Functional characterization of bacterial sRNAs using a network biology approach

Sheetal R. Modi, Diogo M. Camacho, Michael A. Kohanski, Graham C. Walker, and James J. Collins

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 < 0.005) (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), FliH (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 *ryb* 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 RybB in iron homeostasis and GadY in the regulation of acid response has been reported previously (7, 20). Our network analyses correctly characterized...
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., hfg 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 (15061) 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. 2A 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 norflaxacin, 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-
phological phenotype that was different from wild-type (Fig. S3), the double-deletion strain, ∆isrA∆glmZ, did show substantially less filamentation (Fig. 2B and Fig. S3).

We next sought to determine the effects of IsrA and GlmZ on cell survival under DNA damage. We measured cell viability following treatment with norfloxacin, MMC, and UV, and found that the double-deletion strain was significantly less sensitive than wild-type to each of the treatments (Fig. 2C).

Because cells experience low levels of DNA damage under normal physiological conditions (27), we were also interested in determining if the basal mutation rate of the ∆isrA∆glmZ strain differed from wild-type in unperturbed conditions. We found that the mutation rate of the sRNA double mutant is approximately threefold less than that of wild-type (Fig. 2D).

Together, these results suggest that IsrA and GlmZ function as DNA repair regulators, possibly in a redundant manner. We speculate that these two sRNAs act by differentially affecting the regulation of specific genes within the LexA regulon. The SOS system is an important stress response that has been shown to play a central role in a variety of mechanisms, including antibiotic tolerance (17) and antibiotic-resistant gene transfer (29). Our work indicates that the sRNAs IsrA and GlmZ may play critical regulatory roles in this important cellular stress response.

**GcvB Is Involved in the Regulation of Amino Acid Availability.**

We next chose to examine the inferred subnetwork for the GcvB sRNA. GcvB has been shown to regulate peptide transport (30, 31) and acid stress in *E. coli* (32). Functional enrichment of the genes predicted to interact with GcvB in our inferred network (Fig. 3A and Table S2D) revealed that ~50% of them are involved in amino acid transport and metabolic processes, suggesting a broader role for GcvB in nutrient availability.

To assess this predicted role for GcvB, we measured growth rates of strains cultured in minimal media supplemented with different amino acids. Growth experiments in varying nutrient conditions have been used to implicate a number of genes in metabolism and transport. In our study, we compared the doubling times of two mutants, an *gcvB* strain and a strain constitutively expressing *gcvB*, to that of wild-type. We observed that the growth rate of the mutants did not differ from wild-type in unsupplemented conditions (Fig. S4). However, the growth rates in one or both of the *gcvB* strains were significantly different when supplemented with leucine, serine, phenylalanine, or threonine (Fig. 3B). These results support our network-derived hypothesis that GcvB plays a broad role in the regulation of amino acid availability and metabolism.

**GcvB Represses Lrp.**

Among the genes predicted to interact with GcvB is *lrp*, which encodes the Lrp transcription factor, an important regulator of amino acid availability (33). We hypothesized that GcvB regulates amino acid pathways through modulation of Lrp.

Sequence information and corresponding secondary structure have been used to uncover sRNA targets (31). Accordingly, we have used TargetRNA, a sequence-based algorithm with high-performance capabilities (34), to explore the possibility of a physical interaction between GcvB and Lrp. TargetRNA identifies 21 putative targets of GcvB in *E. coli* using default parameters (Fig. S5).

**Fig. 2.** 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 ∆isrA∆glmZ MG1655 (Right) before (100× objective) and during DNA damage treatment (40× 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 ∆isrA∆glmZ 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 ∆isrA∆glmZ MG1655 (red) using a rifampicin-based selection method. Wild-type mutation rate is similar to that previously reported (28).

![Fig. 2](https://example.com/fig2.png)
Although the algorithm does not predict _lp_ to be a target of GcvB, it does predict two targets in our network, _trpE_ and _livK_. The GcvB-<i>trpE</i> and GcvB-<i>livK</i> 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 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 two-fold 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 Lrp 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, _ΔgcvA_, and _Δlp_, with and without glycine addition (Fig. 4B). We found that _gcvB_ expression is significantly lower in _ΔgcvA_ under glycine-rich conditions, confirming known regulation. Interestingly, we also found that _gcvB_ transcript levels are ~30-fold greater in _Δlp_ 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 _ΔgcvA_Δlp_ and _ΔgcvA_ (Fig. S6C) and found that significantly higher levels of _gcvB_ are present in _ΔgcvA_Δlp_ 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 ClustalW (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 conditions.
The algorithm uses mutual information to score the similarity between ex- Network Inference. and indirect sRNA-gene interactions, our methodology can also roles. Although our network map does not discern between direct be used to elucidate bacterial sRNA functional and regulatory In this work, we have shown how a network biology approach can adaptation and resource conservation. Conclusions In this work, we have shown how a network biology approach can to be uncovered by filtering network predictions with sequence-alignment tools or other Direct interactions within our network can be uncovered by targets. The approach described in this work, which relies on compendia of their in- gangenicity or disease states, like cancer. Furthermore, as filters to discover novel sRNAs lead to larger and more extensive networks can be used to guide the discovery and detailed ex- further 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 chal- of their influence is slowly being uncovered. Our work shows that network biology approaches can make significant contributions to these efforts and facilitate the efficient reconstruction of functional and regulatory maps. 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 signification threshold (q < 0.005). The data used as input to the algorithm was an existing compendium of 759 Affymetrix E. coli Antisense2 microarray chips normalized as a group with RMA. The compendium includes arrays from the Many Microbe 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. Rich Defined media (Teknova). Antibiotics were added to the growth media for selection at the following concentrations: chloramphenicol (30 μg/mL; Acros 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 gcv 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 un- regulated and regulated target fusion specific fluorescence. (b) Fold-difference in gcvB expression in ΔgcvA and Δlrp 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.

![Fig. 4.](image)

**A** GFP translational fusions for dppA and lrp (Left) and lrp and lrp-mut (Right). dppA::gfp (plasmid pSK-015) was grown 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 gcv 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 ΔgcvA and Δlrp 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.

![Fig. 4.](image)
DNA damaging agent) and at 1-h intervals (norfloxacin and MMC) or 30-min intervals (UV) following exposure to the DNA damaging agent. Viability of the various strains at each time point was determined by measuring the CFU per milliliter, as described previously (48). Briefly, serially diluted cells were spotted 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)]/0.10 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 OD600 of 0.3 in M9 minimal media (lrp::gfp and lrp-mut:grp) 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 = (regulated target fusion specific fluorescence)/(unregulated target fusion specific fluorescence), where regulated target fusion specific fluorescence = ([fluorescence(gcvB + pZ12-gcvB + target fusion]) – fluorescence(gcvB + pZ12-gcvB + pxG-0)) and unregulated target fusion specific fluorescence = ([fluorescence(gcvB + pZ12-null + target fusion]) – fluorescence(gcvB + pZ12-null + pxG-0)]. The effect of micf expression on lrp::gfp was calculated similarly. See **SI Materials and Methods** for details.

**cDNA Synthesis and qPCR.** Quantitative PCR was performed using the Roche LightCycler 480 and the LightCycler 480 SYBR Green I Master Kit (Roche Applied Science) according to the manufacturer's instructions. Relative quantification of gcvB was determined by the ΔΔCp method using rsrB (165 bp) as a reference gene. Fold-changes were calculated by comparing relative expression across the same conditions in the strains mentioned in the text. See **SI Materials and Methods** and **Table S3** for more details.

**Statistical Analysis.** All data are representative of mean values of replicates, except in Fig. 2D, where the median was used to calculate the mutation rate. Error bars represent ±SE, which was propagated when necessary as described by others (49). Statistical significance was calculated between data sets using a two-tailed t test assuming unequal variance in the population.

**ACKNOWLEDGMENTS.** We thank Susan Gottesman and Gisela Storz for insightful discussions and comments on our work, as well as detailed information on Hfq-dependent small RNAs in Escherichia coli; and Jörg Vogel for the pxG-0, pxG-1, pxG-10, and pSK-015 plasmids. This work was supported by the National Institutes of Health Director's Pioneer Award Program Grant DP1 OD003644 and The Howard Hughes Medical Institute, and National Institutes of Health Grants CA21615 and GM31030 (to G.C.W.). G.C.W. is an American Cancer Society Professor.