapeutic and diagnostic strategies.

Regulatory RNAs have been identified in all living organisms as key players for major adaptive responses and are known to control virulence in major pathogens [9–13]. Our recent data strongly suggest the importance of RNA-based mechanisms for the control of gene expression in *C. difficile* [14,15]. By RNA-seq and differential RNA-seq, we detected more than 200 regulatory RNAs including potential *trans*-encoded small regulatory RNAs (sRNA) located in intergenic regions (IGR), *cis*-antisense RNAs and riboswitches [14,15]. We also detected abundant small RNAs encoded by *C. difficile* CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR associated sequences) system, that might be important for *C. difficile* survival in bacteriophage-rich gut communities and also for the control of genetic exchanges favoured within gut microbiota [15–17]. We recently described type I toxin-antitoxin (TA) modules adjacent to CRISPR arrays with the first functional antisense

RNAs in this pathogen [18]. A total of 13 potential type I TA modules are present in the genome of *C. difficile* reference strain 630 that could also contribute to its fitness inside the host [19,20]. All these studies suggest that the regulatory RNA network controlling *C. difficile* pathophysiology is in many aspects unique and deserves further special attention [14].

Most characterized sRNAs act by base-pairing with their target mRNAs. These interactions usually assisted by RNA chaperone proteins like Hfq for *trans*-encoded riboregulators lead to modulation of mRNA translation and/or stability [21–23]. Hfq is an abundant bacterial RNA-binding protein closely related to the eukaryotic and archaeal Sm and Sm-like protein families forming homohexameric structures [24]. Together with the recently identified ProQ, Hfq is a major RNA chaperone well studied for its implication in the sRNA-mediated regulatory mechanisms, in particular in Gram-negative bacteria [25,26]. In these bacteria, Hfq facilitates the short and imperfect base-pairing interactions of *trans*-encoded sRNAs with their target mRNAs and thus has an important role in stress response and virulence of pathogenic species. Nevertheless, the role of Hfq in Gram-positive bacteria still remains largely questionable. The *C. difficile* genome encodes an Hfq homologue with several unique features among Gram-positive bacteria. This includes the conservation of key amino acid residues for RNA-binding surfaces and the presence of a particular C-terminal region. *C. difficile* Hfq depletion has pleiotropic effects on bacterial physiology with most pronounced impact on sporulation efficiency [27,28]. We recently demonstrated that *C. difficile* Hfq could be involved in the control of metabolic adaptation, biofilm formation, stress responses and sporulation [27]. This is in accordance with large transcriptome changes and altered accumulation of several sRNAs upon Hfq depletion.

We have previously identified nearly 100 potential *trans*-acting regulatory RNAs in the IGR of *C. difficile* genome that could require the RNA chaperone protein Hfq for their action [15]. However, a global view of Hfq-dependent regulatory networks in this species is still lacking. The development of new generation sequencing approaches provides powerful strategies for studying molecular interactions in the cells at the genome level. Here, we used a co-immunoprecipitation coupled to RNA-seq approach (RIP-seq: RNA immunoprecipitation-sequencing) to establish the first comprehensive map of the Hfq-associated RNAs in this important enteropathogen and *in silico* prediction to further associate ncRNA and target mRNA pairs. We found that *C. difficile* Hfq associates with different classes of RNAs, including mRNAs, 5′UTRs (untranslated regions), known regulatory RNAs and potential new sRNAs in IGRs. In particular, enriched RNA molecules include *trans*-encoded sRNAs but also *cis*-antisense RNAs, the toxin mRNA and antitoxin RNAs from all recently identified type I TA modules, a total of 12 CRISPR arrays and numerous riboswitches. The role of RCd1, an sRNA associated with Hfq *in vivo* and *in vitro*, has been analysed. Our results suggest that this Hfq-associated sRNA contributes to the control of late stages of sporulation in *C. difficile*. Altogether, these data provide an essential basis for establishing links between ncRNAs and the phenotypes associated with Hfq in *C. difficile* and constitute a valuable resource for further exploration of post-transcriptional regulatory networks in this emerging enteropathogen.

Material and methods

Plasmid and bacterial strain construction and growth conditions

C. difficile and *Escherichia coli* strains and plasmids used in this study are presented in Supplementary Table S1. *C. difficile* strains were grown anaerobically (5% H_2, 5% CO_2, and 90% N_2) in TY [29] or Brain Heart Infusion (BHI, Difco) media in an anaerobic chamber (Jacomex). When necessary, cefoxitin (Cfx; 25 μg/ml) and thiamphenicol (Tm; 15 μg/ml) were added to *C. difficile* cultures. *E. coli* strains were grown in LB broth [30], and when needed, ampicillin (100 μg/ml) or chloramphenicol (15 μg/ml) was added to the culture medium. The non-antibiotic analogue anhydrotetracycline (ATc) was used for induction of the P_{tet} promoter of pRPF185 vector derivatives in the *C. difficile* [31]. Strains carrying pRPF185 derivatives were generally grown in TY medium in the presence of 250 ng/ml ATc and 7.5 μg/ml Tm for 7.5 h and the induction of gene expression was confirmed by qRT-PCR. The number of spores was estimated in sporulation medium (SM) [32] or BHIS medium [27] with at least three biological replicates [33]. After 24 and 72 h of growth, 1 ml of culture was divided into two samples. To determine the total number of CFU, the first sample was serially diluted and plated on BHIS medium supplemented with 0.1% taurocholate. To determine the number of spores, the vegetative cells of the second sample were heat-killed by incubation for 25 min at 65°C prior to plating on BHIS medium supplemented with 0.1% taurocholate.

All routine plasmid constructions were carried out using standard procedures [34]. All primers used in this study are listed in Supplementary Table S1.

For RIP-seq experiments, the 3xFLAG-tag sequence was introduced to the C-terminal part of *hfq* gene as a translational fusion on pRFP185-derivative plasmid. The expression of 3xFLAG tagged Hfq protein was confirmed by dot blot analysis with anti-FLAG antibodies. As a positive control of CD1234 protein detection by Western blot, the 3xFLAG-tag sequence was introduced to the C-terminal part of *CD1234* gene on pRFP185-derivative plasmid.

To introduce the 3xFLAG-tag at the C-terminal part of the coding region of *CD1234* gene directly upstream the stop codon on the *C. difficile* chromosome, we used an allelic-coupled exchange technique based on a semi-suicidal plasmid vector carrying *E. coli* cytosine deaminase *codA* gene as counter-selection marker [35] that was improved by using an inducible toxin expression from type I toxin-antitoxin module instead of *codA* as counter-selection marker [19]. We used Gibson assembly to construct a plasmid for further conjugation and homologous recombination in *C. difficile*. The resulting derivative plasmid was transformed into the *E. coli* HB101 (RP4) and subsequently mated with *C. difficile* 630Δ*erm*. *C. difficile* transconjugants were selected by sub-culturing on BHI agar containing Tm (15 μg/ml) and Cfx (25 μg/ml).

RNA extraction, quantitative real-time PCR, Northern blot

Total RNA was isolated from *C. difficile* strains grown in TY or SM medium using Trizol (Sigma) or FastRNA PRO BLUE kit (MP Biomedicals) (see [36] for detailed protocol description). The cDNA synthesis by reverse transcription and quantitative real-time PCR analysis were performed as detailed in [36] using standard protocols. No template and no reverse transcriptase controls were systematically included in each experiment. In each sample, the relative expression for a gene was calculated relatively to the 16S rRNA gene or *dnaF* gene (*CD1305*) encoding DNA polymerase III. The relative change in gene expression was recorded as the ratio of normalized target concentrations ($\Delta\Delta$Ct) from at least three independent experiments [37]. For Northern blot analysis, 5 µg of total RNA was separated on a denaturing 6% or 8% polyacrylamide gel containing 8 M urea, and transferred to Hybond-N+ membrane (Amersham) by electroblotting using the Trans-blot cell from Bio-Rad in 1x TBE buffer (89 mM Tris-base, 89 mM boric acid and 2 mM EDTA). Following UV cross-linking of the samples to the membrane, prehybridization was carried out for 2 h at 42°C in 7 mL of prehybridization buffer ULTRAHyb (Ambion). Hybridization was performed overnight at 42°C in the same buffer in the presence of a [γ-^{32}P]-labelled DNA oligonucleotide probe. Alternatively, the probe was synthesized using PCR with 5′end-labelled primer complementary to RNA sequence. After hybridization, membranes were washed twice for 5 min in 50 ml 2X SSC (300 mM sodium chloride and 30 mM sodium citrate) 0.1% sodium dodecyl sulphate (SDS) buffer and twice for 15 min in 50 ml 0.1X SSC 0.1% SDS buffer. Radioactive signal was detected with a Typhoon system (Amersham). The size of the transcripts was estimated by comparison with RNA molecular weight standards (Invitrogen). The relative intensities of the bands from Northern blot analysis via autoradiography were quantified using ImageJ software. For determination of RNA half-lives the *C. difficile* strains were grown in TY medium supplemented with 250 ng/ml ATc and 7.5 µg/ml Tm for 7.5 h at 37°C. Samples were taken at different time points after the addition of 200 µg/ml rifampicin (0, 2, 5, 10, 20, 40, 60 and 120 min) and subjected to RNA extraction and Northern blotting.

RIP-seq experiment

RIP-seq experiment has been performed with four biological replicates as described previously with some modifications [38]. Briefly, Hfq::3xFLAG was expressed during 4.5 h by addition of 250 ng/ml ATc inducer after 3.5 h of culture in TY medium. Cells were suspended in the lysis buffer (50 mM HEPES–KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, protease inhibitor) supplemented with 0.1% SDS and 10% glycerol and lysed using a FastPrep (2 cycles of 45 sec at 6.5 m.s^{-1}). Cell debris were removed by centrifugation (16,000 g for 10 min). Total proteins were quantified with Bradford, and the concentration of protein extract was adjusted to 1 mg/ml. For 3xFLAG co-immunoprecipitation, 600 µl of magnetic beads coated with

3xFLAG antibodies ('Sigma' Anti-FLAG® M2 Magnetic Beads) were added for each sample and incubated during 1.5 h at 4°C on a rotating platform. For Hfq co-immunoprecipitation, 20 µl of serum from rabbit immunized against Hfq [27] was added and pre-immune serum served as a negative control. Hfq was trapped with magnetic beads coated with antibodies against rabbit IgG. Using a magnet, beads were washed five times with cold lysis buffer containing additional 350 mM NaCl. Beads were suspended twice in 500 µl and 200 µl of elution buffer (50 mM Tris, 1 mM EDTA, 1% SDS, pH 8.0 supplemented with 10% of glycerol) and incubated for 15 min at 65°C. A conventional acid phenol-chloroform extraction and ethanol precipitation were performed for the preparation of nucleic acids, followed by DNase treatment. The RNA was then processed into cDNA libraries using Illumina's TruSeq Stranded Total RNA Sample Prep Kit according to manufacturer's instructions with an adaptation of DNA/bead ratio of 1.8 to favour small fragments and sequenced using an Illumina HiSeq 2500 sequencer over 50 cycles (50 bp single-end sequencing).

Western blot

SDS-PAGE and Western immunoblotting were carried out using standard methods in triplicate [27]. Briefly, the strains were grown in SM medium for 12 h in the presence of 250 ng/ml ATc and 7.5 µg/ml Tm. 60 µg or 10 µg of whole cell lysate were loaded on the gels. The lysate of the strain overexpressing CD1234-FLAG from the plasmid served as a positive control for detection with anti-FLAG antibodies. Proteins were separated on 4–20% Bis-Tris polyacrylamide gels in Tris-Glycine SDS buffer. Immunoblotting with anti-SigA antibodies or Coomassie gel staining served as a loading control. The Hfq protein levels were verified by Western blotting in triplicate with polyclonal anti-Hfq antibodies [27]. The strains were grown in TY medium in the presence of 250 ng/ml ATc inducer and whole-cell extracts were loaded on a 4–12% gradient Bis-Tris polyacrylamide gel with MES running buffer. ImageJ software was used to estimate the protein levels.

RNA band-shift assay

Templates for the synthesis of RNA probes were obtained by PCR amplification using the Term and T7 oligonucleotides described in Supplementary Table S1. RNA synthesis, purification, concentration monitoring, 5′-labelling and Cd-HfqHis6 protein purification were performed as described in [27,28]. The RNA-binding capacity of purified Hfq was evaluated by mobility-shift assay using RCd1, *spoIIID* and *CD1234*. Band-shift assays were performed as previously described [39]. Briefly, RNAs were synthesized by T7 RNA polymerase and gel purified. Radiolabelled RNAs were incubated with increasing concentrations of Hfq for 10 min at 37°C in 10 µl of 100 mM NaCl/10 mM Tris-HCl (pH 8.1)/0.25% glycerol/0.5 mM DTT/1 mM EDTA/0.06% Triton X-100, and complexes were separated on 5% polyacrylamide/bisacrylamide (19/1) gels in 0.5× TBE buffer at 100 V at 4°C. The radioactivity level corresponding to the bound fraction of

RNA was plotted versus Hfq concentrations expressed on the basis of the monomer form. The data were fitted to the Hill equation, and half-saturation values ($K_{1/2}$) were estimated using Kaleidagraph for each RNA. When RCd1 was incubated with *spoIIID* or *CD1234*, each RNA was individually incubated in water at 70°C for 1 min and then for 5 min at room temperature. RCd1 was immediately mixed with SpoIIID or CD1234 transcripts or the equivalent amount of yeast RNA, in the presence or the absence of Hfq, in binding buffer and samples were incubated at 20°C for 20 min before loading on native polyacrylamide gels.

In silico analysis description

RIP-seq data analysis

After adaptors trimming using AlienTrimmer [40], sequencing reads were aligned to *C. difficile* strain 630 genome (NC_009089.1 region of the GCF_000009205.2 assembly and downloaded from NCBI) with the Bowtie2 software [41] and using default parameters. RIP-seq coverage visualizations are available through the COV2HTML software [42] at:

https://mmonot.eu/COV2HTML/visualisation.php?str_id=-75000032 (FLAG vs Control) https://mmonot.eu/COV2HTML/visualisation.php?str_id=-75000034 (Hfq vs PI).

Peak detection with MACS2

By retaining the sequencing orientation, we determined the Hfq binding regions from the strand-separated mapping data. We also excluded mapped reads on rRNA and tRNA regions because they are heavily covered and thus bias the peak-calling procedure. We applied MACS2 tool [43] thanks to the docker version (kathrinklee/macs2) with the following settings: the 4 3xFLAG, and 4 control replicas for treatment and control (-t and -c options) respectively, the effective genome size (-g) at 4,232,334 and estimated by the percentage of bases of the genome masked by the repeatmasker tool http://repeatmasker.org, the –broad option adapted to our RIP-seq experiment, –nomodel option, and the -ext_size to 51 option. We doubled those MACS2 runs for wild type strain experiment but with anti-Hfq-antibodies (Hfq630E) and pre-immune serum (PI630E) as couple of treatment and control inputs, respectively (same settings). RNA significant enrichment with Hfq-FLAG or Hfq antibodies was determined by applying the cut-off for broad region in MACS2 version used (macs2.1.0.20140616) that is based on q-value with a cut-off of 0.1 (use –broad option and the default value for the –broad-cut-off option). The significant peaks given by MACS2 were then compared to annotation provided by the MaGe platform [44] (chromosome NC_009089.1 of the *Clostridium difficile* 630 genome) with the intersect tool (-wao options) of the bedtools suite [45]. Results of this analysis with statistically significant enrichment of peaks are displayed in Supplementary Table S3.

Functional clustering with DAVID from KEGG DB

The functional categories of genes were annotated according to the KEGG database [46] based on the DAVID annotation [47] and general categories were assigned with manual curation.

Motifs detection with MEME

With the goal of discovering sequence motifs involved in the binding of Hfq to RNA, we applied software tools from the MEME suite [48] thanks to the docker version (ddiez/meme, v.4.11.1) on different pools of sequences corresponding to: the peaks, the 5′ region of CDS overlapped by a peak (± 100 bp from the translation start site), and the sRNA overlapped by a peak excluding those with multiple copies in the genome or with repeat motifs (CRISPR RNA, TA, c-di-GMP-dependent riboswitches). The fasta formatted sequences needed as inputs for MEME tools were extracted from the chromosome sequence with the faidx tool of the samtools suite [49] following the region locations and their strand. The MEME background model was determined with the fasta-get-markov tool (-m 6 -dna -norc) of the MEME suite. The motif discovery was realized with the meme tool (-mod anr -dna -nmotifs 50, MEME suite) restricting the space of search with the expected motif size corresponding to an RNA–RNA interaction (-minw 4 -maxw 10) or to an Hfq-RNA interaction (-minw 6 -maxw 30).

Prediction of ncRNA-target mRNA interactions with IntaRNA

The predictions of ncRNA-ncRNA interaction or ncRNA–mRNA interaction were made with the IntaRNA tool [50] thanks to the docker version (quay.io/biocontainers/intarna:2.0.4–boost1.61_1, detailed mode output option -D, reporting only one optimal interaction, -n 0). Except for the peak pool, we used the same pools of sequences as for motifs detection as entries for IntaRNA. We set the interaction energy ΔG cut-off to −10 kcal/mol for IntaRNA output presentation. RNAPredator (http://rna.tbi.univie.ac.at/cgi-bin/RNApredator/target_search.cgi) and RNADuplex prediction algorithms available online were also applied for RCd1 target predictions [51,52].

Results

Co-immunoprecipitation analysis of Hfq-binding RNAs in vivo in *C. difficile*

In order to identify all RNAs interacting with Hfq, we performed an Hfq co-immunoprecipitation RIP-seq experiment followed by high-throughput sequencing of RNAs co-purified with Hfq. For this purpose, we have constructed a strain carrying the plasmid pRPF185 with the *hfq* gene in translational fusion with the 3xFLAG tag at the C-terminal end of the protein (pDIA6151) (Supplementary Table S1). This strain named CDIP303 was grown together with the control strain carrying an empty plasmid in TY broth until the late exponential growth phase (8 h of growth). The level of overexpressed FLAG-tagged Hfq was estimated to be at least 10-fold higher as compared to the endogenous Hfq level suggesting that the majority of Hfq proteins in the cell were FLAG-tagged under tested conditions (Supplementary Figure S1).

We then immunoprecipitated Hfq-FLAG with an anti-FLAG antibody and purified associated RNAs. The strain carrying an empty plasmid lacking the 3xFLAG construct was used as a negative control for non-specific binding to anti-FLAG matrix. To assess the impact of 3xFLAG tag on Hfq interactions with RNAs, a control experiment was performed in parallel with the wild-type strain 630Δerm. From the cultures grown 8 h, we immunoprecipitated native Hfq with specific polyclonal antibodies previously used for Western-blot detection of the protein [27]. Pre-immune serum was used for the control sample. Generally, a lower specific signal strength was observed by this approach. The insufficient quality of polyclonal anti-Hfq antibodies led us to use 3xFLAG tagged Hfq for this experiment. The untagged Hfq co-immunoprecipitation (coIP) experiment only serves as an additional control and all further analyses were performed on replicates of Hfq-FLAG-tagged samples. Despite this robust strategy to specifically identify the Hfq-interacting RNAs, we cannot exclude potential limitations due to the pleiotropic impact of Hfq overexpression on *C. difficile* physiology and Hfq low specificity binding. The results from Hfq-untagged samples are shown in IGV visualization to illustrate the consistency between the two experiments for selected RNA partners. The results from Hfq-FLAG-tagged samples in four biological replicates were considered to establish a final list of Hfq ligands.

For each sample, between 27 and 72 million sequence reads were obtained and 97.7 to 99.1% of reads mapped to the *C. difficile* strain 630 genome (Supplementary Table S2). All four biological replicates clustered together for Hfq-FLAG and control samples, respectively, as shown by Principal Component Analysis (Supplementary Figure S2). Sequence reads were then classified according to the strain 630 genome annotation at their mapping sites (Fig. 1). For a small part of reads named ambiguous, no particular category could be assigned. In accordance with the data reported for other bacterial RNA chaperones [53,54], the majority (97.5%) of sequence reads in the control samples corresponded to non-specifically recovered abundant housekeeping rRNAs and

tRNAs. In contrast, about 60% of sequence reads mapped to rRNAs and tRNAs regions for Hfq-FLAG samples, but a significant enrichment was observed for sRNAs, mRNA coding regions (CDS) and new IGR regions (Fig. 1). Similar results were obtained for the untagged wild-type strain (Supplementary Figure S3) with 64% of sequence reads from rRNAs and tRNAs regions versus 86% for control samples, and the remaining reads being specifically enriched for CDS and sRNAs.

Identification of the RNA peaks specifically enriched in Hfq-immunoprecipitated samples

The MACS2 program was used for the analysis of sequencing data to identify peaks of reads that were significantly enriched in the Hfq-FLAG samples as compared to a control lacking FLAG-tag in an independent manner from existing gene annotation [55,56]. A similar analysis was also performed for the wild-type strain experiment with anti-Hfq-antibodies and pre-immune serum as a control. Minimum peak size was set to 51 nt for MACS2 peak calling, corresponding to the size of sequencing reads. The comparative analysis of Hfq-coIP and control samples revealed that the reads corresponding to housekeeping tRNA and rRNA were overrepresented among differentially detected reads and were thus eliminated from further peak calling procedure. Peaks were then mapped to the annotated genome of *C. difficile* strain 630 available on the MaGe platform [44] and their location visualized using Integrative Genomics Viewer (IGV) program [57]. Representative examples of different RNA functional groups are presented in Fig. 2. This analysis identified 808 peaks overlapping with the regions of significant enrichments of reads including 634 peaks for CDS regions and 178 peaks for ncRNAs: 97 known regulatory RNAs of different functional classes and 68 potentially new regulatory RNAs or UTR regions (Supplementary Table S3). Hfq-enriched peaks were categorized into the following groups: mRNAs, sRNAs, antisense RNAs and 5′ regulatory elements. Among these peaks *trans*-encoded sRNAs (Fig. 2A) but also *cis*-antisense RNAs
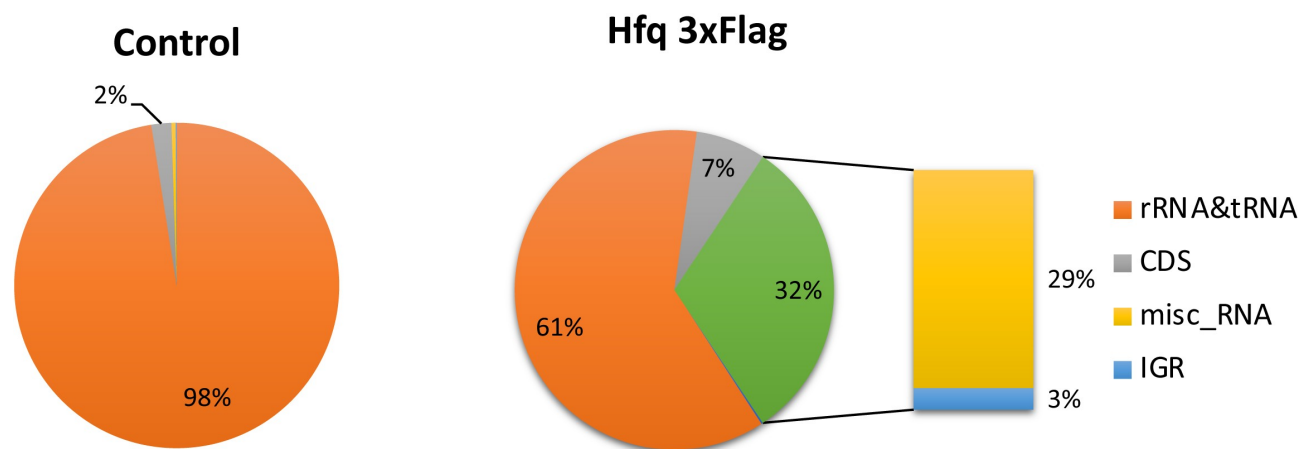
Figure 1. Diagram of relative proportion of different RNA species in Hfq-FLAG coIP and in control sample. All sequences that mapped to the *C. difficile* genome are represented. The rRNA and tRNA abundant housekeeping RNAs are shown in orange, the reads mapping to CDS are shown in grey. The relative proportion of other RNA species including known previously identified regulatory RNAs (named « misc_RNA ») and IGRs for Hfq-3xFLAG coIP samples is detailed in the diagram on the right. Left panel: control coIP, right panel: Hfq 3xFLAG coIP. RIP-seq experiment has been performed with four biological replicates.

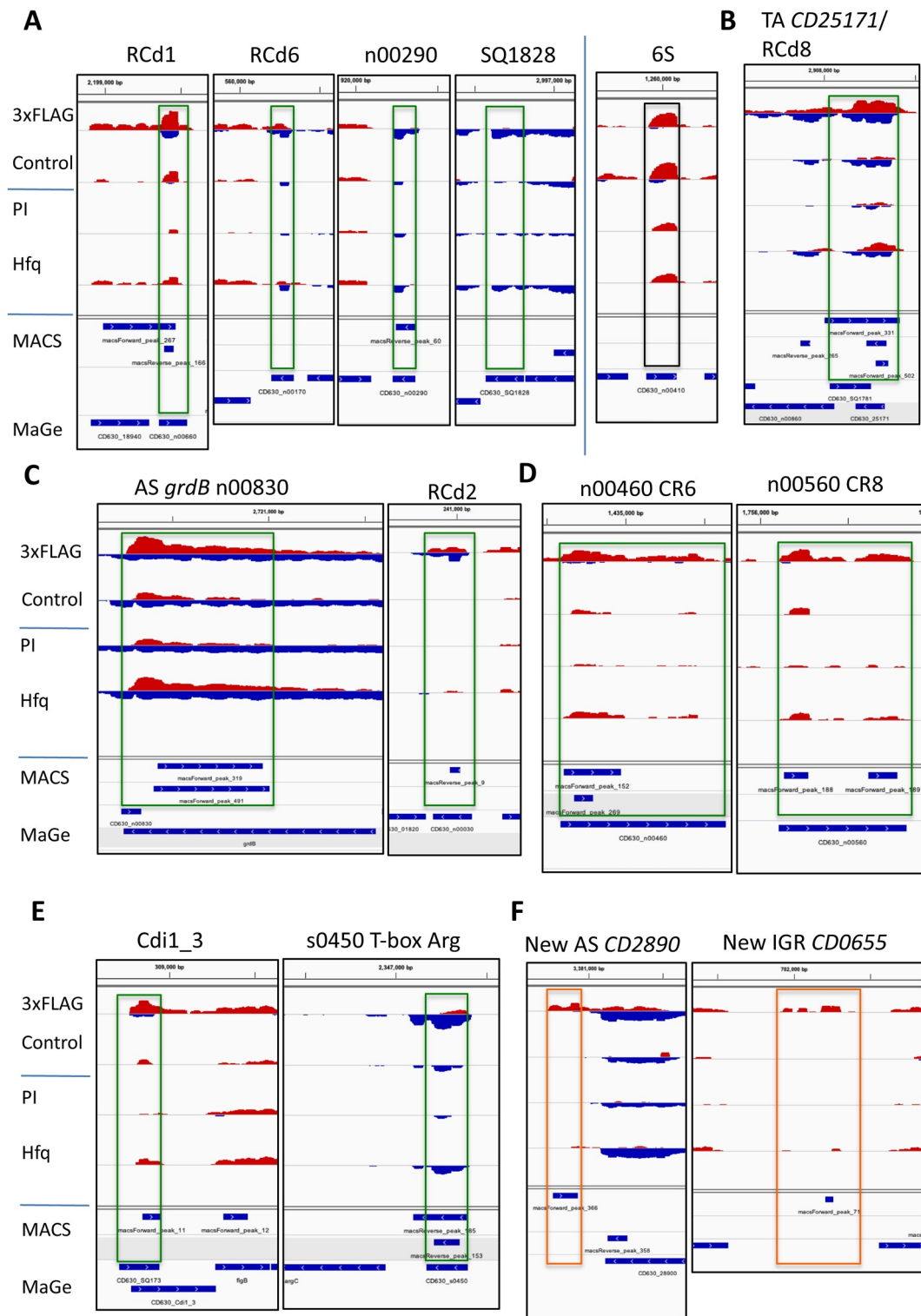**Figure 2.** Visualization of RIP-seq data for regulatory RNAs with IGV. Representative examples from different functional groups of RNAs are presented in A) for sRNAs, B) for TA loci, C) for other antisense RNAs, D) for CRISPR RNAs, E) for riboswitches, F) for MACS peaks without annotation for new ncRNAs in IGR or new antisense RNAs. The results for '3xFLAG' Hfq-coIP sample are compared to the 'Control' sample, while the 'Hfq' represents Hfq-coIP sample compared to the 'PI' pre-immune serum control. The genomic regions for previously identified sRNAs are presented in a green box, newly identified transcribed region are shown in orange box. 6S RNA (an Hfq-independent RNA) serves as a negative control shown in black box. The peaks identified by MACS are shown in blue together with genes from MaGe annotation at the bottom of each panel. In the IGV visualization profiles, '+' strand reads are shown in red and '–' strand reads are shown in blue. Dynamic IGV visualization is presented with adjusted read threshold for each window to compare the Hfq-immunoprecipitated and control samples.

(Fig. 2B-C), the toxin mRNA and antitoxin RNAs from all 13 recently identified type I TA modules (Fig. 2B), a total of 12 CRISPR arrays (Fig. 2D) and numerous riboswitches (Fig. 2E) were significantly enriched. All RNAs specifically referenced in the manuscript are described in Table 1.

A hundred of enriched RIP-seq peaks were detected in antisense orientation to gene annotations in the *C. difficile* 630 genome (Supplementary Table S3). Analysis of these peaks led to the identification of 36 potential new antisense RNAs, including 22 antisense RNAs to CDS, 4 antisense RNAs to previously identified ncRNAs and 10 antisense RNAs to riboswitches responding to SAM, lysine, glycine, FMN, cobalamin and T-boxes. In 7 cases, antisense peaks corresponded to 3′ extensions of antisense RNAs previously identified by transcriptional start site (TSS) mapping and whose 3′end remained undefined. Eighteen peaks in antisense orientation could represent either new RNAs antisense to CDS or transcription readthroughs from adjacent genes. A total of 45 antisense peaks were associated with genomic regions covered by overlapping bi-directional transcriptional units. These included the previously identified type I TA modules and other genes co-transcribed with associated antisense RNAs, as well as a number of overlapping genes in convergent orientation.

In general, RIP-seq results showed a specific increase for particular categories of RNAs for the strain bearing a FLAG tag. Over 75% of Hfq-enriched peaks mapped to a portion of mRNAs and the most abundant class of regulatory RNAs enriched in Hfq-FLAG samples corresponded to antisense RNAs (Supplementary Table S3). These antisense RNAs represented more than one-third of all regulatory RNA-derived peaks. We present below examples of highly enriched RNAs in the Hfq immunoprecipitated samples that belong to different functional classes (Fig. 2).

sRNAs enriched in the Hfq-immunoprecipitated samples. About 100 sRNAs were previously discovered in *C. difficile* that could act as potential *trans*-encoded RNA regulators in concert with the Hfq protein [15]. Of the Hfq-enriched peaks identified in this study, 24 corresponded to these sRNAs (Fig. 2A, Supplementary Table S4). Among previously identified sRNA interacting with Hfq *in vitro* [27], we show that RCd1 and RCd6 bound Hfq with a 254- and 16-fold enrichment, respectively, although the last interaction was not statistically validated (Fig. 2A). RCd1 is a ncRNA conserved in sequenced *C. difficile* strains that has been previously detected in both the historical strain 630 and the epidemic strain R20291 of 027 ribotype by Northern blot and RNA-seq [15]. Further studies demonstrated that the accumulation of RCd1 is negatively affected by the depletion of Hfq and a tight interaction of RCd1 with the Hfq protein has been shown *in vitro* by RNA band-shift experiments [27]. It is interesting to note that the RIP-seq experiment revealed the presence of an RNA antisense to RCd1 with a 78-fold enrichment in the Hfq-FLAG-immunoprecipitated samples (Fig. 2A). We detected this antisense RNA of around 100 nt by Northern blotting but the signal remained low under all conditions tested (Supplementary Figure S4). Two additional sRNAs, CD630_n00290 and SQ1828, whose accumulation was previously shown to be modulated by Hfq depletion [27] were

87- and 220-fold enriched in the Hfq-immunoprecipitated samples with statistical validation for CD630_n00290 (Fig. 2A). Other Hfq-enriched sRNAs are presented in Supplementary Figure S5A. Interestingly, antisense RNA transcription could be detected for several of these enriched peaks, including CD630_n00440, CD630_n00930, and CD630_n00330. By contrast and in accordance with previously reported co-immunoprecipitation results with Hfq from *Salmonella* [54], cDNAs of the abundant, Hfq-independent 6S RNA were found in rather equal numbers in the Hfq-immunoprecipitated and control samples (Fig. 2A).

Antisense RNAs. Remarkably, significantly enriched peaks mapped to the regions transcribed from both sense and antisense strands forming potential RNA duplexes. Data from our RNA-seq and TSS-mapping had revealed the presence of type I TA systems in *C. difficile* with the RNA antitoxin acting as antisense RNA to toxin mRNA [18–20]. We further characterized some of them [18,19] and identified a total of 13 type I TA modules in the reference strain 630 including the previously unannotated TA module with CD630_n00150 antitoxin/ *CD0440.1* toxin genes [20]. Our RIP-seq analysis shows that the toxin and antitoxin RNAs of all these systems bind to Hfq, as is the case for CD2517.1 toxin mRNA and RCd8 antitoxin (2,650-fold enrichment) (Fig. 2B) (see Supplementary Figure S5B for additional examples).

In addition to these 13 antitoxin RNAs, a total of 21 previously identified antisense RNAs was detected as enriched in Hfq coIP samples. Among them, we detected the antisense RNA CD630_n00830, which overlaps the *grdB* gene encoding a glycine reductase (Fig. 2C). This RNA was enriched approximately 2,800-fold for 3xFLAG Hfq compared to the control strain. This peak of 544 nt, associating with a strong TSS identified in our previous work [15], is significantly enriched in both Hfq-FLAG and wild-type strains and covers the 3′-coding part of *grdB* gene in antisense orientation (Fig. 2C). The antisense RNA, RCd2 was also 487-fold enriched in the Hfq coIP sample (Fig. 2C). This RNA is transcribed in an antisense orientation to two adjacent genes: *CD0182* of unknown function and *CD0183* encoding a putative cell wall hydrolase. RCd2 overlaps the 5′-end of the *CD0183* gene. Both sense and antisense reads could be detected in this overlapping region suggesting that these RNAs co-immunoprecipitated with Hfq. Other examples of enriched antisense RNAs were CD630_n00340 (3,027-fold enrichment) overlapping the *CD0898* gene of unknown function, CD630_n00080 (94-fold enrichment) overlapping the *CD0249* (*fliG*) flagellar motor gene, CD630_n00910 (182-fold enrichment) overlapping the *CD2711* gene encoding a LysR family transcriptional regulator and SQ476 (sevenfold enrichment) overlapping the *CD0740* gene encoding a putative pyridoxal phosphate-dependent aminotransferase (Supplementary Figure S5C and Table S3).

Interestingly, RIP-seq analysis revealed a peak corresponding to an antisense RNA overlapping the abundant tmRNA in *C. difficile* (Supplementary Figure S5C). Two other peaks significantly enriched in the Hfq coIP samples were detected antisense to the CD630_s0610 lysine-riboswitch sRNA (2,970-fold enrichment) and to the CD630_s0590 lysine-riboswitch sRNA (3,362-fold enrichment) (Supplementary Figure S5C).

Table 1. Examples of RNAs enriched in Hfq-FLAG RIP-seq samples.

name	strand	annotation	fold_enrichment	-log10(Pval)
ncRNAs				
AS CD630_s0610	+	AS Lysine riboswitch	88.96	5.84172
AS RCd1	-	Antisense ncRNA	78.17	6.83517
AS_CD630_28900	+	Antisense CDS	205	4.12438
CD630_Cdi1_3 (SQ173)	+	GEMM RNA motif	391.6	6.59405
CD630_Cdi1_4	+	GEMM RNA motif	294.48	5.31318
CD630_Cdi1_8	+	GEMM RNA motif	493.76	5.79994
CD630_Cdi1_9 (SQ1656)	+	GEMM RNA motif	309.79	4.25622
CD630_Cdi2_3	-	c-di-GMP-II	4,133.49	3.99184
CD630_n00030 (RCd2)	-	Antisense CDS	487.5	3.80984
CD630_n00080	-	Antisense CDS	94.4	4.5933
CD630_n00150	-	Antisense CDS TA	1,838.83	8.81823
CD630_n00170 (RCd6)	-	ncRNA IGR	17.72	ND
CD630_n00240	+	ncRNA IGR	442.5	5.82709
CD630_n00290	-	ncRNA IGR	87.13	3.85358
CD630_n00330	-	ncRNA IGR	6,627.44	6.46464
CD630_n00340	-	Antisense CDS	3,026.76	5.83478
CD630_n00370 (RCd10)	-	Antisense CDS TA	10,252.1	7.05302
CD630_n00380 (CRISPR 3_4)	+	ncRNA IGR	16.02	4.27916
CD630_n00440	+	ncRNA IGR	558.25	4.85723
CD630_n00460 (CRISPR 6)	+	ncRNA IGR	105.68	5.55447
CD630_n00470	-	ncRNA IGR	11.87	3.52267
CD630_n00500	-	Antisense CDS TA	5018	9.35359
CD630_n00510 (CRISPR 7)	+	ncRNA IGR	14.33	3.92762
CD630_n00560 (CRISPR 8)	+	ncRNA IGR	26.72	5.82053
CD630_n00600 (CRISPR 9)	+	ncRNA IGR	24.99	5.31146
CD630_n00620	+	ncRNA IGR	109.7	5.82838
CD630_n00660 (RCd1)	+	ncRNA IGR	254.73	4.09988
CD630_n00680 (RCd5, Cdi1_6)	+	ncRNA IGR	374.21	6.52781
CD630_n00690 (CRISPR 10)	-	ncRNA IGR	89.22	5.96526
CD630_n00790 (CRISPR 11)	-	ncRNA IGR	48.54	5.5121
CD630_n00830	+	Antisense CDS	2,799.15	4.11757
CD630_n00860 (CRISPR 12)	-	ncRNA IGR	27.35	6.97567
CD630_n00910	-	Antisense CDS	181.93	4.4372
CD630_n00930	-	ncRNA IGR	73.05	5.7913
CD630_n00980 (RCd12)	+	Antisense CDS TA	305.03	5.31318
CD630_n00990 (CRISPR 15_16)	-	ncRNA IGR	16.03	5.45665
CD630_n01000 (RCd9)	+	Antisense CDS TA	5,892.11	8.96925
CD630_n01010 (CRISPR 17)	-	ncRNA IGR	23.98	4.59146
CD630_s0190	+	T-box	486.61	4.9848
CD630_s0250	+	SAM riboswitch	1,188.27	3.74616
CD630_s0300	+	T-box	716.3	4.17452
CD630_s0310	+	T-box	334.8	4.17452
CD630_s0340	+	Lysine riboswitch	329.85	5.61064
CD630_s0360	-	FMN riboswitch	5,577.85	6.65922
CD630_s0450	-	T-box (Arg)	385.26	5.67265
CD630_s0480	+	T-box (Thr)	472.62	4.87999
CD630_s0500	-	Purine riboswitch	119.25	3.2587
CD630_s0510	-	T-box (Arg)	369.65	3.37631
CD630_s0590	-	Lysine riboswitch	3,361.69	8.13716
CD630_SQ1641	+	Antisense CDS TA	3,492.75	8.65138
CD630_SQ1642	+	Antisense CDS TA	193.25	8.65138
CD630_SQ1781 (RCd8)	-	ncRNA IGR TA	2,655	6.97567
CD630_SQ1828	-	ncRNA IGR	219.9	ND
CD630_SQ367	+	ncRNA IGR	41.4	4.0829
CD630_SQ476	+	Antisense CDS	7	4.52273
CD630_SQ808	-	Antisense CDS TA	1,894.75	6.8568
IGR_CD630_06550	+	ncRNA IGR	65	3.90214
mRNAs				
CD630_01640	+	Putative membrane protein	35.44	4.55381
CD630_03160	-	ABC-type transport system, permease	27.42	3.66946
CD630_03180	-	ABC-type transport system, bacitracin/multidrug-family ATP-binding protein	24.05	3.73138
CD630_03650	-	ABC-type transport system, multidrug-family permease Tn916-like, CTn1-Orf10	30.06	3.93697
CD630_04610	-	Two-component sensor histidine kinase	29.69	3.83025
CD630_04620	-	Two-component response regulator	16.13	3.45413
CD630_05760 (virS)	+	Sensor histidine kinase	68.09	4.74074
CD630_06630 (tcdA)	+	Toxin A	23.04	4.44902
CD630_06640 (tcdC)	-	Negative regulator of toxin gene expression	7.49	3.76101
CD630_09562	+	Putative phage protein TA	3.99	3.94307
CD630_10360 (cwp17)	+	Putative N-acetylmuramoyl-L-alanine amidase, autolysin	7	3.43363
CD630_10790	-	Transcriptional regulator, LysR family	27.55	4.45284
CD630_12330 (cwp26)	+	Putative cell surface protein	16.66	4.44377
CD630_14182	+	Putative phage protein TA	119.5	4.8236
CD630_15080	+	Putative iron-sulphur binding protein	24.34	4.03033
CD630_15170 (feoB)	-	Ferrous iron transport protein B	13.78	4.05364
CD630_17630	-	Putative membrane protein	179	4.52408
CD630_18030 (cwp23)	+	Putative cell wall-binding protein	14.11	4.4661
CD630_18290 (kdpD)	+	Two-component sensor histidine kinase KdpD	16.02	3.75086

Table 1. (Continued).

name	strand	annotation	fold_enrichment	-log10(Pval)
CD630_18300 (*kdpE*)	+	Two-component response regulator KdpE	10.55	4.2714
CD630_19090 (*eutP*)	+	Ethanolamine utilization protein, GTPase family	20.15	3.82847
CD630_19100 (*eutV*)	+	Two-component response regulator, Ethanolamine specific	10.55	4.55715
CD630_19740 (*hfq*)	-	Hfq RNA chaperone	105.1	5.05609
CD630_22991	-	Putative phage protein TA	75.25	6.03728
CD630_23050 (*pilW*)	+	Type IV pilus protein	70.28	4.96316
CD630_24830	+	Putative lipoprotein	29.26	4.98239
CD630_24830	+	Putative lipoprotein	29.26	4.98239
CD630_25180 (*cwp29*)	-	Cell surface protein	31.15	4.40865
CD630_26150	+	Transcriptional regulator, TetR family	132.63	4.06302
CD630_27500 (*agrB*)	-	Accessory gene regulator	59.03	3.92598
CD630_28520 (*dltB*)	-	D-alanyl transferase DltB, MBOAT family	22.57	3.81485
CD630_28530 (*dltA*)	-	D-alanine–poly(phosphoribitol) ligase subunit 1	13.99	3.85429
CD630_28540 (*dltD*)	-	D-alanine transferase	29.97	4.40777
CD630_28890	-	Putative phage protein TA	337.02	7.52469
CD630_29750	-	Putative CRISPR-associated Cas2 family protein	38.37	4.59146
CD630_29760	-	Putative CRISPR-associated Cas1 family protein	13.77	3.39713
CD630_30620	-	Transcriptional regulator, RpiR family	57.58	4.15443
CD630_31210	+	Putative flavodoxin/nitric oxide synthase	25.97	4.17693
CD630_31450 (*cbpA*)	-	Surface-exposed adhesin	20.96	3.0304
CD630_32040	-	Putative membrane protein	173.22	4.93516
CD630_32050	+	Putative nitroreductase	16.38	5.73797
CD630_35040 (*pilD*)	-	Type IV prepilin peptidase, A24A family	15.15	4.54123
CD630_35050 (*pilT*)	-	Type IV pilus twitching motility protein	8.24	3.24639
CD630_35110 (*pilC*)	-	Type IV pilus secretion protein	7.45	3.5423

The RNAs from different functional classes referenced in the text are included with the *p*-value from MACS2 analysis and the fold-enrichment in Hfq-FLAG RIP-seq samples as compared to control.

Other examples of potential new antisense RNAs co-immunoprecipitated with Hfq are presented in Supplementary Figure S5C.

CRISPR RNAs. Among the most abundant RNAs detected by our previous TSS-mapping experiment were CRISPR RNAs (crRNAs) [15] from the adaptive immunity CRISPR-Cas system. Here, RIP-sequencing identified enriched peaks (40-fold enrichment in average) for all twelve CRISPR arrays in the *C. difficile* 630 strain. In most cases, the peak covered mainly the 5′-portion of the CRISPR arrays (e.g. CRISPR 6, CRISPR 7, CRISPR 9, CRISPR 11, CRISPR 17 with 106-, 14-, 25-, 48- and 24-fold enrichment) corresponding to the most abundantly expressed leader-proximal part of the locus [15] (Fig. 2D for CRISPR 6 and Supplementary Figure S5D for other examples). Two separate peaks were detected in both 5′- and 3′ parts of CRISPR 8 array (27-fold enrichment) (Fig. 2D) and a longer 441-nt peak almost covering the entire CRISPR 10 array (89-fold enrichment) could be detected (Supplementary Figure S5D). In some cases, a regulatory function has been suggested for crRNAs in addition to defensive function [58,59]. It would be interesting to assess in the future this possibility for crRNA-mediated regulations of gene expression in *C. difficile*.

Riboswitches. Previous deep-sequencing analysis revealed that the *C. difficile* non-coding RNome is characterized by a great number of 66 5′-RNA regulatory elements called riboswitches [15,60]. In particular, 16 riboswitches responding to the signalling molecule c-di-GMP and mediating the coordinated control of motility, biofilm formation, toxin production and other community-behaviour related processes have been identified in this pathogen [15,61–64]. Our RIP-seq data, revealed a total of 27 riboswitches significantly enriched in the Hfq-FLAG sample, including numerous c-di-GMP riboswitches of type I and II, such as Cdi1_3 (391-fold enrichment) upstream of the flagellum operon (Fig. 2E) and

Cdi2_3 (4,133-fold enrichment) upstream of the *CD2831* gene encoding collagen-binding protein precursor (Supplementary Figure S5E). T-boxes responding to the aminoacylation level of tRNA, riboswitches responding to lysine, S-adenosyl-methionine (SAM), purine and flavine mononucleotides were also co-immunoprecipitated with Hfq (Fig. 2E for Arg T-box (385-fold enrichment) and Supplementary Figure S5E). In most cases, the RIP-seq peak was located specifically inside the riboswitch region corresponding to the short prematurely terminated transcript and the downstream mRNA region was not co-immunoprecipitated. These data raise a question on the possible implication of Hfq in riboswitch RNA-related regulatory function *in cis* but also *in trans*.

Potential new transcribed regions. Interestingly, 18 peaks overlapped IGR with no gene annotated in the *C. difficile* genome (Supplementary Table S3). The reannotation of TSS [65] allowed the assignment of 9 of these peaks to 5′UTR regions that were specifically enriched as compared to the coding part of the mRNA and one corresponded to a 3′UTR. Four others were located within IGR and might correspond to novel ncRNAs that were not identified in our previous studies, such as the region between the *CD0655* and *CD0656* genes adjacent to the pathogenicity locus (PaLoc) (Fig. 2F) (see Supplementary Figure S5F for other examples). One peak was assigned to a previously identified *cis*-antisense RNA with a misannotated 3′-end. One new *cis*-antisense RNA could also be detected as well as three additional double-stranded regions with potential antisense transcription.

Thus, in addition to previously identified regulatory RNAs, new potential RNAs were detected through their association with the Hfq protein.

mRNAs. Hfq RIP-seq data showed enrichment for numerous mRNAs (634) (Supplementary Table S3) such as the mRNA of the *CD1517* (*feoB*) gene (14-fold enrichment)
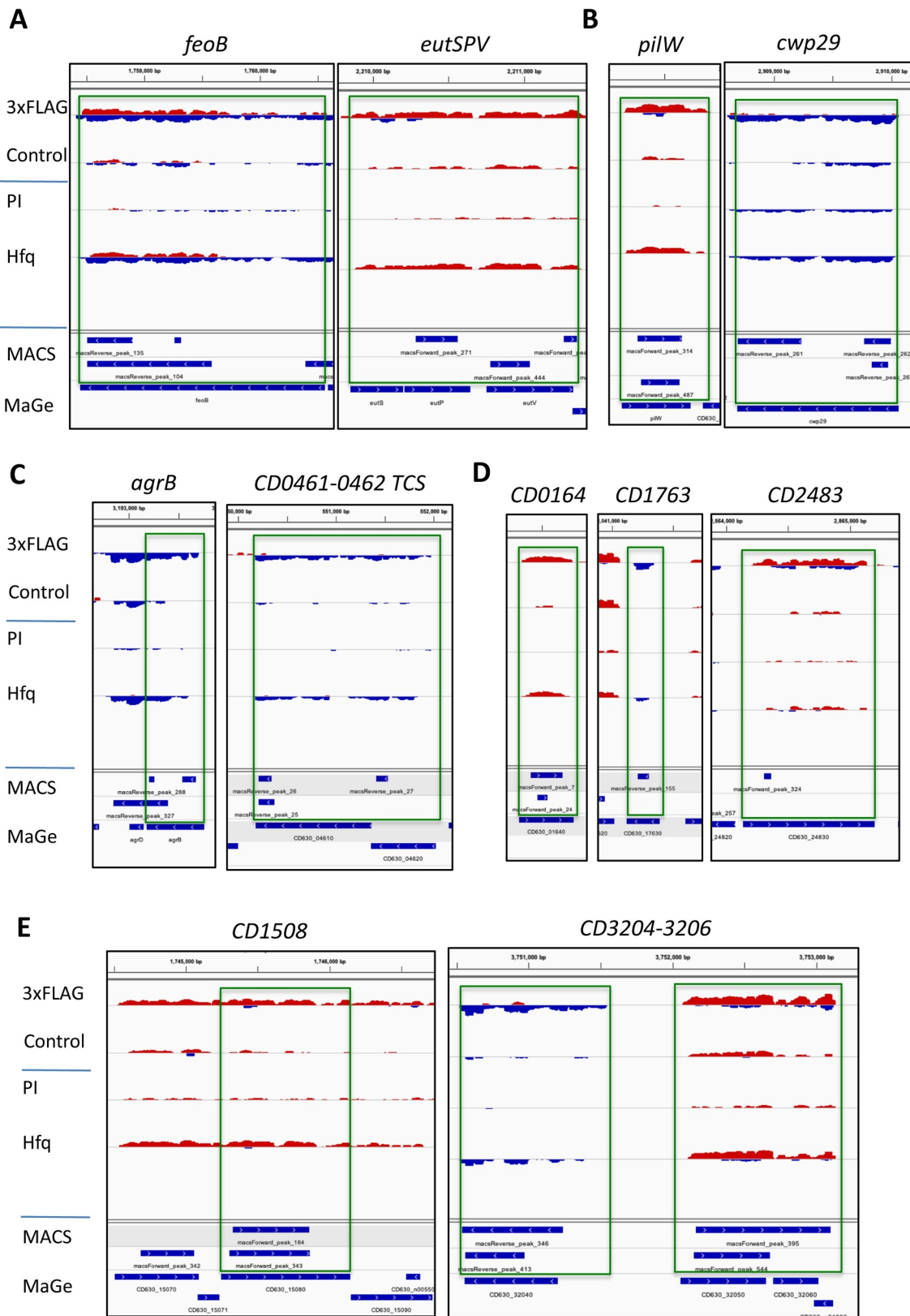
Figure 3. Visualization of RIP-seq data for mRNAs with IGV. Representative examples from different functional groups of mRNAs are presented in A) host-adaptation related genes, B) for type IV pili, adhesion, surface protein-encoding genes, C) for regulatory and signalling pathways including quorum sensing, two-component systems and transcriptional regulatory genes, D) for membrane proteins and transporter genes, E) for metabolism genes. The IGV visualization is presented as in. Fig. 2

encoding an iron transporter or of the *CD1909 (eutP)* gene (20-fold enrichment) encoding an ethanolamine utilization protein (Fig. 3A), whose expression is also modified in our transcriptomic analysis (Supplementary Table S3, Table S5, Table S6, Figure S5F). In addition to these genes, Supplementary Figure S5G shows other examples of virulence- and host-adaptation related genes revealed by the RIP-seq analysis including the *CD0663 (tcdA)* and the *CD0664 (tcdC)* genes (23- and 7.5-fold enrichment) encoding one of the major *C. difficile* virulence factors toxin A and a negative regulator of toxin gene expression, TcdC within the pathogenicity locus (Paloc) region, respectively, as well as the *CD2854-2851 (dltDABC)* operon (20-fold enrichment in average) encoding the D-alanylation proteins that provide protection against antimicrobial peptides [66,67].

Among genes with surface-related functions we observed a significant enrichment of the *CD3504 (pilD), CD3505 (pilT)* and *CD3511 (pilC)* genes of the *CD3513-3503* operon and the *CD2305 (pilW)* gene encoding type IV pili proteins, as well as the genes *CD1036 (cwp17), CD1233 (cwp26), CD1803 (cwp23), CD2518 (cwp29)* and *CD3145 (cbpA)* (15-, 8-, 7-, 70-, 7-, 17-, 14-, 31- and 21-fold enrichment) (Fig. 3B for *pilW* and *cwp29,* Supplementary Figure S5G for other examples).

A number of mRNAs for genes belonging to regulatory and signalling pathways were enriched in the Hfq-immunoprecipitated samples including the quorum sensing gene *agrB,* the two-component system-encoding genes *CD0461-0462* (Fig. 3C) and *CD1829-1830 (kdpDE),* the sensor histidine kinase *CD0576(virS)* as well as the transcriptional regulator genes *CD1079(LysR* family), *CD2615(TetR* family) and *CD3062(RpiR* family) (27-, 30-, 16-, 68-, 133-, 132- and 58-fold enrichment) (Supplementary Figure S5G). Numerous genes encoding membrane proteins and transporters were also detected such as the *CD0164* (Fig. 3D), the *CD2483* lipoprotein, the *CD0316-0318* and the *CD0364-0366* ABC transporter genes (69-, 29-, 27- and 30-fold enrichment) (Supplementary Figure S5G). In line with the suggested pleiotropic role of the RNA chaperone protein in *C. difficile,* numerous mRNAs for metabolism genes (Fig. 3E, Supplementary Figure S5G) were found as enriched in Hfq-immunoprecipitated samples, e.g. the iron-sulphur binding protein gene *CD1508,* the nitroreductase gene *CD3205* (Fig. 3E) and the flavodoxin/nitric oxide synthase gene *CD3121* (Supplementary Figure S5G) (24-, 16- and 26-fold enrichment).

Interestingly, in addition to CRISPR RNAs, the mRNAs for Cas1, Cas2 and Cas4 (CD2975-2976-2977 with 38-fold enrichment) CRISPR-associated adaptation proteins were co-immunoprecipitated with Hfq (Supplementary Figure S5C), further suggesting the potential implication of Hfq in CRISPR-Cas-related regulatory processes.

Hfq autoregulation has previously been shown in *E. coli* where Hfq represses its own translation primarily by binding to the *hfq* 5′UTR mRNA [68]. Accordingly, we observed a significant 105-fold enrichment of *CD1974 hfq* mRNA in the 3xFLAG Hfq-immunoprecipitated sample as compared to the control with a MACS peak covering almost the entire gene being found (Supplementary Figure S5G). Intriguingly, a faint

transcript in antisense orientation was also detected suggesting an additional level of regulation of *hfq* expression. In *Legionella pneumophila,* an anti-*hfq* RNA has recently been discovered as one of the central components of Hfq-dependent regulatory network governing this pathogen differentiation from nonvirulent to virulent bacteria [69]. Our findings thus suggest the existence of a complex autoregulatory loop controlling *hfq* expression in *C. difficile.*

Validation of RIP-seq data. To validate RIP-seq enrichment, we performed qRT-PCR analysis for selected RNAs from different functional classes on the Hfq-FLAG-immunoprecipitated and control samples in at least three biological replicates (Supplementary Table S7). This analysis confirmed the enrichment of 89-fold in average in the Hfq-FLAG-immunoprecipitated samples as compared to the control for all tested Hfq RNA partners. Among sRNAs, this validation included RCd1 (74-fold change) that binds Hfq *in vitro* with high affinity [27], as well as CD630_n00290 and SQ1828 (30- and 24-fold change, respectively) that were differentially expressed in *hfq*-depleted strain [27] (Fig. 2A). A 34-fold enrichment for the antisense RNA RCd2 was confirmed by qRT-PCR (Supplementary Table S7, Fig. 2C). Among CRISPR RNAs, the 61- and 15-fold enrichment of CD630_n00460 CRISPR 6 and CD630_n00560 CRISPR 8 was also validated (Supplementary Table S7, Fig. 2D). For the riboswitch class of Hfq partners, Cdi1_8, a representative of the highly homologous c-di-GMP-responsive riboswitches, was detected as highly enriched (820-fold) in Hfq-FLAG immunoprecipitated samples (Supplementary Figure S5E, Supplementary Table S7). Importantly, an enrichment of 15- and 30-fold was observed for the newly identified transcribed IGR *CD0655* RNA and the new antisense RNA to *CD2890,* respectively (Fig. 2F). In addition to the Northern blot validation of the new antisense RNA to RCd1 revealed by RIP-seq analysis (Supplementary Figure S4), the detection of five ncRNAs enriched in RIP-seq was also validated by Northern blot (Supplementary Figure S6). We have selected pathogenesis- and colonization-related genes *tcdA, tcdC* and *pilA* among mRNA partners of Hfq for qRT-PCR analysis and confirmed their enrichment in the Hfq-FLAG immunoprecipitated samples (31-, 3- and 5-fold enrichment) (Supplementary Figure S5G, Supplementary Table S7).

Our RIP-seq data are in complete accordance with all the previously published results from independent studies by various large-scale and targeted approaches on *C. difficile* ncRNA identification and functional characterization including antisense RNAs within TA systems, CRISPR RNAs, riboswitches and sRNAs. The majority of ncRNAs captured as Hfq ligands have been previously identified by genome-wide approaches including TSS mapping and RNA-seq [15,65] and some of them have been characterized by 5′-RACE, Northern blot, RT-PCR, *in vitro* interaction analysis and RNA stability assays [15,16,18–20,27,65].

In addition, the RIP-seq data on the enriched RNAs associated with Hfq overlapped with the results of our previous transcriptome analysis of the *C. difficile* strain depleted for Hfq under the same growth conditions [27]. This analysis revealed that 53 of the Hfq-bound RNAs (40 mRNAs and 13 ncRNAs, i.e. 7% and 18% of the total number of Hfq-

bound mRNAs and ncRNAs, respectively) were found differentially expressed in the Hfq-depleted strain as compared to the control suggesting a direct implication of Hfq-dependent post-transcriptional mechanisms in these regulations (Supplementary Table S5, Supplementary Figure S7). These genes are implicated in membrane transport, cell wall, regulation processes and sporulation including several members of the late sporulation phase sigma factor SigK regulon. Among ncRNAs, the riboswitches and CRISPR RNA classes were overrepresented.

To get a global view of the functional gene categories subjected to Hfq-dependent control, we analysed the functional categories of the RIP-seq targets using clusters of orthologous groups from the KEGG databases [46]. In accordance with previous phenotypic and transcriptomic analyses, this revealed three main functional categories (Supplementary Table S6, Fig. 4). The most abundant category was composed of 199 genes encoding membrane proteins with additional 45 genes encoding membrane transporters (enrichment score 9.98). Several clusters fitted into the functional category of regulation and signalling processes including a great number of transcriptional regulators, two-component signal transduction systems and diguanylate kinase signalling proteins. A large cluster of 113 genes corresponded to genes of unknown function. Functional clusters for genes encoding metabolic proteins were also overrepresented in our analysis. These data suggest that Hfq could contribute to important regulatory pathways in *C. difficile*.

Effect of RCd1 on the excision of the skin^{Cd} element. One of the sRNAs significantly enriched in Hfq-FLAG coimmunoprecipitated samples, RCd1, has previously been identified as a potential regulator of sporulation [27]. Interestingly, the increase in sporulation efficiency is also among the most pronounced effects observed for the Hfq depletion and *hfq* deletion strains [27,70]. Hfq tightly binds RCd1 *in vitro* and the accumulation of RCd1 is negatively affected by the Hfq

depletion [27]. However, the half-lives of RCd1 transcripts are similar in the wild-type and in the Hfq-depleted strain (Supplementary Figure S8). Since Hfq controls the regulon of the mother cell late sigma factor, SigK [27], we hypothesized that RCd1 also targets directly or indirectly *sigK*.

The *sigK* gene in *C. difficile* is interrupted by a 14 kb prophage-like *skin*^{Cd} element. The *skin* excision thus constitutes an important initial step for the σ^K-dependent cascade activation during sporulation. Two proteins encoded within the *skin* element, CD1231, a specific recombinase, and CD1234, a recombination directionality factor, are required for the *skin* excision [33]. To evaluate if the sporulation control by RCd1 could be linked to the *skin* element, we tested the effect of RCd1 overexpression on sporulation efficiency in a *C. difficile* strain lacking *skin*^{Cd} and therefore having an intact copy of the *sigK* gene [33]. We introduced the plasmid allowing the overexpression of RCd1 in the $\Delta skin^{Cd}$ strain as well as an empty plasmid used as a control. An estimated 3- to 7-fold overexpression of RCd1 was achieved in the presence of ATc inducer throughout the sporulation kinetics (Fig. 6). The number of spores formed in sporulation medium (SM) after 24 h of culture at 37°C was estimated and compared to that of the 630Δ*erm* carrying the same plasmids. While we obtained an important decrease in sporulation efficiency in 630Δ*erm* strain overexpressing RCd1 confirming our previous results [27], similar sporulation efficiencies were observed in the $\Delta skin^{Cd}$ strain overexpressing RCd1 and carrying the control vector (Fig. 5A). Thus, RCd1 loses its inhibitory effect on sporulation when the *skin*^{Cd} element is absent and the *sigK* gene is intact. This suggests that the RCd1 target(s) could be located within the *skin*^{Cd} element or at least be implicated in the *skin*^{Cd} excision. To test whether RCd1 acts on the *skin*^{Cd} excision, we extracted genomic DNA after 24 h of culture in SM medium for the 630Δ*erm* (pDIA6103) and 630Δ*erm* (pDIA6103-RCd1) strains. Using a pair of divergent primers at each end of the
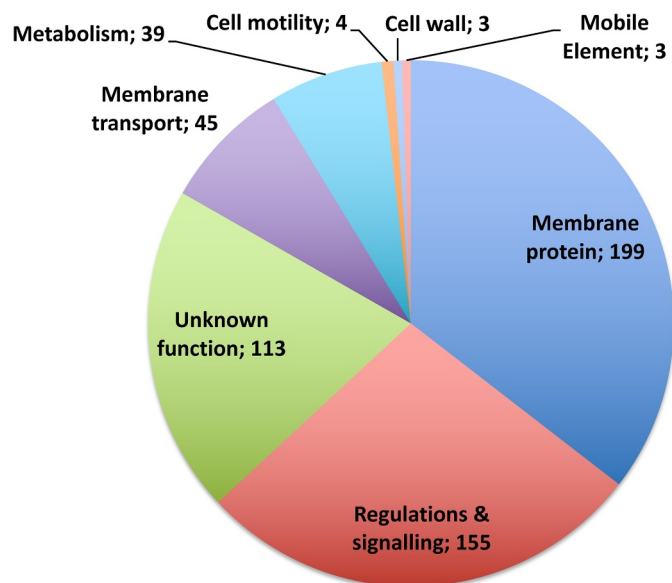

Figure 4. Diagram showing pathway clustering for mRNAs peaks detected by RIP-seq from KEGG database. The functional category of genes is shown with the corresponding gene number.
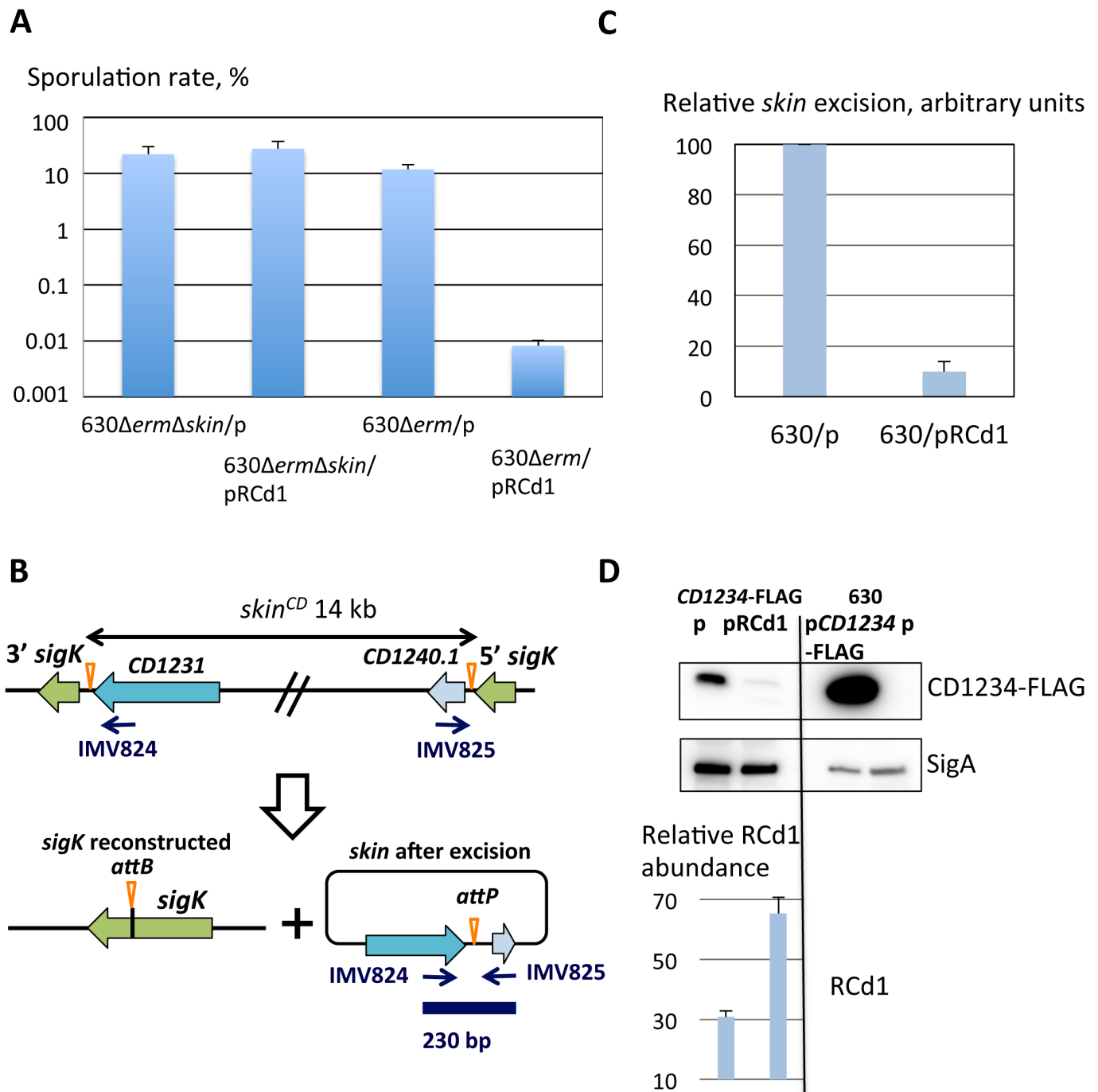
Figure 5. The role of RCd1 in the *skin* element excision. A) Sporulation efficiency by numeration of spores formed by the strains overexpressing RCd1 and their control strains. Sporulation assays were performed in SM after 24 h of culture. Spores were selected by heat treatment at 65℃ for 20 min and their number estimated for strains CDIP590 (630ΔermΔskin/p), CDIP592 (630ΔermΔskin/pRCd1), CDIP219 (630Δerm/p), CDIP258 (630Δerm/pRCd1). The diagram shows the sporulation rate as a proportion in % of spores for the total number of CFU, means of results from at least three independent experiments are presented. Error bars correspond to standard deviations from three biological replicates. Statistical significance for spore formation differences between strains CDIP219 and CDIP258 was evaluated by a Welch two sample *t* test ($p = 0.02$), $N = 3$. **B)** Diagram of the *skin* element integrated into the *sigK* gene and its episomal form shows the location of the oligonucleotides used as primers for PCR (IMV824/IMV825) to distinguish between integrated and episomal *skin* forms. **C)** The PCR results from three independent experiments for relative *skin* excision are presented. **D)** Western-blot detection of FLAG-tagged CD1234 protein in the cell extracts of *C. difficile* strains: lane 1- strain CDIP704 (CD1234-FLAG/p); lane 2- strain CDIP708 (CD1234-FLAG/pRCd1); lane 3- strain CDIP607 (630Δerm/pCD1234-FLAG); lane 4- CDIP369 (630Δerm/p). Immunoblotting with anti-FLAG antibodies detected a major polypeptide of ~11.5 kDa in whole-cell lysates of the strain carrying *CD1234*-3xFLAG gene on the chromosome of *C. difficile* 630Δerm strain but not in extracts of the wild-type strain expressing the non-tagged protein (negative control on lane 4). The representative result from three independent experiments is shown. The strains were grown in SM medium for 12 h in the presence of 250 ng/ml ATc and 7.5 μg/ml Tm. 60 μg (lanes 1–2) or 10 μg (lanes 3–4) of whole-cell lysate were loaded on the gels. The lysate of the strain overexpressing CD1234-FLAG from the plasmid serves as a positive control for detection (lane 3). Proteins were separated on 4–20% Bis-Tris polyacrylamide gels in Tris-Glycine SDS buffer. Immunoblotting with anti-SigA antibodies serves as loading control (bottom panel). The relative RCd1 abundance has been estimated by qRT-PCR, the mean values with standard deviations from three independent experiments are shown.

integrated sequence [33], we then quantified by qPCR the episomal form of the $skin^{Cd}$ element representative of the excision rate in the population. After excision, these two primers converge and amplify a 230 bp fragment on the $skin^{Cd}$ episome (Fig. 5B). We found that the episomal form is 10 times more abundant in 630Δerm (pDIA6103) than in 630Δerm (pDIA6103-RCd1) relative to the reference $dnaF$ gene encoding the DNA polymerase III (Fig. 5C). This indicates that overexpression of RCd1 results in a reduction of excision rate of the $skin^{Cd}$ element. Thus, RCd1 probably inhibits sporulation by preventing reconstitution of the intact $sigK$ gene and therefore SigK production, inhibiting the late sporulation steps dependent on SigK.

Effect of RCd1 on the abundance of potential targets. To assess the effect of RCd1 overexpression on the abundance of potential targets, we evaluated their expression by qRT-PCR analysis. To do so, the 630Δerm strains carrying either a plasmid overexpressing RCd1 under the control of the inducible P_{tet} promoter or an empty plasmid as a control were grown in TY medium until the late exponential phase or for 12 h in SM medium. In agreement with our previous results [27], we observed that the $sigK$ gene was less expressed in the strain overexpressing RCd1 compared to the control (Table 2). We also measured the impact of RCd1 overexpression on genes belonging to the SigK regulon, i.e. $CD0597$ and $CD1613$ genes encoding the spore coat proteins CotJB and CotA, respectively. $CD0597$ and $CD1613$ mRNA levels were reduced when RCd1 was overexpressed (Table 2). Thus, RCd1 is able to modulate the abundance of not only the $sigK$ transcript but also its targets within the SigK regulon.

As RCd1 overexpression had an impact on the $skin^{Cd}$ element excision, we further tested, as potential RCd1 targets, $spoIIID$ controlling $skin^{Cd}$ excision and $sigK$ expression [33] as well as three $skin^{Cd}$ genes. Genes $CD1231$ and $CD1234$ encode essential components of the $skin^{Cd}$ excision complex [33]. We included $CD1233$ encoding the putative cell surface protein Cwp26 since it might be an RCd1 target as predicted by the RNAPredator algorithm. While the expression of $spoIIID$ and $CD1234$ decreased upon RCd1 overexpression, no effect on $CD1231$ and $CD1233$ expression was observed in our conditions (Table 2). We also included the $sigE$ gene encoding the mother cell sigma factor involved in the transcription of $spoIIID$, $sigK$ and $CD1234$, as well as SigE targets, i.e. $CD1192$ encoding the SpoIIIAA protein, $CD1940$ encoding a putative membrane protein and $CD3542$ ($spmA$) encoding the spore maturation protein A. After 12 h of growth in SM

medium, no effect of RCd1 overexpression on the expression of $sigE$ and its target was observed (Supplementary Table S8) positioning $spoIIID$ and $CD1234$ as main regulatory components for RCd1 targeting. By short 7 min pulse induction of RCd1 overexpression, we confirmed a negative effect on expression of both genes (Table 2).

We then analysed a kinetic of expression of RCd1 and its potential targets in SM during sporulation. We harvested culture aliquots every 2 h between 10 h and 24 h. After RNA extraction, the expression profile of RCd1 and selected potential targets was assessed by Northern blot (Fig. 6A). This analysis showed that RCd1 is expressed throughout the sporulation kinetics, being most abundant during the early kinetic time points. Interestingly, as previously observed for the $sigK$ gene [33], a biphasic expression pattern was detected for $spoIIID$ and $CD1234$ genes (Fig. 6A), a first peak of expression occurring at the beginning of sporulation (10 h-14 h) and a second increased expression after 20 h-24 h of growth. We also confirmed the negative effect of RCd1 overexpression on both $spoIIID$ and $CD1234$ during sporulation kinetics (Fig. 6B). RCd1 overexpression was particularly effective at 10 h–12 h of growth and induced the most pronounced negative effect on $spoIIID$ and $CD1234$ potential targets. The complex expression profiles of $spoIIID$ and $CD1234$ result from multiple regulations by specific sigma factors and other regulators during sporulation. We cannot exclude that an additional missing factor controls the action of RCd1 during sporulation.

To further investigate the effect of RCd1 overexpression on CD1234 production and the regulation of the $skin^{Cd}$ excision, we introduced a 3xFLAG tag at the C-terminal part of the CD1234 coding region on the chromosome of $C.$ $difficile$ 630Δerm strain (Supplementary Table S1). Derivative strains carrying a chromosomal $CD1234$-3xFLAG copy and either an empty vector or a plasmid for RCd1 overexpression were then used for Western blot analysis. The 630Δerm strain carrying an empty vector or overexpressing a CD1234-3xFLAG protein fusion from the plasmid served as negative and positive controls, respectively. Based on our expression profile analysis during the sporulation kinetics, we extracted proteins from cells grown for 12 h in SM medium. Western blot analysis with anti-FLAG tag antibodies revealed a sharp decrease in the amount of CD1234-3xFLAG protein in the strain overexpressing RCd1 (Fig. 5D lane 2) compared to the strain carrying the control vector (Fig. 5D lane 1). While the CD1231 recombinase gene is constitutively expressed, the $CD1234$ gene is expressed under the control of both σ^E and

Table 2. Effect of RCd1 overexpression on the abundance of potential targets.

Gene ID	Name	Function	Ratio 630p/630pRCd1		
			TY	TY pulse	SM*
CD1230	sigK	RNA polymerase sigma-K factor SigK	**170 ± 45**	ND	**10.5 ± 1.8**
CD0126	spoIIID	Stage III sporulation protein D	**6.8 ± 1.5**	**5.3 ± 1.3**	**7.8 ± 1.1**
CD0597	cotJB	Spore coat protein	**43 ± 10**	**41 ± 13**	**33 ± 2**
CD1613	cotA	Spore outer coat layer protein CotA	**62 ± 25**	**30 ± 7**	ND
CD1231		Site-specific recombinase for *skin* excision	1.1 ± 0.4	1.3 ± 0.6	0.9 ± 0.4
CD1233	cwp26	Putative cell surface protein	1.2 ± 0.7	0.9 ± 0.3	0.7 ± 0.2
CD1234		Assessory protein for *skin* excision complex	**5.7 ± 2.9**	**8.8 ± 3.5**	**5.8 ± 2.5**

* After 8 h of growth in TY with continuous ATc induction or 'pulse' ATc induction for 7 min or 12 h of growth in SM medium in the presence of ATc inducer.
The effect of RCd1 overexpression has been estimated by qRT-PCR analysis in triplicate as described in Material and Methods.
Mean values ± standard deviations from at least three biological replicates are presented. ND, not determined. The values greater than 2 are shown in green.
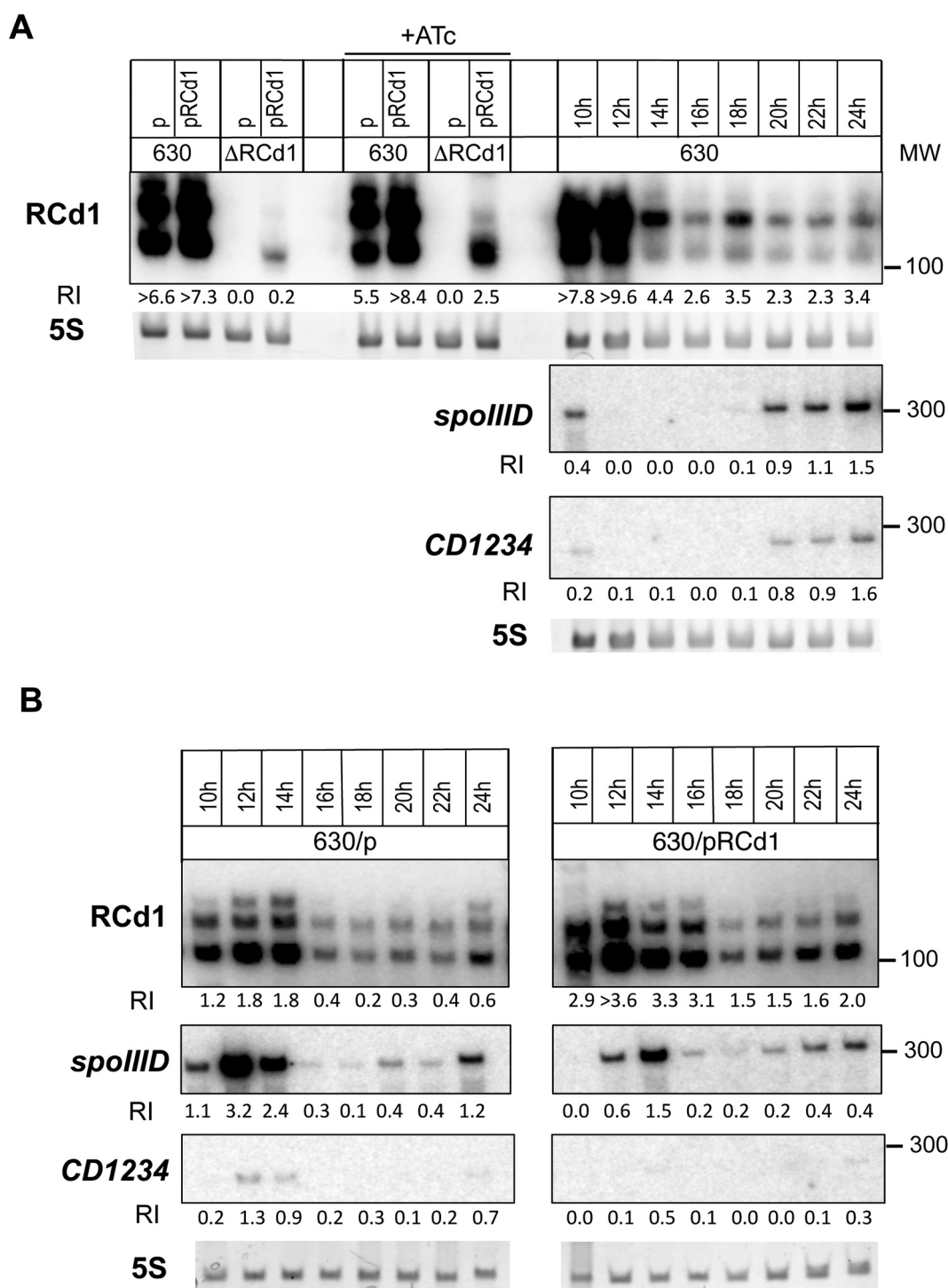
Figure 6. The expression profile of RCd1 and selected potential targets during sporulation. (A) RNA samples were extracted from 630Δerm strain (630), 630Δerm strain carrying an empty vector (630/p) or overexpressing RCd1 from plasmid (630/pRCd1) grown in TY medium supplemented on not with 250 ng/ml ATc (+ATc) or in SM medium for 10 h, 12 h, 14 h, 16 h, 18 h, 20 h, 22 h, 24 h as indicated. The strain deleted for RCd1 (ΔRCd1) was used as a control for specific RCd1 detection. 5S RNA at the bottom serves as loading control. As indicated at the left, the blots were hybridized either with RCd1, spoIIID or CD1234-specific probes. The same 5S control panel is shown when reprobing of the same membrane was performed. Relative intensity of the bands 'RI' normalized to 5S estimated with ImageJ software is indicated underneath each lane, '>' indicates the saturation of the signal. MW, the size of RNA molecular weight standards in nucleotides. (B) Effect of RCd1 overexpression on spoIIID and CD1234 genes during sporulation was monitored in SM medium during sporulation for 10 h, 12 h, 14 h, 16 h, 18 h, 20 h, 22 h, 24 h as indicated.

SpoIIID and is then a crucial target for the timely control of the *skin* excision in the mother cell only [33]. No significantly enriched peak could be detected by MACS for *CD1234* or *spoIIID* in our Hfq RIP-seq data. However, visual inspection of RIP-seq data for the *skin^Cd* region revealed a detectable

signal for *CD1234* gene in both Hfq-FLAG and Hfq-immunoprecipitated samples that was absent in the control samples. An enrichment for *spoIIID* at least in the Hfq-immunoprecipitated samples as compared to the controls was also observed (Supplementary Figure S9). RIP-seq

analysis also revealed peaks enriched in the Hfq-FLAG immunoprecipitated samples for the *CD1230, CD1231* and *CD1233* genes within the *skin* element region.

Altogether these results suggest that RCd1 could act on top of the regulatory cascade triggering the inhibition of the late sporulation stage setup by affecting first the SpoIIID regulator expression and then directly or indirectly the CD1234 component of the *skin^{Cd}* excision complex. The decrease in the *skin^{Cd}* excision rate would then negatively affect the native *sigK* gene recovery thus impacting the whole downstream SigK regulon.

Interactions of RCd1 with potential targets in vitro. Based on our results of the RCd1 overexpression effect on potential target abundance, we selected *spoIIID* and *CD1234* for further interactions analysis. An extended duplex formation between RCd1 and both *spoIIID* and *CD1234* mRNAs could be predicted *in silico* with a ΔG interaction energy estimated to −42.8 and −35.2 kcal/mol, respectively (Supplementary Figure S10). To investigate these interactions *in vitro* in the presence or absence of purified Hfq protein, we then setup an RNA band shift assay. To explore first the role of Hfq, we analysed the RNA-Hfq complex formation *in vitro*. The RNA band shift assays confirmed Hfq binding to RCd1 [27] and showed that Hfq could interact individually with each of the selected mRNAs *spoIIID* and *CD1234* with similar affinity (3.3 nM and 1.8 nM, respectively, Supplementary Figure S11). Incubations of sRNA with mRNA targets revealed shifted bands corresponding to the sRNA-mRNA complex. The addition of Hfq did not allow the appearance of a supershifted band corresponding to sRNA-mRNA-Hfq (Fig. 7). From these experiments, we could conclude that RCd1 may interact with rather low affinity with the selected mRNA targets under *in vitro* conditions and that Hfq is not required for these interactions *in vitro*. Future studies will define the molecular basis for the specific role of Hfq in sporulation control.

Genomic analysis of RCd1 conservation and its co-occurrence with potential sporulation-related targets. The analysis of RCd1 conservation revealed the presence of close homologues only within *C. difficile* species. From 2,700 available *C. difficile* genomes, the sequences with at least 80% of nucleotide identity to RCd1 from the laboratory strain 630Δ*erm* could be found in 99% of analysed *C. difficile* strains (Supplementary Table S9). Multiple alignments of nucleotide sequences of RCd1 homologues showed an extremely high level of conservation (Supplementary Figure S12). We also searched for the co-occurrence of RCd1 homologues with the important regulatory gene *spoIIID* and with the *skin^{Cd}* element including genes involved in the *skin^{Cd}* element excision, *CD1231* and *CD1234*. No clear correlation between the presence of RCd1 and the *skin* element could be observed (Supplementary Figure S13), the presence of a complete copy of the *skin* element being rather variable within *C. difficile* genomes. At an 80% nucleotide sequence identity threshold, *spoIIID* homologues were found in all analysed *C. difficile* genome sequences, while the *CD1231* and *CD1234* genes for the *skin* excision complex were found in 85% of *C. difficile* genomes (Supplementary Table S9). These results suggest that RCd1 could have additional conserved functions in *C. difficile* beyond the control of the *skin^{Cd}* excision.

Discussion

The RNA chaperone Hfq has emerged as an important player in the sRNA-based regulatory mechanisms in Gram-negative bacteria. However, its role remains less characterized in Gram-positive bacteria [71]. Previous studies revealed unique features of the Hfq protein in *C. difficile* suggesting a pleiotropic role in this pathogen unlike the majority of other studied Hfq homologues in Gram-positive bacteria [27]. Transcriptomic analysis revealed that 6% of the genes are affected by Hfq depletion [27]. The *C. difficile* Hfq protein was able to complement several phenotypes of the *hfq* deletion mutant in *E. coli* [28] as also demonstrated for the Hfq protein of *L. monocytogenes*. This latter bacterium is the only firmicute with at least one example of Hfq-dependent riboregulation [72,73]. In the present study, the co-immunoprecipitation analysis showed that Hfq has a great
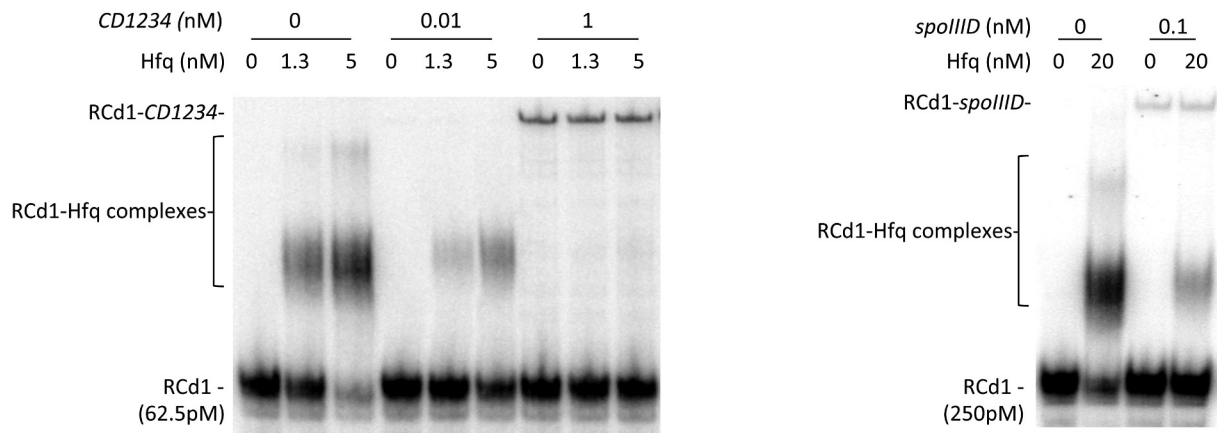

Figure 7. Analysis of the interaction between RCd1 and *spoIIID* or *CD1234* in the presence or the absence of Hfq. Radioactive RCd1 transcript was synthesized *in vitro* and incubated alone or together with *CD1234* or *spoIIID* transcripts with purified Hfq-His$_6$ protein expressed as monomer forms. Brackets show Hfq-RCd1 complexes, the position of unbound RNA and RCd1-potential target complexes is indicated at the left.

number of ribonucleic ligands in *C. difficile*. During Hfq depletion, the abundance of several identified Hfq-binding RNAs transcribed from IGR was shown to be modified, suggesting that Hfq could affect their stability and contribute to their function [27]. Compared to Hfq homologues from other Gram-positive bacteria, structural features of the *C. difficile* Hfq protein, such as the presence of most of the key residues within the RNA-binding sites including two of three conserved arginine residues in the rim region (the outer ring on the surface of Hfq toroid structure as one of four sites for RNA contacts) and the presence of an unusual C-terminal region exceptionally rich in asparagine residues, may explain these unique characteristics. Recent results suggest that this C-terminal part is not essential but contributes to the function of the RNA chaperone [28]. Extensive studies on Hfq-recognition motifs identified U-rich and A-rich or 'ARN' sequences that bind to the proximal and distal face of Hfq, respectively [74]. *In silico* search using MEME for potential RNA motifs recognized by *C. difficile* Hfq within the peaks significantly enriched in RIP-seq samples revealed a rather degenerated motif (Supplementary Figure S14). Presumably, the presence of conserved motifs within several classes of Hfq-associated RNAs and the AT-rich nature of the *C. difficile* genome could make it difficult to precisely define an Hfq-binding motif in interacting RNA species.

To associate the ncRNAs with their potential targets in Hfq-dependent regulations in *C. difficile* we performed an *in silico* target prediction for the RNAs specifically enriched in our RIP-seq analysis using the IntaRNA program [75] (Supplementary Table S10). A majority of Hfq-bound mRNAs was predicted as targets for at least one Hfq-bound ncRNA, suggesting joint ncRNA-mRNA binding to Hfq. To illustrate these results, pairing predictions of several Hfq-bound ncRNAs are shown in Supplementary Figure S15.

This ncRNA-mRNA interaction prediction analysis revealed very large potential targeting capacities of crRNAs, raising the possibility of their regulatory functions in *C. difficile* as suggested in other bacteria [58]. This includes the most abundant crRNAs from CRISPR 16/15 and CRISPR 3/4 arrays located in prophage regions of the strain 630 chromosome (Supplementary Figure S15). No continuous coverage of entire spacer/target was observed and no functional protospacer-adjacent motif (PAM) could be associated with these potential interactions suggesting that the potential regulatory functions could be mediated by different Cas-nuclease-independent mechanisms (Supplementary Figure S15).

ncRNA-ncRNA interactions have recently been identified as additional components of RNA-mediated regulatory networks [76]. Indeed, a ncRNA-sponge function has been demonstrated for several bacterial ncRNAs [77–79]. Thus, in addition to mRNA targets, ncRNAs must be included as potential targets of regulatory RNAs. We thus extended our IntaRNA analysis to ncRNA identified as enriched by RIP-seq (Supplementary Table S11). This analysis revealed more than 300 potential interactions between identified ncRNAs. Intriguingly, the strongest predicted interactions involved CRISPR RNA regions including crRNA spacers targeting riboswitches, IGR sRNAs or antisense RNAs (Supplementary

Figure S16). Obviously, these *in silico* predictions will require experimental validation but still suggest a complex regulatory RNA network in *C. difficile*.

Most of the genes belonging to the SigK regulon are overexpressed in the strain depleted for Hfq suggesting a link between this RNA chaperone and the sporulation control in *C. difficile* [27]. An over-sporulation phenotype is observed for this depleted strain and a Δ*hfq* mutant in *C. difficile* [70]. In line with these data, the entire SigK regulon, *sigK* and *spoIIID* genes are overexpressed in these conditions [27]. SpoIIID, which positively controls *sigK* expression, is located upstream of SigK in the sporulation regulation cascade. The *spoIIID* gene as well as the *sigK* gene are transcribed by the sigma factor SigE [80]. As no global effect of Hfq on the SigE regulon was observed [27], we can exclude SigE. Hfq rather controls the late stages of sporulation by acting either at the SpoIIID or SigK level. From *in silico* target predictions, previous phenotypic and interaction studies and present RIP-seq data, RCd1 appears among Hfq-binding ncRNAs as a good candidate to mediate this Hfq-dependent control. Indeed, RCd1 overexpression has a strong inhibitory effect on the sporulation process and on the SigK regulon expression.

As in *Bacillus subtilis*, the *sigK* gene is interrupted by a *skin*^{Cd} element in most of *C. difficile* strains. The excision of this phage-like element is necessary to trigger SigK production and is carried out by the recombinase CD1231 in complex with the recombination directionality factor, CD1234 [33]. *CD1231* and *CD1234* are the only two genes of the *skin*^{Cd} element conserved in all strains of *C. difficile*. We demonstrated in this work that the deletion of the *skin* element abolished the impact of RCd1 overexpression on sporulation and that RCd1 overexpression negatively affects the *skin* excision in the wild-type strain under sporulation conditions. These results strongly suggest that RCd1 acts at the level of the *skin* excision or upstream of this step. In the regulatory cascade of sporulation, we then propose to include an additional RNA regulatory element, RCd1, that can trigger *sigK* expression and SigK production by promoting the *skin* excision. The excision of the *skin* element in the mother cell occurs at a late stage of the sporulation process and requires the specific expression of *CD1234* under the dual control of SigE and SpoIIID [33]. We propose that RCd1 in concert with Hfq modulates the expression of *spoIIID* but also of *CD1234* either directly or indirectly through a negative control of *spoIIID* expression. An effect on SigE is excluded since neither Hfq-depletion nor RCd1 overexpression affected SigE targets except for *spoIIID* and *CD1234* genes. RCd1 overexpression limits the amount of CD1234 protein, an essential partner of the CD1231 recombinase leading to a decrease in the recovery of a functional *sigK* gene. A direct negative effect of RCd1 on *spoIIID* is also possible and would result, in addition to its effect on CD1234, in a decrease of *sigK* transcription affecting the entire SigK regulon and thus the sporulation efficiency. Our *in vitro* interaction studies showed that RCd1 could bind to *spoIIID* and *CD1234* mRNAs. We were able to show that Hfq could interact with each RNA individually but was not required for RCd1-mRNA interactions *in vitro*. Additional missing factors or particular conditions could affect these interactions during sporulation in *C. difficile*. Hfq could also

act by restructuring the target mRNAs that would lead to inhibition of their translation. An Hfq-independent action of RCd1 could also be envisaged. The signals or regulatory elements affecting the RCd1 action on the *sigK* network as well as the possible role of the 100-nt antisense RNA for RCd1 remain to be studied. A complex regulatory network governs SigK activation during the late stages of sporulation with SpoIIID- and SigK-dependent regulatory feedback loops and the participation of the SigE sigma factor. As for other ncRNAs, RCd1 could have a large spectrum of potential targets contributing to different facets of *C. difficile* physiology. In accordance, RIP-seq and *in silico* gene co-occurrence analyses suggest a larger role of RCd1 in regulatory processes beyond the *skin^{Cd}* excision control.

Several examples of bacterial RNA-based regulation of sporulation have recently been reported. In *Clostridium perfringens*, the VirX ncRNA has been shown to negatively control the sporulation process. However, unlike RCd1 inhibiting late stages of sporulation, VirX inhibits sporulation very early, likely at the Spo0A level. This ncRNA is conserved in several clostridial species but absent in *C. difficile* or *Clostridium tetani* [81]. In *B. subtilis*, several ncRNAs dependent on sporulation sigma factors have been identified, including SurA, SurC and CsfG, but their role has not yet been defined [82–84]. In *Bacillus anthracis*, one of these sporulation-specific sRNA, CsfG, conserved in *Bacillaceae* and some other endospore formers but absent in *C. difficile*, has been shown to bind Hfq1, one of the Hfq homologues present in *B. anthracis* [85] suggesting a possible implication of Hfq in regulatory processes during sporulation. In *B. subtilis*, sporulation was not altered in the Δ*hfq* strain [86], although CsfG was also identified among Hfq-associated sRNAs by a coimmunoprecipitation analysis [87]. A more specialized function connected with stationary phase physiology has recently been suggested for Hfq in *B. subtilis*. However, the Hfq-related stationary fitness phenotype could not be linked to sporulation [88]. As an example of an Hfq-independent RNA contributing to sporulation control, one of the two *B. subtilis* 6S RNAs (6 S-1 RNA) interacting with the housekeeping RNA polymerase in complex with SigA, was shown to be involved in the appropriate timing of the onset of endospore formation [89]. Overall, the diversity of RNA-based mechanisms controlling sporulation largely remains to be explored with specific actors that could be found in different endospore formers.

In addition to classical *trans*-encoded RNAs that were expected to act in concert with Hfq, our RIP-seq analysis revealed new RNA categories of Hfq partners. *Cis*-antisense RNAs were among the most abundant classes of regulatory RNAs enriched in Hfq-immunoprecipitated samples. This RIP-seq analysis allowed us to describe multiple new regions of double-directional transcription in *C. difficile* chromosome suggesting the involvement of Hfq in these complementary RNA interactions. Emerging examples suggest the requirement of Hfq or other RNA chaperones for antisense RNA pairing with their targets [90,91]. For example, Hfq facilitates the interaction of the antisense RNA RNA-OUT with its target RNA-IN in the Tn10/IS10 transposon system [91] and FinO is required for the FinP antisense RNA pairing with *traJ*

mRNA in the F plasmid of *E. coli* [90]. GadY antisense RNA has been shown to interact with Hfq in *E. coli* but the specific role of Hfq in this RNA-mediated regulation remains unclear [92]. In *Rhodobacter sphaeroides*, the Hfq protein binds to and destabilizes both an antisense RNA asPcrL and its target *puf*-mRNA [93].

The mechanisms of Hfq action in the regulations mediated by antisense RNAs remain to be explored. Hfq might restructure one of the interacting RNAs impeding thus its interaction with the partner. We could hypothesize that Hfq binding allows the recruitment of additional factors leading to RNA stabilization or destabilization. For example, in *E. coli*, Hfq was shown to promote the RNase E-mediated degradation of Hfq-bound sRNAs [94]. Furthermore, RNase III being a double-strand specific RNase, the recruitment of Hfq could interfere with RNase III-mediated degradation of RNAs [93,95]. In some cases, the action of antisense RNAs on *trans*-encoded targets could potentially require Hfq for efficient base-pairing [96]. Even in the case of an extended complementarity for antisense RNA base-pairing with their targets *in cis*, the RNA–RNA interaction could be initiated within a short *seed* region as for Hfq-dependent *trans*-encoded sRNAs [97]. A short loop–loop interaction within a so-called *kissing* complex is generally required to initiate the duplex formation of antisense RNAs with their targets [98].

The first functional antisense RNAs described in *C. difficile* act as antitoxins within type I TA systems [18–20]. In the present study, both toxin mRNAs and antitoxin antisense RNAs for all 13 type I TA modules were highly enriched in Hfq-immunoprecipitated samples. The possible role of Hfq in stabilizing type I TA interactions had already been suggested in our previously reported work in *C. difficile* [18]. Similarly, antitoxins and toxins from type I TA modules were enriched in Hfq coIP samples in *B. subtilis* and Hfq was able to bind antitoxins *in vitro* [87]. While Hfq stabilized the SR5 antitoxin in this bacterium, it did not affect the *bsrE* toxin RNA of the type I TA module [99] and was generally not required for the function of type I TA systems [86,100]. Further studies will clarify if Hfq could in some way contribute to the function of TA in *C. difficile*. Interestingly, a new role for *B. subtilis* CsrA protein promoting the base-pairing interactions between the sRNA SR1 and its target *ahrC* mRNA has been demonstrated in a recent work [101]. These findings suggest that at least in some cases other RNA-binding proteins could replace well-known RNA chaperones for sRNA-mediated regulations.

In light of our RIP-seq analysis, a potential role of Hfq in uncommon regulatory functions associated with crRNAs from CRISPR arrays and with prematurely terminated riboswitch transcripts could be suggested. Several examples of a noncanonical role of such ncRNAs have been reported in other bacteria [102]. However, the implication of Hfq protein has never been explored in these regulatory processes. Indeed, a SAM-riboswitch-associated prematurely terminated transcript has been shown to act *in trans* as an independent regulatory RNA controlling the expression of the *prfA* gene encoding the major virulence regulator in *L. monocytogenes* [102]. The detection of multiple riboswitches from different functional groups as Hfq-binding RNAs in the present study and their large targeting potential

leads to a tempting speculation about their regulatory actions through base-pairing that need to be further explored. Similarly, a large class of regulatory RNA elements located within 5′-leader regions has been revealed among Hfq-associated RNAs in *B. subtilis* [87]. These *cis*-acting regulatory RNAs respond to a variety of ligands including metabolites, tRNAs and metal ions generally through the mechanism of premature termination of transcription in many Gram-positive bacteria. Their association with Hfq opens an interesting question about the possible sRNA-like action of these abundant terminated transcripts.

The targeting of endogenous RNAs has recently been demonstrated for the Cas9 protein of *Campylobacter jejuni* guided by native crRNA through imperfect complementarity [58]. Other examples of CRISPR-Cas regulatory function include the down-regulation of expression of the immunostimulatory BLP lipoprotein in *Francisella novicida* by type II-B Cas9 through tracrRNA in association with another small CRISPR-Cas-associated RNA (scaRNA) by an unknown mechanism [103–105]. The great potential of *C. difficile* crRNA targeting revealed in our work is in line with these findings, further suggesting a potential regulatory role for CRISPR-Cas systems in bacterial pathogens. Indeed, potential functions beyond defence and in relation with virulence control in several pathogens emerged from several recent studies [59,106–110]. However, the exact mechanism implicating crRNA targeting for regulatory purposes remains the subject of debate. For example, in contrast to the previously reported regulatory potential for type I-F CRISPR-Cas in *Pseudomonas aeruginosa* [106], more recent studies did not find evidence of RNA targeting [111,112]. The possible implication of the RNA chaperone Hfq in crRNA interactions with potential mRNA targets implies that crRNAs could be additional components in the regulatory RNA networks and suggests an alternative mechanism for crRNA regulatory actions independent from the Cas machinery. Discontinuous base-pairing of spacer-containing regions with potential mRNA targets and the absence of functional PAM in the proximity of spacer-targeted zone are further in favour of the particular mechanism for potential crRNA regulatory functions reminiscent to the *trans*-encoded sRNA-mediated action.

In Gram-negative bacteria, the Hfq protein generally binds to AU-rich single-stranded regions upstream of the rho-independent terminator in sRNAs and could also interact with the 3′ poly(U) tail [74]. The AT-rich nature of the *C. difficile* genome could provide a large high-affinity interaction potential for efficient Hfq protein association with ncRNAs. Widespread post-transcriptional mechanisms of Hfq-dependent ncRNA action are linked to the mRNA RBS binding leading to the repression of translational initiation. Our RIP-seq analysis combined with *in silico* base-pairing prediction for Hfq-associated RNAs revealed a number of potential ncRNA-mRNA interactions that could fit this mechanism. Interestingly, potentially accessible short C-rich motifs could be involved in these interactions in *C. difficile*. In *Staphylococcus aureus*, another Gram-positive pathogen with a low GC-genome content, a class of ncRNAs has been identified that carries a conserved unpaired C-rich motif for targeting mRNA RBS regions and that could

repress the translation by preventing the formation of a ribosomal initiation complex [113–115]. Even though Hfq was not shown to be involved in these regulatory mechanisms in *S. aureus*, the presence of C-rich motifs could constitute a general hallmark for efficient mRNA targeting by ncRNAs.

Conclusion

The present study provides the first genomic map of RNAs interacting with Hfq in *C. difficile* and paves the way for future detailed analysis of RNA-based mechanisms in this human enteropathogen. The RIP-seq data analysis expands our knowledge on the great interaction potential of the RNA chaperone Hfq in a Gram-positive bacterium opening interesting perspectives for future studies. This analysis leads to the discovery of a number of new antisense RNAs and IGR ncRNAs further enriching the repertoire of regulatory RNAs in *C. difficile*. Together with the previous genome-wide identification of ncRNAs by TSS mapping, this global approach further confirms the importance of RNA-based mechanisms in *C. difficile*. The discovery of regulatory RNA functions still stays in its early stage in *C. difficile*. The identification of specific roles of Hfq in RNA-based regulations in *C. difficile* certainly requires further investigations. The present work provides an essential molecular basis on the repertoire of Hfq-binding RNAs and their interaction potential that will foster future studies on the post-transcriptional regulatory mechanisms to better understand the emergence and success of this important human enteropathogen.

Accession numbers

RIP-Seq data have been deposited in ENA with accession no. PRJEB39335.

ORCID

Johann Peltier http://orcid.org/0000-0003-3207-9465
Eliane Hajnsdorf http://orcid.org/0000-0002-8249-2007
Marc Monot http://orcid.org/0000-0003-0738-7335
Bruno Dupuy http://orcid.org/0000-0001-7176-8453
Isabelle Martin-Verstraete http://orcid.org/0000-0002-8376-7690
Daniel Gautheret http://orcid.org/0000-0002-1508-8469
Claire Toffano-Nioche http://orcid.org/0000-0003-4134-6844
Olga Soutourina http://orcid.org/0000-0001-6439-7228

References

[1] Carroll KC, Bartlett JG. Biology of *Clostridium difficile*: implications for epidemiology and diagnosis. Annu Rev Microbiol. 2011;65(1):501–521.

[2] Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nat Rev Microbiol. 2009;7(7):526–536.

[3] Seekatz AM, Young VB. *Clostridium difficile* and the microbiota. J Clin Invest. 2014;124(10):4182–4189.

[4] Janoir C. Virulence factors of *Clostridium difficile* and their role during infection. Anaerobe. 2016;37:13–24.

[5] Just I, Selzer J, Wilm M, et al. Glucosylation of Rho proteins by *Clostridium difficile* toxin B. Nature. 1995;375(6531):500–503.

[6] Smits WK, Lyras D, Lacy DB, et al. *Clostridium difficile* infection. Nat Rev Dis Primers. 2016;2:16020.

[7] Vedantam G, Clark A, Chu M, et al. *Clostridium difficile* infection: toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. Gut Microbes. 2012;3(2):121–134.

[8] Deakin LJ, Clare S, Fagan RP, et al. The *Clostridium difficile spo0A* gene is a persistence and transmission factor. Infect Immun. 2012;80(8):2704–2711.

[9] Chakravarty S, Masse E. RNA-dependent regulation of virulence in pathogenic bacteria. Front Cell Infect Microbiol. 2019;9:337.

[10] Heroven AK, Nuss AM, Dersch P. RNA-based mechanisms of virulence control in Enterobacteriaceae. RNA Biol. 2017;14(5):471–487.

[11] Quereda JJ, Cossart P. Regulating bacterial virulence with RNA. Annu Rev Microbiol. 2017;71(1):263–280.

[12] Svensson SL, Sharma CM. Small RNAs in bacterial virulence and communication. Microbiol Spectr. 2016;4(3).

[13] Westermann AJ. Regulatory RNAs in virulence and host-microbe interactions. Microbiol Spectr. 2018;6(4).

[14] Soutourina O. RNA-based control mechanisms of *Clostridium difficile*. Curr Opin Microbiol. 2017;36:62–68.

[15] Soutourina OA, Monot M, Boudry P, et al. Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. PLoS Genetics. 2013;9(5):e1003493. .

[16] Boudry P, Semenova E, Monot M, et al. Function of the CRISPR-cas system of the human pathogen *Clostridium difficile*. MBio. 2015;6:e01112–01115.

[17] Maikova A, Severinov K, Soutourina O. New insights into functions and possible applications of *Clostridium difficile* CRISPR-Cas system. Front Microbiol. 2018;9:1740.

[18] Maikova A, Peltier J, Boudry P, et al. Discovery of new type I toxin-antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. Nucleic Acids Research. 2018;46(9):4733–4751.

[19] Peltier J, Hamiot A, Garneau J, et al. Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in *Clostridioides difficile*. Commun Biol. 2020;3(1):718.

[20] Soutourina O. Type I toxin-antitoxin systems in lostridia. Toxins (Basel). 2019;11(5):253.

[21] Kavita K, de Mets F, Gottesman S. New aspects of RNA-based regulation by Hfq and its partner sRNAs. Curr Opin Microbiol. 2018;42:53–61.

[22] Santiago-Frangos A, Woodson SA. Hfq chaperone brings speed dating to bacterial sRNA. Wiley Interdiscip Rev RNA. 2018;9(4):e1475.

[23] Updegrove TB, Zhang A, Storz G. Hfq: the flexible RNA matchmaker. Curr Opin Microbiol. 2016;30:133–138.

[24] Hajnsdorf E, Boni IV. Multiple activities of RNA-binding proteins S1 and Hfq. Biochimie. 2012;94(7):1544–1553.

[25] Melamed S, Adams PP, Zhang A, et al. RNA-RNA interactomes of proQ and Hfq reveal overlapping and competing roles. Mol Cell. 2020;77(411–425):e417.

[26] Olejniczak M, Storz G. ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers? Mol Microbiol. 2017;104(6):905–915.

[27] Boudry P, Gracia C, Monot M, et al. Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen *Clostridium difficile*. Journal of Bacteriology. 2014;196(18):3234–3248.

[28] Caillet J, Gracia C, Fontaine F, et al. *Clostridium difficile* Hfq can replace *Escherichia coli* Hfq for most of its function. RNA. 2014;20(10):1567–1578.

[29] Dupuy B, Sonenshein AL. Regulated transcription of *Clostridium difficile* toxin genes. Mol Microbiol. 1998;27(1):107–120.

[30] Bertani G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. Journal of Bacteriology. 1951;62:293–300.

[31] Fagan RP, Fairweather NF. *Clostridium difficile* has two parallel and essential Sec secretion systems. J Biol Chem. 2011;286(31):27483–27493.

[32] Wilson KH, Silva J, Fekety FR. Fluorescent-antibody test for detection of *Clostridium difficile* in stool specimens. J Clin Microbiol. 1982;16(3):464–468.

[33] Serrano M, Kint N, Pereira FC, et al. A recombination directionality factor controls the cell type-specific activation of sigmaK and the fidelity of spore development in *Clostridium difficile*. PLoS Genetics. 2016;12(9):e1006312.

[34] Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual, second edition. In: Cold spring harbor laboratory Press. N. Y: Cold Spring Harbor; 1989.

[35] Cartman ST, Kelly ML, Heeg D, et al. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdc genotype and toxin production. Appl Environ Microbiol. 2012;78(13):4683–4690.

[36] Peltier J, Soutourina O. Identification of c-di-GMP-responsive riboswitches. Methods Mol Biol. 2017;1657:377–402.

[37] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta c(t)) method. Methods. 2001;25(4):402–408.

[38] Sahr T, Buchrieser C. Co-immunoprecipitation: protein-RNA and protein-DNA interaction. Methods Mol Biol. 2013;954:583–593.

[39] Folichon M, Allemand F, Regnier P, et al. Stimulation of poly(A) synthesis by *Escherichia coli* poly(A)polymerase I is correlated with Hfq binding to poly(A) tails. Febs J. 2005;272(2):454–463.

[40] Criscuolo A, Brisse S. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. Genomics. 2013;102(5–6):500–506.

[41] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359.

[42] Monot M, Orgeur M, Camiade E, et al. COV2HTML: a visualization and analysis tool of bacterial next generation sequencing (NGS) data for postgenomics life scientists. OMICS. 2014;18(3):184–195.

[43] Liu T. Use model-based analysis of ChIP-seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods Mol Biol. 2014;1150:81–95.

[44] Vallenet D, Labarre L, Rouy Z, et al. MaGe: a microbial genome annotation system supported by synteny results. Nucleic Acids Res. 2006;34(1):53–65.

[45] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842.

[46] Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–462.

[47] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

[48] Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server):W202–208.

[49] Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(2078):–2079.

[50] Mann M, Wright PR, Backofen R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. Nucleic Acids Res. 2017;45(W1):W435–W439.

[51] Eggenhofer F, Tafer H, Stadler PF, et al. RNApredator: fast accessibility-based prediction of sRNA targets. Nucleic Acids Res. 2011;39(suppl_2):W149–154.

[52] Lorenz R, Bernhart SH, Siederdissen HZ, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6(1):26.

[53] Heidrich N, Bauriedl S, Barquist L, et al. The primary transcriptome of Neisseria meningitidis and its interaction with the RNA chaperone Hfq. Nucleic Acids Res. 2017;45(10):6147–6167.

[54] Sittka A, Lucchini S, Papenfort K, et al. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet. 2008;4(8):e1000163.

[55] Assis NG, Ribeiro RA, da Silva LG, et al. Identification of Hfq-binding RNAs in Caulobacter crescentus. RNA Biology. 2019;16(6):719–726.

[56] Feng J, Liu T, Qin B, et al. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012;7(9):1728–1740.

[57] Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–26.

[58] Dugar G, Leenay RT, Eisenbart SK, et al. CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the Campylobacter jejuni Cas9. Mol Cell. 2018;69(893–905):e897.

[59] Westra ER, Buckling A, Fineran PC. CRISPR-Cas systems: beyond adaptive immunity. Nat Rev Microbiol. 2014;12:317–326.

[60] Rosinski-Chupin I, Soutourina O, Martin-Verstraete I. Riboswitch discovery by combining RNA-seq and genome-wide identification of transcriptional start sites. Methods Enzymol. 2014;549:3–27.

[61] Bordeleau E, Fortier LC, Malouin F, et al. c-di-GMP turn-over in Clostridium difficile is controlled by a plethora of diguanylate cyclases and phosphodiesterases. PLoS Genet. 2011;7(3):e1002039.

[62] McKee RW, Harvest CK, Tamayo R. Cyclic diguanylate regulates virulence factor genes via multiple riboswitches in Clostridium difficile. mSphere. 2018;3.

[63] Purcell EB, McKee RW, McBride SM, et al. Cyclic diguanylate inversely regulates motility and aggregation in Clostridium difficile. Journal of Bacteriology. 2012;194(13):3307–3316.

[64] Sudarsan N, Lee ER, Weinberg Z, et al. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science. 2008;321(5887):411–413.

[65] Soutourina O, Dubois T, Monot M, et al. Genome-wide transcription start site mapping and promoter assignments to a sigma factor in the human enteropathogen Clostridioides difficile. Front Microbiol. 2020;11:1939.

[66] McBride SM, Sonenshein AL. The dlt operon confers resistance to cationic antimicrobial peptides in Clostridium difficile. Microbiology (Reading). 2011;157(5):1457–1465.

[67] Woods EC, Nawrocki KL, Suarez JM, et al. The Clostridium difficile dlt pathway is controlled by the extracytoplasmic function sigma factor σ v in response to lysozyme. Infect Immun. 2016;84(6):1902–1916.

[68] Morita T, Aiba H. Mechanism and physiological significance of autoregulation of the Escherichia coli hfq gene. RNA. 2019;25(2):264–276.

[69] Oliva G, Sahr T, Rolando M, et al. A unique cis-encoded small noncoding RNA is regulating Legionella pneumophila Hfq expression in a life cycle-dependent manner. mBio. 2017;8(1):e02182-16.

[70] Maikova A, Kreis V, Boutserin A, et al. Using an endogenous CRISPR-cas system for genome editing in the human pathogen Clostridium difficile. Applied and Environmental Microbiology. 2019;85(20):e01416-19.

[71] Sobrero P, Valverde C. The bacterial protein Hfq: much more than a mere RNA-binding factor. Crit Rev Microbiol. 2012;38(4):276–299.

[72] Nielsen JS, Larsen MH, Lillebaek EM, et al. A small RNA controls expression of the chitinase ChiA in Listeria monocytogenes. PloS One. 2011;6(4):e19019.

[73] Nielsen JS, Lei LK, Ebersbach T, et al. Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in Listeria monocytogenes. Nucleic Acids Research. 2010;38:907–919.

[74] Vogel J, Luisi BF. Hfq and its constellation of RNA. Nat Rev Microbiol. 2011;9(8):578–589.

[75] Busch A, Richter AS, Backofen R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics. 2008;24(24):2849–2856.

[76] Carrier MC, Lalaouna D, Masse E. Broadening the definition of bacterial small RNAs: characteristics and mechanisms of action. Annu Rev Microbiol. 2018;72(1):141–161.

[77] Figueroa-Bossi N, Bossi L. Sponges and predators in the small RNA world. Microbiol Spectr. 2018;6(4):6.

[78] Lalaouna D, Carrier MC, Semsey S, et al. A 3′ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. Mol Cell. 2015;58(3):393–405.

[79] Miyakoshi M, Chao Y, Vogel J. Cross talk between ABC transporter mRNA s via a target mRNA-derived sponge of the Gcv B small RNA. Embo J. 2015;34(11):1478–1492.

[80] Saujet L, Pereira FC, Serrano M, et al. Genome-wide analysis of cell type-specific gene transcription during spore formation in Clostridium difficile. PLoS Genetics. 2013;9(10):e1003756.

[81] Ohtani K, Hirakawa H, Paredes-Sabja D, et al. Unique regulatory mechanism of sporulation and enterotoxin production in Clostridium perfringens. Journal of Bacteriology. 2013;195(12):2931–2936.

[82] Marchais A, Duperrier S, Durand S, et al. CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers. RNA Biol. 2011;8(3):358–364.

[83] Schmalisch M, Maiques E, Nikolov L, et al. Small genes under sporulation control in the Bacillus subtilis genome. J Bacteriol. 2010;192(20):5402–5412.

[84] Silvaggi JM, Perkins JB, Losick R. Genes for small, noncoding RNAs under sporulation control in Bacillus subtilis. Journal of Bacteriology. 2006;188(2):532–541.

[85] Panda G, Tanwer P, Ansari S, et al. Regulation and RNA-binding properties of Hfq-like RNA chaperones in Bacillus anthracis. Biochim Biophys Acta. 2015;1850(9):1661–1668.

[86] Silvaggi JM, Perkins JB, Losick R. Small untranslated RNA antitoxin in Bacillus subtilis. Journal of Bacteriology. 2005;187(19):6641–6650.

[87] Dambach M, Irnov I, Winkler WC. Association of RNAs with Bacillus subtilis Hfq. PLoS One. 2013;8(2):e55156.

[88] Rochat T, Delumeau O, Figueroa-Bossi N, et al. Tracking the elusive function of Bacillus subtilis Hfq. PLoS One. 2015;10(4):e0124977.

[89] Cavanagh AT, Wassarman KM. 6S-1 RNA function leads to a delay in sporulation in Bacillus subtilis. Journal of Bacteriology. 2013;195:2079–2086.

[90] Mark Glover JN, Chaulk SG, Edwards RA, et al. The FinO family of bacterial RNA chaperones. Plasmid. 2015;78:79–87.

[91] Ross JA, Ellis MJ, Hossain S, et al. Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. RNA. 2013;19(5):670–684.

[92] Opdyke JA, Kang JG, Storz G. GadY, a small-RNA regulator of acid response genes in Escherichia coli. Journal of Bacteriology. 2004;186:6698–6705.

[93] Reuscher CM, Klug G. Antisense RNA asPcrL regulates expression of photosynthesis genes in Rhodobacter sphaeroides by promoting RNase III-dependent turn-over of puf mRNA. RNA Biol. 2020.

[94] De Lay N, Schu DJ, Gottesman S. Bacterial small RNA-based negative regulation: hfq and its accomplices. J Biol Chem. 2013;288(12):7996–8003.

[95] Court DL, Gan J, Liang YH, et al. RNase III: genetics and function; structure and mechanism. Annu Rev Genet. 2013;47(1):405–431.

[96] Pichon C, Du Merle L, Lequeutre I, et al. The AfaR small RNA controls expression of the AfaD-VIII invasin in pathogenic *Escherichia coli* strains. Nucleic Acids Res. 2013;41(10):5469–5482.

[97] Gorski SA, Vogel J, Doudna JA. RNA-based recognition and targeting: sowing the seeds of specificity. Nat Rev Mol Cell Biol. 2017;18(4):215–228.

[98] Wagner EG, Blomberg P, Nordstrom K. Replication control in plasmid R1: duplex formation between the antisense RNA, CopA, and its target, CopT, is not required for inhibition of RepA synthesis. Embo J. 1992;11(3):1195–1203.

[99] Muller P, Jahn N, Ring C, et al. A multistress responsive type I toxin-antitoxin system: bsrE/SR5 from the *B.subtilis* chromosome. RNA Biol. 2016;13(5):511–523.

[100] Jahn N, Preis H, Wiedemann C, et al. BsrG/SR4 from *Bacillus subtilis* the first temperature-dependent type I toxin-antitoxin system. Mol Microbiol. 2012;83(3):579–598.

[101] Muller P, Gimpel M, Wildenhain T, et al. A new role for CsrA: promotion of complex formation between an sRNA and its mRNA target in *Bacillus subtilis*. RNA Biology. 2019;16(7):972–987.

[102] Loh E, Dussurget O, Gripenland J, et al. A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. Cell. 2009;139:770–779.

[103] Sampson TR, Napier BA, Schroeder MR, *et al*. A CRISPR-Cas system enhances envelope integrity mediating antibiotic resistance and inflammasome evasion. Proc Natl Acad Sci U S A. 2014;111(30):11163–11168. .

[104] Sampson TR, Saroj SD, Llewellyn AC, et al. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature. 2013;497(7448):254–257.

[105] Sampson TR, Saroj SD, Llewellyn AC, et al. Author Correction: a CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature. 2019;570(7760):E30–E31.

[106] Li R, Fang L, Tan S, et al. Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. Cell Res. 2016;26(12):1273–1287.

[107] Muller-Esparza H, Randau L. Commentary: type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. Front Microbiol. 2017;8:319.

[108] Sampson TR, Weiss DS. Alternative roles for CRISPR/Cas systems in bacterial pathogenesis. PLoS Pathog. 2013;9(10):e1003621.

[109] Sampson TR, Weiss DS. CRISPR-Cas systems: new players in gene regulation and bacterial physiology. Front Cell Infect Microbiol. 2014;4:37.

[110] Strutt SC, Torrez RM, Kaya E, et al. RNA-dependent RNA targeting by CRISPR-Cas9. Elife. 2018;7(7). DOI:10.7554/eLife.32724

[111] Buyukyoruk M, Wiedenheft B. Type I-F CRISPR-Cas provides protection from DNA, but not RNA phages. Cell Discov. 2019;5(1):54.

[112] Hoyland-Kroghsbo NM, Munoz KA, Bassler BL. Temperature, by controlling growth rate, regulates CRISPR-Cas activity in *Pseudomonas aeruginosa*. mBio. 2018;9(6):e02184-18.

[113] Geissmann T, Chevalier C, Cros MJ, *et al*. A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. Nucleic Acids Res. 2009;37(21):7239–7257.

[114] Rochat T, Bohn C, Morvan C, *et al*. The conserved regulatory RNA RsaE down-regulates the arginine degradation pathway in *Staphylococcus aureus*. Nucleic Acids Research. 2018;46(17):8803–8816.

[115] Romilly C, Lays C, Tomasini A, et al. A non-coding RNA promotes bacterial persistence and decreases virulence by regulating a regulator in *Staphylococcus aureus*. PLoS Pathog. 2014;10(3):e1003979.