Functional characterization of bacterial sRNAs using a network biology approach

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1104318108">http://dx.doi.org/10.1073/pnas.1104318108</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Proceedings of the National Academy of Sciences (PNAS)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Jan 12 20:18:48 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/70019">http://hdl.handle.net/1721.1/70019</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Functional characterization of bacterial sRNAs using a network biology approach

Sheetal R. Modi1,2, Diogo M. Camacho3, Michael A. Kohanski4,5, Graham C. Walker6, and James J. Collins7,8,9

Small RNAs (sRNAs) are important components of posttranscriptional regulation. These molecules are prevalent in bacterial and eukaryotic organisms, and involved in a variety of responses to environmental stresses. The functional characterization of sRNAs is challenging and requires highly focused and extensive experimental procedures. Here, using a network biology approach and a compendium of gene expression profiles, we predict functional roles and regulatory interactions for sRNAs in Escherichia coli. We experimentally validate predictions for three sRNAs in our inferred network: IsrA, GlmZ, and GcvB. Specifically, we validate a predicted role for IsrA and GlmZ in the SOS response, and we expand on current knowledge of the GcvB sRNA, demonstrating its broad role in the regulation of amino acid metabolism and transport. We also show, using the inferred network coupled with experiments, that GcvB and Lrp, a transcription factor, repress each other in a mutually inhibitory network. This work shows that a network-based approach can be used to identify the cellular function of sRNAs and characterize the relationship between sRNAs and transcription factors.

Results and Discussion

Small RNA Regulatory Network Inference. We developed a computational biology approach to characterize functional roles for sRNAs in bacteria (see Fig. S1 and SI Materials and Methods for a more complete overview of this method). As a first step, we used the Context Likelihood of Relatedness (CLR) algorithm (16) to infer the sRNA regulatory network in E. coli. The CLR algorithm is an inference approach based on mutual information and allows for the identification of regulatory relationships between biomolecular entities. This algorithm previously has been used to infer transcriptional regulatory networks (17) by examining the functional relationships between transcription factors and target genes. We applied the CLR algorithm to an existing compendium of E. coli microarrays collected under different experimental conditions (Table S1A) to reverse engineer and analyze the regulatory subnetworks for Hfq-dependent sRNAs. This process allowed us to infer potential targets of each of these sRNAs with a highly significant false-discovery rate (FDR)-corrected P-value ($q < 0.05$) (18). The inferred network (Fig. 1 and Table S2A) consists of 459 putative direct and indirect targets for the Hfq-dependent sRNAs, including sRNA–sRNA interactions as well as a number of genes predicted to be coregulated by two sRNAs.

A cellular regulatory scheme in which each transcription factor regulates at least one sRNA (4) has been hypothesized. It is therefore interesting to note that 10 of the sRNAs in the network are predicted to interact with at least one transcription factor, although directionality of regulation is not implied. Transcriptional regulators in the network include LexA (SOS response), FlhD (chemotaxis), and GadE, GadW, and GadX (acid stress response). These network results indicate that regulation of the associated cellular processes may involve a complex interplay between sRNAs, transcription factors, and their respective targets.

We subsequently performed pathway enrichment for each of the inferred sRNA subnetworks, either by Gene Ontology (GO) term enrichment analysis or by using gene function information obtained from EcoCyc (19). These analyses allowed us to classify subnetworks according to function, and thereby implicate the sRNAs as regulators of specific cellular processes (Fig. 1). We were able to identify functional enrichment for seven of the inferred sRNA subnetworks: iron homeostasis (under ryhB regulation), amino acid metabolism (gcvB), motility and chemotaxis (micF), pH adaptation (gadY), DNA repair (glmZ and isrA), protection and adaptation to stress (cyaR), and extracellular transport (dicF). The involvement of RyhB in iron homeostasis and GadY in the regulation of acid response has been reported previously (7, 20). Our network analyses correctly characterized


The authors declare no conflict of interest.

This Direct Submission article had a prearranged editor. Freely available online through the PNAS open access option.

Data deposition: The microarray data reported in this paper have been deposited at www.bu.edu/abl.

1S.R.M. and D.M.C. contributed equally to this work.

2To whom correspondence should be addressed. E-mail: jcollins@bu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1104318108/-/DCSupplemental.

these functions and additionally, suggested important roles for sRNAs in other cellular responses. Network topology also shows connectivity between functional processes. Although direct connections between functional processes may be tenuous, this predicted architecture shows that expression of intermediary genes varies significantly with multiple sRNAs, alluding to themes of overlapping sRNA regulation to coordinate global behavior. The functional annotations in our network, made possible by the identification of a large number of putative targets for Hfq-dependent sRNAs in *E. coli*, provide a basis for further exploration of the functional roles of sRNAs.

The compendium of microarray expression profiles used to reconstruct our regulatory network encompasses broad perturbations, such as different growth conditions and stress inductions (Table S1A). Including chips with sRNA-related genetic perturbations (e.g., *hfq* mutants) and additional environmental perturbations that increase the expression landscape of the cell would improve the algorithm’s performance (Table S1B and Fig. S2). Furthermore, because our approach relies on RNA expression data, our approach is limited to predicting regulation that affects transcript levels. This finding could explain the absence of functional predictions for the *oxyS* and *spf* subnetworks (Fig. 1), as these sRNAs are known to regulate translation of their targets (21–23). Incorporating data at the translational level, such as from 2D gels and mass spectrometry profiling, would improve the predictive power of our approach in the discovery of regulatory roles for sRNAs.

**IsrA and GlmZ Are Involved in the DNA Damage Response.** To assess the validity of our network approach, we chose to explore the predicted involvement of the GlmZ and IsrA sRNAs in the cellular response to DNA damage. GlmZ is known to activate GlmS, a protein involved in the biosynthesis of amino sugars (constituents of the cell wall) to regulate expression based on the availability of external sugars (24). In contrast, no information has been published on IsrA (IS061) since its discovery in a bioinformatics-based screen (25). Our network results show that ~15% of the putative targets for these two sRNAs are involved in the DNA damage response, with 53% of these genes being under the regulation of the LexA repressor protein (Fig. 24 and Table S2 B and C).

To investigate the predicted role of these sRNAs, we treated *E. coli* cultures with DNA-damaging agents, specifically, the gyrase inhibitor norfl Roxacin, mitomycin C (MMC), and UV radiation, and observed their morphology. It is known that the SOS response induces filamentous growth, which is considered to be indicative of the state of DNA damage (26). Although the single-gene deletion strains, ΔisrA and ΔglmZ, did not exhibit a mor-
A functional role for IsrA and GlmZ in the DNA damage response. (A) Inferred network connections for IsrA and GlmZ. Of the identified interactions, 18 are involved in DNA damage pathways. Approximately 50% of these DNA repair genes are members of the LexA regulon. (B) Representative micrographs of MG1655 (Left) and ΔisrA ΔglmZ MG1655 (Right) before (100x objective) and during DNA damage treatment (40x objective). Images show cells during norfloxacin treatment (125 ng/mL, T = 3 h), MMC treatment (2 μg/mL, T = 2 h), and repeated UV exposure (100 J/m², T = 1.5 h). See Materials and Methods for treatment details and Fig. S3 for full micrograph images. (C) Log change in colony-forming units per milliliter (CFU/mL) during DNA damage exposure. Survival of MG1655 (blue diamonds) and ΔisrA ΔglmZ MG1655 (red squares) following exposure to norfloxacin (125 ng/mL), MMC (2 μg/mL), and repeated UV exposure (100 J/m²). In this and all other figures, error bars represent ± SE. (D) Basal mutation rate (mutations per cell per generation) for MG1655 (blue) and ΔisrA ΔglmZ MG1655 (red) using a rifampicin-based selection method. Wild-type mutation rate is similar to that previously reported (28).
Although the algorithm does not predict \_lp\_ to be a target of GcvB, it does predict two targets in our network, \_trpE\_ and \_lvk\_. The GcvB-\_trpE\_ and GcvB-\_lvk\_ binding regions predicted by TargetRNA overlap and together span positions 65 to 91 on the GcvB transcript. We searched for homologies of this region’s complement within 100 bp of the \_lp\_ translational start and identified a putative binding site for GcvB in the \_lp\_ 5’ UTR (Fig. S6A).

To examine this putative direct interaction between GcvB and Lrp, we used an \_lp\_ gene fusion to GFP to function as a reporter of translational control by GcvB. Our translational fusion consisted of the 5’ UTR of Lrp and the first 15 amino acids fused to the N terminal of GFP, and was constructed using the modular plasmid system described by Urban and Vogel (11). This system was designed to confirm sRNA-mediated control of mRNA targets through its ability to uncouple both species from the chromosomal regulatory network and to reliably suppress pleiotropic effects of sRNA expression on target fusion transcription. Expressing GcvB, we observed an approximately twofold decrease in fluorescence of the \_lp:\_gfp\_ fusion compared with a control plasmid (Fig. 4A, Left). We used a \_dppA:\_gfp\_ fusion as a positive control for our expression system (Fig. 4A, Left) and obtained results for this known GcvB target that were consistent with those previously reported (11). As a control to address potential indirect regulation of \_lp:\_gfp\_, we experimentally demonstrated that the MicF sRNA, which is not predicted to interact with Lrp, had no effect on the Lrp fusion (Fig. S6B).

To obtain additional evidence of the interaction between GcvB and Lrp, we mutated the predicted binding region in the 5’ UTR of \_lp\_. Four base-pair mutations were made to our target fusion—specifically, A(-9)T, C(-8)G, A(-7)T, and A(-6)T—where base position is with respect to the \_lp\_ translational start. These mutations eliminated GcvB repression of the \_lp\_ transcript (Fig. 4A, Right). Taken together, these results demonstrate the direct posttranscriptional repression of Lrp by GcvB and offer an sRNA-transcription-factor regulation scheme for the control of amino acid availability.

**gcvB Is Regulated by Lrp.** Analysis of our microarray compendium revealed that expression of \_gcvB\_ and \_lp\_ are anticorrelated, independent of growth phase, suggesting that these genes function in a complex regulatory circuit. Mutually regulating elements endow networks with interesting properties, such as bistability and memory (35, 36). There is precedence for this type of motif at the posttranscriptional level in eukaryotes (37), and many other network architectures have been demonstrated in bacterial sRNA regulation (38, 39). However, mutually inhibitory networks involving sRNAs have not been found in bacteria. We hypothesized that Lrp and GcvB function together in a mutually inhibitory network for controlled pathway regulation of cellular amino acid availability.

Building on our results that establish GcvB regulation of Lrp (Fig. 4A), we sought to explore Lrp regulation of GcvB. In elucidating the relationship of Lrp to gcvB, it was important to do so within the context of known regulation. GcvB is activated by the glycine cleavage system regulator, GcvA, under glycine-rich conditions (40). This interaction is dependent upon Lrp binding and is negatively regulated by GcvR when glycine is limiting (41). Using quantitative PCR, we measured relative expression levels of \_gcvB\_ in wild-type, \_gcvA\_, and \_lp\_, with and without glycine addition (Fig. 4B). We found that \_gcvB\_ expression is significantly lower in \_gcvA\_ under glycine-rich conditions, confirming known regulation. Interestingly, we also found that \_gcvB\_ transcript levels are ~30-fold greater in \_lp\_ compared with wild-type, independent of glycine addition. Because Lrp is a central regulator of cellular processes, it is possible that Lrp mediates negative regulation of \_gcvB\_ indirectly. To investigate dependence on GcvA, we compared relative \_gcvB\_ expression in \_gcvA\_Δ\_lp\_ and \_gcvA\_ (Fig. S6C) and found that significantly higher levels of \_gcvB\_ are present in \_gcvA\_Δ\_lp\_ strain, demonstrating that Lrp regulation is not mediated exclusively through GcvA.

We next used sequence analysis to look for evidence that may suggest a direct interaction of Lrp on \_gcvB\_. We used ClustalW2 (42) to search for known Lrp-binding consensus sequences (43, 44) in the 500-bp region upstream of \_gcvB\_. Homology results indicate that there are two putative binding sites for Lrp in this region (Fig. S6D). These data suggest a direct interaction of Lrp on \_gcvB\_; however, an indirect regulation scheme remains possible and cannot be excluded.

Collectively, our analyses indicate that Lrp and GcvB repress each other (directly or indirectly) in a mutually inhibitory network (Fig. 4C). We speculate that E. coli can use this dual repression scheme to create a controlled response to changing
amino acid availability in the environment, allowing for robust adaptation and resource conservation.

Conclusions

In this work, we have shown how a network biology approach can be used to elucidate bacterial sRNA functional and regulatory roles. Although our network map does not discern between direct and indirect sRNA-gene interactions, our methodology can also be used to improve current methods of sRNA target identification. Direct interactions within our network can be uncovered by filtering network predictions with sequence-alignment tools or other target detection methods (14), as we illustrate for GcvB-mediated regulation of Lrp. Examining expression levels of predicted sRNA targets in relevantly perturbed hfq mutant strains may provide additional confirmation of direct interactions.

The approach described in this work, which relies on compendia of expression data, can be readily extended to other organisms and used to characterize sRNAs in pathogens as well as microRNAs in eukaryotes. Efforts along these lines could enhance our understanding of the posttranscriptional regulatory events that lead to pathogenicity or disease states, like cancer. Furthermore, as efforts to discover novel sRNAs lead to larger and more extensive RNA-seq expression datasets, our network biology approach could enrich the information found in these expression profiles to infer function of newly discovered sRNAs.

The present study also highlights how large-scale biomolecular networks can be used to guide the discovery and detailed experimental investigation of small-scale networks. In our case, a large-scale, reconstructed sRNA regulatory network enabled us to uncover an intriguing mutually inhibitory network made up of a small RNA and a transcription factor.

Understanding the regulatory roles of noncoding RNAs in prokaryotic and eukaryotic regulation presents an exciting challenge. Reverse genetics methodologies aspire to illuminate the regulatory roles of noncoding RNAs in pathogenicity or disease states, like cancer. Furthermore, as efforts to discover novel sRNAs lead to larger and more extensive RNA-seq expression datasets, our network biology approach could enrich the information found in these expression profiles to infer function of newly discovered sRNAs.

Materials and Methods

Network Inference. The sRNA network was inferred using the CLR algorithm. The algorithm uses mutual information to score the similarity between expression levels of two genes in a set of microarrays and applies an adaptive background correction step to eliminate false correlations and indirect influences (16). A gene pair is predicted to interact if their mutual information score is larger than an FDR-corrected score (18) at a given significance threshold (q < 0.005). The data used as input to the algorithm was an existing compendium of 759 Affymetrix E. coli Antisense microarray chips normalized as a group with RMA. The compendium includes arrays from the Microbiome Microarray Database (E. coli_v3_build_3), as well as 235 arrays run in-house, with experiments involving antibiotic treatment, biofilm growth, different growth media, acid shifts, anaerobic growth, as well as various perturbations of coding genes (Table S1A).

To gain insight into the functional roles of Hfq-dependent sRNAs, we performed pathway enrichment for each of the inferred sRNA subnetworks, either by GO term enrichment analysis (P value < 0.05, minimum GO term depth of 3) or by using gene function information obtained from EcoCyc (19). See SI Materials and Methods for more details and references on network analysis.

Media and Growth Conditions. Cultures were grown at 37 °C in Luria-Bertani broth (Fisher Scientific), M9 minimal media supplemented with 0.4% glucose (Fisher Scientific), or E. coli Rich Defined media (Teknova). Antibiotics were added to the growth media for selection at the following concentrations: chloramphenicol (30 μg/mL), Acrös Organics) and ampicillin (100 μg/mL; Fisher Scientific). Amino acids in rich media to model conditions in which it was originally tested (11). Lrp:gfp and Lrp-mut:gfp were grown in M9 minimal media to assess the effects of gcvB expression in nutrient-limiting conditions. Unregulated target fusion specific fluorescence (expressing control vector) is shown in gray, and regulated target fusion specific fluorescence (expressing gcvB) is shown in orange. See Materials and Methods and SI Materials and Methods for details on fluorescence measurements and calculations. Asterisks represent significant (P < 0.05) differences between unregulated and regulated target fusion specific fluorescence. (B) Fold-difference in gcvB expression in ∆gcvA and ∆lrp relative to wild-type during growth in M9 minimal media. Blue bars represent relative expression when exogenous glycine was absent, and red bars represent relative expression when glycine (300 μg/mL) was added to the media. Error bars represent propagated error measures. (C) GcvB-Lrp regulatory subnetwork resulting from translational fusion, expression data, and known regulatory interactions. Unsequenced GcvA activates gcvB expression when glycine is present. GcvB directly represses Lrp and Lrp directly or indirectly represses gcvB.

Strains and Plasmids. All experiments were performed with E. coli MG1655 (ATCC 700926)-derived strains (Table S3A). Gene deletions were derived either from p1 phage transduction from the Keio collection (45) or by the PCR-based phage-λ recombine method (46). Expression vectors of gcvB were derived from the pZ system (47). Translational fusion-related plasmids, pXG-1, pXG-0, pXG-10, and pSK-015, were used for constructing and examining the target fusions mentioned above, and were kindly provided by Jörg Vogel, Institute for Molecular Infection Biology, University of Würzburg, Würzburg, Germany. Site-directed mutagenesis was performed using the Phusion Site-Directed Mutagenesis kit (New England Biolabs). See SI Materials and Methods for details on plasmid construction.

DNA Damage Sensitivity Assays. Cultures of the various strains were grown in 25 mL LB medium in 250 mL flasks to an OD600 of 0.3 (time 0), at which time they were exposed to three different DNA damaging agents (norfoxacin, MMC, and UV light). For norfoxacin-treated cultures, norfoxacin was added at a concentration of 125 ng/mL for MMC-treated cultures, MMC was added at a concentration of 2 μg/mL. UV treatment was delivered using a Stratalinker UV box (Stratagene) such that cultures were exposed to 100 J/m² of UV radiation every 30 min over the course of 2 h. Strain viability was assessed by collecting aliquots of the cultures at time 0 (just before exposure to the
DNA damaging agent) and at 1-h intervals (norfloxicin and MMC) or 30-min intervals (UV) following exposure to the DNA damaging agent. Viability of the various strains was determined by measuring the CFU per milliliter, as described previously (48). Briefly, serially diluted cells were spot plated on LB-agar plates and grown overnight at 37 °C. Colonies were counted at those dilutions with ~10 to 50 cells, and CFU per milliliter was calculated using the following formula: CFU/mL = [(# of colonies) ÷ (dilution factor)] × 0.01 mL. Average CFU per milliliter was determined based on the results of three biological replicates.

Translational Fusion Experiments and Calculations. At inoculation, cultures were induced with 1 mM IPTG (Invitrogen) for gcvB expression and grown to an OD_{600} of 0.3 in M9 minimal media (lrp::gfp and lrp-mut::gfp) or EZ rich media (pSK-015). Fluorescence measurements were taken, and relative fluorescence values were calculated as previously described (11) using the following formula: fold-change mediated by sRNA = (fluorescence of regulated target fusion specific fluorescence)(unregulated target fusion specific fluorescence), where regulated target fusion specific fluorescence = (fluorescence(gcvB + pZA12-gcvB + target fusion) – fluorescence(gcvB + pZA12-gcvB + pxg-0)) and unregulated target fusion specific fluorescence = (fluorescence(gcvB + pZA12-null + target fusion) – fluorescence(gcvB + pZA12-null + pxg-0)). The effect of micf expression on lrp::gfp was calculated similarly. See SI Materials and Methods for more details.