Stochastic Analysis Probes the Role of the Chaperone Hfq in *E. coli* Sugar Shock Response

David M. Bianchi[1,2], **Troy A. Brier**[1,2], **Anustup Poddar**[2,3,4], **Muhammad S. Azam**[5,§], **Carin K. Vanderpool**[5], **Taekjip Ha**[2,3,4], **Zaida Luthey–Schulten**[1,2,*]

***For correspondence:** zan@illinois.edu

Present address: [§]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA

[1]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA; [2]Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL, USA; [3]Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA; [4]Howard Hughes Medical Institute, USA; [5]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Abstract Small RNAs (sRNAs) play a crucial role in the regulation of bacterial gene expression by silencing the translation of target mRNAs. SgrS is an sRNA that relieves glucose–phosphate stress, or “sugar shock” in *E. coli*. The power of single cell measurements is their ability to obtain population level statistics that illustrate cell–to–cell variation. Here, we utilize single molecule super–resolution microscopy in single *E. coli* cells coupled with stochastic modeling to analyze glucose–phosphate stress regulation by SgrS. We present a kinetic model that captures the combined effects of transcriptional regulation, gene replication and chaperone mediated RNA silencing in the SgrS regulatory network. This more complete kinetic description, simulated stochastically, recapitulates experimentally observed cellular heterogeneity and is used to describe the dynamics of SgrS stabilization by the chaperone protein Hfq.

Introduction

The ability of living cells to modulate their gene expression in response to changing environmental conditions is critical to their growth and continued development. Many bacteria use the phosphoenolpyruvate phosphotransferase (PTS) system to transport and phosphorylate incoming sugars to prepare them for subsequent glycolytic metabolism. The uptake of phosphosugars must be balanced with their breakdown in order to prevent metabolic stress. In *E. coli*, a stress response induced by unbalanced glucose–phosphate transport and metabolism or “sugar shock", is referred to as glucose–phosphate stress response. A primary activity of this stress response is RNA silencing of *ptsG*, a gene coding for the glucose transport protein of the same name (also known as EIICBGlc in *E. coli*), by the small RNA (sRNA) SgrS. Small RNAs are usually non–coding RNA molecules that act by base pairing with target messengers to regulate translation or mRNA stability and have been observed across all domains of life (***Babski et al., 2014***). *sgrS* is upregulated by a transcriptional activator (SgrR) when the cell is under a state of glucose–phosphate stress. SgrS regulates *ptsG* post-transcriptionally by a mechanism where SgrS binds to *ptsG* messenger RNA (mRNA) and prevents its translation to protein by blocking access of the ribosome to the mRNA (***Vanderpool and Gottesman, 2004***; ***Maki et al., 2010***). This also enhances the co-degradation of *ptsG* mRNA and SgrS via enzymes responsible for the removal of bulk RNA such as ribonuclease E (RNase E) (***Kawamoto et al., 2006***; ***Maki et al., 2010***). The co-degradation reduces the number of PtsG sugar transporter

proteins that are produced and thus reduces the impact of glucose–phosphate stress, since fewer transport proteins are available to bring sugar into the cell.

SgrS and *ptsG* mRNA associate via complementary base pairing that occludes the ribosome binding site on the mRNA. Recently, this mechanism has been analyzed in conjunction with binding of the Sm-like chaperone protein Hfq to SgrS, which stabilizes the sRNA, and facilitates the interaction between the sRNA and mRNA (***Ishikawa et al., 2012***). Hfq also promotes SgrS–dependent regulation of other targets involved in sugar shock such as *manXYZ*, and *yigL* in *E. coli*. In this study, we focus only on the primary regulatory target *ptsG* and do not consider the other targets of SgrS regulon, which are described in ***Bobrovskyy et al.*** (***2019***).

Previous experimental and theoretical work (***Jones et al., 2014***; ***Peterson et al., 2015***) has demonstrated the necessity of accounting for gene replication over the course of the cell cycle in order to capture the population variation observed in messenger RNA abundance. The additional noise emanating from transcription at multiple gene loci manifests itself in observed broad mRNA copy number distributions in a population of cells. This work also demonstrated that including the effect of gene regulation by transcription factors can be critical in order to appropriately describe stochastic dynamics. The effect of transcriptional regulation is apparent in the SgrS–*ptsG* mRNA system, where the expression of SgrS is maintained by the regulator SgrR, which activates *sgrS* and autorepresses its own expression during glucose-phosphate stress conditions (***Vanderpool and Gottesman, 2004***, ***2007b***).

Recently, ***Fei et al.*** (***2015***) presented a deterministic kinetic model of the SgrS mediated regulation of *ptsG* mRNA in *E. coli*. Using single-molecule fluorescence experiments (smFISH and STORM), SgrS and *ptsG* mRNA copy numbers in cells were measured, which produced distributions of RNA at various time points after the induction of sugar stress across a population of fast-growing *E. coli*. However, it is important to note that both the *ptsG* mRNA and the SgrS regulating it are present in low copy number (a few to tens of particles) and therefore exhibit intrinsically noisy behavior in both their gene expression and regulatory behaviors. For this reason it is most appropriate to treat the regulatory network via stochastic simulation in order to quantify the variation that is observed across a population of cells, which has been demonstrated previously (***Elowitz et al., 2002***; ***Raser, 2005***; ***Earnest et al., 2018***).

Here, we have developed a stochastic model, to our knowledge the first of its kind for RNA silencing, that captures the mRNA and sRNA distributions observed in a population of hundreds of *E. coli* cells. The stochastic model additionally incorporates the following features that extend the platform given by ***Fei et al.*** (***2015***): (1) accounting for gene replication, (2) transcriptional gene regulation of *sgrS* by its activator SgrR and (3) explicit representation of the SgrS stabilization via the Hfq chaperone protein. This model robustly describes experimentally observed RNA distributions, closely matching regulatory dynamics from immediately after induction until a steady state is reached 20 minutes later. We also utilize this model to analyze the effects of the size of the pool of Hfq chaperone protein available, the rate of Hfq binding to SgrS to stabilize it and enhance its binding to the target, *ptsG* mRNA, and the effect of an SgrS point mutation in the SgrS-Hfq binding region on the regulatory network.

Model

The previous kinetic model for SgrS regulation of *ptsG* mRNA (***Fei et al., 2015***) utilized simple mass-action kinetics to describe the target search process and modeled gene expression as a constitutive process, with RNA species originating from a single gene copy. Despite its simplicity, this model captures average regulatory network behavior and also gives insight into many of the parameters required for the more descriptive stochastic model that is the focus of this work. For example, since a binding rate for SgrS to *ptsG* mRNA was established in ***Fei et al.*** (***2015***) we are now able to complexify the model by the addition of the chaperone protein Hfq, which allowed us to predict (by fitting to the experimental data) the size of the pool of Hfq available to stabilize SgrS and the rate at which it binds the sRNA.

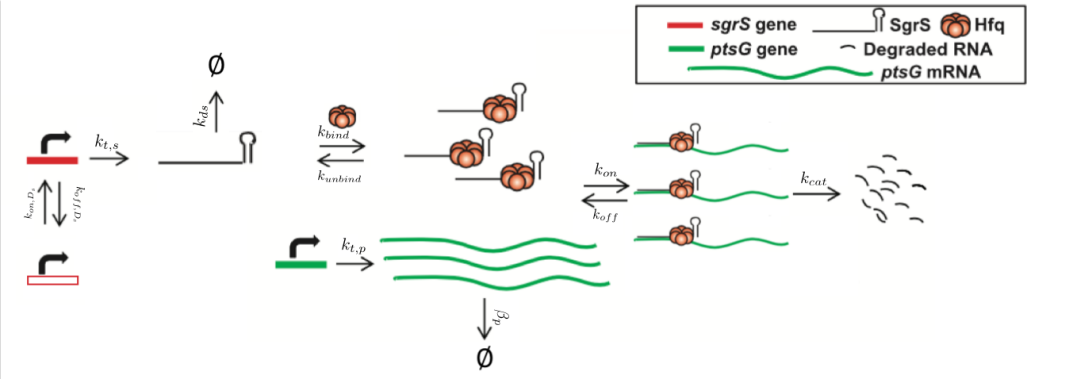


Figure 1. Schematic of the kinetic model as described in the text. The RNA species are transcribed from a sampled genome state with *sgrS* capable of switching between an on and off state. Explicitly represented Hfq can bind and unbind with SgrS, and then the Hfq–SgrS complex binds (and potentially unbinds) with *ptsG* mRNA. All degradation events are carried out by the enzyme RNase E. See ***Figure 4*** for the kinetic equations described above.

The kinetic model was implemented and solved stochastically as a well-mixed Chemical Master Equation (CME) in the Lattice Microbes (LM) simulation software suite (***Roberts et al., 2013***; ***Hallock et al., 2014***; ***Hallock and Luthey-Schulten, 2016***; ***Peterson et al., 2013***). The corresponding rate constants (***Table 1***) were adapted from the kinetic model described above (***Figure 1***). One important feature added to the model is the explicit presence of the chaperone protein Hfq, which has been shown to both stabilize SgrS (substantially increasing its half-life) and facilitate the association of SgrS to *ptsG* mRNA (***Hopkins et al., 2011***; ***Wagner, 2013***; ***Vanderpool and Gottesman, 2004***; ***Santiago-Frangos and Woodson, 2018***). In order to capture the cell-to-cell heterogeneity due to the small number of particles (*e.g.*, gene copies) involved in transcription, it is critical to account for transcriptional regulation of the genes involved in the glucose-phosphate stress response. For this reason, we include the transcriptional activation of *sgrS* by the transcription factor SgrR, which has been shown to upregulate expression in the presence of αMG (***Vanderpool and Gottesman, 2007a***, ***2004***). Regulation of *ptsG* by the transcriptional repressor Mlc was not included in the model since repression is relieved in the presence of glucoside sugars. With αMG present, Mlc is sequestered at the membrane by binding the EIIB subunit of the PtsG transporter protein complex (***Lee, 2000***; ***Seitz et al., 2003***; ***Nam et al., 2008***), relieving repression and resulting in high levels of *ptsG* transcriptional activity (***Balasubramanian and Vanderpool, 2013***). Since the decay time of PtsG proteins is expected to be approximately on the order of eight hours (***Maier et al., 2011***), much longer than the timescale of mRNA decay, Mlc repressors are likely still sequestered by the transporters at the membrane and have little effect on the SgrS regulatory process. Rates for the association of the Hfq-SgrS complex to *ptsG* mRNA (k_{on}) and the dissociation of the Hfq-SgrS-*ptsG* mRNA complex (k_{off}) were obtained from ***Fei et al.*** (***2015***), which did not include Hfq explicitly but provides the corresponding association and dissociation reaction rates. The value for the co-degradation rate of SgrS and *ptsG* mRNA from the Hfq-SgrS-*ptsG* mRNA complex by RNase E (k_{cat}) is also obtained from ***Fei et al.*** (***2015***) (see Section 'Methods and Materials').

Calculation of Gene Copy Number

Finally, and critically, in order to appropriately capture regulatory effects on gene expression of SgrS and *ptsG* mRNA, it is important to account for gene duplication, as we have previously shown (***Peterson et al., 2015***). As illustrated by ***Jones et al.*** (***2014***) since the approximate time to replicate the *E. coli* genome (approximately 40 minutes) (***Cooper and Helmstetter, 1968***) is longer than the fast-growing *E. coli* cell division time of 20 minutes (or the 35 minutes observed in our experiments), the cell has nested replication forks that are already replicating the genomes of daughter and granddaughter cells prior to cell division. In particular, genes close to the origin of

replication are likely to have multiple copies present over much of the cell cycle. This phenomenon has been shown previously for genes near the origin in *E. coli* by both isotopic labeling of nucleotides and imaging of fluorescent chromosome markers (***Cooper and Helmstetter, 1968***; ***Youngren et al., 2014***). Due to the position of *sgrS* (only 6° away along the circular chromosome) very near to the origin of replication, it is likely that multiple gene copies are accessible for transcription over the course of the cell cycle. About half-way between the origin and terminus of replication (at approximately 90°) *ptsG* is also likely to have multiple gene copies present at some point over the course of the cell cycle, although at lower copy number than *sgrS*. ***Figure 2*** depicts the two genes and their location along the circular *E. coli* genome.

The experimentally measured cells were unsynchronized and should have multiple replication forks present over the course of the 20 minutes post–induction, our measurement window. To account for gene duplication effects in the population of unsynchronized cells, we sample the percentage of the cellular population in either a low or high gene state, which corresponds to the expected distribution of the number of genes present over the course of the cell cycle after induction. In this way, we effectively flip a coin to decide whether a simulation replicate corresponding to an individual experimentally imaged *E. coli* cell has 2 copies (low gene state) of or 4 copies (high gene state) of *sgrS* and similarly 1 or 2 copies of *ptsG*. This allows us to account for the effect of gene duplication in generating mRNA noise over the heterogeneous population of hundreds of *E. coli* cells that were observed experimentally. We assume that all gene copies are transcribed independently from one another and at the same rate, a notion that ***Wang et al.*** (***2019***) has recently examined in *E. coli* under various growth conditions. Under similar growth conditions to ours (MOPS glucose-based medium with a doubling time of 35 minutes, (see Section 'Methods and Materials')), the data from ***Wang et al.*** (***2019***) suggest that transcription does appear to be independent and uncorrelated between copies of the same gene (***Wang et al., 2019***).

Figure 3 illustrates the reasoning for the specific choices of high and low state gene copy numbers for *ptsG* and *sgrS* in an *E. coli* cell growing faster than the expected time necessary for replication (approximately 40 minutes, compared to an experimentally observed generation time of approximately 35 minutes) (***Cooper and Helmstetter, 1968***; ***Youngren et al., 2014***).

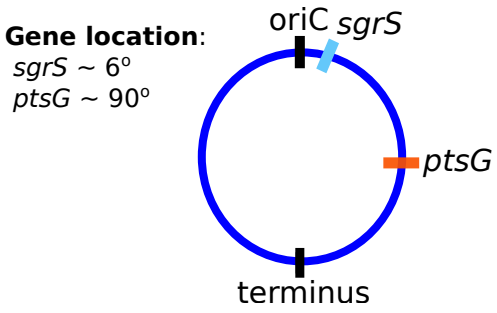


Figure 2. The gene location for SgrS and *ptsG* mRNA relative to the origin of replication (*oriC*) are shown on the circular chromosome of the *E. coli* cells used for this study. As it is closer to the origin of replication *sgrS* (cyan) is likely to be present in higher gene copy number than *ptsG* (in orange), which is farther away from the *oriC* as shown.

Stochastic simulations were performed by sampling the CME for the model given in ***Figure 1*** with the widely used Gillespie Direct Method of the Stochastic Simulation Algorithm (SSA), which is available through the publicly available Lattice Microbes (LM) software suite (version 2.3 was used) and its python interface pyLM (***Roberts et al., 2013***; ***Hallock et al., 2014***; ***Hallock and Luthey-Schulten, 2016***; ***Peterson et al., 2013***). We ran 2000 replicate simulations for 25 minutes after αMG induction of glucose-phosphate stress in order to match the corresponding 20-minute smFISH-STORM experiments. Initial conditions for basal SgrS (1-3 copies) and *ptsG* mRNA (30-40 copies) copy number were sampled from the experimentally measured distributions and rounded to the nearest integer particle number. Simulations were computed on a local cluster containing AMD Opteron Interlagos cores.

SgrS Regulatory Network Kinetic Model

The kinetic model describing the reactions characterizing the *E. coli* glucose–phosphate response network by the small RNA SgrS is given in ***Figure 4***. Simulation files are available in Jupyter Notebook

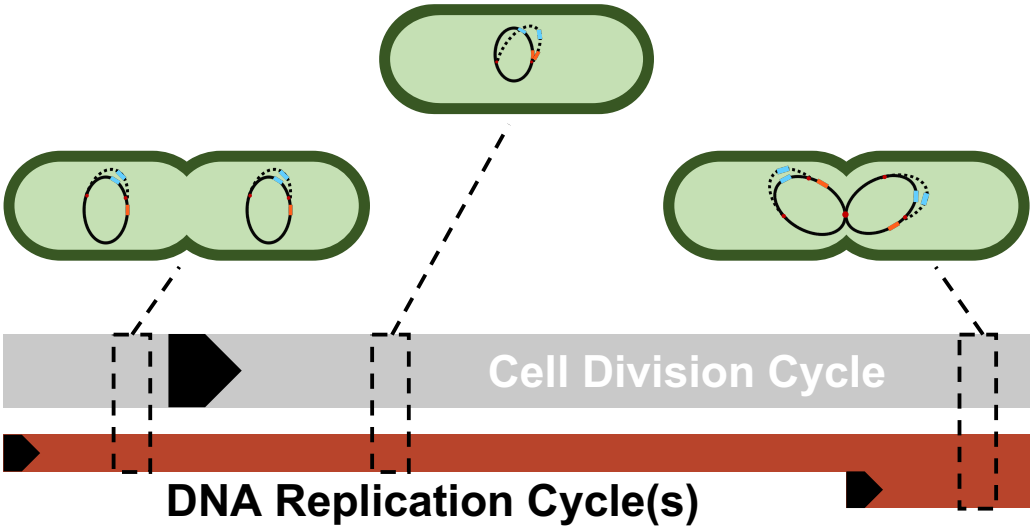


Figure 3. A simplified depiction of possible cellular states throughout a single DNA replication cycle. Each cell shows a snapshot of the gene state of a cell given its progression through the DNA replication and cell division cycle. Due to the difference in lengths of the cell division cycle (~35 mins) and DNA replication cycle (~40 mins), DNA replication and cell division are not completely in sync. Multiple replication forks (red dots) can form on the genome in order to ensure DNA is duplicated properly in these fast-growing cells. As a result, genes closer to the origin such as *sgrS* (blue) are duplicated in the same timeframe that replication is initiated (resulting in 2 or 4 gene copies), while genes closer to the terminus such as *ptsG* (orange) are replicated in the C period, the period when a majority of DNA is duplicated (resulting in 1 or 2 gene copies). The black arrows denote the start of a cycle.

format to be simulated via the Lattice Microbes (LM) Software Package at **Add link to jupyter notebook**.

Results

Figure 5 demonstrates the ability of our newly developed kinetic model to capture the average cellular copy number of SgrS and *ptsG* mRNA over the course of the 20 minute period post-induction. The overlap of the interquartile range of both the experimental and simulated cellular populations demonstrates the agreement over a variety of cells (at different gene states (*i.e* high/low copy number), and RNA expression levels).

The ability of our improved kinetic model to capture population-level statistics of single cell copy number distributions of SgrS and *ptsG* mRNA is demonstrated in ***Figure 6***. Kernel Density Estimates (KDE), which are used to estimate the probability densities of distributions of approximately 100–200 experimentally measured cells and 2000 simulated cells are displayed, along with dashed vertical lines giving the average RNA copy numbers observed. KDEs were utilized to provide a reasonable comparison to the experimental values despite the fact that there were a relatively low number of cells measured at each time point (approximately 100–200) compared to the number of replicates required for appropriate stochastic simulation (2000). The distributions obtained from both experiment and the kinetic model show strong agreement (especially in the case of *ptsG* mRNA), which can be seen quantitatively by the starred line showing the Kulback–Leibler Divergence (KL Divergence) in ***Figure 7***. The KL Divergence (***Equation 2***), which was minimized to fit to experimental RNA distributions over all time points, is a statistical measure used to characterize the difference between a probability distribution (the KDE of simulated cells) and a reference distribution (the KDE of experimentally measured cells).

The parameters obtained from the fitting process give some insight into the role of stabilization by Hfq in the SgrS-*ptsG* mRNA target search process and the role of transcriptional regulation

by SgrR in the regulatory network. The pseudo first order rate of Hfq binding to SgrS (k_{bind}) is $0.063\ s^{-1}$, while the degradation rate of SgrS (k_{ds}), obtained from Δ*hfq* strain experiments (described in Section 'Methods and Materials'), is $0.022\ s^{-1}$. The available Hfq pool size of 250 predicted by fitting to the kinetic model seems reasonable in that average proteomics values have been found to be on the order 1500 (***Taniguchi et al., 2010***; ***Santiago-Frangos and Woodson, 2018***) and unique sRNAs have been shown to be bound to 10 to 1000 copies of Hfq in *E. coli* (***Melamed et al., 2020***) (**Appendix 1–Effects of Varying Available Hfq Pool Size**). Additionally, the aforementioned SgrS-Hfq binding rate k_{bind} corresponds well to experimentally measured *in vitro* values for sRNA-Hfq binding for sRNA of its approximate size (***Santiago-Frangos and Woodson, 2018***; ***Fender et al., 2010***; ***Hopkins et al., 2011***). If the pseudo first order rate for k_{bind} reported in ***Table 1*** is converted to a bulk second order rate by incorporating the Hfq concentration at the predicted available pool size of 250, we obtain a binding rate of $1.5 \times 10^5\ M^{-1}\ s^{-1}$. This value agrees better with the reported value of approximately (***Santiago-Frangos and Woodson, 2018***) $10^6\ M^{-1}\ s^{-1}$ for long RNAs binding to Hfq (***Fender et al., 2010***; ***Lease and Woodson, 2004***) than $10^8\ M^{-1}\ s^{-1}$ reported for short, unstructured RNAs binding to Hfq (***Hopkins et al., 2011***)). Since SgrS is a relatively long sRNA (sRNA have typically been found to be between 37–300 nt (***Wang et al., 2015a***)) with a length of 227 nucleotides, the slow sRNA–Hfq binding rate obtained by fitting seems appropriate. This type of slow sRNA association process has been suggested to be characterized by RNA restructuring (by which Hfq remodels sRNA regions in order to make its secondary structure more accessible for target mRNA base pairing) (***Antal et al., 2004***; ***Bordeau and Felden, 2014***; ***Soper et al., 2011***; ***Soper and Woodson, 2008***), which has been proposed to occur for SgrS (***Maki et al., 2010***). k_{bind} is also much greater than the Hfq–SgrS unbinding rate (k_{unbind}) of $0.0018\ s^{-1}$ which was obtained from fitting to the degradation rate of SgrS in a cell where Hfq was expressed (distinct from the Δ*hfq*

$$\mathbf{D_{on,p_{\{1,2\}}}} \xrightarrow{\mathbf{k_{t,p}}} \boldsymbol{ptsG}\ \mathbf{mRNA} \qquad \textbf{rxn 1.1}$$

$$\boldsymbol{ptsG}\ \mathbf{mRNA} \xrightarrow{\beta_{\mathrm{p}}} \emptyset \qquad \textbf{rxn 1.2}$$

$$\mathbf{D_{off,s}} \underset{\mathbf{k_{off,D_s}}}{\overset{\mathbf{k_{on,D_s}}}{\rightleftharpoons}} \mathbf{D_{on,s}} \qquad \textbf{rxn 2.0}$$

$$\mathbf{D_{on,s_{\{2,4\}}}} \xrightarrow{\mathbf{k_{t,s}}} \mathbf{SgrS} \qquad \textbf{rxn 2.1}$$

$$\mathbf{SgrS} \xrightarrow{\mathbf{k_{ds}}} \emptyset \qquad \textbf{rxn 2.2}$$

$$\mathbf{Hfq} + \mathbf{SgrS} \underset{\mathbf{k_{unbind}}}{\overset{\mathbf{k_{bind}}}{\rightleftharpoons}} \mathbf{Hfq{:}SgrS} \qquad \textbf{rxn 3}$$

$$\mathbf{Hfq{:}SgrS} + \boldsymbol{ptsG}\ \mathbf{mRNA} \underset{\mathbf{k_{off}}}{\overset{\mathbf{k_{on}}}{\rightleftharpoons}} \mathbf{Hfq{:}SgrS{:}}\boldsymbol{ptsG}\ \mathbf{mRNA} \qquad \textbf{rxn 4}$$

$$\mathbf{Hfq{:}SgrS{:}}\boldsymbol{ptsG}\ \mathbf{mRNA} \xrightarrow{\mathbf{k_{cat}}} \mathbf{Hfq} \qquad \textbf{rxn 5}$$

Figure 4. Kinetic Equations of the SgrS regulatory network. $D_{on,p_{1,2}}$ refers to the gene (or DNA) for *ptsG* in 1 (low state) or 2 (high state) copies and $D_{on,s_{2,4}}$ corresponds to the gene for *sgrS* in 2 (low state) or 4 (high state) copies. $D_{on,s}$ corresponds to *sgrS* when it is in the "ON" state due to activated or solute bound transcriptional activator SgrR being bound (***Vanderpool and Gottesman, 2007b***). k_{ds} corresponds to the experimentally measured degradation rate of SgrS when cellular Hfq is not present and k_{unbind} corresponds to the experimentally measured degradation of SgrS when Hfq was present.

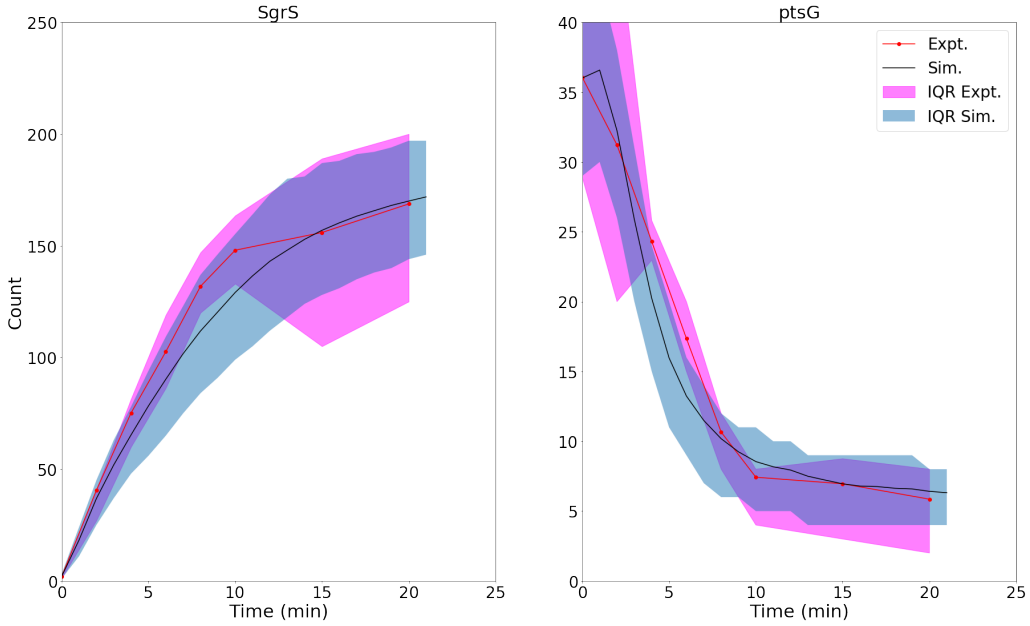


Figure 5. Average time trace and interquartile range (IQR) of labeled SgrS (left) and *ptsG* mRNA (right) from both 85–169 cells from smFISH experiments (red) and 2000 replicates from kinetic model simulations (blue). The kinetic model shows strong agreement, especially at long times (10-20 minutes) after induction and captures overall response behavior. An available pool of 250 Hfq and the kinetic parameters given in ***Table 1*** were utilized. Results considering both lower and higher available Hfq pools are discussed in **Appendix 1–Figure 1**.

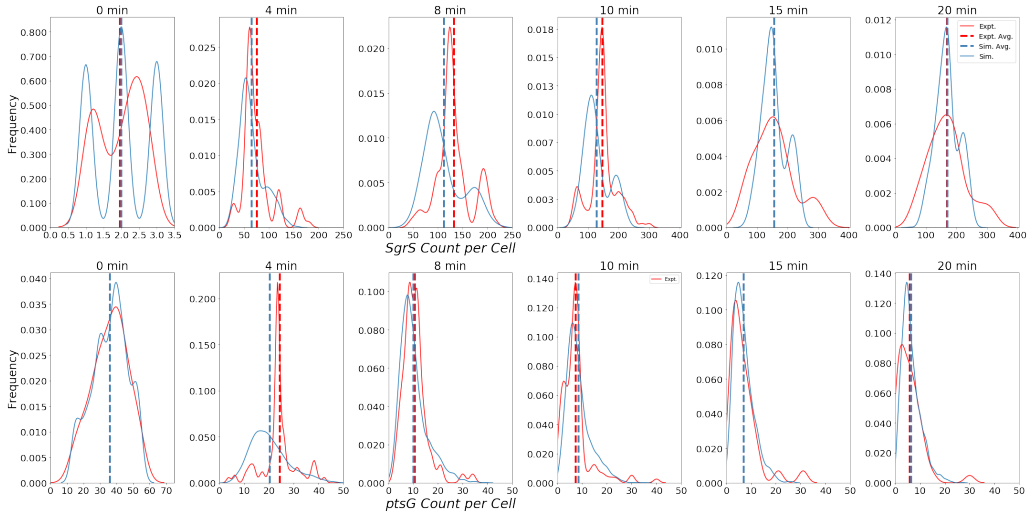


Figure 6. Distributions of Wild-Type SgrS (top) and *ptsG* mRNA (bottom) at various time points from 0 to 20 minutes post-induction. Data from smFISH-STORM experiments (red, 100–200 cells per time point) and stochastic simulations (blue, 2000 cells per time point) are shown as kernel density estimates. The effect of number of cell replicates is studied further in **Appendix 1–Figure 3**. Average copy number at each time point is are displayed with dashed vertical lines.

Table 1. The list of parameters used for the kinetic model. The % in each gene state refers to percentage of cells with the gene being in a low or high gene copy state as described in . $a)$ k_{bind} is a Pseudo first order rate accounting for the average expected pool size of Hfq participating in SgrS stabilization and enhancement (250). When converted to the corresponding bulk second order rate with 250 Hfq present k_{bind} agrees well with the range of Hfq binding rates measured for other sRNA reviewed in (***Santiago-Frangos and Woodson, 2018***) and discussed further in Results

Parameter	Value	Unit	Source
$k_{t,p}$	0.12	s^{-1}	Experimentally Measured
β_p	3.7×10^{-3}	s^{-1}	Experimentally Measured
k_{on,D_s}	3.0×10^{-2}	s^{-1}	fit
k_{off,D_s}	9.5×10^{-3}	s^{-1}	fit
$k_{t,s}$	0.33	s^{-1}	fit
k_{ds}	0.022	s^{-1}	Δ*hfq* decay rate of SgrS
k_{bind}	0.063^a	s^{-1}	fit
k_{unbind}	0.0018	s^{-1}	SgrS decay rate
k_{on}	3.1×10^{-4}	$molec^{-1}s^{-1}$	***Fei et al.*** (***2015***)
k_{off}	0.22	s^{-1}	***Fei et al.*** (***2015***)
k_{cat}	0.3	s^{-1}	***Fei et al.*** (***2015***)
% high, low gene state *sgrS*	25, 75	%	fit
% high, low gene state *ptsG*	46, 54	%	fit
Hfq pool size (available to SgrS Regulon)	250	$molec$	fit

rate) by assuming that Hfq–SgrS unbinding is the rate-limiting step in the degradation of free SgrS represented in ***Figure 4*** Reaction 2.2. These results seem reasonable in that SgrS should associate with Hfq at a rate comparable to its degradation as well as that SgrS-Hfq binding should happen at a significantly higher rate than their dissociation for sRNA chaperone stabilization by Hfq to be effective.

The kinetic values for transcriptional regulation by the activator SgrR also seem reasonable with a $k_{on,Ds}$ of $3.0 \times 10^{-2}\ s^{-1}$ and a $k_{off,Ds}$ of $9.5 \times 10^{-3}\ s^{-1}$. The gene switching parameters correspond to *sgrS* activation via SgrR binding occurring approximately 30 seconds after initiation of induction, with all *sgrS* genes assumed to start in the “OFF" state (the effect of starting genes in the “OFF" versus the “ON" state is given in **Appendix 1–Figure 2**). This seems reasonable since SgrS sRNA moves from a basal level of a few copies to greater than 40 copies on average in two minutes time (***Figure 5***). The fact that $k_{on,Ds}$ is 3 times greater than $k_{off,Ds}$ means that activation happens more frequently than deactivation from unbinding of SgrR. The relative behavior is somewhat expected as sugar shock has been induced and SgrR is believed to be transformed to its active conformation as a transcriptional factor for *sgrS* by binding to a small molecule at its C-terminus (***Vanderpool and Gottesman, 2004***, ***2007b***). While the available evidence suggests that the activity of SgrR due to solute binding rather than *sgrR* expression affects activation of *sgrS*, it has been demonstrated that SgrR is negatively autoregulated (***Vanderpool and Gottesman, 2007b***) which may lead to a ceiling on the level of *sgrS* activation that can occur even after glucose-phosphate stress is fully induced. Thus, we incorporate constant rates of $k_{on,Ds}$ and $k_{off,Ds}$ for *sgrS* activation in our model, instead of a time variant rate constant for either parameter.

Comparison of Goodness of Fit Based on Model Complexity

To illustrate the improvement of the kinetic model to describe cellular populations, we compare simulation results sequentially as each level of complexity (*i.e.*, transcriptional regulation by SgrR, gene replication, and stabilization by the chaperone protein Hfq) is added to the original reduced model presented in ***Fei et al.*** (***2015***). ***Figure 7*** demonstrates the improvement in descriptiveness at both an average and population level with progression to a more fine–grained kinetic model. The

relative error of the average copy number of SgrS and *ptsG* mRNA gives the capability of the model to reproduce experiments on an average level, while the Kulback–Leibler Divergence (KL Divergence) shows the agreement between the experimentally observed and simulation distributions of RNA copy numbers at a series of times from 0 to 20 minutes post induction.

The Relative Error used to illustrate the agreement between the experimentally measured average RNA copy number and the theoretical value is given by:

$$\eta = \left| \frac{Exp_{avg} - Sim_{avg}}{Exp_{avg}} \right| \tag{1}$$

The KL Divergence used to compare agreement between experimental and simulated distributions is given by:

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \tag{2}$$

where P(i) is the continuous probability distribution given by the Gaussian KDE of the experimental copy number distribution of RNA (SgrS or *ptsG* mRNA) and Q(i) is the analogous simulated RNA copy number distribution.

It is clear that the decrease in the KL Divergence (***Equation 2***), describing the ability of the kinetic model to accurately describe heterogeneity, is most substantial in the final model presented in this work (star markers). Accounting for transcriptional regulation by SgrR, ongoing gene replication, and the stabilizing effect of Hfq allows for a more faithful description of the noise observed in a cellular population in the process of sugar shock response.

Characterizing the Effects of SgrS Point Mutation on Association to Hfq and ptsG *mRNA*

The stochastic model we have presented can also be utilized to characterize the effects of *sgrS* point mutations on the regulatory network as a whole. The polyU tail region of *sgrS* comprising the final 8 residues of the 5' end (all of which are uridine in the sRNA) has previously been shown to be an important site for Hfq recruitment (***Otaka et al., 2011***). When the polyU tail is truncated or similarly disrupted, there is a noticeable decrease in SgrS regulatory efficiency. With this in mind, we used the previously defined kinetic model (See ***Figure 4***) to characterize the effect of a point mutation resulting in a U to G change in SgrS at position 224 (in the polyU tail region, hereafter referred to as U224G) of the sRNA on regulatory kinetics. This point mutation is well downstream of the seed region (nucleotides 168–187) where SgrS-*ptsG* mRNA base pairing occurs (***Maki et al., 2010***; ***Bobrovskyy and Vanderpool, 2014***) so it should not directly interfere with sRNA–mRNA interactions. It is also important to consider the possible structural effects arising from polyU tail mutation. Through *in silico* folding (**Appendix 1–Figure 5**) with the RNA structure prediction tool mFold (***Zuker, 2003***), we confirmed that the stability of the U224G with a ΔG of $-17.60\ kcal/mol$ is unchanged from the predicted wild-type value of $-17.60\ kcal/mol$, indicating that sRNA structure is conserved and the measured wild-type ΔHfq degradation rate (see Section 'Methods and Materials') is appropriate for use in fitting the U224G mutant data (as a rate for ***Figure 1***, **rxn 2.2**).

We then fit to the experimentally measured SgrS and *ptsG* mRNA distributions using the previously determined kinetic model. A robust fit describing both average behavior as well as population level variation (***Figure 8***, **Appendix 1–Figure 4**) was achieved primarily by modulating the rates of SgrS to Hfq binding and unbinding and the *ptsG* mRNA annealing rates k_{on} and k_{off} (which were also free parameters in this treatment) to a much lesser extent, which further demonstrates the role of the polyU tail in Hfq chaperone recruitment.

The changes in the kinetic parameters of the model used to fit mutant U224G relative to the wild-type cells (WT) illustrate the effects of this mutation on SgrS-Hfq association, relative to the subsequent annealing of SgrS to its target *ptsG* mRNA (***Table 2***).

The 48% decrease in the SgrS-Hfq binding rate k_{bind} and 66% increase in the unbinding rate of the sRNA and chaperone complex k_{unbind} highlight the effects of polyU tail disruption, and support

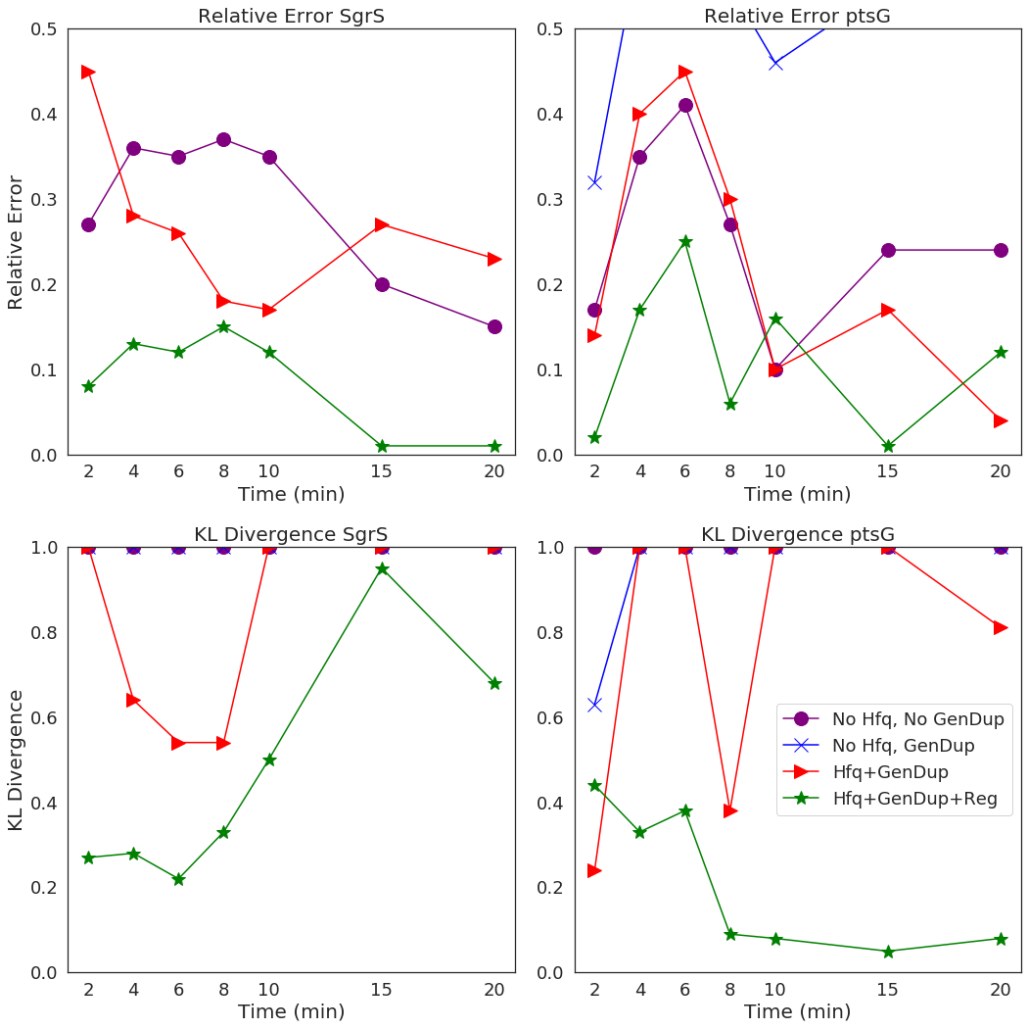


Figure 7. Statistical analysis of the agreement of sRNA and mRNA copy number between experiment and theory on both an average (Relative Error) and distribution (Kulback–Leibler: KL Divergence) level. KL Divergence values for the model with no Hfq stabilization nor Gene Duplication are not shown as the values obtained are at 1.0, corresponding to significant disagreement in that model variant and experiment. GeneDup refers to a model with Gene Duplication for both SgrS and *ptsG* implemented and Reg refers to a model with transcriptional regulation of SgrS by SgrR in place. The green line (with star markers) indicates the full kinetic model used for this study, which provides the best fit to both average and population level data for both SgrS and *ptsG* mRNA.

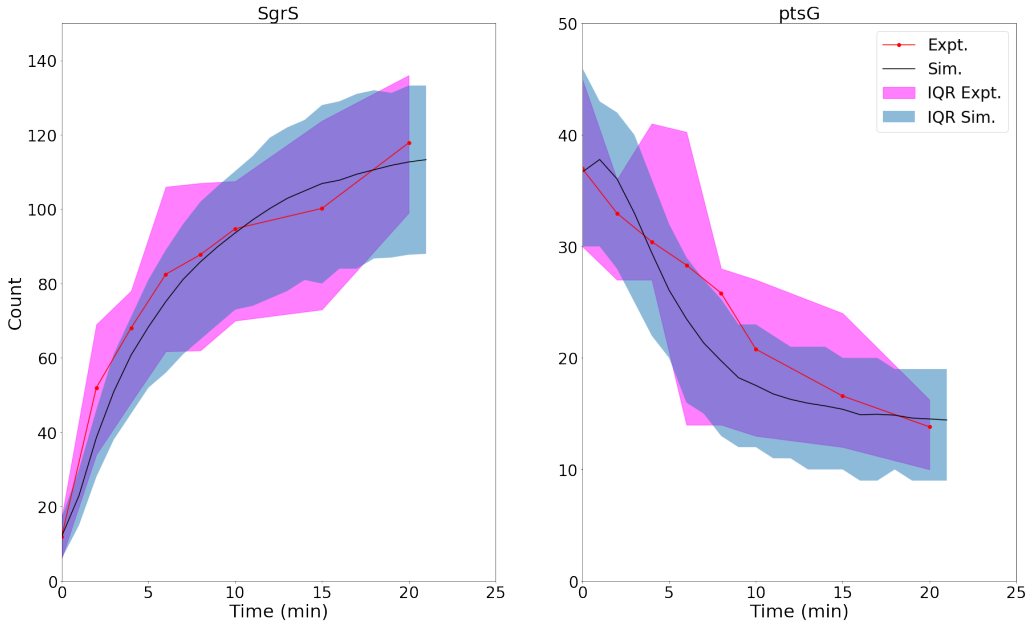


Figure 8. For U224G mutant cells, average time trace and interquartile range (IQR) of labeled SgrS (left) and *ptsG* mRNA (right) from both 85-169 cells from smFISH experiments (red) and 2000 replicates from kinetic model simulations (blue). The kinetic model shows strong agreement, especially at long times (10-20 minutes) after induction and captures overall response behavior. An available pool of 250 Hfq and the kinetic parameters given in ***Table 1*** were utilized, other than changes to SgrS-Hfq binding and unbinding rates and *ptsG* mRNA annealing and dissociation rates given in ***Table 2***.

previous conclusions that this is an important site for Hfq stabilization of SgrS (***Otaka et al., 2011***), and the regulatory efficiency of the network as a whole. The smaller relative changes in the SgrS-*ptsG* mRNA annealing rates k_{on} and k_{off} by 32% and 22% respectively may be due to altered interactions with Hfq that impair Hfq–dependent annealing of SgrS and *ptsG* mRNA (**Appendix 1 – Effects of SgrS Point Mutation on Regulatory Kinetics**). In light of the previously discussed slow SgrS–Hfq association process, it is reasonable that RNA restructuring of Hfq may be disrupted by mutation U224G, thus leading to slower and weaker annealing to *ptsG* mRNA. One possible explanation for the disturbance of regulation in mutant U224G is the disruption of orderly transcription termination (the polyU tail at the 3′ end of *sgrS*). Such readthrough transcription has previously been ascribed to decrease the efficiency of SgrS binding to Hfq (***Morita et al., 2015***, ***2017***).

Discussion

The construction of a stochastic kinetic model including gene replication, transcriptional regulation, and the role of the Hfq chaperone protein demonstrates the utility of combining single cell experiments with stochastic modeling. The SgrS Regulatory Network is a noisy system characterized by small numbers of sRNA and mRNA, as well as gene copy numbers that vary from cell-to-cell. This leads to the population level heterogeneity that can then be used to parameterize a kinetic model for analysis of the role of specific molecular actors, such as the chaperone Hfq, and the effects of point mutation on sRNA silencing of mRNA.

The average number of Hfq hexamers present in an *E. coli* cell has been reported to be on the order of 1400 to 10000 (2 μM - 15 μM) (***Santiago-Frangos and Woodson, 2018***; ***Taniguchi et al., 2010***; ***Wiśniewski and Rakus, 2014***; ***Mancuso et al., 2012***; ***Wang et al., 2015b***). It is worth noting that an extensive microfluidic-based, single-cell proteomics study that analyzed over 4000 individual *E. coli* cells grown in similar media conditions as our study (***Taniguchi et al., 2010***) found a mean Hfq level of 1500. Additional immunoprecipitation and sequencing studies have shown the number of various individual mRNAs and sRNAs being bound to Hfq to range from 10s to 1000 in *E.*

Table 2. The list of kinetic parameters for SgrS-Hfq association (k_{bind} and k_{unbind}) and annealing with *ptsG* mRNA (k_{on} and k_{off}) for wild-type (WT) cells as well as SgrS mutant U224G (Reactions in ***Figure 4***). The substantial decrease in the values of k_{bind} and k_{unbind} demonstrate the disruption of Hfq binding that accompanies the mutation in the polyU tail, which has been observed previously (***Otaka et al., 2011***). The smaller relative changes in the *ptsG* mRNA annealing rates may be due to disruption of RNA restructuring (***Antal et al., 2004***; ***Bordeau and Felden, 2014***; ***Soper et al., 2011***; ***Soper and Woodson, 2008***) of SgrS by Hfq that hampers association to the mRNA target.

Parameter	Mutant	Value	% Difference from WT
k_{bind}	U224G	$0.033\ s^{-1}$	−48%
	WT	$0.063\ s^{-1}$	
k_{unbind}	U224G	$0.003\ s^{-1}$	+66%
	WT	$0.0018\ s^{-1}$	
k_{on}	U224G	$2.1 \times 10^{-4}\ molec^{-1} s^{-1}$	−32%
	WT	$3.1 \times 10^{-4}\ molec^{-1} s^{-1}$	
k_{off}	U224G	$0.27\ s^{-1}$	+22%
	WT	$0.22\ s^{-1}$	

coli (***Melamed et al., 2020***). Thus, our prediction (from fitting) that a pool of approximately 250 Hfq (0.5 μM) are available to bind with SgrS sRNA at any time in the simulation of sugar shock regulation seems reasonable.

In addition, our approach allowed us to characterize the rate of Hfq-SgrS association compared to values for reported for Hfq stabilization of other regulatory sRNAs. If the pseudo first order Hfq binding rate k_{bind} reported in ***Table 1*** is converted to a bulk second order rate we obtain a binding rate of $1.5 \times 10^5\ M^{-1}\ s^{-1}$ which agrees reasonably well with the reported value (***Santiago-Frangos and Woodson, 2018***) of approximately $10^6\ M^{-1}\ s^{-1}$ for long RNAs binding to Hfq (***Fender et al., 2010***; ***Lease and Woodson, 2004***) (compared to the value of to $10^8\ M^{-1}\ s^{-1}$ for short, unstructured RNAs binding to Hfq (***Hopkins et al., 2011***)). SgrS is a relatively long sRNA with a length of 227 nucleotides (sRNAs have been observed with 37-300 nt (***Wang et al., 2015a***)), therefore the slow sRNA-Hfq binding process that we describe does seem likely and could be due to RNA restructuring of SgrS (***Maki et al., 2010***; ***Antal et al., 2004***; ***Bordeau and Felden, 2014***; ***Soper et al., 2011***; ***Soper and Woodson, 2008***) by Hfq in order to promote binding with *ptsG* mRNA. It is thought that cellular sRNA and mRNA are present in large excess over Hfq (***Wagner, 2013***), so nearly all cellular Hfq hexamers are thought to be bound to RNA. Since cellular mRNA in *E. coli* are thought to be on the order of approximately 2000-8000 copies (***Bartholomäus et al., 2016***) (much greater than the highest measured SgrS sRNA value of 200) the available Hfq pool size that we present is representative of the relative competitiveness (and time-dependent cycling) of SgrS for Hfq relative to its other RNA targets.

The study of mutant U224G shows the importance of Hfq stabilization in the SgrS regulatory network as a whole and seems to corroborate previous findings (***Otaka et al., 2011***) that highlight the importance of the polyU tail for Hfq association with SgrS. The substantial decrease of the Hfq-SgrS binding rate and increase in the related unbinding rate relative to the *ptsG* mRNA annealing rates further down the network obtained from fitting confirms this point (***Table 2***). The changes in the SgrS-*ptsG* mRNA annealing rates k_{on} and k_{off} seem to support conclusions from the wild-type cells that Hfq-SgrS binding may result in some restructuring of the sRNA that makes this a slow process. This may explain the lower efficiency in *ptsG* mRNA association observed in mutant U224G, since Hfq cannot bind SgrS as effectively due to mutation at the polyU tail. Therefore, the restructuring of SgrS by Hfq necessary to facilitate *ptsG* mRNA association is also hampered, resulting in slower and less stable mRNA binding (a decrease in k_{on} and an increase in k_{off}).

In conclusion, by incorporating gene replication, stabilization by the chaperone protein Hfq, and

transcriptional gene regulation of *sgrS* we have developed a kinetic model capable of describing the cellular heterogeneity observed in the *E. coli* sugar shock response network. Stochastic simulation of the kinetic model allows us to take full advantage of the single-molecule fluorescence data that illustrates cell-to-cell variability in a collection of hundreds of cells. While the post-transcriptional regulation and silencing of *ptsG* mRNA by the sRNA is the critical feature, accounting for gene replication, transcriptional regulation, and stabilization gives a more robust picture of the regulatory network as a whole. In addition, complexifying the model highlights the importance of stabilization by Hfq and chaperone proteins in general in RNA silencing networks and allowed for a prediction of the rate of association of SgrS and Hfq, the effective available Hfq pool size for the SgrS regulon under sugar stress conditions, as well as an analysis of an SgrS point mutation in one of the presumed Hfq binding modules (the polyU tail). The model presented in this work establishes a framework for models analyzing the other regulatory targets of SgrS along with spatially-resolved models describing SgrS target search kinetics.

Methods and Materials

Wild type *E. coli* cells (DJ480) were grown overnight at 37 °C, 250 rpm in LB Broth. The SgrS U224G mutant was grown in LB Broth with 50 μg/ml spectinomycin (Spec) (Sigma-Aldrich). The next day, overnight cultures were diluted 100-fold into MOPS EZ rich defined medium with 0.2% glucose and the cells were grown until OD_{600} reached 0.15–0.25. α-methyl D-glucopyranoside (αMG) (Sigma Aldrich) was then added to provoke glucose-phosphate stress and induce a SgrS expression response. Specific volumes of liquid were removed from the culture at 0, 2, 4, 6, 8, 10, 15, and 20 minutes after induction and mixed with formaldehyde (Fisher Scientific) to a final concentration of 4% for cell fixation prior to single molecule experiments.

Following fixation, the cells were incubated and washed, before being permeabilized with 70% ethanol, to allow for fluorescence *in situ* hybridization (FISH). Stellaris Probe Designer was used to design the smFISH oligonucleotide probes that were ordered from Biosearch Technologies (https://www.biosearchtech.com/). Each sRNA was labeled with 9 Alexa Fluor 647 probes while each *ptsG* mRNA was labeled with 28 CF 568 probes. The labeled RNA molecules were then imaged via the super-resolution technique STORM (Stochastic Optical Reconstruction Microscopy). A density-based clustering analysis algorithm (DBSCAN) (***Daszykowski et al., 2001***) was utilized to calculate RNA copy numbers. The algorithm used was the same as previously published (***Fei et al., 2015***), but the Nps and Eps values were updated for the SgrS and *ptsG* mRNA images, since CF 568 was used instead of Alexa Fluor 568 and a 405 nm laser to reactivate the dyes. The SgrS (9 probes labeled with AlexaFluor 647) images were clustered using $Nps = 3$ and $Eps = 15$ and the *ptsG* mRNA (28 probes labeled with CF 568) images were clustered using $Nps = 10$ and $Eps = 25$ and these numbers were empirically chosen. A MATLAB code was used for cluster analysis.

The raw data was acquired using the Python-based acquisition software and it was analyzed using a data analysis algorithm which was based on work previously published by ***Babcock et al. (2013)***. The peak identification and fitting were performed using the method described previously. The z-stabilization was done by the CRISP system and the horizontal drift was calculated using Fast Fourier Transformation (FFT) on the reconstructed images of subsets of the super-resolution image, comparing the center of the transformed images and corrected using linear interpolation.

The *ptsG* mRNA degradation rates were calculated via a rifampicin-chase experiment. The wild type (DJ480) *E. coli* cells and Δ*hfq* mutant strain SA1816 [DJ480, *laclg*, *tetR*, *spec*, Δ*hfq::kan*] cells were grown in LB Broth with the respective antibiotics at 37 °C, 250 rpm overnight. They were used to calculate the RNA degradation rates. The Δ*hfq::kan* allele was moved to create strain SA1816 constructed by P1 transduction (***Miller, 1972***). When the OD_{600} reached 0.15–0.25, rifampicin (Sigma-Aldrich) was added to cultures to a final concentration of 500 μg/ml. The cells were labeled by smFISH probes and analyzed by the same process described above, taking the time of rifampicin addition or αMG removal as the 0 time point. Aliquots were taken after 0,2,4,6,8,10,15, and 20 minutes (0,2,4,6, and 8 minutes for ΔHfq strains). For the purpose of background subtraction, ΔSgrS

and Δ*ptsG* mRNA strains were grown, labeled with probes and imaged in the same manner to be used for the measurement of the background signal due to the non-specific binding of Alexa Fluor 647 and CF 568. The natural logs of the RNA copy numbers were plotted against time and the slope of the linear fitting was used to calculate the RNA lifetime and then the degradation rates. SgrS degradation rates were obtained from ***Fei et al.*** (***2015***), where they were measured by stopping the transcription of *sgrS* by removing αMG from the media and then were imaged and analyzed to calculate the degradation rates in the same manner as was described for *ptsG* mRNA. The values for k_{cat}, k_{on}, and k_{off} for WT cells were confirmed to be within the errors reported for the values given in (***Fei et al., 2015***) by fitting to the experimentally measured RNA counts with the simplified model given in that work. The transcription rate of *ptsG* was determined using $k_{t.p} = \beta_p \times [p]_0$, (as described in ***Fei et al.*** (***2015***)), where $[p]_0$ was the average initial level of *ptsG* mRNA before stress induction. The transcription rate obtained was unchanged between the wild-type and the U224G mutant cells.

Acknowledgements

This work was supported by grants from National Institutes of Health (NIGMS Grant R01 GM112659 and R35 GM122569) and through the National Science Foundation Physics Frontiers Center: “The Center for the Physics of Living Cells" (CPLC) (NSF PHY 1430124).

Competing interests

The authors declare that no competing interests exist.

References

Antal M, Bordeau V, Douchin V, Felden B. A Small Bacterial RNA Regulates a Putative ABC Transporter. Journal of Biological Chemistry. 2004 Dec; 280(9):7901–7908. https://doi.org/10.1074/jbc.m413071200, doi: 10.1074/jbc.m413071200.

Babcock HP, Moffitt JR, Cao Y, Zhuang X. Fast compressed sensing analysis for super-resolution imaging using L1-homotopy. Optics Express. 2013 Nov; 21(23):28583. https://doi.org/10.1364/oe.21.028583, doi: 10.1364/oe.21.028583.

Babski J, Maier LK, Heyer R, Jaschinski K, Prasse D, Jäger D, Randau L, Schmitz RA, Marchfelder A, Soppa J. Small regulatory RNAs in Archaea. RNA Biology. 2014 May; 11(5):484–493. https://doi.org/10.4161/rna.28452, doi: 10.4161/rna.28452.

Balasubramanian D, Vanderpool CK. Deciphering the Interplay between Two Independent Functions of the Small RNA Regulator SgrS in Salmonella. Journal of Bacteriology. 2013 Aug; 195(20):4620–4630. https://doi.org/10.1128/jb.00586-13, doi: 10.1128/jb.00586-13.

Bartholomäus A, Fedyunin I, Feist P, Sin C, Zhang G, Valleriani A, Ignatova Z. Bacteria differently regulate mRNA abundance to specifically respond to various stresses. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016 Mar; 374(2063):20150069. https://doi.org/10.1098/rsta.2015.0069, doi: 10.1098/rsta.2015.0069.

Bobrovskyy M, Azam MS, Frandsen JK, Zhang J, Poddar A, Ma X, Henkin TM, Ha T, Vanderpool CK. Determinants of target prioritization and regulatory hierarchy for the bacterial small RNA SgrS. Molecular Microbiology. 2019 Aug; https://doi.org/10.1111/mmi.14355, doi: 10.1111/mmi.14355.

Bobrovskyy M, Vanderpool CK. The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria. Frontiers in Cellular and Infection Microbiology. 2014 May; 4. https://doi.org/10.3389/fcimb.2014.00061, doi: 10.3389/fcimb.2014.00061.

Bordeau V, Felden B. Curli synthesis and biofilm formation in enteric bacteria are controlled by a dynamic small RNA module made up of a pseudoknot assisted by an RNA chaperone. Nucleic Acids Research. 2014 Jan; 42(7):4682–4696. https://doi.org/10.1093/nar/gku098, doi: 10.1093/nar/gku098.

Cooper S, Helmstetter CE. Chromosome replication and the division cycle of Escherichia coli. Journal of Molecular Biology. 1968 Feb; 31(3):519–540. doi: 10.1016/0022-2836(68)90425-7.

Daszykowski M, Walczak B, Massart D. Looking for natural patterns in data. Chenom Intell Lab Syst. 2001; 56:83–92.

Earnest TM, Cole JA, Luthey-Schulten Z. Simulating Biological Processes: Stochastic Physics from Whole Cells to Colonies. Rep Prog Phys. 2018 apr; 81(5):052601. doi: 10.1088/1361-6633/aaae2c.

Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic Gene Expression in a Single Cell. Science. 2002 American Association for the Advancement of Science; 297(5584):1183–1186. doi: 10.1126/science.1070919.

Fei J, Singh D, Zhang Q, Park S, Balasubramanian D, Golding I, Vanderpool CK, Ha T. Determination of in vivo target search kinetics of regulatory noncoding RNA. Science. 2015 Mar; 347(6228):1371–1374. doi: 10.1126/science.1258849.

Fender A, Elf J, Hampel K, Zimmermann B, Wagner EGH. RNAs actively cycle on the Sm-like protein Hfq. Genes & Development. 2010 Dec; 24(23):2621–2626. https://doi.org/10.1101/gad.591310, doi: 10.1101/gad.591310.

Hallock MJ, Stone JE, Roberts E, Fry C, Luthey-Schulten Z. Simulations of reaction diffusion processes over biologically-relevant size and time scales using multi-GPU workstations. Parallel Comput. 2014; 40:86–99. doi: 10.1016/j.parco.2014.03.009.

Hallock MJ, Luthey-Schulten Z. Improving reaction kernel performance in Lattice Microbes: particle-wise propensities and run-time generated code. In: *Parallel and Distributed Processing Symposium Workshop (IPDPSW), 2016 IEEE International*; 2016. p. 428–343.

Hopkins JF, Panja S, Woodson SA. Rapid binding and release of Hfq from ternary complexes during RNA annealing. Nucleic Acids Research. 2011 Mar; 39(12):5193–5202. https://doi.org/10.1093/nar/gkr062, doi: 10.1093/nar/gkr062.

Ishikawa H, Otaka H, Maki K, Morita T, Aiba H. The functional Hfq-binding molecule of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3' poly(U) tail. RNA. 2012 May; 18(5):1062–1074. doi: 10.1261/rna.031575.111.

Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. Science. 2014 Dec; 346(6216):1533–1536. doi: 10.1126/science.1255301.

Kawamoto H, Koide Y, Morita T, Aiba H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. Molecular Microbiology. 2006 Aug; 61(4):1013–1022. doi: 10.1111/j.1365-2958.2006.05288.x.

Lease RA, Woodson SA. Cycling of the Sm-like Protein Hfq on the DsrA Small Regulatory RNA. Journal of Molecular Biology. 2004 Dec; 344(5):1211–1223. https://doi.org/10.1016/j.jmb.2004.10.006, doi: 10.1016/j.jmb.2004.10.006.

Lee SJ. Signal transduction between a membrane-bound transporter, PtsG, and a soluble transcription factor, Mlc, of Escherichia coli. The EMBO Journal. 2000 Oct; 19(20):5353–5361. https://doi.org/10.1093/emboj/19.20.5353, doi: 10.1093/emboj/19.20.5353.

Maier T, Schmidt A, Güell M, Kühner S, Gavin AC, Aebersold R, Serrano L. Quantification of mRNA and protein and integration with protein turnover in a bacterium. Molecular Systems Biology. 2011 Jan; 7(1):511. https://doi.org/10.1038/msb.2011.38, doi: 10.1038/msb.2011.38.

Maki K, Morita T, Otaka H, Aiba H. A minimal base–pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA. Molecular Microbiology. 2010 Mar; 76(3):782–792. doi: 10.1111/j.1365-2958.2010.07141.x.

Mancuso F, Bunkenborg J, Wierer M, Molina H. Data extraction from proteomics raw data: An evaluation of nine tandem MS tools using a large Orbitrap data set. Journal of Proteomics. 2012 Sep; 75(17):5293–5303. https://doi.org/10.1016/j.jprot.2012.06.012, doi: 10.1016/j.jprot.2012.06.012.

Melamed S, Adams PP, Zhang A, Zhang H, Storz G. RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles. Molecular Cell. 2020 Jan; 77(2):411–425.e7. https://doi.org/10.1016/j.molcel.2019.10.022, doi: 10.1016/j.molcel.2019.10.022.

Miller JH. Experiments in Molecular Genetics. xvi ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 1972.

Morita T, Ueda M, Kubo K, Aiba H. Insights into transcription termination of Hfq-binding sRNAs of Escherichia coli and characterization of readthrough products. RNA. 2015 Jun; 21(8):1490–1501. https://doi.org/10.1261/rna.051870.115, doi: 10.1261/rna.051870.115.

Morita T, Nishino R, Aiba H. Role of the terminator hairpin in the biogenesis of functional Hfq-binding sRNAs. RNA. 2017 Jun; 23(9):1419–1431. https://doi.org/10.1261/rna.060756.117, doi: 10.1261/rna.060756.117.

Nam TW, Jung HI, An YJ, Park YH, Lee SH, Seok YJ, Cha SS. Analyses of Mlc-IIBGlc interaction and a plausible molecular mechanism of Mlc inactivation by membrane sequestration. Proceedings of the National Academy of Sciences. 2008 Mar; 105(10):3751–3756. https://doi.org/10.1073/pnas.0709295105, doi: 10.1073/pnas.0709295105.

Otaka H, Ishikawa H, Morita T, Aiba H. PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. Proceedings of the National Academy of Sciences. 2011 Jul; 108(32):13059–13064. https://doi.org/10.1073/pnas.1107050108, doi: 10.1073/pnas.1107050108.

Peterson JR, Hallock MJ, Cole JA, Luthey-Schulten ZA. A Problem Solving Environment for Stochastic Biological Simulations. In: *PyHPC 2013* Supercomputing 2013; 2013. doi: 10.13140/2.1.3207.7440.

Peterson JR, Cole JA, Fei J, Ha T, Luthey-Schulten ZA. Effects of DNA replication on mRNA noise. PNAS. 2015 Dec; 112(52):15886–15891. doi: 10.1073/pnas.1516246112.

Raser JM. Noise in Gene Expression: Origins, Consequences, and Control. Science. 2005 sep; 309(5743):2010–2013. doi: 10.1126/science.1105891.

Roberts E, Stone JE, Luthey-Schulten Z. Lattice Microbes: high-performace stochastic simulation method for the reaction-diffusion master equation. J Comp Chem. 2013; 3:245–255. doi: 10.1002/jcc.23130.

Santiago-Frangos A, Woodson SA. Hfq chaperone brings speed dating to bacterial sRNA. Wiley Interdisciplinary Reviews: RNA. 2018 Apr; 9(4):e1475. https://doi.org/10.1002/wrna.1475, doi: 10.1002/wrna.1475.

Seitz S, Lee SJ, Pennetier C, Boos W, Plumbridge J. Analysis of the Interaction between the Global Regulator Mlc and EIIBGlcof the Glucose-specific Phosphotransferase System inEscherichia coli. Journal of Biological Chemistry. 2003 Jan; 278(12):10744–10751. https://doi.org/10.1074/jbc.m212066200, doi: 10.1074/jbc.m212066200.

Soper TJ, Doxzen K, Woodson SA. Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. RNA. 2011 Jun; 17(8):1544–1550. https://doi.org/10.1261/rna.2767211, doi: 10.1261/rna.2767211.

Soper TJ, Woodson SA. The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. RNA. 2008 Jul; 14(9):1907–1917. https://doi.org/10.1261/rna.1110608, doi: 10.1261/rna.1110608.

Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science. 2010 Jul; 329(5991):533–538. https://doi.org/10.1126/science.1188308, doi: 10.1126/science.1188308.

Vanderpool CK, Gottesman S. The Novel Transcription Factor SgrR Coordinates the Response to Glucose–Phosphate Stress. Journal of Bacteriology. 2007 Jan; 189(6):2238–2248. https://doi.org/10.1128/jb.01689-06, doi: 10.1128/jb.01689-06.

Vanderpool CK, Gottesman S. The Novel Transcription Factor SgrR Coordinates the Response to Glucose-Phosphate Stress. Journal of Bacteriology. 2007 Jan; 189(6):2238–2248. doi: 10.1128/jb.01689-06.

Vanderpool CK, Gottesman S. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. Molecular Microbiology. 2004 Oct; 54(4):1076–1089. doi: 10.1111/j.1365-2958.2004.04348.x.

Wagner EGH. Cycling of RNAs on Hfq. RNA Biology. 2013 Apr; 10(4):619–626. https://doi.org/10.4161/rna.24044, doi: 10.4161/rna.24044.

Wang J, Liu T, Zhao B, Lu Q, Wang Z, Cao Y, Li W. sRNATarBase 3.0: an updated database for sRNA-target interactions in bacteria. Nucleic Acids Research. 2015 Oct; 44(D1):D248–D253. https://doi.org/10.1093/nar/gkv1127, doi: 10.1093/nar/gkv1127.

Wang M, Zhang J, Xu H, Golding I. Measuring transcription at a single gene copy reveals hidden drivers of bacterial individuality. Nature Microbiology. 2019 Sep; https://doi.org/10.1038/s41564-019-0553-z, doi: 10.1038/s41564-019-0553-z.

Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. PROTEOMICS. 2015 Mar; 15(18):3163–3168. https://doi.org/10.1002/pmic.201400441, doi: 10.1002/pmic.201400441.

Wiśniewski JR, Rakus D. Quantitative analysis of the Escherichia coli proteome. Data in Brief. 2014 Dec; 1:7–11. https://doi.org/10.1016/j.dib.2014.08.004, doi: 10.1016/j.dib.2014.08.004.

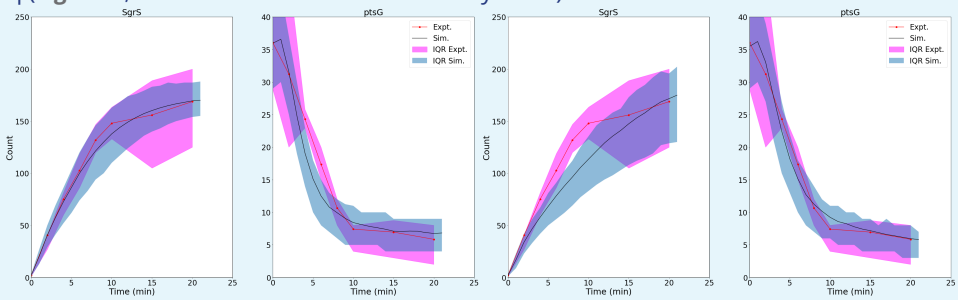
Youngren B, Nielsen HJ, Jun S, Austin S. The multifork Escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. Genes & Development. 2014 Jan; 28(1):71–84. doi: 10.1101/gad.231050.113.

Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research. 2003 Jul; 31(13):3406–3415. https://doi.org/10.1093/nar/gkg595, doi: 10.1093/nar/gkg595.

Appendix 1

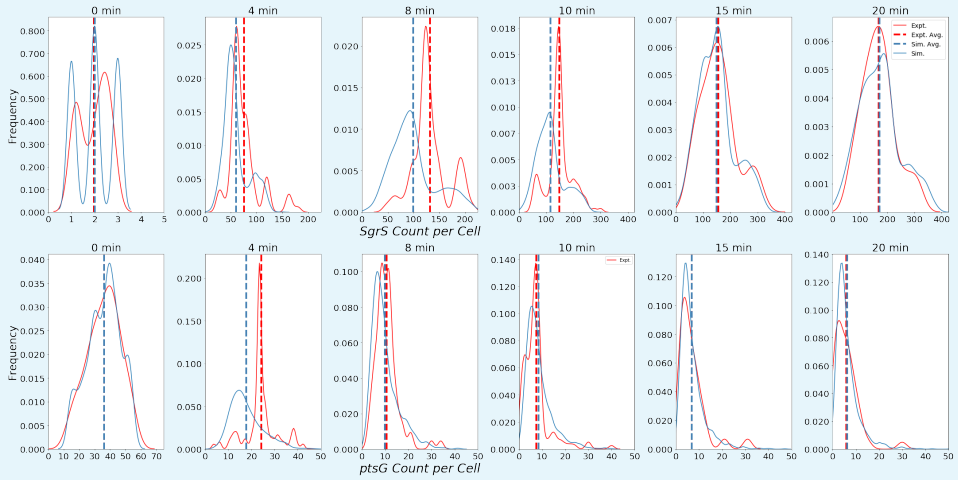
Effects of Varying Available Hfq Pool Size

The available pool of Hfq utilized in the model represents the fraction of cellular Hfq hexamers bound to SgrS as opposed to other targets and thus the relative binding strength of SgrS compared to other RNAs stabilized by the chaperone. Previous work (***Melamed et al., 2020***) has shown that the typical number of Hfq bound to a given sRNA varies widely across sRNA species. If an even smaller pool of cellular Hfq is assumed to be available for SgrS binding under sugar shock conditions the average behavior of SgrS experimentally observed can be more exactly captured (**Appendix 1–Figure 1**). However, this comes at a loss of the population level noise observed in the measured RNA distributions because fewer SgrS can be stabilized and so it decays on a much faster timescale, resulting in a loss of cell-to-cell variation. When additional Hfq is added to the available pool such as the 800 available in the simulations shown in **Appendix 1–Figure 1** the opposite behavior can be seen. SgrS exhibits greater population level heterogeneity, but with a less robust fit to the average behavior that is experimentally observed. We propose that this creates more noise because SgrS is less likely to be present in its free form an decays more slowly when it is associated with *ptsG* mRNA and Hfq (k_{on} is small relative to k_{ds}) than it would when it is not stabilized by Hfq (***Figure 4***, **rxn 2.2** versus rxn 4 followed by rxn 5).



Appendix 1 Figure 1. Left: Trace and interquartile range (IQR) of SgrS sRNA and *ptsG* mRNA mRNA where simulations include a smaller pool of 200 Hfq available (versus 250 in main text simulations). While averages can be more tightly fit, the population level variation observed for SgrS is minimized even further from what is observed experimentally, including at long times post-induction. **Right**: A similar plot of Trace and IQR with Hfq available pool size equal to 800. Here the population level variation is larger (especially at long times post induction), but the initial average traces are less well captured.

Effects of Initial Gene State



Appendix 1 Figure 2. Distributions of Wild-Type SgrS (top) and *ptsG* mRNA (bottom) at various time points from 0 to 20 minutes post-induction. Data from smFISH–STORM experiments (red, 100-200 cells per time point) and stochastic simulations (blue, 2000 cells per time point) are shown as kernel density estimates. Average copy number at each time point is displayed with dashed vertical lines.

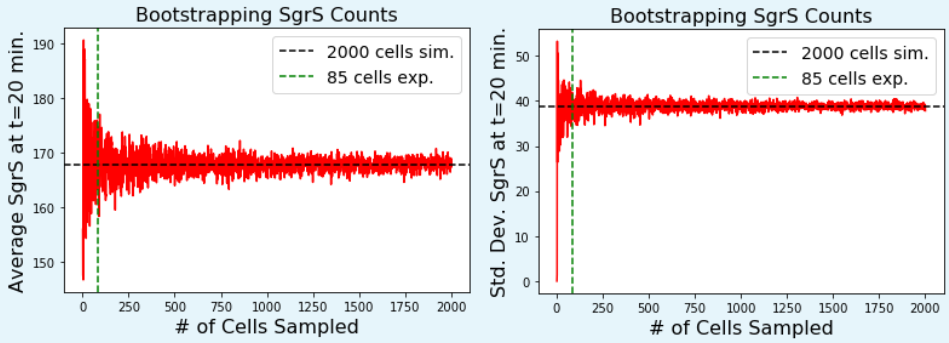
Of interest from a more technical standpoint, is the state of the *sgrS* genes at time = 0 minutes in the simulation. While, in principle these genes should be in the “OFF" state and unable to be transcribed since induction has yet to begin it is interesting to understand the effects of initial gene state on population level noise. Consider the following example, when all SgrS genes begin in “ON" state. While the average behavior a times from 4 to 10 minutes is poorly captured, the RNA distributions are well-described at 15 and 20 minutes post-induction (**Appendix 1–Figure 2**). This model assumes an immediate switch of the *sgrS* genes due to induction. While unrealistic when taken at face value, it is reasonable to assume the induction occurs on the order of seconds, since the amount of SgrS increases by a factor of 10 from its basal value by 2 minutes post-induction (***Figure 5***) and since binding of the SgrR activator for *sgrS* is mediated by binding to a small molecule (*i.e* glucose-6 phosphate), which presumably takes some interval of time. The smaller $k_{on,Ds}$ and $k_{off,Ds}$ values ($2.0 \times 10^{-3}\ s^{-1}$ and $6.5 \times 10^{-4}\ s^{-1}$ versus $3.0 \times 10^{-2}\ s^{-1}$ and $9.5 \times 10^{-3}\ s^{-1}$, respectively) used in ***Figure 4*** Rxn 2.0 relative to those given in ***Table 1*** then lead to a wider range of population distributions at late times due to longer dwell times (*i.e.* up to 5 minutes) for the *sgrS* gene in the “OFF" state compared to the a typical dwell time of less than 1 minute in the “OFF" state when the more appropriate regulatory values ($3.0 \times 10^{-2}\ s^{-1}$ and $9.5 \times 10^{-3}\ s^{-1}$) are used for $k_{on,Ds}$ and $k_{off,Ds}$ respectively.

Effects of Increased Cell Replicate Number

The number of *E. coli* cells that are simulated or have their RNA distributions experimentally measured is of great importance when considering a process characterized by stochasticity. A certain number of cells must be observed to accurately capture both the average behavior and cell-to-cell variability that emanates from a kinetic regulatory system (***Taniguchi et al., 2010***; ***Elowitz et al., 2002***; ***Raser, 2005***).

Appendix 1–Figure 3 shows the effect of number of cells measured on the average and standard deviation of the SgrS simulated at 20 minutes post-induction. The bootstrapping technique presented allows for the selection of an individual *E. coli* cellular replicate with replacement up to N cells. The vertical dashed line in each figure shows the expected average and standard deviation values produced from bootstrapping with N=85, the number of cells

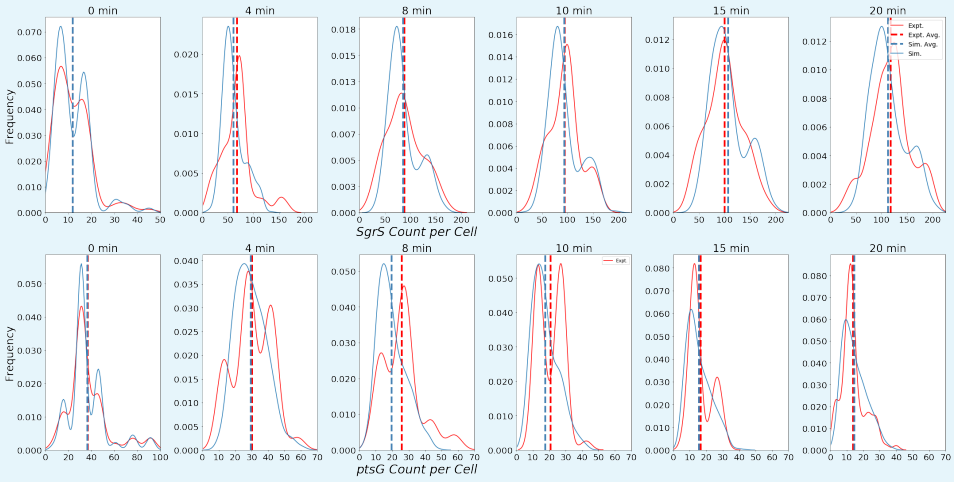
experimentally measured at time 20 minutes post-induction. This highlights the possible error in both mean copy number (5-10 copies) or population level variation (5-10 copies) that could be accrued due to insufficient sampling.



Appendix 1 Figure 3. Bootstrapping of SgrS sRNA simulated at 20 minutes post sugar shock induction. The x axis gives the number of samples taken (N) with replacement out of a total 2000 independent simulation trajectories in the bootstrapping procedure. The vertical dashed line at N=85 shows the number of cells experimentally imaged at this time point. It takes thousands of simulated cells before the SgrS mean and population level variation noise begin to relax to the calculated values.

Effects of SgrS Point Mutation On Regulatory Kinetics

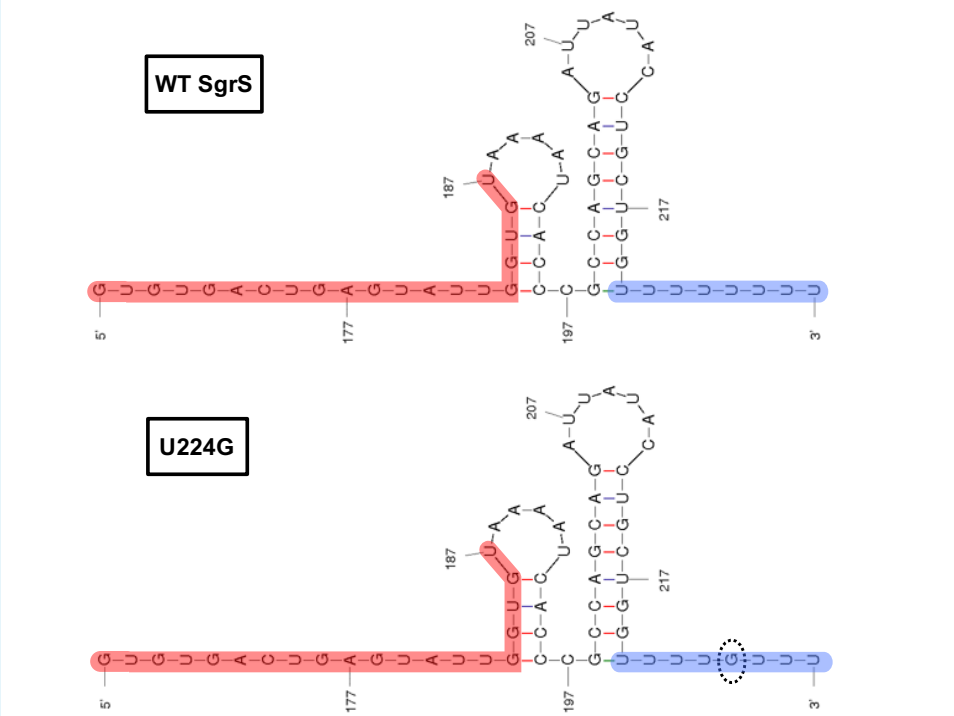
In order to fit to mutant U224G the same parameters were utilized other than the SgrS-*ptsG* mRNA binding and unbinding rates k_{bind} and k_{unbind} and the *ptsG* mRNA association rates k_{on} and k_{off}. The same gene state (high versus low gene copy number) percentages for *sgrS* and *ptsG* as for the wild-type cells were used as well as the same "available" Hfq pool size of 250 hexamers. The distributions (as kernel density estimates) shown in **Appendix 1–Figure 4** for both SgrS and *ptsG* mRNA were obtained via the fitting process.



Appendix 1 Figure 4. Distributions of polyU tail mutant U224G for SgrS (top) and *ptsG* mRNA (bottom) at various time points from 0 to 20 minutes post-induction. Data from smFISH–STORM experiments (red, 100-200 cells per time point) and stochastic simulations (blue, 2000 cells per time point) are shown as kernel density estimates. Average copy number at each time point is are displayed with dashed vertical lines.

In order to attempt focus on a point mutation that primarily showed a disruption in SgrS-Hfq association we sought a mutant in which SgrS secondary structure would not be significantly disrupted, leading to a higher free degradation rate of SgrS. Via *in silico* folding using the RNA structure prediction tool mFold (***Zuker, 2003***), we confirmed that the stability

of the U224G with a ΔG of $-17.60\ kcal/mol$ is unchanged from the predicted wild-type value of $-17.60\ kcal/mol$, indicating that sRNA structure is conserved. The predicted U224G mutant structure also shows similar shape to that of the wild-type (**Appendix 1–Figure 5**) and does not show any additional stem loop formation as did other mutants. Thus, an assumption that the measured wild-type ΔHfq degradation rate (see Section 'Methods and Materials') is appropriate for use in fitting the U224G mutant data is reasonable.



Appendix 1 Figure 5. Flattened predicted sRNA structures for wild-type (WT) SgrS (left) as well as the U224G mutant (right) studied in this work obtained via mFold *in silico* folding. **Red**: the SgrS-*ptsG* mRNA baseparing region, **Blue**: the polyU tail, with the mutated residue circled in the U224G structure. The predicted structures show similar conformation as well as identical free energies ($-17.60\ kcal/mol$), indicating that SgrS secondary structure is likely not significantly destabilized by the U224G point mutation in the polyU tail.