Global discovery and functional characterization of Hfq-associated sRNA-target networks in *C. difficile*

Globale Identifizierung und funktionelle Charakterisierung von Hfq-assoziierten sRNA-Zielnetzwerken in *C. difficile*

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and Immunity

submitted by

Manuela Fuchs

from Offenburg

Würzburg 2023

This document is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0):
http://creativecommons.org/licenses/by-nc-nd/4.0 This CC license does not apply to third party material (attributed to another source) in this publication.

Submitted on:

Members of the Thesis Committee

Chairperson: Prof. Thomas Dandekar

Primary Supervisor: Jun. Prof. Franziska Faber

Supervisor (Second): Dr. Ana Rita Brochado

Supervisor (Third): Jun. Prof. Alexander Westermann

Supervisor (Fourth): Dr. Meina Neumann-Schaal

Summary

The anaerobic Gram-positive human pathogen *Clostridioides difficile* (*C. difficile*) has emerged as the leading cause of nosocomial antibiotic-associated diarrhoea. Various virulence factors and traits influence disease progression and severity, including toxin expression and spore formation. Small regulatory RNAs (sRNAs) are well-known post-transcriptional regulators of virulence and stress associated pathways in many pathogenic bacteria. In Gram-negative species, sRNA-dependent post-transcriptional regulation is frequently mediated by the RNA chaperone Hfq, which facilitates sRNA-mRNA base-pairing. Despite its importance in Gram-negative bacteria, comparatively little is known about the various aspects of post-transcriptional regulation in Gram-positive species. Initial data has indicated an important function of Hfq in mediating regulation of several infection-associated pathways in *C. difficile*, as well as the existence of a large post-transcriptional network. However, a global identification of Hfq-associated RNAs and their impact on *C. difficile* virulence and colonization remains unexplored.

In this work, dRNA-seq (differential RNA sequencing) and RNAtag-seq were applied to first define the global transcriptome architecture of *C. difficile*, followed by Hfq RIP-seq (RNA immunoprecipitation followed by RNA-seq) and RIL-seq (RNA interaction by ligation and sequencing) to characterize the Hfq-mediated sRNA interactome on a transcriptome-wide scale. These approaches resulted in the annotation of > 60 novel sRNAs. Notably, it not only revealed 50 Hfq-bound sRNAs, but also > 1000 mRNA-sRNA interactions, confirming Hfq as a global RNA matchmaker in *C. difficile*. Similar to its function in Gram-negative species, deletion of Hfq resulted in decreased sRNA half-lives, providing evidence that Hfq affects sRNA stability in *C. difficile*. Finally, several sRNAs and their function in various infection relevant conditions were characterized. The sRNA nc085 directly interacts with the two-component response regulator *eutV*, resulting in regulation of ethanolamine utilization, an abundant intestinal carbon and nitrogen source known to impact *C. difficile* pathogenicity. Meanwhile, SpoY and SpoX regulate translation of the master regulator of sporulation *spo0A in vivo*, thereby affecting sporulation initiation. Furthermore, SpoY and SpoX deletion significantly impacts *C. difficile* gut colonization and spore burden in a mouse model of *C. difficile* infection.

Overall, this work provides evidence for extensive Hfq-dependent post-transcriptional regulation affecting physiology and virulence in a Gram-positive pathogen. Although the work presented in this thesis has barely scratched the surface of sRNA-mediated regulation in *C. difficile*, the RIL-seq data may serve as basis for future mechanistic studies of RNA-based gene regulation in *C. difficile.*

Zusammenfassung

Der anaerobe Gram-positive humanpathogene Erreger *Clostridioides difficile* (*C. difficile*) gilt als Hauptursache für nosokomiale Antibiotika-assoziierte Diarrhöe. Verschiedene Virulenzfaktoren und -eigenschaften beeinflussen das Fortschreiten und den Schweregrad der Krankheit, darunter Toxinexpression und Sporenbildung. Kleine regulatorische RNAs (sRNAs) sind bekannte post-transkriptionelle Regulatoren von Virulenz- und Stress-assoziierten Stoffwechselwegen in vielen pathogenen Bakterien. In Gram-negativen Arten wird sRNA-abhängige post-transkriptionelle Regulierung häufig durch das RNA-Chaperon Hfq vermittelt, welches die sRNA-mRNA-Basenpaarung erleichtert. Trotz ihrer Bedeutung in Gram-negativen Bakterien ist vergleichsweise wenig über die verschiedenen Aspekte der post-transkriptionellen Regulation in Gram-positiven Arten bekannt. Erste Daten deuten auf eine wichtige Funktion von Hfq bei der Regulierung verschiedener infektionsassoziierter Signalwege in *C. difficile* hin, sowie auf die Existenz eines umfangreichen post-transkriptionellen Netzwerks. Eine globale Identifizierung von Hfq-assoziierten RNAs und deren Einfluss auf die Virulenz von und Kolonisierung durch *C. difficile* ist jedoch bisher noch nicht erfolgt.

In dieser Arbeit wurde dRNA-seq (differentielle RNA-Sequenzierung) und RNAtag-seq angewandt, um zunächst die globale Transkriptom-Architektur von *C. difficile* zu definieren. Anschließend wurde Hfq RIP-seq (RNA-Immunpräzipitation gefolgt von RNA-seq) und RIL-seq (RNA-Interaktion durch Ligation und Sequenzierung) durchgeführt, um das Hfq-vermittelte sRNA-Interaktom auf globaler Ebene zu charakterisieren. Diese Ansätze führten zur Annotation von > 60 neuen sRNAs. Darüber hinaus wurden 50 Hfq-gebundene sRNAs, sowie > 1000 mRNA-sRNA-Interaktionen identifiziert, wodurch Hfq als globaler RNA-Matchmaker in *C. difficile* bestätigt wurde. Analog zu seiner Funktion in Gram-negativen Arten, führte die Deletion von Hfq zu verringerten sRNA-Halbwertszeiten, was darauf hindeutet, dass Hfq die sRNA-Stabilität in *C. difficile* beeinflusst. Schließlich wurden mehrere sRNAs und ihre Funktion unter verschiedenen infektionsrelevanten Bedingungen charakterisiert. Die sRNA nc085 interagiert direkt mit dem Zweikomponenten-Regulator *eutV*, was zu einer Regulierung der Ethanolaminverwertung führt. Als häufig vorkommenden Kohlenstoff- und Stickstoffquelle im Darm, kann Ethanolamin die Pathogenität von *C. difficile* beeinflussen. SpoY und SpoX regulieren dagegen die Translation des Hauptregulators der Sporulation *spo0A in vivo* und damit die Sporulationsinitiation. Darüber hinaus hat die Deletion von SpoY und SpoX signifikante Auswirkungen auf die Besiedlung des Darms mit *C. difficile* sowie die Sporenbelastung in einem Mausmodell der *C. difficile*-Infektion.

Insgesamt liefert diese Arbeit Beweise für eine umfassende Hfq-abhängige post-transkriptionelle Regulierung, die die Physiologie und Virulenz eines Gram-positiven Erregers beeinflusst. Auch wenn mit dieser Arbeit die Charakterisierung der sRNA-vermittelten Regulation in *C. difficile*

gerade erst begonnen hat, können die RIL-seq-Daten als Grundlage für zukünftige mechanistische Studien der RNA-basierten Genregulation in *C. difficile* herangezogen werden.

List of Publications

The following list shows all publications that were authored or co-authored as a result of this PhD thesis in chronological order.

1. Lamm-Schmidt, V., **Fuchs, M.**, Sulzer, J., Gerovac, M., Hör, J., Dersch, P., Vogel, J., & Faber, F. Grad-seq identifies KhpB as a global RNA-binding protein in *Clostridioides difficile* that regulates toxin production. *MicroLife*, 2(April), 1–21 (2021).

   https://doi.org/10.1093/femsml/uqab004

2. **Fuchs, M.**, Lamm-Schmidt, V., Sulzer, J., Ponath, F., Jenniches, L., Kirk, J. A., Fagan, R. P., Barquist, L., Vogel, J., & Faber, F. An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important Gram-positive bacterium. *Proceedings of the National Academy of Sciences*, 118(25) (2021).

   https://doi.org/10.1073/pnas.2103579118

3. **Fuchs, M.**, Lamm-Schmidt, V., Lenče, T., Sulzer, J., Bublitz, A., Wackenreuter, J., Gerovac, M., Strowig, T., & Faber, F. A network of small RNAs regulates sporulation initiation in *Clostridioides difficile*. *The EMBO Journal*, e112858 (2023).

   https://doi.org/10.15252/embj.2022112858

Table of Contents

List of Figures

Abbreviation Index

Abbreviation	
asRNA	antisense RNA
B. subtilis	*Bacillus subtilis*
BAs	bile acids
C. difficile	*Clostridioides difficile*
C. sardiniense	*Clostridium sardiniense*
CDC	Centers for Disease Control and Prevention
CDI	*C. difficile* infection
CDS	coding sequence
CDT	*C. difficile* transferase or binary toxin
CDTloc	CDT locus
CLASH	UV cross-linking, ligation, and sequencing of hybrids
CLIP-seq	cross-linking and immunoprecipitation followed by deep sequencing
CRISPR	clustered regularly interspaced short palindromic repeats
CTD	C-terminal domain
dRNA-seq	differential RNA sequencing
E. coli	*Escherichia coli*
E. faecalis	*Enterococcus faecalis*
e.g.	*exempli gratia*; for example
et al.	*et alii*; and others
FC	fold change
FMT	faecal microbiota transplantation
GRIL-seq	global small non-coding RNA target identification by ligation and sequencing
Hi-GRIL-seq	high-throughput GRIL-seq
IGR	intergenic region
in silico	performed on computer or *via* computer simulation
in vitro	performed in a test tube, culture dish, or elsewhere outside a living organism
in vivo	performed in whole, living organisms or cells
L. monocytogenes	*Listeria monocytogenes*

Abbreviation	
MAPS	MS2-affinity purification coupled with RNA sequencing
mRNA	messenger RNA
ncRNAs	non-coding RNAs
OMP	outer membrane protein
ORF	open reading frame
P. aeruginosa	*Pseudomonas aeruginosa*
P. bifermentans	*Paraclostridium bifermentans*
PaLoc	pathogenicity locus
PE	pulse-expression
RBP	RNA binding protein
RBS	ribosome binding site
RIL-seq	RNA interaction by ligation and sequencing
RIP-seq	RNA immunoprecipitation followed by RNA-seq
RNA	ribonucleic acid
RNAP	RNA polymerase
RNA-seq	RNA sequencing
RT	ribotype
rut site	Rho utilisation site
S. aureus	*Staphylococcus aureus*
S. enterica	*Salmonella enterica*
S. typhimurium	*Salmonella typhimurium*
SCFAs	short chain fatty acids
SD	Shine-Dalgarno
sRNA	small regulatory RNA
TSS	transcriptional start site
TTS	transcriptional termination site
UTR	untranslated region
V. cholerae	*Vibrio cholerae*
Y. pestis	*Yersinia pestis*

1 Introduction

1.1 A brief summary of ncRNA-mediated regulation in bacteria

Bacteria have evolved numerous mechanisms to regulate transcription and translation in response to ever-changing environmental conditions. In the past two decades, regulatory non-coding RNAs (ncRNAs) have been identified as important factors that enable a tight and rapid control of gene expression, influencing transcription, translation, and stability of their mRNA targets [1,2]. Although varying nomenclatures can be found in the literature, regulatory ncRNAs are most commonly grouped into *cis*- and *trans*-encoded ncRNAs, depending on their genomic location in relation to their mRNA target. Accordingly, *cis*-encoded ncRNAs are located in direct proximity to their target and include riboswitches, RNA thermometers and antisense RNAs (asRNA) [3]. Riboswitches and RNA thermometers typically reside in the 5' untranslated region (UTR) of their respective target gene and fold into complex secondary and tertiary structures. Binding of small molecules (for riboswitches) or changes in temperature (for RNA thermometers) lead to conformational rearrangements and consequently modulation of transcription or translation of the associated gene [4,5]. Similar to riboswitches and RNA thermometers, asRNAs are co-localized to their target gene although transcribed from the opposite strand. Perfect sequence complementarity allows base-pairing of asRNA to their target mRNA, affecting gene expression and/or mRNA stability [6].

In contrast to *cis*-encoded ncRNA, *trans*-encoded ncRNAs and their targets are encoded in different genomic locations and encompass clustered regularly interspaced short palindromic repeat (CRISPR) RNAs and small regulatory RNAs (sRNAs). CRISPR RNAs are part of the bacterial adaptive immune system and facilitate targeting of foreign genetic material by endonucleases [7]. sRNAs, similar to asRNAs, interact and regulate their target *via* direct base-pairing, most commonly affecting the translational efficiency and stability of targeted mRNAs [8]. However, in contrast to asRNAs, sRNAs and their targets exhibit only partial sequence complementarity and frequently require an RNA binding protein (RBP) to facilitate the interaction [8,9]. A more detailed introduction to the different aspects of sRNA-mediated regulation is given in the following chapters.

1.2 sRNA in bacteria – an introduction

1.2.1 The past and present of sRNA identification and annotation

The first evidence of a *trans*-encoded sRNA regulating gene expression in bacteria dates back to the 1980s, when Mizuno and colleagues described the post-transcriptional regulation of the outer

membrane porin OmpF by MicF, a 93 nt long sRNA in *Escherichia coli* (*E. coli*) [10,11]. Their results revealed that regulation of *ompF* occurs by binding of MicF to the ribosome binding site (RBS) of the *ompF* mRNA, thereby inhibiting ribosome binding and consequently translation of the outer membrane porin [10]. Similar to MicF, identification of sRNAs in the early days of sRNA research was mostly a by-product of analysing operon expression in specific stress conditions [10,12,13]. However, with the development of computational techniques during the early 2000s, systematic predictions of sRNAs based on known sRNA features became available, allowing a more directed identification [14–18]. Furthermore, the discovery of sRNA-binding proteins, such as Hfq (see 1.2.4, 1.3 and 1.4), provided the possibility of co-purifying unknown sRNA binding partners *via* protein co-immunoprecipitation [19,20]. Coupled with supplementary experiments, such as microarray detection or RNA labelling, these experiments accelerated the discovery and annotation of sRNAs, albeit mainly focused on the Gram-negative model organism *E. coli* [19,21].

Finally, the rise of high-throughput sequencing techniques, in particular RNA deep sequencing (RNA-seq) has also opened a new chapter of sRNA discovery. Combining RNA-seq with protein co-immunoprecipitation experiments (e.g. RIP-seq or CLIP-seq, described in 1.4.2) or computational analyses has greatly improved our ability to rapidly identify and annotate potential sRNA candidates [22–25]. Most notably the introduction of differential RNA-seq (dRNA-seq) in 2010 by Sharma and colleagues has resulted in extensive annotation of sRNAs in numerous bacterial and several archaeal species, including human pathogens such as *Campylobacter jejuni*, *Yersinia pestis* (*Y. pestis*), *Staphylococcus aureus* (*S. aureus*), *Mycobacterium tuberculosis* and *Vibrio cholerae* (*V. cholerae*) [26–33]. Originally designed to describe the primary transcriptome of the gastric pathogen *Helicobacter pylori*, the method relies on selectively sequencing primary transcripts (harbouring a 5'-triphosphate) through prior degradation of processed transcripts (5'-monophosphate or 5'-hydroxyl) with terminator exonuclease [31,32]. As a consequence, transcription start sites (TSSs) can be mapped genome wide, allowing the annotation of 5'UTRs, operon structures as well as sRNAs [31,32]. Taken together, these methods have significantly contributed and continue to contribute to our understanding of sRNA expression in bacteria.

1.2.2 Reservoirs of bacterial sRNAs – IGRs and beyond

During the early days of sRNA discovery, sRNAs were characterized as 50 to 400 nt long non-coding RNA fragments that are not translated into proteins [19]. Additionally, they were thought to be highly conserved and mostly encoded in IGRs with their own dedicated promoter and a strong Rho-independent terminator [19,21]. However, in the ensuing 20 years of research on sRNA-mediated regulation, many of these assumptions have been revised resulting in a more nuanced definition [33]. Although intergenic sRNAs, such as the previously mentioned MicF, remain the largest group of functionally characterized sRNAs, it became apparent that sRNAs can be transcribed or processed from nearly any part of the bacterial genome (Figure 1) [10,33].

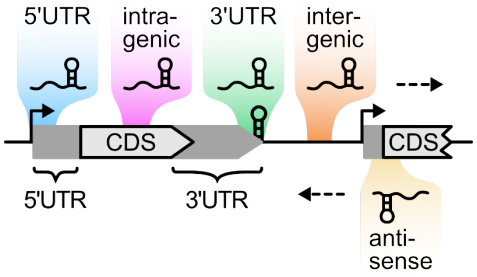


Figure 1: Reservoirs of bacterial sRNAs. Bacterial sRNAs can be transcribed or processed from nearly any part of the bacterial genome. Transcriptional start sites, intrinsic terminators, and untranslated regions are indicated.

1.2.2.1 5'UTR-encoded sRNAs

Usually known as sources for *cis*-acting regulatory ncRNAs, mRNA 5'UTRs have been shown to also encode *trans*-acting sRNAs that are either intrinsically terminated or processed to form a distinct transcript [33,34]. For example, RhlS, an sRNA discovered in *Pseudomonas aeruginosa* (*P. aeruginosa*), was shown to share its TSS with the *rhlI* mRNA encoded downstream of RhlS, while terminating at a distinct transcription termination site (TTS) 34 nt upstream of the *rhlI* start codon [35]. In contrast, srn135, an sRNA that regulates pilus expression in *Streptococcus pneumoniae* (*S. pneumoniae*), is a product of several processing events within the *rrgA* 5'UTR, although the relevant ribonucleases remain unknown [36]. Moreover, there has been evidence that transcripts originating from riboswitches might also act *in trans* in addition to their *cis*-regulatory function [15,37,38]. These include EutX from *Enterococcus faecalis* (*E. faecalis*) and Rli55 from *Listeria monocytogenes* (*L. monocytogenes*), both of which simultaneously encode and are regulated by an adenosylcobalamin-responsive riboswitch [38,39]. In the absence of adenosylcobalamin, the full-length sRNAs are transcribed and able to sequester a two-component response regulator that is necessary for ethanolamine utilization. In contrast, when adenosylcobalamin is available, the riboswitches produce a short truncated sRNA isoform that can no longer bind the two-component response regulator, resulting in expression of the ethanolamine utilization operon. As adenosylcobalamin is a necessary cofactor for ethanolamine metabolism, the sRNAs prevent transcription of genes relevant for ethanolamine utilization in unfavourable conditions.

1.2.2.2 3'UTR-encoded sRNAs

Similar to 5'UTRs, 3'UTRs were found to be potential reservoirs for sRNAs that are either processed (also referred to as type II 3'UTR-derived sRNA) or independently transcribed from the associated mRNA (also referred to as type I 3'UTR-derived sRNA) [33,40–43]. In both cases, sRNA and mRNA usually share the same Rho-independent terminator consisting of a stable stem-loop followed by a poly-U stretch [33,40–43]. Stable transcripts originating from mRNA 3'UTRs were already detected during initial screens for sRNA candidates in *E. coli*, which was soon followed by functional characterizations of several 3'UTR-encoded sRNAs [15,40,41]. Examples are DapZ, encoded in the *dapB* 3'UTR in *Salmonella typhimurium* (*S. typhimurium*), which is involved in the regulation of amino acid synthesis, or CpxQ, a highly conserved sRNA located in the *cpxP* 3'UTR in *E. coli*

which is relevant for the inner membrane stress response [44,45]. While DapZ is independently transcribed under the control of a promoter located within the *dapB* coding sequence (CDS), CpxQ is processed from the *cpxP* mRNA in an RNase E-dependent manner [44,45]. Interestingly, Hfq co-immunoprecipitation approaches proved to be particularly successful in identifying 3'UTR-encoded sRNAs [41,44]. This can be explained by the necessity of a Rho-independent terminator structure at the sRNA 3'end for Hfq binding (see 1.3.1), which is a common feature of 3'UTR-encoded sRNA and a consequence of their genomic location [41,42,46].

1.2.2.3 Emerging classes of sRNAs

Besides sRNAs located in mRNA 5'- and 3'UTRs, there is also evidence of antisense sRNAs that regulate targets *in trans*, in addition to their *cis*-regulatory function [33,47]. The *E. coli* sRNA ArrS, for example, is usually required for RNase III-mediated processing of the antisense-encoded *gadE* mRNA [48,49]. However, data obtained from Hfq-associated sRNA-target interaction studies suggest additional *trans*-encoded targets of ArrS, although verification of these interactions is still missing [50]. Several global approaches including transcriptome wide TSS and 3'end mapping have also provided evidence of intragenic sRNAs that reside entirely within a CDS [34,50,51]. For example, the sRNA FtsO is part of the CDS of the cell division protein FtsI in *E. coli* and most likely processed from the *ftsI* mRNA [34].

Furthermore, analysis of various sRNA sequences revealed the existence of small open reading frames (ORFs), contradicting the original assumption that sRNAs are not translated into proteins [52–54]. Several of these sRNAs, termed "dual-function" sRNAs have been investigated in more detail including SR1, a highly conserved 205 nt long sRNA discovered in *Bacillus subtilis* (*B. subtilis*) that also encodes the protein SR1P [55–57]. While SR1 regulates arginine metabolism, SR1P positively regulates the glycolytic enzyme GapA.

As a whole, these examples illustrate the broad variety of sRNAs that have been discovered so far. Although the most prevalent sources of sRNAs have been covered, additional sources such as tRNA precursors have been discussed in the literature and indicate yet unknown reservoirs of sRNAs [33].

1.2.3 Regulation of sRNA biogenesis

Originally, sRNAs were characterized as separate transcripts that are regulated by specific transcription factors [19,21]. However, with the discovery of additional sRNA reservoirs, such as 5'UTRs, 3'UTRs or CDSs, it has become apparent that alternative factors must be involved in generating sRNAs [33]. Besides dedicated promoters and Rho-independent terminators, these factors include ribonucleases which mediate cleavage of sRNAs from their associated mRNA transcripts and can further process sRNAs into different isoforms with varying regulatory

functions and stability [33,47]. A general overview of known regulatory processes that affect sRNA biogenesis is given below.

1.2.3.1 Regulation of transcription initiation

Transcription of sRNAs from a dedicated promotor is highly dependent on the promoter sequence and the associated transcription factor, as is the case for any transcriptional process [58]. Accordingly, regulation by specialized transcription factors and response regulators of two-component systems allows directed expression of sRNAs in response to specific stress-related or nutritional signals [33,58]. Indeed, strongly induced expression in response to specific stress conditions is a common feature of many sRNAs and is often indicative of potential related sRNA targets [33,47,59,60]. Prominent example are the MicA, RybB, and MicL sRNAs in *E. coli*, all of which are transcriptionally regulated by the alternative sigma factor RpoE (σ^E) as part of the envelope stress response [61–63]. Upon expression, MicA, RybB, and MicL post-transcriptionally downregulate more than 30 mRNAs including all major outer membrane proteins (OMPs), thereby preventing an accumulation of unassembled OMPs in the periplasm [61,63,64]. In addition, sRNAs may also control expression of their own transcription factor, thus generating a positive or negative feedback loop [65,66]. For example, transcription of the MicF sRNA is repressed by the transcription factor Lrp, a metabolic regulator that is upregulated during nutrient-poor conditions [67]. Simultaneously, MicF inhibits translation initiation of the *lrp* mRNA in nutrient-rich conditions, creating a double-negative feedback loop that supports adaptation to nutrient availability in *E. coli* [68].

1.2.3.2 Regulation of transcription termination

Transcription termination of sRNAs predominantly occurs in a Rho-independent manner through formation of a stable stem-loop structure that causes RNA polymerase (RNAP) pausing and subsequent dissociation [69,70]. Consequently, many sRNAs fold into a stable stem-loop followed by a poly-U stretch at their 3'end, a typical feature of Rho-independent terminators [70]. Interestingly, these features are of particular importance for Hfq-dependent sRNAs (also see 1.2.4) [46,71]. In fact, studies in *E. coli* revealed that hairpin-loop strength as well as the length of the adjacent poly-U stretch are critical for Hfq binding and thus sRNA function of Hfq-dependent sRNAs [46,71].

As discussed above, sRNAs can be encoded at almost any genomic location, including 5'UTRs and ORFs, raising the question of how transcription termination and thus expression of the sRNA as opposed to the full-length transcript is regulated [70]. Research investigating transcription termination of the SgrS and RybB sRNAs in *E. coli* suggested that stress conditions might impact transcription termination, although the exact mechanism remains elusive [72]. Furthermore, specific protein factors such as the transcription elongation factor NusA, which affects termination efficiency at Rho-independent terminators, have been proposed as potential regulators of this process [70,73,74]. More recently, Morita and colleagues identified three protein factors, *cspD*, *ygjH*,

and *rof*, and one sRNA, CyaR, as attenuation factors in a multicopy screen developed to evaluate read-through at the SgrS terminator [75]. Nevertheless, regulation and thus efficiency of transcription termination of sRNAs under various growth and stress conditions is poorly understood and needs to be addressed further in the future.

1.2.3.3 sRNA processing

Although sRNA expression is often dependent on dedicated TSSs and/or intrinsic Rho-independent terminators, sRNAs can also be processed at the 5' and/or the 3'end from the associated full length mRNA [33]. In these cases, transcription of the processed sRNA logically depends on transcription of the full-length precursor RNA and the associated promoter. Enzymatic processing usually results in a 5'-monophosphate, as compared to the characteristic 5'-triphosphate group of primary transcripts [76]. This is mechanistically significant, as the 5'-monophosphate end can allosterically activate RNase E-mediated degradation of sRNAs and sRNA-bound target RNAs [42,76,77]. Furthermore, some sRNAs exist in multiple isoforms wherein the short isoform is processed from a longer isoform with potentially different regulatory functions. These include the *Salmonella enterica* (*S. enterica*) sRNA SdsR that is transcribed from a σ^S-dependent promoter and cleaved by RNase E into a shorter isoform with a distinct target spectrum [78]. In Gram-negative species such as *E. coli* and *S. enterica*, biogenesis and processing of sRNAs has been extensively studied and is mostly performed by the single-strand specific endonuclease RNase E [51,79,80]. However, the degradation machineries of Gram-negative and Gram-positive bacteria vary remarkably, with RNase E being entirely absent in Gram-positive species [81,82]. Alternative mechanisms have been proposed, including cleavage mediated by the double-strand specific endonuclease RNase III, as shown for RsaC in *S. aureus* [41,83]. Nevertheless, mechanistic studies are scarce, and the process is significantly less well understood, calling for further exploration.

1.2.3.4 Factors affecting sRNA stability

Finally, sRNA stability and thus turnover is another important feature of sRNA-mediated regulation; understanding how sRNA levels are controlled is an integral part of understanding their regulatory capacity [84,85]. The process of sRNA degradation is complex and seems to vary depending on the condition and type of sRNA as well as the bacterial species[33,84,86]. Several RNases have been implicated in sRNA degradation, most notably RNase E, RNase III and PNPase [77,84,87]. In *E. coli*, the single-strand specific endonuclease RNase E and double-strand specific RNase III have been shown to co-degrade some Hfq-dependent sRNAs and their bound mRNA targets, while others are stabilized by Hfq binding [86,88,89]. PNPase on the other hand is a 3' to 5' exoribonuclease that recognizes natural hairpins followed by a single-stranded 3'end, a common feature of sRNAs and a consequence of Rho-independent transcription termination [90]. Although PNPase was found to bind both Hfq-associated as well as Hfq-independent sRNAs, RNase activity seems to be

restricted to unbound sRNAs that are not protected by Hfq binding [90,91]. Furthermore, several reports in *E. coli* indicate that PNPase may also have a stabilizing effect on some sRNAs, contradicting its RNase function [90,92].

More recently, sRNA sequestration by other sRNAs was reported to impact sRNA turnover [93]. An example from Gram-positive bacteria is RosA, an sRNA sponge described to interact with two additional sRNAs, RosX and FsrA, in *B. subtilis* [94]. While the interaction of RosA and the FsrA sRNA merely results in sequestration of FsrA, interaction with RoxS leads to destabilization and degradation of RoxS [94]. Accordingly, sRNA sponges not only prevent sRNAs from regulating their targets but can also result in degradation of the sRNA [93].

1.2.4 Target regulation by sRNAs

Trans-encoded sRNAs regulate their mRNA targets by direct base-pairing *via* a short sequence that is required for target recognition [95–97]. This so-called "seed sequence" typically involves 6-10 bases with partial complementary to the target mRNA [95,98]. Although longer seed sequences have been reported, only a few nucleotides appear to be critical for regulation [96,99]. As a result, a single sRNA can usually base-pair with multiple mRNA targets, using the same seed sequence [98,100,101]. In fact, Faigenbaum-Romm and colleagues demonstrated that mutating the sRNA seed region can result in a shift in bound mRNA targets [102]. Furthermore, some sRNAs such as the *E. coli* GcvB sRNA contain several seed sequences that interact with distinct mRNA targets respectively [103,104]. Due to the low sequence complementarity of these interactions, they are often supported by RNA chaperones, most notably Hfq (see 1.3) [84]. Depending on the sRNA, the mRNA, and the interaction site, the regulatory outcome can vary considerably, ranging from a positive or negative effect on translation to regulation of transcription termination [103]. A more detailed description of the most common regulatory mechanisms will be provided below.

1.2.4.1 Negative regulation of transcript stability and translation

The most prevalent mechanism of sRNA-mediated regulation is inhibition of mRNA translation by binding of the sRNA at or near the RBS and start codon region of an mRNA target (Figure 2A) [8]. In an attempt to define an "inhibitory window" for sRNA binding, Bouvier and colleagues performed systematic antisense scanning of the 5' CDS of diverse mRNAs in *S. enterica* [105]. Their results revealed that binding of an sRNA up to the fifth codon of the CDS can sterically interfere with binding of the 30S ribosomal subunits and thus with translation initiation [105]. Naturally, seed sequences targeting the RBS region are often C-rich and thus complementary to the G-rich Shine-Dalgarno (SD) sequence of prokaryotic mRNAs [106]. First reported for OxyS-mediated repression of the transcriptional activator *fhlA* in *E. coli*, variations of this C-rich seed sequence have since been found in sRNAs of multiple species, including the Gram-positive *L. monocytogenes* and *S. aureus* [107–109].

Once translationally inhibited by an sRNA, mRNAs are often subject to degradation by RNases [110]. Ribosome occupancy of an mRNA usually protects mRNAs from cleavage by RNases and affects mRNA turnover [111]. By binding to the RBS, sRNAs inhibit the assembly of new ribosomes, thus leaving the mRNA vulnerable to RNase-mediated degradation (Figure 2A) [111]. In addition, several experiments investigating sRNA-mediated decay of mRNA targets revealed that sRNAs in complex with the RNA chaperone Hfq can actively recruit RNase E and thus accelerate mRNA decay in Gram-negative species [76,112]. While not always the case, cleavage of the mRNA target can also lead to co-degradation of bound sRNAs [89,113]. However, most research focusing on sRNA-target degradation has been performed in Gram-negative model organisms, namely *E. coli* and *S. enterica.* Since ribonucleases encoded by Gram-negative bacteria differ considerably from those of Gram-positive species, any comparison is only possible to a limited extent [82].

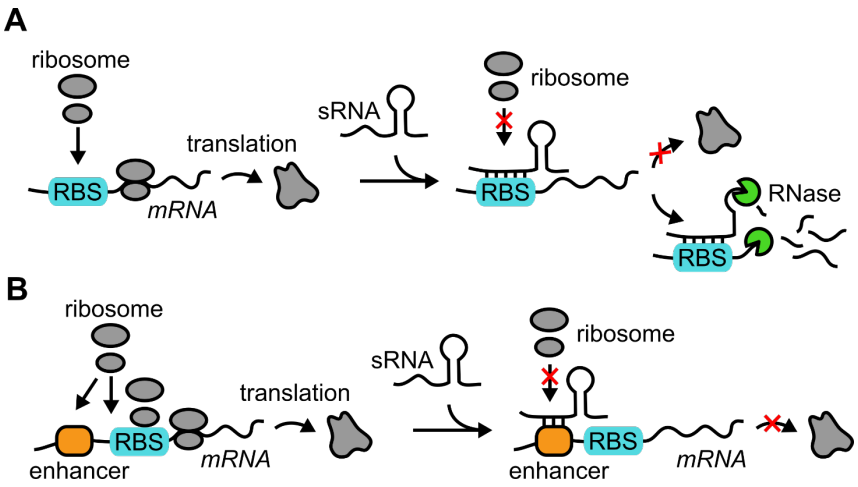


Figure 2: Common mechanisms of sRNA-mediated negative regulation of translation and transcript stability. (A) sRNA binding to the RBS and start codon region of mRNAs targets can inhibit ribosome binding and thus translation. Translation inhibition is often accompanied by RNase-mediated cleavage of the mRNA target and/or the base-pairing sRNA. **(B)** Similarly, sRNA binding to enhancer elements that serve to enhance ribosome binding to the RBS can result in inhibition of ribosome binding and consequently translation.

Interestingly, sRNA binding to the RBS region is not the only sRNA-mediated mechanism resulting in translational inhibition or transcript instability[114,115]. Additional, although less common mechanisms include sRNA binding further upstream in the 5'UTR, binding within the CDS or even the 3'UTR of an mRNA target [43,115]. For example, GcvB, a highly conserved sRNA in Gram-negative species, inhibits translation of the ABC-transporter GltI by binding to a translational enhancer element further upstream in the *gltI* 5'UTR [116,117]. A similar mechanism was recently reported for the *S. enterica* sRNA SgrS in inhibiting *manY* translation [118,119]. By sequestering the enhancer element, SgrS inhibits binding of the ribosomal subunit protein S1, which would usually render the SD sequence more accessible to ribosome binding and thus enhance translational efficiency

(Figure 2B) [118]. Examples of sRNA binding the mRNA CDS include MicC and SgrS, both of which bind the *ompD* mRNA downstream of the "inhibitory window" within the CDS [120,121]. Both interactions directly result in RNase E-dependent accelerated mRNA decay rather than inhibition of translation. Finally, there are some reports of sRNAs binding mRNA 3'UTRs with a negative effect on mRNA translation [43]. In *S. aureus*, interaction of the RsaI sRNA and the *icaR* 3'UTR negatively impacts translation of the transcriptional repressor IcaR, however the mechanism of regulation is not well understood [122]. Interestingly, the *icaR* 3'UTR contains an anti-SD motif that can bind the *icaR* 5'UTR-located SD sequence and thus inhibit translation [123]. Accordingly, RsaI could potentially contribute to this interaction by stabilizing the duplex formation [122].

1.2.4.2 Positive regulation of translation and transcript stability

Although less common, there is a growing number of sRNA-mRNA interactions that result in an increase in mRNA translation and/or stability. Similar to the mechanism described above, the regulatory outcome is usually based on altered accessibility to ribosome binding or RNase-mediated cleavage, albeit with opposing effects (Figure 3) [124]. As is the case for negative regulation, sRNAs targeting the mRNA 5'UTR are the most common and examples include the sRNA Rli27 in *L. monocytogenes*, the MicF sRNA in *E. coli* and the *S. enterica* sRNA RydC [125,126]. Rli27 interacts with the long 5'UTR of Lmo0514, a protein upregulated upon *L. monocytogenes* colonization of eukaryotic cells [125]. By binding to the *lmo0514* 5'UTR, Rli27 most likely liberates the *lmo0514* RBS that is otherwise occluded by an intrinsic inhibitory structure formed by the long *lmo0514* 5'UTR, thus resulting in increased protein levels of Lmo0514 (Figure 3A) [125]. More recently, Carrier and colleagues suggested that binding of the *E. coli* sRNA MicF to the *oppA* mRNA 5'UTR promotes translation initiation by rendering a translational enhancer element more accessible to ribosome binding [127]. Interestingly, this interaction seems to be dependent on the presence of GvcB and Hfq, even though Hfq-mediated GvcB-*oppA* binding has been reported to inhibit *oppA* translation. Finally, base-pairing of the *S. enterica* sRNA RydC stabilizes its target, the *cfa* mRNA, potentially by inhibiting RNase-mediated degradation rather than increasing mRNA translation [126].

Reports of sRNAs targeting CDSs or mRNA 3'UTRs with a positive effect on mRNA translation and/or stability are rare, however several examples have been published. In *S. typhimurium*, the aforementioned sRNA SgrS was reported to stabilize *yigL*, the second gene in a bicistronic operon, following initial RNase E-mediated cleavage in the upstream gene *pldB* [128]. By binding the cleaved mRNA intermediate within the remaining CDS of *pldB*, SgrS interferes with further 5' to 3' degradation by RNase E, leading to increased YigL production. Furthermore, a recent publication by Abdulla and colleagues reported the interaction of two sRNA, SdsR and Spot 42, with the 3'UTR of the *hilD* mRNA, a regulator of *Salmonella* pathogenicity [129]. In both cases, sRNA binding to the

310 nt long 3'UTR interferes with RNaseE-dependent degradation and thus results in increased *hilD* mRNA levels (Figure 3B).

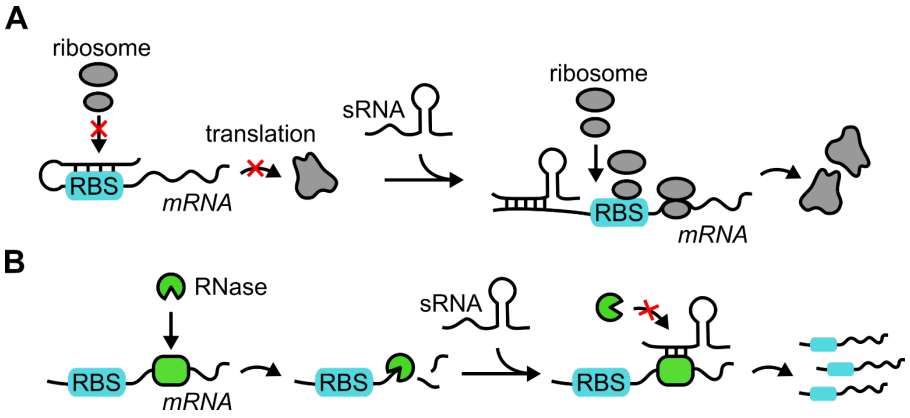


Figure 3: Common mechanisms of sRNA-mediated positive regulation of translation and transcript stability. (A) sRNA binding to the target mRNA can liberate the RBS that is otherwise occluded by an intrinsic inhibitory structure, ultimately leading to increased protein levels. **(B)** sRNA binding can interfere with RNase-mediated cleavage of mRNAs, resulting in increased mRNA stability.

1.2.4.3 Regulation of Rho-dependent transcription termination

Although sRNAs are predominantly known as post-transcriptional regulators, a number of publications reported an additional regularity function at the level of transcription by affecting Rho-dependent transcription termination [70,130]. Rho is an ATP-dependent hydrolase/RNA translocase in bacteria, that can destabilize the RNAP elongation complex following recognition of C-rich Rho utilisation (rut) sites in the nascent mRNA transcript and thus terminate transcription [131]. Actively translating ribosomes can inhibit this process by blocking Rho access to rut sites, as transcription and translation are dynamically coupled in bacteria [130,132]. While these events frequently occur at the 3'end of an mRNA, they are also relevant for transcription termination in long 5' leader sequences or differential expression of individual genes within a polycistronic operons (also known as discoordinate expression) [70].

By inhibiting ribosome binding and thus ribosome occupancy of the nascent RNA transcript, sRNAs can indirectly open up rut sites to Rho binding (Figure 4A) [70,130]. In turn, sRNA-mediated upregulation of translation can have the opposite effect, blocking Rho binding and subsequent transcription termination. This mechanism was first described in connection with the *S. enterica* sRNA ChiX that downregulates translation of the *chiPQ* mRNA by binding the *chiPQ* 5'UTR and blocking ribosome binding [133]. Additionaly, ChiX-mediated occlusion of the RBS exposes an intragenic rut site to Rho binding, resulting in premature transcription termination within the *chiP* CDS [133]. Furthermore, sRNAs can also operate as anti-termination factors by directly blocking Rho loading and/or translocation along the mRNA (Figure 4B) [130]. Sedlyarova and colleagues

reported that base-pairing of three Hfq-dependent sRNAs, DsrA, RprA, and ArcZ with the long *rpoS* 5'UTR in *E. coli* inhibits Rho-dependent transcription termination within the *rpoS* 5'UTR [134]. Considering the high number of Rho-dependent transcription termination events in long 5' UTRs in *E. coli*, they speculated that this mechanism might be more common than originally anticipated [134]. However, most research regarding Rho-dependent transcription termination has been performed in *E. coli* and *S. enterica* and comparatively little is known about Rho function in Gram-positive species [130].

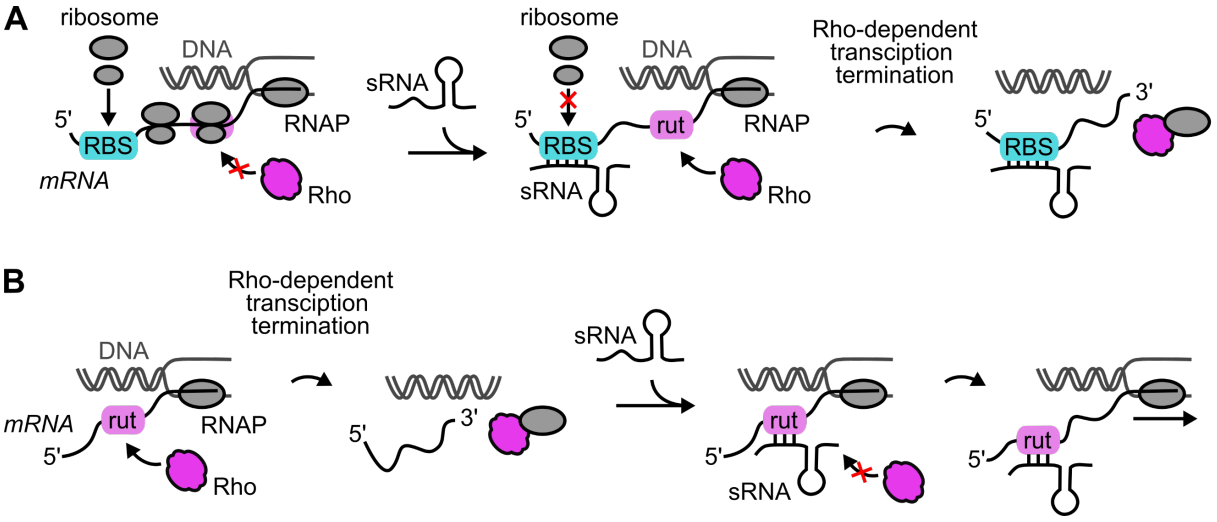


Figure 4: sRNA-mediated regulation of Rho-dependent transcription termination. (A) sRNA binding to the RBS and start codon region of mRNAs targets can inhibit ribosome binding. Consequently, translating ribosomes no longer shield rut sites from Rho binding, resulting in Rho-dependent transcription termination. **(B)** By binding in proximity of or directly to rut sties, sRNAs can inhibit Rho binding, resulting in anti-termination.

1.2.4.4 sRNAs and their function in regulatory networks

The past two decades of research on sRNA-mediated regulation have highlighted the importance of sRNAs in transcriptional networks, also referred to as regulons [135–137]. While regulons were historically described as a set of genes that are controlled by the same transcriptional regulator, it is now clear that many regulons also include one or more sRNAs and thus indirectly any mRNA target of the regulated sRNA [47,65]. In fact, sRNAs are now considered the non-coding arm (with proteins referred to as the coding arm) in many stress and virulence regulons, providing a rapid and cost effective response to external signals [47,136,138]. While transcriptional regulators exclusively affect transcription, sRNAs can supplement other regulatory processes by directly affecting translation and stability of (pre-)existing mRNAs [66,135,139]. Accordingly, sRNA-mediated positive regulation could potentially enforce expression of targets in a specific regulon, while negative regulation of mRNA translation and stability could promote the downregulation of genes that are disadvantageous in a given growth condition [47]. This is particularly advantageous in the

case of genes whose mRNAs have long half-lives and are thus still available for translation, even if transcription of the corresponding gene has been shut off [65].

Furthermore, transcriptional regulators, and as a consequence their regulon, can be a target of sRNA-mediated regulation themselves [65]. In fact, the same transcriptional regulator is frequently targeted by multiple sRNAs in response to varying signal inputs, creating so called "regulatory hubs" [136]. A well known example is $FlhD_2C_2$, the master regulator of flagella genes in *E. coli* that is targeted by five sRNAs (ArcZ, OmrA, OmrB, OxyS, and McaS) [140]. Expression of these sRNAs and thus base-pairing with the *flhDC* 5'UTR is dependent on various signals, including osmolarity, oxidative stress, and glucose availability, resulting in either negative or positive regulation of motility. Similarly, biofilm formation is a central target of sRNA-mediated regulation through base-pairing of seven sRNAs (OmrA/B, McaS, RprA, RydC, GcvB, and RybB) with the *csgD* 5'UTR, encoding the central regulator of curli formation in *E. coli* [141]. Expression of these sRNAs and thus *csgD* regulation depends on various environmental signals including osmolarity and carbon limitation. Interestingly, sRNAs can also regulate translation of their own transcription factor, thereby creating feedback loops that delay or promote a regulatory response [65,66,142]. Well-known examples are the *E. coli* sRNAs OmrA/OmrB, both of which repress translation of their own transcriptional activator OmpR, resulting in a direct negative feedback loop [143].

Considering that the same mRNA target can be regulated by multiple sRNAs, the question arises as to how certain interactions are prioritized over others. Several factors have been discussed to contribute to these "regulatory hierarchies", most notably "availability" of either sRNA or target mRNA. Accordingly, the transcription rate of a specific sRNA in relation to its target mRNA as well as a competing sRNA highly affects the regulatory success [65]. However, other factors, such as the presence of additional target mRNAs or sRNA sponges that sequester a specific sRNA (see 1.2.3.4), competition for RNA-binding proteins such as Hfq (see 1.3.2) as well as target affinity have been shown to affect the regulatory outcome of sRNA-mRNA interactions and thus contribute to these complex regulatory networks [66,144,145].

1.3 Hfq – an important factor of sRNA-mediated regulation

As discussed above, interactions between *trans*-acting sRNAs and their target mRNA generally occur *via* a short stretch of base-pairing nucleotides with partial complementarity (see 1.2.4) [2]. Consequently, many sRNA-mRNA interactions require the assistance of an RNA chaperone for successful base-pairing [84]. During the past two decades of research on sRNA-mediated regulation, several such RNA chaperones have been identified with varying function and target spectra, most notably Hfq, ProQ and CsrA [9,84,146,147].

CsrA is the founding member of the Csr/Rsm class of proteins that is conserved in various bacterial phyla and binds to GGA motifs within the loop of short hairpins [148,149]. Initially known for its function in carbon storage and glycogen production, CsrA is now recognized as a global regulator, influencing translation initiation, transcript stability, sRNA abundance and sRNA-target interaction [148,150]. ProQ was established as a global RNA chaperone in 2016 and is a FinO domain-containing protein commonly present in Proteobacteria [151–153]. In *E. coli* and *S. enterica* ProQ was found to be associated with a large suite of cellular transcripts and dozens of sRNAs, of which many are stabilized upon ProQ binding [152,154,155]. Interestingly, ProQ preferentially recognizes and binds double-stranded, highly structured RNA rather than a specific sequence motif [146,154,156]. Most recently, pull-down experiments in *S. pneumoniae* and *Clostridioides difficile* (*C. difficile*) added the KH domain proteins KhpA and KhpB to the list of sRNA-associate RBPs [157–159]. However, very little is known about the RNA binding mechanisms or the regulatory function of these interactions [159].

Among the known RNA chaperones named above, by far the most studied is the Sm-like protein Hfq, originally identified in 1968 as an *E. coli* host factor required for replication of the RNA phage Qβ [160,161]. It became a focus of research on post-transcriptional regulation during the early 2000s, when several reports indicated its association with known sRNAs in *E. coli* [16,161–163]. Following two decades of research, Hfq is now widely perceived as the core component of global post-transcriptional networks that facilitates the base-pairing interactions of sRNAs with *trans*-encoded target mRNAs [161]. Loss of Hfq not only compromises bacterial fitness in various stress conditions but also virulence [164,165]. In particular, Gram-negative pathogens such as *S. typhimurium*, *V. cholerae*, uropathogenic *E. coli* or *P. aeruginosa* are highly attenuated in infection models upon deletion of Hfq [164,166–169]. Despite its importance in these species, Hfq is not ubiquitous throughout the bacterial domain [170]. Some genera such as *Enterococcus*, *Fusobacterium* or *Bacteroides* neither encode Hfq, nor any of the other known RBPs despite the presence of a large post-transcriptional network [171–173]. Other bacteria, in particular Gram-positive species including *B. subtilis* and *S. aureus*, harbor Hfq, however its function remains enigmatic as sRNA regulation in these organisms largely occurs independent of Hfq [55,174–176]. Accordingly, most of the studies investigating Hfq function were conducted in *E. coli* or *S. enterica* and caution should be taken in generalizing these findings.

1.3.1 Hfq interacts with RNA at distinct interaction sites

To be functional, Hfq characteristically folds into a ring-like quaternary structure consisting of six identical protomers [177–179]. Each monomer encodes the relatively conserved bipartite Sm motif denoting it as a member of the Sm-like protein family, followed by a highly variable disordered C-terminus [178,180,181]. Fully assembled, the homohexameric Hfq complex interacts with RNA at four different sites: the proximal and the distal faces, the lateral rim region, and the C-terminal domain (CTD) that protrudes from the rim region (Figure 5A) [182,183].

The Hfq proximal surface contains the highly conserved interaction site that is ubiquitously present in Sm proteins and allows the sequence-specific interaction with U-rich single stranded RNA [184]. As described above (1.2.2.2), poly-U tails are a common feature of sRNA 3'ends that are transcriptionally terminated in a Rho-independent manner [71]. Accordingly, mutations in the proximal binding site of Hfq usually result in decreased sRNA function and stability [185,186]. Furthermore, research on the SgrA sRNA in *E. coli* indicates that the length of the poly-U stretch at the sRNA 3'end is important for successful Hfq binding [46,71]. While a length of seven or more uridines resulted in a stable interaction, shortening the poly-U tail below four uridines completely abolished Hfq binding and thus sRNA function [46,71].

In contrast to the proximal surface, the distal surface is relevant for mRNA binding and interacts with single-stranded A-rich sequences [185,187] However, the actual sequence length and composition recognized by Hfq seems to vary between species. In Proteobacteria, Hfq typically binds a triplet (AAN)x motif with 2-4 repeats, although 6 repeats can be accommodated on the ring-shaped hexamer [187,188]. In contrast, experiments performed in *S. aureus*, *B. subtilis* and *L. monocytogenes* suggest that Hfq in Firmicutes recognizes an (AN)x motif with up to 12 repeats [188–191]. This interaction can also result in rearrangements of the mRNA secondary structure, rendering it more accessible to sRNA binding [192–194].

While the proximal and distal face are important for sRNA and mRNA binding, semi-conserved arginine patches on the lateral rim were reported to interact with the sRNA "body", including the seed region [186,195,196]. Consequently, the sRNA is positioned in proximity to the mRNA target, thereby accelerating and facilitating the RNA-RNA interaction [186,195,196]. Interestingly, the amino acid composition of the Hfq rim region seems to correlate with the reported role of Hfq in sRNA-mediated regulation [197]. While Hfq homologs from Gram-negative species such as the *E. coli* Hfq encode up to three arginines, this number is lower in species where Hfq is reportedly less important for sRNA-mediated regulation [197].

Although the general mode-of-action described above applies to most sRNA-mRNA interactions mediated by Hfq, this is not universally true for all sRNAs. Mutation studies investigating interactions at the distal face and lateral rim in *E. coli* found that some sRNAs bind both, proximal and distal face, while their mRNA target interacts with the rim region [86,186]. Based on these binding properties sRNAs were classified in two classes, I and II, wherein class I sRNAs interact with the proximal site and lateral rim, and the less common class II sRNAs with the proximal and distal face of Hfq (Figure 5B) [86].

Among the four binding sites of Hfq, the disordered C-terminus is the least conserved and varies considerably across species in length as well as in amino acid composition [178]. In *E. coli,* the tip of the Hfq CTD is characterized by an enrichment of acidic residues. According to a model published

by Santiago-Frangos and colleges, these residues can transiently bind the basic residues located at the rim region and thereby actively compete against nucleic acid binding [182,198]. These interactions compete with non-specific RNA binding and an accelerated release of double-stranded sRNA-mRNA pairs, promoting a rapid cycling of RNA on Hfq [198–200]. This model was expanded by a more recent study which tested Hfq function upon combining C-terminal length mutants with mutations in the remaining RNA binding sites [201]. Based on the updated model, an arginine residue located within the CTD contacts the Hfq distal face, in addition to the rim interaction of the acidic tip, with various effects depending on the individual sRNA [201]. In either case, these models discuss the C-terminal function of the *E. coli* Hfq and should not be generalized, given the high amino acid divergence in Hfq C-termini across species [182].

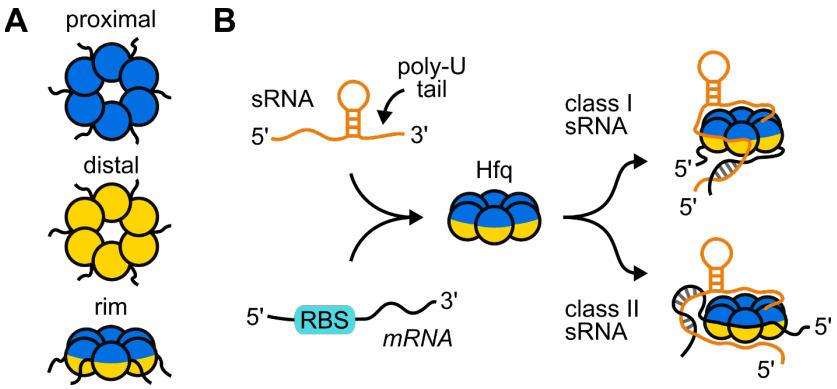


Figure 5: Hfq interacts with RNA at distinct interaction sites. (A) Hfq forms a homohexameric ring with four distinct interaction sites: proximal, distal and rim surfaces, as well as the CTD that protrudes from the rim region. **(B)** Depending on the sRNA class, Hfq binds sRNAs and their target mRNAs at different sites. While class I sRNAs bind the proximal and rim sites of Hfq and their mRNA targets the distal site, class II sRNAs bind the proximal and distal surfaces of Hfq, and their mRNA targets the rim region.

1.3.2 sRNAs compete for Hfq binding

Given the large number of reported sRNA-mRNA interactions facilitated by Hfq, it became apparent early on that Hfq availability must be a limiting factor for Hfq-dependent sRNA-mediated regulation in bacterial cells [161,202,203]. In *E. coli*, Hfq strictly autoregulates its own translation to maintain its cellular concentration [204]. Obviously, this raises the question of how Hfq availability is controlled to allow efficient, fast acting regulatory responses. Experiments in *E. coli* have suggested a combination of RNA expression and Hfq binding affinity as potential factors that lead to prioritized binding of one RNA over another [102,205,206]. Rapid dissociation of non-specific RNA and double-stranded sRNA-mRNA pairs has been discussed as another important aspect which contributes to Hfq availability [200,207]. As mentioned above, the Hfq CTD has been implicated in the

latter by actively displacing bound RNAs from the hexameric ring, resulting in dynamic RNA-cycling on Hfq [198–200].

Recently performed live cell imaging experiments in *E. coli* have further contributed to these hypotheses and revealed that Hfq is predominantly bound by mRNA under normal growth conditions [208]. This study also indicated that Hfq could recruit RNase E to degrade bound mRNA transcripts independent of any sRNA interaction, thereby recycling the mRNA binding site. Additional interactions with sRNAs resulted either in active replacement of the mRNA or simultaneous binding of both RNAs *via* distinct Hfq binding sites. The outcome was highly dependent on the sRNA type (class I or class II sRNA) and confirmed previously published data regarding RNA interaction at the different Hfq binding surfaces (see 1.3.1). Similar results were obtained by Roca and colleagues, using a single-molecule fluorescence platform to visualize simultaneous binding of multiple RNAs to Hfq [209]. Once bound, sRNAs rarely dissociate autonomously, but were rather actively replaced when challenged by another sRNA on the same binding face of Hfq, leading to a rapid exchange of bound RNAs.

1.3.3 Additional regulatory functions of Hfq

Although best known for its function in promoting base-pairing between sRNA and their target mRNAs, Hfq has been implicated in additional regulatory mechanisms [210]. Several publications have reported that Hfq can directly inhibit mRNA translation and stability in an sRNA-independent mechanism by binding mRNA 5'UTRs and blocking ribosome binding [211,212]. Examples include the interactions of Hfq with the 5'UTR of its own transcript as well as the *mutS* and *cirA* 5'UTR in *E. coli* [213–215].

Besides sRNAs and mRNAs, Hfq was also found to interact with rRNA, tRNA and DNA [210]. Accordingly, recent work in *E. coli* suggested that Hfq binds rRNA and, in cooperation with RNase R, is required for rRNA maturation, quality control and ribosomal assembly with secondary effects on translation efficiency and fidelity, effectively making it an important ribosome biogenesis factor [216,217]. Similarly, Hfq was found to bind tRNAs in *E. coli*, potentially affecting tRNA modification and thereby also contributing to translational fidelity [218]. Furthermore, DNA was found to interact with the Hfq CTD, thereby inducing self-assembly of Hfq into amyloid-like structures with potential effects on DNA compaction [219].

In addition to nucleic acid substrates, Hfq was repeatedly shown to interact with other proteins, most importantly, proteins involved in RNA degradation and transcription [210]. As mentioned above (1.2.3.4), RNase E and PNPase, both members of the RNA degradation machinery, have been reported to interact with Hfq. In both cases, complex formation impacts stability of Hfq-bound RNAs [84,90,208,220]. Lastly, Rabhi and colleagues demonstrated that Hfq can physically interact with

the transcription termination factor Rho. This interaction impairs Rho-binding to additional transcription termination factors and thus its terminating function [221].

In summary, these examples illustrate the wide variety of Hfq regulated processes and could potentially explain why some bacterial species express Hfq with seemingly no effect on sRNA-mediated post-transcriptional regulation [55,174,175].

1.4 Identifying sRNA-mRNA interactions

To fully understand the function of sRNAs in bacterial cell physiology, it is essential to both identify mRNA targets and characterize the regulatory implications of their target interactions. Although a vast number of sRNAs have been discovered in various species (see 1.1), only a fraction of those have been functionally characterized to this date. In addition, most research regarding sRNA-mediated regulation has been focused on the Gram-negative model species *E. coli* and *S. enterica*; comparatively little is known about the characteristics and function of sRNA in Gram-positive bacteria [147].

1.4.1 sRNA-centric investigation of sRNA-mRNA interactions on a global scale

Early attempts at identifying sRNA-target interactions were based on bioinformatic predictions followed by experimental confirmation of individual interactions using genetic and biochemical approaches [222]. Later on, the development of several micro-array-based techniques enabled global searches for sRNA-mRNA interaction partners until they were eventually replaced by modern genome-wide RNA-seq-based methods [23]. Accordingly, one of the most straight forward-approaches encompasses pulse-expressing an sRNA of interest for a short duration followed by RNA-seq and transcriptome analysis. As a result, transcriptome changes are ideally primarily the result of direct target interactions rather than secondary effects that might arise from prolonged high sRNA levels [23]. Alternatively, bait-based techniques such as MAPS (MS2-affinity purification coupled with RNA sequencing) and GRIL-seq (global small non-coding RNA target identification by ligation and sequencing) can be applied to identify direct mRNA targets of a specific sRNA [23]. For MAPS, an MS2-tagged sRNA is expressed *in vivo*, followed by cell lysis, affinity chromatography and sequencing of bound RNA species [223]. Although very similar to MAPS, GRIL-seq involves the co-expression of T4 RNA ligase, resulting in the ligation of the sRNA and its targets prior to cell lysis [224]. However, each method described above is limited to the analysis of a single sRNA and its target spectrum rather than the whole range of sRNA-mRNA interactions in a bacterial cell.

1.4.2 RBP-centric investigation of sRNA-mRNA interactions on a global scale

An important step forward in characterizing sRNA-target networks on a global scale was the discovery of RNA chaperones such as Hfq (see 1.3) and their subsequent use for co-

immunoprecipitation approaches [23]. In its most basic version, termed RIP-seq (RNA immunoprecipitation followed by RNA-seq) the RBP of interest is tagged, and expressed *in vivo*, allowing immunoprecipitation, purification and sequencing of bound RNAs (Figure 6) [225]. Several modifications including UV-crosslinking to covalently link protein-RNA interactions as well as RNase treatment of RBP bound species further refined the technique and was termed CLIP-seq (cross-linking and immunoprecipitation followed by deep sequencing) [24]. Due to the covalent linkage between proteins and RNA, very stringent purification conditions can be applied, resulting in significantly reduced background signals. RNase treatment prevents co-purification of additional cross-linked proteins and results in RNase digest of all, but the RBP protected RNA fragments, allowing identification of the RBP binding site. Although effective at capturing RBP-

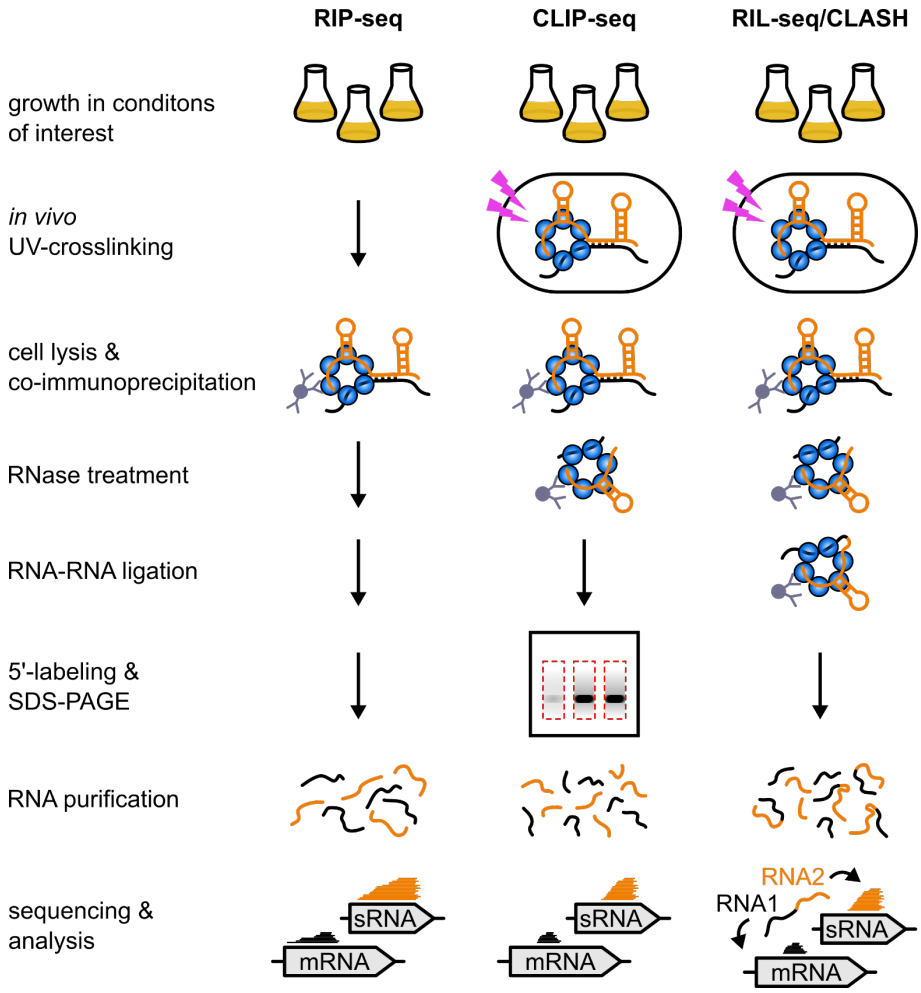


Figure 6: RBP-centric methods for identifying sRNA-mRNA interactions. RIP-seq employs expression of a tagged RBP of interest, followed by immunoprecipitation, purification and sequencing of bound RNAs. In addition, CLIP-seq employs UV-crosslinking to covalently link protein-RNA interactions as well as RNase treatment of RBP bound species to reduce background signals and narrow down the RBP binding site. Furthermore, RIL-seq and CLASH allow capturing of direct sRNA-target interactions through proximity ligation of RBP-bound RNA-RNA interactions.

bound sRNAs and mRNAs in a given condition, both RIP-seq and CLIP-seq fail at identifying direct sRNA-mRNA interaction partners. RIL-seq (RNA interaction by ligation and sequencing) and CLASH (UV cross-linking, ligation, and sequencing of hybrids) overcome this limitation by adding a ligation step following immunoprecipitation [50,226]. Accordingly, RBP-bound RNAs are ligated in a proximity-dependent manner, forming chimeric RNA-RNA duplexes that once sequenced, map to distinct genomic loci, revealing direct RNA interactions. Both methods are fairly similar, however, CLASH uses a stringent purification protocol, while RIL-seq is performed in native conditions and instead employs a more elaborate statistical analysis following sequencing.

1.4.3 Global, bait-independent investigation of sRNA-mRNA interactions

While all of the methods described above are effective at identifying sRNA targets on a global scale, they are restricted to either target interactions of a specific sRNA or RBP-associated interactions. Nonetheless, several methods have been developed recently that focus on identifying RNA-RNA interactions independent of a specific sRNA or RBP. For example Hi-GRIL-seq (high-throughput GRIL-seq), a variation of the aforementioned GRIL-seq, applies sequencing of the whole RNA-pool and thus any RNA-RNA hybrid following rRNA depletion, rather than enriching a specific sRNA [227]. Alternatively, a modified CLASH approach employs AMT (4'-aminomethyl trioxsalen) cross-linking of RNA-RNA interaction *in vivo* followed by ligation of interacting RNAs, thereby enabling RNA-seq of hybrid molecules on a genome-wide scale [228]. Finally, several studies have highlighted the advantage of using combinational approaches instead of relying on a single method to gain a more comprehensive understanding of sRNA-dependent regulatory networks [229]. In particular integrating computational analysis with multiple experimental datasets can enhance the specificity and add valuable information that might otherwise be missed [229].

1.5 *Clostridioides difficile* – an introduction

C. difficile is a Gram-positive obligate anaerobe and spore forming opportunistic pathogen. Since its initial discovery as the causative agent of antibiotic-associated pseudomembranous colitis in 1978, *C. difficile* has become the major cause of nosocomial antibiotic-associated infection in the developed world [230–232]. According to a recent report from the Centers for Disease Control and Prevention (CDC), *C. difficile* caused more than 224,000 infections (approx. incidence rate of 69 hospitalised CDI cases per 100,000 people), 13,000 deaths, and estimated healthcare costs of $1 billion in 2017 within the United States alone [230]. In Germany, the incidence rate within the same timeframe was even higher (approx. 105 cases per 100,000 people) with 87,000 infections and 1700 deaths, highlighting the need for new treatment options [233].

1.5.1 The *C. difficile* infection cycle

C. difficile transmission and infection usually occurs *via* the fecal-oral route through the ingestion of *C. difficile* spores [234]. In contrast to *C. difficile* vegetative cells, metabolically dormant spores are highly resistant to environmental stressors and antimicrobials [235]. Historically, *C. difficile* infection (CDI) is regarded as a nosocomial infection primarily occurring in healthcare settings through patient-to-patient transmission of spores, contaminated equipment, or healthcare personnel. However, community acquired infections have been increasingly reported as well, with potential *C. difficile* sources ranging from asymptomatic carriers and outpatient care facilities to food, animals or environmental reservoirs (Figure 7) [236]. Once ingested, *C. difficile* spores transit through the stomach into the small intestine where spores can germinate under favourable conditions (see 1.5.3) and eventually disseminate to the colon [234,237]. Stress induced toxin secretion by *C. difficile* vegetative cells in the colon then triggers a complex cascade of host cellular responses that eventually result in clinical manifestation of CDI [234]. Symptoms can range from asymptomatic carriage or mild diarrhoea to severe fulminant life-threatening colitis including toxic megacolon, bowel perforation, and sepsis [234]. Importantly, the same stress factors that trigger toxin production also induce endospore formation. This process marks the final step of the CDI cycle, resulting in excretion of *C. difficile* spores into the environment (Figure 7).

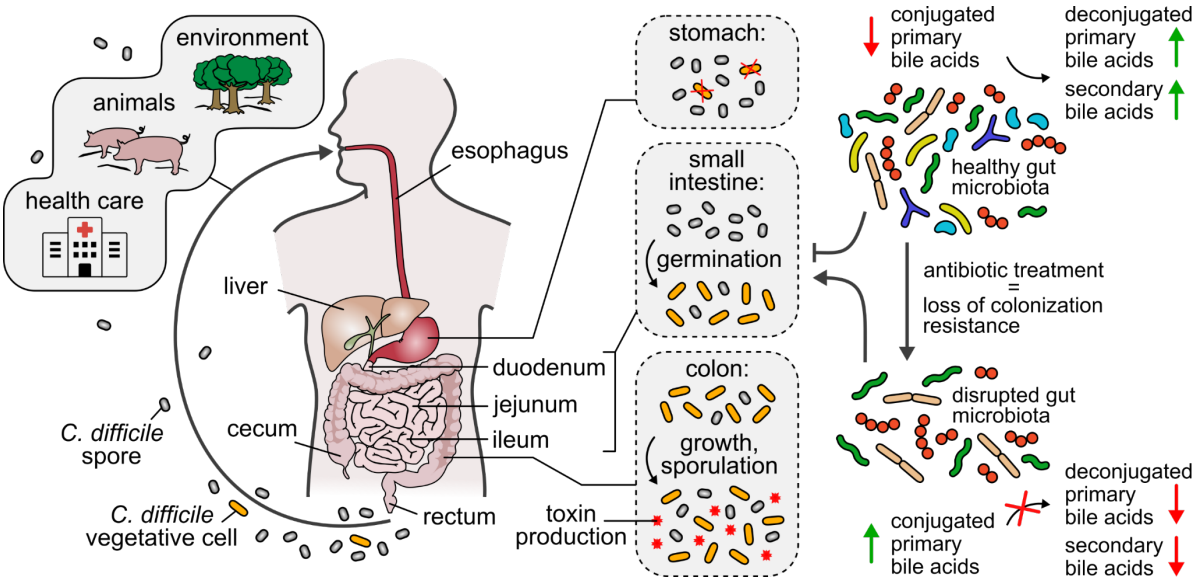


Figure 7: The *C. difficile* infection cycle. *C. difficile* transmission occurs through ingesting of spores that can be acquired from a number of sources including healthcare facilities, asymptomatic carriers, animals or the environment. In contrast to vegetative cells, spores are highly resistant to environmental stressors such as atmospheric oxygen and can survive the acidic environment in the human stomach. While a healthy gut microbiota usually confers colonization resistance to *C. difficile*, antibiotic treatment can disrupt the gut microbiota, creating conditions that favour *C. difficile* spore germination in the small intestine and dissemination of vegetative cells. Following colonization of the colon, *C. difficile* vegetative cells can produce toxins and eventually spores that are excreted by the patient during infection, restarting the infection cycle.

1.5.2 CDI risk factors and treatment options

Several risk factors that have been associated with the development of CDI, including advanced age (>65 years), the use of proton pump inhibitors, various underlying diseases, immunosuppression as well as previous hospitalization [238]. However, the single most important risk factor remains antibiotic administration, in particular treatment with broad-spectrum antibiotics that reach high concentrations in the gut [238]. While a healthy gut microbiota usually confers resistance to *C. difficile* colonization (see 1.5.3 for a more detailed definition), antibiotic administration can perturb this balance resulting in dysbiosis of the endogenous intestinal microbiota [239]. Due to the ensuing loss of colonization resistance, *C. difficile* can germinate and proliferate in the human intestinal tract, ultimately leading to CDI (Figure 7) [239].

Ironically, antibiotics are still the first-line treatment option for CDI, most notably vancomycin, metronidazole, and more recently fidaxomicin [240,241]. Although potentially successful, antibiotic treatment is also associated with increased drug resistance and disease relapse [242]. In fact, chances of a CDI recurrence within 8 weeks after recovery are 20-30% and increase with every new round of infection, a condition termed recurrent CDI [242]. A recently developed alternative to antibiotic therapy is faecal microbiota transplantation (FMT). In contrast to antimicrobials, this method is based on the collection and processing of faecal material from healthy donors that is then administered into the recipient's intestinal tract [241]. Live biotherapeutic products that consist of defined bacterial consortia and thus avoid unknown risk elements associated with donor samples are also being investigated [243]. Rather than directly targeting *C. difficile*, both approaches focus on re-establishing colonization resistance. FMT in particular has been applied with considerable success to patients with recurrent infections (70% - 90% success rate) [241,244]. Additional therapeutic approaches that are being developed include phage therapy, antibody therapy and novel small molecule antimicrobials. While there are FDA approved FMT (Rebyota) and monoclonal antibody (Bezlotoxumab) products, there is no live biotherapeutic product or phage therapy option available yet [245,246].

1.5.3 The gut microbiota and beyond - factors that impact CDI

As discussed above, ingestion of *C. difficile* spores does not necessary result in CDI. In fact, the human immune system and most importantly the gut microbiota usually prevent CDI, a phenomenon coined “colonization resistance” [237,247,248]. The ability of a healthy and diverse microbiota to inhibit colonization and overgrowth by invading microbes was first described in the 1960s and became a major research focus soon after [249,250]. It has since been discovered that colonization resistance is dependent on a wide range of factors including direct competition between bacteria for physical space and available nutrients in the gut [250]. Furthermore, the innate and adaptive immune response provided by the host also limit colonization by *C. difficile* through secreting antimicrobial molecules or restricting nutrient availability [250]. Of note, none of these

factors are necessarily specific to *C. difficile* and will most likely also affect other aspects of both, host-microbe and microbe-microbe interactions in the human gut.

One of the most important classes of molecules that not only shape the human gut microbiota, but also significantly affect *C. difficile* germination and growth are bile acids (BAs) [237,251]. Depending on their modification status, BAs are usually grouped into primary and secondary BAs. Primary BAs are produced from cholesterol in the liver and conjugated with taurine or glycine. Conjugated primary BAs are then secreted into the intestine to promote fat and cholesterol absorption during digestion, where approximately 95% are eventually reabsorbed [252]. The remaining proportion however, is deconjugated and potentially transformed into secondary BAs through dehydroxylation [251,252]. Both transformation steps are mediated by microbiota-derived enzymes in the small intestine and, to a larger extent, the colon [251,252]. In addition, recently discovered amino acid conjugations of BAs indicated that some members of the microbiota are able to conjugate BAs, further expanding the already complex pool of known BAs [253,254]. Generally, microbiota-derived chenodeoxycholic acid-based deconjugated primary BAs and secondary BAs have an inhibitory effect on *C. difficile* germination and/or growth [237,255]. Furthermore, several of the newly discovered microbial conjugated BAs seem to have a negative impact on *C. difficile* colonization [256]. In contrast, host-derived cholic acid-based primary BAs, both conjugated and unconjugated, seem to act as a germinant for *C. difficile* spores [237,257]. Concordantly, several landmark studies revealed that higher levels of microbiota-derived secondary BAs, as found in a healthy gut environment, protect against CDI [237,258–260]. In turn, antibiotic-induced dysbiosis of the gut microbiota and the resulting higher levels of conjugated primary BAs correlate with a *C. difficile* susceptible state [237,251,258–260].

Besides secondary BAs, several microbial-derived short chain fatty acids (SCFAs), such as butyrate and acetate, have also been associated with CDI [261,262]. Fachi and colleagues demonstrated that microbiota-produced butyrate protects the host from CDI-mediated damage by boosting the expression of tight junctions and consequently improving intestinal barrier function in infected mice [263]. Similar to butyrate, acetate has been shown to enhance the host immune response by promoting the activation of neutrophils and ILC3s in a mouse model of CDI [264].

In addition to BAs and SCFAs, nutrient availability is another important aspect that affects colonization resistance to *C. difficile* [237,265–267]. As a member of the amino acid fermenting *Clostridia*, *C. difficile* mainly relies on Stickland metabolism for energy production. Accordingly, availability of amino acids needed for the oxidative and reductive branches of Stickland metabolism has been associated with CDI [268–270]. Furthermore, *C. difficile* can use the Wood-Ljungdahl pathway as an alternative electron sink in the absence of amino acid Stickland acceptors to support the fermentation of carbohydrates; a mechanism that allows *C. difficile* to adapt to decreasing nutrient availability during CDI [265,271].

Additional nutrients that have been implicated in CDI are host membrane-derived ethanolamine and carbohydrates such as sorbitol, trehalose, mannose and sialic acids [265,272–277]. In particular sorbitol metabolism is tightly connected with *C. difficile* virulence [273]. A study published by Pruss and Sonnenburg demonstrated that *C. difficile* uses both, diet and host-derived sorbitol to support its growth. Moreover, *C. difficile* also induces sorbitol production and release by the host through toxin-mediated inflammation in a mouse model of CDI [273]. Interestingly, haem is another host-derived metabolite that is exploited upon toxin-mediated inflammation by *C. difficile* [278,279]. However, in contrast to sorbitol, *C. difficile* utilizes haem as a protective measure rather than a nutrient source, by incorporation it into an oxidative stress defence system [278,279].

Considering the intimate connection between nutrient availability and *C. difficile* colonization, it is not surprising that members of the microbiota competing for the same nutrients can inhibit CDI [269,280]. Conversely, microbes that produce nutrients with a growth-promoting effect on *C. difficile* increase virulence of the pathogen [277,280,281]. Two examples are *Paraclostridium bifermentans* (*P. bifermentans*) and *E. faecalis* [269]. Using defined co-colonization experiments with *C. difficile* and *P. bifermentans* in gnotobiotic mice, Girinathan and colleagues demonstrated that *P. bifermentans* can reduce CDI disease severity [269]. As an amino acid fermenting species, *P. bifermentans* depletes nutrients that are preferred by *C. difficile*, thus partially protecting against CDI [269]. In contrast, Smith and colleagues reported a positive correlation between *C. difficile* pathogenesis and the enrichment of Enterococci in multiple mouse models of infection and patients with CDI [281]. According to their results, *E. faecalis* enhances *C. difficile* fitness *via* cross-feeding of ornithine, a fermentable amino acids catabolized by *C. difficile* [281].

In summary, these examples highlight the multitude of factors in the human gut that, when taken together, tip the scale towards a state that either inhibits or favours CDI. However, many aspects remain unknown, as the complexity of these factors also impede our understanding of the relationship between *C. difficile* and the gut environment during CDI.

1.5.4 Virulence factors and traits that contribute to CDI

C. difficile harbours various virulence factors and traits that influence disease progression and severity [245]. Combined with the extraordinary genome plasticity exhibited by *C. difficile*, these factors create a highly diverse landscape of *C. difficile* isolates, that vary considerably in their epidemic and virulence potential [282]. So far, the most commonly used typing technique to differentiate *C. difficile* isolates remains PCR ribotyping; a method that allows clustering of *C. difficile* isolates into “ribotypes” (RT) based on variations in the intergenic spacer region of the 16S-23S rRNA [282]. Although the geographic distributed of RTs is quite diverse, some RTs dominate depending on the region (RT027 & 014-020 in America & Europe, RT017 in Asia and Middle East, RT084 in Africa, and RT009 in Oceania) [282]. In particular, RT027, which first emerged in a severe

outbreaks of CDI in North America and Europe in 2003, has been subject to many scientific studies due to its hypervirulent nature [283,284]. Features that affect overall virulence include variations in toxin production, sporulation rates, and antibiotic resistances and will be discussed below [245].

The primary mediators of CDI-induced inflammation are the large clostridial toxins, toxin A (TcdA) and toxin B (TcdB). Both toxins are encoded on the pathogenicity locus (PaLoc) in *C. difficile*, along with the sigma- and anti-sigma factors *tcdR* and *tcdC,* which regulate toxin expression, as well as the holing like protein *tcdE* [245,285,286]. Activity of TcdR and thus transcription of *tcdA* and *tcdB* is modulated by various regulators (e.g. Ccpa, CodY, PrdR, RstA & Rex) in response to environmental cues, most notably nutrient availability [287]. In contrast, mechanistic details of TcdC function are still missing. However, mutations in *tcdC* have been associated with increased virulence, most likely due to abrogation of negative regulation of the PaLoc locus [245]. While TcdR and TcdC regulate toxin transcription, TcdE is involved in secretion of TcdA and TcdB [287]. Once secreted, both toxin can be internalized by gut epithelial cells, where they glycosylate and thus inactivate small Rho GTPases, ultimately resulting in loss of intestinal barrier function and cell death [286]. However, *C. difficile* displays a considerable toxin heterogeneity, ranging from strains that encode only one toxin or a truncated version to strains that are completely nontoxigenic by lacking the whole PaLoc locus [270]. In particular, TcdB is highly variable with considerable effects on receptor specificity, translocation ability, inflammatory responses and pathological outcome [286]. Interestingly, approximately 20% of *C. difficile* strains express a third toxin, known as *C. difficile* transferase or binary toxin (CDT), encoded by the CDT Locus (CDTloc) along with *cdtR*, a regulatory protein [286,288]. In contrast to the glycosyltransferases TcdA and TcdB, CDT functions as a ADP-ribosyltransferase and catalyses the depolymerisation of actin upon endocytosis by host cells [286,288]. Although expression of CDT has been linked to increased disease severity, the role of CDT in CDI remains controversial [288]. In addition to toxin production, several surface proteins serve as virulence factors by facilitating *C. difficile* adherence to the gut epithelium and activation of the immune response [245]. Examples include the surface layer protein SlpA as well as various cell wall proteins, such as Cwp2, Cwp66, Cwp19, Cwp22 and Cwp V [245].

Resistance to multiple antibiotics is another common feature of many epidemic isolates and has been shown to significantly contribute to *C. difficile* pathogenesis and spread [289]. As discussed above, *C. difficile* is a genetically highly diverse species, shaped by horizontal gene transfer and large-scale recombination events [290,291]. This is also reflected in the frequent acquisition of genetic elements or alterations of antibiotic target sites that result in gain of antibiotic resistances by *C. difficile* [292]. In addition, biofilm formation by *C. difficile* has been associated with antimicrobial resistance [292]. In fact, biofilm formation has gained recent interest as a potential virulence trait of *C. difficile*, not only because of its role in antimicrobial resistance, but also as a factor that contributes to recurrent CDI [293]. Similar to other biofilm forming bacteria, the switch between a

motile and sessile lifestyle in *C. difficile* is regulated by the second messenger molecule c-di-GMP [294]. Additional known regulators of biofilm formation in *C. difficile* partially overlap with those for toxin production and sporulation and include CcpA, CodY, RstA, Spo0A and SigL/RpoN [294]. However, similar to the previously discussed virulence factors, biofilm formation is highly strain dependent and the role of biofilm in the CDI cycle has yet to be identified [294].

Lastly, the ability to form endospores is a crucial virulence trait enabling *C. difficile* transmission and survival in hostile conditions, such as exposure to atmospheric oxygen, the acidic stomach environment or treatment with antimicrobials [245]. Accordingly, mutants defective in spore formation are incapable of persisting within or transmission between hosts [295]. In addition, Castro-Córdova and colleagues recently demonstrated that *C. difficile* spores can enter intestinal epithelial cells, thereby contributing to recurrent CDI [296].The complex process of spore formation is generally conserved in sporulating Firmicutes and is centred on the integration of environmental and nutritional signals *via* the response regulator Spo0A [295,297]. However, *C. difficile* lacks many of the known conserved regulatory mechanisms that activate Spo0A, including the phosphorelay system that integrates environmental signals to active Spo0A in the model organism *B. subtilis* [298]. Nevertheless, several regulators of sporulation initiation have been identified, including the aforementioned catabolite control protein CcpA and the branched-chain amino acid sensor CodY [299]. In addition, three orphan histidine kinases (PtpABC), the small phosphatase Spo0E and the YicC-like protein CD25890 (CDIF630_02843) can negatively impact Spo0A function, although the exact mechanisms remain to be identified [299]. SigH, SinR and RstA are the only known positive regulators of Spo0A, with SigH regulating *spo0A* transcription and the latter two acting on Spo0A function *via* unknown mechanisms [299]. As mentioned above, several of these regulators are also involved in regulation of toxin production and motility, highlighting the tightly linked control of sporulation motility and toxin production in *C. difficile*.

Once the initiation decision has been made, phosphorylated Spo0A (Spo0A-P) transcriptionally regulates a set of early sporulation genes. Activation of these genes induces the asymmetric division of a cell into a larger mother cell and smaller forespore [300]. This is followed by engulfment of the forespore by the mother cell, resulting in two membranes that encase the forespore. A thick layer of peptidoglycan (cortex) between those membranes, as well as outer layers of coat proteins and exosporium add to the protection of the newly forming spore [301]. Once all layers are completed, the mother cell lyses, releasing the mature and metabolically inactive spore into the environment. Each of these steps is tightly coordinated and executed in a sequential manner governed by four compartment specific sigma factors: sigma factors, σ^E, σ^K (mother cell) σ^F, and σ^G (forespore) [300,302,303]. Germination and thus transformation into metabolically active cells is initiated in response to germinants (e.g. secondary BAs) indicative of an environment conductive to vegetative cell survival and growth [304].

1.5.5 sRNA-mediated regulation in *C. difficile*

Ongoing research in *E. coli*, *S. enterica* and other species has solidified the importance of sRNAs and their protein chaperones in mediating adaptive processes including stress response, host colonization and infection [1,47]. However, studies focused on sRNA-mediated regulation in Gram-positive pathogens including *C. difficile* have only recently gained traction [147]. Early computational and RNA-seq/dRNA-seq-based approaches in *C. difficile* identified >100 putative sRNAs, hinting at an extensive post-transcriptional network [305–307]. Furthermore, depletion of the RNA-chaperone Hfq profoundly impacts gene expression and bacterial physiology in *C. difficile*, contrary to reports from other Gram-positive species [174,175,308,309]. Boudry and colleges demonstrated that depletion of Hfq in *C. difficile* results in altered growth on various carbon and nitrogen sources, decreased stress tolerance and increased sporulation and biofilm formation [308]. In addition, Caillet and colleagues reported that *C. difficile* Hfq can largely complement an *E. coli* Δ*hfq* mutant, indicating similarities in chaperone function [310]. Nevertheless, very little is known about sRNA-mediated regulatory mechanisms in *C. difficile*. Although Hfq seems to bind and impact the stability of several sRNAs the role of Hfq in mediating sRNA-mRNA interactions in *C. difficile* remains unknown [308,311]. Moreover, RCd1, an sRNA that inhibits the production of the late mother cell-specific sigma factor σ^K, remains the only functionally characterized sRNA in *C. difficile,* revealing a paucity of knowledge that clearly warrants further investigation [311].

1.6 Aim of this thesis

The obligate anaerobic, Gram-positive, spore forming human pathogen *C. difficile* has emerged as the leading cause of antibiotic-associated diarrhoea [230–232]. Although numerous studies have significantly expanded our knowledge of CDI mechanisms, little is known about the bacterium's transcriptome architecture and mechanisms of post-transcriptional regulation. Initial data identified >100 sRNAs in *C. difficile*, suggesting the existence of a large post-transcriptional network [305–307]. Furthermore, depletion of the RNA-chaperone Hfq broadly impacts gene expression and bacterial physiology similar to its function in many Gram-negative bacteria [174,175,308,309]. However, conditions relevant for expression and targets interaction of sRNAs as well as the importance of Hfq in facilitating these interactions remain fundamental open questions. Accordingly, the aim of this thesis was to identify Hfq-associated sRNA-target interactions on a global scale and to functionally characterize their impact on *C. difficile* virulence and colonization. To achieve these goals, the following objectives were formulated:

1. Application of transcription start site and termination mapping to generate a single-nucleotide–resolution RNA map of *C. difficile*, allowing the annotation of ncRNAs including potential sRNAs.

2. Application of RIP-seq and RIL-seq to investigate the function of Hfq in sRNA-mediated regulation in *C. difficile* and to map the Hfq-mediated sRNA-target interactome on a transcriptome-wide scale.

3. Functional characterization of sRNA-mediated processes and their impact on *C. difficile* virulence and colonization.

Successful completion of these objectives will provide insights into the scope and importance of post-transcriptional regulation in this important human pathogen. In addition, global datasets that result from this thesis will be publicly available as a valuable foundation for future mechanistic studies of RNA-based gene regulation in *C. difficile*.

2 Publications

2.1 Statement of individual author contributions and of legal second publication rights to manuscripts included in the dissertation:

Manuscript 1:
Fuchs, M., Lamm-Schmidt, V., Sulzer, J., Ponath, F., Jenniches, L., Kirk, J. A., Fagan, R. P., Barquist, L., Vogel, J., & Faber, F. An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important Gram-positive bacterium. *Proceedings of the National Academy of Sciences*, 118(25) (2021). https://doi.org/10.1073/pnas.2103579118

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	FF	MF	VL-S	LB	JV
Methods Development	MF	FF	VL-S	RF	JK
Data Collection	MF	FF	VL-S	JS	
Data Analysis and Interpretation	MF	FF	VL-S	FP	LJ
Manuscript Writing					
Writing of Introduction	FF	MF	VL-S		
Writing of Materials & Methods	MF	VL-S	FF	JK	
Writing of Discussion	FF	MF	VL-S		
Writing of First Draft	FF	MF	VL-S		

Manuscript 2:
Fuchs, M., Lamm-Schmidt, V., Lenče, T., Sulzer, J., Bublitz, A., Wackenreuter, J., Gerovac, M., Strowig, T., & Faber, F. A network of small RNAs regulates sporulation initiation in *Clostridioides difficile*. *The EMBO Journal*, e112858 (2023). https://doi.org/10.15252/embj.2022112858

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	FF	MF	TS	AB	
Methods Development	MF	AB	TL	JS	
Data Collection	MF	AB	TL	JS	JW
Data Analysis and Interpretation	MF	VL-S	AB	TL	JW
Manuscript Writing					
Writing of Introduction	MF	FF	TL	JW	JS
Writing of Materials & Methods	MF	VL-S	TL	JS	AB
Writing of Discussion	MF	FF	TL	JW	JS
Writing of First Draft	MF	FF	TL	JW	JS

If applicable, the doctoral researcher confirms that she/he has obtained permission from both the publishers (copyright) and the co-authors for legal second publication.

The doctoral researcher and the primary supervisor confirm the correctness of the above mentioned assessment.

Manuela Fuchs	08/05/2023	Würzburg	
Doctoral Researcher's Name	Date	Place	Signature
Jun. Prof. Franziska Faber	08/05/2023	Würzburg	
Primary Supervisor's Name	Date	Place	Signature

2.2 Statement of individual author contributions to figures/tables of manuscripts included in the dissertation:

Manuscript 1:
Fuchs, M., Lamm-Schmidt, V., Sulzer, J., Ponath, F., Jenniches, L., Kirk, J. A., Fagan, R. P., Barquist, L., Vogel, J., & Faber, F. An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important Gram-positive bacterium. *Proceedings of the National Academy of Sciences*, 118(25) (2021). https://doi.org/10.1073/pnas.2103579118

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	MF	VL-S	FF		
2	MF	VL-S	FF		
3	MF	VL-S	FF		
4	MF	VL-S	FF		
5	MF	VL-S			
6	MF	VL-S			

Manuscript 2:
Fuchs, M., Lamm-Schmidt, V., Lenče, T., Sulzer, J., Bublitz, A., Wackenreuter, J., Gerovac, M., Strowig, T., & Faber, F. A network of small RNAs regulates sporulation initiation in *Clostridioides difficile*. *The EMBO Journal*, e112858 (2023). https://doi.org/10.15252/embj.2022112858

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	MF	VL-S			
2	MF	VL-S			
3	MF	VL-S			
4	MF				
5	MF				
6	TL	JS	MF		
7	MF	JW			
8	AB	MF			
9	MF				

I also confirm my primary supervisor's acceptance.

Manuela Fuchs	08/05/2023	Würzburg	
Doctoral Researcher's Name	Date	Place	Signature

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

Manuela Fuchs [1§], Vanessa Lamm-Schmidt [1§], Johannes Sulzer [1], Falk Ponath [2], Laura Jenniches [2], Joseph A. Kirk [4], Robert P. Fagan [4], Lars Barquist [2,3], Jörg Vogel [1,2], Franziska Faber [1, *]

1. Institute for Molecular Infection Biology (IMIB), Faculty of Medicine, University of Würzburg, D-97080 Würzburg, Germany
2. Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany
3. Faculty of Medicine, University of Würzburg, D-97080 Würzburg, Germany
4. The Florey Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN, UK

§ authors contributed equally

* Correspondence:

Franziska Faber: phone +49-931-3186280; email: franziska.faber@uni-wuerzburg.de

Published in: Proceedings of the National Academy of Sciences, 118(25).

This is an open access article under the terms of the Creative Commons Attribution NonCommercial-NoDerivs License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is non-commercial, and no modifications or adaptations are made (CC BY-NC-ND 4.0).

Competing Interest Statement: We declare no competing interests.

Classification: Biological Sciences, Microbiology

Keywords: *Clostridioides difficile*, differential RNA-seq, transcriptional termination site, small RNA, Hfq

Manuscript modifications: The Materials and Methods section is part of the Supplementary information in the published paper and has been moved to the manuscript section for the purpose of this thesis. The supplementary information is restricted to supplementary figures. All remaining supplementary information, including references for the Materials and Methods section, tables, and datasets are available online.

2.3.1 Abstract

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 post-transcriptional 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 non-coding regulators, including 42 sRNAs. Our results indicate functionality of many conserved riboswitches and predict novel *cis*-regulatory RNA elements upstream of MDR-type 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 post-transcriptional regulatory networks in gram-positive pathogens remain controversial. Using 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 their expression. As a proof-of-concept, we show that 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 ribo-regulation with implications for the infection process and provides evidence for a global role of Hfq in post-transcriptional regulation in a gram-positive bacterium.

2.3.2 Significance Statement

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. 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 novel non-coding 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.

2.3.3 Introduction

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* (*C. difficile*) has become the leading cause of antibiotic-associated diarrhea over the past two decades (1). Increasing numbers of multi-resistant 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 sporulation control by several global metabolic regulators including CcpA, CodY, Rex and PrdR (5, 6). Furthermore, several specialized and general sigma factors including TcdR (6), Spo0A (7), SigD (8, 9), SigH (10) and SigB (11) 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-seq-based annotations of the global transcriptome architecture, which have accelerated research in the gram-positive pathogens *Listeria* (12), *Staphylococcus* (13) and *Streptococcus* (14), have only recently become available for *C. difficile* (15).

This paucity of global knowledge about RNA output in *C. difficile* readily extends to post-transcriptional 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 (16, 17). Specifically, deletion of *hfq* increases sporulation (18), a crucial pathogenic feature of this bacterium that enables transmission between hosts. In gram-negative bacteria, Hfq commonly exerts global post-transcriptional control by facilitating short base pairing interactions of small regulatory RNAs (sRNAs) with *trans*-encoded target mRNAs (19, 20) 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 (21), but comparative analyses of a wild-type and *hfq* knockout strain revealed only moderate effects on sRNA and mRNA transcript levels (22) and the absence of any significant growth defect (23) which lead to the conclusion that Hfq plays a minor role in post-transcriptional regulation in *B. subtilis*.

Previous efforts using *in silico* methods (24) and RNA-sequencing (25) have predicted >100 sRNA candidates in *C. difficile*, which suggests the existence of a large post-transcriptional network. 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 post-transcriptional control in *C. difficile* are *cis*-regulatory RNA elements. Pioneering work has deciphered the function of genetic switches in the 5' untranslated region (5' UTR) of the *flgB* operon, containing the early-stage flagellar genes, and the cell wall protein encoding gene *cwpV* (26-28). Moreover, cyclic di-GMP responsive riboswitches were shown to regulate biofilm formation and toxin production (9, 29-31). 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 post-transcriptional regulation in this important human pathogen.

In the present study, we applied recently developed methods of bacterial RNA biology (32) to construct a global atlas of transcriptional and post-transcriptional control in *C. difficile*. We provide the first 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 post-transcriptional regulation and lay the foundation for future mechanistic studies of RNA-based gene regulation in *C. difficile*.

2.3.4 Results

2.3.4.1 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 broth, and late-exponential growth in Brain-heart-infusion broth (Fig. 1A). The resulting genome-wide maps provide single-nucleotide resolution transcriptional start sites (TSSs) and termination sites (TTS) (Fig. 1B). We have used this information to annotate 5' and 3' untranslated regions (UTRs) and operon structures, to correct previous ORF annotations, to add previously overlooked small genes and to annotate sRNA loci. Inspired by other gene expression databases such as SalCom (33), AcinetoCom (34) and Theta-Base (35), we have launched the interactive web browser 'Clost-Base' (https://www.helmholtz-hiri.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 non-coding 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.

2.3.4.2 Genome-wide annotation of TSSs

Differential RNA-seq (dRNA-seq) (36, 37) 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- cDNA libraries indicate RNA processing sites (Fig. 1B). The *C. difficile* 630 genome comprises annotations for 3,778 coding sequences as well as tRNAs, 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 half of the genome (Dataset S1).

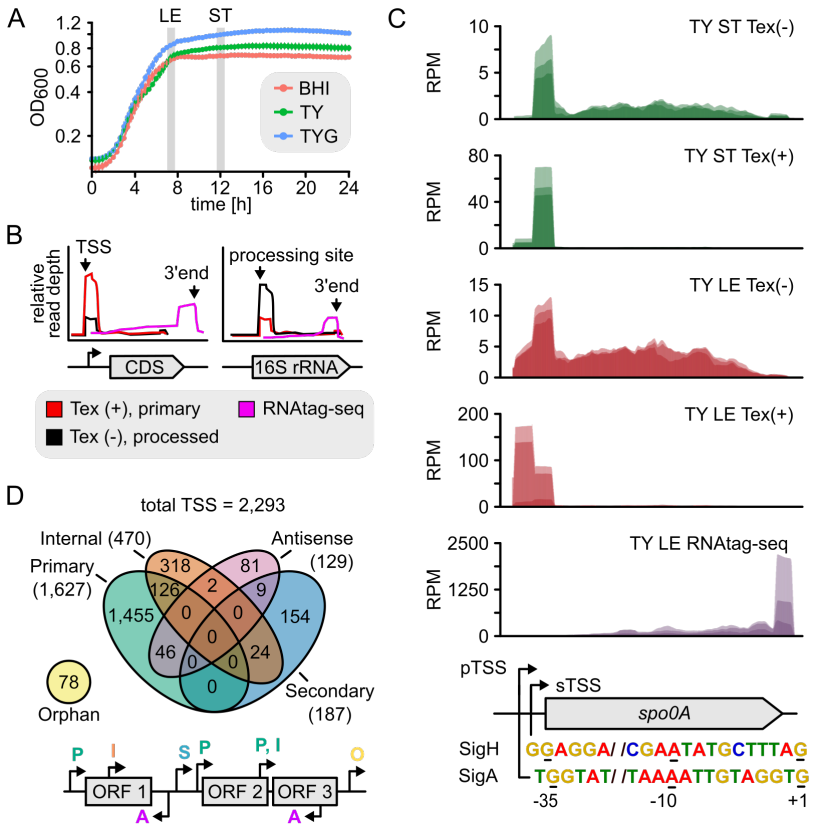


Figure 1. Global RNA-seq approaches for high-resolution transcriptome mapping. (A) Sequencing samples were generated from strain *C. difficile* 630 grown to late-exponential and stationary phase in Tryptone-yeast broth, as well as late-exponential phase in Brain-heart-infusion broth. Error bars show standard deviation of three biological replicates. **(B)** Read profiles generated with dRNA-seq and RNAtag-seq allow the annotation of transcriptional start sites (TSSs) and transcriptional termination sites (TTSs). For dRNA-seq, one fraction of total RNA is treated with terminator exonuclease (TEX+), which specifically degrades processed transcripts carrying a 5'-P or 5'OH. The other fraction remains untreated. This differential treatment results in a relative read enrichment for primary transcripts (5'-PPP) in the TEX+ treated libraries allowing the identification of TSSs. For RNAtag-seq, adapters for sequencing are ligated to RNA 3' ends thereby capturing 3' ends of transcripts. **(C)** Benchmarking of dRNA-seq approach. Two experimentally determined growth-phase-dependent TSSs for *spo0A* are consistent with dRNA-seq-based identification of *spo0A* associated TSSs. Abbreviations: LE, late-exponential growth phase; ST, stationary growth phase; CDS, coding sequence; TSS, transcriptional start site; TTS, transcript termination site. **(D)** Venn diagram showing distribution of TSSs among classes. TSS classification is based on expression strength and genomic location: primary (P), secondary (S), internal (I), antisense (A) and orphan (O).

Benchmarking our global data with previous studies of individual genes (10, 27, 38), our dRNA-seq results are consistent with published TSSs of *sigH*, *cwpV*, *cwp19*, and *spo0A* (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, for example, among the 1627 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 (39), 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 mis-annotated 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 belongs to a conserved family of ATPases involved in plasmid 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, GUG) and a ribosome-binding site (RBS) downstream of the iTSS identified a new start codon 43 nt downstream of the existing ORF annotation, placing a RBS in optimal distance to the new start codon and generating a 5' UTR length of 29 nt (Fig. S1A). This new annotation is in line with earlier studies in *Streptococcus pyogenes* which suggested this internal start codon is used *in vivo* (40) and accordingly used the shorter variant to solve the respective protein crystal structure (PDB id: 2oze). Similarly, we propose re-annotations in eight additional cases, including the *spaFEGRK* operon encoding an antibiotic/multidrug-family ABC transport system (Dataset S2).

2.3.4.3 Promoter architectures

C. difficile encodes 14 sigma factors, among them consensus sequences have been proposed for the vegetative SigA (25), the general stress response SigB (11, 41), the major transition phase SigH (10) and the sporulation-specific SigK (42) 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, that combined transcriptional start site 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 (15).

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 (Fig. S1B, Dataset S1 and S3). In line with recent data (15), 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, 43, 44). Similarly, we identify many genes previously predicted to be regulated by SigH (10, 15). Further, we identified 52 genes associated with a SigK promoter signature, including known genes of the SigK regulon such as *sleC* and *cdeC* (45). 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 (Fig. S1C).

2.3.4.4 Global mapping of transcript ends

To map transcription termination sites (TTS), we adopted the RNAtag-seq protocol (46), 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 non-coding 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, Dataset S4). In addition, a decrease in folding energy (Fig. S2A) and a spike in GC content at preceding TTS positions (Figure 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 (47). 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. Re-analysis of the original dataset for *C. difficile* confirms our

experimental results that intrinsic terminators rarely overlap with stop codons (Fig. S2B) suggesting that runaway transcription is not a determining feature of gene expression in *C. difficile*.

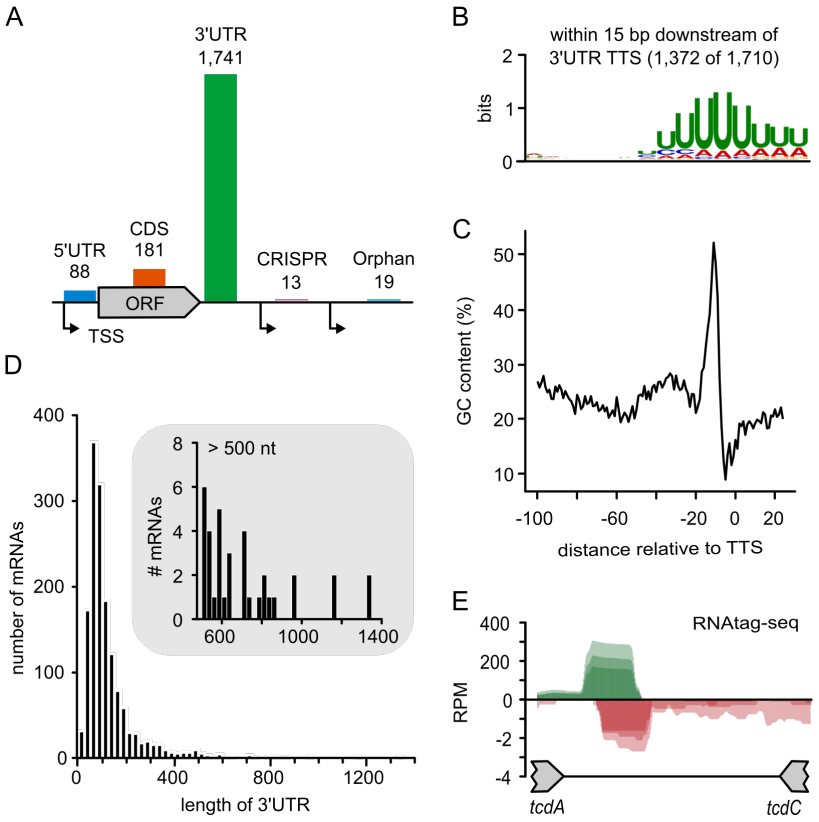


Figure 2. Global features associated with transcription termination sites (TTS). (A) Classification of TTS based on their genomic location: 3' UTR (downstream of CDS or non-coding gene), 5' UTR (within the 5' UTR of a coding sequence), CDS (within coding sequence), orphan (downstream of orphan TSS), CRISPR (associated with CRISPR array). Numbers indicate amount of annotated TTSs for each class. **(B)** Consensus motif associated with 3' UTR located TTSs downstream of coding sequences was identified using the MEME suite. The identified motif is located within 15 bp downstream of detected 3' UTR TTS positions. **(C)** Calculation of average GC content (%) at each indicated position relative to 3' UTR located TTSs reveals a spike in GC content immediately upstream of the TTS, indicating a stable stem structure. **(D)** Frequencies of 3' UTR lengths based on 1,741 detected 3' UTR TTSs. **(E)** Reads in the RNAtag-seq libraries reveal overlapping TTSs for the convergently transcribed genes *tcdA* and *tcdC*.

We further detected many overlapping termination events between convergently transcribed genes, such as for the toxin gene *tcdA* (*toxA* in CP010905.2 which was used in this study) and the negative regulator of toxin gene expression *tcdC* (Fig. 2E), which is also a pervasive transcriptome feature in *E. coli* (48). 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).

2.3.4.5 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 new biological functions (49, 50). In many cases, they are predicted to be membrane proteins, containing an alpha-helical transmembrane domain (51). Focusing on oTSS in particular, we searched for novel ORFs based on the following criteria (i) presence of a start and stop codon (ii) presence of a RBS within 15 bp upstream of the start codon and (iii) sequence conservation in other *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 *C. difficile* type-I toxin-antitoxin systems (52, 53). Among the remaining sORFs four are high confidence candidates; one being a conjugal transfer protein with annotation in other *C. difficile* strains; two candidates that each have a Shine-Dalgarno (SD) sequence and a predicted α-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 (48, 54, 55) but predicted operons in *C. difficile* are still largely based on computational inference from other bacterial species (56). Here, our transcriptome-based annotation predicts 440 operons (Dataset S1) including 64 novel operons such as the di-cistronic transcript comprising spermine/spermidine acetyltransferase *bltD* and CDIF630_01360, a putative N-acetyltransferase. Further, we identify 109 sub-operon structures which, in 29 cases, uncouple the last or the last 2 genes from upstream genes within an operon. One example is *acpS-ndoA* where the last two genes encoding a type-II toxin-antitoxin system (*ndoA* and *ndoA1*) are uncoupled from the full operon.

2.3.4.6 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 *B. subtilis* and *Listeria monocytogenes*, the majority of experimentally mapped 5' UTR lengths ranged from 20 to 60 nt (Fig. 3A, Dataset S1) (12, 36, 57). A likely RBS site (aGGAGg) was detected in ~ 90% of these 5' UTRs (Fig. 3A, inlet). Only six mRNAs appeared leaderless having a 5' UTR of <10 nt including the *spoVAE* gene within the tricistronic *spoVACDE* operon (58).

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 (59) (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 *speF* riboswitch previously only documented in gram-negative alpha-proteobacteria 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 (Fig. S3).

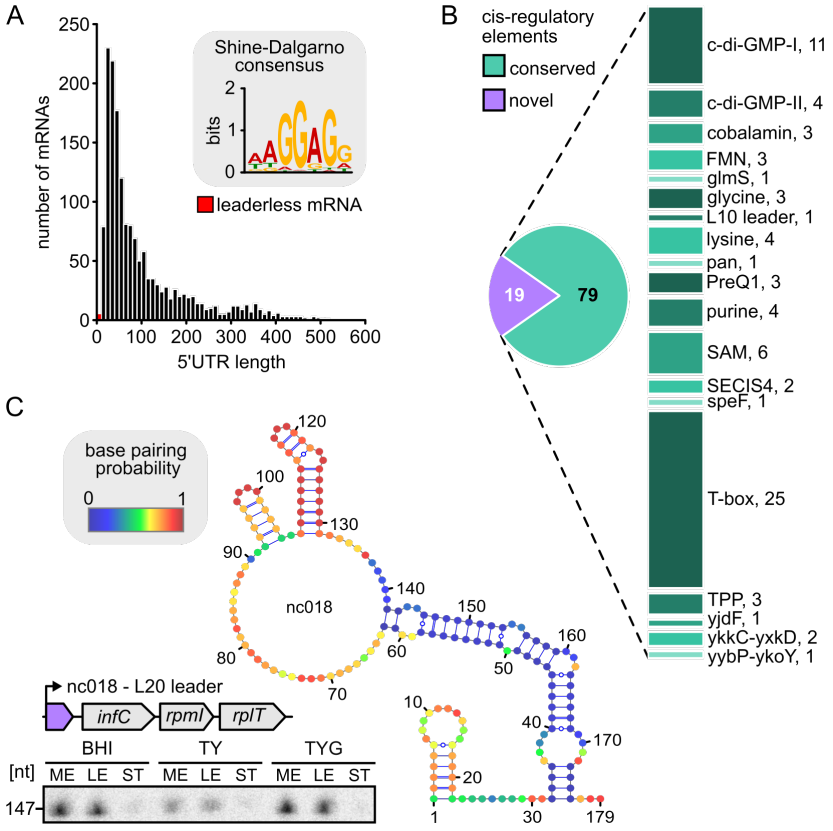


Figure 3. Discovery of *cis*-regulatory elements in 5' UTR regions. (A) Frequencies of 5' UTR lengths based on 1,646 primary and secondary TSSs. Red bars indicate six leaderless mRNAs with a 5' UTR length of <10 nt. The inset shows the predicted SD sequence motif of *C. difficile* 630. **(B)** Classification of premature transcription termination (PTT) events in 5' UTRs of mRNAs. The majority is associated with conserved RNA families. The remaining 19 PTT events lack homology to known riboregulators and are classified novel putative regulators. **(C)** Putative L20 leader in the 5' UTR of the *infC* operon. Top: Predicted secondary structure for CDIF630nc_018 using RNAfold. Nucleotides are colored according to base-pairing probabilities. Bottom: schematic overview of the *infC* operon organization including 5' UTR region and expression analysis of CDIF630nc_018 by northern blot using a radioactively labeled DNA probe. Total RNA was extracted at mid-exponential (ME), late-exponential (LE) and stationary (ST) growth phases from *C. difficile* 630 grown in BHI, TY and TYG media.

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* (60, 61). This autoregulatory structure is also found in other low-GC gram-positive bacteria (Rfam family RF00558) but does not seem to be conserved on the primary sequence level in *C. difficile*. Nevertheless, secondary structure prediction reveals extensive interactions between distant bases that is reminiscent of the antiterminator conformation of the L20 leader described in *B. subtilis* (Fig. 3C). Further PTT-associated-genes encoded PTS systems (*bglF*, *bglF1*, *bglG3* and *bglG4*) known to be regulated by antiterminator proteins in *B. subtilis* (62), transcriptional regulators (CDIF630_00097, CDIF630_02384, CDIF630_02922), MDR-type ABC transporters (CDIF630_03083, CDIF630_02847, 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 (Fig. S3).

2.3.4.7 The sRNA landscape of *C. difficile* in changing environmental conditions and during infection

Our primary transcriptome analysis identified 42 novel transcripts that lacked an internal open reading frame, qualifying them as potential small regulatory RNAs (sRNA) (Dataset S5). A classification based on their genomic location (Fig. 4A) revealed the largest group to be 3' UTR-derived sRNAs (18), followed by those encoded *cis*-antisense either to a gene, a 5'/3' UTR or another sRNA (13). In comparison, only few sRNAs were located in intergenic regions (8) or derived from the 5' UTR of mRNAs (3). Sequence conservation beyond *C. difficile* was extremely rare (Fig. S4A, Dataset S7) whereas most are highly conserved among clinical strains of *C. difficile* covering 5 major genomic clades and diverse toxin repertoires (Fig. S4B) (63). 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 non-coding 5'UTR 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 half of them are downregulated 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 (Fig. S5).

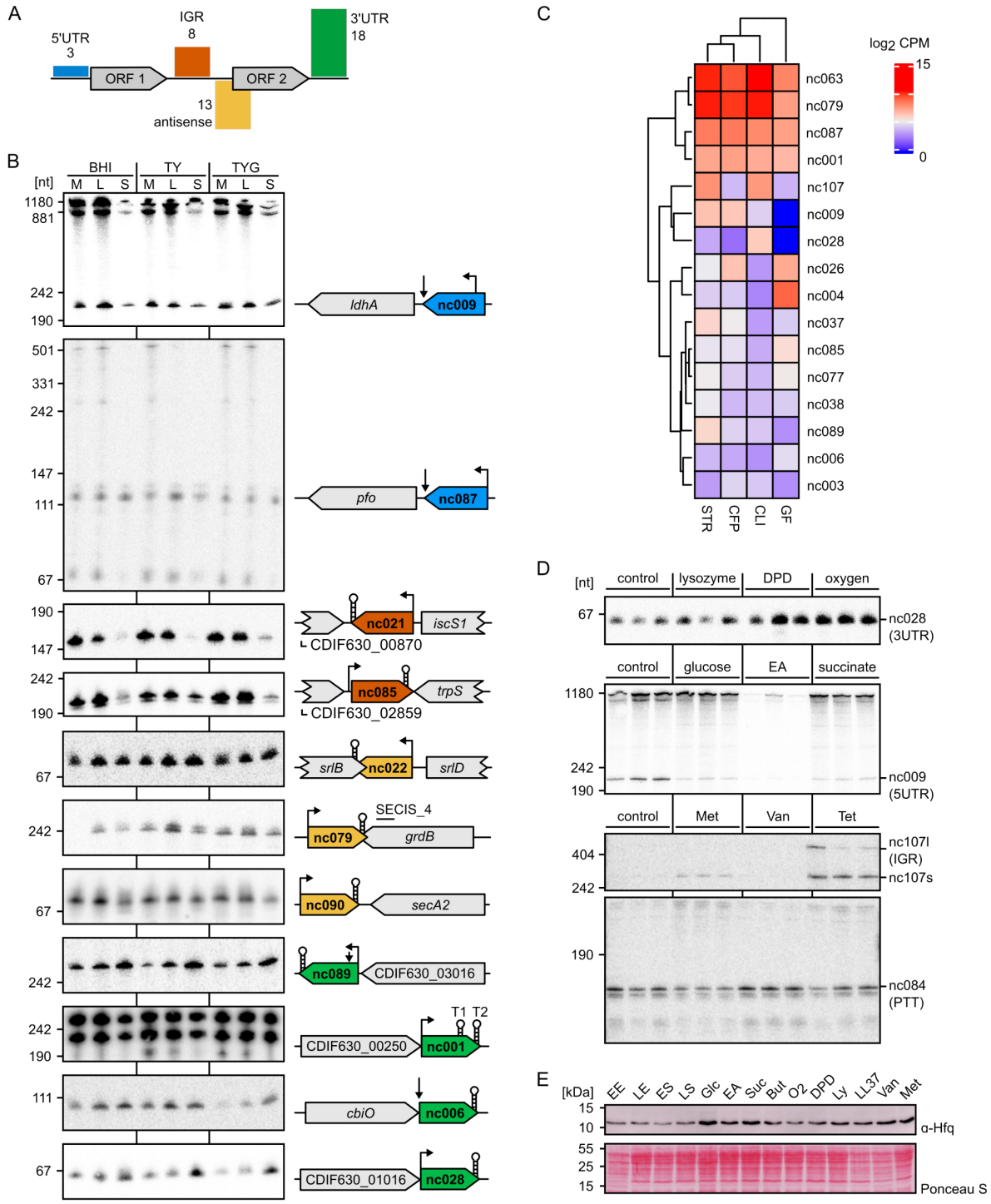


Figure 4. *C. difficile* 630 sRNA landscape in changing environmental conditions. (A) Classification of annotated sRNA candidates based on their genomic location. Numbers indicate amount of annotated sRNAs for each class. **(B)** Expression profiles of representative candidates for each sRNA class. Total RNA was extracted at mid-exponential (ME), late-exponential (LE) and stationary (ST) growth phases from *C. difficile* 630 grown in TY, TYG and BHI medium and analyzed by northern blot using radioactively labeled DNA probes. **(C)** Heatmap of $\log_2$ 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 (64) and re-analyzed 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 mid-exponential 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 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 (Ly), LL-37, vancomycin (Van), Metronidazole (Met) or in 1% oxygen atmosphere (O_2). Precise concentrations used are listed under materials and methods.

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 pre-treated with different antibiotics (64). Re-analysis of the data revealed expression of 16 sRNAs (cutoff = 50 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 (Figure 4C, 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, butyrate), oxygen stress, iron limitation as well as antimicrobial compounds (lysozyme, LL-37, vancomycin, metronidazole, tetracycline). Within this panel, we observed regulation for several sRNAs or PTTs (Fig. 4D); for example, 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.

2.3.4.8 *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 (65). However, whether Hfq functions as a central RNA-binding protein (RBP) 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 3×FLAG tagged *hfq* (Hfq-FLAG) under its native promoter to draft the spectrum of Hfq-bound RNAs *in vivo*. Bacteria were grown to mid-exponential, late-exponential and stationary phase (Fig. S6A) and subjected to the RIP-seq protocol (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 (Fig. S6C). In addition, co-purification of Hfq-bound sRNAs was confirmed by northern blot analysis of lysates and eluate fractions from Hfq-FLAG and Hfq-ctrl cultures (Fig. S6C).

The majority of RNA species co-immunoprecipitating with Hfq-3×FLAG were mRNAs (Fig. 5A, 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 PTT (66/124 5'UTRs) (Fig. 5A). Many mRNAs showed growth-phase-dependent association with Hfq (Fig. S7B, 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 (*spo0A*, *spoVB*, *oppB*, *sigG* and *sspA2*), toxin production (*tcdE*), motility (*flgB*, *fliJ*, *flhB* 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) (Fig. S7C, Dataset S6).

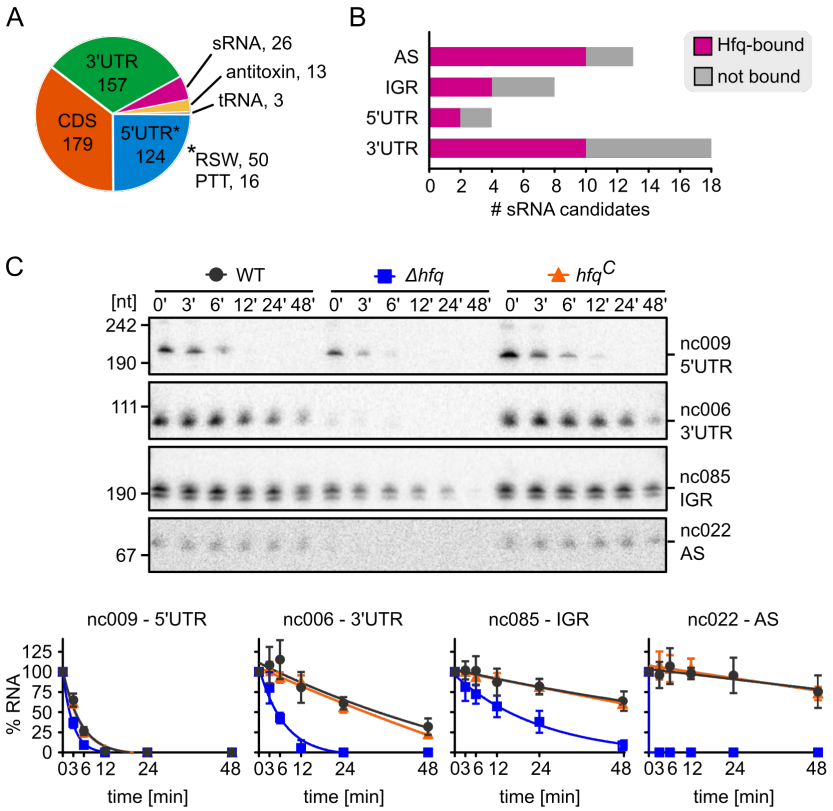


Figure 5. The spectrum of Hfq-associated RNA ligands and Hfq impact on RNA stability. (A) Pie chart for Hfq RIP-seq showing the relative amount of Hfq-associated sequences mapping to different RNA classes. Hfq RIP-seq was performed on *C. difficile* 630 WT and C-terminally 3xFLAG tagged *hfq* expressed from a plasmid, grown to mid-exponential, late-exponential and early-stationary phases in TY medium in two independent experiments. **(B)** Bar chart showing the fraction of sRNAs for each class that were bound by Hfq *in vivo*. **(C)** Hfq acts as a stabilizing factor for its ncRNA ligands *in vivo*. Top: Total RNA was extracted from WT, Δ*hfq*, and the complemented *hfq* mutant (*hfq*^{C}) in late-exponential phase after addition of rifampicin at the indicated time points and analyzed by Northern blotting using radioactively labeled DNA probes. A representative of three independent northern blots is shown. Abbreviations: RSW (riboswitch), PTT (premature transcription termination), AS (antisense), IGR (intergenic)

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 enrichment of many 3' UTRs and of 5' UTRs harboring a riboswitch or PTT 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 (Fig. S7A).

Importantly, many predicted sRNAs associated with Hfq, including the previously identified Hfq-binding sRNA CDIF630nc_070 (Fig. 5A, Dataset S6) (25). In addition, all but one type-I antitoxin transcript were enriched (Fig. 5A, Dataset S6). 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 half of all IGR, 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 (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 (Fig. S6D).

2.3.4.9 Hfq acts as a stabilizing factor for its noncoding RNA ligands

One major function of gram-negative Hfq homologues is the stabilization of their sRNA ligands, which appears to be the exception in gram-positive bacteria (66-68). To test if transcript stabilities of Hfq-associating ncRNAs were affected by Hfq, we generated a *hfq* deletion (Δ*hfq*) in *C. difficile* 630 using a newly developed homologous recombination vector that enables effective counterselection on standard laboratory media using the tightly regulated *E. coli* endoribonuclease *mazF* (for detail see *SI Appendix*). To complement *hfq in cis (hfq*^{C}*)*, we used an additional novel vector system which enables efficient insertion of DNA by homologous recombination between *pyrE* (CD630_01870) and CD630_01880.

C. difficile WT, Δ*hfq* and *hfq*^{C} 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 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 sRNAs were decreased, an effect that could be restored in the *hfq*^{C} strain (Fig. 5C, Fig. S8A, B). 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 (Fig. S8B).

2.3.4.10 A role of CDIF630nc_085 in the regulation of ethanolamine utilization

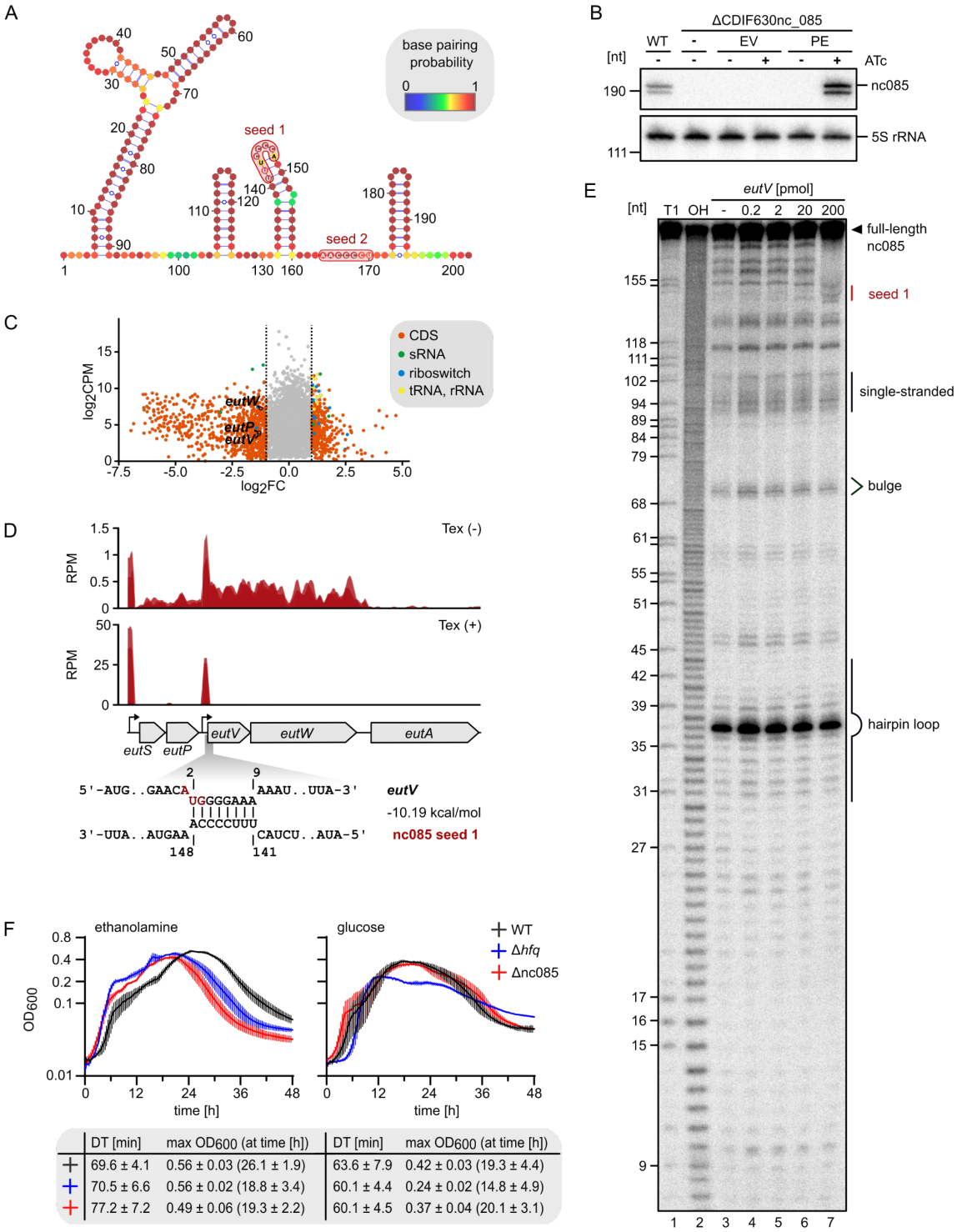


Figure 6. CDIF630nc_085 regulates ethanolamine utilization in *C. difficile*. (A) Predicted secondary structure for CDIF630nc_085 (nc085) using RNAfold. Nucleotides are colored according to base-pairing probabilities. Two seed regions predicted to interact with differentially expressed mRNA targets are 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 puls-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. Genes that were differentially expressed upon pulse expression of nc085 (FDR <=0.1, log_2FC >=1 OR log_2FC <=-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. Red letters indicate the start codon. **(E)** In-line probing of 0.2 pmol of ^{32}P-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, Δ*hfq* and Δnc085 in CDMM supplemented with either glucose (5 mM) or ethanolamine (50 mM). Growth experiments were performed in two independent experiments with 3 biological replicates each (n = 6). Error bars show standard error of the mean. Bottom table: listed are the calculated doubling times (DT) ± standard deviation as well as the maximum OD_{600} and the timepoint when it was reached during the experiment, for each strain and each growth condition. Due to the complex, diauxic growth behaviour of all strains in the analyzed growth conditions, doubling times were calculated only for the first exponential phase.

To gain first insight into sRNA functions, we chose the *bona fide* intergenic sRNA CDIF630nc_085 (short nc085) (Fig. 4B) for further 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 hair-pins in the center of the sRNA, and a Rho-independent terminator at the 3' end (Figure 6A). To identify potential mRNA targets of nc085, we pulse-expressed the sRNA from an anhydrotetracycline-inducible plasmid in a ΔCDIF630nc_085 background (Fig. 6B) and determined differentially expressed genes using RNA-seq analysis (Fig. 6C and Dataset S11). We observed many sporulation-associated genes to be down-regulated upon nc085 expression (Dataset S11). In addition, gene cluster analysis revealed ABC transporters and two-component systems as the largest group of regulated genes (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 (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 hair-pin 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 Δ*hfq* mutant in *C. difficile* minimal medium (CDMM) supplemented either with glucose or ethanolamine. Cultivation in glucose-supplemented medium resulted in

identical growth for WT and Δnc085, whereas Δ*hfq* had an extended lag phase and grew to lower maximum optical densities. In contrast, when grown in CDMM supplemented with ethanolamine, both Δnc085 and Δ*hfq* 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. 6F, table). 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* (69).

2.3.5 Discussion

In the present study, we have applied global approaches of bacterial RNA biology (32) to capture the transcriptome architecture of *C. difficile* and unravel the scope of post-transcriptional regulation in this important human pathogen.

2.3.5.1 A gram-positive Hfq homologue 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 *Listeria monocytogenes*, *B. subtilis* and *Staphylococcus* species. For example, the sRNA RsaE, that is conserved in the order Bacillales, has been shown to coordinate central carbon and amino acid metabolism in *Staphylococcus aureus* (70, 71) and also to promote biofilm formation in *Staphylococcus epidermidis* by supporting extracellular DNA release and the production of polysaccharide-intercellular adhesin (PIA) (72). 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 (73).

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 (16). 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 (16). Taken together, the regulatory activities of Hfq in *C. difficile* strongly resemble the situation in Gram-negative model organisms, such as *E. coli* and *S. enterica*, and open avenues for the study of sRNA-based gene regulation in this important pathogen.

2.3.5.2 Ethanolamine utilization is governed by sRNA regulation in gram-positive pathogens

We have taken a first step towards this goal by identifying the ethanolamine utilization pathway to be regulated by the intergenic 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 (69, 74-76). Interestingly, the ethanolamine utilization (*eut*) gene cluster is regulated by a small regulatory RNA in both *Enterococcus faecalis* (EutX) and *Listeria monocytogenes* (Rli55). In both species, the sRNA contains a classic AdoCbl-responsive riboswitch in it's 5' portion that allows expression of the full-length sRNA only in the absence of vitamin B_{12} (77, 78). This full length transcript harbours a dual hairpin structure in its 3' part that serves as a binding platform for the ethanolamine-specific, RNA-binding response regulator EutV (79) thereby sequestering EutV from binding to its target sequences in the 5' UTRs of *eut* genes. *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-expresssion 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.

2.3.6 Materials and Methods

2.3.6.1 Bacterial strains and growth conditions

A complete list of *C. difficile* and *E. coli* strains that were used in this study is provided in Table S1. *C. difficile* cultures were routinely grown anaerobically inside a Coy chamber (85% N_2, 10% H_2 and 5% CO_2) in Brain Heart Infusion (BHI) broth or on BHI agar plates (1.5% agar) unless indicated otherwise. When necessary, antibiotics were added to the medium at the following concentrations: thiamphenicol 15 µg/ml, cycloserine 250 µg/ml, cefoxitin 8 µg/ml. *E. coli* cultures were propagated aerobically in Luria-Bertani (LB) broth (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl) or on LB agar plates (1.5% agar) supplemented with chloramphenicol (20 µg/ml) as appropriate. *E. coli* strain Top10 (Invitrogen) was used as a recipient for all cloning procedures, and *E. coli* CA434 (HB101 carrying the IncPβ conjugative plasmid R702) was used as donor strain for conjugation of plasmids into *C. difficile* 630.

For dRNA-seq analysis biological triplicates of *C. difficile* 630 were grown to late-exponential and early-stationary phase in Tryptone-yeast broth, and to late-exponential phase in Brain-heart-infusion broth. For RNAtag-seq analysis biological triplicates of *C. difficile* 630 were grown to late-

exponential phase in Tryptone-yeast broth. Subsequently, 4 OD units were mixed with stop mix, snap frozen in liquid nitrogen and stored at -80 °C until RNA extraction (hot phenol protocol).

For profiling sRNA expression in a variety of environmental conditions, *C. difficile* 630 was grown in TY broth to mid-exponential phase before being exposed for 20 min to the following nutrient and stress conditions: glucose (0.5%), ethanolamine (0.5 %), succinate (0.1%), lysozyme (50 µg/ml), LL-37 (2 µg/ml), metronidazole (2 µg/ml), vancomycin (2 µg/ml), oxygen (atmospheric oxygen concentration), iron limitation (200 µM 2,2'-Dipyridine) and butyrate (5 mM). Subsequently, 4 OD units were mixed with stop mix, snap frozen in liquid nitrogen and stored at -80 °C until RNA extraction (hot phenol protocol).

For profiling Hfq expression, bacterial cultures were started in biological duplicates in TY medium supplemented with above nutrients and reagents. Samples for western blot analysis were harvested in late exponential phase. The following conditions were adapted to allow for efficient growth of *C. difficile*: LL-37 (1 µg/ml), metronidazole (0.5 µg/ml), vancomycin (0.5 µg/ml), oxygen (cultures were incubated in a defined atmosphere of 5% CO_2 and 1% oxygen inside a phcbi MCO-170M multi-gas incubator). Equal OD units for all cultures were harvested *via* centrifugation for 5 min at 5,000 x g. The pellet was frozen over night at -20°C. Cells were resuspended in 1 x PBS and incubated for 40 mins at 37°C which leads to consistent cell lysis (1).

For growth in defined *C. difficile* minimal medium (CDMM), medium was prepared as described by Neumann-Schaal *et al.* (2) and glucose or ethanolamine were added individually to serve as carbon sources. Bacteria were sub-cultured in biological triplicates 2 x in CDMM containing 5 mM glucose before inoculation of main cultures. Main cultures contained either 5 mM glucose or 50 mM ethanolamine, and growth was measured using a Synergy™ H1 microplate reader.

2.3.6.2 Plasmid construction

All plasmids and DNA oligonucleotides used in this study are listed in Tables S2 and S3, respectively. Plasmids were propagated in *E. coli* TOP10 according to standard procedures (3).

pJAK184 – a novel plasmid for generating gene deletions by homologous recombination

Efficient homologous recombination in *C. difficile* relies on effective counterselection markers to enable screening for plasmid loss. Cytosine deaminase (encoded by *codA*) is a commonly used marker but requires use of defined media, which is time consuming to make and results in slow growth, particularly of mutants with lower fitness. To enable selection on standard laboratory media, an alternative recombination vector was constructed using a tightly regulated *E. coli mazF* (encoding endoribonuclease) in place of *codA*. BamHI, SacI, and KpnI restriction sites were first removed from pMTL-SC7215 (4) by inverse PCR cloning (RF851/RF852). The resulting plasmid was linearised by PCR using oligonucleotides RF1065 and RF1066, adding new terminal SacI and

BamHI sites. Homology arms upstream and downstream of the *slpA* gene were amplified using oligonucleotides RF1025/RF1067 and RF1026/RF1068, respectively, joined by SOEing PCR and combined with the linearised vector by SacI/BamHI restriction ligation resulting in pJAK112. The *xylB* promoter and adjacent *xylR* (Pxyl) were amplified from *C. difficile* strain R20291 using oligonucleotides RF1589 and RF1590 and inserted into pRPF185 using KpnI/SacI restriction-ligation. *E. coli mazF* was codon optimised for *C. difficile*, synthesised (FragmentGENE by Genewiz) and inserted into the resulting plasmid *via* SacI/BamHI restriction-ligation. Pxly-*mazF* was then amplified using oligonucleotides RF1704 and RF1705 and inserted into pJAK112 using NotI/XhoI restriction-ligation resulting in pJAK183. The SacI site upstream of *mazF* was then removed by inverse PCR (RF1718/RF1719) resulting in pJAK184. pJAK112 and pJAK184 have been deposited with Addgene (#167280 and #167281 respectively).

Synthesised *mazF* sequence:

GCATTCGAGCTCCTGCAGTAAAGGAGAAAATTTTATGGTAAGTAGATATGTACCAGATATGGGAGATT
TAATATGGGTTGATTTTGATCCAACAAAAGGTAGTGAACAAGCAGGACATAGACCAGCTGTTGTATTA
AGTCCATTTATGTATAATAATAAAACAGGTATGTGTTTATGTGTTCCTTGTACAACTCAATCAAAAGG
ATATCCATTTGAAGTTGTTTTATCAGGTCAAGAAAGAGATGGAGTAGCATTAGCTGATCAAGTAAAAA
GTATAGCATGGAGAGCAAGAGGAGCAACTAAAAAAGGAACAGTTGCACCAGAAGAATTACAATTAAT
TAAAGCAAAAATTAATGTATTAATTGGATAAATCAACAACTCTCCTGGCGCACCGGATCCTAT

pJAK080 – a novel plasmid for gene complementation by homologous recombination

pJAK080 was constructed to enable efficient insertion of cargo DNA by homologous recombination between *pyrE* (CD630_01870) and CD630_01880. The plasmid was initially designed to insert a tetracycline-inducible copy of *mreB2* as proof of principle but can be easily modified to replace the cargo gene (SacI/BamHI), promoter (KpnI/SacI) or both (KpnI/BamHI). *mreB2* was PCR amplified using oligonucleotides RF627 and RF628 and inserted downstream of Ptet in pRPF185 (1) using SacI/BamHI restriction-ligation. Ptet-*mreB2* and homology arms upstream and downstream of the genome insertion site were amplified by PCR using oligonucleotides RF839 and RF840, RF837 and RF838, and RF841 and RF842 respectively. The three resulting fragments were assembled by Gibson assembly into pMTL-SC7315 (4), linearised by PCR using oligonucleotides RF311 and RF312. Extraneous BamHI, SacI, and KpnI restriction sites in the plasmid backbone were removed as above, resulting in pJAK080. pJAK080 has been deposited with Addgene (#167279).

pFF-10 and pFF-12 - plasmids for expression of WT and 3X-FLAG-tagged *hfq* for RIP-seq

hfq including its native 5' UTR and native promoter as determined by transcriptional start site mapping (5) was PCR amplified from *C. difficile* 630 using either FFO-122 and FFO-136 or FFO-

122 and FFO-207. The resulting fragments were purified from 1% agarose gels with NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel) and cloned into SacI/BamHI-digested pRPF185 (1), generating pFF-10, encoding *hfq* including 5' UTR and native promoter, and pFF-12, encoding a C-terminally 3X-FLAG-tagged *hfq* including 5' UTR and native promoter, respectively.

pFF-50 and pFF-112 - plasmids for generating deletions of *hfq* and CDIF630nc_085

For deletion of *hfq*, allelic exchange cassettes were designed with approx. 1.2 kb of homology to the chromosomal sequence flanking the up- and downstream regions of the *hfq* coding region. Homology regions were amplified *via* high fidelity PCR with 5% DMSO using FFO-370/-371 and FFO-369/-372. The resulting fragments were purified from 1% agarose gels with NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel). Ligation into PCR-linearized pJAK184 (FFO-362/-363) was done *via* Gibson Assembly (Gibson Assembly ® Master Mix, New England BioLabs) according to the manufacturer's instructions, resulting in pFF-50.

Plasmid assembly for the deletion of CDIF630nc_085 was performed similarly, using FFO-670/-671 and FFO-672/-673 for amplification of homology regions. However, instead of pJAK184, pMTL-SC7315 (4) was used as vector backbone, yielding pFF-112.

pFF-102 - plasmid for complementation of Δ*hfq in trans*

hfq including its native promoter, 5' UTR and terminator was amplified from *C. difficile* 630 using FFO-601 and FFO-602. The resulting fragment was cloned into PCR linearized pJAK80 (FFO-599/-600), resulting in pFF-102. PCR amplification, purification and Gibson assembly were performed as described for PFF-50 and pFF-112.

pFF-114 and pFF-14 - plasmids for pulse-expression of CDIF630nc_085

CDIF630nc_085 was PCR amplified from *C. difficile* 630 using FFO-678 and FFO-679. The resulting fragment was purified from a 1% agarose gel with NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel) and cloned into a PCR linearized pRPF185 (1). PCR linearization was achieved using FFO-323/-324, enabling Gibson cloning directly adjacent to the P_{tet} promoter and *slpA* terminator region and yielding pFF-114.

The empty control plasmid pFF-14 was derived from pRPF185 by PCR linearization using FFO-151 and FFO-152, removing *gusA* and introducing an additional BamHI site. Following BamHI digest, the linearized plasmid was purified from a 1% agarose gel with NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel) and re-ligated using T4 DNA ligase (NEB), according to the manufacturer's instructions.

pFF-115 and pFF-124 - plasmids for *in vitro* transcription of CDIF630nc_085 and *eutV*

CDIF630nc_085 and *eutV* (5'UTR and first 100 nt of coding region) were amplified *via* Phusion high fidelity PCR using FFO-703/-704 and FFO-836/-837 respectively, adding a 5' overhang including the T7-promoter sequence (5'- GTTTTTTTTAATACGACTCACTATAGGG) to both fragments. The resulting PCR products were purified from 1% agarose gels with NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel). Prior to Strataclone TA-cloning, 3'-adenine overhangs were added to both fragments using Taq Polymerase (Biozym). Subsequently fragments were cloned into Strataclone TA-cloning vector and transformed into StrataClone SoloPack competent cells according to the manufacturer's protocol (Strataclone PCR Cloning Kit).

pET-21b - plasmid for recombinant expression of *C. difficile* 630 Hfq-6xHis

hfq coding region was PCR amplified from *C. difficile* 630 using FFO-298 and FFO-299. The resulting fragment was purified and cloned into the PCR linearized pET-21b vector (Novagen, Inc.) *via* Gibson cloning. PCR linearization was achieved using FFO-296 and FFO-297, gel purification and Gibson Assembly was performed as described before.

2.3.6.3 Plasmid conjugation

For conjugation purposes, plasmids were transformed into *E. coli* CA434 (HB101 carrying the IncPβ conjugative plasmid R702) as follows: 80 µl of electro competent *E. coli* CA434 were mixed with 100 – 500 ng of plasmids in a pre-chilled electroporation cuvette and transformed by delivering a pulse of electricity (1.8 kV, 200 Ω, 4 - 5 sec). Cells were recovered for 4 h at 37 °C in 1 ml LB. Colonies harbouring the plasmid were selected on LB with chloramphenicol (20 µg/ml). Conjugation was performed according to Kirk and Fagan, 2016 (6). In short: 200 µl of *C. difficile* 630 overnight cultures were incubated at 37 °C for 2 min. Simultaneously, 1 ml of overnight *E. coli* conjugant donor culture was harvested by centrifugation at 4000 x g for 2 min. *E. coli* pellets were then transferred into the anaerobic workstation and gently resuspended in pre-incubated 200 µl *C. difficile* 630 culture. Following resuspension, the cell suspension was pipetted onto well-dried, non-selective BHI agar plates (10 × 10 µl spots), allowed to dry and incubated for 8 h at 37 °C. Growth was harvested using 900 µl of TY broth, serially diluted and spread on plates containing either cycloserine (control), or cycloserine and thiamphenicol, to select for transconjugants. Plates were incubated for between 24 and 72 h, until colonies were apparent.

2.3.6.4 Generation of deletion- and complemented strains

Gene deletions were constructed *via* homologous recombination as previously published (4). In short, plasmids were transformed into *E. coli* CA434 and conjugated in *C. difficile* 630 as described above. Following successful conjugation, colonies were screened for the first recombination event *via* PCR. Positive recombinants were broadly streaked on non-selective BHI. Following incubation for 2 – 3 days, plates were harvested using 900 µl 1xPBS. Subsequently, 50 µl of a 10^{-4} and 10^{-5} dilution of the mixture were streaked either on CDMM supplemented with 50 µg/ml

Fluorocytosine for pMTL-SC7315-derived plasmids, or on TY containing 4% w/v xylose for pJAK184-derived plasmids. Once growth occurred, 8 – 15 colonies were re-streaked to purity and tested for secondary recombination events *via* PCR. Additionally, strains were tested for plasmid loss on selective plates containing 15 µg/ml thiamphenicol. Finally, successful deletions and complementations were confirmed *via* Sanger sequencing.

2.3.6.5 Expression and purification of recombinant *C. difficile* 630 Hfq-6xHis

For recombinant expression of *C. difficile* 630 Hfq-6xHis, FFS-263 was grown in 500 ml LB till OD_{600}=0.6. The culture was induced with 0.25 mM IPTG and incubated over night at 18 °C. For cell lysis, cells were harvested *via* centrifugation for 30 min at 25,000 x g and 4 °C, resuspended in 10 ml lysis buffer per g cell pellet (20 nM Tris-HCl pH 8.0, 500 nM NaCl, 10% glycerol, 0.1% Triton X-100, 20 nM Inidazole, protease inhibitors) and lysed using sonication. The cell lysate was then incubated at 85 °C for 15 min before centrifugation at 25,000 x g for 30 min at 4 °C. Purification *via* IMAC was performed using a 0.5 ml Ni-NTA resin, equilibrated to lysis buffer, and incubated with cell lysate supernatant for 1 h at room temperature, followed by gravity flow column chromatography using 6 x 0.5 ml IMAC elution buffer (20 nM Tris-HCl pH 8.0, 500 nM NaCl, 10% glycerol, 0.1% Triton X-100, 20 nM Inidazole). Appropriate fractions were pooled, concentrated to <500 µl and centrifuged for 30 min at 16,000 x g and room temperature. Finally, a second purification step was performed using a Superdex® 200 10/300 column (SEC buffer: 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol, 1 mM EDTA) applying a flow rate of 0.5 ml/min at room temperature, resulting in purified Hfq.

Purified Hfq was used for antibody production, performed by Davids Biotechnologie GmbH.

2.3.6.6 Rifampicin assay

Cultures were grown in TY broth in biological triplicates to late exponential phase (OD_{600} = 1) before adding 200 µg/ml of rifampicin (Merck) to arrest transcription. Samples were collected at indicated timepoints, mixed with 1/5 volume of stop mix (95% EtOH, 5% phenol) and snap frozen in liquid nitrogen. Until further processing, samples were stored at -80 °C. RNA was extracted using the hot phenol method as described below. RNA was analysed *via* Northern blot and signal strengths were quantified with ImageJ (7). GraphPad Prism 9 was used for calculation of ncRNA half-lifes, by applying the one phase decay equation.

2.3.6.7 Pulse-expression, RNA-sequencing and data analysis

C. difficile 630 WT, FFS-329 as well as FFS-329 harbouring pFF-114 and pFF-14 (FFS-333 and FFS-332) were grown in TY broth in biological triplicates to late exponential phase (OD_{600} = 1). Cultures of FFS-333 and FFS- 332 were then split in half and one of each was induced for 20 min by adding 100 ng/ml anhydrotetracycline (Merck). Following pulse-expression, all cultures were stopped by adding 1/5 volume of stop mix (95% EtOH, 5% phenol) and snap frozen in liquid

nitrogen. Until further processing, samples were stored at -80 °C. RNA was extracted using the hot phenol method as described below. Successful pulse-expression was confirmed *via* Northern blot analysis.

Sequencing was accomplished by Vertis Biotechnology GmbH. In brief, RNA was analyzed on a Shimadzu MultiNA microchip. Ribosomal RNA was depleted using an in-house developed depletion protocol. Fragmentation was done with ultrasound (1 pulse of 30 s at 4 °C). 3'end adapter was ligated and first-strand cDNA synthesis was performed using M-MLV reverse transcriptase with 3' adapter as primer. cDNA was purified and 5'Illumina TruSeq adapter was ligated to the 3' end of the antisense cDNA. Amplification of cDNA (10-20 ng/µl) was done by PCR with high fidelity polymerase and 15 PCR cycles. Final purification of cDNA was done with Agencourt AMPure XP kit. The samples were pooled in approximately equimolar amounts and size fractionated (range 200 – 600 bp) by agarose gel electrophoresis. Sequencing was done on an Illumina NextSeq 500 system using 75 bp read length.

Reads were filtered and trimmed with BBDuk. Mapping and read counting was done with READemption pipeline 0.4.3 (8). Normalization and enrichment analysis was done with edgeR version 3.28.1 (9). To calculate normalization factors the trimmed means of M values (TMM) method was used between pulse expression and empty vector libraries. All features that had more than 10 read counts in at least two libraries were included in the analysis. Transcripts with $\log_2$ fold change of >1 or <-1 and an FDR of <0.1 were considered as significant. KEGG/GO enrichment analysis was done with clusterProfiler (10).

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 GSE155167.

2.3.6.8 Hot phenol extraction of total RNA

Total RNA was extracted using the hot phenol protocol. Bacterial cultures were grown to the desired OD_{600}, mixed with 0.2 volumes of STOP solution (95% ethanol, 5% phenol) and snap-frozen at -80 °C if not directly processed. The bacterial solution was centrifuged for 20 min, 4500 rpm at 4 °C and the supernatant was completely discarded. Cells were suspended in 600 µl of 10 mg/ml lysozyme in TE buffer (pH 8.0) and incubated at 37 °C for 10 min. Next, 60 µl of 10% w/v SDS was added and everything mixed by inversion. Samples were incubated in a water bath at 64 °C, 1-2 min before adding 66 µl 3 M NaOAc, pH 5.2. Next, 750 µl acid phenol (Roti-Aqua phenol) was added, followed by incubation for 5 min. at 64 °C. Samples were briefly placed on ice to cool before centrifugation for 15 min, 13,000 rpm at 4 °C. The aqueous layer was transferred into a 2 ml phase lock gel tube (Eppendorf), 750 µl chloroform (Roth, #Y015.2) was added, and everything centrifuged for 12 min, 13,000 rpm at room temperature. For ethanol precipitation,

the aqueous layer was transferred to a fresh tube, 2 volumes of 30:1 mix (EtOH:3 M NaOAc, pH 6.5) was added and incubated overnight at -20 °C. Precipitated RNA was harvested by centrifugation, washed with cold 75% v/v ethanol and air-dried. DNA contaminations were removed by incubating RNA samples with DNaseI, 10x DNase buffer (Merck) and RNase inhibitor (Takara Bio) for 1 h at 37 °C. Finally, RNA was re-extracted using a single phenol-chloroform extraction step (ROTI®Phenol/Chloroform/Isoamylalkohol). Purified RNA was resuspended in 50 µl RNase-free water and stored at -80 °C.

2.3.6.9 Library preparation for differential RNA-seq (dRNA-seq)

Library preparation for dRNA-seq was accomplished by Vertis Biotechnology AG. In brief, total RNA was analyzed on a Shimadzu MultiNA microchip electrophoresis system. 23S/16S ratio for all samples was 1.3. RNA was fragmented *via* ultrasound (4 pulses a 30 sec, 4 °C) and subsequently treated with T4 Polynucleotide Kinase (NEB). Half of the samples were then treated with terminator exonuclease (TEX) for dRNA-seq.

For the cDNA synthesis, the RNA fragments were poly(A)-tailed and 5'PPP structures were removed with RNA 5' Polyphosphatase (Epicentre). The RNA sequencing adapter with the barcodes were ligated to the 5'-monophosphate of the fragments. First strand cDNA was synthesized with an oligo(dT)-adapter primer and M-MLV reverse transcriptase. Amplification of cDNA was done *via* PCR to an approximate amount of 10 - 20 ng/µl. cDNA was purified with the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analyzed with Shimadzu MultiNA microchip electrophoresis. Equimolar amounts of the samples were pooled for sequencing. cDNAs had a size between 200 and 550 bp. The library pool was fractionated *via* differential clean-up with the Agencourt AMPure kit. The cDNA pool was checked with capillary electrophoresis as stated above. Libraries were sequenced on an Illumina NextSeq 500 system using 75 bp read length.

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 GSE155167.

2.3.6.10 Library preparation for RNAtag-Seq protocol

Total RNA quality was checked using a 2100 Bioanalyzer with the RNA 6000 Nano kit (Agilent Technologies) and rRNA was detected. The RIN for all samples was >6.5. Equal amounts of samples (~500 ng) were used for the preparation of cDNA libraries with the RNAtag-Seq protocol as previously published by Shishkin *et al.* (11) with minor modifications.

Briefly, the RNA samples were fragmented with FastAP buffer at 94 °C for 3 min and were dephoshorylated using the FastAP enzyme (Thermo Scientific) at 37 °C for 30 min followed by

bead purification with 2 volumes of Agencourt RNAClean XP beads. Fragmentation profiles were checked using RNA 6000 pico kit (Agilent) with a 2100 Bioanalyzer. The RNA fragments were ligated with 3´-barcoded adaptors at 22 °C for 1 h and 30 min using T4 RNA Ligase (NEB). Barcoded RNA samples were pooled together, the ligase was inhibited by RTL buffer (Qiagen RNeasy Min Elute Cleanup Kit) and purified with RNA Clean & Concentrator-5 column (Zymo). rRNA were depleted from pools using Ribo-Zero (Bacteria) Kit (Illumina) and purified with RNA Clean & Concentrator-5 column (Zymo). The profiles before and after rRNA depletion were analysed with RNA 6000 pico kit (Agilent) with a 2100 Bioanalyzer. The rRNA depleted RNA were reverse transcribed and the first strand of the cDNA were synthesized using a custom AR2 oligo (Sigma-Aldrich) and the AffinityScript multiple temperature cDNA synthesis kit (Agilent) at 55 °C for 55 min. The RNA was degraded with 1 M NaOH at 70 °C for 12 min, the reactions were neutralized with Acetic acid and were purified with 2 volumes of MagSi-NGS^{PREP} Plus beads (AMSBIO). A second 3Tr3 adaptor were ligated to the cDNA using T4 RNA Ligase (NEB) overnight at 22 °C followed with two bead purification steps with 2 volumes of MagSi-NGS^{PREP} Plus beads (AMSBIO). A PCR enrichment test was performed in order to determine the number of PCR cycles that are necessary for each pool followed by a bead purification step and a QC with DNA HS kit (Agilent). Then, the final PCR was performed with the number of cycles determined from the previous step using P5 and P7 primers. Two sequential bead purification steps were performed with 1.5 and 0.7 respectively.

Libraries were quantified with the Qubit 3.0 Fluometer (ThermoFisher) and the library quality and size distribution (~480 bp peak size) was checked using a 2100 Bioanalyzer with the High Sensitivity DNA kit (Agilent). Sequencing of pooled libraries, spiked with 5% PhiX control library, was performed with ~8 million reads / sample in single-end mode on the NextSeq 500 platform (Illumina) with the High Output Kit v2.5 (75 Cycles). We noticed a strong enrichment of reads mapping to native 3' ends of transcripts in *C. difficile* 630 libraries that were prepared for the RNAtag-Seq protocol. We cannot fully explain this observation but we suspect a combination of inefficient RNA fragmentation and native 3' end stability of transcripts to contribute to this outcome for RNAtag-Seq libraries prepared from *C. difficile* 630 total RNA.

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 GSE155167.

2.3.6.11 Hfq co-immunoprecipitation and RNA-sequencing (RIP-seq)

Overnight cultures of FFS-38 and FFS-50 were prepared in biological duplicates in buffered TY and used for inoculation of pre-cultures. Once an OD_{600} of 0.5 was reached, each replicate was used for inoculation of three separate flasks. Main cultures were grown until either mid-exponential

(OD_{600}=0.5), late-exponential (OD_{600}=0.9) or stationary (3 h post entry) growth phase. For each condition, 50 OD were harvested by centrifugation for 20 min at 4000 rpm at 4 °C, snap frozen and resuspended in 800 µl lysis buffer (20 mM TRIS pH 8.0, 1 mM $MgCl_2$, 150 mM KCl, 1 mM DTT). Subsequently, each sample was mixed with 1 µl DNase I (Fermentas, 1 U/µl) and 800 µl of 0.1 mm glass beads and lysed in a RETSCH's Mixer Mill (30 Hz, 10 min, 4°C). Bacterial lysates were cleared by centrifugation for 10 min at 14,000 x g and 4 °C. Approximately 900 µl supernatant was transferred to a new tube and incubated with 25 µl (1/2 * OD in µl) of mouse-anti-FLAG antibody (clone M2, Merck/Sigma-Aldrich #F1804) at 4 °C for 45 min (rocking). Immunoprecipitation of FLAG-tagged Hfq was performed by incubating each sample with 75 µl pre-washed (3 x resuspended in 1 ml lysis buffer and centrifuged at 10,000 rpm for 1 min) protein A sepharose beads (Merck/Sigma-Aldrich #P6649) for 45 min at 4 °C (rocking). Beads and captured proteins were washed 5 x with 500 µl lysis buffer (mixed by inversion and centrifuged at 10,000 rpm for 1 min) and resuspended in 500 µl lysis buffer. Finally, RNA, co-immunoprecipitated with antibodies and protein A sepharose beads, was eluted by adding the same volume of Phenol:Chloroform:Isoamylalcohol (25:24:1, pH 4.5, Roth). The solution was mixed for 20 s, transferred to PLG tubes (Eppendorf) and incubated at room temperature for 3 min. Following centrifugation (30 min, 15,200 x g, 15 °C) the aqueous phase was transferred to new tubes and RNA was precipitated over night at -20 °C by adding 30 µg Glycoblue and 800 µl isopropanol. Precipitated RNA was pelleted (centrifugation for 45 min, 15,200 rpm, 4 °C), washed with 80% and 100% ethanol (centrifugation for 10 min, 15,200 rpm, 4 °C) and resuspended in 15,5 µl nuclease-free water (65 °C, 1 min, 600 rpm). For DNase treatment, purified RNA was incubated with 2 µl DNase I, 0.5 µl RNase inhibitor and 2 µl 10 x DNase buffer for 30 min at 37 °C. To each sample, 100 µl nuclease free water was added and DNase treated RNA was purified by Phenol:Chloroform:Isoamylalcohol treatment, as described before. Precipitation was performed overnight at -20 °C by mixing the samples with 3 x the volume of sodium-acetate/ethanol (1:30), followed by centrifugation and washing with 80% and 100% ethanol. Finally, air-dried pellets were resuspended in 20 µl nuclease-free water (65 °C, 1 min, 600 rpm).

RNA quality was controlled using a 2100 Bioanalyzer and the RNA 6000 Pico kit (Agilent Technologies). Since rRNA was detectable, the RIN for all samples was >7.5. Equal amounts of RNA (~50 ng) were used for preparation of cDNA libraries with the NEBNext Multiplex Small RNA Library Prep kit for Illumina (NEB) with some minor changes to the manufacturers' instructions. RNA samples were fragmented with Mg^{2+} at 94 °C for 2 min, 45 s using the NEBNext Magnesium RNA Fragmentation Module (NEB) followed by RNA purification with the Zymo RNA Clean & Concentrator kit. Fragmented RNA was dephosphorylated at the 3' end, phosphorylated at the 5' end and decapped using 10 U T4-PNK +/-, 40 nmol ATP and 5 U RppH respectively (NEB). After each enzymatic treatment, RNA was purified with the Zymo RNA Clean & Concentrator kit. The

RNA fragments were ligated for cDNA synthesis to 3' SR adapters and 5' SR adapters and diluted 1:5 with nuclease-free water before use. PCR amplification was performed to add Illumina adaptors and indices to the cDNA for 16 cycles, using 1:5 diluted primers. Barcoded DNA Libraries were purified using magnetic MagSi-NGSPREP Plus beads (AMSBIO) at a 1.8 ratio of beads to sample volume. Finally, libraries were quantified with the Qubit 3.0 Fluorometer (ThermoFisher) and the library quality and size distribution (~230 bp peak size) was checked using a 2100 Bioanalyzer with the High Sensitivity DNA kit (Agilent). Sequencing of pooled libraries, spiked with 5% PhiX control library, was performed with ~5 million reads / sample in single-end mode on the NextSeq 500 platform (Illumina), using the High Output Kit v2.5 (75 cycles). Demultiplexed FASTQ files were generated with bcl2fastq2 v2.20.0.422 (Illumina). Reads were trimmed for the NEBNext adapter sequence using Cutadapt version 2.5 with default parameters. In addition, Cutadapt was given the –nextseq-trim=20 switch to handle two colour sequencing chemistry. Reads that were trimmed to length=0 were discarded.

The READemption pipeline version 4.5 (8) with segemehl version 0.2 (12) was used for read mapping and calculation of read counts. Reads were mapped on CP010905.2 with additional annotations for ncRNAs, 3' UTRs and 5' UTRs. Normalization and enrichment analysis was done with the edgeR version 3.28.1 (9). To calculate normalization factors the trimmed means of M values (TMM) method was used between flag-tagged and non-flag-tagged Hfq libraries of each growth phase individually. All features that considered more than 10 read counts in at least two replicates were included in the analysis. Only features with a Benjamini-Hochberg corrected P-value (FDR) of ≤0.1 and a $\log_2$ fold-change of ≥2 were considered as significantly enriched.

To find a consensus motif for enriched sequences, the peak calling pipeline PEAKachu (https://github.com/tbischler/PEAKachu) was used to define regions of read enrichment. PEAKachu was run on adaptive mode with deseq as normalization method. Peaks with a minimum of $\log_2$ fold change of 2 and maximum of 0.1 adjusted p-value were considered as significant. The resulting peak table was manually curated. Peak sequences were extracted and used as input for CMfinder 0.3 (13). Resulting motifs were depicted with R2R (14).

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 GSE155167.

2.3.6.12 Re-Analysis of *in vivo* RNA-seq data

RNA-sequencing data from Jenior *et al.* (15) were retrieved from NCBIs Sequence Read Archive (SRR5124212 – SRR5124216) and converted to fastq format with "fastq-dump". Mapping was done as described above with READemption pipeline in paired-end mode against CP010905.2 annotation with additional annotations for regulatory RNAs and UTRs. Resulting BAM files were

treated similarly to Jenior *et al.* (15), by subsequent removal of optical and PCR duplicates using Picard MarkDuplicates (*http://broadinstitute.github.io/picard/*) with lenient stringency. Read counts were then calculated with READemption and simply converted to counts per million (CPM) for comparison of treatments. For heatmaps CPMs were $\log_2$ transformed with a pseudocount of 1.

2.3.6.13 Northern blotting

RNA samples were mixed with equal amounts of gel loading buffer II, boiled at 98 °C for 5 min and cooled down on ice before loading on a denaturing 6% polyacrylamide gel in TBE buffer containing 7 M urea. Gels were transferred onto Hybond+ membranes (GE Healthcare Life Sciences) at 4 °C with 50 V (~100 W) for 1 h. For hybridization with P^{32}-labeled DNA oligonucleotides, membranes were incubated over night at 42 °C in Roti® Hybri-Quick Buffer (Roth). Finally, membranes were washed three times with decreasing concentrations of SSC buffer (5 x, 1 x and 0.5 x) before imaging on a Typhoon FLA 7000 phosphor imager.

2.3.6.14 Western blotting

To verify the expression and successful pulldown of FLAG-tagged Hfq, 9 µl lysate and 50 µl of resuspended beads were mixed with 81 µl and 50 µl 1 x protein loading dye respectively and boiled for 5 min at 98 °C. Following incubation, 20 µl of each sample was loaded and separated on a 15% SDS–polyacrylamide gel followed by transfer of proteins to a PVDF membrane. For detection of FLAG-tagged proteins, the membrane was blocked in TBS-T with 5% milk powder for 1 h at room temperature and washed 3 x in TBS-T for 10 min. Subsequently the membrane was incubated over night at 4 °C with anti-FLAG antibody (Sigma) diluted 1:1,000 in TBS-T with 3% BSA and washed again 3 x in TBS-T for 10 min. Following the last washing step the membrane was incubated for 1 h at room temperature with anti-mouse-HRP antibody (ThermoScientific) diluted 1:10,000 in TBS-T with 3% BSA and finally washed 3 x in TBS-T for 10 min before adding ECL substrate for detection of HRP activity using a CCD camera (ImageQuant, GE Healthcare).

For profiling Hfq expression in several growth and stress conditions, cell lysates were mixed with equal amounts of 2 x protein loading dye, boiled for 5 min at 95 °C before loading 0.5 OD on a 15% SDS–polyacrylamide gel for subsequent western blotting as described above. Equal loading of protein samples was confirmed *via* Ponceau S staining of the PVDF membrane. Staining was reversed by washing the stained membrane with 0.1 M NaOH for 1 min. The stained and destained membrane was then processed as described above using an anti-Hfq antibody (Davids Biotechnologie) and a secondary anti-rabbit-HRP antibody (GE Healthcare).

2.3.6.15 In vitro transcription and radiolabelling of RNA

For *in vitro* transcription of CDIF630nc_085 and *eutV*, pFF-115 and pFF-124 were used for template generation *via* Phusion High-Fidelity PCR. Resulting PCR products were purified from

1% agarose gels with NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel) to prevent the production of side products during *in vitro* transcription. *In vitro* transcription was performed in 40 µl reactions using the Invitrogen MEGAscript T7 Transcription Kit (Thermofisher Scientific) according to the manufacturer's protocol. The RNA product was then purified from a denaturing Urea PAGE with 6% polyacrylamide and 7 M Urea as follows: the gel was stained for 10 min in ethidium bromide (Carl Roth) and washed in water before imaging using a Intas Gel Doc systems.

The desired bands were cut out in small pieces and moved into a 2 ml tubes. For RNA elution, 750 µl RNA elution buffer (0.1 M NaAc, 0.1% SDS, 10 mM EDTA) was added and the samples were incubated at 4 °C and 1000 rpm over night. Gel pieces were removed by centrifugation at 5,000 x g and 4 °C for 1 min. Supernatants were transferred to new tubes and RNA extraction was performed using a single phenol-chloroform extraction step (ROTI® Phenol/Chloroform/Isoamylalkohol). Purified RNA was resuspended in RNase-free water and stored at -80 °C.

For radioactive labelling, 50 pmol of *in vitro* transcribed CDIF630nc_085 was dephosphorylated using 25 U of calf intestine alkaline phosphatase (NEB) in a 50 µL reaction volume and incubated for 1 h at 37 °C. RNA was extracted using a single phenol-chloroform extraction step (ROTI® Phenol/Chloroform/Isoamylalkohol) and resuspended in RNase-free water. Subsequently, the dephosphorylated and purified RNA (20 pmol) was 5' end-labeled (20 µCi of ^{32}P-γATP) using 1 U of polynucleotide kinase (NEB) for 1 h at 37 °C in a 20 µL reaction volume. Finally, the labelled RNA was purified on a G-50 column (GE Healthcare) and extracted from a PAA gel as above following visualization on a phosphorimager (FLA-3000 Series, Fuji).

2.3.6.16 In-line probing

In-line probing assays were performed by incubating 0.2 pmol of labelled CDIF630nc_085 either alone or with increasing concentrations of *in vitro* transcribed *eutV* (0.2, 2, 20 and 200 pmol) for 40 h at room temperature in 1 x in-line probing buffer (100 mM KCl, 20 mM $MgCl_2$, 50 mM Tris-HCl, pH 8.3). The RNase T1 ladder was prepared by incubating 0.2 pmol of labelled CDIF630nc_085 in 8 µL of 1 x sequencing buffer (Ambion) at 95 C for 1 min followed by the addition of 1 µL RNase T1 (0.1 U µL^{-1}) and incubation at 37 °C for 5 min. The alkaline hydrolysis ladder was prepared by incubating 0.2 pmol labelled CDIF630nc_085 with 9 µL 1 x alkaline hydrolysis buffer (Ambion) and incubated at 95 °C for 5 min. All reactions were stopped by the addition of 10 µL 2 x colorless gel-loading solution (10 M urea, 1.5 mM EDTA) and stored on ice.

Samples were resolved on a 10% (vol/vol) PAA-7 M urea sequencing gel pre-run for 30 min, 45 W prior to sample loading. The gel was dried for 2 h on a Gel Dryer 583 (Bio-Rad) and visualized after appropriate exposure on a phosphorimager (FLA-3000 Series, Fuji).

2.3.6.17 Annotation of TSS, TTS, sRNAs and operons

For the prediction of transcriptional start sites (TSSs) and small regulatory RNAs (sRNAs), we used the modular, command-line tool ANNOgesic (16). All parameters were kept at the default setting, unless stated otherwise. The TSS prediction algorithm was trained using parameters that were derived from manual curation of the first 200 kb of the genome. Secondary TSSs were excluded if a primary TSS was within seven nucleotides distance. For annotation of 5' untranslated regions (5' UTRs) the default setting for maximum length was changed to 500 nt.

For the prediction of sRNAs, we used the default settings, that is, the transcript must have a predicted TSS or processing site (PS) at their 5' end, form a stable secondary structure (calculated with RNAfold from Vienna RNA package), and a length between 30 to 500 nt. For 5' UTR-derived sRNAs the transcripts must have a PS at their 3' end. For 3' UTR-derived sRNA the transcript can either have a TSS or PS at their 5' end and share its terminator with the parental gene.

Prediction of transcriptional termination sites (TTSs) was accomplished manually. Read coverage tracks of RNAtag-Seq libraries were visualized with Integrative Genomics Viewer (17) and scanned for enriched regions. Termination sites were defined as the first position with less than half of the maximal reads of the entire enriched region.

For operon annotation, we used TSS and TTS to define transcription unit boundaries, and only those with a detected TSS were considered. The resulting operon predictions were then compared to the list of (mostly) computationally predicted operons available for the *C. difficile* 630 reference genome (NC_009089) on the BioCyc database (18) and (re-)classified according to the current operon status into four categories: confirmed, extended, new or shorter.

Read coverage profiles were generated with (19).

2.3.6.18 Prediction of promoter and terminator motifs, novel ORFs

Motif search and generation of sequence logos was accomplished with MEME version 5.1.1 (20) and CentriMO version 5.1.1 (21). For determination of a SigA promoter motif, 100 nt upstream of each TSS was extracted and uploaded to MEME. The number of motifs to be found by MEME was set to 10 and only the given strand was searched. All other settings were left at default. The resulting SigA motif matrix was then uploaded to CentriMo along with the extracted promoter regions. Again, only the given strand was searched and all other settings were left at default. Based on the CentriMo output, those promoter sequences harboring a SigA promoter motif were uploaded to MEME once more to refine the SigA promoter motif. In contrast to SigA, sequences known to harbor either a SigB, SigH or SigK promoter motif were uploaded for the initial MEME search instead of extracted promoter sequences generated in this study. The input sequences were based on the following studies: SigB (22); SigH (23); SigK (24). The thereby generated motif matrices were uploaded to CentriMo together with all of our extracted promoter regions. Those

sequences predicted to contain a SigB, SigH or SigK motif were uploaded to MEME to generate a refined promoter motif based on our promoter sequences. Following the MEME search, some sequences were excluded manually based on the following criteria: promoter location within the uploaded sequence has to be at the 3' end; at least one G within 3 nucleotides next to the -10 box (only SigB); G at position 25, 26 or 27 (only SigH); ATA at position 23-25 and at least one A and C between position 2-7 (only SigK). Finally, the manually curated promoter sequences were uploaded to MEME to generate the final promoter motif.

For 3' UTR motif search 15 nt downstream of all termination sites were extracted to search for the presence of a common poly-U stretch. The command line version of MEME was executed with default options except for the minimal motif length, which was set to 4. Only the given strand was used for the search.

Prediction of novel ORFs was accomplished manually. In order to identify novel ORFs we focused on oTSS, - TSS not associated with any coding or non-coding sequence. To avoid any miss-annotation we applied the following criteria: (i) presence of a start and stop codon (ii) presence of a ribosome binding site within 15 bp upstream of a start codon (iii) sequence conservation in other *Clostridioides* strains.

2.3.6.19 Prediction of RNA folding and sRNA target mRNAs

Secondary structures of ncRNAs were predicted with the RNAfold WebServer (25) and visualized with VARNA (26).

To predict folding of TTS the sequence 60 nt upstream and 10 nt downstream of the termination site was extracted. RNAfold version 2.2.7 of the Vienna RNA Package 2 was used with default settings to predict secondary structures.

To find potential stem loop structures in annotated 3' UTRS 100 nt upstream and 25 nt downstream of each termination site was extracted. A sliding window approach (window size: 25 nt, step: 5 nt) was used to calculate local minimal free energies (MFE) with RNAfold for each sequence. To calculate Z-scores, the MFE of each window was compared to the MFEs of 1000 randomly shuffled sequences with the same dinucleotide frequency.

Potential sRNA-target interactions were predicted using CopraRNA (27-29). Target prediction was performed for all sRNAs that had an associated Hfq peak in our RIP-seq dataset. As input, three homologous sRNA sequences were uploaded for each sRNA including the *C. difficile* 630 homologue. Due to the low sequence conservation, only homologues present in other *C. difficile* strains could be selected as input sequences, however, we focused on those sequences that showed at least some sequence variation when compared to *C. difficile* 630. The target region was specified as 200 nt upstream and 100 nt downstream of annotated start codons. Default settings

were used. CopraRNA predictions for each sRNA candidate were then compared with mRNA candidates that co-immunoprecipitated with Hfq to filter for high probability targets.

Potential targets of CDIF630nc_085 were predicted using IntaRNA (27, 28). NZ_CP010905 was selected as Target NCBI RefSeq ID. Extracted sequences were set to 75 nt upstream and 200 nt downstream of start codons. All other settings were left as default. Targets predicted by IntaRNA were then compared with CDIF630nc_085 pulse-expression results for selection of high probability targets.

2.3.6.20 Conservation Analysis

Conservation Analysis of potential ncRNAs was conducted *via* homology search with profile Hidden Markov Models (HMM). First, profile HMMs were constructed with the hmmbuild tool of the HMMER 3 package (30). Subsequently, an iterative homology search with the nhmmer tool was done with an E-value and inclusion threshold of 0.001. The underlying genome database was constructed with an in-house python script that generates a file consisting of all Refseq assemblies of the phylum Firmicutes (TaxID 1239) or Clostridioides (TaxID 1496) which are considered as completely assembled. Percent identity of all hits was calculated *via* the esl-alipid tool which is also a part of the HMMER 3 package.

To calculate the conservation of UTR regions, orthologs were assigned for different strains of *Clostridioides* by using the tool Proteinortho (Version 6) (31). The ortholog assignment was based on all protein sequences of annotated genes of 21 different strains of *Clostridioides* representing all major clades and different toxin inventories (32). Proteinortho was used with default options. For further analysis, only the genes with exactly one ortholog in each of the 21 strains were considered.

Next, fasta files were generated for each ortholog group. The files contained the nucleotide sequence of 50 nt upstream of the start position for 5' UTR analysis or 50 nt downstream of the stop position for 3' UTR analysis of each ortholog. Only the orthologs of genes were considered for which we annotated a 3' UTR or 5' UTR in this work.

For each group of orthologs a multiple sequence alignment was generated with Clustal Omega (1.2.4) with default options (33). Percent identities were calculated as stated above with esl-alipid.

2.3.7 Data Availability

All RNA-sequencing data are available at the NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE155167. Plasmids pJAK184 and pJAK080 have been deposited with Addgene (167279–167281).

2.3.8 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 IZKF at the University of Würzburg (project Z-6). We would also like to acknowledge the recombinant protein expression facility of the Rudolf-Virchow-Center for expression and purification of Hfq. We thank the Vogel Stiftung Dr. Eckernkamp for supporting F.P. with a Dr. Eckernkamp Fellowship. J.S. was supported by a DFG grant (FA 1113/2-1). J.A.K. was supported by the Wellcome Trust (grant number 204877/Z/16/Z).

2.3.9 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.J., L.B., and F.F. analyzed data; and M.F., V.L.-S., and F.F. wrote the paper.

2.3.10 References

1. W. K. Smits, D. Lyras, D. B. Lacy, M. H. Wilcox, E. J. Kuijper, *Clostridium difficile* infection. Nat Rev Dis Primers 2, 16020 (2016).
2. Z. Peng et al., Update on antimicrobial resistance in *Clostridium difficile*: resistance mechanisms and antimicrobial susceptibility testing. J Clin Microbiol 55, 1998-2008 (2017).
3. S. Khanna, D. N. Gerding, Current and future trends in *clostridioides* (*clostridium*) *difficile* infection management. Anaerobe 10.1016/j.anaerobe.2019.04.010 (2019).
4. B. Guery, T. Galperine, F. Barbut, *Clostridioides difficile*: diagnosis and treatments. Bmj 366, l4609 (2019).
5. L. Bouillaut, T. Dubois, A. L. Sonenshein, B. Dupuy, Integration of metabolism and virulence in *Clostridium difficile*. Res Microbiol 166, 375-383 (2015).
6. I. Martin-Verstraete, J. Peltier, B. Dupuy, The regulatory networks that *Control Clostridium difficile* toxin synthesis. Toxins (Basel) 8 (2016).
7. L. J. Pettit et al., Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. BMC Genomics 15, 160 (2014).
8. I. El Meouche et al., Characterization of the SigD regulon of *C. difficile* and its positive control of toxin production through the regulation of tcdR. PLoS One 8, e83748 (2013).
9. R. W. McKee, M. R. Mangalea, E. B. Purcell, E. K. Borchardt, R. Tamayo, The second messenger cyclic Di-GMP regulates *Clostridium difficile* toxin production by controlling expression of sigD. J Bacteriol 195, 5174-5185 (2013).
10. L. Saujet, M. Monot, B. Dupuy, O. Soutourina, I. Martin-Verstraete, The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. J Bacteriol 193, 3186-3196 (2011).
11. N. Kint et al., The alternative sigma factor sigma(B) plays a crucial role in adaptive strategies of *Clostridium difficile* during gut infection. Environ Microbiol 19, 1933-1958 (2017).

12. O. Wurtzel et al., Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. Molecular systems biology 8, 583 (2012).

13. U. Mader et al., *Staphylococcus aureus* transcriptome architecture: from laboratory to infection-mimicking conditions. PLoS Genet 12, e1005962 (2016).

14. I. Warrier et al., The transcriptional landscape of *Streptococcus pneumoniae* TIGR4 reveals a complex operon architecture and abundant riboregulation critical for growth and virulence. PLoS pathogens 14, e1007461 (2018).

15. O. Soutourina et al., Genome-wide transcription start site mapping and promoter assignments to a sigma factor in the human enteropathogen *Clostridioides difficile*. Frontiers in microbiology 11, 1939 (2020).

16. P. Boudry et al., Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen *Clostridium difficile*. J Bacteriol 196, 3234-3248 (2014).

17. J. Caillet, C. Gracia, F. Fontaine, E. Hajnsdorf, *Clostridium difficile* Hfq can replace *Escherichia coli* Hfq for most of its function. RNA 20, 1567-1578 (2014).

18. A. Maikova, V. Kreis, A. Boutserin, K. Severinov, O. Soutourina, Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. Appl Environ Microbiol 85 (2019).

19. E. Holmqvist, J. Vogel, RNA-binding proteins in bacteria. Nat Rev Microbiol 16, 601-615 (2018).

20. K. Kavita, F. de Mets, S. Gottesman, New aspects of RNA-based regulation by Hfq and its partner sRNAs. Curr Opin Microbiol 42, 53-61 (2018).

21. M. Dambach, I. Irnov, W. C. Winkler, Association of RNAs with *Bacillus subtilis* Hfq. PLoS One 8, e55156 (2013).

22. H. Hammerle et al., Impact of Hfq on the *Bacillus subtilis* transcriptome. PLoS One 9, e98661 (2014).

23. T. Rochat et al., Tracking the elusive function of *Bacillus subtilis* Hfq. PLoS One 10, e0124977 (2015).

24. Y. Chen, D. C. Indurthi, S. W. Jones, E. T. Papoutsakis, Small RNAs in the genus *Clostridium*. MBio 2, e00340-00310 (2011).

25. O. A. Soutourina et al., Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. PLoS Genet 9, e1003493 (2013).

26. B. R. Anjuwon-Foster, R. Tamayo, A genetic switch controls the production of flagella and toxins in *Clostridium difficile*. PLoS Genet 13, e1006701 (2017).

27. J. E. Emerson et al., A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. Mol Microbiol 74, 541-556 (2009).

28. D. Trzilova, B. R. Anjuwon-Foster, D. Torres Rivera, R. Tamayo, Rho factor mediates flagellum and toxin phase variation and impacts virulence in *Clostridioides difficile*. PLoS pathogens 16, e1008708 (2020).

29. R. W. McKee, C. K. Harvest, R. Tamayo, Cyclic diguanylate regulates virulence factor genes *via* multiple riboswitches in *Clostridium difficile*. mSphere 3 (2018).

30. E. Bordeleau, L. C. Fortier, F. Malouin, V. Burrus, c-di-GMP turn-over in *Clostridium difficile* is controlled by a plethora of diguanylate cyclases and phosphodiesterases. PLoS Genet 7, e1002039 (2011).

31. J. Peltier et al., Cyclic diGMP regulates production of sortase substrates of *Clostridium difficile* and their surface exposure through ZmpI protease-mediated cleavage. J Biol Chem 290, 24453-24469 (2015).

32. J. Hor, S. A. Gorski, J. Vogel, Bacterial RNA biology on a genome scale. Mol Cell 70, 785-799 (2018).

33. C. Kroger et al., An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. Cell Host Microbe 14, 683-695 (2013).

34. C. Kröger et al., The primary transcriptome, small RNAs and regulation of antimicrobial resistance in *Acinetobacter baumannii* ATCC 17978. Nucleic Acids Res 46, 9684-9698 (2018).

35. D. Ryan, L. Jenniches, S. Reichardt, L. Barquist, A. J. Westermann, A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe *Bacteroides thetaiotaomicron*. Nature communications 11, 3557 (2020).

36. C. M. Sharma et al., The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464, 250-255 (2010).

37. C. M. Sharma, J. Vogel, Differential RNA-seq: the approach behind and the biological insight gained. Curr Opin Microbiol 19, 97-105 (2014).

38. S. Wydau-Dematteis et al., Cwp19 Is a novel lytic transglycosylase involved in stationary-phase autolysis resulting in toxin release in *Clostridium difficile*. MBio 9 (2018).

39. J. T. Wade, D. C. Grainger, Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. Nat Rev Microbiol 12, 647-653 (2014).

40. F. Pratto et al., *Streptococcus pyogenes* pSM19035 requires dynamic assembly of ATP-bound ParA and ParB on parS DNA during plasmid segregation. Nucleic Acids Res 36, 3676-3689 (2008).

41. N. Kint et al., The sigma(B) signalling activation pathway in the enteropathogen *Clostridioides difficile*. Environ Microbiol 21, 2852-2870 (2019).

42. L. Saujet et al., Genome-wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. PLoS Genet 9, e1003756 (2013).

43. M. Serrano et al., A recombination directionality factor controls the cell type-specific activation of sigmaK and the fidelity of spore development in *Clostridium difficile*. PLoS Genet 12, e1006312 (2016).

44. E. C. Woods, A. N. Edwards, K. O. Childress, J. B. Jones, S. M. McBride, The *C. difficile clnRAB* operon initiates adaptations to the host environment in response to LL-37. PLoS pathogens 14, e1007153 (2018).

45. K. Pishdadian, K. A. Fimlaid, A. Shen, SpoIIID-mediated regulation of sigmaK function during *Clostridium difficile* sporulation. Mol Microbiol 95, 189-208 (2015).

46. A. A. Shishkin et al., Simultaneous generation of many RNA-seq libraries in a single reaction. Nature methods 12, 323-325 (2015).

47. G. E. Johnson, J. B. Lalanne, M. L. Peters, G. W. Li, Functionally uncoupled transcription-translation in *Bacillus subtilis*. Nature 585, 124-128 (2020).

48. X. Ju, D. Li, S. Liu, Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. Nature microbiology 4, 1907-1918 (2019).

49. M. R. Hemm, J. Weaver, G. Storz, *Escherichia coli* small proteome. EcoSal Plus 9 (2020).

50. P. Garai, A. Blanc-Potard, Uncovering small membrane proteins in pathogenic bacteria: regulatory functions and therapeutic potential. Mol Microbiol 10.1111/mmi.14564 (2020).

51. G. Storz, Y. I. Wolf, K. S. Ramamurthi, Small proteins can no longer be ignored. Annual review of biochemistry 83, 753-777 (2014).

52. A. Maikova et al., Discovery of new type I toxin-antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. Nucleic Acids Res 10.1093/nar/gky124 (2018).

53. O. Soutourina, Type I toxin-antitoxin systems in *Clostridia*. Toxins (Basel) 11 (2019).

54. B. Yan, M. Boitano, T. A. Clark, L. Ettwiller, SMRT-Cappable-seq reveals complex operon variants in bacteria. Nature communications 9, 3676 (2018).

55. S. Sáenz-Lahoya et al., Noncontiguous operon is a genetic organization for coordinating bacterial gene expression. Proc Natl Acad Sci U S A 116, 1733-1738 (2019).

56. P. D. Karp et al., The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform 20, 1085-1093 (2019).

57. I. Irnov, C. M. Sharma, J. Vogel, W. C. Winkler, Identification of regulatory RNAs in *Bacillus subtilis*. Nucleic Acids Res 38, 6637-6651 (2010).

58. M. L. Donnelly, K. A. Fimlaid, A. Shen, Characterization of *Clostridium difficile* spores lacking either SpoVAC or dipicolinic acid synthetase. J Bacteriol 198, 1694-1707 (2016).

59. I. Kalvari et al., Non-coding RNA analysis using the Rfam database. Current protocols in bioinformatics 62, e51 (2018).

60. N. Choonee, S. Even, L. Zig, H. Putzer, Ribosomal protein L20 controls expression of the *Bacillus subtilis infC* operon *via* a transcription attenuation mechanism. Nucleic Acids Res 35, 1578-1588 (2007).

61. A. M. Babina, D. J. Parker, G. W. Li, M. M. Meyer, Fitness advantages conferred by the L20-interacting RNA *cis*-regulator of ribosomal protein synthesis in *Bacillus subtilis*. Rna 24, 1133-1143 (2018).

62. Y. Fujita, Carbon catabolite control of the metabolic network in *Bacillus subtilis*. Bioscience, biotechnology, and biochemistry 73, 245-259 (2009).

63. T. Riedel et al., High metabolic versatility of different toxigenic and non-toxigenic *Clostridioides difficile* isolates. Int J Med Microbiol 10.1016/j.ijmm.2017.05.007 (2017).

64. M. L. Jenior, J. L. Leslie, V. B. Young, P. D. Schloss, *Clostridium difficile* colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. mSystems 2 (2017).

65. 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 (2020).

66. 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).

67. 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).

68. J. K. Christiansen et al., Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. Rna 12, 1383-1396 (2006).

69. 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).

70. 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).

71. 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).

72. S. M. K. Schoenfelder et al., The small non-coding RNA RsaE influences extracellular matrix composition in *Staphylococcus epidermidis* biofilm communities. PLoS pathogens 15, e1007618 (2019).

73. S. Pitman, K. H. Cho, The Mechanisms of virulence regulation by small noncoding RNAs in low GC Gram-positive pathogens. International journal of molecular sciences 16, 29797-29814 (2015).

74. 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).

75. K. G. Kaval et al., Loss of ethanolamine utilization in *Enterococcus faecalis* increases gastrointestinal tract colonization. mBio 9 (2018).

76. 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).

77. S. DebRoy et al., Riboswitches. A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. Science 345, 937-940 (2014).

78. J. R. Mellin et al., Riboswitches. Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA. Science 345, 940-943 (2014).

79. M. F. Del Papa, M. Perego, Ethanolamine activates a sensor histidine kinase regulating its utilization in *Enterococcus faecalis*. J Bacteriol 190, 7147-7156 (2008).

2.3.11 Supplementary Information

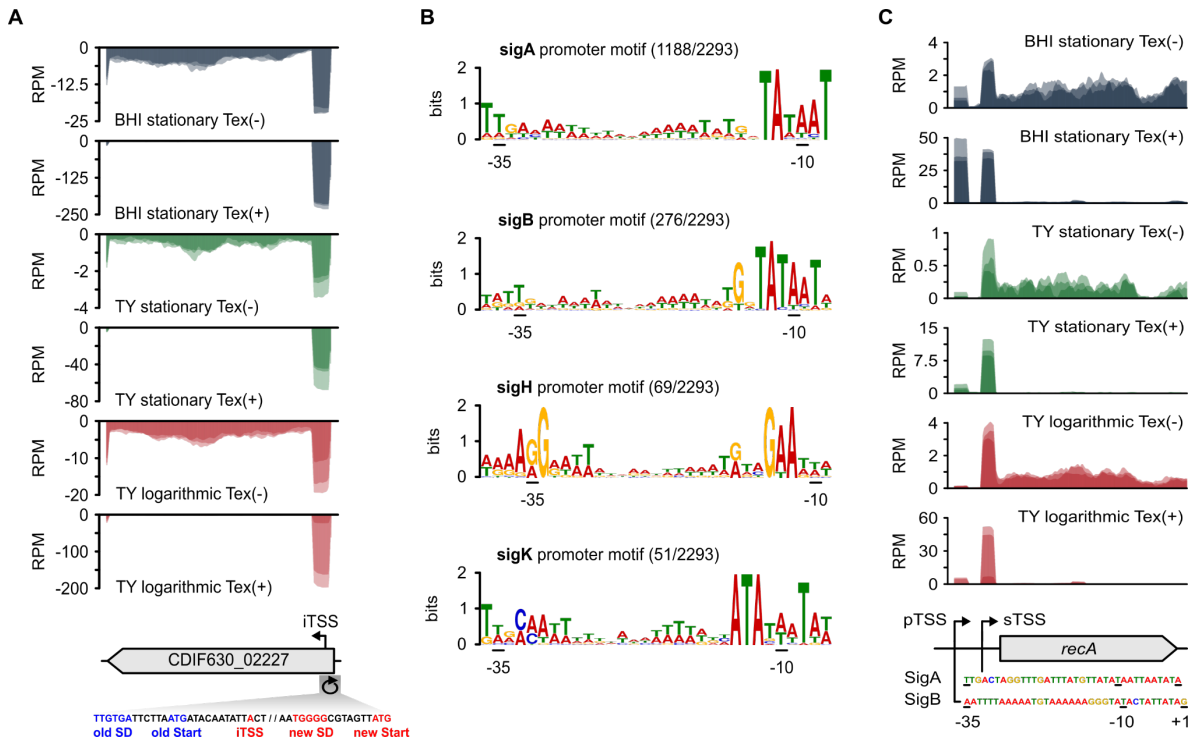


Figure S1. Transcriptome-based ORF annotation and global promoter features associated with transcriptional start sites. (A) Re-annotation of CDIF630_02227 based on a newly identified internal TSS 30 bases downstream of its annotated start codon. Reads of dRNA-seq libraries (TEX+/TEX-) mapping to the CDIF630_02227 region are shown above the sequence covering the region around the old and new start codon. The old and new AUGs and associated Shine-Dalgarno (SD) sequences are indicated. For better visibility the sequence is shown in reversed orientation (indicated by a circular arrow). **(B)** Promoter regions associated with detected TSSs were analyzed using the MEME suite. Promoter motifs were recovered for SigA, SigH, SigB and SigK. **(C)** Two alternative promoter sequences for SigA and SigB are associated with a pTSS and sTSS of recA, respectively.

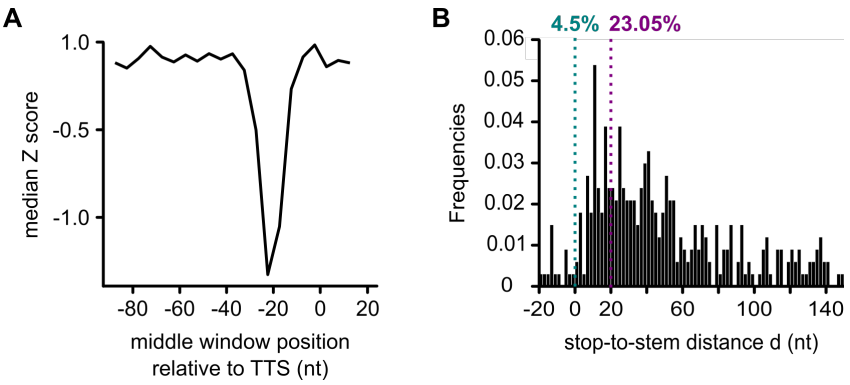


Figure S2. Global features associated with transcription termination sites (TTS). (A) Z-score of minimal free energies (MFE) of 3' regions. MFE was calculated using a 25 nt sliding window and sliding steps of 5 nt, and compared against randomly shuffled sequences. The resulting median Z score is plotted against the middle window position relative to the detected TTS. **(B)** Genome-wide distribution of stop-to-stem distances for high-confidence intrinsic terminators (n=380), based on data published in (34). ORF-overlapping terminators (d ≤ 0 nt) account for 4.5%, and terminators within 20 nt to a preceding stop codon account for 23% of all terminators.

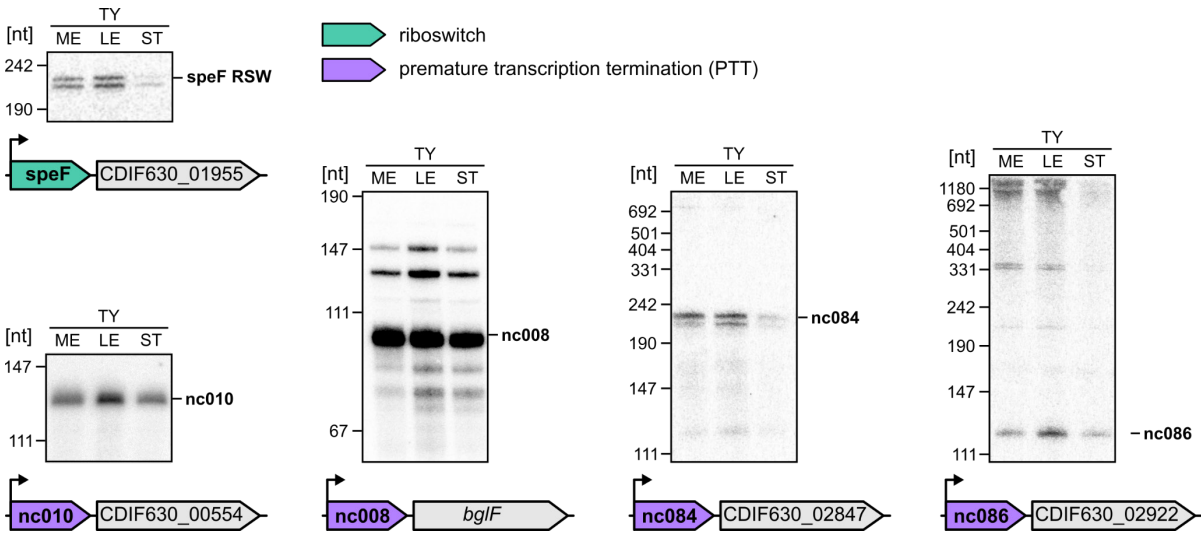


Figure S3. Expression analysis of novel 5' UTR *cis*-regulatory elements. Total RNA was extracted at mid-exponential, late-exponential and stationary growth from *C. difficile* 630 grown in TY medium and analyzed by northern blot using radioactively labeled DNA probes specific for a putative speF riboswitch (turquois) and several PTT events (purple) within 5' UTR regions of the indicated gene.

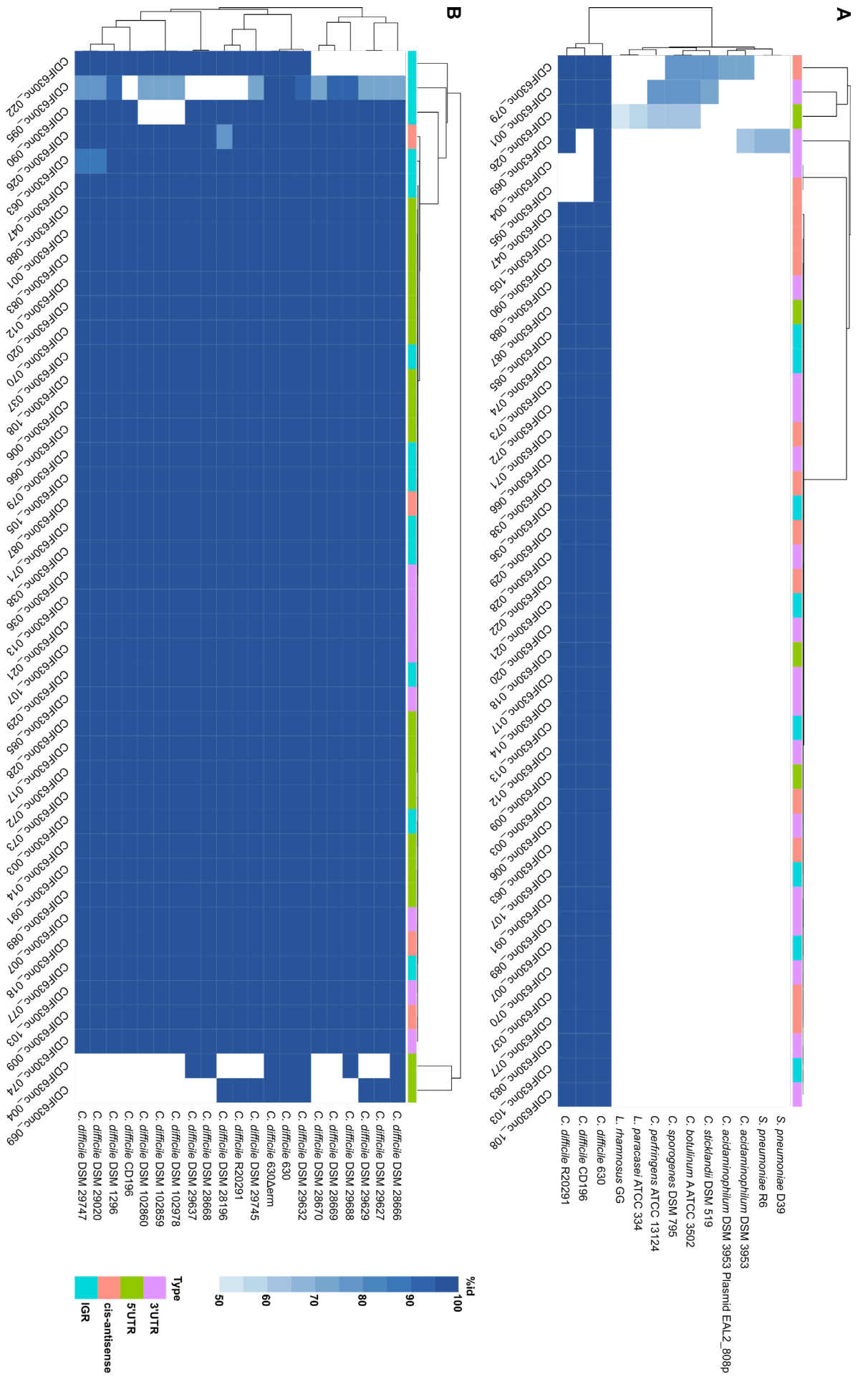

A
B
S. pneumoniae D39
S. pneumoniae R6
C. acidaminophilum DSM 3953
C. acidaminophilum DSM 3953 Plasmid EAL2_808p
C. sticklandii DSM 519
C. botulinum A ATCC 3502
C. sporogenes DSM 795
C. perfringens ATCC 13124
L. paracasei ATCC 334
L. rhamnosus GG
C. difficile 630
C. difficile CD196
C. difficile R20291
C. difficile DSM 28666
C. difficile DSM 29627
C. difficile DSM 29629
C. difficile DSM 29688
C. difficile DSM 28669
C. difficile DSM 28670
C. difficile DSM 29632
C. difficile 630
C. difficile 630Δerm
C. difficile DSM 29745
C. difficile R20291
C. difficile DSM 28196
C. difficile DSM 28668
C. difficile DSM 29637
C. difficile DSM 102978
C. difficile DSM 102859
C. difficile DSM 102860
C. difficile CD196
C. difficile DSM 1296
C. difficile DSM 29020
C. difficile DSM 29747
%id
100
90
80
70
60
50
Type
3'UTR
5'UTR
cis-antisense
IGR

Figure S4. Conservation of *C. difficile* 630 sRNAs. Conservation of sRNAs was analyzed with profile Hidden Markov Models against a genome database of Refseq assemblies of the phylum Firmicutes. Percent identities of 3' UTR, 5' UTR, IGR and *cis*-antisense sRNA for selected strains of the phylum Firmicutes **(A)** and for a set of clinical *C. difficile* strains representing 5 major genomic clades and diverse toxin repertoires **(B)** are shown. For several hits in one genome, only the hit with the highest percent identity is shown. Rows and columns were hierarchically clustered with UPGMA (unweighted pair group method with arithmetic mean) algorithm.

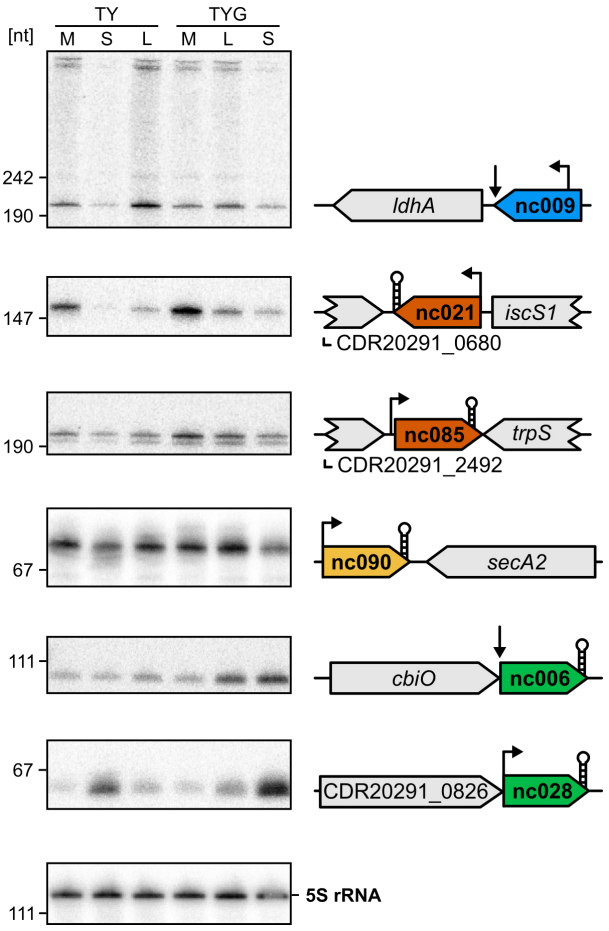


Figure S5. sRNA expression profiles are conserved in the hypervirulent RT029 isolate *C. difficile* R20291. Expression profiles of representative candidates for each sRNA class. Total RNA was extracted at mid-exponential (M), late-exponential (L) and stationary (S) growth phases from *C. difficile* R20291 grown in TY and TYG medium and analyzed by northern blot using radioactively labeled DNA probes. Color coding is based on the genomic location of each sRNA: blue=5' UTR, red=IGR, yellow=antisense and green=3' UTR located (see Figure 4). Note that late-exponential (L) and stationary (S) growth in TY are loaded in reversed order.

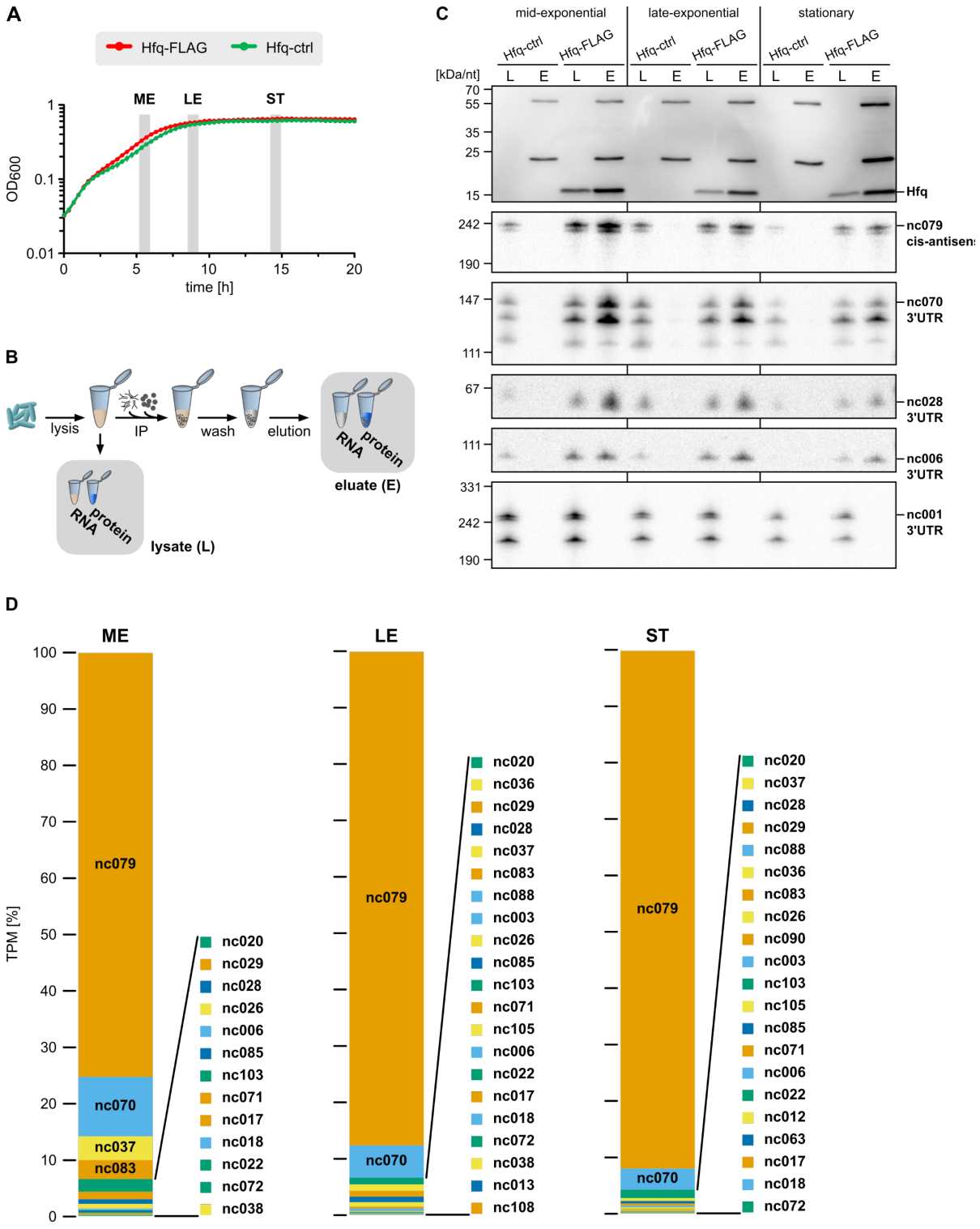


Figure S6. Hfq RIP-seq and the landscape of Hfq-associated ncRNA ligands. (A) Hfq RIP-seq was performed on *C. difficile* 630 WT and C-terminally 3xFLAG tagged hfq strains grown to mid-exponential, late-exponential and early-stationary phases in TY medium in two independent experiments. **(B)** Overview of RIP-seq workflow. **(C)** Pull-down of 3xFLAG tagged Hfq was validated by Western blot using anti-FLAG antibody. Hfq-associated RNAs were validated by northern blot using specific DNA probes. CDIF630nc_001 served as negative control since it was not enriched in the pull-down. **(D)** Distribution of reads matching experimentally annotated ncRNA candidates in Hfq-FLAG cDNA libraries at different phases of growth. Percentage indicates the reads (in transcripts per million) of a given sRNA compared to all sRNAs in a cDNA library.

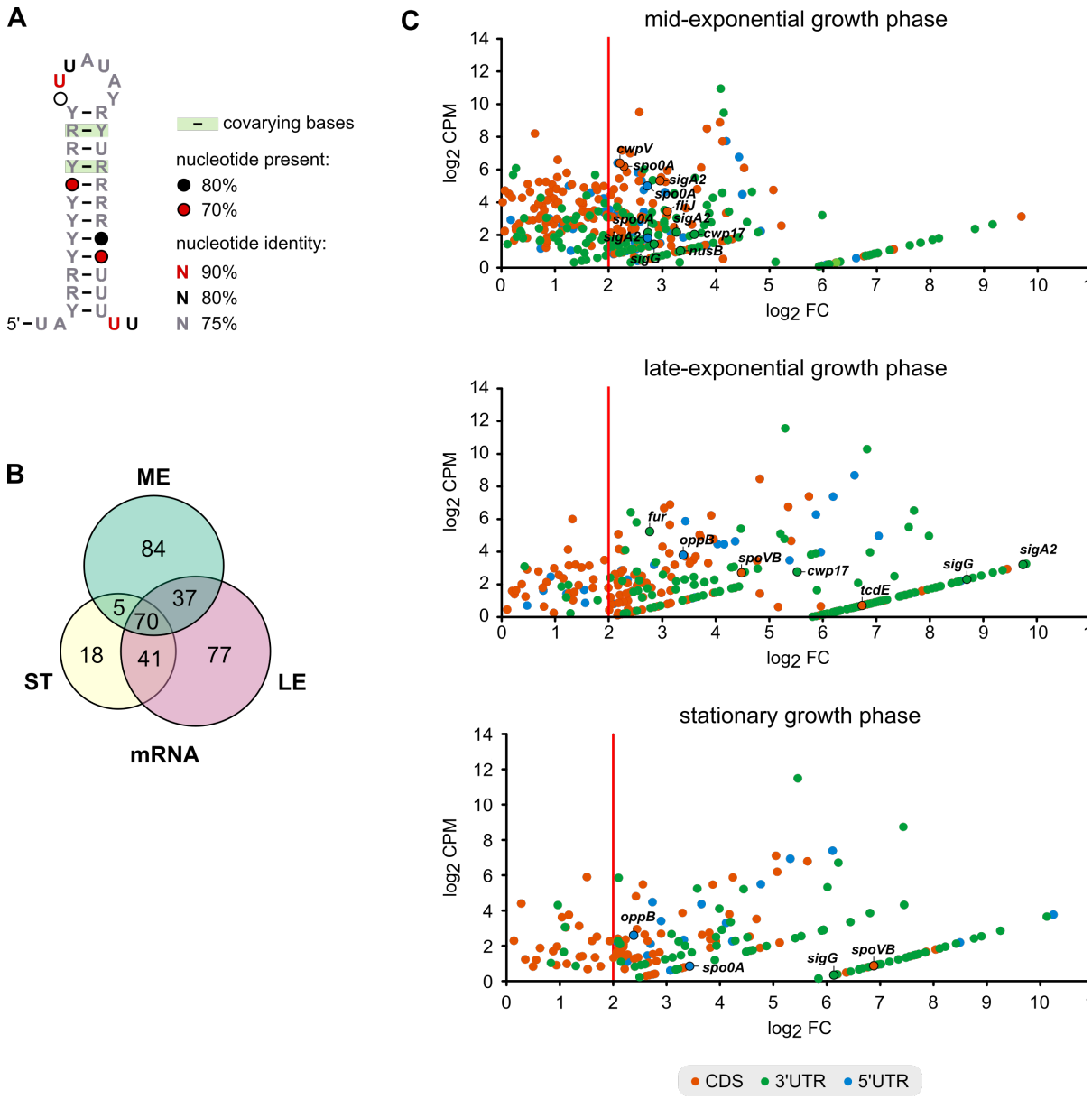


Figure S7. Consensus motif and spectrum of Hfq-associated mRNAs in *C. difficile* 630 across different growth phases. (A) Consensus motif generated by CMfinder based on all Hfq-associated peaks. **(B)** Number of Hfq-associated mRNAs at different growth phases. **(C)** Scatter-plot analysis of RIP-seq results for mRNAs that were enriched in each analyzed growth phase (log_2 FC ≥2; cDNA read ≥10; Benjamini-Hochberg corrected P-value ≤0.1) in the 3xFLAG tagged Hfq samples. Selected genes encoding for virulence-related proteins are labelled.

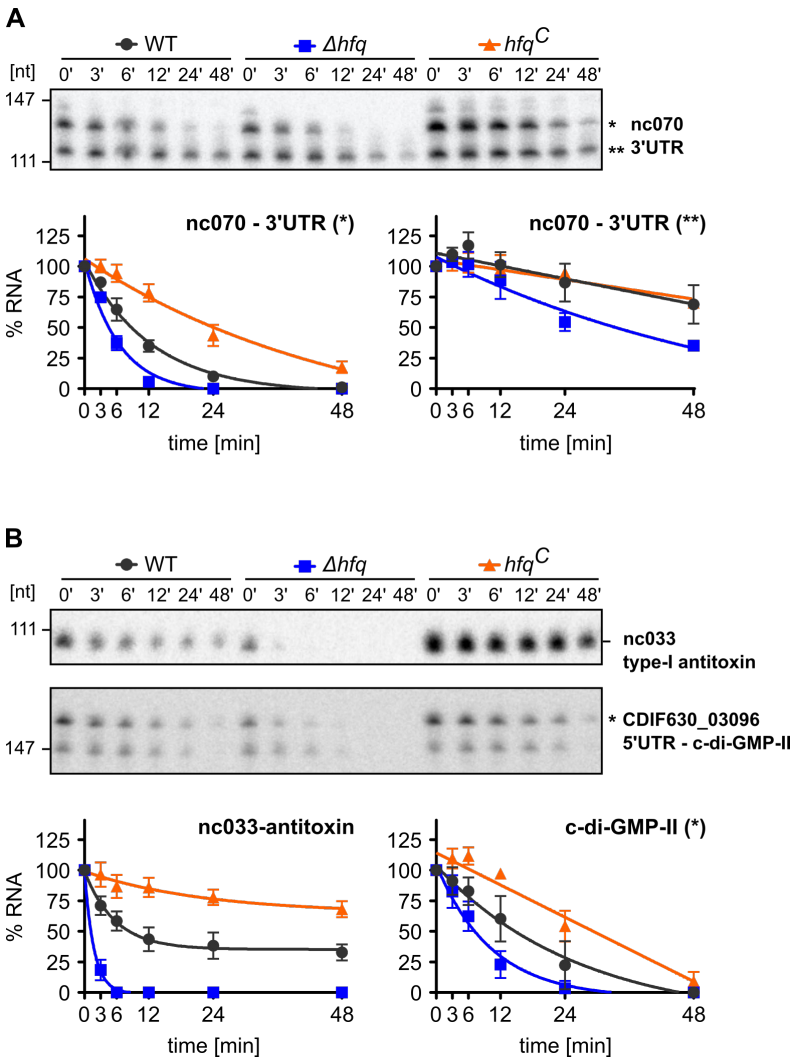


Figure S8. Stability of Hfq-associated ncRNA *in vivo*. Total RNA was extracted from WT, Δhfq, and complemented hfq C strains in late-exponential phase after inhibition of transcription by addition of rifampicin and analyzed by Northern blotting using radioactively labeled DNA probes against **(A)** selected sRNA ligands **(B)** Hfq-associated type-I antitoxin and c-diGMP-II riboswitch. Left: A representative of three independent Northern blot experiments is shown. Right: Relative sRNA levels compared to timepoint 0 (addition of rifampicin) were calculated based on three biological experiments. Error bars indicate standard deviation

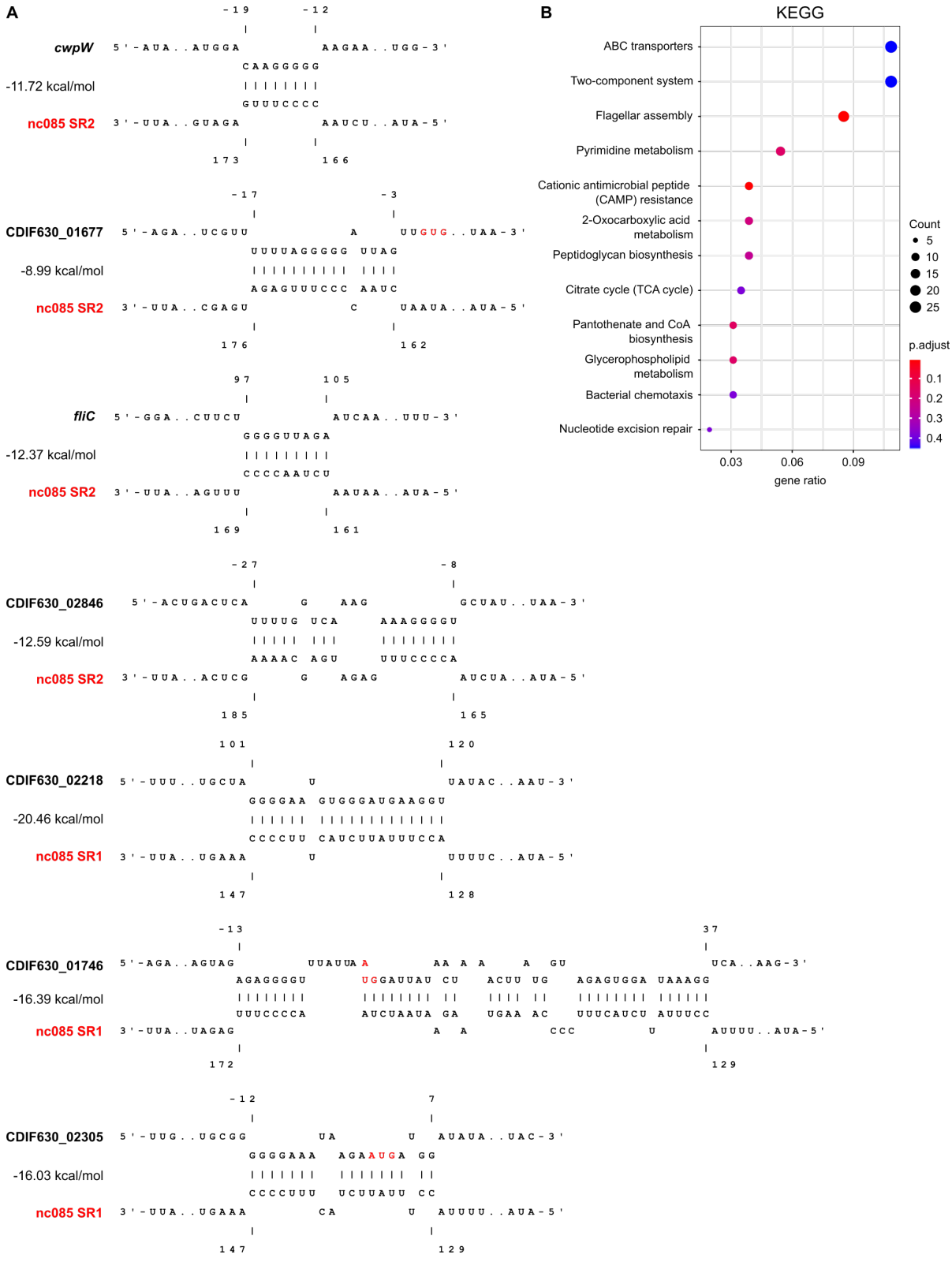


Figure S9. Putative target spectrum of CDIF630nc_085 identified by pulse-expression. (A) Selection of *in silico* predicted interaction sites (intaRNA) of CDIF630nc_085 with targets that were differentially regulated upon pulse expression of CDIF630nc_085. Interacting regions of mRNAs are labeled in relation to the start codon (highlighted in red). Examples of mRNAs interacting with either SR1 or SR2 of CDIF630nc_085 are given. **(B)** KEGG enrichment analysis for genes that were differentially expressed upon pulse expression of CDIF630nc_085.

2.4 A network of small RNAs regulates sporulation initiation in *Clostridioides difficile*

Manuela Fuchs [1,2], Vanessa Lamm-Schmidt [1,2], Tina Lenče [2], Johannes Sulzer [2], Arne Bublitz [3], Janet Wackenreuter [1], Milan Gerovac [2], Till Strowig [3,4], Franziska Faber [1,2] *

1. Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany
2. Julius-Maximilians-University of Würzburg (JMU), Faculty of Medicine, Institute for Molecular Infection Biology (IMIB), Germany
3. Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany
4. German Center for Infection Research (DZIF), partner site Hannover-Braunschweig, Germany

* Correspondence:

Franziska Faber: phone +49-931-3186280; email: franziska.faber@uni-wuerzburg.de

Published in: The EMBO Journal, e112858.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).

Competing Interest Statement: We declare no competing interests.

Keywords: *Clostridioides difficile*, RIL-seq, small RNA, Hfq, Spo0A

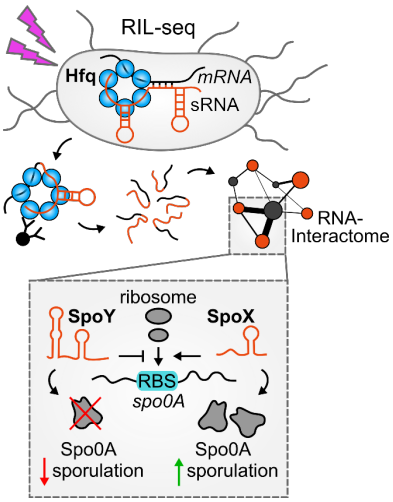
Manuscript modifications: The supplementary information is restricted to supplementary figures. All remaining supplementary information, including tables and datasets are available online.

2.4.1 Abstract

The obligate anaerobic, enteric pathogen *Clostridioides difficile* persists in the intestinal tract by forming antibiotic-resistant endospores that contribute to relapsing and recurrent infections. Despite the importance of sporulation for *C. difficile* pathogenesis, environmental cues, and molecular mechanisms regulating sporulation initiation remain ill defined. Here, using RIL-seq to capture the Hfq-dependent RNA-RNA interactome, we discovered a network of small RNAs that bind to mRNAs encoding sporulation-related genes. We show that two of these small RNAs, SpoX and SpoY, regulate translation of the master regulator of sporulation, Spo0A, in an opposing manner, which ultimately leads to altered sporulation rates. Infection of antibiotic-treated mice with SpoX and SpoY deletion mutants revealed a global effect on gut colonization and intestinal sporulation. Our work uncovers an elaborate RNA-RNA interactome controlling the physiology and virulence of *C. difficile* and identifies a complex post-transcriptional layer in the regulation of spore formation in this important human pathogen.

2.4.2 Synopsis

RIL-seq has revealed the RNA chaperone Hfq-dependent interactome of small RNAs and their targets in gram-negative bacteria. Here, application of RIL-seq in the human pathogen *C. difficile* uncovers the first Hfq-dependent RNA interactome in a gram-positive bacterium and identifies a pair of small regulatory RNAs that modulate sporulation initiation.



- RIL-seq in *C. difficile* strain 630 identifies Hfq-associated sRNA-mRNA pairs.
- RIL-seq analysis identifies the master regulator of sporulation, Spo0A, as a central target of sRNA-based regulation.
- The small RNAs SpoY and SpoX act as negative and positive regulators, respectively, of sporulation by base-pairing with the *spo0A* mRNA.
- Deletion of either SpoY or SpoX affects gut colonization in a mouse model of *C. difficile*-induced colitis.

2.4.3 Introduction

Since its discovery as a causative agent of antibiotic-associated pseudomembranous colitis, *Clostridioides difficile* (*C. difficile*) has emerged as the leading cause of nosocomial antibiotic-associated disease in the developed world (1–3). Several virulence traits contribute to disease severity of *C. difficile* infections (CDI), including exotoxin production and spore formation (4). In particular, spores are a key element in host transmission and disease recurrence, due to their

resistance to conventional antibiotics, disinfectants and other environmental stressors (5–7). Hence, understanding the environmental signals and molecular mechanisms that control spore formation in this important human pathogen is essential for the development of alternative treatment options.

Spore formation has been studied extensively in a variety of sporulating bacteria and represents an energetically costly, morphogenic process that is irreversible beyond a certain point in spore development (8,9). In particular, sporulation initiation is tightly controlled through the integration of environmental and nutritional signals that mediate the post-translational activation of the master regulator of sporulation, Spo0A (10,11). However, *C. difficile* lacks many of the known conserved regulatory mechanisms that activate Spo0A, rendering sporulation initiation a poorly understood process in this gram-positive pathogen (12). Once activated, phosphorylated Spo0A-P acts as a transcriptional regulator that induces the expression of a set of early sporulation genes. This ultimately leads to the hierarchical activation of four compartment-specific sigma factors – σ^E/σ^K in the mother cell and σ^F/σ^G in the forespore – and culminates in the formation of a metabolically dormant spore (13).

Most recently, post-transcriptional regulation mediated by the RNA binding protein (RBP) Hfq has been implicated in modulating sporulation in *C. difficile* (14,15). Boudry *et al.* demonstrated that depletion of Hfq leads to the upregulation of several sporulation related genes *as well as an* increased sporulation rate (14,15). Hfq is known for its ability to facilitate base-pairing between small regulatory RNAs (sRNAs) and their target mRNAs, leading to altered translational efficiency and mRNA stability (16). Similar to its extensively studied gram-negative counterparts, Hfq immunoprecipitation followed by sequencing of bound RNA species (RIP-seq) in *C. difficile* uncovered a vast number of sRNAs and mRNAs bound by Hfq (17,18). Furthermore, several sRNAs, not only in *C. difficile* but also in other spore-forming Firmicutes, have been associated with the sporulation process, mostly through RNA-seq and microarray-based expression profiles (14,19–21). However, only a few of these sRNAs have been functionally described. *In C. difficile,* sRNA RCd1, that inhibits the production of the late mother cell-specific sigma factor σ^K, remains the only sporulation-associated sRNA characterized to this date (17), revealing a paucity of knowledge that clearly warrants further investigation.

Global approaches such as RIP-seq are powerful tools in discovering RBP-bound sRNAs or mRNAs (18,22). However, they rely on additional experimental and computational assays to identify directly interacting sRNA-target pairs (23). Melamed and colleagues circumvented this difficulty by introducing RIL-seq (RNA interaction by ligation and sequencing) to the field of bacterial RNA-biology (24,25). Similarly to CLASH and hiCLIP, RIL-seq relies on ligation of RBP-bound RNA pairs and thereby directly captures and identifies interaction partners (26,27).

In the present study, we applied Hfq RIL-seq to *C. difficile,* which led to the discovery of an extensive Hfq-mediated sRNA-target network. Among the identified sRNA-mRNA interactions were several sRNAs bound to the *spo0A* mRNA, encoding the master regulator of sporulation. We show that two of these sRNAs, SpoY and SpoX, regulate *spo0A* translation in an opposite manner *in vivo*, resulting in altered sporulation rates. Furthermore, SpoY and SpoX deletion significantly impacts *C. difficile* gut colonization and spore burden in a mouse model of *C. difficile* infection. Overall, we provide the first example of sRNAs regulating sporulation initiation by fine-tuning *spo0A* translation, which adds a new layer of post-transcriptional regulation to the complex process of sporulation initiation in this important human pathogen.

2.4.4 Results

2.4.4.1 Hfq is a global RNA binding protein that mediates sRNA-mRNA interactions in *C. difficile*

To better understand the impact of post-transcriptional regulation on sporulation, we performed Hfq RIL-seq in *C. difficile* in sporulating conditions (25). *C. difficile* 630 cells expressing a chromosomally FLAG-tagged Hfq variant (Hfq-FLAG, n=4) were harvested during the transition phase (early stationary phase), when *C. difficile* shifts to a non-growing state, accompanied by sporulation to ensure survival in nutrient limiting conditions (28,29). Harvested cells were UV-crosslinked to stabilize *in vivo* protein-RNA interactions, followed by cell lysis and Hfq co-immunoprecipitation. Identification of Hfq-associated RNA-RNA interaction partners was achieved by ligation of Hfq-bound RNA pairs ("chimeras"), followed by RNA purification, sequencing, and computational analysis using a previously published primary transcriptome annotation of *C. difficile* 630 (Figure 1A) (18). *C. difficile* 630 expressing native Hfq (WT) served as a control and was treated similarly (n=4). Analysis of the RIL-seq data revealed a high number of Hfq-bound single and chimeric fragments with a considerable enrichment of chimeric reads in the Hfq-FLAG strain, when compared to the WT (Figure 1B). The list of chimeras was manually curated and further reduced to statistically relevant interactions (Odds ratio ≥1 and p-value <0.05) that are represented by at least 25 chimeric fragments (25). All remaining interactions are listed in Dataset EV2&3. The resulting RIL-seq network is publicly available and explorable in an RNA-RNA interactome browser (https://resources.helmholtz-hiri.de/rilseqcd/, Appendix Figure S2). In accordance with existing *E. coli* and *S. enterica* RIL-seq data, most chimeras (67%) consisted of mRNA-sRNA interactions, with mRNAs (5'UTR, CDS or 3'UTR) at position 1 (RNA1/5'end) and sRNAs at position 2 (RNA2/3'end), as shown in Figure 1C&D and Appendix Figure S1B. Although most sRNAs were predominantly found at position 2, such as nc159, a few sRNAs showed a clear preference for position 1, including nc083 (Appendix Figure S1B). However, of all chimeric fragments mapping to sRNAs more than 90% mapped to RNA2 (Figure 1D). This

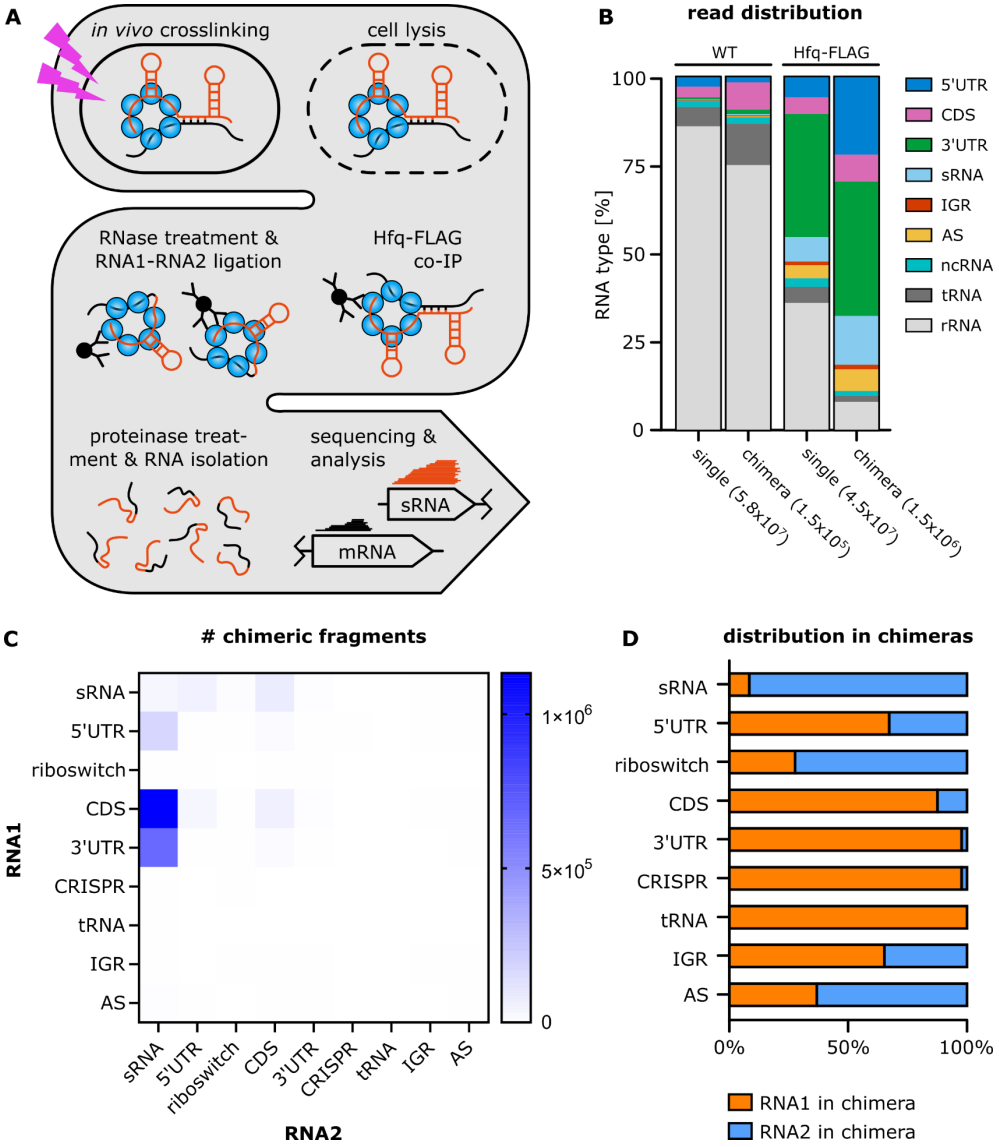


Figure 1: RIL-seq establishes Hfq as a platform for RNA-RNA interactions in *C. difficile*. (A) Schematic representation of the RIL-seq workflow. **(B)** Distribution of all reads, single and chimeric, across all RNA classes, comparing Hfq-FLAG and control (WT) strain (n=4 each). ncRNAs include riboswitches, tmRNA, SRP RNA, RNase P RNA, and 6S RNA. All chimeras were included, without filtering for statistical significance or manual curation. **(C&D)** Distribution of RNA classes in chimeric fragments based on Dataset EV3, where RNA1 constitutes the 5'end and RNA2 the 3'end of a chimera. Only statistically relevant interactions (Fisher's exact test ≤ 0.05) that are represented by at least 25 chimeric fragments are included.

position bias reflects the mechanism by which most sRNAs bind Hfq in gram-negative species. Interactions generally occur between the proximal face of Hfq and the distinct intrinsic terminator and poly-U tail that characterizes most sRNAs, ultimately rendering the sRNA 3'end inaccessible to proximity ligation (Figure 1A) (30). Accordingly, our data imply that *C. difficile* Hfq might employ binding mechanisms similar to those described for *S. enterica* and *E. coli* in facilitating sRNA target interactions (30). While the majority of sRNAs were found ligated to CDSs (56%), a surprisingly high number (33%) interacted with mRNA 3'UTRs (Figure 1C). Recently published

C. difficile Hfq RIP-seq data revealed similar distributions of Hfq-bound RNA species, however, RIP-seq does not allow identification of direct interaction partners (18). In contrast, sRNA-mRNA chimeras in *E. coli* and *S. enterica* were clearly dominated by sRNAs interacting with CDSs or 5'UTRs, while sRNA-3'UTR ligations were barely found (24,31). Although bacterial 5'UTRs have long been described as the prototypical target of sRNA-mediated post-transcriptional regulation, there are examples of sRNAs targeting mRNA 3'UTRs, including sRNA Spot42 targeting the 310-nt long *hilD* 3'UTR, a transcriptional regulator of virulence in *S. enterica* (32–34). Indeed, research on *S. aureus* suggests that long 3'UTRs in particular might be an underrated source of regulatory elements that impact transcript stability and translation (32,35). Considering that in *C. difficile* 42% of all annotated 3'UTRs are longer than 100 nt, they might constitute a source of regulatory elements targeted by sRNAs (18).

2.4.4.2 Hfq RIL-seq identifies novel sRNA candidates

Recent publications suggest that RIL-seq network data can be exploited to identify new sRNAs by taking into account unique features of sRNAs in general and sRNA RIL-seq chimeras in particular (31,36). Accordingly, a high number of chimeric fragments mapping to a single RNA has been identified as a promising indicator of potential new sRNAs (36). This is reflected in the formation of "interaction hubs" that consist of a dominating, single RNA interacting with a large number of unique RNAs (31). By mapping all chimeric fragments to the *C. difficile* genome we could identify 24 interaction hubs formed by sRNA candidates that were previously unknown (n=15) or non-validated (n=9) sRNAs (Figure 2A, Dataset EV4) (18,37,38). Subsequent northern blot analysis of sRNA expression during different growth stages (Figure 2B) confirmed the expression of six out of eight tested sRNA candidates (Figure 2A C). Expression profiles of these sRNAs indicated expression mainly during late exponential/early stationary growth phase (Figure 2A C), coinciding with the growth stage selected for our RIL-seq experiment (early stationary). Accordingly, performing RIL-seq in distinct growth conditions has the potential to uncover novel sRNA candidates that have evaded previous detection approaches such as RIP-seq, due to its unique ability to reveal both Hfq-association and RNA-RNA interaction (24).

2.4.4.3 RIL-seq data suggests sRNA-mediated discoordinate regulation of operons in *C. difficile*

In addition to the vast number of chimeras representing sRNA-mRNA pairs (n=1046), sRNA-encompassing interactions also included chimeras consisting of sRNA-sRNA (n=39) and sRNA-intergenic region (IGR) ligations (n=24, Dataset EV2). While sRNA-sRNA pairs have been discussed as a pool of potential sRNA sponges, sRNA-IGR chimeras have not been investigated previously (24,31,39). A detailed analysis of those interactions revealed that in several cases (n=8, 33%) "IGRs" represented non-coding regions in polycistronic mRNAs (Figure 2D). To further

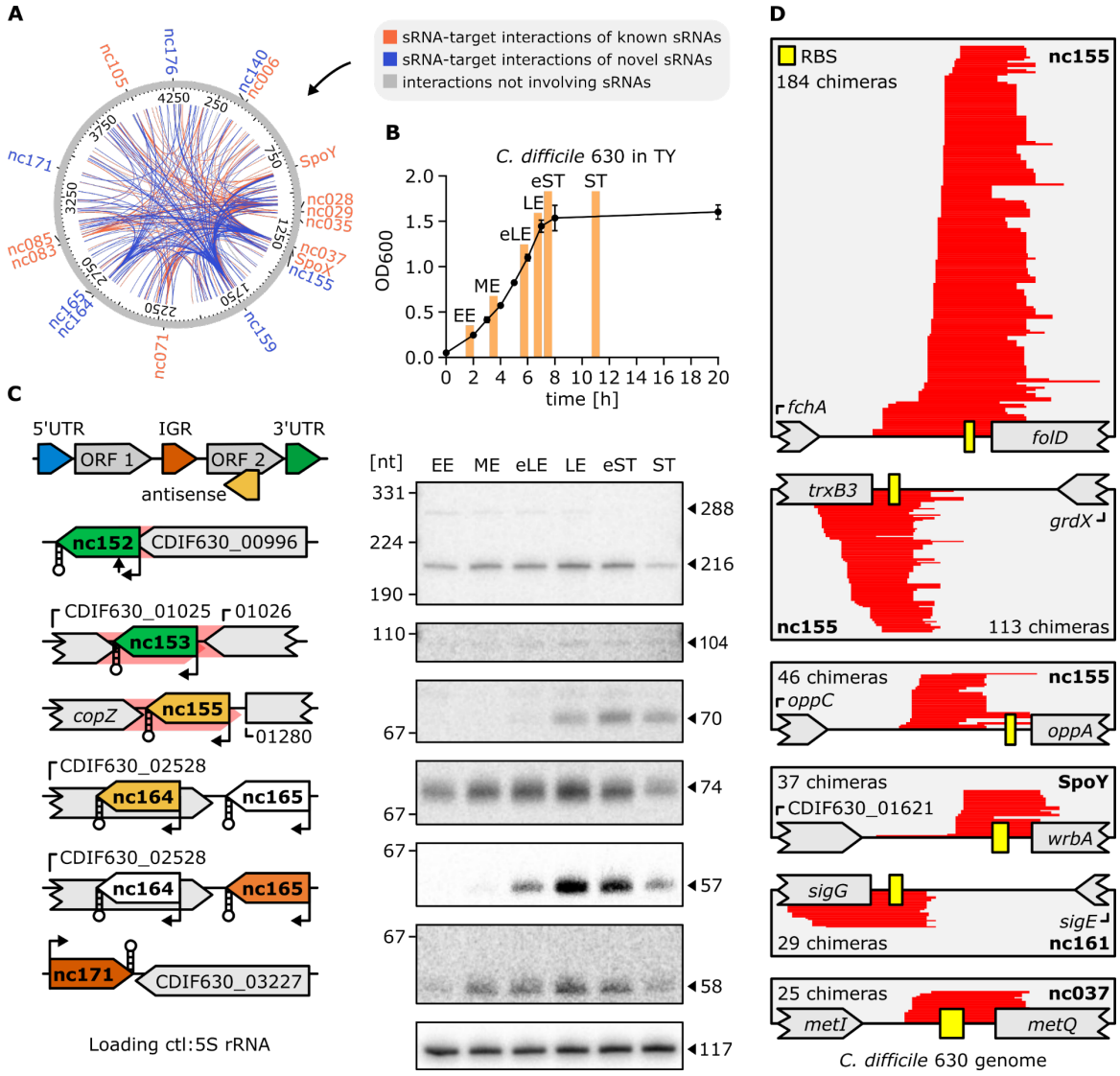


Figure 2: RIL-seq analysis facilitates annotation of novel sRNAs and reveals sRNA-mediated regulation of polycistronic transcripts. (A) Circos plot of all RIL-seq interactions that are represented by ≥ 200 chimeric fragments, mapped to the *C. difficile* 630 chromosome. Interaction hubs characterized by sRNAs with ≥ 20 unique interactions are labelled. Interactions involving known sRNAs are marked in orange, while interactions involving new sRNAs are highlighted in blue. All other interactions are grey. **(B&C)** Northern blot validation of new sRNAs. **(B)** Samples were taken in early exponential (EE), mid-exponential (ME), early late exponential (eLE), late exponential (LE), early stationary (eST), and stationary (ST) phase of growth in TY medium. The error bars represent the mean ± SD of four biological replicates **(C)** sRNAs are color-coded according to their genomic location. Stem-loop structures indicate intrinsic terminators, arrows indicate TSSs or processing sites, and red shadings untranslated regions [312]. A representative image of three independent experiments is shown. **(D)** Coverage plots of several sRNA-IGR chimeras, where the IGR is located within poly-cistronic operons. Chimeric reads are highlighted in red; the RBS is depicted as a yellow box. The number of chimeras covering each interaction is provided and the respective interacting sRNA is highlighted in bold.

understand the impact of sRNA-mediated regulation on operon expression in *C. difficile*, we compared our RIL-seq data with previously published operon annotations (18). We found that 383 RIL-seq chimeras mapped to 170 out of 400 known operons. Of these 383 chimeras, 98 constituted sRNAs interacting with intra-operon ribosome binding site (RBS) regions (25 nt up-

and 20 nt downstream of the respective start codon), indicating potentially widespread coordinate and discoordinate regulation of polycistronic mRNAs in *C. difficile* (Dataset EV5). While sRNA-mediated coordinate regulation of entire operons is more common, there are several reports of sRNAs targeting individual genes in polycistronic mRNAs, thereby only affecting a subset of genes within an operon (discoordinate regulation) (40,41). For instance, in *E. coli*, RyhB targets the *iscRSUA* operon, selectively inhibiting translation of *iscS* and resulting in the degradation of the *iscSUA* part (42,43). Hence, discoordinate regulation of operons allows bacteria to selectively produce operon components, *e,g*. when only a specific gene product is needed in a given condition (40). In line, our RIL-seq dataset comprised chimeras formed by the RBS region of the sporulation-specific sigma factor *sigG* and the newly annotated sRNA CDIF630nc_161 (Figure 2D). *sigG* encodes the forespore-specific late sporulation sigma factor σ^G and constitutes the last gene in an operon formed by two additional sporulation specific genes, including *sigE* directly upstream of *sigG* (44). In contrast to σ^G, the *sigE*-encoded sigma factor σ^E is active in the mother cell during early sporulation. Accordingly, both sigma factors not only operate during different stages of sporulation, but also in different compartments, and consequently require a tight regulation (44). Hence, sRNA-mediated post-transcriptional regulation might finetune the sequential expression of both sigma factors to ensure correct spore development. Nevertheless, a more detailed analysis is needed to fully understand the nature and extent of these regulatory events in *C. difficile*, not only on sporulation but on cellular processes in general.

2.4.4.4 The master regulator of sporulation, Spo0A, is a central target of sRNA-based regulation

Interestingly, *sigG* was only one among several sporulation specific genes enriched in our RIL-seq dataset. Additional sporulation related genes included *sigE*, *sleB*, *spoIIAB*, *spoIVA* and *spo0A*, encoding the master regulator of sporulation (Dataset EV2) (10). The latter was of particular interest since chimeras comprising *spo0A* and the sRNA nc020 as well as *spo0A* and sRNA nc038 were among the top 5 most abundant RIL-seq interactions in the entire dataset (>20,000 chimeras each). We decided to investigate these interactions in more detail and renamed both sRNAs to SpoY (nc020) and SpoX (nc038), to reflect their involvement in sporulation. SpoY is a 5'UTR-derived sRNA, sharing its transcription start site with CDIF630_00827, which encodes a protein of unknown function (Figure 3A&B). Northern blot analysis indicated a complex expression profile with the highest SpoY expression during the early- and mid-exponential growth phases, decreasing levels towards early stationary growth and increasing expression following entry into stationary phase when grown in TY (Figure 3C, growth phases are indicated in Figure 2B). We also monitored expression in additional media that are known to impact sporulation frequency, including BHI, TYG, TYF and 70:30 medium, which revealed expression of SpoY over a broad range

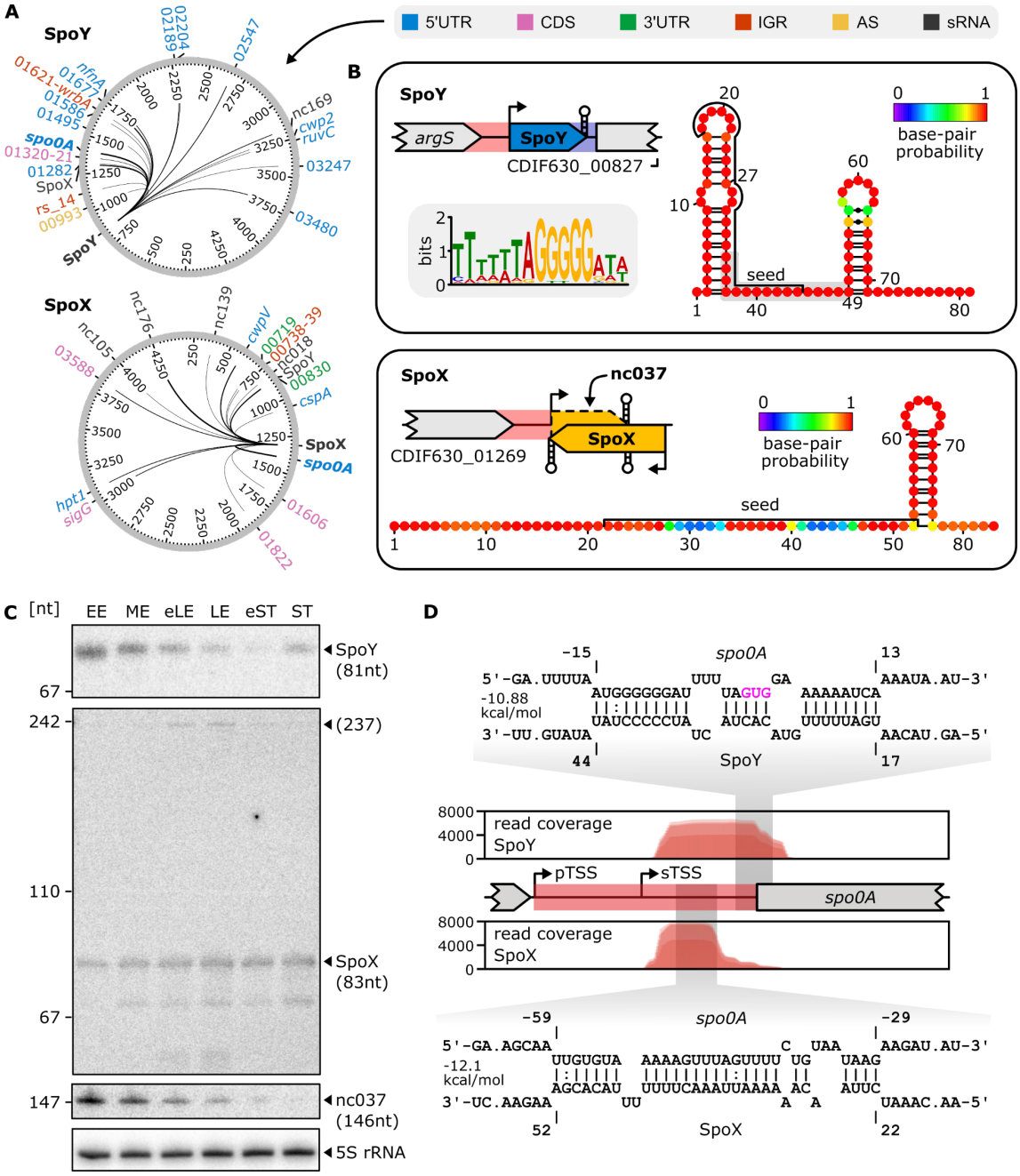


Figure 3: RIL-seq reveals *spo0A* as a target of sRNA-mediated post-transcriptional regulation. (A) Circos plots highlighting the SpoY and SpoX interactome, only interactions supported by ≥25 chimeras are included. Target types are discriminated by color. Edge strength correlates with the number of chimeras supporting an individual interaction. **(B)** Genomic location and predicted secondary structure (RNAfold (86)) of SpoY and the short SpoX isoform are provided. Seed regions relevant for *spo0A* interaction were predicted *in silico* (IntaRNA (47)) and are labeled in the secondary structure. For both sRNAs, target sequences were extracted from the RIL-seq data and uploaded to MEME (45), resulting in the successful identification of a common SpoY target motif (present in 24/28 target sequences). **(C)** Northern blot validation of SpoY, SpoX, and nc037 expression in early exponential (EE), mid-exponential (ME), early late exponential (eLE), late exponential (LE), early stationary (eST), and stationary (ST) phase of growth in TY medium. 5S rRNA served as a loading ctl. A representative image of three independent experiments is shown. **(D)** Read coverage of *spo0A* by SpoY-*spo0A* (top) and SpoX-*spo0A* (bottom) chimeric reads. Base pairing information and location of predicted interaction sites (IntaRNA (47)) are highlighted. The *spo0A* nucleotide position is calculated relative to the *spo0A* start codon (highlighted in pink). The sigA-dependent primary transcription start site (pTSS) and sigH-dependent secondary TSS are marked.

of conditions (Appendix Figure S3). MEME analysis of SpoY RIL-seq chimeras, including the *spo0A* interaction, suggested that SpoY preferably binds mRNA 5'UTRs at a G-rich target motif that resembles the RBS (Figure 3A&B, Figure EV1A&B) (45). Indeed, i*n silico* predictions of RNA-RNA interactions performed with IntaRNA indicated that SpoY binding blocks the *spo0A* RBS, which typically leads to translational inhibition (Figure 3D) (46,47).

In contrast to SpoY, SpoX is encoded partially antisense to another putative sRNA, nc037 (Figure 3B). Interestingly, previously published RNA-seq data revealed an intrinsic terminator within the SpoX sequence, resulting in a short (83 nt) and long isoform (237 nt, Figure 3B, Figure EV1B) (18). The long isoform is encoded antisense to nc037, while the short isoform terminates prior to the overlapping region. According to northern blot analysis, the short SpoX isoform is far more prevalent and expressed through all growth phases, while the long isoform appears barely expressed and is only detectable during late exponential growth phase (Figure 3C). Additional expression profiling in various media further indicated that SpoX expression tends to be upregulated in stationary phase (Appendix Figure S3). Interestingly, expression of the putative sRNA nc037 in TY is mostly anticorrelated (Figure 3C) to the expression of SpoX, potentially influencing expression of the long SpoX isoform. Further analyses will be necessary to assess the impact of nc037 on SpoX expression. The target spectrum of SpoX is diverse, including several CDSs and sRNAs in addition to the *spo0A* 5'UTR (Figure 3A, Figure EV1A). Consequently, identification of a conserved target motif using MEME was not successful (45). According to IntaRNA analysis and the peak profile of SpoX-*spo0A* chimeric reads, SpoX binds further upstream in the *spo0A* 5'UTR (Figure 3D) (47). *In silico* predictions of the secondary structures of SpoX and the *spo0A* mRNA upon duplex formation, suggested that SpoX-*spo0A* base pairing disrupts the *spo0A* 5'UTR secondary structure, potentially rendering the RBS more accessible to ribosome binding (Figure EV2A&B). In-line probing using ^{32}P-*spo0A* in combination with SpoX further supported this hypothesis (Figure EV2C). Of note, the predicted SpoX interaction site (seed region) involved in *spo0A* base-pairing is located at the 5' end of the SpoX sRNA and therefore present in both SpoX isoforms (Figure 3B&D, Figure EV1B) (47). However, SpoX-*spo0A* chimeric reads solely mapped to the short isoform, which suggested that the short rather than the long version of SpoX predominantly binds *spo0A* (Figure EV1B). In summary, while the RIL-seq data revealed that both, SpoX and SpoY interact with *spo0A*, their distinct interaction sites and chimeric read profiles suggested that the regulatory mechanisms applied by SpoX and SpoY differ.

2.4.4.5 SpoY and SpoX directly bind the spo0A mRNA *in vitro* and *in vivo*

To confirm the *in silico* predicted sRNA-*spo0A* base-pairing, we performed electrophoretic mobility shift assays (EMSAs), combining either SpoY or SpoX with *spo0A* (5'UTR and first 69 nt of CDS, Appendix Figure S4). Considering the 5' location of the SpoX seed region we decided to use

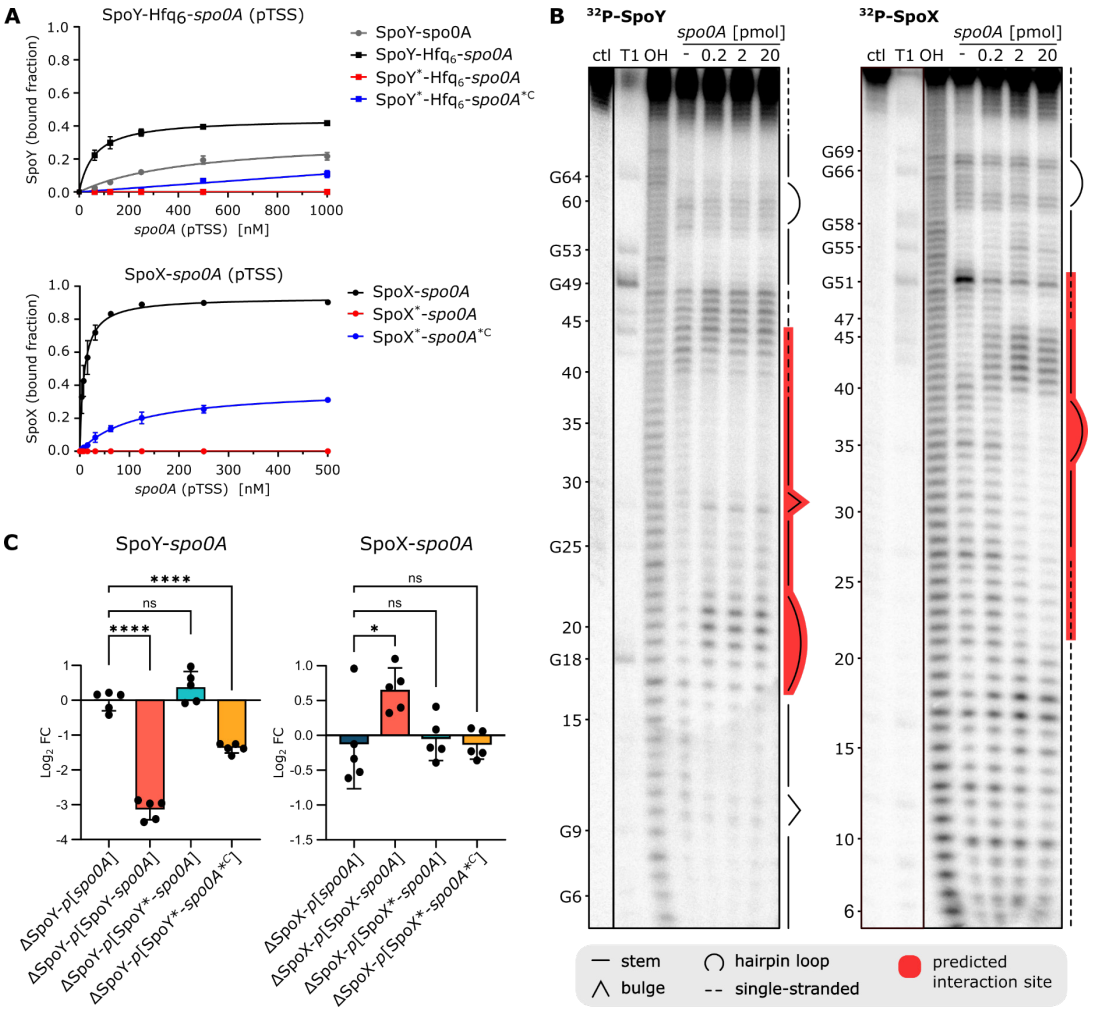


Figure 4: SpoY and SpoX directly interact with spo0A *in vitro* and *in vivo*. (A) Quantification of EMSAs (error bars represent the mean ± SD of n=3 biological replicates, Appendix Figure S4B&C) performed with either ^{32}P-labeled SpoY or SpoX (short isoform) with increasing concentrations of the spo0A target region, respectively. Purified Hfq was added to facilitate SpoY-spo0A complex formation. Mutating the respective sRNA seed region (Appendix Figure S4A, SpoY*/SpoX*) abolished the interaction, while introducing compensatory mutations into the *spo0A* target region (*spo0A*^{*C}) slightly rescued the complex formation. **(B)** In-line probing of 0.2 pmol of 32P-labeled SpoY and SpoX (short isoform) in the absence (lane 4) or presence of increasing concentrations (lane 5-7) of the *spo0A* long 5'UTR (starting from pTSS) and first 69 nt of the CDS. RNase T1 and alkali-digested (OH) SpoY and SpoX serve as ladders respectively. Secondary structure and predicted seed region are highlighted. A representative image of three independent experiments is shown. **(C)** mCherry fluorescence of translational fusion constructs (error bars represent the mean ± SD of n=5 biological replicates, Appendix Figure S5A) expressed in the respective sRNA knock-out background. Fluorescence intensity was normalized to that of the respective *p*[*spo0A*] ctl. Mutating the sRNA seed regions (Appendix Figure S4A, SpoY*/SpoX*) abolished the impact on *spo0A* translation, while introducing compensatory mutations into the *spo0A* target region (*spo0A*^{*C}) partially rescued the effect. The long SpoX isoform was used for reporter assays. Ordinary one-way ANOVA with Dunnett's multiple comparison test was used to calculate statistical significance. Not significant (ns) $P > 0.05$; (∗) $P \leq 0.05$; (∗∗∗∗) $P \leq 0.0001$.

the short SpoX isoform for *in vitro* experiments. For *spo0A,* we initially tested two 5'UTR lengths that correspond to the previously published sigA-dependent (long 5' UTR) and sigH-dependent (short 5' UTR) transcription start sites of *spo0A* (Figure 3D) (18,28). However, since EMSAs

combining SpoY or SpoX with either the long or the short *spo0A* 5'UTR showed similar results, we focused on the long 5'UTR for further experiments (Appendix Figure S4D). Complex formation of SpoY-*spo0A* required high concentrations of *spo0A*, but clearly improved upon addition of purified Hfq (Figure 4A, Appendix Figure S4B). The complex of SpoX and *spo0A* formed more efficiently, resulting in an apparent K_D of 8.7 nM (Figure 4A, Appendix Figure S4C). Mutating the respective sRNA seed regions (SpoY*/SpoX*) completely abolished the interaction in both cases, while introducing compensatory mutations in the respective *spo0A* target regions (*spo0A**C) restored the complex formation, albeit not to WT levels (Figure 4A, Appendix Figure S4A-C). In line probing analysis further corroborated these results. As shown in Figure 4B, duplex formation of 5'-end-labeled SpoY with *spo0A* (long 5'UTR and first 69 nt of CDS) protected SpoY from cleavage at positions 30-41, partially confirming the predicted interaction site. Base-pairing was even more apparent for SpoX, where a clear concentration-dependent effect could be observed, protecting SpoX from spontaneous cleavage at positions 22-39 & 47-52 upon duplex formation with *spo0A* (Figure 4B, Figure 3D, Figure EV2B). Based on these results we conclude that SpoY and SpoX interact with *spo0A in vitro via* direct base-pairing at distinct target sites in the *spo0A* mRNA. To further validate these interactions and their impact on *spo0A* translation *in vivo*, we designed a translational reporter system, expressing the long 5' UTR and first 20 aa of *spo0A* fused to mCherry (Appendix Figure S5A). The *spo0A* fusion construct was expressed alone (*p*[*spo0A*]) or in combination with either sRNA (*p*[SpoY- or SpoX-*spo0A*]) in the respective sRNA deletion mutant. In contrast to the *in vitro* approaches described above, the long SpoX isoform was used for all reporter assays to fully reflect the *in vivo* situation, including potential regulation by nc037. Constitutive co-expression of SpoY and the mCherry fusion construct significantly decreased fluorescence as compared to the ΔSpoY-*p*[*spo0A*] control (ctl), demonstrating that SpoY inhibits *spo0A* translation (Figure 4C). Mutating the SpoY seed region (SpoY*) eliminated this inhibitory effect, while introducing the corresponding compensatory mutations in *spo0A* (*spo0A*C*) restored the phenotype (Figure 4C, Appendix Figure S5A). Interestingly, SpoX had the opposite effect on *spo0A* translation, as co-expression of SpoX and the mCherry fusion construct resulted in an increase of *spo0A* translation and consequently mCherry fluorescence (Figure 4C). *spo0A* translation was restored to WT levels when co-expressing a SpoX* seed region mutant, however, introducing compensatory mutations in the *spo0A* 5'UTR did not restore the positive effect on translation. It is possible that the compensatory mutations interfere with the *spo0A* 5'UTR secondary structure, thereby preventing SpoX-mediated opening of the *spo0A* 5'UTR to ribosome binding, as suggested above (Figure EV2B). Overall, we were able to confirm that SpoY and SpoX directly base-pair with the *spo0A* mRNA *in vivo*, resulting in translational repression of *spo0A* by SpoY and increased translation of *spo0A* upon interaction with SpoX.

2.4.4.6 Post-transcriptional regulation of spo0A by SpoY and SpoX has opposing effects on sporulation

Considering the evident alteration of *spo0A* translation, we hypothesized that SpoY and SpoX impact native Spo0A protein levels *in vivo*. Accordingly, we performed western blot analysis to compare Spo0A protein levels in a WT strain (*p*[ctl]) to those in the SpoY and SpoX deletion mutants (e.g., ΔSpoY-*p*[ctl]), or to strains constitutively over-expressing the respective sRNA (e.g., ΔSpoY-*p*[SpoY]). Although there was no effect on Spo0A in a SpoY deletion mutant, overexpression of SpoY resulted in a significant decrease in Spo0A protein levels (~2.5-fold, Figure 5A), confirming that SpoY inhibits *spo0A* translation. In contrast, deleting SpoX slightly decreased Spo0A levels, while over-expression of SpoX restored the Spo0A signal to WT levels.

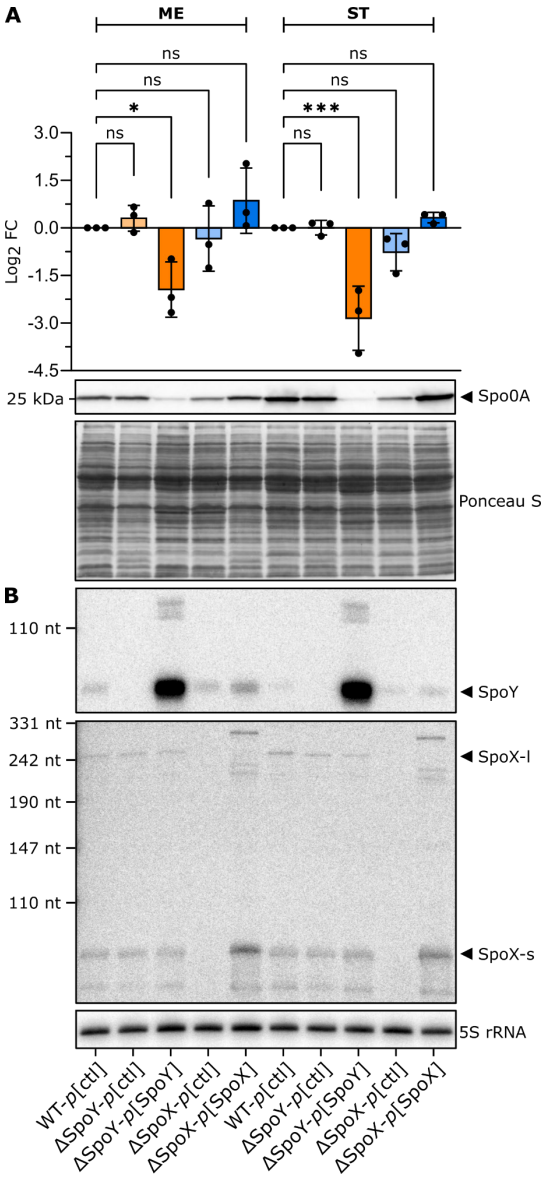


Figure 5: sRNA-mediated regulation of *spo0A* affects Spo0A levels. (A) Western blot analysis comparing Spo0A protein levels in a WT strain (*p*[ctl]), sRNA knock-out mutants (ΔSpoY/ΔSpoX-*p*[ctl]) and strains constitutively expressing the respective sRNA (ΔSpoY/ΔSpoX-*p*[ΔSpoY/ΔSpoX]). Equal optical density (OD) units of total cell lysates (error bars represent the mean ± SD of n=3 biological replicates) were loaded from strains gown till mid-exponential (ME) or stationary (ST) phase of growth in TY. Band intensities were measured and Log_2 fold changes were calculated relative to the respective WT. Western blot membranes were incubated with anti-Spo0A antibody. To calculate statistical significance, an ordinary one-way ANOVA with Dunnett's multiple comparison test was applied. Not significant (ns) $P > 0.05$; (∗) $P \leq 0.05$; (∗∗∗) $P \leq 0.001$. For western blot analysis Ponceau S staining of the blotting membrane served as loading control and for northern blot analysis 5S rRNA. **(B)** Northern blot analysis validating sRNA expression in the respective growth conditions.

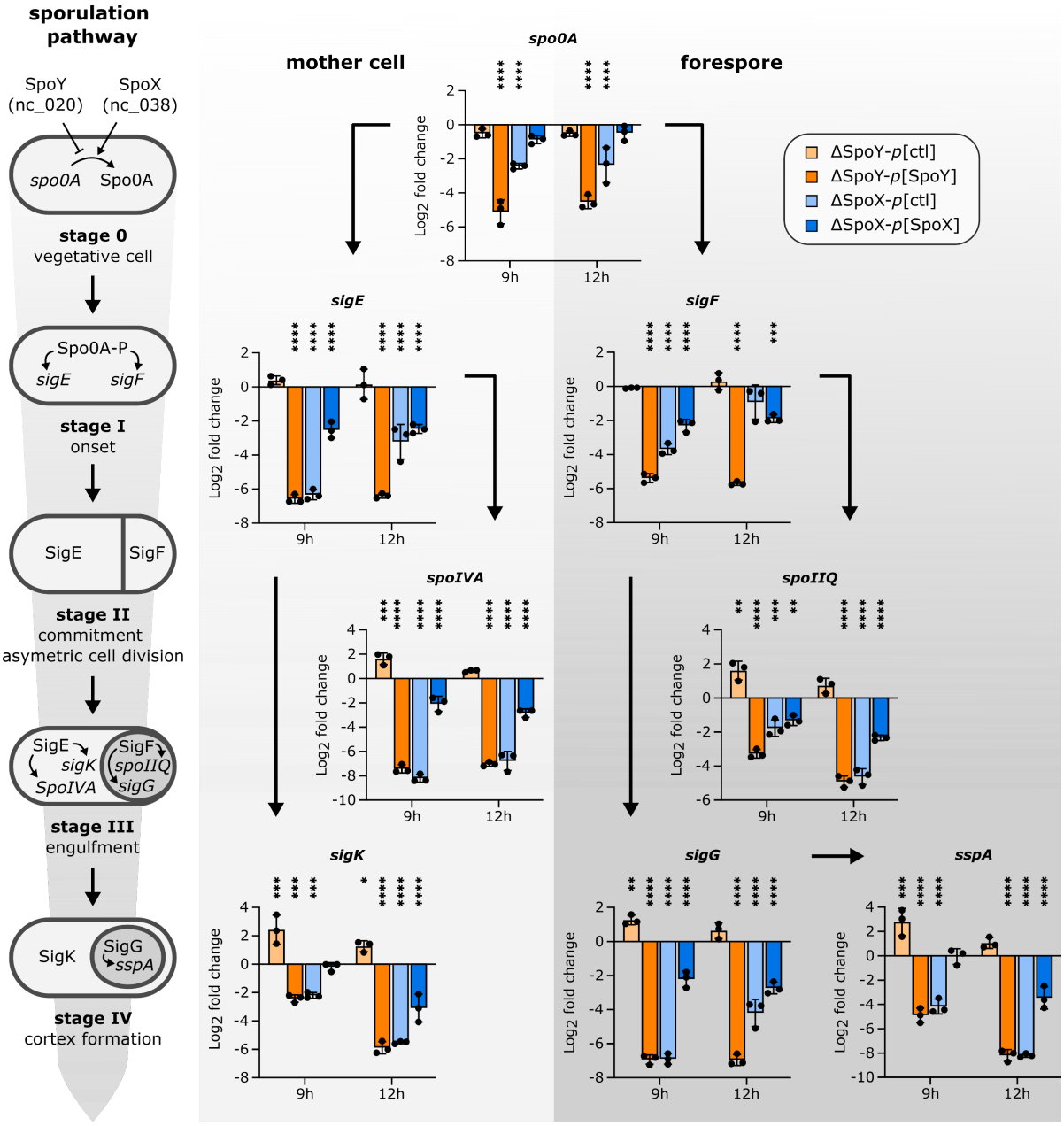


Figure 6: sRNA-mediated post-transcriptional regulation of *spo0A* affects sporulation-specific gene expression. Transcript levels of genes encoding sporulation-specific sigma factors and their respective regulon in sRNA knock-out mutants (ΔSpoY/ΔSpoX-*p*[ctl]) and strains constitutively expressing the respective sRNA (ΔSpoY/ΔSpoX-*p*[ΔSpoY/ΔSpoX]) relative to *C. difficile* 630 WT (*p*[ctl]). A schematic representation of the first four stages of sporulation in *C. difficile* is given on the left. RNA was extracted from samples (n = 3 biological replicates) taken at 9 h and 12 h post induction of sporulation on 70:30 sporulation plates. Log_2 fold changes were calculated relative to the WT. Statistical significance was determined using two-way ANOVA with Dunnett's multiple comparison test. $P > 0.05$; (∗) $P \leq 0.05$; (∗∗) $P \leq 0.01$; (∗∗∗) $P \leq 0.001$; (∗∗∗∗) $P \leq 0.0001$. Error bars represent the mean ± SD. Validation of SpoX and SpoY expression for the respective strains and time points by northern blot analysis is shown in Appendix Figure S6A.

Corresponding sRNA expression was confirmed *via* NB analysis (Figure 5B). These results corroborated our model, in which SpoX positively impacts Spo0A translation by base-pairing to the *spo0A* 5'UTR. Although SpoX and SpoY regulate *spo0A* translation, changes in Spo0A levels

might not directly translate into changes in Spo0A activity, as Spo0A requires additional activation *via* phosphorylation (Spo0A-P) (8). To evaluate if SpoX and SpoY-mediated changes in Spo0A levels correlated with Spo0A-P activity, we analyzed transcript levels of several sporulation-specific genes that operate downstream of Spo0A-P (Figure 6). Specifically, we measured transcript levels of *sigE*, *sigF*, *spoIV* (σ^E regulon), *spoIIQ* (σ^F regulon), *sigK*, *sigG*, and *sspA* (σ^G regulon) as well as *spo0A via* qRT-PCR (44,48). In accordance with our previous results, SpoY overexpression had a negative effect on transcript levels of all tested genes, whereas SpoY deletion had either no effect (*sigE*, *sigF*) or resulted in an increase of transcript abundance. In contrast, deletion of SpoX reduced transcript levels of all tested genes, while SpoX overexpression partially restored transcript abundance to WT levels. Taken together, the observed changes in expression of sporulation-specific genes suggested that modulation of Spo0A levels by SpoY leads to an

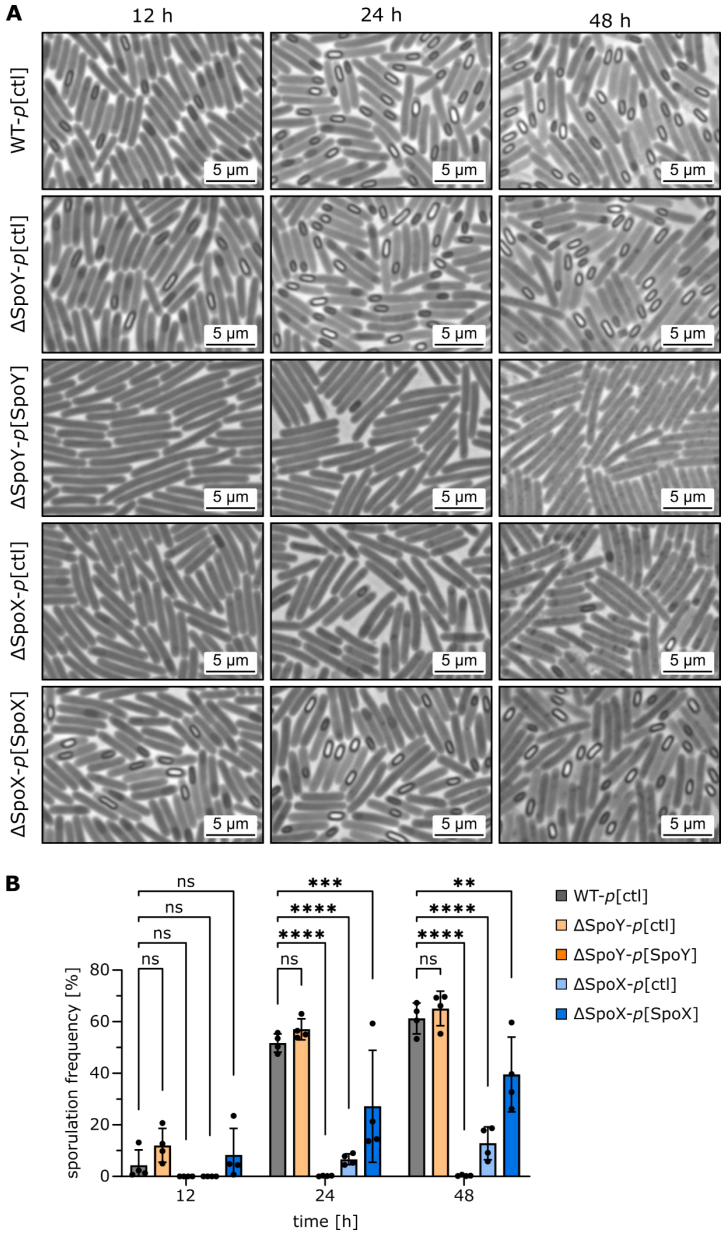


Figure 7: sRNA-mediated regulation of spo0A affects sporulation frequencies. (A) Representative phase-contrast images (n=4) of a WT strain (*p*[ctl]), sRNA knock-out mutants (ΔSpoY/ΔSpoX-*p*[ctl]) and strains constitutively expressing the respective sRNA (ΔSpoY/ΔSpoX-*p*[ΔSpoY/ΔSpoX]) at 6 h, 12 h, 24 h and 48 h post inoculation of 70:30 liquid sporulation medium. **(B)** Sporulation frequencies calculated based on phase-contrast microscopy. Error bars represent the mean ± SD of n=4 biological replicates (A). 2-way ANOVA with Dunnett's multiple comparison test was used to calculate statistical significance. Not significant (ns) $P > 0.05$; (**) $P \leq 0.01$; (***) $P \leq 0.001$; (****) $P \leq 0.0001$.

overall downregulation of the sporulation cascade, while activity of SpoX results in upregulation of spore formation. Interestingly, SpoY overexpression and SpoX deletion also resulted in a decrease in *spo0A* transcript levels, potentially by affecting ribosome occupancy of the *spo0A* mRNA (Figure 6). By altering ribosome occupancy of mRNA targets, sRNA-mediated regulation can indirectly affect RNase-mediated cleavage and thus mRNA stability (49,50).

To further assess the impact of SpoY and SpoX on sporulation, we determined sporulation frequencies of WT and mutant strains by phase contrast microscopy. As shown in Figure 7A&B, SpoY over-expression and SpoX deletion resulted in significantly reduced sporulation frequencies during late time points (12 h & 24 h). These observations were confirmed by CFU-based calculation of sporulation frequencies (Appendix Figure S6B). Hence, our data show that SpoY and SpoX not only affect *spo0A* translation in an inverse manner, but consequently influence gene expression of sporulation-specific genes and ultimately sporulation frequencies.

2.4.4.7 SpoY and SpoX impact *C. difficile* gut colonization in a mouse model of *C. difficile* infection

Considering the marked impact of SpoY and SpoX deletion on sporulation in *C. difficile*, as well as their extended interactome, we decided to monitor the effect of SpoY and SpoX deletion in a mouse model of *C. difficile* infection (Figure 8A). Overall, a delayed onset of disease in mice challenged with the sRNA deletion strains was observed, as ΔSpoY and ΔSpoX infected mice showed a delayed body weight loss compared to mice infected with *C. difficile* WT (Figure 8B). Nevertheless, colon shortening on day 7 was equally severe in mice challenged with *C. difficile* WT, ΔSpoY, or ΔSpoX suggesting similar levels of toxin production and consequently disease severity over the course of infection (Figure 8C). Initial colonization was comparable in ΔSpoY, ΔSpoX and WT treated mice, as no difference in vegetative cells or spores was observed on day 1 post infection (Figure 8D). However, lower CFUs of spores and vegetative cells were recovered from feces of ΔSpoY and ΔSpoX infected mice at days 3, 5, and 7 post infection, compared to mice challenged with *C. difficile* WT. Accordingly, mice infected with ΔSpoY or ΔSpoX strains exhibited an accelerated *C. difficile* clearance rate (Figure 8E). This was particularly evident in ΔSpoY treated mice starting from day 3, while the bacterial burden in mice infected with ΔSpoX only decreased on day 5 (vegetative cells) and day 7 (vegetative cells and spores) compared to WT treated mice. Generally, the lower spore counts in ΔSpoY or ΔSpoX infected mice were paralleled by lower vegetative cell counts, revealing a global effect of SpoY and SpoX deletion on *C. difficile* gut colonization rather than on sporulation alone (Figure 8D). In support of this hypothesis, our RIL-seq dataset suggested base-pairing of SpoY with the *cwp2* mRNA as well as base-pairing of SpoX with the *cwpV* mRNA (Dataset EV2), each encoding cell wall proteins with predicted virulence functions (51,52). Using our translational reporter system, we were able to validate these predicted RIL-seq targets.

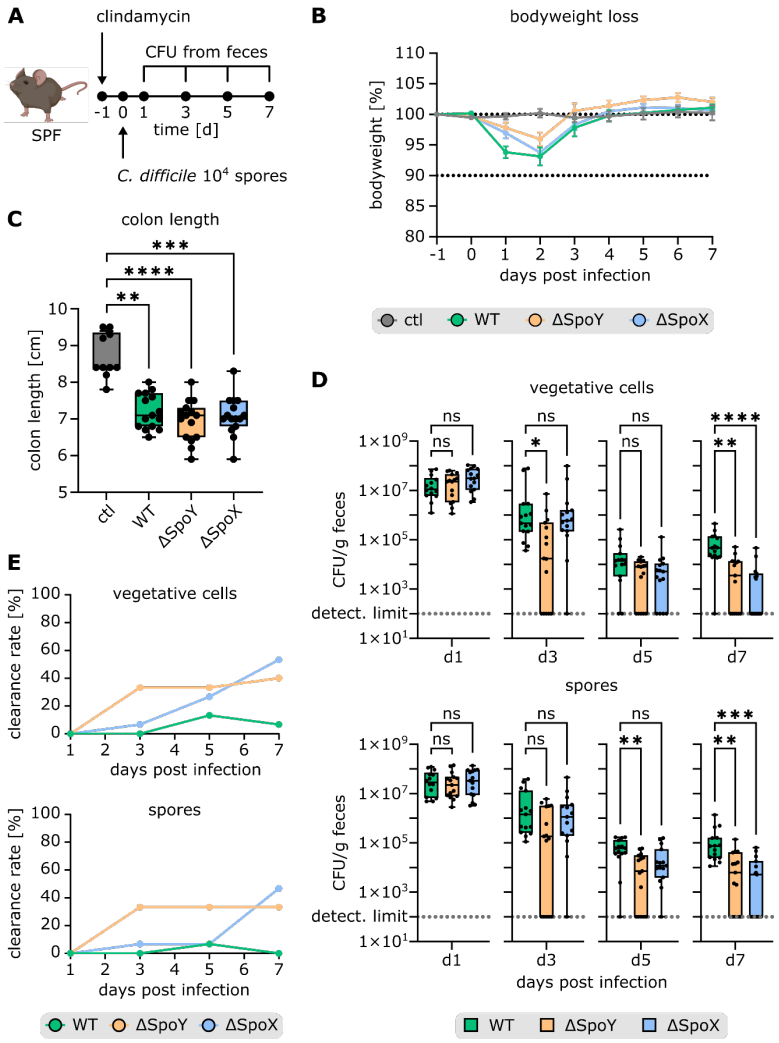


Figure 8: SpoY and SpoX deletion affects *C. difficile* gut colonization and spore burden in a mouse model of *C. difficile* infection. (A) Schematic representation of the mouse model of *C. difficile* infection (partially created with BioRender.com). SPF mice were treated with clindamycin 24 h prior to infection, administered *via* intraperitoneal injection to induce susceptibility to *C. difficile* infection (85). Following antibiotic treatment, groups of mice were infected with 10^4 spores of *C. difficile* 630 WT (n=15), ΔSpoY (n=15), or ΔSpoX (n=15) *via* oral gavage. Mice were monitored for disease between days 0 and 7 post infection. Faecal samples were collected at indicated time points to determine pathogen burden. **B, C (B)** Body weight loss over the course of 7 days (data points and error bars represent mean ± standard error of the mean; the dotted lines serve as a visual guideline) as well as **(C)** final colon length at day 7 post infection of uninfected mice (n=10) and mice infected with *C. difficile* 630 WT (n=15), ΔSpoY (n=15), or ΔSpoX (n=15) are indicated (box-and-whisker plots show minimum, first quartile, median, third quartile, and maximum). **(D)** Comparison of CFUs of *C. difficile* vegetative cells and spores in fecal pellets of mice at different time points during infection with *C. difficile* 630 WT (n=15), ΔSpoY (n=15), or ΔSpoX (n=15). Replicates with CFUs below the detection limit were set to 100. Box-and-whisker plots show minimum, first quartile, median, third quartile, and maximum. **(E)** Clearance rate (number of mice below detection limit divided by the total number of mice) of SPF mice infected with *C. difficile* 630 WT (n=15), ΔSpoY (n=15), or ΔSpoX (n=15). Data information: Kruskal-Wallis was performed with Dunnett's multiple comparison test to calculate statistical significance in (C)&(D). Not significant (ns) $P > 0.05$; (*) $P \leq 0.05$; (**) $P \leq 0.01$; (***) $P \leq 0.001$; (****) $P \leq 0.0001$.

Quantification of fluorescence signals of the respective mCherry fusion constructs revealed that SpoY inhibits *cwp2* translation, while SpoX acts as a positive regulator of *cwpV* translation (Figure EV4A&B). Taken together, the impact of SpoY and SpoX on intestinal pathogenesis suggested that their regulatory functions extend beyond regulating Spo0A protein levels and likely include additional regulatory targets that contribute to intestinal colonization.

2.4.5 Discussion

A plethora of research in human pathogens such as *Pseudomonas aeruginosa*, *S. enterica*, and *Vibrio cholerae* has highlighted the importance of sRNAs in regulating virulence pathways (53). To uncover sRNA-mediated regulatory mechanisms that might shape *C. difficile* virulence, we performed RIL-seq during the onset of sporulation in *C. difficile* (28). Endospore formation has been extensively studied, particularly in the model organism *B. subtilis and* is a tightly regulated process, defined by several sequential morphological stages (Figure 6) (54). *A*lthough the sporulation cascade is generally conserved between *C. difficile* and other endospore-forming Firmicutes, there are some striking differences, most notably regarding sporulation initiation (13). In *B. subtilis*, environmental signals that induce sporulation are channeled through a complex phosphorelay system consisting of several sensor kinases and phosphotransferases, culminating in the activation of Spo0A (55). Phosphorylated Spo0A-P then initiates the sporulation process by activating the transcription of several key sporulation-specific genes (55). Unlike *B. subtilis*, *C. difficile* does not encode an apparent intermediate phosphorelay system (8). Although three putative sensor histidine kinases (PtpA-C) have been described to directly influence Spo0A phosphorylation, the overall process of Spo0A activation remains barely understood and points towards additional unknown mechanisms regulating Spo0A activity in *C. difficile* (8,56,57).

In this study, we identified sRNA-mediated post-transcriptional regulation of *spo0A* translation as a new mechanism contributing to sporulation initiation in *C. difficile* (Figure 9A-C). There are a few examples of sRNA-mediated regulation of sporulation in endospore-forming Firmicutes. In *B. subtilis* sRNA SR1 inhibits translation of the histidine kinases *kinA* that transmits environmental signals, eventually resulting in phosphorylation of Spo0A (58). Another example is the *virX* sRNA in *C. perfringens that* negatively regulates sporulation by repressing transcription of the early forespore-specific σ factor σ^F (59). Furthermore, sRNA RCd1 in *C. difficile* inhibits production of the late mother cell-specific σ factor σ^K by preventing the excision of the prophage-like element that interrupts the *sigK* gene (17). Here, we characterized two novel sRNAs, SpoY and SpoX, that function by directly binding and regulating the *spo0A* mRNA (Figure4, Figure 3D). We could show that SpoY operates by base-pairing with the *spo0A* RBS region, thereby inhibiting *spo0A* translation, resulting in reduced sporulation frequencies when overexpressed (Figure 7, Figure 9A). In contrast, SpoX interaction with the *spo0A* 5'UTR results in an upregulation of *spo0A*

translation and consequently sporulation, most likely through a change of the *spo0A* 5'UTR secondary structure upon base-pairing with SpoX (Figure 9B). Analogous mechanisms have been described in the literature; a well-known example is the positive regulation of *rpoS,* the key regulator of general stress responses in *E. coli* (60). In this case, several sRNAs (DsrA, RprA, and ArcZ) base-pair with the *rpoS* 5' leader to expose the *rpoS* translational start site that is otherwise blocked by an inhibitory stem-loop structure (60).

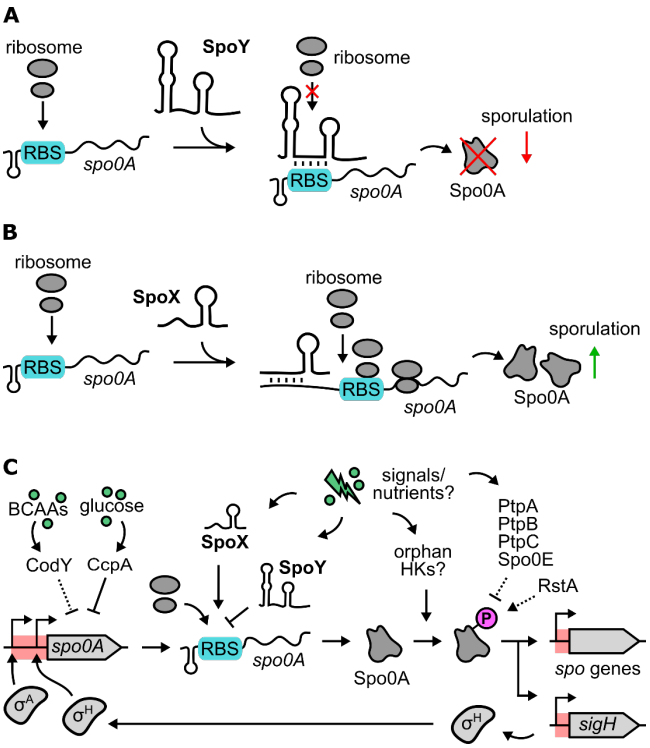


Figure 9: SpoY and SpoX add a post-transcriptional layer to the complex regulatory network shaping sporulation initiation in *C. difficile*. (A&B) Model of the SpoY and SpoX-mediated regulation of *spo0A* translation. SpoY base-pairing with the *spo0A* RBS (highlighted in blue) inhibits ribosome binding, ultimately resulting in translational repression of *spo0A* and decreased sporulation. In contrast, SpoX base-pairing with the *spo0A* 5'UTR most likely results in a conformational change of the *spo0A* 5'UTR that renders the RBS more accessible, ultimately leading to increased translation and sporulation. **(C)** Current model of transcriptional, post-transcriptional and post-translational regulators affecting initiation of sporulation in *C. difficile* (12). Arrows indicate positive regulation and blunt ends negative regulation. Dotted lines indicate an indirect regulation or regulation where the mechanism of action is still unknown, whereas solid lines indicate verified direct interactions. Red shades correspond to untranslated regions. Phosphorylated Spo0A is highlighted by a pink "P". Branched-chain amino acids are abbreviated as "BCAAs" and histidine kinases as "HKs".

Of note, chimeras formed by *spo0A* and SpoX or SpoY are not the only *spo0A* interactions present in the dataset, albeit the most enriched. In fact, we found nine additional chimeras consisting of the *spo0A* mRNA ligated to an sRNA (n=8) or another mRNA (n=1, Figure EV3). Of these nine interactions, eight displayed a chimeric read profile comparable to SpoY, suggesting a similar mode of action. Accordingly, these sRNAs might substitute for the SpoY function, potentially explaining the minor sporulation phenotype of a SpoY deletion strain. Interestingly, *in silico* predicted interaction sites overlap with the RIL-seq peak profile for five of the detected *spo0A*-sRNA interactions, further corroborating the RIL-seq results (Figure EV3) (47). There are known examples of mRNAs that are directly targeted by multiple sRNAs, most of which encode key regulators. For instance, *flhDC*, the master regulator of flagellar genes in *E. coli* interacts with five sRNAs (ArcZ, OmrA, OmrB, OxyS, and McaS) that base-pair with the *flhDC* 5'UTR resulting in either

negative or positive regulation of motility (61). Similarly, biofilm formation is a central target of sRNA-mediated regulation through base-pairing of seven sRNAs (OmrA/B, McaS, RprA, RydC, GcvB, and RybB) with the *csgD* 5'UTR, encoding a central regulator of curli formation in *E. coli* (62). sRNA transcription in turn is regulated in response to specific environmental or nutritional signals, allowing targeted mRNA regulation in relevant conditions. For example OmrA/B are regulated by the EnvZ/OmpR two-component system in response to high osmolarity while McaS is activated by cAMP-CRP in response to carbon limitation (63,64).

In line with these examples, *C. difficile* Spo0A might constitute another key regulator that is a central target of sRNA-mediated regulation. Accordingly, the various sRNAs interacting with *spo0A* may serve to integrate different environmental signals to modulate Spo0A expression, thus partially replacing the missing phosphorelay system that translates environmental cues into Spo0A activity in *B. subtilis* (8,55). This hypothesis is further corroborated by the distinct expression profiles of SpoY and SpoX, which points to *spo0A* regulation during different growth phases (Figure 3C). SpoY most likely suppresses *spo0A* translation during early growth stages, in conditions that favor active growth rather than sporulation. In contrast, SpoX expression indicates that it exerts its positive effect on *spo0A* translation mostly in late exponential and early stationary phase when the sporulation process initiates. Given the asynchronous nature of sporulation initiation in *C. difficile*, identifying the specific environmental cues and transcription factors that regulate expression of both sRNAs and thus *spo0A* will likely necessitate single-cell-based approaches to fully understand the nuanced post-transcriptional regulation of *spo0A*. Accordingly, methods such as scRNA-seq could be applied to dissect the various physiological cellular states that co-exist within a population and often mask distinct regulatory events in individual cells that undergo sporulation.

Interestingly, some of the sRNAs interacting with *spo0A* also formed chimeras with SpoX, including SpoY, nc105 and nc176 (Figure EV1A). It is possible that SpoX not only upregulates *spo0A* translation, but also sponges sRNAs that would otherwise inhibit *spo0A* translation. This could explain why the deletion of SpoX has such a pronounced effect on sporulation frequencies (Figure 7). Alternatively, it is equally likely that SpoY, nc105 and nc176 sRNA regulate SpoX activity, preventing positive regulation of sporulation in conditions favoring active proliferation. sRNA sponges generally act by sequestering a target sRNA, thereby preventing the sRNA-target interaction, an effect that depends on the stoichiometry between the sponge, sRNA, and mRNA (65). Reports of sRNAs simultaneously acting as sRNA sponge and mRNA regulator have been published previously, supporting this hypothesis (65). For example ArcZ and CyaR, both known regulators of the *rpoS* mRNA, also interact with each other, as ArcZ overexpression reduces CyaR steady state levels and upregulates CyaR targets (66). Further research is necessary to verify these

SpoX-sRNA interactions and decipher if, when, and how these interactions impact Spo0A activity (65).

As discussed above, our study indicates a complex network of post-transcriptional regulatory interactions that shape *C. difficile* virulence. Not surprisingly, our mouse experiments have revealed a global impact of SpoX and SpoY on *C. difficile* gut colonization. Interestingly, this impact appears to be in contrast to the opposing effects on sporulation, which we observed under isolated *in vitro* conditions that lack complex gut interactions. Of note, Spo0A has been implicated in pathways other than sporulation, as Spo0A inactivation also results in altered toxin production and biofilm formation in *C. difficile* (67–69). Hence, sRNA-mediated regulation of *spo0A* might impact additional processes besides sporulation that could not be covered in this work and will require further investigation (67,68,70). Moreover, it is important to consider that both sRNAs interact with additional targets (Figure EV1A). For example, SpoY also inhibits translation of *cwp2*, a cell wall protein known to affect cellular adherence *in vitro* (52). Furthermore, SpoX positively regulates translation of *cwpV*, which encodes a cell wall protein described to promote cell aggregation (Figure EV4) (51). Consequently, deleting or overexpressing either sRNA *in vivo* most likely affects processes beyond sporulation and might explain the global effect of both sRNA deletion strains in the mouse model of *C. difficile* infection (Figure 8).

In summary, the application of Hfq RIL-seq to *C. difficile* has revealed a global view of extensive Hfq-mediated RNA interactions (Figure 1B, Dataset EV2) (30). Although we have barely scratched the surface of sRNA-mediated regulation in *C. difficile*, our RIL-seq data represents a starting point for the characterization of additional processes modulated by sRNAs. In this work, we uncovered a new layer of post-transcriptional regulation in *C. difficile* hinting at a complex sRNA network regulating sporulation in this important human pathogen. Given the low conservation of mechanisms governing sporulation initiation these results might open an interesting avenue for potential therapeutic targets to counteract CDI. In fact, the use of antisense nucleic acids to selectively target species in a microbial community has gained attention as a promising alternative to conventional antibiotics (71,72). Accordingly, mimicking or blocking the activity of sRNAs using antisense nucleic acid derivatives might represent an interesting alternative, especially considering the contribution of antibiotics to CDI recurrence (7,53).

2.4.6 Materials and Methods

2.4.6.1 Bacterial strains and growth conditions

A complete list of all *C. difficil*e and *E. coli* strains used in this study is provided in Appendix table S1. *C. difficile* 630 cultures were routinely grown anaerobically inside a Coy chamber (85% N_2, 10% H_2 and 5% CO_2) in Brain Heart Infusion (BHI) broth or on BHI agar plates (1.5% agar) unless stated otherwise. If necessary, antibiotics were added to the medium at the following

concentrations: thiamphenicol (TAP) 15 µg/ml, cycloserine (CS) 250 µg/ml, cefoxitin (FOX) 8 µg/ml. *E. coli* cultures were propagated aerobically in Luria-Bertani (LB) broth (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl) or on LB agar plates (1.5% agar) supplemented with chloramphenicol (CHL, 20 µg/ml). *E. coli* strain Top10 (Invitrogen) served as a recipient for all cloning procedures, and *E. coli* CA434 (HB101 carrying the IncPβ conjugative plasmid R702) was used as donor strain for plasmid conjugations into *C. difficile* 630.

2.4.6.2 Plasmid construction

All plasmids and DNA oligonucleotides used in this study are listed in Appendix table S2&3, respectively. *E. coli* TOP10 was used for plasmid propagation according to standard procedures (73).

pFF-53 - plasmid for generating a *hfq*::3×FLAG strain

For insertion of a C-terminal *hfq::*3×FLAG-tag, allelic exchange cassettes were designed with approximately 1.2 kb of homology to the chromosomal sequence flanking the up- and downstream regions of the *hfq* stop codon. Both homology regions were PCR-amplified from *C. difficile* 630 using high fidelity Fusion Polymerase (Mobidiag) with 5% DMSO, FFO-364/-365 and FFO-368/-369. The resulting fragments were gel purified with NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel). The 3×FLAG-tag (DYKDHDGDYKDHDIDYKDDDDK) was similarly amplified and purified with FFO-366/-367, using the previously published pFF-12 as a template (18). Insert assembly and ligation into PCR-linearized pJAK184 (FFO-362/-363) was achieved *via* Gibson Assembly (Gibson Assembly ® Master Mix, New England BioLabs) according to the manufacturer's instructions, resulting in pFF-53 (18).

pFF-162 / -163 / -164 / -245 / -166 / -248 / -247 / -167 – plasmids for *in vitro* transcriptions

SpoY (FFO-958/-959), SpoY* (FFO-958/-960), SpoX (short isoform, FFO-961/-962), SpoX* (short isoform, FFO-1261/-1262), *spo0A* 5'UTR starting from pTSS and first 69 nt of coding region (FFO-964/-966) and *spo0A* 5'UTR starting from sTSS and first 69 nt of coding region (FFO-965/-966) were PCR-amplified from *C. difficile* 630 using Fusion Polymerase (Mobidiag), adding a 5' overhang comprising the T7-promoter sequence (5'-GTTTTTTTTTAATACGACTCACTATAGGG). For inserting SpoY* compensatory mutations in *spo0A*, *spo0A*C* was amplified in two parts using FFO-964/-1268 and FFO-1269/-966, before joining both fragments *via* SOEing PCR with FFO-964/-966. SpoX* compensatory mutations were inserted similarly using FFO-964/-1259 and FFO-1267/-966, followed by SOEing PCR with FFO-964/-966. In each case, PCR products were gel purified as described above. Subsequently, 3'-adenine overhangs were added to all PCR product using Taq Polymerase (Biozym). The resulting fragments were cloned into the StrataClone TA-cloning vector and transformed into StrataClone SoloPack competent cells according to the manufacturer's protocol (StrataClone PCR Cloning Kit, Agilent), resulting in pFF-162 (SpoY), pFF-

163 (SpoY*), pFF-164 (SpoX), pFF-245 (SpoX*), pFF-166 (*spo0A* with 5'UTR starting from pTSS), pFF-248 (*spo0A*^{C}* SpoY*), pFF-247 (*spo0A*^{C}* SpoX*) and pFF-167 (*spo0A* with 5'UTR starting from sTSS).

pFF-170 / -171 – plasmids for generating SpoY and SpoX deletion mutants

For deletion of SpoY and SpoX, allelic exchange cassettes were designed with approximately 1.2 kb of homology to the chromosomal sequence flanking the deletion sites of SpoY and SpoX. To avoid polar effects on genes or sRNAs encoded adjacent to SpoY or antisense to SpoX (Figure 3B), the deleted region was restricted to nucleotide 11-46 in case of SpoY, and nucleotide 1-83 of SpoX in addition to 40 nt upstream of SpoX encompassing the SpoX promoter region. Homology arms were PCR amplified, and gel purified as described above, using FFO-977/-978 and FFO-979/-980 for the SpoY homology arms, and FFO-985/-986 and FFO-987/-988 for amplification of the SpoX homology regions, respectively. The homology arms were joined *via* SOEing PCR resulting in one large fragment encompassing both homology regions, and a BamHI/SacI restriction site at the 5'/3'end for both, SpoY (FFO-977/-980) and SpoX (FFO-985/-988). Following restriction digest using BamHI and SacI, the fragments were mixed in a 3:1 ratio with an equally digested and gel purified pJAK112 and ligated overnight at 4 °C using T4 DNA ligase (Thermo Scientific), resulting in pFF-170 (SpoY deletion) and pFF-171 (SpoX deletion) (18).

pFF-185 / -186 / -191 / -254 / -285 / -187 / -192 / -260 / -289 / -207 / -344 / -345 / -346 / -347 - translational fusion reporter

To discern the impact of sRNA-target interactions on target translation, we designed a translational fusion system based on the previously published pDSW1728 mCherryOpt plasmid, that was initially designed to study gene expression (74). The full reporter system (RS) constitutively co-expresses an sRNA controlled by an *fdxA* promoter and the target 5'UTR fused to mCherryOpt, controlled by a *cwp2* promoter. Regulation of target translation *via* sRNAs is measured by comparing fluorescence of the strain expressing the full RS (*p*[sRNA-*target*]) to a control strain only expressing the sRNA (*p*[sRNA]) or the target 5'UTR fused to mCherryOpt (*p*[*target*]). All plasmids were designed to allow easy exchange of individual components *via* restriction digestion and are illustrated in Appendix Figure S5A. sRNA expression was verified *via* northern blot analysis (Appendix Figure S5B&C). For generating a *p*[*spo0A*] plasmid, the *cwp2* promoter and the *spo0A* 5'UTR (starting from the pTSS and including 60 nt of CDS) were PCR amplified and gel purified as described above from *C. difficile* 630. FFO-1004/-1000 and FFO-1001/-1002 were used, respectively, thereby adding NheI/XhoI restriction at the 5'/3'end of the *cwp2* promoter, and a XhoI/SacI restriction site at the 5'/3'end of the *spo0A* 5'UTR, respectively. mCherryOpt was PCR amplified from pDSW1728 with FFO-1056/-1057 starting with the second codon of the mCherryOpt CDS, adding a SacI restriction site directly upstream and preserving the

BamHI restriction site at the 3'end. All components were subjected to restriction digest using the appropriate restriction enzymes and ligated into the NheI/BamHI digested pDSW1728 vector as described above, resulting in pFF-185 (*p*[*spo0A*]). To generate a *p*[sRNA] plasmid, the *fdxA* promoter and SpoY were PCR amplified from *C. difficile* 630 with FFO-995/-1005 and FFO-1006/-1007, respectively, inserting an NheI restriction site at the *fdxA* 5'end and a SpoY overlapping region at the 3'end. Inserting a restriction site upstream of SpoY was avoided to preserve the sRNA primary and secondary structure. Accordingly, an *fdxA* overlapping region was added to the SpoY 5'end and an XbaI restriction site at the 3'end followed by the *slpA* terminator and a BamHI restriction site to prevent readthrough. Both fragments were joined *via* SOEing PCR with FFO-995/-1007, NheI/BamHI digested and ligated into the NheI/BamHI digested pDSW1728 vector, resulting in pFF-186 (*p*[SpoY]). Finally, the *p*[sRNA-*target*] plasmid was generated by PCR amplifying the *cwp2-spo0A* 5'UTR-mCherryOpt construct from pFF-185, using FFO-999 and FFO-1057, thereby exchanging the 5' NheI with an XbaI restriction site followed by the *slpA* terminator. The resulting fragment was digested with XbaI/BamHI and ligated into the equally digested pFF-186 yielding pFF-191 (*p*[SpoY-*spo0A*]). All remaining plasmids were generated by exchanging either sRNA or target of the plasmids described above. For SpoX constructs, *fdxA* and SpoX were PCR amplified using FFO-995/-1008 and FFO-1009/-1010. Both fragments were joined *via* SOEing PCR with FFO-995/-1010, digested with NheI/XbaI and ligated into NheI/XbaI digested pFF-186 and pFF-191, resulting in pFF-187 (*p*[SpoX]) and pFF-192 (*p*[SpoX-*spo0A*]). For SpoY* constructs, SpoY* was PCR amplified from pFF-163 with FFO-1006/-1007 and joined *via* SOEing PCR with the previously amplified *fdxA* promoter, using FFO-995/-1007. The SOEing product was NheI/XbaI digested and ligated into the equally digested pFF-191, yielding pFF-254 (*p*[SpoY*-*spo0A*]). Next *spo0A* harboring compensatory mutations was PCR amplified from pFF-248, using FFO-1001/-1002, digested with XhoI/SacI, and ligated into XhoI/SacI digested pFF-254 resulting in pFF-285 (*p*[SpoY*-*spo0A*C*]). Generation of SpoX* constructs was achieved by first amplifying SpoX* in two fragments, inserting the seed region mutations with FFO-1264/-1262 and FFO-1263/-1010. Both fragments were then joined *via* SOEing PCR with FFO-1264/-1010, followed by a second SOEing PCR, combining the previously amplified *fdxA* promoter and the full length SpoX*, using FFO-995/-1010. The resulting product was NheI/XbaI digested and ligated into a similarly digested pFF-191, yielding pFF-260 (*p*[SpoX*-*spo0A*]). Finally, *spo0A* harboring compensatory mutations was PCR amplified from pFF-247, using FFO-1001/-1002, digested with XhoI/SacI, and ligated into XhoI/SacI digested pFF-260, resulting in pFF-289 (*p*[SpoX*-*spo0A*C*]). In addition to the reporter system constructs, an empty control vector was generated by linearizing pFF-191 *via* PCR with FFO-994/-1205, adding an additional NheI restriction site at the 5'end. The resulting product was NheI digested and re-ligated yielding pFF-207 (*p*[ctl]). For generating a RS constitutively expressing either *cwp2* alone or co-expressing SpoY and *cwp2, cwp2* (5'UTR and 75 nt of CDS) was PCR amplified and gel purified as described above from *C. difficile* 630. FFO-

1354/-1355 was used, adding a XhoI/SacI restriction site at the 5'/3'end of the *cwp2* 5'UTR and start of CDS. The resulting product was XhoI/SacI digested and ligated into a similarly digested pFF-191 and pFF-185, yielding pFF-344 (*p*[SpoY-*cwp2*]) and pFF-345 (*p*[*cwp2*]). Finally, a RS constitutively expressing either *cwpV* alone or co-expressing SpoX (long isoform) and *cwpV, cwpV* (5'UTR and 75 nt of CDS) was PCR amplified and gel purified from *C. difficile* 630 using FFO-1469/-1470, adding a XhoI/SacI restriction site at the 5'/3'end of the *cwpV* 5'UTR and start of CDS. The resulting product was XhoI/SacI digested and ligated into a similarly digested pFF-192 and pFF-185, yielding pFF-346 (*p*[SpoX-*cwpV*]) and pFF-347 (*p*[*cwpV*]).

2.4.6.3 Plasmid conjugation

For conjugation purposes, plasmids were transformed using 80 µl of electro competent *E. coli* CA434 mixed with 100 – 500 ng of plasmids in a pre-chilled electroporation cuvette. Following electroporation with 1.8 kV, 200 Ω, 4 - 5 sec, cells were recovered for 4 h at 37 °C in 1 ml LB. Colonies harbouring the plasmid were selected on LB supplemented with CHL and confirmed *via* colony PCR. Conjugation was performed according to Kirk and Fagan (2016), as published previously (18,75).

2.4.6.4 Generation of deletion and insertion strains

Gene deletions were constructed using homologous recombination as previously published (18,76). In short, plasmids were conjugated in *C. difficile* 630 strains as described above. Following conjugation, colonies were screened for the first recombination event *via* PCR. Positive recombinants were streaked on non-selective BHI, followed by incubation for 2 – 3 days. Growth was harvested using 900 µl 1xPBS, and 50 µl of a 10^{-4} and 10^{-5} dilution of the mixtures were streaked either on CDMM supplemented with 50 µg/ml fluorocytosine for pJAK112-derived plasmids, or on TY containing 4% w/v xylose for pJAK184-derived plasmids. Once colonies appeared, 8 – 15 were re-streaked to purity and tested for secondary recombination events *via* PCR. To test for plasmid loss, colonies were simultaneously streaked on selective plates containing TAP. Sanger sequencing was applied to confirm successful deletions and insertions.

2.4.6.5 RIL-seq (RNA interaction by ligation and sequencing)

RIL-seq experimental procedure

RIL-seq was performed, following the original protocol published by Melamed and collogues with minor alterations (24,25). Briefly, *C. difficile* 630 WT and Hfq-FLAG were grown in sterile filtered TY in four biological replicates until OD_{600} of ~1.2 (transition phase) (28,29). Of each replicate, 100 ODs were harvested (4,500 x g, 15 min at 4 °C), resuspended in 10 ml ice-cold 1xPBS and irradiated with UV light (254 nm, 80,000 mJ/cm2). Following centrifugation (4,500 x g, 15 min at 4 °C), pellets were resuspended in 800 µl of ice-cold wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 0.05% Tween, 1:200 diluted protease inhibitor cocktail set III (Calbiochem)) supplemented with

0.1 U/µL of recombinant RNase inhibitor (wash buffer-RIn, Takara). Mechanical cell lysis was achieved by mixing each sample with 0.1 mm glass beads and grinding in a Retsch mixer mill at a frequency of 30/s for 5 min. The grinding was repeated four times; after each step the adaptors were placed on ice for 2 min. The resulting supernatant was transferred to a new tube (17,000 x g, 2 min at 4 °C), 400 µl wash buffer-RIn was added to the glass beads and the grinding was repeated once more. For Hfq co-immunoprecipitations, the accumulated lysates were incubated with protein A/G magnetic beads (Thermo-Fisher) pre-coupled with anti-Flag antibody (M2 monoclonal antibody, Sigma-Aldrich) for 2 h at 4°C with rotation, followed by three washing steps with wash buffer-RIn. Next, samples were treated with 480 µl of wash buffer supplemented with RNase A/T1 (1:520, Thermo Fisher Scientific) for 5 min at 22 °C, trimming any exposed RNA ends. The process was stopped by washing each sample with 200 µl of wash buffer supplemented with SUPERase In RNase inhibitor (final concentration of 0.1 U/µL, Ambion) three times for 5 min at 4 °C. Following PNK treatment for 2 h at 22 °C (1x PNK buffer A, 1 mM ATP, 1 U/µl recombinant RNase inhibitor, 0.5 U/µl T4 PNK (New England BioLabs)), samples were washed again three times with wash buffer-RIn. Hfq-bound RNAs were then proximity ligated with T4 RNA ligase I (1x T4 RNA ligase buffer, 9% DMSO, 1 mM ATP, 20% PEG 8000, 0.6 U/µl recombinant RNase inhibitor, 2.7 U/µl T4 RNA ligase I (New England BioLabs)) overnight at 22°C with agitation, followed by 3 washing steps with wash buffer-RIn at 4 °C. Finally, RNA was eluted by incubating the beads with Proteinase K (50 mM Tris-HCl pH 7.8, 50 mM NaCl, 1% SDS, 5 mM EDTA pH 8.0, 5 mM β-mercaptoethanol, 0.1 U/µl recombinant RNase inhibitor, 0.33 mg/ml Proteinase K (Thermo Fisher Scientific)) for 2 h at 55 °C. RNA purification was achieved using TRIzol LS according to the manufacturer's instructions. Purified RNA was resuspended in 7 µL of nuclease-free water and quality controlled on a Bioanalyzer Pico RNA chip.

cDNA library preparation and sequencing were performed by Vertis Biotechnologie AG. First, oligonucleotide adapters were ligated to the RNA 5' and 3'ends followed by first-strand cDNA synthesis using M-MLV reverse transcriptase and the 3' adapters as primers. The resulting cDNAs were PCR-amplified using a high-fidelity DNA polymerase for 14-16 cycles. Next, cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and quality controlled by capillary electrophoresis on a Shimadzu MultiNA microchip electrophoresis system. For Illumina NextSeq sequencing, the samples were pooled in approximately equimolar amounts. The resulting cDNA pool was then size fractionated in the size range of 170 – 400 bp by polyacrylamide gel electrophoresis and paired-end sequenced on an Illumina HighSeq system using 2 x 150 bp read lengths.

RIL-seq data analysis

Sequencing and mapping results are listed in Dataset EV1. Data analysis was performed as described in Melamed *et al*., with a few modifications (24,25). Briefly, raw reads were trimmed

and quality filtered with BBDuk (min. phred-score of 20, min. read length of 25). First mapping was conducted with BWA provided by the RIL-seq computational pipeline (77). The map_chimeric_fragments.py script provided by the RIL-seq pipeline was used to classify fragments into single and chimeric. All parameters were set to default. Fragments that mapped within a distance of 1,000 nt or within the same transcript were considered single, whereas fragments that mapped to two different loci were considered chimeric. To test whether two fragments mapped within the same transcript, an additional annotation file for CP010905.2 was used. To decide if the replicates can be considered reproducible, their correlation was computed as described in Melamed *et al*, by comparing the numbers of mapped fragments in corresponding genomic windows between each pair of libraries, for single and chimeric fragments, respectively (25). To be able to use the ribozero option of RILseq_significant_regions.py, which excludes rRNAs from the analysis, the necessary BioCyc database was generated with Pathway Tools based on the CP010905.2 annotation (78). The high correlation coefficient of all *hfq*::3xFLAG pairs (r≥0,79) allowed us to unify the replicates into a single dataset (Appendix Figure S1A). Fisher's exact test was applied to assign an odds ratio and a p-value to each chimera. Chimeras with a min. odds ratio of 1 and a p-value < 0.05 were considered significant and termed S-chimeras. In addition, only S-chimeras covered by ≥ 25 chimeric reads were considered for further analysis. The final output table was merged with the CP010905.2 annotation manually curated by assigning each interaction partner to one of the following categories: sRNA, 5'UTR, riboswitch, coding sequence (CDS), 3'UTR, CRISPR, tRNA, IGR, or anti-sense (AS), resulting in Dataset EV2 & 3 (18). Pairs can appear more than once if corresponding chimeric reads span multiple regions, such as an mRNA 5'UTR and CDS. Only counting each pair once, the dataset consisted of 1,569 unique interactions (1,198,921 chimeric reads) in the Hfq-FLAG strain and 6 interactions (461 chimeric reads) in the WT, yielding similar numbers to published RIL-seq data (Dataset EV1) (24,31).

To analyze intra-operon RBS overlaps of interactions, an inhouse python script was used. First a database of all intra-operon RBSs was build, based on the operon table published in Fuchs *et al* (18). The table was filtered for operons with primary TSSs, excluding the first gene of an operon. Intra-operon RBS regions were defined as 25 nt upstream and 20 nt downstream of the respective start codon of the gene. RNA1 and RNA2 of all S-chimeras were searched for overlaps with intra-operon RBS regions. For this purpose, the coordinates listed in the S-chimera table were used whereas the coordinate of either RNA1 start of last read and RNA2 start of first read were extended 100 nt in the respective direction. Additionally, the final output of interactions overlapping with intra-operon RBS regions was curated manually (Dataset EV5).

RIL-seq data visualization

To count how many chimeric or single reads overlap with given features in the CP010905.2 annotation, the script count_chimeric_reads_per_gene.py of the RILseq computational pipeline

was used with slight modifications. We modified the script in a way that it would consider a list of feature categories instead of a single one. The following features were considered for the counting: CDS, 3'UTR, 5'UTR, ncRNA (which includes riboswitches, tmRNA, SRP_RNA, RNase_P_RNA, 6S RNA), sRNAs, rRNA, tRNA, antisense and intergenic regions. A fragment was counted as intergenic if it did not overlap with any of the other features. The minimal overlap between a fragment and a gene was set to 5. If any fragment overlapped with two features, both were counted.

To create images of specific interactions, bed files of chimeric fragments of single interactions were generated using the script generate_BED_file_of_endpoints.py of the RIL-seq computational pipeline. The bed files were visualized with IGV 2.12.3, and further processed with Inkscape 0.92.4 for the respective figure (Figure 2D). For coverage plots these bed files were first converted into coverage files with bedtools genomecov, followed by plot generation using the R package Gviz (*eg.* Figure 3D) (79,80).

For circos plot visualization of sRNA networks, data were obtained by using the script plot_circos_plot.py from the RIL-seq computational pipeline. Circos plots were generated with Circos (81). To avoid overloading the plots, only a fraction of the interactions were shown, as indicated in the respective figure description. For a more detailed visualization of all SpoY and SpoX target interactions Cytoscape 3.9.1 was used. All targets (nodes) were included, with targets supported by ≥25 chimeras marked by a solid line, while targets supported by <25 chimeras were highlighted with a dashed line. Target types were discriminated by color as indicated in the figure legend. Edge strengths correlate with the total number of chimeras supporting an individual interaction as listed in Dataset EV2.

2.4.6.6 Hot phenol extraction of total RNA

Total RNA was extracted using the hot phenol protocol. Bacterial cultures were grown to the desired OD_{600}, mixed with 0.2 volumes of STOP solution (95% ethanol, 5% phenol) and snap-frozen in liquid nitrogen. Once thawed on ice, the cell suspension was centrifuged for 20 min, 4500 rpm at 4 °C and the supernatant discarded. For cell lysis, pellets were suspended in 600 µl of 10 mg/ml lysozyme in TE buffer (pH 8.0) and incubated at 37 °C for 10 min. Next, 60 µl of 10% w/v SDS was added respectively, samples were mixed by inversion, and incubated in a water bath at 64 °C, 1-2 min before adding 66 µl 3 M NaOAc, pH 5.2. Phase separation was induced by mixing samples with 750 µl of acid phenol (Roti-Aqua phenol), followed by incubation for 6 min at 64 °C, while regularly inverting the tubes. Samples were briefly placed on ice to cool before centrifugation for 15 min, 13,000 rpm at 4 °C. The aqueous layer was transferred into a 2 ml phase lock gel tube (Eppendorf) and mixed with 750 µl chloroform (Roth) by shaking, followed by centrifugation for 12 min, 13,000 rpm at room temperature. For ethanol precipitation, the aqueous layer was transferred to a new tube, 2 volumes of a 30:1 EtOH:3 M NaOAc, pH 6.5 mix

was added and incubated overnight at -20 °C. Finally, samples were centrifuged, washed with cold 75% v/v ethanol and air-dried for 15 min. Precipitated RNA was resuspended in 50 µl nuclease-free water and stored at -80 °C.

2.4.6.7 Northern blotting

Cells were grown in TY (Figure 2, Figure 3, Figure 5, Appendix Figure S5), BHI, TY, TY supplemented with 0.5% glucose (TYG) or 0.5% fructose (TYF) and 70:30 sporulation medium (Appendix Figure S3), or on 70:30 sporulation plates (Appendix Figure S6) as described in the respective figure legends. RNA was purified as described above. Samples were mixed with equal amounts of gel loading buffer II (95% formamide, 18 mM EDTA, 0,025% SDS, 2% bromphenolblue), boiled at 98 °C for 5 min and cooled down on ice before loading on a denaturing 6% polyacrylamide gel containing 7 M urea. RNA was separated for 1 h and 50 min, 300 V and transferred onto a Hybond-N+ membrane (GE Healthcare Life Sciences) at 4 °C for 1 h, 50 V (~100 W) followed by UV irradiation (0.12 J/cm^2). Once cross-linked, membranes were pre-hybridized 1 h in ROTI Hybri-Quick Buffer (Roth) before adding ^{32}P-labeled DNA oligonucleotides. 5'-labeling was performed by incubating 10 pmol oligonucleotide with 1 µL of ^{32}P-γ-ATP (10 µCi/µL) and 5 U T4 Polynucleotide Kinase (Thermo Fisher Scientific) for 1 h at 37°C in a 10 µL reaction. Labeled oligonucleotides were purified using microspin G-25 columns (GE Healthcare) according to the manufacturer's instructions. Following hybridization overnight at 42 °C, membranes were washed three times with decreasing concentrations (5x, 1x and 0.5x) of SSC buffer (20x SSC: 3 M NaCl, 0.3 M sodium citrate, pH 7.0). Air dried membranes were then exposed onto a phosphor screen for 1-7 days and signals were visualized on a Typhoon FLA 7000 phosphor imager. Following signal detection, the membranes were stripped (0.1% SDS in freshly boiled water, 15 min), and incubated with ROTI Hybri-Quick Buffer (Roth) before adding a new ^{32}P-labeled DNA oligonucleotide.

2.4.6.8 *In vitro* transcription and radiolabeling of RNA

For *in vitro* transcription of SpoY, SpoX, *spo0A* and associated mutants, pFF-162/-163/-164/-245/-166/-248/-247/-167 and corresponding primers were used for template generation *via* Phusion High-Fidelity PCR. Resulting PCR products were purified from 1% agarose gels with NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel) to prevent the production of side products during *in vitro* transcription. *In vitro* transcription was performed using the Invitrogen MEGAscript T7 Transcription Kit (Thermofisher Scientific) in 40 µl reactions according to the manufacturer's protocol followed by DNase treatment (1µl TURBO DNase, 2U/µl for 15 min at 37 °C). Resulting RNA fragments were separated on a denaturing urea PAGE with 6% polyacrylamide and 7 M urea, followed by ethidium bromide (Carl Roth) staining for 10 min and imaging using an Intas Gel Doc system. Bands of correct size were cut out in small pieces and transferred into 2 ml tubes. For RNA elution, 750 µl RNA elution buffer (0.1 M NaAc, 0.1% SDS,

10 mM EDTA) was added and the samples were incubated at 4 °C and 1000 rpm overnight. Following centrifugation at 5,000 x g and 4 °C for 1 min, the supernatants were transferred to new tubes and RNA extraction was performed using a single phenol-chloroform extraction step (ROTI phenol/chloroform/isoamylalkohol). Purified RNA was resuspended in 20 µl RNase-free water and stored at -80 °C. For radioactive labeling, 50 pmol of *in vitro* transcribed SpoY, SpoY*, SpoX, SpoX* or *spo0A* (5'UTR starting from pTSS and start of CDS) was dephosphorylated using 25 U of calf intestine alkaline phosphatase (NEB) in a 50 µL reaction volume and incubated for 1 h at 37 °C. RNA was extracted again, using a single phenol-chloroform extraction step (ROTI phenol/chloroform/isoamylalkohol) and resuspended in 16 µl RNase-free water. Subsequently, 20 pmol of dephosphorylated and purified RNA was 5' end-labeled (20 µCi of ^{32}P-γATP) using 1 U of Polynucleotide Kinase (NEB) for 1 h at 37 °C in a 20 µL reaction volume. Finally, the labeled RNA was purified on a G-50 column (GE Healthcare) according to manufacturer's instructions and extracted from a polyacrylamide gel as described above following visualization on a Phosphorimager (FLA-3000 Series, Fuji). Purified RNA was resuspended in 10 µl RNase-free water and stored at -80 °C for up to 2 weeks.

2.4.6.9 EMSA (electrophoretic mobility shift assays)

EMSAs were performed by incubating 0.04 pmol of radio-labelled sRNA either alone or with increasing concentrations of *in vitro* transcribed mRNA. Prior to incubation, labeled sRNA and unlabeled mRNA were denatured at 95 °C for 1 min and chilled on ice for 5 min. All components were mixed to a final concentration of 1x structure buffer (10 mM Tris-HCl pH 7.0, 0.1 M KCl, 10 mM $MgCl_2$), 0.004 pmol/µl sRNA, 0.1 µg/µl yeast RNA (Ambion) and mRNA ranging from 0-1,000 pmol/µl. For EMSAs analyzing SpoY interactions with *spo0A* (5'UTR starting from pTSS), 500 pmol/µl purified *C. difficile* Hfq_6 was added as well (18). Reactions were incubated at 37°C for 1 h, stopped by adding 3 µl of 5x native loading dye (0.5x TBE, 50% glycerol, 0.2% xylene cyanol, 0.2% bromophenol blue) and directly loaded on a native 6% polyacrylamide gel at 4 °C in 0.5% TBE at 300 V for 3-4 h. The gel was dried for 1 h at 80 °C on a Gel Dryer 583 (Bio-Rad) and visualized after appropriate exposure on a Phosphorimager (FLA-3000 Series, Fuji).

2.4.6.10 In-line probing

In line probing exploits the natural instability of unpaired RNA that leads to differential degradation according to its structure, allowing elucidation of secondary structure information. In-line probing assays were performed by incubating 0.2 pmol of labeled RNA (sRNA or *spo0A* mRNA starting from pTSS) either alone or with increasing concentrations of *in vitro* transcribed *spo0A* (0.2, 2, and 20 pmol) or SpoX (0.2, and 2 pmol) for 40 h at room temperature in 1x in-line probing buffer (100 mM KCl, 20 mM $MgCl_2$, 50 mM Tris-HCl, pH 8.3). Both, sRNAs and mRNA were denatured at 95 °C for 1 min and chilled on ice for 5 min before assembling the reactions. Ladders

were prepared directly prior to loading. For the RNase T1 ladder 0.2 pmol of labeled RNA was incubated with 8 µL of 1x sequencing buffer (Ambion) at 95 °C for 1 min followed by the addition of 1 µL RNase T1 (0.1 U/ µl) and incubation at 37 °C for 5 min. The alkaline hydrolysis ladder was prepared by incubating 0.2 pmol labeled sRNA with 9 µL of 1x alkaline hydrolysis buffer (Ambion) and incubated at 95 °C for 5 min. All reactions were stopped by the addition of 10 µL of 2x colourless gel-loading solution (10 M urea, 1.5 mM EDTA) and stored on ice. 0.2 pmol of labeled RNA mixed with 10 µL 2x colourless gel-loading solution served as a control. Samples were resolved on a 10% (Figure 4B) or 6% (Figure EV2C; vol/vol) polyacrylamide, 7 M urea sequencing gel pre-run for 30 min, 45 W prior to sample loading. The gel was dried for 2 h on a Gel Dryer 583 (Bio-Rad) and visualized after appropriate exposure on a Phosphorimager (FLA-3000 Series, Fuji).

2.4.6.11 Reporter system assay

Single colonies of *C. difficile* 630 ΔSpoY harbouring pFF-185/-186/-191/-254/-285/-344, or -345 (FFS-536/-535/-537/-779/-798/-929/-930) and *C. difficile* ΔSpoX harbouring pFF-185/-187/-192/-260/-289/-346, or -347 (FFS-539/-538/-540/-785/-802/-931/-932) were used to inoculate overnight cultures in biological triplicates in sterile filtered TY supplemented with TAP. Main cultures were inoculated by diluting overnight cultures 1:330 in sterile filtered TY supplemented with TAP and grown till ME growth phase (OD_{600} of ~0.5) before harvesting 0.07 ODs respectively in a 96-well plate (5 min 4500 x g). Cell pellets were resuspended in 200 µl 4% PFA and incubated at room temperature for 30 min in the dark. Following cell fixation, samples were washed three times in 200 µl 1x PBS, resuspended in 30 µl 1x PBS and incubated overnight at 4 °C in the dark, allowing full maturation of mCherry. Subsequently, samples were diluted in 1x PBS to a final volume of 200 µl and mCherry fluorescence was detected using an Agilent NovoCyte Flow Cytometer. The sample acquisition threshold was set to 5,000, ungated in the FSC-H channel, and a maximum of 100,000 events. Three parameters were recorded for each particle, including FSC-H, SSC-H and PE-Texas Red-H. sRNA expression was verified *via* northern blot analysis of samples grown in the same growth conditions (Appendix Figure S5B&C).

2.4.6.12 Western blotting

To test Spo0A expression in a SpoY and SpoX deletion mutant and corresponding overexpression strains, FFS-591, FFS-593, FFS-535, FFS-594 and FFS-538 were inoculated into sterile filtered TY supplemented with TAP in biological triplicates from single colonies. Main cultures were inoculated by diluting overnight cultures to OD_{600} = 0.05 in sterile filtered TY containing TAP. For western blot analysis samples were taken at mid-exponential (5.5 h post inoculation, OD_{600} of ~ 0.5) and stationary phase (9 h post inoculation, OD_{600} of ~ 1.3) by harvesting 2 OD units *via* centrifugation for 5 min at 5,000 x g. Pellets were frozen overnight at -20°C, followed by cell resuspension in 50 µl 1x PBS and incubation for 50 mins at 37°C, leading to cell lysis (82).

Subsequently, cell lysates were mixed with equal amounts of 2x protein loading dye and boiled for 5 min at 95 °C. Of each sample, 0.3 OD (15 µl) was loaded and separated on a 15% SDS–polyacrylamide gel followed by transfer of proteins to a Protran 0.2 µm NC membrane (Amersham) at 4 °C for 1.5 h, 340 mA using a semi-dry blotting system. Equal loading of protein samples was confirmed *via* Ponceau S staining (Sigma-Aldrich) for 4 min. Staining was reversed by washing the stained membrane with 0.1 M NaOH for 1 min, followed by blocking in TBS-T with 5% powdered milk for 1 h at room temperature. Subsequently the membrane was incubated overnight at 4 °C with anti-Spo0A antibody (kindly provided by Amy Shen (83)) diluted 1:5,000 in TBS-T with 5% powdered milk and washed again 3x in TBS-T for 10 min. Following the last washing step the membrane was incubated for 1 h at room temperature with anti-mouse-HRP antibody (RRID: AB_228307; Thermo Scientific # 31430) diluted 1:10,000 in TBS-T with 5% powdered milk and finally washed 3x in TBS-T for 10 min before adding ECL substrate (Amersham) for detection of HRP activity using a CCD camera (ImageQuant, GE Healthcare). sRNA expression was verified *via* northern blot analysis of samples grown in the same growth conditions (Figure 5B).

2.4.6.13 RT-qPCR analysis of sporulation genes

Glycerol stocks of *C. difficile* strains (FFS-535, FFS-538, FFS-591, FFS-593, and FFS-594) were inoculated onto BHIS plates (BHI agar containing 5 g/L yeast extract (Roth) and 0.1% sterile filtered cysteine) supplemented with taurocholate (TA, 0.1% w/v) and TAP. Single colonies were then inoculated into liquid BHIS-TA-TAP media in biological triplicates and grown overnight. These cultures were diluted to OD_{600} = 0.05 with BHIS-TA-TAP media and grown till early stationary phase. Pre-cultures were then diluted using BHIS-TAP to OD_{600} = 0.05 and grown as main cultures until an OD_{600} of ~0.5. To induce sporulation, 120 µl of the main cultures were spread onto 70:30 agar plates (70% SMC media and 30% BHIS media) supplemented with TAP. Sporulating cultures were collected at 9 h and 12 h after plating using phosphate-buffered saline (PBS) and flash frozen immediately upon the addition of 0.2 volumes ice-cold STOP mix (5% water-saturated phenol (pH < 7.0) in ethanol). Total RNA was then isolated from collected samples using the hot-phenol extraction procedure as described above. For DNA removal, 5 µg of total RNA was treated with DNase I (Thermo Scientific) for 1 h at 37 °C and then further purified using a single phenol-chloroform extraction step (ROTI Phenol/Chloroform/Isoamylalkohol). Purified RNA was resuspended in 50 µl RNase-free water and stored at -80 °C. Reverse transcription was performed using the M-MLV Reverse Transcriptase Kit (Invitrogen) as per manufacturer's instructions, with 1 µg of DNase I treated, purified total RNA and Random Hexamer Primer (Invitrogen). cDNA was then diluted 20-fold and 1 µL were used for each qPCR reaction along with 20 nM of gene-specific oligonucleotides in a 10 µL reaction mix. qPCR was performed with Takyon™ No ROX SYBR 2x MasterMix blue dTTP (Eurogentec) reagent in

technical duplicates using the QuantStudio™ 5 Real-Time PCR cycler (Thermo Fisher Scientific) and following conditions: TakyonTM activation at 95 °C for 3 min; DNA denaturing at 95 °C for 10 sec; 40 cycles of annealing and extension at 60 °C for 60 sec, followed by melting curve denaturation at 95 °C for 15 sec and melting curve analysis at 55-95 °C using "step and hold" with 0.5 °C and 10 sec of incubation per step. Transcript levels were normalized to 5S rRNA and are displayed as ΔΔCt values, representing Log_2-fold change relative to FFS-591 (WT-*p*[ctl]). All oligonucleotide sequences used for RT-qPCR are listed in Appendix table S7. sRNA expression was verified *via* northern blot analysis of samples grown in the same growth conditions (Appendix Figure S6A).

2.4.6.14 Sporulation frequencies

Sporulation assays were performed in 70:30 sporulation broth medium according to Edwards *et al* with minor alterations (84). In short, *C. difficile* cultures (FFS-535, FFS-538, FFS-591, FFS-593, and FFS-594) were started in four biological replicates in BHIS medium supplemented with 0.1% taurocholate (TA) and 0.2% fructose until mid-log phase ($OD_{600} \leq 0.9$). Cultures were then back-diluted in 70:30 medium to $OD_{600} = 0.01$ and monitored for the production of spores. At each timepoint (6 h, 12 h, 24 h and 48 h post inoculation), samples were taken, serially diluted, and spotted (10 µl spots in technical triplicates) on BHIS-TA plates and incubated for 24 to 48 h to enumerate total number of CFU (spores and vegetative cells). Simultaneously, 500 µl from each culture was removed, mixed 1:1 with 95% EtOH and incubated for 30 min to kill all vegetative cells. EtOH-treated samples were then serially diluted, similarly plated, and incubated, representing the spore CFU. The sporulation frequency was determined by dividing the number of spores by the total number of CFUs at each time point (spore ratio), multiplied by 100 (percentage of spores formed).

For phase-contrast microscopy *C. difficile* strains were grown in 70:30 sporulation medium as described above in four biological replicates. At 12 h, 24 h and 48 h post inoculation, 1 ml of culture was removed from the anaerobic chamber, centrifuged at full speed for 30 s, and resuspended in ~10-30 µl of supernatant. Microcopy slides were prepared by placing 2 µl of the concentrated cultures onto thin 1% agarose pads that were applied directly to the surface of the slide. Phase-contrast microscopy was performed using a HC PLAN FLUOTAR 100x/1.32 PH3 oil immersion objective on a LEICA DM2500 microscope. Two fields of view for each strain and replicate were acquired and used to calculate the percentage of spores (the number of spores divided by the total number of spores, pre-spores, and vegetative cells; 300 cells per field of view were analyzed).

2.4.6.15 Murine model of *C. difficile* infection

All animal experiments were performed in agreement with the guidelines of the Helmholtz Centre for Infection Research (HZI), Brunswick, Germany, the national animal protection law (TierSchG), the animal experiment regulations (TierSchVersV), and the recommendations of the Federation of European Laboratory Animal Science Association (FELASA). Mice experiments were approved by the Lower Saxony State Office for Nature, Environment and Consumer Protection (LAVES), Oldenburg, Lower Saxony, Germany; permit No. 33.19-42502-04-19/3126.

C57BL/6N SPF mice were maintained (including housing) at the animal facilities of the HZI under enhanced specific pathogen-free (SPF) conditions for at least two weeks before the start of the experiment. Female mice aged between 12-14 weeks were used. Sterilized food and water were provided *ad libitum*. Mice were kept under a strict 12-hour light cycle (lights on at 7:00 am and off at 7:00 pm) and housed in groups of up to six mice per cage. All mice were euthanized by asphyxiation with CO_2 and cervical dislocation.

Infection experiments were performed with *C. difficile* 630 strain FFS-01 (WT), FFS-491 (ΔSpoY) and FFS-492 (ΔSpoX). SPF mice were weighted and treated with 10 mg/kg clindamycin 24 h prior to infection, administered *via* intraperitoneal injection to induce susceptibility to *C. difficile* infection (85). Spores were heat-treated at 65 °C for 20 min before infection to kill remaining vegetative cells. Mice were infected with 10^4 *C. difficile* spores in 200 µL 1x PBS administered *via* oral gavage. Following infection, mice were monitored and scored daily for symptoms of clinically severe CDI including behavior, posture, fur and skin, provoked behavior, weight loss and feces consistency. Mice showing signs of CDI were monitored twice a day and euthanized after losing 20% of their initial weight or developing severe clinical signs of features listed above.

For quantification of bacterial burden, fresh fecal samples were collected at different time points, their weight recorded, supplemented with 1.0 mm diameter zirconium/glass beads in 1 mL 1x PBS and subsequently homogenized for 50 sec with Mini-Beadbeater-96 (Biospec). To determine colony forming units (CFUs), serial dilutions of homogenized samples were plated on bioMérieuxTM *C. difficile* agar. For the quantification of spores, aliquots of homogenized samples were incubated 65 °C for 20 min to kill remaining vegetative cells and plated on bioMérieux™ *C. difficile* agar pretreated with 0.1% taurocholic acid to induce germination. Plates were cultured at 37 °C for 48 h in anaerobic jars before counting. CFUs of *C. difficile* were calculated by normalization to feces weight.

2.4.6.16 Prediction of RNA folding and sRNA target interactions

Secondary structures of sRNAs were predicted with the RNAfold WebServer, while RNAcofold was used to predict secondary structures of single stranded sRNA and mRNA sequences upon duplex formation (46,86). In both cases, structures were visualized with VARNA (87).

Potential sRNA-target interactions were predicted using IntaRNA by uploading either the SpoY, SpoX or *spo0A* sequence in combination with interaction partners revealed by RIL-seq analysis (Dataset EV2) (47). Default settings were used.

2.4.6.17 Prediction of sRNA target motif

Motif search and generation of sequence logos was accomplished with MEME version 5.4.1 (45). For both sRNAs, all target sequences were extracted from the RIL-seq data, including target interactions supported by <25 chimeras (Dataset EV2). For this purpose, the coordinates listed in the S-chimera table for "start of first read" and "start of last read" were used and extended 50 nt downstream. The resulting sequences were uploaded to MEME for target motif identification (SpoY = 28 targets, SpoX = 42 targets). The number of motifs to be found by MEME was set to 5 and only the given strand was searched. All other settings were left at default.

2.4.6.18 Quantification and statistical analysis

Statistical analysis of RIL-seq results is described above and was performed according to the original protocol (24,25). Quantification and analysis of western blot, northern blot and EMSA signals as well as in-line probing and microscopy images was performed with ImageJ (88). GraphPad Prism 9 was used for all statistical analyses and data visualizations, in combination with Inkscape 0.92.4 (88). Sample sizes and detailed descriptions of statistical analyses are indicated in the figure legends and method section for each experiment separately.

2.4.7 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 accession number GSE213005. The RIL-seq dataset can be accessed *via* an RNA-RNA interactome browser (https://resources.helmholtz-hiri.de/rilseqcd/). Phase-contrast microscopy images (Figure 7) are available on BioImage Archive under accession number S-BIAD622.

2.4.8 Acknowledgments

We would like to thank Gianluca Matera for assisting with the RIL-seq experiment and Aimee Shen for providing anti-Spo0A antibodies. We are also very grateful to Anke Sparmann and Jörg Vogel for providing critical and insightful feedback on the manuscript. T.L. was supported by Bavarian bayresq.net. J.S. was supported by the Deutsche Forschungsgemeinschaft under Grant FA 1113/2-1. M.G. was supported by the European Social Fund (ESF) under Grant 823905. Open Access funding enabled and organized by Projekt DEAL.

2.4.9 Author Contributions

Manuela Fuchs: Conceptualization; data curation; formal analysis; validation; investigation; visualization; methodology; writing – original draft; writing – review and editing. Vanessa Lamm-

Schmidt: Data curation; formal analysis; visualization; methodology; writing – review and editing. Tina Lenče: Data curation; formal analysis; validation; investigation; methodology; writing – review and editing. Johannes Sulzer: Formal analysis; validation; investigation; methodology; writing – review and editing. Arne Bublitz: Data curation; formal analysis; validation; investigation; methodology; writing – review and editing. Janet Wackenreuter: Data curation; formal analysis; validation; investigation; methodology; writing – review and editing. Milan Gerovac: Software; visualization; writing – review and editing. Till Strowig: Conceptualization; investigation; methodology; writing – review and editing. Franziska Faber: Conceptualization; supervision; funding acquisition; methodology; writing – original draft; writing – review and editing.

2.4.10 References

1. Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States, 2019. CDC (2019). doi:10.15620/cdc:82532
2. European Centre for Disease Prevention and Control. Healthcare-associated infections: *Clostridium difficile* infections. ECDC (2018).
3. Bartlett, J. G., Chang, T. W., Gurwith, M., Gorbach, S. L. & Onderdonk, A. B. Antibiotic-associated pseudomembranous colitis due to toxin-producing *Clostridia*. N. Engl. J. Med. 298, 531–534 (1978).
4. Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. *Clostridium difficile* infection. Nat. Rev. Dis. Prim. 2, 16020 (2016).
5. Peng, Z. et al. Update on antimicrobial resistance in *Clostridium difficile*: resistance mechanisms and antimicrobial susceptibility testing. J. Clin. Microbiol. 55, 1998–2008 (2017).
6. O'Grady, K., Knight, D. R. & Riley, T. V. Antimicrobial resistance in *Clostridioides difficile*. Eur. J. Clin. Microbiol. Infect. Dis. (2021). doi:10.1007/s10096-021-04311-5
7. Zhu, D., Sorg, J. A. & Sun, X. *Clostridioides difficile* biology: sporulation, germination, and corresponding therapies for *C. difficile* infection. Front. Cell. Infect. Microbiol. 8, 29 (2018).
8. Edwards, A. N. & McBride, S. M. Initiation of sporulation in *Clostridium difficile*: a twist on the classic model. FEMS Microbiol. Lett. 358, 110–118 (2014).
9. Shen, A., Edwards, A. N., Sarker, M. R. & Paredes-Sabja, D. Sporulation and germination in clostridial pathogens. in Gram-Positive Pathogens, Third Edition 7, 903–926 (American Society of Microbiology, 2019).
10. Deakin, L. J. et al. The *Clostridium difficile spo0A* gene is a persistence and transmission factor. Infect. Immun. 80, 2704–2711 (2012).
11. Rosenbusch, K. E., Bakker, D., Kuijper, E. J. & Smits, W. K. *C. difficile* 630Δ*erm* Spo0A regulates sporulation, but does not contribute to toxin production, by direct high-affinity binding to target DNA. PLoS One 7, e48608 (2012).

12. Lee, C. D. et al. Genetic mechanisms governing sporulation initiation in *Clostridioides difficile*. Curr. Opin. Microbiol. 66, 32–38 (2022).

13. Shen, A., Edwards, A. N., Sarker, M. R. & Paredes-Sabja, D. Sporulation and germination in Clostridial pathogens. in Gram-Positive Pathogens, Third Edition 7, 903–926 (American Society of Microbiology, 2019).

14. Boudry, P. et al. Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen *Clostridium difficile*. J. Bacteriol. 196, 3234–3248 (2014).

15. Maikova, A., Kreis, V., Boutserin, A., Severinov, K. & Soutourina, O. Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. Appl. Environ. Microbiol. 85, (2019).

16. Holmqvist, E. & Vogel, J. RNA-binding proteins in bacteria. Nat. Rev. Microbiol. (2018). doi:10.1038/s41579-018-0049-5

17. Boudry, P. et al. Identification of RNAs bound by Hfq reveals widespread RNA partners and a sporulation regulator in the human pathogen *Clostridioides difficile*. RNA Biol. 1–22 (2021). doi:10.1080/15476286.2021.1882180

18. Fuchs, M. et al. An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium. Proc. Natl. Acad. Sci. 118, (2021).

19. Marchais, A., Duperrier, S., Durand, S., Gautheret, D. & Stragier, P. CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers. RNA Biol. 8, 358–364 (2011).

20. Silvaggi, J. M., Perkins, J. B. & Losick, R. Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. J. Bacteriol. 188, 532–541 (2006).

21. Schmalisch, M. et al. Small genes under sporulation control in the *Bacillus subtilis* genome. J. Bacteriol. 192, 5402–5412 (2010).

22. Lamm-Schmidt, V. et al. Grad-seq identifies KhpB as a global RNA-binding protein in *Clostridioides difficile* that regulates toxin production. microLife 2, 1–21 (2021).

23. Hör, J., Gorski, S. A. & Vogel, J. Bacterial RNA biology on a genome scale. Mol. Cell 70, 785–799 (2018).

24. Melamed, S. et al. Global mapping of small RNA-target interactions in bacteria. Mol. Cell 63, 884–897 (2016).

25. Melamed, S. et al. Mapping the small RNA interactome in bacteria using RIL-seq. Nat. Protoc. 13, 1–33 (2018).

26. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153, 654–665 (2013).

27. Kwok, C. K. Dawn of the *in vivo* RNA structurome and interactome. Biochem. Soc. Trans. 44, 1395–1410 (2016).

28. Saujet, L., Monot, M., Dupuy, B., Soutourina, O. & Martin-Verstraete, I. The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. J. Bacteriol. 193, 3186–96 (2011).

29. Hofmann, J. D. et al. Metabolic reprogramming of *Clostridioides difficile* during the stationary phase with the induction of toxin production. Front. Microbiol. 9, 1970 (2018).

30. Park, S. et al. Dynamic interactions between the RNA chaperone Hfq, small regulatory RNAs and mRNAs in live bacterial cells. Elife 10, (2021).

31. Matera, G. et al. Global RNA interactome of *Salmonella* discovers a 5' UTR sponge for the MicF small RNA that connects membrane permeability to transport capacity. Mol. Cell (2022). doi:10.1016/j.molcel.2021.12.030

32. Menendez-Gil, P. & Toledo-Arana, A. Bacterial 3' UTRs: a useful resource in post-transcriptional regulation. Front. Mol. Biosci. 7, 617633 (2021).

33. Bronesky, D. et al. A multifaceted small RNA modulates gene expression upon glucose limitation in *Staphylococcus aureus*. EMBO J. 38, (2019).

34. El Mouali, Y. et al. CRP-cAMP mediates silencing of *Salmonella* virulence at the post-transcriptional level. PLOS Genet. 14, e1007401 (2018).

35. Ruiz de los Mozos, I. et al. Base pairing interaction between 5'- and 3'-UTRs controls icaR mRNA translation in *Staphylococcus aureus*. PLoS Genet. 9, e1004001 (2013).

36. Bar, A., Argaman, L., Altuvia, Y. & Margalit, H. Prediction of novel bacterial small RNAs from RIL-seq RNA–RNA interaction data. Front. Microbiol. 12, 635070 (2021).

37. Chen, Y., Indurthi, D. C., Jones, S. W. & Papoutsakis, E. T. Small RNAs in the genus *clostridium*. MBio 2, e00340-10 (2011).

38. Soutourina, O. A. et al. Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. PLoS Genet. 9, e1003493 (2013).

39. Małecka, E. M., Sobańska, D. & Olejniczak, M. Bacterial chaperone protein Hfq facilitates the annealing of sponge RNAs to small regulatory RNAs. J. Mol. Biol. 167291 (2021). doi:10.1016/j.jmb.2021.167291

40. Balasubramanian, D. & Vanderpool, C. K. New developments in post-transcriptional regulation of operons by small RNAs. RNA Biol. 10, 337–341 (2013).

41. Rice, J. B., Balasubramanian, D. & Vanderpool, C. K. Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. Proc. Natl. Acad. Sci. 109, E2691–E2698 (2012).

42. Møller, T., Franch, T., Udesen, C., Gerdes, K. & Valentin-Hansen, P. Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. Genes Dev. 16, 1696–1706 (2002).

43. Desnoyers, G., Morissette, A., Prévost, K. & Massé, E. Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. EMBO J. 28, 1551–1561 (2009).

44. Saujet, L., Pereira, F. C., Henriques, A. O. & Martin-Verstraete, I. The regulatory network controlling spore formation in *Clostridium difficile*. FEMS Microbiol. Lett. 358, 1–10 (2014).

45. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings. Int. Conf. Intell. Syst. Mol. Biol. 2, 28–36 (1994).

46. Bernhart, S. H. et al. Partition function and base pairing probabilities of RNA heterodimers. Algorithms Mol. Biol. 1, 3 (2006).

47. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. Nucleic Acids Res. 45, W435–W439 (2017).

48. Fimlaid, K. A. et al. Global analysis of the sporulation pathway of *Clostridium difficile*. PLoS Genet. 9, e1003660 (2013).

49. Deana, A. & Belasco, J. G. Lost in translation: the influence of ribosomes on bacterial mRNA decay. Genes Dev. 19, 2526–2533 (2005).

50. Prevost, K., Desnoyers, G., Jacques, J.-F., Lavoie, F. & Masse, E. Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. Genes Dev. 25, 385–396 (2011).

51. Reynolds, C. B., Emerson, J. E., de la Riva, L., Fagan, R. P. & Fairweather, N. F. The *Clostridium difficile* cell wallprotein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. PLoS Pathog. 7, e1002024 (2011).

52. Bradshaw, W. J., Kirby, J. M., Roberts, A. K., Shone, C. C. & Acharya, K. R. Cwp2 from *Clostridium difficile* exhibits an extended three domain fold and cell adhesion *in vitro*. FEBS J. 284, 2886–2898 (2017).

53. Westermann, A. J. Regulatory RNAs in virulence and host-microbe interactions. in Regulating with RNA in Bacteria and Archaea 6, 305–337 (American Society of Microbiology, 2019).

54. Errington, J. Regulation of endospore formation in *Bacillus subtilis*. Nat. Rev. Microbiol. 1, 117–126 (2003).

55. Stephenson, K. & Hoch, J. A. Evolution of signalling in the sporulation phosphorelay. Mol. Microbiol. 46, 297–304 (2002).

56. Childress, K. O. et al. The phosphotransfer protein CD1492 represses sporulation initiation in *Clostridium difficile*. Infect. Immun. 84, 3434–3444 (2016).

57. Edwards, A. N., Wetzel, D., DiCandia, M. A. & McBride, S. M. Three orphan histidine kinases inhibit *Clostridioides difficile* sporulation. J. Bacteriol. e0010622 (2022). doi:10.1128/jb.00106-22

58. Ul Haq, I., Brantl, S. & Müller, P. A new role for SR1 from *Bacillus subtilis* : regulation of sporulation by inhibition of kinA translation. Nucleic Acids Res. 49, 10589–10603 (2021).

59. Ohtani, K. et al. Unique regulatory mechanism of sporulation and enterotoxin production in *Clostridium perfringens*. J. Bacteriol. 195, 2931–6 (2013).

60. Kim, W. & Lee, Y. Mechanism for coordinate regulation of rpoS by sRNA-sRNA interaction in *Escherichia coli*. RNA Biol. 17, 176–187 (2020).

61. De Lay, N. & Gottesman, S. A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. Mol. Microbiol. 86, 524–538 (2012).

62. Mika, F. & Hengge, R. Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. RNA Biol. 11, 494–507 (2014).

63. Guillier, M. & Gottesman, S. Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. Mol. Microbiol. 59, 231–247 (2006).

64. Thomason, M. K., Fontaine, F., De Lay, N. & Storz, G. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. Mol. Microbiol. 84, 17–35 (2012).

65. Denham, E. L. The Sponge RNAs of bacteria –how to find them and their role in regulating the post-transcriptional network. Biochim. Biophys. Acta - Gene Regul. Mech. 1863, 194565 (2020).

66. Iosub, I. A. et al. Hfq CLASH uncovers sRNA-target interaction networks linked to nutrient availability adaptation. Elife 9, (2020).

67. Underwood, S. et al. Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production. J. Bacteriol. 191, 7296–7305 (2009).

68. Dawson, L. F., Valiente, E., Faulds-Pain, A., Donahue, E. H. & Wren, B. W. Characterisation of *Clostridium difficile* biofilm formation, a role for Spo0A. PLoS One 7, e50527 (2012).

69. Pettit, L. J. et al. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. BMC Genomics 15, 160 (2014).

70. Taggart, M. G. et al. Biofilm regulation in *Clostridioides difficile*: novel systems linked to hypervirulence. PLOS Pathog. 17, e1009817 (2021).

71. Mondhe, M., Chessher, A., Goh, S., Good, L. & Stach, J. E. M. Species-selective killing of bacteria by antimicrobial peptide-PNAs. PLoS One 9, e89082 (2014).

72. Sully, E. K. & Geller, B. L. Antisense antimicrobial therapeutics. Curr. Opin. Microbiol. 33, 47–55 (2016).

73. Green, M. R. & Sambrook, J. Molecular cloning: a laboratory manual. (Cold Spring Harbor Laboratory Press,U.S., 2012).

74. Ransom, E. M., Ellermeier, C. D. & Weiss, D. S. Use of mCherry red fluorescent protein for studies of protein localization and gene expression in *Clostridium difficile*. Appl. Environ. Microbiol. 81, 1652–1660 (2015).

75. Kirk, J. A. & Fagan, R. P. Heat shock increases conjugation efficiency in *Clostridium difficile*. Anaerobe 42, 1–5 (2016).

76. Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T. & Minton, N. P. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. Appl. Environ. Microbiol. 78, 4683–4690 (2012).

77. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).

78. Karp, P. D. et al. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Brief. Bioinform. 22, 109–126 (2021).

79. Hahne, F. & Ivanek, R. Visualizing genomic data using Gviz and Bioconductor. in Methods in molecular biology (Clifton, N.J.) 1418, 335–351 (2016).

80. Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinforma. 47, 11.12.1-34 (2014).

81. Krzywinski, M. et al. Circos: An information aesthetic for comparative genomics. Genome Res. 19, 1639–1645 (2009).

82. Fagan, R. P. & Fairweather, N. F. *Clostridium difficile* has two parallel and essential Sec secretion systems. J. Biol. Chem. 286, 27483–27493 (2011).

83. Putnam, E. E., Nock, A. M., Lawley, T. D. & Shen, A. SpoIVA and SipL are *Clostridium difficile* spore morphogenetic proteins. J. Bacteriol. 195, 1214–1225 (2013).

84. Edwards, A. N., Nawrocki, K. L. & McBride, S. M. Conserved oligopeptide permeases modulate sporulation initiation in *Clostridium difficile*. Infect. Immun. 82, 4276–4291 (2014).

85. Theriot, C. M., Bowman, A. A. & Young, V. B. Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for *Clostridium difficile* spore germination and outgrowth in the large intestine. mSphere 1, (2016).

86. Lorenz, R. et al. ViennaRNA package 2.0. Algorithms Mol. Biol. 6, 26 (2011).

87. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. Bioinformatics 25, 1974–1975 (2009).

88. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675 (2012).

89. Emerson, J. E. et al. A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. Mol. Microbiol. 74, 541–556 (2009).

2.4.11 Supplementary Information

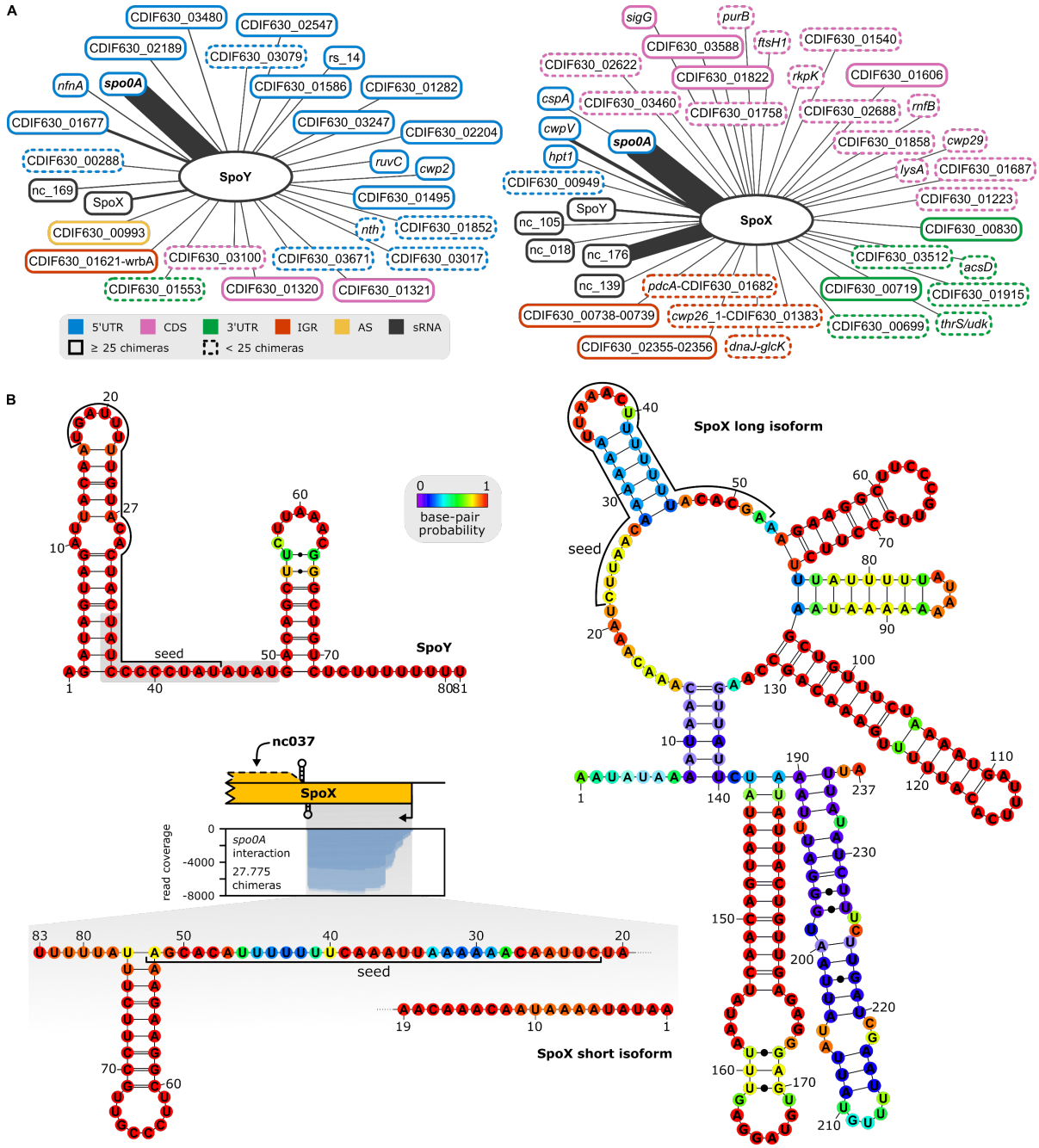


Figure EV1: RIL-seq reveals *spo0A* as a target of sRNA-mediated post-transcriptional regulation. **(A)** Target network of SpoY and SpoX, targets supported by ≥25 chimeras are marked by a solid line, while targets supported by <25 chimeras are highlighted with a dashed line. Target types are discriminated by color. Edge strength correlates with the number of chimeras supporting an individual interaction. **(B)** Predicted secondary structure (RNAfold (86)) for SpoY and both isoforms of SpoX are provided. Seed regions relevant for *spo0A* interaction were predicted *in silico* (IntaRNA (47)) and emphasized in the secondary structure. Read coverage of SpoX by SpoX-*spo0A* chimeric reads is highlighted in relation to the SpoX-encoding region.

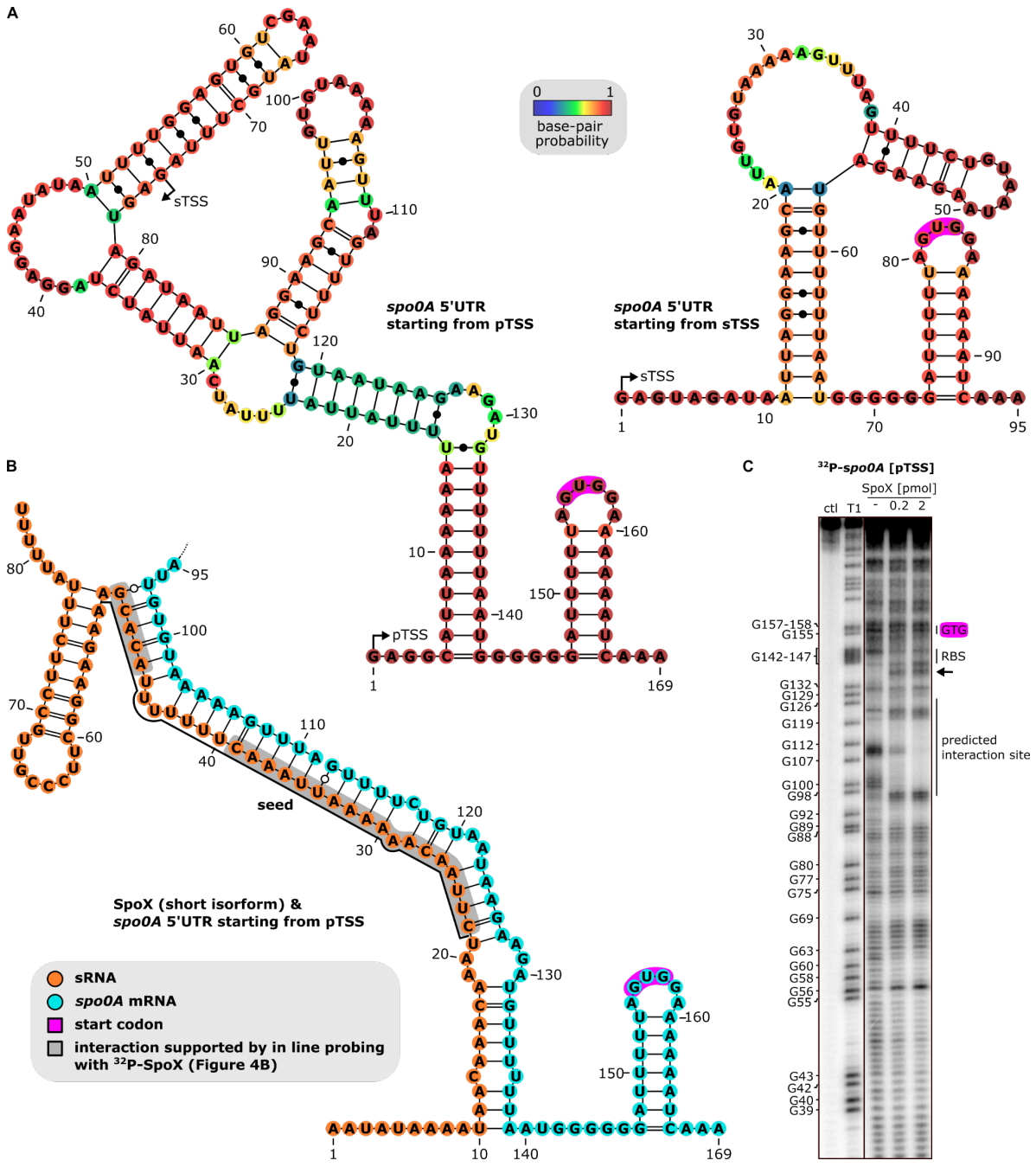


Figure EV2: SpoX binding potentially renders the *spo0A* RBS region more accessible. (A) Predicted secondary structures of the *spo0A* 5'UTR and beginning of CDS, starting from the primary transcription start site (pTSS) and secondary TSS (RNAfold (86)). The start codon is highlighted in pink. **(B)** Predicted secondary structures of the *spo0A* 5'UTR and beginning of CDS upon dimer formation with SpoX (shown is the short isoform). Resulting dimer for the depicted region looks identical, independent of which SpoX isoform or *spo0A* 5'UTR length was used (RNAcofold (46)). sRNA and mRNA are highlighted in orange and blue, respectively. sRNA-target base-pairings supported by in-line probing (Figure 4B) are shaded in grey. **(C)** In-line probing of 0.2 pmol of ^{32}P-labeled spo0A (starting from pTSS) in the absence (lane 3) or presence of increasing concentrations (lane 4&5) of SpoX (short isoform). RNase T1 digested spo0A serves as a ladder. Start codon and predicted seed region are highlighted. A representative image of three independent experiments is shown.

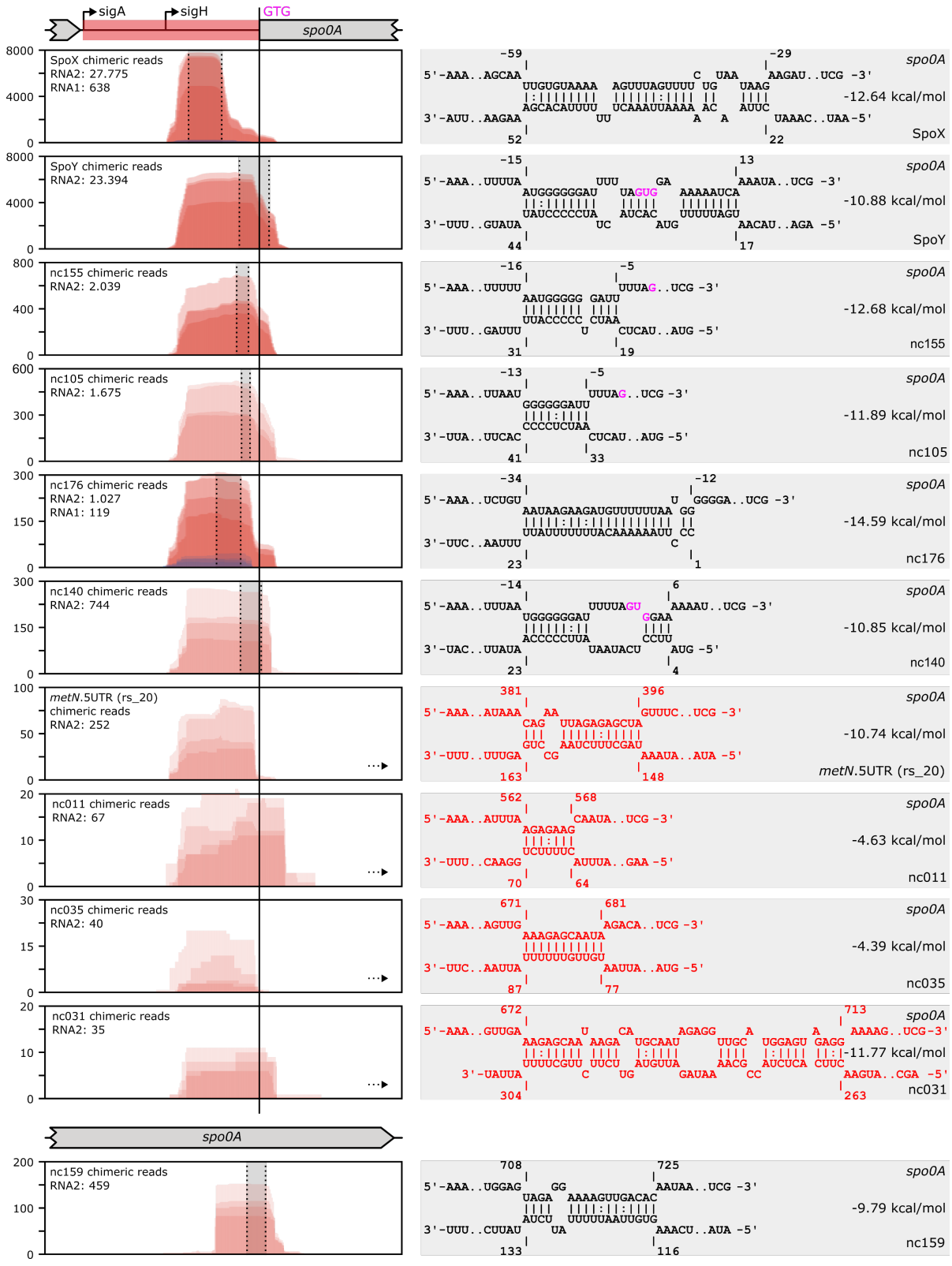


Figure EV3: *spo0A* is a target of extensive sRNA-mediated post-transcriptional regulation. On the left site, read coverage (y-axis) of *spo0A* by chimeric reads of all sRNA-*spo0A* interactions detected by RIL-seq analysis is depicted. The *spo0A* 5'UTR position including pTSS (sigA) and sTSS (sigH), start codon and coding sequence are marked on the x-axis. Chimeric reads covering *spo0A* were predominantly found at position 1 (RNA1) in a chimera and are color-coded in red. Chimeric reads found at position 2 (RNA2) are marked in blue. The number of chimeric reads covering each interaction is provided on the left. On the right site, base pairing information and location of the predicted binding sites (IntaRNA (47)) for each interaction are highlighted. *In silico* predictions that do not overlap with RIL-seq data are marked in red,

and location of the interaction in relation to the RIL-seq peak is indicated by an arrow. Predicted interactions overlapping with RIL-seq data are shaded in gray and superimposed over the coverage plots. The *spo0A* nucleotide position is calculated relative to the *spo0A* start codon (highlighted in pink).

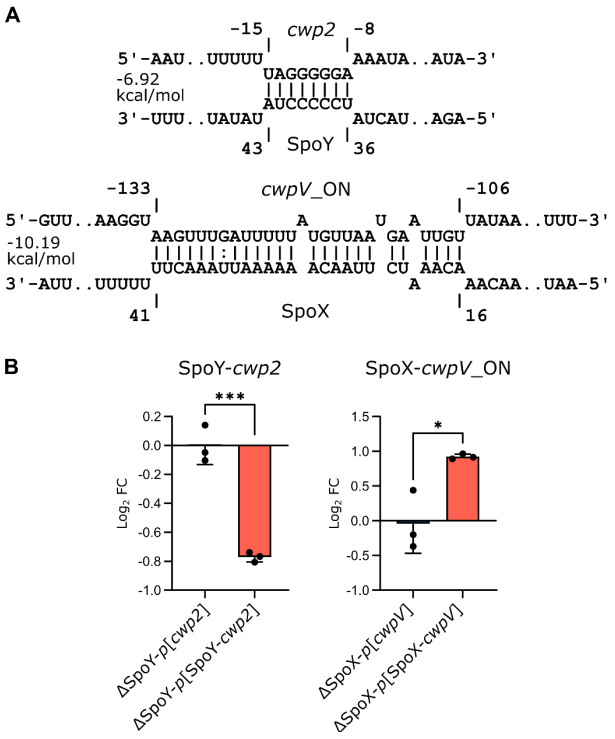
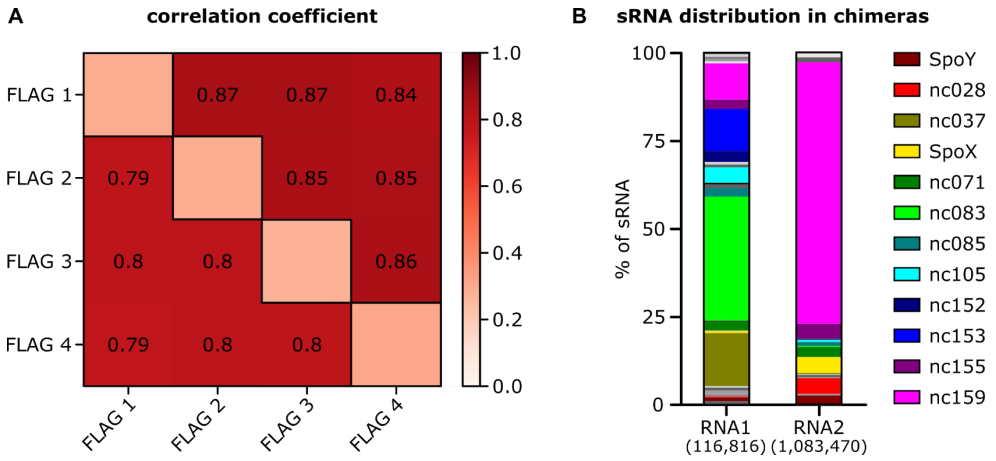
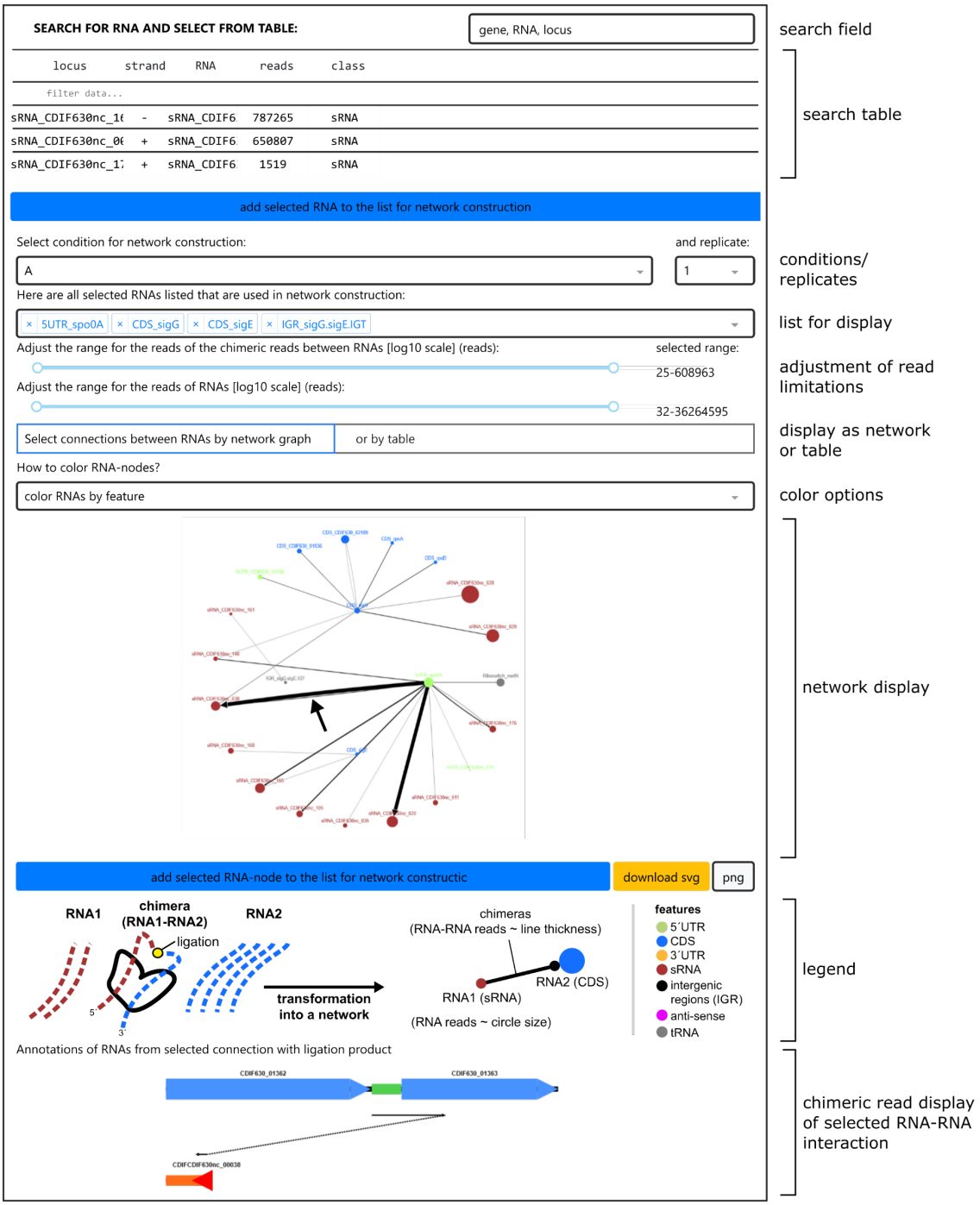


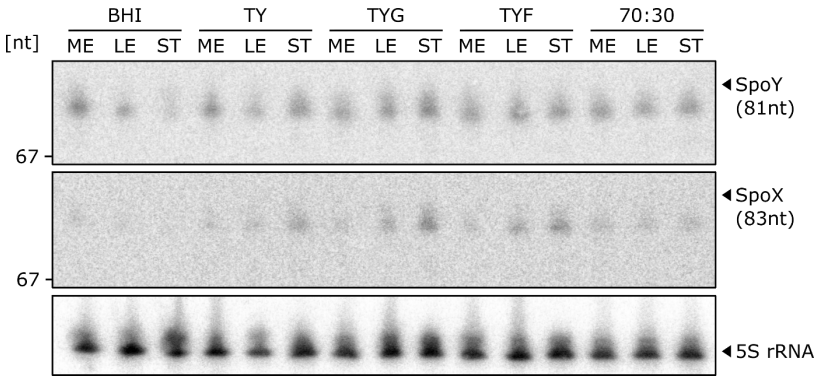
Figure EV4: SpoY and SpoX target additional genes besides *spo0A*. (A) Base pairing information and location of the predicted binding sites (IntaRNA (47)) for SpoY-*cwp2* and SpoX-*cwpV* with the invertible region within the *cwpV* 5'UTR in the ON orientation, allowing CwpV expression (89). The nucleotide positions for *cwpV* and *cwp2* are calculated relative to the respective start codons. **(B)** mCherry fluorescence of translational fusion constructs (error bars represent the mean ± SD of n=3 biological replicates, Appendix Figure S5A) expressed in the respective sRNA knock-out background. Fluorescence intensity was normalized to that of the respective *p*[*cwp2*]/*p*[*cwpV*] ctl. Unpaired t test was used to calculate statistical significance. Not significant (ns) $P > 0.05$; (∗) $P \leq 0.05$; (∗∗∗) $P \leq 0.001$.



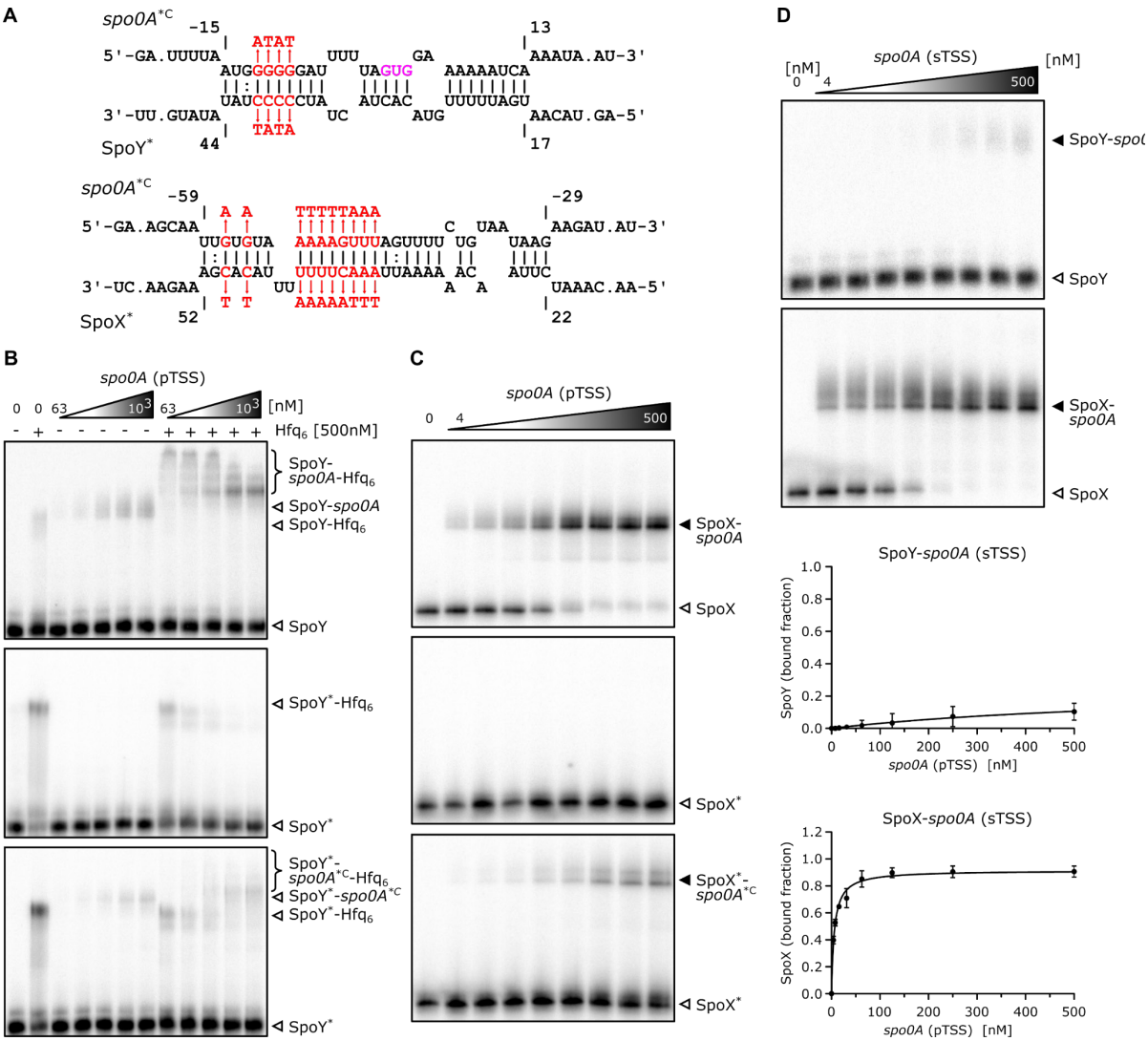
Appendix Figure S1: Hfq serves as a platform for RNA-RNA interactions in *C. difficile*. (A) Replicate reproducibility calculated as correlation coefficient by comparing the numbers of mapped fragments in corresponding genomic windows between each pair of libraries, for single (below diagonal) and chimeric (above diagonal) fragments, respectively. **(B)** Distribution of sRNAs in chimeric fragments, where RNA1 constitutes the 5'end and RNA2 the 3'end of a chimera (n=4). sRNAs that are present in ≥ 1.5% off all chimeras in either RNA1 or RNA2 are highlighted in order of genomic location. A total of 116,816 chimeric reads mapped to sRNAs in RNA1 and 1,083,470 to sRNAs in RNA2.



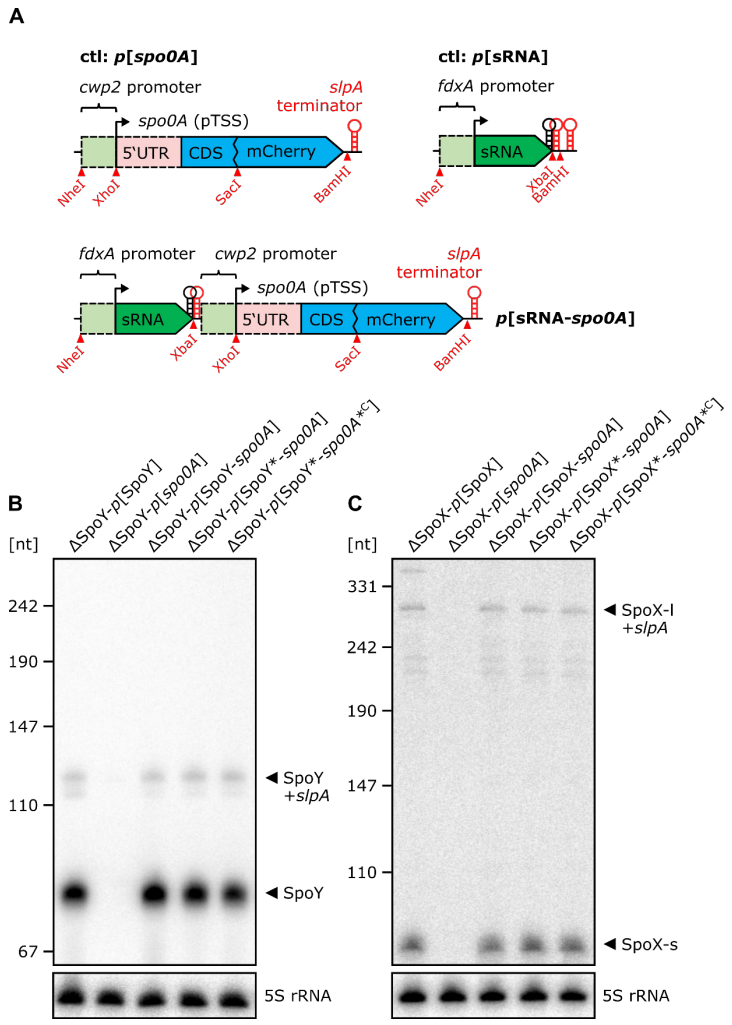
Appendix Figure S2: Rilseqcd is a web-browser that allows easy access to our RIL-seq data and an interactive search for RNA-RNA interactions. Screenshot of the RIL-seq browser accessible *via* https://resources.helmholtz-hiri.de/rilseqcd/. Details explaining the available options are given on the right. So far only one condition and one replicate are available, the latter because all four replicates have been pooled into a single dataset. Specific targets can be searched and added to the network display either *via* the search bar and table at the top, or by directly typing into the "list for display" field. If no targets are selected, a network of all detected interactions will be shown. By clicking on specific interaction in the "network display", a schematic representation of the selected RNA-RNA interaction will appear on the bottom.



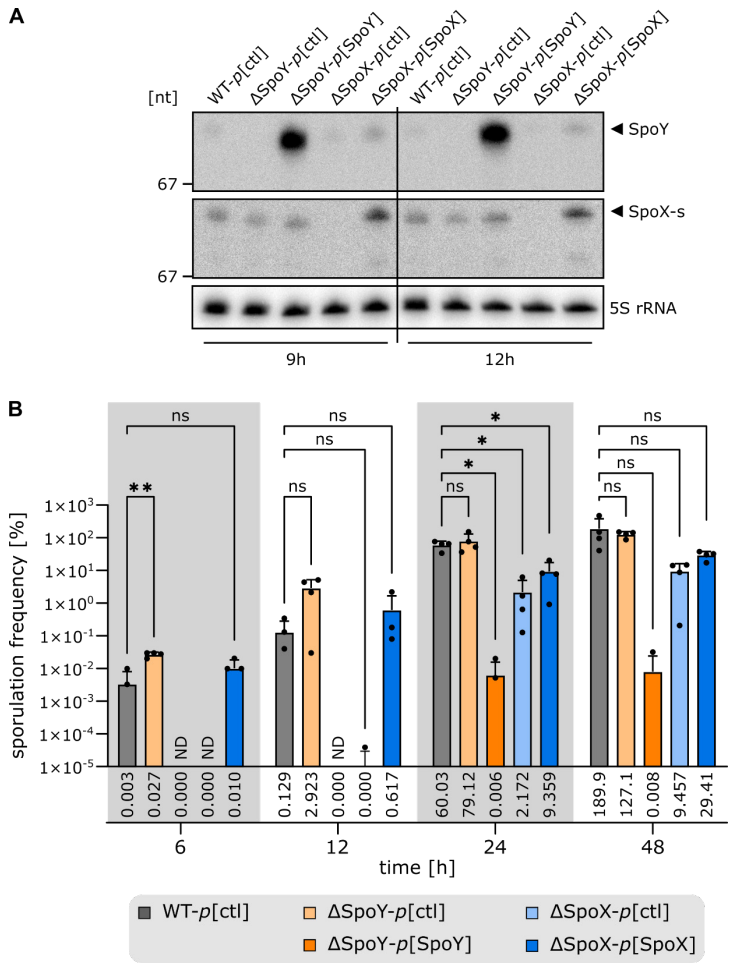
Appendix Figure S3: SpoY and SpoX expression in selected growth conditions. Northern blot validation of SpoY and SpoX expression in mid-exponential (ME), late exponential (LE) and stationary (ST) phase of growth in either BHI, TY, TY supplemented with 0.5% glucose (TYG) or 0.5% fructose (TYF) or 70:30 sporulation medium respectively. 5S rRNA served as a loading ctl. A representative image of three independent experiments is shown.



Appendix Figure S4: SpoY and SpoX directly interact with the *spo0A* mRNA *in vitro*. (A) *In silico* predicted SpoY-*spo0A* and SpoX-*spo0A* interaction sites (IntaRNA (47)). Mutations introduced in the sRNA seed region as well as compensatory mutations in the *spo0A* target region are highlighted in red. The *spo0A* nucleotide position is calculated relative to the *spo0A* start codon (highlighted in pink). **(B-C)** EMSAs performed with either ^{32}P-labeled SpoY **(B)** or SpoX (short isoform) **(C)** with increasing concentrations of the long *spo0A* 5'UTR and first 69 nt of CDS, respectively. Purified Hfq was added to facilitate SpoY-*spo0A* complex formation. Mutating the respective sRNA seed region (SpoY*/SpoX*) abolished the interaction, while introducing compensatory mutations into the *spo0A* target region (*spo0A**C) slightly rescued the complex formation. A representative image of three independent experiments is shown, respectively. **(D)** EMSAs and corresponding quantifications (n=3) were performed with either ^{32}P-labeled SpoY or SpoX (short isoform) with increasing concentrations of the short *spo0A* 5'UTR and first 69 nt of CDS, respectively**.**



Appendix Figure S5: SpoY and SpoX directly interact with the *spo0A* mRNA *in vivo*. **(A)** Schematic representation of translational fusion constructs designed for *in vivo* reporter system assays. Restriction sites that allow easy exchange of each component individually are annotated. "sRNA" refers to either SpoY or SpoX (long isoform. mCherry fused to the *spo0A* 5'UTR (starting from the pTSS) and beginning of CDS serves as a readout. For Figure EV4, *spo0A* was replaced by either *cpw2* or *cwpV* 5'UTR and first 20 aa of CDS. **(B-C)** Northern blot validation of sRNA expression from reporter constructs grown in TY till ME growth phase in the respective sRNA deletion mutant. A representative image of three independent experiments is shown, respectively.



Appendix Figure S6: sRNA-mediated regulation of *spo0A* affects sporulation specific genes and sporulation frequencies. (A) Northern blot validation of sRNA expression in conditions used for Figure 6. RNA was extracted from samples (n = 3) taken at 9 h and 12 h post induction of sporulation on 70:30 sporulation plates. **(B)** Sporulation frequencies (n=4) of a WT strain (*p*[ctl]), sRNA knock-out mutants (ΔSpoY/ΔSpoX-*p*[ctl]) and strains constitutively expressing the respective sRNA (ΔSpoY/ΔSpoX-*p*[ΔSpoY/ΔSpoX]) at 6 h, 12 h, 24 h and 48 h post inoculation of 70:30 liquid sporulation medium. ND: not determined – no viable spores. 2-way ANOVA with Dunnett's multiple comparison test was used to calculate statistical significance. Not significant (ns) $P > 0.05$; (∗) $P \leq 0.05$; (∗∗) $P \leq 0.01$.

3 Discussion

An extensive discussion of results published as part of this PhD thesis can be found in the associated publications (see 2.3.5 & 2.4.5) [312,313]. This applies in particular to the functional characterization of sRNAs and sRNA-target interactions. Thus, this chapter will focus on aspects that have been touched upon only marginally, and highlight general conclusions drawn from analysing both publications as a whole rather than separately.

3.1 Hfq and its function in *C. difficile*

Research in human pathogens such as *S. enterica, V. cholerae, P. aeruginosa* and *S. aureus* has highlighted the importance of post-transcriptional regulation in mediating adaptive processes during host colonization and infection [1,47]. RBPs are known to impact and enable these interactions. In particular, the role of Hfq as an RNA matchmaker facilitating sRNA-mRNA base-paring is well established [9]. Despite its importance in Gram-negative species, Hfq function in Gram-positive species has remained enigmatic as sRNA regulation in these organisms largely occurs independently of Hfq [174–176,314]. RIP-seq analysis performed by us and others uncovered a vast number of sRNAs and mRNAs bound by Hfq in *C. difficile* [311,312]. These results were corroborated by our RIL-seq data which not only identified multiple Hfq-bound sRNAs but extensive Hfq-associated sRNA-mRNA interactions [313]. Furthermore, our results indicated that deletion of *hfq* leads to decreased stability and steady-state levels of several Hfq-associated sRNAs. Decreased sRNA stability upon Hfq deletion is a common phenotype in Gram-negative species and thus suggest a conserved Hfq function in the Gram-positive *C. difficile* [312]. This hypothesis is further supported by a previous publication that reported altered growth on various carbon and nitrogen sources, decreased stress tolerance, and increased biofilm formation following Hfq depletion in *C. difficile* [308,312]. High conservation of residues relevant for RNA binding in *C. difficile* compared to Gram-negative model organisms might explain the similarities in Hfq function (Figure 8) [186,197]. This is particularly true for conserved arginines on the rim of Hfq that catalyze base-pair formation between RNAs [196,197]. By measuring RNA annealing activity of Hfq, Zheng and colleagues revealed that the number of arginine residues at the Hfq rim seems to correlate with its ability to facilitate sRNA-mRNA base-pairing [196,197]. In Gram-negative species, the Hfq rim contains up to three arginine residues, forming a basic patch. In contrast, *B. subtilis* and *S. aureus*, both species in which Hfq is apparently not involved in post-transcriptional regulation, encode only one or none respectively (Figure 8) [174,197]. However, the Hfq rim motif in *C. difficile* mirrors that of the Gram-negative *P. aeruginosa,* a species in which Hfq-mediated post-transcriptional regulation is well-documented (Figure 8) [197,315,316]. Nevertheless, there are several differences between the *C. difficile* Hfq and its Gram-negative counterparts, most notably the

poorly conserved CTD [199]. In *E. coli*, the Hfq CTD consists of multiple acidic residues that compete with RNA binding at the rim-region and thereby promoting a rapid cycling of RNA on Hfq (see 1.3.1) [198–200]. Additional functions in DNA binding, distinguishing among different RNAs, stabilization of the Hfq hexamer, oligomerization, cell localization, and phase separation have been reported as well [317–321]. However, the Hfq CTD in *C. difficile* is considerably shorter and encodes mostly asparagine residues, indicating potential differences in function that require further investigation (Figure 8) [310].

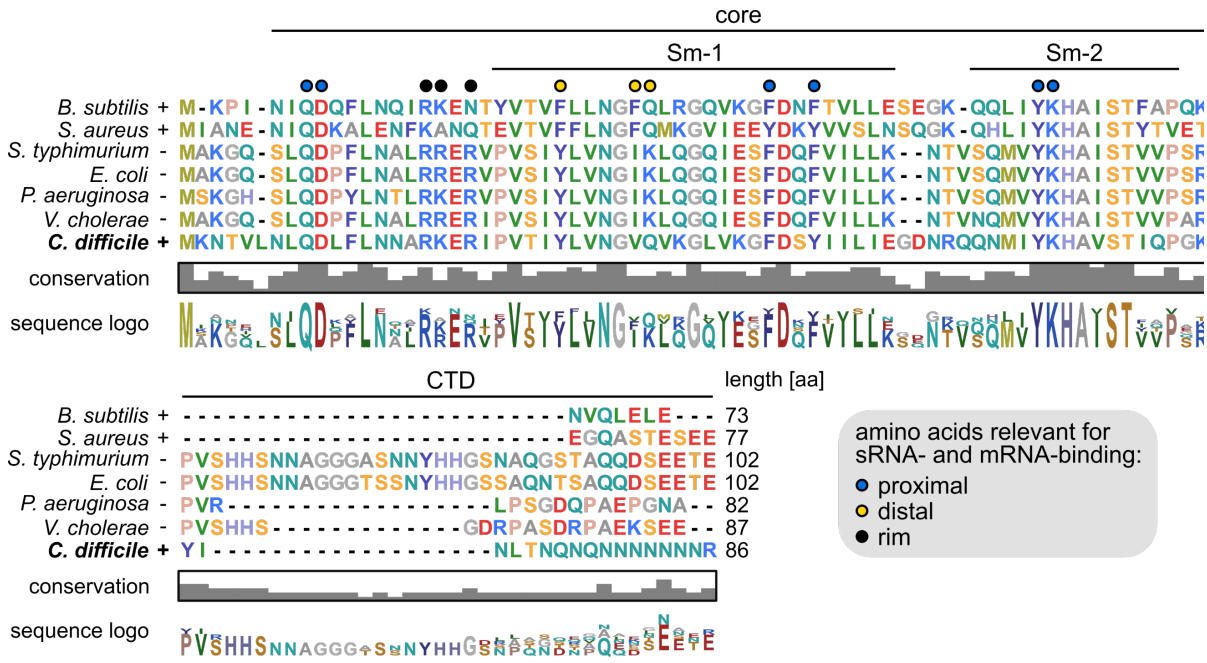


Figure 8: Amino acid sequence alignment of several Hfq homologues. Primary structure conservation of the *B. subtilis*, *S. aureus*, *S. typhimurium*, *E. coli*, *P. aeruginosa*, *V. cholerae*, and *C. difficile* Hfq are shown. Gram-differentiation of each species, Sm-1, Sm-2 as well as the core region and CTD are indicated. Coloured dots above the alignment mark amino acid residues that have been shown to be relevant for RNA-binding in *E. coli* [186,197].

3.2 mRNA 3'UTRs – a neglected target of sRNA-mediated regulation?

Besides differences in the Hfq primary structure, we also observed differences in the composition of Hfq-bound RNAs when comparing Hfq immunoprecipitation data from *C. difficile* to data obtained from Gram-negative bacteria. Single fragments as well as chimeric reads derived from Hfq RIL-seq in *E. coli* and *S. enterica* were clearly dominated by reads mapping to sRNAs, followed by CDS and 5'UTRs, whereas mRNA 3'UTRs were barely represented [50,322,323]. In contrast, the fraction of chimeric reads mapping to 3'UTRs in *C. difficile* was surprisingly large, exceeding the fraction of reads mapping to 5'UTRs (Figure 9) [313]. We have speculated that mRNA 3'UTRs might constitute a neglected source of regulatory elements that are targeted by sRNAs in *C. difficile*, reflected by the high number of sRNA-3'UTR chimeric fragments [312,313]. This hypothesis was

further supported by our RNAtag-seq data which revealed a median 3'UTR length of >100 nt in *C. difficile* [312]. Other Gram-positive species, including *S. aureus* and *L. monocytogenes*, harbour similarly long 3'UTRs, whereas in *E. coli* only 15% of 3'UTRs exceed 100 nt [123,324,325]. Considering that the average size of Rho-independent transcription terminators ranges from 40-50 nt, long 3'TURs might have additional regulatory functions, outside of transcription termination, that contribute to gene expression [43,123]. This includes the potential role of 3'UTRs in the regulation of mRNA translation and decay [43,123,326,327]. In *S. aureus,* the long 3'UTR of *icaR* contains an anti-SD motif that can bind the 5'UTR located SD sequence, forming a double strand that inhibits translation and induces RNase III-mediated double strand cleavage [123]. A similar 5'-3'UTR interaction was recently described for the *hbs* mRNA of *B. subtilis* [328]. However, instead of inducing degradation, the interaction inhibits RNase Y-mediated cleavage of the *hbs* 5'UTR by blocking the RNase Y cleavage site [328]. Furthermore, 3'UTRs can serve as a direct entry point for ribonucleases to initiate mRNA degradation as observed for the *hmsT* mRNA in *Y. pestis* or the *hilD* mRNA in *S. enterica* [329,330]. Interestingly, sRNA-mRNA interactions affecting these regulatory functions have already been reported (see 1.2.4). RsaI, an sRNA encoded by *S. aureus* interacts with the *icaR* 3'UTR potentially stabilizing the *icaR* 5'-3' duplex formation to downregulate *icaR* translation [122]. In contrast, SdsR and Spot 42 binding to the *hilD* 3'UTR prevent RNaseE-dependent degradation, resulting in increased *hilD* mRNA levels in *Salmonella* [129]. Taken together, these examples highlight the possibility of similar sRNA-3'UTR interactions occurring in *C. difficile,* particularly given the widespread presence of long 3'UTRs in this bacterium [43,312]. However, future analyses will be necessary to fully understand the mechanism of these interactions as well as their prevalence in *C. difficile*.

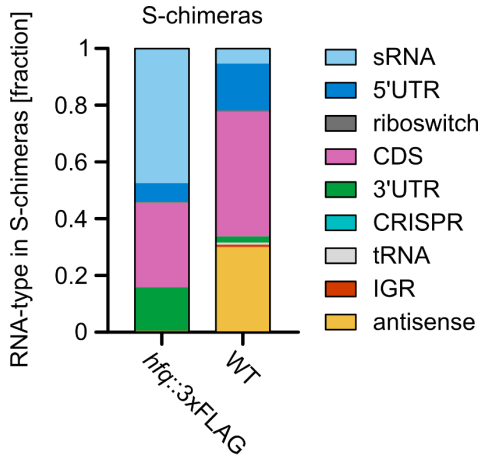


Figure 9: RIL-seq RNA distribution in S-chimeras. Fractions of sequenced chimeric fragments corresponding to S-chimeras (remaining chimeric fragments following filtering for statistical significance and manual curation) plotted by RNA type for the WT ctrl, and *C. difficile hfq*::3×FLAG strain [313].

3.3 Riboswitches – more than just *cis*-regulators?

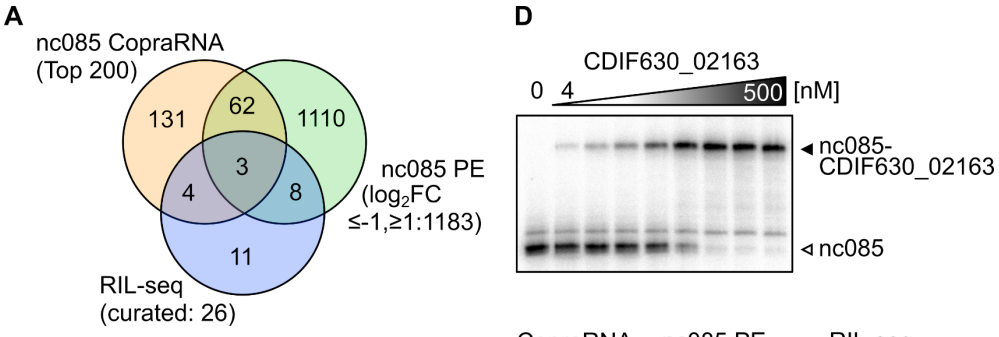
In contrast to 3'UTRs, the regulatory capacities of 5'UTRs are well documented in bacteria [331,332]. Accordingly, sRNAs targeting 5'UTRS, in particular the ribosome binding site or start codon

region, are a prevalent regulatory paradigm (see 1.2.4.) [3]. Interestingly, analysis of our Hfq RIP-seq data revealed that approximately 40% of the 5'UTR-located peaks harbour a riboswitch, resulting in enrichment of 50 out of 79 annotated riboswitches in *C. difficile* [312]. Similar results were obtained from independent Hfq pull-down experiments in *C. difficile* and *B. subtilis* [311,333]. The authors initially speculated that extensive riboswitch binding by Hfq in *B. subtilis* might occur non-specifically due to high cellular concentrations of stable riboswitch-derived fragments that compete with sRNAs for Hfq binding [333]. However, there are examples of riboswitches that operate as *trans*-regulators on a post-transcriptional level in addition to regulation of their *cis*-encoded target mRNA [334,335]. SreA and SreB, two SAM riboswitches encoded in *L. monocytogenes*, downregulate expression of the virulence regulator *prfA* by binding to its 5'UTR independent of their *cis*-regulatory function [37]. A similar dual-function has been reported for the adenosylcobalamine-responsive riboswitches Rli55 and EutX, encoded by *L. monocytogenes* and *E. faecalis*, respectively (see 1.2.2.1) [38,39]. However, rather than regulating gene expression of a *trans*-encoded target, Rli55 and EutX sequester the two-component response regulator EutV in the absence of adenosylcobalamine [38,39]. These examples point toward a potential function in post-transcriptional regulation of *trans*-encoded targets by riboswitches [334,335]. Supporting this hypothesis, our RIL-seq data revealed multiple chimeras consisting of a riboswitch and either an sRNA or mRNA [313]. This not only suggesting a *trans*-regulatory function of these riboswitches, but also as a potential role of Hfq in mediating them. Interestingly, riboswitches are more prevalent among Bacilli and Clostridia than any other bacterial class including γ-Proteobacteria [336]. This might explain why reports of dual-functioning riboswitches have been restricted to Bacilli, despite extensive research on sRNA-mediated post-transcriptional regulation in γ-Proteobacteria (e.g. *E. coli*) [336]. Taken together, these results point towards a mechanism in which ligand sensing by riboswitches can result in regulation of *cis*-encoded targets as well as *trans*-encoded targets, be they mRNAs, proteins or other sRNAs, on a larger scale than originally anticipated [335]. Nonetheless, further analyses are necessary to fully understand the extend of these interactions as well as the role of Hfq in mediating them.

3.4 Identification and characterization of sRNA-target interactions

Besides identifying Hfq-associated RNA-RNA interactions, RIL-seq network data can be mined for potential new sRNA candidates, as highlighted by previous publications (see 1.2.1) [50,337]. Similarly, we were able to expand the set of annotated sRNAs in *C. difficile* by 24 new candidates, resulting in a total of 66 sRNAs when combined with our dRNA-seq data [312,313]. However, multiple publications have illustrated that experimental conditions strongly affect sRNA expression and Hfq-dependent and independent sRNA-target interactions [50,146,338] Analysis of the expression of newly identified sRNAs *via* Northern blot experiments support this observation [313]. Verified

sRNAs were either strongly expressed throughout growth or during late-exponential and early stationary phase of growth, which corresponded to the growth condition that was chosen for RIL-seq analysis [313]. Accordingly, applying RIL-seq or dRNA-seq to additional growth or stress conditions in *C. difficile* will most likely expand the scope of known sRNAs even further, while also aiding in identifying their regulatory function [27,146,339]. Rather than relying on a single dataset, several studies have also highlighted the advantages of combining multiple experimental and computational approaches to identify sRNA-target interactions and sRNA function [229,340]. Georg and colleagues, for example, combined proteomics and MAPS data from *S. aureus* with computational sRNA-target predictions *via* CopraRNA to identify high probability targets for the RsaA sRNA [229]. We attempted a similar approach by combining *in silico* predictions of nc085 targets (CopraRNA) with our RIL-seq and nc085 pulse-expression datasets to find promising nc085 interaction partners (Figure 10A). Indeed, we were able to identify three targets (CDIF630_00926 / 02163 & 01745) that were present in all datasets. The nc085-CDIF_02163 interaction was additionally verified *via* electrophoretic mobility shift assays (Figure 10B-D).



B

locus tag:	function:	CopraRNA kcal/mol	nc085 PE log_2FC	RIL-seq # chimeras
CDIF630_00926	hypothetical protein	-16.15	1.03	8,592
CDIF630_02163	4Fe-4S binding protein	-16.14	1.59	45
CDIF630_01745	esterase	-12.91	-1.27	132

C

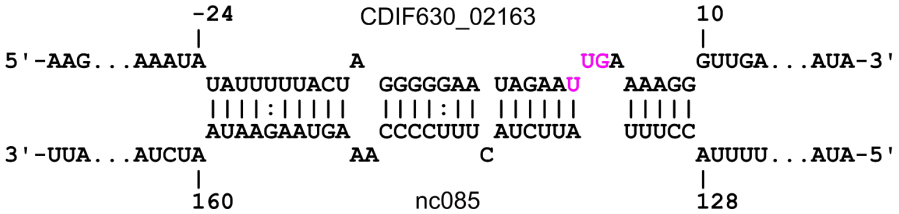


Figure 10: Dataset integration facilitates identification of sRNA targets. (A) Venn diagram showing the overlap of the top 200 CopraRNA target predictions, targets that were differentially expressed (log_2 fold change (FC) of ≤-1 or ≥1) upon pulse-expression (PE) of nc085 and RIL-seq chimeras containing nc085 [312,313,341]. **(B)** List of genes that were found in all three datasets. Energy score according to CopraRNA, PE log_2 FC and the number of RIL-seq chimeric fragments for each target are indicated. **(C)** nc085-CDIF630_02163 base-pairing information predicted using IntaRNA. Position within the 5'UTR of CDIF630_02163 is calculated relative to the start codon (highlighted in pink) [342]. **(D)** Electrophoretic mobility shift assay performed with ^{32}P-labeled nc085 and increasing concentrations of the CDIF630_02163 5'UTR and start of CDS.

Although further analyses are clearly necessary to functionally characterize this interaction (including the potential involvement of Hfq), these results highlight the value of integrating multiple datasets to identify high probability targets of a given sRNA. Interestingly, the nc085 target *eutV*, a positive regulator of ethanolamine utilization that we initially characterized in more detail, was not among the nc085 RIL-seq interactions [312,313]. However, as discussed above, this might simply be due to the chosen experimental conditions. Indeed, previously published data indicated that *eut* gene expression decreases during the stationary growth phase and without ethanolamine supplementation, potentially explaining its absence in the RIL-seq dataset [272,313].

Besides nc085 and its role in ethanolamine utilization, we have functionally characterized two additional *C. difficile* sRNAs in more detail: SpoY and SpoX [312,313]. Both sRNAs regulate translation of the master regulator of sporulation *spo0A* by binding to the *spo0A* 5'UTR and start of CDS [313]. However, once translated, Spo0A requires activation *via* phosphorylation to be able to transcriptionally regulate its target genes, raising the question of how post-transcriptional regulation of *spo0A* and Spo0A-P function are connected (see 1.5.4) [300]. Recently published work on post-transcriptional regulation of the response regulator o*mpR* in *E. coli* investigated a similar situation. In this case, the authors could show that levels of OmpR-P (the active form of OmpR) were insensitive to regulation of *ompR* translation by sRNAs [343]. However, while expression of several OmpR-P targets remained unaffected by OmrA/B-mediated regulation of *ompR*, others, including the sRNAs OmrA/B were sensitive to changes in OmpR levels [343]. According to the authors, either unphosphorylated OmpR or a heterodimer consisting of OmpR-OmpR-P might still be able to regulate transcription of the latter group [343]. Interestingly, analysis of functional residues of Spo0A indicated that loss of phosphorylation resulted in a reduced, but not total loss of sporulation in *C. difficile*, suggesting the existence of a similar mechanism [344]. Alternatively, sRNA-mediated regulation of *spo0A* might affect Spo0A as well as Spo0A-P levels, contrary to the mechanism described in *E. coli.* Although we did not analyse if SpoY and SpoX-mediated regulation of *spo0A* affects the pool of available Spo0A-P, we have shown that SpoY and SpoX-mediated post-transcriptional regulation of *spo0A* also affects transcript levels of Spo0A-P regulated genes. This indicates that sRNA-mediated regulation of *spo0A* translates into Spo0A-P function, contrary to sRNA-mediated regulation of *ompR* [313]. Either way, additional experiments are necessary to decipher when and how SpoY and SpoX-mediated regulation of *spo0A* translation affects levels of Spo0A-P in addition to Spo0A.

As discussed previously (2.4.4.6), deletion of SpoX resulted in significant downregulation of sporulation in *C. difficile* [313]. In contrast, SpoY deletion did not affect sporulation frequencies, although SpoY overexpression also resulted in downregulation of sporulation [313]. Similarly, nc085 deletion did not cause a significant phenotype, and only marginally affected growth in CDMM supplemented with ethanolamine [312]. Interestingly, SpoY and nc085 are no exception in this

regard, as sRNA deletions rarely result in a measurable phenotype [1,345,346]. Several hypotheses have been proposed that either individually or taken together might explain this phenomenon. The simplest answer is that sRNA-mediated regulation is only of limited importance in most cases, and rather serves to fine-tune bacterial adaptation. Since sRNA activity is often restricted to very specific conditions, it is also possible that the correct conditions were simply not tested when analysing a specific sRNA mutant. Another potential reason is a redundancy in function, as multiple sRNAs often target the same mRNA, creating so called regulatory hubs (see 1.2.4.4) [345,346]. In this case, the collective function of these sRNAs might be essential, while individual interactions are dispensable. Post-transcriptional regulation of *spo0A* might be one such example in *C. difficile* [313]. As indicated by our RIL-seq data, SpoX and SpoY were not the only sRNAs capable of binding the *spo0A* mRNA [313]. In fact, we identified four additional sRNA-*spo0A* interactions which were predicted to occur at the *spo0A* RBS region and thus potentially mimic the SpoY-mediated inhibition of *spo0A* translation. This suggests a functional redundancy of SpoY and thus potentially explains the lack of a SpoY deletion phenotype [313]. However, further characterization of these sRNAs is necessary to fully understand which factors coordinate their expression and function and thus post-transcriptional regulation of *spo0A*.

Figure 11: nc085, SpoY and SpoX promoter region. The nc085, SpoY and SpoX promoter regions were compared with published promoter motifs [307]. Potential matching motifs are highlighted. The transcriptional start sites (+1) as well as -10 and -35 positions are indicated.

As mentioned before, sRNA expression is generally known to be extensively regulated at the transcriptional level by various transcription factors (e.g. sigma factors or two component systems) in response to environmental conditions (see 1.2.3.1) [65,136,138,314]. Although a variety of transcription factors have been annotated in *C. difficile*, identification and analysis of regulated targets is primarily limited to coding genes and does not allow inference of sRNA regulation [303,307,347,348]. Identification of specific promoter motifs upstream of sRNAs using computational approaches is equally ineffective due to their scarcity and high variability. This is highlighted by our own only partially successful attempts at identifying transcription factors that regulate sRNA expression in *C. difficile*, including expression of nc085, SpoX, and SpoY (Figure 11) [307,312,313].

Experimental approaches utilizing defined transcription factor mutants or over-expression strains in addition to specific stress conditions might help to better understand which transcription factors and consequently environmental signals control sRNA expression in *C. difficile*.

3.5 Ribonucleases and their function in sRNA-mediated regulation

Although many sRNAs are transcribed from dedicated promoters (e.g. SpoY and SpoX), they can also be generated through processing of longer transcripts (see 1.2.3). Our dRNA-seq analysis identified multiple sRNAs that are derived from 5' or 3'UTRs *via* processing of a precursor RNA (e.g. nc009, nc087) [312]. Logically, these processed sRNAs are transcriptionally regulated in tandem with their respective precursor RNA, while simultaneously requiring a ribonuclease for processing. However, relevant ribonucleases that facilitate this process in *C. difficile* remain unknown. As mentioned previously, biogenesis and processing of sRNAs in Gram-negative species is predominantly performed by the single-strand specific endonuclease RNase E and the 3'→5 exonuclease PNPase (see 1.2.3.3) [51,79,80]. However, RNase E is absent in most Gram-positive species, including *C. difficile*, *S. aureus*, and *B. subtilis*, and publications investigating alternative mechanisms are still rare [82]. Generally, RNA degradation within most Gram-positive bacteria is based on initial endonucleolytic cleavage by either RNase Y or RNase III, wherein RNase Y functions as an RNase E orthologue. This is followed by 3'→5' and 5'→3' exonucleolytic cleavage mediated by PNPase and RNase R as well as RNase J1/J2 [82,349,350]. Accordingly, RNaseY has been reported to process the RoxS sRNA in *B. subtilis*, resulting in a shorter isoform with altered target-binding efficiencies [350–352]. Measurements of RNA-stability in an *S. aureus* RNase Y mutant also indicated a potential function in sRNA turnover [353]. Furthermore, the double-strand specific endonuclease RNase III was shown to mediated cleavage of several sRNAs in *S. aureus*, including the 3'UTR-derived sRNA RsaC and the 5'UTR-derived sRNA teg49 [83,354]. Indeed, RNase III seems to have a prominent function not only in sRNA biogenesis, but also in sRNA-mediated regulation of target mRNAs in *S. aureus* [355]. ArtR induced degradation of the *sarT* mRNA by RNase III upon ArtR binding to the *sarT* 5'UTR in *S. aureus*, is only one of several examples [356]. In fact, two recently published studies that performed RNase III-CLASH in *S. aureus*, identified hundreds of novel condition-specific sRNA-RNA interactions. This suggests that RNase III plays a much larger role in sRNA-mediated regulation in *S. aureus* than anticipated [327,357].

Based on these reports, it is likely that RNase Y and RNase III function in a similar fashion in *C. difficile,* by supporting sRNA biogenesis as well as target degradation upon sRNA binding. Indeed, the decrease in *spo0A* mRNA levels that we observed upon SpoY binding might be an example of sRNA induced degradation of a target mRNA [313]. Similarly, pulse-expression of nc085 negatively impacted *eutV* transcript levels, suggesting RNase-mediated mRNA degradation

following sRNA binding [312]. However, further studies are necessary to fully understand the contribution of each respective ribonucleases in sRNA biogenesis and sRNA-mediated regulation in *C. difficile*.

4 Conclusion & Outlook

For decades, proteins were considered the most prominent factors in regulating adaptation to changing environmental conditions in bacteria [11]. However, the past twenty years of research have challenged this paradigm to the point that sRNAs are emerging as a distinct class of important post-transcriptional regulators [1,358]. Nevertheless, there is a general lack of knowledge pertaining sRNA-mediated regulation in many Gram-positive species, despite the importance of sRNAs in mediating multiple physiological and virulence-related processes in Gram-negative species [1,358]. Furthermore, sRNA-mediated regulation in Gram-negative bacteria frequently depends on RNA chaperones, most notably Hfq, that facilitate base-paring of sRNAs to their mRNA targets [84]. However, the function and importance of these chaperones in Gram-positive bacteria, particularly Hfq, has long been a topic of debate. Accordingly, the goal of this thesis was to define the scope of sRNA-mediated regulation with an emphasis on virulence and colonization in *C. difficile*, and to determine the impact of Hfq in mediating these sRNA-target interactions.

In line with this goal, we successfully applied dRNA-seq and RIL-seq to identify a multitude of sRNAs expressed during various growth stages of *C. difficile* in rich media; three of which we functionally characterized in more detail [312,313]. While the sRNA nc085 is involved in ethanolamine uptake, an abundant intestinal nutrient that impacts *C. difficile* pathogenesis, SpoY and SpoX regulate sporulation initiation and thus an essential component of the CDI cycle [245,272,312,313]. Although we have barely scratched the surface of sRNA-mediated regulation in *C. difficile*, our RIL-seq dataset represents a starting point for the characterization of additional regulatory processes as highlighted by the following examples.

According to our RIL-seq results, SpoY and SpoX are only two of multiple sRNAs interacting not only with the *spo0A* 5'UTR, but also with each other. This points to a more complex network of sRNAs regulating this important sporulation factor in *C. difficile* [313]. Investigating the timing and manner of these interactions will be vital to fully comprehend the poorly understood process of sporulation initiation in this important pathogen.

In contrast to Hfq co-immunoprecipitation data from Gram-negative species, our RIP-seq, as well as our RIL-seq data pointed towards a considerable enrichment of mRNA 3'UTRs and riboswitches [312,313]. However, the relevance and consequence of these interactions in post-transcriptional regulation in *C. difficile* remains unknown.

RIL-seq analysis has previously proven successful in identifying various sRNA sponges in *E. coli* (RbsZ & PspH), *S. enterica* (OppX), and *V. cholerae* (QrrX) [50,146,322,338]. These results imply that RIL-seq data can be used to identify alternative sRNA classes such sRNA sponges and potentially dual-functioning sRNA in *C. difficile* (see 1.2.3.4 & 1.2.2.3). While

the former is indicated by multiple sRNA-sRNA interactions enriched by Hfq, the latter is implied by read profiles that resemble an sRNA, but overlap with very short CDSs in the *C. difficile* genome [313].

Taken together, these examples highlight the impressive breadh of post-transcriptional regulatory processes that can be investigated by applying RIL-seq analysis. Performing RIL-seq in stress and infection relevant conditions will most likely expand the scope of known sRNAs even further, while also aiding in identifying their regulatory function in a stress-related or infection context. In addition, further advanced techniques, such as dualRNA-seq or tripleRNA-seq can be applied to simultaneously decipher transcriptomic changes in *C. difficile* as well as the host and members of the microbiota [359–361]. Indeed, application of dualRNA-seq to various cell lines infected with *S. typhimurium* resulted in the identification of the PinT sRNA as an important regulator of several *S. typhimurium* virulence genes [359].

Besides the identification of various RNA-RNA interactions, our RIL-seq data also established the importance of Hfq as an RNA-chaperone in *C. difficile*. Contrary to reports on Hfq function in other Gram-positive bacteria, previously published data indicated enrichment of multiple sRNAs as well as a large impact on gene expression and bacterial physiology upon Hfq depletion [308,311]. Our RIP-seq and RIL-seq data further corroborated these results, as Hfq co-immunoprecipitation not only enriched various sRNAs and mRNAs, but also hundreds of RNA-RNA interactions, most of which comprised an sRNA [312,313]. Taken together with our rifampicin assays confirming destabilisation of multiple sRNAs in an Hfq knockout, these results suggest a vital role of the *C. difficile* Hfq in post-transcriptional regulation that mirrors its function in Gram-negative species [313]. Nevertheless, several questions regarding the Hfq mode-of-action in *C. difficile* remain open. Although most amino acids relevant for RNA binding and mediating RNA-RNA interactions in *E. coli* are conserved in the *C. difficile* Hfq, there are key differences. Most notably the amino acid composition of the poorly conserved C-terminal domain varies considerably between both species, indicating a difference in function. Measuring binding activities of various proximal, distal, rim and C-terminal Hfq mutants in *C. difficile* may provide a better understanding of these differences [186].

Interestingly, Hfq is not the only RBP associated with post-transcriptional regulation in *C. difficile*. Both the carbon storage regulator CsrA as well as the KH-domain containing protein KhpB also function as post-transcriptional regulators in *C. difficile* [158,362,363]. KhpB seems be of particular importance, as RIP-seq analysis reported hundreds of mRNA and various sRNA bound by KhpB [158]. Although there is an overlap in Hfq and KhpB enriched targets, KhpB predominantly binds CDSs [158]. Based on this observation, the authors speculated that unlike Hfq, KhpB has, at most, a marginal function in facilitating sRNA-target interactions [158]. Nevertheless, further experiments are necessary to fully decipher the mechanisms by which KhpB alone as well as in concert with

Hfq and other RBPs mediates post-transcriptional regulation in *C. difficile*. This also applies to the function of various ribonucleases in sRNA biogenesis and sRNA-mediated mRNA degradation. CLASH experiments in *S. aureus* have highlighted the importance of RNase III, in sRNA-mediated regulation in a Gram-positive species [327,357]. A similar approach might clarify how various nucleases integrate into post-transcriptional regulatory networks in *C. difficile*.

Taken together, this thesis was successful in establishing the function and importance of sRNAs and the RNA-chaperone Hfq in post-transcriptional regulation in *C. difficile*. In addition, this work represents a valuable foundation for future studies of post-transcriptional regulatory processes in this important human pathogen. Nevertheless, many significant questions remain open, and in-depth functional characterizations of sRNAs are still restricted to a few examples. For most sRNAs encoded by *C. difficile*, we do not currently understand their regulatory function or physiological relevance. A precise knowledge of these mechanisms might eventually help in developing targeted treatment options to combat CDI. In fact, several studies have already highlighted the potential of sRNAs, either as drug targets to reduce the fitness of a specific pathogen, or as a promising therapeutic option in the form of asRNAs that bind and downregulate specific mRNA targets [1,364]. The latter in particularly has gained increasing attention as a promising alternative to conventional antibiotics by selectively targeting species in a microbial community, an aspect that is of particular importance in combating CDI and recurrent CDI [364–367].

Bibliography

1. Westermann, A. J. Regulatory RNAs in virulence and host-microbe interactions. in *Regulating with RNA in Bacteria and Archaea* vol. 6 305–337 (American Society of Microbiology, 2019).

2. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628 (2009).

3. Storz, G. & Papenfort, K. *Regulating with RNA in Bacteria and Archaea*. (ASM Press, 2018).

4. Ariza-Mateos, A., Nuthanakanti, A. & Serganov, A. Riboswitch mechanisms: new tricks for an old dog. *Biochem.* **86**, 962–975 (2021).

5. Abduljalil, J. M. Bacterial riboswitches and RNA thermometers: Nature and contributions to pathogenesis. *Non-coding RNA Res.* **3**, 54–63 (2018).

6. Georg, J. & Hess, W. R. *cis*-antisense RNA, another level of gene regulation in bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 286–300 (2011).

7. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).

8. Wagner, E. G. H. & Romby, P. Small RNAs in bacteria and archaea. in *Advances in Genetics* vol. 90 133–208 (2015).

9. Holmqvist, E. & Vogel, J. RNA-binding proteins in bacteria. *Nat. Rev. Microbiol.* (2018) doi:10.1038/s41579-018-0049-5.

10. Mizuno, T., Chou, M. Y. & Inouye, M. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc. Natl. Acad. Sci.* **81**, 1966–1970 (1984).

11. Delihas, N. Discovery and characterization of the first non-coding RNA that regulates gene expression, *micF* RNA: A historical perspective. *World J. Biol. Chem.* **6**, 272 (2015).

12. Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L. & Storz, G. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* **90**, 43–53 (1997).

13. Bouché, F. & Bouché, J.-P. Genetic evidence that DicF, a second division inhibitor encoded by the *Escherichia coli dicB* operon, is probably RNA. *Mol. Microbiol.* **3**, 991–994 (1989).

14. Chen, S. *et al.* A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems* **65**, 157–177 (2002).

15. Vogel, J. *et al.* RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* **31**, 6435–43 (2003).

16. Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G. & Gottesman, S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* **15**, 1637–1651 (2001).

17. Rivas, E., Klein, R. J., Jones, T. A. & Eddy, S. R. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr. Biol.* **11**, 1369–1373 (2001).

18. Argaman, L. *et al.* Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.* **11**, 941–950 (2001).

19. Vogel, J. & Sharma, C. M. How to find small non-coding RNAs in bacteria. *Biol. Chem.* **386**, 1219–1238 (2005).

20. Zhang, A. *et al.* Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.* **50**, 1111–1124 (2003).

21. Montzka Wassarman, K. Small RNAs in *Escherichia coli*. *Trends Microbiol.* **7**, 37–45 (1999).

22. Barquist, L. & Vogel, J. Accelerating discovery and functional analysis of small RNAs with new technologies. *Annu. Rev. Genet.* **49**, 367–394 (2015).

23. Hör, J., Gorski, S. A. & Vogel, J. Bacterial RNA biology on a genome scale. *Mol. Cell* **70**, 785–799 (2018).

24. Lee, F. C. Y. & Ule, J. Advances in CLIP technologies for studies of protein-RNA interactions. *Mol. Cell* **69**, 354–369 (2018).

25. Andresen, L. & Holmqvist, E. CLIP-Seq in Bacteria: global recognition patterns of bacterial RNA-binding proteins. in *Methods in Enzymology* 127–145 (2018). doi:10.1016/bs.mie.2018.08.008.

26. Bohn, C. *et al.* Experimental discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of central metabolism. *Nucleic Acids Res.* **38**, 6620–6636 (2010).

27. Dugar, G. *et al.* High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genet.* **9**, e1003495 (2013).

28. Beauregard, A. *et al.* Identification and characterization of small RNAs in *Yersinia pestis*. *RNA Biol.* **10**, 397–405 (2013).

29. Cortes, T. *et al.* Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* **5**, 1121–1131 (2013).

30. Papenfort, K., Förstner, K. U., Cong, J.-P., Sharma, C. M. & Bassler, B. L. Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proc. Natl. Acad. Sci.* **112**, E766-75 (2015).

31. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**, 250–255 (2010).

32. Sharma, C. M. & Vogel, J. Differential RNA-seq: the approach behind and the biological insight gained. *Curr. Opin. Microbiol.* **19**, 97–105 (2014).

33. Adams, P. P. & Storz, G. Prevalence of small base-pairing RNAs derived from diverse genomic loci. *Biochim. Biophys. Acta - Gene Regul. Mech.* 194524 (2020) doi:10.1016/j.bbagrm.2020.194524.

34. Adams, P. P. *et al.* Regulatory roles of *Escherichia coli* 5' UTR and ORF-internal RNAs detected by 3' end mapping. *Elife* **10**, (2021).

35. Thomason, M. K. *et al.* A *rhlI* 5′ UTR-derived sRNA regulates RhlR-dependent quorum sensing in *Pseudomonas aeruginosa*. *MBio* **10**, (2019).

36. Acebo, P. *et al.* A small non-coding RNA modulates expression of pilus-1 type in *Streptococcus pneumoniae*. *Microorganisms* **9**, 1883 (2021).

37. Loh, E. *et al.* A *trans*-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell* **139**, 770–779 (2009).

38. DebRoy, S. *et al.* A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. *Science (80-.).* **345**, 937–940 (2014).

39. Mellin, J. R. *et al.* Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA. *Science (80-.).* **345**, 940–943 (2014).

40. Kawano, M. Detection of 5'- and 3'-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.* **33**, 1040–1050 (2005).

41. Ponath, F., Hör, J. & Vogel, J. An overview of gene regulation in bacteria by small RNAs derived from mRNA 3′ ends. *FEMS Microbiol. Rev.* **46**, (2022).

42. Miyakoshi, M., Chao, Y. & Vogel, J. Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Curr. Opin. Microbiol.* **24**, 132–139 (2015).

43. Menendez-Gil, P. & Toledo-Arana, A. Bacterial 3′UTRs: a useful resource in post-transcriptional regulation. *Front. Mol. Biosci.* **7**, 617633 (2021).

44. Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C. M. & Vogel, J. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).

45. Chao, Y. & Vogel, J. A 3′ UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. *Mol. Cell* **61**, 352–363 (2016).

46. Morita, T., Nishino, R. & Aiba, H. Role of the terminator hairpin in the biogenesis of functional Hfq-binding sRNAs. *RNA* **23**, 1419–1431 (2017).

47. Hör, J., Matera, G., Vogel, J., Gottesman, S. & Storz, G. *Trans*-acting small RNAs and their effects on gene expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus* **9**, (2020).

48. Aiso, T., Murata, M. & Gamou, S. Transcription of an antisense RNA of a *gadE* mRNA is regulated by GadE, the central activator of the acid resistance system in *Escherichia coli*. *Genes to Cells* **16**, 670–680 (2011).

49. Aiso, T., Kamiya, S., Yonezawa, H. & Gamou, S. Overexpression of an antisense RNA, ArrS, increases the acid resistance of *Escherichia coli*. *Microbiology* **160**, 954–961 (2014).

50. Melamed, S. *et al.* Global mapping of small RNA-target interactions in bacteria. *Mol. Cell* **63**, 884–897 (2016).

51. Dar, D. & Sorek, R. Bacterial noncoding RNAs excised from within protein-coding transcripts. *MBio* **9**, (2018).

52. Raina, M., King, A., Bianco, C. & Vanderpool, C. K. Dual-function RNAs. *Microbiol. Spectr.* **6**, (2018).

53. Gimpel, M. & Brantl, S. Dual-function small regulatory RNAs in bacteria. *Mol. Microbiol.* **103**, 387–397 (2017).

54. Vanderpool, C. K., Balasubramanian, D. & Lloyd, C. R. Dual-function RNA regulators in bacteria. *Biochimie* **93**, 1943–1949 (2011).

55. Heidrich, N., Chinali, A., Gerth, U. & Brantl, S. The small untranslated RNA SR1 from the *Bacillus subtilis* genome is involved in the regulation of arginine catabolism. *Mol. Microbiol.* **62**, 520–536 (2006).

56. Heidrich, N., Moll, I. & Brantl, S. *In vitro* analysis of the interaction between the small RNA SR1 and its primary target *ahrC* mRNA. *Nucleic Acids Res.* **35**, 4331–4346 (2007).

57. Gimpel, M., Heidrich, N., Mäder, U., Krügel, H. & Brantl, S. A dual-function sRNA from *B. subtilis*: SR1 acts as a peptide encoding mRNA on the *gapA* operon. *Mol. Microbiol.* **76**, 990–1009 (2010).

58. Helmann, J. D. Where to begin? Sigma factors and the selectivity of transcription initiation in bacteria. *Mol. Microbiol.* **112**, 335–347 (2019).

59. Fröhlich, K. S. & Gottesman, S. Small regulatory RNAs in the Enterobacterial response to envelope damage and oxidative stress. *Microbiol. Spectr.* **6**, (2018).

60. Charbonnier, M., González-Espinoza, G., Kehl-Fie, T. E. & Lalaouna, D. Battle for metals: regulatory RNAs at the front line. *Front. Cell. Infect. Microbiol.* **12**, 952948 (2022).

61. Johansen, J., Rasmussen, A. A., Overgaard, M. & Valentin-Hansen, P. Conserved small non-coding RNAs that belong to the σE regulon: role in down-regulation of outer membrane proteins. *J. Mol. Biol.* **364**, 1–8 (2006).

62. Mutalik, V. K., Nonaka, G., Ades, S. E., Rhodius, V. A. & Gross, C. A. Promoter strength properties of the complete sigma E regulon of *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.* **191**, 7279–7287 (2009).

63. Guo, M. S. *et al.* MicL, a new σ E -dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.* **28**, 1620–1634 (2014).

64. Gogol, E. B., Rhodius, V. A., Papenfort, K., Vogel, J. & Gross, C. A. Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc. Natl. Acad. Sci.* **108**, 12875–12880 (2011).

65. Brosse, A. & Guillier, M. Bacterial small RNAs in mixed regulatory networks. *Microbiol. Spectr.* **6**, (2018).

66. Nitzan, M., Rehani, R. & Margalit, H. Integration of bacterial small RNAs in regulatory networks. *Annu. Rev. Biophys.* **46**, 131–148 (2017).

67. Delihas, N. & Forst, S. MicF : an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *J. Mol. Biol.* **313**, 1–12 (2001).

68. Holmqvist, E., Unoson, C., Reimegård, J. & Wagner, E. G. H. A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. *Mol. Microbiol.* **84**, 414–427 (2012).

69. Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.* **85**, 319–347 (2016).

70. Chen, J., Morita, T. & Gottesman, S. Regulation of transcription termination of small RNAs and by small RNAs: molecular mechanisms and biological functions. *Front. Cell. Infect. Microbiol.* **9**, 201 (2019).

71. Otaka, H., Ishikawa, H., Morita, T. & Aiba, H. PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc. Natl. Acad. Sci.* **108**, 13059–13064 (2011).

72. Morita, T., Ueda, M., Kubo, K. & Aiba, H. Insights into transcription termination of Hfq-binding sRNAs of *Escherichia coli* and characterization of readthrough products. *RNA* **21**, 1490–1501 (2015).

73. Ami, V. K. G., Balasubramanian, R. & Hegde, S. R. Genome-wide identification of the context-dependent sRNA expression in *Mycobacterium tuberculosis*. *BMC Genomics* **21**, 167 (2020).

74. Mondal, S., Yakhnin, A. V., Sebastian, A., Albert, I. & Babitzke, P. NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat. Microbiol.* **1**, 15007 (2016).

75. Morita, T. *et al.* Identification of attenuators of transcriptional termination: implications for RNA regulation in *Escherichia coli*. *MBio* e0237122 (2022) doi:10.1128/mbio.02371-22.

76. Bandyra, K. J. *et al.* The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol. Cell* **47**, 943–953 (2012).

77. Schilder, A. & Görke, B. Role of the 5′ end phosphorylation state for small RNA stability and target RNA regulation in bacteria. *Nucleic Acids Res.* (2023) doi:10.1093/nar/gkad226.

78. Fröhlich, K. S., Haneke, K., Papenfort, K. & Vogel, J. The target spectrum of SdsR small RNA in *Salmonella*. *Nucleic Acids Res.* **44**, gkw632 (2016).

79. Chao, Y. *et al. In vivo* cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Mol. Cell* **65**, 39–51 (2017).

80. Hoyos, M., Huber, M., Förstner, K. U. & Papenfort, K. Gene autoregulation by 3' UTR-derived bacterial small RNAs. *Elife* **9**, (2020).

81. Condon, C. & Putzer, H. The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res.* **30**, 5339–5346 (2002).

82. Durand, S., Tomasini, A., Braun, F., Condon, C. & Romby, P. sRNA and mRNA turnover in Gram-positive bacteria. *FEMS Microbiol. Rev.* **39**, 316–330 (2015).

83. Lalaouna, D. *et al.* RsaC sRNA modulates the oxidative stress response of *Staphylococcus aureus* during manganese starvation. *Nucleic Acids Res.* **47**, 9871–9887 (2019).

84. Quendera, A. P. *et al.* RNA-binding proteins driving the regulatory activity of small non-coding RNAs in bacteria. *Front. Mol. Biosci.* **7**, 78 (2020).

85. Viegas, S. C. & Arraiano, C. M. Regulating the regulators: How ribonucleases dictate the rules in the control of small non-coding RNAs. *RNA Biol.* **5**, 230–243 (2008).

86. Schu, D. J., Zhang, A., Gottesman, S. & Storz, G. Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J.* **34**, 2557–2573 (2015).

87. Cameron, T. A., Matz, L. M. & De Lay, N. R. Polynucleotide phosphorylase: not merely an RNase but a pivotal post-transcriptional regulator. *PLOS Genet.* **14**, e1007654 (2018).

88. Afonyushkin, T., Vecerek, B., Moll, I., Bläsi, U. & Kaberdin, V. R. Both RNase E and RNase III control the stability of *sodB* mRNA upon translational inhibition by the small regulatory RNA RyhB. *Nucleic Acids Res.* **33**, 1678–89 (2005).

89. Massé, E., Escorcia, F. E. & Gottesman, S. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev.* **17**, 2374–2383 (2003).

90. Bandyra, K. J., Sinha, D., Syrjanen, J., Luisi, B. F. & De Lay, N. R. The ribonuclease polynucleotide phosphorylase can interact with small regulatory RNAs in both protective and degradative modes. *RNA* **22**, 360–372 (2016).

91. Andrade, J. M., Pobre, V., Matos, A. M. & Arraiano, C. M. The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *RNA* **18**, 844–855 (2012).

92. De Lay, N. & Gottesman, S. Role of polynucleotide phosphorylase in sRNA function in *Escherichia coli*. *RNA* **17**, 1172–1189 (2011).

93. Denham, E. L. The sponge RNAs of bacteria –how to find them and their role in regulating the post-transcriptional network. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1863**, 194565 (2020).

94. Durand, S. *et al.* Identification of an RNA sponge that controls the RoxS riboregulator of central metabolism in *Bacillus subtilis*. *Nucleic Acids Res.* **49**, 6399–6419 (2021).

95. Gorski, S. A., Vogel, J. & Doudna, J. A. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol.* **18**, 215–228 (2017).

96. Kawamoto, H., Koide, Y., Morita, T. & Aiba, H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol. Microbiol.* **61**, 1013–1022 (2006).

97. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43**, 880–891 (2011).

98. Papenfort, K., Bouvier, M., Mika, F., Sharma, C. M. & Vogel, J. Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc. Natl. Acad. Sci.* **107**, 20435–20440 (2010).

99. Rutherford, S. T., Valastyan, J. S., Taillefumier, T., Wingreen, N. S. & Bassler, B. L. Comprehensive analysis reveals how single nucleotides contribute to noncoding RNA function in bacterial quorum sensing. *Proc. Natl. Acad. Sci.* **112**, E6038–E6047 (2015).

100. Balbontín, R., Fiorini, F., Figueroa-Bossi, N., Casadesús, J. & Bossi, L. Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in *Salmonella enterica*. *Mol. Microbiol.* **78**, 380–394 (2010).

101. Beisel, C. L. & Storz, G. The base-pairing RNA Spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Mol. Cell* **41**, 286–297 (2011).

102. Faigenbaum-Romm, R. *et al.* Hierarchy in Hfq chaperon occupancy of small RNA targets plays a major role in their regulation. *Cell Rep.* **30**, 3127-3138.e6 (2020).

103. Coornaert, A., Chiaruttini, C., Springer, M. & Guillier, M. Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS Genet.* **9**, e1003156 (2013).

104. Sharma, C. M. *et al.* Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol. Microbiol.* **81**, 1144–1165 (2011).

105. Bouvier, M., Sharma, C. M., Mika, F., Nierhaus, K. H. & Vogel, J. Small RNA binding to 5′ mRNA coding region inhibits translational initiation. *Mol. Cell* **32**, 827–837 (2008).

106. Shine, J. & Dalgarno, L. The 3′-Terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci.* **71**, 1342–1346 (1974).

107. Altuvia, S., Zhang, A., Argaman, L., Tiwari, A. & Storz, G. The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J.* **17**, 6069–6075 (1998).

108. Sievers, S. *et al.* A multicopy sRNA of *Listeria monocytogenes* regulates expression of the virulence adhesin LapB. *Nucleic Acids Res.* **42**, 9383–9398 (2014).

109. Geissmann, T. *et al.* A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic Acids Res.* **37**, 7239–7257 (2009).

110. Dutta, T. & Srivastava, S. Small RNA-mediated regulation in bacteria: a growing palette of diverse mechanisms. *Gene* **656**, 60–72 (2018).

111. Deana, A. & Belasco, J. G. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev.* **19**, 2526–2533 (2005).

112. Ikeda, Y., Yagi, M., Morita, T. & Aiba, H. Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in *Escherichia coli*. *Mol. Microbiol.* **79**, 419–432 (2011).

113. Sinha, D. & De Lay, N. R. Target recognition by RNase E RNA-binding domain AR2 drives sRNA decay in the absence of PNPase. *Proc. Natl. Acad. Sci.* **119**, e2208022119 (2022).

114. Felden, B. & Augagneur, Y. Diversity and versatility in small RNA-mediated regulation in

bacterial pathogens. *Front. Microbiol.* **12**, 719977 (2021).

115. Fröhlich, K. S. & Papenfort, K. Regulation outside the box: new mechanisms for small RNAs. *Mol. Microbiol.* mmi.14523 (2020) doi:10.1111/mmi.14523.

116. Takahashi, S., Furusawa, H., Ueda, T. & Okahata, Y. Translation enhancer improves the ribosome liberation from translation initiation. *J. Am. Chem. Soc.* **135**, 13096–13106 (2013).

117. Sharma, C. M., Darfeuille, F., Plantinga, T. H. & Vogel, J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.* **21**, 2804–2817 (2007).

118. Azam, M. S. & Vanderpool, C. K. Translation inhibition from a distance: the small RNA SgrS silences a ribosomal protein S1-dependent enhancer. *Mol. Microbiol.* **114**, 391–408 (2020).

119. Rice, J. B., Balasubramanian, D. & Vanderpool, C. K. Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. *Proc. Natl. Acad. Sci.* **109**, E2691–E2698 (2012).

120. Fröhlich, K. S., Papenfort, K., Berger, A. A. & Vogel, J. A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. *Nucleic Acids Res.* **40**, 3623–3640 (2012).

121. Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C. D. & Vogel, J. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat. Struct. Mol. Biol.* **16**, 840–846 (2009).

122. Bronesky, D. *et al.* A multifaceted small RNA modulates gene expression upon glucose limitation in *Staphylococcus aureus*. *EMBO J.* **38**, (2019).

123. Ruiz de los Mozos, I. *et al.* Base pairing interaction between 5′- and 3′-UTRs controls *icaR* mRNA translation in *Staphylococcus aureus*. *PLoS Genet.* **9**, e1004001 (2013).

124. Papenfort, K. & Vanderpool, C. K. Target activation by regulatory RNAs in bacteria. *FEMS Microbiol. Rev.* **39**, 362–378 (2015).

125. Quereda, J. J., Ortega, Á. D., Pucciarelli, M. G. & García-del Portillo, F. The *Listeria* small RNA Rli27 regulates a cell wall protein inside eukaryotic cells by targeting a long 5′-UTR variant. *PLoS Genet.* **10**, e1004765 (2014).

126. Fröhlich, K. S., Papenfort, K., Fekete, A. & Vogel, J. A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J.* **32**, 2963–2979 (2013).

127. Carrier, M.-C., Lalaouna, D. & Massé, E. Hfq protein and GcvB small RNA tailoring of *oppA* target mRNA to levels allowing translation activation by MicF small RNA in *Escherichia coli*. *RNA Biol.* **20**, 59–76 (2023).

128. Papenfort, K., Sun, Y., Miyakoshi, M., Vanderpool, C. K. & Vogel, J. Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* **153**, 426–437 (2013).

129. Abdulla, S. Z. *et al.* Small RNAs activate *Salmonella* pathogenicity island 1 by modulating mRNA stability through the *hilD* mRNA 3′ untranslated region. *J. Bacteriol.* e0033322

(2022) doi:10.1128/jb.00333-22.

130. Bossi, L., Figueroa-Bossi, N., Bouloc, P. & Boudvillain, M. Regulatory interplay between small RNAs and transcription termination factor Rho. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1863**, 194546 (2020).

131. Hao, Z., Svetlov, V. & Nudler, E. Rho-dependent transcription termination: a revisionist view. *Transcription* **12**, 171–181 (2021).

132. Grylak-Mielnicka, A., Bidnenko, V., Bardowski, J. & Bidnenko, E. Transcription termination factor Rho: a hub linking diverse physiological processes in bacteria. *Microbiology* **162**, 433–447 (2016).

133. Bossi, L., Schwartz, A., Guillemardet, B., Boudvillain, M. & Figueroa-Bossi, N. A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes Dev.* **26**, 1864–1873 (2012).

134. Sedlyarova, N. *et al.* sRNA-mediated control of transcription termination in *E. coli*. *Cell* **167**, 111-121.e13 (2016).

135. Shimoni, Y. *et al.* Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol. Syst. Biol.* **3**, 138 (2007).

136. Mandin, P. & Guillier, M. Expanding control in bacteria: interplay between small RNAs and transcriptional regulators to control gene expression. *Curr. Opin. Microbiol.* **16**, 125–132 (2013).

137. Papenfort, K. & Melamed, S. Small RNAs, large networks: posttranscriptional regulons in gram-negative bacteria. *Annu. Rev. Microbiol.* **77**, (2023).

138. Piattelli, E., Peltier, J. & Soutourina, O. Interplay between regulatory RNAs and signal transduction systems during bacterial infection. *Genes (Basel).* **11**, 1209 (2020).

139. Hussein, R. & Lim, H. N. Direct comparison of small RNA and transcription factor signaling. *Nucleic Acids Res.* **40**, 7269–7279 (2012).

140. De Lay, N. & Gottesman, S. A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. *Mol. Microbiol.* **86**, 524–538 (2012).

141. Mika, F. & Hengge, R. Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. *RNA Biol.* **11**, 494–507 (2014).

142. Beisel, C. L. & Storz, G. Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol. Rev.* **34**, 866–882 (2010).

143. Guillier, M. & Gottesman, S. The 5′ end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Res.* **36**, 6781–6794 (2008).

144. Fei, J. *et al.* Determination of *in vivo* target search kinetics of regulatory noncoding RNA. *Science (80-.).* **347**, 1371–1374 (2015).

145. Bossi, L. & Figueroa-Bossi, N. Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. *Nat. Rev. Microbiol.* **14**, 775–784 (2016).

146. Melamed, S., Adams, P. P., Zhang, A., Zhang, H. & Storz, G. RNA-RNA interactomes of ProQ and Hfq reveal overlapping and competing roles. *Mol. Cell* (2019) doi:10.1016/j.molcel.2019.10.022.

147. Jørgensen, M. G., Pettersen, J. S. & Kallipolitis, B. H. sRNA-mediated control in bacteria: an increasing diversity of regulatory mechanisms. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1863**, 194504 (2020).

148. Romeo, T. & Babitzke, P. Global regulation by CsrA and its RNA antagonists. in *Regulating with RNA in Bacteria and Archaea* vol. 6 341–354 (American Society of Microbiology, 2019).

149. Romeo, T., Gong, M., Liu, M. Y. & Brun-Zinkernagel, A. M. Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J. Bacteriol.* **175**, 4744–4755 (1993).

150. Müller, P., Gimpel, M., Wildenhain, T. & Brantl, S. A new role for CsrA: promotion of complex formation between an sRNA and its mRNA target in *Bacillus subtilis*. *RNA Biol.* 1–16 (2019) doi:10.1080/15476286.2019.1605811.

151. Olejniczak, M. & Storz, G. ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers? *Mol. Microbiol.* **104**, 905–915 (2017).

152. Smirnov, A. *et al.* Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl. Acad. Sci.* **113**, 11591–11596 (2016).

153. Holmqvist, E., Berggren, S. & Rizvanovic, A. RNA-binding activity and regulatory functions of the emerging sRNA-binding protein ProQ. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1863**, 194596 (2020).

154. Holmqvist, E., Li, L., Bischler, T., Barquist, L. & Vogel, J. Global maps of ProQ binding *in vivo* reveal target recognition *via* RNA structure and stability control at mRNA 3' Ends. *Mol. Cell* **70**, 971-982.e6 (2018).

155. Westermann, A. J. *et al.* The major RNA-binding protein ProQ impacts virulence gene expression in *Salmonella enterica* serovar typhimurium. *MBio* **10**, (2019).

156. Smirnov, A., Wang, C., Drewry, L. L. & Vogel, J. Molecular mechanism of mRNA repression *in trans* by a ProQ-dependent small RNA. *EMBO J.* **36**, 1029–1045 (2017).

157. Zheng, J. J., Perez, A. J., Tsui, H.-C. T., Massidda, O. & Winkler, M. E. Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in *Streptococcus pneumoniae* D39. *Mol. Microbiol.* **106**, 793–814 (2017).

158. Lamm-Schmidt, V. *et al.* Grad-seq identifies KhpB as a global RNA-binding protein in *Clostridioides difficile* that regulates toxin production. *microLife* **2**, 1–21 (2021).

159. Olejniczak, M., Jiang, X., Basczok, M. M. & Storz, G. KH domain proteins: another family of bacterial RNA matchmakers? *Mol. Microbiol.* (2021) doi:10.1111/mmi.14842.

160. Franze de Fernandez, M. T., Eoyang, L. & August, J. T. Factor fraction required for the

synthesis of bacteriophage Qbeta-RNA. *Nature* **219**, 588–590 (1968).

161. Vogel, J. & Luisi, B. F. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 (2011).

162. Zhang, A. *et al.* The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J.* **17**, 6061–6068 (1998).

163. Sledjeski, D. D., Whitman, C. & Zhang, A. Hfq is necessary for regulation by the untranslated RNA DsrA. *J. Bacteriol.* **183**, 1997–2005 (2001).

164. Chao, Y. & Vogel, J. The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.* **13**, 24–33 (2010).

165. Sobrero, P. & Valverde, C. The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.* **38**, 276–299 (2012).

166. Sonnleitner, E. *et al.* Reduced virulence of a *hfq* mutant of *Pseudomonas aeruginosa* O1. *Microb. Pathog.* **35**, 217–228 (2003).

167. Kulesus, R. R., Diaz-Perez, K., Slechta, E. S., Eto, D. S. & Mulvey, M. A. Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. *Infect. Immun.* **76**, 3019–3026 (2008).

168. Ding, Y., Davis, B. M. & Waldor, M. K. Hfq is essential for *Vibrio cholerae* virulence and downregulates σE expression. *Mol. Microbiol.* **53**, 345–354 (2004).

169. Sittka, A., Pfeiffer, V., Tedin, K. & Vogel, J. The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol. Microbiol.* **63**, 193–217 (2007).

170. Katsuya-Gaviria, K., Paris, G., Dendooven, T. & Bandyra, K. J. Bacterial RNA chaperones and chaperone-like riboregulators: behind the scenes of RNA-mediated regulation of cellular metabolism. *RNA Biol.* **19**, 419–436 (2022).

171. Ponath, F. *et al.* RNA landscape of the emerging cancer-associated microbe *Fusobacterium nucleatum*. *Nat. Microbiol.* **6**, 1007–1020 (2021).

172. Michaux, C. *et al.* Single-nucleotide RNA maps for the two major nosocomial pathogens *Enterococcus faecalis* and *Enterococcus faecium*. *Front. Cell. Infect. Microbiol.* **10**, 600325 (2020).

173. Prezza, G. *et al.* Comparative genomics provides structural and functional insights into *Bacteroides* RNA biology. *Mol. Microbiol.* **117**, 67–85 (2022).

174. Christopoulou, N. & Granneman, S. The role of RNA-binding proteins in mediating adaptive responses in Gram-positive bacteria. *FEBS J.* febs.15810 (2021) doi:10.1111/febs.15810.

175. Rochat, T. *et al.* Tracking the elusive function of *Bacillus subtilis* Hfq. *PLoS One* **10**, e0124977 (2015).

176. Bohn, C., Rigoulay, C. & Bouloc, P. No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. *BMC Microbiol.* **7**, 10 (2007).

177. Møller, T. *et al.* Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol. Cell* **9**, 23–30 (2002).

178. Sun, X., Zhulin, I. & Wartell, R. M. Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res.* **30**, 3662–71 (2002).

179. Schumacher, M. A., Pearson, R. F., Møller, T., Valentin-Hansen, P. & Brennan, R. G. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.* **21**, 3546–56 (2002).

180. Wilusz, C. J. & Wilusz, J. Lsm proteins and Hfq: Life at the 3' end. *RNA Biol.* **10**, 592–601 (2013).

181. Weichenrieder, O. RNA binding by Hfq and ring-forming (L)Sm proteins: a trade-off between optimal sequence readout and RNA backbone conformation. *RNA Biol.* **11**, 537–549 (2014).

182. Santiago-Frangos, A. & Woodson, S. A. Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip. Rev. RNA* **9**, e1475 (2018).

183. Updegrove, T. B., Zhang, A. & Storz, G. Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.* **30**, 133–138 (2016).

184. Sauer, E. & Weichenrieder, O. Structural basis for RNA 3′-end recognition by Hfq. *Proc. Natl. Acad. Sci.* **108**, 13065–13070 (2011).

185. Mikulecky, P. J. *et al. Escherichia coli* Hfq has distinct interaction surfaces for DsrA, *rpoS* and poly(A) RNAs. *Nat. Struct. Mol. Biol.* **11**, 1206–1214 (2004).

186. Zhang, A., Schu, D. J., Tjaden, B. C., Storz, G. & Gottesman, S. Mutations in interaction surfaces differentially impact *E. coli* Hfq association with small RNAs and their mRNA targets. *J. Mol. Biol.* **425**, 3678–3697 (2013).

187. Link, T. M., Valentin-Hansen, P. & Brennan, R. G. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc. Natl. Acad. Sci.* **106**, 19292–19297 (2009).

188. Robinson, K. E., Orans, J., Kovach, A. R., Link, T. M. & Brennan, R. G. Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res.* **42**, 2736–2749 (2014).

189. Kovach, A. R., Hoff, K. E., Canty, J. T., Orans, J. & Brennan, R. G. Recognition of U-rich RNA by Hfq from the Gram-positive pathogen *Listeria monocytogenes*. *RNA* **20**, 1548–1559 (2014).

190. Horstmann, N., Orans, J., Valentin-Hansen, P., Shelburne, S. A. & Brennan, R. G. Structural mechanism of *Staphylococcus aureus* Hfq binding to an RNA A-tract. *Nucleic Acids Res.* **40**, 11023–11035 (2012).

191. Someya, T. *et al.* Crystal structure of Hfq from *Bacillus subtilis* in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Res.* **40**, 1856–1867 (2012).

192. Wroblewska, Z. & Olejniczak, M. Hfq assists small RNAs in binding to the coding sequence of *ompD* mRNA and in rearranging its structure. *RNA* **22**, 979–994 (2016).

193. Soper, T. J. & Woodson, S. A. The *rpoS* mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA* **14**, 1907–1917 (2008).

194. Peng, Y., Curtis, J. E., Fang, X. & Woodson, S. A. Structural model of an mRNA in complex with the bacterial chaperone Hfq. *Proc. Natl. Acad. Sci.* **111**, 17134–17139 (2014).

195. Sauer, E., Schmidt, S. & Weichenrieder, O. Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc. Natl. Acad. Sci.* **109**, 9396–9401 (2012).

196. Panja, S., Schu, D. J. & Woodson, S. A. Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic Acids Res.* **41**, 7536–7546 (2013).

197. Zheng, A., Panja, S. & Woodson, S. A. Arginine patch predicts the RNA annealing activity of Hfq from Gram-negative and Gram-positive bacteria. *J. Mol. Biol.* **428**, 2259–2264 (2016).

198. Santiago-Frangos, A., Jeliazkov, J. R., Gray, J. J. & Woodson, S. A. Acidic C-terminal domains autoregulate the RNA chaperone Hfq. *Elife* **6**, (2017).

199. Santiago-Frangos, A., Kavita, K., Schu, D. J., Gottesman, S. & Woodson, S. A. C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA. *Proc. Natl. Acad. Sci.* **113**, E6089–E6096 (2016).

200. Fender, A., Elf, J., Hampel, K., Zimmermann, B. & Wagner, E. G. H. RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev.* **24**, 2621–2626 (2010).

201. Kavita, K. *et al.* Multiple *in vivo* roles for the C-terminal domain of the RNA chaperone Hfq. *Nucleic Acids Res.* (2022) doi:10.1093/nar/gkac017.

202. Hussein, R. & Lim, H. N. Disruption of small RNA signaling caused by competition for Hfq. *Proc. Natl. Acad. Sci.* **108**, 1110–1115 (2011).

203. Moon, K. & Gottesman, S. Competition among Hfq-binding small RNAs in *Escherichia coli*. *Mol. Microbiol.* **82**, 1545–1562 (2011).

204. Morita, T. & Aiba, H. Mechanism and physiological significance of autoregulation of the *Escherichia coli* hfq gene. *RNA* **25**, 264–276 (2019).

205. Beisel, C. L., Updegrove, T. B., Janson, B. J. & Storz, G. Multiple factors dictate target selection by Hfq-binding small RNAs. *EMBO J.* **31**, 1961–1974 (2012).

206. Małecka, E. M., Stróżecka, J., Sobańska, D. & Olejniczak, M. Structure of bacterial regulatory RNAs determines their performance in competition for the chaperone protein Hfq. *Biochemistry* **54**, 1157–1170 (2015).

207. Wagner, E. G. H. Cycling of RNAs on Hfq. *RNA Biol.* **10**, 619–626 (2013).

208. Park, S. *et al.* Dynamic interactions between the RNA chaperone Hfq, small regulatory RNAs and mRNAs in live bacterial cells. *Elife* **10**, (2021).

209. Roca, J., Santiago-Frangos, A. & Woodson, S. A. Diversity of bacterial small RNAs drives competitive strategies for a mutual chaperone. *Nat. Commun.* **13**, 2449 (2022).

210. dos Santos, R. F., Arraiano, C. M. & Andrade, J. M. New molecular interactions broaden the functions of the RNA chaperone Hfq. *Curr. Genet.* (2019) doi:10.1007/s00294-019-00990-y.

211. Ellis, M. J., Trussler, R. S. & Haniford, D. B. Hfq binds directly to the ribosome-binding site of IS 10 transposase mRNA to inhibit translation. *Mol. Microbiol.* **96**, 633–650 (2015).

212. Sonnleitner, E. *et al.* Interplay between the catabolite repression control protein Crc, Hfq and RNA in Hfq-dependent translational regulation in *Pseudomonas aeruginosa*. *Nucleic Acids Res.* **46**, 1470–1485 (2018).

213. Chen, J. & Gottesman, S. Hfq links translation repression to stress-induced mutagenesis in *E. coli*. *Genes Dev.* **31**, 1382–1395 (2017).

214. Salvail, H., Caron, M.-P., Bélanger, J. & Massé, E. Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. *EMBO J.* **32**, 2764–2778 (2013).

215. Vecerek, B., Moll, I. & Bläsi, U. Translational autocontrol of the *Escherichia coli hfq* RNA chaperone gene. *RNA* **11**, 976–84 (2005).

216. Andrade, J. M., dos Santos, R. F., Chelysheva, I., Ignatova, Z. & Arraiano, C. M. The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity. *EMBO J.* **37**, e97631 (2018).

217. dos Santos, R. F., Andrade, J. M., Pissarra, J., Deutscher, M. P. & Arraiano, C. M. Hfq and RNase R mediate rRNA processing and degradation in a novel RNA quality control process. *MBio* **11**, (2020).

218. Lee, T. & Feig, A. L. The RNA binding protein Hfq interacts specifically with tRNAs. *RNA* **14**, 514–523 (2008).

219. Malabirade, A. *et al.* Revised role for Hfq bacterial regulator on DNA topology. *Sci. Rep.* **8**, 16792 (2018).

220. Morita, T., Maki, K. & Aiba, H. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.* **19**, 2176–2186 (2005).

221. Rabhi, M. *et al.* The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J.* **30**, 2805–2816 (2011).

222. Pichon, C. & Felden, B. Small RNA gene identification and mRNA target predictions in bacteria. *Bioinformatics* **24**, 2807–2813 (2008).

223. Lalaouna, D., Prévost, K., Eyraud, A. & Massé, E. Identification of unknown RNA partners using MAPS. *Methods* **117**, 28–34 (2017).

224. Han, K., Tjaden, B. & Lory, S. GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by *in vivo* proximity ligation. *Nat. Microbiol.* **2**, 16239 (2017).

225. Sittka, A. *et al.* Deep sequencing analysis of small noncoding RNA and mRNA targets of the

global post-transcriptional regulator, Hfq. *PLoS Genet.* **4**, e1000163 (2008).

226. Waters, S. A. *et al.* Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *EMBO J.* **36**, 374–387 (2017).

227. Zhang, Y.-F. *et al.* Probing the sRNA regulatory landscape of *P. aeruginosa* : post-transcriptional control of determinants of pathogenicity and antibiotic susceptibility. *Mol. Microbiol.* **106**, 919–937 (2017).

228. Liu, T. *et al.* Detecting RNA-RNA interactions in *E. coli* using a modified CLASH method. *BMC Genomics* **18**, 343 (2017).

229. Georg, J. *et al.* The power of cooperation: experimental and computational approaches in the functional characterization of bacterial sRNAs. *Mol. Microbiol.* **113**, 603–612 (2020).

230. Centers for Disease Control and Prevention. *Antibiotic resistance threats in the United States, 2019*. *CDC* https://stacks.cdc.gov/view/cdc/82532 (2019) doi:10.15620/cdc:82532.

231. European Centre for Disease Prevention and Control. Healthcare-associated infections: *Clostridium difficile* infections. *ECDC* (2018).

232. Bartlett, J. G., Chang, T. W., Gurwith, M., Gorbach, S. L. & Onderdonk, A. B. Antibiotic-associated pseudomembranous colitis due to toxin-producing *Clostridia*. *N. Engl. J. Med.* **298**, 531–534 (1978).

233. Brestrich, G. *et al.* Epidemiology of *Clostridioides difficile* infections in Germany, 2010–2019: a review from four public databases. *Infect. Dis. Ther.* 1–16 (2023) doi:10.1007/s40121-023-00785-2.

234. Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. *Clostridium difficile* infection. *Nat. Rev. Dis. Prim.* **2**, 16020 (2016).

235. Paredes-Sabja, D., Shen, A. & Sorg, J. A. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends Microbiol.* **22**, 406–416 (2014).

236. Fu, Y., Luo, Y. & Grinspan, A. M. Epidemiology of community-acquired and recurrent *Clostridioides difficile* infection. *Therap. Adv. Gastroenterol.* **14**, 175628482110162 (2021).

237. Aguirre, A. M. & Sorg, J. A. Gut associated metabolites and their roles in *Clostridioides difficile* pathogenesis. *Gut Microbes* **14**, 2094672 (2022).

238. De Roo, A. C. & Regenbogen, S. E. *Clostridium difficile* infection: an epidemiology update. *Clin. Colon Rectal Surg.* **33**, 049–057 (2020).

239. Theriot, C. M., Bowman, A. A. & Young, V. B. Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for *Clostridium difficile* spore germination and outgrowth in the large intestine. *mSphere* **1**, (2016).

240. Clancy, C. J., Buehrle, D., Vu, M., Wagener, M. M. & Nguyen, M. H. Impact of revised Infectious Diseases Society of America and Society for Healthcare Epidemiology of America clinical practice guidelines on the treatment of *Clostridium difficile* infections in the United States. *Clin. Infect. Dis.* **72**, 1944–1949 (2021).

241. Romero-Rodríguez, A., Martínez de la Peña, C., Troncoso-Cotal, S., Guzmán, C. & Sánchez, S. Emerging alternatives against *Clostridioides difficile* infection. *Anaerobe* **78**, 102638 (2022).

242. Chilton, C. H., Pickering, D. S. & Freeman, J. Microbiologic factors affecting *Clostridium difficile* recurrence. *Clin. Microbiol. Infect.* **24**, 476–482 (2018).

243. Dsouza, M. *et al.* Colonization of the live biotherapeutic product VE303 and modulation of the microbiota and metabolites in healthy volunteers. *Cell Host Microbe* **30**, 583-598.e8 (2022).

244. Seekatz, A. M., Safdar, N. & Khanna, S. The role of the gut microbiome in colonization resistance and recurrent *Clostridioides difficile* infection. *Therap. Adv. Gastroenterol.* **15**, 175628482211343 (2022).

245. Buddle, J. E. & Fagan, R. P. Pathogenicity and virulence of *Clostridioides difficile*. *Virulence* (2022) doi:10.1080/21505594.2022.2150452.

246. Walter, J. & Shanahan, F. Fecal microbiota-based treatment for recurrent *Clostridioides difficile* infection. *Cell* **186**, 1087 (2023).

247. Nibbering, B., Gerding, D. N., Kuijper, E. J., Zwittink, R. D. & Smits, W. K. Host immune responses to *Clostridioides difficile*: toxins and beyond. *Front. Microbiol.* **12**, 804949 (2021).

248. Van der Waaij, D., Berghuis-de Vries, J. M. & Lekkerkerk-van der Wees, J. E. C. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. *J. Hyg. (Lond).* **69**, 405–411 (1971).

249. Bohnhoff, M. & Miller, C. P. Enhanced susceptibility to *Salmonella* infection in streptomycin-treated mice. *J. Infect. Dis.* **111**, 117–127 (1962).

250. Caballero-Flores, G., Pickard, J. M. & Núñez, G. Microbiota-mediated colonization resistance: mechanisms and regulation. *Nat. Rev. Microbiol.* (2022) doi:10.1038/s41579-022-00833-7.

251. Collins, S. L., Stine, J. G., Bisanz, J. E., Okafor, C. D. & Patterson, A. D. Bile acids and the gut microbiota: metabolic interactions and impacts on disease. *Nat. Rev. Microbiol.* **21**, 236–247 (2023).

252. Shulpekova, Y. *et al.* A recent ten-year perspective: bile acid metabolism and signaling. *Molecules* **27**, 1983 (2022).

253. Quinn, R. A. *et al.* Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* **579**, 123–129 (2020).

254. Lucas, L. N. *et al.* Dominant bacterial phyla from the human gut show widespread ability to transform and conjugate bile acids. *mSystems* **6**, e0080521 (2021).

255. Sorg, J. A. & Sonenshein, A. L. Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination. *J. Bacteriol.* **191**, 1115–1117 (2009).

256. Foley, M. H. *et al.* Bile salt hydrolases shape the bile acid landscape and restrict *Clostridioides difficile* growth in the murine gut. *Nat. Microbiol.* (2023)

doi:10.1038/s41564-023-01337-7.

257. Sorg, J. A. & Sonenshein, A. L. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J. Bacteriol.* **190**, 2505–2512 (2008).

258. Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nat. Commun.* **5**, 3114 (2014).

259. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517**, 205–208 (2015).

260. Giel, J. L., Sorg, J. A., Sonenshein, A. L. & Zhu, J. Metabolism of bile salts in mice influences spore germination in *Clostridium difficile*. *PLoS One* **5**, e8740 (2010).

261. Ouyang, Z.-R., Niu, X.-R., Wang, W.-G. & Zhao, J.-H. The role of short-chain fatty acids in *Clostridioides difficile* infection: A review. *Anaerobe* **75**, 102585 (2022).

262. Hryckowian, A. J. *et al.* Microbiota-accessible carbohydrates suppress *Clostridium difficile* infection in a murine model. *Nat. Microbiol.* **3**, 662–669 (2018).

263. Fachi, J. L. *et al.* Butyrate protects mice from *Clostridium difficile*-induced colitis through an HIF-1-dependent mechanism. *Cell Rep.* **27**, 750-761.e7 (2019).

264. Fachi, J. L. *et al.* Acetate coordinates neutrophil and ILC3 responses against *C. difficile* through FFAR2. *J. Exp. Med.* **217**, (2020).

265. Marshall, A., McGrath, J. W., Graham, R. & McMullan, G. Food for thought - the link between *Clostridioides difficile* metabolism and pathogenesis. *PLOS Pathog.* **19**, e1011034 (2023).

266. Fletcher, J. R., Erwin, S., Lanzas, C. & Theriot, C. M. Shifts in the gut metabolome and *Clostridium difficile* transcriptome throughout colonization and infection in a mouse model. *mSphere* **3**, (2018).

267. Gomez, S. Y., Patel, J. & Lopez, C. A. What's metal got to do with it? Transition metals in *Clostridioides difficile* infection. *Curr. Opin. Microbiol.* **65**, 116–122 (2022).

268. Aguirre, A. M. *et al.* Bile acid-independent protection against *Clostridioides difficile* infection. *PLOS Pathog.* **17**, e1010015 (2021).

269. Girinathan, B. P. *et al. In vivo* commensal control of *Clostridioides difficile* virulence. *Cell Host Microbe* (2021) doi:10.1016/j.chom.2021.09.007.

270. Rupnik, M. & Janezic, S. An update on *Clostridium difficile* toxinotyping. *J. Clin. Microbiol.* **54**, 13–18 (2016).

271. Gencic, S. & Grahame, D. A. Diverse energy-conserving pathways in *Clostridium difficile*: growth in the absence of amino acid stickland acceptors and the role of the Wood-Ljungdahl pathway. *J. Bacteriol.* **202**, (2020).

272. Nawrocki, K. L., Wetzel, D., Jones, J. B., Woods, E. C. & McBride, S. M. Ethanolamine is a valuable nutrient source that impacts *Clostridium difficile* pathogenesis. *Environ. Microbiol.* **20**, 1419–1435 (2018).

273. Pruss, K. M. & Sonnenburg, J. L. *C. difficile* exploits a host metabolite produced during toxin-mediated disease. *Nature* **593**, 261–265 (2021).

274. Hasan, M. K., Dhungel, B. A. & Govind, R. Characterization of an operon required for growth on cellobiose in *Clostridioides difficile*. *Microbiology* **167**, (2021).

275. Engevik, M. A. *et al.* Mucin-degrading microbes release monosaccharides that chemoattract *Clostridioides difficile* and facilitate colonization of the human intestinal mucus layer. *ACS Infect. Dis.* **7**, 1126–1142 (2021).

276. Collins, J. *et al.* Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature* **553**, 291–294 (2018).

277. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–99 (2013).

278. Knippel, R. J. *et al. Clostridioides difficile* senses and hijacks host heme for incorporation into an oxidative stress defense system. *Cell Host Microbe* (2020) doi:10.1016/j.chom.2020.05.015.

279. Knippel, R. J. *et al.* Heme sensing and detoxification by HatRT contributes to pathogenesis during *Clostridium difficile* infection. *PLOS Pathog.* **14**, e1007486 (2018).

280. Lesniak, N. A. *et al.* The gut bacterial community potentiates *Clostridioides difficile* infection severity. *MBio* e0118322 (2022) doi:10.1128/mbio.01183-22.

281. Smith, A. B. *et al. Enterococci* enhance *Clostridioides difficile* pathogenesis. *Nature* (2022) doi:10.1038/s41586-022-05438-x.

282. Mengoli, M. *et al.* Make it less difficile: understanding genetic evolution and global spread of *Clostridioides difficile*. *Genes (Basel).* **13**, 2200 (2022).

283. Clements, A. C., Magalhães, R. J. S., Tatem, A. J., Paterson, D. L. & Riley, T. V. *Clostridium difficile* PCR ribotype 027: assessing the risks of further worldwide spread. *Lancet Infect. Dis.* **10**, 395–404 (2010).

284. Knight, D. R., Elliott, B., Chang, B. J., Perkins, T. T. & Riley, T. V. Diversity and evolution in the genome of *Clostridium difficile*. *Clin. Microbiol. Rev.* **28**, 721–741 (2015).

285. Kuehne, S. A. *et al.* The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* **467**, 711–713 (2010).

286. Kordus, S. L., Thomas, A. K. & Lacy, D. B. *Clostridioides difficile* toxins: mechanisms of action and antitoxin therapeutics. *Nat. Rev. Microbiol.* (2021) doi:10.1038/s41579-021-00660-2.

287. Majumdar, A. & Govind, R. Regulation of *Clostridioides difficile* toxin production. *Curr. Opin. Microbiol.* **65**, 95–100 (2022).

288. Martínez-Meléndez, A., Cruz-López, F., Morfin-Otero, R., Maldonado-Garza, H. J. & Garza-González, E. An update on *Clostridioides difficile* binary toxin. *Toxins (Basel).* **14**, 305 (2022).

289. Imwattana, K., Rodríguez, C., Riley, T. V & Knight, D. R. A species-wide genetic atlas of antimicrobial resistance in *Clostridioides difficile*. *Microb. Genomics* **7**, (2021).

290. He, M. *et al.* Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl. Acad. Sci.* **107**, 7527–7532 (2010).

291. Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**, 779–786 (2006).

292. Peng, Z. *et al.* Update on antimicrobial resistance in *Clostridium difficile*: resistance mechanisms and antimicrobial susceptibility testing. *J. Clin. Microbiol.* **55**, 1998–2008 (2017).

293. Frost, L. R., Cheng, J. K. J. & Unnikrishnan, M. *Clostridioides difficile* biofilms: a mechanism of persistence in the gut? *PLOS Pathog.* **17**, e1009348 (2021).

294. Taggart, M. G. *et al.* Biofilm regulation in *Clostridioides difficile*: novel systems linked to hypervirulence. *PLOS Pathog.* **17**, e1009817 (2021).

295. Deakin, L. J. *et al.* The *Clostridium difficile spo0A* gene is a persistence and transmission factor. *Infect. Immun.* **80**, 2704–2711 (2012).

296. Castro-Córdova, P. *et al.* Entry of spores into intestinal epithelial cells contributes to recurrence of *Clostridioides difficile* infection. *Nat. Commun.* **12**, 1140 (2021).

297. Rosenbusch, K. E., Bakker, D., Kuijper, E. J. & Smits, W. K. *C. difficile* 630Δ*erm* Spo0A regulates sporulation, but does not contribute to toxin production, by direct high-affinity binding to target DNA. *PLoS One* **7**, e48608 (2012).

298. Stephenson, K. & Hoch, J. A. Evolution of signalling in the sporulation phosphorelay. *Mol. Microbiol.* **46**, 297–304 (2002).

299. Lee, C. D. *et al.* Genetic mechanisms governing sporulation initiation in *Clostridioides difficile*. *Curr. Opin. Microbiol.* **66**, 32–38 (2022).

300. Shen, A., Edwards, A. N., Sarker, M. R. & Paredes-Sabja, D. Sporulation and germination in Clostridial pathogens. in *Gram-Positive Pathogens, Third Edition* vol. 7 903–926 (American Society of Microbiology, 2019).

301. Zeng, J., Wang, H., Dong, M. & Tian, G.-B. *Clostridioides difficile* spore: coat assembly and formation. *Emerg. Microbes Infect.* 1–28 (2022) doi:10.1080/22221751.2022.2119168.

302. Pereira, F. C. *et al.* The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. *PLoS Genet.* **9**, e1003782 (2013).

303. Saujet, L. *et al.* Genome-wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. *PLoS Genet.* **9**, e1003756 (2013).

304. Baloh, M. & Sorg, J. A. *Clostridioides difficile* spore germination: initiation to DPA release. *Curr. Opin. Microbiol.* **65**, 101–107 (2022).

305. Chen, Y., Indurthi, D. C., Jones, S. W. & Papoutsakis, E. T. Small RNAs in the genus clostridium. *MBio* **2**, e00340-10 (2011).

306. Soutourina, O. A. *et al.* Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genet.* **9**, e1003493 (2013).

307. Soutourina, O. *et al.* Genome-wide transcription start site mapping and promoter assignments to a sigma factor in the human enteropathogen *Clostridioides difficile*. *Front. Microbiol.* **11**, 1939 (2020).

308. Boudry, P. *et al.* Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen *Clostridium difficile*. *J. Bacteriol.* **196**, 3234–3248 (2014).

309. Maikova, A., Kreis, V., Boutserin, A., Severinov, K. & Soutourina, O. Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. *Appl. Environ. Microbiol.* **85**, (2019).

310. Caillet, J., Gracia, C., Fontaine, F. & Hajnsdorf, E. *Clostridium difficile* Hfq can replace *Escherichia coli* Hfq for most of its function. *RNA* **20**, 1567–1578 (2014).

311. Boudry, P. *et al.* Identification of RNAs bound by Hfq reveals widespread RNA partners and a sporulation regulator in the human pathogen *Clostridioides difficile*. *RNA Biol.* 1–22 (2021) doi:10.1080/15476286.2021.1882180.

312. Fuchs, M. *et al.* An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium. *Proc. Natl. Acad. Sci.* **118**, (2021).

313. Fuchs, M. *et al.* A network of small RNAs regulates sporulation initiation in *Clostridioides difficile*. *EMBO J.* 2022.10.17.512509 (2023) doi:10.15252/embj.2022112858.

314. Pitman, S. & Cho, K. The mechanisms of virulence regulation by small noncoding RNAs in low GC Gram-positive pathogens. *Int. J. Mol. Sci.* **16**, 29797–29814 (2015).

315. Santoro, S. *et al.* Multifaceted interplay between Hfq and the small RNA GssA in *Pseudomonas aeruginosa*. *MBio* **14**, e0241822 (2023).

316. Trouillon, J., Han, K., Attrée, I. & Lory, S. The core and accessory Hfq interactomes across *Pseudomonas aeruginosa* lineages. *Nat. Commun.* **13**, 1258 (2022).

317. Sarni, S. H. *et al.* Intrinsically disordered interaction network in an RNA chaperone revealed by native mass spectrometry. *Proc. Natl. Acad. Sci.* **119**, e2208780119 (2022).

318. Goldberger, O., Szoke, T., Nussbaum-Shochat, A. & Amster-Choder, O. Heterotypic phase separation of Hfq is linked to its roles as an RNA chaperone. *Cell Rep.* **41**, 111881 (2022).

319. Beaufay, F. *et al.* Polyphosphate drives bacterial heterochromatin formation. *Sci. Adv.* **7**, eabk0233 (2021).

320. Fortas, E. *et al.* New insight into the structure and function of Hfq C-terminus. *Biosci. Rep.* **35**, 1–9 (2015).

321. Malabirade, A. *et al.* Membrane association of the bacterial riboregulator Hfq and functional perspectives. *Sci. Rep.* **7**, 10724 (2017).

322. Matera, G. *et al.* Global RNA interactome of *Salmonella* discovers a 5′ UTR sponge for the MicF small RNA that connects membrane permeability to transport capacity. *Mol. Cell* (2022) doi:10.1016/j.molcel.2021.12.030.

323. Holmqvist, E. *et al.* Global RNA recognition patterns of post-transcriptional regulators Hfq

and CsrA revealed by UV crosslinking *in vivo*. *EMBO J.* **35**, 991–1011 (2016).

324. Dar, D. & Sorek, R. High-resolution RNA 3′-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res.* **46**, 6797–6805 (2018).

325. Toledo-Arana, A. *et al.* The Listeria transcriptional landscape from saprophytism to virulence. *Nature* **459**, 950–956 (2009).

326. Ren, G.-X., Guo, X.-P. & Sun, Y.-C. Regulatory 3′ untranslated regions of bacterial mRNAs. *Front. Microbiol.* **8**, 1276 (2017).

327. Mediati, D. G. *et al.* RNase III-CLASH of multi-drug resistant *Staphylococcus aureus* reveals a regulatory mRNA 3′UTR required for intermediate vancomycin resistance. *Nat. Commun.* **13**, 3558 (2022).

328. Braun, F., Durand, S. & Condon, C. Initiating ribosomes and a 5′/3′-UTR interaction control ribonuclease action to tightly couple *B. subtilis* hbs mRNA stability with translation. *Nucleic Acids Res.* **45**, 11386–11400 (2017).

329. Zhu, H., Mao, X.-J., Guo, X.-P. & Sun, Y.-C. The *hmsT* 3′ untranslated region mediates c-di-GMP metabolism and biofilm formation in *Yersinia pestis*. *Mol. Microbiol.* **99**, 1167–1178 (2016).

330. López-Garrido, J., Puerta-Fernández, E. & Casadesús, J. A eukaryotic-like 3′ untranslated region in *Salmonella enterica hilD* mRNA. *Nucleic Acids Res.* **42**, 5894–5906 (2014).

331. Duval, M., Simonetti, A., Caldelari, I. & Marzi, S. Multiple ways to regulate translation initiation in bacteria: mechanisms, regulatory circuits, dynamics. *Biochimie* **114**, 18–29 (2015).

332. Chiaruttini, C. & Guillier, M. On the role of mRNA secondary structure in bacterial translation. *WIREs RNA* **11**, e1579 (2020).

333. Dambach, M., Irnov, I. & Winkler, W. C. Association of RNAs with *Bacillus subtilis* Hfq. *PLoS One* **8**, e55156 (2013).

334. De Lay, N. R. & Garsin, D. A. The unmasking of 'junk' RNA reveals novel sRNAs: from processed RNA fragments to marooned riboswitches. *Curr. Opin. Microbiol.* **30**, 16–21 (2016).

335. Mellin, J. R. & Cossart, P. Unexpected versatility in bacterial riboswitches. *Trends Genet.* **31**, 150–156 (2015).

336. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995–1011 (2017).

337. Bar, A., Argaman, L., Altuvia, Y. & Margalit, H. Prediction of novel bacterial small RNAs from RIL-seq RNA–RNA interaction data. *Front. Microbiol.* **12**, 635070 (2021).

338. Huber, M. *et al.* An RNA sponge controls quorum sensing dynamics and biofilm formation in *Vibrio cholerae*. *Nat. Commun.* **13**, 7585 (2022).

339. Popitsch, N., Bilusic, I., Rescheneder, P., Schroeder, R. & Lybecker, M. Temperature-

dependent sRNA transcriptome of the Lyme disease spirochete. *BMC Genomics* **18**, 28 (2017).

340. Arrieta-Ortiz, M. L. *et al.* Inference of bacterial small RNA regulatory networks and integration with transcription factor-driven regulatory networks. *mSystems* **5**, (2020).

341. Wright, P. R. *et al.* CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res.* **42**, W119–W123 (2014).

342. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.* **45**, W435–W439 (2017).

343. Brosse, A., Korobeinikova, A., Gottesman, S. & Guillier, M. Unexpected properties of sRNA promoters allow feedback control *via* regulation of a two-component system. *Nucleic Acids Res.* **44**, gkw642 (2016).

344. DiCandia, M. A. *et al.* Identification of functional Spo0A residues critical for sporulation in *Clostridioides difficile*. *J. Mol. Biol.* **434**, 167641 (2022).

345. Goldberger, O., Livny, J., Bhattacharyya, R. & Amster-Choder, O. Wisdom of the crowds: a suggested polygenic plan for small-RNA-mediated regulation in bacteria. *iScience* **24**, 103096 (2021).

346. Rachwalski, K., Ellis, M. J., Tong, M. & Brown, E. D. Synthetic genetic interactions reveal a dense and cryptic regulatory network of small noncoding RNAs in *Escherichia coli*. *MBio* **13**, e0122522 (2022).

347. Chandra, H., Sorg, J. A., Hassett, D. J. & Sun, X. Regulatory transcription factors of *Clostridioides difficile* pathogenesis with a focus on toxin regulation. *Crit. Rev. Microbiol.* 1–16 (2022) doi:10.1080/1040841X.2022.2054307.

348. Kint, N. *et al.* The alternative sigma factor σ B plays a crucial role in adaptive strategies of *Clostridium difficile* during gut infection. *Environ. Microbiol.* **19**, 1933–1958 (2017).

349. Mediati, D. G., Lalaouna, D. & Tree, J. J. Burning the candle at both ends: have exoribonucleases driven divergence of regulatory RNA mechanisms in bacteria? *MBio* **12**, e0104121 (2021).

350. Bechhofer, D. H. & Deutscher, M. P. Bacterial ribonucleases and their roles in RNA metabolism. *Crit. Rev. Biochem. Mol. Biol.* **54**, 242–300 (2019).

351. Durand, S. *et al.* A nitric oxide regulated small RNA controls expression of genes involved in redox homeostasis in *Bacillus subtilis*. *PLOS Genet.* **11**, e1004957 (2015).

352. Broglia, L. *et al.* An RNA-seq based comparative approach reveals the transcriptome-wide interplay between 3′-to-5′ exoRNases and RNase Y. *Nat. Commun.* **11**, 1587 (2020).

353. Marincola, G. *et al.* RNase Y of *Staphylococcus aureus* and its role in the activation of virulence genes. *Mol. Microbiol.* **85**, 817–832 (2012).

354. Manna, A. C. *et al.* Small RNA teg49 is derived from a *sarA* transcript and regulates virulence genes independent of SarA in *Staphylococcus aureus*. *Infect. Immun.* **86**, (2018).

355. Lioliou, E. *et al. In vivo* mapping of RNA–RNA interactions in *Staphylococcus aureus* using the endoribonuclease III. *Methods* **63**, 135–143 (2013).

356. Xue, T., Zhang, X., Sun, H. & Sun, B. ArtR, a novel sRNA of *Staphylococcus aureus*, regulates α-toxin expression by targeting the 5′ UTR of *sarT* mRNA. *Med. Microbiol. Immunol.* **203**, 1–12 (2014).

357. McKellar, S. W. *et al.* RNase III CLASH in MRSA uncovers sRNA regulatory networks coupling metabolism to toxin expression. *Nat. Commun.* **13**, 3560 (2022).

358. Djapgne, L. & Oglesby, A. G. Impacts of small RNAs and their chaperones on bacterial pathogenicity. *Front. Cell. Infect. Microbiol.* **11**, 604511 (2021).

359. Westermann, A. J. *et al.* Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature* **529**, 496–501 (2016).

360. Seelbinder, B. *et al.* Triple RNA-seq reveals synergy in a human virus-fungus co-infection model. *Cell Rep.* **33**, 108389 (2020).

361. Westermann, A. J., Barquist, L. & Vogel, J. Resolving host–pathogen interactions by dual RNA-seq. *PLOS Pathog.* **13**, e1006033 (2017).

362. Gu, H. *et al.* Carbon storage regulator CsrA plays important roles in multiple virulence-associated processes of *Clostridium difficile*. *Microb. Pathog.* **121**, 303–309 (2018).

363. Zhu, D., Wang, S. & Sun, X. FliW and CsrA govern flagellin (FliC) synthesis and play pleiotropic roles in virulence and physiology of *Clostridioides difficile* R20291. *Front. Microbiol.* **12**, 735616 (2021).

364. Eichner, H., Karlsson, J. & Loh, E. The emerging role of bacterial regulatory RNAs in disease. *Trends Microbiol.* **30**, 959–972 (2022).

365. Zhu, D., Sorg, J. A. & Sun, X. *Clostridioides difficile* biology: sporulation, germination, and corresponding therapies for *C. difficile* infection. *Front. Cell. Infect. Microbiol.* **8**, 29 (2018).

366. Mondhe, M., Chessher, A., Goh, S., Good, L. & Stach, J. E. M. Species-selective killing of bacteria by antimicrobial peptide-PNAs. *PLoS One* **9**, e89082 (2014).

367. Sully, E. K. & Geller, B. L. Antisense antimicrobial therapeutics. *Curr. Opin. Microbiol.* **33**, 47–55 (2016).

Acknowledgements

Science is never the work of one, but the work of many. It begins with the numerous efforts of people you build your work on, to the many colleagues and friends who support you along the journey, all the way to the people who continue the work once you're gone. Over the years, a lot of colleagues and collaborators have contributed to this work, even if not mentioned specifically, without whom this thesis wouldn't have been possible, Thank you!

First and foremost, I'd like to thank my supervisor Franziska Faber for giving me the opportunity to write my thesis in her group. Her optimism and passion for science as well as her never-ending support as a group leader is truly inspiring and something I will always be deeply thankful for.

I'd also like to extend my thanks to the members of my thesis committee, Alexander Westermann, Ana Rita Brochado, and Meina Neumann-Schaal for their insightful suggestions and constructive criticism of this work.

A lot of thanks go the current and former members of the Faber lab - in particular, Tina, Vanessa, Johannes, Janet, Jessi, Chris, Thao, Kerstin & Anna for all their support, and for creating such a positive and supportive work environment. I also want to acknowledge the friends I've made throughout my years in Würzburg, for their help, their scientific insight, and for upholding my mental health by making the last several years so enjoyable, particularly Austin, Elise, Gianluca P, Falk, Gohar, Daniel, Sarah, Gianluca M, Elisa, Yan, Svetlana, Esther, Valentina, Shuba, Mindy and all the others, that I couldn't mention by name. Special thanks to Austin and Falk for taking the time to read through this work and for their insightful comments. Furthermore, I'd like to thank Hilde and Elisa for keeping the administrative site running smoothly, the Graduate School of Life Sciences for their organization of the graduate program and for further funding as well as Sci-Hub for providing free and unrestricted access to scientific knowledge to everyone.

Lastly, thanks to my friends and family at home for always offering a different perspective and to Christian for his support in everything.

Curriculum Vitae

Affidavit / Eidesstattliche Erklärung

I hereby confirm that my thesis entitled "Global discovery and functional characterization of Hfq-associated sRNA-target networks in *C. difficile*" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, 08/05/2023

Place, Date | Signature

Hiermit erkläre ich an Eides statt, die Dissertation „Globale Identifizierung und funktionelle Charakterisierung von Hfq-assoziierten sRNA-Zielnetzwerken in *C. difficile*" eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, 08/05/2023

Ort, Datum | Unterschrift