The Molecular Biophysics of Evolutionary and Physiological Adaptation

Thesis by
Griffin Daniel Chure

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
Defended May 1, 2020

© 2020

Griffin Daniel Chure
ORCID: 0000-0002-2216-2057

ACKNOWLEDGEMENTS

This is a test acknowledgements file

ABSTRACT

This is a test abstract file.

PUBLISHED CONTENT AND CONTRIBUTIONS

Chure, G., Kaczmarek, Z.A., and Phillips, R. (2019b). Physiological Adaptability and Parametric Versatility in a Simple Genetic Circuit. BioRxiv 2019.12.19.878462.

Hirokawa, S., Chure, G., Belliveau, N.M., Lovely, G.A., Anaya, M., Schatz, D.G., Baltimore, D., and Phillips, R. (2019). Sequence-Dependent Dynamics of Synthetic and Endogenous RSSs in V(D)J Recombination. BioRxiv 791954.

Chure, G., Razo-Mejia, M., Belliveau, N.M., Einav, T., Kaczmarek, Z.A., Barnes, S.L., and Phillips, R. (2019a). Predictive Shifts in Free Energy Couple Mutations to Their Phenotypic Consequences. Proceedings of the National Academy of Sciences 116.

Phillips, R., Belliveau, N.M., Chure, G., Garcia, H.G., Razo-Mejia, M., and Scholes, C. (2019). Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. Annu. Rev. Biophys. 48, 121–163.

Chure, G.* , Lee, H.J.* , Rasmussen, A., and Phillips, R. (2018). Connecting the Dots between Mechanosensitive Channel Abundance, Osmotic Shock, and Survival at Single-Cell Resolution. Journal of Bacteriology 200. * contributed equally

Razo-Mejia, M.* , Barnes, S.L.* , Belliveau, N.M.* , Chure, G.* , Einav, T.* , Lewis, M., and Phillips, R. (2018). Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. Cell Systems 6, 456-469.e10. * contributed equally

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Number   *Page*

THROUGH THE INTRAMOLECULAR GRAPE VINE: SIGNAL PROCESSING VIA ALLOSTERIC TRANSCRIPTION FACTORS

A version of this chapter originally appeared as Razo-Mejia, M.* , Barnes, S.L.* , Belliveau, N.M.* , Chure, G.* , Einav, T.* , Lewis, M., and Phillips, R. (2018). Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. Cell Systems 6, 456-469.e10. M.R.M, S.L.B, N.M.B, G.C., and T.E. contributed equally to this work from the theoretical underpinnings to the experimental design and execution.

1.1   Introduction

Understanding how organisms sense and respond to changes in their environment has long been a central theme of biological inquiry. At the cellular level, this interaction is mediated by a diverse collection of molecular signaling pathways. A pervasive mechanism of signaling in these pathways is allosteric regulation, in which the binding of a ligand induces a conformational change in some target molecule, triggering a signaling cascade (Lindsley and Rutter, 2006). One of the most important examples of such signaling is offered by transcriptional regulation, where a transcription factor's propensity to bind to DNA will be altered upon binding to an allosteric effector.

Despite allostery's ubiquity, we lack a formal, rigorous, and generalizable framework for studying its effects across the broad variety of contexts in which it appears. A key example of this is transcriptional regulation, in which allosteric transcription factors can be induced or corepressed by binding to a ligand. An allosteric transcription factor can adopt multiple conformational states, each of which has its own affinity for the ligand and for its DNA target site. *In vitro* studies have rigorously quantified the equilibria of different conformational states for allosteric transcription factors and measured the affinities of these states to the ligand (Harman, 2001; Lanfranco et al., 2017). In spite of these experimental observations, the lack of a coherent quantitative model for allosteric transcriptional regulation has made it

impossible to predict the behavior of even a simple genetic circuit across a range of regulatory parameters.

The ability to predict circuit behavior robustly— that is, across both broad ranges of parameters and regulatory architectures —is important for multiple reasons. First, in the context of a specific gene, accurate prediction demonstrates that all components relevant to the gene's behavior have been identified and characterized to sufficient quantitative precision. Second, in the context of genetic circuits in general, robust prediction validates the model that generated the prediction. Possessing a validated model also has implications for future work. For example, when we have sufficient confidence in the model, a single data set can be used to accurately extrapolate a system's behavior in other conditions. Moreover, there is an essential distinction between a predictive model, which is used to predict a system's behavior given a set of input variables, and a retroactive model, which is used to describe the behavior of data that has already been obtained. We note that even some of the most careful and rigorous analysis of transcriptional regulation often entails only a retroactive reflection on a single experiment. This raises the fear that each regulatory architecture may require a unique analysis that cannot carry over to other systems, a worry that is exacerbated by the prevalent use of phenomenological functions (e.g. Hill functions or ratios of polynomials) that can analyze a single data set but cannot be used to extrapolate a system's behavior in other conditions (Poelwijk et al., 2011; Rogers et al., 2015; Rohlhill et al., 2017; Setty et al., 2003).

This work explores what happens when theory takes center stage, namely, we first write down the equations governing a system and describe its expected behavior across a wide array of experimental conditions, and only then do we set out to experimentally confirm these results. Building upon previous work (**???**; Brewster et al., 2014; Garcia and Phillips, 2011) and the work of Monod, Wyman, and Changeux (Monod et al., 1965), we present a statistical mechanical rendering of allostery in the context of induction and corepression (shown schematically in Fig. [1.1] (A) and henceforth referred to as the MWC model) and use it as the basis of parameter-free predictions which we then test experimentally. More specifically, we study the simple repression motif – a widespread bacterial genetic regulatory architecture in which binding of a transcription factor occludes binding of an RNA poly-

merase, thereby inhibiting transcription initiation. The MWC model stipulates that an allosteric protein fluctuates between two distinct conformations – an active and inactive state – in thermodynamic equilibrium (Monod et al., 1965). During induction, for example, effector binding increases the probability that a repressor will be in the inactive state, weakening its ability to bind to the promoter and resulting in increased expression. To test the predictions of our model across a wide range of operator binding strengths and repressor copy numbers, we design an *E. coli* genetic construct in which the binding probability of a repressor regulates gene expression of a fluorescent reporter.

In total, the work presented here demonstrates that one extremely compact set of parameters can be applied self-consistently and predictively to different regulatory situations including simple repression on the chromosome, cases in which decoy binding sites for repressor are put on plasmids, cases in which multiple genes compete for the same regulatory machinery, cases involving multiple binding sites for repressor leading to DNA looping, and induction by signaling (**???**; **???**; **???**; Boedicker et al., 2013; Garcia and Phillips, 2011). Thus, rather than viewing the behavior of each circuit as giving rise to its own unique input-output response, the MWC model provides a means to characterize these seemingly diverse behaviors using a single unified framework governed by a small set of parameters.

1.2   Results

Characterizing Transcription Factor Induction using the Monod-Wyman-Changeux (MWC) Model

We begin by considering a simple repression genetic architecture in which the binding of an allosteric repressor occludes the binding of RNA polymerase (RNAP) to the DNA . When an effector (hereafter referred to as an "inducer" for the case of induction) binds to the repressor, it shifts the repressor's allosteric equilibrium towards the inactive state as specified by the MWC model . This causes the repressor to bind more weakly to the operator, which increases gene expression. Simple repression motifs in the absence of inducer have been previously characterized by an equilibrium model where the probability of each state of repressor and RNAP promoter occupancy is dictated by the Boltzmann distribution (we note that non-equilibrium models of simple repression have been shown to have the same

(A)

induction

examples from E. coli		
transcription factor	allosteric effector	role
LacI	allolactose	metabolism
TetR	tetracycline	antibiotic resistance
NagC	GlcNAc	catabolism

RNA polymerase	
inactive repressor	
active repressor	
allosteric effector	

corepression

examples from E. coli		
transcription factor	allosteric effector	
IclR	glyoxylate	me
PurR	purines	ca
TrpR	tryptophan	ca

(B)

(C)

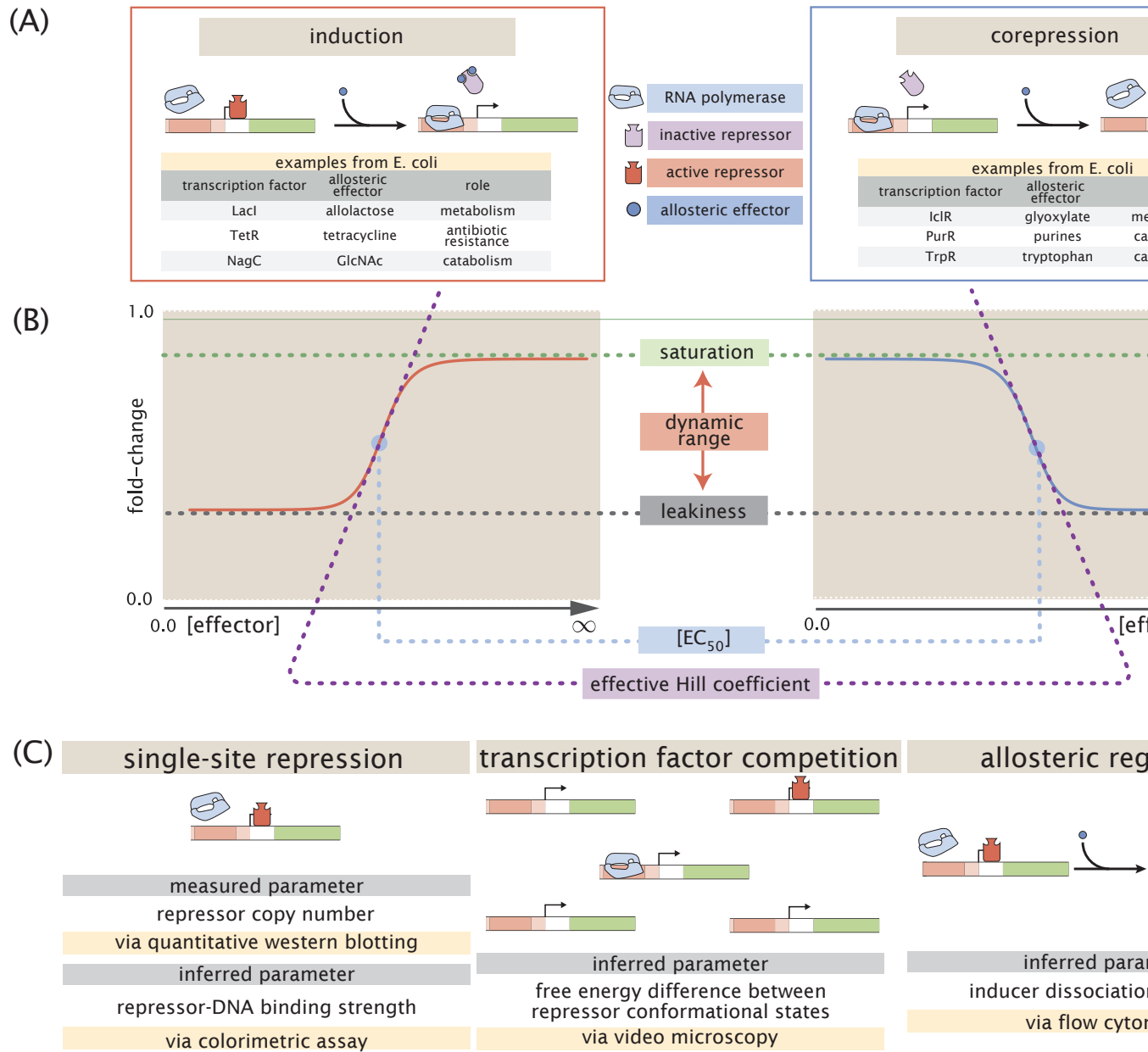| single-site repression	transcription factor competition	allosteric reg
measured parameter		
repressor copy number		
via quantitative western blotting		
inferred parameter	inferred parameter	inferred para
repressor-DNA binding strength	free energy difference between repressor conformational states	inducer dissociatio
via colorimetric assay	via video microscopy	via flow cytor

Figure 1.1: (A) We consider a promoter regulated solely by an allosteric repressor. When bound, the repressor prevents RNAP from binding and initiating transcription. Induction is characterized by the addition of an effector which binds to the repressor and stabilizes the inactive state (defined as the state which has a low affinity for DNA), thereby increasing gene expression. In corepression, the effector stabilizes the repressor's active state and thus further reduces gene expression. We list several characterized examples of induction and corepression that support different physiological roles in *E. coli* . (B) A schematic regulatory response of the two architectures shown in Panel (A) plotting the fold-change in gene expression as a function of effector concentration, where fold-change is defined as the ratio of gene expression in the presence versus the absence of repressor. We consider the following key phenotypic properties that describe each response curve: the minimum response (leakiness), the maximum response (saturation), the difference between the maximum and minimum response (dynamic range), the concentration of ligand which generates a fold-change halfway between the minimal and maximal response (EC_{50}), and the log-log slope at the mid-

functional form that we derive below). We extend these models to consider allostery by accounting for the equilibrium state of the repressor through the MWC model.

Thermodynamic models of gene expression begin by enumerating all possible states of the promoter and their corresponding statistical weights. As shown in [1.2], the promoter can either be empty, occupied by RNAP, or occupied by either an active or inactive repressor. The probability of binding to the promoter will be affected by the protein copy number, which we denote as P for RNAP, R_A for active repressor, and R_I for inactive repressor. We note that repressors fluctuate between the active and inactive conformation in thermodynamic equilibrium, such that R_A and R_I will remain constant for a given inducer concentration . We assign the repressor a different DNA binding affinity in the active and inactive state. In addition to the specific binding sites at the promoter, we assume that there are N_{NS} non-specific binding sites elsewhere (i.e. on parts of the genome outside the simple repression architecture) where the RNAP or the repressor can bind. All specific binding energies are measured relative to the average non-specific binding energy. Thus, $\Delta \varepsilon_P$ represents the energy difference between the specific and non-specific binding for RNAP to the DNA. Likewise, $\Delta \varepsilon_{RA}$ and $\Delta \varepsilon_{RI}$ represent the difference in specific and non-specific binding energies for repressor in the active or inactive state, respectively.

Thermodynamic models of transcription posit that gene expression is proportional to the probability that the RNAP is bound to the promoter p_{bound}, which is given by

$$p_{\text{bound}} = \frac{\frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_P}}{1 + \frac{R_A}{N_{NS}} e^{-\beta \Delta \varepsilon_{RA}} + \frac{R_I}{N_{NS}} e^{-\beta \Delta \varepsilon_{RI}} + \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_P}}, \tag{1.1}$$

with $\beta = \frac{1}{k_B T}$ where k_B is the Boltzmann constant and T is the temperature of the system. As $k_B T$ is the natural unit of energy at the molecular length scale, we treat the products $\beta \Delta \varepsilon_j$ as single parameters within our model. Measuring p_{bound} directly is fraught with experimental difficulties, as determining the exact proportionality between expression and p_{bound} is not straightforward. Instead, we measure the fold-change in gene expression due to the presence of the repressor. We define fold-change as the ratio of

Figure 1.2: States and weights for the simple repression motif.** RNAP (light blue) and a repressor compete for binding to a promoter of interest. There are R_A repressors in the active state (red) and R_I repressors in the inactive state (purple). The difference in energy between a repressor bound to the promoter of interest versus another non-specific site elsewhere on the DNA equals $\Delta\varepsilon_{RA}$ in the active state and $\Delta\varepsilon_{RI}$ in the inactive state; the P RNAP have a corresponding energy difference $\Delta\varepsilon_P$ relative to non-specific binding on the DNA. N_{NS} represents the number of non-specific binding sites for both RNAP and repressor. A repressor has an active conformation (red, left column) and an inactive conformation (purple, right column), with the energy difference between these two states given by $\Delta\varepsilon_{AI}$. The inducer (blue circle) at concentration c is capable of binding to the repressor with dissociation constants K_A in the active state and K_I in the inactive state. The eight states for a dimer with $n = 2$ inducer binding sites are shown along with the sums of the active and inactive states.

gene expression in the presence of repressor relative to expression in the absence of repressor (i.e. constitutive expression), namely,

$$\text{fold-change} \equiv \frac{p_{\text{bound}}(R > 0)}{p_{\text{bound}}(R = 0)}. \tag{1.2}$$

We can simplify this expression using two well-justified approximations: (1) $\frac{P}{N_{NS}}e^{-\beta\Delta\varepsilon_P} \ll 1$ implying that the RNAP binds weakly to the promoter ($N_{NS} = 4.6 \times 10^6$, $P \approx 10^3$, $\Delta\varepsilon_P \approx -2$ to $-5\,k_B T$, so that $\frac{P}{N_{NS}}e^{-\beta\Delta\varepsilon_P} \approx 0.01$) and (2) $\frac{R_I}{N_{NS}}e^{-\beta\Delta\varepsilon_{RI}} \ll 1 + \frac{R_A}{N_{NS}}e^{-\beta\Delta\varepsilon_{RA}}$ which reflects our assumption that the inactive repressor binds weakly to the promoter of interest. Using these approximations, the fold-change reduces to the form

$$\text{fold-change} \approx \left(1 + \frac{R_A}{N_{NS}}e^{-\beta\Delta\varepsilon_{RA}}\right)^{-1} \equiv \left(1 + p_A(c)\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{RA}}\right)^{-1}, \tag{1.3}$$

where in the last step we have introduced the fraction $p_{\text{act}}(c)$ of repressors in the active state given a concentration c of inducer, such that $R_A(c) = p_{\text{act}}(c)R$. Since inducer binding shifts the repressors from the active to the inactive state, $p_{\text{act}}(c)$ grows smaller as c increases.

1.3 References

'WATER, WATER EVERYWHERE, NOR ANY DROP TO DRINK': HOW BACTERIA ADAPT TO CHANGES IN OSMOLARITY

A version of this chapter was published as Chure, G.* , Lee, H.J.* , Rasmussen, A., and Phillips, R. (2018). *Connecting the Dots between Mechanosensitive Channel Abundance, Osmotic Shock, and Survival at Single-Cell Resolution.* Journal of Bacteriology 200. (* contributed equally). G.C., H.J.L, and R.P. designed and planned experiments. G.C. and H.J.L performed experiments. H.J.L constructed bacterial strains. A.R. performed electrophysiology experiments. G.C. performed data analysis and figure generation. G.C. and R.P. wrote the manuscript

2.1 Introduction

Changes in the extracellular osmolarity can be a fatal event for the bacterial cell. Upon a hypo-osmotic shock, water rushes into the cell across the membrane, leaving the cell with no choice but to equalize the pressure. This equalization occurs either through damage to the cell membrane (resulting in death) or through the regulated flux of water molecules through transmembrane protein channels (Fig 1A). Such proteinaceous pressure release valves have been found across all domains of life, with the first bacterial channel being described in 1987 (**???**). Over the past thirty years, several more channels have been discovered, described, and (in many cases) biophysically characterized. *E. coli*, for example, has seven of these channels (one MscL and six MscS homologs) which have varied conductance, gating mechanisms, and expression levels. While they have been the subject of much experimental and theoretical dissection, much remains a mystery with regard to the roles their abundance and interaction with other cellular processes play in the greater context of physiology (**???**; **???**; **???**; **???**; **???**; **???**; **???**).

Of the seven channels in *E. coli*, the mechanosensitive channel of large conductance (MscL) is one of the most abundant and the best characterized. This channel has a large conductance (3 nS) and mediates the flux of water

molecules across the membrane via a ~3 nm wide pore in the open state (**???; ???**). Molecular dynamics simulations indicate that a single open MscL channel permits the flux of 4×10^9 water molecules per second, which is an order of magnitude larger than a single aquaporin channel (BNID 100479) (**???; ???**). This suggests that having only a few channels per cell could be sufficient to relieve even large changes in membrane tension. Electrophysiological experiments have suggested a small number of channels per cell (**???; ???**), however, more recent approaches using quantitative Western blotting, fluorescence microscopy, and proteomics have measured several hundred MscL per cell (**???; ???; ???**). To further complicate matters, the expression profile of MscL appears to depend on growth phase, available carbon source, and other environmental challenges (**???, @schmidt2016; ???; ???**). While there are likely more than just a few channels per cell, why cells seem to need so many and the biological rationale behind their condition-dependent expression both remain a mystery.

While their biochemical and biophysical characteristics have received much attention, their connection to cell survival is understudied. Drawing such a direct connection between channel copy number and survival requires quantitative *in vivo* experiments. To our knowledge, the work presented in van den Berg et al. 2016 (**???**) is the first attempt to simultaneously measure channel abundance and survivability for a single species of mechanosensitive channel. While the measurement of channel copy number was performed at the level of single cells using super-resolution microscopy, survivability after a hypo-osmotic shock was assessed in bulk plating assays which rely on serial dilutions of a shocked culture followed by counting the number of resulting colonies after incubation. Such bulk assays have long been the standard for querying cell viability after an osmotic challenge. While they have been highly informative, they reflect only the mean survival rate of the population, obfuscating the variability in survival of the population. The stochastic nature of gene expression results in a noisy distribution of MscL channels rather than a single value, meaning those found in the long tails of the distribution have quite different survival rates than the mean but are lost in the final calculation of survival probability.

In this work, we present an experimental system to quantitatively

probe the interplay between MscL copy number and survival at single-cell resolution, as is seen in (**???**)B. We generated an *E. coli* strain in which all seven mechanosensitive channels had been deleted from the chromosome followed by a chromosomal integration of a single gene encoding an MscL-super-folder GFP (sfGFP) fusion protein. To explore copy number regimes beyond those of the wild-type expression level, we modified the Shine-Dalgarno sequence of this integrated construct, allowing us to cover nearly three decades of MscL copy number. To probe survivability, we exposed cells to a large hypo-osmotic shock at controlled rates in a flow cell under a microscope, allowing the observation of the single-cell channel copy number and the resulting survivability of single cells. With this large set of single cell measurements, we approach the calculation of survival probability in a manner that is free of binning bias which allows the reasonable extrapolation of survival probability to copy numbers outside of the observed range. In addition, we show that several hundred channels are needed to convey high rates of survival and observe a minimum number of channels needed to permit any degree of survival.

2.2 Results

Quantifying the single-cell MscL copy number

The principal goal of this work is to examine the contribution of a single mechanosensitive channel species to cell survival under a hypo-osmotic shock. While this procedure could be performed for any species of channel, we chose MscL as it is the most well characterized and one of the most abundant species in *E. coli*. To probe the contribution of MscL alone, we integrated an *mscL* gene encoding an MscL super-folder GFP (sfGFP) fusion into a strain in which all seven known mechanosensitive channel genes were deleted from the chromosome (**???**). Chromosomal integration imposes strict control on the gene copy number compared to plasmid borne expression systems, which is important to minimize variation in channel expression across the population and provide conditions more representative of native cell physiology. Abrogation of activity, mislocalization, or cytotoxicity are all inherent risks associated with creating chimeric reporter constructs. In Supplement A, we carefully dissect the functionality of this protein through electrophysiology (Fig. S1), measure the rate of fluorophore maturation (Fig. S2), and quantify potential aggregates (Figs. S3 and S4). To
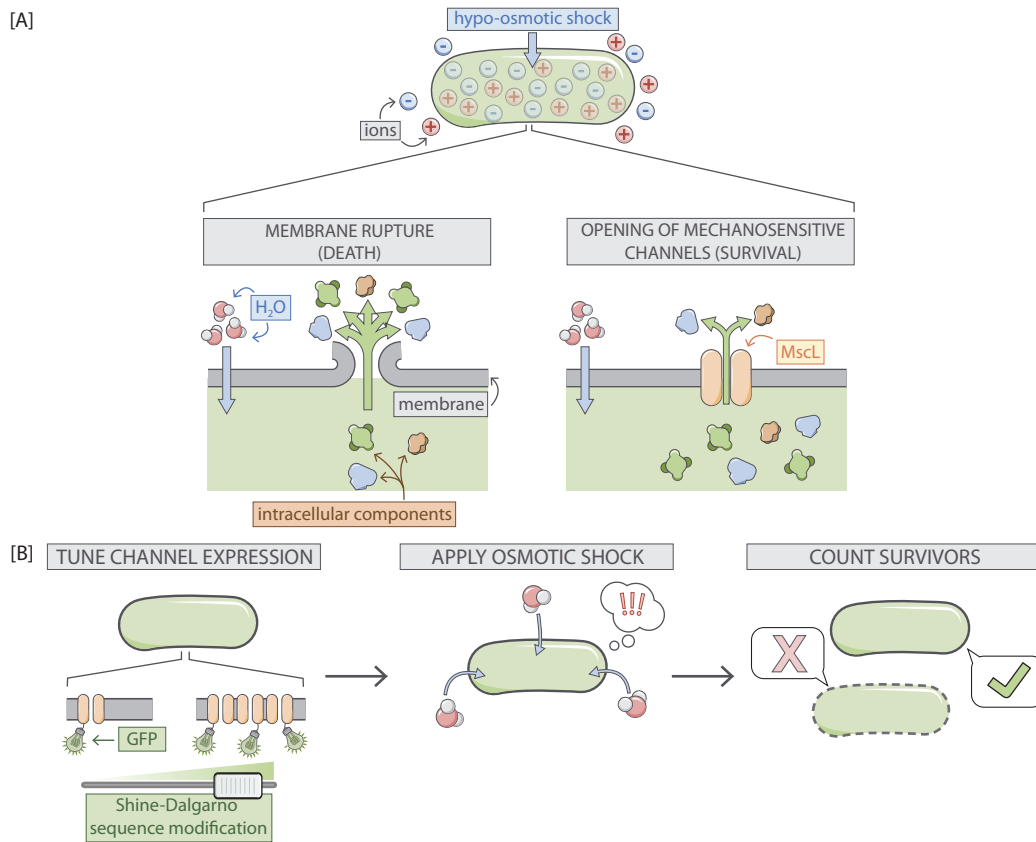
Figure 2.1: Role of mechanosensitive channels during hypo-osmotic shock. (A) A hypo-osmotic shock results in a large difference in the osmotic strength between the intracellular and extracellular spaces. As a result, water rushes into the cell to equalize this gradient increasing the turgor pressure and tension in the cell membrane. If no mechanosensitive channels are present and membrane tension is high (left panel), the membrane ruptures releasing intracellular content into the environment resulting in cell death . If mechanosensitive channels are present (right panel) and membrane tension is beyond the gating tension, the mechanosensitive channel MscL opens, releasing water and small intracellular molecules into the environment thus relieving pressure and membrane tension. (B) The experimental approach undertaken in this work. The number of mechanosensitive channels tagged with a fluorescent reporter is tuned through modification of the Shine-Dalgarno sequence of the *mscL* gene. The cells are then subjected to a hypo-osmotic shock and the number of surviving cells are counted, allowing the calculation of a survival probability.

the best of our knowledge, the MscL-sfGFP fusion protein functions identically to the wild-type, allowing us to confidently draw conclusions about the physiological role this channel plays in wild-type cells.

To modulate the number of MscL channels per cell, we developed a series of mutants which were designed to decrease the expression relative to wild-type. These changes involved direct alterations of the Shine-Dalgarno sequence as well as the inclusion of AT hairpins of varying length directly upstream of the start codon which influences the translation rate and hence the number of MscL proteins produced ((**???**)A). The six Shine-Dalgarno sequences used in this work were chosen using the RBS binding site strength calculator from the Salis Laboratory at the Pennsylvania State University (**???**; Salis et al., 2009). While the designed Shine-Dalgarno sequence mutations decreased the expression relative to wild-type as intended, the distribution of expression is remarkably wide spanning an order of magnitude.

To measure the number of MscL channels per cell, we determined a fluorescence calibration factor to translate arbitrary fluorescence units per cell to protein copy number. While there have been numerous techniques developed over the past decade to directly measure this calibration factor, such as quantifying single-molecule photobleaching constants or measuring the binomial partitioning of fluorescent proteins upon cell division (**???**; **???**), we used *a priori* knowledge of the mean MscL-sfGFP expression level of a particular *E. coli* strain to estimate the average fluorescence of a single channel. In Bialecka-Fornal et al. 2012 (**???**), the authors used single-molecule photobleaching and quantitative Western blotting to probe the expression of MscL-sfGFP under a wide range of growth conditions. To compute a calibration factor, we used the strain MLG910 (*E. coli* K12 MG1655 ϕ(mscL-sfGFP)) as a "standard candle", highlighted in yellow in (**???**)B. This standard candle strain was grown and imaged in identical conditions in which the MscL count was determined through fluorescence microscopy. The calibration factor was computed by dividing the mean total cell fluorescence by the known MscL copy number, resulting in a measure of arbitrary fluorescence units per MscL channel. Details regarding this calculation and appropriate propagation of error as well as its sensitivity to varying growth media can be found in the Materials & Methods as well as Supplement B (Fig. S5 - S8).

While it is seemingly straightforward to use this calibration factor to determine the total number of channels per cell for wild-type or highly expressing strains, the calculation for the lowest expressing strains is complicated by distorted cell morphology. We observed that as the channel copy number decreases, cellular morphology becomes increasingly aberrant with filamentous, bulging, and branched cells becoming more abundant (Fig. S7A). This morphological defect has been observed when altering the abundance of several species of mechanosensitive channels, suggesting that they play an important role in general architectural stability (**???; ???**). As these aberrant morphologies can vary widely in size and shape, calculating the number of channels per cell becomes a more nuanced endeavor. For example, taking the total MscL copy number for these cells could skew the final calculation of survival probability as a large but severely distorted cell would be interpreted as having more channels than a smaller, wild-type shaped cell (Fig. S7B). To correct for this pathology, we computed the average expression level per unit area for each cell and multiplied this by the average cellular area of our standard candle strain which is morphologically indistinguishable from wild-type *E. coli*, allowing for the calculation of an effective channel copy number. The effect of this correction can be seen in Fig. S7C and D, which illustrate that there is no other correlation between cell area and channel expression.

Our calculation of the effective channel copy number for our suite of Shine-Dalgarno mutants is shown in (**???**)B. The expression of these strains cover nearly three orders of magnitude with the extremes ranging from approximately four channels per cell to nearly one thousand. While the means of each strain are somewhat distinct, the distributions show a large degree of overlap, making one strain nearly indistinguishable from another. This variance is a quantity that is lost in the context of bulk scale experiments but can be accounted for via single-cell methods.

Control of MscL expression and calculation of channel copy number. (A) Schematic view of the expression modifications performed in this work. The beginning portion of the native *mscL* sequence is shown with the Shine-Dalgarno sequence, spacer region, and start codon shaded in red, green, and blue, respectively. The Shine-Dalgarno sequence was modified through the Salis lab Ribosomal Binding Strength calculator (**???**; Salis et al., 2009).

The wild-type sequence (MLG910) is shown in black with mutations for the other four Shine-Dalgarno mutants highlighted in red. Expression was further modified by the insertion of repetitive `AT` bases into the spacer region, generating hairpins of varying length which acted as a thermodynamic barrier for translation initiation. (B) Variability in effective channel copy number is computed using the standard candle. The boxes represent the interquartile region of the distribution, the center line displays the median, and the whiskers represent 1.5 times the maximum and minimum of the interquartile region. Individual measurements are denoted as black points. The strain used for calibration of channel copy number (MLG910) is highlighted in yellow.

Performing a single-cell hypo-osmotic challenge assay

To measure the channel copy number of a single cell and query its survival after a hypo-osmotic shock, we used a custom-made flow cell in which osmotic shock and growth can be monitored in real time using video microscopy ((**???**)A). The design and characterization of this device has been described in depth previously and is briefly described in the Materials & Methods (**???**). Using this device, cells were exposed to a large hypo-osmotic shock by switching between LB Lennox medium supplemented with 500 mM NaCl and LB Lennox media alone. All six Shine-Dalgarno modifications shown in (**???**)B (excluding MLG910) were subjected to a hypo-osmotic shock at controlled rates while under observation. After the application of the osmotic shock, the cells were imaged every sixty seconds for four to six hours. Each cell was monitored over the outgrowth period and was manually scored as either a survivor, fatality, or inconclusive observation. The criteria used for scoring death were the same as those previously described in Bialecka-Fornal et al. 2015 (**???**). Survivors were defined as cells that underwent multiple divisions post-shock. To qualify as survivors, cells must undergo at least two divisions, although more typically, four to eight divisions are observed without any signs of slowing down. Imaging is stopped when the survivors cells begin to go out of focus or overlap each other. Survivors do not show any sign of ceasing division. More information regarding this classification can be found in the Materials and Methods as well as Supplement C (Fig. S9 - S10 and Table S1 - S2). The brief experimental protocol can be seen in (**???**)B.

Experimental approach to measuring survival probability. (A) Layout of a home-made flow cell for subjecting cells to osmotic shock. Cells are attached to a polyethylenimine functionalized surface of a glass coverslip within the flow chamber by loading a dilute cell suspension through one of the inlets. (B) The typical experimental procedure. Cells are loaded into a flow chamber as shown in (A) and mounted to the glass coverslip surface. Cells are subjected to a hypo-osmotic shock by flowing hypotonic medium into the flow cell. After shock, the cells are monitored for several hours and surviving cells are identified.

Due to the extensive overlap in expression between the different Shine-Dalgarno mutants (see (**???**)B), computing the survival probability by treating each mutant as an individual bin obfuscates the relationship between channel abundance and survival. To more thoroughly examine this relationship, all measurements were pooled together with each cell being treated as an individual experiment. The hypo-osmotic shock applied in these experiments was varied across a range of 0.02 Hz (complete exchange in 50 s) to 2.2 Hz (complete exchange in 0.45 s). Rather than pooling this wide range of shock rates into a single data set, we chose to separate the data into "slow shock" (< 1.0 Hz) and "fast shock" ($\geq$ 1.0 Hz) classes. Other groupings of shock rate were explored and are discussed in Supplement D (Fig. S11 and S12). The cumulative distributions of channel copy number separated by survival are shown in (**???**). In these experiments, survival was never observed for a cell containing less than approximately 100 channels per cell, indicated by the red stripe in (**???**). This suggests that there is a minimum number of channels needed for survival on the order of 100 per cell. We also observe a slight shift in the surviving fraction of the cells towards higher effective copy number, which matches our intuition that including more mechanosensitive channels increases the survival probability.

Distributions of survival and death as a function of effective channel number. (A) Empirical cumulative distributions of channel copy number separated by survival (green) or death (purple) after a slow (< 1.0 Hz) osmotic shock. (B) The empirical cumulative distribution for a fast ($\geq$ 1.0 Hz) osmotic shock. Shaded green and purple regions represent the 95% credible region of the effective channel number calculation for each cell. Shaded red stripe signifies the range of channels in which no survival was observed.

Prediction of survival probability as a function of channel copy number

There are several ways by which the survival probability can be calculated. The most obvious approach would be to group each individual Shine-Dalgarno mutant as a single bin and compute the average MscL copy number and the survival probability. Binning by strain is the most frequently used approach for such measurements and has provided valuable insight into the qualitative relationship of survival on other physiological factors (**???**; **???**). However the copy number distribution for each Shine-Dalgarno mutant ((**???**)B) is remarkably wide and overlaps with the other strains. We argue that this coarse-grained binning negates the benefits of performing single-cell measurements as two strains with different means but overlapping quartiles would be treated as distinctly different distributions.

Another approach would be to pool all data together, irrespective of the Shine-Dalgarno mutation, and bin by a defined range of channels. Depending on the width of the bin, this could allow for finer resolution of the quantitative trend, but the choice of the bin width is arbitrary with the *a priori* knowledge that is available. Drawing a narrow bin width can easily restrict the number of observed events to small numbers where the statistical precision of the survival probability is lost. On the other hand, drawing wide bins increases the precision of the estimate, but becomes further removed from a true single-cell measurement and represents a population mean, even though it may be a smaller population than binning by the Shine-Dalgarno sequence alone. In both of these approaches, it is difficult to extrapolate the quantitative trend outside of the experimentally observed region of channel copy number. Here, we present a method to estimate the probability of survival for any channel copy number, even those that lie outside of the experimentally queried range.

To quantify the survival probability while maintaining single-cell resolution, we chose to use a logistic regression model which does not require grouping data into arbitrary bins and treats each cell measurement as an independent experiment. Logistic regression is an inferential method to model the probability of a Boolean or categorical event (such as survival or death) given one or several predictor variables and is commonly used in medical statistics to compute survival rates and dose response curves (**???**; **???**). The primary assumption of logistic regression is that the log-odds

probability of survival p_s is linearly dependent on the predictor variable, in our case the log channels per cell N_c with a dimensionless intercept β_0 and slope β_1,

$$\log \frac{p_s}{1 - p_s} = \beta_0 + \beta_1 \log N_c. \tag{2.1}$$

Under this assumption of linearity, β_0 is the log-odds probability of survival with no MscL channels. The slope β_1 represents the change in the log-odds probability of survival conveyed by a single channel. As the calculated number of channels in this work spans nearly three orders of magnitude, it is better to perform this regression on $\log N_c$ as regressing on N_c directly would give undue weight for lower channel copy numbers due to the sparse sampling of high-copy number cells. The functional form shown in (**???**) can be derived directly from Bayes' theorem and is shown in Supplement E. If one knows the values of β_0 and β_1, the survival probability can be expressed as

$$p_s = \frac{1}{1 + N_c^{-\beta_1} e^{-\beta_0}}. \tag{2.2}$$

In this analysis, we used Bayesian inferential methods to determine the most likely values of the coefficients and is described in detail in the Supplement E (Fig. S13 and S14).

The results of the logistic regression are shown in (**???**). We see a slight rightward shift the survival probability curve under fast shock relative to the slow shock case, reaffirming the conclusion that survival is also dependent on the rate of osmotic shock (**???**). This rate dependence has been observed for cells expressing MscL alongside other species of mechanosensitive channels, but not for MscL alone. This suggests that MscL responds differently to different rates of shock, highlighting the need for further study of rate dependence and the coordination between different species of mechanosensitive channels. (**???**) also shows that several hundred channels are required to provide appreciable protection from osmotic shock. For a survival probability of 80%, a cell must have approximately 500 to 700 channels per cell for a fast and slow shock, respectively. The results from the logistic regression are showed as continuous colored curves. The individual cell measurements separated by survival and death are shown at the top and bottom of each plot, respectively, and are included to provide a sense of sampling density.

Over the explored range of MscL copy number, we observed a maximum of 80% survival for any binning method. The remaining 20% survival may be attained when the other species of mechanosensitive channels are expressed alongside MscL. However, it is possible that the flow cell method performed in this work lowers the maximal survival fraction as the cells are exposed to several, albeit minor, mechanical stresses such as loading into the flow cell and chemical adherence to the glass surface. To ensure that the results from logistic regression accurately describe the data, we can compare the survival probabilities to those using the binning methods described earlier (red and black points, (**???**)). Nearly all binned data fall within error of the prediction (see Materials & Methods for definition of error bar on probability), suggesting that this approach accurately reflects the survival probability and gives license to extrapolate the estimation of survival probability to regions of outside of our experimentally explored copy number regime.

Thus far, we've dictated that for a given rate of osmotic shock (i.e. "fast" or "slow"), the survival probability is dependent only on the number of channels. In Fig. S13, we show the result of including other predictor variables, such as area and shock rate alone. In such cases, including other predictors resulted in pathological curves showing that channel copy number is the most informative out of the available predictor variables.

Probability of survival as a function of MscL copy number. (A) Estimated survival probability for survival under slow shock as a function of channel copy number. (B) The estimated survival probability of survival under a fast shock as a function of channel copy number. Solid curves correspond to the most probable survival probability from a one-dimensional logistic regression. Shaded regions represent the 95% credible regions. Points at the top and bottom of plots represent individual cell measurements which survived and perished, respectively. The red and black points correspond to the survival probability estimated via binning by Shine-Dalgarno sequence and binning by groups of 50 channels per cell, respectively. Horizontal error bars represent the standard error of the mean from at least 25 measurements. Vertical error bars represent the certainty of the probability estimate given n survival events from N total observations.

2.3   Discussion

One of the most challenging endeavors in the biological sciences is linking the microscopic details of cellular components to the macro-scale physiology of the organism. This formidable task has been met repeatedly in the recent history of biology, especially in the era of DNA sequencing and single molecule biochemistry. For example, the scientific community has been able to connect sickle-cell anemia to a single amino acid substitution in Hemoglobin which promotes precipitation under a change in O_2 partial pressure (**???**; **???**; **???**). Others have assembled a physical model that quantitatively describes chemosensation in bacteria (**???**) in which the arbiter of sensory adaptation is the repeated methylation of chemoreceptors (**???**; **???**; **???**; **???**). In the past ~50 years alone, numerous biological and physical models of the many facets of the central dogma have been assembled that give us a sense of the interplay between the genome and physiology. For example, the combination of biochemical experimentation and biophysical models have given us a picture of how gene dosage affects furrow positioning in *Drosophila* (**???**), how recombination of V(D)J gene segments generates an extraordinarily diverse antibody repertoire (**???**; **???**; **???**), and how telomere shortening through DNA replication is intrinsically tied to cell senescence (**???**; **???**), to name just a few of many such examples.

By no means are we "finished" with any of these topics. Rather, it's quite the opposite in the sense that having a handle on the biophysical knobs that tune the behavior opens the door to a litany of new scientific questions. In the case of mechanosenstaion and osmoregulation, we have only recently been able to determine some of the basic facts that allow us to approach this fascinating biological phenomenon biophysically. The dependence of survival on mechanosensitive channel abundance is a key quantity that is missing from our collection of critical facts. To our knowledge, this work represents the first attempt to quantitatively control the abundance of a single species of mechanosensitive channel and examine the physiological consequences in terms of survival probability at single-cell resolution. Our results reveal two notable quantities. First, out of the several hundred single-cell measurements, we never observed a cell which had less than approximately 100 channels per cell and survived an osmotic shock, irrespective of the shock rate. The second is that between 500 and 700 channels per cell are needed to provide $\geq 80\%$ survival, depending on the shock rate.

Only recently has the relationship between the MscL copy number and the probability of survival been approached experimentally. In van den Berg et al. (2016), the authors examined the contribution of MscL to survival in a genetic background where all other known mechanosensitive channels had been deleted from the chromosome and plasmid-borne expression of an MscL-mEos3.2 fusion was tuned through an IPTG inducible promoter (**???**). In this work, they measured the single-cell channel abundance through super-resolution microscopy and queried survival through bulk assays. They report a nearly linear relationship between survival and copy number, with approximately 100 channels per cell conveying 100% survival. This number is significantly smaller than our observation of approximately 100 channels as the *minimum* number needed to convey an y observable degree of survival.

The disagreement between the numbers reported in this work and in van den Berg et al. may partially arise from subtle differences in the experimental approach. The primary practical difference is the magnitude of the osmotic shock. van den Berg et al. applied an approximately 600 mOsm downshock in bulk whereas we applied a 1 Osm downshock, which would lead to lower survival (**???**). In their work, the uncertainty in both the MscL channel count and survival probability is roughly 30% (Fig. S14). Given this uncertainty, it is reasonable to interpret that the number of channels needed for complete protection from osmotic downshock is between 100 and 250 per cell. The uncertainty in determining the number of channels per cell is consistent with the observed width of the channel number distribution of the Shine-Dalgarno sequence mutants used in this work ((**???**)B). A unique property of the single-cell measurements performed in this work allow is the direct observation of survival or death of individual cells. We find that morphological classification and classification through a propidium iodide staining agree within 1% (Supplement C). The bulk plating assays, as are used in van den Berg et al., rely on colony formation and outgrowth to determine survival probability. As is reported in their supplemental information, the precision in this measurement is around 30% (Fig. S14). Accounting for this uncertainty brings both measurements within a few fold where we still consistently observe lower survival for a given channel number. This remaining disagreement may be accounted for by systematic uncertainty in both experimental methods.

For example, variation in the length of outgrowth, variable shock rate, and counting statistics could bias towards higher observed survival rates in ensemble plating assays. During the outgrowth phase, the control sample not exposed to an osmotic shock is allowed to grow for approximately 30 minutes in a high-salt medium before plating. The shocked cells, however, are allowed to grow in a low-salt medium. We have found that the difference between the growth rates in these two conditions can be appreciable (approximately 35 minutes versus 20 minutes, respectively) as can be seen in Fig S2. Cells that survived an osmotic shock may have a growth advantage relative to the control sample if the shock-induced lag phase is less than the outgrowth, leading to higher observed survival rates (**???**). This is one possible explanation for the survival rates which are reported in excess of 100%. Cells that survived an osmotic shock may have a growth advantage relative to the normalization sample if the shock-induced lag phase is less than the outgrowth, leading to higher observed survival rates, even surpassing 100%. We have performed these assays ourselves and have observed survival rates above of 100% (ranging from 110% to 125%) with an approximate 30% error (see Fig. S3 in Bialecka-Fornal et al. 2012 (**???**)) which we concluded to arise from differences in growth rate. We also note that survival rates greater than 100% are observed in van den Berg et al. (Fig. S14). For strains that have survival rates between 80% and 100% the uncertainty is typically large, making it difficult to make precise statements regarding when full survival is achieved.

It has been shown that there is a strong inverse relationship between the rate of osmotic shock and survival probability (**???**). Any experiment in which the shock was applied more slowly or quickly than another would bias toward higher or lower survivability, respectively. The shocks applied in bulk assays are often performed manually which can be highly variable. We note that in our experiments, we frequently observe cells which do not separate and form chains of two or more cells (Fig. S9 and S10). In plating assays, it is assumed that colonies arise from a single founding cell however a colony formed by a cluster of living and dead cells would be interpreted as a single surviving cell, effectively masking the death of the others in the colony forming unit. This too could bias the measurement toward higher survival rates. Single-cell shock experiments can also have systematic errors which can bias the results towards lower survival rates. Such errors

are associated with handling of the cells such as shear damage from loading into the flow cell, adhering the cells to the coverslip, and any chemical perturbations introduced by the dye used to measure the shock rate.

Despite these experimental differences, the results of this work and van den Berg et al., are in agreement that MscL must be present at the level of 100 or more channels per cell in wild-type cells to convey appreciable survival. As both of these works were performed in a strain in which the only mechanosensitive channel was MscL, it remains unknown how the presence of the other channel species would alter the number of MscL needed for complete survival. In our experiments, we observed a maximum survival probability of approximately 80% even with close to 1000 MscL channels per cell. It is possible that the combined effort of the six other mechanosensitive channels would make up for some if not all of the remaining 20%. To explore the contribution of another channel to survival, van den Berg et al. also queried the contribution of MscS, another mechanosensitive channel, to survival in the absence of any other species of mechansensitive channel. It was found that over the explored range of MscS channel copy numbers, the maximum survival rate was approximately 50%, suggesting that different mechanosensitive channels have an upper limit to how much protection they can confer. Both van den Berg et al. and our work show that there is still much to be learned with respect to the interplay between the various species of mechanosensitive channel as well as their regulation.

Recent work has shown that both magnitude and the rate of osmotic down shock are important factors in determining cell survival (**???**). In this work, we show that this finding holds true for a single species of mechanosensitive channel, even at high levels of expression. One might naïvely expect that this rate-dependent effect would disappear once a certain threshold of channels had been met. Our experiments, however, show that even at nearly 1000 channels per cell the predicted survival curves for a slow (< 1.0 Hz) and fast (≥ 1.0 Hz) are shifted relative to each other with the fast shock predicting lower rates of survival. This suggests either we have not reached this threshold in our experiments or there is more to understand about the relationship between abundance, channel species, and the shock rate.

Some experimental and theoretical treatments suggest that only a few copies of MscL or MscS should be necessary for 100% protection given our

knowledge of the conductance and the maximal water flux through the channel in its open state (**???**; **???**). However, recent proteomic studies have revealed average MscL copy numbers to be in the range of several hundred per cell, depending on the condition, as can be seen in Table 1 (**???**; **???**; Li et al., 2014). Studies focusing solely on MscL have shown similar counts through quantitative Western blotting and fluorescence microscopy (**???**). Electrophysiology studies have told another story with copy number estimates ranging between 4 and 100 channels per cell (**???**; **???**, @booth2005). These measurements, however, measure the active number of channels. The factors regulating channel activity in these experiments could be due to perturbations during the sample preparation or reflect some unknown mechanism of regulation, such as the presence or absence of interacting cofactors (**???**). The work described here, on the other hand, measures the *maximum* number of channels that could be active and may be able to explain why the channel abundance is higher than estimated by theoretical means. There remains much more to be learned about the regulation of activity in these systems. As the *in vivo* measurement of protein copy number becomes accessible through novel single-cell and single-molecule methods, we will continue to collect more facts about this fascinating system and hopefully connect the molecular details of mechanosensation with perhaps the most important physiological response – life or death.

2.4 Materials & Methods

Bacterial strains and growth conditions

The bacterial strains are described in Table S1. The parent strain for the mutants used in this study was MJF641 (**???**), a strain which had all seven mechanosensitive channels deleted. The MscL-sfGFP coding region from MLG910 (**???**) was integrated into MJF641 by P1 transduction, creating the strain D6LG-Tn10. Selection pressure for MscL integration was created by incorporating an osmotic shock into the transduction protocol, which favored the survival of MscL-expressing stains relative to MJF641 by ~100-fold. Screening for integration candidates was based on fluorescence expression of plated colonies. Successful integration was verified by sequencing. Attempts to transduce RBS-modified MscL-sfGFP coding regions became increasingly inefficient as the targeted expression level of MscL was reduced. This was due to the decreasing fluorescence levels and survival

rates of the integration candidates. Consequently, Shine-Dalgarno sequence modifications were made by inserting DNA oligos with lambda Red-mediated homologous recombination, i.e., recombineering (???). The oligos had a designed mutation ((???)) flanked by ~25 base pairs that matched the targeted MscL region (Table S2). A two-step recombineering process of selection followed by counter selection using a *tetA-sacB* gene fusion cassette (???) was chosen because of its capabilities to integrate with efficiencies comparable to P1 transduction and not leave antibiotic resistance markers or scar sequences in the final strain. To prepare the strain D6LG-Tn10 for this scheme, the Tn10 transposon containing the *tetA* gene needed to be removed to avoid interference with the *tetA-sacB* cassette. Tn10 was removed from the middle of the *ycjM* gene with the primer Tn10delR (Table S2) by recombineering, creating the strain D6LG (SD0). Counter selection against the *tetA* gene was promoted by using agar media with fusaric acid (???; ???). The *tetA-sacB* cassette was PCR amplified out of the strain XTL298 using primers MscLSPSac and MscLSPSacR (Table S2). The cassette was integrated in place of the spacer region in front of the MscL start codon of D6LG (SD0) by recombineering, creating the intermediate strain D6LTetSac. Positive selection for cassette integration was provided by agar media with tetracycline. Finally, the RBS modifying oligos were integrated into place by replacing the *tetA-sacB* cassette by recombineering. Counter selection against both *tetA* and *sacB* was ensured by using agar media with fusaric acid and sucrose (???), creating the Shine-Dalgarno mutant strains used in this work.

Strain cultures were grown in 5 mL of LB-Lennox media with antibiotic (apramycin) overnight at $37°$C. The next day, 50 μL of overnight culture was inoculated into 5 mL of LB-Lenox with antibiotic and the culture was grown to OD600nm ~0.25. Subsequently, 500 μL of that culture was inoculated into 5 mL of LB-Lennox supplemented with 500mM of NaCl and the culture was regrown to OD600nm ~0.25. A 1 mL aliquot was taken and used to load the flow cell.

Flow cell

All experiments were conducted in a home-made flow cell as is shown in (???)A. This flow cell has two inlets which allow media of different osmolarity to be exchanged over the course of the experiment. The imaging region

is approximately 10 mm wide and 100 μm in depth. All imaging took place within 1 – 2 cm of the outlet to avoid imaging cells within a non-uniform gradient of osmolarity. The interior of the flow cell was functionalized with a 1:400 dilution of polyethylenimine prior to addition of cells with the excess washed away with water. A dilute cell suspension in LB Lennox with 500 mM NaCl was loaded into one inlet while the other was connected to a vial of LB medium with no NaCl. This hypotonic medium was clamped during the loading of the cells.

Once the cells had adhered to the polyethylenimine coated surface, the excess cells were washed away with the 500 mM NaCl growth medium followed by a small (~20 μL) air bubble. This air bubble forced the cells to lay flat against the imaging surface, improving the time-lapse imaging. Over the observation period, cells not exposed to an osmotic shock were able to grow for 4 – 6 divisions, showing that the flow cell does not directly impede cell growth.

Imaging conditions

All imaging was performed in a flow cell held at 30°C on a Nikon Ti-Eclipse microscope outfitted with a Perfect Focus system enclosed in a Haison environmental chamber (approximately 1°C regulation efficiency). The microscope was equipped with a 488 nm laser excitation source (CrystaLaser) and a 520/35 laser optimized filter set (Semrock). The images were collected on an Andor iXon EM+ 897 EMCCD camera and all microscope and acquisition operations were controlled via the open source μManager microscope control software (**???**). Once cells were securely mounted onto the surface of the glass coverslip, between 15 and 20 positions containing 5 to 10 cells were marked and the coordinates recorded. At each position, a phase contrast and GFP fluorescence image was acquired for segmentation and subsequent measurement of channel copy number. To perform the osmotic shock, LB media containing no NaCl was pulled into the flow cell through a syringe pump. To monitor the media exchange, both the high salt and no salt LB media were supplemented with a low-affinity version of the calcium-sensitive dye Rhod-2 (250 nM; TEF Labs) which fluoresces when bound to Ca^{2+}. The no salt medium was also supplemented with 1μM $CaCl_2$ to make the media mildly fluorescent and the exchange rate was calculated by measuring the fluorescence increase across an illuminated sec-

tion of one of the positions. These images were collected in real time for the duration of the shock. The difference in measured fluorescence between the pre-shock images and those at the end of the shock set the scale of a 500 mM NaCl down shock. The rate was calculated by fitting a line to the middle region of this trace. Further details regarding this procedure can be found in Bialecka-Fornal et al. 2015 (**???**).

Image Processing

Images were processed using a combination of automated and manual methods. First, expression of MscL was measured via segmenting individual cells or small clusters of cells in phase contrast and computing the mean pixel value of the fluorescence image for each segmented object. The fluorescence images were passed through several filtering operations which reduced high-frequency noise as well as corrected for uneven illumination of the excitation wavelength.

Survival or death classification was performed manually using the Cell-Profiler plugin for ImageJ software (NIH). A survivor was defined as a cell which was able to undergo at least two division events after the osmotic down shock. Cell death was recognized by stark changes in cell morphology including loss of phase contrast through ejection of cytoplasmic material, structural decomposition of the cell wall and membrane, and the inability to divide. To confirm that these morphological cues corresponded with cell death, we probed cell viability on a subset of our strains after osmotic shock through staining with propidium iodide, a DNA intercalating dye commonly used to identifying dead cells (LIVE/DEAD BacLight Bacterial Cell Viability Assay, Thermo Fisher). We found that our classification based on morphology agreed with that based off of staining within 1%. More information regarding these experiments can be found in the Supplement C. Cells which detached from the surface during the post-shock growth phase or those which became indistinguishable from other cells due to clustering were not counted as survival or death and were removed from the dataset completely. A region of the cell was manually marked with 1.0 (survival) or 0.0 (death) by clicking on the image. The `xy` coordinates of the click as well as the assigned value were saved as an `.xml` file for that position.

The connection between the segmented cells and their corresponding

manual markers was automated. As the manual markings were made on the first phase contrast image after the osmotic shock, small shifts in the positions of the cell made one-to-one mapping with the segmentation mask non-trivial. The linkages between segmented cell and manual marker were made by computing all pairwise distances between the manual marker and the segmented cell centroid, taking the shortest distance as the true pairing. The linkages were then inspected manually and incorrect mappings were corrected as necessary.

All relevant statistics about the segmented objects as well as the sample identity, date of acquisition, osmotic shock rate, and camera exposure time were saved as `.csv` files for each individual experiment. A more in-depth description of the segmentation procedure as well as the relevant code can be accessed as a Jupyter Notebook at (`http://rpgroup.caltech.edu/mscl_survival`).

Calculation of effective channel copy number

To compute the MscL channel copy number, we relied on measuring the fluorescence level of a bacterial strain in which the mean MscL channel copy number was known via fluorescence microscopy (**???**). *E. coli* strain MLG910, which expresses the MscL-sfGFP fusion protein from the wild-type SD sequence, was grown under identical conditions to those described in Bialecka-Fornal et al. 2015 in LB Miller medium (BD Medical Sciences) to an OD_{600nm} of ~0.3. The cells were then diluted ten fold and immobilized on a rigid 2% agarose substrate and placed onto a glass bottom petri dish and imaged in the same conditions as described previously.

Images were taken of six biological replicates of MLG910 and were processed identically to those in the osmotic shock experiments. A calibration factor between the average cell fluorescence level and mean MscL copy number was then computed. We assumed that all measured fluorescence (after filtering and background subtraction) was derived from the MscL-sfGFP fusion,

$$\langle I_{\text{tot}} \rangle = \alpha \langle N \rangle, \tag{2.3}$$

in which α is the calibration factor and $\langle N \rangle$ is the mean cellular MscL-sfGFP copy number as reported in Bialecka-Fornal et al. 2012 (**???**). To correct for errors in segmentation, the intensity was computed as an areal density $\langle I_A \rangle$ and was multiplied by the average cell area $\langle A \rangle$ of the population. The

calibration factor was therefore computed as

$$\alpha = \frac{\langle I_A \rangle \langle A \rangle}{\langle N \rangle}. \tag{2.4}$$

We used Bayesian inferential methods to compute this calibration factor taking measurement error and replicate-to-replicate variation into account. The resulting average cell area and calibration factor was used to convert the measured cell intensities from the osmotic shock experiments to cell copy number. The details of this inference are described in depth in the supplemental information (*Standard Candle Calibration*).

Logistic regression

We used Bayesian inferential methods to find the most probable values of the coefficients β_0 and β_1 and the appropriate credible regions and is described in detail in the supplemental information (*Logistic Regression*). Briefly, we used Markov chain Monte Carlo (MCMC) to sample from the log posterior distribution and took the most probable value as the mean of the samples for each parameter. The MCMC was performed using the Stan probabilistic programming language (**???**) and all models can be found on the GitHub repository (`http://github.com/rpgroup-pboc/mscl_survival`).

Calculation of survival probability error

The vertical error bars for the points shown in (**???**) represent our uncertainty in the survival probability given our measurement of n survivors out of a total N single-cell measurements. The probability distribution of the survival probability p_s given these measurements can be written using Bayes' theorem as

$$g(p_s \mid n, N) = \frac{f(n \mid p_s, N) g(p_s)}{f(n \mid N)}, \tag{2.5}$$

where g and f represent probability density functions over parameters and data, respectively. The likelihood $f(n \mid p_s, N)$ represents the probability of measuring n survival events, given a total of N measurements each with a probability of survival p_s. This matches the story for the Binomial distribu-

tion and can be written as

$$f(n \mid p_s, N) = \frac{N!}{n!(N-n)!} p_s^n (1-p_s)^{N-n}. \tag{2.6}$$

To maintain maximal ignorance we can assume that any value for p_s is valid, such that is in the range [0, 1]. This prior knowledge, represented by $g(p_s)$, can be written as

$$g(p_s) = \begin{cases} 1 & 0 \le p_s \le 1 \\ 0 & \text{otherwise} \end{cases}. \tag{2.7}$$

We can also assume maximal ignorance for the total number of survival events we could measure given N observations, $f(n \mid N)$. Assuming all observations are equally likely, this can be written as

$$f(n \mid N) = \frac{1}{N+1} \tag{2.8}$$

where the addition of one comes from the possibility of observing zero survival events. Combining (**???**; **???**; **???**), the posterior distribution $g(p_s \mid n, N)$ is

$$g(p_s \mid n, N) = \frac{(N+1)!}{n!(N-n)!} p_s^n (1-p_s)^{N-n}. \tag{2.9}$$

The most probable value of p_s, where the posterior probability distribution given by (**???**) is maximized, can be found by computing the point at which derivative of the log posterior with respect to p_s goes to zero,

$$\frac{d \log g(p_s \mid n, N)}{d p_s} = \frac{n}{p_s} - \frac{N-n}{1-p_s} = 0. \tag{2.10}$$

Solving (**???**) for p_s gives the most likely value for the probability,

$$p_s^* = \frac{n}{N}. \tag{2.11}$$

So long as $N \gg n p_s^*$, (**???**) can be approximated as a Gaussian distribution with a mean p_s^* and a variance $\sigma_{p_s}^2$. By definition, the variance of a Gaussian

distribution is computed as the negative reciprocal of the second derivative of the log posterior evaluated at $p_s = p_s^*$,

$$\sigma_{p_s}^2 = -\left(\left.\frac{d^2 \log g(p_s \mid n, N)}{dp_s^2}\right|_{p_s = p_s^*}\right)^{-1}. \tag{2.12}$$

Evaluating (**???**) yields

$$\sigma_{p_s}^2 = \frac{n(N-n)}{N^3}. \tag{2.13}$$

Given (**???**) and (**???**), the most-likely survival probability and estimate of the uncertainty can be expressed as

$$p_s = p_s^* \pm \sigma_{p_s}. \tag{2.14}$$

Data and software availability

All raw image data is freely available and is stored on the Caltech-DATA Research Data Repository (**???**). The raw Markov chain Monte Carlo samples are stored as `.csv` files on CaltechDATA (**???**). All processed experimental data, Python, and Stan code used in this work are freely available through our GitHub repository (`http://github.com/rpgroup-pboc/mscl_survival`)(**???**) accessible through DOI: 10.5281/zenodo.1252524. The scientific community is invited to fork our repository and open constructive issues.

2.5 Acknowledgements

Boedicker, J.Q., Garcia, H.G., and Phillips, R. (2013). Theoretical and Experimental Dissection of DNA Loop-Mediated Repression. Physical Review Letters *110*, 018101.

Brewster, R.C., Weinert, F.M., Garcia, H.G., Song, D., Rydenfelt, M., and Phillips, R. (2014). The Transcription Factor Titration Effect Dictates Level of Gene Expression. Cell *156*, 1312–1323.

Garcia, H.G., and Phillips, R. (2011). Quantitative dissection of the simple repression input-output function. Proceedings of the National Academy of Sciences *108*, 12173–12178.

Harman, J.G. (2001). Allosteric regulation of the cAMP receptor protein. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology *1547*, 1–17.

Lanfranco, M.F., Gárate, F., Engdahl, A.J., and Maillard, R.A. (2017). Asymmetric configurations in a reengineered homodimer reveal multiple subunit communication pathways in protein allostery. Journal of Biological Chemistry *292*, 6086–6093.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. Cell *157*, 624–635.

Lindsley, J.E., and Rutter, J. (2006). Whence cometh the allosterome? Proceedings of the National Academy of Sciences *103*, 10533–10535.

Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: A plausible model. Journal of Molecular Biology *12*, 88–118.

Poelwijk, F.J., de Vos, M.G.J., and Tans, S.J. (2011). Tradeoffs and Optimality in the Evolution of Gene Regulation. Cell *146*, 462–470.

Rogers, J.K., Guzman, C.D., Taylor, N.D., Raman, S., Anderson, K., and Church, G.M. (2015). Synthetic biosensors for precise gene control and real-time monitoring of metabolites. Nucleic Acids Research *43*, 7648–7660.

Rohlhill, J., Sandoval, N.R., and Papoutsakis, E.T. (2017). Sort-Seq Approach to Engineering a Formaldehyde-Inducible Promoter for Dynamically Regulated Escherichia coli Growth on Methanol. ACS Synthetic Biology *6*, 1584–1595.

Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. Nature Biotechnology *27*, 946–950.

Setty, Y., Mayo, A.E., Surette, M.G., and Alon, U. (2003). Detailed map of a cis-regulatory input function. Proceedings of the National Academy of Sciences *100*, 7702–7707.

Chapter 3

QUESTIONNAIRE

Chapter 4

CONSENT FORM