The EMBO Journal (2012) 31, 1961–1974 | © 2012 European Molecular Biology Organization | All Rights Reserved 0261-4189/12
www.embojournal.org

THE EMBO JOURNAL

Multiple factors dictate target selection by Hfq-binding small RNAs

Chase L Beisel[1,]*, Taylor B Updegrove, Ben J Janson and Gisela Storz*

Cell Biology and Metabolism Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA

Hfq-binding small RNAs (sRNAs) in bacteria modulate the stability and translational efficiency of target mRNAs through limited base-pairing interactions. While these sRNAs are known to regulate numerous mRNAs as part of stress responses, what distinguishes targets and non-targets among the mRNAs predicted to base pair with Hfq-binding sRNAs is poorly understood. Using the Hfq-binding sRNA Spot 42 of *Escherichia coli* as a model, we found that predictions using only the three unstructured regions of Spot 42 substantially improved the identification of previously known and novel Spot 42 targets. Furthermore, increasing the extent of base-pairing in single or multiple base-pairing regions improved the strength of regulation, but only for the unstructured regions of Spot 42. We also found that non-targets predicted to base pair with Spot 42 lacked an Hfq-binding site, folded into a secondary structure that occluded the Spot 42 targeting site, or had overlapping Hfq-binding and targeting sites. By modifying these features, we could impart Spot 42 regulation on non-target mRNAs. Our results thus provide valuable insights into the requirements for target selection by sRNAs.

The EMBO Journal (2012) **31,** 1961–1974. doi:10.1038/emboj.2012.52; Published online 2 March 2012
Subject Categories: RNA
Keywords: base-pairing; *Escherichia coli*; Spot 42; target prediction

Introduction

Small RNAs (sRNAs) are critical regulators of bacterial responses to changes in the environment (Waters and Storz, 2009). sRNAs act via a range of mechanisms, including base-pairing with mRNAs and modulating the activity of proteins. Most sRNAs form limited base-pairing interactions with mRNAs to modulate mRNA stability and translation. In enteric bacteria, gene regulation by this prevalent class of sRNAs (dubbed Hfq-binding sRNAs) requires the binding activity of the RNA chaperone protein Hfq. Hfq protects unpaired Hfq-binding sRNAs from RNase attack, melts intramolecular structures and facilitates base-pairing interactions between bound sRNAs and mRNAs, and in addition, recruits RNases to degrade base-paired sRNA:mRNA complexes (Vogel and Luisi, 2011). Despite the capacity for positive and negative regulation, Hfq-binding sRNAs predominantly repress gene expression.

Ongoing characterization of Hfq-binding sRNAs has revealed that sRNAs often regulate multiple mRNAs as part of environmental responses (Storz *et al*, 2011). For instance, the sRNA GcvB directly represses the expression of at least 21 genes when amino acids are overabundant (Urbanowski *et al*, 2000; Pulvermacher *et al*, 2009a, b; Sharma *et al*, 2007, 2011), while the σ^E-regulated sRNAs RybB and MicA together directly repress the expression of at least 23 genes in response to membrane stress (Papenfort *et al*, 2006, 2010; Gogol *et al*, 2011). Within a given set of sRNA targets, the strength of regulation varies considerably; studies using translational reporter fusions with known targets have reported strengths of regulation varying from less than two-fold to over 40-fold (De Lay and Gottesman, 2009; Durand and Storz, 2010; Sharma *et al*, 2011; Beisel and Storz, 2011a).

The factors that separate strongly regulated targets, weakly regulated targets, and non-targets of sRNAs are only beginning to be understood. One emerging factor is Hfq binding to both sRNAs and mRNAs. Recent crystal structures of the donut-shaped Hfq hexamer have indicated that two distinct surfaces (the proximal side and the distal side) bind specific sequences in sRNAs and mRNAs (Schumacher *et al*, 2002; Link *et al*, 2009). The proximal side of Hfq binds U-rich sequences often present in sRNAs, while the distal side binds repeats of the triplet A-R-N (R is a purine and N is any ribonucleotide). The few target mRNAs in which Hfq binding has been demonstrated contain binding sites for the distal side of Hfq (e.g., *ompA*, *ompD*, *rpoS*, *sodB*) (Moll *et al*, 2003; Geissmann and Touati, 2004; Soper and Woodson, 2008; Pfeiffer *et al*, 2009), although it remains unclear whether all mRNAs must bind Hfq to be regulated by Hfq-binding sRNAs.

Another factor required for regulation by Hfq-binding sRNAs is base-pairing between sRNAs and target mRNAs. sRNA-based regulation may require as few as six base pairs (Kawamoto *et al*, 2006) but interactions in excess of 40 base pairs have been predicted (Møller *et al*, 2002). Within sRNAs, base-pairing regions tend to be highly conserved and unstructured (Peer and Margalit, 2011). Some sRNAs such as RybB appear to contain only one base-pairing region (Papenfort *et al*, 2010), while other sRNAs such as FnrS, GcvB, and Spot 42 contain multiple base-pairing regions (Durand and Storz, 2010; Sharma *et al*, 2011; Beisel and Storz, 2011a). Within target mRNAs, base-pairing regions (or sRNA targeting sites) often are concentrated around the ribosome-binding site and start codon, although targeting sites have been identified substantially upstream of the ribosome-binding site or in the coding sequence (Sharma *et al*, 2007; Bouvier *et al*, 2008; Pfeiffer *et al*, 2009).

*Corresponding author. CL Beisel or G Storz, Cell Biology and Metabolism Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, 18 Library Drive, MSC 5430, Bethesda, MD 20892-5430, USA. Tel.: +1 301 402 0968;
Fax: +1 301 402 0078; E-mail: cbeisel@ncsu.edu or storz@helix.nih.gov
[1]Present address: Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695-7905, USA.

Received: 12 September 2011; accepted: 8 February 2012; published online: 2 March 2012

Various algorithms have been developed for the prediction of genomic targets of Hfq-binding sRNAs. The first available algorithm (*TargetRNA*) scores target genes based on the extent of base-pairing (Tjaden, 2008). More recent algorithms (*intaRNA*, *RNApredator*) also account for the energetic cost of disrupting secondary structures within sRNAs and target mRNAs (Busch *et al*, 2008; Eggenhofer *et al*, 2011). sRNA target prediction algorithms have the capacity to identify known sRNA targets, although the number of known targets represents a mere fraction of the total number predicted by these algorithms. Assessing the accuracy of these prediction algorithms and identifying factors important in target regulation is critical for improving target prediction, understanding the evolution of sRNA regulatory networks, and advancing the design of synthetic regulatory RNAs.

Here, we employed the Hfq-binding sRNA Spot 42 to elucidate factors that define targets of this sRNA. Spot 42 is upregulated in the presence of a preferred carbon source and represses numerous metabolic genes through three of its conserved, unstructured regions (Møller *et al*, 2002; Beisel and Storz, 2011a). Using these regions of Spot 42 in *TargetRNA* predictions followed by assays of reporter fusions, we found that only the unstructured regions of Spot 42 contributed to regulation. We additionally found that non-targets predicted to base pair with the unstructured regions of Spot 42 lacked an Hfq-binding site, folded into a secondary structure that occluded the Spot 42 targeting site, or had overlapping Hfq-binding and Spot 42 targeting sites. Our results reveal critical factors for identifying targets of Hfq-binding sRNAs and begin to establish core principles underlying strong regulation by sRNAs.

Results

Computational search using the unstructured regions of Spot 42 reveals additional targets

We began by searching for mRNAs that potentially base pair with Spot 42. Using *TargetRNA* with a standard parameter set, we generated a list of the 10 top-scoring mRNAs containing putative Spot 42 targeting sites within 45 nucleotides upstream and 25 nucleotides downstream of annotated start codons (Table I). This search yielded one known target, *galK*, in line with the original identification of this target based on its extensive complementarity to Spot 42 (Møller *et al*, 2002). To assess whether any of the other nine genes are regulated by Spot 42, we fused the annotated 5′ end (or at least 200 nucleotides upstream of the start codon for genes encoded in operons) through the ~14th codon of each gene to a *lacZ* reporter. Overexpression of Spot 42 led to repression of two of the 10 reporter fusions (*galK*, *puuE*) beyond what was observed for an empty plasmid (>1.2-fold) (Table I). Negligible repression of the other eight reporters by Spot 42 suggests that these genes are not targets, although the possibility exists that generation of the *lacZ* fusions compromised regulation. These results indicate that *TargetRNA* can identify genes regulated by Spot 42, albeit with low accuracy.

In our previous characterization of Spot 42, we found that genes repressed following Spot 42 overexpression were predicted to base pair with three regions (I–III) of Spot 42 (Figure 1A) (Beisel and Storz, 2011a). These same regions were predicted to base pair with the *galK* (regions II and III) and *puuE* (region III) mRNAs, where mutational analysis of the *puuE* fusion confirmed that region III is critical for regulation (Figure 1B). Secondary structure prediction and *in vitro* structural probing suggested that these three regions of Spot 42 are unstructured (Møller *et al*, 2002). The unstructured regions of sRNAs generally may be responsible for target regulation, as a recent bioinformatics analysis showed that the conserved, unstructured regions of Hfq-binding sRNAs tend to contribute to base-pairing with target mRNAs (Peer and Margalit, 2011). We hypothesized that utilizing these three unstructured regions of Spot 42 rather than the full-length sRNA would improve the accuracy of target prediction.

We repeated the target search using the unstructured regions of Spot 42 as the sole input into *TargetRNA*. We then considered the five top-scoring genes for each unstructured region (Table II). This list partially overlapped with the list using full-length Spot 42, where *galK* and *usg* were within the top five for region II. Of the 15 genes in Table II, three (*nanC*, *srlA*, *galK*) previously were shown to be direct targets of Spot 42 and one (*nanT*) was shown to be regulated by Spot 42 with no evidence for direct base-pairing (Møller *et al*, 2002; Beisel and Storz, 2011a).

We generated *lacZ* translational fusions with the 15 top-scoring genes and again performed β-galactosidase assays to assess repression by Spot 42. Ten of the 15 gene fusions listed in Table II were repressed following Spot 42 overexpression (compared to 2/10 fusions generated based on predictions with full-length Spot 42). The fusions showed varying basal levels of expression, which may reflect differences in mRNA levels and/or translation. We conducted mutational analysis on three of the regulated fusions (*ascF*, *nanT*, *fucP*) to determine whether the predicted base-pairing interactions are responsible for the observed regulation (Figure 1B–E). Mutations in the region implicated in base-pairing disrupted repression while compensatory mutations restored regulation, confirming that Spot 42 base pairs with these fusions through the predicted interactions. We additionally evaluated whether endogenous expression of Spot 42 can alter mRNA levels of these new target genes. Quantitative real-time PCR analysis was performed on WT and Δ*spf* cells grown in M9 minimal media supplemented with glucose to induce Spot 42 expression. Among the genes tested, two of the five regulated as *lacZ* fusions (*glpF*, *paaK*) were significantly upregulated in the Δ*spf* strain. The other three genes may be regulated at the level of translation or are not measurably regulated by Spot 42 under the conditions tested. In contrast, all three of the genes not regulated as *lacZ* fusions (*usg*, *moeA*, *entB*) were not upregulated in the Δ*spf* strain (Supplementary Figure S1). Together, these results demonstrate that focussing on the unstructured regions of sRNAs can improve the prediction of direct targets.

Increased base-pairing in the unstructured regions of Spot 42 strengthens regulation

Hfq-binding sRNAs display large differences in the strength of regulation of target mRNAs, even for targets regulated by the same sRNA. One explanation for the variation in regulatory strength is the extent of base-pairing, where more extensive base-pairing is thought to lead to increased regulation (Mitarai *et al*, 2007, 2009). We tested how increasing the extent of base-pairing affects regulation of three weakly regulated targets: *gltA* (region I), *srlA* (region II), and *fucP*

 ©2012 European Molecular Biology Organization

Table I Gene targets predicted by *TargetRNA* using the full length of Spot 42

Gene	Predicted base-pairing[a]	Basal levels (MU)[b]	Fold regulation[c]	Hfq binding[d]	Microarray[e]
1. *puuE*	20 UGUUCUAUCUUUCAGACCUUUU-ACU--UCACG--UAAUCGGA 57 ||:|| || ||||| |||||| ||| ||||| ||||||| +7 ACGAG-UACAAAGUGUGGAAAACUGAGAAGUGCCAAUUAGCCU −35	6510 ± 130	5.1 ± 0.1	Y	A
2. *yeaO*[f]	20 UGUUCUAUCUUUCA-----GACCUUUUACUUCAC-GUA--AUCGGAU-UUGGCUGAAU 68 |||||| :|||||| || | |||| |||| :|| |:|||| ||:|| |||| −7 ACAAGAGGGAAAGUGCGUACUAAACAAUGCAGUGAUAUCGUGGCCUAUAAUCG-CUUA −63	2800 ± 60	0.78 ± 0.01	Y	1.2 ± 0.5
3. *yjaH*	27 UCUUUCAGACCUUUUACU-UCACGUAAUCGGAUUUGGCUGAAUAUUUUAG 75 || || |:|| ||:||| :|||:|||| |:||::||| | |||:|| +17 AGCAAUUUUGACAAGUGACGGUGCGUUAGAAUGAAUUGACGCAGAAAGUC −33	ND	–	–	−1.2 ± 0.1
4. *usg*	19 AUGUUCUAUCUUUCAGACCU-----UUUACUUCACGUAAU--CGGAUUUGGCUG--AAUA 69 ||||||:| |:||||||| | ||||||:|| | ||| |::| :| ||| ||| +20 UACAAGGUCGGAAGUCUGUACCACCAAAUGAGGU-CUUUAUCGUUUCGAGGGACGGUUAC −39 70 UUUUAG 75 || :|| −40 AAUGUC − 45	961 ± 38	1.2 ± 0.0	–	NP
5. *galK*	20 UGUUCUAUCUUUCAGACCUUUUACUU-CACGUAAUCGGAUUUG 61 |||| | ||||||||| : || ||| ||| | :|||||:| +21 ACAAAAAAGAAAGUCUGAGUAAAGAAUGUG-A--GGCCUAAGC −19	2850 ± 90	2.3 ± 0.1	Y	−2.7 ± 0.5
6. *pntB*	26 AUCUUUCAGACCU-U-UUACUUCACGUAAUCGGAUUUGGCUGAAUAUUUU-AGC 76 |||:|:||||| | | |||::| |||| ||| || |||:||||| ||| +13 UAGGAGGUCUGUAUACAAUGGGGA--AUUA----AAAACGCCUUGUAAAAGUCG −35	78 ± 2	0.90 ± 0.04	–	−1.0 ± 0.1
7. *ykfG*	4 GGGUACAGAGGUAAGAUGUUCUAUCU 29 ||||| :||| ||| |||| :||| +24 CCCAUUUUUCCUUUCGACAAAGUAGA −2	1350 ± 20	0.74 ± 0.01	–	A
8. *ydhU*	30 UUCAGACCUUUUACUUCACGUAAUCGGAU-UUGGCUGAA 67 |||| || |||| ||| ||||::|| |||:|:||| +5 AAGUAUGUAAAAGGAAAA--AUUAGUUUACAACUGGCUU − 32	453 ± 21	0.69 ± 0.03	–	1.2 ± 0.1
9. *erpA*	26 AUCUUUCAGACCUUUUACUUCA--CGUAAUCGG-AUUUGGCUGAAUAUUUUAGCC 77 ||| :||| | |||| ||:| || ||||| | |||| | | |||:||:|| +9 UAGUGAGUAU--AAAACGAGGUUUGCUGUAGCCGUUAACC--CGUUUAAGAUUGG −42	76 ± 2	0.87 ± 0.02	Y	−1.0 ± 0.2
10. *yciS*	26 AUCUUUCAGACCUUUUACUUCACGUAAUCGGAUUUGGCUGAAUAUUUUA 74 || ||||| | ||||||| ||| || | || |: :||||| ||| +9 UAUAAAGUGU----AAUGAAGGGCAAUACACCAAUCUCGCUUAUUUAAU −36	ND	–	Y	1.1 ± 0.2

[a]Base-pairing predictions using the full-length Spot 42 sequence with the terminator deleted. Numbers designate the location in Spot 42 (top) or in each gene in relation to the start codon (bottom). Nucleotides highlighted in grey designate the unstructured regions of Spot 42 (Figure 1A).
[b]β-Galactosidase activity for cells containing the *lacZ* fusion and pSpot42 grown in the absence of IPTG. Values represent the average and standard deviation from measurements of three independent colonies. MU, Miller units; ND, no detectable β-galactosidase activity.
[c]Fold regulation defined as the ratio of β-galactosidase activities for cells containing the *lacZ* fusion and pSpot42 incubated in the absence or presence of IPTG.
[d]Genes whose mRNAs were enriched following Hfq co-immunoprecipitation (A Zhang, unpublished data). Y, enrichment score above 2.0 for at least one of the three tested conditions; —, enrichment score <2.0 for all tested conditions. Potential binding sites for the distal side of Hfq and the enrichment score for each gene are shown in Supplementary Table S1. Enrichment scores were computed as described in Supplementary data.
[e]Fold regulation of mRNA levels following Spot 42 overexpression reported previously (Beisel and Storz, 2011a). Reported values are the average and standard deviation from three independent experiments. Positive values represent activation and negative values represent repression. A, mRNA signal absent in any of the three experiments. NP, gene not probed on the microarray chip.
[f]NCBI RefSeq (used by *TargetRNA*) and the NCBI database entry for MG1655 (NC_000913) annotate different start codons for *yeaO*.

(region III). Specifically, we inserted up to six nucleotides in each *lacZ* fusion either upstream or downstream of the targeting site to extend the predicted base-pairing (Figure 2A–C). The inserted nucleotides extended base-pairing through either the remainder of the unstructured region or into the structured region of Spot 42.

We found that extending base-pairing through the remainder of the unstructured region substantially improved regulation (*gltA_L*, *srlA_R*, *fucP_R*). In contrast, extending base-pairing into the structured region did not improve regulation (*gltA_R*, *fucP_L*). For *srlA*, extending base-pairing into the structured region of Spot 42 (*srlA_L*) improved regulation less than what was observed when base-pairing was extended through the remainder of the unstructured region (*srlA_R*), although interpretation of this result is complicated by the necessity of having the start codon interrupt the extended pairing. For all constructs, the measured strength of regulation did not correlate with the predicted increase in free energy (Supplementary Figure S2), suggesting that regulation by Hfq-binding sRNAs is kinetically driven rather than thermodynamically driven. Overall, these results indicate that (i) the unstructured regions of Spot 42 are critical for target regulation and (ii) extending base-pairing through these regions and not the structured regions of Spot 42 can improve the strength of repression.

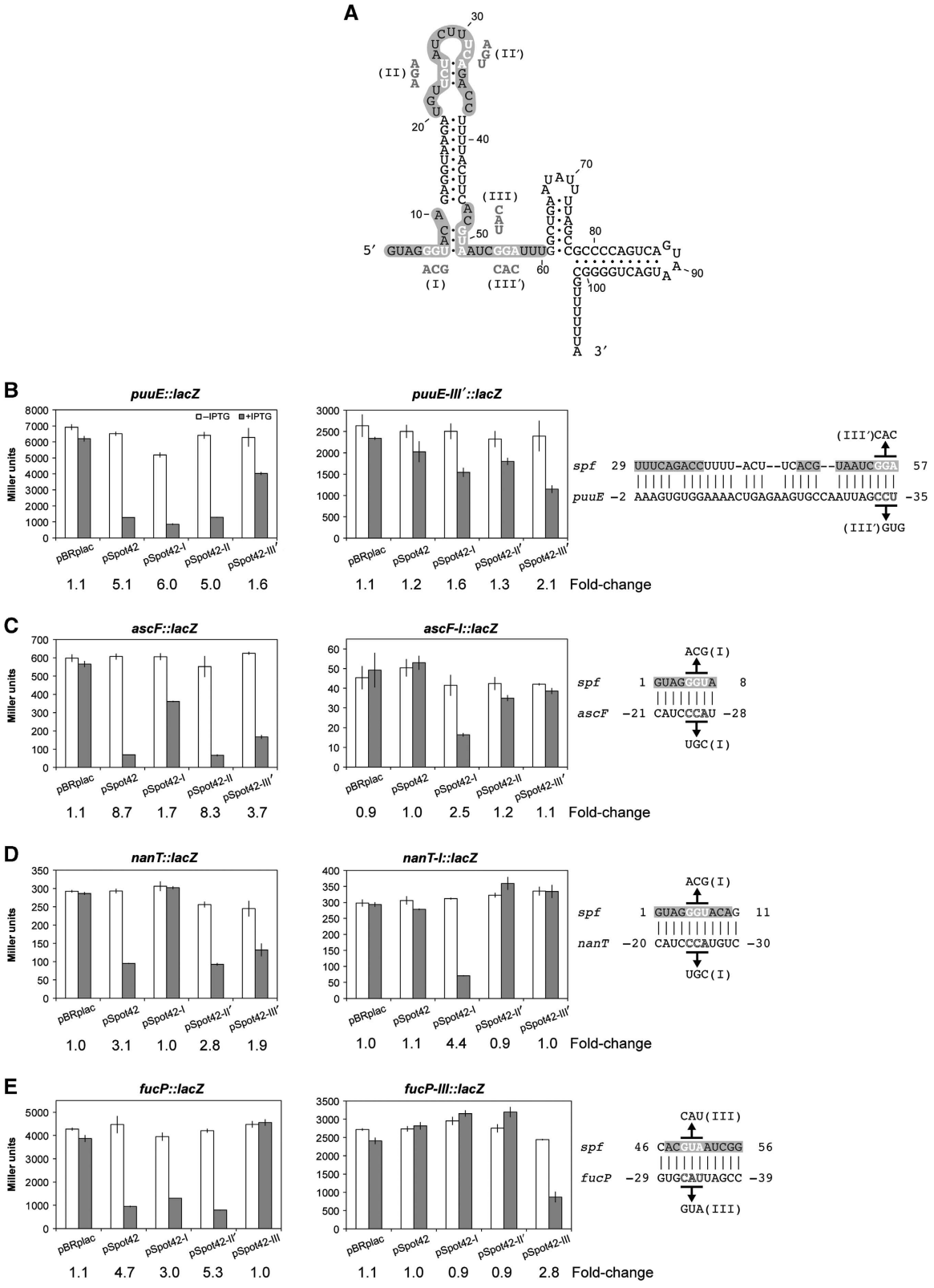
Base-pairing through multiple regions of Spot 42 strengthens regulation

Generally, individual unstructured regions of Hfq-binding sRNAs are involved in base-pairing interactions. However, for sRNAs with multiple unstructured regions, more than one region could be involved in base-pairing with individual mRNAs. To assess whether Spot 42 employs multiple unstructured regions to regulate individual targets, we employed *TargetRNA* and the folding algorithm *NUPACK* to identify genes containing more than one putative Spot 42 targeting site. We maintained two criteria: (i) *TargetRNA* predicts that two unstructured regions of Spot 42 each form at least six base pairs with the target mRNA and (ii) the folding algorithm *NUPACK* predicts the same base-pairing interactions as *TargetRNA*. Using this approach, we identified four target mRNAs that have the potential to base pair with two unstructured regions of Spot 42: *nanC* (regions I and III), *galK* (regions II and III), *sthA* (regions I and III), and *ascF* (regions I and III) (Figure 3A). Mutational analysis of Spot 42 and the *nanC*, *sthA*, and *ascF* fusions in this work and previous work supported multi-site pairing, as mutations in individual base-pairing sites partially reduced repression while mutations in both sites (one site mutated in Spot 42 and the other site mutated in the target fusion) eliminated regulation (Figures 1C and 3A) (Beisel and Storz, 2011a).

©2012 European Molecular Biology Organization

In most cases (e.g., *nanC*), one site predominantly contributed to regulation. We observed that deletion of *hfq* greatly compromised repression of the most strongly regulated target, *nanC*, suggesting that Hfq is required even when multiple sRNA targeting sites are present (Supplementary Figure S3).

To assess whether Spot 42 can base pair with these targets through two regions, we performed *in vitro* structural probing with RNase T1, lead, and RNase V1 on Spot 42 complexed with the *nanC* mRNA. The altered cleavage patterns in the presence of unlabelled *nanC* mRNA supported base-pairing between



 ©2012 European Molecular Biology Organization

regions I and III of Spot 42 and the *nanC* mRNA (Figure 3B). Altered cleavage also was observed outside of regions I and III, which may be attributed to more extended base-pairing and/or Spot 42 undergoing conformational changes upon pairing with the *nanC* mRNA. The cleavage pattern of radiolabelled *nanC* mRNA incubated with unlabelled Spot 42 similarly supported Spot 42 base-pairing with the two predicted targeting sites (Figure 3C). These results indicate that multiple unstructured regions of Spot 42 can base pair with multiple targeting sites in a particular mRNA.

Mutational analysis of the *nanC*, *sthA*, and *ascF* fusions demonstrated that the presence of an additional targeting site improved the strength of regulation. We thus asked whether regulation could be strengthened in single-site targets by introducing additional targeting sites for Spot 42. To address this, we focussed on the *srlA* and *fucP* fusions that only base pair with regions II and III of Spot 42, respectively (Figure 1E) (Beisel and Storz, 2011a). For the *srlA* fusion, we inserted 11 nucleotides that are complementary to region III of Spot 42 (*srlA + III*) upstream of the original targeting site (Figure 4A). For the *fucP* fusion, we mutated 11 nucleotides to be complementary to region I of Spot 42 (*fucP + I*) downstream of the original targeting site (Figure 4B). For both fusions, introduction of the additional targeting site substantially improved regulation from 2.8- to 27-fold for *srlA* and from 4.7- to 13-fold for *fucP*, an effect that was compromised in an

Table II Gene targets predicted by *TargetRNA* using the three unstructured regions of Spot 42

Gene	Predicted base-pairing[a]	Basal levels (MU)[b]	Fold regulation[c]	Hfq binding[d]	Microarray[e]
Region I	1 GUAGGGUACAG 11				
1. *nanT*	–20 CAUCCCAUGUC –30	293 ± 6	3.1 ± 0.0	–	–4.3 ± 1.7
2. *nanC*	–7 CAUCCUAUGUC –17	2210 ± 200	43 ± 4	Y	–3.5 ± 1.8
3. *paaK*	–2 CAUCCCAUG –10	759 ± 15	6.2 ± 0.7	Y	–1.6 ± 0.6
4. *atoD*	–29 CCCAUGUC –36	2270 ± 110	5.7 ± 0.8	Y	1.3 ± 0.5
5. *ascF*	–21 CAUCCCAU –28	607 ± 16	8.7 ± 0.1	Y	A
Region II	19 AUGUUCUAUCUUUCAGACCU 38				
1. *usg*	+20 UACAAGGUCGGAAGUCUG +3	961 ± 38	1.2 ± 0.0	–	NP
2. *entB*	–3 ACAAGAGAGGAAGUC –17	986 ± 59	0.94 ± 0.02	Y	1.2 ± 0.0
3. *galK*	+21 ACAAAAAAGAAAGUCUG +5	2850 ± 90	2.3 ± 0.1	Y	–2.7 ± 0.5
4. *srlA*	–1 ACAAGAGAGGAAGUC –15	7430 ± 340	2.8 ± 0.1	Y	–12.6 ± 15.9
5. *caiA*	+5 AGGUAGAAAGU-UGGA –10	4870 ± 430	1.6 ± 0.3	–	1.1 ± 0.2
Region III	41 CACGUAAUCGGAUUUG 56				
1. *fucP*	–29 GUGCAUUAGCCAAUAC –44	4470 ± 370	4.7 ± 0.5	–	–1.7 ± 0.6
2. *moeA*	–23 UUAGCUUAAAC –33	1500 ± 140	1.0 ± 0.1	–	–1.2 ± 0.4
3. *glpF*	+3 GUAUUAGCCUA –8	5950 ± 490	11.2 ± 0.5	Y	–1.7 ± 1.0
4. *lon*	–28 GUCCAUUAGUCUA –40	3250 ± 70	0.95 ± 0.03	–	1.1 ± 0.2
5. *lacY*	+4 UGUAUUA-CCUAAA –9	60 ± 7	0.72 ± 0.10	Y	NP

[a]Base-pairing predictions for the five top-scoring gene targets from *TargetRNA* using unstructured region I, II, or III of Spot 42. Nucleotides highlighted in grey designate the unstructured regions (Figure 1A). Numbers designate the location in Spot 42 (top) or in each gene in relation to the start codon (bottom).
[b]β-Galactosidase activity for cells containing the *lacZ* fusion and pSpot42 grown in the absence of IPTG. Values represent the average and standard deviation from measurements of three independent colonies. MU, Miller units.
[c]Fold regulation defined as the ratio of β-galactosidase activities for cells containing the *lacZ* fusion and pSpot42 incubated in the absence or presence of IPTG.
[d]Genes whose mRNAs were enriched following Hfq co-immunoprecipitation (A Zhang, unpublished data). Y, enrichment score above 2.0 for at least one of the three tested conditions; —, enrichment score <2.0 for all tested conditions. Potential binding sites for the distal side of Hfq and the enrichment score for each gene are shown in Supplementary Table S1. Enrichment scores were computed as described in Supplementary data.
[e]Fold regulation of mRNA levels following Spot 42 overexpression reported previously (Beisel and Storz, 2011a). Reported values are the average and standard deviation from three independent experiments. Positive values represent activation and negative represent repression. A, mRNA signal absent in any of the three experiments. NP, gene not probed on the microarray.

Figure 1 Mutational analysis of base-pairing interactions between Spot 42 and selected target mRNAs. (**A**) Secondary structure of Spot 42 supported by *in vitro* structural probing data (Møller *et al*, 2002). The three unstructured regions (I–III) are highlighted in grey, while Spot 42 mutations are shown in white. The base pairs between nucleotides 7–9 and nucleotides 23–25 likely are unpaired most of the time. (**B**–**E**) β-Galactosidase assay results for *lacZ* translational fusion strains carrying the empty vector pBRplac or different Spot 42 expression plasmids. Genes tested as *lacZ* fusions are (**B**) *puuE*, (**C**) *ascF*, (**D**) *nanT*, and (**E**) *fucP*. See Supplementary Table S1 for the sequences of the gene fragment included in each *lacZ* fusion. Strains harbouring each plasmid were induced with 0.2% L-arabinose with or without 1 mM IPTG for 1 h before assaying each culture. The fold-change is the ratio of β-galactosidase activity of cells grown in the absence and presence of IPTG. Error bars represent the standard deviation from measurements of three independent colonies. Base-pairing interactions predicted by *TargetRNA* are on the right. Nucleotides in the unstructured regions of Spot 42 are highlighted in grey. The start codon of the fusion is in green, where the number of nucleotides upstream (−) and downstream (+) of the start codon is indicated. Sequences above the predicted base-pairing interactions correspond to the indicated pSpot42 mutations, while sequences below the predicted base-pairing interactions correspond to the compensatory mutations in the target gene fusions.

©2012 European Molecular Biology Organization

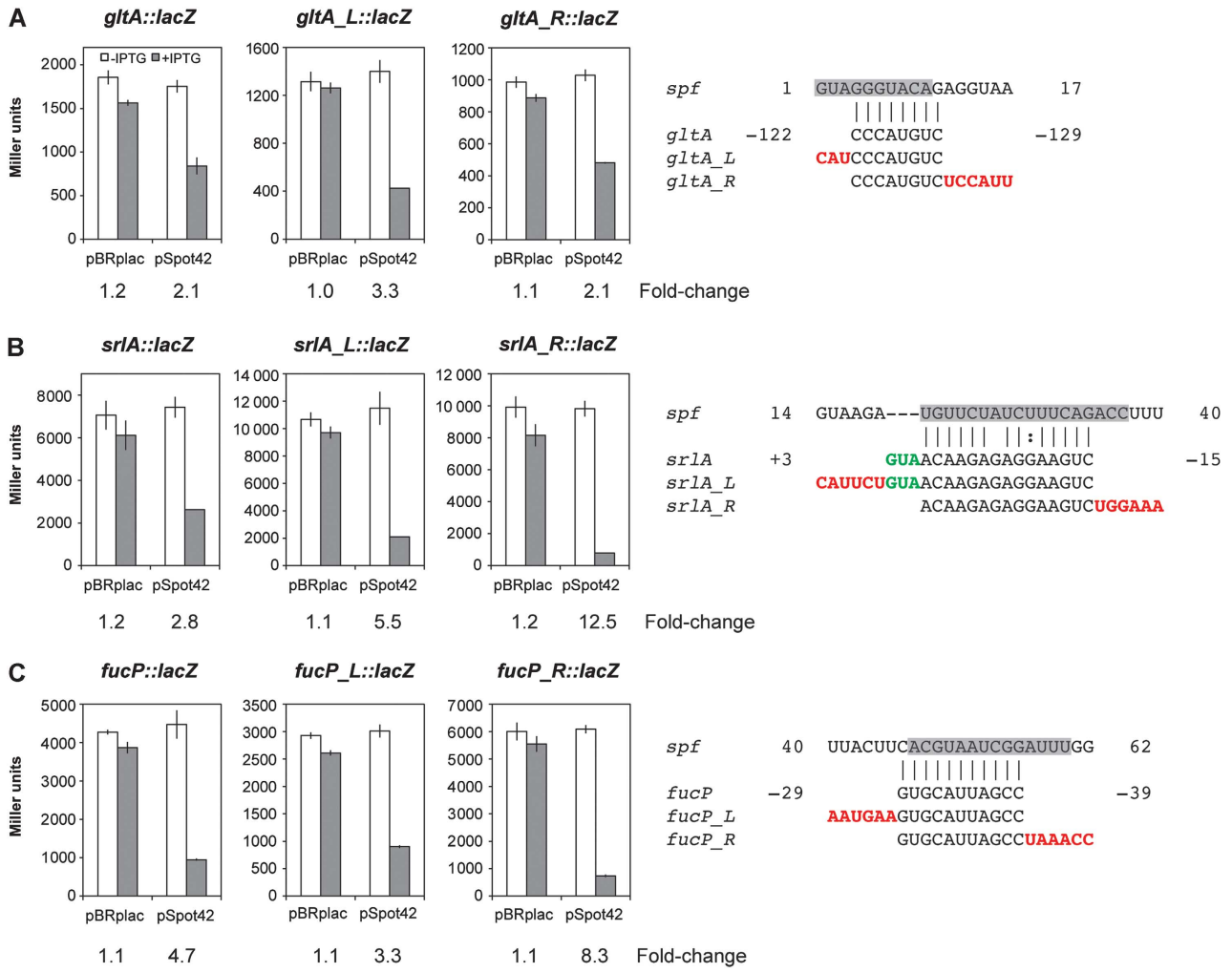


Figure 2 Improved regulation with extended base-pairing interactions in the unstructured regions of Spot 42. Up to six nucleotides were inserted immediately downstream (L) or upstream (R) of each targeting site in the *lacZ* fusions with (**A**) *gltA*, (**B**) *srlA*, and (**C**) *fucP*. Inserted nucleotides (in red) were designed to extend base-pairing between Spot 42 and each fusion for all three regions of Spot 42 (*gltA*, region I; *srlA*, region II; *fucP*, region III). See Figure 1 for a description of the β-galactosidase assay conditions, fold-change, and colouring and numbering of nucleotides in the predicted base-pairing interactions. Error bars represent the standard deviation from measurements of three independent colonies.

hfq-deletion strain (*srlA* + *III*; Supplementary Figure S3). Spot 42 variants containing mutations in either base-pairing region reduced but did not eliminate repression (Figure 4), suggesting that both regions contribute to target regulation. Mutations in both base-pairing regions (one in Spot 42 and the other in the fusion) either eliminated (*fucP-III* + *I*) or greatly reduced (*srlA-II* + *III*) repression. The residual repression of *srlA-II* + *III* by pSpot42-III may be attributed to a persisting potential for base-pairing even after mutations were introduced into regions II and III. These results demonstrate that base-pairing through multiple unstructured regions of an sRNA can improve the strength of target regulation.

Multiple factors separate targets and non-targets of Spot 42

Spot 42 expression had negligible effects on four of the gene fusions (*usg*, *moeA*, *lon*, *entB*) listed in Table II despite strong basal expression and putative base-pairing near the ribosome-binding site of each fusion. We sought to determine why Spot 42 did not have an effect on these target fusions and whether regulation could be activated.

One potential barrier to regulation was insufficient Hfq binding to the fusion mRNA. We began with the *usg* fusion, which lacks a recognizable binding site for the distal side of Hfq (Supplementary Table S1). The *usg* mRNA also was not enriched following co-immunoprecipitation of *Escherichia coli* mRNAs bound to Hfq (A Zhang, unpublished data) (Table I). To introduce Hfq binding, we inserted the 5′ end of *srlA* containing a putative binding site for the distal side of Hfq immediately upstream of the ribosome-binding site in the *usg* fusion (Figure 5A). The resulting *srlA–usg* mRNA was modestly enriched following co-immunoprecipitation of *E. coli* mRNAs bound to Hfq and showed increased binding to Hfq *in vitro* similar to that observed for *srlA* (Figure 5C; Supplementary Figure S4). Insertion of the 5′ end of *srlA* also imparted 3.7-fold repression of the *srlA–usg* fusion by Spot 42 (Figure 5D) that was lost in an *hfq*-deletion strain (Supplementary Figure S3). The predicted base-pairing interactions were responsible for regulation

 ©2012 European Molecular Biology Organization

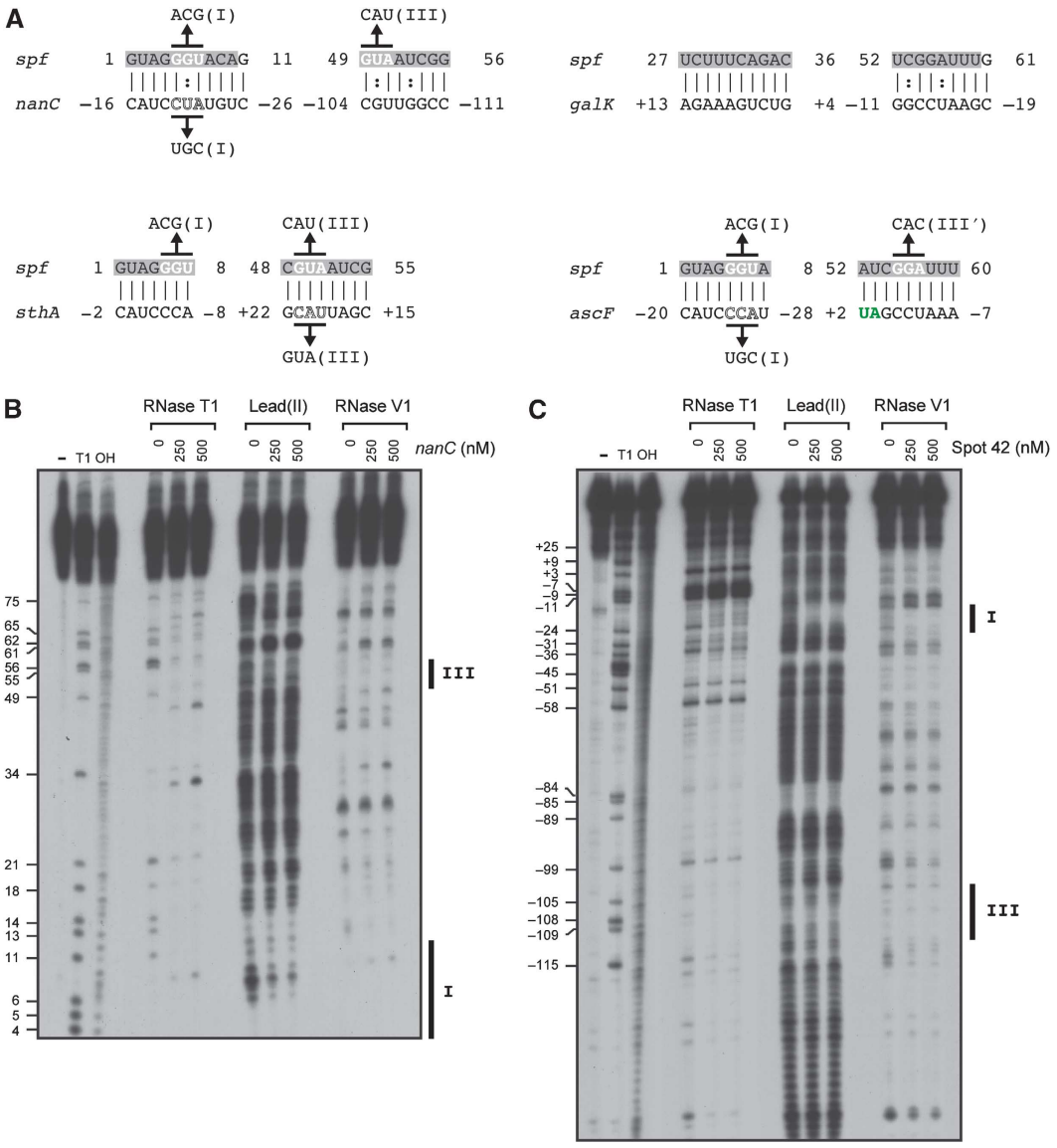


Figure 3 Base-pairing between multiple unstructured regions of Spot 42 and individual target mRNAs. Four Spot 42 target genes that met the requirements for multi-site pairing described in the main text were identified: *nanC*, *galK*, *sthA*, and *ascF*. (**A**) Predicted base-pairing interactions between Spot 42 and each target mRNA. See Figure 1 for a description of the colouring and numbering of the predicted base-pairing interactions. Results of β-galactosidase assays for fusions and Spot 42 variants containing the indicated mutations are available either in this work (*ascF*; Figure 1B) or in our previous work (*nanC*, *sthA*) (Beisel and Storz, 2011a). (**B**) *In vitro* structural probing of labelled Spot 42 hybridized with the unlabelled *nanC* mRNA. 5′ Radiolabelled Spot 42 (20 nM) was incubated at 37°C with 0.1 μg/μl yeast RNA and various concentrations of an unlabelled 192-nucleotide portion of the *nanC* mRNA (0, 250, 500 nM). Concentrations of unlabelled *nanC* mRNAs were selected based on gel shift assays (Supplementary Figure S6A). Incubated RNAs were treated with RNase T1 (cleavage of single-stranded G residues), lead acetate (cleavage of single-stranded nucleotides), or RNase V1 (cleavage of double-stranded and stacked single-stranded nucleotides) and resolved by denaturing PAGE. Untreated Spot 42 (−), denatured Spot 42 treated with RNase T1 (T1), and alkaline hydrolysis of Spot 42 (OH) were resolved as references. Regions I and III of Spot 42 are indicated by black bars to the right of the gel image. Numbering to the left of the gel image indicates the position relative to the 5′ G in Spot 42. (**C**) *In vitro* structural probing of labelled *nanC* mRNA hybridized to unlabelled Spot 42. 5′ ^{32}P-radiolabelled *nanC* mRNA (20 nM) was incubated as described in (**B**) with various concentrations of unlabelled Spot 42 (0, 250, 500 nM). Concentrations of unlabelled Spot 42 were selected based on gel shift assays (Supplementary Figure S6B). Predicted targeting sites for regions I and III of Spot 42 are indicated by black bars to the right of the gel image. Numbering to the left of the gel image indicates the position upstream (−) and downstream (+) of the start codon. The RNase V1 cleavage products have a 3′-hydroxyl, which results in slightly reduced mobility compared with cleavage products with a 3′-phosphate resulting from RNase T1 digestion and alkaline hydrolysis. Figure source data can be found in Supplementary data.

of the *srlA–usg* fusion, as a mutation in the implicated region of Spot 42 disrupted repression while a compensatory mutation in the fusion restored regulation (Figure 5B and D). These results indicate that mRNAs repressed by Hfq-binding sRNAs require a binding site for the distal side of Hfq.

Next, we assessed why Spot 42 had a negligible effect on the *moeA* fusion. Similar to *usg*, the *moeA* fusion lacks a recognizable binding site for the distal side of Hfq (Supplementary Table S1) and the *moeA* mRNA was not enriched following co-immunoprecipitation of Hfq-bound mRNAs (Table II). However, unlike the *usg* fusion, introduction

©2012 European Molecular Biology Organization

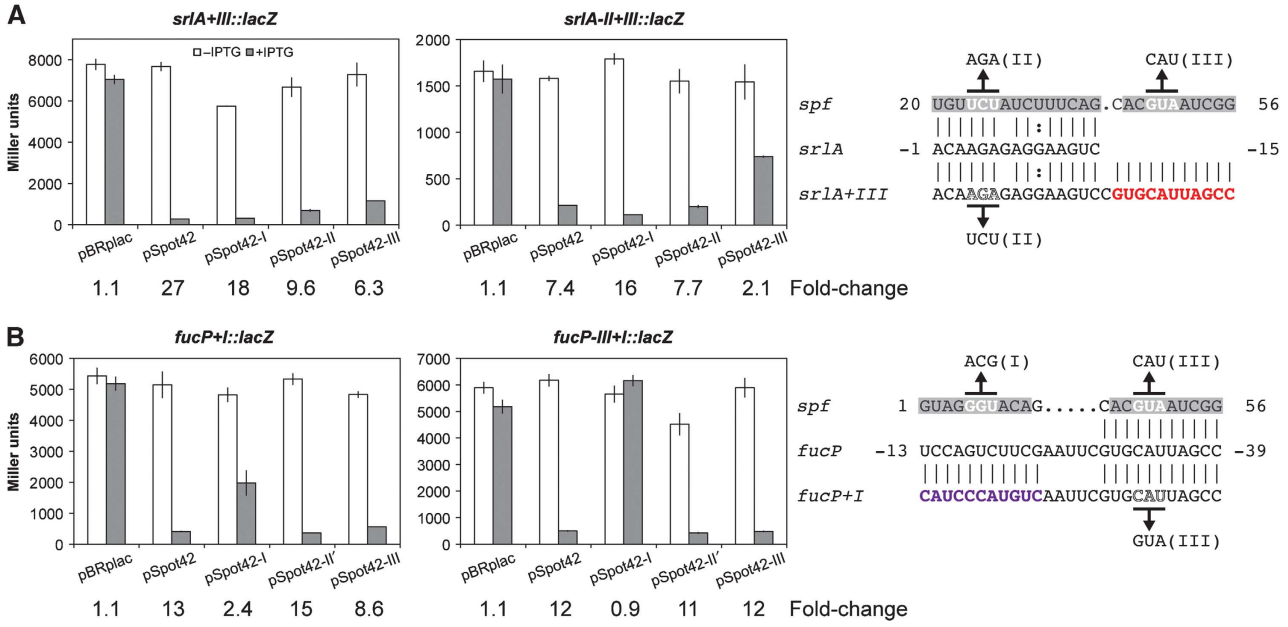


Figure 4 Improved regulation with base-pairing through multiple unstructured regions of Spot 42. Eleven nucleotides were either inserted (in red) or mutated (in purple) in the (**A**) *srlA* or (**B**) *fucP* fusions to create a new targeting site for Spot 42 (*srlA* + *III::lacZ*, *fucP* + *I::lacZ*). Indicated mutations were introduced at site II in *srlA* + *III::lacZ* and site III in *fucP* + *I::lacZ* (to give *srlA-II* + *III::lacZ* and *fucP-III* + *I::lacZ*, respectively). See Figure 1 for a description of the β-galactosidase assay conditions, fold-change, and colouring and numbering of the predicted base-pairing interactions. Error bars represent the standard deviation from measurements of three independent colonies.

of the 5′ end of *srlA* immediately upstream of the predicted Spot 42 targeting site (Figure 6A) did not impart regulation by Spot 42 (Figure 6D). Secondary structure predictions revealed that two different sequences within the 5′ untranslated region of *moeA* may mask the putative targeting site (Figure 6B). Masking the sRNA targeting site previously was shown to reduce and even eliminate regulation of the *sodB* mRNA by the sRNA RyhB (Hao *et al*, 2011).

We hypothesized that the *moeA* fusion fails to be regulated by Spot 42 for two reasons: lack of an Hfq-binding site and occlusion of the targeting site. To assess whether targeting site occlusion prevents regulation of the *moeA* fusion containing an Hfq-binding site (*srlA–moeA*), we introduced two modifications predicted to free the targeting site: mutation of two nucleotides downstream of the targeting site (*srlA–moeA1*) or placement of *lacZ* immediately downstream of the start codon (*srlA–moeA2*) (Figure 6A and B). In conjunction with the Hfq-binding site, the two modifications imparted repression by Spot 42 either individually or when combined (*srlA–moeA1,2*) that was lost in an *hfq*-deletion strain (*srlA–moeA1,2*; Supplementary Figure S3). Furthermore, mutations in the unstructured region of Spot 42 implicated in base-pairing eliminated repression, supporting the predicted base-pairing interactions (Figure 6C and D). Relieving occlusion of the targeting site in the *moeA* fusion lacking an Hfq-binding site (*moeA2*) did not impart regulation (Supplementary Figure S5A and B). Regulation of the *lon* fusion by Spot 42 appears to be hindered by a similar structural barrier, as introduction of the 5′ end of *srlA* did not impart regulation by Spot 42 and *NUPACK* predicted extensive base-pairing between the 5′ end of *srlA* and the putative Spot 42 targeting site (Supplementary Figure S5C–E). We thus conclude that occlusion of the targeting site can prevent sRNA-based repression.

Finally, we investigated why Spot 42 had a negligible effect on the *entB* fusion. Unlike *usg* and *moeA*, the lack of an Hfq-binding site does not appear to be the culprit: the *entB* fusion contains two putative binding sites for the distal side of Hfq (Supplementary Table S1) and the *entB* mRNA was strongly enriched following co-immunoprecipitation of Hfq-bound mRNAs (Table I). In addition, the lack of regulation likely is not due to occlusion of the Spot 42 targeting site, as *NUPACK* predicts that the Spot 42 targeting site is not as structured as the *srlA–moeA* and the *srlA–lon* fusions (Figure 6B; Supplementary Figure S5D and G). Thus, we predicted that the *entB* fusion lacks an additional factor important for regulation by Spot 42.

We first focussed on the putative upstream Hfq-binding site (Figure 6E). If this is the principal site of Hfq binding, then Hfq stimulation of Spot 42 pairing could be impeded by sequences upstream of this site or by the large stretch of 31 nucleotides separating this putative Hfq-binding site and the Spot 42 targeting site. We tested these possibilities by placing the transcriptional start site 18 nucleotides upstream of this putative Hfq-binding site or shortening the distance between this site and the Spot 42 targeting site to 12 nucleotides or 4 nucleotides (Supplementary Figure S5F and G). Overexpression of Spot 42 had a negligible effect on the expression of all three constructs (Supplementary Figure S5H), suggesting that the putative upstream Hfq-binding site is not involved in target regulation. What prevents this site from contributing to regulation by Spot 42 remains unclear, although other factors important for Hfq binding (e.g., the presence of an adjacent hairpin) and function may remain to be elucidated.

We next focussed on the putative downstream Hfq-binding site (Figure 6E), which directly overlaps with the Spot 42 targeting site. If Hfq must bind the mRNA for sRNA-based

 ©2012 European Molecular Biology Organization

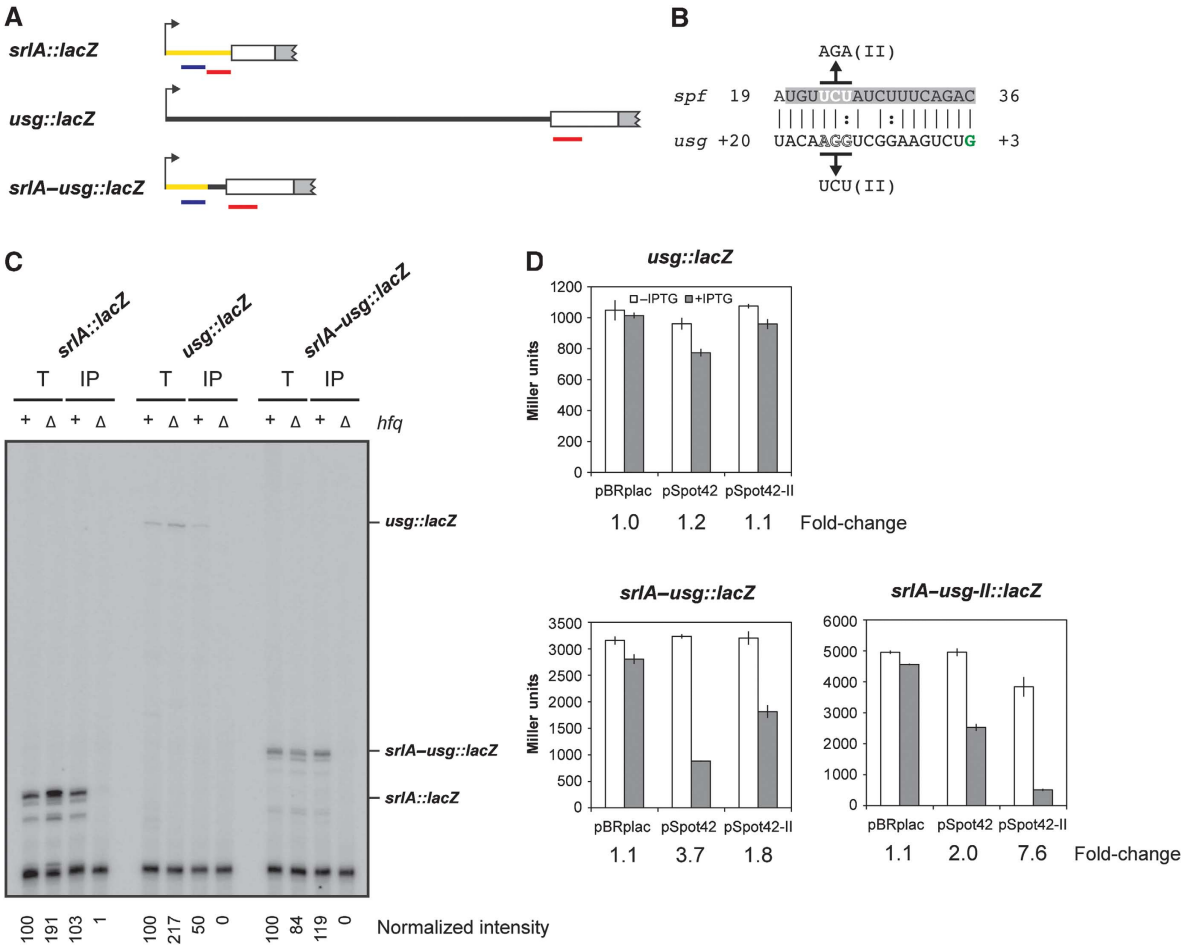


Figure 5 Gene regulation by Spot 42 conferred through insertion of an Hfq-binding site. (**A**) Schematic representation of the *srlA* and *usg* fusions. Putative Hfq-binding sites are in blue, Spot 42 targeting sites are in red, the coding region of each gene included in the fusion is a white box, the 5′ portion of *lacZ* is a grey box, and the 5′ untranslated region of *srlA* included in each fusion is coloured yellow. To generate *srlA–usg::lacZ*, the first 26 nucleotides of the *srlA* mRNA were fused immediately upstream of the ribosome-binding site in *usg::lacZ*. (**B**) Base-pairing interactions between Spot 42 and the *usg* mRNA predicted by *TargetRNA*. The indicated mutations are contained in pSpot42-II (above) and *srlA–usg-II::lacZ* (below). (**C**) Primer extension analysis of *srlA*, *usg*, and *srlA–usg* fusion mRNAs co-immunoprecipitated with Hfq. Total RNA (T, 2 μg) or RNA eluted from Hfq immunoprecipitated with α-Hfq antibodies (IP, 0.2 μg) from wild-type (+) or *hfq*-deletion (Δ) strains were reverse-transcribed with a 5′ radiolabelled *lacZ*-specific primer and resolved by denaturing PAGE. The anticipated primer extension product for each strain is indicated on the right. The intensity of bands at these locations were quantified using a phosphorimager and normalized to the band for total RNA from the wild-type strain following subtraction of background intensity. Similar results were obtained using gene-specific primers. The non-specific bands at the bottom of the gel indicate equal loading of each sample. (**D**) β-Galactosidase assay results for the *usg* fusions. See Figure 1 for a description of the β-galactosidase assay conditions and fold-change. Error bars represent the standard deviation from measurements of three independent colonies. Figure source data can be found in Supplementary data.

regulation to occur, then Hfq binding could be preventing base-pairing between Spot 42 and the *entB* mRNA. We tested this hypothesis by introducing either a separate Hfq-binding site or a separate Spot 42 targeting site. To introduce a separate Hfq-binding site, we introduced the 5′ of *srlA* immediately upstream of the predicted Spot 42 targeting site, which replaced the 5′ end of *entB* and the putative upstream Hfq-binding site (*srlA–entB*; Figure 6E). To introduce a separate Spot 42 targeting site, we mutated the first 11 nucleotides downstream of the start codon to be complementary to region I of Spot 42 (*entB+I*; Figure 6E). The resulting fusions showed 4.2- and 1.7-fold repression by Spot 42, respectively, which was disrupted when the implicated region of pairing in Spot 42 was mutated (Figure 6F and G). In addition, regulation of the *srlA–entB* fusion by Spot 42 was disrupted when *hfq* was deleted (Supplementary Figure S3). These observations suggest that, while Hfq is required for Spot 42 to associate with the *entB* mRNA, the Hfq-binding and sRNA targeting sites cannot be overlapping.

In total, our results suggest that mRNAs containing a putative sRNA targeting site require additional factors to undergo sRNA-based repression, including an unstructured sRNA targeting site and a non-overlapping Hfq-binding site.

Discussion

In this study, we provide evidence that Spot 42 directly represses seven genes beyond those previously identified: *ascF*, *atoD*, *caiA*, *fucP*, *nanT*, *paaK*, and *puuE*. Besides *nanT*, these genes were not identified by our previous microarray

©2012 European Molecular Biology Organization

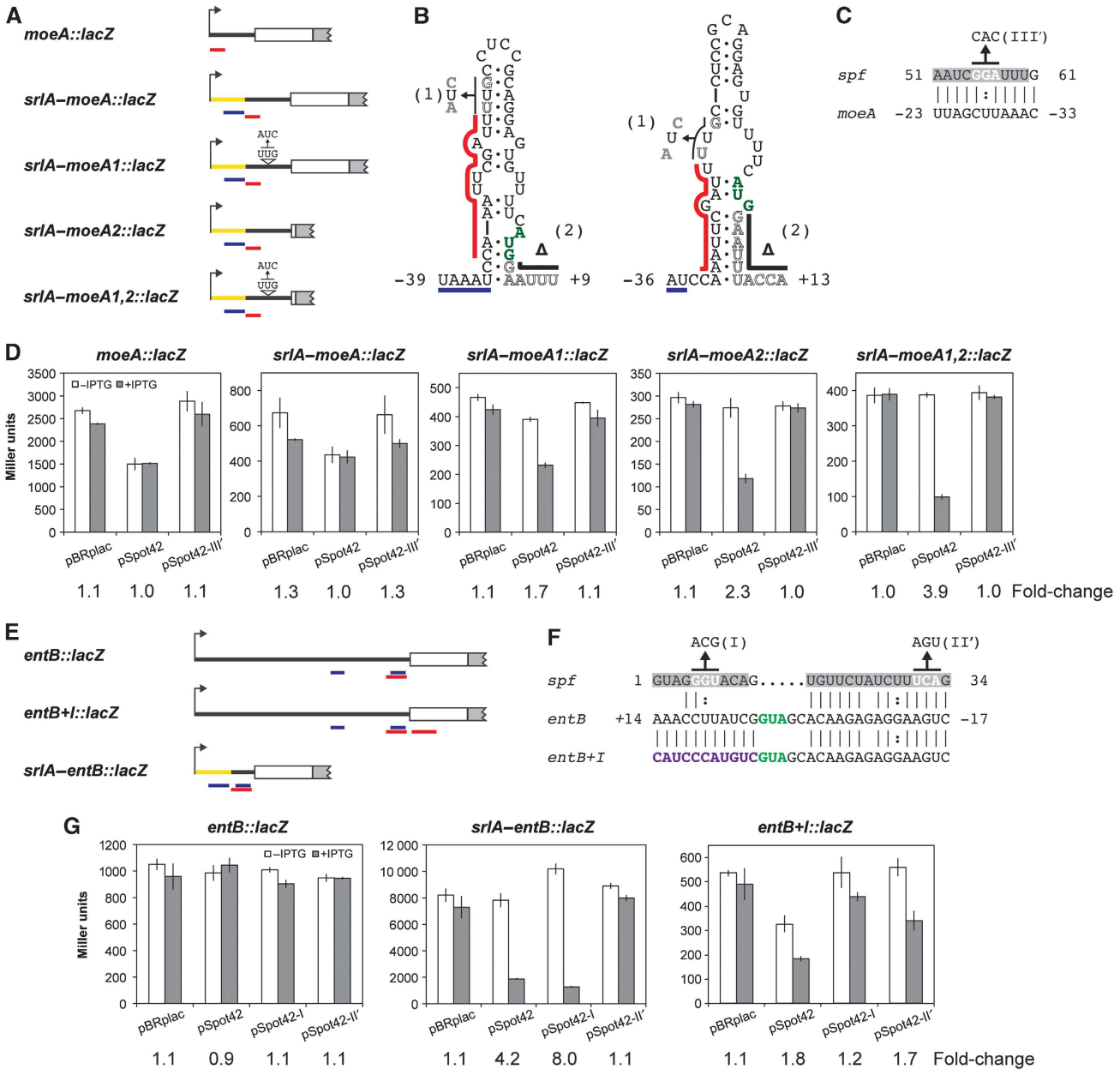


Figure 6 Gene regulation by Spot 42 conferred by freeing the Spot 42 targeting site or separating the Hfq-binding site and Spot 42 targeting site. (**A**) Schematic representation of the *moeA* fusions. See Figure 5A for a description of the colouring. *moeA::lacZ* was generated by fusing *lacZ* to the 14th codon of *moeA*, *srlA–moeA::lacZ* was generated by fusing the 5′ end of *srlA* to the transcriptional start site of *moeA::lacZ*, *srlA–moeA1::lacZ* was generated by introduction of two point mutations in the 5′ untranslated region of *moeA*, and *srlA–moeA2::lacZ* was generated by fusing *lacZ* to the *srlA–moeA* start codon. (**B**) Secondary structures of *srlA–moeA* predicted by *NUPACK* and *mfold*. The mutations contained within *srlA–moeA1::lacZ* (1), *srlA–moeA2::lacZ* (2), or *srlA–moeA1,2::lacZ* (both 1 and 2) are bordered by black lines. Nucleotides bordered by a red line are the Spot 42 targeting site, nucleotides bordered by a blue line are the Hfq-binding site, and nucleotides in green are the start codon. Indicated is the number of nucleotides upstream (–) and downstream (+) of the start codon. (**C**) Base-pairing interactions between Spot 42 and the *moeA* mRNA predicted by *TargetRNA*. See Figure 1 for a description of the colouring and numbering of the predicted base-pairing interactions. (**D**) β-Galactosidase assay results for the *moeA* fusions. See Figure 1 for a description of the β-galactosidase assay conditions and fold-change. (**E**) Schematic representation of the *entB* fusions. See Figure 5A for a description of colouring. (**F**) Predicted base-pairing interactions between Spot 42 and either *entB* or *entB* mutated to include a targeting site for region I of Spot 42 (*entB + I*). See Figure 1 for a description of the colouring and numbering of the predicted base-pairing interactions. Mutated nucleotides are in purple. (**G**) β-Galactosidase assay results for the *entB* fusions. See Figure 1 for a description of the β-galactosidase assay conditions and fold-change. Error bars in (**D**, **G**) represent the standard deviation from measurements of three independent colonies. Different fusions showed differing levels of basal expression, which for *moeA* and *srlA–moeA* (**D**) as well as *entB* and *srlA–entB* (**G**) are reflected in differing mRNA levels (Supplementary Figure S7). Although there also was some variation in the basal levels between experiments, the fold-change was very consistent.

analysis (Beisel and Storz, 2011a) possibly due to the inability to detect the mRNAs (*ascF*, *puuE*), the short time of Spot 42 induction, strong repression from endogenous Spot 42 expression, or sole regulation of target genes at the level of translation (Tables I and II). However, the genes identified in this study fit the broad role of Spot 42 in catabolite repression (Beisel and Storz, 2011b). Many of the genes encode enzymes and transporters responsible for the consumption of carbon

 ©2012 European Molecular Biology Organization

sources previously associated with Spot 42 (*nanT*, *N*-acetylneuraminic acid; *fucP*, L-fucose) as well as additional carbon sources (*atoD*, acetoacetate; *paaK*, phenylacetate; *glpF*, glycerol; *ascF*, β-glucosides). One of the newly identified genes, *fucP*, is encoded in the same operon as the previously identified Spot 42 target *fucI*, indicating that Hfq-binding sRNAs can base pair with multiple genes within an operon. In addition, four of the identified genes (*ascF*, *caiA*, *glpF*, *paaK*) appear to be transcriptionally activated by CRP—the repressor of Spot 42 expression (Weissenborn *et al*, 1992; Buchet *et al*, 1999; Ferrández *et al*, 2000; Ishida *et al*, 2009)—and thus are targeted by the CRP-Spot 42 feedforward loop (Beisel and Storz, 2011a). These results suggest that Spot 42 has an even greater impact on cellular metabolism than reported in our previous study (Beisel and Storz, 2011a). Since we restricted our analysis to the five top-scoring genes for each unstructured region of Spot 42, more targets of this sRNA likely await discovery.

Our findings indicate that the accuracy of sRNA target prediction should be improved by incorporating restrictions for both sRNAs and target mRNAs. First, utilizing only unstructured regions of each sRNA should improve target prediction because of the importance of these regions for base-pairing with target mRNAs. Second, unstructured regions are more likely to be targeting sites in mRNAs. Third, mRNAs should contain at least three repeats of the A-R-N motif that binds the distal side of Hfq. In support of this requirement, all of the Spot 42 targets we have identified contain at least three repeats located within 14 nucleotides of the targeting site (excluding intervening hairpins; Supplementary Table S1). Fourth, mRNAs should not contain overlapping Hfq-binding sites and sRNA targeting sites to prevent Hfq from blocking access of the sRNA. Both *intaRNA* and *RNApredator* already account for the second criterion for target prediction, where use of *intaRNA* improved the identification of Ysr1 targets in *Prochlorococcus* MED4 (Richter *et al*, 2010). All prediction programmes could readily be modified to incorporate the other criteria.

Other factors potentially important for target regulation remain to be investigated systematically, including the spacing between the Hfq-binding and sRNA targeting sites, the exact composition of an Hfq-binding site (e.g., the presence of an adjacent hairpin), and the presence of an A nucleotide often immediately downstream of the targeting site (present in 9 out of the 12 Spot 42 targets with validated base-pairing interactions) (Papenfort *et al*, 2010). Since we focussed our search around ribosome-binding sites, additional criteria may need to be established to identify bona fide targeting sites well upstream of the ribosome-binding site or in coding and 3′ untranslated regions.

We demonstrated that for some mRNA targets, Spot 42 regulation involves multiple unstructured regions. This parallels the previous prediction that DsrA base pairs near the start codon and the stop codon of a subset of mRNAs (Lease and Belfort, 2000) as well as the recent observation that GcvB employs two of its unstructured regions to regulate the target *cycA* (Sharma *et al*, 2011). By studying both natural and synthetic targets with multi-site pairing, we found that base-pairing through two targeting sites improved the strength of regulation. We anticipate that multi-site pairing provides a strategy for sRNAs to tune regulatory strength. Since numerous Hfq-binding sRNAs contain multiple unstructured regions (Peer and Margalit, 2011), multi-site pairing may represent a common strategy in sRNA-based regulation.

While we demonstrated that both sites in the multi-site targets contribute to regulation, it is unclear whether Spot 42 base pairs with both sites simultaneously. This question is particularly relevant to the *ascF* and *sthA* fusions, which are capable of forming bimolecular pseudoknots with Spot 42. One possibility is that two Spot 42 molecules bind to one mRNA molecule, leading to increased regulation. However, recent evidence suggests that Hfq can accommodate only one sRNA:mRNA pair (Updegrove *et al*, 2011). Another possibility is that one Spot 42 molecule base pairs with one targeting site at a time. By providing two sites, Spot 42 would have greater avidity for the mRNA, leading to greater association and target regulation. Further biochemical analyses will be required to fully understand how sRNAs interact with their target mRNAs *in vivo*, whether through single or multiple targeting sites.

Recent work by Shi and coworkers has suggested that Hfq binding may not be essential for sRNA-based regulation (Hao *et al*, 2011). They showed that truncation of RyhB to the base-pairing region encoded in a stem-loop together with the transcriptional terminator relieved the need for Hfq for both sRNA stability and destabilization of the *sodB*, *fumA*, and *sdhD* mRNAs. Although this suggests that Hfq-binding sRNAs may act independently of Hfq, the truncated version of RyhB was overexpressed and departs heavily from the original RyhB sequence. The truncated RyhB RNA perceivably acts through kissing hairpin interactions similar to many natural and synthetic antisense RNAs (Heidrich and Brantl, 2003; Isaacs *et al*, 2004; Nakashima and Tamura, 2009; Lucks *et al*, 2011). Even though the truncated RyhB deviates from natural Hfq-binding sRNAs, its ability to repress selected targets raises the question why cells would employ Hfq in sRNA-based regulation. Hfq may relieve sequence restrictions in the base-pairing regions of sRNAs often observed for antisense RNAs (Heidrich and Brantl, 2003), thus easing multi-gene targeting by Hfq-binding sRNAs. In addition, Hfq binding may limit the number of mRNAs encountered by sRNAs, thereby restraining off-target effects. Therefore, we posit that Hfq-binding sites are a standard component in mRNAs naturally targeted by Hfq-binding sRNAs.

Finally, our findings inform the design of synthetic Hfq-binding sRNAs and target mRNAs. Previous studies have suggested that a base-pairing region, a binding site for the proximal side of Hfq, and a transcriptional terminator are sufficient for the construction of Hfq-binding sRNAs (Papenfort *et al*, 2010). While this may be sufficient for one base-pairing region, our study demonstrates that Hfq-binding sRNAs can be designed to base pair through multiple regions as long as each region is unstructured. Introduction of additional base-pairing regions could facilitate coordinated regulation of numerous target genes and allow heightened repression. Our study also provides clear guidelines for the design of target mRNAs. Designed targets should be equipped with an Hfq-binding site (such as the 5′ end of *srlA*) and an adjacent but non-overlapping targeting site, where both sites are unstructured. Target regulation can be improved by introducing additional targeting sites, where the exact order and configuration and these sites and the Hfq-binding site appear to be less critical. Using these guidelines, conceivably

©2012 European Molecular Biology Organization

any gene—whether endogenous or heterologous—could be converted into a potent sRNA target.

Materials and methods

Computational prediction of Spot 42 targets, base-pairing interactions, and RNA secondary structures

We searched for targets of Spot 42 using the sRNA target prediction algorithm *TargetRNA* (snowwhite.wellesley.edu/targetRNA/). See Supplementary data for a description of the parameter set employed and the identification of multiple targeting sites in individual mRNAs.

Secondary structure predictions were performed using the folding algorithms *NUPACK* (http://www.nupack.org) (Zadeh *et al*, 2011) and *mfold* (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form) (Zuker, 2003) using default parameters. The two structures of the *srlA–moeA* fusion reported in Figure 6B represent the minimal free energy structure and the most stable suboptimal structure predicted by *NUPACK* and *mfold*.

Plasmid and strain construction

Oligonucleotides, plasmids, and strains used in this study are listed in Supplementary Table S2. Plasmids pBRplac, pSpot42, pSpot42-I, pSpot42-II, and pSpot42-III were reported previously (Beisel and Storz, 2011a). Plasmids pSpot42-II′ and pSpot42-III′ were constructed using the Gibson assembly method (Gibson, 2011). See Supplementary data for a detailed description of the procedure.

All strains were derived from *E. coli* strain K-12 substrain MG1655. MG1655 Δ*spf* was generated by P1 transduction of Δ*spf::kan*^{R} from NM525 Δ*spf::kan*^{R} (Beisel and Storz, 2011a) and excision of the *kan*^{R} resistance cassette with plasmid pCP20 (Cherepanov and Wackernagel, 1995). The *lacZ* fusions were generated in PM1205 Δ*spf::kan*^{R} as described in detail previously (Mandin and Gottesman, 2009). Briefly, each gene was amplified by PCR from the MG1655 genome with flanking ends complementary to the P_{BAD} promoter (5′ end) and the *lacZ* coding region (3′ end). The amplified DNA template was recombined in place of the *cat-sacB* cassette using mini-λ-mediated recombination (Court *et al*, 2003). Desired recombinants were identified by selective growth on M9 sucrose plates. Mutant constructs were generated by amplifying each gene in two fragments by PCR from the associated PM1205 strain. The two fragments were assembled by PCR to generate the final DNA template for recombination into PM1205 Δ*spf::kan*^{R}. Transductions were confirmed by PCR and all recombination events were confirmed by sequencing.

Growth conditions

All strains were grown by shaking at 250 r.p.m. at 37°C unless noted otherwise. Strains were grown in Luria-Bertani (LB) media (1% bacto-tryptone, 0.5% yeast extract, and 1% NaCl) or M9 minimal media (1 × M9 salts, 10 μg/ml thiamine, 2 mM $MgSO_4$, and 0.1 mM $CaCl_2$) supplemented with 0.2% glucose. Cell density was determined by measuring A_{600} using an Ultrospec 3000 UV/Vis spectrophotometer (Pharmacia Biotech).

β-Galactosidase assays

β-Galactosidase assays were performed as described previously (Beisel and Storz, 2011a). Three separate colonies were grown overnight in LB, back-diluted to $A_{600}=0.01$ in the same media and grown to $A_{600}=\sim 0.1$. L-arabinose and IPTG then were added to each culture to final concentrations of 0.2% and 1 mM, respectively, as indicated. Cells were assayed for β-galactosidase activity (Miller, 1977) after an additional 1 h of growth when cultures attained $A_{600}=0.4$–0.6. The A_{600} and A_{420} of the cultures were measured using an Ultrospec 3000 UV/Vis spectrophotometer (Pharmacia Biotech).

RNA radiolabelling and in vitro structural probing

Template DNA for T7 transcription was amplified from the genomic DNA of the corresponding PM1205 strain by PCR with primers containing the T7 promoter (GTTTTTTTTAATACGACTCACTATAGG). T7 transcription was conducted using the MegaShortscript T7 Transcription Kit (Ambion) according to the manufacturer's instructions. Transcribed RNAs were checked for complete synthesis. The 5′ ends of RNAs were ^{32}P-radiolabelled as described previously, after which the RNAs were gel purified (Sittka *et al*, 2007).

In vitro structural probing was performed in 10 μl reactions similarly to previous work (Sharma *et al*, 2007). Radiolabelled RNA (~0.2 pmol) was denatured at 95°C for 1 min and chilled on ice for 5 min, followed by the addition of 1 × structure buffer (Ambion), 0.1 μg/μl yeast RNA, and the indicated concentration of unlabelled RNA. Concentrations of unlabelled RNAs were selected based on gel shift assay results (Supplementary Figure S6). Following incubation at 37°C for 1 h, 2 μl of RNase T1 (0.01 U/μl; Ambion) or 2 μl RNase V1 (0.0001 U/μl; Ambion) was added to the corresponding reaction and incubated for 6 min. For the lead cleavage reactions, 2 μl of fresh lead(II) acetate (25 mM) was added and the samples were incubated for 1.5 min at 37°C. Reactions were stopped with the addition of 20 μl of Inactivation/Precipitation buffer (Ambion) and vortexing. Reactions were precipitated out of solution at −80°C for 15 min, spun down and washed with 70% EtOH, and finally RNA pellets suspended in 7 μl loading buffer II (95% formamide, 18 mM EDTA, 0.025% SDS, 0.2% xylene cyanol, 0.2% bromophenol blue; Ambion) and placed on ice.

To generate the RNase T1 ladder, radiolabelled RNA (~0.4 pmol) was combined with 1 × Sequencing Buffer (Ambion), denatured at 95°C for 1 min, and incubated with RNase T1 (1 μl, 0.1 U/μl) at 37°C for 5 min. For the hydroxyl ladder, radiolabelled RNA (~0.4 pmol) was combined with 1 × Alkaline Hydrolysis Buffer (Ambion) and incubated at 90°C for 5 min. Both ladder reactions were stopped with the addition of 12 μl of loading buffer II. All samples were denatured at 95°C for 3 min and 3 μl resolved on a 6% denaturing polyacrylamide/7 M urea gel in 1 × TBE. Gels were dried and exposed to BioMax XAR film (Kodak).

Hfq co-immunoprecipitation

Hfq co-immunoprecipitation was performed similarly to previous work (Zhang *et al*, 1998). Briefly, cultures were grown to mid-log phase and incubated in 0.2% L-arabinose for 1 h as described in the β-galactosidase assays. Cultures then were pelleted, resuspended in lysis buffer (20 mM Tris–HCl/pH 8.0, 150 mM KCl, 1 mM $MgCl_2$, and 1 mM DTT, 0.2 U RNaseOUT (Ambion)), and lysed by vortexing with glass beads for 10 min. Cell lysates then were used to extract total RNA or immunoprecipitate Hfq. To immunoprecipitate Hfq, 200 μl cell lysate was combined with 24 mg of Protein A Sepharose CL-4B beads (Amersham Biosciences) complexed with 20 μl of α-Hfq serum, 200 μl of Net2 Buffer (50 mM Tris–HCl/pH 7.4, 150 mM NaCl, and 0.05% Triton X-100), and 1 μl RNaseOUT. The mixture was incubated at 4°C for 2 h with rotation then washed 5 × with 1.5 ml Net2 Buffer. Following the washes, the beads were combined with 400 μl of Net2 Buffer, 50 μl of 3 M NaOAc, 5 μl of 10% SDS, and 600 μl of Phenol:Chloroform:Isoamyl Alcohol (Ambion) and RNA was ethanol precipitated. Total RNA (2 μg) or co-immunoprecipitated RNA (0.2 μg) then was used for primer extension assay. Primer extension analysis was performed as described previously (Zhang *et al*, 1998). The products (5 μl) were resolved by denaturing PAGE. Gels were dried and either exposed to BioMax XAR film or to a phosphor screen and quantified using a phosphorimager.

Additional methods

See Supplementary data for detailed descriptions of plasmid construction, sRNA target predictions, enrichment scores calculated from the Hfq co-immunoprecipitation data, quantitative real-time PCR, gel shift assays for RNA hybridization and Hfq binding, primer extension analysis, and free energy calculations.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank B Tjaden, S Gottesman, and members of the Storz laboratory for helpful discussions and critical reading of the manuscript. We are grateful to C Sharma for technical assistance, to A Zhang for providing the *hfq* mutant allele, purified Hfq and α-Hfq serum, and to MK Thomason and A Zhang for sharing unpublished data. Work carried out in the laboratory of GS was supported by the Intramural Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development. CLB is a Gordon and Betty Moore Foundation Fellow of the Life Sciences Research Foundation.

 ©2012 European Molecular Biology Organization

Author contributions: CLB, TBU, and GS designed the experiments; CLB, TBU, and BJJ performed the experiments; CLB and GS analysed the data; and CLB and GS wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Beisel CL, Storz G (2011a) 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

Beisel CL, Storz G (2011b) Discriminating tastes: physiological contributions of the Hfq-binding small RNA Spot 42 to catabolite repression. *RNA Biol* **8:** 766–770

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

Buchet A, Nasser W, Eichler K, Mandrand-Berthelot MA (1999) Positive co-regulation of the *Escherichia coli* carnitine pathway *cai* and *fix* operons by CRP and the CaiF activator. *Mol Microbiol* **34:** 562–575

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

Cherepanov PP, Wackernagel W (1995) Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158:** 9–14

Court DL, Swaminathan S, Yu D, Wilson H, Baker T, Bubunenko M, Sawitzke J, Sharan SK (2003) Mini-lambda: a tractable system for chromosome and BAC engineering. *Gene* **315:** 63–69

Durand S, Storz G (2010) Reprogramming of anaerobic metabolism by the FnrS small RNA. *Mol Microbiol* **75:** 1215–1231

Eggenhofer F, Tafer H, Stadler PF, Hofacker IL (2011) RNApredator: fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res* **39:** W149–W154

Ferrández A, García JL, Díaz E (2000) Transcriptional regulation of the divergent *paa* catabolic operons for phenylacetic acid degradation in *Escherichia coli*. *J Biol Chem* **275:** 12214–12222

Geissmann TA, Touati D (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J* **23:** 396–405

Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments. *Meth Enzymol* **498:** 349–361

Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA (2011) Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc Natl Acad Sci USA* **108:** 12875–12880

Hao Y, Zhang ZJ, Erickson DW, Huang M, Huang Y, Li J, Hwa T, Shi H (2011) Quantifying the sequence-function relation in gene silencing by bacterial small RNAs. *Proc Natl Acad Sci USA* **108:** 12473–12478

Heidrich N, Brantl S (2003) Antisense-RNA mediated transcriptional attenuation: importance of a U-turn loop structure in the target RNA of plasmid pIP501 for efficient inhibition by the antisense RNA. *J Mol Biol* **333:** 917–929

Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22:** 841–847

Ishida Y, Kori A, Ishihama A (2009) Participation of regulator AscG of the beta-glucoside utilization operon in regulation of the propionate catabolism operon. *J Bacteriol* **191:** 6136–6144

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

De Lay N, Gottesman S (2009) The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J Bacteriol* **191:** 461–476

Lease RA, Belfort M (2000) A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. *Proc Natl Acad Sci USA* **97:** 9919–9924

Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci USA* **106:** 19292–19297

Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP (2011) Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc Natl Acad Sci USA* **108:** 8617–8622

Mandin P, Gottesman S (2009) A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system. *Mol Microbiol* **72:** 551–565

Miller J (1977) *Experiments in Molecular Genetics*, 3rd edn Cold Spring Harbor: Cold Spring Harbor Laboratory

Mitarai N, Andersson AMC, Krishna S, Semsey S, Sneppen K (2007) Efficient degradation and expression prioritization with small RNAs. *Phys Biol* **4:** 164–171

Mitarai N, Benjamin J-AM, Krishna S, Semsey S, Csiszovszki Z, Massé E, Sneppen K (2009) Dynamic features of gene expression control by small regulatory RNAs. *Proc Natl Acad Sci USA* **106:** 10655–10659

Moll I, Leitsch D, Steinhauser T, Bläsi U (2003) RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep* **4:** 284–289

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

Nakashima N, Tamura T (2009) Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. *Nucleic Acids Res* **37:** e103

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

Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JCD, Vogel J (2006) σ^E-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol Microbiol* **62:** 1674–1688

Peer A, Margalit H (2011) Accessibility and evolutionary conservation mark bacterial small-RNA target-binding regions. *J Bacteriol* **193:** 1690–1701

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

Pulvermacher SC, Stauffer LT, Stauffer GV (2009a) Role of the sRNA GcvB in regulation of *cycA* in *Escherichia coli*. *Microbiology* **155:** 106–114

Pulvermacher SC, Stauffer LT, Stauffer GV (2009b) The small RNA GcvB regulates *sstT* mRNA expression in *Escherichia coli*. *J Bacteriol* **191:** 238–248

Richter AS, Schleberger C, Backofen R, Steglich C (2010) Seed-based IntaRNA prediction combined with GFP-reporter system identifies mRNA targets of the small RNA Yfr1. *Bioinformatics* **26:** 1–5

Schumacher MA, Pearson RF, Møller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J* **21:** 3546–3556

Sharma CM, Darfeuille F, Plantinga TH, Vogel J (2007) 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

Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf H-J, Hinton JCD, Vogel J (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* **81:** 1144–1165

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

Soper TJ, Woodson SA (2008) The *rpoS* mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA* **14:** 1907–1917

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

©2012 European Molecular Biology Organization

Tjaden B (2008) TargetRNA: a tool for predicting targets of small RNA action in bacteria. *Nucleic Acids Res* **36:** W109–W113

Updegrove TB, Correia JJ, Chen Y, Terry C, Wartell RM (2011) The stoichiometry of the *Escherichia coli* Hfq protein bound to RNA. *RNA* **17:** 489–500

Urbanowski ML, Stauffer LT, Stauffer GV (2000) The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol Microbiol* **37:** 856–868

Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* **9:** 578–589

Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* **136:** 615–628

Weissenborn DL, Wittekindt N, Larson TJ (1992) Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. *J Biol Chem* **267:** 6122–6131

Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, Dirks RM, Pierce NA (2011) NUPACK: analysis and design of nucleic acid systems. *J Comput Chem* **32:** 170–173

Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, Storz G (1998) The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J* **17:** 6061–6068

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31:** 3406–3415

 ©2012 European Molecular Biology Organization