An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium

Manuela Fuchs1,2, Vanessa Lamm-Schmidtb,d, Johannes Sulzerd, Falk Ponathb, Laura Jennichesb, Joseph A. Kirkc, Robert P. Faganc, Lars Barquistbd, Jörg Vogelhb, and Franziska Fabera,2

1Institute for Molecular Infection Biology, Faculty of Medicine, University of Würzburg, 97080 Würzburg, Germany; 2Helmholtz Institute for RNA-based Infection Research, Helmholtz Centre for Infection Research, 97080 Würzburg, Germany; 3The Florey Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, United Kingdom; and 4Faculty of Medicine, University of Würzburg, 97080 Würzburg, Germany

Edited by Gisela Storz, National Institute of Child Health and Human Development, Bethesda, MD, and approved May 6, 2021 (received for review February 24, 2021)

The gram-positive human pathogen *Clostridioides difficile* has emerged as the leading cause of antibiotic-associated diarrhea. However, little is known about the bacterium’s transcriptome architecture and mechanisms of posttranscriptional control. Here, we have applied transcription start site and termination mapping to generate a single-nucleotide-resolution RNA map of *C. difficile* 5′ and 3′ untranslated regions, operon structures, and noncoding regulators, including 42 sRNAs. Our results indicate functionality of many conserved riboswitches and predict cis-regulatory RNA elements upstream of multidrug resistance (MDR)-type ATP-binding cassette (ABC) transporters and transcriptional regulators. Despite growing evidence for a role of Hfq in RNA-based gene regulation in *C. difficile*, the functions of Hfq-based posttranscriptional regulatory networks in gram-positive pathogens remain controversial. Using Hfq immunoprecipitation followed by sequencing of bound RNA species (RIP-seq), we identify a large cohort of transcripts bound by Hfq and show that absence of Hfq affects transcript stabilities and steady-state levels. We demonstrate sRNA expression during intestinal colonization by *C. difficile* and identify infection-related signals impacting its expression. As a proof of concept, we show that the utilization of the abundant intestinal metabolite ethanolamine is regulated by the Hfq-dependent sRNA CDIF630nc_085. Overall, our study lays the foundation for understanding clostridial riboregulation with implications for the infection process and provides evidence for a global role of Hfq in posttranscriptional regulation in a gram-positive bacterium.

Antibiotic-resistant bacteria are a major global threat to human health, endangering our ability to perform a range of modern medical interventions. The obligate anaerobe, spore-forming *Clostridioides difficile* has become the leading cause of antibiotic-associated diarrhea over the past two decades (1). Increasing numbers of multiresistant clinical isolates (2) and recurrent infections are key challenges in the treatment of *C. difficile* infections (CDI), often leaving fecal microbiota transfer as the only clinical option (3, 4).

The clinical challenges posed by CDI have prompted much effort to understand how this pathogen regulates virulence in response to environmental conditions. As a result, there is a comprehensive body of literature focusing on toxin production and sporeulation control by several global metabolic regulators including CcpA, CodY, Rex, and PrdR (5, 6). Furthermore, several specialized and general sigma factors including TcdR (6), SpotA (7), SigD (8, 9), SigH (10), and SigB (11–13) are linked to virulence and metabolism, although the exact molecular mechanisms often remain unknown. Most of this knowledge has been accumulated through detailed studies of individual genes and promoters, whereas RNA-sequencing (RNA-seq)–based annotations of the global transcriptome architecture, which have accelerated research in the gram-positive pathogens *Listeria* (14), *Staphylococcus* (15), and *Streptococcus* (16), have only recently become available for *C. difficile* (17).

This paucity of global knowledge about RNA output in *C. difficile* readily extends to posttranscriptional control of gene expression. The bacterium is of particular scientific interest, being the only gram-positive species thus far in which deletion of *hfq* seems to have a large impact on gene expression and bacterial physiology (18–20). Specifically, deletion of *hfq* increases sporeulation (21), a crucial pathogenic feature of this bacterium that enables transmission between hosts. In gram-negative bacteria, Hfq commonly exerts global posttranscriptional control by facilitating short base-pairing interactions of small regulatory RNAs (sRNAs) with trans-encoded messenger RNAs (mRNAs) (22, 23), but its role in gram-positive bacteria remains controversial. For example, functional studies on Hfq in *Bacillus subtilis*, a model bacterium for Firmicutes, revealed in vivo association with a subset of sRNA (24), but comparative analyses of a wild-type and *hfq* knockout strain revealed only moderate effects on sRNA and mRNA transcript levels (25) and the absence of any significant growth defect (26), which lead to the conclusion that Hfq plays a minor role in posttranscriptional regulation in *B. subtilis*.

**Microbiology**

**Significance**

*Clostridioides difficile* is the leading cause of healthcare-associated diarrhea worldwide following antibiotic treatment. Consequently, there is medical need for novel antibacterial agents acting against *C. difficile* that leave the resident microbiota unharmed. The development of such narrow-spectrum antibiotics requires precise knowledge of the mechanisms that fine-tune gene expression to orchestrate the genomic output at each locus in the genome. We address this issue by defining the global transcriptome architecture of *C. difficile* including noncoding regulatory elements, many of which are expressed during gut colonization. Our analysis of these regulators provides evidence for the global function of Hfq in sRNA binding and stabilization in a gram-positive bacterium.

Author contributions: M.F., V.L.-S., L.B., J.V., and F.F. designed research; M.F., V.L.-S., and J.S. performed research; J.A.K. and R.P.F. contributed new reagents/analytic tools; M.F., V.L.-S., F.P., L.B., and F.F. analyzed data; and M.F., V.L.-S., and F.F. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1M.F. and V.L.-S. contributed equally to this work.


to whom correspondence may be addressed. Email: franziska.faber@uni-wuerzburg.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2103579118/-/DCSupplemental.

Published June 15, 2021.
Previous efforts using in silico methods (27) and RNA-seq (28) have predicted >100 sRNA candidates in *C. difficile* and identified widespread RNA binding by Hfq (20), which suggests the existence of a large posttranscriptional network in *C. difficile*. However, under which conditions these sRNAs are expressed, which targets they regulate, and whether they depend on Hfq remain fundamental open questions.

Another important feature of posttranscriptional control in *C. difficile* are cis-regulatory RNA elements. Pioneering work has deciphered the function of genetic switches in the 5′ untranslated region (UTR) of the fglB operon, containing the early-stage flagellar genes, and the cell wall protein-encoding gene *cwpV* (29–31). Moreover, cyclic di-GMP–responsive riboswitches were shown to regulate biofilm formation and toxin production (9, 32–34). However, other members of the many different riboswitch classes have not been systematically identified. Combined with the nascent stage of sRNA biology in *C. difficile*, this argues that global approaches are needed to understand the full scope of posttranscriptional regulation in this important human pathogen.

In the present study, we applied recently developed methods of bacterial RNA biology (35, 36) to construct a global atlas of transcriptional and posttranscriptional control in *C. difficile*. We provide an example for a gram-positive bacterium with Hfq-dependent sRNA-based gene regulation and identify the ethanolamine utilization pathway as a target of sRNA-mediated regulation in *C. difficile*. Overall, we provide evidence for extensive Hfq-dependent posttranscriptional regulation and lay the foundation for future mechanistic studies of RNA-based gene regulation in *C. difficile*.

Results

High-Resolution Transcriptome Maps of *C. difficile* 630. For generating high-resolution transcriptome maps of *C. difficile*, we chose the reference strain 630 (DSM 27543, CP010905.2). Being widely used by the *C. difficile* community, this strain offers the most comprehensive genome annotation. Using two global RNA-seq approaches, we analyzed RNA samples from three different conditions: late-exponential and early stationary growth in tryptone-yeast (TY) broth and late-exponential growth in brain heart infusion (BHI) broth (Fig. 1A). The resulting genome-wide maps provide single-nucleotide–resolution transcriptional start sites (TSSs) and termination sites (TTSs) (Fig. 1B). We have used this information to annotate 5′ and 3′ UTRs and operon structures, to correct previous open reading frame (ORF) annotations, to add previously overlooked small genes, and to annotate sRNA loci. Inspired by other gene expression databases such as SalCom (37), Acinet-TOcom (38), and Theta-Base (39), we have launched the interactive web browser “Clost-Base” (https://www.helmholtz-hire.de/en/datasets/clostridium) for our transcriptome data. This online resource for *C. difficile* allows visualization of the transcriptomic data in the context of annotated coding and noncoding genes as well as transcript features (e.g., TSS) that we have experimentally determined in this work. The browser enables search queries and retrieval of primary sequences for any annotated feature in the database to serve as a resource for the *C. difficile* community.

Genome-wide Annotation of TSSs. Differential RNA-seq (dRNA-seq) (40, 41) was performed to capture 5′ ends of transcripts. In brief, one-half of the sample remains untreated (Tex−) to capture both, primary (5′ -PPP) and processed (5′-P or 5′OH) 5′ ends of transcripts; the other half is treated with terminator exonuclease (Tex+), leading to specific degradation of processed (5′-P or 5′OH) RNAs, thus enriching primary transcripts and enabling TSS annotation (Fig. 1B). Conversely, relative read enrichment in the Tex− complementary DNA (cDNA) libraries indicate RNA processing sites (Fig. 1B). The *C. difficile* 630 genome comprises annotations for 3,778 coding sequences (CDS) as well as transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and the housekeeping RNAs 6S RNA, RNaseP, tmRNA, and SRP. By identifying 2,293 TSSs, we were able to define transcriptional units of individual genes and polycistronic operons for approximately one-half of the genome (Dataset S1).

Benchmarking our global data with previous studies of individual genes (10, 30, 42), our dRNA-seq results are consistent with published TSSs of *sigH*, *cwpV*, *cwp19*, and *spodA* (Fig. 1C). For annotation purposes, we assigned TSSs to one of five classes according to their genomic location and expression level: pTSS (primary TSS of a gene or operon), sTSS (secondary TSS showing lower expression level compared to pTSS for the same gene or operon), iTSS (internal TSS located inside a gene), aTSS (antisense to a gene within 100-nt distance), and oTSS (orphan TSS, no nearby gene) (Fig. 1D). Naturally, some of these TSS annotations overlap (e.g., among the 1,627 pTSS, 126 are located within a gene (iTSS), and 46 are transcribed antisense (aTSS) to a gene). However, in contrast to other bacteria where antisense transcription is a pervasive transcriptome feature (43), it only accounts for ~6% of TSS events in *C. difficile*.

Internal TSSs were abundant and accounted for 20% of all start sites, many of which seem to uncouple downstream genes within operons. However, they may also indicate misannotated translational start codons or protein isoforms. One example is the putative hydrolase (CDIF630_02227) encoded between *ermB1* and *ermB2* of the erythromycin resistance cassette, which has been shown to be involved in ATP-binding cassette (ABC) transport partitioning. In addition to a very weakly expressed primary TSS upstream of the annotated start codon, we detected a strongly expressed iTSS located 14 nt inside the coding sequence. Manual checking for start codons (AUG, UUG, and GUG) and a ribosome-binding site (RBS) downstream of the iTSS identified a start codon 43 nt downstream of the existing ORF annotation, placing an RBS in optimal distance to the start codon and generating a 5′ UTR length of 29 nt (SI Appendix, Fig. S1A). This annotation is in line with earlier studies in *Streptococcus pyogenes*, which suggested this internal start codon is used in vivo (44) and accordingly used the shorter variant to solve the respective protein crystal structure (PDB ID: 2oe2). Similarly, we propose reannotations in eight additional cases, including the *spaPEGK* operon encoding an antibiotic/multidrug–family ATP-binding cassette (ABC) transport system (Dataset S2).

Promoter Architectures. *C. difficile* encodes 14 sigma factors, among them, consensus sequences have been proposed for the vegetative SigA (28), the general stress response SigB (11–13), the major transition phase SigH (10), and the sporulation-specific SigK (45) sigma factor, mostly based on comparative transcriptome analyses of respective deletion strains. In addition, a recently published strategy for genome-wide identification of promoter sequences, which combined TSS mapping and positional weight matrices based on transcriptome profiles of sigma factor mutants, expanded several existing sigma factor regulons and defined novel regulons, including one for SigL (17).

Here, we applied multiple expectation maximizations for motif elicitation (MEME)-based searches upstream of all TSS leading to the confirmation of published consensus sequences for SigA, SigB, SigH, and SigK (SI Appendix, Fig. S1B and Datasets S1 and S3). In line with recent data (17), the majority of detected TSS (1,188/2,293) were associated with a SigA-type promoter (Dataset S3), including the known TSSs for *sigH* and *clnR* (10, 46, 47). Similarly, we identify many genes previously predicted to be regulated by SigH (10, 17). Furthermore, we identified 52 genes associated with a SigK promoter signature, including known genes of the SigK regulon such as *sleC* and *cdeC* (48). Comparing our promoter predictions with previous studies showed partial overlaps which we attribute to differential computational strategies for promoter assignment as well as growth conditions used (e.g., SigK promoter mutant analyses were performed in sporulation conditions...
using specific sporulation medium). Overall, our results add confidence to existing regulon predictions and extend the list of high-confidence candidates that have been repeatedly identified to be associated with a certain sigma factor (indicated in Dataset S3).

We further identified 39 genes with alternative TSSs that were associated with two different promoter sequences. For example, RecA, which controls the DNA damage response by homologous recombinational repair of damaged DNA, is associated with a SigA- and SigB-type promoter supporting its role in stress responses (SI Appendix, Fig. S1C).

**Global Mapping of Transcript Ends.** To map TTSs, we adopted the RNAtag-seq protocol (49), a technique utilizing initial adapter ligation to exposed RNA 3′ ends. This approach yielded 2,042 experimentally determined TTSs, which were assigned to the following classes: 3′ UTR (downstream of CDS or noncoding RNA), 5′ UTR (between the TSS and the start codon of CDS), CDS (within a coding sequence), orphan (downstream of an orphan TSS), and CRISPR (associated with a CRISPR array) (Fig. 2A).

The majority of TTSs mapped to the region downstream of CDSs. Analysis of primary sequences and secondary structures around these TTSs reveals a downstream poly-U tail motif for the majority (75%) of them (Fig. 2B and Dataset S4). In addition, a decrease in folding energy (SI Appendix, Fig. S2A) and a spike in guanine-cytosine (GC) content at preceding TTS positions (Fig. 2C) indicate stable stem–loop structures, pointing to an important role for intrinsic transcription termination in *C. difficile*.

Our genome-wide map of 3′ UTRs (Dataset S4) reveals that only ~7% of 3′ UTR TTSs are in close distance to a preceding stop codon (<20 nt), whereas 42% of them are >100 nt away (Fig. 2D). This contrasts a recent study in *B. subtilis* that revealed a large fraction of intrinsic terminators overlapping with coding sequences (50). The ability of these terminators to retain efficient transcription termination activity was assigned to runaway transcription of the RNA polymerase, leading to insensitivity to translation-based transcriptional regulation. We explored the original publication, in which a variety of genomes were analyzed in silico, with respect to the presence and location of intrinsic terminators. Reanalysis of the original dataset for *C. difficile* confirms our experimental results that intrinsic terminators rarely overlap with stop codons (SI Appendix, Fig. S2B), suggesting that runaway transcription is not a determining feature of gene expression in *C. difficile*.
We further detected many overlapping termination events between convergently transcribed genes, such as for the toxin gene \textit{tcdA} (\textit{toxA} in CP010905.2 which was used in this study) and the negative regulator of toxin gene expression \textit{tcdC} (Fig. 2E), which is also a pervasive transcriptome feature in \textit{Escherichia coli} (51). Overall, our UTR annotations revealed an abundance of overlapping transcripts between convergently transcribed genes (246 gene pairs or 32.06% of transcripts with TTSs), whereas overlaps between divergently transcribed genes were rare (8 or 1.17% of transcripts with TSSs) (Dataset S8).

**Annotation of Small ORFs and Operons.** High-resolution transcriptome maps allow the identification of ORFs and operon structures that have been overlooked in automated genome annotations. This includes so-called small ORFs (sORFs) of usually 50 amino acids or less, an emerging class of bacterial genes with an unfolding spectrum of biological functions (52, 53). In many cases, they are predicted to be membrane proteins, containing an alpha-helical transmembrane domain (54). Focusing on oTSS in particular, we searched for novel ORFs based on the following criteria: 1) presence of a start and stop codon, 2) presence of an RBS within 15 base pairs (bp) upstream of the start codon, and 3) sequence conservation in other \textit{Clostridioides} strains. Based on this approach, we identified 12 sORF candidates, seven of which are predicted to contain a transmembrane helix (Dataset S2). Among the identified candidates, six are toxins and part of previously identified \textit{C. difficile} type-I toxin–antitoxin systems (55, 56).

Among the remaining sORFs, four are high-confidence candidates: one being a conjugal transfer protein with annotation in other \textit{C. difficile} strains; two candidates that each have a Shine–Dalgarno (SD) sequence and a predicted \(\alpha\)-helical transmembrane domain; and one sORF associated with a 160-nt long 5\('\)UTR region harboring a c-di-GMP-I riboswitch.

Transcriptome studies have revealed operon structures more complex than previously anticipated (51, 57, 58), but predicted operons in \textit{C. difficile} are still largely based on computational inference from other bacterial species (59). Here, our transcriptome-based annotation predicts 440 operons (Dataset S1), including 64 previously unknown operons such as the dicistronic transcript comprising spermine/spermidine acetyltransferase \textit{bltD} and CDIF630_01360, a putative N-acetyltransferase. Furthermore, we identify 109 suboperon structures which, in 29 cases, uncouple the last or the last two genes from upstream genes within an operon. One example is \textit{acps–ndoA}, in which the last two genes encoding a type-II toxin–antitoxin system (\textit{ndoA} and \textit{ndoA1}) are uncoupled from the full operon.

**5\('\)UTRs and Associated Regulatory Elements.** Bacterial 5\('\)UTRs can influence gene expression, usually through embedded cis-regulatory elements such as riboswitches, RNA thermometers, and genetic switches. Similar to \textit{B. subtilis} and \textit{Listeria monocytogenes}, the majority of experimentally mapped 5\('\)UTR lengths ranged from 20 to 60 nt (Fig. 3A and Dataset S1) (14, 40, 60). A likely RBS (aGGAGg) was detected in \(~\)90% of these 5\('\)UTRs (Fig. 3A, Inset).
Only six mRNAs appeared leaderless, having a 5' UTR of <10 nt, including the spoVAE gene within the tricistronic spoVACDE operon (61).

Many genes (561) were associated with surprisingly long 5' UTRs of >100 nt. Among those having a very long 5' UTR of >300 nt, we identified 77 Rfam-predicted riboswitch candidates (62) (Fig. 3B). Premature transcription termination (PTT) was evident for 57 of them, indicating their ON state in which transcription of the parental gene is repressed (Dataset S5). Interestingly, we detected a spoF riboswitch previously only documented in gram-negative alphaproteobacteria associated with CDIF630_01955, encoding a methyltransferase domain protein. As before, the associated termination site suggests that the riboswitch is functional, which is further supported by a 220-nt Northern blot signal (SI Appendix, Fig. S3).

We further detected 17 putative PTT events in 5' UTRs lacking similarities to conserved riboregulators. A potential generation from mRNA processing is unlikely in these cases, since the characteristic read enrichment in the untreated (TEX-) cDNA library, usually associated with processing sites, is missing. One such PTT event is located in the 5' UTR of the infC- rpmI-rplT operon whose expression is controlled through an L20 leader via transcription attenuation in B. subtilis (63, 64). This autoregulatory structure is also found in other low-GC gram-positive bacteria (Rfam family 61).

Attenuation in B. subtilis (65), transcriptional regulators (CDIF630_00097, CDIF630_02384, and CDIF630_02922), multidrug resistance (MDR)-type ABC transporters (CDIF630_03083, CDIF630_02847, and CDIF630_03664), and aroF (Dataset S5). Northern blot validation for a selection of candidates confirmed the RNA-seq-predicted transcript sizes and revealed larger bands in several cases that are likely corresponding to the full-length parental gene or its degradation products (SI Appendix, Fig. S3).

The RNA Landscape of C. difficile in Changing Environmental Conditions and during Infection. Our primary transcriptome analysis identified 42 transcripts that lacked an internal ORF, qualifying them as potential sRNAs (Dataset S5). A classification based on their genomic location (Fig. 4A) revealed the largest group to be 3' UTR-derived sRNAs (18 3' UTR), followed by those encoded cis-antisense either to a gene, a 5'/3' UTR, or another sRNA (13 AS). In comparison, only few sRNAs were located in intergenic regions (8 IGR) or derived from the 5' UTR of mRNAs (5 5' UTR). Sequence conservation beyond C. difficile was extremely rare (SI Appendix, Fig. S4A and Dataset S7), whereas most are highly conserved among clinical strains of C. difficile covering five major genomic clades and...
diverse toxin repertoires (SI Appendix, Fig. S4B) (66). Some of the sRNAs that were absent in a subset of strains were either transposon-associated sRNAs (CDIF630nc_004 and CDIF630nc_069) or located on a prophage (CDIF630nc_095). Of note, this high level of sequence conservation within the species generally extended to noncoding 5′ and 3′ UTR regions (92% of 5′ UTRs with >90% identity and 90% of 3′ UTRs with >90% identity, Dataset S9).

Northern blot validation revealed expression throughout exponential growth for most sRNA candidates (Fig. 4B). About one-half of them are down-regulated after entry into stationary phase, while three candidates (CDIF630nc_028, CDIF630nc_089, and CDIF630nc_105) showed a marked accumulation. In agreement with the expression profile, CDIF630nc_028 has a predicted promoter sequence for SigB, while CDIF630nc_089 is associated with a SigH promoter. Moreover, Northern blot validation of sRNA expression in the hypervirulent RT027 isolate R20291 revealed identical expression profiles in both strains for the selected candidates (SI Appendix, Fig. S5).

To gain insight into sRNA activities during infection, we took advantage of published in vivo transcriptome data from C. difficile–infected ex-germ–free mice as well as conventional mice pretreated with different antibiotics (67). Reanalysis of the data

![Fig. 4. C. difficile 630 sRNA landscape in changing environmental conditions. (A) Classification of annotated sRNA candidates based on their genomic location. The numbers indicate amount of annotated sRNAs for each class. (B) Expression profiles of representative candidates for each sRNA class. Total RNA was extracted at midexponential (ME), late-exponential (LE), and stationary (ST) growth phases from C. difficile 630 grown in TY, TY supplemented with 0.5% glucose (TYG), and BHI medium and analyzed by Northern blot using radioactively labeled DNA probes. (C) Heatmap of log2 CPM values for newly annotated sRNAs during C. difficile 630 intestinal colonization of different antibiotic-treated mouse models (Streptomycin, STR; Cefoperazone, CFP; Clindamycin, CLI) and exGerm-free mice (GF). Published data were taken from ref. 67 and reanalyzed using the updated genome annotation. Only sRNAs with a CPM value ≥50 were considered to be expressed. (D) Expression profiles of sRNAs that are differentially expressed in response to various nutritional and stress conditions. Total RNA was extracted from midexponential TY cultures of C. difficile 630 WT exposed to TY (control), lysozyme, iron limitation (2,2′-Dipyridine, DPD), oxygen, glucose, ethanolamine (EA), succinate, metronidazole (Met), vancomycin (Van), and tetracycline (Tet) for 20 min and analyzed by Northern blot using radioactively labeled DNA probes. (E) Western blot analysis of Hfq protein levels across different growth phases and growth media. Equal optical density (OD) units of total cell lysates of C. difficile 630 WT were loaded. Western blot membranes were incubated with anti-Hfq antibody. Ponceau S staining of the blotting membrane served as loading control. Total cell extracts of equal OD units were prepared from C. difficile 630 WT grown in TY medium to early exponential (EE), late exponential (LE), early stationary (ES), and late stationary (LS) phase and from bacteria grown to late-exponential phase in TY medium supplemented with glucose (Glc), ethanolamine (EA), succinate (Suc), butyrate (But), 2,2′-Dipyridine (DPD), lysozyme (Lys), LL-37, vancomycin (Van), metronidazole (Met), or in 1% oxygen atmosphere (O2). Precise concentrations used are listed under SI Appendix, Materials and Methods.](https://doi.org/10.1073/pnas.2103579118)
revealed expression of 16 sRNAs (cutoff = 50 counts per million mapped reads [CPM]), including the most broadly conserved sRNAs CDIF630nc_001, CDIF630nc_026, and CDIF630nc_079 as well as the SigB- and SigH-associated CDIF630nc_028 and CDIF630nc_089, respectively (Fig. 4C and Dataset S10). To identify infection-related signals regulating sRNA expression, we exposed *C. difficile* to a panel of nutritional and stress conditions including intestinal metabolites (glucose, ethanolamine, succinate, and butyrate), oxygen stress, iron limitation, as well as antimicrobial compounds (lysozyme, LL-37, vancomycin, metronidazole, and tetracycline). Within this panel, we observed regulation for several sRNAs or PTTs (Fig. 4D), e.g., SigB-regulated CDIF630nc_028 responded to iron limitation and oxygen stress, suggesting a potential role during infection. To complement the picture, we also quantified native protein levels of the sRNA chaperone Hfq during growth in these conditions (Fig. 4E). We observed similar Hfq levels not only in different growth phases but also under a variety of nutritional and stress conditions, demonstrating that Hfq is constitutively produced in changing environmental conditions.

**C. difficile** Hfq Is a Global RNA Chaperone. In many gram-negative model organisms, sRNAs require the RNA chaperone Hfq to facilitate their interaction with target mRNAs (68). However, whether Hfq functions as a central RNA-binding protein (RBPs) in *C. difficile* remains unknown. Therefore, we performed Hfq immunoprecipitation followed by sequencing of bound RNA species (RIP-seq) in a strain expressing C-terminally 3xFLAG-tagged hfq (Hfq-FLAG) under its native promoter to draft the spectrum of Hfq-bound RNAs in vivo. Bacteria were grown to midexponential, late-exponential, and stationary phase (SI Appendix, Fig. S6A) and subjected to the RIP-seq protocol (SI Appendix, Fig. S6B). Western blot validation using monoclonal FLAG antibody confirmed expression of Hfq throughout all growth phases and specific enrichment of Hfq-FLAG from *C. difficile* 630 lysate (SI Appendix, Fig. S6C). In addition, copurification of Hfq-bound sRNAs was confirmed by Northern blot analysis of lysates and eluate fractions from Hfq-FLAG and Hfq-ctrl cultures (SI Appendix, Fig. S6C).

The majority of RNA species coimmunoprecipitating with Hfq-3xFLAG were mRNAs (Fig. 5A and Dataset S6). Hfq bound mostly to the 5′ and 3′ UTRs of target mRNAs, and of note, many enriched 5′ UTRs harbored a riboswitch or PTTr (66/124 5′ UTRs) (Fig. 5A) as recently observed in an independent Hfq RIP-seq analysis (20). Many mRNAs showed growth phase–dependent association with Hfq (SI Appendix, Fig. S7 B and C) and mainly included genes involved in the biosynthesis of amino acids, membrane transport, signal transduction, and translation. Most interestingly, we also found several mRNAs encoding proteins involved in sporulation (spoIIA, spoIVB, oppB, sigG, and spa42), toxin production (tdcE), motility (flbD, flbL, flbH, and cheW1), quorum sensing (luxS and agrB), the cell wall protein–encoding cwpV, as well as several adhesins potentially involved in biofilm formation (CDIF630_03429, CDIF630_01459, and CDIF630_03096) (SI Appendix, Fig. S7C and Dataset S6). The enrichment of many 5′ UTRs could suggest that Hfq is recognizing the SD or ATG sequence region, but lack of clear Hfq peaks at SD or start codon sites does not sufficiently support this assumption at this point. In addition, the high number of many 3′ UTRs and of 5′ UTRs harboring a riboswitch or PTTr event could point to a potential role of terminator hairpins in Hfq binding. An RNA motif search using CMfinder identified an RNA structure reminiscent of a terminator hairpin that was enriched in Hfq-bound transcripts, albeit nucleotide identity was low (SI Appendix, Fig. S7A).

Importantly, many predicted sRNAs associated with Hfq, including the previously identified Hfq-binding sRNA CDIF630nc_070 (also known as Red1, Fig. 5.4 and Dataset S6) (18, 28). In addition, all but one type-I antitoxin transcript were enriched (Fig. 5A and Dataset S6), similar to a recent RIP-seq analysis of Hfq in *C. difficile* (20). Classifying Hfq-associated sRNAs with respect to their genomic location revealed that the majority of cis-antisense–encoded sRNAs in *C. difficile* were bound by Hfq, as well as approximately one-half of all intergenic, 5′ UTR–, and 3′ UTR–encoded sRNAs (Fig. 5B). Additionally, many sRNAs were bound across all three growth phases, with CDIF630nc_070 and CDIF630nc_079 being the two top-enriched sRNAs (SI Appendix, Fig. S6D and Dataset S6). However, a few sRNAs displayed growth phase–dependent association with Hfq, such as CDIF630nc_090, which was only enriched in stationary phase (SI Appendix, Fig. S6D).

Hfq Acts as a Stabilizing Factor for its Noncoding RNA Ligands. One major function of gram-negative Hfq homologs is the stabilization of their sRNA ligands, which appears to be the exception in gram-positive bacteria (69–71). To test if transcript stabilities of Hfq–associating noncoding RNAs (ncRNAs) were affected by Hfq, we generated a hfq deletion (Δhfq) in *C. difficile* 630 using a homologous recombination vector we developed for effective counterselection on standard laboratory media using the tightly regulated *E. coli* endoribonuclease mazF (for details, see SI Appendix). To complement hfq in cis (hfqC), we established an additional vector system which enables efficient insertion of DNA by homologous recombination between pyrE (CD630_01870) and CD630_01880. *C. difficile* wild type (WT), Δhfq, and hfqC strains were grown to late-exponential phase, and rifampicin was added to arrest transcription. Northern blotting of RNA samples collected at several timepoints after rifampicin addition was performed to

---

Fuchs et al.  
An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium  
PNAS | 7 of 11  
https://doi.org/10.1073/pnas.2103579118
quantify transcript levels for several sRNAs as well as a Hfq-bound riboswitch and type-I antitoxin. Remarkably, half-lives and, in some cases, the steady-state levels of all analyzed Hfq-associated sRNAs (CDIF630nc_006, CDIF630nc_085, and CDIF630nc_022) were decreased, an effect that could be restored in the hfq− strain (Fig. 5C and SI Appendix, Fig. S8A and B) and that was not observed for the Hfq-independent sRNA CDIF630nc_009. In addition, half-lives were also reduced for the tested type-I antitoxin as well as the c-di-GMP-II riboswitch, suggesting a potential dual function for the latter cis-acting regulator (SI Appendix, Fig. S8B).

A Role of CDIF630nc_085 in the Regulation of Ethanolamine Utilization. To gain insight into sRNA functions, we chose the bona fide intergenic sRNA CDIF630nc_085 (short nc085) (Fig. 4B) for further

---

Fig. 6. CDIF630nc_085 regulates ethanolamine utilization in C. difficile. (A) Predicted secondary structure for CDIF630nc_085 (nc085) using RNAfold. Nucleotides are colored according to base-pairing probabilities. Two seed regions predicted to interact with differentially expressed mRNA targets are located in an extended small stem loop (seed region 1) and a single-stranded stretch (seed region 2) in the 3′ region of the sRNA, respectively. (B) Northern blot validation of nc085 pulse expression in a nc085 deletion strain. sRNA expression was transiently induced for 20 min by addition of 100 ng/mL anhydrotetracycline (ATc) to late-exponential cultures of C. difficile WT, Δnc085 (−), Δnc085 harboring the empty pRPF185 backbone (EV), and Δnc085 harboring the pulse-expression vector carrying nc085 (PE) under the control of an ATc-inducible promoter. (C) RNA was extracted from ATc-induced cultures harboring the empty vector (EV) or the pulse-expression plasmid (PE) and analyzed by RNA-seq. The genes that were differentially expressed upon pulse expression of nc085 (false discovery rate ≤ 0.1, log2 fold change ≥ 1, or log2FC ≤ −1) are colored according to their transcript class. (D) Read profiles from untreated (Tex−), Tex-treated (Tex+), and RNAtag-seq libraries mapping to the eut region. TSSs were identified upstream of eutS and eutV. An RNA duplex formed by nc085 with the start codon region of eutV was predicted using IntaRNA. The red letters indicate the start codon. (E) In-line probing of 0.2 pmol of 32P-labeled nc085 in the absence (lane 3) or presence of either 0.2 pmol (lane 4), 2 pmol (lane 5), 20 pmol (lane 6), or 200 pmol (lane 7) eutV target region. Partially RNase T1– (lane 1; T1) or alkali-digested (lane 2; OH) nc085 served as ladders. Seed 1 (red) denotes the seed region that is protected from cleavage changes in the presence of the target. A representative image of two independent experiments is shown. (F) Growth of C. difficile 630 WT, Δhfq, and Δnc085 in CDMM supplemented with either glucose (5 mM) or ethanolamine (50 mM). Growth experiments were performed in two independent experiments with three biological replicates each (n = 6). The error bars show SEM. Bottom table: listed are the calculated doubling times (DT) ± SD as well as the maximum optical densities (OD600) and the timepoint when they were reached during the experiment, for each strain and each growth condition. Due to the complex, diauxic growth behavior of all strains in the analyzed growth conditions and doubling times were calculated only for the first exponential phase.

---
characterization because its expression during intestinal colonization (Fig. 4C) suggested a potential virulence function. Secondary structure prediction showed a moderately folded structure with an extended stem loop at the 5′ end, two hairpin pairs in the center of the sRNA, and a Rho-independent terminator at the 3′ end (Fig. 6A). To identify potential mRNA targets of nc085, we pulsed the sRNA from an anhydrotracycline-inducible plasmid in a ΔCDF1630nc_085 background (Fig. 6B) and determined differentially expressed genes using RNA-seq analysis (Fig. 6C and Dataset S1). We observed many sporulation-associated genes to be down-regulated upon nc085 expression (Dataset S1A). In addition, gene cluster analysis revealed ABC transporters and two-component systems as the largest group of regulated genes (SI Appendix, Fig. S9B). Searching for base-pairing regions within the 5′ UTRs of differentially expressed genes identified two potential seed regions within nc085 (Fig. 6A, SR1 and SR2), both complementarity to several target mRNAs (SI Appendix, Fig. S9A).

Among them was eutV, which is part of a two-component system encoded by the eutVW operon, a positive regulator of ethanolamine utilization. Our transcriptome data identified low-level constitutive expression across all growth conditions and a TSS upstream of the eutVW operon (Fig. 6D, data only shown for late-exponential TY condition). Moreover, the predicted base-pairing region within the eutV 5′ UTR comprised the start codon site of eutV (Fig. 6D).

To test the prediction experimentally, we performed in-line probing of nc085 with or without the eutV 5′ UTR. Although high concentrations of the target mRNA were required, changes in the protection from cleavage in the presence of eutV were observed (Fig. 6E, lane 7) and provided support for the computationally predicted base-pairing region in nc085. Considering the location of the seed region within a hairpin structure, this interaction may need to be facilitated by Hfq in vivo. To explore the impact of nc085 on ethanolamine utilization, we compared growth of C. difficile WT, Δnc085, and the Δhfq mutant in C. difficile minimum medium (CDMM) supplemented either with glucose or ethanolamine. Cultivation in glucose-supplemented medium resulted in identical growth for WT and Δnc085, whereas Δhfq had an extended lag phase and grew to lower maximum optical densities. In contrast, when grown in CDMM supplemented with ethanolamine, both Δnc085 and Δhfq exhibited accelerated growth in comparison to the WT (Fig. 6F).

Specifically, both strains entered exponential growth faster and reached their maximum optical densities at earlier timepoints (Fig. 6 F, Bottom). Together, these data identify a regulatory function of nc085 in the utilization of the abundant intestinal carbon and nitrogen source ethanolamine, which was shown to impact pathogenicity of C. difficile (72).

Discussion
In the present study, we have applied global approaches of bacterial RNA biology (35) to capture the transcriptome architecture of C. difficile and unravel the scope of posttranscriptional regulation in this important human pathogen.

A Gram-Positive Hfq Homolog that Impacts sRNA Stabilities and Function. It is undisputed that sRNAs are central players in the regulation of physiological and virulence pathways in Gram-positive bacteria including L. monocytogenes, B. subtilis, and Staphylococcus species. For example, the sRNA RsaE, which is conserved in the order Bacillales, has been shown to coordinate central carbon and amino acid metabolism in Staphylococcus aureus (73, 74) and also to promote biofilm formation in Staphylococcus epidermidis by supporting extracellular DNA release and the production of polysaccharide-intercellular adhesive (75). However, the function of Hfq in facilitating the interactions between sRNA regulators and their target mRNAs in gram-positive bacteria remains elusive since most of the known sRNAs seem to exert their regulatory activity independent of this global RBP (76).

In the present study, our characterization of Hfq–sRNA interactions in C. difficile not only revealed a large network of Hfq-binding sRNAs along with potential target mRNAs but also provides clear evidence for the impact of Hfq on sRNA steady-state levels and stabilities in vivo. Our data align with initial studies using a Hfq-depleted C. difficile strain that identified pleiotropic effects of Hfq depletion on gene expression using microarray analysis (18). In line with these observations, absence of Hfq impacts growth on various carbon and nitrogen sources, as shown in our study, and results in increased sensitivity to stresses, sporulation rates, and biofilm formation (18). Taken together, the regulatory activities of Hfq in C. difficile strongly resemble the situation in gram-negative model organisms such as E. coli and Salmonella enterica and open avenues for the study of sRNA-based gene regulation in this important pathogen.

Ethanolamine Utilization Is Governed by sRNA Regulation in Gram-Positive Pathogens. We have taken a first step toward the characterization of sRNA-mRNA interactions in C. difficile by identifying the ethanolamine utilization pathway to be regulated by the intragenic sRNA nc085. Ethanolamine is an abundant intestinal nutrient that is exploited by both gram-negative and gram-positive enteric pathogens, including C. difficile, during intestinal colonization (72, 77–79). Interestingly, the ethanolamine utilization (eut) gene cluster is regulated by a small regulatory RNA in both Enterococcus faecalis (EutX) and L. monocytogenes (Rli55). In both species, the sRNA contains a classic AdoCbl-responsive riboswitch in its 5′ portion that allows expression of the full-length sRNA only in the absence of vitamin B12 (80, 81). This full-length transcript harbors a dual-hairpin structure in its 3′ part that serves as a binding platform for the ethanolamine-specific, RNA-binding response regulator EutV (82), thereby sequestering EutV from binding to its target sequences in the 5′ UTRs of eut genes. An in silico search for an AdoCbl riboswitch failed to identify such a conserved element within nc085, and secondary structures of the three sRNAs show a less-complex folding pattern for nc085 that is missing the dual hairpin for EutV binding. Overall, this suggests that the mechanism of eut gene regulation by nc085 differs from EutX and Rli55. Currently, down-regulation of the eutVW operon upon nc085 pulse expression together with a predicted target site comprising the start codon of eutV suggests a classical mode-of-action whereby nc085 binding interferes with translation initiation. Consequently, deleting nc085 should relieve the repression and increase the pool of EutVW that can sense ethanolamine and activate the eut gene cluster. Aligning with this model, we observed early onset of growth for the nc085 deletion strain in CDMM with ethanolamine as the sole carbon source (Fig. 6F).

Overall, our findings further confirm that ethanolamine utilization is a conserved target of sRNA-based regulation in gram-positive pathogens.

Materials and Methods
Comprehensive descriptions of all materials and methods that were used in this study are provided in SI Appendix, Materials and Methods.

Data Availability. All RNA-sequencing data are available at the National Center for Biotechnology Information Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo) under the accession number GSE151567. Plasmids pKJAK112, pJAK184, and pJAK080 have been deposited with Addgene (167279–167281).

ACKNOWLEDGMENTS. We thank the Core Unit SysMed at the University of Würzburg for technical support with RIP-seq and RNAtag-seq data generation. This work was supported by the Interdisciplinary Center for Clinical Research (IZKF) at the University of Würzburg (Project No. 2-6). We also
The research conducted by the authors, which includes work with the Virchow Center for the expression and purification of Hfq, was supported by the German Research Foundation (DFG Grant No. FA 11132-1). J.A.K. was supported by the Wellcome Trust (Grant No. 208477/Z/16/Z).

A recent genetic study of Clostridioides difficile disease reveals broad activity of Hfq in a clinically important gram-positive bacterium.
67. M. L. Jenier, J. L. Leslie, V. B. Young, P. D. Schloss, Clostridium difficile colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. mSystems 2, e00063-17 (2017).
68. J. Hör, G. Matera, J. Vogel, S. Gottesman, G. Storz, Trans-acting small RNAs and their effects on gene expression in Escherichia coli and Salmonella enterica. Ecosal Plus 9, 10.1128/ecosalplus.ESP-0030-2019 (2020).
69. I. Ul Haq, P. Müller, S. Brantl, Intermolecular communication in Bacillus subtilis: RNA-RNA, RNA-protein and small protein-protein interactions. Front. Mol. Biosci. 7, 178 (2020).
70. J. S. Nielsen et al., Defining a role for hfq in gram-positive bacteria: Evidence for hfq-dependent antisense regulation in Listeria monocytogenes. Nucleic Acids Res. 38, 907–919 (2010).
71. J. K. Christiansen et al., Identification of small Hfq-binding RNAs in Listeria monocytogenes. RNA 12, 1383–1396 (2006).
72. K. L. Nawrocki, D. Wetzel, J. B. Jones, E. C. Woods, S. M. McBride, Ethanolamine is a valuable nutrient source that impacts Clostridium difficile pathogenesis. Environ. Microbiol. 20, 1419–1435 (2018).
73. C. Bohn et al., Experimental discovery of small RNAs in Staphylococcus aureus reveals a riboregulator of central metabolism. Nucleic Acids Res. 38, 6620–6636 (2010).
74. T. Rochat et al., The conserved regulatory RNA RsAE down-regulates the arginine degradation pathway in Staphylococcus aureus. Nucleic Acids Res. 46, 8803–8816 (2018).
75. S. M. K. Schoenfelder et al., The small non-coding RNA RsAE influences extracellular matrix composition in Staphylococcus epidermidis biofilm communities. PLoS Pathog. 15, e1007618 (2019).
76. S. Pitman, K. H. Cho, The mechanisms of virulence regulation by small noncoding RNAs in low GC gram-positive pathogens. Int. J. Mol. Sci. 16, 29797–29814 (2015).
77. Y. Bertin et al., Enterohaemorrhagic Escherichia coli gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content. Environ. Microbiol. 13, 365–377 (2011).
78. K. G. Kaval et al., Loss of ethanolamine utilization in Enterococcus faecalis increases gastrointestinal tract colonization. mBio 9, e00790-18 (2018).
79. P. Thiennimitr et al., Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. Proc. Natl. Acad. Sci. U.S.A. 108, 17480–17485 (2011).
80. S. DebRoy et al., Riboswitches. A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. Science 345, 937–940 (2014).
81. J. R. Mellin et al., Riboswitches. Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA. Science 345, 940–943 (2014).
82. M. F. Del Papa, M. Perego, Ethanolamine activates a sensor histidine kinase regulating its utilization in Enterococcus faecalis. J. Bacteriol. 190, 7147–7156 (2008).