TITLE: Critical factors for precise and efficient RNA cleavage by RNase Y in *Staphylococcus aureus*

Authors: Alexandre Le Scornet[1], Ambre Jousselin[1], Kamila Baumas[1], Gergana Kostova[2], Sylvain Durand[2], Leonora Poljak[1], Roland Barriot[1], Eve Coutant[1], Céline Péllisier[1], Gladys Munoz[1], Ciarán Condon[2], Peter Redder[1]*

1) Laboratoire de Microbiologie et Génétique Moléculaires (LMGM), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Toulouse III – Paul Sabatier, Toulouse, France.
2) UMR8261 CNRS, Université Paris Diderot (Sorbonne Paris Cité), Institut de Biologie Physico-Chimique, Paris, France.

*) Corresponding author email: peter.redder@univ-tlse3.fr

Abstract

Cellular processes require precise and specific gene regulation, in which continuous mRNA degradation is a major element. The mRNA degradation mechanisms should be able to degrade a wide range of different RNA substrates with high efficiency, but should at the same time be limited, to avoid killing the cell by elimination of all cellular RNA. RNase Y is a major endoribonuclease in found most Firmicutes, including *Bacillus subtilis* and *Staphylococcus aureus*. However, the molecular interactions that direct RNase Y to cleave the correct RNA molecules at the correct position remain unknown. In this work we have identified transcripts that are homologs in *S. aureus* and *B. subtilis*, and are RNase Y targets in both bacteria. Two such transcript pairs were used as models to show a functional overlap between the *S. aureus* and the *B. subtilis* RNase Y, which highlighted the importance of the nucleotide sequence of the RNA molecule itself in the RNase Y targeting process. We then truncated one model transcript to identify 103 nucleotides that are sufficient for this targeting, and detailed analyses of this region revealed sequence and structural elements that drive cleavage efficiency and fine-tune the positioning of the cleavage to a specific nucleotide bond.

Introduction:

RNA turnover is a central element in the gene regulation of all organisms, but especially so in bacteria where mRNA half-lives are typically 2-5 minutes, but can be as low as 18 seconds

(Steglich et al. 2010; Anderson et al. 2010). These short RNA lifespans allow the cells to control gene expression at the transcriptome level, since a simple pause in the synthesis of a given mRNA will rapidly lead to a complete depletion of this mRNA. The RNA decay machinery is continuously active to balance out RNA synthesis, but that does not mean that all RNAs are treated equally by the RNA decay machinery. Ribosomal RNA, tRNAs, several small regulatory RNAs, and even a few mRNAs exhibit half-lives that are longer than the generation time of the cells (Durand et al. 2015; Khemici et al. 2015; Hamouche et al. 2021), strongly suggesting that specific factors protect a certain RNAs from degradation or, alternatively, target specific RNAs for degradation. The RNases can either degrade an RNA from the 5' or 3' ends (5' and 3' exoribonucleases, respectively) or cleave the RNA internally (endoribonucleases), in which case the endoribonuclease has as many potential substrates as there are phophodiester bonds in the RNA molecule. Endoribonucleases cleavage events can roughly be classified into three types: i) Cleavage leads to degradation of the entire RNA molecule, ii) Cleavage leads to two stable RNA molecules, iii) Cleavage leads to rapid degradation of either the upstream or downstream fragment, whereas the remaining fragment is more stable. Indeed, there are multiple examples of mRNAs that are cleaved to produce a mature mRNA or regulatory RNA. For example, the mRNA of the *saePQRS* virulence regulation operon in *Staphylococcus aureus* is cleaved by RNase Y between the *saeP* and *saeQ* coding sequences, leading to rapid degradation of the *saeP* fragment and a stable *saeQRS* portion of the mRNA. Consequently, the *saeQRS* mRNA has a much higher steady state level than *saeP* mRNA, even though they are transcribed (and regulated) simultaneously (Meinken et al. 2003; Marincola et al. 2012). Importantly, if such cleavage events maintain the correct homeostasis between the proteins coded by the operons, then there is an evolutionary pressure to conserve both the elements that position the cleavage at the correct nucleotide and the elements that ensure that it happens with the appropriate efficiency. If the positioning is different, then the resulting cleavage products may have altered stability; if the cleavage efficiency is too high or too low, then the relative abundance of the two cleavage products may be inappropriate, leading to misregulation of the gene products.

RNase Y represents one of the three major families of bacterial RNA decay systems (the two others being RNase E and RNase J). RNase Y is ubiquitous in Firmicutes (Gram positive bacteria) and is scattered in most other bacterial families, with the notable exceptions of Cyanobacteria, Alpha-, Beta- and Gamma-proteobacteria (Laalami et al. 2014). Even prior to the discovery of the ribonuclease activity of RNase Y (Shahbabian et al. 2009), the gene encoding this enzyme (*rny/cvfA*) had been determined to be crucial for *S. aureus* virulence in silkworms (Kaito et al. 2005), a finding that was further confirmed by a mouse study (Marincola et al. 2012). RNase Y from *S. aureus* and *Bacillus subtilis* are highly similar (68% identity and 78% similarity) and have an N-terminal membrane anchoring alpha-helix, followed by a disordered coiled-coil domain, an RNA binding KH domain, a HD domain containing the active site and a C-terminal region of unknown function (Figure S2) (Lehnik-Habrink et al. 2011; Redder 2018a). The membrane anchor is not essential for enzymatic activity, but deleting it does lead to slower cell division in *S. aureus*. This effect is presumably due to RNase Y activity at inappropriate intracellular locations, since mutants with the RNase Y active-site inactivated, or indeed with the whole RNase Y gene deleted, exhibit growth similar to the wild-type strain, although the virulence is greatly diminished in both mouse and silkworm infection models (Kaito et al. 2005; Marincola et al. 2012; Khemici et al. 2015). In contrast to the minor effect in *S. aureus*, deletion of RNase Y in *B. subtilis* results in growth defects and severe physiological changes to the bacteria (Figaro et al. 2013). Deletion of RNase Y in *S. aureus* increases the

half-lives of about 15% of the transcripts (Khemici et al. 2015). This relatively small number indicates that mechanisms are in place to select the RNase Y target RNAs among the global population of RNA in the cell. Furthermore, the RNAs are not cleaved at random positions, but are consistently cleaved at specific sites (Marincola et al. 2012; Khemici et al. 2015; Marincola and Wolz 2017). Finally, it has been observed for multiple targets that RNase Y does not cleave all the molecules of a given RNA and that its efficiency is limited by unknown factors.

RNase Y interacts with several other proteins, and while different studies sometimes find different interaction partners, in *B. subtilis* there is for example some evidence for interactions with the 3' exonuclease PNPase, whereas this interaction was not identified in *S. aureus* (Lehnik-Habrink et al. 2011; Roux et al. 2011). All reports from *S. aureus* and *B. subtilis* agree that both the glycolytic enzyme enolase and the DEAD-box RNA helicase CshA interact with RNase Y (Lehnik-Habrink et al. 2010, 2011; Roux et al. 2011; Giraud et al. 2015; Redder 2018a). However, these interactions are thought to be transitory, since several attempts to isolate an RNase Y-based degradosome complex have failed.

In an effort to understand the role of RNase Y within the larger context of RNA decay, transcriptomic assays have been performed in *S. aureus*, *B. subtilis* and *Streptococcus pyogene*s, to identify the exact cleavage sites of RNase Y. Briefly, these assays compare RNA ends found in wild-type and RNase Y mutant strains; RNA ends that appear in the wild-type but not in the RNase Y mutant are presumed to be generated by an RNase Y cleavage. This approach has led to the identification of 99, 27 and 190 precise RNase Y cleavage sites in *S. aureus*, *B. subtilis* and *S. pyogenes*, respectively (Khemici et al. 2015; DeLoughery et al. 2018; Broglia et al. 2020). Comparisons of the sequences surrounding the RNase Y cleavage positions in *S. aureus* and *S. pyogenes* revealed a strong preference for a purine immediately upstream of the cleaved phosphodiester bond, with a preference for G over A. No additional consensus sequence elements could be identified at the primary sequence level, but *in silico* analyses of the putative secondary structures revealed a tendency to find potential hairpin structures in the vicinity of (but not overlapping with) the cleavage sites (Khemici et al. 2015; Broglia et al. 2020). These findings were in concordance with data from the *S. aureus sae* operon, where RNase Y cleaves 7 nucleotides upstream of a 21 bp *in silico*-predicted intra-molecular duplex, formed by the two complementary regions. Deletions of either complementary region abolished RNase Y cleavage, as did point mutations in one of the regions that would prevent duplex formation (Marincola et al. 2012; Marincola and Wolz 2017). Marincola and Wolz also showed that a 13 bp deletion encompassing the cleavage position generated a new cleavage site was 13 nt upstream of the native site, and a similar shift of 20 nt was observed with a deletion of 20 bp. This suggested a molecular ruler mechanism, where the cleavage position was chosen based on the distance from a feature downstream, potentially the base of duplex between the complementary regions (Marincola and Wolz 2017). Nevertheless, other questions remained unanswered: are the nucleotides that generate affinity for RNase Y identical to those that direct RNase Y to cleave at a specific position, and are the *sae* complementary regions sufficient to cause RNase Y cleavage of other RNA molecules?

In this study we explore the relationship between RNase Y and several different target RNAs, and we establish that the nucleotide sequence surrounding an RNase Y cleavage site is sufficient for cleavage and can convert a non-substrate RNA into a substrate. We also use systematic mutational analyses to determine which nucleotides in the vicinity of the cleavage

site are required for efficient cleavage and which nucleotides determine the exact cleavage position. These findings are shown to be general within the Firmicutes, since homologous operons are cleaved by RNase Y in both *S. aureus* and *B. subtilis*, and both RNases and RNA substrates from one bacterium can replace their equivalent in the other. Finally, we show that the *S. aureus gapR* transcript has evolved to maintain a precise cleavage position, but with low cleavage efficiency, and that the efficiency can be modulated by individual point mutations around the cleavage site.

Results:

Identifying model RNase Y cleavage sites

Endoribonucleolytic cleavage events can either be useful to the cell (increase fitness), detrimental to the cell (decrease fitness), or be neutral (no change in fitness). There is an evolutionary pressure to avoid RNase Y target sequences that are detrimental to the cell, and such RNA sequences are therefore not expected to be found in a strain possessing the enzyme. To distinguish between neutral cleavage sites and beneficial cleavage sites, we argued that a useful cleavage site/event would be much more likely to be evolutionary conserved, since a neutral event can be lost by genomic drift without any fitness loss to the cell.

To identify evolutionary conserved cleavage sites, of RNase Y from *S. aureus* and *B. subtilis*, (SaY and BsY, respectively), we first compared the genes with known RNase Y cleavage sites in both bacteria (Khemici et al. 2015; DeLoughery et al. 2018), and identified the transcript-pairs Sa-*gapR*/Bs-*cggR* and Sa-*glnR*/Bs-*glnR* that are conserved in both organisms and cleaved by their respective RNase Ys at similar positions (Figure 1). The Sa-*gapR*/Bs-*cggR* operons encode a sugar-binding transcriptional regulator, glyceraldehyde-3-phosphate dehydrogenase, enolase and other enzymes in the glycolytic pathway (Meinken et al. 2003; Commichau et al. 2009), whereas the Sa-*glnR*/Bs-*glnR* operon encodes glutamine synthetase repressor and glutamine synthetase. To be able to modify these metabolic transcripts without altering the growth of the cell, we cloned the regions surrounding the RNase Y cleavage sites in front of the transcriptional terminator in the *E. coli-S. aureus* shuttle-vector pEB01 (Figure 1B). The pSaGap construct was designed with an 864 bp (288 codons) deletion within the Sa-*gapR* open reading frame. This in-frame deletion preserved the start and stop codons, but shortened the produced ΔGapR peptide to 51 amino acids, thus ensuring that no functional GapR could be produced from the multi-copy plasmid to interfere with cell physiology, as well as reducing the transcript to a manageable length for cleavage detection by both Northern blotting and EMOTE (Exact Mapping Of Transcriptome Ends, see Materials and Methods). In the same way, the pBsCgg construct was made by deleting 708 bp (236 codons) from the *cggR* open reading frame leaving a short (106-codon) ΔCggR peptide. A similar approach was used to study cleavage of the conserved RNase Y target *glnR*. pSaGln and pBsGln were constructed with 150 bp (50 codons) and 189 bp (63 codons) deletions respectively in the *glnR* open reading frames.

The resulting constructs were introduced into *S. aureus* wild-type (WT) and RNase Y deletion mutants (ΔY). RNA was prepared from the transformed *S. aureus* strains, and the cleavage (or lack thereof) was monitored by Northern blotting. We were able to detect the full-length transcripts and the smaller RNase Y cleavage products for all four constructs (Figure 1C) using the probe P1, which hybridises to the vector-derived portion of the mRNA upstream of the transcription terminator (Figure 1B). Lack of downstream cleavage product in the ΔY strain confirmed that the cleavage was indeed RNase Y-dependent.

It has previously been shown that the upstream cleavage product of the *B. subtilis cggR* operon mRNA is quickly degraded (Commichau et al. 2009), and this was also the case for the equivalent upstream cleavage product from the pSaGap transcript, which we could only detect when using a strain where the 3' exonuclease PNPase was deleted (Supplementary Figure S1). The rapid degradation of the upstream cleavage product in a WT strain led us to use the downstream product (and probe P1) as an indicator for cleavage in the rest of this study.

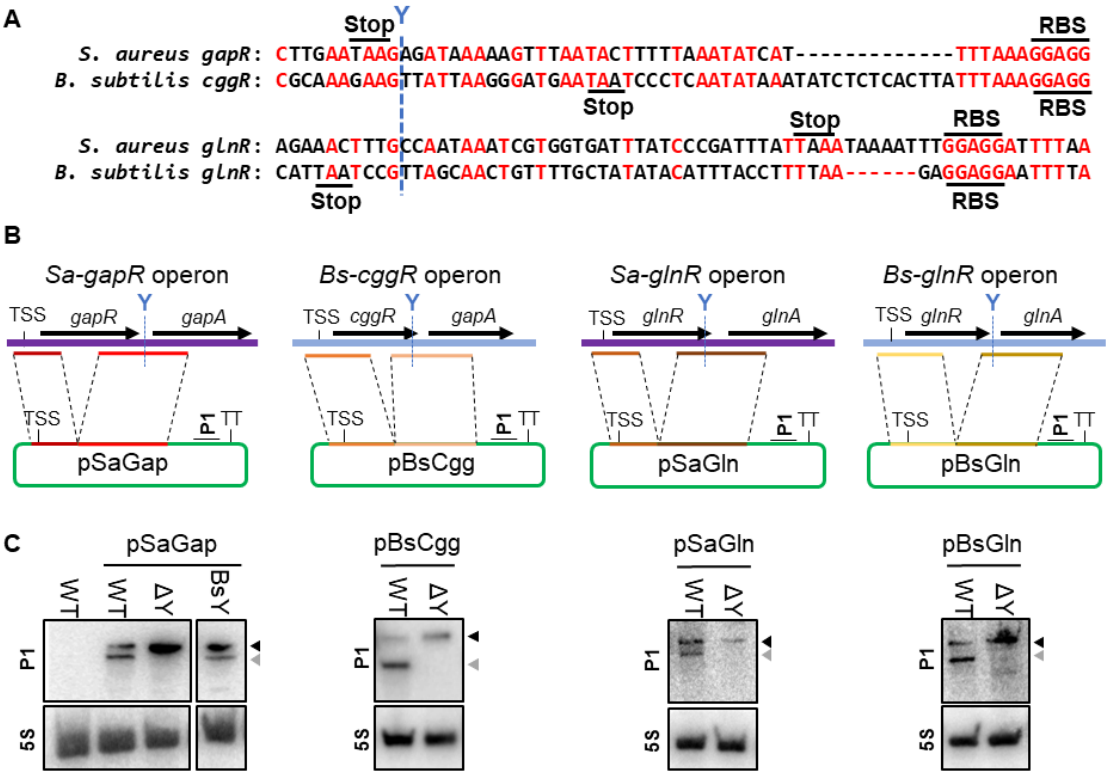

Figure 1. RNase Y cleaves the model transcripts.
A) Sequence alignment of *S. aureus* and *B. subtilis* transcripts with the published RNase Y cleavage sites indicated (Meinken et al. 2003; Khemici et al. 2015; DeLoughery et al. 2018).
B) Schematic representation of the *Sa-gapR*, *Sa-glnR*, *Bs-cggR* and *Bs-glnR* operons in *S. aureus* and *B. subtilis*. TSS indicates the transcription start site, and the native RNase Y cleavage sites are indicated by blue dotted lines. Coloured lines show the regions cloned in front of the transcription terminator (TT) of the pEB01 vector (green). The location of the P1 probe is indicated.
C) The four constructs from panel B were transformed into the WT and ΔY strains, and RNase Y cleavage was detected by Northern blotting using the P1 probe to detect the full-length (black arrowhead) and the smaller RNase Y cleavage product (grey arrowhead). A probe

against 5S rRNA (5S) is used as loading control. pSaGap was also introduced into the *S. aureus* strain BsY, which expresses *B. subtilis* RNase Y from the *S. aureus rny* promoter at the native *S. aureus* chromosomal *rny* locus.

All necessary information for cleavage selectivity is found within the target RNA nucleotide sequence.

The RNase Y proteins from *S. aureus* and *B. subtilis* are highly similar, but not identical (Figure S2). This difference could potentially explain why the conserved RNase Y cleavage positions in *S. aureus* and *B. subtilis* only align roughly. We therefore expressed the *cggR* transcript (originating from *B. subtilis*) in *S. aureus*, and observed by EMOTE that the *cggR* RNA is cleaved at exactly the same site by SaY as it is by BsY (Figure 1 and Figure S3). In addition, we replaced the SaY open reading frame on the *S. aureus* chromosome with the BsY open reading frame, and the resulting mutant (ΔSaY::BsY) cleaved both the pBsCgg and pSaGap mRNAs at their respective native sites. These results rule out that a hypothetical organism-specific "RNase Y associated factor" is responsible for guiding the SaY and BsY to their respective sites, since any organism-specific factor would be absent during exogenous expression. The cleavage-site difference between the pSaGap and pBsCgg transcripts is therefore purely due to the differences in nucleotide sequence between the two RNA molecules. Moreover, to ensure that the *Sa-gapR*/*Bs-cggR* pair was not a unique case, we verified that the pSaGln and pBsGln mRNAs are also both cleaved at their native sites, irrespective of the origin of the RNase Y (Figure 1C and S3).

S. aureus RNase Y is functional in *B. subtilis*.

We next wanted to examine whether SaY and BsY were interchangeable in *B. subtilis*. We therefore ectopically expressed SaY and BsY in a *B. subtilis* RNase Y mutant and, after addition of rifampicin to block transcription, examined two well-characterised target RNAs, *glnAR* and *atpIBE*, by Northern blotting. In the absence of RNase Y, full-length transcripts were stabilized and the processed species no longer detected. In Δ*rny* cells ectopically expressing either SaY or BsY, processed forms of the *glnAR* and *atpIBE* transcripts accumulated with similar stabilities to those observed in a wild-type *B. subtilis* strain, while the full-length mRNA disappeared rapidly (Figure 2). Thus, similar to what we observed in *S. aureus*, the enzymes can complement each other in *B. subtilis*.

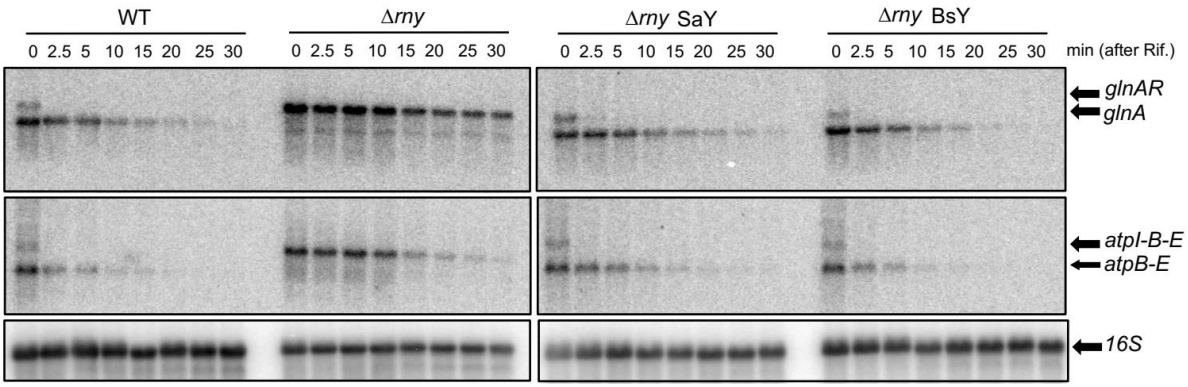

Figure 2. *S. aureus* RNase Y complements a *B. subtilis* Δ*rny* strain.

Northern blots of total *B. subtilis* RNA, showing the disappearance of the uncleaved transcripts (*glnAR* and *atpIB-E*) after rifampicin addition, accompanied by the appearance of the downstream cleavage products (*glnA* and *atpB-E*). WT: *B. subtilis* wild-type strain. Δ*rny*: *B. subtilis* with the RNase Y gene deleted. Δ*rny* SaY: Δ*rny* strain with *S. aureus* RNase Y expressed from the *amyE* locus. Δ*rny* BsY: Δ*rny* strain with *B. subtilis* RNase Y expressed from the *amyE* locus. 16S rRNA was used as loading control.

Translation is not required for RNase Y cleavage of the pSaGap transcript.

The *Bs-cggR* and *Sa-glnR* transcripts are cleaved within an ORF shortly before a stop codon and *gapR* and *Bs-glnR*, are cleaved just after a stop codon but within the ~15 nt region covered by the ribosome (Figure 1A). It was therefore possible that the translational machinery influenced RNase Y target site selection. We therefore constructed two variants of pSaGap that avoid nucleotide changes in close vicinity to the cleavage site but nevertheless modify ribosome positioning/occupancy: pSaGap[NoStart] with the *gapR* start codon mutated, and pSaGap[Δ128] with a 128 nt deletion which causes the ribosome to encounter a stop codon five nucleotides upstream of the original *gapR* stop codon. Northern blot analyses of transcripts from both vectors revealed a prominent cleaved product (Figure S4A) and EMOTE confirmed that the cleavage of both transcripts still occurs precisely at the native position (Figure S4B). We therefore conclude that translation does not influence the RNase Y cleavage site position of the pSaGap transcript.

G, A and U are all accepted upstream of the cleavage site.

We originally detected a preference for a guanidine residue immediately upstream of the RNase Y cleavage sites in *S. aureus*, and this preference has since been observed for RNase Y from *Streptococcus pyogenes* as well (Khemici 2015; Broglia 2018; Broglia 2020). To gauge the importance of this nucleotide in RNase Y site selectivity, we generated constructs pSaGap[G258U], pSaGap[G258C] and pSaGap[G258A], where this nucleotide was mutated to uridine, cytosine and adenosine, respectively. Northern blots performed on these strains revealed that RNase Y-dependent cleavage was reduced but still detectable in each of the three mutants (Figure 3A). To determine whether the RNase Y cleavages were at the exact same sites as in the pSaGap transcript, we complemented the Northern blot data with EMOTE targeted to the transcripts expressed from the vector (Redder 2018b). Targeted EMOTE is

similar to standard 5'-RACE protocols in that an oligo is ligated to the 5'-end of the RNA, followed by reverse transcription with a transcript-specific primer and subsequent PCR amplification. However, instead of cloning the amplified cDNA into a vector, the PCR product is directly sequenced with Illumina sequencing, to obtain the exact 5'-end of more than 5000 individual cleaved molecules per RNA sample (in contrast to the 10-20 molecules typically analysed when cloning is used for 5' RACE).

The EMOTE data revealed that RNase Y cleavage of the pSaGap[G258C] transcript preferentially occurred immediately upstream or downstream of the native cleavage site, whereas in pSaGap[G258A] and pSaGap[G258U], cleavage was still preferentially at the native site, but frequently occurred a few nucleotides upstream or downstream (Figure 3B). This pattern was also observed in the ΔSaY::BsY strain, where the *B. subtilis* RNase Y cleaves exactly after the G, but somewhat imprecisely when the G is substituted by a U (Figure 3C). These results confirmed that the G is not required for cleavage *per se*, but that it plays an important role in avoiding heterogeneity of the RNase Y cleavage position.

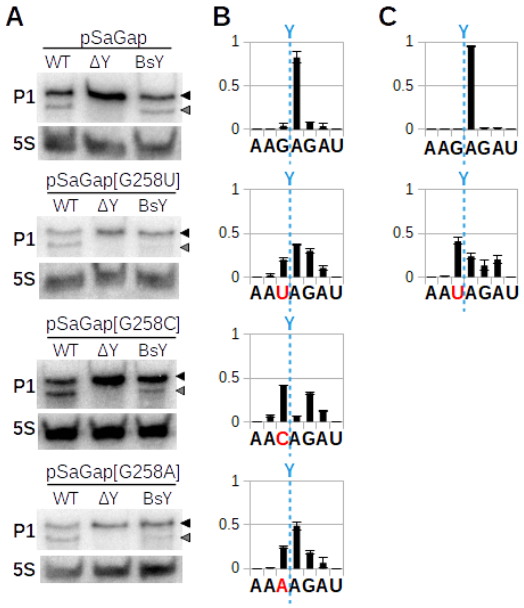

Figure 3, The upstream G promotes precise cleavage.

A) Northern blots using probe P1 on transcripts from wild-type pSaGap, and versions where the G upstream of the cleavage site has been mutated to U, C and A (pSaGap[G258U], pSaGap[G258C] and pSaGap[G258A]), respectively. Each plasmid was introduced into the WT, ΔY and BsY strains. The full-length transcript and the downstream cleavage product are indicated by black and grey arrowheads, respectively. A 5S rRNA probe was used as loading control.

B) EMOTE analysis of the exact cleavage position for *S. aureus* RNase Y. The sequence below each graph shows the nucleotides surrounding the natively preferred cleavage site (blue dotted line). The mutated nucleotide for each plasmid is shown in red. Columns indicate the

proportion of molecules with 5'-ends at each position within the shown window. Two biological replicates were analysed, and the error-bars indicate the standard deviation.

C) Same as panel B, but where *B. subtilis* RNase Y has replaced *S. aureus* RNase Y at the native *rny* locus.

Four sectors are required for efficient RNase Y cleavage of the *gapR* transcript.

Previous studies have indicated that RNA secondary structures are important for RNase Y cleavage (Meinken et al. 2003; Marincola and Wolz 2017), and we therefore used the mFold software (Zuker 2003) to predict the secondary structures surrounding the RNase Y cleavage site. Based on this prediction, we divided the region around the RNase Y cleavage site in the pSaGap transcript into six sectors, named I to VI (Figure 4A). We then examined which sectors of the pSaGap transcript were required for the RNase Y cleavage.

Deletion of the large sector I, ending 14 nt upstream of the cleavage site, did not measurably alter the cleavage (Figure 4). We therefore went on to generate a construct where both sectors I and II are deleted (pSaGap[ΔIΔII]), and observed that this abolished cleavage of the resulting transcript (Figure S5A). Since sector II includes the G immediately upstream of the RNase Y cleavage site, it was possible that the absence of this G prevented cleavage. We therefore constructed pSaGap[ΔIΔII+G], which was identical to pSaGap[ΔIΔII], except that the G was left intact. However, reintroducing the G was not sufficient to re-establish cleavage (Figure S5A), showing that the 12 first nucleotides of sector II contain elements essential for efficient RNase Y cleavage. This was confirmed when no cleavage product could be detected by Northern blotting of pSaGap[ΔII+G] transcript (Figure 4B).

A similar approach was taken at the 3' end of the pSaGap construct, where sector VI was deleted, either alone (pSaGap[ΔVI], Figure 4B), or in combination with sector I (pSaGap[ΔIΔVI], Figure S5B). Neither of these deletions prevented RNase Y cleavage (Figures 4B and S5B). However, in a construct where both sectors V and VI were deleted (pSaGap[ΔVΔVI]), cleavage was undetected by Northern blotting and barely detectable by EMOTE (Figure 4B and 4C). This result suggested that sector V might be important for cleavage, but to our surprise a cleavage product was visible by Northern blot for a transcript with only sector V deleted (pSaGap[ΔV]) (Figure 4B). Furthermore, EMOTE confirmed that the cleavage of the pSaGap[ΔV] transcript was at the native position (Figure 4C).

Sector IV is predicted to form a very weak (ΔG of -5 kcal/mol) hairpin three nucleotides downstream of the cleavage site (Figure 4A). A cleaved pSaGap transcript was no longer detectable by Northern blotting when sector IV was deleted (pSaGap[ΔIV]), but careful analysis of the corresponding EMOTE data revealed a weak cleavage at the exact position of the native RNase Y cleavage site. This cleavage was not observed in the ΔY strain, showing that it is an RNase Y-dependent event.

Overall, our deletion analysis suggested that sectors II, III and IV together with either Sector V or VI would be sufficient to define an RNase Y cleavage site.

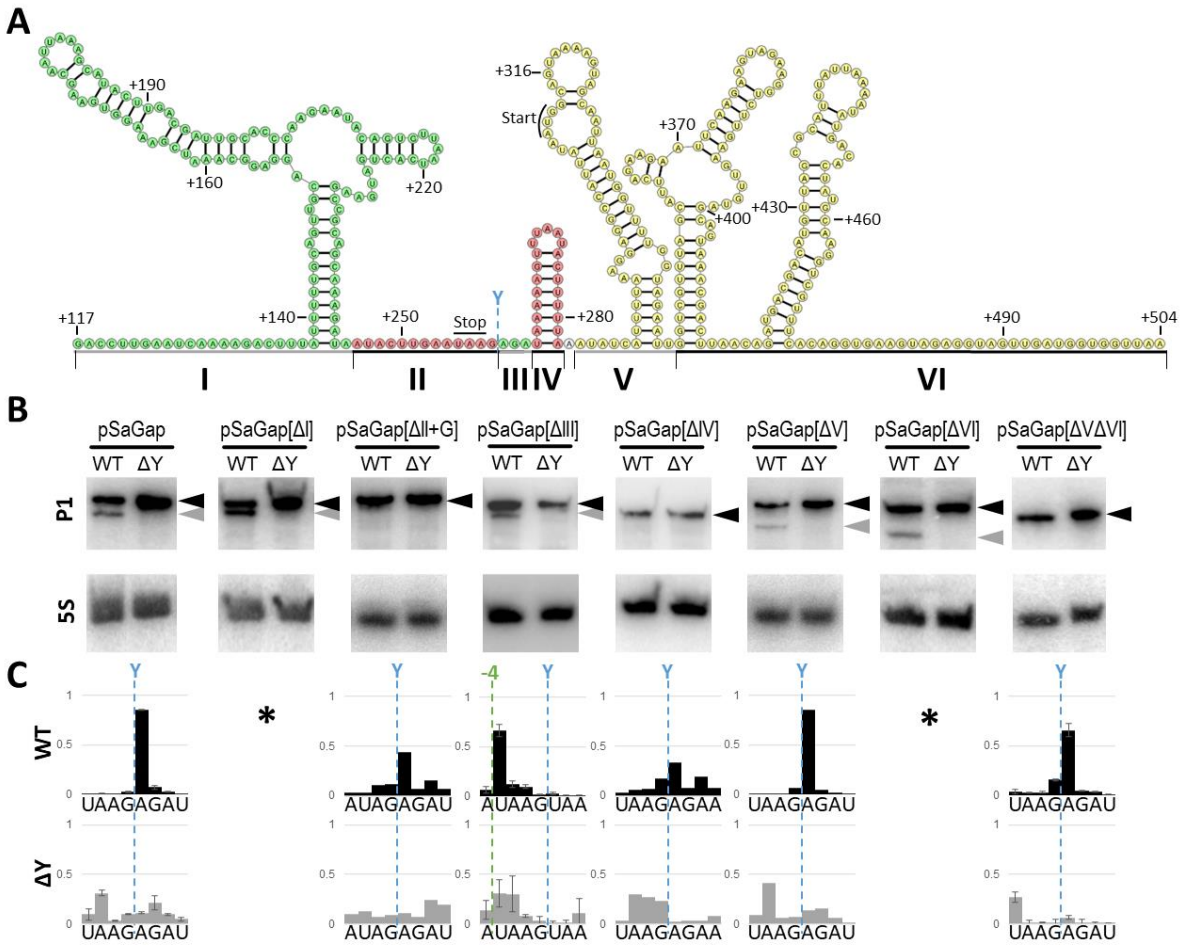

Figure 4, Deleting Sectors of the pSaGap transcript in the region surrounding the RNase Y cleavage.

A) Predicted secondary structures of positions +117 to +504 in the pSaGap vector transcript. We divided the region into six sectors (I-VI), shown with black and grey lines below the nucleotide sequence. The native RNase Y cleavage site (Y-cut) is located between sector II and sector III, and the positions of the stop codon of *gapR*$^{\Delta 864nt}$ and the start codon of *gapA* are indicated. Green sectors can be deleted without losing cleavage; red sectors cannot be deleted; yellow sectors can be deleted under certain conditions.

B) Northern blots of the pSaGap transcripts deleted of the various sectors. Black arrowheads denote full-length transcripts and grey arrowheads denote the downstream cleavage products. 5S rRNA was used as loading control.

C) EMOTE data showing the exact location of 5' ends in the strains used for Northern blotting in panel B. Asterisks indicate that EMOTE data was obtained for the vector where both sector I and sector VI were deleted, and that no change in cleavage site was detected (see Figure S5B). The green "-4" indicates the 4 nt upstream shift of the cleavage position in the pSaGap[ΔIII] transcript. Note that the EMOTE data is presented as proportions of RNA

molecules with a given 5' end on the Y-axis (number of reads detected at a specific position divided by the total number of reads detected within the chosen window), and that the number of detected molecules is much larger in the WT strain than in the ΔY strain.

Sectors II to V are sufficient for RNase Y cleavage of an exogenous transcript.

We next wanted to use this knowledge to generate an RNase Y cleavage site in a previously uncleaved transcript. We generated a vector which expressed the heterologous *E. coli fliM* transcript (encoding a flagellar motor protein) with and without the 103 nts of sector II, III, IV and V (Figure 5A). We could detect the full-length *fliM* transcript in both *S. aureus* WT and ΔY strains, as a single uncut RNA by Northern blot (Figure 5B). When the chimeric *fliM*::II-V transcript from the WT strain was examined we could observe both the full-length transcript and a cleavage product of the expected size. However, we were equally able to observe the cleaved *fliM*::II-V product in the ΔY strain, suggesting that the cleavage was not RNase Y dependent (Figure 5C). To understand this unexpected result, we examined the *fliM*::II-V RNA by EMOTE, and discovered two different cleavage sites in the WT strain, only 2 nucleotides apart (Figure 5D). One of these cleavages was unexpectedly detected in both WT and ΔY and is responsible for the cleaved product observed by Northern blot for *fliM*::II-V in the ΔY strain. This means that the insertion of sectors II to V into the *fliM* sequence generated a cleavage site for an as yet unidentified RNase. The other cleavage was located at the exact position where RNase Y normally cleaves the *Sa-gapR* transcript, and this cleavage was RNase Y dependent, since it disappeared in the ΔY strain (Figure 5C). Thus, the addition of sectors II-V was sufficient to generate an RNase Y cleavage site in an exogenous RNA.

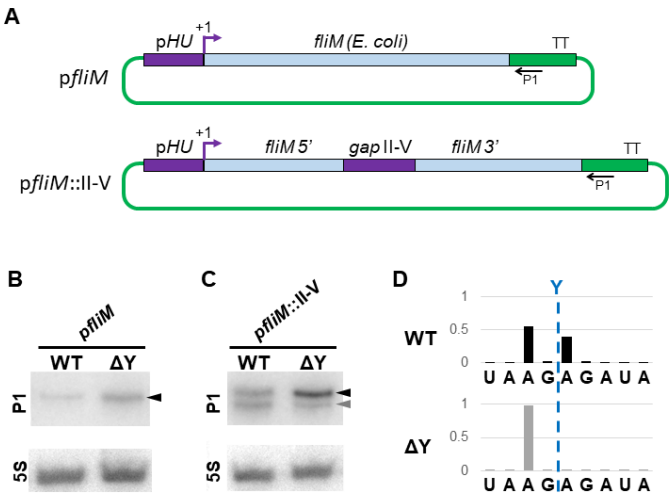

Figure 5, Generating an RNase Y cleavage site in an exogenous transcript.

A) Overview of the two vectors p*fliM* and p*fliM*::II-V, where regions from *S. aureus* are in purple, regions from *E. coli* are in light blue and regions in green are from the vector backbone.

TT indicates the transcription terminator and P1 shows the location where the P1 probe hybridises. *gap*II-V indicates the 103 bp from sector II to V in the *Sa-gapR* operon.

B) Northern blot of the *fliM* transcript. The full-length uncleaved RNA is indicated by a black arrowhead. 5S rRNA was used as loading control.

C) Northern blot of the *fliM*::II-V transcript, where the uncleaved RNA and downstream cleavage products are indicated by black and grey arrowheads, respectively. 5S rRNA was used as loading control.

D) EMOTE data for WT and ΔY strains carrying the p*fliM*::II-V vector. The data from the WT strain (black) shows two distinct peaks at the native RNase Y cleavage site (Y) as well as two nucleotides upstream, whereas the data from the ΔY strain (grey) only shows a single peak two nucleotides upstream of the native RNase Y site.

Altering the RNase Y cleavage site selection

Sector III consists of the three nucleotides which separates the RNase Y cleavage site from the predicted hairpin of sector IV. Deleting Sector III (pSaGap[ΔIII]) did not substantially alter the size or intensity of the cleaved fragment in a Northern blot, but EMOTE revealed that the cleavage site had shifted four nucleotides upstream from the native site (Figure 4C). This suggests that downstream of a cleavage site there is a reference point for measuring the cleavage position. This reference point must be somewhere within sector IV, since removal of sector V from the pSaGap did not change the cleavage position (Figure 4).

Exchanging the predicted sector IV hairpins from *Bs-cggR* and *Sa-gapR* alters the cleavage pattern.

Similar to what we observe for the *Sa-gapR* transcript, the *Bs-cggR* RNA is also predicted to form a hairpin structure immediately downstream of the cleavage site (Figure 6B), but with the key difference that the *Bs-cggR* cleavage site is only a single nucleotide upstream of the predicted hairpin. We exploited this minor difference in cleavage position, and examined cleavage in a construct where sector IV from pSaGap was exchanged with the corresponding hairpin from *Bs-cggR* (Figure 6C; pSaGap[ΔIV::cggHP]). The hypothesis was that if the position of the base of the hairpin was the defining factor for RNase Y cleavage site selection, then the transcript from pSaGap[ΔIV::cggHP] would be cleaved at the same site as for pSaGap. Alternatively, if the defining factor was not the transition from single-stranded RNA to double stranded RNA, but instead further within the hairpin, then the cleavage site should be identical to that from the pBsCgg transcript. Northern blotting showed that the pSaGap[ΔIV::cggHP] transcript was cleaved efficiently, demonstrating that an exogenous hairpin (from *B. subtilis*) was able to replace the factors in sector IV that are required for efficient cleavage (Figure 6D). EMOTE data furthermore revealed that the major cleavage position was shifted downstream by two nucleotides to coincide with the native cleavage position for the pBsCgg transcript (Figure 6 panels B and E). This demonstrates that hairpin composition is a major factor in cleavage positioning. However, residual cleavage was also observed at the native pSaGap position and the intermediate position, suggesting that other

factors in the pSaGap[ΔIV::cggHP] transcript are still able to guide RNase Y to cleave at the native pSaGap position, and these factors could be the start of the hairpin duplex or the composition of one of the other sectors in the pSaGap operon.

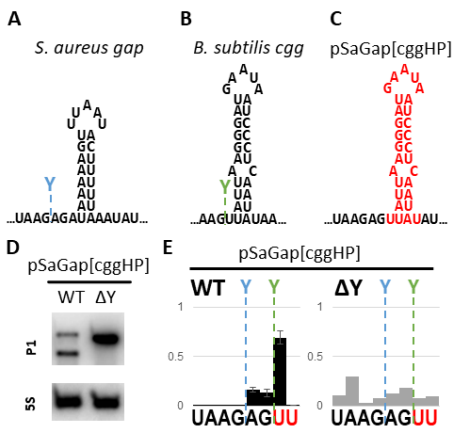

Figure 6, RNase Y cleavage site selection depends on hairpin composition.

A) *S. aureus gapR* hairpin structure immediately downstream of the cleavage site. B) *B. subtilis Bs-cggR* hairpin structure immediately downstream of the cleavage site, with the RNase Y cleavage position shown in green. C) The hairpin structure immediately downstream of the cleavage site in the pSaGap construct was exchanged with the hairpin from *Bs-cggR* to obtain pSaGap[ΔIV::cggHP]. Mutated nucleotides are shown in red. D) Northern blot to detect the RNase Y-dependent cleavage of the pSaGap[ΔIV::cggHP] transcript. E) EMOTE signal from the pSaGap[ΔIV::cggHP] transcript. The native cleavage positions of pSaGap and pBsCgg are shown in blue and green, respectively. Mutated nucleotides are in red.

Role of the putative downstream hairpin

In addition to the two predicted hairpins detected in *Sa-gapR* and *Bs-cggR*, hairpins were also predicted to form in Sa-*glnR* and Bs-*glnR*, two and three nucleotides downstream of the native cleavage site respectively. The hairpins are all AU rich, but with one or a cluster of G-C base-pairs close to the loop (Figures 7Ai, 7Bi and 7Ci, S6A). Although not all RNase Y cleavage sites are followed by hairpin structures (Khemici et al. 2015), in light of the shifts in cleavage position observed for pSaGap[ΔIII] and pSaGap[ΔIV::cggHP] (green lines in figures 4C and 6E, respectively), we nevertheless decided to examine whether modification of these predicted hairpins would alter RNase Y cleavage. We first examined whether the sequence of the loop had an impact on cleavage. The 5 nucleotide UUAAU loop sequence was inverted to AAUUA (pSaGap[InvLoop], Figure S6B), to alter the sequence while maintaining the minimal potential for additional duplex formation within the transcript. This change in the loop sequence did not prevent RNase Y from cleaving the transcript (Figure S6C), and we therefore focussed on the hairpin stem.

Hairpin G-C base-pairing promotes cleavage

For pSaGap, pBsCgg and pSaGln, the G-C pairing(s) in the stem of the potential hairpin were systematically mutated to destabilise the secondary structure, and Northern blotting showed impaired RNase Y cleavage of these transcripts (Figure 7, panels Aii, Aiv, Bii and Cii). In contrast, mutations restoring the integrity of the putative hairpin with complementary mutations all restored cleavage (Figure 7, panels Aiii, Biii and Ciii). Importantly, faint bands corresponding in size to the cleaved product could be detected in the strains with pSaGap[G269C], pSaGap[C277G] and pBsCgg[GtoC] constructs (Figure 7, panels Aii, Aiv and Bii), indicating that while the G-C base-pairings are important for cleavage of the pSaGap and pBsCgg transcripts, there are other element(s) that can permit RNase Y cleavage.

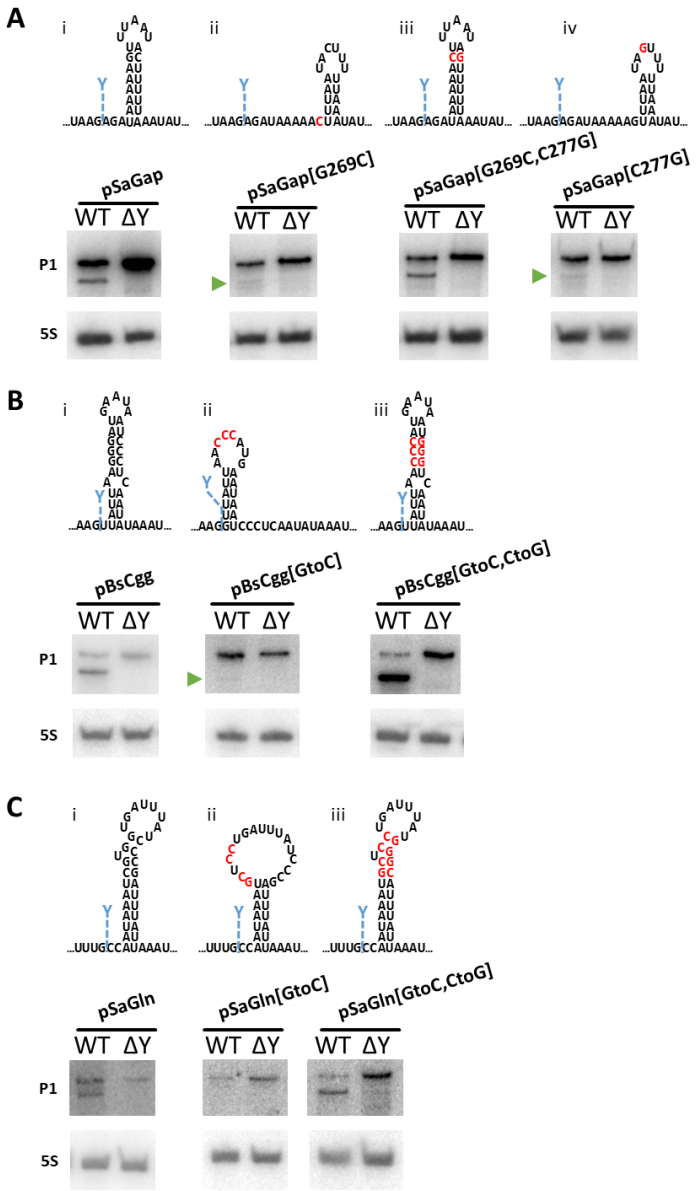

Figure 7, Hairpin G-C base-pairing promotes cleavage in pSaGap, pBsCgg and pSaGln transcripts.

A) Putative secondary structures of the pSaGap sector IV hairpin with mutated G-C base-pair (mutated nucleotides in red). Northern blots revealed efficient cleavage when the G-C pair was inverted (pSaGap[G269C,C277G]) but only weak cleavage when either the G or the C were mutated (green arrowheads). EMOTE data for these constructs can be found in Figure S7.
B) Putative secondary structures of the pBsCgg downstream hairpin with mutated G-C base-pairs (mutated nucleotides in red). Northern blots revealed efficient cleavage by *S. aureus* RNase Y when the G-C pairs were inverted (pBsCgg[GtoC,CtoG]) but only weak cleavage when the three Gs were mutated to Cs (green arrowhead).
C) Putative secondary structures of the pSaGln downstream hairpin with mutated G-C base-pairs (mutated nucleotides in red). Northern blots revealed efficient cleavage when the G-C pairs were inverted (pSaGln[GtoC,CtoG]) but only weak cleavage when CGUGG was mutated to GCUCC (pSaGln[GtoC]). EMOTE data for these constructs can be found in Figure S7.

The composition of the hairpin stem affects both cleavage efficiency and site selection.

Since the previous experiment clearly demonstrated that the G-C base-pair in pSaGap is a key factor in RNase Y cleavage, we generated mutants where the duplex of the stem was either fully inverted or where all but the G-C pair was inverted (Figure 8A). These transcripts are still predicted to form hairpin structures with identical base-composition, but RNase Y cleavage efficiency was reduced (Figure 8C) and an EMOTE assay revealed that the cleavage position was frequently shifted two nucleotides downstream (Figure 8B). A similar two-nucleotide shift could also be observed when a single U on the downstream “arm” of the stem was mutated into an A (pSaGap[U282A]). In contrast, neither the mutation of the opposing A to a U (pSaGap[A264U]) nor the double mutation to regain base-pairing (pSaGap[A264U,U282A]), gave a strong shift in cleavage site positioning (Figure 8B).

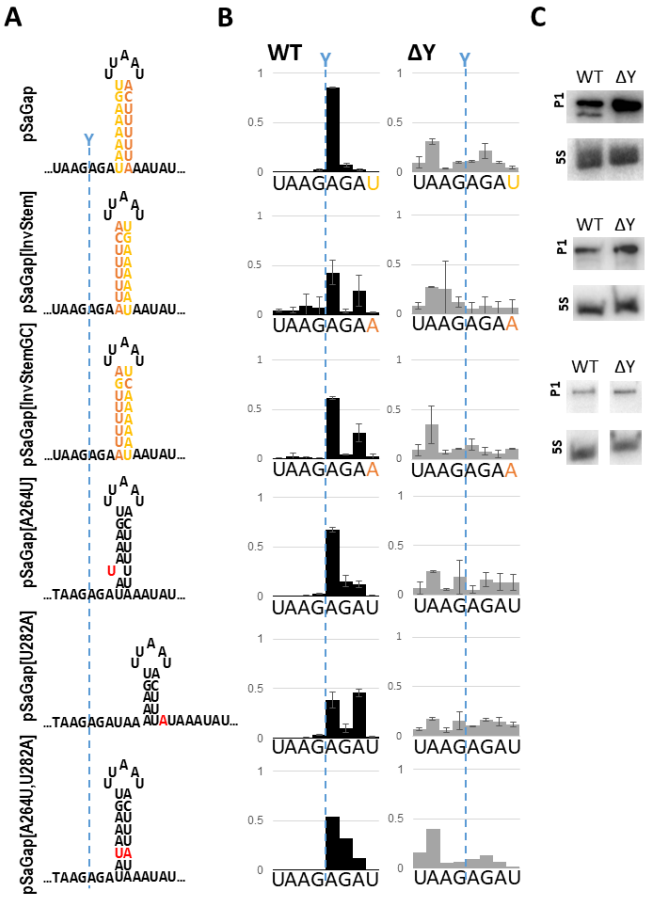

Figure 8, Modifications of the putative hairpin stem in sector IV.

A) Putative secondary structures of sector IV and derivatives. pSaGap: shown with the upstream and downstream hairpin "arms" in yellow and orange, respectively. pSaGap[InvStem]: the "arms" of the hairpin inverted. pSaGap[InvStemGC]: the "arms" of the hairpin, excluding the G-C base-pair, inverted. pSaGap[A264U]: third nucleotide of the upstream "arm" mutated to U, disrupting the potential base-pairing with U282 of the downstream "arm". pSaGap[U282A]: the third nucleotide in the downstream "arm" mutated to an A, disrupting the potential base-pairing with A264 of the upstream "arm". pSaGap[A264U,U282A]: Both A264 and U282 are mutated as above, resulting in a reconstituted putative base-pairing. B) EMOTE cleavage profiles of the mutants illustrated in panel A. C) Northern blots of pSaGap, pSaGap[InvStem] and pSaGap[InvStemGC].

Mutations immediately downstream of the *Sa-gapR* cleavage site modulate the cleavage efficiency.

Having previously observed that deleting the three nucleotides immediately downstream of the cleavage site (i.e. sector III) results in an upstream shift (Figure 4), we wondered whether the nucleotide sequence of sector III influences the cleavage efficiency. We therefore

generated a library of pSaGap plasmids expressing the 64 possible nucleotide combinations that can be generated within sector III (Figure 9A). The library was transformed into the *S. aureus* wild-type strain, and the relative abundance of cleaved RNA for each of the trinucleotide combinations was measured with EMOTE. The EMOTE signal was normalised by the relative abundance of each of the 64 plasmids in our library (we assumed that mutation of three nucleotides in the middle of the pSaGap transcript would not significantly modify transcription rates). The wild-type trinucleotide sequence (AGA) turned out to be one of the worst sequences for RNase Y cleavage efficiency, and transcripts with other NGA trinucleotides are also poorly cleaved (Figure 9B). The cleavage efficiency of transcripts with GNN was also generally low, although this could be due to biases in the detection method (see discussion). To support the EMOTE data with an alternative method, we generated individual pSaGap variants with low efficiency cleavage (pSaGap[IIIGGA] and pSaGap[IIICGA]), very low cleavage efficiency (pSaGap[IIIAUU]) and high cleavage efficiency (pSaGap[IIICUA]). The cleavage of these transcripts was examined by Northern blotting. No cleavage product was detected for the pSaGap[IIIAUU] transcript, whereas the other three transcripts were cleaved (Figure 9C).

Overall, most trinucleotide combinations gave transcripts that were cleaved more efficiently than the wild-type transcript, with some combinations (AUA, CUA, CUC, CUG and UUG) reaching >9-fold higher efficiency. The AUU and GUU trinucleotides can hypothetically extend the stem of the putative hairpin in sector IV, which would localise the cleavage site to the base of the stem and potentially prevent cleavage (Figure 9D). Both AUU and GUU transcripts are indeed cleaved with poor efficiency, and although AUU was not statistically different from wild-type in our EMOTE data, we confirmed the lack of cleavage by Northern blot (Figure 9C). Neither the CUU nor the UUU transcripts are cleaved with lower efficiency than the AGA transcript (Figure 9B), suggesting that a 2 base-pair extension of the stem is insufficient to block cleavage.

The base of the stem is one of the possible reference-points for the molecular ruler mechanism that help determine the cleavage position (see Figures 6C and E) and extending the stem by modifying Sector III will move the base of the stem further upstream, thus potentially shifting the cleavage position (Figures 9D and S8A). However, when we examined the EMOTE data upstream of the native RNase Y cleavage, we observed no shift in cleavage position for trinucleotides that extend the stem by one or two base-pairs (NVU and YUU trinucleotides, respectively) (Figures S8B and C).

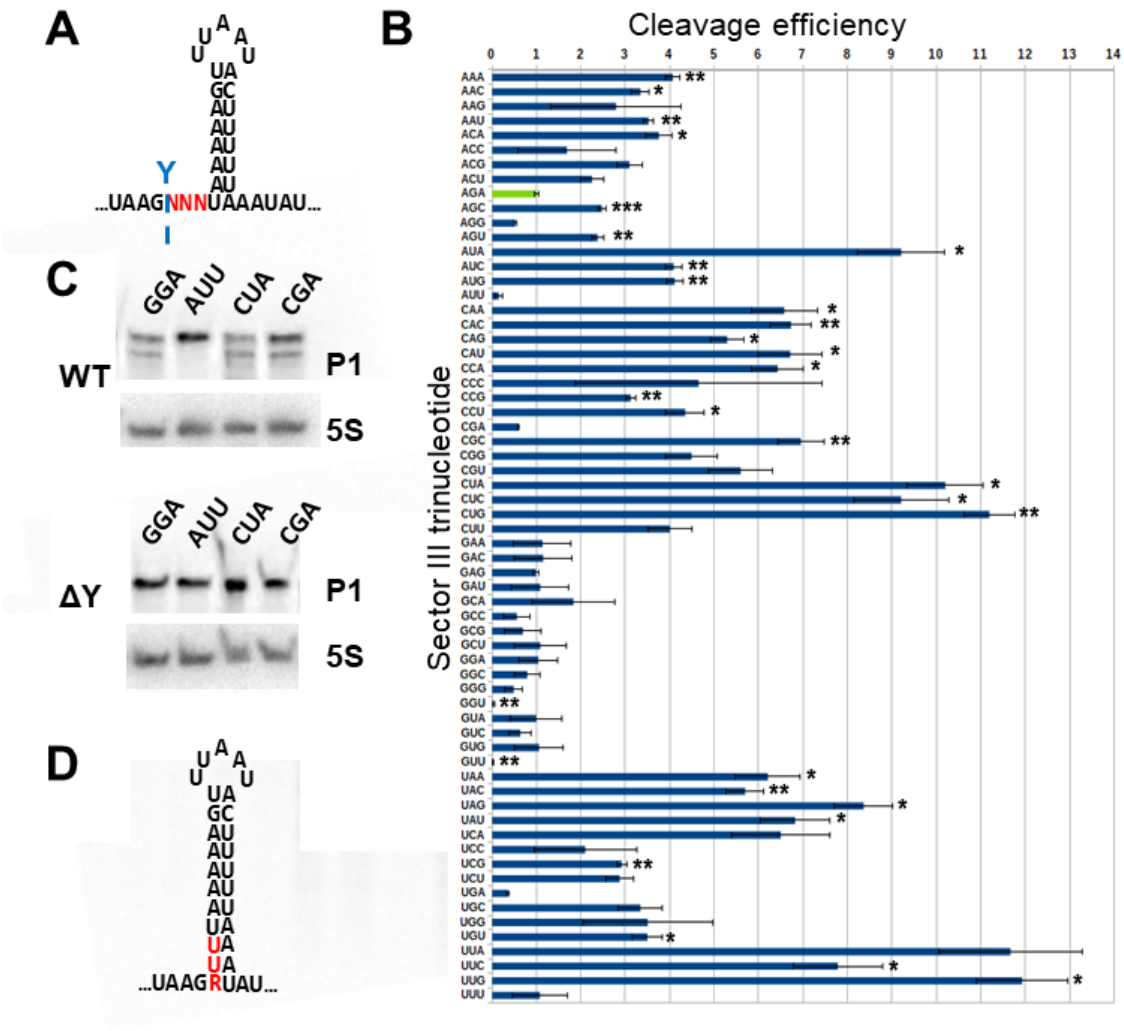

Figure 9, The three nucleotides downstream of the cleavage position strongly influence cleavage efficiency.

A) The putative hairpin in the *Sa-gapR* transcript, with the randomly mutagenised trinucleotide marked with red N's. The native RNase Y cleavage site is indicated by a light blue Y.

B) The relative cleavage efficiency, measured with EMOTE, of each of the 64 possible combinations for the three nucleotides downstream of the RNase Y cleavage site. The wild-type trinucleotide (AGA) is shown in green, and all other sequences were compared to the wild-type cleavage efficiency, using T-tests with correction for multiple testing. * = $p<0.05$, ** = $p<0.01$ and *** = $p<0.001$. The experiment was repeated four times.

C) Northern blot of constructs with selected trinucleotides in sector III (pSaGap[IIIGGA], pSaGap[IIIAUU], pSaGap[IIICUA] and pSaGap[IIICGA]) showing the lack of cleavage with AUU.

D) The potential for prolonging the hairpin stem when the trinucleotide is either AUU or GUU. R indicates a purine.

Discussion

Advantages of *in vivo* cleavage assays.

RNase Y operates within the complex environment of the intracellular space and with the physical constraints inherent in its membrane localisation. Furthermore, the cell contains a large diversity of potential RNA substrates that can compete for binding to RNase Y. Of these, only an estimated ~15% are actually cleaved by RNase Y and, importantly, are not cleaved at random phosphodiester bonds, but rather at specific positions (Khemici et al. 2015). *In vitro* assays can generate precise kinetic measurements of the enzymatic cleavage, but such data might not be relevant in context of complex intracellular conditions, and true cleavage efficiency of a given substrate must therefore be evaluated *in vivo*. In this study we isolated total RNA from *S. aureus* cultures expressing different versions of the same transcript and then used Northern blotting to detect both the full-length and cleaved substrate. This provided preliminary quantification of *in vivo* cleavage efficiency for each transcript version but did not map the exact cleavage position. We therefore complemented the Northern blots with EMOTE assays, which not only identify the exact cleavage position for individual RNA molecules but also allow us to detect RNAs with low concentrations *in vivo* (Figures 3, 4, 6 and S7). Importantly, in cultures expressing a library of similar transcripts, EMOTE could be used to simultaneously identify the exact cleavage position of each RNA molecule and to quantitatively compare the efficiencies with which they are cleaved *in vivo* (Figure 9).

Putative role of RicF/YlbF.

We showed that sectors II-V of the *Sa-gapR* transcript are sufficient to promote an RNase Y cleavage in the otherwise non-targeted *fliM* mRNA (Figure 5). This means that sectors II-V contain all the elements needed to define a precise RNase Y cleavage position, but also that the exogenous p*fliM*::II-V mRNA is presumably able to find RNase Y at the membrane. It is currently unknown whether this is caused by random diffusion of the mRNA or is facilitated by specific proteins.

B. subtilis RNase Y (BsY) is fully functional in *S. aureus*, cleaving the pSaGap transcript at the same position as SaY, showing that the two RNase Y enzymes are sufficiently similar to exhibit the same preferences for target RNAs as well as for the cleavage position on the RNA (Figure S3). RNase Y cleavage of the *Sa-gapR* operon and the *Bs-cggR* operon has been shown by DeLoughery and co-workers (DeLoughery et al. 2018) to be dependent on YlbF (annotated as RicF in *S. aureus*). Taken together, this strongly suggests that whatever requirement BsY has for YlbF can be fulfilled by RicF in *S. aureus*, and vice versa, since SaY is functional in *B. subtilis* (Figure 2). Yet, this does not explain why RNase Y requires RicF/YlbF for cleaving the majority of RNAs, while a minority of RNAs can be cleaved by RNase Y in absence of RicF/YlbF (DeLoughery et al. 2018). However, since neither imprecise nor novel cleavage positions appeared in *ylbF* and *ricF* mutants (DeLoughery et al. 2018), they may not be involved in specifying the exact RNase Y cleavage position on the RNA molecules. It is possible that RicF/YlbF aids in localising specific RNA molecules near the membrane, in which case the nucleotides recognised by RicF/YlbF must be located within the pSaGap transcript, and potentially in one of the sectors that we have established as essential for RNase Y cleavage. However, it is also possible that RicF/YlbF changes the conformation of RNase Y

to recognise transcripts where cleavage depends on RicF/YlbF, a conformational change which is apparently not required for the transcripts that are cleaved in absence of RicF/YlbF. Further studies will be needed to determine what transcript elements render RicF/YlbF unnecessary.

Cleavage efficiency factors.

Upstream factors.

Deleting the 12 nucleotides of sector II (leaving the G immediately upstream of the cleavage site) prevented cleavage (ΔII+G in Figure 4). We considered the possibility that this was because the sector II deletion brings the putative secondary RNA structure of sector I very close to the cleavage site, which potentially could prevent RNase Y access. However, when we removed the secondary structure by deleting both sector I and sector II, we did not restore cleavage, leaving us to conclude that the sequence immediately upstream of the cleavage site (i.e. sector II) is a key element for efficient cleavage by RNase Y.

Mutations in the sequence immediately downstream of the cleavage site can fine-tune efficiency.

Endoribonucleolytic cleavages can be maturation events, where the only objective of the cleavage is to obtain a final cleaved and stable RNA, or homeostasis-maintaining events, where the cleavage event switches the RNA from one function to another (or initiates the degradation of one or both of the generated fragments). Maturation events are expected to be highly efficient to avoid wasting energy on non-functional immature RNAs, whereas homeostasis-maintaining cleavage events are expected to be less efficient, since both uncleaved and cleaved RNA have a function for the cell. In our various Northern blots from wild-type cells, we are always able to observe both the uncleaved (full-length) and cleaved transcripts (Figures 1C and 2), suggesting that these cleavages are maintaining homeostasis. The ratio between full-length and cleavage product for the *Sa-gapR* transcript appears to have evolved to be in favour of the full-length (Figure 1C), and the native trinucleotide AGA found immediately downstream of the pSaGap cleavage site is indeed among the sequences that give the lowest cleavage efficiency (Figure 9B).

Transcripts with sector III trinucleotides that start with G (i.e. GNN) all appear to be cleaved with poor efficiency (Figure 9B). We do not believe that this is due to bias in the EMOTE protocol, since we have never observed an anti-G bias in the many other EMOTE assays carried out in our lab over the years. It is possible that the GNN trinucleotides shift the cleavage position one nucleotide downstream, which will prevent the EMOTE assay from attributing the cleaved molecule to a specific trinucleotide sequence. However, a Northern blot analysis of pSaGap[IIIGGA] revealed a cleavage product with low intensity, supporting the EMOTE data (Figure 9C), and we therefore conclude that a guanidine immediately after the cleavage position lowers the efficiency of cleavage by RNase Y.

The only trinucleotides that result in lower cleavage efficiency than AGA are CGA, GGU, AUU and GUU. In contrast, there are many trinucleotides that increase cleavage efficiency, with

some combinations (AUA, CUA, CUC, CUG and UUG) reaching >9-fold higher than the wild-type sequence. A common factor for these efficiently cleaved transcripts is that they contain a trinucleotide with U at the second position (NUN). However, this pattern is then modified by other factors, such as G at the first position (GUN trinucleotides are not efficiently cleaved) and whether the trinucleotide can extend the putative hairpin (e.g. AUU and GUU are the most inefficiently cleaved transcripts, presumably because they can both extend the putative hairpin).

Secondary structure downstream of cleavage site.

Our mutational analyses of the putative hairpins in the pSaGap, pBsCgg and pSaGln transcripts revealed that the putative G-C base-pairs are crucial for cleavage efficiency, but that the orientation of the G-C base-pairs appear to be of little importance (Figure 7). This suggests that in order to be cleaved, the RNA needs to form a secondary structure that is strengthened by these G-C base-pairs, and it is therefore highly likely that the putative hairpins form *in vivo*, and that they are required for RNase Y to cleave the RNA molecules where they form.

Although the G-C pair in the pGap transcript can be inverted without affecting cleavage, inverting the remaining base-pairs of the hairpin stem almost completely abolishes cleavage (Figure 8), indicating that a distinct primary nucleotide sequence is required for RNase Y cleavage, and that this distinct sequence must furthermore be in the context of a double-stranded RNA. The loop-sequence of the hairpin does not appear to be of importance, since inversion of the loop does not detectably affect cleavage efficiency (Figure S6B and C). Elongation of the stem by two base-pairs has no detectable effect on cleavage efficiency, but if the stem is extended by three base-pairs, then cleavage efficiency is reduced drastically (Figures 9B and 9D) (Note that these effects could also be directly due to the changes in the primary sequence needed to obtain such stem extensions).

However, our reporter system is set up in a living cell, which is a dynamic system where the stem duplex might continuously form and then dissociate (and this is especially true with energetically weak duplexes like the stem of the sector IV hairpin in the *Sa-gapR* transcript). A given RNA molecule might therefore be cleaved or not depending on the current state of the duplex, and the duplex/single-strand equilibrium may have evolved to ensure the correct cleavage efficiency for the RNA.

Redundancy between sectors V and VI.

Deleting either sector V or sector VI individually had no detectable impact on cleavage efficiency. However, when both sectors were deleted, then it became impossible to detect any cleavage products by Northern blotting, and it required the highly sensitive EMOTE assay to detect any cleavage at all. It thus appears that sector V and VI both contain sequence elements that promote RNase Y cleavage, but that these elements are mutually redundant.

Cleavage positioning factors.

The upstream nucleotide fine-tunes the site selectivity.

A guanosine has been observed upstream of most RNase Y cleavage sites in both *S. aureus* and *S. pyogenes*. However, our results show that (at least for the *Sa-gapR* transcript) this is

only a preference and not a strict requirement, and that the main role of the guanosine is to direct all RNase Y cleavage events towards the same phosphodiester bond (Figure 3). Such precision is presumably advantageous for RNA maturation and homeostasis cleavages, where the two resulting RNA cleavage products should have a sequence that protects them (or not) against exoribonucleases.

A reference-point for a molecular ruler mechanism is located in sector IV.

Deletion of the three nucleotides in sector III caused the cleavage site to shift upstream by four nucleotides (Figure 4), indicating that the RNase Y cleavage site is selected by measuring a defined distance from a specific reference point downstream of the cleavage site. Within a margin of 1 nt, this consistent with the molecular ruler mechanism proposed by Marincola and co-workers (2017). The 1 nt difference could be due to an imprecise ruler mechanism and/or to the presence of a non-preferred uridine at three nucleotides upstream of the native site (position +255). The reference point for the molecular ruler must be within sector IV (since deletion of sector V does not shift the cleavage position, see Figure 4C). When the putative hairpin in sector IV was replaced by the putative hairpin from the *Bs-cggR* operon (pSaGap[cggHP]), then the major cleavage site was shifted 2 nts downstream to the native cleavage site of the *Bs-cggR* transcript (green line in Figure 6E), strongly suggesting that elements within the *Sa-gapR* and *Bs-cggR* hairpins serve as reference point for the ruler mechanism. We moreover analysed EMOTE data from variations of sector III sequences that are predicted to move the stem-bottom one or two nucleotides upstream (NVU and YUU trinucleotides, respectively). All of these exhibited a clear major cleavage position at the native position, with no secondary cleavage positions (Figure S8), and we conclude that the transition from single stranded to duplex RNA is not the reference point for the molecular ruler mechanism.

The composition of the hairpin stem also affects cleavage position. As mentioned in the previous section, inverting the stem almost completely removes cleavage, but the weak remaining cleavage is a split between the native cleavage site, and a site two nucleotides downstream (Figure 8B). This effect also manifests when a single “U” in the downstream “arm” is mutated to an “A” (pSaGap[U282A], Figure 8). However, when both the “U” and opposed “A” are mutated to regain a potential for base-pairing (pSaGap[A264U,U282A]), then RNase Y cleavage site is once again almost exclusively at the native site, which points to the downstream “hairpin-arm” as the reference point for the molecular ruler mechanism, with U282 as one of the key elements.

Finally, while this study strongly suggests that positioning is determined by a molecular ruler mechanism that measures from a feature in sector IV, we note that when sector IV is deleted we nevertheless observe a very weak residual RNase Y-dependent cleavage at the native position (Figure 4C). These events are only detectable with the highly sensitive EMOTE method, but their existence indicate that there is partial redundancy between sector IV and one or more of the other sectors with regards to cleavage positioning.

Seen in the light of global RNA decay, the precise positioning of an endoribonucleolytic cleavage can seem unimportant, since almost any cleavage will accomplish the goal of providing access-points for exoribonucleases to finish the degradation. Such low fidelity was indeed observed for *E. coli* RNase E, where the cleavage motif is gn¦wUu and each RNA molecule is cleaved multiple times (Chao et al. 2017). RNase Y appears to have more stringent cleavage criteria, since transcriptomic studies in *S. aureus*, *B. subtilis* and *S. pyogenes* identify

a limited number of RNase Y cleavage positions (99, 27 and 190 positions, respectively) (Khemici et al. 2015; DeLoughery et al. 2018; Broglia et al. 2020). Nevertheless, a recent study showed that some of the phenotypes of a *B. subtilis* RNase Y mutant can be complemented by expressing RNase E from *E. coli*, and that this complementation was highly dependent on anchoring RNase E to the membrane (Laalami et al. 2021). Importantly, these complemented strains were unable to properly degrade the *Bs-cggR* transcript and no version of RNase E could generate the stable downstream cleavage product observed in this and other studies (Meinken et al. 2003; Commichau et al. 2009; Laalami et al. 2021). This suggests that RNase E can only complement the general RNA decay function of RNase Y, leading to the partial rescue of the *Δrny* growth defects observed by Laalami and co-workers (2021). For normal efficient growth however, it appears that *B. subtilis* requires the type of precise cleavage events that we examine in this study.

Designing modules for use in synthetic biology.

Biotechnological applications that seek to optimise product synthesis require tuneable and modular elements for gene expression regulation. For example, several RNA sequence modules have been developed for the yeast ribonuclease Rnt1p, in order to precisely control protein expression levels (Babiskin and Smolke 2011). In a similar manner, introducing the RNase Y cleavage site in a transcript could prove useful in bacterial synthetic biology applications, where precise cleavage of the transcript can be designed to prevent translation, generate a small regulatory RNA or, as is the case in both the *Sa-gapR* and *Sa-glnR* operons, differentially express proteins encoded on the same transcript. Here we show that sectors II-V generate an *in vivo* RNase Y cleavage site (Figure 5), and that *in vivo* cleavage efficiency can be fine-tuned by single nucleotide changes in sector III (Figure 9). Combining these observations should make it possible to generate modules that can optimise gene expression profiles for synthetic biology projects in *B. subtilis*, *S. aureus* and possibly other bacteria that carry RNase Y.

Materials and Methods:

Growth of strains and media

Unless stated otherwise, *S. aureus* was grown in MHB liquid medium (Mueller Hinton Broth, Becton Dickinson) or agar plates at 37°C. MH medium was systematically supplemented with uracil (20 mg/l) as the parental strain PR01 is Δ*pyrFE* (Redder and Linder 2012), and 10 mg/l chloramphenicol was used to select for transformants and maintain vectors. *Escherichia coli* was grown in LB medium with 100 mg/l ampicillin to select for plasmids.

Vector design and construction

The *Sa-gap/Bs-cgg* and *Sa-gln/Bs-gln* operon pairs were identified as evolutionary conserved RNase Y targets by aligning (using protein BLAST) the proteins encoded downstream of the RNase Y cleavage sites described by DeLoughery and co-workers (DeLoughery et al. 2018) with the translated *S. aureus* genome. The genome positions of the *S. aureus* hits were then

compared to the list of *S. aureus* RNase Y cleavage sites described by Khemici and co-workers (Khemici et al. 2015).

All expression vectors constructed in this study are based on plasmid pEB01, an *E. coli-S. aureus* shuttle-vector which carries a transcription terminator sequence downstream of the polylinker (Oun et al. 2013). pSaGap and pBsCgg were constructed by PCR-amplifying the regions indicated in Figure 1 with primers that added restriction sites for SalI, KpnI and AscI (See Table S2). After digestion with the relevant restriction enzymes, the inserts were ligated into the SalI and AscI sites of pEB01 and transformed into *E. coli*. pGlnSa and pGlnBs were constructed using In-Fusion cloning assembly (Takara). Specific modifications in the vector sequences were introduced via divergent PCR with 5' phosphorylated primers, followed by blunt-end ligation with T4 DNA ligase (Thermo Fischer) and transformation into *E. coli*. All vectors-inserts were verified by Sanger sequencing before transformation into *S. aureus* strains WT and ΔY (and BsY where relevant). Please see Table S1 for strain-list.

Constructing random library of Sector III

The three base-pairs of pSaGap Sector III were randomised by polymerase chain reaction (PCR) using the pSaGap construct as template and primers 634 and 636 listed in Table S2. The resulting plasmid library of 64 (4^3) potential variants was transformed first into *E. coli* DH5α then into *S. aureus* WT (generating WT+pSaGap[IIIrandom]). More than 6400 CFUs were recovered from each transformation to ensure that all 64 variants are represented in the library.

RNA extraction

S. aureus cultures were inoculated from overnight cultures prepared from a single colony and diluted at 1/100 into fresh medium, 8 ml culture were harvested at OD_{600}~0.4 and immediately mixed with 40 mL ice cold ethanol–acetone mix (1:1). After centrifugation (10 min at 3500g), the cells were washed once in 1ml TE buffer (10 mM Tris pH8, 1mM EDTA) and suspended in a mix of 90 µL TE buffer + 10 µL lysostaphin 35520 U/ml (EUROMEDEX) + 2,5 µL RNasin Plus 40 U/µl (Promega). Samples were incubated at 37°C for 10 min before addition of 300 µl TRI-Reagent (MRC). A Phase Lock Gel tube (5Prime) was used to obtain a pure aqueous phase, in which 20 µg of glycogen (Roche) and 150 µl isopropanol was added to precipitate the RNA over-night at -20°C. RNA pellets were washed with 900µL cold 75% ethanol, re-suspended in 20µL 0.1X TE and stored at -80°C.

Analysing RNase Y cleavage of ectopic transcripts

The cleavage (or lack thereof) was studied on total RNA isolated from *S. aureus* strains carrying vectors expressing the various transcripts of interest (see Table 1). Northern blotting was carried out as described in (Khemici et al. 2015), and the targeted EMOTE was a variation of the EMOTE protocol described in (Redder 2018b).

Northern blotting

Briefly, 4µg total RNA were separated by PAGE in a 6% polyacrylamide 8M urea gel, transferred to a Hybond-N+ membrane (Amersham), then UV crosslinked 120 J/cm² (Hoefer, UVC 500). The size marker (RiboRuler Low range, Thermo Scientific) as well as ribosomal RNAs were visualised with Methylene blue. Blots were hybridized with ^{32}P labelled synthetic DNA probes P1, P2 or 5S (Table S2) and revealed using a phosphorimager (Fuji).

Targeted EMOTE

Targeted EMOTE is a variation of the original EMOTE described in detail previously (Redder 2018b). Specificities of the targeted EMOTE are described here.

Ligation was carried out using 4 ug of total RNA with 1 ul of 100uM BioRp8 in a 10 ul volume and placed at 85°C for 2 minutes and flash cooled in ice water. A mix of 3µl water, 2 µl of 10X RNA ligase buffer (NEB), 2 µl 10 mM ATP, 1 µl of RNasinPlus 40 U/µl (Promega), 1µl RNA ligase 10 000 U/µl (NEB) and 1 µl RppH 5000 U/µl (NEB) was added to the RNA/oligo mix. Ligation was performed in the presence of RppH in order to detect the primary transcripts for normalisation purposes. The ligation mix was placed for 3 hours at 37°C. RNA from the sample was purified by ethanol precipitation.
Reverse transcription was carried out as described in (Redder., 2018) using the plasmid-specific pEB-TT-R3-illuB primer to specifically reverse transcribe the exogenous transcripts.

Illumina sequencing and analyses of EMOTE data

Illumina sequencing consisted of 50 nucleotide single-end sequencing, except for the results presented in Figure 9 where 100 nt reads were required. The sequencing was carried out at the genomics platform of University of Geneva (https://lims.ige3.genomics.unige.ch/). An R script was used to process the raw sequencing data: Each read was demultiplexed, using EMOTE barcodes to assign each read to its original sample. Reads were then mapped with Bowtie onto the sequence of the vector that had provided the examined transcript (for example pSaGap, pBsCgg, pSaGap[InvStem], etc.). A table was then generated which indicates the exact 5' nucleotide position of each detected RNA molecule (Prados et al. 2016).

In order to avoid bias generated by preferential PCR amplification of one cDNA over another, the EMOTE protocol employs a Unique Molecular Identifier (UMI) in the oligo that is ligated to the 5' ends of the RNA samples. The UMI has 2187 (3^7) possible combinations, and since all experiments generated > 5000 reads for a given RNA cleavage (and usually much more than 5000 reads), we used a subsampling (also referred to as down-sampling) approach to get below the saturation level of the UMI. Briefly, for each position analysed in this work, if the number of UMIs was >150, data was subsampled by a 'dilution factor' until the output was in the 50-150 interval. Subsampled UMI counts were then multiplied by the "dilution factor" in order to get a calculated UMI count for the original sample. This UMI count then represents the number of detected individual RNA molecules with a given 5' end.

For each EMOTE graphical representation presented, we calculated the total number of detected RNA molecules with a 5' end within the shown bracket. Each column indicates the percentage of individual RNA molecules with a 5' end at the given position. A tall column for nucleotide 'X' therefore indicates that 'X' was the 5' nucleotide in a majority of RNA molecules.

Analysis of trinucleotide variants in Sector III

Total DNA was isolated from *S. aureus* WT cultures carrying the pSaGap[IIIrandom], and PCR was used to amplify the trinucleotide locus from the library cultures (in two biological replicates). The PCR products were Illumina sequenced and the relative abundances of each of the 64 expression vectors in our libraries were determined by counting the number of reads with each trinucleotide combination (AAA, AAC, etc.).
Total RNA was also isolated from *S. aureus* WT cultures carrying the pSaGap[IIIrandom], and used for targeted EMOTE sequencing (in four biological replicates). The EMOTE sequencing reads were sorted into 64 separate files according to the trinucleotide they contain (AAA, AAC, etc.). For each set of reads, the calculated number of individual RNA molecules (see above) at the native RNase Y cleavage position (+259) was normalised with the relative abundance of each corresponding expressing vector in the library (e.g. number of RNA molecules with AAA / number of vectors with AAA).
A Bilateral paired Student test (T-Test) was then performed to determine whether RNase Y cleaved the transcript with a given trinucleotide with a different efficiency than the transcript with a WT trinucleotide sequence (AGA). The calculated p-values were then adjusted with Bonferroni correction for multiple testing.

Acknowledgments

We would like to thank Manuel Campos for scientific discussions, Pierre Genevaux and Raffaele Ieva for critical reading of the manuscript, and the iGE3 Genomics Platform at University of Geneva, Switzerland, for help and patience with development of novel sequencing protocols. This work was supported by grants from Société Académique de Genève and Axes Thématiques Prioritaires de l'Université Paul Sabatier.

Data availability

The Illumina sequencing data from this study are available in the SRA database, accession number PRJNA1005908.

References

Anderson KL, Roux CM, Olson MW, et al (2010) Characterizing the effects of inorganic acid and alkaline shock on the Staphylococcus aureus transcriptome and messenger RNA turnover. FEMS Immunol Med Microbiol 60:208–250. https://doi.org/10.1111/j.1574-695X.2010.00736.x

Babiskin AH, Smolke CD (2011) A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. Mol Syst Biol 7:471. https://doi.org/10.1038/msb.2011.4

Broglia L, Lécrivain A-L, Renault TT, et al (2020) An RNA-seq based comparative approach reveals the transcriptome-wide interplay between 3'-to-5' exoRNases and RNase Y. Nat Commun 11:1587. https://doi.org/10.1038/s41467-020-15387-6

Chao Y, Li L, Girodat D, et al (2017) In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. Mol Cell 65:39–51. https://doi.org/10.1016/j.molcel.2016.11.002

Commichau FM, Rothe FM, Herzberg C, et al (2009) Novel activities of glycolytic enzymes in Bacillus subtilis: interactions with essential proteins involved in mRNA processing. Mol Cell Proteomics 8:1350–1360. https://doi.org/10.1074/mcp.M800546-MCP200

DeLoughery A, Lalanne J-B, Losick R, Li G-W (2018) Maturation of polycistronic mRNAs by the endoribonuclease RNase Y and its associated Y-complex in Bacillus subtilis. Proc Natl Acad Sci USA 115:E5585–E5594. https://doi.org/10.1073/pnas.1803283115

Durand S, Braun F, Lioliou E, et al (2015) A nitric oxide regulated small RNA controls expression of genes involved in redox homeostasis in Bacillus subtilis. PLoS Genet 11:e1004957. https://doi.org/10.1371/journal.pgen.1004957

Figaro S, Durand S, Gilet L, et al (2013) Bacillus subtilis mutants with knockouts of the genes encoding ribonucleases RNase Y and RNase J1 are viable, with major defects in cell morphology, sporulation, and competence. J Bacteriol 195:2340–2348. https://doi.org/10.1128/JB.00164-13

Giraud C, Hausmann S, Lemeille S, et al (2015) The C-terminal region of the RNA helicase CshA is required for the interaction with the degradosome and turnover of bulk RNA in the opportunistic pathogen Staphylococcus aureus. RNA Biol 12:658–674. https://doi.org/10.1080/15476286.2015.1035505

Hamouche L, Poljak L, Carpousis AJ (2021) Ribosomal RNA degradation induced by the bacterial RNA polymerase inhibitor rifampicin. RNA 27:946–958. https://doi.org/10.1261/rna.078776.121

Kaito C, Kurokawa K, Matsumoto Y, et al (2005) Silkworm pathogenic bacteria infection model for identification of novel virulence genes. Mol Microbiol 56:934–944. https://doi.org/10.1111/j.1365-2958.2005.04596.x

Khemici V, Prados J, Linder P, Redder P (2015) Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation. PLoS Genet 11:e1005577. https://doi.org/10.1371/journal.pgen.1005577

Laalami S, Cavaiuolo M, Roque S, et al (2021) Escherichia coli RNase E can efficiently replace RNase Y in Bacillus subtilis. Nucleic Acids Res 49:4643–4654. https://doi.org/10.1093/nar/gkab216

Laalami S, Zig L, Putzer H (2014) Initiation of mRNA decay in bacteria. Cell Mol Life Sci 71:1799–1828. https://doi.org/10.1007/s00018-013-1472-4

Lehnik-Habrink M, Newman J, Rothe FM, et al (2011) RNase Y in Bacillus subtilis: a Natively disordered protein that is the functional equivalent of RNase E from Escherichia coli. J Bacteriol 193:5431–5441. https://doi.org/10.1128/JB.05500-11

Lehnik-Habrink M, Pförtner H, Rempeters L, et al (2010) The RNA degradosome in Bacillus subtilis: identification of CshA as the major RNA helicase in the multiprotein complex. Mol Microbiol 77:958–971. https://doi.org/10.1111/j.1365-2958.2010.07264.x

Marincola G, Schäfer T, Behler J, et al (2012) RNase Y of Staphylococcus aureus and its role in the activation of virulence genes. Mol Microbiol 85:817–832. https://doi.org/10.1111/j.1365-2958.2012.08144.x

Marincola G, Wolz C (2017) Downstream element determines RNase Y cleavage of the saePQRS operon in Staphylococcus aureus. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx296

Meinken C, Blencke H-M, Ludwig H, Stülke J (2003) Expression of the glycolytic gapA operon in Bacillus subtilis: differential syntheses of proteins encoded by the operon. Microbiology (Reading) 149:751–761. https://doi.org/10.1099/mic.0.26078-0

Oun S, Redder P, Didier J-P, et al (2013) The CshA DEAD-box RNA helicase is important for quorum sensing control in Staphylococcus aureus. RNA Biol 10:157–165. https://doi.org/10.4161/rna.22899

Redder P (2018a) Molecular and genetic interactions of the RNA degradation machineries in Firmicute bacteria. Wiley Interdiscip Rev RNA. https://doi.org/10.1002/wrna.1460

Redder P (2018b) Mapping 5'-Ends and Their Phosphorylation State With EMOTE, TSS-EMOTE, and nEMOTE. Meth Enzymol 612:361–391. https://doi.org/10.1016/bs.mie.2018.08.011

Redder P, Linder P (2012) New range of vectors with a stringent 5-fluoroorotic acid-based counterselection system for generating mutants by allelic replacement in Staphylococcus aureus. Appl Environ Microbiol 78:3846–3854. https://doi.org/10.1128/AEM.00202-12

Roux CM, DeMuth JP, Dunman PM (2011) Characterization of components of the Staphylococcus aureus mRNA degradosome holoenzyme-like complex. J Bacteriol 193:5520–5526. https://doi.org/10.1128/JB.05485-11

Shahbabian K, Jamalli A, Zig L, Putzer H (2009) RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. EMBO J 28:3523–3533. https://doi.org/10.1038/emboj.2009.283

Steglich C, Lindell D, Futschik M, et al (2010) Short RNA half-lives in the slow-growing marine cyanobacterium Prochlorococcus. Genome Biology 11:R54. https://doi.org/10.1186/gb-2010-11-5-r54

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415. https://doi.org/10.1093/nar/gkg595